

Protein Analysis

SMART

(<http://smart.embl-heidelberg.de/>)

- `The Simple Modular Architecture Research Tool (SMART) is a database of protein families implicated in cellular signaling, extracellular domains, and chromatin function.`

2- SMART

(<http://smart.embl-heidelberg.de/>)

- `The Simple Modular Architecture Research Tool (SMART) is a database of protein families implicated in cellular signaling, extracellular domains, and chromatin function.`
- SMART allows the **identification** and **annotation** of genetically mobile domains and the analysis of domain architectures.

1-Finding the name and the domains of the protein from an unknown aa sequence

Unknown amino acid sequence
(Download from opencourses)

Opencourses or MS Teams(Files) file name:

In class Practicals – Unknown Sequences



BIOL312_UnknownAASeq_SMART.docx

```
MNRRDFIKNTAIASAASVAGLSVPSSMLGAQEEDWKWDKAVCRFCGTGCGIMIARKDGKIVATKGDPAAP
VNRGLNCIKGYFNAKIMYGEDRLVMPLLRMNEKGEFDKKGKFKQQVSWQRAFDEMEKQFKKAYNELGVTGI
GIFGSGQYTIQEGYAALKLAKAGFRTNNIDPNARHCMASAVVGFMTQTFGVDEPSGICYDDIELDTIITWG
ANMAEMHPILWSRVSDRKLSDKVKVNLSTFSNRTSNIADIEIIFKPNTDLAIWNYIAREIVYNHPEA
MDMKFIKDHCVFATGYADIGYGMRRNPNHPKFKSEKDTVEKENVITLDDEEATSLSYLGVKAGDKFEMK
HQQVADKNWEISFDEFKGLAPYTTLEYTARVAKGDDNESLEDFKKKLQELANLYIEKNRKKVVSFWTMGFN
QHTRGSWVNEQAYMVHFLGKQAKPGSGAFSLTGQPSACGTAREVGTFSHRLPADMVVANPKHREISEKI
WKVPAKTINPKPGSPYLNIMRDLEDGKIKFAWVQVNNPWQNTANANHWIAAAREMDNFIVVSDCYPGISA
KVADLILPSAMIYEKKGAYGNAERRTQHWKQQVLPVGAAMSDTWQILEFAKRFLKEVWKEQKVDNKLTL
PSVLEEAKAMGYSEDDTLFDVLFANKEAKSFNPNDIAIKGFDNTDVKGDERKIQGSDGKEFTGYGFFVQK
YLWEEYRKFGGLGHGHDLADFDTYHKVRGLRWPVNGKETQWRFNTKFDYYAKKAAPNSDFAFYGDFNKML
TNGDLIAPKDEKEHSIKNKAKIFFRPFMKAPERPSKEYPFWLATGRVLEHWHSGTMTMRVPELYRAVPEA
LCYMSEKDGEKLGDLNQGDLVWVESRRGKVKARVDMRGRNKPPVGLVYVPWFDENVYINKVTLDATCPLSK
QTDFKKCAVKIYKA
```

Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART analysis service.

Sequence ID or ACC

Examples: #1, #2



Protein sequence

```
MNRRDFIKNTAIAASAASVAGLSVPSMGLGAQEEDWIKWDAVC
RFCGTGCGIIMARKDGKIVATKGDPAAP
VNRGLNCIKGYFNAKIMYGEDRLVMPLLRMNEKGEFDKKGKF
QQVSMQRAFDEMEKQFKKAYNELGVTGI
GIFGSGQYTIQEGYAALKLAKAGFRTNINIDPNARHCMAAVV
GFMQTFGVDEPSGCYDDIELTDITITWG
ANMAEMHPILWSRVSDRKLNLKDKVKNLSTFSNRTSNIAD
IEIIFKPNTDLAIWNYIAREIVYNHPEA
```

Examples: #1, #2



1. Paste unknown aa sequence

Sequence SMART

Reset

HMMER searches of the SMART database occur by default. You may also find:

- [Outlier homologues](#) and homologues of known structure
- [PFAM domains](#)
- [signal peptides](#)
- [internal repeats](#)

2. Select these features

Domains detected by SMART

Search domain and protein annotation

Search Examples: #1, #2



- [Browse](#) the database of all available domains in the SMART database
- [Download](#) domain descriptions in tab delimited plain text

Architecture analysis

You can search for proteins with combinations of [specific domains](#) in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains.

Domain selection

Examples: #1, #2



GO terms query

Examples: #1, #2



Taxonomic selection

If you wish to restrict your domain architecture query to a particular species or taxonomic class, start typing its name in the box, and select a match from the popup list.

Architecture query

Reset

You can try an [Advanced Query](#) if you're familiar with SQL.

metaSMART

metaSMART is a novel integral part of SMART, dedicated to the exploration of protein domains and domain architectures in various metagenomics datasets. At the moment, the following datasets are publicly available through metaSMART:

- Sargasso Sea ([PubMed](#))
- Minnesota farm soil ([PubMed](#))
- Acid mine drainage biofilm ([PubMed](#))
- "Whale fall" carcasses ([PubMed](#))

[Access metaSMART](#)

Domains within *Campylobacter jejuni subsp. jejuni* NCTC 11168 = ATCC 700819 protein **NAPA_CAMJE (Q9PPD9)**

Name of the unknown aa sequence

Periplasmic nitrate reductase

+ = - SAVE



Domains of the protein

Information Architecture Interactions Pathways Orthology

Length	924 aa
Source database	UniProt
Identifiers	Q0PAB0, NAPA_CAMJR, Q5HV12, A0A0W8KX23_CAMJU, A0A0W8KX23, A0A2K1WME0_CAMJU, A0A2K1WME0, A0A218KX30_CAMJU, A0A218KX30, A0A0B6U490_CAMJU, A0A0B6U490, A5KI63_CAMJU, A5KI63, A0A0E1ZM91_CAMJU, A0A0E1ZM91, A0A059I056_CAMJU, A0A059I056, A0A059HWQ1_CAMJU, A0A059HWQ1, NAPA_CAMJE, Q9PPD9

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
Molybdop_Fe4S4	35	89	3.36e-12
Pfam:Molybdopterin	92	612	4.3e-70
Pfam:Molydop_binding	810	918	4.2e-25

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

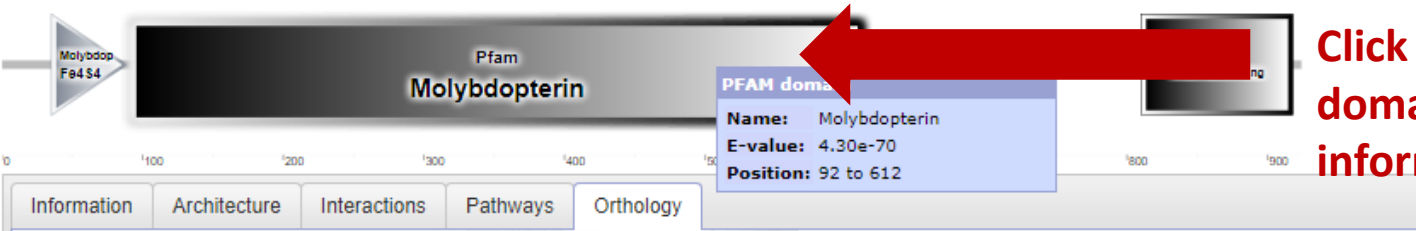
Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
Pfam:Molybdop...	35	89	3.9e-19	overlap

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Domains within *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 = ATCC 700819 protein [NAPA_C](#)

Periplasmic nitrate reductase



Click on the domain for detailed information

Selected feature details

This is a Pfam domain. Please see the [Molybdopterin](#) entry in Pfam for full annotation.

Position: 92 to 612
E-value: 4.3e-70 (HMMER3)

Accession: [PF00384](#)

Description:

Interpro abstract (IPR006656): This domain is found in a number of molybdopterin-containing oxidoreductases, tungsten formylmethanofuran dehydrogenase subunit d (FwdD) and molybdenum formylmethanofuran dehydrogenase subunit (FmdD); ...[\(full abstract\)](#)

GO process: [oxidation-reduction process \(GO:0055114\)](#)

GO function: [oxidoreductase activity \(GO:0016491\)](#)

Pfam domain sequence (521 aa):

```
RLVMPLLRMNEKGEFDKKGKQQVSWQRAFDEMEKQFKKAYNELGVTGIGIFGSGQYTIQ
EGYAALKLAKAGFRITNNIDPNARHCMASAVVGFMTQFGVDEPSGCYDDIELTDTIITWGA
NMAEMHPILWSRVSDRKLNSLDKVKVNLSTFSNRTSNIADIEIIFKPNLDLAIWNYIAR
EIVYNHPEAMDKFIKDHCVFATGYADIGYGMRRNPNHKPKFKESEKDTVEKENVITLDDE
EATLSYLGVKAGDKFEMKHQGVADKNWEISFDEFKKGLAPYTTLEYTARVAKGDDNESLE
DFKKLQELANLYIEKNRNVVSWTGMFNQHTRGSWVNEQAYMVHFLLGKQAKPGSGAFS
LTGQPSACGTAREVGTFSHRLPADMVVANPKHREISEKIWKVPAKTINPKPGSPYLNIMR
DLEDGKIKFAWVQVNNPWQNTANANHWIAAAREMDFIVVSDCYPGISAKVADLILPSAM
IYEKVGAYGNAERRTQHWKQVLPVGAAMSDTQIILEFAKR
```

tein are listed below.

during aerobic growth
n membrane and peri

during aerobic growth
n membrane and peri

during aerobic growth
n membrane and peri

during aerobic growth
n membrane and peri

ed cutoffs are not shown
brane > Coiled coil >

Featur

Name Start End E-value Name Start

SMART SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK keywords... Search SMART

Domains within *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 = ATCC 700819 protein **NAPA_CAMJE** (Q9PPD9)

Periplasmic nitrate reductase

+ = - SAVE

Molybdop Fe4S4

Pfam Molybdopterin

Pfam Molybdop_binding

0 100 200 300 400 500 600 700 800 900

Information Architecture **Interactions** Pathways Orthology

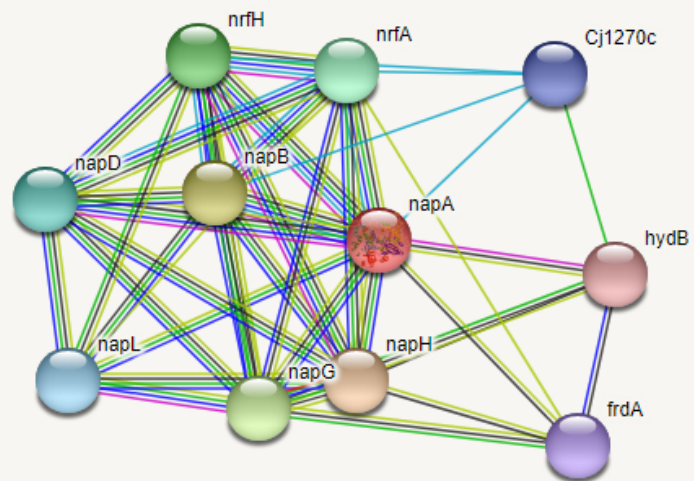
Predicted functional partners

The network on the left comes from STRING, a database of known and predicted protein interactions. Displayed here is the evidence view, where different line colors represent the types of evidence for the association.

Protein **NAPA_CAMJE** is shown as **napA** in the network.

Click the image to open the corresponding annotation page in STRING, where you can explore the network in detail.

- Click the image to open the corresponding annotation page in STRING, where you can explore the network in detail.





- Viewers >
- Legend v
- Settings >
- Analysis >
- Exports >
- Clusters >
- + More
- Less

Nodes:



Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color

-  *colored nodes: query proteins and first shell of interactors*
-  *white nodes: second shell of interactors*

Node Content



-  *empty nodes: proteins of unknown 3D structure*
-  *filled nodes: some 3D structure is known or predicted*

Edges:




Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.




Known Interactions

-  *from curated databases*
-  *experimentally determined*

Predicted Interactions

-  *gene neighborhood*
-  *gene fusions*
-  *gene co-occurrence*

Others

-  *textmining*
-  *co-expression*
-  *protein homology*

Your Input:

 napA

Periplasmic nitrate reductase; Catalytic subunit of the periplasmic nitrate reductase (NAP). Only expressed at high levels during aerobic growth. NapAB complex receives electrons from the membrane-anchored tetraheme protein NapC, thus allowing electron flow between membrane and periplasm. Essential function for nitrate assimilation and may have a role in anaerobic metabolism (924 aa)

neighborhood
 gene fusion
 gene co-occurrence
 co-expression
 textmining
 protein homology
 protein fusion

Domains within *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 = ATCC 700819 protein **NAPA_CAMJE** (Q9PPD9)

Periplasmic nitrate reductase



[Information](#) [Architecture](#) [Interaction](#) [Pathways](#) [Orthology](#)

Metabolic pathways

Protein **NAPA_CAMJE** is possibly involved in the following metabolic pathways:

[map00910](#): Nitrogen metabolism

These assignments are based on similarity to the following KEGG orthologous groups:

[K02567](#): NAPA (periplasmic nitrate reductase NapA [EC:1.7.99.4])



The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity**. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
Molybdop_Fe4S4	35	89	3.36e-12
Pfam:Molybdopterin	92	612	4.3e-70
Pfam:Molydop_binding	780	850	1.2e-25

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
Pfam:Molybdop...	35	89	3.9e-19	overlap

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

2- Comparing the domains of a specific protein in different species by using SMART

- KRAS human: P01116
- KRAS mouse: P32883
- KRAS wild turkey: P79800

Compare following for each entry;

- Name, number and length of the domain(s)
- Interactions
- Pathways
- PTMs

Domains within *Homo sapiens* protein **RASK_HUMAN** (P01116)

GTPase KRas



0 100

Information	Architecture	Interactions	Pathways	PTMs	Orthology
Length	189 aa				
Source database	UniProt				
Identifiers	RASK_HUMAN, P01116, ENSP00000256078.4, ENSP00000256078, A8K8Z5, B0LPF9, P01118, Q96D10, I1SRC5_HUMAN, I1SRC5, A0A221ZRZ1_HUMAN, A0A221ZRZ1, L7RSL8_HUMAN, L7RSL8, A0A2I3T6U4_PANTR, A0A2I3T6U4, F6SFG7_MACMU, F6SFG7, U3BT07_CALJA, U3BT07, G3QYD4_GORGO, G3QYD4, A0A2K5WC05_MACFA, A0A2K5WC05, G1PTH5_MYOLU, G1PTH5, A0A0D9R5Y6_CHLSB, A0A0D9R5Y6, A0A2K6LSR6_RHIBE, A0A2K6LSR6, A0A2K6EAW1_MACNE, A0A2K6EAW1, A0A2K5LFF0_CERAT, A0A2K5LFF0, A0A2K6QVV1_RHIRO, A0A2K6QVV1, A0A2K5DW76_AOTNA, A0A2K5DW76, A0A2I3NH01_PAPAN, A0A2I3NH01, A0A2K6SVB2_SAIIB, A0A2K6SVB2, A0A2K5HQC3_COLAP, A0A2K5HQC3, A0A2K5RJ06_CEBCA, A0A2K5RJ06, A0A2K5XYR5_MANLE, A0A2K5XYR5, A0A287B854_PIG, A0A287B854, G1QZK0_NOMLE, G1QZK0, G5BQA1_HETGA, G5BQA1, H0VHH8_CAVPO, H0VHH8				
Source gene	ENSG00000133703				
Alternative splicing	A0A024RAV5_HUMAN , G3V5T7_HUMAN , RASK_HUMAN , G3V4K2_HUMAN				

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART** > **PFAM** > **PROSPERO repeats** > **Signal peptide** > **Transmembrane** > **Coiled coil** > **Unstructured regions** > **Low complexity**. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
RAS	1	166	9.11e-123

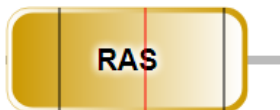
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
Pfam:Arf	1	163	6.8e-10	overlap
Pfam:Roc	5	120	1.2e-22	overlap
Pfam:MMR_HS...	5	117	0.0000041	overlap
Pfam:Ras	5	165	1.3e-59	overlap
low complexity	7	15	N/A	overlap

Domains within *Mus musculus* protein **RASK_MOUSE** (P32883)

GTPase KRas



0 100

Information **Architecture** Interactions Pathways PTMs Orthology

Length	189 aa
Source database	UniProt
Identifiers	RASK_MOUSE, P32883, ENSMUSP00000107339.1, ENSMUSP00000107339, P04200, P08643, RASK_RAT, P08644, P46203, P97914, Q0VDV7_MOUSE, Q0VDV7
Source gene	ENSMUSG00000030265
Alternative splicing	Q5J7N1_MOUSE , A0A0N4SVY1_MOUSE , RASK_MOUSE, E9Q8V2_MOUSE , ENSMUSP00000118251.1

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity**. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
RAS	1	166	3.7e-123

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
low complexity	7	15	N/A	overlap

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Domains within *Meleagris gallopavo* protein **RASK_MELGA** (P79800)

GTPase KRas



0 100

Information	Architecture	Interactions	Pathways	PTMs	Orthology
Length	188 aa				
Source database	UniProt				
Identifiers	RASK_MELGA, P79800, ENSMGAP00000017540.1, ENSMGAP00000017540, K7G9J5_PELSI, K7G9J5, A0A091V1Z7_NIPNI, A0A091V1Z7, A0A091HPB7_CALAN, A0A091HPB7, A0A093H8V6_STRCA, A0A093H8V6, A0A091G8P9_9AVES, A0A091G8P9, A0A093P8Q0_PYGAD, A0A093P8Q0, A0A210MKD7_COLL1, A0A210MKD7, A0A226MTE1_CALSU, A0A226MTE1, A0A091QP67_LEPDC, A0A091QP67, A0A099ZCM6_TINGU, A0A099ZCM6, A0A093BCR9_CHAPE, A0A093BCR9, A0A210TUI3_LIMLA, A0A210TUI3, A0A0A0A7C2_CHAVO, A0A0A0A7C2, A0A091JH63_EGRGA, A0A091JH63, A0A0Q3X8E7_AMAAE, A0A0Q3X8E7, U3JMM7_FICAL, U3JMM7, A0A093BWM4_TAUER, A0A093BWM4, H0ZPS0_TAEGU, H0ZPS0, A0A1D5NY37_CHICK, A0A1D5NY37, A0A091L889_CATAU, A0A091L889, A0A087R1L3_APTFO, A0A087R1L3, A0A1U7S121_ALLSI, A0A1U7S121, A0A091PTT1_HALAL, A0A091PTT1, A0A091SIA8_9AVES, A0A091SIA8, A0A091QXA5_9GRUI, A0A091QXA5, U3J8L1_ANAPL, U3J8L1, A0A087VBQ1_BALRE, A0A087VBQ1, A0A091FA76_CORBR, A0A091FA76, A0A093F8S8_TYAL, A0A093F8S8, A0A091XRZ7_OPIHO, A0A091XRZ7, A0A226P486_COLVI, A0A226P486, A0A218UH66_9PASE, A0A218UH66				
Source gene	ENSMGAG00000013673				
Alternative splicing	RASK_MELGA, ENSMGAP00000014451.2				

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity**. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
RAS	1	166	2.62e-123

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
Pfam:Arf	1	163	1e-10	overlap
Pfam:MMR_HS...	5	117	0.0000033	overlap
Pfam:Ras	5	165	1.5e-59	overlap
Pfam:Roc	5	120	1.1e-22	overlap

3-Finding domains with similar domain architecture

The image shows a screenshot of the SMART database website. The main navigation bar includes 'HOME', 'SETUP', 'FAQ', 'ABOUT', 'GLOSSARY', 'WHAT'S NEW', and 'FEEDBACK'. The 'SMART MODE:' dropdown is set to 'NORMAL GENOMIC'. The 'Architecture analysis' section is highlighted with a red border and contains the following content:

Architecture analysis

You can search for proteins with combinations of [specific domains](#) in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains.

Domain selection

Examples: #1, #2 [?](#)

GO terms query

Examples: #1, #2 [?](#)

Taxonomic selection

If you wish to restrict your domain architecture query to a particular species or taxonomic class, start typing its name in the box, and select a match from the popup list.

Architecture query **Reset**

You can try an [Advanced Query](#) if you're familiar with SQL.

metaSMART

metaSMART is a novel integral part of SMART, dedicated to the exploration of protein domains and domain architectures in various metagenomics datasets. At the moment, the following datasets are publicly available through metaSMART:

- Sargasso Sea ([PubMed](#))
- Minnesota farm soil ([PubMed](#))
- Acid mine drainage biofilm ([PubMed](#))
- "Whale fall" carcasses ([PubMed](#))

[Access metaSMART](#)

Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART analysis service.

Sequence ID or ACC

Examples: #1, #2 [?](#)

Protein sequence

Examples: #1, #2 [?](#)

Sequence SMART **Reset**

HMMER searches of the SMART database occur by default. You may also find:

- [Outlier homologues](#) and homologues of known structure
- [PFAM](#) domains
- [signal peptides](#)
- [internal repeats](#)

Domains detected by SMART

Search domain and protein annotation

Search Examples: #1, #2 [?](#)

- [Browse](#) the database of all available domains in the SMART database
- [Download](#) domain descriptions in tab delimited plain text

i- Find the domains of P00533 (EGFR Human)

SMART SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK Search SMART

Domains within *Homo sapiens* protein **EGFR_HUMAN** (P00533)

Epidermal growth factor receptor

+ = - Introns SAVE

Information Architecture Interactions Pathways PTMs Orthology

Length 1210 aa

Source database UniProt

Identifiers EGFR_HUMAN, P00533, ENSP00000275493.2, ENSP00000275493, O00688, O00732, P06268, Q14225, Q68GS5, Q92795, Q9BZS2, Q9GZX1, Q9H2C9, Q9H3C9, Q9UMD7, Q9UMD8, Q9UMG5, F1JTL6_HUMAN, F1JTL6, C9JYS6_HUMAN, C9JYS6, Q75MF2_HUMAN, Q75MF2, H9F140_MACMU, H9F140

Source gene ENSG00000146648

Alternative splicing ENSP00000395243.3, A0A1W2PRR9_HUMAN, Q504U8_HUMAN, P00533-4, P00533-3, P00533-2, EGFR_HUMAN, ENSP00000413354.1

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity. In either case, features not shown in the above diagram are marked as 'overlap' in the right side table below.

Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
low complexity	7	25	N/A
FU	228	270	0.0000536
FU	496	547	6.8e-7
FU	552	601	6.2e-10
transmembrane region	646	668	N/A
low complexity	675	692	N/A
TyrKc	712	968	1e-128

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
low complexity	651	666	N/A	overlap
transmembrane region	773	795	N/A	overlap

Click on a row to highlight the feature in the diagram above. Click the feature name for more information.

ii- Find the domains TyrKc and FU in any other species

The image shows the SMART database website interface. At the top left is the SMART logo. To its right is a 'SMART MODE:' selector with 'NORMAL' selected and 'GENOMIC' as an option. Further right is the acronym 'SMART' (Simple Modular Architecture Research Tool) and a search box with 'keywords...' and a 'Search SMART' button. Below the header is a navigation menu with links: HOME, SETUP, FAQ, ABOUT, GLOSSARY, WHAT'S NEW, FEEDBACK. The main content area is divided into two columns. The left column is titled 'Sequence analysis' and contains a form for 'Sequence ID or ACC' and 'Protein sequence'. The right column is titled 'Architecture analysis' and contains a 'Domain selection' form with a text input field containing 'TyrKc and FU', a 'GO terms query' form, and a 'Taxonomic selection' form. Red arrows point to the 'Domain selection' input field and the 'Architecture query' button. Red text next to the arrows says 'Write the desired domain names in query box'. At the bottom of the page are sections for 'Domains detected by SMART' and 'metaSMART'.

SMART

SMART MODE:
NORMAL
GENOMIC

Simple
Modular
Architecture
Research
Tool

keywords...
Search SMART

Letunic et al. (2014) *Nucleic Acids Res* doi: 10.1093/nar/gku949
Letunic et al. (2017) *Nucleic Acids Res* doi: 10.1093/nar/gkx922

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Sequence analysis

You may use either a [Uniprot/Ensembl](#) sequence identifier (ID) / accession number (ACC) or the protein sequence itself to perform the SMART analysis service.

Sequence ID or ACC

Examples: #1, #2 ?

Protein sequence

Examples: #1, #2 ?

HMMER searches of the SMART database occur by default. You may also find:

- [Outlier homologues](#) and homologues of known structure
- [PFAM domains](#)
- [signal peptides](#)
- [internal repeats](#)

Architecture analysis

You can search for proteins with combinations of [specific domains](#) in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains.

Domain selection

?

GO terms query

Examples: #1, #2 ?

Taxonomic selection

If you wish to restrict your domain architecture query to a particular species or taxonomic class, start typing its name in the box, and select a match from the popup list.

You can try [Advanced Query](#) if you're familiar with SQL.

Domains detected by SMART

metaSMART

Proteins in all species with TyrKc and FU

There are 2868 proteins matching your query. Select an option below and press the action button. To include only a subset of proteins, make a selection; otherwise, select all proteins in all its subnodes.

Action display SMART bubblograms of all or selected proteins

Selection 2868 proteins



Click to `Display proteins`

- Eukaryota (2868)**
 - Metazoa (2850)**
 - Annelida (4)
 - Arthropoda (488)
 - Brachiopoda (2)
 - Chordata (2102)
 - Cnidaria (4)
 - Echinodermata (8)
 - Hemichordata (2)
 - Mollusca (24)
 - Nematoda (146)
 - Orthonectida (1)
 - Platyhelminthes (60)
 - Porifera (3)
 - Priapulida (2)
 - Tardigrada (4)
 - undefined kingdom (18)
 - undefined phylum (18)

Protein [EGFR_HUMAN](#)

Description Epidermal growth factor receptor

Species *Homo sapiens*

1 100 200



Protein [EGFR_MACMU](#)

Description Epidermal growth factor receptor

Species *Macaca mulatta*

1 100 200



Protein [EGFR_MOUSE](#)

Description Epidermal growth factor receptor

Species *Mus musculus*

1 100 200



Protein [ERBB2_HUMAN](#)

Description Receptor tyrosine-protein kinase erbB-2

Species *Homo sapiens*

1 100 200



Protein Data Bank (PDB)

(www.rcsb.org/pdb/ or www.pdb.org)

- A broad range of primary structural data is collected, such as atomic coordinates, chemical structures of cofactors, and descriptions of the crystal structure. The PDB then validates structures by assessing the quality of the deposited models and by how well they match experimental data.

Search for `EGF`/PDB ID: 2KV4

www.rcsb.org/pdb/home/home.do#Category-welcome

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 117438 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands **Go**

Advanced Search | Browse by Annotations

PDB-101 WORLDWIDE PDB EMDatabank NUCLEIC ACID DATABASE StructuralBiology Knowledgebase

f t y a i

Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Wellcome Trust Image Awards

Illustrating Ebola

March Molecule of the Month

RAF Protein Kinases

Latest Entries

As of Tuesday Mar 29



New Features

February 2016 Release

Redesigned Search Results Page
New Organization. Improved
Layout. Clean. Usable. Faster.

News

Publications

Paper Published in
Journal of Chemical
Education

An overview of RCSB
PDB resources is part

Contact Us