# HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

keyword search





### Powering down the Pfam website

On October 5th, we will start redirecting the traffic from Pfam ( $\underline{pfam.xfam.org}$ ) to InterPro ( $\underline{www.ebi.ac.uk/interpro}$ ). The Pfam website will be available at  $\underline{legacy.pfam.xfam.org}$  until January 2023, when it will be decommissioned. You can read more about the sunset period in our  $\underline{blog.post}$ .

# **Pfam Help**

O architectures O sequences O interactions O species O structures

#### Summary

#### Changes

### **Getting Started**

Training

FAQ

Glossary

Scores

**Citing Pfam** 

Pfam and Wikipedia

**Linking to Pfam** 

# Guide to Graphics

RESTful interface

Pfam database

**FTP** site

Privacy

**Team Members** 

**Contact Us** 

# Jump to... 🌵

enter ID/acc Go

# A Guide to the Pfam Domain Graphics

One of the visualisations provided by the Pfam website is a graphical representation of the features found within a sequence, termed *domain graphics*. There are a variety of different shapes and styles and each one has a particular meaning. This page gives an in-depth description of the elements of Pfam domain graphics.

The library that generates the images in this page and throughout the Pfam site uses a  $\underline{\rm JSON}$  string to describe the domain graphic. Each of the example graphics in this page is followed by a link that can be used to show the JSON snippet that produced it.

# **Generating graphics**

You can try generating your own graphics using the <u>domain graphics generator</u>. The JSON descriptions in this page can be pasted directly into the generator to produce the graphics that you see here.

You can also generate the domain graphics for specific sequences, using the <u>UniProt graphics generator</u>.

## Using the domain graphics code

Finally, if you would like to use the javascript library in your own site, we have put together <u>an example page</u>, showing how to set up the library and its dependencies. Look at the source code of the page for an explanation.

### The sequence

The base sequence, undecorated by any domains or features, is represented by a plain grey bar:

```
Hide JSON
{
    "length" : "400"
}
```

The length of the domain graphic that is drawn is proportional to the length of the sequence itself. The graphics in this page are drawn with a X-scale of 0.5 pixels per amino-acid, so that a 400 residue sequence will result in a 200 pixel-wide image. Any domains or features which are drawn on the sequence are also scaled by the same factor.

back to top

## Pfam-A

The high quality, curated *Pfam-A* domains are classified into one of six different types: *family, domain, coiled-coil, disordered, repeat* and *motif* (more details). These different classification types are rendered slightly differently.

### Family/domain

It is possible for a sequence to match either the full length of a Pfam HMM (a full length match), or to match a portion of an HMM (a fragment match). The two types of match are rendered differently.

Both family and domain entries are rendered as rectangles with curved ends when the sequence is a full length match. Different types of domain are displayed with different colours. When the domain image is long enough, the domain name is shown within the domain itself. In most cases, you can click on the domains to visit the "family page" for that domain. Moving the mouse over the domain image should also display a tooltip showing the domain name, as well as the start and end positions of the domain.

```
},
{
    "type" : "pfama",
    "text" : "LongFamilyNamesNotShown",
    "colour" : "#399",
    "display" : true,
    "startStyle" : "straight",
    "endStyle" : "straight",
    "start" : "210",
    "end" : "250",
    "aliStart" : "215",
    "aliEnd" : "245"
}
```

From Pfam 24.0 onwards, Pfam has been generated using <u>HMMER3</u>, which introduces the concept of "<u>envelope coordinates</u>" for a match. Envelope regions are represented in domain graphics as lighter coloured regions. The graphic above shows short envelope regions at the ends of both domains.

When the sequence does not match the full length of the HMM that models a Pfam entry, matching domain fragments are shown. When a sequence match does not pass through the first position in the HMM, the N-terminal side of the domain graphic is drawn with a jagged edge instead of a curved edge. Similarly, when a sequence match does not pass through the last position of the HMM, the C-terminal side of the domain graphic is drawn with a jagged edge. In some rarer cases, the sequence match may not pass through either of the first or last positions of the HMM, in which case both sides are drawn with jagged edges. Examples of all three cases are shown here:

### back to top

# Repeat/motif

Repeats and motifs are types of Pfam domain which do not form independently folded units. In order to distinguish them from domains of type family and domain, repeats and motifs are represented by rectangles with straight edges. As for families and domains, partial matches are represented with jagged edges.

```
"endStyle" : "straight",
"start" : "82",
"end" : "118"
                "type" : "pfama",
"text" : "HEAT",
"colour" : "#1fc01f",
"display": "true",
"startStyle" : "straight",
"endStyle" : "straight",
"start" : "120",
"end" : "155"
         },
                 "type" : "pfama",
"text" : "HEAT",
"colour" : "#1fc01f",
"display": "true",
"startStyle" : "straight",
"endStyle" : "straight",
"start" : "159",
"end" : "195"
         }
1
```

#### back to top

#### Discontinuous nested domains

Some domains in Pfam are disrupted by the insertion of another domain (or domains) within them. A number of names have been given to this arrangement: discontinuous (referring to the outer domain), inserted or nested (both referring to the inner domain). For example, in many sequences containing an IMPDH domain, the IMPDH domain is continuous along the primary sequence. However, in some cases the linear sequence of the IMPDH domain is broken by the insertion of a <u>CBS domain</u>, as shown below.

Where three-dimensional structures are available for representatives of a Pfam domain, it is generally clear that the three-dimensional arrangement of the domain containing the nested domain is maintained. Typically the nested domain is found inserted within a surface exposed loop, having little or no effect on the structure of the other domain. Such an arrangement explains why and how these nested domains can be functionally tolerated.

To represent this arrangement of domain graphically, the discontinuous domain is represented in two parts (as shown below). These two parts are joined by a line bridging them.

```
Hide JSON
{
     "length" : "200",
"regions" : [
         {
  "type" : "pfama",
  "TMPDH",
              "text" : "IMPDH"
              "colour": "#1fc01f",
"display": "true",
             "startStyle": "curved",
"endStyle": "jagged",
"start": "5",
"end": "80"
         },
              "type" : "pfama",
              "text" : "CBS",
"colour" : "#c00f0f",
"display": "true",
              "startStyle" : "curved",
"endStyle" : "curved",
"start" : "81",
              "end" : "135"
              "type" : "pfama",
"text" : "IMPDH",
             "text": "IMPDH",
"colour": "#1fc01f",
"display": "true",
"startStyle": "jagged",
"endStyle": "curved",
"start": "136",
"end": "197"
         }
      "markups" : [
         {
             "type" : "Nested",
"colour" : "#000000",
"display" : true,
"v_align" : "top",
"start" : "76",
"end" : "136"
         }
    ]
}
```

# Other sequence motifs

In addition to domains, smaller sequences motifs are represented by the domain graphics. Currently the following motifs are represented: signal peptides, low complexity regions, coiled-coils and transmembrane regions. These usually take lower prority than other regions that are drawn and they are therefore often obscured by, for example, a Pfam-A graphic being drawn over the top of them. An example of each motif is shown here.

```
Hide JSON
{
    "length" : "200",
"motifs" : [
         {
             "type" : "sig_p",
"colour" : "#ff9c00",
"display" : true,
             "start" : 1,
"end" : 27
              "type" : "low_complexity",
             "colour": "#0FF",
"display": true,
"start": 39,
             "end" : 47
             "type" : "low_complexity",
"colour" : "#0FF",
"display" : true,
             "start" : 67,
"end" : 76
             "type" : "coiled_coil",
"colour" : "#9cff00",
"display" : true,
"start" : 103,
"end" : 123
             "type" : "transmembrane",
"colour" : "#F00",
"display" : true,
"start" : 155,
"end" : 175
             "type" : "transmembrane",
"colour" : "#F00",
"display" : true,
"start" : 180,
"end" : 195
        }
}
```

## back to top

# Signal peptides

Signal peptides are short regions (<60 residues long) found at the N-terminus of proteins, which direct the post-translational transport of a protein and are subsequently removed by peptidases. More specifically, a signal peptide is characterised by a short hydrophobic helix (approximately 7-15 residues). This helix is preceded by a slight positively charged region of highly variable length (approximately 1-12 residues). Between the hydrophobic helix and the cleavage site is a somewhat polar and uncharged region, of between 3 and 8 amino-acids. In Pfam, we use Phobius for the prediction of signal peptides and represent them graphically by a small orange box.

**J. Mol. Biol.** (2004) 338(5):1027-36

### back to top

### Low complexity regions

Low complexity regions are regions of biased sequence composition, usually comprised of different types of repeats. These regions have been shown to be functionally important in some proteins, but they are generally not well understood and are masked out to focus on globular domains within the protein.

Within Pfam, we use **SEG** to calculate low complexity regions in Pfam. The presence of a low complexity region is indicated by a cyan rectangle.

<u>A global compositional complexity measure for biological sequences: AT-rich and GC-rich genomes encode less complex proteins</u>: H. Wan and J.C. Wootton

Comput. Chem. (2000) 24(1):71-94

Non-globular domains in protein sequences: automated segmentation using complexity measures : J.C. Wootton Comput. Chem. (1994) 18(3):268-85

back to top

### **Disordered regions**

We use the IUPred method for the prediction of disordered regions in the query sequence. The <u>IUPred server</u> provides more detailed disorder prediction results than currently offered here.

Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Z. Dosztanyi, B. Msszaros, I. Simon

**Brief Bioinform** (2010) 11:225-43

<u>IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.</u> Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon

Bioinformatics (2005) 21:3433-3434

### back to top

#### Coiled-coils

Coiled coils are motifs found in proteins that structurally form alpha-helices that wrap or wind around each other. Normally, two to three helices are involved, but cases of up to seven alpha-helices have been reported. Coilded-coild are found in a wide variety of proteins, many functionally very important. In Pfam we use <a href="ncoils">ncoils</a> , to identify these motifs. Coiled-coils are represented by a small lime-green rectangle.

#### back to top

### Transmembrane regions

Integral membrane proteins contain one or more *transmembrane regions* that are comprised of an alpha-helix that passes through or "spans" a membrane. Transmembrane helices are quite variable in length, with the average being about 20 amino-acids in length. Again, <a href="Phobius">Phobius</a> is used for the prediction of transmebrane regions, which are represented by a red rectangle.

#### back to top

### **Other Sequence features**

Below is a demonstration of how disulphide bridges and active residues are representated in Pfam. Each of these features can appear above or below the sequence, but in this case the disulphide bridges are shown above the sequence and the active site residues below the line.

```
╗ 1
                   Peptidase C1
                                                     Hide JSON
"length" : "400",
"regions" : [
      "colour": "#1fc01f",
"endStyle": "curved",
"startStyle": "curved",
"display": true,
       "end": "104",
"href": "/family/Inhibitor_I29",
"text": "Inhibitor_I29",
       "metadata" : {
   "scoreName" : "e-value",
   "score" : "1.3e-38",
           "description" : "Inhibitor_I29",
"accession" : "PF08246",
           "end" : "104",
"database" : "pfam",
"identifier" : "Inhibitor_I29",
           "type" : "Domain",
           "start" : "48"
        "type" : "pfama",
       "start" : "48"
       "colour" : "#c00f0f",
"endStyle" : "curved",
"startStyle" : "curved",
       "display": true,
"end": "343",
"href": "/family/Peptidase_C1",
"text": "Peptidase_C1",
       "modelLength" : "307",
       "metadata" : {
   "scoreName" : "e-value",
   "score" : "1.3e-38",
           "description" : "Peptidase_C1",
"accession" : "PF00112",
           "end" : "343",
           "database" : "pfam",
"identifier" : "Peptidase_C1",
           "type" : "Domain",
"start" : "134"
       "type" : "pfama",
"start" : "134"
```

```
"markups" : [
          "lineColour": "#CCC",
"colour": "#CCC",
"display": true,
"end": "196",
"v_align": "top",
"metadata": {
   "database": "pfam",
   "type": "Disulphide, 155-196",
   "end": "196",
   "start": "155"
},
           "type" : "Disulphide",
"start" : "155"
          "lineColour": "#CCC",
"colour": "#CCC",
"display": true,
"end": "228",
"v_align": "top",
"metadata": {
   "database": "pfam",
   "type": "Disulphide, 189-228",
   "end": "228",
   "start": "189"
           "type" : "Disulphide",
"start" : "189"
           "lineColour" : "#CCC",
"colour" : "#CCC",
"display" : true,
"end" : "333",
             "v_align" : "top",
"metadata" : {
   "database" : "pfam",
                  "type": "Disulphide, 286-333",
"end": "333",
"start": "286"
           "type" : "Disulphide",
"start" : "286"
           "lineColour": "#000",
"colour": "#F36",
"display": true,
"residue": "C",
"headStyle": "diamond",
"v_align": "bottom",
"type": "Active site",
"metadata": {
            "metadata" : {
  "database" : "pfam",
  "description" : "Active site, C158",
  "start" : "158"
             "start" : "158"
          "lineColour": "#900",
"colour": "#90C",
"display": true,
"residue": "H",
"headStyle": "diamond",
"v_align": "bottom",
"type": "Pfam predicted active site, H292",
"metadata": {
   "database": "pfam",
   "description": "Pfam predicted active site, H292",
   "start": "292"
},
           },
             "start" : "292"
           "lineColour": "#000",
"colour": "#F6F",
"display": true,
"residue": "N",
"headStyle": "diamond",
"v_align": "bottom",
"type": "Pfam predicted active site, N308",
"metadata": {
   "database": "pfam",
   "description": "Pfam predicted active site, N308",
"start": "308"
},
               "start" : "308"
```

### back to top

### Disulphide bridges

Disulphide bridges play a fundamental role in the folding and stability of some proteins. They are formed by covalent bonding between the thiol groups from two cysteine residues. The disulphide bridge annotations used in Pfam come from <a href="UniProt">UniProt</a> and are represented by a solid bridge-shaped line. When mutliple disulphide bonds occur, the heights of the bridges are adjusted to avoid overlaps between them. Inter-protein disulphides are represented by single vertical lines. As always, moving the mouse over the "bridge graphic" shows the details of the bond in a tooltip.

#### back to top

#### Active site residues

Within an enyzme, a small number of residues are directly involved in catalysis of a reaction. These are termed active site residues. Within Pfam there are three categories of active site: those that are experimentally determined, those that are predicted by UniProt and those predicted by Pfam. All three types are represented by a "lollipop" with a diamond head. The head is coloured red, pink and purple for each of the three types respectively.

Pfam-predicted active sites are determined by using the experimental data and transferring these annotations through a Pfam alignment.

#### back to top

### "Lollipops"

A wide range of different lollipop styles can be create by combining different line and head colours with different drawing styles. The lollipop head can be drawn as a square, circle or diamond, as a simple coloured bar, or as an arrow (pointing away from the sequence) or a "pointer" (an arrow pointing towards the sequence).

```
"display" : true,
    "v_align" : "top",
    "headStyle" : "arrow",
    "type" : "Green arrow, above sequence",
    "start" : "100"
},
{
    "lineColour" : "#666",
    "colour" : "#08F",
    "display" : true,
    "v_align" : "bottom",
    "headStyle" : "pointer",
    "type" : "Blue pointer, below sequence",
    "start" : "120"
},
{
    "lineColour" : "#666",
    "colour" : "#F80",
    "display" : true,
    "v_align" : "top",
    "headStyle" : "line",
    "v_align" : "top",
    "headStyle" : "line",
    "type" : "Orange line, above sequence",
    "start" : "140"
}
}
```

### back to top

# **Tooltips**

If appropriate metadata are present in the sequence description, the domain graphics library can also add tooltips to the image. The example below is a "live" domain graphic and its description includes the necessary metadata for generating tooltips; move your mouse over the various domains and sequence features to see them.

```
Peptidase S8
                                                                                                                         PA
Hide JSON
        "length" : "950",
"regions" : [
                 {
                         "modelStart" : "5",
"modelEnd" : "292",
"colour" : "#2dcf00",
"endStyle" : "jagged",
"startStyle" : "jagged",
"display" - t "jagged",
                         "display": true,
"end": "361",
"aliEnd": "361",
"href": "/family/PF00082",
"text": "Peptidase_58",
"modelloogth": "3897"
                         "text": "Peptidase_S8",
"modelLength": "307",
"metadata": {
   "scoreName": "e-value",
   "score": "1.3e-38",
   "description": "Subtilase family",
   "accession": "PF00082",
                                   accession : "Pr00082",
"end" : "587",
"database" : "pfam",
"aliEnd" : "573",
"identifier" : "Peptidase_S8",
"type" : "Peptidase_S8",
                                    "type" : "Domain",
"aliStart" : "163",
"start" : "159"
                         "type" : "pfama",
"aliStart" : "163",
"start" : "159"
                         "modelStart" : "5",
"modelEnd" : "292",
"colour" : "#2dcf00",
"endStyle" : "jagged",
"startStyle" : "jagged",
"display" : true,
"end" : "587",
"aliEnd" : "573",
"href" : "/family/PF00082",
"text" : "Peptidase_S8",
"modellength" : "307"
                         "text" : "Peptidase_S8",
"modelLength" : "307",
"metadata" : {
    "scoreName" : "e-value",
    "score" : "1.3e-38",
    "description" : "Subtilase family",
    "accession" : "PF00082",
    "end" : "587",
    "database" : "pfam",
```

```
"aliEnd" : "573",
"identifier" : "Peptidase_S8",
                         "type" : "Domain",
"aliStart" : "163",
"start" : "159"
                "type" : "pfama",
"aliStart" : "470",
"start" : "470"
             "modelStart" : "12",
"modelEnd" : "100",
"colour" : "#ff5353",
"endStyle" : "curved",
"startStyle" : "jagged",
"display" : true,
"end" : "469",
"aliEnd" : "469",
"href" : "/family/PF02225",
"text" : "PA",
"modelLength" : "100",
"metadata" : {
    "scoreName" : "e-value",
    "score" : "7.le-09",
    "description" : "PA domain",
    "accession" : "PF02225",
    "end" : "469",
    "database" : "pfam",
    "aliEnd" : "469",
    "identifier" : "PA",
    "type" : "Family",
    "aliStart" : "385",
    "start" : "362"
},
"type" : "pfama"
                "type" : "pfama",
"aliStart" : "385",
"start" : "362"
              "modelStart" : "1",
"modelEnd" : "112",
"colour" : "#5b5bff",
"endStyle" : "curved",
"startStyle" : "curved",
"discalage" : true
                "display": true,
"end": "726",
"aliEnd": "726",
"href": "/family/PF06280",
"text": "DUF1034",
                "text": "DUF1054",
"modelLength": "112",
"metadata": {
    "scoreName": "e-value",
    "score": "1.1e-13",
    "description": "Fn3-like domain (DUF1034)",
    "accession": "PF06280",
"accession": "PF06280",
                        "end": "726",
"database": "pfam",
"aliEnd": "726",
"identifier": "DUF1034",
                         "type" : "Domain",
"aliStart" : "613",
"start" : "613"
                "type" : "pfama",
"aliStart" : "613",
"start" : "613"
      }
],
"markups" : [
               "lineColour": "#ff0000",
"colour": "#000000",
"display": true,
"end": "470",
"v_align": "top",
"metadata": {
   "database": "pfam",
   "type": "Link between discontinous regions",
   "end": "470",
   "start": "361"
                          "start" : "361'
                "type" : "Nested",
"start" : "361"
               "lineColour" : "#333333",
"colour" : "#e469fe",
"display" : true,
"residue" : "S",
"headStyle" : "diamond",
"v_align" : "top",
"type" : "Pfam predicted active site",
```

```
"metadata" : {
   "database" : "pfam",
   "description" : "S Pfam predicted active site",
   "start" : "538"
            "start" : "538"
          "lineColour" : "#333333",
"colour" : "#e469fe",
"display" : true,
"residue" : "D",
"headStyle" : "diamond",
"v_align" : "top",
"type" : "Pfam predicted active site",
"metadata" : {
   "database" : "pfam",
   "description" : "D Pfam predicted active site",
"start" : "185"
},
            "start" : "185"
     },
          "lineColour": "#333333",
"colour": "#e469fe",
"display": true,
"residue": "H",
"headStyle": "diamond",
"v_align": "top",
"type": "Pfam predicted active site",
"metadata": {
            "metadata" : {
   "database" : "pfam",
   "description" : "H Pfam predicted active site",
   "start" : "235"
          },
"start" : "235"
    }
],
 "metadata" : {
   "database" : "uniprot",
   "identifier" : "0560V8_CRYNE",
   "organism" : "Cryptococcus neoformans (Filobasidiella neoformans)",
   "description" : "Putative uncharacterized protein",
   "taysid" : "E007"
     "taxid" : "5207",
     "accession" : "Q560V8"
 "motifs" : [
  "colour" : "#ffa500",

"metadata" : {
    "database" : "Phobius",
    "type" : "sig_p",
    "end" : "23",
    "start" : "1"
           "type" : "sig_p",
"display" : true,
           "end" : 23,
"start" : 1
           "colour" : "#00ffff",
"metadata" : {
    "database" : "seg",
    "type" : "low_complexity",
    "score" : "2.5100",
    "end" : "21",
    "start" : "3"
          f,
"type" : "low_complexity",
"display" : false,
"end" : 21,
"start" : 3
          "colour": "#86bcff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "1.4900",
    "end": "156",
    "start": "134"
}
           },
"type" : "low_complexity",
           "display" : true,
"end" : "156",
"start" : "134"
          "colour" : "#00ffff",
"metadata" : {
   "database" : "seg",
   "type" : "low_complexity",
```

```
"score" : "2.0200",
"end" : "187",
"start" : "173"
                       "type" : "low_complexity",
                     "type": "low_comp
"display": false,
"end": "187",
"start": "173"
                    "colour": "#00ffff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "2.0800",
    "end": "218",
    "start": "207"
                     },
"type" : "low_complexity",
"display" : false,
"end" : "218",
"start" : "207"
                    "colour": "#00ffff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "2.1300",
    "end": "231",
    "start": "220"
                    },
"type" : "low_complexity",
"display" : false,
"end" : "231",
"start" : "220"
              {
                    "colour": "#00ffff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "2.0000",
    "end": "554",
    "start": "538"
                     "type" : "low_complexity",
"display" : false,
"end" : "554",
"start" : "538"
              },
                    "colour": "#86bcff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "1.9100",
    "end": "590",
    "start": "578"
                     },
"type" : "low_complexity",
"display" : true,
"end" : "590",
"start" : 588
                     "colour": "#00ffff",
"metadata": {
    "database": "seg",
    "type": "low_complexity",
    "score": "1.7600",
    "end": "831",
    "start": "822"
                     },
"type" : "low_complexity",
"display" : false,
"end" : "831",
"start" : "822"
      ]
}
back to top
```



Pfam is an Elixir service **Read more** 

Comments or questions on the site? Send a mail to  ${\bf pfam-help@ebi.ac.uk.}$ 

European Molecular Biology Laboratory