

PFAM domains

Domains – modular building blocks

Domains reappear in different proteins.

Are often first identified from protein structure (common folds)

Homologous domains frequently have related functions.

New domain combinations can yield new functions.



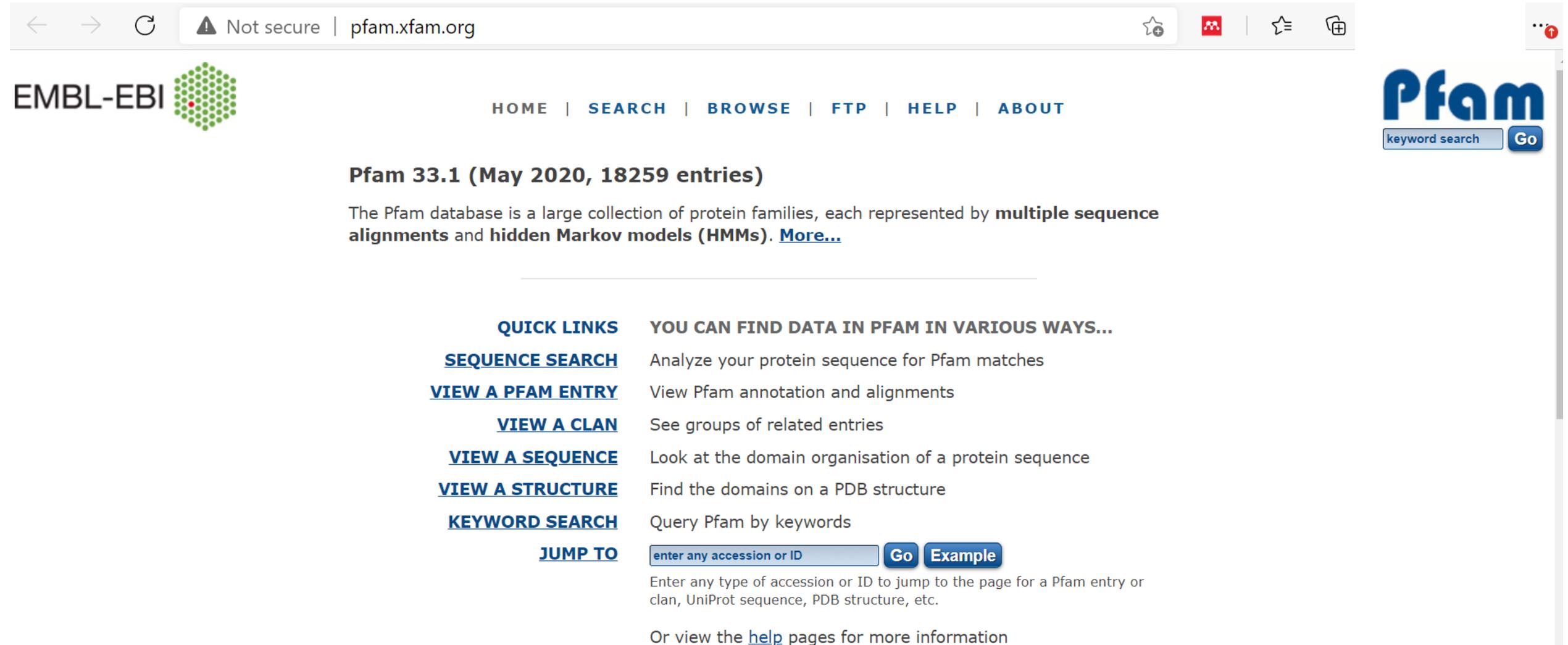
Domains – modular building blocks

Have conserved sequence motifs (or patterns of sequence conservation)

PFAM differentiates between DOMAINS as structural units and FAMILIES as collection of related protein regions. Both are modular building blocks and identifying domains and/or families help us understand a protein's function.



Pfam is the main protein domain database



The image shows a screenshot of the Pfam website homepage. At the top, there is a browser address bar showing 'pfam.xfam.org' with a 'Not secure' warning. The website header includes the EMBL-EBI logo on the left and the Pfam logo on the right, which includes a 'keyword search' box and a 'Go' button. A navigation menu in the center contains links for HOME, SEARCH, BROWSE, FTP, HELP, and ABOUT. The main content area features a section titled 'Pfam 33.1 (May 2020, 18259 entries)' with a brief description of the database. Below this, there are two columns of quick links and a 'JUMP TO' section with a search input field and a 'Go' button. The 'JUMP TO' section includes an 'Example' button and a note about entering accession or ID numbers.

EMBL-EBI 

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam
keyword search **Go**

Pfam 33.1 (May 2020, 18259 entries)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

QUICK LINKS	YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
SEQUENCE SEARCH	Analyze your protein sequence for Pfam matches
VIEW A PFAM ENTRY	View Pfam annotation and alignments
VIEW A CLAN	See groups of related entries
VIEW A SEQUENCE	Look at the domain organisation of a protein sequence
VIEW A STRUCTURE	Find the domains on a PDB structure
KEYWORD SEARCH	Query Pfam by keywords
JUMP TO	<input type="text" value="enter any accession or ID"/> Go Example

Enter any type of accession or ID to jump to the page for a Pfam entry or clan, UniProt sequence, PDB structure, etc.

Or view the [help](#) pages for more information

Pfam uses Hidden Markov models (HMMs) to find domains in protein sequences

- A HMM is made based on a SEED alignment.
- The **SEED alignment** is a small subset of sequences (from a larger protein family alignment) that are used to make a representative position specific PROFILE HMM.

Family: ANF_receptor (PF01094)

423 architectures 25562 sequences 4 interactions 711 species 632 structures

Alignments

We store a range of different sequence alignments for each Pfam-A family. In addition to the seed alignment, from which the family is built, we provide the full alignment, generated by searching the sequence database (retrieved from UniProtKB sequence database, the NCBI TrEMBL database, and the Meta-Transcriptome database). We generate alignments using four [representative proteomes](#)¹ (RP) sets, the UniProt database. [More...](#)

View options

We make a range of alignments for each Pfam-A family. You can see a description of each alignment [above](#). You can view these alignments in various ways but please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (71)	Full (25562)	Representative proteomes				UniProt (40484)	NCBI (177829)	Meta (2724)
			RP15 (4281)	RP35 (10298)	RP55 (19045)	RP75 (26920)			
Jalview	✓	✓	✓	✓	✓	✓	✓	✓	
HTML	✓	—	×	×	×	×	×	×	
PP/heatmap	× ₁	—	×	×	×	×	×	×	

¹Cannot generate PP/Heatmap alignments for seeds; no PP data available

Key: ✓ available, × not generated, — not available.

Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures

Jump to...

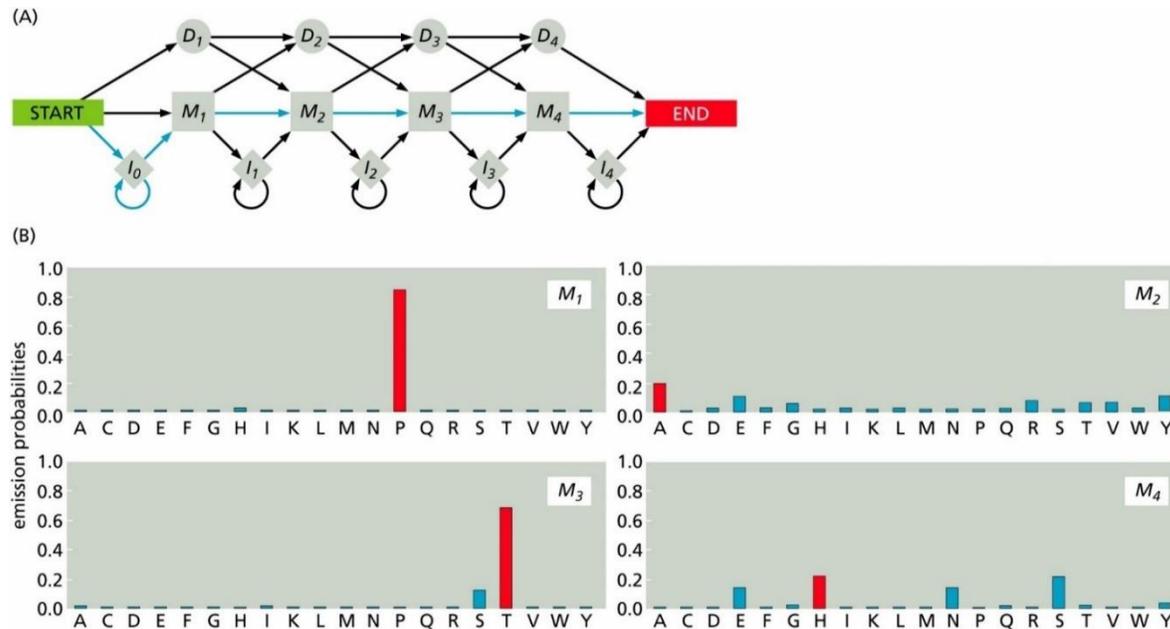
enter ID/acc **Go**

Note: A purple box highlights the text "This seed alignment is not to be confused with the seed alignment from BLAST" with a red triangle pointing to the 'Seed' column in the table.

- PFAM Seed alignment:

The representative alignment for a family/domain. Used to make a Hidden Markov Model (HMM). The HMM is used to find other sequences that belong to this group.

Profile HMM



M – match (the amino acid residue in the sequence matches the HMM)
 I – insert
 D – delete

Each event has an output (emission) probability (even I and D).

- PFAM Full alignments:

The alignment generated from searching a sequence database using the HMM.

Start with representative set of known members (sequences known to have a domain)

Make a seed alignment

Build HMM from seed alignment

Search database - were all known members not in seed found?

Add missed members to seed

Align found members to HMM to make a full alignment

No

Yes

PFAM workflow for making HMMs for domains

Pfam example

Based on typing in the UniProt code: Q02846

Protein: *GUC2D_HUMAN* (Q02846)

1 architecture 1

Summary

This is the summary of UniProt entry [GUC2D_HUMAN](#) (Q02846).

Description:	Retinal guanylyl cyclase 1 EC=4.6.1.2
Source organism:	Homo sapiens (Human) (NCBI taxonomy ID 9606)
Length:	1103 amino acids
Reference Proteome:	✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This sna that, although some UniProt entries may be removed *after* a Pfam release, these entries will not be removed from Pfa

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking below gives the domain boundaries for each of the domains. [More...](#)

[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
low_complexity	n/a	22	49
transmembrane	n/a	38	63
low_complexity	n/a	69	87
Pfam	ANF_receptor	72	400
low_complexity	n/a	135	146
low_complexity	n/a	156	173
low_complexity	n/a	357	370
disorder	n/a	436	439
low_complexity	n/a	457	483
transmembrane	n/a	465	487
disorder	n/a	523	527
disorder	n/a	534	535
disorder	n/a	539	540
Pfam	PK_Tyr_Ser-Thr	554	805
disorder	n/a	693	695
Pfam	HNOBA	808	865
coiled_coil	n/a	817	844
low_complexity	n/a	830	844
disorder	n/a	838	840
disorder	n/a	847	849
disorder	n/a	863	864
Pfam	Guanylate_cyc	871	1058
disorder	n/a	1065	1103

Pfam example

Based on performing a sequence search in PFAM with the FASTA sequence for Q02846.

Sequence search results

[Show](#) the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)



[Show](#) the search options and sequence that you submitted.

[Return](#) to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

[Show](#) or [hide](#) all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
				Start	End	Start	End	From	To			
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	3	182	183	210.2	1.8e-62
ANF_receptor	Receptor family ligand binding region	Domain	CL0144	72	400	73	394	2	333	354	145.7	1.8e-42
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	555	805	578	800	49	254	259	93.8	1.0e-26
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	180	221	221	25.4	8.3e-06

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas **Alignment coordinates** show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).

HMM is the part of the HMM that is matched from the sequence.

Pfam example

Based on performing a sequence search in PFAM with the FASTA sequence for Q02846.

Only the 4th domain is found to be close to complete (complete means from 1 to whatever the HMM length is). For domains 1-3, the entire HMM is not found and these are shown with jagged edges in the domain cartoon. The cartoon is jagged on the side that is not complete.

Sequence search results

[Show](#) the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)



[Show](#) the search options and sequence that you submitted.

[Return](#) to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

[Show](#) or [hide](#) all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
				Start	End	Start	End	From	To			
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	3	182	183	210.2	1.8e-62
ANF_receptor	Receptor family ligand binding region	Domain	CL0144	72	400	73	394	2	333	354	145.7	1.8e-42
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	555	805	578	800	49	254	259	93.8	1.0e-26
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	180	221	221	25.4	8.3e-06

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas **Alignment coordinates** show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).

HMM is the part of the HMM that is matched from the sequence.

Pfam example

>sp|Q02846|GUC2D_HUMAN Retinal guanylyl cyclase 1

Sequence search results

[Show](#) the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)



[Show](#) the search options and sequence that you submitted.

[Return](#) to the search form to look for Pfam domains on a new sequence.

Only the 4th domain is found to be close to complete (complete means from 1 to whatever the HMM length is). For domains 1-3, the entire HMM is not found and these are shown with jagged edges in the domain cartoon. The cartoon is jagged on the side that is not complete.

Significant Pfam-A Matches

[Show](#) or [hide](#) all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
ANF_receptor	Receptor family ligand binding region	Family	CL0144	72	400	73	394	2	331	352	146.0	1.4e-42	n/a	Show
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	554	805	577	800	48	255	260	94.1	7.5e-27	n/a	Show
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	180	221	221	25.4	7.9e-06	n/a	Show
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	3	182	183	210.2	1.7e-62	n/a	Show

Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.

European Molecular Biology Laboratory

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas **Alignment coordinates** show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).

HMM is the part of the HMM that is matched from the sequence.

Pfam example

>sp|Q02846|GUC2D_HUMAN Retinal guanylyl cyclase 1
<http://pfam.xfam.org/family/PF00211.19>

Sequence search results

[Show](#) the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)



[Show](#) the search options and sequence that you submitted.

[Return](#) to the search form to look for Pfam domains on a new sequence.

The screenshot shows the Pfam website interface. At the top, there is a navigation menu with links for HOME, SEARCH, BROWSE, FTP, HELP, and ABOUT. The main heading is 'Family: Guanylate_cyc (PF00211)'. Below this, there are statistics: 783 architectures, 16138 sequences, 2 interactions, 2387 species, and 179 structures. The 'Summary' section is active, showing the title 'Summary: Adenylate and Guanylate cyclase catalytic domain'. Below this, there are tabs for 'No Wikipedia article', 'Pfam', and 'InterPro'. The 'InterPro' tab is selected, showing the entry 'InterPro entry IPR001054'. The description states: 'Guanylate cyclases (EC) catalyse the formation of cyclic GMP (cGMP) from GTP. cGMP acts as an intracellular messenger, activating cGMP-dependent kinases and regulating cGMP-sensitive ion channels. The role of cGMP as a second messenger in vascular smooth muscle relaxation and retinal photo-transduction is well established. Guanylate cyclase is found both in the soluble and particulate fractions of eukaryotic cells. The soluble and plasma membrane-bound forms differ in structure, regulation and other properties [PUBMED:1349465, PUBMED:1356629, PUBMED:1680765, PUBMED:1982420]. Most currently known plasma membrane-bound forms are receptors for small polypeptides. The soluble forms of guanylate cyclase are cytoplasmic heterodimers having alpha and beta subunits. In all characterised eukaryote guanylyl- and adenylyl cyclases, cyclic nucleotide synthesis is carried out by the conserved class III cyclase domain.' Below this, there is a 'Gene Ontology' section with a table of terms:

Gene Ontology	Term
Molecular function	phosphorus-oxygen lyase activity (GO:0016849)
Biological process	cyclic nucleotide biosynthetic process (GO:0009190)
	intracellular signal transduction (GO:0035556)

Domain organization:

How many different domain architecture are this domain found in? Here the first 5 of 783 domain architectures (or domain organizations) are shown.



Family: *Guanylate_cyc* (PF00211)

783 architectures 16138 sequences 2 interactions 2387 species 179 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

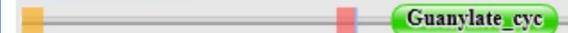
Jump to...

Domain organisation

Below is a listing of the unique domain organisations or architectures in which this domain is found. [More...](#)

There are 4208 sequences with the following architecture: *Guanylate_cyc*

[H2BSX8_9FLAO](#) [Gillisia limnaea DSM 15749] Adenylate/guanylate cyclase {ECO:0000313|EMBL:EHQ01508.1} (619 residues)



[Show](#) all sequences with this architecture.

There are 800 sequences with the following architecture: *CHASE2*, *Guanylate_cyc*

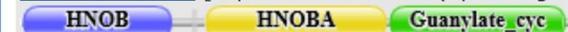
[X6GGD4_9RHIZ](#) [Mesorhizobium sp. L48C026A00] Uncharacterized protein {ECO:0000313|EMBL:ESZ18435.1} (639 residues)



[Show](#) all sequences with this architecture.

There are 643 sequences with the following architecture: *HNOB*, *HNOBA*, *Guanylate_cyc*

[W5MNT1_LEPOC](#) [Lepisosteus oculatus (Spotted gar)] Uncharacterized protein {ECO:0000313|Ensembl:ENSLOCP00000010040} (612 residues)



[Show](#) all sequences with this architecture.

There are 640 sequences with the following architecture: *ANF_receptor*, *Pkinase_Tyr*, *Guanylate_cyc*

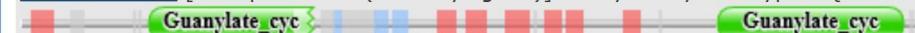
[GUC2G_MOUSE](#) [Mus musculus (Mouse)] Guanylate cyclase 2G EC=4.6.1.2 (1100 residues)



[Show](#) all sequences with this architecture.

There are 625 sequences with the following architecture: *Guanylate_cyc* x 2

[L5KRF9_PTEAL](#) [Pteropus alecto (Black flying fox)] Adenylate cyclase type 3 {ECO:0000313|EMBL:ELK14032.1} (1008 residues)



Alignments: Shows the seed alignment and full alignment plus other alignments based on different criteria. Here the seed alignment used to make the HMM contained 20 sequences but it found >16000 sequences in Swissprot (the full alignment).



Family: *Guanylate_cyc* (PF00211)

783 architectures 16138 sequences 2 interactions 2387 species 179 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to...

enter ID/acc

Alignments

We store a range of different sequence alignments for families. As well as the seed alignment from which the family is built, we provide the full alignment, generated by searching the sequence database ([reference proteomes](#)) using the family HMM. We also generate alignments using four [representative proteomes](#) (RP) sets, the UniProtKB sequence database, the NCBI sequence database, and our metagenomics sequence database. [More...](#)

View options

We make a range of alignments for each Pfam-A family. You can see a description of each [above](#). You can view these alignments in various ways but please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (20)	Full (16138)	Representative proteomes				UniProt (35875)	NCBI (64098)	Meta (3136)
			RP15 (5149)	RP35 (9203)	RP55 (14591)	RP75 (21280)			
Jalview	✓	✓	✓	✓	✓	✓	✓	✓	
HTML	✓	—	✗	✗	✗	✗	✗	✗	
PP/heatmap	✗ ₁	—	✗	✗	✗	✗	✗	✗	

¹Cannot generate PP/Heatmap alignments for seeds; no PP data available

Key: ✓ available, ✗ not generated, — not available.

Format an alignment

Seed (20)	Full (16138)	Representative proteomes				UniProt (35875)	NCBI (64098)	Meta (3136)
		RP15 (5149)	RP35 (9203)	RP55 (14591)	RP75 (21280)			

Clan: Related domains but with specific seed alignments

Family: *Guanylate_cyc* (PF00211)

 783 architectures

 16138 sequences

 2 interactions

 2387 species

 179 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to... 

enter ID/acc

Pfam Clan

This family is a member of clan [Nucleot_cyclase](#) (CL0276), which has the following description:

This superfamily includes adenylyl cyclase and the GGDEF domain [1].

The clan contains the following 5 members:

[DUF3692](#)

[EAL](#)

[GCH_III](#)

[GGDEF](#)

[Guanylate_cyc](#)

HMM logo: Stacks are scaled based on how much the site varies (the sequence entropy). Letters are scaled based on letter frequency.

Family: *Guanylate_cyc* (PF00211)

 783 architectures

 16138 sequences

 2 interactions

 2387 species

 179 structures

- Summary
- Domain organisation
- Clan
- Alignments
- HMM logo**
- Trees
- Curation & model
- Species
- Interactions
- Structures

Jump to... 

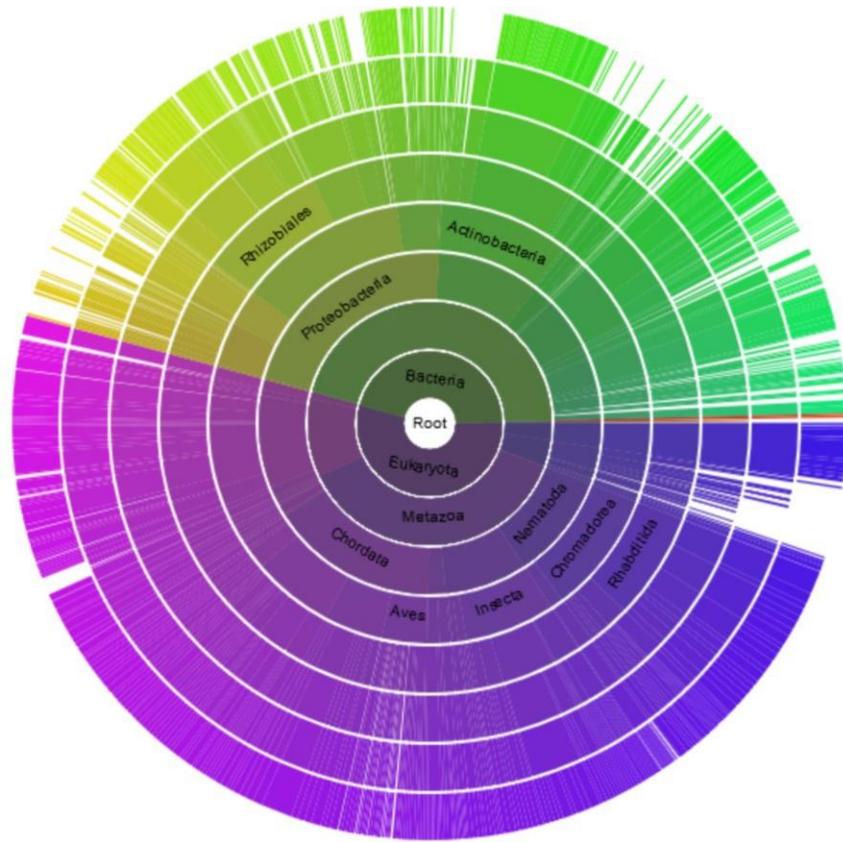
enter ID/acc

HMM logo

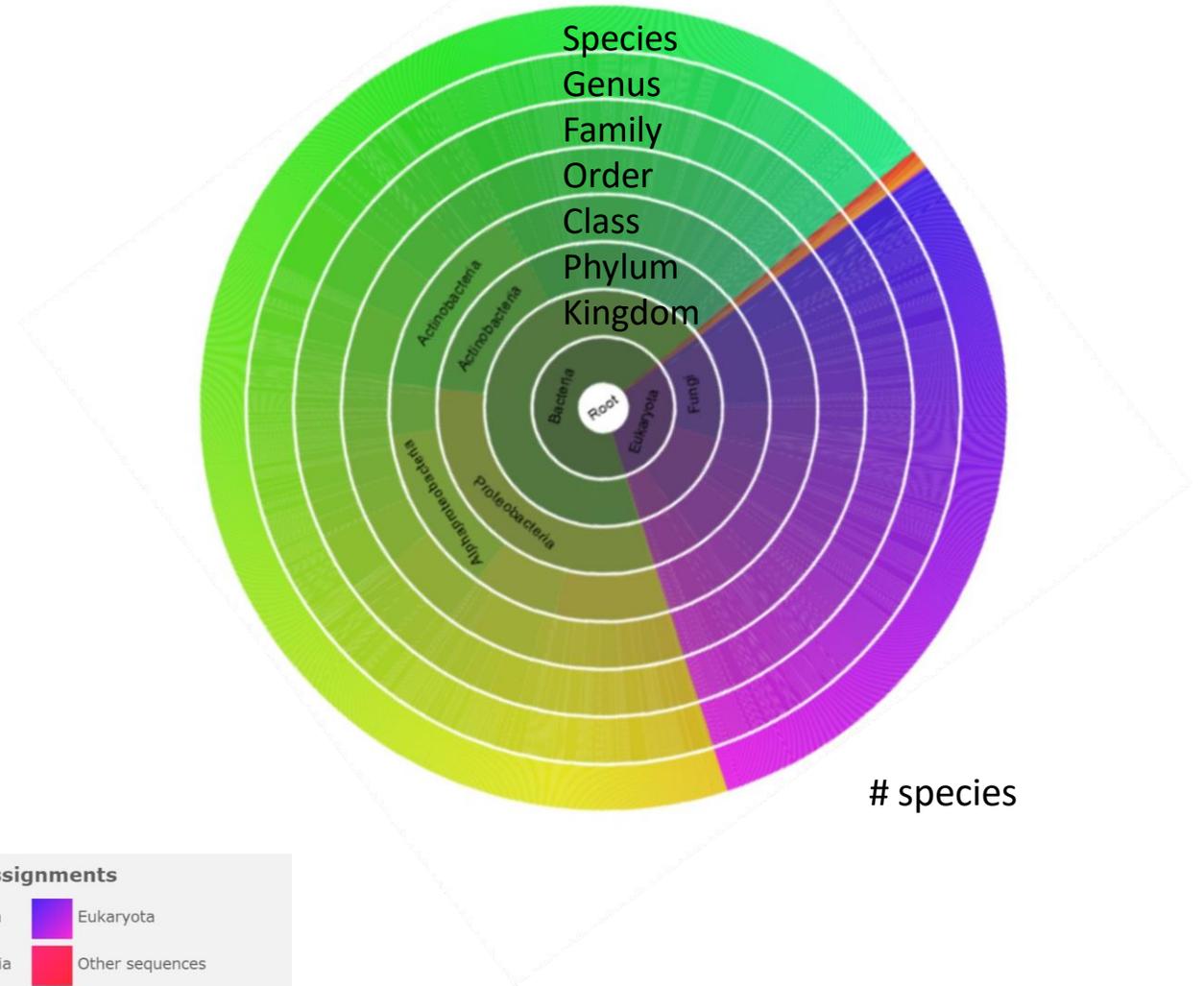
HMM logos is one way of visualising profile HMMs. Logos provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them [here](#). [More...](#)



Species: The SUNBURST TREE – Overview over which species the domain has been found in and how many sequences per species



sequences



species

Interactions: Known interacting partners

Family: *Guanylate_cyc* (PF00211)

 783 architectures

 16138 sequences

 2 interactions

 2387 species

 179 structures

- Summary
- Domain organisation
- Clan
- Alignments
- HMM logo
- Trees
- Curation & model
- Species

Interactions

Structures

Jump to... 

enter ID/acc

Interactions

There are **2** interactions for this family. [More...](#)

[G-alpha](#)

[Guanylate_cyc](#)



Any questions?

KEGG
adds
CONTEXT





KEGG ▾

Search

Help

» Japanese

KEGG Home

[Release notes](#)

[Current statistics](#)

KEGG Database

[KEGG overview](#)

[Searching KEGG](#)

[KEGG mapping](#)

[Color codes](#)

KEGG Objects

[Pathway maps](#)

[Brite hierarchies](#)

[KEGG DB links](#)

KEGG Software

[KEGG API](#)

[KGML](#)

KEGG FTP

[Subscription](#)

[Background info](#)

[GenomeNet](#)

[DBGET/LinkDB](#)

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

See [Release notes](#) (January 1, 2021) for new and updated features.

New article [KEGG: integrating viruses and cellular organisms](#)

🟡 Main entry point to the KEGG web service

[KEGG2](#)

[KEGG Table of Contents](#) [[Update notes](#) | [Release history](#)]

🟡 Data-oriented entry points

[KEGG PATHWAY](#)

[KEGG pathway maps](#)

[KEGG BRITE](#)

[BRITE hierarchies and tables](#)

[KEGG MODULE](#)

[KEGG modules](#)

[KEGG ORTHOLOGY](#)

[KO functional orthologs](#) [[Annotation](#)]

[KEGG GENOME](#)

[Genomes](#) [[Pathogen](#) | [Virus](#) | [Plant](#)]

[KEGG GENES](#)

[Genes and proteins](#) [[SeqData](#)]

[KEGG COMPOUND](#)

[Small molecules](#)

[KEGG GLYCAN](#)

[Glycans](#)

[KEGG REACTION](#)

[Biochemical reactions](#) [[RModule](#)]

Classification

[Pathway](#)

[Brite](#)

[Brite table](#)

[Module](#)

[KO \(Function\)](#)

[Organism](#)

[Virus](#)

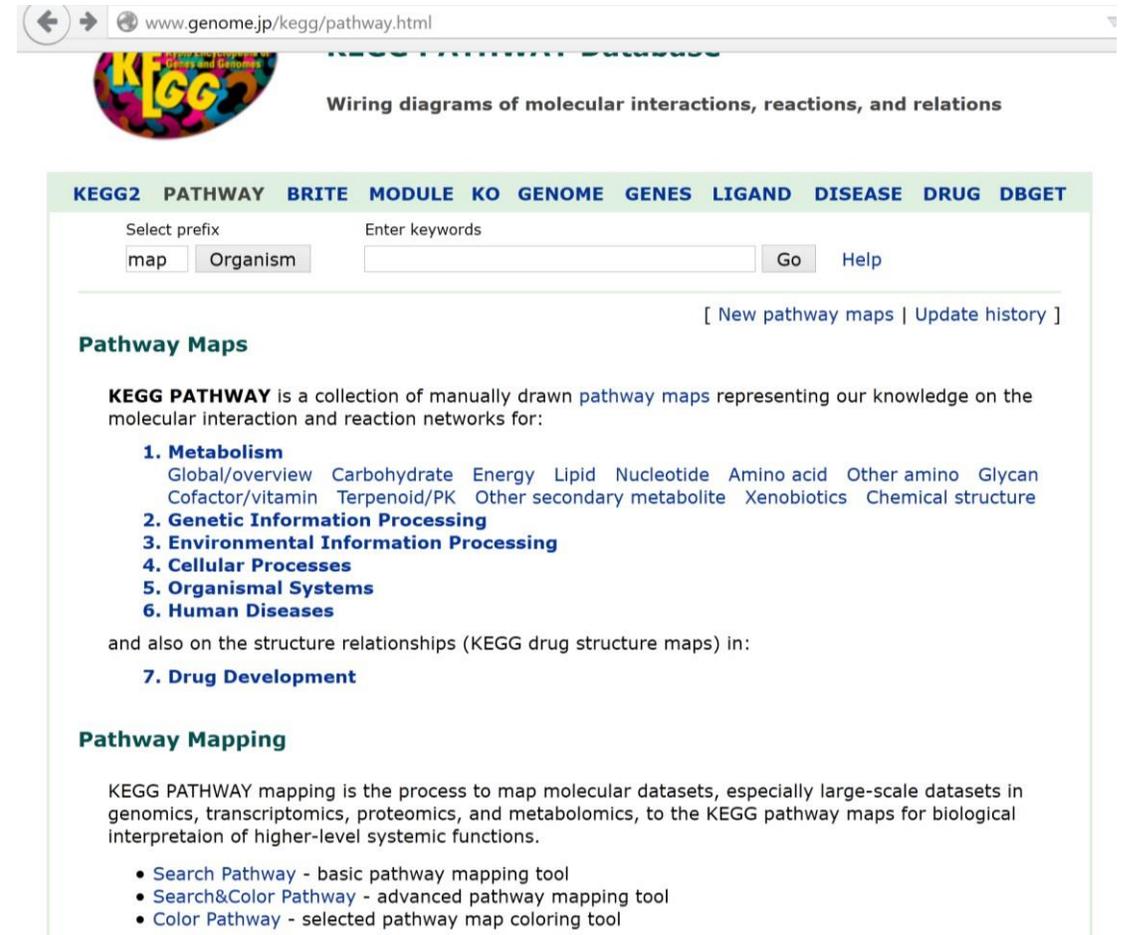
[Compound](#)

[Network](#)

[Disease \(ICD\)](#)

PATHWAY CONTEXT

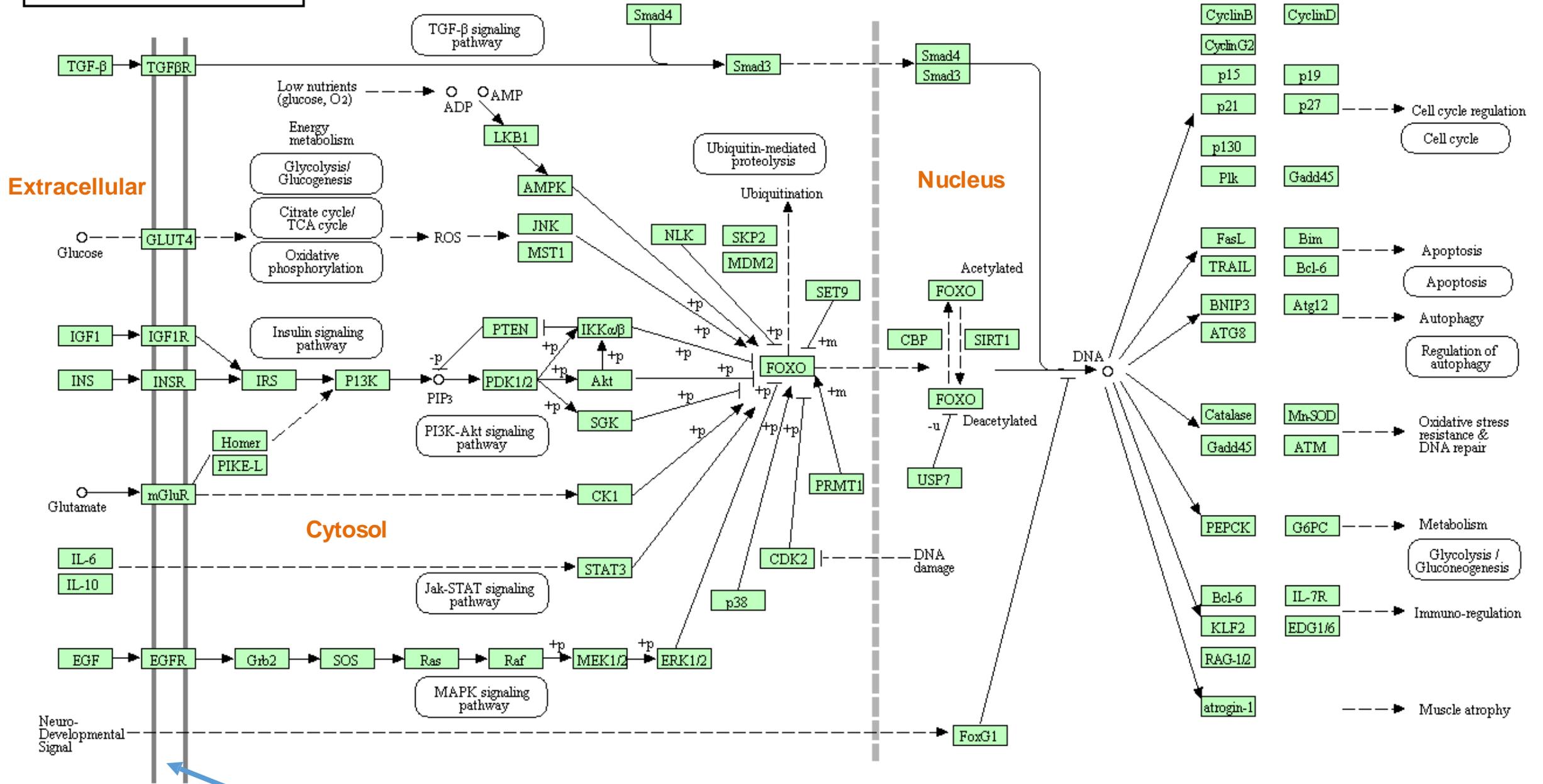
- In which pathway does the protein and its homologs act?
- Are they enzymes? Where is the substrate coming from and where is the product going?
- Are they involved in signaling? Which proteins do they signal to/from?
- What species have a certain pathways?
- What genes/proteins of a pathway does a species have?



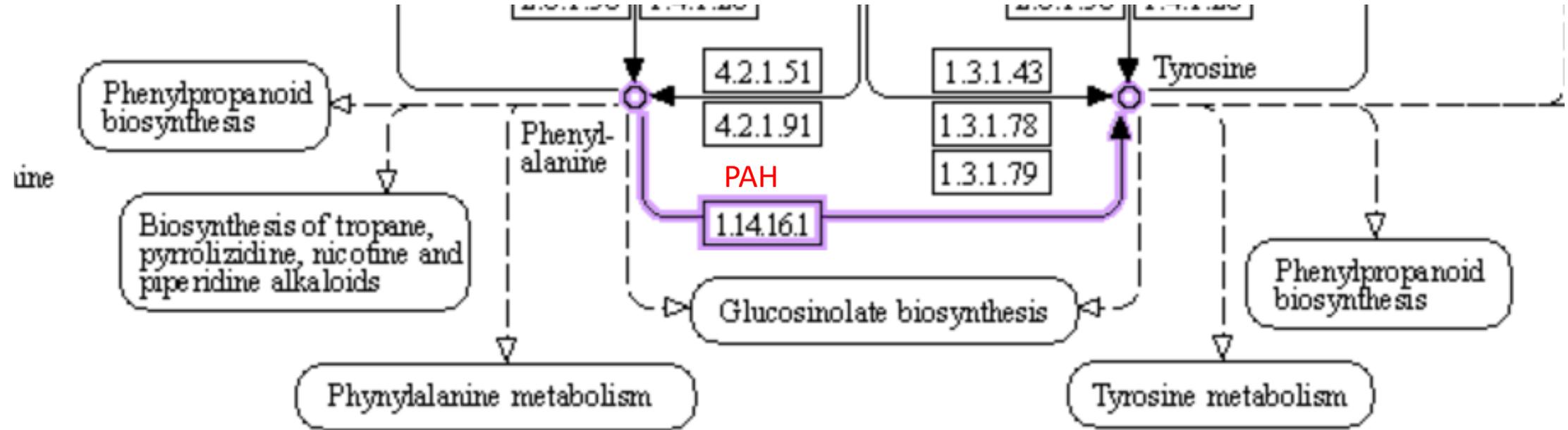
The screenshot shows the KEGG Pathway website. The browser address bar displays "www.genome.jp/kegg/pathway.html". The KEGG logo is visible, along with the tagline "Wiring diagrams of molecular interactions, reactions, and relations". A navigation menu includes "KEGG2", "PATHWAY", "BRITE", "MODULE", "KO", "GENOME", "GENES", "LIGAND", "DISEASE", "DRUG", and "DBGET". Below the menu is a search area with a "Select prefix" dropdown set to "map" and "Organism", and an "Enter keywords" input field. There are "Go" and "Help" buttons. A link for "[New pathway maps | Update history]" is present. The main content area is titled "Pathway Maps" and describes KEGG PATHWAY as a collection of manually drawn pathway maps. It lists seven categories: 1. Metabolism, 2. Genetic Information Processing, 3. Environmental Information Processing, 4. Cellular Processes, 5. Organismal Systems, 6. Human Diseases, and 7. Drug Development. A "Pathway Mapping" section explains the process of mapping molecular datasets to KEGG pathway maps and lists three tools: Search Pathway, Search&Color Pathway, and Color Pathway.

KEGG – Kyoto Encyclopedia of Genes and
Genomes <http://www.kegg.jp/>

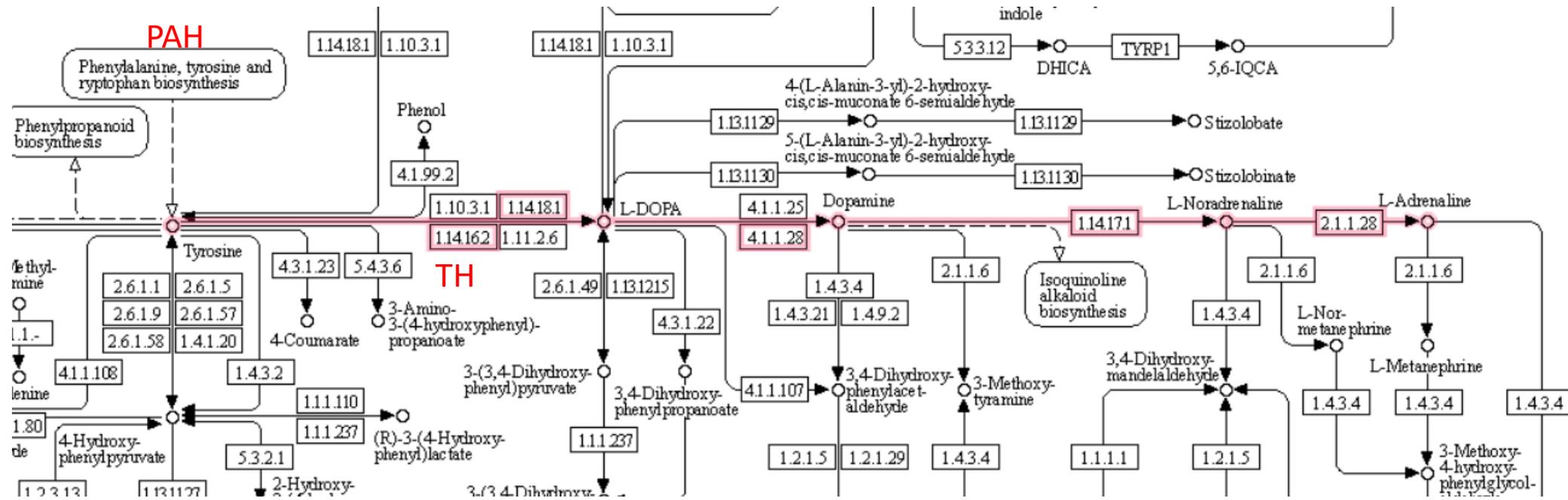
FOXO SIGNALING PATHWAY



More about the aromatic amino acid hydroxylases...



More about the aromatic amino acid hydroxylases...





Search MedlinePlus

About MedlinePlus

Health Topics Drugs & Supplements Genetics Medical Tests Videos & Tutorials



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>

Get the latest research information from NIH: <https://covid19.nih.gov>

Learn more about COVID-19 and you from HHS: <https://combatcovid.hhs.gov>

Home → Medical Tests → Phenylketonuria (PKU) Screening

Phenylketonuria (PKU) Screening

What is a PKU screening test?

A PKU screening test is a blood test given to newborns 24–72 hours after birth. PKU stands for **phenylketonuria**, a rare disorder that prevents the body from properly breaking down a substance called phenylalanine (Phe). Phe is part of **proteins** that are found in many foods and in an artificial sweetener called aspartame.

If you have PKU and eat these foods, Phe will build up in the blood. High levels of Phe can permanently damage the **nervous system** and brain, causing a variety of health problems. These include **seizures**, psychiatric problems, and severe intellectual disability.

PKU is caused by a **genetic mutation**, a change in the normal function of a gene. Genes are the basic units of heredity passed down from your mother and father. For a child to get the disorder, both the mother and father must pass down a mutated PKU gene.

Although PKU is rare, all newborns in the United States are required to get a PKU test.

