

PART 1 - Introduction to Online Gene Study

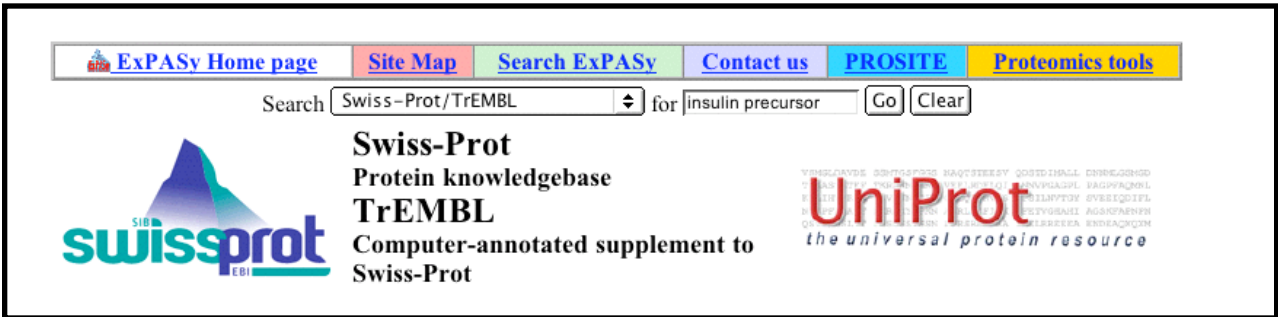
Insulin

A. Learn about Insulin

1. Insulin History
<http://www.schoolscience.co.uk/content/4/biology/abpi/history/history12.html>
2. Insulin Structure
http://c4.cabrillo.cc.ca.us/projects/insulin_tutorial/tutorial/
3. Protein Spotlight
http://www.expasy.org/spotlight/back_issues/2001/04/insulin_protein.shtml

B. Investigate the Protein Sequence for Insulin

1. Visit the Swiss-Prot Protein Knowledgebase <http://us.expasy.org/sprot/> This site is an annotated protein sequence database that includes descriptions of the functions of proteins and more.
2. Enter 'insulin precursor' into the Search box at the top of the page.



ExpASY Home page Site Map Search ExpASY Contact us PROSITE Proteomics tools

Search Swiss-Prot/TrEMBL for insulin precursor Go Clear

Swiss-Prot
Protein knowledgebase
TrEMBL
Computer-annotated supplement to
Swiss-Prot

UniProt
the universal protein resource

3. On the results page, scroll down and click on INS HUMAN.

[INS_GORGO \(Q0YK55\)](#)
Insulin precursor. {GENE: Name=INS} - Gorilla gorilla gorilla (Lowland gorilla)

[INS_HORSE \(P01310\)](#)
Insulin precursor. {GENE: Name=INS} - Equus caballus (Horse)

[INS_HUMAN \(P01308\)](#)
Insulin precursor. {GENE: Name=INS} - Homo sapiens (Human)

[INS_LOPPI \(P01341\)](#)
Insulin precursor. {GENE: Name=INS} - Lophius piscatorius (Allmouth goosefish) (Anglerfish), Lophius americanus (American goosefish) (Anglerfish)

4. On this page, look at the **Comments** section. What is the purpose of this protein? Defects in Insulin cause *familial hyperproinsulinemia*. What does this mean?

Comments

- **FUNCTION:** Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.
- **SUBUNIT:** Heterodimer of a B chain and an A chain linked by two disulfide bonds.
- **SUBCELLULAR LOCATION:** Secreted.
- **DISEASE:** Defects in INS are the cause of familial hyperproinsulinemia [MIM:176730].
- **PHARMACEUTICAL:** Available under the names Humulin or Humalog (Eli Lilly) and Novolin (Novo Nordisk). Used in the treatment of diabetes. Humalog is an insulin analog with 52-Lys-Pro-53 instead of 52-Pro-Lys-53.
- **SIMILARITY:** Belongs to the insulin family.
- **DATABASE:** NAME=Insulin at Eli Lilly; NOTE=Clinical information on Eli Lilly insulin products; WWW="<http://www.lillyDiabetes.com/Products/PatientInfo.cfm>".
- **DATABASE:** NAME=Protein Spotlight; NOTE=Issue 9 of April 2001; WWW="<http://www.expasy.org/spotlight/articles/sptlt009.html>".

5. Scroll down to the **Features** section. Locate the sections that contain variants. Some of these cause diseases such as diabetes. Click on one of the variants. What changes in the protein result from this mutation?

Features						
	Feature table viewer				Feature aligner	
Key	From	To	Length	Description	FTId	
SIGNAL	1	24	24			
CHAIN	25	54	30	Insulin B chain.		
PROPEP	57	87	31	C peptide.		
CHAIN	90	110	21	Insulin A chain.		
DISULFID	31	96		Interchain.		
DISULFID	43	109		Interchain.		
DISULFID	95	100				
VARIANT	34	34	*	H -> D (in familial hyperproinsulinemia; Providence).	VAR_003971	
VARIANT	48	48	*	F -> S (associated with diabetes mellitus type-II; Los-Angeles).	VAR_003972	
VARIANT	49	49	*	F -> L (in Chicago).	VAR_003973	
VARIANT	89	89	*	R -> H (in familial hyperproinsulinemia; impairs posttranslational cleavage).	VAR_003974	
VARIANT	89	89	*	R -> L (in familial hyperproinsulinemia; Kyoto).	VAR_003975	
VARIANT	92	92	*	V -> L (in Wakayama).	VAR_003976	
TURN	32	32	1			
HELIX	33	46	14			
STRAND	48	50	3			
HELIX	91	95	5			
TURN	96	97	2			
HELIX	102	108	7			
STRAND	109	109	1			

6. Click the back button to return to the Human Insulin page. Scroll all the way to the bottom and look at **Sequence Information**. How long is the protein molecule (how many amino acids)? Click on the FASTA format link to get the sequence in a format we can use.

Sequence information						
Length: 110 AA [This is the length of the unprocessed precursor]		Molecular weight: 11981 Da [This is the MW of the unprocessed precursor]		CRC64: C2C3B23B85E520E5 [This is a checksum on the sequence]		
10	20	30	40	50	60	
MALWMRLPL	LALLALWGPD	PAAAFVNQHL	CGSHLVEALY	LVCGERGFFY	TPKTRREAED	
70	80	90	100	110		
LQVGQVELGG	GPGAGSLQPL	ALEGLSLQKRG	IVEQCCTSIC	SLYQLENYCN		
					P01308 in FASTA format	

7. Highlight and copy the amino acid sequence:

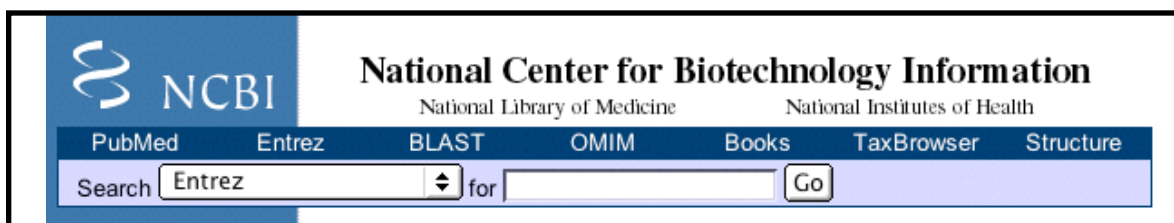
```
>sp|P01308|INS_HUMAN Insulin precursor - Homo sapiens (Human).
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGLSLQKRGIVEQCCTSICSLYQLENYCN
```

Visit this link to find out what all these letters stand for:

<http://www.bio.davidson.edu/courses/Molbio/aatable.html>

C. Finding the Insulin Gene

1. Go to the **National Center for Biotechnology Information** (NCBI) home page at <http://www.ncbi.nlm.nih.gov/>. This is the home of the national archive of gene sequences called GenBank. This database contains virtually all sequenced genes. We will compare the human insulin protein sequence to everything previously sequenced.



The BLAST search pages allow you to select from several different programs. Below is a table of these programs.

□

Program	Description
blastp	Compares an amino acid query sequence against a protein sequence database.
blastn	Compares a nucleotide query sequence against a nucleotide sequence database.
blastx	Compares a nucleotide query sequence translated in all reading frames against a protein sequence database. You could use this option to find potential translation products of an unknown nucleotide sequence.
tblastn	Compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.
tblastx	Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. Please note that the tblastx program cannot be used with the nr database on the BLAST Web page because it is computationally intensive.

- At the top of the page, click on “BLAST” and then on the next page “Protein-protein (blastp)”. If we were using the DNA (nucleotide) sequence, we could use the blastn.

Protein

- [Protein-protein BLAST \(blastp\)](#)
- [PHI- and PSI-BLAST](#)
- [Search for short, nearly exact matches](#)
- [Search the conserved domain database \(rpsblast\)](#)
- [Search by domain architecture \(cdart\)](#)

- Paste in the protein sequence you copied in step 7 of part B. Click **BLAST**. The algorithm is now comparing your query sequence to all others in the database to find the most likely matches. This may take a few minutes.

[Search](#)

```
>sp|P01308|INS_HUMAN Insulin precursor - Homo sapiens (Human).
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

[Set subsequence](#) From: To:

[Choose database](#)

[Do CD-Search](#)


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

- Click **Format** to see if the results are ready. After clicking once, the results page will refresh itself until the search is complete – this will also take a few minutes.

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

Query = sp|P01308|INS_HUMAN Insulin precursor - Homo sapiens (Human). (110 letters)

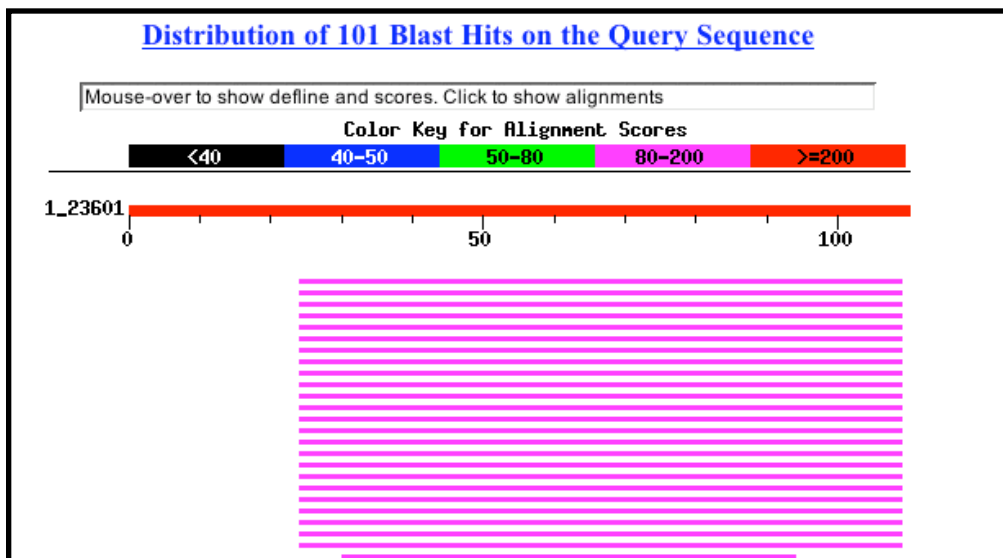
Putative conserved domains have been detected, click on the image below for detailed results.



The request ID is

Format! or **Reset all**

5. The **Color Key for Alignment Scores** shows how well the sequences found by BLAST match the sequence you entered.



You can scroll down to see a list of the proteins with their **E-values**. E-Values are a mathematical representation of how related other proteins are to your search sequence. The smaller the number, the closer the match.

Sequences producing significant alignments:	Score (bits)	E Value
gi 30584395 gb AAP36446.1 Homo sapiens insulin [synthetic ...	169	1e-41
gi 23986711 gb AAN39451.1 insulin [Homo sapiens] >gi 22901...	169	1e-41
gi 22901134 gb AAN06933.1 insulin precursor [Pan troglodyt...	169	1e-41
gi 208668 gb AAA72172.1 synthetic preproinsulin	169	2e-41
gi 22901146 gb AAN06937.1 insulin precursor [Pongo pygmaeu...	169	2e-41
gi 22809 emb CAA43405.1 Preproinsulin [Cercopithecus aethi...	167	4e-41
gi 266373 sp P30406 INS_MACFA Insulin precursor >gi 86627 p...	167	5e-41
gi 58103 emb CAA23424.1 unnamed protein product [synthetic...	162	2e-39
gi 223965 prf 1006230A insulin,pro-	160	7e-39

6. Below this list are the actual **sequences**. The top row is the insulin sequence you searched for. The bottom row contains the matches found in the database. The middle row shows the parts that match. The percent that match is also given. Compare the sequences at the top of the list to the ones at the bottom. What other proteins are similar to insulin? What animals have insulin proteins similar to human insulin?

```

>gi|30584395|gb|AAP36446.1| Homo sapiens insulin [synthetic construct]
Length = 111

Score = 169 bits (429), Expect = 1e-41
Identities = 86/86 (100%), Positives = 86/86 (100%)

Query: 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGVELGSSQPEALLEG 84
          FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGVELGSSQPEALLEG 84
Sbjct: 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGVELGSSQPEALLEG 84

Query: 85 SLQKRGIVEQCCTSICSLYQLENYCN 110
          SLQKRGIVEQCCTSICSLYQLENYCN
Sbjct: 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

```

```

>gi169306|pir||IPPG   insulin precursor - pig
      Length = 84

Score = 115 bits (289), Expect = 2e-25
Identities = 72/86 (83%), Positives = 73/86 (84%), Gaps = 2/86 (2%)

Query: 25  FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84
          FVNQHLCGSHLVEALYLVCGERGFFYTPK RREAE+ Q G VE   G G G LQ LALEG
Sbjct: 1  FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVE--LGGGLGGLQALALEG 58

Query: 85  SLQKRGIVEQCCTSICSLYQLENYCN 110
          QKRGIVEQCCTSICSLYQLENYCN
Sbjct: 59  PPQKRGIVEQCCTSICSLYQLENYCN 84

```

Visit this site for a comparison of human insulin precursor to three other species <http://www.uni-mainz.de/~cfrosch/bc4s/example.html> (scroll down to the insulin discussion).

7. You can also go back to the main BLAST page (<http://www.ncbi.nlm.nih.gov/BLAST/>) and choose an animal under the Genomes section to search for your insulin protein sequence.

Genomes

- [Chicken, cow, pig, dog, sheep, cat](#) **NEW**
- [Environmental samples](#)
- [Human, mouse, rat](#)
- [Fugu rubripes, zebrafish](#)
- [Insects, nematodes, plants, fungi, malaria](#)
- [Microbial genomes, other eukaryotic genomes](#)

Or visit <http://www.ensembl.org/> to compare other diseases, proteins or sequences between species. This site will also show the base pair sequence and chromosome location

Species - Ensembl v24		
Human	pre!	NCBI 34 Jul 04
Mouse		NCBI m33 Jul 04
Zebrafish		WTSI Zv4 Sep 04
Rat		RGSC 3.1 Jul 04
Chicken		WASHUC1 Jul 04
Mosquito		MOZ 2 Apr 04
Fugu		Fugu v2.0 May 04
Fruitfly		BGDP 3.1 Jul 03
Chimp		CHIMP1 May 04
Honeybee		Amel1.1 Sep 04
Tetradon		TETRAODON7 Sep 04

For more activities download *Bioinformatics and the Human Genome Project*

http://www.bsos.org/page.asp?id=Curriculum_DevelopmentHigh_School_9-12Bioinformatics