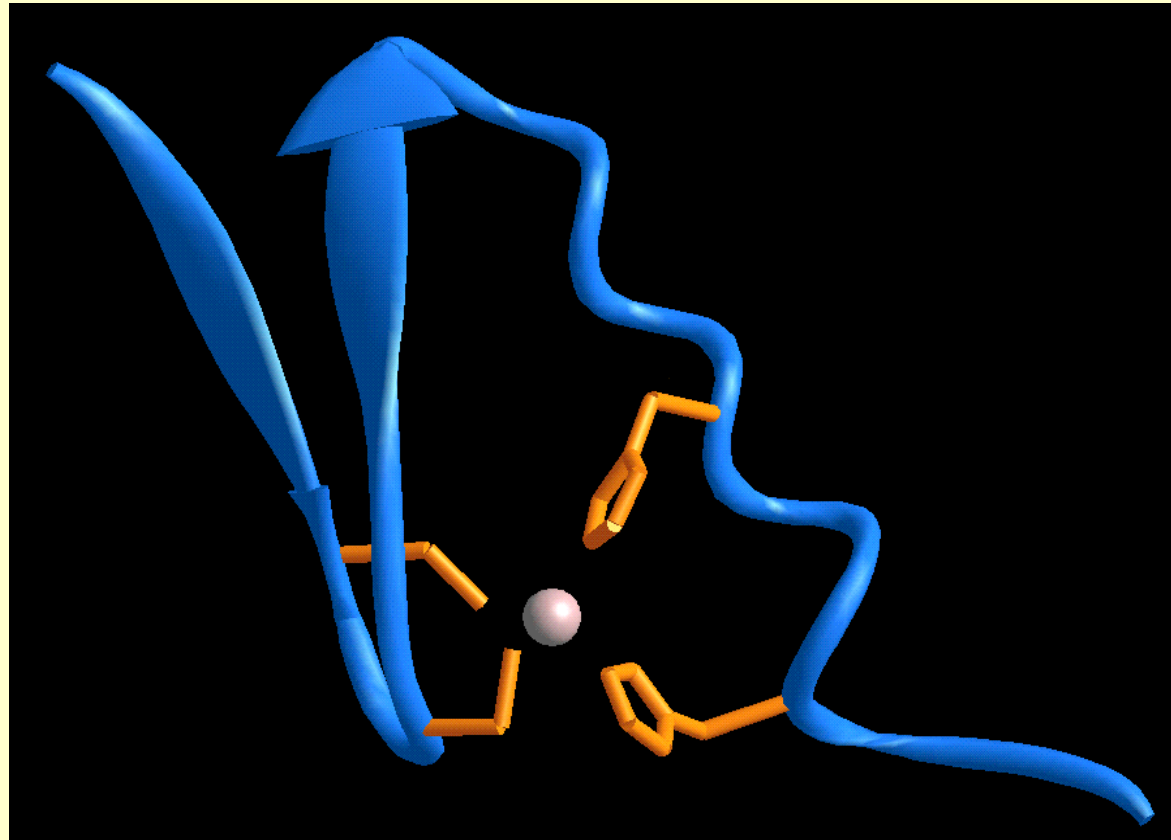


# Bioinformatics II

<http://biochem158.stanford.edu/bioinformatics.html>

Genomics, Bioinformatics & Medicine

<http://biochem158.stanford.edu/>



Doug Brutlag

Professor Emeritus of Biochemistry & Medicine  
Stanford University School of Medicine

# Discovering Function from Protein Sequence

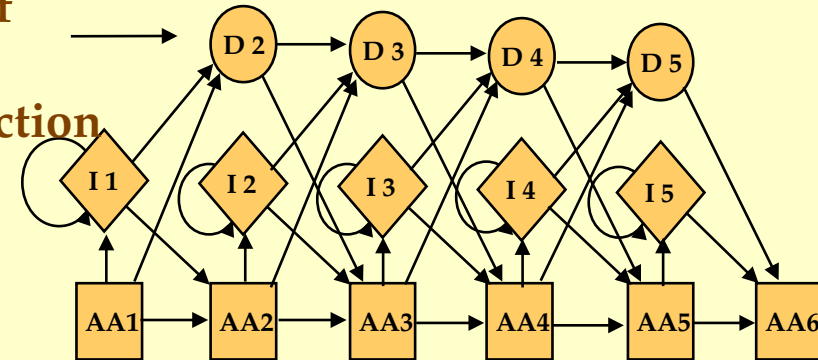
BLOCKS, PRINTs, PSSMS or Weight Matrices

	Position											
	1	2	3	4	5	6	7	8	9	10	11	12
A	2	1	3	13	10	12	67	4	13	9	1	2
R	7	5	8	9	4	0	1	16	7	0	1	0
N	0	8	0	1	0	0	0	2	1	1	10	0
D	0	1	0	1	13	0	0	12	1	0	4	0
C	0	0	1	0	0	0	0	0	0	2	2	1
Q	1	1	21	8	10	0	0	7	6	0	0	2
E	2	0	0	9	21	0	0	15	7	3	3	0
G	9	7	1	4	0	0	8	0	0	0	46	0
H	4	3	1	1	2	0	0	2	2	0	5	0
I	10	0	11	1	2	10	0	4	9	3	0	16
L	16	1	17	0	1	31	0	3	11	24	0	14
K	3	4	5	10	11	1	1	13	10	0	5	2
M	7	1	1	0	0	0	0	0	5	7	1	8
F	4	0	3	0	0	4	0	0	0	10	0	0
P	0	6	0	1	0	0	0	0	0	0	0	0
S	1	17	0	8	3	1	3	0	2	2	2	0
T	5	22	3	11	1	5	0	2	2	2	0	5
W	2	0	0	0	0	0	0	0	0	1	0	1
Y	1	0	4	2	0	1	0	0	2	4	0	1
V	6	3	1	1	2	15	0	0	2	12	0	28

Consensus Sequences  
or Sequence Motifs  
Zinc Finger (C2H2 type)  
C X{2,4} C X{12} H X{3,5} H

Profiles, PSI-BLAST  
Hidden Markov Models

Sequences of  
Common  
Structure or Function



Sequence Similarity

	10	20	30	40	50
Query	HLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGN				
	:	: :	:	:	:
Database	VLSPADKTNVKAAWGKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-----DLSHGHS				
	10	20	30	40	50



# Swiss Institute of Bioinformatics

<http://www.isb-sib.ch/>



Swiss Institute of Bioinformatics

Fellowship ▾

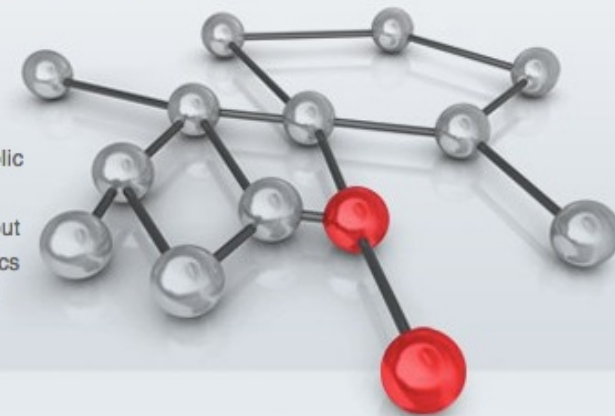
Home | Contact Us | Jobs | Finding People

[About SIB](#) [Groups](#) [Services](#) [Education](#) [Fellowship Program](#) [Research](#) [For the media](#)

[What is Bioinformatics](#)

# WELCOME

The SIB Swiss Institute of Bioinformatics is an academic, non-profit foundation recognised of public utility and established in 1998. SIB coordinates research and education in bioinformatics throughout Switzerland and provides high quality bioinformatics services to the national and international research community.



## Latest News

- 09 Oct 2012  
**Protein Spotlight Update: nature's flaws**  
Nothing is perfect. And nature is no exception....
- 11 Sep 2012  
**Protein Spotlight Update: life's boundaries**  
There is only one way of propagating the specie...
- 11 Sep 2012  
**SIB Fellowship program**  
Do you want to contribute to the challenges of ...

All

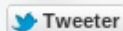
## Conferences & Events

- 13 Jan-31 Dec 2012-Lausanne, Switzerland  
**EPFL Life Science Seminars**
- 13 Jan-31 Dec 2012-Lausanne, Switzerland  
**The CIG Seminars & Workshops**
- 17-22 Sep 2012-Kyiv, Ukraine  
**Mol.Onco-2012 Symposium**
- 20-21 Sep 2012-Barcelona, Spain  
**Chromosomes, Stem Cells and Disease**

All

## Quick Links

- [About SIB](#)
- [Group Leaders](#)
- [ExPASy: SIB Bioinformatics Resource Portal](#)
- [Latest Jobs](#)
- [Publications](#)
- [Finding People](#)
- [List of courses](#)





# Expasy Bioinformatics Resource Portal

<http://expasy.org/>



ExpASY  
Bioinformatics Resource Portal

SIB Fellowship ▼

[Home](#) [About](#) [Contact](#)

Query all databases

search help

## Visual Guidance

### Categories

- proteomics
- genomics
- structural bioinformatics
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- drug design

## Resources A..Z

## Links/Documentation

ExpASY is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see **Categories** in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

### Featuring today

#### HCD/CID spectra merger

A tool to combine HCD and CID MS/MS spectra  
[\[details\]](#)



### Popular resources

- UniProtKB
- SWISS-MODEL
- STRING
- PROSITE

### Latest News



#### Updates to HAMAP and its website - 2012-10-12

The scope of HAMAP has been extended to classify and annotate also proteins from eukaryotic species. Additionally, our website has been updated and we provide now distinct pages for HAMAP family profiles on the HAMAP website. More information on: <http://hamap.expasy.org>.

#### UniProt Knowledgebase release 2012\_09 - 2012-10-03

UniProtKB/SwissProt Release of 03-Oct-2012 contains 538,010 sequence entries...[More](#).  
UniProtKB/TrEMBL Release of 03-Oct-2012 contains 26,079,526 sequence entries...[More](#)



### How to use this portal?

- Features and updates
- New to ExpASY
- Experienced ExpASY users: what is different





# Expasy Bioinformatics Resource Portal

<http://expasy.org/>



ExpASY

Bioinformatics Resource Portal

[Home](#) [About](#) [Contact](#)

query    [help](#)

## Visual Guidance

## Categories

### proteomics

protein sequences and identification  
mass spectrometry and 2-DE data  
protein characterisation and function  
families, patterns and profiles  
post-translational modification  
protein structure  
protein-protein interaction  
similarity search/alignment

### genomics

structural bioinformatics

systems biology

phylogeny/evolution

population genetics

transcriptomics

biophysics

imaging

IT infrastructure

drug design

## Resources A..Z

## Links/Documentation

## Databases

- UniProtKB • functional information on proteins • [\[more\]](#)
- UniProtKB/Swiss-Prot • protein sequence database • [\[more\]](#)
- SWISS-MODEL Repository • protein structure homology models • [\[more\]](#)
- STRING • protein-protein interactions • [\[more\]](#)
- PROSITE • protein domains and families • [\[more\]](#)
- ViralZone • portal to viral UniProtKB entries • [\[more\]](#)
- neXtProt • human proteins • [\[more\]](#)

- ENZYME • nomenclature of enzymes • [\[more\]](#)
- GlycoSuiteDB • glycan database • [\[more\]](#)
- GPSDB • gene and protein synonyms • [\[more\]](#)
- HAMAP • Microbial proteome annotation in UniProtKB • [\[more\]](#)
- MIAPEGelDB • MIAPE document edition • [\[more\]](#)
- MyHits • protein domains database and tools • [\[more\]](#)
- PaxDb • protein abundance database • [\[more\]](#)
- Prolune • Popular science articles (in French) • [\[more\]](#)
- Protein Model Portal • structural information for a protein • [\[more\]](#)
- Protein Spotlight • Informally written reviews on proteins • [\[more\]](#)
- SugarBind • pathogen sugar-binding • [\[more\]](#)
- SWISS-2DPAGE • proteins on 2-D and SDS PAGE maps • [\[more\]](#)
- SwissVar • variants in UniProtKB entries • [\[more\]](#)
- TCS • interaction specificity in two-component

## Tools

- SWISS-MODEL Workspace • structure homology-modeling • [\[more\]](#)
- SwissDock • protein ligand docking server • [\[more\]](#)
- AACompSim • amino acid composition comparison • [\[more\]](#)
- AllAll • protein sequences comparisons • [\[more\]](#)
- Biochemical Pathways • Biochemical Pathways • [\[more\]](#)
- BLAST • sequence similarity search • [\[more\]](#)
- boxshade • MSA pretty printer • [\[more\]](#)
- Compute pI/MW • theoretical pI and Mw computation • [\[more\]](#)
- Dotlet • sequence similarity plots • [\[more\]](#)
- EMBnet services • bioinformatics tools and databases • [\[more\]](#)
- FindMod • protein post-translational modifications • [\[more\]](#)
- FindPept • peptide identification from unspecific cleavage • [\[more\]](#)
- GlycanMass • oligosaccharide structure mass calculation • [\[more\]](#)
- GlycoMod • oligosaccharide structure prediction • [\[more\]](#)
- HAMAP • Microbial proteome annotation in UniProtKB • [\[more\]](#)
- HamapScan • scan sequences against HAMAP • [\[more\]](#)
- HCD/CID spectra merger • combine HCD and CID MS/MS spectra • [\[more\]](#)
- ImageMaster / Melanie • software for 2-D PAGE analysis • [\[more\]](#)
- IsotopIdent • theoretical isotopic distribution • [\[more\]](#)

Welcome to the new UniProt website! We hope you enjoy the new design. If you're not quite ready yet, you can still [go back to the old site](#).

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

## UniProtKB

Swiss-Prot  
(546,439)

Manually annotated and reviewed.

TrEMBL  
(83,955,074)

Automatically annotated and not reviewed.

## UniRef

Sequence clusters



## UniParc

Sequence archive



## Proteomes



## Supporting data

Literature citations



Taxonomy



Subcellular locations



Cross-ref. databases



Diseases

XXX

Keywords



## News



Small is beautiful (and useful) | Evidences in the UniProtKB flat file format  
[UniProt release 2014\\_09](#)

Ubiquitin caught at its own game | New human variant types available on the FTP site  
[UniProt release 2014\\_08](#)

Lark or owl? PER3 is the answer | Cross-

[News archive](#)



# UniProt Knowledge Base Advanced Search

<http://www.uniprot.org/>



UniProt

BLAST Align Upload

## Results

Filter by<sup>i</sup>

Reviewed

Searching in UniProtKB

AND

Organism [OS] human

Protein name [DE] opsin

All

25

Help Contact

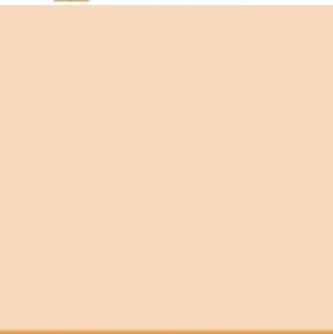
for UniProtKB

Basket

+

ⓧ

🔍





# UniProt Human Opsin Entries

<http://www.uniprot.org/>

UniProtKB  Advanced

BLAST Align Upload Lists Help Contact

## Results

### Filter by<sup>i</sup>

Reviewed (7)  
Swiss-Prot

Unreviewed (9)  
TrEMBL

Popular organisms  
Human (16)

### Search terms

Filter "opsin" as:

Filter "human" as:

### View by

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class
- Pathway

### UniRef

Your results in

Columns BLAST Align Download Add to basket

1 to 16 of 16 Show

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	P04001	OPSG_HUMAN		Medium-wave-sensitive opsin 1	OPN1MW, GCP OPN1MW2	Homo sapiens (Human)	364	
<input type="checkbox"/>	P04000	OPSR_HUMAN		Long-wave-sensitive opsin 1	OPN1LW, RCP	Homo sapiens (Human)	364	
<input type="checkbox"/>	P03999	OPSB_HUMAN		Short-wave-sensitive opsin 1	OPN1SW, BCP	Homo sapiens (Human)	348	
<input type="checkbox"/>	P08100	OPSD_HUMAN		Rhodopsin	RHO, OPN2	Homo sapiens (Human)	348	
<input type="checkbox"/>	Q9H1Y3	OPN3_HUMAN		Opsin-3	OPN3, ECPN	Homo sapiens (Human)	402	
<input type="checkbox"/>	Q9UHM6	OPN4_HUMAN		Melanopsin	OPN4, MOP	Homo sapiens (Human)	478	
<input type="checkbox"/>	Q6U736	OPN5_HUMAN		Opsin-5	OPN5, GPR136, PGR12, TMEM13	Homo sapiens (Human)	354	
<input type="checkbox"/>	Q0PJU0	Q0PJU0_HUMAN		Opsin 1 (Cone pigments), short-wave...	OPN1SW, hCG_41660	Homo sapiens (Human)	348	
<input type="checkbox"/>	B1B1F0	B1B1F0_HUMAN		Long-wave-sensitive opsin 1	OPN1LW, Z68193.1-002	Homo sapiens (Human)	164	
<input type="checkbox"/>	J3KPQ2	J3KPQ2_HUMAN		Opsin 5	OPN5, hCG_1642475	Homo sapiens (Human)	353	
<input type="checkbox"/>	Q6GMT1	Q6GMT1_HUMAN		Opsin 3	OPN3,	Homo sapiens	142	



# UniProt Human Opsin Entries Reviewed

<http://www.uniprot.org/>

UniProtKB

organism:human AND name:opsin AND reviewed:yes

Advanced



BLAST Align Upload Lists

Help Contact

Show help for UniProtKB

Basket

## Results

### Filter by<sup>i</sup>

Reviewed (7)   
 Swiss-Prot

Popular organisms   
 Human (7)

### Search terms

Filter "opsin" as:   
 protein name

Filter "human" as:   
 organism

### View by

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class
- Pathway

Columns BLAST Align Download Add to basket

1 to 7 of 7

Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
P04001	OPSG_HUMAN	Medium-wave-sensitive opsin 1	OPN1MW, GCP OPN1MW2	Homo sapiens (Human)	364
P04000	OPSR_HUMAN	Long-wave-sensitive opsin 1	OPN1LW, RCP	Homo sapiens (Human)	364
P03999	OPSB_HUMAN	Short-wave-sensitive opsin 1	OPN1SW, BCP	Homo sapiens (Human)	348
P08100	OPSD_HUMAN	Rhodopsin	RHO, OPN2	Homo sapiens (Human)	348
Q9H1Y3	OPN3_HUMAN	Opsin-3	OPN3, ECPN	Homo sapiens (Human)	402
Q9UHM6	OPN4_HUMAN	Melanopsin	OPN4, MOP	Homo sapiens (Human)	478
Q6U736	OPN5_HUMAN	Opsin-5	OPN5, GPR136, PGR12, TMEM13	Homo sapiens (Human)	354

1 to 7 of 7

Show 25



UniProt

BLAST Align Upload Lists Help Contact

## P04001 - OPSG\_HUMAN Basket

**Protein** | **Medium-wave-sensitive opsin 1**

**Gene** | **OPN1MW** [more](#)

**Organism** | *Homo sapiens (Human)*

**Status** | Reviewed - - Experimental evidence at protein level<sup>i</sup>

Display None

[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

[Comment \(0\)](#) [Feedback](#) [Help video](#)

- FUNCTION
- NAMES & TAXONOMY
- SUBCELL. LOCATION
- PATHOL./BIOTECH
- PTM / PROCESSING
- EXPRESSION
- INTERACTION
- STRUCTURE
- FAMILY & DOMAINS
- SEQUENCE
- CROSS-REFERENCES
- PUBLICATIONS
- ENTRY INFORMATION
- MISCELLANEOUS

### Function<sup>i</sup>

Visual pigments are the light-absorbing molecules that mediate vision. They consist of an apoprotein, opsin, covalently linked to cis-retinal.

**Absorption<sup>i</sup>**  
Abs(max)=530 nm

### GO - Molecular function<sup>i</sup>

▶ G-protein coupled receptor activity Source: UniProtKB-KW ▶ photoreceptor activity Source: ProtInc

### GO - Biological process<sup>i</sup>

▶ G-protein coupled receptor signaling pathway Source: ProtInc

▶ phototransduction, visible light Source: Reactome ▶ positive regulation of cytokinesis Source: UniProt

▶ protein-chromophore linkage Source: UniProtKB-KW ▶ retinoid metabolic process Source: Reactome

▶ visual perception Source: ProtInc

Complete GO annotation...

### Keywords - Molecular function<sup>i</sup>

G-protein coupled receptor, Photoreceptor protein, Receptor, Retinal protein, Transducer

### Keywords - Biological process<sup>i</sup>

Sensory transduction, Vision

### Keywords - Ligand<sup>i</sup>

Chromophore

[Top](#)






# UniProt Human Opsin PTM Processing


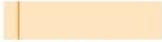



<http://www.uniprot.org/uniprot/P04001>

## PTM / Processing<sup>i</sup>

### Molecule processing

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Chain <sup>i</sup>	1 – 364	364	Medium-wave-sensitive opsin 1		PRO_0000197785	 Add  BLAST

### Amino acid modifications

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Glycosylation <sup>i</sup>	34 – 34	1	N-linked (GlcNAc...)  Curated			
Disulfide bond <sup>i</sup>	126 ↔ 203		 PROSITE-ProRule annotation			
Modified residue <sup>i</sup>	312 – 312	1	N6-(retinylidene)lysine			

### Post-translational modification<sup>i</sup>

Phosphorylated on some or all of the serine and threonine residues present in the C-terminal region.

### Keywords - PTM<sup>i</sup>

Disulfide bond, Glycoprotein, Phosphoprotein

### Proteomic databases

PaxDb <sup>i</sup>	P04001.
PRIDE <sup>i</sup>	P04001.

### PTM databases

PhosphoSite <sup>i</sup>	P04001.
--------------------------	---------





# Blast UniProt Human Opsin OPN1MW Entry

<http://www.uniprot.org/uniprot/P04001>



UniProtKB ▾

Advanced ▾



BLAST Align Upload Lists

Help Contact

## How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g.P00750 or A4\_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

[? Help](#)

[▶ Tutorials and Videos](#)

[↓ Downloads](#)

# BLAST

```
>sp|P04001|OPSG_HUMAN Medium-wave-sensitive opsin 1 OS=Homo sapiens GN=OPN1MW PE=1 SV=1
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNNSNSTRGPFEGPNYHIAPRWVYHLTSVWM
IFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASTISVFNQVYGYFV
LGHPMCYLEGTVSLCGITGLWSLAIISWERWMVCKPFGNVRFDKLAIVGIAFSWIWA
AVWTAPPIFGWSRYWPHGLKTSYCPDVFSGSSYPGVQSYMIVLMVTCCITPLSIIIVLCYL
QVWLAIKRAVAKQKQKESSTQKAEKEVTRMVVVMVLAFCFCWGPYAFFACFAAANPGYPFH
PLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDGSELSSASKTEVSSVSS
VSPA
```

Target database<sup>i</sup>

UniProtKB/Swiss-Prot ▾

E-Threshold<sup>i</sup>

10 ▾

Matrix<sup>i</sup>

Auto ▾

Filtering<sup>i</sup>

None ▾

Gapped<sup>i</sup>

yes ▾

Hits<sup>i</sup>

1000 ▾

Run Blast in a separate window.

[Run BLAST](#) [Clear](#)

# Blast UniProt Human OPN1MW Results

<http://www.uniprot.org/uniprot/P04001>

## Graphical overview

Color code for identity 0-100% =

Accession	Entry name	0Query hit364	0Match hit (sqrt scale)706	Name (Organism)
<input type="checkbox"/> Query				
201202022EFL3U01X8				
<input type="checkbox"/> P04001	OPSG_HUMAN			Medium-wave-sensitive opsin 1 (Homo sapiens)
<input type="checkbox"/> P04000	OPSR_HUMAN			Long-wave-sensitive opsin 1 (Homo sapiens)
<input type="checkbox"/> P34989	OPSL_CALJA			Opsin, longwave 563 nm (Callithrix jacchus)
<input type="checkbox"/> O35478	OPSG_SCICA			Medium-wave-sensitive opsin 1 (Sciurus carolinensis)
<input type="checkbox"/> O18913	OPSR_FELCA			Long-wave-sensitive opsin 1 (Felis catus)
<input type="checkbox"/> O18914	OPSR_CANFA			Long-wave-sensitive opsin 1 (Canis familiaris)
<input type="checkbox"/> Q9BG17	OPSR_BOVIN			Long-wave-sensitive opsin 1 (Bos taurus)
<input type="checkbox"/> Q95170	OPSR_CAPHI			Long-wave-sensitive opsin 1 (Capra hircus)
<input type="checkbox"/> Q9R024	OPSG_CAVPO			Medium-wave-sensitive opsin 1 (Cavia porcellus)
<input type="checkbox"/> O18910	OPSG_RABIT			Medium-wave-sensitive opsin 1 (Oryctolagus cuniculus)
<input type="checkbox"/> O35476	OPSG_RAT			Medium-wave-sensitive opsin 1 (Rattus norvegicus)
<input type="checkbox"/> O35599	OPSG_MOUSE			Medium-wave-sensitive opsin 1 (Mus musculus)
<input type="checkbox"/> P41592	OPSR_ANOCA			Red-sensitive opsin (Anolis carolinensis)
<input type="checkbox"/> P22329	OPSR_CHICK			Red-sensitive opsin (Gallus gallus)
<input type="checkbox"/> O12948	OPSR_XENLA			Red-sensitive opsin (Xenopus laevis)
<input type="checkbox"/> P87367	OPSR_ORYLA			Red-sensitive opsin (Oryzias latipes)
<input type="checkbox"/> P32313	OPSR_CARAU			Red-sensitive opsin (Carassius auratus)
<input type="checkbox"/> P35358	OPSG_GECGE			Green-sensitive opsin P521 (Gecko gecko)
<input type="checkbox"/> Q9W6A7	OPSR1_DANRE			Red-sensitive opsin-1 (Danio rerio)
<input type="checkbox"/> Q8AYN0	OPSR2_DANRE			Red-sensitive opsin-2 (Danio rerio)
<input type="checkbox"/> P22332	OPSR_ASTFA			Red-sensitive opsin (Astyanax fasciatus)
<input type="checkbox"/> P22331	OPSG2_ASTFA			Green-sensitive opsin-2 (Astyanax fasciatus)
<input type="checkbox"/> P22330	OPSG1_ASTFA			Green-sensitive opsin-1 (Astyanax fasciatus)
<input type="checkbox"/> O18912	OPSR_HORSE			Long-wave-sensitive opsin 1 (Equus caballus)
<input type="checkbox"/> O18911	OPSG_ODOVI			Medium-wave-sensitive opsin 1 (Odocoileus virginianus virginianus)
<input type="checkbox"/> P51476	OPSP_COLLI			Pinopsin (Columba livia)

## Detailed BLAST results [Customize](#)

- › Show hits with [3D data only](#).
- › Show hits from [complete proteomes only](#).

Alignments	Entry	Entry name	Status	Protein names <a href="#">...</a>	Organism	Length	Identity	Score	E-value	Gene names
------------	-------	------------	--------	-----------------------------------	----------	--------	----------	-------	---------	------------

# Entrez Gene search for Human Opsin

<http://www.ncbi.nlm.nih.gov/gene>

Gene

Gene

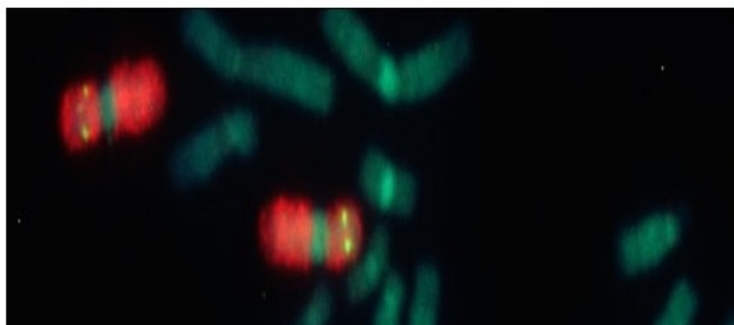
human opsin



Search

Advanced

Filters activated: Current only. [Clear all](#)



## Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to gene-specific phenotype-, and locus-specific resources worldwide.

### Using Gene

[Gene Quick Start](#)

[FAQ](#)

[Download/FTP](#)

[RefSeq Mailing List](#)

[Gene News](#)

[Factsheet](#)

### Gene Tools

[Submit GeneRIFs](#)

[Submit Correction](#)

[Statistics](#)

[BLAST](#)

[Genome Workbench](#)

[Splign](#)

### Other Resources

[HomoloGene](#)

[OMIM](#)

[RefSeq](#)

[RefSeqGene](#)

[UniGene](#)

[Protein Clusters](#)





# Entrez Gene search for Human Opsin

<http://www.ncbi.nlm.nih.gov/gene/?term=human+opsin>

Gene

Gene



human opsin

[Save search](#) [Advanced](#)

[Show additional filters](#)

**Display Settings:**  Tabular, 20 per page, Sorted by Relevance

**Send to:**

[Clear all](#)

**Results: 1 to 20 of 574**

<< First < Prev Page 1 of 29 Next > Last >>

**i** Filters activated: Current only. [Clear all](#) to show 580 items.

**Gene sources**

Genomic

**Categories**

Alternatively spliced

Annotated genes

Protein-coding

Pseudogene

**Sequence content**

CCDS

Ensembl

RefSeq

RefSeqGene

**Status**

[clear](#)

**Current only**

**Chromosome locations**

Select ...

[Clear all](#)

[Show additional filters](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">RHO</a> ID: 6010	rhodopsin [ (human)]	Chromosome 3, NC_000003.12 (129528639..129535344)	CSNBAD1, OPN2, RP4	180380
<input type="checkbox"/> <a href="#">OPN1LW</a> ID: 5956	opsin 1 (cone pigments), long-wave-sensitive [ (human)]	Chromosome X, NC_000023.11 (154144224..154159032)	hCG_41347, CBBM, CBP, COD5, RCP, ROP	300822
<input type="checkbox"/> <a href="#">OPN4</a> ID: 94233	opsin 4 [ (human)]	Chromosome 10, NC_000010.11 (86654557..86666460)	MOP	606665
<input type="checkbox"/> <a href="#">OPN1SW</a> ID: 611	opsin 1 (cone pigments), short-wave-sensitive [ (human)]	Chromosome 7, NC_000007.14 (128772489..128775790, complement)	BCP, BOP, CBT	613522
<input type="checkbox"/> <a href="#">OPN1MW</a> ID: 2652	opsin 1 (cone pigments), medium-wave-sensitive [ (human)]	Chromosome X, NC_000023.11 (154182596..154196861)	CBBM, CBD, COD5, GCP, GOP1, OPN1MW	300821

# Entrez Gene search for Human MW1 Opsin

<http://www.ncbi.nlm.nih.gov/gene/2652>

Display Settings:  Full Report

Send to:

## OPN1MW opsin 1 (cone pigments), medium-wave-sensitive [ *Homo sapiens* (human) ]

Gene ID: 2652, updated on 26-Dec-2014


### Summary

<b>Official Symbol</b>	OPN1MW provided by <a href="#">HGNC</a>
<b>Official Full Name</b>	opsin 1 (cone pigments), medium-wave-sensitive provided by <a href="#">HGNC</a>
<b>Primary source</b>	<a href="#">HGNC:HGNC:4206</a>
<b>See related</b>	<a href="#">HPRD:02365</a> ; <a href="#">MIM:300821</a>
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<a href="#">Homo sapiens</a>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
<b>Also known as</b>	CBD; GCP; GOP; CBBM; COD5; OPN1MW1
<b>Summary</b>	This gene encodes for a light absorbing visual pigment of the opsin gene family. The encoded protein is called green cone photopigment or medium-wavelength sensitive opsin. Opsins are G-protein coupled receptors with seven transmembrane domains, an N-terminal extracellular domain, and a C-terminal cytoplasmic domain. The long-wavelength opsin gene and multiple copies of the medium-wavelength opsin gene are tandemly arrayed on the X chromosome and frequent unequal recombination and gene conversion may occur between these sequences. X chromosomes may have fusions of the medium- and long-wavelength opsin genes or may have more than one copy of these genes. Defects in this gene are the cause of deutanopic colorblindness. [provided by RefSeq, Mar 2009]



# Entrez Gene search for Opsin MW1 FASTA Protein

<http://www.ncbi.nlm.nih.gov/protein/219521160?report=fasta>

 NCBI Resources  How To

Protein

Protein

[Limits](#) [Advanced](#)

[Display Settings:](#)  FASTA

[Send to:](#)

## Opsin 1 (cone pigments), medium-wave-sensitive [Homo sapiens]

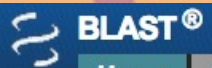
GenBank: AAI71755.1

[GenPept](#) [Graphics](#)

```
>gi|219520771|gb|AAI71755.1| Opsin 1 (cone pigments), medium-wave-sensitive [Homo sapiens]
```

```
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNSNSTRGPFEGPNYHIAPRWVYHLTSVWMIFVVIASVFT  
NGLVLAATMKFKLRHPLNWILVNLAVADLAETVIASTISVVNQVYGYFVLGHPCVLEGYTVSLCGITG  
LWSLAIISWERWLVVCKPFGNVRFDKLAIVGIAFSWIWAAVWTAPPIFGWSRYWPHGLKTSCGPDVFSG  
SSYPGVQSYMIVLMVTCCITPLSIIIVLCYLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMVLAFCFC  
WGPYAFFACFAAANPGYPFHPLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDGSELSSA  
SKTEVSSVSSVSPA
```





Basic Local Alignment Search Tool

- Home
- Recent Results
- Saved Strategies
- Help

My NCBI   
[\[Sign In\]](#) [\[Register\]](#)

► NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** DELTA-BLAST, a more sensitive protein-protein search

## BLAST Assembled Genomes

Find Genomic BLAST pages:

Enter organism name or id—completions will be suggested

- [Human](#)
- [Rabbit](#)
- [Zebrafish](#)
- [Mouse](#)
- [Chimp](#)
- [Clawed frog](#)
- [Rat](#)
- [Guinea pig](#)
- [Arabidopsis](#)
- [Cow](#)
- [Fruit fly](#)
- [Rice](#)
- [Pig](#)
- [Honey bee](#)
- [Yeast](#)
- [Dog](#)
- [Chicken](#)
- [Microbes](#)

## Basic BLAST

Choose a BLAST program to run.

### [nucleotide blast](#)

Search a **nucleotide** database using a **nucleotide** query  
*Algorithms: blastn, megablast, discontinuous megablast*

### [protein blast](#)

Search **protein** database using a **protein** query  
*Algorithms: blastp, psi-blast, phi-blast, delta-blast*

### [blastx](#)

Search **protein** database using a **translated nucleotide** query

### [tblastn](#)

Search **translated nucleotide** database using a **protein** query

### [tblastx](#)

Search **translated nucleotide** database using a **translated nucleotide** query

## Your Recent Results **New!**

 [All Recent results...](#)

## News

### [Find Genomic BLAST pages](#)

You can now find Genomic BLAST pages using the search box from the BLAST homepage.

Thu, 02 Oct 2014 11:00:00 EST

 [More BLAST news...](#)

## Tip of the Day

 [More tips...](#)



NCBI/ BLAST/ blastp suite **Standard Protein BLAST**

blastn blastp **blastx** tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

[Reset page](#) [Bookmark](#)

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

```
>sp|P04001|OPSG_HUMAN Medium-wave-sensitive opsin 1 OS=Homo sapiens  
GN=OPN1MW PE=1 SV=1  
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNNSNSTRGPFEGPNYHIAPRWVYHLTSVWM  
IFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASISVVNQVYGYFV  
LGHPMCVLEGYTVSLCGITGLWSLAIISWERWMVVCCKPFGNVRFDAKLAIVGIAFSWIWA
```

Query subrange [?](#)

From

To

Or, upload file

No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Database

[?](#)

Organism  
Optional

Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude  
Optional

Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query  
Optional

Enter an Entrez query to limit search [?](#)

**Program Selection**

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm [?](#)

**BLAST**

Search **database UniProtKB/Swiss-Prot(swissprot)** using **Blastp (protein-protein BLAST)**

Show results in a new window

[Algorithm parameters](#)

Note: Parameter values that differ from the default are highlighted in yellow and marked with **◆** sign

[General Parameters](#)

**BLAST**

Search **database UniProtKB/Swiss-Prot(swissprot)** using **Blastp (protein-protein BLAST)**

Show results in a new window

**Algorithm parameters**

**Note: Parameter values that differ from the default are highlighted**

**General Parameters**

**Max target sequences**    
Select the maximum number of aligned sequences to display

**Short queries**  Automatically adjust parameters for short input sequences

**Expect threshold**

**Word size**

**Max matches in a query range**

**Scoring Parameters**

**Matrix**

**Gap Costs** Existence: 11 Extension: 1

**Compositional adjustments**

**Filters and Masking**

**Filter**  Low complexity regions

**Mask**  Mask for lookup table only  
 Mask lower case letters

**BLAST**

Search **database UniProtKB/Swiss-Prot(swissprot)** using **Blastp (protein-protein BLAST)**

Show results in a new window



# NCBI BLAST Results

<http://blast.ncbi.nlm.nih.gov/>

## OPN1MW human opsin (364 letters)

**RID** [CE2V56ER015](#) (Expires on 01-28 10:52 am)

**Query ID** lcl|75188  
**Description** OPN1MW human opsin  
**Molecule type** amino acid  
**Query Length** 364

**Database Name** swissprot  
**Description** Non-redundant UniProtKB/SwissProt sequences  
**Program** BLASTP 2.2.30+ [▶ Citation](#)

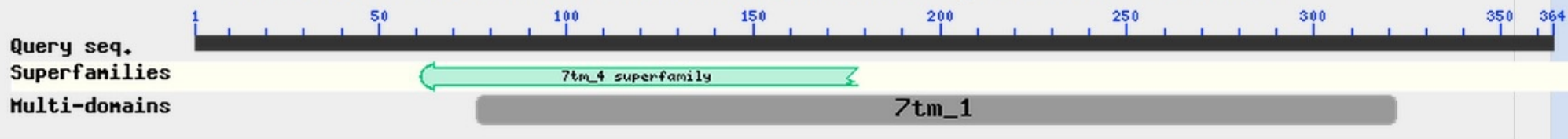
Other reports: [▶ Search Summary](#) [[Taxonomy reports](#)] [[Distance tree of results](#)] [[Multiple alignment](#)]

**New** DELTA-BLAST, a more sensitive protein-protein search

### Graphic Summary

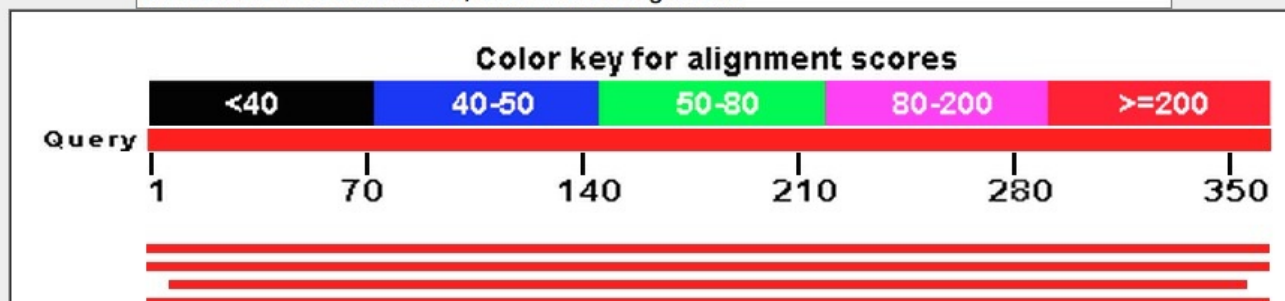
#### Show Conserved Domains

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



#### Distribution of 100 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments



# NCBI BLAST Results

<http://blast.ncbi.nlm.nih.gov/>

## Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

 [Alignments](#)
 [Download](#)
 [GenPept](#)
 [Graphics](#)
 [Distance tree of results](#)
 [Multiple alignment](#)


	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	750	750	100%	0.0	100%	<a href="#">P04001.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Long-wave-sensitive opsin 1; AltName: Full=Red cone photoreceptor pigment; AltName: Full=Red cone photoreceptor pigment</a>	712	712	100%	0.0	96%	<a href="#">P04000.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Opsin, longwave 563 nm [Callithrix jacchus]</a>	696	696	96%	0.0	94%	<a href="#">P34989.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	681	681	100%	0.0	90%	<a href="#">O35478.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Long-wave-sensitive opsin 1; AltName: Full=Red cone photoreceptor pigment; AltName: Full=Red cone photoreceptor pigment</a>	681	681	100%	0.0	90%	<a href="#">O18913.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Long-wave-sensitive opsin 1; AltName: Full=Red cone photoreceptor pigment; AltName: Full=Red cone photoreceptor pigment</a>	676	676	100%	0.0	90%	<a href="#">O18914.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Long-wave-sensitive opsin 1; AltName: Full=Red cone photoreceptor pigment; AltName: Full=Red cone photoreceptor pigment</a>	674	674	100%	0.0	89%	<a href="#">Q9BGI7.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Long-wave-sensitive opsin 1; AltName: Full=Red cone photoreceptor pigment; AltName: Full=Red cone photoreceptor pigment</a>	672	672	100%	0.0	89%	<a href="#">Q95170.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	670	670	100%	0.0	89%	<a href="#">Q9R024.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	662	662	100%	0.0	88%	<a href="#">O18910.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	660	660	98%	0.0	89%	<a href="#">O35476.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Medium-wave-sensitive opsin 1; AltName: Full=Green cone photoreceptor pigment</a>	656	656	98%	0.0	88%	<a href="#">O35599.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Red-sensitive opsin; AltName: Full=Red cone photoreceptor pigment [Anolis carolinensis]</a>	626	626	97%	0.0	81%	<a href="#">P41592.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Red-sensitive opsin; AltName: Full=Iodopsin; AltName: Full=Red cone photoreceptor pigment</a>	613	613	92%	0.0	86%	<a href="#">P22329.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Red-sensitive opsin; AltName: Full=KFH-R; AltName: Full=Red cone photoreceptor pigment</a>	599	599	98%	0.0	79%	<a href="#">P87367.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Red-sensitive opsin; AltName: Full=Red cone photoreceptor pigment [Xenopus laevis]</a>	599	599	97%	0.0	78%	<a href="#">O12948.1</a>



**Sequences producing significant alignments:**

Select: [All](#) [None](#) Selected:0

 [Alignments](#)
 [Download](#)
 [GenPept](#)
 [Graphics](#)
 [Distance tree of results](#)
 [Multiple alignment](#)


	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">RecName: Full=C-X-C chemokine receptor type 1; Short=CXC-R1; Short=CXCR-1; AltName: Full=</a>	32.0	32.0	30%	7.8	24%	<a href="#">P55920.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 2T8 [Homo sapiens]</a>	32.0	32.0	78%	7.8	19%	<a href="#">A6NH00.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=N-arachidonyl glycine receptor; Short=NAGly receptor; AltName: Full=G-protein co</a>	32.0	32.0	57%	7.9	21%	<a href="#">Q14330.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 6F1; AltName: Full=Olfactory receptor OR1-38 [Homo sapiens]</a>	32.0	32.0	35%	8.0	25%	<a href="#">Q8NGZ6.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Sphingosine 1-phosphate receptor 4; Short=S1P receptor 4; Short=S1P4; AltName</a>	32.0	32.0	53%	8.1	21%	<a href="#">O95977.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=C-X-C chemokine receptor type 3; Short=CXC-R3; Short=CXCR-3; AltName: Full=</a>	32.0	32.0	70%	8.2	23%	<a href="#">Q9JII9.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 2A7; AltName: Full=Olfactory receptor OR7-18 [Homo sapiens]</a>	31.6	31.6	31%	8.3	21%	<a href="#">Q96R45.3</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 1073; AltName: Full=Putative gustatory receptor PTE45 [Rattus</a>	31.6	31.6	27%	8.3	22%	<a href="#">P35898.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 2A4; AltName: Full=Olfactory receptor 2A10; AltName: Full=Olfa</a>	31.6	31.6	31%	8.5	21%	<a href="#">O95047.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 469; AltName: Full=Olfactory receptor 204-21 [Mus musculus]</a>	31.6	31.6	21%	8.5	19%	<a href="#">Q8VF66.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 4A47; AltName: Full=Olfactory receptor OR11-113 [Homo sapien]</a>	31.6	31.6	60%	8.5	19%	<a href="#">Q6IF82.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 8G2; AltName: Full=Olfactory receptor 8G4; AltName: Full=Olfa</a>	31.6	31.6	21%	8.8	21%	<a href="#">Q15614.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=C-C chemokine receptor type 7; Short=C-C CKR-7; Short=CC-CKR-7; Short=CCR</a>	32.0	32.0	32%	8.8	26%	<a href="#">Q5MD62.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Probable G-protein coupled receptor 34 [Mus musculus]</a>	31.6	31.6	93%	9.0	20%	<a href="#">Q9R1K6.1</a>
<input type="checkbox"/>	<a href="#">RecName: Full=C-X-C chemokine receptor type 1; Short=CXC-R1; Short=CXCR-1; AltName: Full=</a>	31.6	31.6	30%	9.0	24%	<a href="#">P25024.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Probable G-protein coupled receptor Mth-like 11; AltName: Full=Protein methusela</a>	32.0	32.0	24%	9.2	31%	<a href="#">P83118.2</a>
<input type="checkbox"/>	<a href="#">RecName: Full=Olfactory receptor 5H14 [Homo sapiens]</a>	31.6	31.6	30%	9.2	22%	<a href="#">A6NHG9.1</a>



# Discovering Function from Protein Sequence

BLOCKS, PRINTs, PSSMS or Weight Matrices

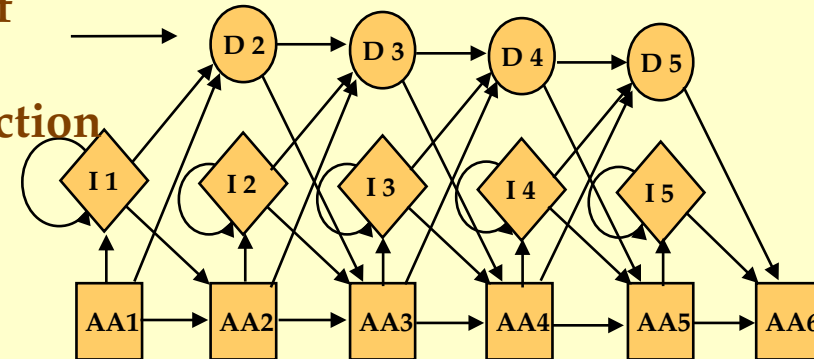
	Position											
	1	2	3	4	5	6	7	8	9	10	11	12
A	2	1	3	13	10	12	67	4	13	9	1	2
R	7	5	8	9	4	0	1	16	7	0	1	0
N	0	8	0	1	0	0	0	2	1	1	10	0
D	0	1	0	1	13	0	0	12	1	0	4	0
C	0	0	1	0	0	0	0	0	0	2	2	1
Q	1	1	21	8	10	0	0	7	6	0	0	2
E	2	0	0	9	21	0	0	15	7	3	3	0
G	9	7	1	4	0	0	8	0	0	0	46	0
H	4	3	1	1	2	0	0	2	2	0	5	0
I	10	0	11	1	2	10	0	4	9	3	0	16
L	16	1	17	0	1	31	0	3	11	24	0	14
K	3	4	5	10	11	1	1	13	10	0	5	2
M	7	1	1	0	0	0	0	0	5	7	1	8
F	4	0	3	0	0	4	0	0	0	10	0	0
P	0	6	0	1	0	0	0	0	0	0	0	0
S	1	17	0	8	3	1	3	0	2	2	2	0
T	5	22	3	11	1	5	0	2	2	2	0	5
W	2	0	0	0	0	0	0	0	0	1	0	1
Y	1	0	4	2	0	1	0	0	2	4	0	1
V	6	3	1	1	2	15	0	0	2	12	0	28

Consensus Sequences  
or Sequence Motifs

Zinc Finger (C2H2 type)  
C X{2,4} C X{12} H X{3,5} H

Profiles, PSI-BLAST  
Hidden Markov Models

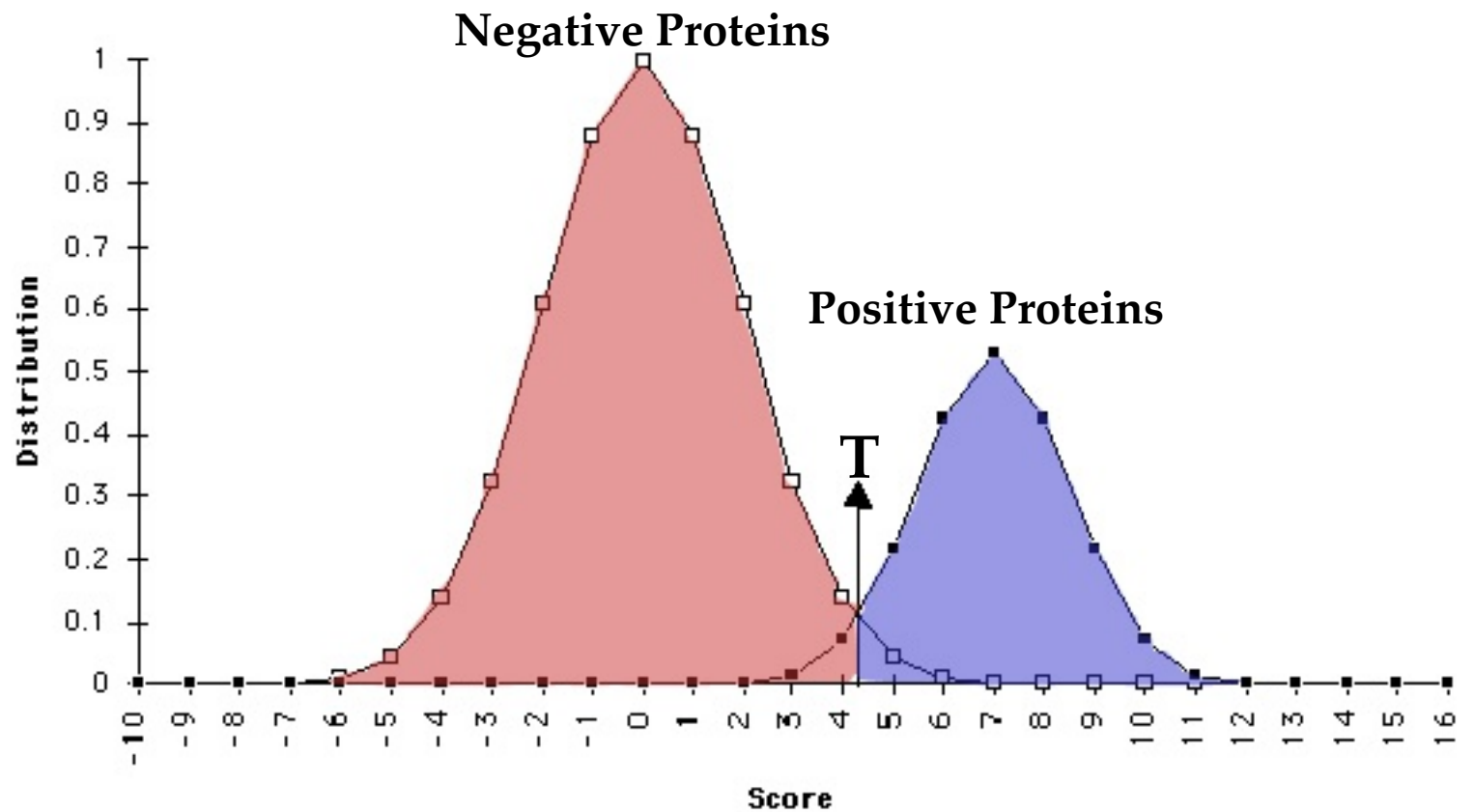
Sequences of  
Common  
Structure or Function



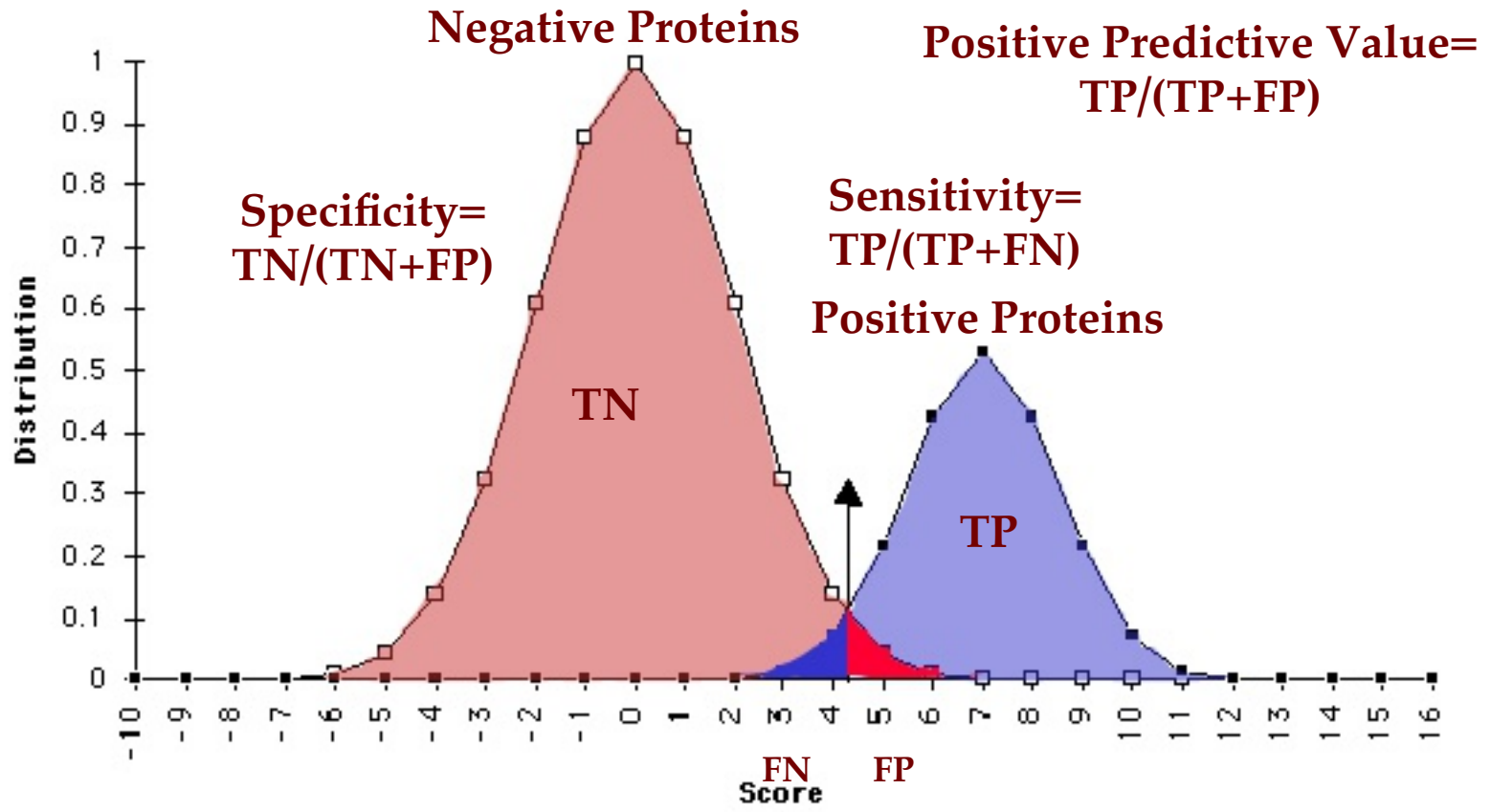
Sequence Similarity

	10	20	30	40	50
Query	HLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGN				
	:  : :    :              :        :				
Database	VLSPADKTNVKAAWGKVG AHAGEYGAELERMFLSFPTTKTYFPHF-----DLSHG S				
	10	20	30	40	50

# Evaluation of PSSMs, Profiles and HMMs



# Evaluation of Profiles





user: GUEST width: 600

### Tools

- Search ...
  - Pattern Search
  - BLASTP/PSI-BLAST
  - PFSEARCH (profile)
  - HMMER3 (profile-HMM)
- Motif Scan
- Query ...
  - by Protein
  - by Motif
- Align...
  - MAFFT
  - TCOFFEE
  - Profile Align
- Classify ...
  - JACOP
  - MkDom2
- Tools ...
  - Reformat MSA
  - Reformat SEQ
  - Dotlet

### Hub

### Results

### Misc

### Deprecated

**Hits** is a free database devoted to protein domains. It is also a collection of tools for the investigation of the relationships between protein sequences and motifs described on them. These motifs are defined by an heterogeneous collection of predictors, which currently includes regular expressions, generalized profiles and hidden Markov models.

### Last news

- Dec 2014** Update [mafft](#) to version 7.187
- Apr 2014** Update [BLASTP/PSI-BLAST](#) to BLAST+ version 2.2.28
- Sep 2013** Update [mafft](#) to version 7.058b
- Sep 2013** Update [hmmer3](#) to version 3.1b1
- Aug 2013** Link to new Toffee server
- May 2011** New Web interface
- Apr 2010** Add [hmmer3](#) service.
- Jun 2009** Add the CS-BLAST tool
- Oct 2008** Update of the Jalview applet to release 2.4
- Aug 2008** Add a text search to find sequence, motif or taxonomy matches.

[See all news](#)



# MyHits Local Motifs Query

<http://myhits.isb-sib.ch/>

## Motif Scan

search help

user: GUEST width: 600  
 log in settings

Motif scanning means finding all known motifs that occur in a sequence. This form lets you paste a protein sequence, select the collections of motifs to scan for, and launch the search. A [document](#) deals with the interpretation of the match scores. You should consult the home pages of [Prosite](#) on ExPASy, [Pfam](#) and [InterPro](#) for additional information.

If your proteins of interest are already in the sequence databases (see [list](#)), the [Query by Protein](#) form is much faster, and the [Protein Hub](#) provides a collection of tools that you might find useful.

- Tools**
- Search ...
  - Pattern Search
  - BLASTP/PSI-BLAST
  - PFSEARCH (profile)
  - HMMER3 (profile-HMM)
- Motif Scan
- Query ...
  - by Protein
  - by Motif
- Align...
  - MAFFT
  - TCOFFEE
  - Profile Align
- Classify ...
  - JACOP
  - MkDom2
- Tools ...
  - Reformat MSA
  - Reformat SEQ
  - Dotlet
- Hub**
- Results**
- Misc**
- Deprecated**

Protein Identifiers  
 or Protein Sequence

examples

clear input

```
>gi|219521160|gb|AAI71754.1| Opsin 1 (cone pigments),
medium-wave-sensitive [Homo sapiens]
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYNTSNSTRGPFEGPNYHIAPR
WVYHLTSVWMIFVVIASVFT
NGLVLAATMKFKLRHPLNWILVNLAVADLAETVIASISVNVQVYGYFVL
GHPMCYLEGYTVSLCGITG
LWSLAIISWERWLVVCKPFGNVRFDKLAIVGIAFSWIWAAVWTAPPIFGWS
RYWPHGLKTCGPDVFSG
```

- mot\_source
- perox - PeroxiBase profiles
  - hamap - HAMAP profiles
  - pat - PROSITE patterns
  - freq\_pat - PROSITE patterns (frequent match producers)
  - prf - PROSITE profiles
  - pre - More profiles
  - pfam\_fs - Pfam HMMs (local models)
  - pfam\_ls - Pfam HMMs (global models)

search

The scan might take a few minutes.  
 reset page



# MyHits Motif Scan Hits Summary

<http://myhits.isb-sib.ch/>

**Query Protein** temporarily stored [here](#).

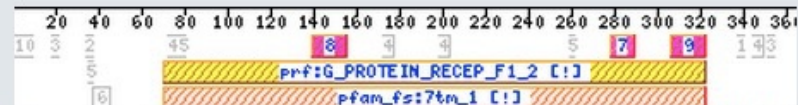
**Database of motifs** PROSITE patterns (frequent match producers) [freq\_pat], PROSITE patterns [pat], Pfam HMMs (local models) [pfam\_fs], More profiles [pre], PROSITE profiles [prf].

searching PROSITE patterns  
 searching PROSITE patterns (frequent match producers)  
 searching PROSITE profiles  
 searching More profiles  
 searching Pfam HMMs (local models)  
 postprocessing

## Summary

**Original output** [pat](#), [freq\\_pat](#), [prf](#), [pre](#), [pfam fs](#).

**Matches map**  
 (features from query are above the ruler, matches of the motif scan are below the ruler)



**Legends:** 1, freq\_pat:AMIDATION [?]; 2, freq\_pat:ASN\_GLYCOSYLATION [?]; 3, freq\_pat:CK2\_PHOSPHO\_SITE [?]; 4, freq\_pat:MYRISTYL [?]; 5, freq\_pat:PKC\_PHOSPHO\_SITE [?]; 6, freq\_pat:TYR\_PHOSPHO\_SITE [?]; 7, pat:EGF\_2 [!]; 8, pat:G\_PROTEIN\_RECEP\_F1\_1 [!]; 9, pat:OPSIN [!]; 10, prf:ALPHA\_BOX [?].

## List of matches

FT	MYHIT	337	340	freq_pat:AMIDATION [?]
FT	MYHIT	34	37	freq_pat:ASN_GLYCOSYLATION [?]
FT	MYHIT	18	21	freq_pat:CK2_PHOSPHO_SITE [?]
FT	MYHIT	351	354	freq_pat:CK2_PHOSPHO_SITE [?]
FT	MYHIT	72	77	freq_pat:MYRISTYL [?]
FT	MYHIT	172	177	freq_pat:MYRISTYL [?]
FT	MYHIT	198	203	freq_pat:MYRISTYL [?]
FT	MYHIT	344	349	freq_pat:MYRISTYL [?]
FT	MYHIT	35	37	freq_pat:PKC_PHOSPHO_SITE [?]
FT	MYHIT	78	80	freq_pat:PKC_PHOSPHO_SITE [?]
FT	MYHIT	259	261	freq_pat:PKC_PHOSPHO_SITE [?]
FT	MYHIT	37	45	freq_pat:TYR_PHOSPHO_SITE [?]
FT	MYHIT	278	289	pat:EGF_2 [!]
FT	MYHIT	139	155	pat:G_PROTEIN_RECEP_F1_1 [!]
FT	MYHIT	306	322	pat:OPSIN [!]
FT	MYHIT	1	9	prf:ALPHA_BOX [?]
FT	MYHIT	70	322	prf:G_PROTEIN_RECEP_F1_2 [!]
FT	MYHIT	70	322	pfam_fs:7tm_1 [!]



# MyHits PTM Motif Hits

<http://myhits.isb-sib.ch/>

## Detail of matches

### match detail

1  
KK  
XGRR  
I:::  
FGKK

### match score

Status: ?  
pos.: **337-340**

### motif information

freq\_pat:AMIDATION  
*Amidation site.*  
[ [entry](#) ]Legends: 1,  
amidation.

[^ image ^](#)

1  
Y Y  
W W  
V V  
T T  
S S  
R R  
R R  
R R  
N N  
N N  
M M  
M M  
L L  
L L  
K K  
K K  
I I  
H H  
G G  
G G  
F F  
F F  
E E  
E E  
D D  
D D  
C T C  
NASA  
I:::  
NSTR

Status: ?  
pos.: **34-37**

freq\_pat:ASN\_GLYCOSYLATION  
*N-glycosylation site.*  
[ [entry](#) ]Legends: 1,  
carbohydrate.

[^ image ^](#)

1  
T E  
SXXD  
:::  
SYED

Status: ?  
pos.: **18-21**

freq\_pat:CK2\_PHOSPHO\_SITE  
*Casein kinase II  
phosphorylation site.*

1  
T E  
SXXD  
:::  
SKTE

Status: ?  
pos.: **351-354**

[ [entry](#) ]Legends: 1,  
phosphorylation.

[^ image ^](#)

```

disulfide
disulfide
disulfide
  W
  PY
CXCXXGFXXXXC
| | :: |
CFCWGPYAFFAC

```

Status: **!**  
pos.: **278-289**

pat:EGF\_2  
*EGF-like domain  
signature 2.*  
[ [entry](#) ]

[^ image ^](#)

```

Y YY Y
W WW W
V VV V
T TT T
Y S SS S
W R QQ R
C V N PP Q
W T M NN C N
Y S L MM A M
FEQ K LL T L
MDN IAI I S K
VPM HGHH W I
ICL GGGTCY HH
LNI FNFFFNF SG
AAG EMEEMAMH CF M
TTF DVDDVTN WD V
SSC CICCISIE YC I
GGAXALALGLDRFAXL
::: :::::|:::
TGLWSLAIISWERWLVV

```

Status: **!**  
pos.: **139-155**

pat:G\_PROTEIN\_RECEP\_F1\_1  
*G-protein coupled receptors  
family 1 signature.*  
[ [entry](#) ]

[^ image ^](#)

```

retinal
Y
W
V
V
T
S
R
R
Q
P
N
M
L
K
I
H
F
E
D
C
A
T
S
T
E
P
Y
LPXAXSKSGSXXDAXXI
::: :|::: :
LPAFFAKSATIYNPVIY

```

Status: **!**  
pos.: **306-322**

pat:OPSIN  
*Visual pigments  
(opsins) retinal binding  
site.*  
[ [entry](#) ]

[^ image ^](#)

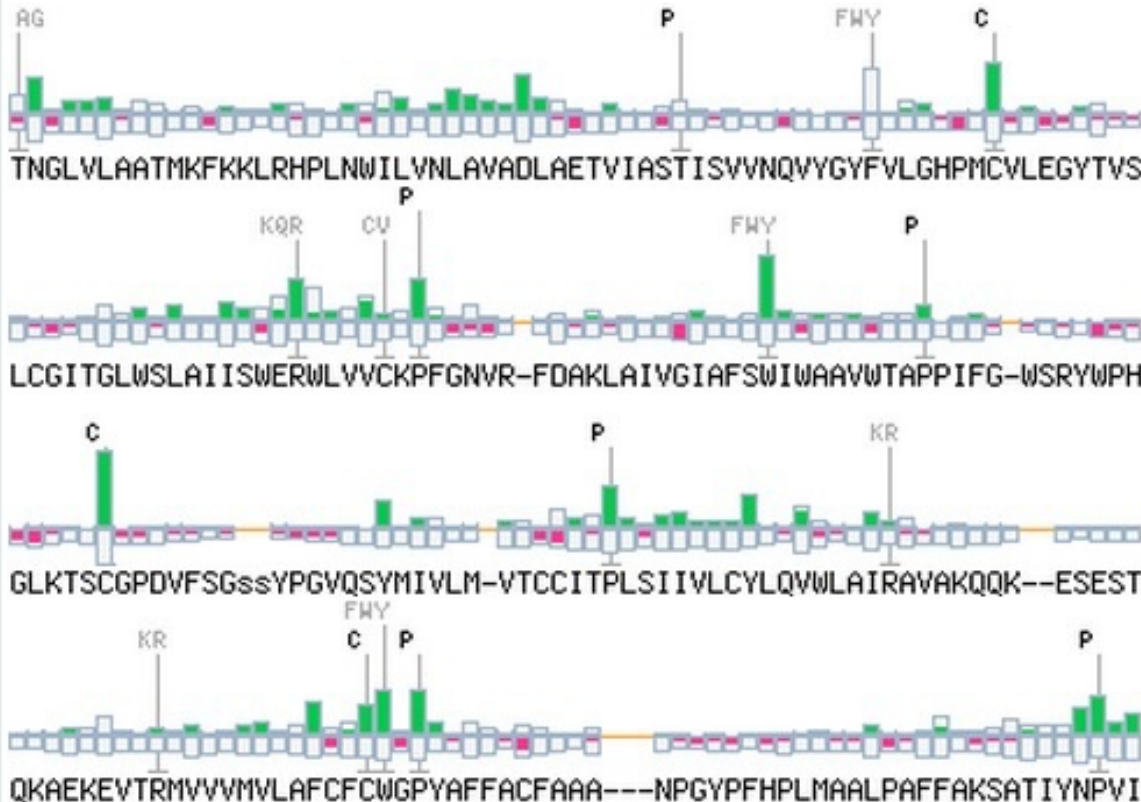
# MyHits Profile Hits (Cont.)



Status: **?**  
 pos.: **1-9**  
 raw-score = **191**  
 N-score = **7.485**  
 E-value = **0.69**

prf:ALPHA\_BOX  
*Alpha box DNA-binding domain profile.*  
[\[ entry \]](#)  
[\[ graphics \]](#)

[^ image ^](#)

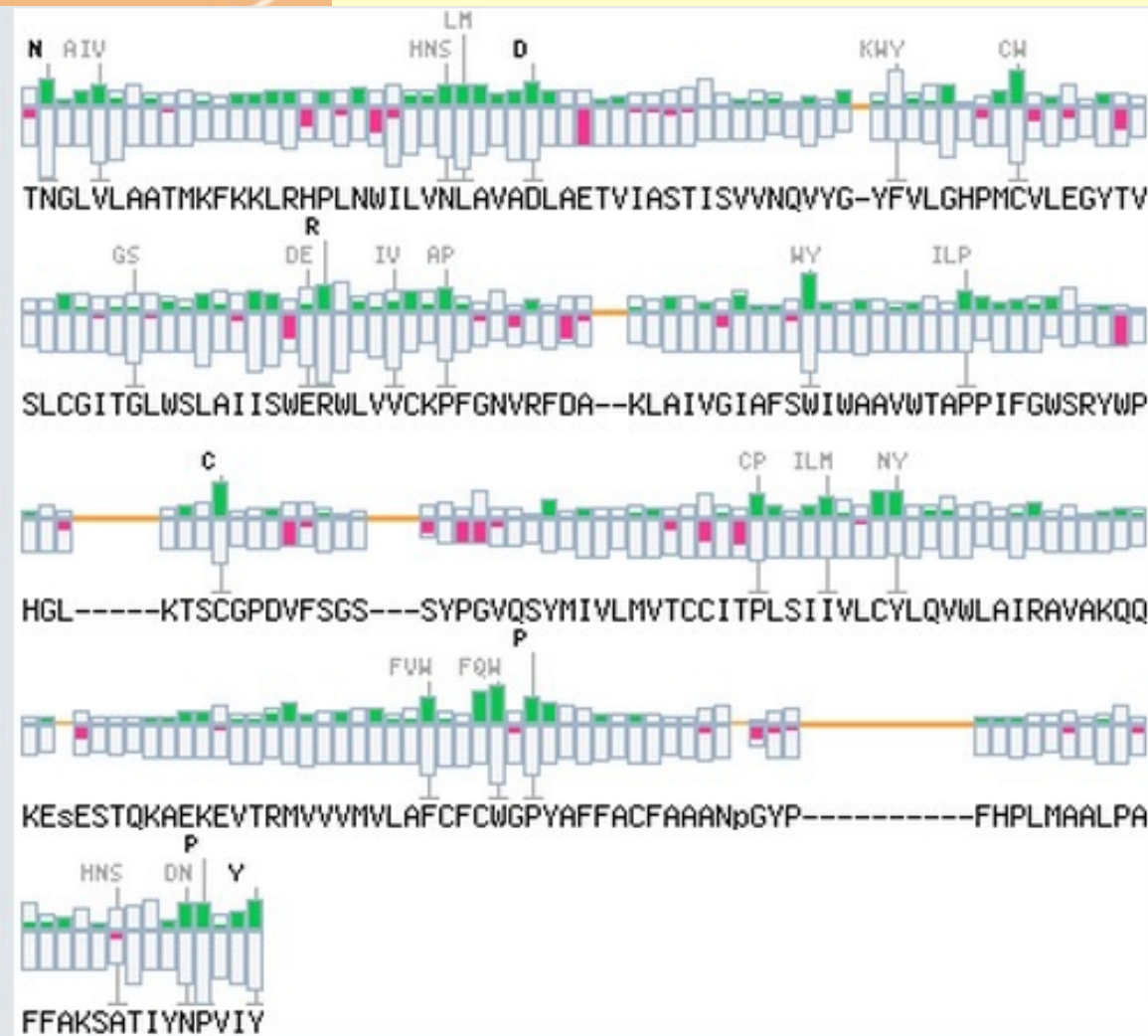


Status: **!**  
 pos.: **70-322**  
 raw-score = **1565**  
 N-score = **33.331**  
 E-value = **9.9e-27**

prf:G\_PROTEIN\_RECEP\_F1\_2  
*G-protein coupled receptors family 1 profile.*  
[\[ entry \]](#)  
[\[ graphics \]](#)



# MyHits Profile Hits (Cont.)





Status: **!**  
 pos.: **70-322**  
 raw-score = **167.5**  
 N-score = **63.946**  
 E-value = **2.4e-57**

pfam\_fs:7tm\_1  
 7 transmembrane  
 receptor (rhodopsin  
 family)  
 [ [entry](#) ]

[^ image ^](#)

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N.  
 PROSITE, a protein domain database for functional characterization and annotation.  
*Nucleic Acids Res.* 2010; **38**(Database issue):D161-6. [ [RIS](#) ]

EMBL-EBI  Se

 **InterPro**  
Protein sequence analysis & classification

Search InterPro...  
Examples: IPR020405, kinase, P51587, PF025

Home Search Release notes Download About InterPro Help Contact

## InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

### Analyse your protein sequence

```
>gil219521160|gb|AAI71754.1| Opsin 1 (cone pigments), medium-wave-sensitive [Homo sapiens]
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNNSNSTRGRPFEGPNYHIAPRW
VYHLTSVWMIFVVIASVFT
NGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASISVVNQVYGYFVLGHP
MCVLEGYTVSLCGITG
```

|

### Documentation

[About InterPro](#): core concepts, update frequency, how to cite, team and consortium members.

[FAQs](#): what are entry types and why are they important, interpreting results, downloading InterPro?

[Web services documentation](#)

### Protein focus



[Don't blame the cat - the toxoplasmosis effect](#)

The parasite

*Toxoplasma gondii*, the causative agent of toxoplasmosis, manipulates the behavior of rodents to ensure reaching its definitive host: the cat. Here we explore the biological mechanisms deployed to 'control' the host, and how this game of cat and mouse could be affecting you.

[HTML](#) | [All protein focus](#)

### Publications



[The InterPro protein families database: the classification resource after 15 years](#)

Our latest paper describing new developments on the InterPro website (*Nucleic Acids Research*, Dec 2014).

[HTML](#) | [PDF \(3,4Mb\)](#) | [All publications](#)

Search

Examples: IPR020405, kinase, P51587, PF02932, GO:0007165



LOADING...

## Your job is currently running... please be patient

The result of your job will appear in this browser window. This page refreshes automatically every 20 seconds.

You may bookmark this page to view your results later if you wish. Results are stored for 7 days.

Job ID: [iprscan5-S20150127-180017-0873-50865222-pg](#)



## Overview

Similar proteins

Structures

## Filter view on

### Entry type

- F** Family
- D** Domains
- R** Repeats
- S** Site

### Status

- ?** Unintegrated

### Colour by

[help](#)

- domain relationship
- source database

## Protein

Export Select format

# AAI71754.1

**Length** 364 amino acids

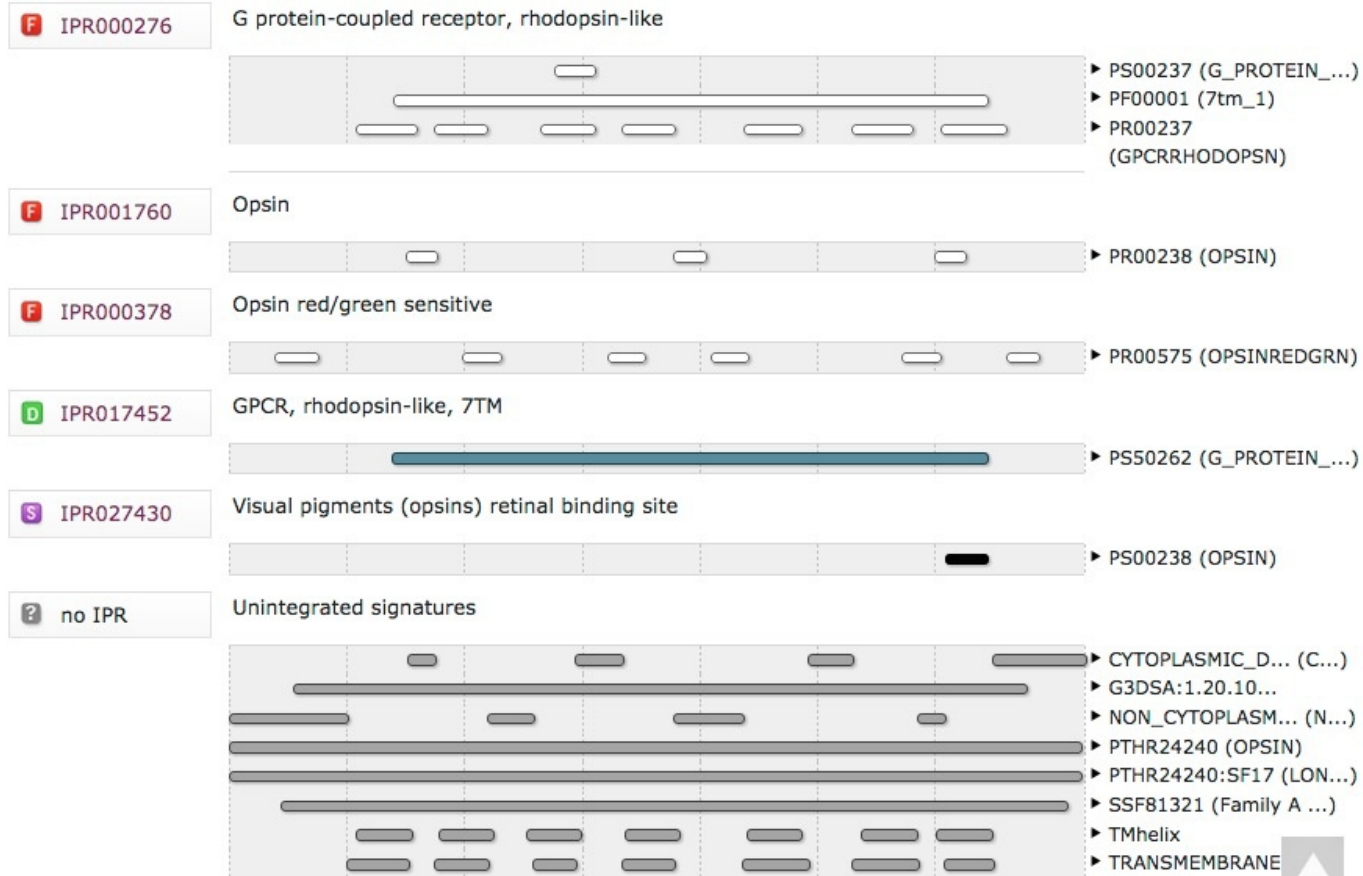
## Protein family membership

- F** G protein-coupled receptor, rhodopsin-like (IPR000276)
  - F** Opsin (IPR001760)
    - F** Opsin red/green sensitive (IPR000378)

## Domains and repeats



## Detailed signature matches



## GO term prediction

### Biological Process

- [GO:0007186](#) G-protein coupled receptor signaling pathway
- [GO:0007601](#) visual perception
- [GO:0007602](#) phototransduction

### Molecular Function

- [GO:0004930](#) G-protein coupled receptor activity

### Cellular Component

- [GO:0016021](#) integral component of membrane

# GO: Gene Ontology Database

<http://geneontology.org/>

## Search GO data

opn1mw

Search

## Enrichment analysis (beta)

Your gene IDs here...

biological process

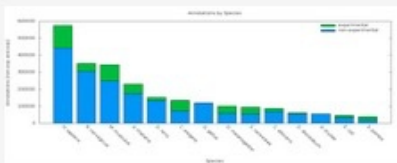
H. sapiens

Submit

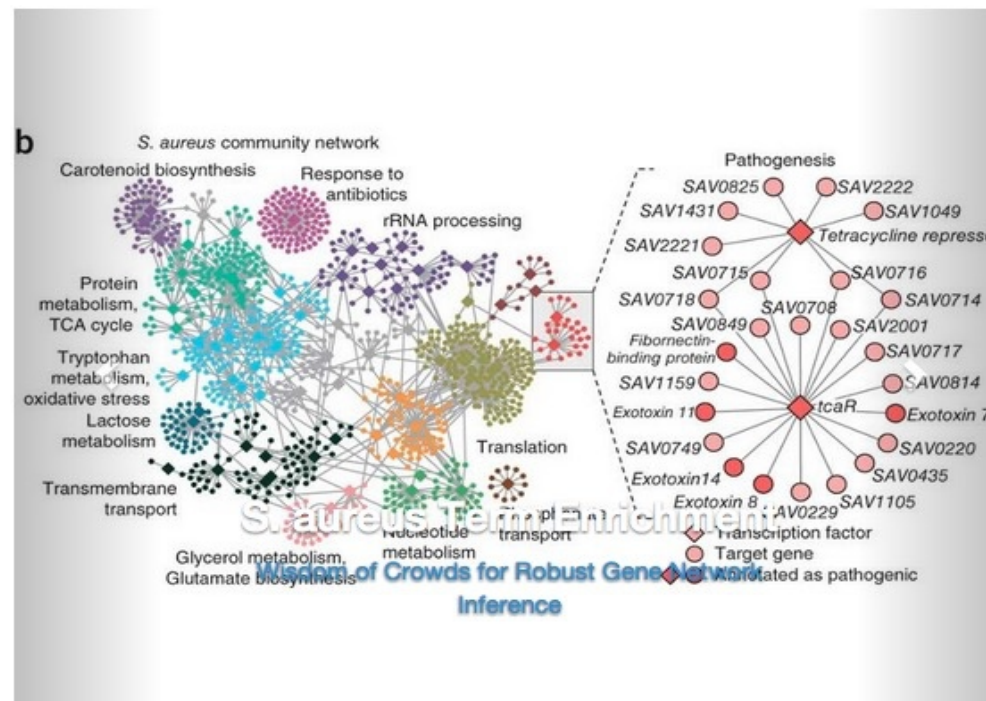
Advanced options

Powered by PANTHER

## Statistics



## Gene Ontology Consortium



## What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Ten quick tips for using the Gene Ontology **Important**
- Gene Ontology tools
- Enrichment analysis
- Downloads

Search



## Highlighted GO term

Representing "phases" in GO biological process

The GOC has recently introduced a new term **biological phase** (GO:0044848), as a direct subclass of biological process. This class represents a distinct period or stage during which biological processes can occur.

[more](#)

## Random FAQs

- What is annotation?
- What are mappings?
- What is "GO slimming" ?

[View all FAQs](#)

## On the web

GO term enrichment

New GO paper: TermGenie, a web-application for pattern-based ontology class gene...

New GO bioenergy paper - making



# GO: Gene Ontology for Opsin OPN1MW

<http://geneontology.org/>



## Search Directory

### Text search document selection

The following results were found for **opn1mw** using a general search over all text fields.

To narrow your search, select the type of document that you would like to search for and continue narrowing your search from the linked search page.

Ontology classes for GO.

0

Genes and gene products associated with GO terms.

12

Associations between GO terms and genes or gene products.

160



## Found entities

Total: 12; showing 1-10 Results count 10



<input type="checkbox"/>	Acc	Name	Taxon	PANTHER family	Type	Source	Direct annotation	Synonyms
<input type="checkbox"/>	<a href="#">Opn1mw4</a>	opsin 1 (cone pigments), medium-wave-sensitive, 4	Danio rerio	opsin pthr24240	gene_product	ZFIN	signal transduction protein-chromophore linkage <a href="#">more...</a>	
<input type="checkbox"/>	<a href="#">Opn1mw1</a>	opsin 1 (cone pigments), medium-wave-sensitive, 1	Danio rerio	opsin pthr24240	gene_product	ZFIN	protein-chromophore linkage signal transduction <a href="#">more...</a>	
<input type="checkbox"/>	<a href="#">Opn1mw</a>	Medium-wave-sensitive opsin 1	Rattus norvegicus		protein	UniProtKB	protein-chromophore linkage photoreceptor activity <a href="#">more...</a>	OPSG_RAT Opn1mw Gcp
<input type="checkbox"/>	<a href="#">Opn1mw</a>	opsin 1 (cone pigments), medium-wave-sensitive	Rattus norvegicus	opsin pthr24240	gene	RGD	protein-chromophore linkage biological_process <a href="#">more...</a>	
<input type="checkbox"/>	<a href="#">Opn1mw</a>	opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan)	Mus musculus	opsin pthr24240	protein	MGI	protein binding positive regulation of cytokinesis <a href="#">more...</a>	Gcp green long wavelength sensitive cone opsin <a href="#">more...</a>
<input type="checkbox"/>	<a href="#">OPN1MW</a>	Medium-wave-sensitive opsin 1	Cavia porcellus		protein	UniProtKB	plasma membrane	OPSG_CAVPO OPN1MW GCP
<input type="checkbox"/>	<a href="#">OPN1MW</a>	Medium-wave-sensitive opsin 1	Sciurus carolinensis		protein	UniProtKB	plasma membrane	OPSG_SCICA OPN1MW GCP
<input type="checkbox"/>	<a href="#">OPN1MW</a>	Medium-wave-sensitive opsin 1	Odocoileus virginianus virginianus		protein	UniProtKB	plasma membrane	OPSG_ODOVI OPN1MW GCP
<input type="checkbox"/>	<a href="#">OPN1MW</a>	Medium-wave-sensitive opsin 1	Oryctolagus cuniculus		protein	UniProtKB	plasma membrane	OPSG_RABIT OPN1MW GCP
<input type="checkbox"/>	<a href="#">OPN1MW2</a>	Medium-wave-sensitive opsin 1	Homo sapiens		protein	UniProtKB	G-protein coupled receptor signaling pathway G-protein coupled receptor activity <a href="#">more...</a>	H0Y642_HUMAN OPN1MW2 OPN1MW

# GO: Gene Ontology for Opsin OPN1MW

<http://geneontology.org/>

## Search Annotations

Information about Annotations search [?](#)

Free-text filtering

Your search is pinned to these filters

- + document\_category: annotation

User filters

- + bioentity: UniProtKB:H0Y642

- ▶ Source
- ▶ Assigned by
- ▶ Ontology (aspect)
- ▶ Evidence type
- ▶ PANTHER family
- ▶ Qualifier
- ▶ Taxon
- ▶ Direct annotation
- ▶ Inferred annotation
- ▶ Annotation extension

### Found entities

Total: 5; showing 1-5 Results count

<input type="checkbox"/>	Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Source	Taxon	Evidence	Evidence with	PANTHER family	Isoform	Reference
<input type="checkbox"/>	OPN1MW2	Medium-wave-sensitive opsin 1		G-protein coupled receptor signaling pathway		UniProtKB	Homo sapiens	IEA	UniProtKB-KW:KW-0297			GO_REF:0000038
<input type="checkbox"/>	OPN1MW2	Medium-wave-sensitive opsin 1		phototransduction		UniProtKB	Homo sapiens	IEA	InterPro:IPR000378			GO_REF:0000002
<input type="checkbox"/>	OPN1MW2	Medium-wave-sensitive opsin 1		G-protein coupled receptor activity		UniProtKB	Homo sapiens	IEA	UniProtKB-KW:KW-0297			GO_REF:0000038
<input type="checkbox"/>	OPN1MW2	Medium-wave-sensitive opsin 1		integral component of membrane		UniProtKB	Homo sapiens	IEA	UniProtKB-KW:KW-0812			GO_REF:0000038
<input type="checkbox"/>	OPN1MW2	Medium-wave-sensitive opsin 1		visual perception		UniProtKB	Homo sapiens	IEA	InterPro:IPR000378			GO_REF:0000002



# Bioinformatics Homework

<http://biochem118.stanford.edu/bioinformatics.html>

---

## Homework Assignment

- 1) Select a **protein** from [OMIM](#) or from [Entrez Gene](#) or from [UniProt](#) concerning the disease of interest to you. Copy and save the FASTA format of the protein file.
- 2) Search your protein for motifs with the [MyHits](#) Motif Scan Query. Be sure to Include Prosite Patterns, Prosite Frequent Patterns, Prosite Profiles, Profiles, Pfam HMMs (local Models) in your search. Please send me the MyHits you think are biologically significant and at least 1 or 2 hits which you think are not statistically or biologically significant. Please note that only the Profiles have expectation values. The Patterns do not have a measure of statistical significance.
- 3) Search your protein for blocks using the [InterPro](#) database. Please send me a few of the InterPro domains hits you think are significant and at least 1 or 2 hits which you think are not statistically or biologically significant. Please note that the default graphic output of InterPro does not list expectation values. You must switch to the Tabular view to obtain the statistical significance.
- 4) Search your protein for homology using the [BLAST](#) method. Please report two or three hits which are both statistically and biologically significant. Also report two or three hits which you think are neither statistically nor biologically significant. If your protein family is very large, you may have to ask BLAST to return more hits to find statistically insignificant hits.

# Statistical vs. Biological Significance

---

## Assignment

First, for each search (MyHits, InterPro and BLAST hit), I would like you to report some significance hits and describe why you think they are significant both statistically and biologically; also report some statistically insignificant hits (and why) and are any of your statistically insignificant hits, still significant biologically). To remind you what I said in class: a statistically significant find in the database search is always biologically significant, but a biologically significant result in the search is not necessarily always statistically significant.

## Statistical significance and expectation values.

Statistical significance is determined by the expectation value which gives you a measure of how likely this finding is based on pure chance. A finding with an E-value of 1 or greater is not significant because it could occur by pure chance. A finding with an E-value less than  $10^{-3}$  (one chance in a thousand) is generally considered statistically significant (unless of course you are doing a 1,000 searches!). So the lower the expectation value, the more significant the finding. Findings between  $10^{-3}$  and 1 are in the so called twilight zone and require some further analysis or experiments to determine their validity.

# Statistical vs. Biological Significance (cont.)

---

## InterPro

Unlike most of the other methods, InterPro sets a very high level of significance for a finding before it will report it. This means that you will usually not find any statistically insignificant hits for this particular search.

## Biological Significance

In order to determine biological significance you must read the biological properties (ontology terms are the most useful) of your protein and the biological properties of your findings. The findings may be significant because the finding defines a very closely related protein family (opsins for example) or a very broad family (G-coupled protein receptors or 7-transmembrane proteins) or a common structure (protein fold) or a specific function (retinal binding site) or a very specific catalytic activity. You should describe in words the level of the biological significance.



# Statistical vs. Biological Significance (cont.)

---

## MyHits

If you ask MyHits to return PATTERNs as well as motifs, you will notice that PATTERNs do not have E-values associated with them so there is no easy way to judge statistical significance. With pattern findings you are left only with judging biological significance. Also none of the Frequent patterns from MyHits are statistically significant.

## BLAST

If you do not have any insignificant hits from the BLAST search, it means that your protein family is very large and you have to ask BLAST to return more results using the Advanced Options at the bottom of the form. Only when you see hits with E-values  $> 0.001$  do you have insignificant findings.