

# Goals

- “Understand” annotation transfer by homology
- Know what protein family databases are and why they are useful

# Outline

- Homology
- Exercise 1 and 2: homology-based function annotation transfer
- Protein domains
- Protein families
- Protein family databases
- Team exercise: how to build a new (Pfam) protein family

# Homology

## Definition:

Two proteins are **homologous** if they share a common ancestor, i.e. they are evolutionary related

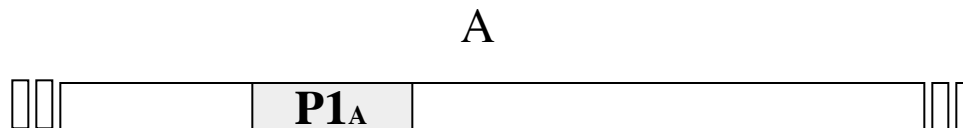
# Origins of homology in proteins

# Origin of homology in proteins

- Speciation (orthology)
- Gene duplication (paralogy)
- Horizontal gene transfer (xenology)
- Whole genome duplication (ohnology)
- Gametology, Synology

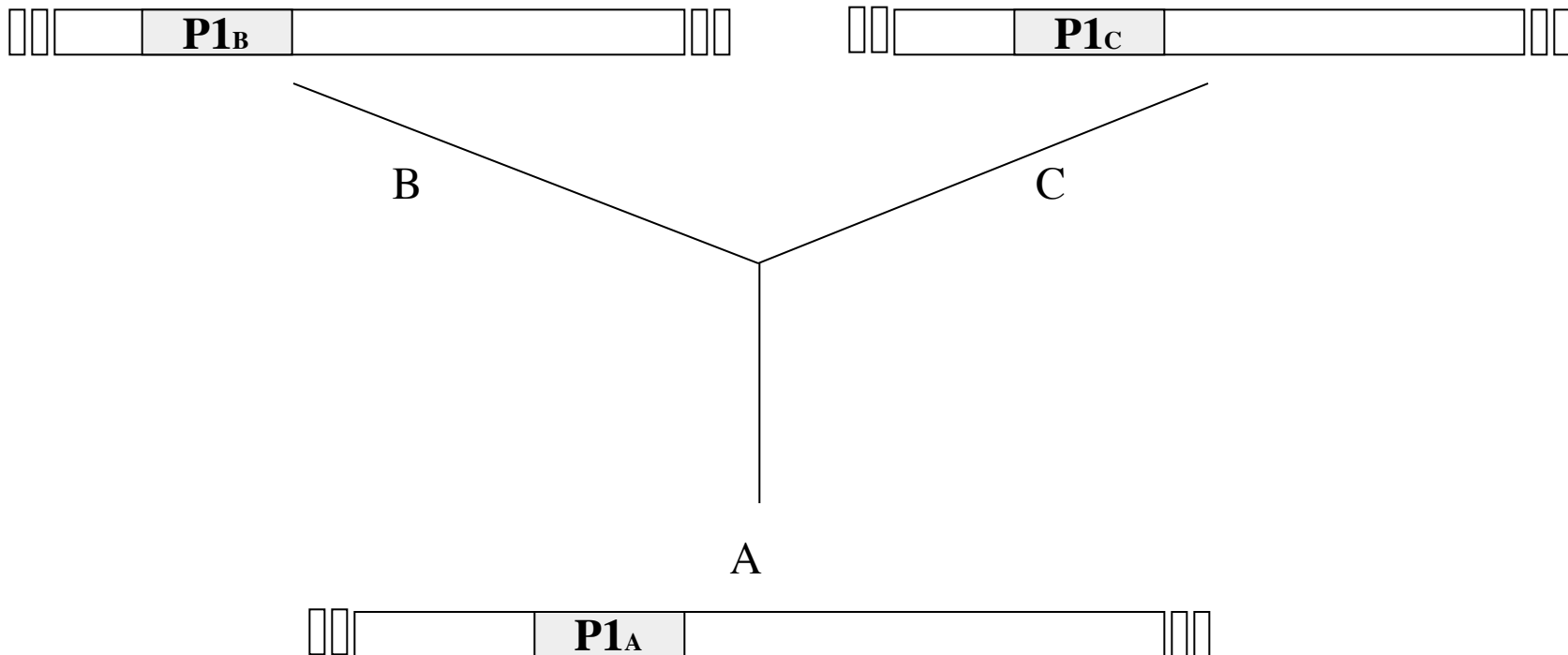
# Origin of homology in proteins

Marco Punta



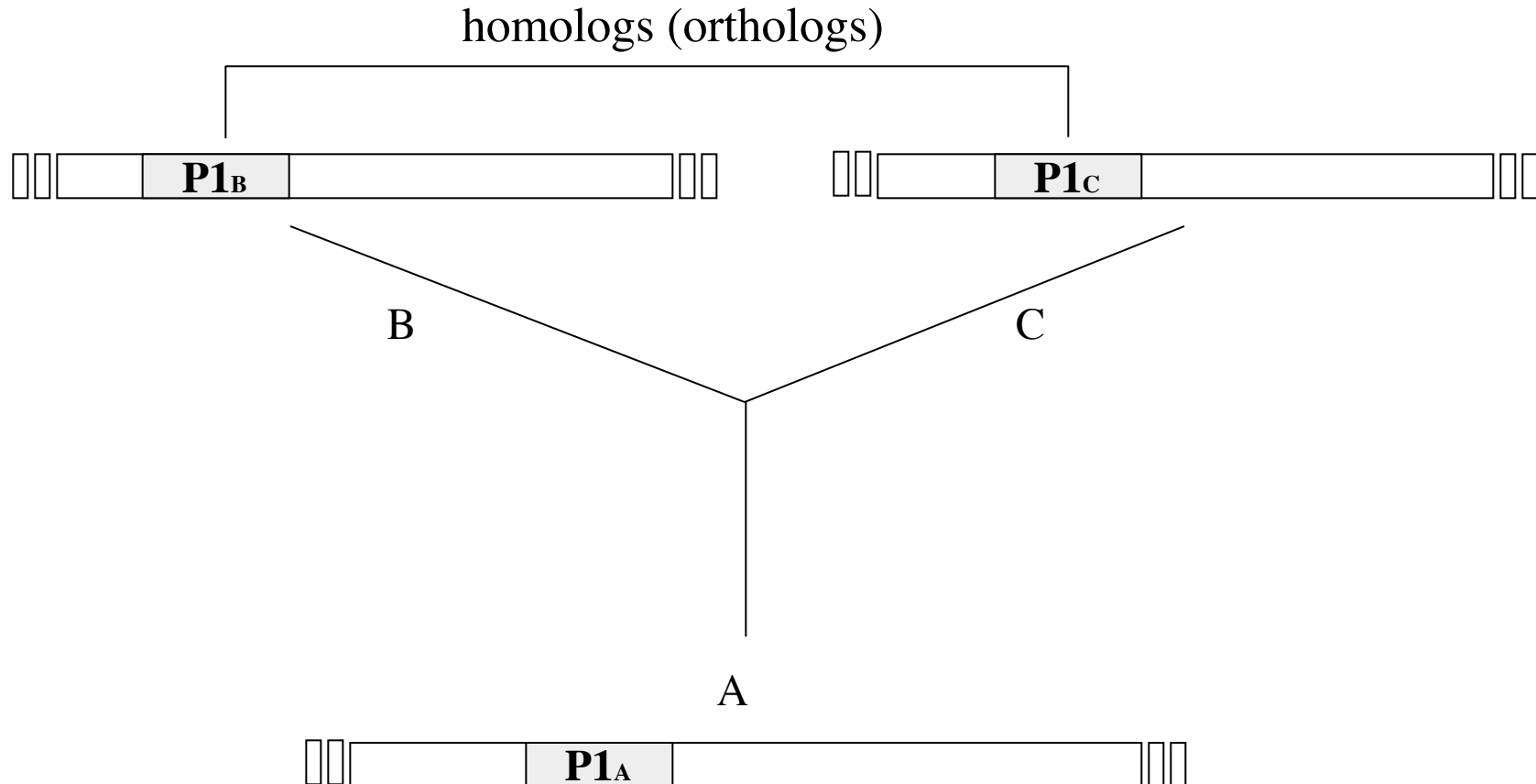
# Origin of homology in proteins

Marco Punta

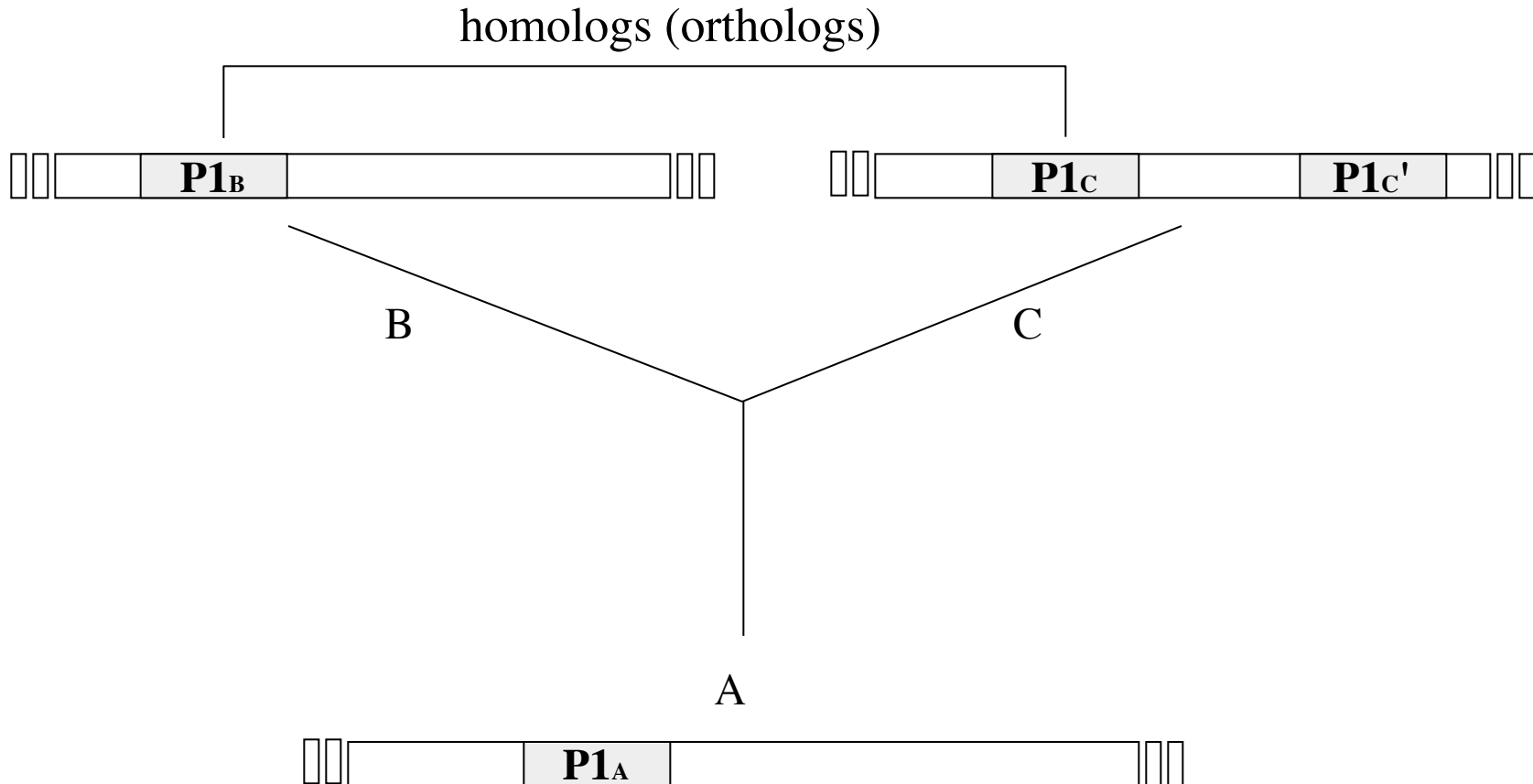




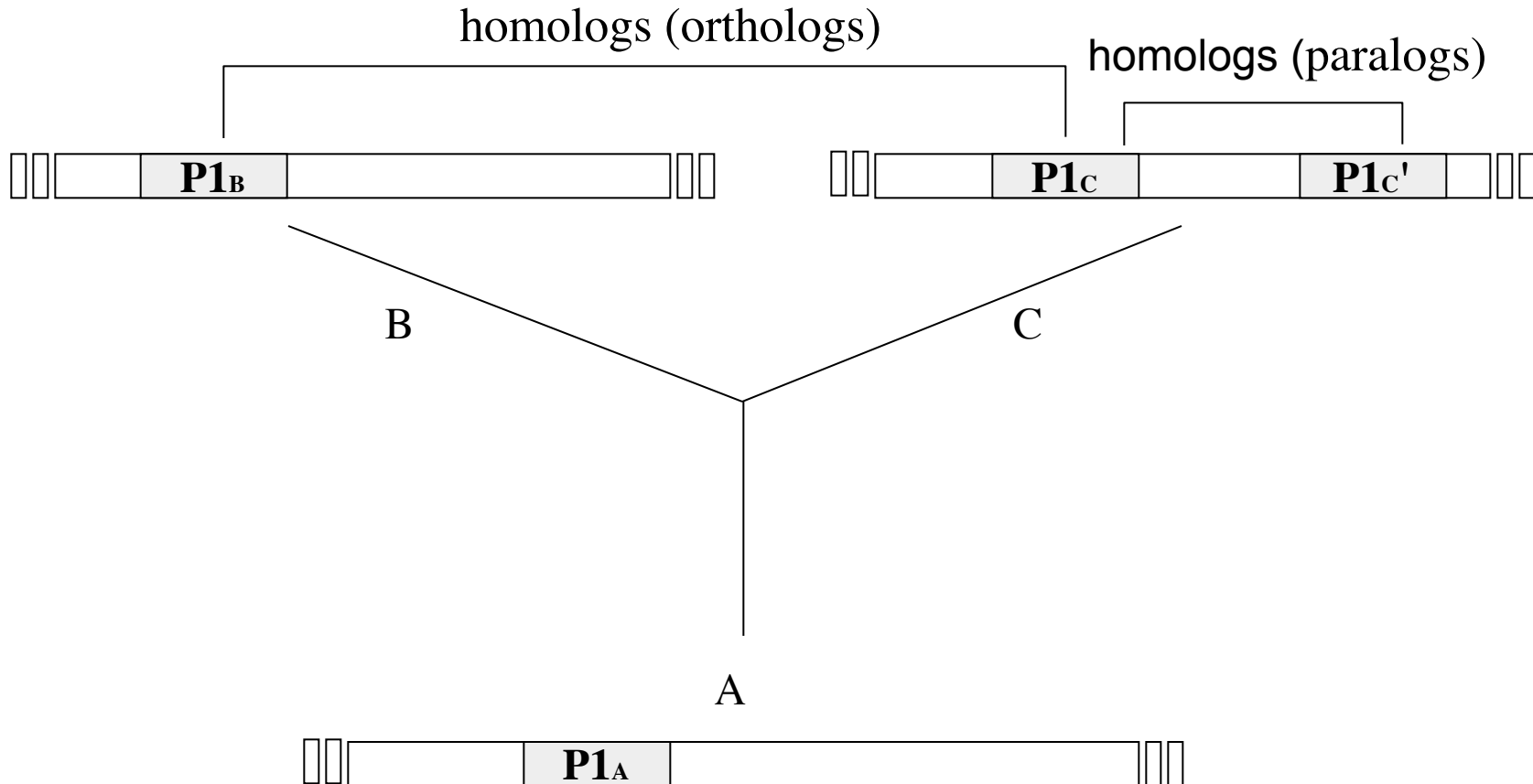
# Origin of homology in proteins



# Origin of homology in proteins

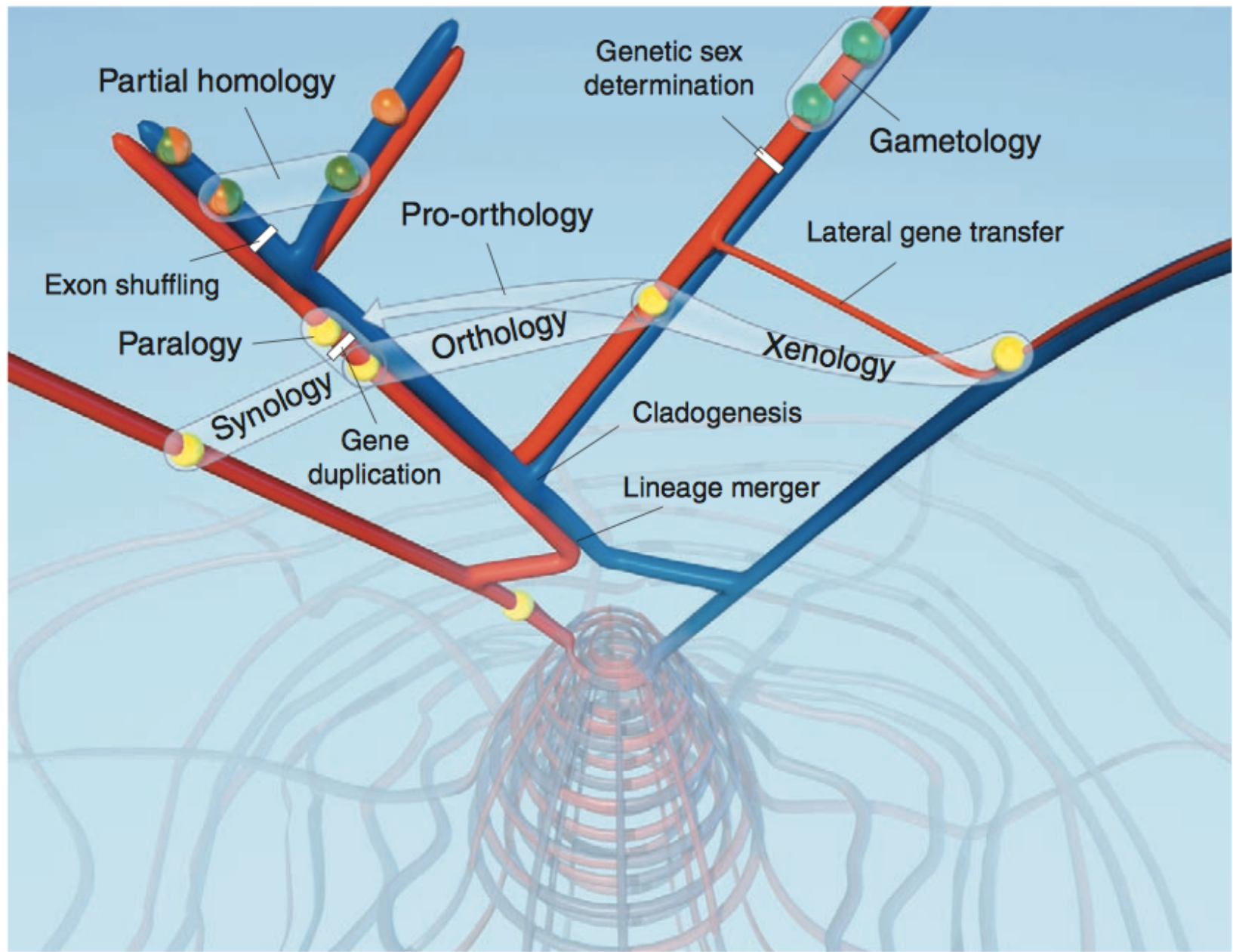


# Origin of homology in proteins



# Origin of homology in proteins

- Speciation (orthology)
- Gene duplication (paralogy)
- Horizontal gene transfer (xenology)
- Whole genome duplication (ohnology)
- Gametology, Synology



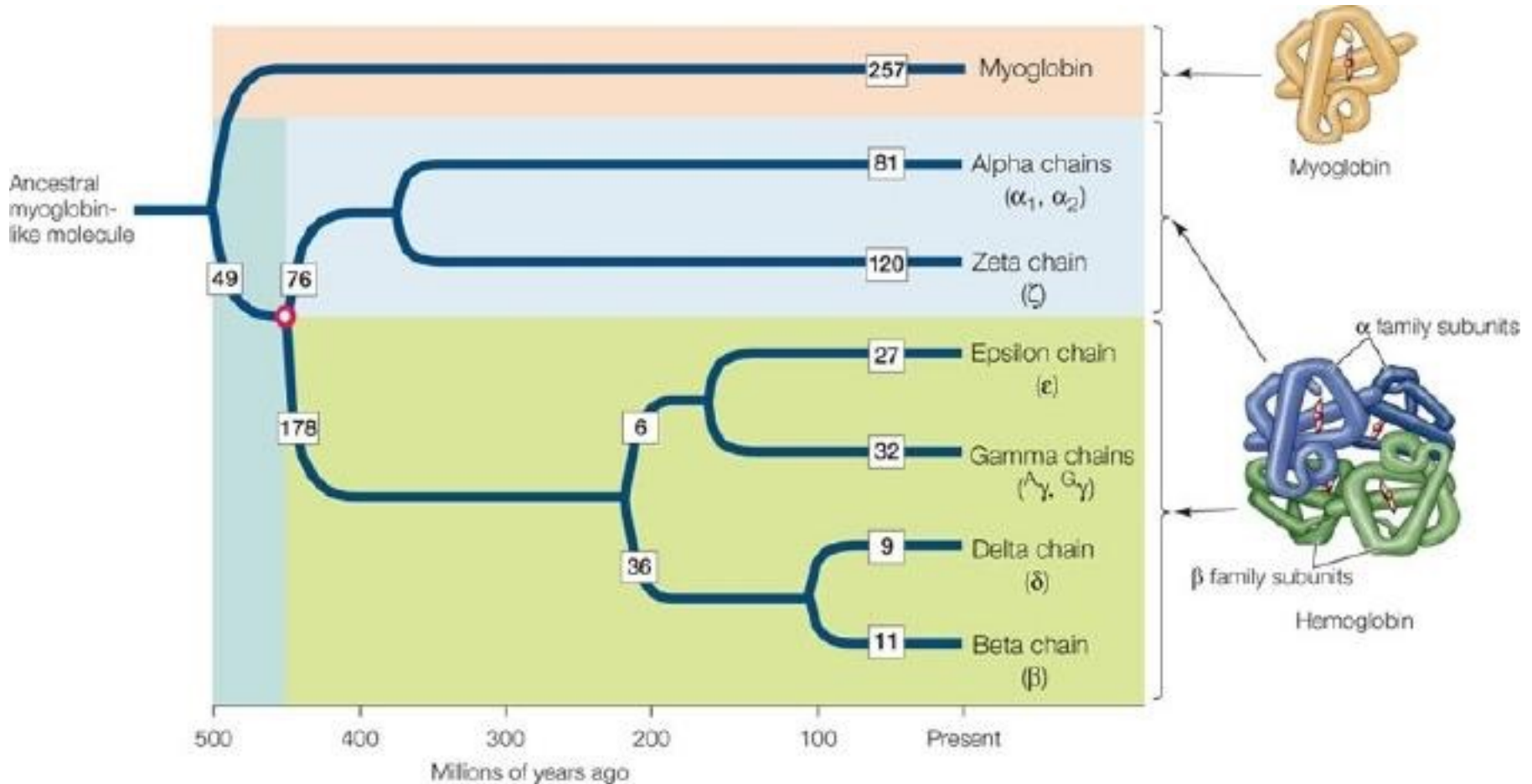
Mindell and Meyer Trends in Ecology and Evolution 2001

# Protein Families

## Definition:

We call 'family' a group of evolutionary related proteins and/or protein regions

# Globins in Human





# Homology: why bother?

# Mind the gap!

Marco Punta

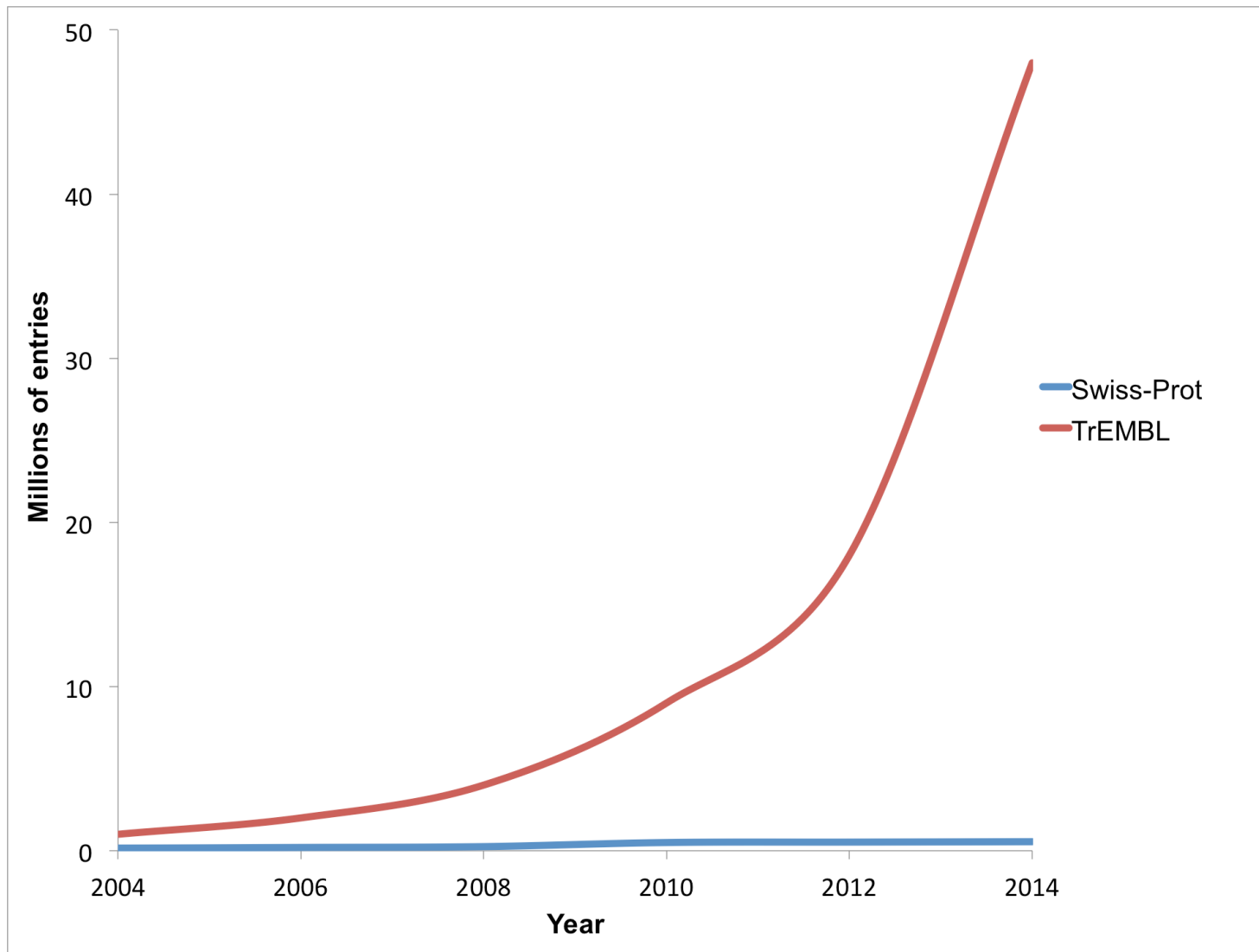


Figure courtesy of Alex Mitchell (EMBL-EBI)

EMBO Workshop, Norwich, 2015

# Mind the gap!

Marco Punta

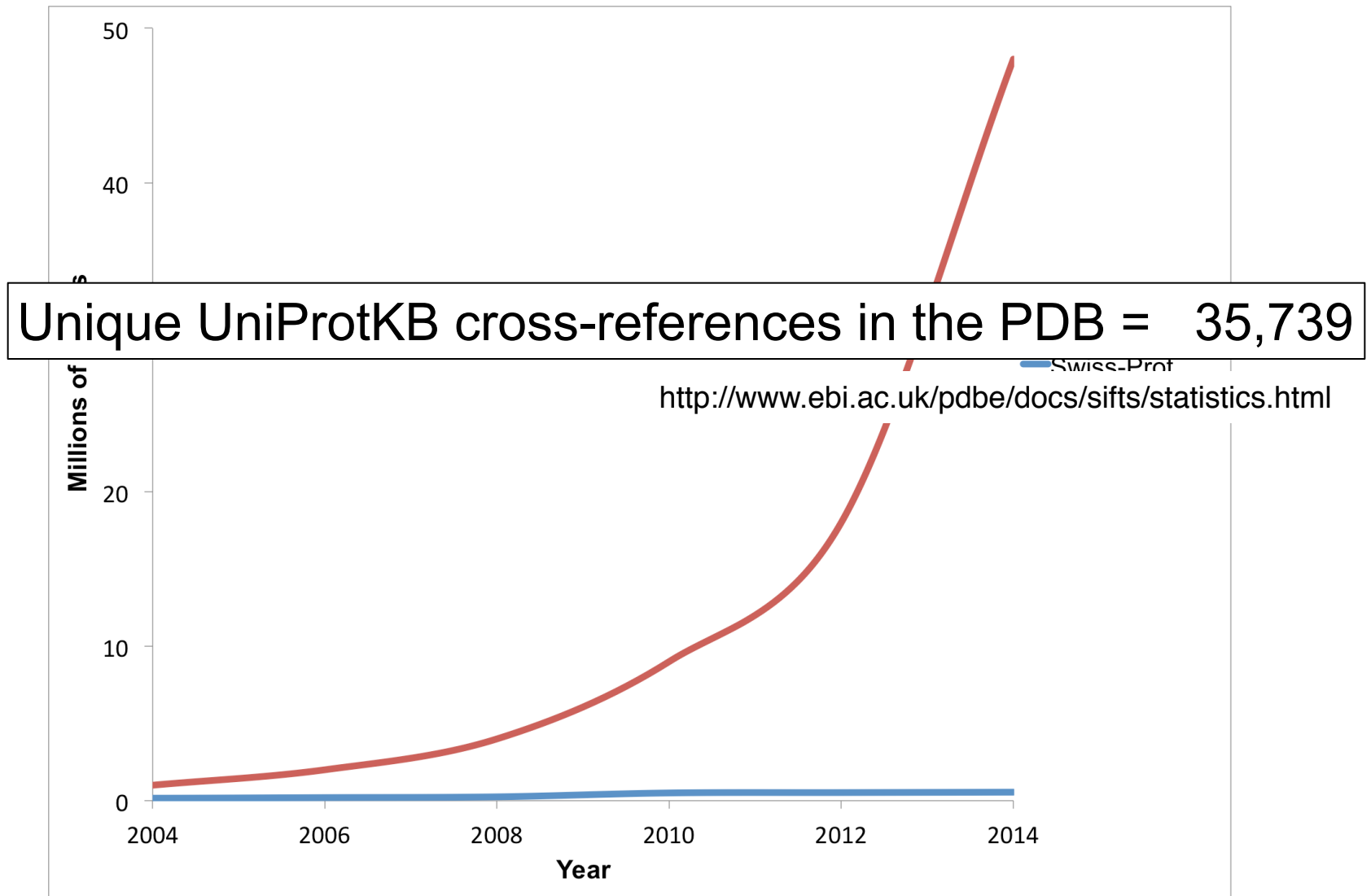


Figure courtesy of Alex Mitchell (EMBL-EBI)

EMBO Workshop, Norwich, 2015

# Homologous protein regions have a similar (core) structure!

Chotia and Lesk *EMBO J* (1986)

# Homology Modelling

Marco Punta

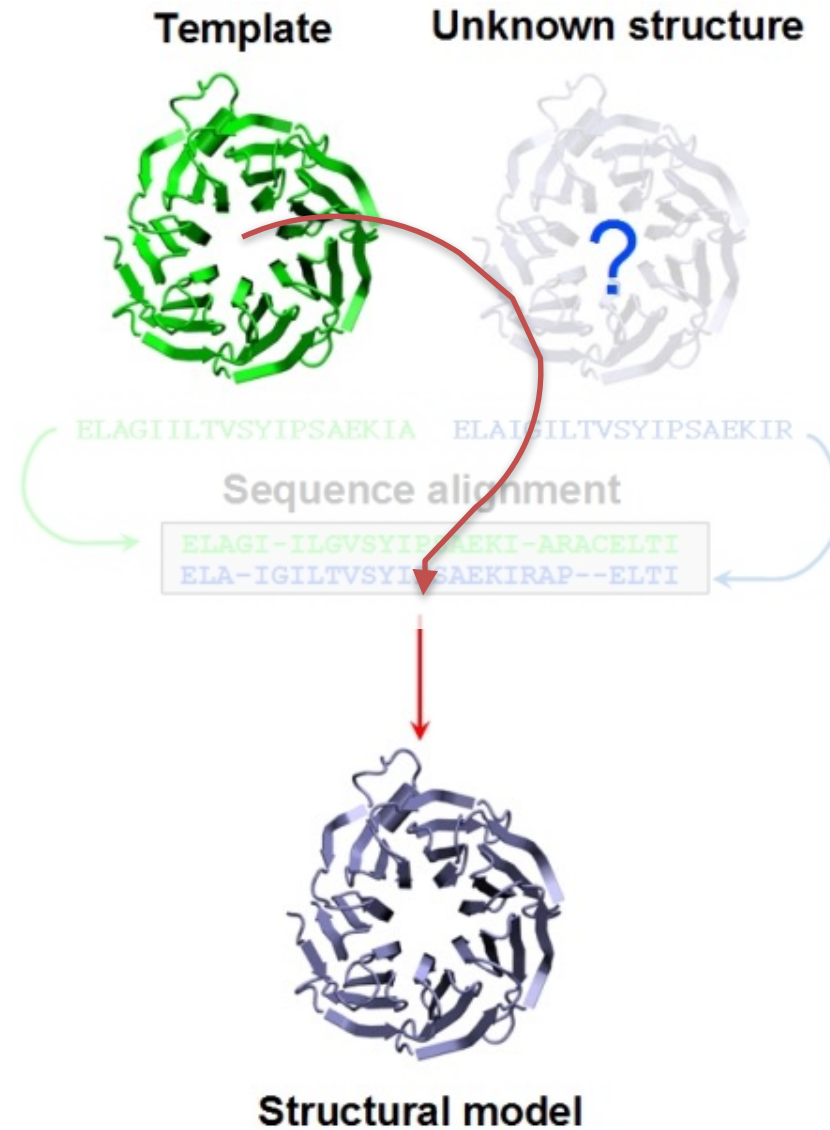
**Template**



**Unknown structure**



# Homology Modelling



# Homology: why bother?

Marco Punta

Do homologous protein regions  
perform a similar function?

## The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase

Heidi L. Schubert<sup>1</sup>, Keith S. Wilson<sup>1</sup>, Evelyne Raux<sup>2</sup>, Sarah C. Woodcock<sup>2</sup> and Martin J. Warren<sup>2</sup>

**Biosynthesis of the corrin ring of vitamin B<sub>12</sub> requires the action of six S-adenosyl-L-methionine (AdoMet) dependent transmethylases, closely related in sequence. The first X-ray structure of one of these, cobalt-precorrin-4 transmethylase, CbiF, from *Bacillus megaterium* has been determined to a resolution of 2.4 Å. CbiF contains two  $\alpha/\beta$  domains forming a trough in which S-adenosyl-L-homocysteine (AdoHcy) binds. The location of AdoHcy and a number of conserved residues, helps define the precorrin binding site. A second crystal form determined at 3.1 Å resolution highlights the flexibility of two loops around this site. CbiF employs a unique mode of AdoHcy binding and represents a new class of transmethylase.**



## The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase

Heidi L. Schubert<sup>1</sup>, Keith S. Wilson<sup>1</sup>, Evelyne Raux<sup>2</sup>, Sarah C. Woodcock<sup>2</sup> and Martin J. Warren<sup>2</sup>

**Biosynthesis of the corrin ring of vitamin B<sub>12</sub> requires the action of six S-adenosyl-L-methionine (AdoMet) dependent transmethylases, closely related in sequence. The first X-ray structure of one of these, cobalt-precorrin-4 transmethylase, CbiF, from *Bacillus megaterium* has been determined to a resolution of 2.4 Å. CbiF contains two  $\alpha/\beta$  domains forming a trough in which S-adenosyl-L-homocysteine (AdoHcy) binds. The location of AdoHcy and a number of conserved residues, helps define the precorrin binding site. A second crystal form determined at 3.1 Å resolution highlights the flexibility of two loops around this site. CbiF employs a unique mode of AdoHcy binding and represents a new class of transmethylase.**

## The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase

Heidi L. Schubert<sup>1</sup>, Keith S. Wilson<sup>1</sup>, Evelyne Raux<sup>2</sup>, Sarah C. Woodcock<sup>2</sup> and Martin J. Warren<sup>2</sup>

**Biosynthesis of the corrin ring of vitamin B<sub>12</sub> requires the action of six S-adenosyl-L-methionine (AdoMet) dependent transmethylases, closely related in sequence. The first X-ray structure of one of these, cobalt-precorrin-4 transmethylase, CbiF, from *Bacillus megaterium* has been determined to a resolution of 2.4 Å. CbiF contains two  $\alpha/\beta$  domains forming a trough in which S-adenosyl-L-homocysteine (AdoHcy) binds. The location of AdoHcy and a number of conserved residues, helps define the precorrin binding site. A second crystal form determined at 3.1 Å resolution highlights the flexibility of two loops around this site. CbiF employs a unique mode of AdoHcy binding and represents a new class of transmethylase.**

## The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase

Heidi L. Schubert<sup>1</sup>, Keith S. Wilson<sup>1</sup>, Evelyne Raux<sup>2</sup>, Sarah C. Woodcock<sup>2</sup> and Martin J. Warren<sup>2</sup>

**Biosynthesis of the corrin ring of vitamin B<sub>12</sub> requires the action of six S-adenosyl-L-methionine (AdoMet) dependent transmethylases, closely related in sequence. The first X-ray structure of one of these, cobalt-precorrin-4 transmethylase, CbiF, from *Bacillus megaterium* has been determined to a resolution of 2.4 Å. CbiF contains two  $\alpha/\beta$  domains forming a trough in which S-adenosyl-L-homocysteine (AdoHcy) binds. The location of AdoHcy and a number of conserved residues, helps define the precorrin binding site. A second crystal form determined at 3.1 Å resolution highlights the flexibility of two loops around this site. CbiF employs a unique mode of AdoHcy binding and represents a new class of transmethylase.**

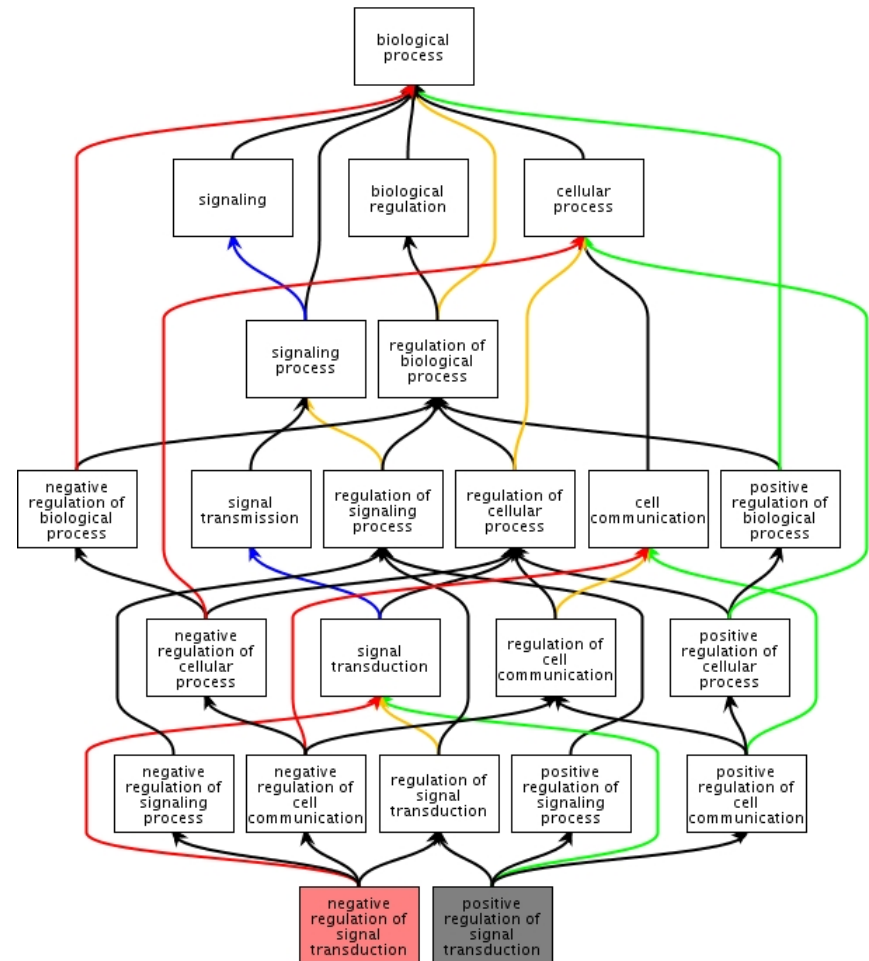
## The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase

Heidi L. Schubert<sup>1</sup>, Keith S. Wilson<sup>1</sup>, Evelyne Raux<sup>2</sup>, Sarah C. Woodcock<sup>2</sup> and Martin J. Warren<sup>2</sup>

**Biosynthesis of the corrin ring of vitamin B<sub>12</sub> requires the action of six S-adenosyl-L-methionine (AdoMet) dependent transmethylases, closely related in sequence. The first X-ray structure of one of these, cobalt-precorrin-4 transmethylase, CbiF, from *Bacillus megaterium* has been determined to a resolution of 2.4 Å. CbiF contains two  $\alpha/\beta$  domains forming a trough in which S-adenosyl-L-homocysteine (AdoHcy) binds. The location of AdoHcy and a number of conserved residues, helps define the precorrin binding site. A second crystal form determined at 3.1 Å resolution highlights the flexibility of two loops around this site. CbiF employs a unique mode of AdoHcy binding and represents a new class of transmethylase.**

# The Gene Ontology (GO)

- A way to capture biological knowledge in a written and computable form
- A set of concepts and their relationships to each other

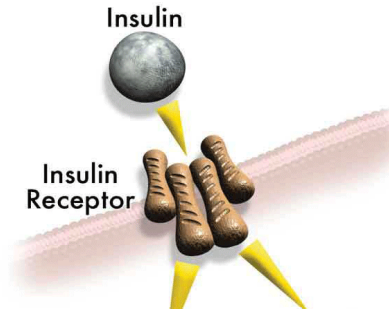


[www.ebi.ac.uk/QuickGO](http://www.ebi.ac.uk/QuickGO)

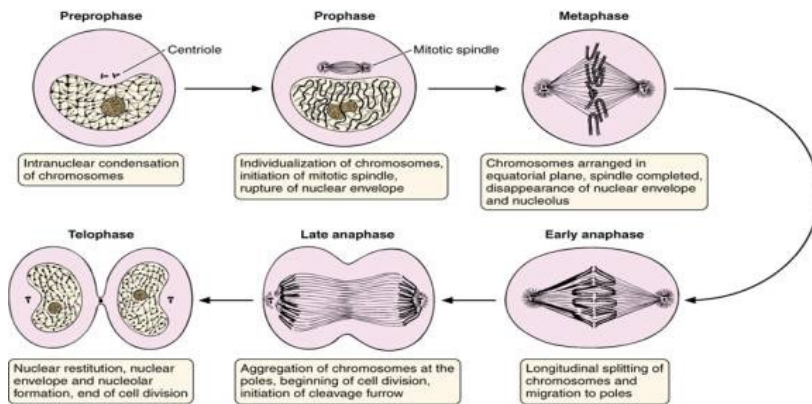
# GO: 3 ontologies in 1

## 1. Molecular Function

An elemental activity or task or job



- protein kinase activity
- insulin receptor activity



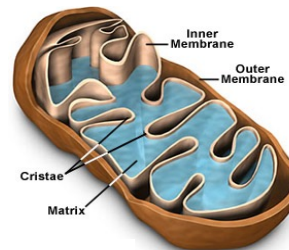
## 2. Biological Process

A commonly recognised series of events

- cell division

## 3. Cellular Component

Where a gene product is located



- mitochondrion
- mitochondrial matrix
- mitochondrial inner membrane

# CbiF GO annotation

Marco Punta

Database	Gene Product ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Form ID
<b>Process</b>													
UniProtKB	O87696	cbiF		<a href="#">GO:0006779</a>	porphyrin-containing compound biosynthetic process	P	IEA	InterPro2GO	InterPro:IPR003043	1404	20150919	InterPro	
UniProtKB	O87696	cbiF		<a href="#">GO:0008152</a>	metabolic process	P	IEA	InterPro2GO	InterPro:IPR000878 InterPro:IPR014776 InterPro:IPR014777	1404	20150919	InterPro	
UniProtKB	O87696	cbiF		<a href="#">GO:0009236</a>	cobalamin biosynthetic process	P	IEA	InterPro2GO	InterPro:IPR006362	1404	20150919	InterPro	
UniProtKB	O87696	cbiF		<a href="#">GO:0009236</a>	cobalamin biosynthetic process	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0169	1404	20150919	UniProt	
UniProtKB	O87696	cbiF		<a href="#">GO:0009236</a>	cobalamin biosynthetic process	P	IEA	UniPathway2GO	UniPathway:UPA00148	1404	20150912	UniProt	
UniProtKB	O87696	cbiF		<a href="#">GO:0032259</a>	methylation	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0489	1404	20150919	UniProt	
UniProtKB	O87696	cbiF		<a href="#">GO:0055114</a>	oxidation-reduction process	P	IEA	InterPro2GO	InterPro:IPR003043	1404	20150919	InterPro	
<b>Function</b>													
UniProtKB	O87696	cbiF		<a href="#">GO:0008168</a>	methyltransferase activity	F	IEA	InterPro2GO	InterPro:IPR000878 InterPro:IPR003043 InterPro:IPR014776 InterPro:IPR014777	1404	20150919	InterPro	
UniProtKB	O87696	cbiF		<a href="#">GO:0008168</a>	methyltransferase activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0489	1404	20150919	UniProt	
UniProtKB	O87696	cbiF		<a href="#">GO:0016740</a>	transferase activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0808	1404	20150919	UniProt	
UniProtKB	O87696	cbiF		<a href="#">GO:0043115</a>	precorrin-2 dehydrogenase activity	F	IEA	InterPro2GO	InterPro:IPR003043	1404	20150919	InterPro	
UniProtKB	O87696	cbiF		<a href="#">GO:0046026</a>	precorrin-4 C11-methyltransferase activity	F	IEA	InterPro2GO	InterPro:IPR006362	1404	20150919	InterPro	

Database	Gene Product ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Referen
UniProtKB	P02144		MB	<a href="#">GO:0001666</a>	response to hypoxia	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0006810</a>	transport	P	IEA	UniProt Prot ent
UniProtKB	P02144		MB	<a href="#">GO:0007507</a>	heart development	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0009725</a>	response to hormone	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0015671</a>	oxygen transport	P	IEA	InterPro
UniProtKB	P02144		MB	<a href="#">GO:0015671</a>	oxygen transport	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0015671</a>	oxygen transport	P	IEA	UniProt Prot ent
UniProtKB	P02144		MB	<a href="#">GO:0031444</a>	slow-twitch skeletal muscle fiber contraction	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0042542</a>	response to hydrogen peroxide	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0043353</a>	enucleate erythrocyte differentiation	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0050873</a>	brown fat cell differentiation	P	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0005344</a>	oxygen transporter activity	F	IEA	UniProt Prot ent
UniProtKB	P02144		MB	<a href="#">GO:0005506</a>	iron ion binding	F	IEA	InterPro
UniProtKB	P02144		MB	<a href="#">GO:0019825</a>	oxygen binding	F	IEA	InterPro
UniProtKB	P02144		MB	<a href="#">GO:0019825</a>	oxygen binding	F	IEA	Ensemb
UniProtKB	P02144		MB	<a href="#">GO:0020037</a>	heme binding	F	IEA	InterPro
UniProtKB	P02144		MB	<a href="#">GO:0046872</a>	metal ion binding	F	IEA	UniProt Prot ent
UniProtKB	P02144		MB	<a href="#">GO:0070062</a>	extracellular vesicular exosome	C	IDA	PMID:23



Database	Gene Product ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Form ID
<b>Process</b>													
JniProtKB	P02144	MB		<a href="#">GO:0001666</a>	response to hypoxia	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0006810</a>	transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0813	9606	20140913	UniProt	
JniProtKB	P02144	MB		<a href="#">GO:0007507</a>	heart development	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0009725</a>	response to hormone	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0015671</a>	oxygen transport	P	IEA	InterPro2GO	InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02144	MB		<a href="#">GO:0015671</a>	oxygen transport	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0015671</a>	oxygen transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02144	MB		<a href="#">GO:0031444</a>	slow-twitch skeletal muscle fiber contraction	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0042542</a>	response to hydrogen peroxide	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0043353</a>	enucleate erythrocyte differentiation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0050873</a>	brown fat cell differentiation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl	
<b>Function</b>													
JniProtKB	P02144	MB		<a href="#">GO:0005344</a>	oxygen transporter activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02144	MB		<a href="#">GO:0005506</a>	iron ion binding	F	IEA	InterPro2GO	InterPro:IPR000971 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02144	MB		<a href="#">GO:0019825</a>	oxygen binding	F	IEA	InterPro2GO	InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02144	MB		<a href="#">GO:0019825</a>	oxygen binding	F	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl	
JniProtKB	P02144	MB		<a href="#">GO:0020037</a>	heme binding	F	IEA	InterPro2GO	InterPro:IPR000971 InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02144	MB		<a href="#">GO:0046872</a>	metal ion binding	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0479	9606	20140913	UniProt	
<b>Component</b>													
JniProtKB	P02144	MB		<a href="#">GO:0070062</a>	extracellular vesicular exosome	C	IDA	PMID:23533145		9606	20140714	UniProt	

Database	Gene Product ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Form ID
<b>Process</b>													
JniProtKB	P02008	HBZ		<a href="#">GO:0000122</a>	negative regulation of transcription from RNA polymerase II promoter	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000020531	9606	20140913	Ensembl	
JniProtKB	P02008	HBZ		<a href="#">GO:0006810</a>	transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0813	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		<a href="#">GO:0015671</a>	oxygen transport	P	IEA	InterPro2GO	InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		<a href="#">GO:0015671</a>	oxygen transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		<a href="#">GO:0043249</a>	erythrocyte maturation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000020531	9606	20140913	Ensembl	
<b>Function</b>													
JniProtKB	P02008	HBZ		<a href="#">GO:0005344</a>	oxygen transporter activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		<a href="#">GO:0005344</a>	oxygen transporter activity	F	TAS	PMID:7555018		9606	20030904	PINC	
JniProtKB	P02008	HBZ		<a href="#">GO:0005506</a>	iron ion binding	F	IEA	InterPro2GO	InterPro:IPR000971 InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		<a href="#">GO:0005515</a>	protein binding	F	IPI	PMID:11159543	UniProtKB:P68871	9606	20140914	IntAct	
JniProtKB	P02008	HBZ		<a href="#">GO:0005515</a>	protein binding	F	IPI	PMID:6683087	UniProtKB:P68871	9606	20140914	IntAct	
JniProtKB	P02008	HBZ		<a href="#">GO:0019825</a>	oxygen binding	F	IEA	InterPro2GO	InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		<a href="#">GO:0020037</a>	heme binding	F	IEA	InterPro2GO	InterPro:IPR000971 InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		<a href="#">GO:0046872</a>	metal ion binding	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0479	9606	20140913	UniProt	
<b>Component</b>													
JniProtKB	P02008	HBZ		<a href="#">GO:0005833</a>	hemoglobin complex	C	IEA	InterPro2GO	InterPro:IPR002338 InterPro:IPR002340	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		<a href="#">GO:0005833</a>	hemoglobin complex	C	TAS	PMID:7555018		9606	20030904	PINC	
JniProtKB	P02008	HBZ		<a href="#">GO:0070062</a>	extracellular vesicular exosome	C	IDA	PMID:23533145		9606	20140714	UniProt	

Database ID	Gene Product	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Form ID	
<b>Process</b>														
JniProtKB	P02144	MB		GO:0001666	response to hypoxia	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0006810	transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0813	9606	20140913	UniProt		
JniProtKB	P02144	MB		GO:0007507	heart development	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0009725	response to hormone	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0015671	oxygen transport	P	IEA	InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro			
JniProtKB	P02144	MB		GO:0015671	oxygen transport	P	IEA	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl			
JniProtKB	P02144	MB		GO:0015671	oxygen transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt		
JniProtKB	P02144	MB		GO:0031444	slow-twitch skeletal muscle fiber contraction	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0042542	response to hydrogen peroxide	P	IEA	Ensembl Compara	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0043353	enucleate erythrocyte differentiation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl		
JniProtKB	P02144	MB		GO:0050873	brown fat cell differentiation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000125995	9606	20140913	Ensembl		
<b>Function</b>														
JniProtKB	P02144	MB		GO:0005344	oxygen transporter activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt		
JniProtKB	P02144	MB		GO:0005506	iron ion binding	F	IEA	InterPro:IPR000971 InterPro:IPR012292	9606	20140913	InterPro			
JniProtKB	P02144	MB		GO:0019825	oxygen binding	F	IEA	InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro			
JniProtKB	P02144	MB		GO:0019825	oxygen binding	F	IEA	Ensembl:ENSRNOP00000006184	9606	20140913	Ensembl			
JniProtKB	P02144	MB		GO:0020037	heme binding	F	IEA	InterPro:IPR000971 InterPro:IPR002335 InterPro:IPR012292	9606	20140913	InterPro			
JniProtKB	P02144	MB		GO:0046872	metal ion binding	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0479	9606	20140913	UniProt		
<b>Component</b>														
JniProtKB	P02144	MB		GO:0070062	extracellular vesicular exosome	C	IDA	PMID:23533145		9606	20140714	UniProt		

SAME

SAME

Database ID	Gene Product	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Form ID
<b>Process</b>													
JniProtKB	P02008	HBZ		GO:0000122	negative regulation of transcription from RNA polymerase II promoter	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000020531	9606	20140913	Ensembl	
JniProtKB	P02008	HBZ		GO:0006810	transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0813	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		GO:0015671	oxygen transport	P	IEA	InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro		
JniProtKB	P02008	HBZ		GO:0015671	oxygen transport	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		GO:0043249	erythrocyte maturation	P	IEA	Ensembl Compara	Ensembl:ENSMUSP00000020531	9606	20140913	Ensembl	
<b>Function</b>													
JniProtKB	P02008	HBZ		GO:0005344	oxygen transporter activity	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0561	9606	20140913	UniProt	
JniProtKB	P02008	HBZ		GO:0005344	oxygen transporter activity	F	TAS	PMID:7555018		9606	20030904	PINC	
JniProtKB	P02008	HBZ		GO:0005506	iron ion binding	F	IEA	InterPro:IPR000971 InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro		
JniProtKB	P02008	HBZ		GO:0005515	protein binding	F	IPI	PMID:11159743	UniProtKB:P68871	9606	20140914	IntAct	
JniProtKB	P02008	HBZ		GO:0005515	protein binding	F	IPI	PMID:11159743	UniProtKB:P68871	9606	20140914	IntAct	
JniProtKB	P02008	HBZ		GO:0019825	oxygen binding	F	IEA	InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro		
JniProtKB	P02008	HBZ		GO:0020037	heme binding	F	IEA	InterPro2GO	InterPro:IPR000971 InterPro:IPR002338 InterPro:IPR002340 InterPro:IPR012292	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		GO:0046872	metal ion binding	F	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)	UniProtKB-KW:KW-0479	9606	20140913	UniProt	
<b>Component</b>													
JniProtKB	P02008	HBZ		GO:0005833	hemoglobin complex	C	IEA	InterPro2GO	InterPro:IPR002338 InterPro:IPR002340	9606	20140913	InterPro	
JniProtKB	P02008	HBZ		GO:0005833	hemoglobin complex	C	TAS	PMID:7555018		9606	20030904	PINC	
JniProtKB	P02008	HBZ		GO:0070062	extracellular vesicular exosome	C	IDA	PMID:23533145		9606	20140714	UniProt	

SAME

SAME

# Do homologous protein regions perform a similar function?

Homologous proteins may share a number of functional features, however:

- functional drift can lead to different functions or aspects of function
- while functional similarity generally correlates with evolutionary distance, no distance is safe for inferring function (very closely related proteins can have slightly to radically different functions)

# We can integrate homology with other information, for example:

Marco Punta

- Functional motifs
- Conservation of functional residues
- Genomic context (mostly in bacteria)

If structure available:

- structural motifs
- Electrostatic, cavities, etc.

# Detecting homology

# From sequence

Sequences of homologous proteins are related by an evolutionary process, they diverged from a common ancestor.

Modern day homologous proteins have evolved from the same sequence via a number of events (mutations, insertions, deletions, duplications,...)

ALHWRAALAA TVLLVIVLLAGS WLAVLAE

ALHWKAAGAATVLLVIVLLAGSYLAVLAE



ALHWRAAGAATVLLVIVLLAGSYLAVLAE

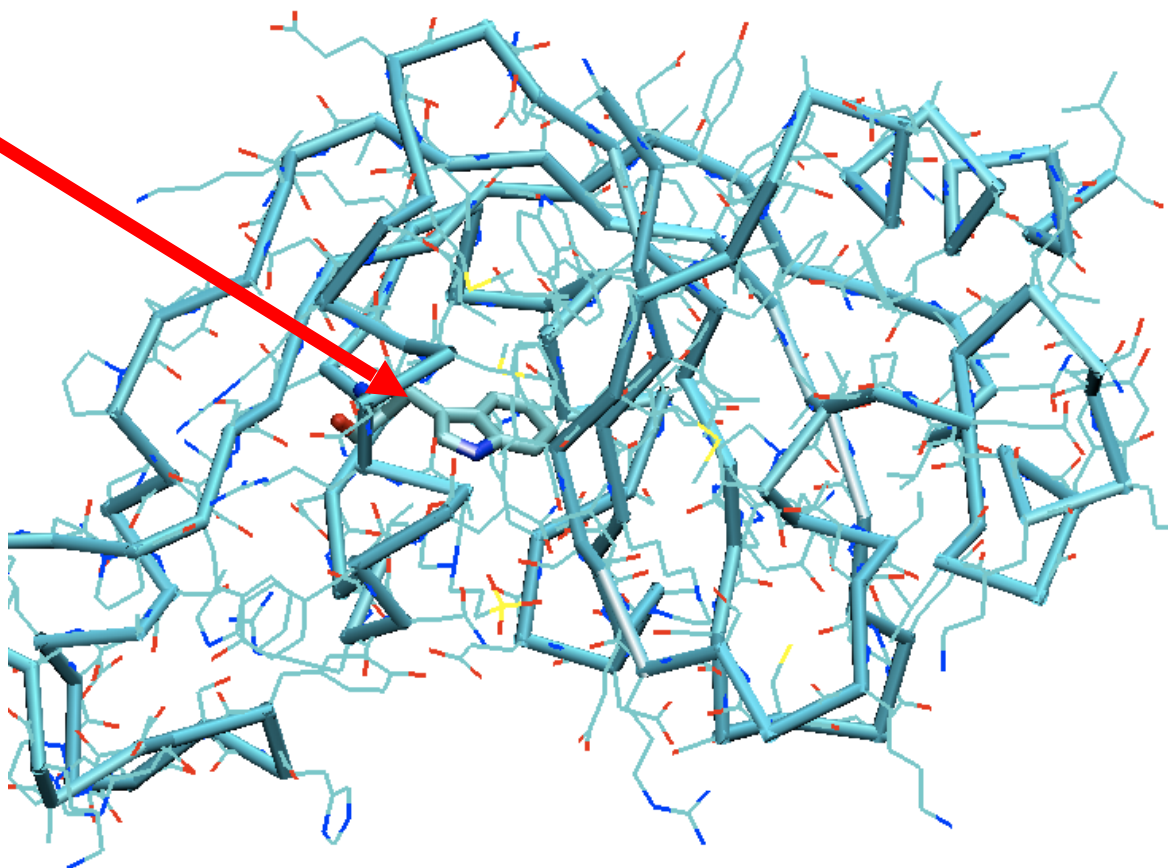
Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLI R LRFK GHPETLEKFDKFKHLKSEDEM KASE 60  
MGLSDGEWQLVLNVWGKVEAD GHGQEVLI LFK HPETL KFDKFK LKSE MK SE  
Mouse: 1 MGLSDGEWQLVLNVWGKVEADLAGHGQEVLI GLFKTHPETLDKFDKFKNLKSEEDMKGSE 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120  
DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL H  
Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEQPLAQSHATKHKIPVKYLEFISEIIIEVLKCRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154  
GDFGADAQGAM KALELFR D A YKELGFQG  
Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

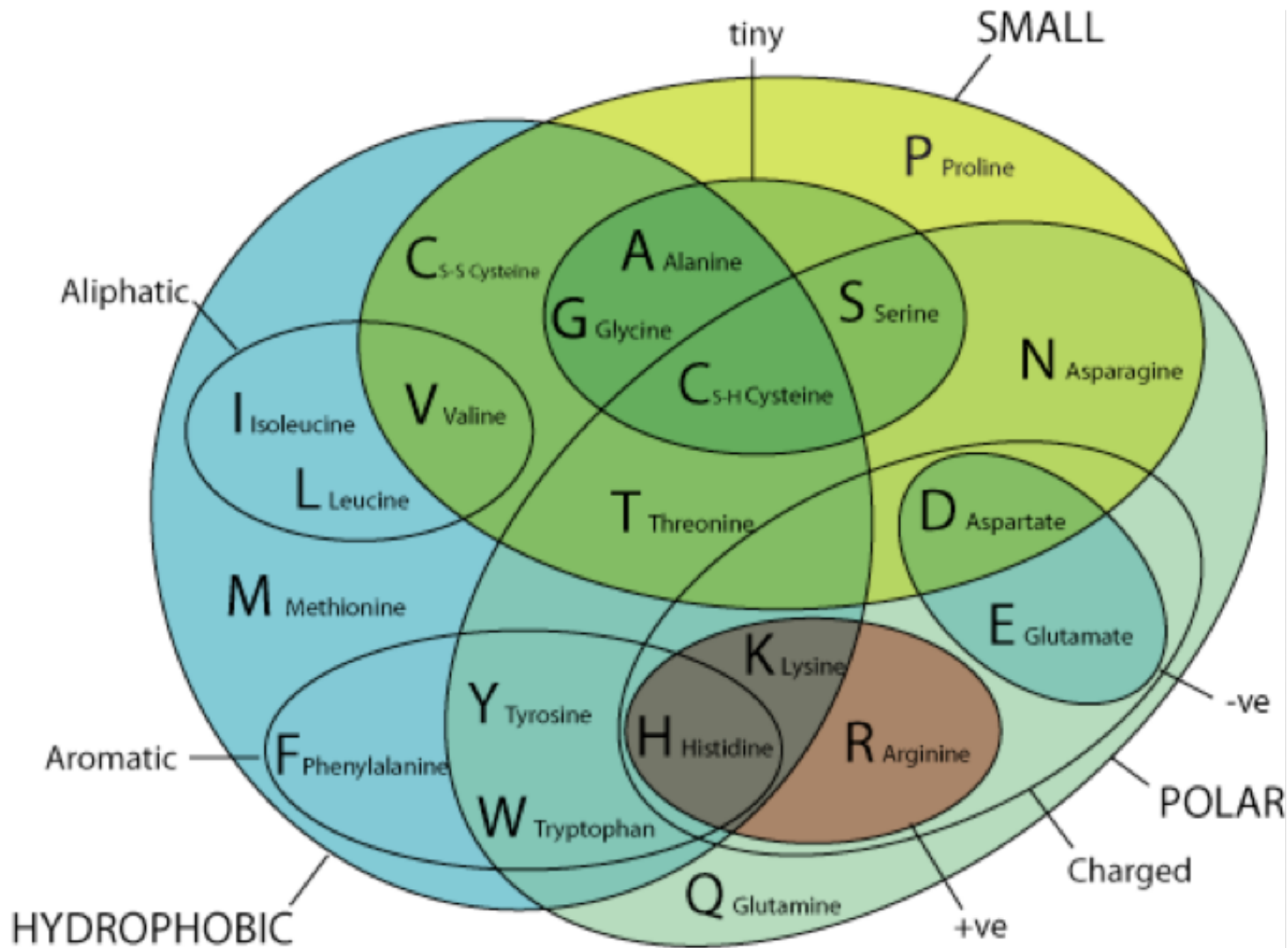
# Protein structural and functional constraints

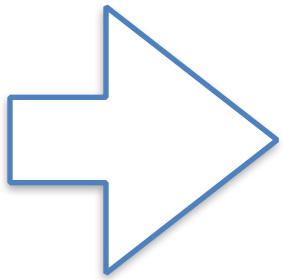
Trp (W)





# aa physico-chemical properties





If divergence not too large we can hope to use sequence similarity to detect homology (excess sequence similarity\*  
-> homology)

# BLOSUM62 matrix

<b>Ala</b>	4																			
<b>Arg</b>	-1	5																		
<b>Asn</b>	-2	0	6																	
<b>Asp</b>	-2	-2	1	6																
<b>Cys</b>	0	-3	-3	-3	9															
<b>Gln</b>	-1	1	0	0	-3	5														
<b>Glu</b>	-1	0	0	2	-4	2	5													
<b>Gly</b>	0	-2	0	-1	-3	-2	-2	6												
<b>His</b>	-2	0	1	-1	-3	0	0	-2	8											
<b>Ile</b>	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
<b>Leu</b>	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
<b>Lys</b>	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
<b>Met</b>	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
<b>Phe</b>	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
<b>Pro</b>	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
<b>Ser</b>	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
<b>Thr</b>	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
<b>Trp</b>	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
<b>Tyr</b>	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
<b>Val</b>	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
	<b>Ala</b>	<b>Arg</b>	<b>Asn</b>	<b>Asp</b>	<b>Cys</b>	<b>Gln</b>	<b>Glu</b>	<b>Gly</b>	<b>His</b>	<b>Ile</b>	<b>Leu</b>	<b>Lys</b>	<b>Met</b>	<b>Phe</b>	<b>Pro</b>	<b>Ser</b>	<b>Thr</b>	<b>Trp</b>	<b>Tyr</b>	<b>Val</b>

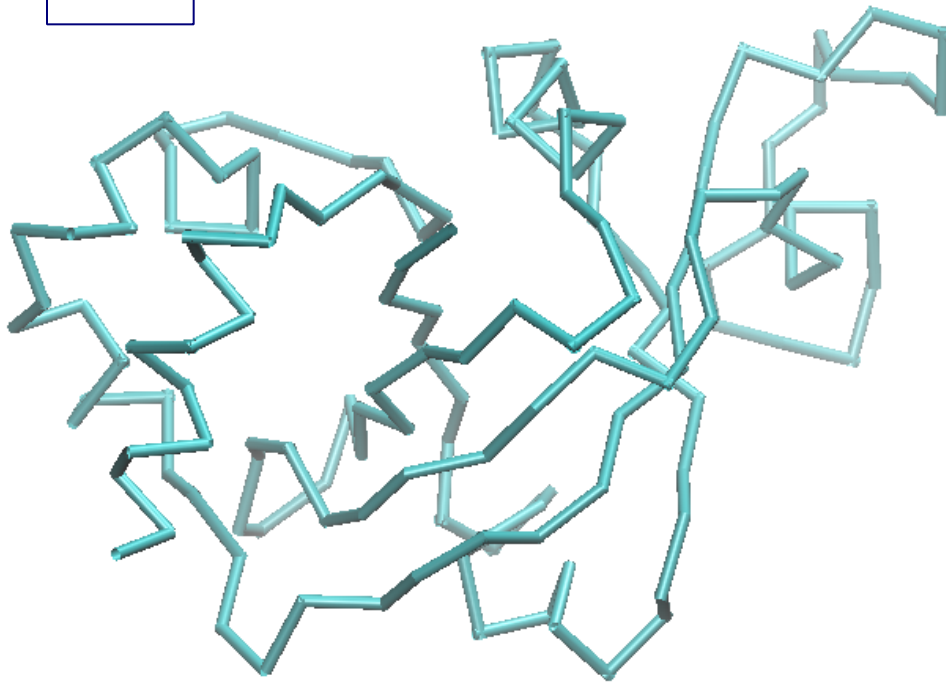
# Sequence alignment, what we need:

- Scoring system => empirically derived substitution matrices (PAMs, BLOSUMs,...)
- Efficient way to find highest scoring alignments => dynamic programming (Needleman-Wunsch, Smith-Waterman,...)
- Way to decide whether top score is high enough to infer homology (significance) => E-value, ...

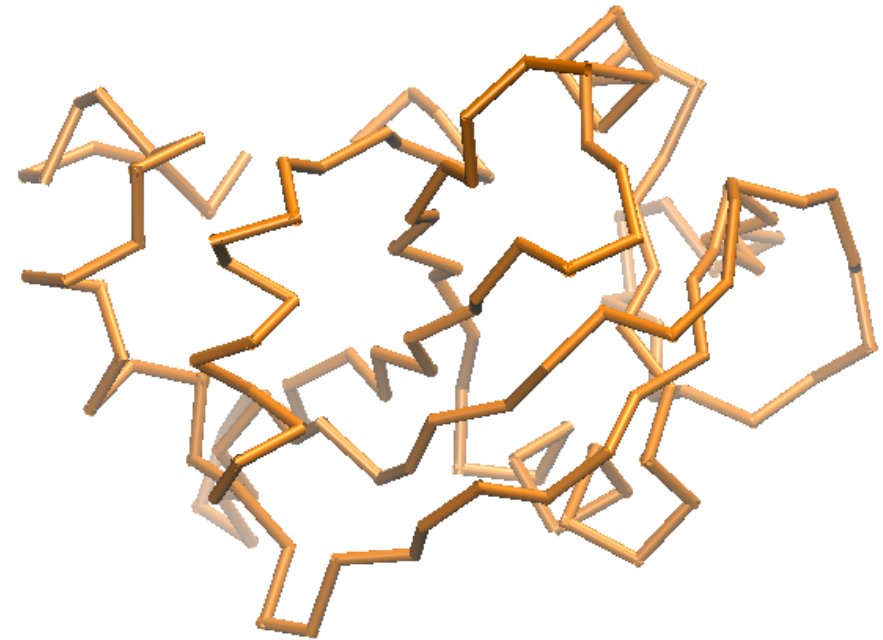
# From structure

Marco Punta

2G2X



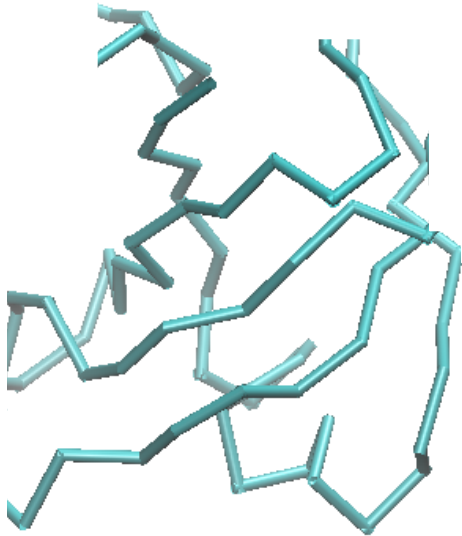
2P5D



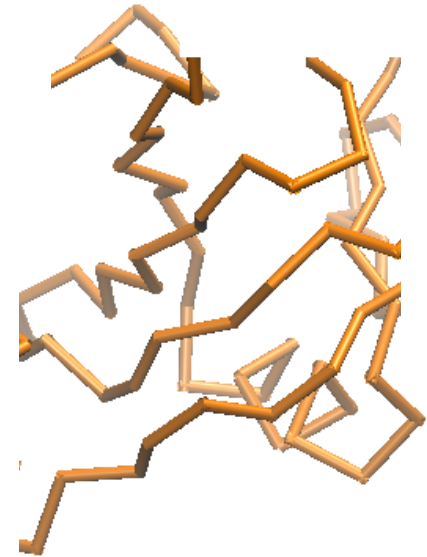
# Structural similarity

Marco Punta

2G2X



2P5D

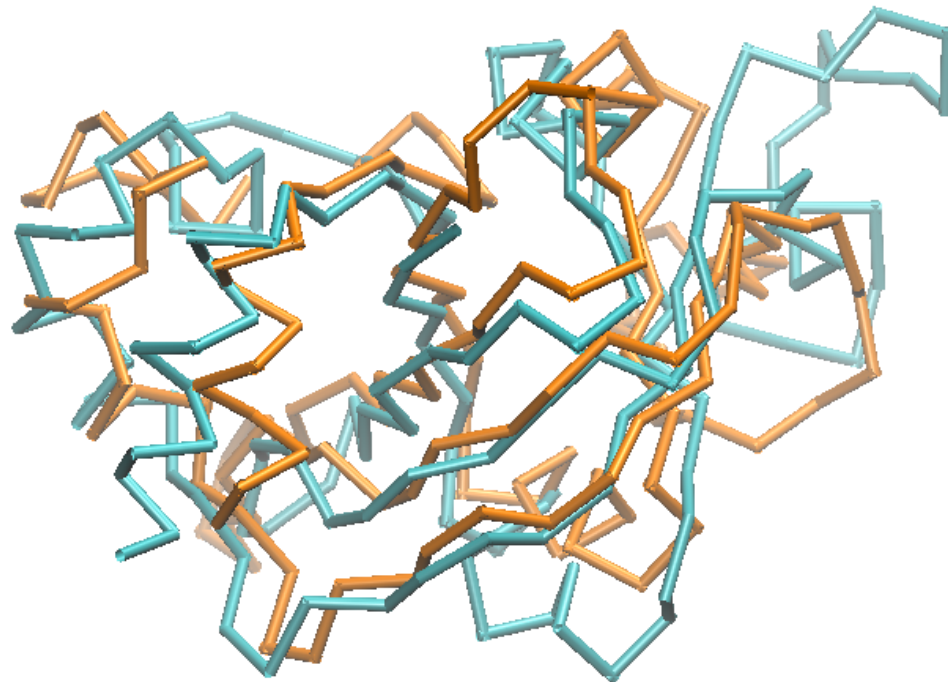


# Structural similarity

Marco Punta

2G2X

2P5D



Z-score = 12.2

RMSD = 2.9

Lali = 122

%id = 20

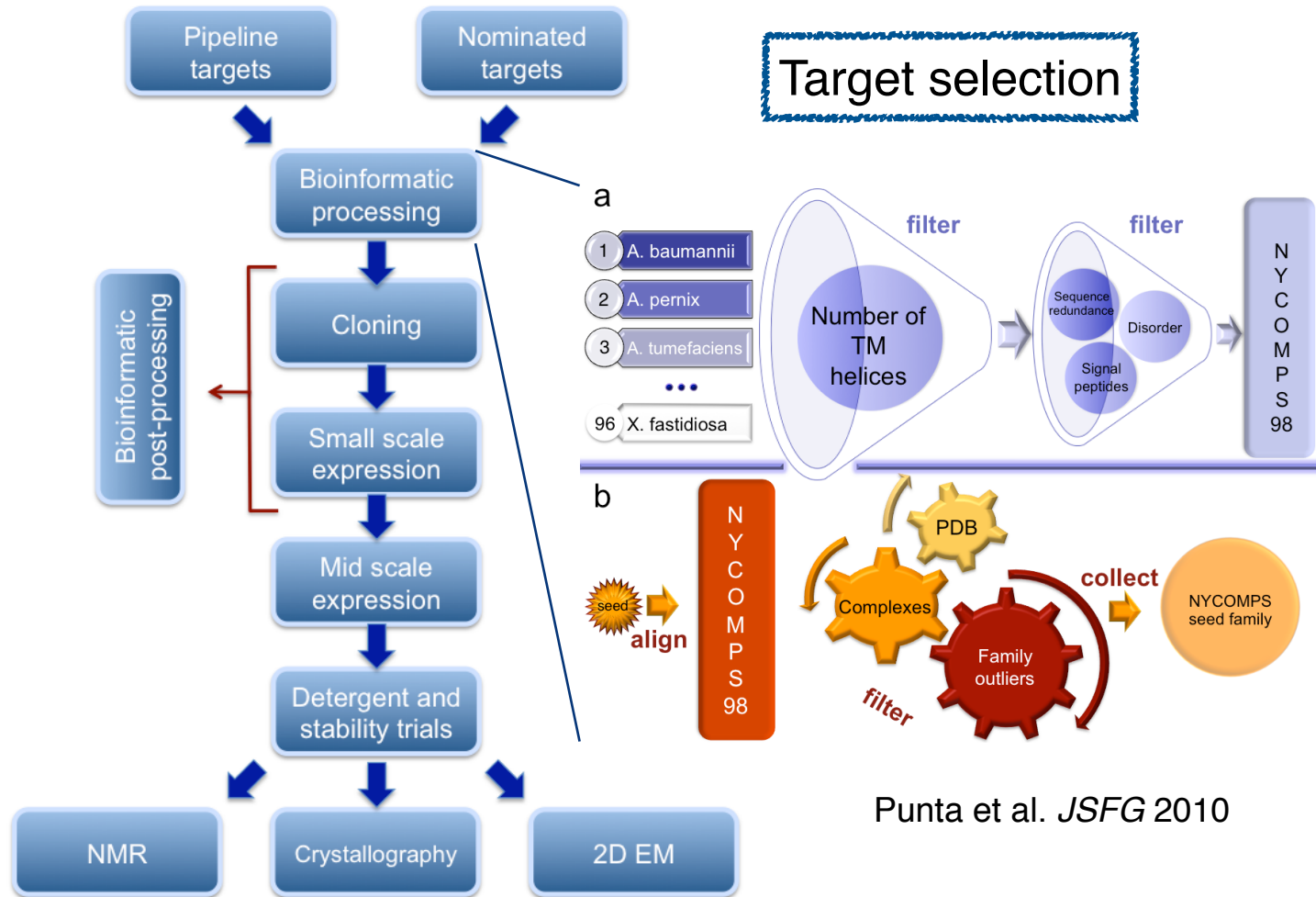
DALI: [http://ekhidna.biocenter.helsinki.fi/dali\\_lite/start](http://ekhidna.biocenter.helsinki.fi/dali_lite/start)

# Exercise

Homology-based function annotation transfer #1



# NYCOMPS pipeline



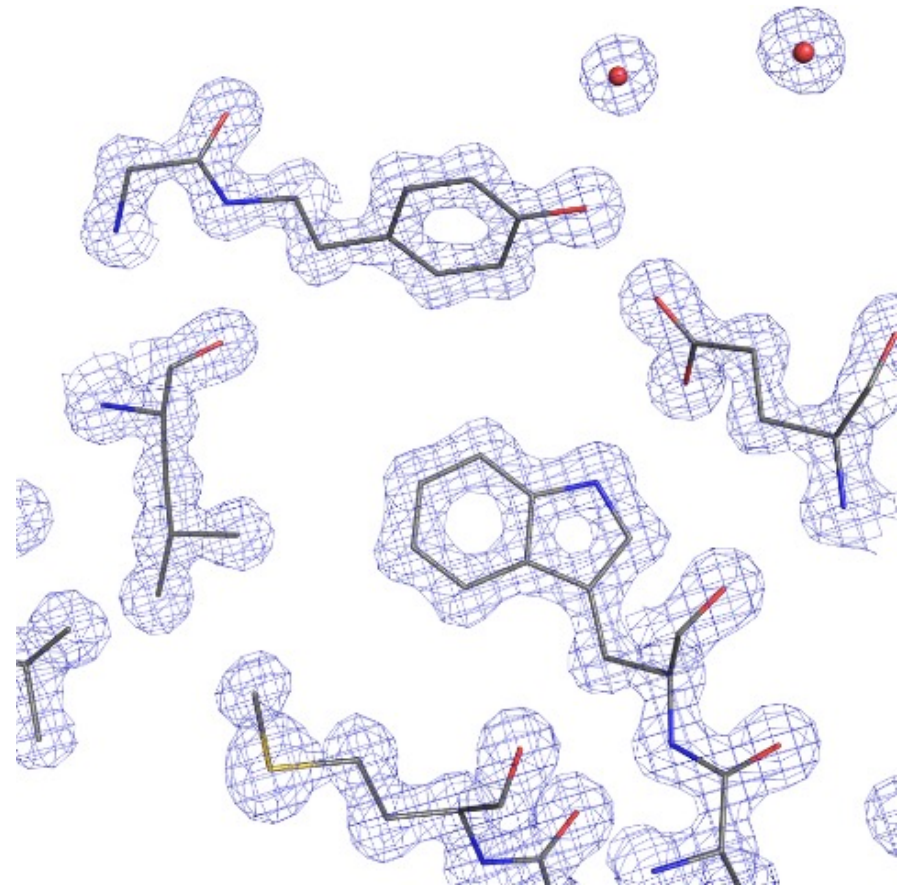
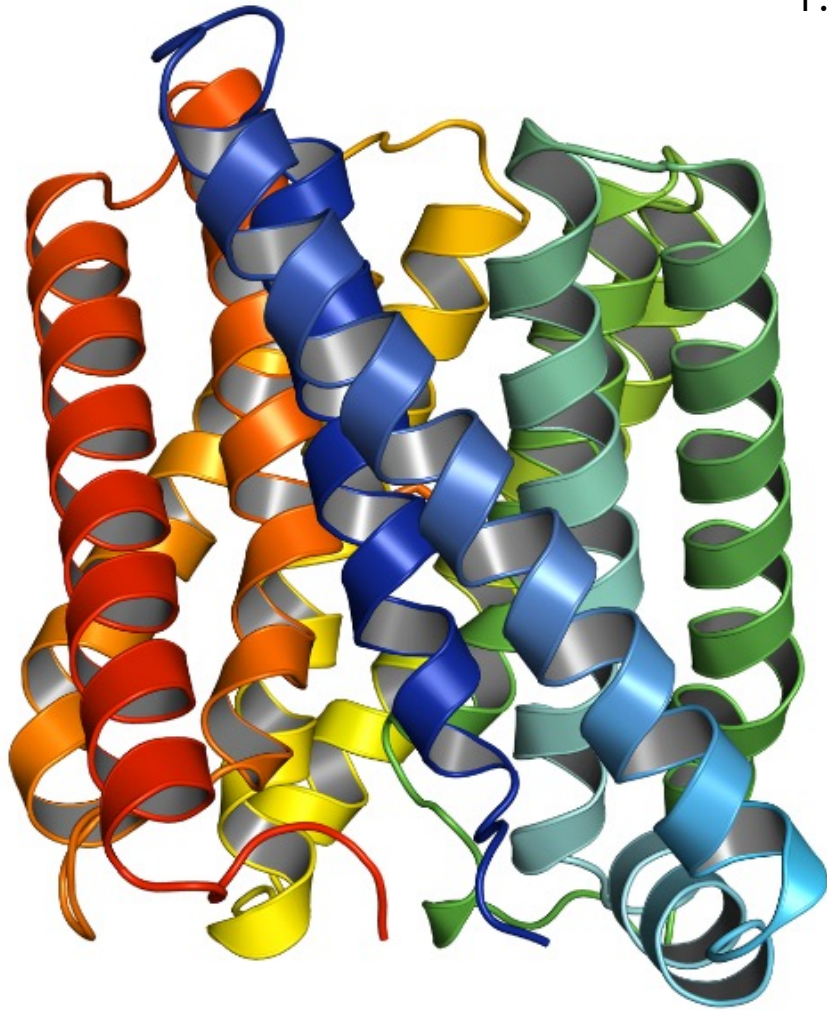
Love et al. *JSFG* 2010

Punta et al. *JSFG* 2010

H. influenzae protein [3M71] ← PDB id

Marco Punta


1.20 Å



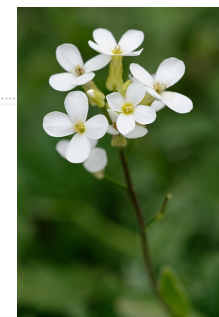
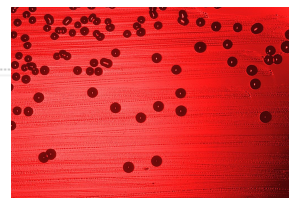
Chen et al. *Nature* 467 (2010)

# Alignment

Q9LD83		SLAC1_ARATH - Guard cell S-type anion channel SLA... - Arabidopsis thal...	
<b>E-value:</b> 3e-10		<b>Positives :</b> 41.0%	
<b>Score:</b> 160		<b>Query Length:</b> 328	
<b>Ident.:</b> 22.0%		<b>Match Length:</b> 556	

P44741	20	PFPL--PTGYFGIPLGLAALS LawFHLE-----NLFP AARMVSDVLGIVASAVWILFILM	72
Q9LD83	183	PF L P G FGI LGL++ ++ W L N +++ V+ + + V +	242
P44741	73	YAYKLRYYFEEVRAEYHSPVRF SFIALIPITMLVG---DILYRWNPLIAEVL I WIGTIG	129
Q9LD83	243	YILKCIFYFEAVKREYFHPVRVNF FFAPWVVCMLAISVPPMFSPNRKYLHPA IWCVFMG	302
P44741	130	QLLFSTLRVSELWQGGVFEQ--KSTHPSFYLP A V AANFTSASSLALLGYHDLGYL PFGAG	187
Q9LD83	303	F L++ W G + K +PS +L +V NF A + +G+ ++ + G	361
P44741	188	MIAWIIFEPVLLQHLRISSLEPQFRATMG I V L A P A F V C V S A Y L S I N H G E V D T L A K I L W G Y	247
Q9LD83	362	+++ L Q L S P+ + + A S + +G+ D ++ +	421
P44741	248	GFLQLFLLRLFPWIVEKGLNIGLWAFS FGLASMAN SATAFY----HGNVLQGVSI FAFV	303
Q9LD83	422	L+ + ++ W+++F + + A+ AT Y G + +++	480
P44741	304	FSNMIGLLVLMTI 317	
Q9LD83	481	S M+ +L + T+ 494	



E-value is the number of matches with a given score (or higher) that we expect to occur by chance.


This depends on database size!

For an alignment with score  $S$  and  $E\text{-value}=1$ , we expect to have by chance 1 match with the same or higher score.

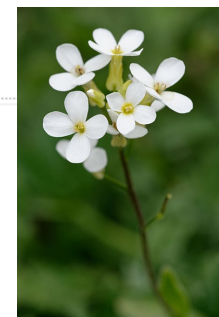
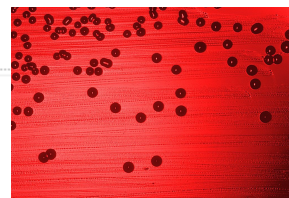
For an alignment with score  $S$  and  $E\text{-value}=1$ , we expect to have by chance 1 match with the same or higher score. If  $E\text{-value}$  is 0.001 then we expect by chance 0.001 matches with the same or higher score.

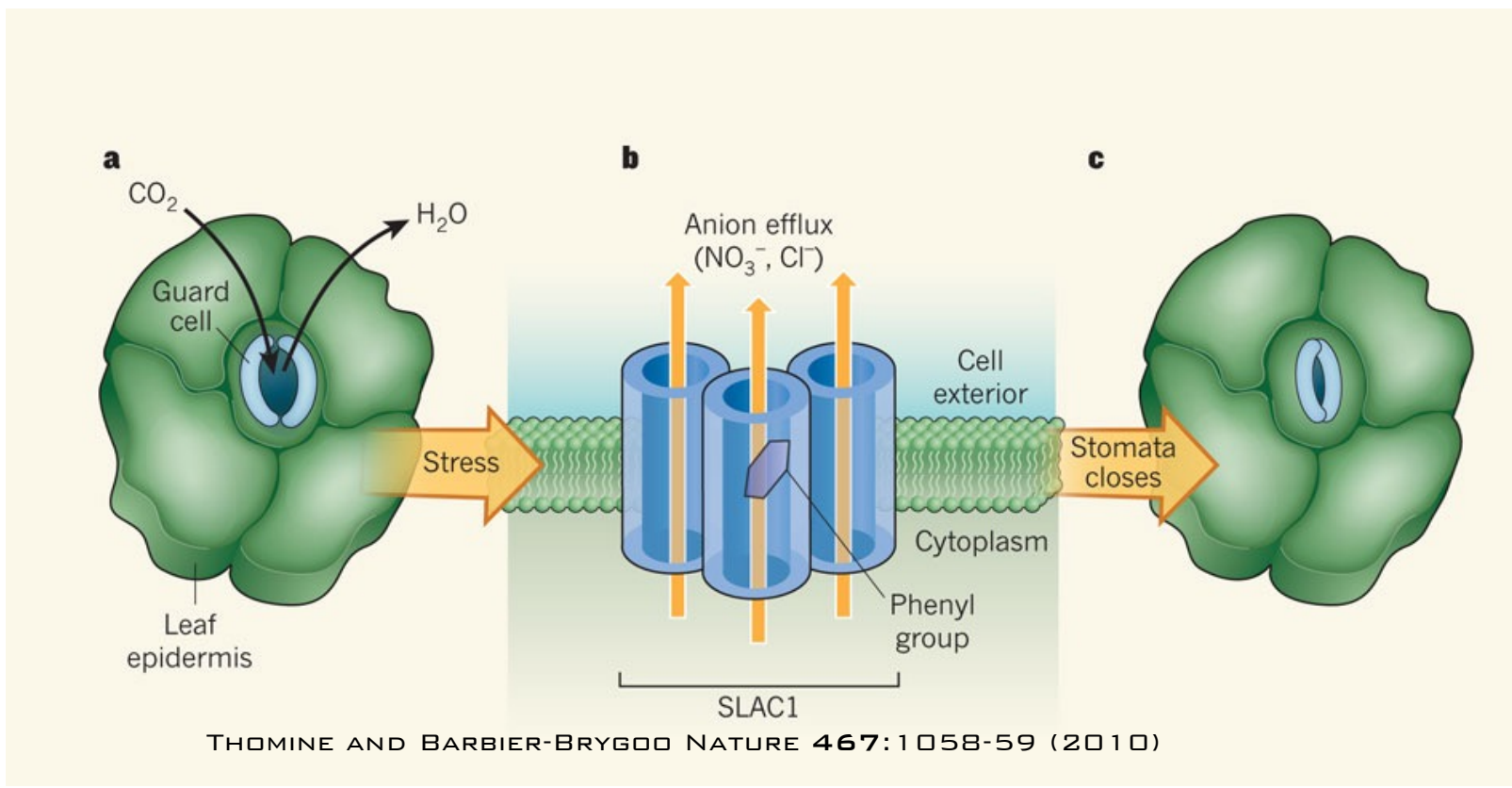
# Alignment

Q9LD83		SLAC1_ARATH - Guard cell S-type anion channel SLA... - Arabidopsis thal...	
<b>E-value:</b> 3e-10		<b>Positives :</b> 41.0%	
<b>Score:</b> 160		<b>Query Length:</b> 328	
<b>Ident.:</b> 22.0%		<b>Match Length:</b> 556	

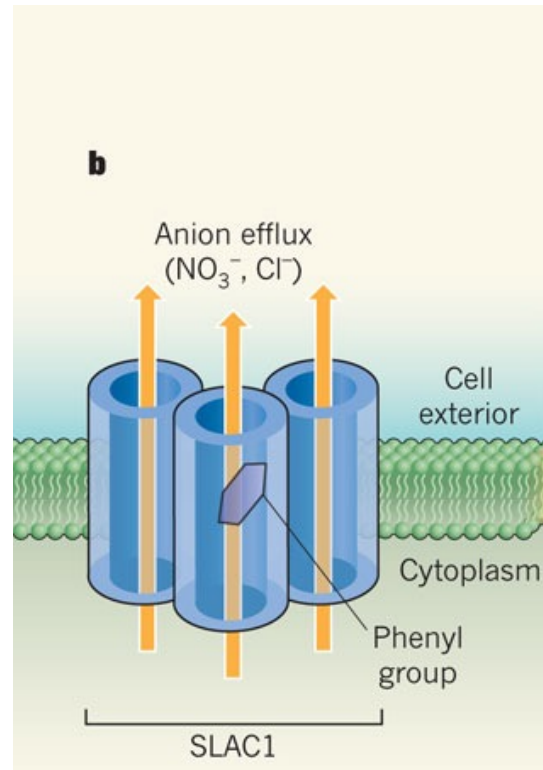
P44741	20	PFPL--PTGYFGIPLGLAALS LawFHLE-----NLFP AARMVSDVLGIVASAVWILFILM	72
Q9LD83	183	PF L P G FGI LGL++ ++ W L N +++ V+ + + V +	242
P44741	73	YAYKLRYFEEVRAEYHSPVRF SFIALIPITMLVG---DILYRWNPLIAEVL I WIGTIG	129
Q9LD83	243	YILKCIFYFEAVKREYFHPVRVNF FFAPWVCMFLAISVPPMFS PNRKYLHPA IWCVFMG	302
P44741	130	QLLFSTLRVSELWQGGVFEQ--KSTHPSFYLP A V A A N F T S A S S L A L L G Y H D L G Y L P F G A G	187
Q9LD83	303	F L++ W G + K +PS +L +V NF A + +G+ ++ + G	361
P44741	188	MIAWIIFEPVLLQHLRISSELPQFRATMGIVLAPAFV CV S A Y L S I N H G E V D T L A K I L W G Y	247
Q9LD83	362	+++ L Q L S P+ + + A S + +G+ D ++ +	421
P44741	248	GFLQLFLLRLFPWIVEKGLNIGLWAFS FGLASMAN SATAFY----HGNVLQGV S I F A F V	303
Q9LD83	422	L+ + ++ W+++F + + A+ AT Y G + +++	480
P44741	304	FSNVMIGLLV LMTI 317	
Q9LD83	481	S M+ +L + T+ 494	





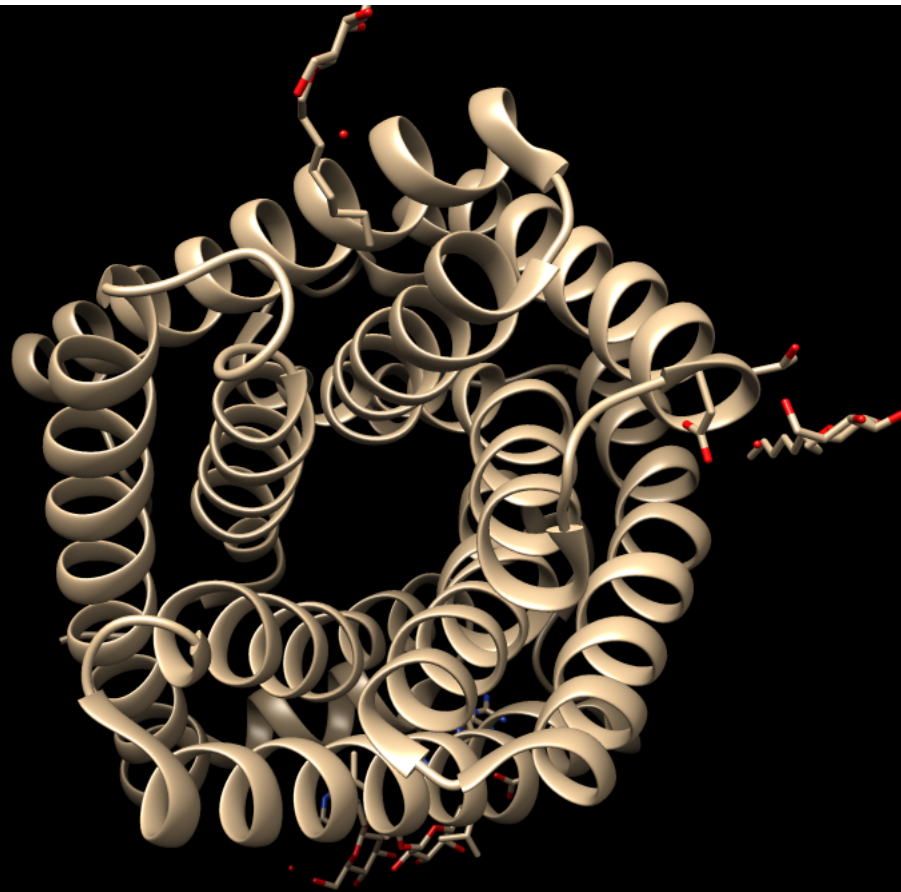
THOMINE AND BARBIER-BRYGDOO NATURE **467**:1058-59 (2010)





1. OPEN Chimera

2. File -> Open "3M71.pdb"

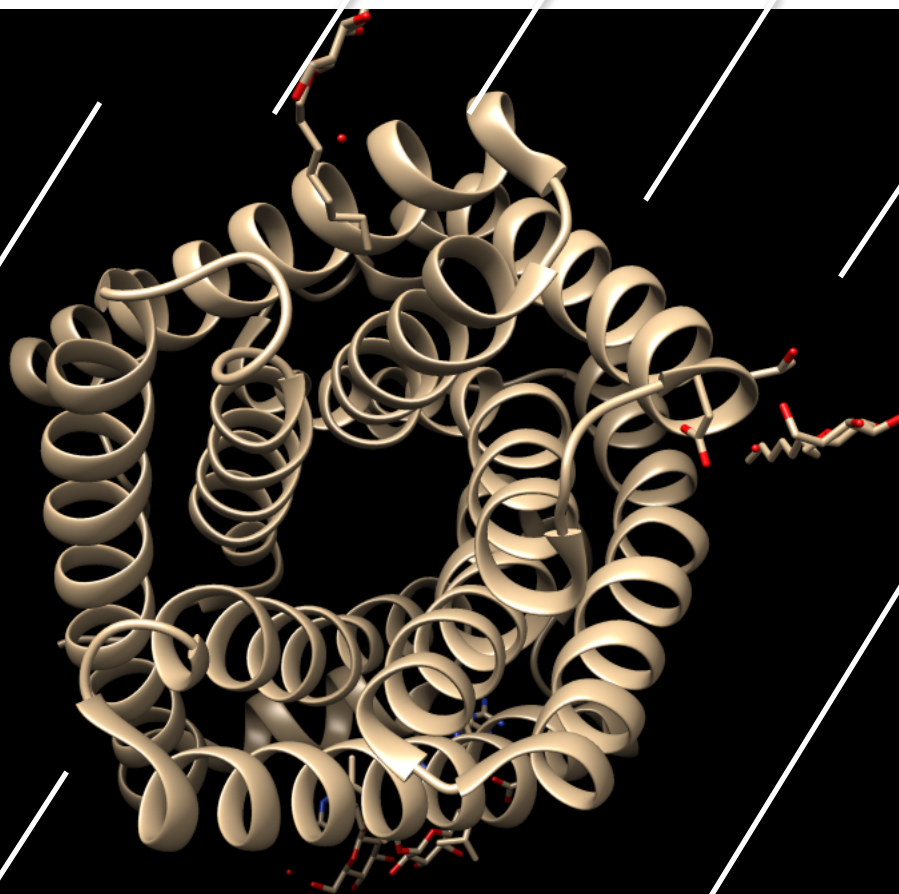


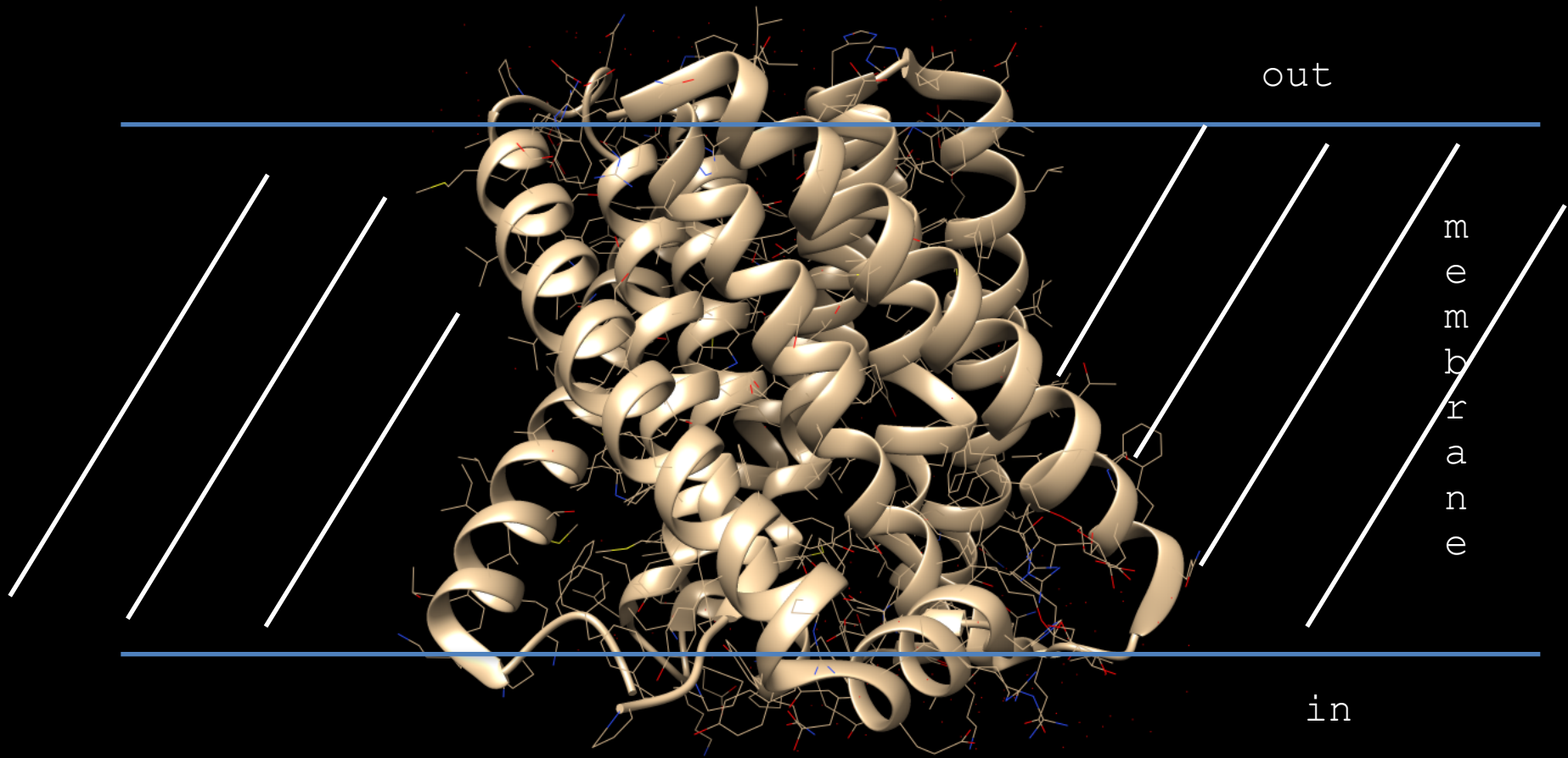
out

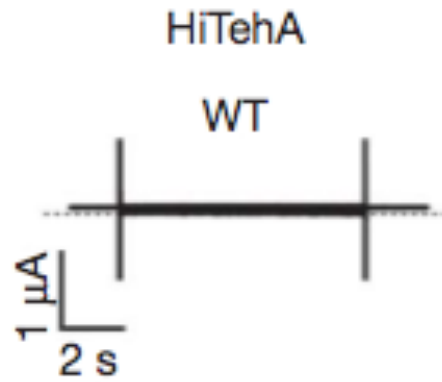
out

out

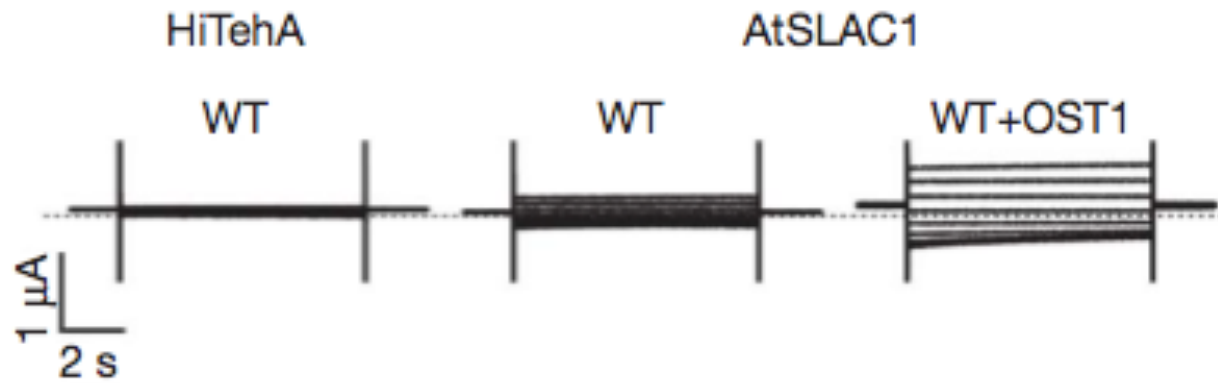
out













1. Actions -> Atoms/Bonds -> wire

2. Actions -> Atoms/Bonds -> show

1. Actions -> Atoms/Bonds -> wire

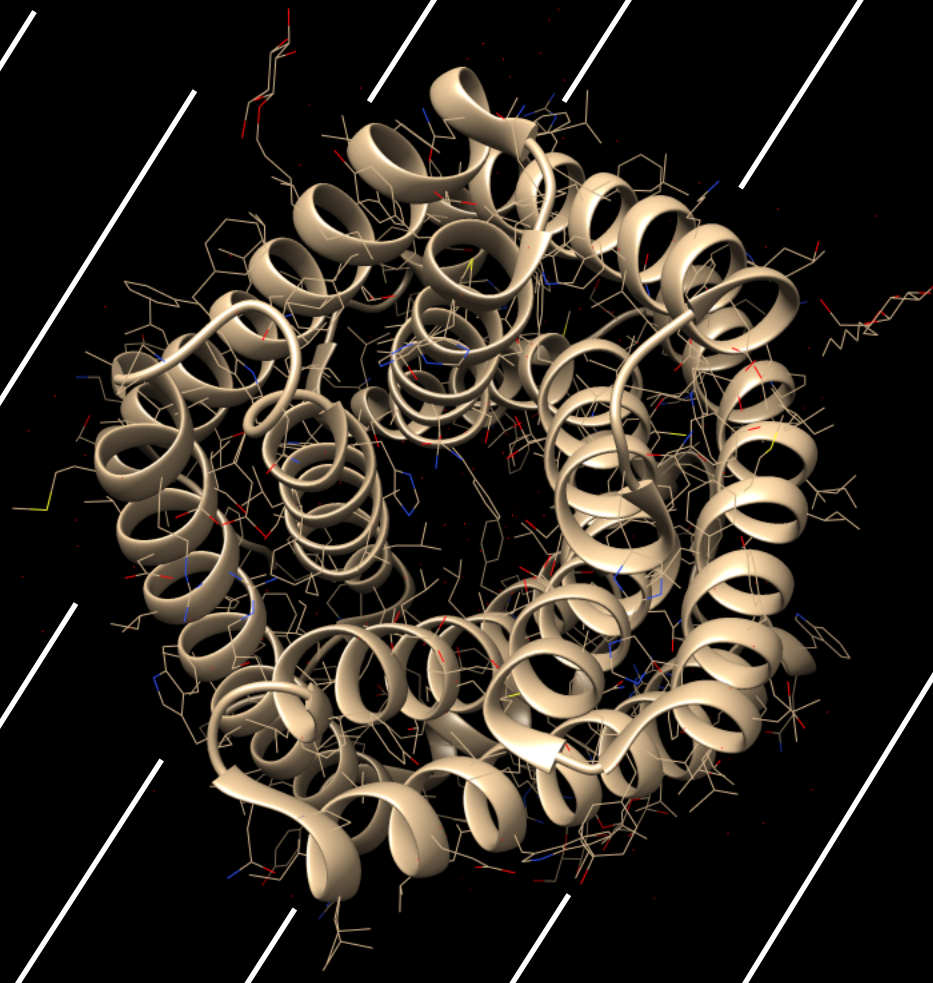
2. Actions -> Atoms/Bonds -> show

out

out


out

out

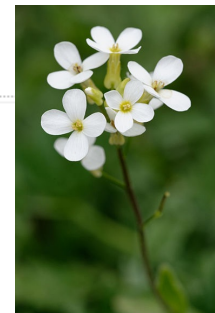
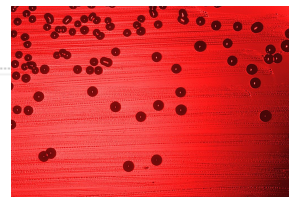


# Alignment

Q9LD83		SLAC1_ARATH - Guard cell S-type anion channel SLA... - Arabidopsis thal...	
<b>E-value:</b> 3e-10		<b>Positives :</b> 41.0%	
<b>Score:</b> 160		<b>Query Length:</b> 328	
<b>Ident.:</b> 22.0%		<b>Match Length:</b> 556	

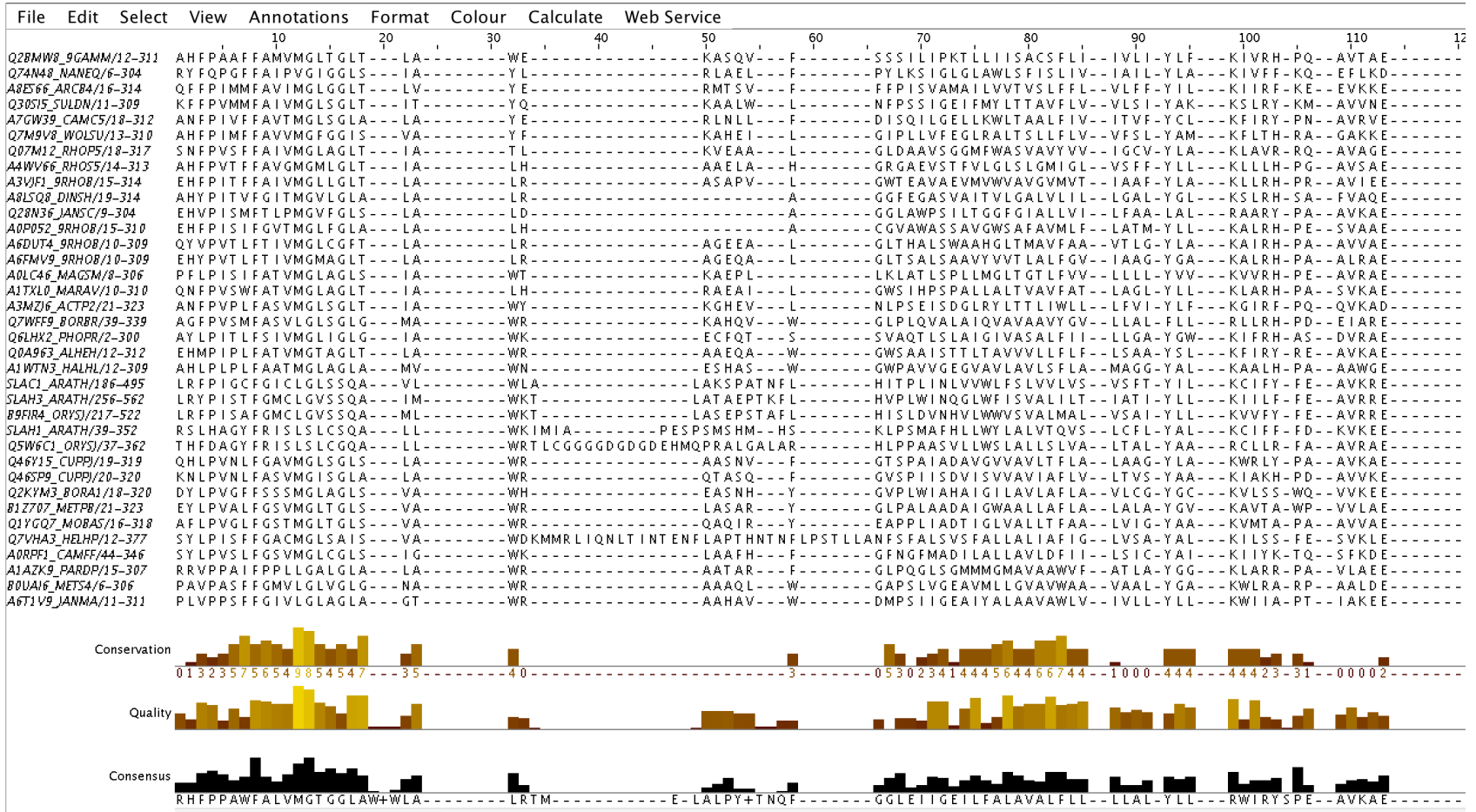
  


P44741	20	PFPL--PTGYFGIPLGLAALS LawFHLE-----NLFP AARMVSDVLGIVASAVWILFILM	72
Q9LD83	183	PF L P G FGI LGL++ ++ W L N +++ V+ + + V +	242
P44741	73	YAYKLRYFFEEVRAEYHSPVRF SFIALIPITMLVG---DILYRWNPLIAEVL I WIGTIG	129
Q9LD83	243	YILKCIFYFEAVKREYFHPVRVNF FFAPWVCMFLAISVPPMFSPNRKYLHPA IWCVFMG	302
P44741	130	QLLFSTLRVSELWQGGVFEQ--KSTHPSFYLP A V A A N F T S A S S L A L L G Y H D L G Y L P F G A G	187
Q9LD83	303	PYFFLELKIYQWL SGGKRR LCKVANPSSHL--SVVGNFV GAILASKV G W D E V A K F L W A V G	361
P44741	188	MIAWIIFEPVLLQHLRIS SLEPQFRATMGIVLAPAFVCV S A Y L S I N H G E V D T L A K I L W G Y	247
Q9LD83	362	FAHYLVVFTLYQRLPTSEALPKELHPVYSMFIAAPSAAS IAWNTIY G Q F D G C S R T C F F I	421
P44741	248	GFLQLFLLRLFPWIVEKGLNIGLWAF S FCLASMAN SATAFY----HGNVLQGV S I F A F V	303
Q9LD83	422	ALFLYISLVARINFFTGFKFSVAWWSYTFMTT--ASVATIKYAEAVPGYPSRALALTLSF	480
P44741	304	FSNMIGLLV LMTI 317	
Q9LD83	481	ISTAMVCVLFVSTL 494	



1. OPEN Jalview

2. File -> Input Alignment -> From File "PF03595\_seed.txt"

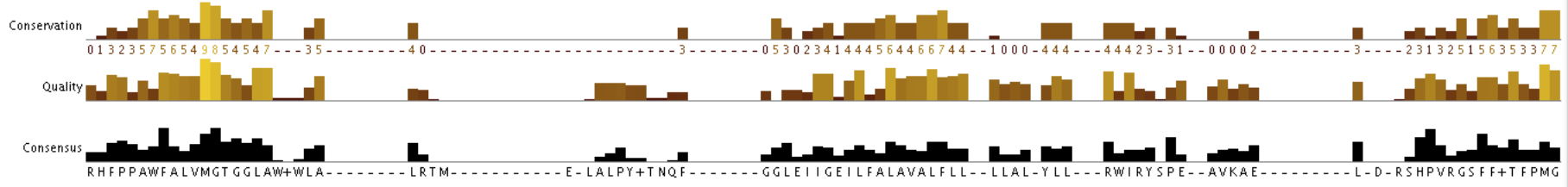


# 1. Colour -> BLOSUM62

(View 1)

Edit Select View Format Colour Calculate Help

	10	20	30	40	50	60	70	80	90	100	110	120	130	140
QHK5_NITHX/19-344	RRLSPAY	FALV	MG	GIVS	IA	SE								
CM02_KLULA/14-410	EQFS	PFW	IT	CMGT	GISA	SI	LH							
PUS3_MYCUA/88-426	GNIGPNW	EASV	MG	GIVA	VA	GA								
EMF6_MYCPE/17-349	SHLP	IAYT	GLS	L	GGGG	IGNCLS	L							
TRG5_9MICO/34-371	AF	IGNPW	FAAV	MG	GIVA	NA	VA							
X021_NEUCS/27-379	T	CYQWT	FT	MT	MGVA	NV	LH							
BQ77_DEBHA/44-436	KG	FT	PGL	FVT	MG	PVSS	CI	LY						
KNW1_CRYNJ/86-454	LN	IS	PAF	FS	NMG	GITS	IL	LY						
ZWF9_YEAS7/11-410	RQ	FD	PF	MM	VG	GISS	NI	LY						
8507_GRABC/25-365	RQ	FT	PNW	FAAT	MG	GILA	LT	LP						
5WM7_MANSM/7-307	F	P	IP	VNY	FS	MV	LAGL	LA	WR					
MZ16_ACTP2/21-323	AN	FV	PL	FAS	V	MGLS	GLT	IA	WY					
J820_SULAC/13-338	SK	LL	PSY	FAS	V	M	GIF	IA	FF					
JEY2_NITOC/23-349	KD	LS	PAS	F	GM	V	MAT	GIVS	LA	AH				
AH1_ARATH/39-352	RS	LHAGY	ER	IS	LS	CS	QA	LL	WK	IMIA				
51A3_NEUCR/181-523	KH	FT	WAWY	T	LC	M	ST	GGLS	LL	IA				
T4T9_BURTA/21-360	RQ	FT	PNW	FAMS	MG	NG	I	V	LV					
Z707_METPB/21-323	EY	L	P	V	A	L	G	S	VA	WR				
FRV1_CANGA/10-406	R	D	E	F	F	M	V	M	A	S	G	I	S	S
Q217_9BACT/9-334	K	T	L	H	P	A	V	F	I	M	S	T	G	I
N826_MYCGA/11-350	E	Q	V	L	A	L	S	G	L	T	G	I	G	L
34M0_LACC3/6-302	K	R	V	P	L	M	A	G	L	T	L	G	L	S
FD48_9GAMM/10-310	K	L	L	R	T	P	V	A	G	L	A	L	G	I
CHP7_ASPCL/17-368	F	K	E	S	P	Q	W	L	V	P	Q	G	S	I
N172_CORDI/18-353	P	G	S	G	P	V	W	F	S	V	M	G	T	G
CKM4_PASMU/9-309	F	P	L	I	N	Y	G	I	L	L	S	A	L	S
51K0_NEUCR/119-539	E	G	C	L	S	R	E	T	W	P	T	S	L	L
FM12_COREF/44-374	P	P	A	G	P	A	W	A	G	S	L	M	G	T
XPG1_PSENY/16-355	R	H	E	T	P	N	W	F	A	V	T	M	G	T
VKFB_ACTS2/5-305	T	P	L	P	V	N	Y	E	S	I	V	L	G	L
38CS_ENTFA/10-307	R	K	V	P	I	P	C	G	L	I	L	G	M	S
BQ78_DEBHA/45-438	KG	FT	P	A	V	G	M	G	T	G	V	S	S	C
S2V5_NEUCR/87-465	M	H	F	T	F	A	W	Y	T	V	T	L	S	T
T1V9_JANMA/11-311	P	L	V	P	P	S	F	G	I	V	L	G	L	A
R616_ASPNC/19-376	S	S	V	A	W	G	W	Y	A	I	S	L	S	W
M1Z3_CORGL/17-324	P	P	P	G	S	W	A	G	S	L	M	G	I	S
76_METJA/13-336	K	N	F	V	P	S	W	A	V	M	G	T	G	I
6Y15_CURPJ/19-319	Q	H	L	V	N	L	G	A	V	M	G	L	S	L
LC46_MAGSM/8-306	P	F	L	I	S	I	F	A	T	V	M	G	L	A
R414_MYCS2/18-309	R	R	V	K	P	N	V	F	A	V	M	A	T	G
8N36_JANCS/9-304	E	H	V	I	S	M	E	T	L	P	M	G	V	F



sequence position 46 8.587444

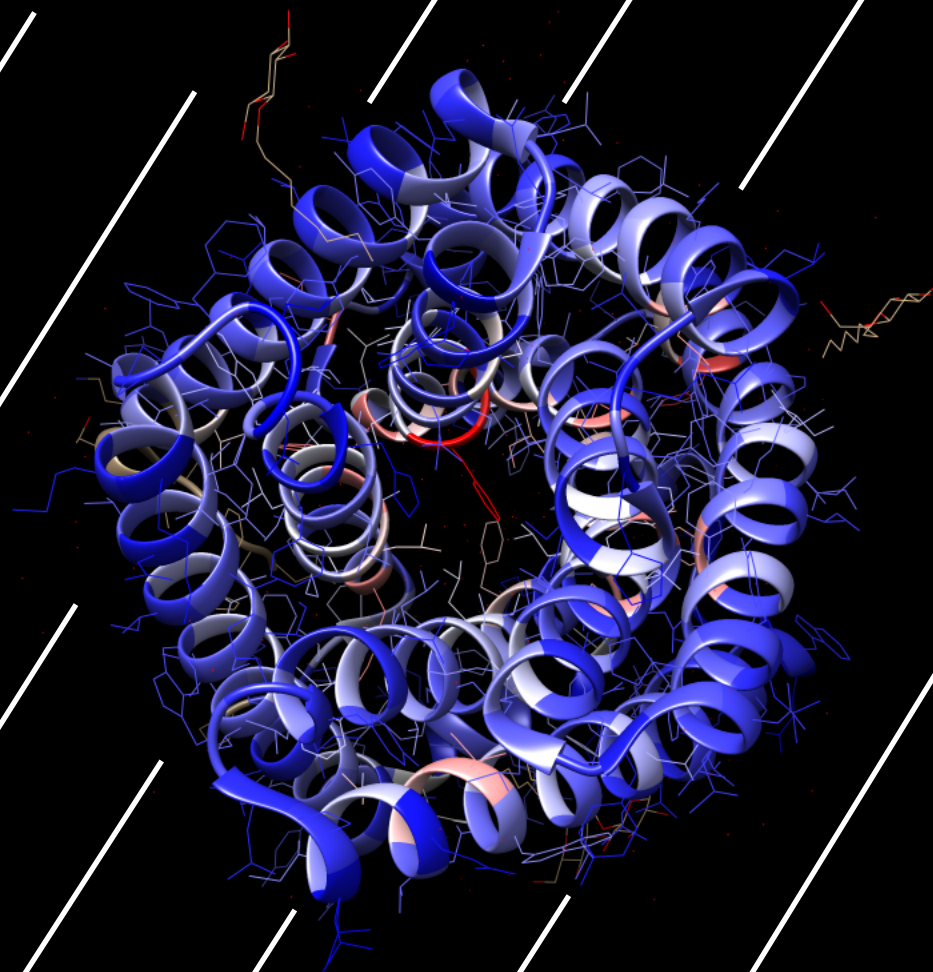
1. Tools-> Sequence -> Multialign Viewer
2. Choose "PF03595\_seed.txt"
3. Select Aligned FASTA
4. Structure -> Render by Conservation

out

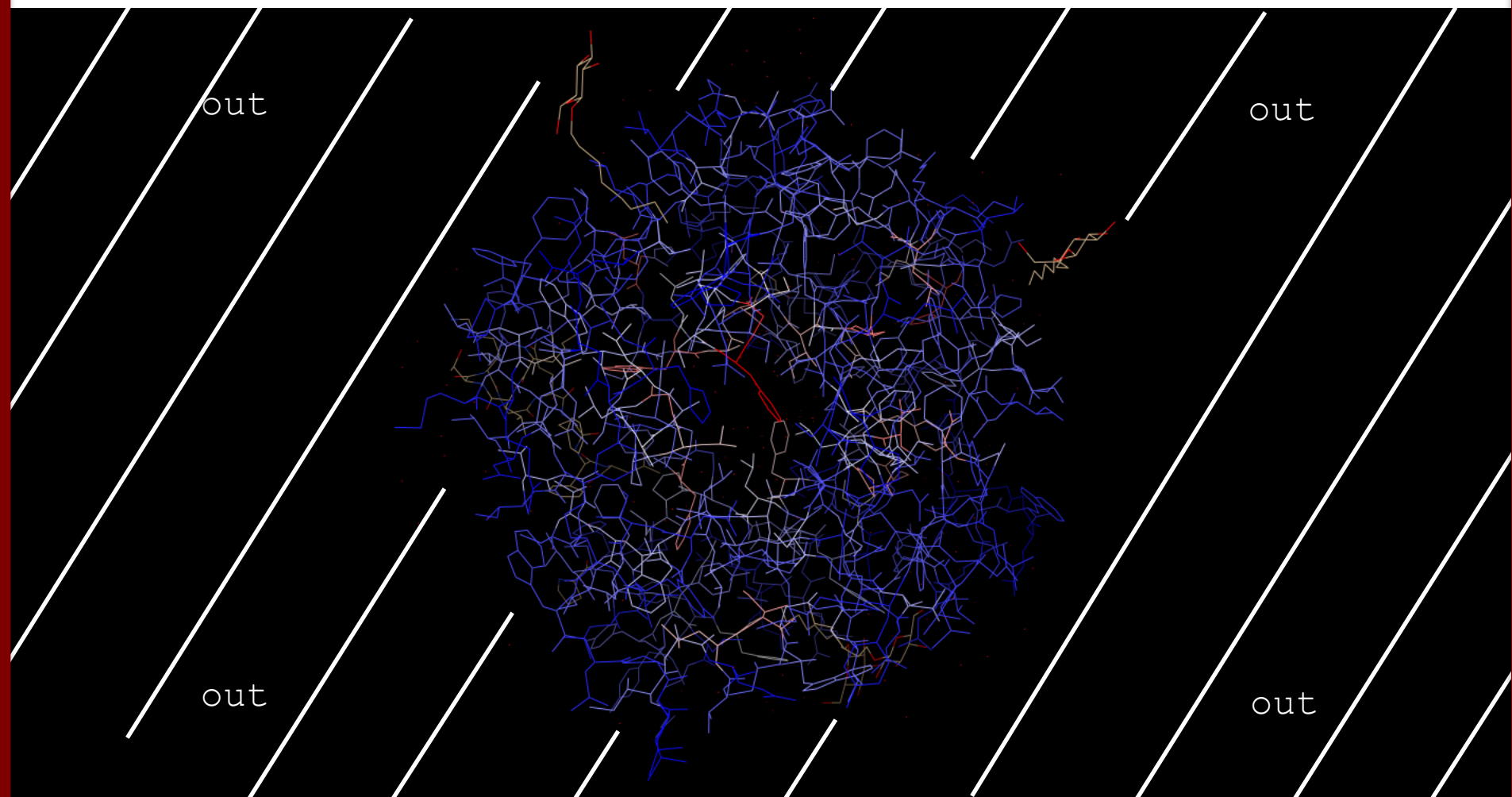
out

out

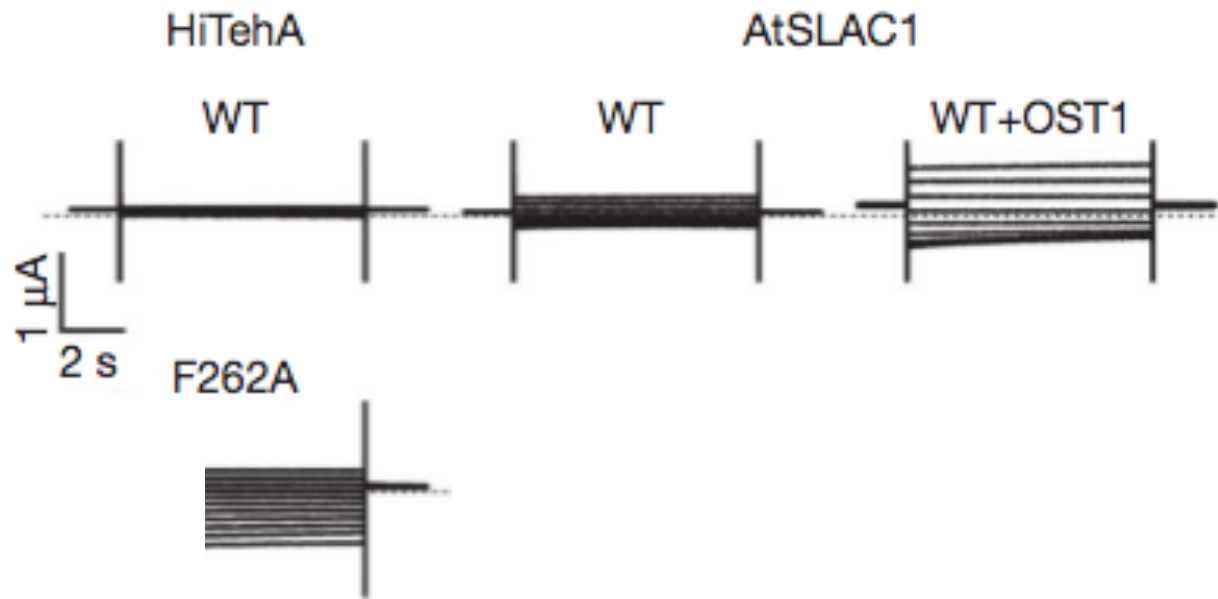
out

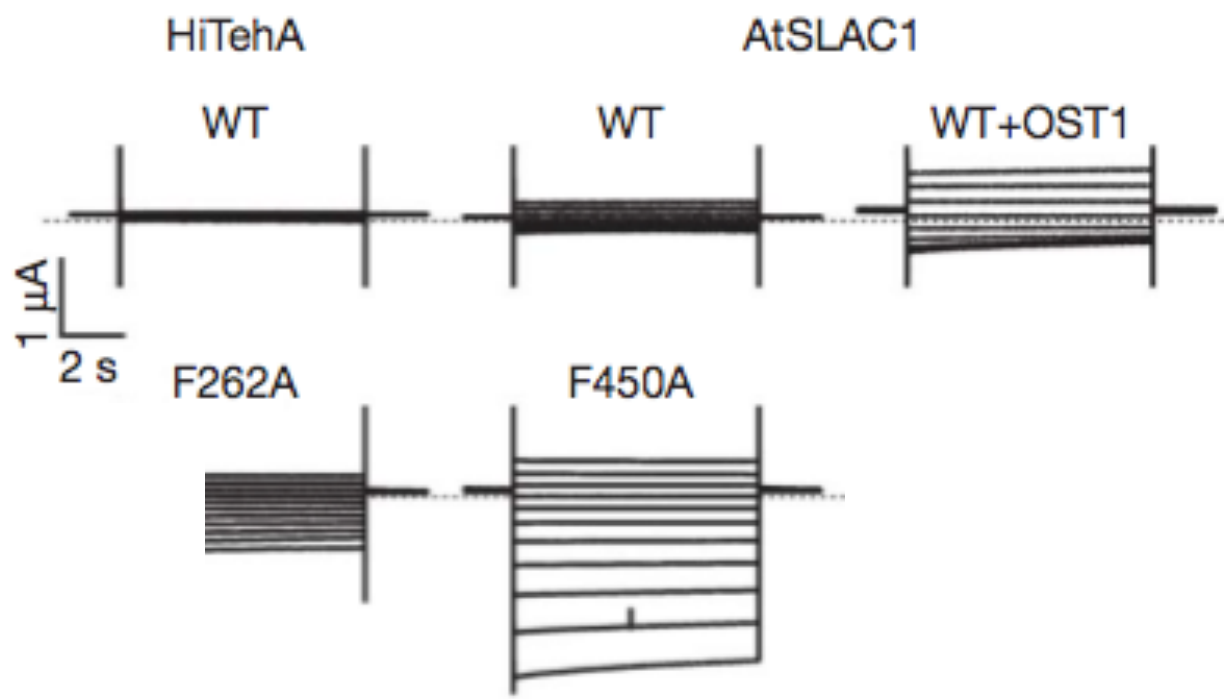


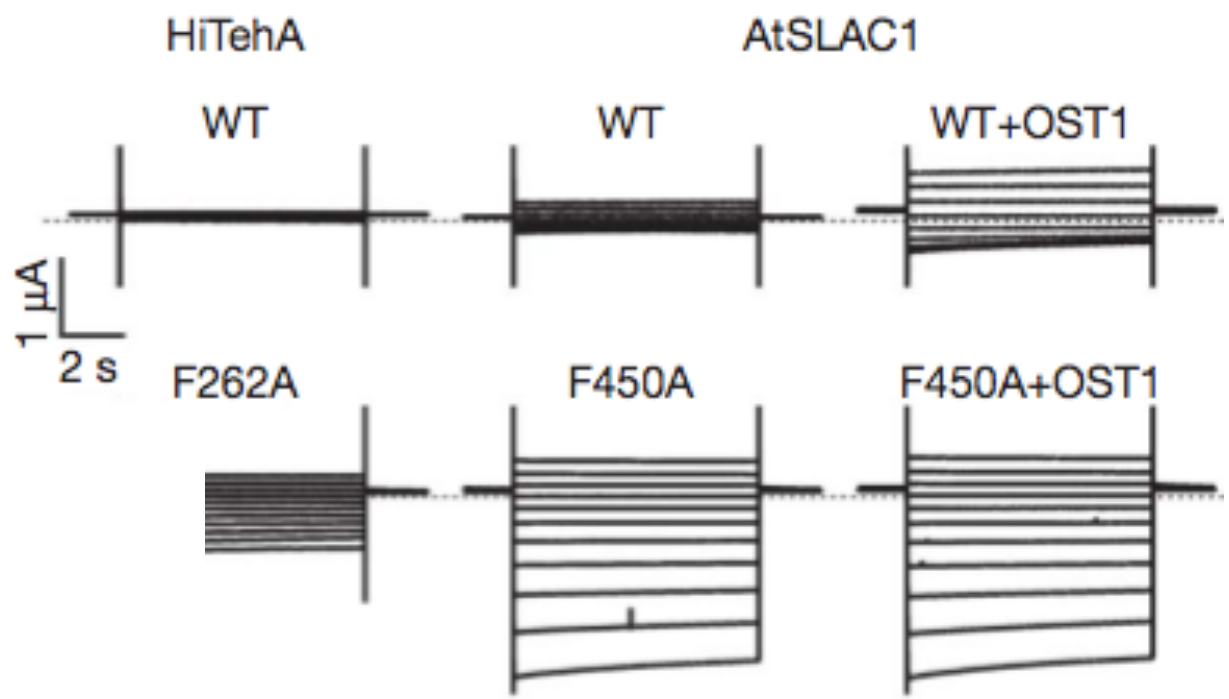
1. Actions-> Ribbon -> hide



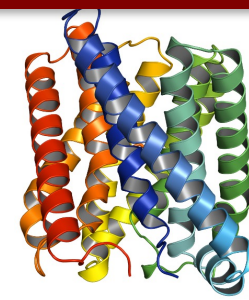




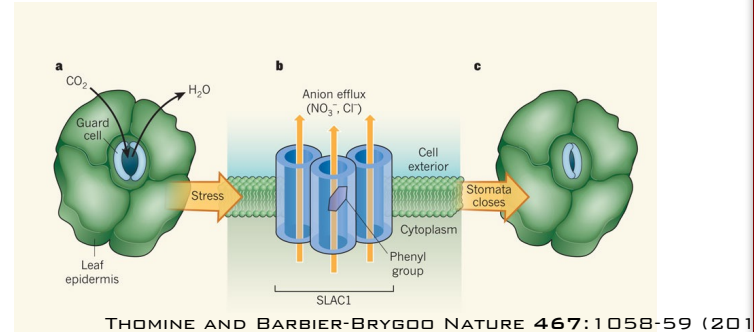




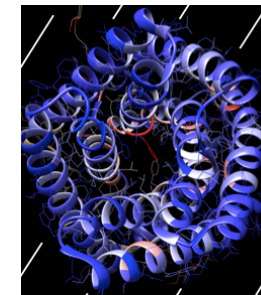
*H. influenzae* protein structure



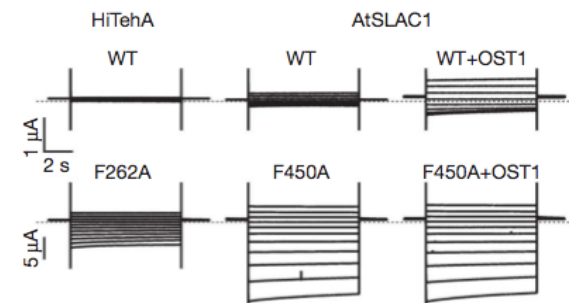
Functional hypothesis via homology to SLAC1



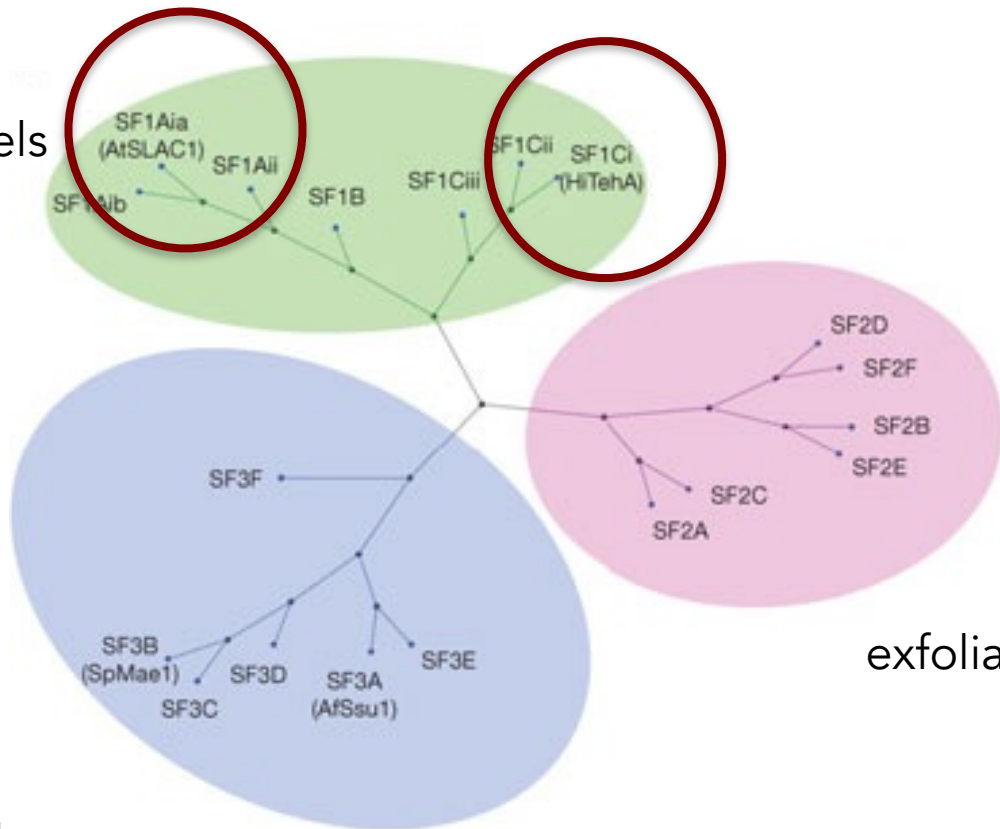
Identification potential functional residues using sequence conservation across the family and structural knowledge



Suggested experiments to test functional hypothesis



Anion channels



malate uptake transporter  
sulphite efflux pump

exfoliative toxins

# Exercise

Homology-based function annotation transfer #2

> more [Exercise\\_2/blast-2-seqs.link](#)

>cGMP-gated cation channel alpha-1

MKLSMKNNIINTQQSFVTMPNVIVPDIEKEIRRMENGACSSFSEDDDSASTSEESENE  
NP HARGSF SYKSLRKGGPSQREQYLPGAIALFNVNNSNKDQEP EKKKKKKKEKKS  
KSDDN ENKNDPEKKKKKKDKEKKKKKEEKSKDKKEEEKKEVVVIDPSGNTY  
YNWLF C I T L P V M Y N W T M V I A R A C F D E L Q S D Y L E Y W L I L D Y V S D I V Y L I D M F V R T R T G Y L E Q G L L V K E E L K L I N K Y  
K S N L Q F K L D V L S L I P T D L L Y F K L G W N Y P E I R L N R L L R F S R M F E F F Q R T E T R T N Y P N I F R I  
S N L V M Y I V I I I H W N A C V F Y S I S K A I G F G N D T W V Y P D I N D P E F G R L A R K Y V Y S L Y W S T L T L  
T T I G E T P P P V R D S E Y V F V V V D F L I G V L I F A T I V G N I G S M I S N M N A A R A E F Q A R I D A I K Q Y  
M H F R N V S K D M E K R V I K W F D Y L W T N K K T V D E K E V L K Y L P D K L R A E I A I N V H L D T L K K V R I F  
A D C E A G L L V E L V L K L Q P Q V Y S P G D Y I C K K G D I G R E M Y I I K E G K L A V V A D D G V T Q F V V L S D  
G S Y F G E I S I L N I K G S K A G N R R T A N I K S I G Y S D L F C L S K D D L M E A L T E Y P D A K T M L E E K G K  
Q I L M K D G L L D L N I A N A G S D P K D L E E K V T R M E G S V D L L Q T R F A R I L A E Y E S M Q Q K L K Q R L T  
K V E K F L K P L I D T E F S S I E G P G A E S G P I D S T



>mystery protein

MGNGSVKPKHSHKHPDGHSGNLTTDALRNKVTELERELRRKDAEIQEREYHLKELREQLSK  
QTVAIAELTEELQNKCIQLNKLQDVVHMQGGSPLOASPDKVPLEVHRKTSGLVSLHSRRG  
AKAGVSAEPTTRTYDLNKPPEFSFEKARVRKDSSEKKLITDALNKNQFLKRLDPQQIKDM  
VECMYGRNYQQGSYIIKQGEPGNHIFVLAEGRLEVFQGEKLLSSIPMWTTFGELAILYNC  
TRTASVKAITNVKTWALDREVFQONIMRRTAQARDEQYRNFLRSVSLLKNLPEDKLTKIID  
CLEVEYYDKGDYIIREGEEGSTFFILAKGKVKVTQSTEGHDQPQLIKTLQKGEYFGEKAL  
ISDDVRSANIIAEENDVACLVIDRETFNQTVGTFEELQKYLEGYVANLNRDDEKRHAKRS  
MSNWKLSKALSLEMIQLKEKVARFSSSSPFQNLLEIIATLGVGGFGRVELVKVKNENVAFA  
MKCIRKKHIVDTKQQEHVYSEKRILEELCSPFIVKLYRTFKDNKYVYMLLEACLGGELWS  
ILRDRGSFDEPTSKFCVACVTEAFDYLHRLGIIYRDLKPENLILDAEGYLKLVDFGFAKK  
IGSGQKTWTFCGTPEYVAPEVILNKGHDFSVDVFWSLGILVYELLTGNPPFSGVDQMMTYN  
LILKGIEKMDFPRKITRRPEDLIRRLCRQNPTERLGNLKNGINDIKKHRWLNGFNWEGLK  
ARSLPSPLQRELKGPIDHSYFDKYPPEKGMPPDELSGWDKDF

## Align Sequences Protein BLAST

[blastn](#)**[blastp](#)**[blastx](#)[tblastn](#)[tblastx](#)BLASTP programs search protein subjects using a protein query. [more...](#)

## Enter Query Sequence


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

```
TTIGETPPPVRDSEYVFFVVVDFLIGVLIFATIVGNIGSMISNMNAARAEFQARIDAIKQY
MHFRNVSKDMEKRVIKWFYDLWTNKKTVDEKEVLKYLDPDKLRAEIAINVHLDTLKKVRI
ADCEAGLLVELVLKLPQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDGVTQFVVVLS
DSYFGEISILNIKSGKAGNRRTANIKSIGYSDLFCLSKDDLMEALTEYPDAKTMLEEK
GKQILMKDGLLDLNIANAGSDPKDLEEKVTRMEGVSVDLLQTRFARILAEYESMQQK
LQRLTKVEKFLKPLIDTEFSSIEGPGAESGPIDST
```


From

To

Or, upload file

Browse... No file selected. 

Job Title

Enter a descriptive title for your BLAST search  Align two or more sequences 

## Enter Subject Sequence


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

```
MSNWKLSKALSLEMIQLKEKVARFSSSSPFQNLIIATLGVGGFGRVELVKVKNENVAFA
MKCIRKKHIVDTKQQEHVYSEKRILEELCSPIVKLYRTFFKDNKYVYMLLEACLGGELWS
ILRDRGSFDEPTSKFCVACVTEAFDYHLHRLGIIYRDLKPENLILDAEGYLKLVDFGFAK
IGSQKTWTFCGTPEYVAVEVILNKGHDFSVDFWSLGLVYELLTGNPPFSGVDQMMTYN
LILKGIKMDFPKIRTRRPEDLIRRLCRQNPTERLGNLKNKINDIKKRWLNGFNWEG
LARSLSPLQRELKGPIDHSYFDKYPPEKGMPPDELSEGWDKDF
```

From

To

Or, upload file

Browse... No file selected. 

## Alignments

Download [Graphics](#) Sort by:

unnamed protein product

Sequence ID: lcl|Query\_22995 Length: 762 Number of Matches: 3

Range 1: 281 to 383 [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
41.2 bits(95)	1e-07	Compositional matrix adjust.	30/110(27%)	54/110(49%)	11/110(10%)

Query	474	LKKVRIFADCEAGLLVELVLKLPQVYSPGDYICKKGDIGREMYIIKEGKLAVV---- <td>529</td>	529
		L+ V + + L +++ L+ + Y GDYI ++G+ G +I+ +GK+ V	
Sbjct	281	LRSVSLKLNLPEDKLTKIIDCLEVEYYDKGDYIIREGEEGSTFFILAKGKVKVTQSTEGH	340
Query	530	DGVTQFVVLSDGSYFGEISILNIKSGKAGNRRRTANIKSIGYSDLFCLSKD	579
		D L G YFGE +++ + + R+ANI + +D+ CL D	
Sbjct	341	DQPQLIKTLQKGEYFGEKALI-----SDDVRSANIIA-EENDVACLVID	383

Range 2: 161 to 260 [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Method	Identities	Positives	Gaps
38.1 bits(87)	1e-06	Compositional matrix adjust.	26/108(24%)	53/108(49%)	8/108(7%)

Query	472	DTLKKVRIFADCEAGLLVELVLKLPQVYSPGDYICKKGDIGREMYIIKEGKLAVVADDG	531
		D L K + + + ++V + + Y G YI K+G+ G ++++ EG+L V +	
Sbjct	161	DALNKNQFLKRLDPQIQKDMVECMYGRNYQQGSYIIKQGEPGNHIFVLAEGRLEVFQGEK	220
Query	532	VTQFVVLSDGSYFGEISILNIKSGKAGNRRRTANIKSIGYSDLFCLSKD	579
		+ + + + FGE++IL RTA++K+I + L ++	
Sbjct	221	LLSSIPM--WTTFGELAIL-----YNCTRTASVKAITNVKTWALDRE	260

Range 3: 593 to 649 [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

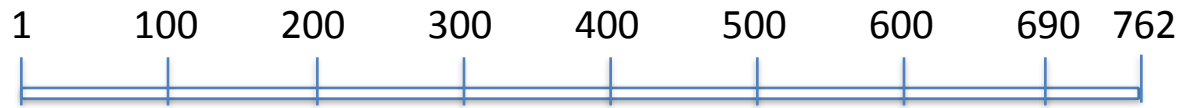
Score	Expect	Method	Identities	Positives	Gaps
22.3 bits(46)	0.081	Compositional matrix adjust.	17/58(29%)	27/58(46%)	7/58(12%)

Query	317	VFYSISKAIGFGNDTWVY---PDINDPEFGRLARKYVYSL-YWS--TLTLTTIGETPP	368
		V + +K IG G TW + P+ PE L + + +S+ +WS L + PP	
Sbjct	593	VDFGFAKKIGSGQKTWTF CGTPEYVAPEV-ILNKGHDFSVDFWSL GILVYELLTGNPP	649

Mystery protein is a cGMP-gated cation channel?

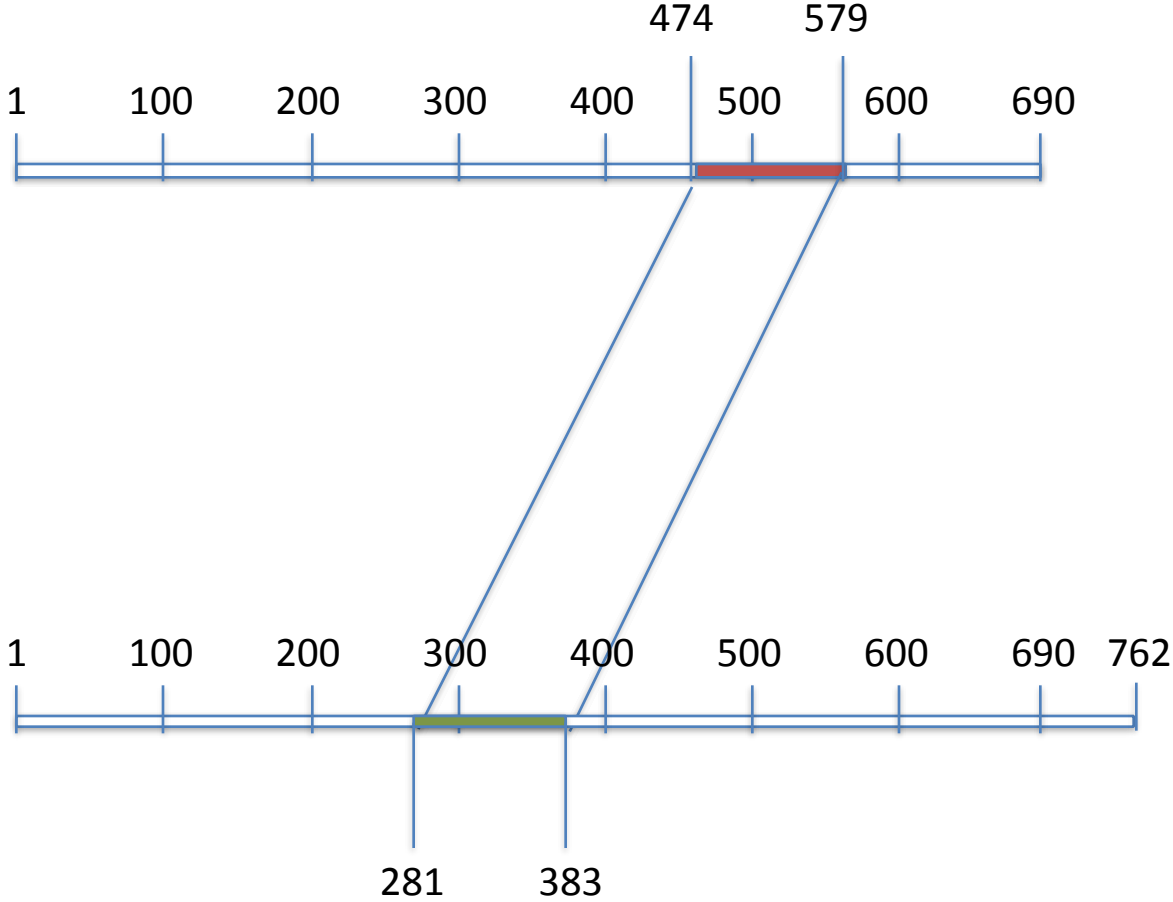
cGMP-gated cation channel alpha-1

P29973 (CNGA1\_HUMAN)



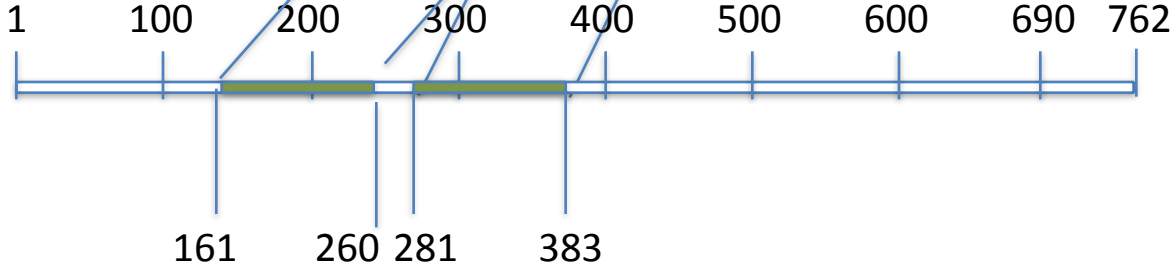
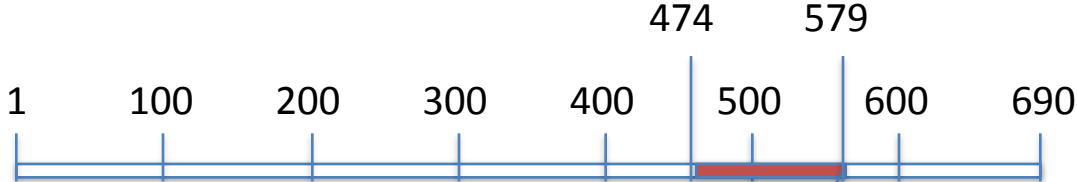
cGMP-gated cation channel alpha-1

P29973 (CNGA1\_HUMAN)



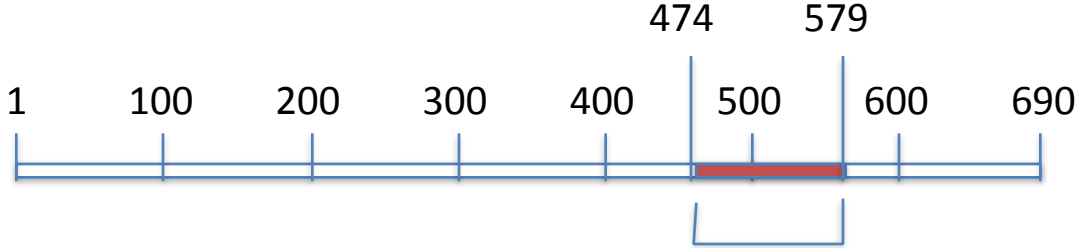
cGMP-gated cation channel alpha-1

P29973 (CNGA1\_HUMAN)

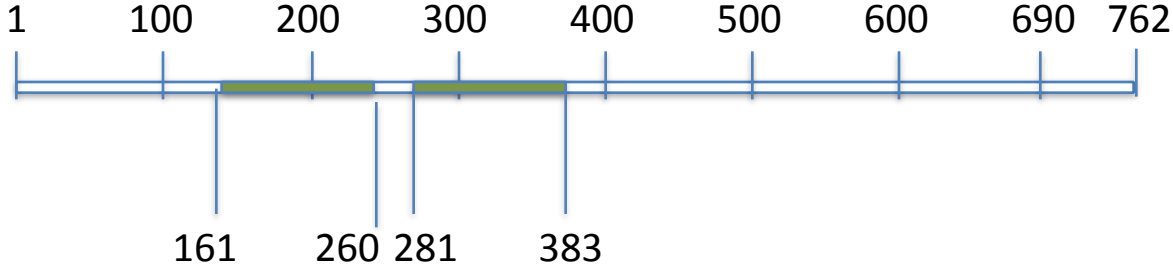


cGMP-gated cation channel alpha-1

P29973 (CNGA1\_HUMAN)



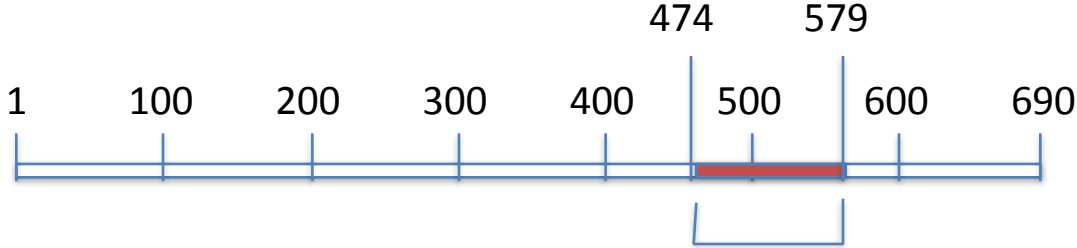
cGMP binding domain



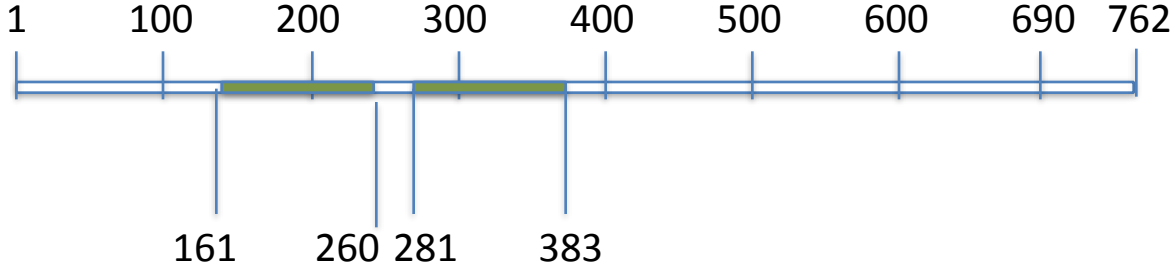


cGMP-gated cation channel alpha-1

P29973 (CNGA1\_HUMAN)

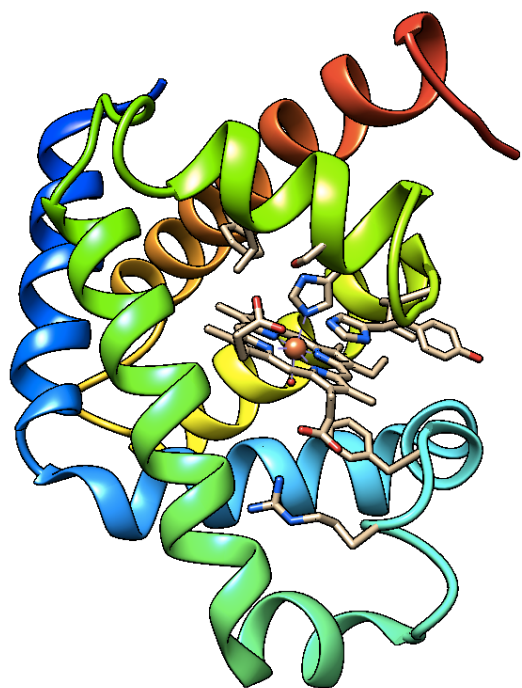


cGMP binding domain

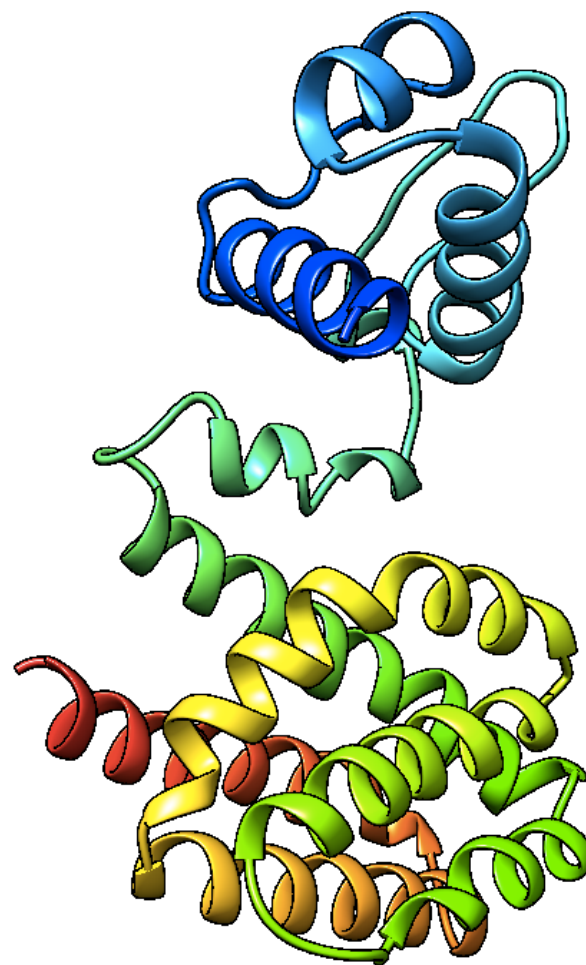


cGMP binding domains?

Mystery protein is a cGMP-dependent protein kinase 2  
Q13237 (KGP2\_HUMAN)



1MBN



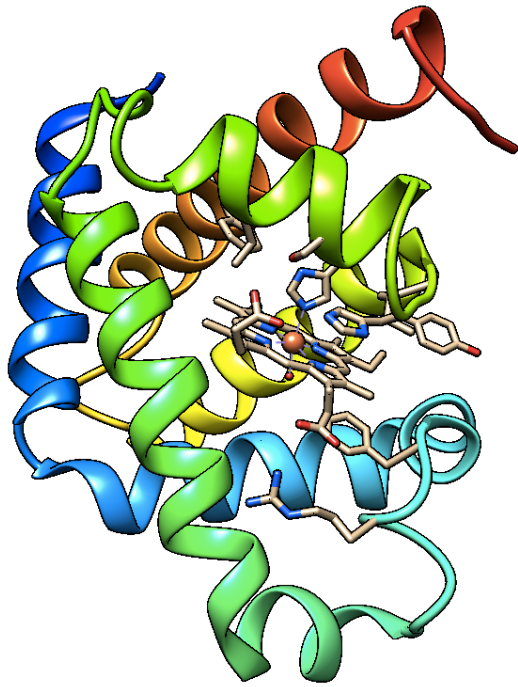
1HW2

Colour Scheme:

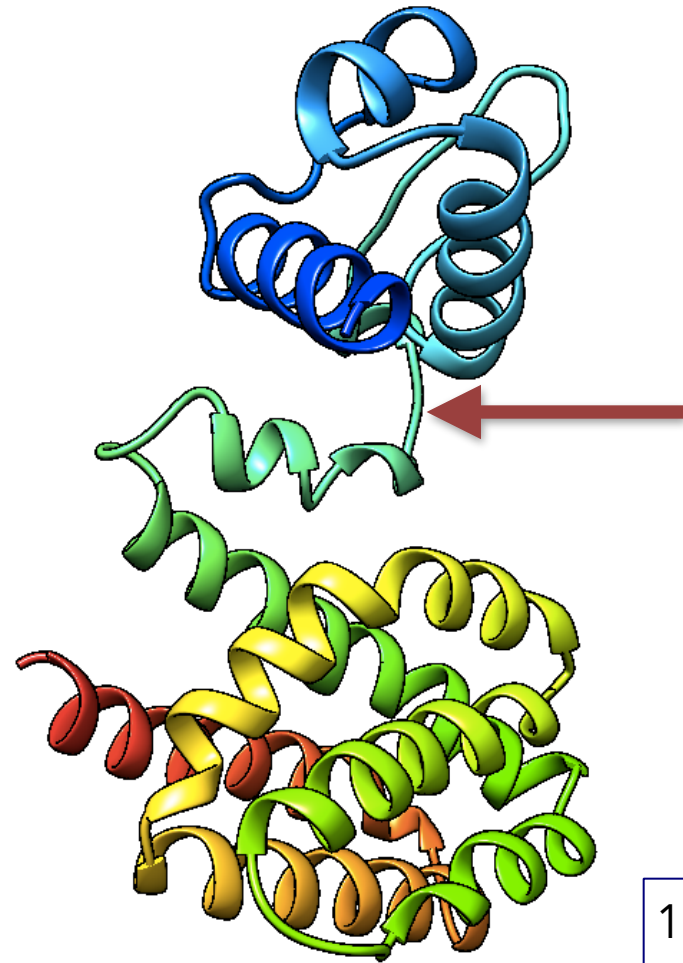


N'

C'



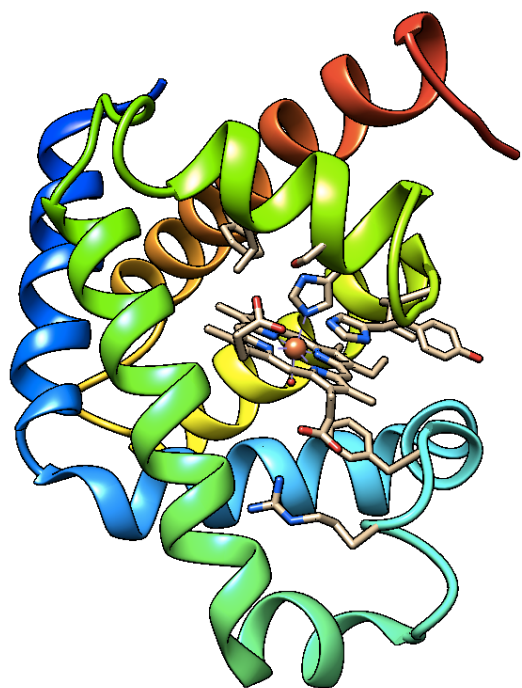
1MBN



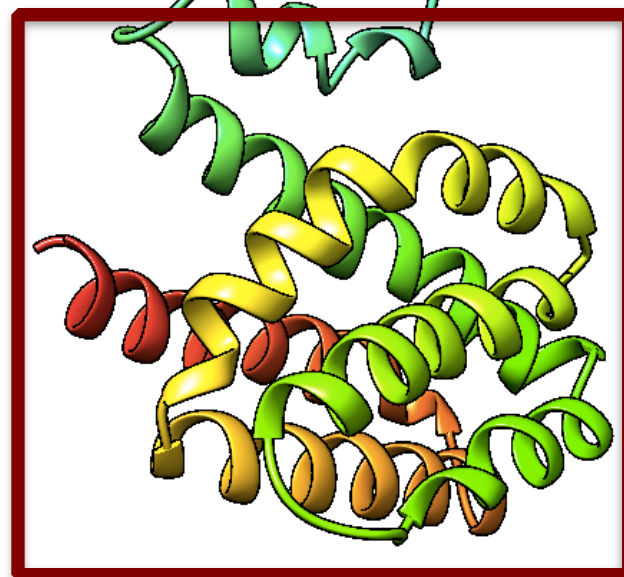
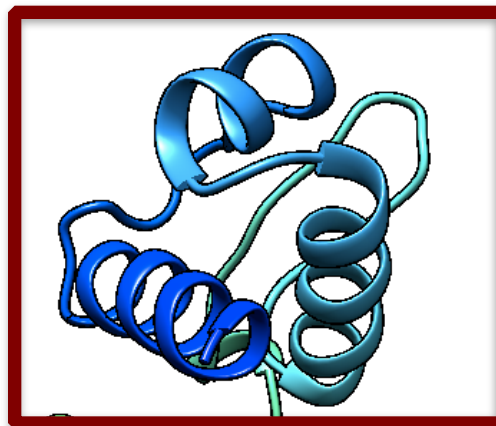
1HW2

Colour Scheme:





1MBN



1HW2

Colour Scheme:



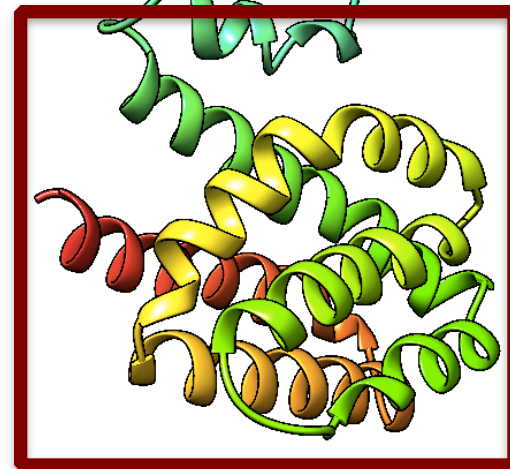
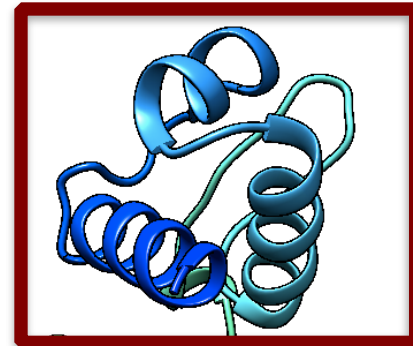
N'

C'

Definition (Wikipedia):

A protein domain is a conserved part of a given protein sequence and structure that can evolve, function, and exist independently of the rest of the protein chain. A domain forms a compact three-dimensional structure and often can be independently stable and folded.

(Marco): in proteins individual domains can be combined to perform complex functions.

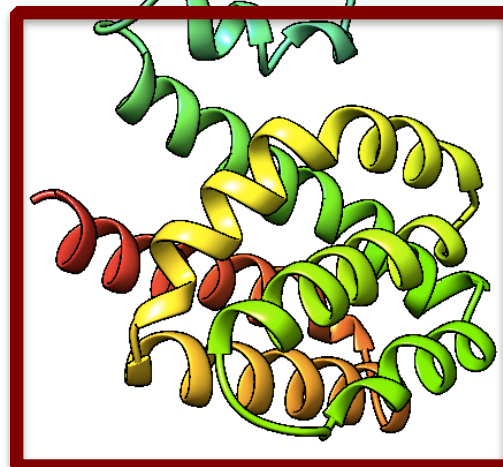
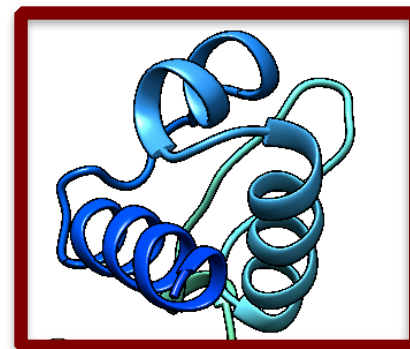
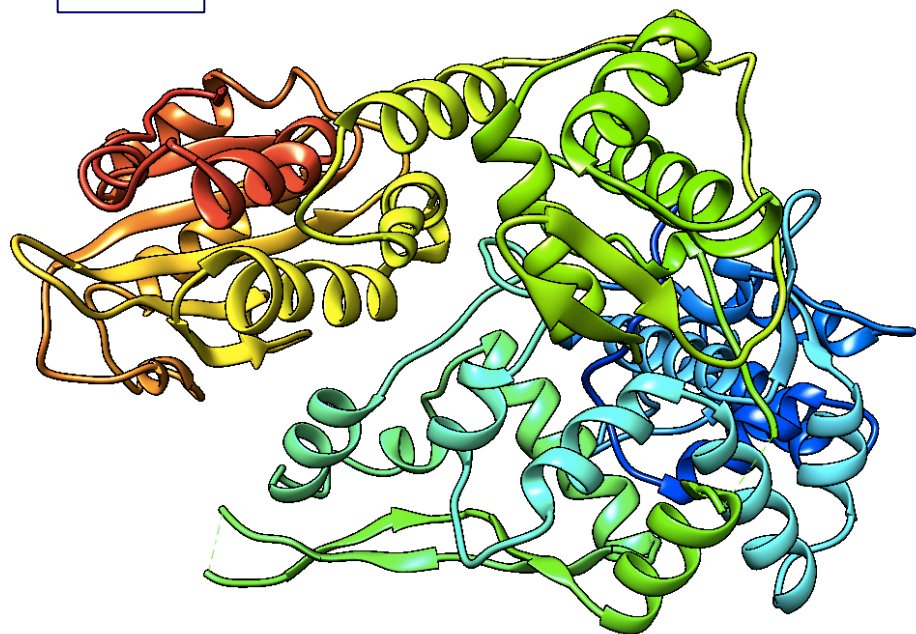


1HW2

Colour Scheme:



1FOK



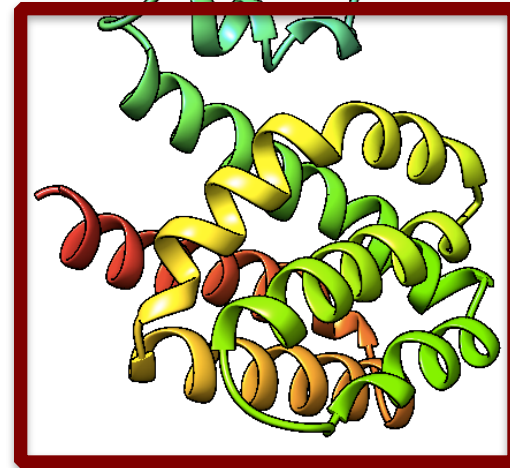
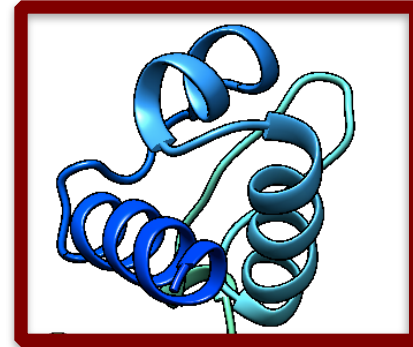
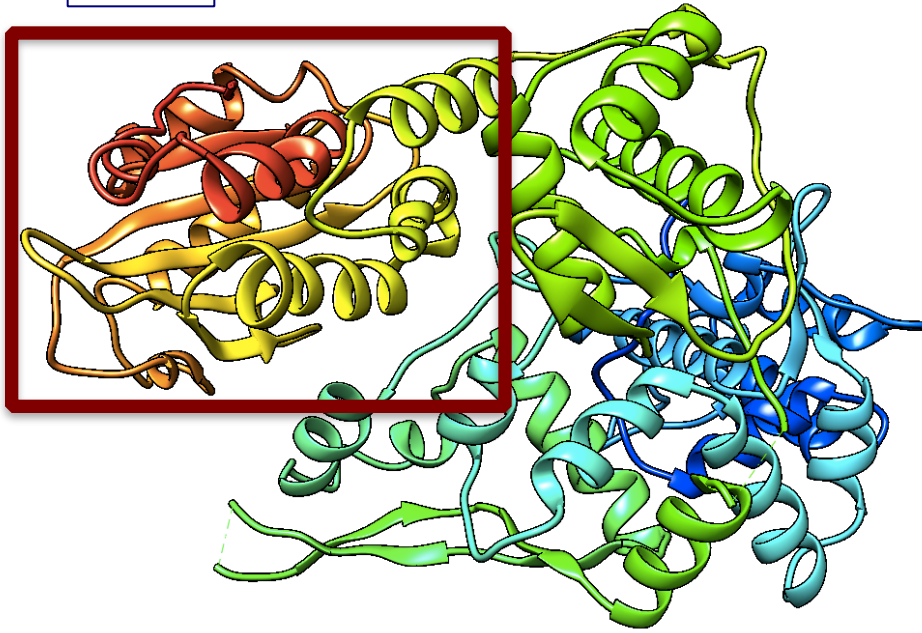
1HW2

Colour Scheme:





1FOK



1HW2

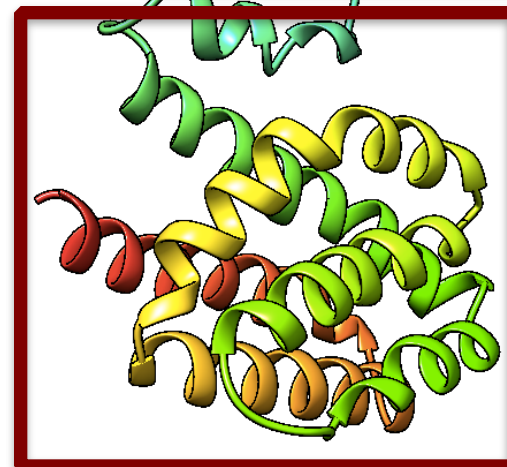
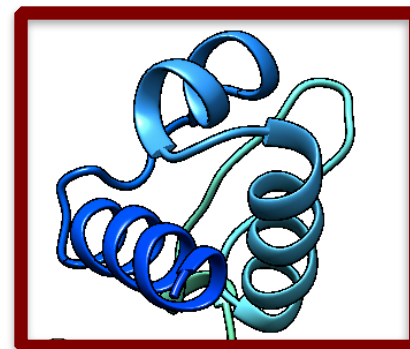
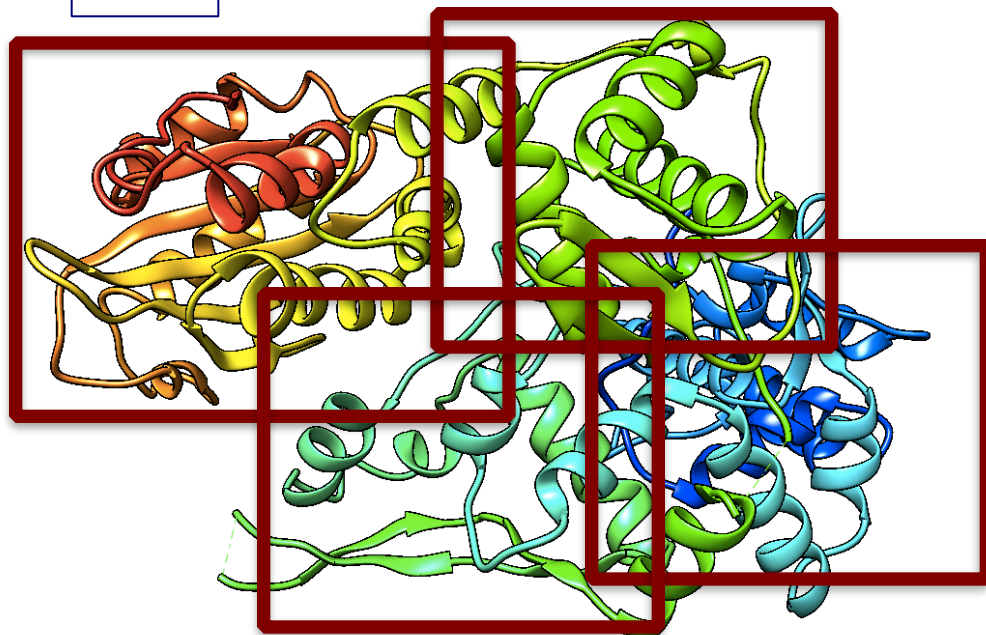
Colour Scheme:



N'

C'

1FOK



1HW2

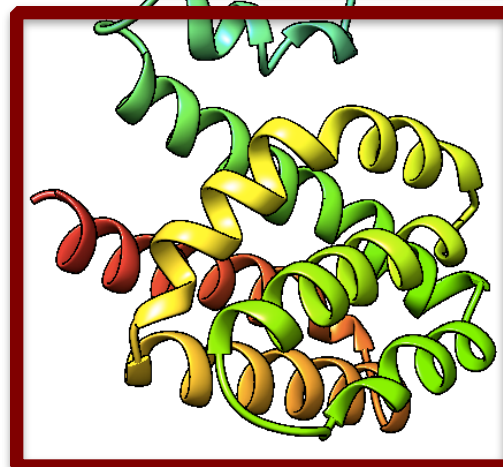
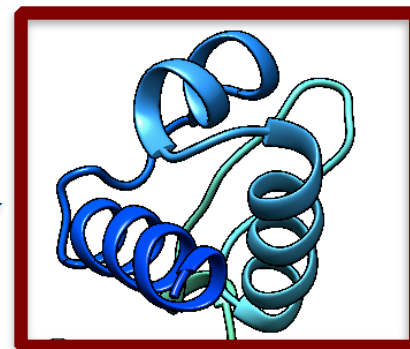
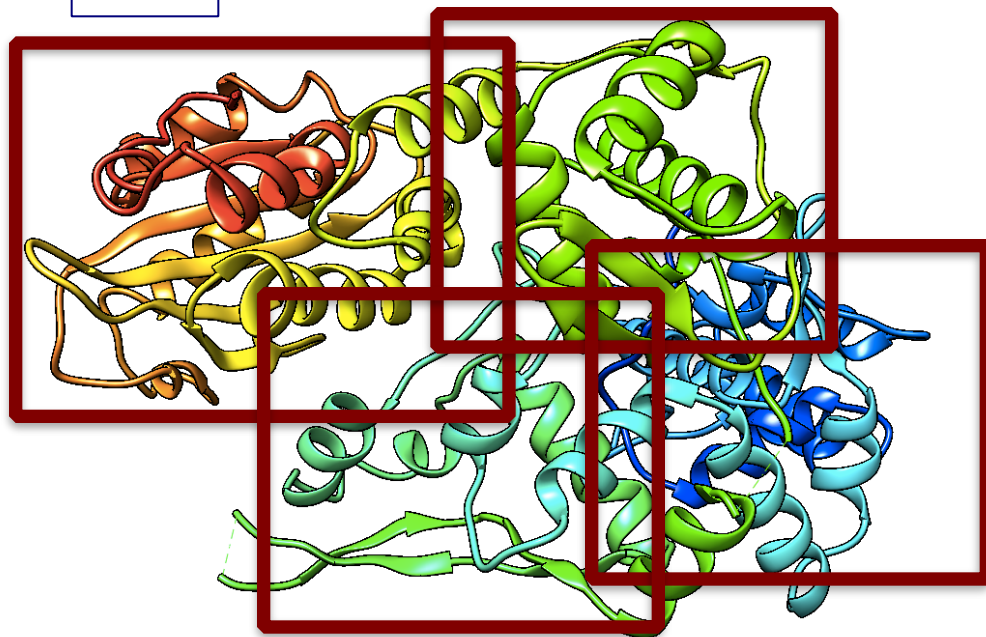
Colour Scheme:



N'

C'

1FOK

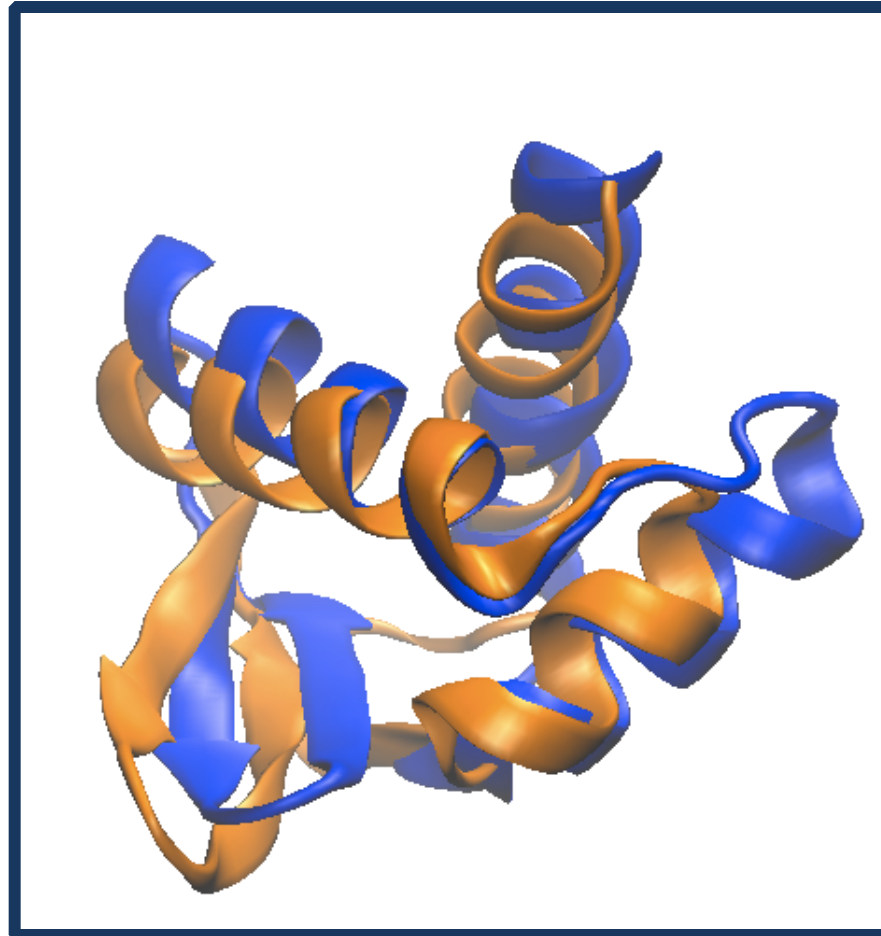


1HW2

Colour Scheme:

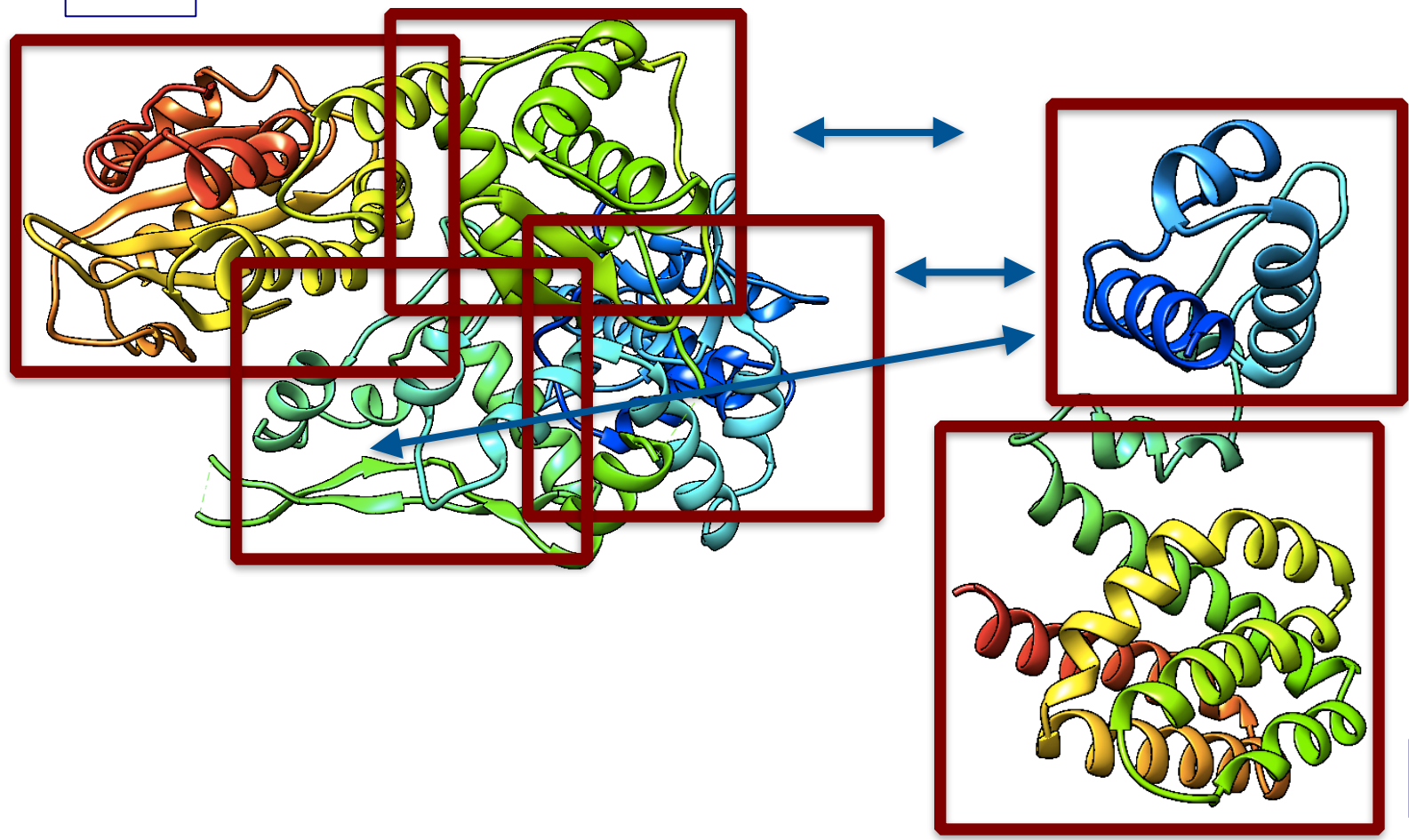


## Winged helix domain (WHD)



Z-score = 4.0  
%id = 8%  
RMSD = 2.7Å

1FOK

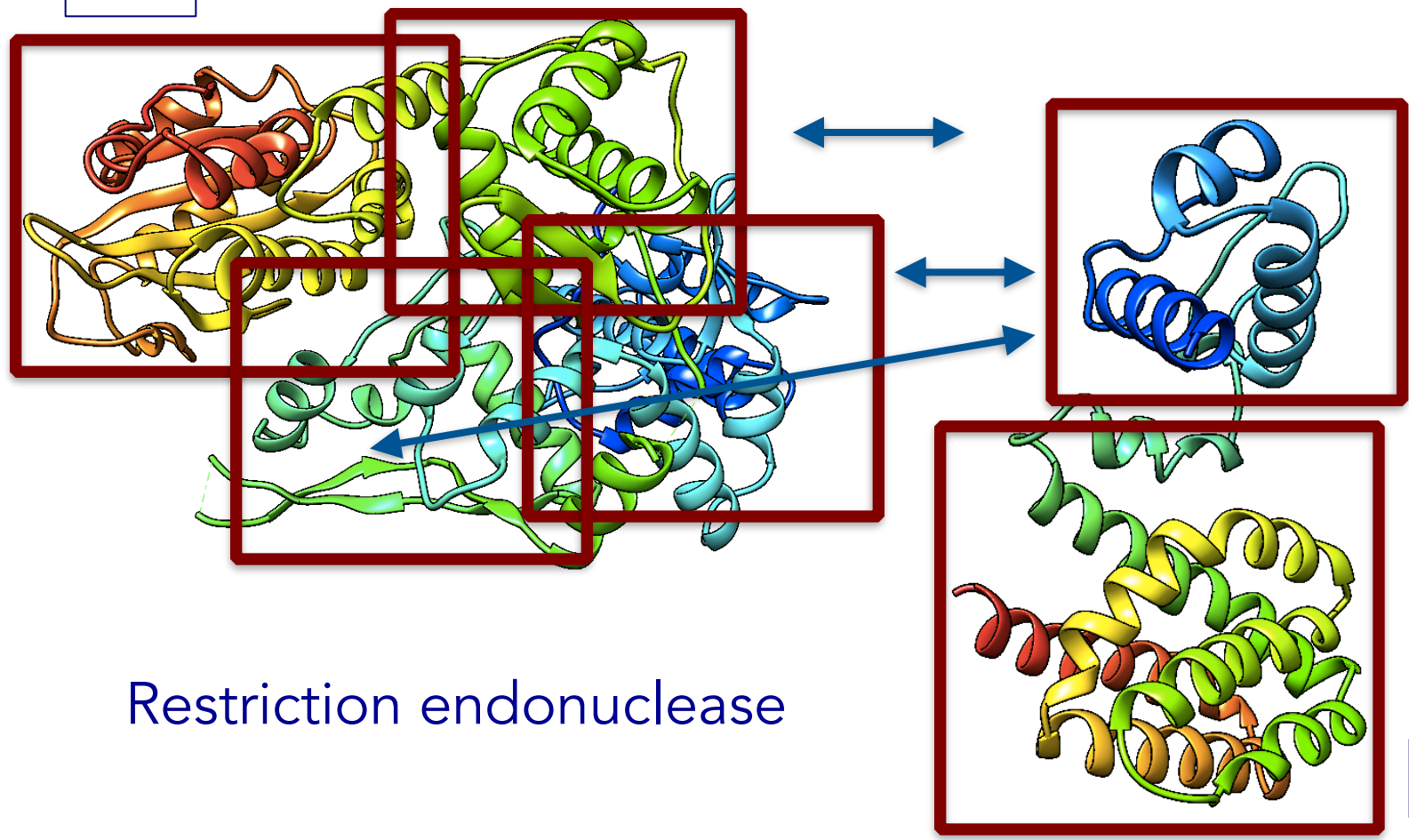


1HW2

Colour Scheme:



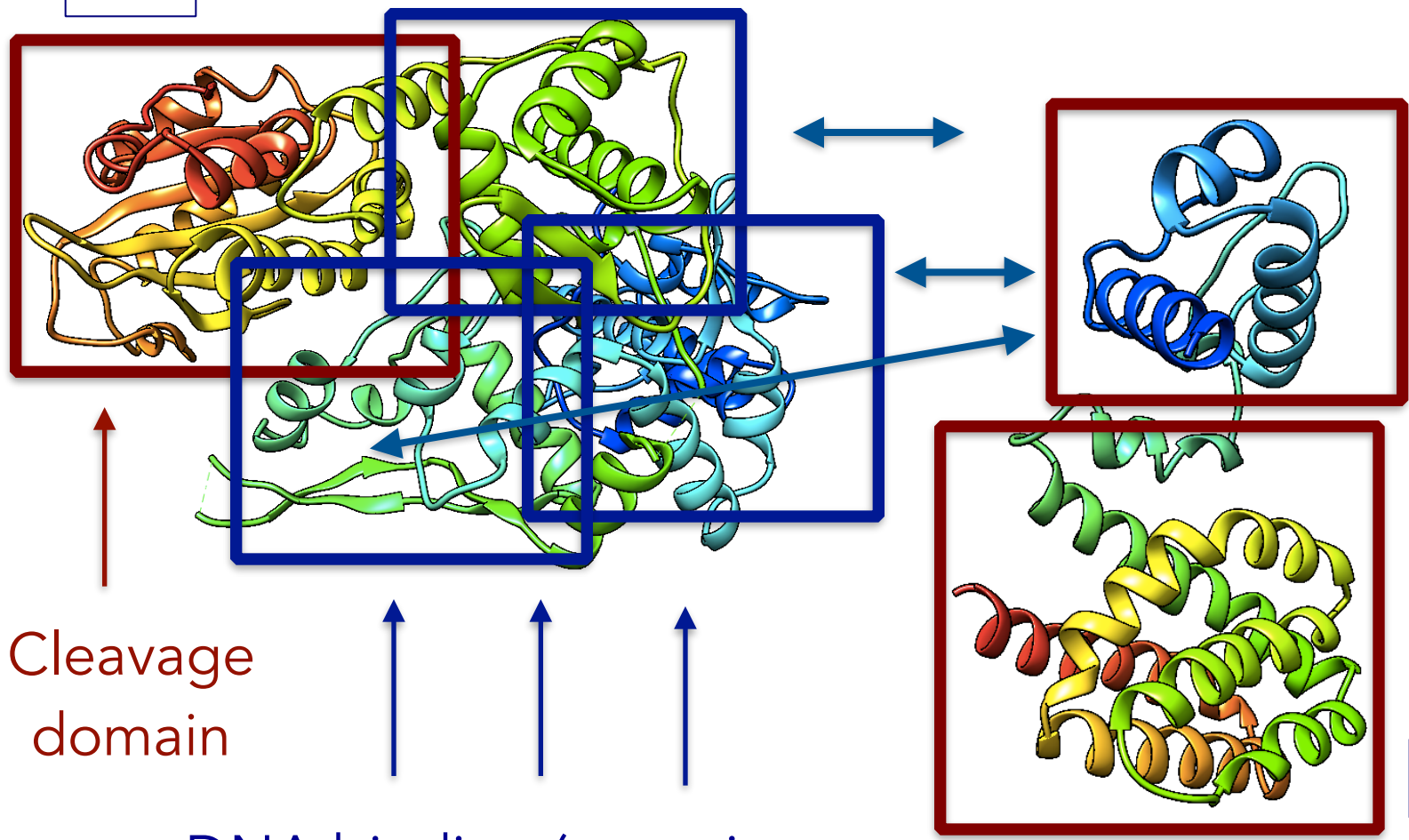
1FOK



1HW2

Restriction endonuclease

1FOK



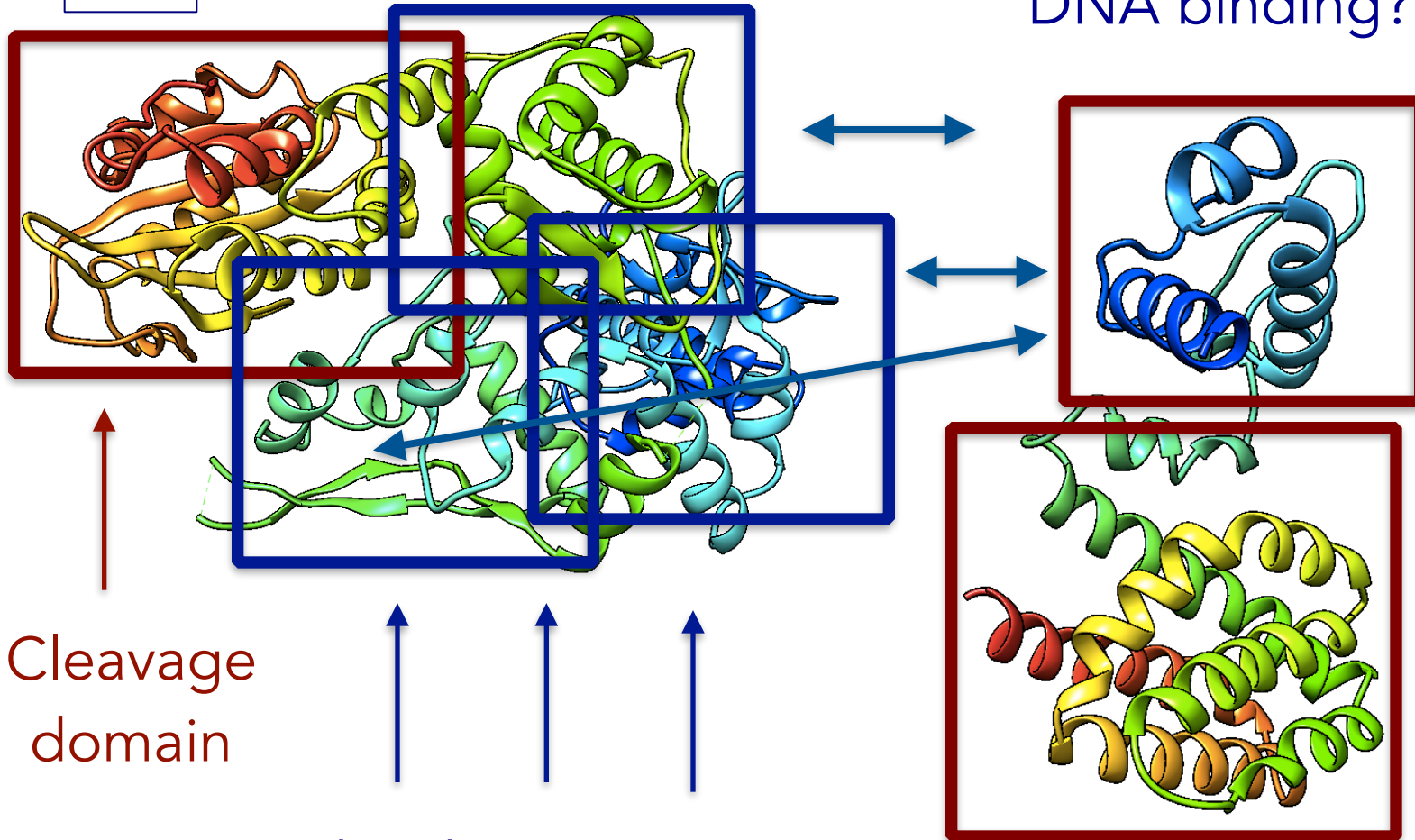
Cleavage domain

DNA binding (targeting to a specific DNA sequence)

1HW2

1FOK

DNA binding?



Cleavage domain

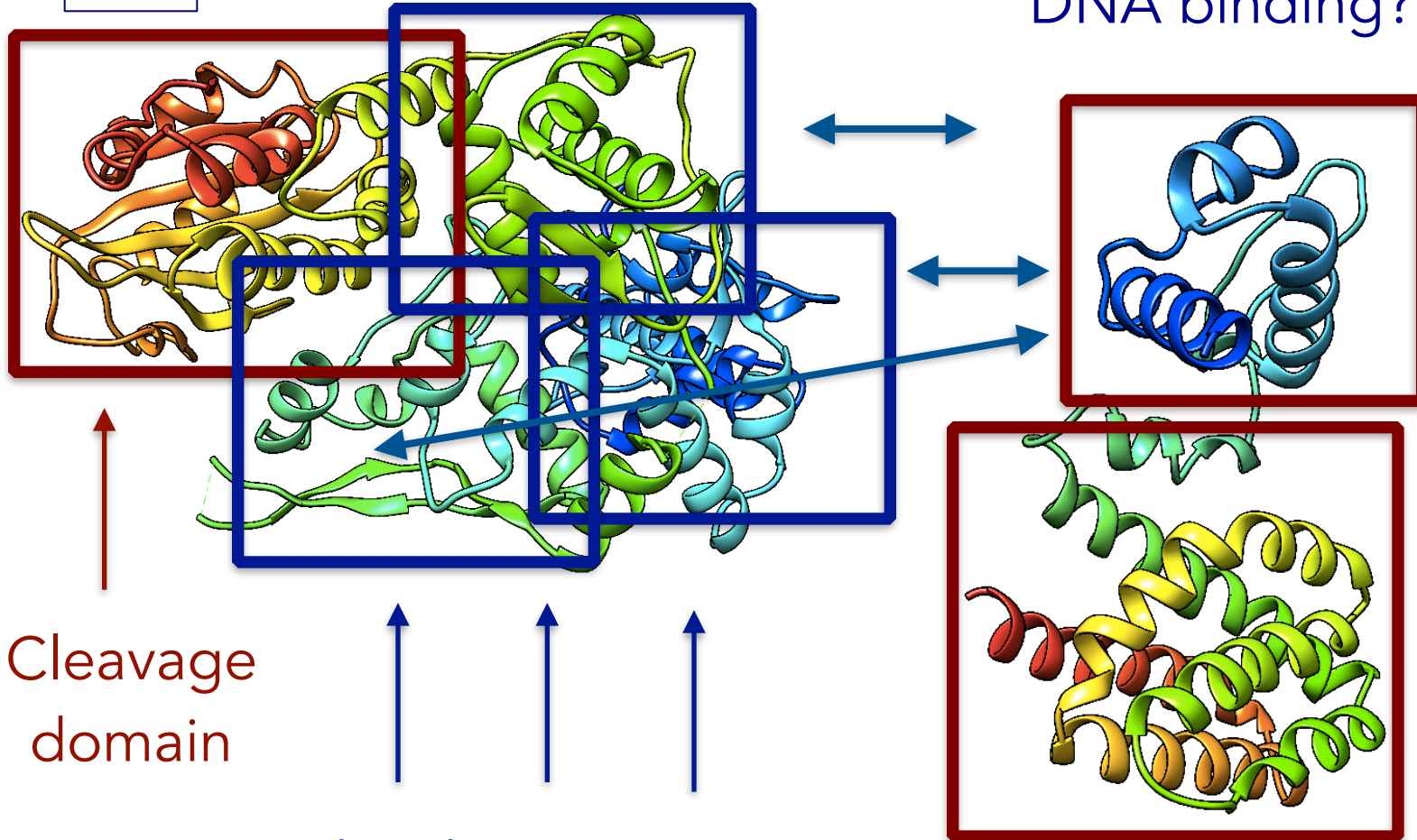
DNA binding (targeting to a specific DNA sequence)

1HW2



1FOK

DNA binding?



Cleavage domain

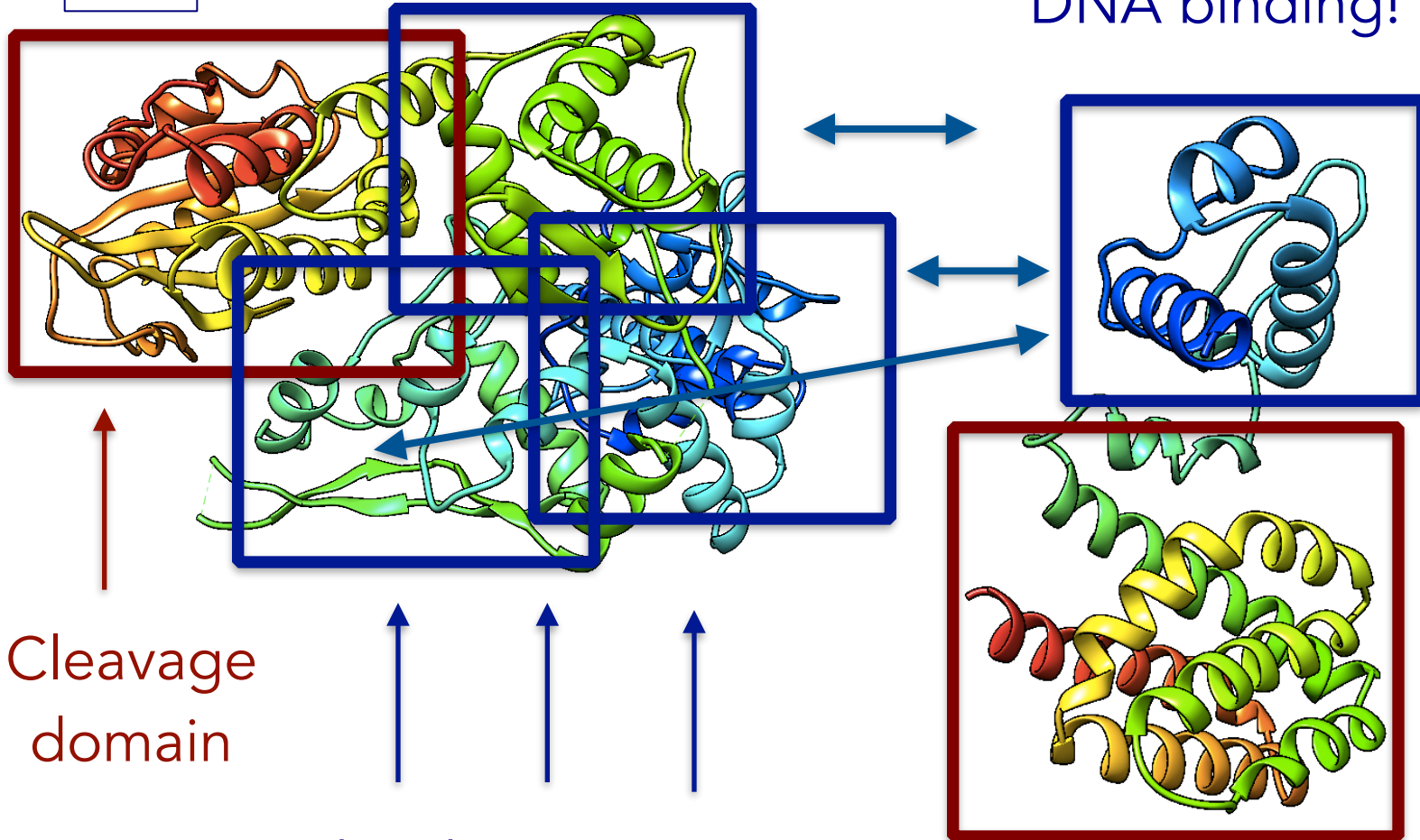
DNA binding (targeting to a specific DNA sequence)

1HW2

?

1FOK

DNA binding!



Cleavage domain

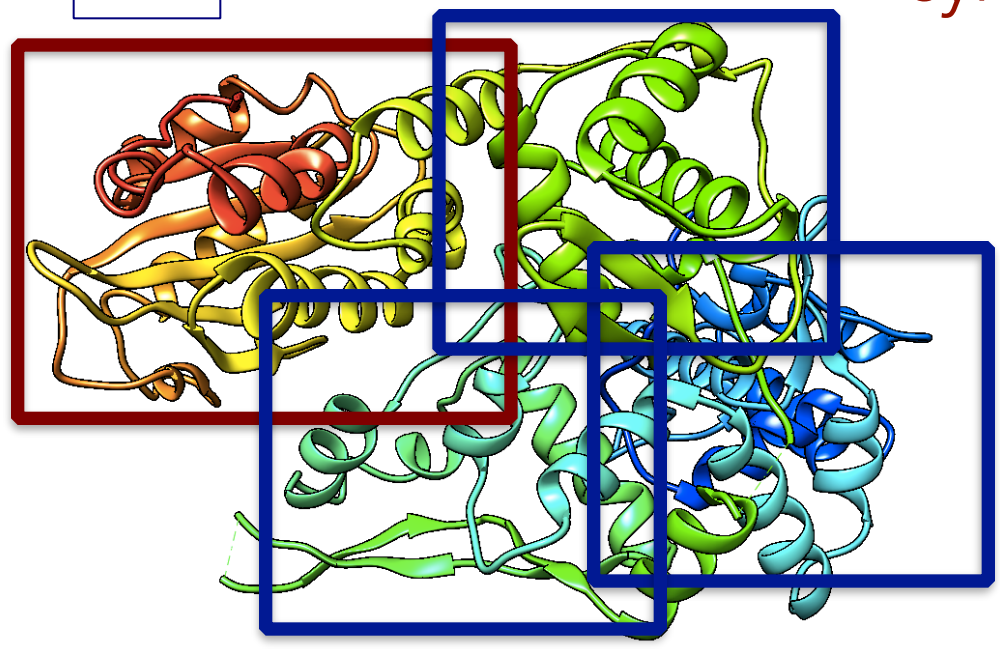
DNA binding (targeting to a specific DNA sequence)

1HW2

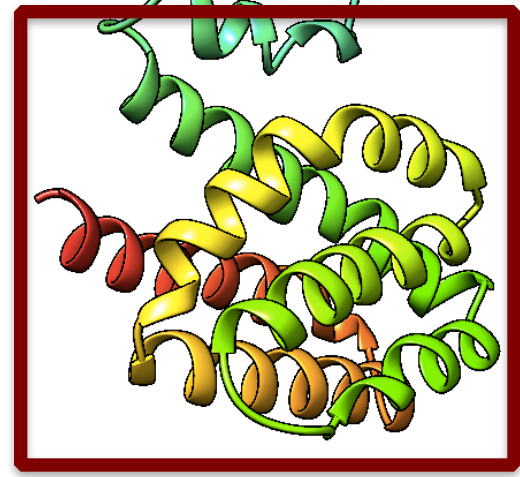
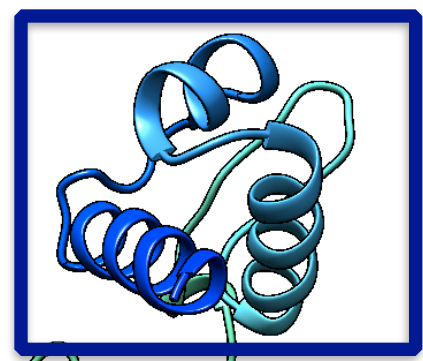
acyl-CoA binding domain controls affinity

1FOK

“syntactical change”



Restriction endonuclease



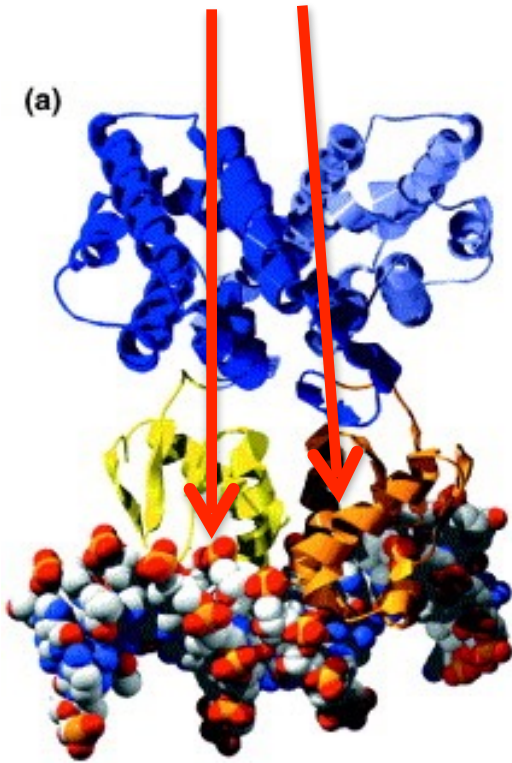
1HW2

Transcription factor

# Semantic change

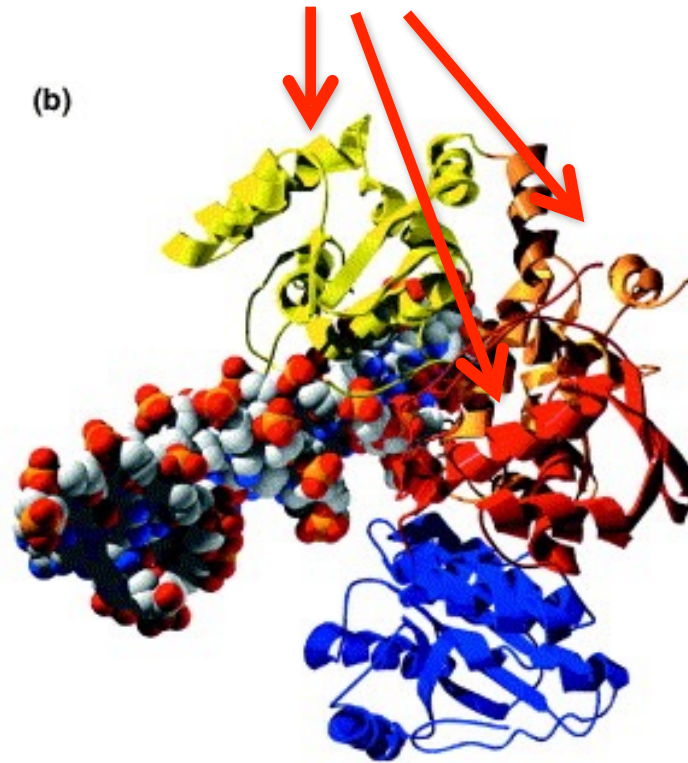
Marco Punta

DNA binding



Transcription  
factor

DNA binding



Restriction  
endonuclease

substrate specificity pocket



Human methionine  
aminopeptidase 2

Current Opinion in Structural Biology

# “syntactical change”

DNA sequence  
recognised

Restriction endonuclease

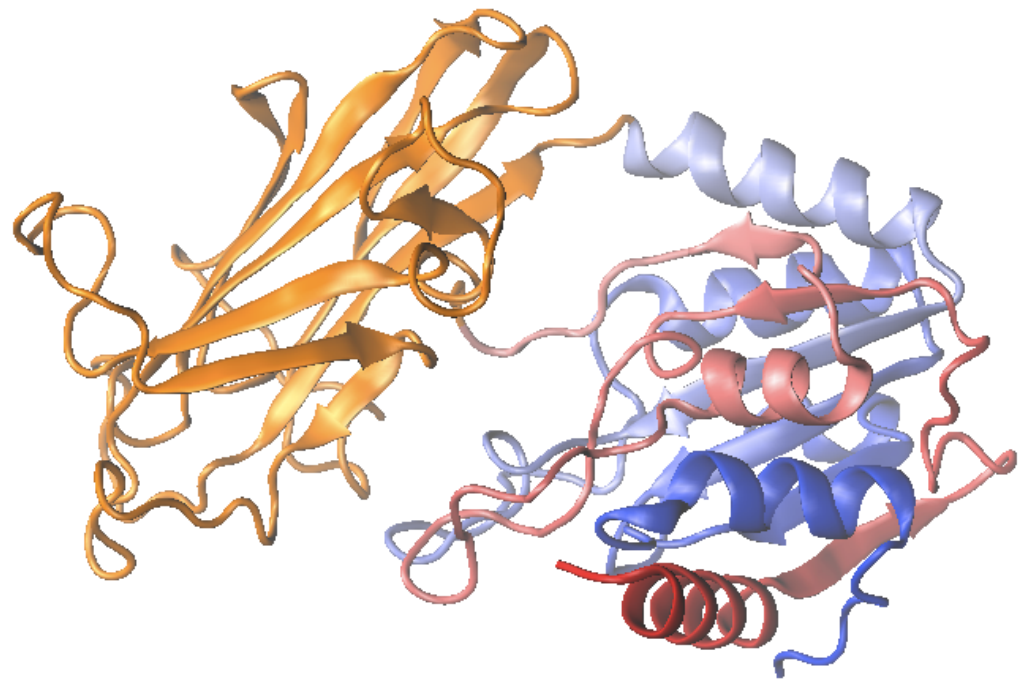
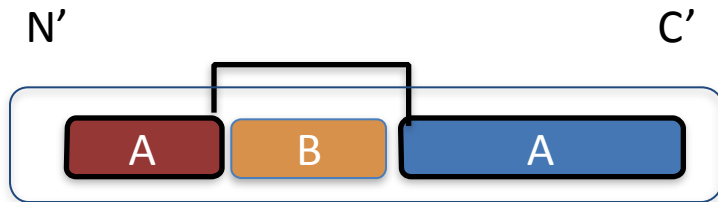
5'-GGATG-3'

Transcription factor

5'-TGGNNNNCCA-3'

# “Nested” domains

Marco Punta



3ABZ-truncated

# Function annotation transfer by homology

Homologous proteins may share a number of functional features, however:

- functional drift can lead to radically different functions
- while functional similarity correlate with function, no similarity threshold is safe for transfer
- if more than one functional domain is present annotation transfer can be attempted only between domains that are homologous and NOT for the full-length protein function

# Protein families

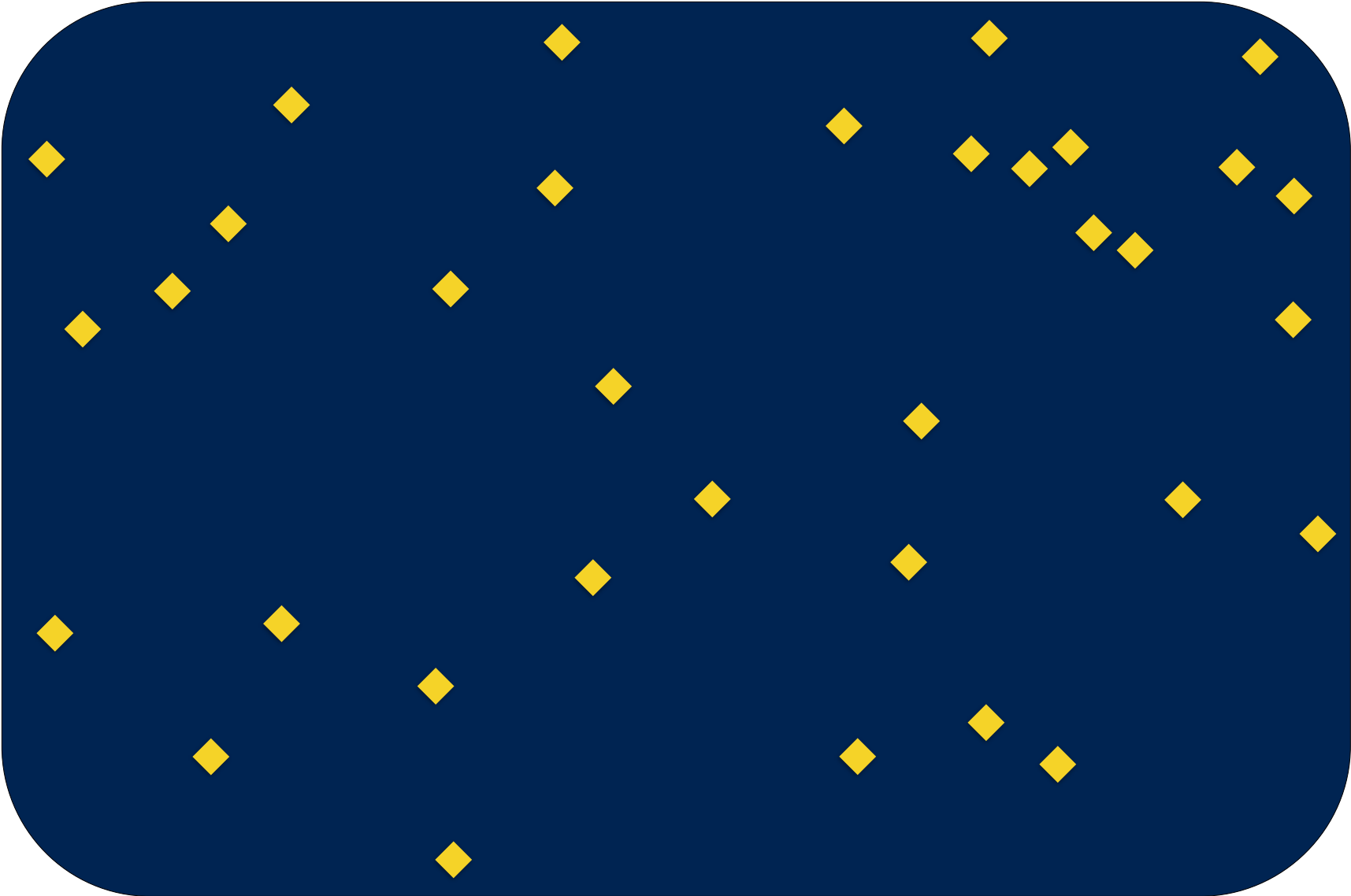
- Members will be structurally similar
- Members may share aspects of function
- Also, the whole set of members may reveal elements of protein and organism evolution



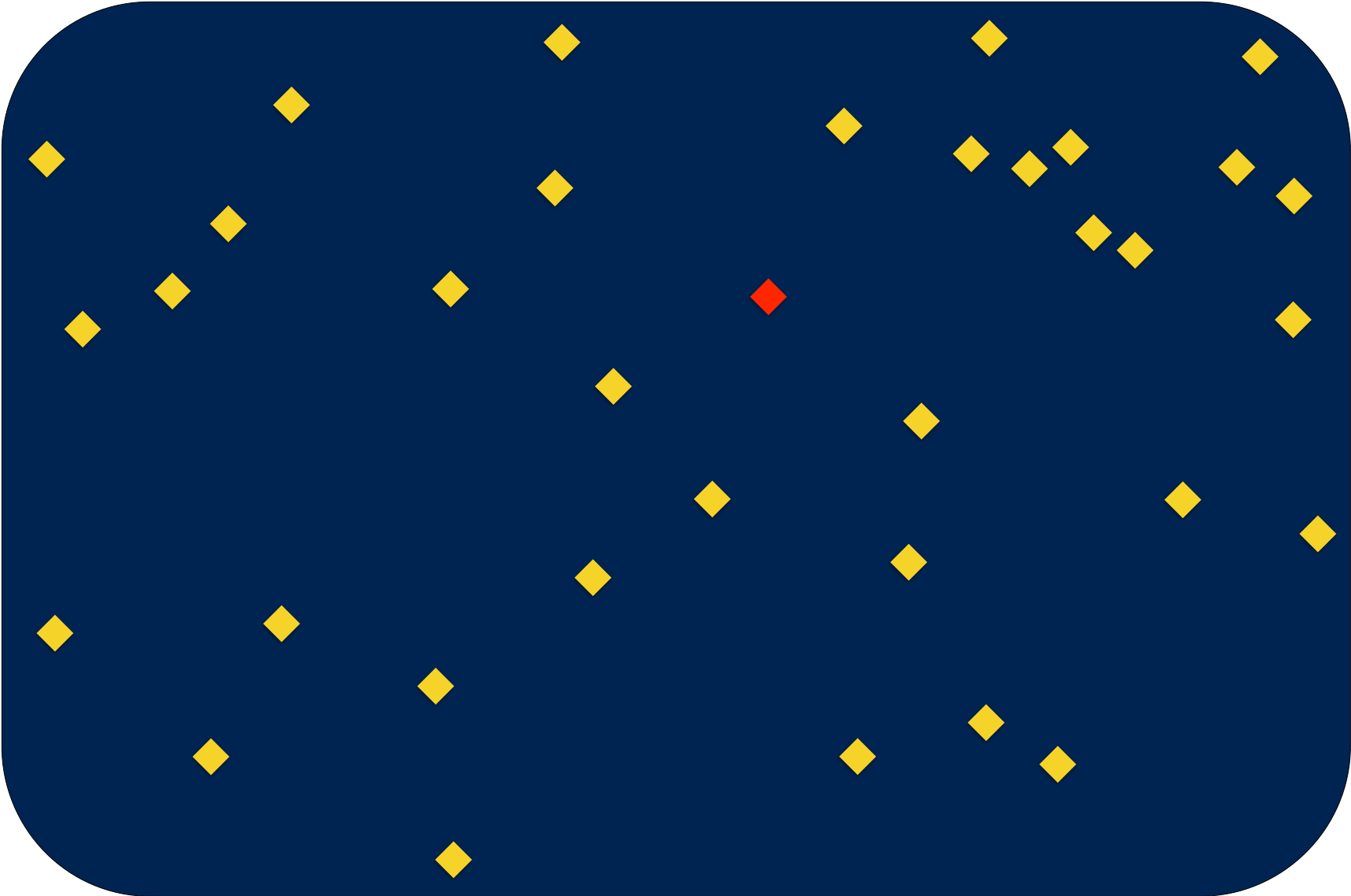
# The sequence space



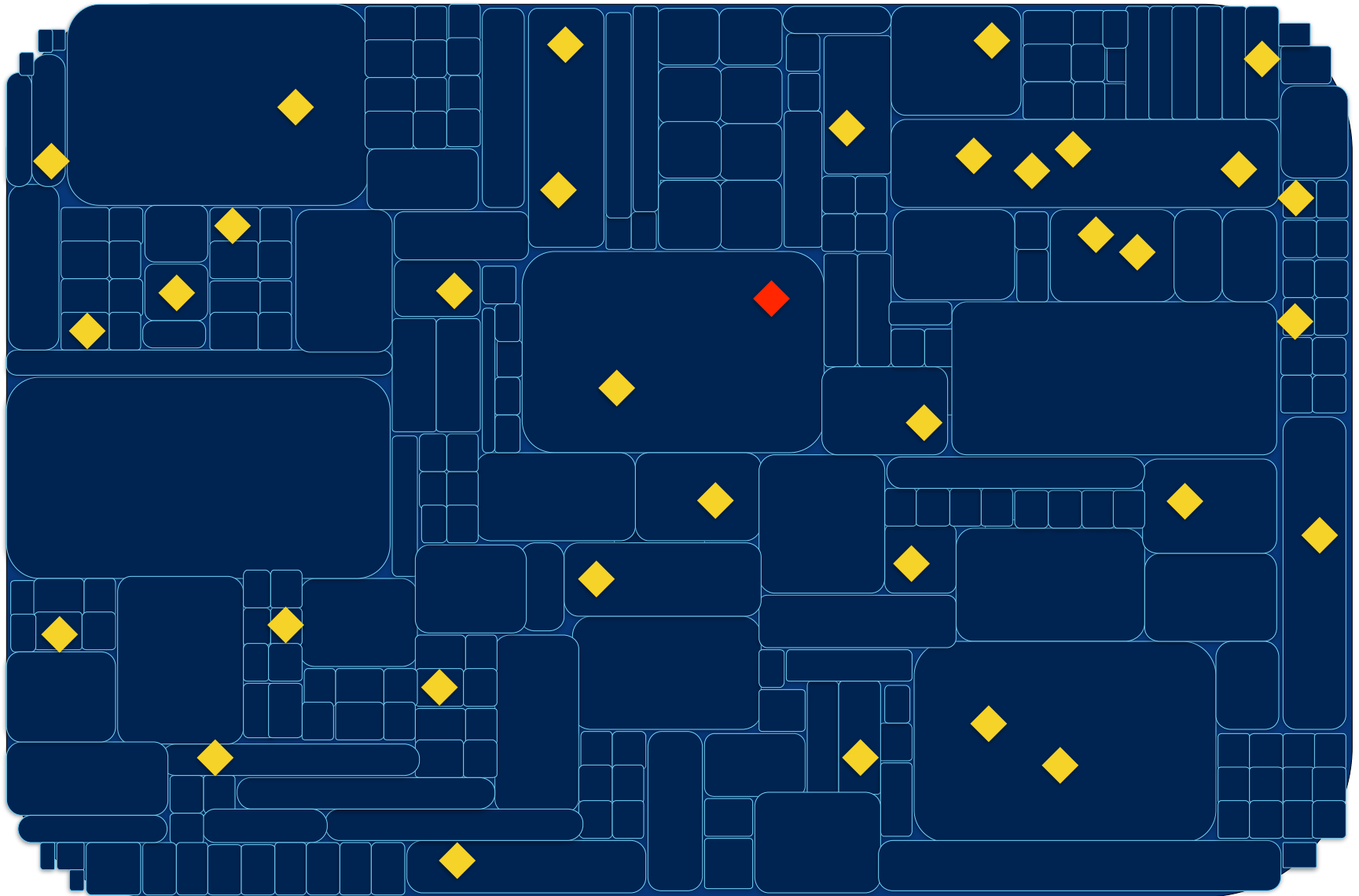
# The sequence space and annotated proteins



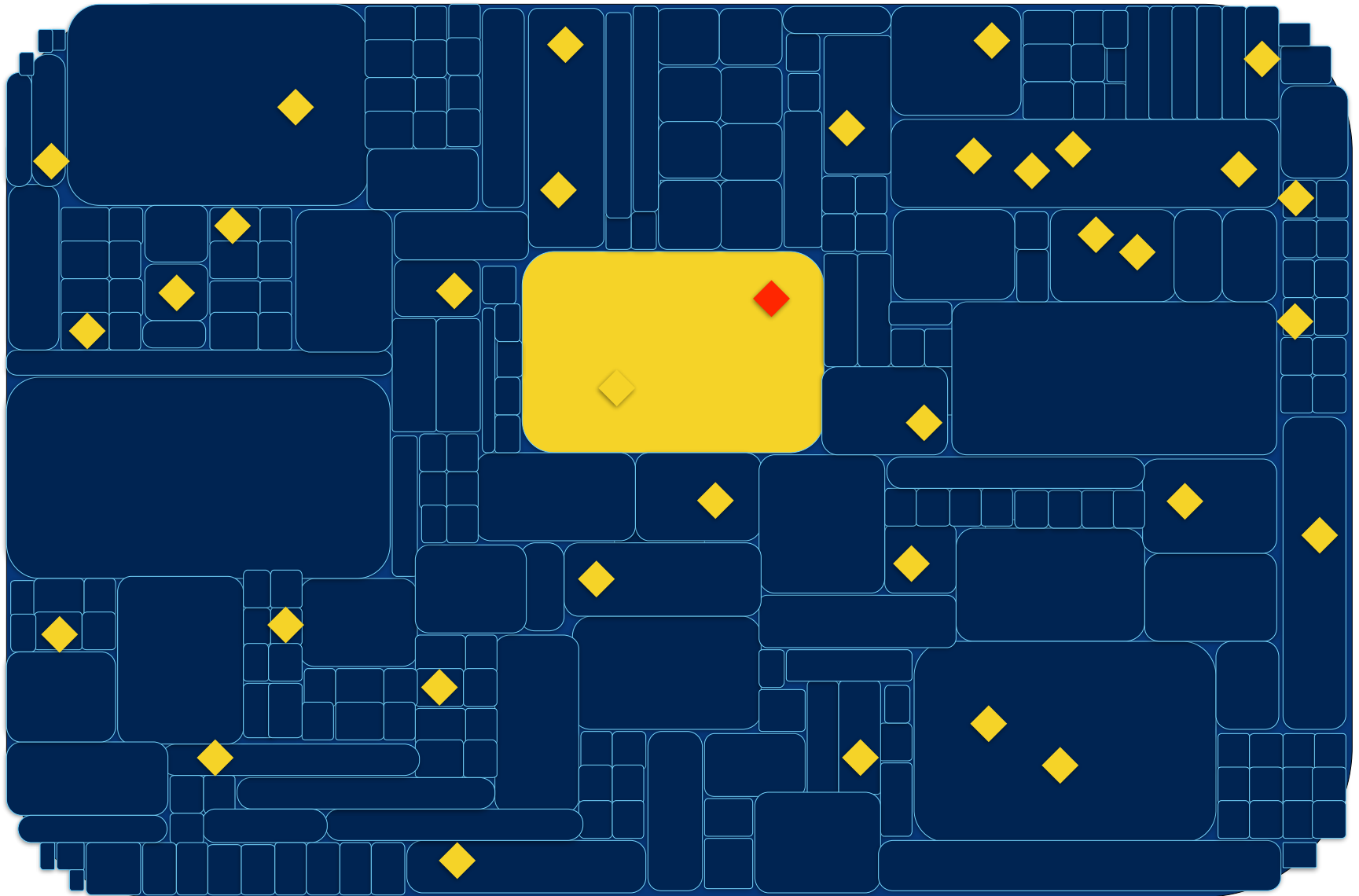
# The sequence space and annotated proteins



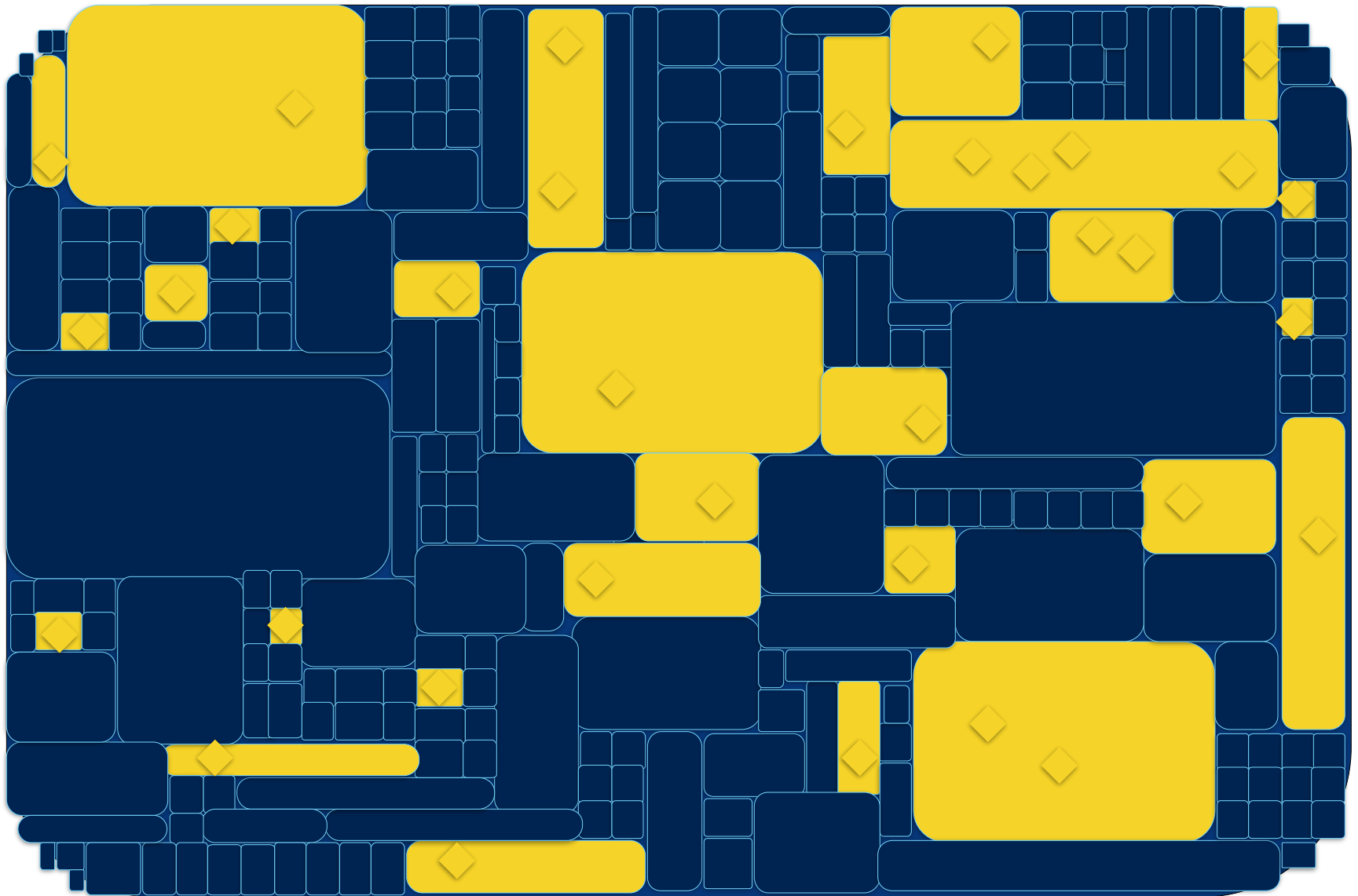
# Protein families



# Annotation transfer by homology

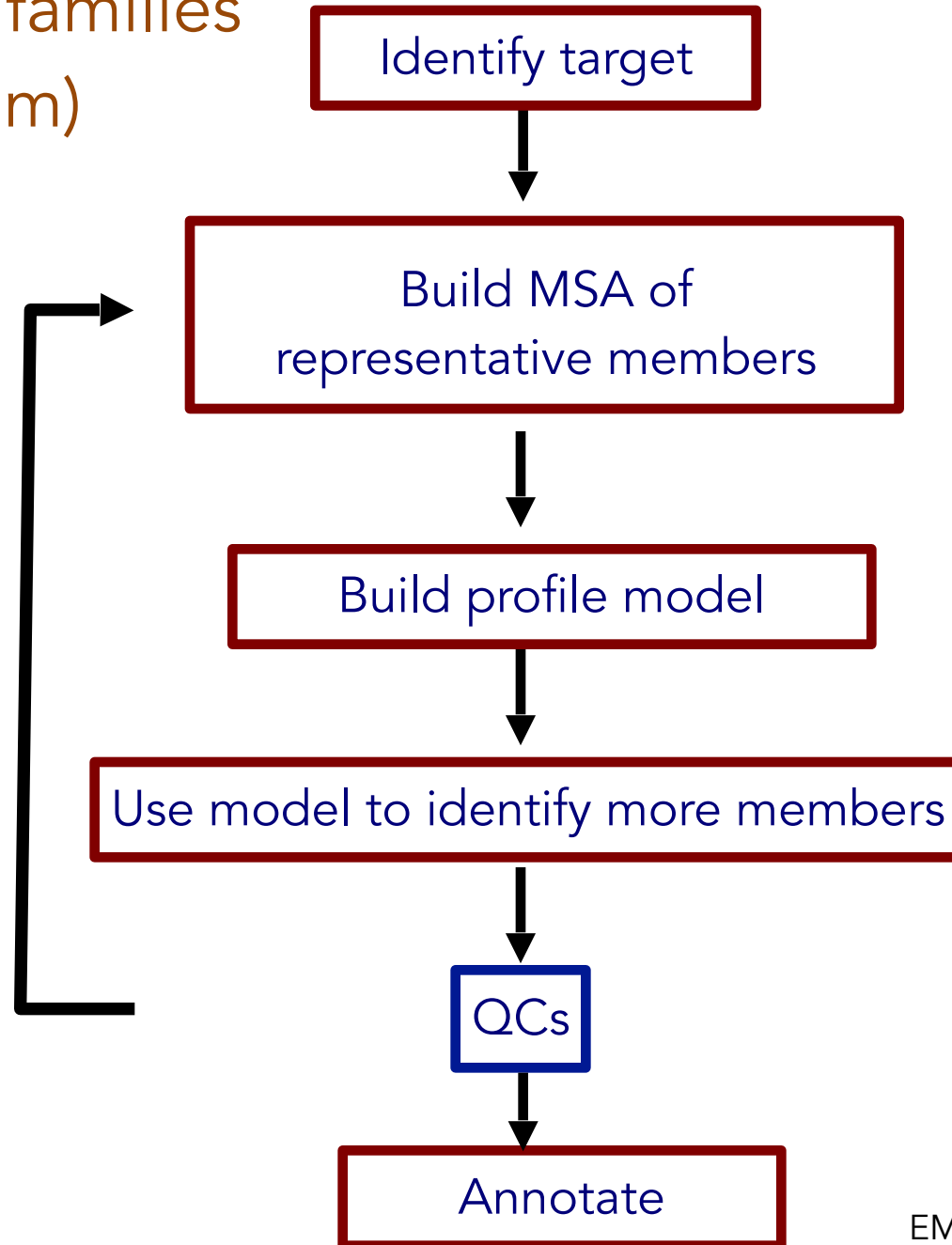


# Annotation transfer by homology



# Protein family databases

# Building families (Pfam)





Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIIRLFKGHPEKFDKFKHLKSEDEMKASE 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154

# Family power

10 20 30 40 50 60 70 80 90 100 110 120  
 HBAZ\_CAPHI/7-107 ERTIILSLWSK-I ST-----QADVIGTETLERLFS CYPQAKTY FPHF--DLH-----SGSAQLRAHGSKVVAAVGDAVKS I-----D-NVTSALS KLS ELHAYVL---RVDPVNFKFLSHCL  
 HBA3\_PLEWA/7-107 EKALVVG LCGK- I SG-----HCDALGGEALDRLEFASFGQTRTYFSHF--DLS-----PGSADVKRHHGKVLSAIGEAAKHI-----D-SMDQALS KLS DLHAYNL---RVDPGNFQLLSHCI  
 HBA\_CATCL/6-107 DKADV K IAWAK- I SP-----RADEI GAELGRMLTVYPQTKYFAHW-ADLS-----PGSGPVKHKGVIMGAI GDAVTKF-----D-DLLGGLS LSELHASKL---RVDPSPFKILANCI  
 HBB\_HETPO/7-106 ELHEITTTWKS- I-----DKHSLGAKALARMFIVYPWTTRYFGNL-KEFT-----ACSYGVKEHAKKVTGALGVAVTHL-----G-DVKSQFTDLSKKHAEEL---HVDVESFKLLAKCF  
 HBB\_SQUAC/7-107 EKALVNAVWTK-T-----DHQAVVAKALERLFFVYPWTQTYFVKFNGKHF-----ASDSTVQTHAGKVVSA LTVAYNH I-----D-DVKPHFVELSKKHYEEL---HVDPENFKLLANCL  
 HBB1\_CYGMA/8-112 ELTIINDIFSH-L-----DYDDIGPKALSRLIVYPWTRHFSFGF-GNLYNAEAIIGNANVAAGIKVHLGLDRGLKNM-----D-NIVDAYAELSTLHSEKL---HVDPDFNFKLLSDCI  
 HBB1\_XENBO/7-111 DRQLINSTWGK-V-----CAKTIGKEALGRLLWMTYPTWQRYFSSFGNLSADAVFHNEAAVAAGKEKVVTSIGEAIKHM-----D-DIKGYYAQLSKYHSETL---HVDPCNFKRFGGCL  
 HBB\_LUTCT/1-105 GGS DVSAFLAK-V-----DKRAVGGEARLRLIVYPWTQRYFSTF-GNLGSDAISHSNKVLAHQQRVLDSEIEGLKHP-----Z-BLKAYYAKLSERHSGEL---HVDPANFYRLGNVL  
 HBB\_LEPPA/7-111 EKQYIVSVFSK-I-----DVDHVGANTLERVLI VFPWTKRYFNSFGDLSPGA I KHNKVS AHGRKVLAAIIECTRFH-----G-NIKGHLANLSHLHSEKL---HVDPHNFRVLGQCL  
 HBB2\_XENLA/8-112 EKAAITSVWQK-V-----NVEHDGHDALGRLLIVYPWTQRYFNSFGNLSNSAAVAGNAKVAHQKVVLSAVGNAISHI-----D-SVKSSLQQLSKIHAT EL---FVDPENFKRFGGVL  
 HBB\_ALLMI/7-111 ERKFIVDLWAK-V-----DVAQCGADALSRLMIVYPWKRRYFEHF-GKMCNAHDILHNSKVQEHGKVLASFGAEVAKHL-----D-NIKGHFANLSKLHCEKF---HVDPENFKLLGDI I  
 HBB0\_MOUSE/8-112 EKAAITSIWDK-V-----DLEKVGGETLGRLLIVYPWTQRFDDKF-GNLSAQAIMGNPRIKAHGKVVLTSLGLAVKNM-----D-NLKETFAHLS ELHCDKL---HADPENFKLLGNML  
 HBBN\_AMMLE/2-106 BKALITGFWSK-V-----KVBZVGAZALGRLLVVPWTZRFFZHF-GBLSSABAVMBAKVKAHGKVVLSFBSBGLKHL-----B-BLKGAFASLSZLHCBKL---HVPZBFRLLGBVL  
 HBA\_LEPPA/7-108 DEVLIKEAWGL-L-H-----QIPNAGGEALARMFSCYPGTKSYFPHFGHDFS-----ANNEKVKHHGKVVDAIGQGVQHL-----H-DLSSCLHTLSEKHAREL---MVPD CNFYLI EAI  
 HBA1\_TORMA/6-107 NKKA IKNLLQK-IHS-----QTEVLGAELARLF ECHPQTKSYFPKF-SGFS-----ANDKRVKHHGALV LKALVDTNHL-----D-DLPHHLNKLAEKHGKGL---LVDPHNFKLFSDCI  
 HBA\_SQUAC/6-107 DKTAIKHLTGS-LRT-----NAEAWGAESLARMFATTPSTKTYFSKF-TDFS-----ANGKRVKAHGGKVLNAVADATDHL-----D-NVAGHLDP LAVLHGTTL---CVDPHNFPLLTQCI  
 HBA\_HETPO/13-114 DRAE LAALS KV-LAQ-----NAEAFGAELARMFTVY AATKSYFKDY-KDF-----AAAPS IKAHGAKVVTALAKACDHL-----D-DLKTHLHLKATFHGSEL---KVDPANFYLSYCL  
 GLB1\_TYLHE/7-110 QR I KVKQQAQ-VYSV-----GESRTDFAIDVFNMFRTNPD RS-LFNRVNGDNV-----YSPEFKAHMVRVFAFGFDLISVYL---DDKPVLDQALAHYA AFHKQFG---TIP---FKAFGQTM  
 GLB4\_LUMTE/11-120 DRREIRHIWDD-VWSSS-FTDRRVAIVRAVFDL LFKHYPTSKALFERVKIDEP-----ESGEFKSHLVRVANGLDLLINLL---DDTLVLQSHLGH LADQHIQRK---GVTKEYFRGIG EAF  
 GLB3\_TYLHE/8-117 DRHEVLDNWKG-IWSAE-FTGRRVAIGQAI FQELFALDPNAKGVFGRVNVD-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEEELKHLARQHRERS---GVKAVYFDEMEKAL  
 GLB4\_TYLHE/8-117 DRREEVQALWRS-IWSAE-DTGRRTLIGRLLFEELFIDGATKGLFKRVNVDVT-----HSP EEAHVLRVNGLDL LIGVL---GDSD-TLNSLIDHLAEQHKARA---GFKTVYFKEFGKAL  
 GLB2\_TYLHE/9-115 QR LKVKKQWAK-AYGV-----GHERVELGIALWKS MFAQDNDARDLFRKRVHGEDV-----HSPAF EAHMARVFNGLDRVIVSSL---TDEPV LNAQLEHLRQQH I KLG---ITGHMFNLMRTGL  
 GLB2\_LUMTE/8-114 EGLKVKSEWGR-AYGS-----GHDREAFSQA IWRATFAQVPE SRS LFRKRVHGDVT-----SHPAF IAHAEVRLGGLDL IASTL---DQPATLKEELDHLQVQHEGRK---IPDNYDAFKTAI  
 GLB\_TUBTU/6-112 QRFKVKHQWAE-AFGT-----SHHRLDFGLKLWNSIFRDAPEIRGLFKRVGDG-N-----AYS AEF EAHAEVRLGG LDMTISL---DDQAAFDAQLAHLKLSQAERN---IKADY YGVFVNL  
 GLB3\_LAMSP/7-113 QRLKVKRQWAE-AYGS-----GNDR EEF GHFIWTHVFKDAPSDARLFRKRVGDN I-----HTPAFRAHATRVLGG LDMC IALL---DDEGV LNTQLAHLASQHS SRG---VSAQYD VVEHSV  
 GLB\_PAREP/8-117 QDILLLKELGPH-V-DT---PAHIVETGLGAYHALFT AHPQYI I HFSRL-EG-HTIENVMQS EGIKHYARTL EAI VHM LKEI---SND AEVKKIAAQY GKDHTSRK---VTKDEFMSGEP I F  
 Q21978\_CAEL/165-283 SCEVVADSWRL-VESRSSAETSACFGIDLYKRVFESKIPMLRPLFG-L-SESDDVFDLPDNHPVRRRHARLFTSILHISVKNV---DELEAQVAPT VFKYGERHYRPI T PHMT EENVRV FCAQ I  
 GLB\_PSED/C/21-134 TREL CMKSL EHA-KVGT---SKEAKQDGLYKRVFEHY PAMKKYFKHR---ENYTPADVQKDPPIKQGN ILLACHVLCATY---DDRETFDAYVGLMARHERDHV---KIPNDVWNHFEWHF  
 GLB\_ASCSU/21-134 TREL CMKSL EHA-KVGT---SNEARQDGLDLYKHMFEHY PPLRKYFKNR---EYTAEDVQNDPFFAKQGGQ ILLACHVLCATY---DDRET FNAYTRELDRHARDHV---HMPEVWTFDWKLF  
 GLB\_C\_NIPBR/21-135 DVK--KHTVES-MKAVP-VGRDKAQNGIDFYKFFFTHHKDLRKF FKG A---ENFGADDVQKSKRF EKQGTALLLAVHVLANVY---DNQAVFHG FVRELMNRHEKRGVDPK LWKIFDDVWVVP  
 GLB\_C\_CAEL/10-119 DLC-VKSL EGR-MVGT E---AQNI-ENGNAFYRYFFTNFPDLRVYFKGA---EKY TADDVQKSERFDKQGGRI L LACHLLANVY---TNEEVFKGYVREINRHR IYK---MDPALWMAFFT V F  
 GLB2\_NIPBR/16-114 P I S K A Q Q --- A Q --- V G K D F Y K F F T N H P D L R K Y F K G A --- E N F T A D D V Q K S D R F E K L G S G L L L S V H I L A N T F --- D N E D V F R A F C R E T I D R H V G R G --- L D P A L W K A F W S V W  
 GLB\_H\_TRICQ/30-132 D Y V P L G S T P E K - L --- E N G R E F Y K Y F F T N H P D L R K Y F K G A --- E T F T A D D I A K S D R F E K L G N Q L L S V H L A D T Y --- D N E M I F R A F C R E T I D R H V D R G --- L D P K L W K F W S I Y  
 Q20638\_CAEL/74-184 E K E L L R R T W S D - E F D --- N L Y E L G S A I Y C Y I F D H N P N C K Q L F P - F - I S K Y Q G D E W K E S K F R S Q A L K F V Q T L A Q V V K N I Y H M E R T E S F L Y M V G Q K H V K F A D R G --- F K H E Y W D I F O D A M  
 Q19601\_CAEL/105-215 E K I L L E Q S W R K - T R K --- T G A D H I G S K I F M V L T A Q P D I K A I F G - L - E K I P T G R L K Y D N P R F R Q H A L V Y T T K L D F V I R N L --- D Y P G K L E V Y F E N L G R H V A M Q G - R G F E P G Y W E T F A E C M  
 Q18311\_CAEL/32-140 T K R L V I Q E W P R - V L A --- D C P E L T E I W H K S A T R S T S I K L A F G - I - A E - N - E S P M Q N A A F L G L S S T I Q A F F Y K L I T Y E - L - N D D Q V R E A C E Q L G R A H V D F I S - R G F N S H F W D I F L V C M



# Family power

Human: 1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLI R L R L F K G H P E T L E K F D K F H L K S E D E M K A S E 60  
 MGLSDGEWQLVLNVWGKVEAD GHGQEVLI LFK HPETL KFDKFF LKSE MK SE

Mouse: 1 MGLSDGEWQLVLNVWGKVEADLAGHGQEVLI G L F K T H P E T L D K F D K F H N L K S E E D M K G S E 60

Human: 61 DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120  
 DLKKHG TVLTALG ILKKKG H AEI PLAQSHATKHKIPVKYLEFISE II VL H

Mouse: 61 DLKKHGCTVLTALGTILKKKGQHAAEQPLAQSHATKHKIPVKYLEFISEIIIEVLKCRH 120

Human: 121 PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154  
 GDFGADAQGAM KALELFR D A YKELGFQG

Mouse: 121 SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG 154

# Family power

	10	20	30	40	50	60	70	80	90	100	110	120																																																																																																
HBAZ_CAPHI/7-107	ERTI	LSLWSK	I	ST	----	QADV	IGTET	LERL	F	SCY	P	QAKT	F	F	F	----	DLH	----	S	G	S	A	L	R	A	H	G	S	K	V	V	A	A	V	G	D	A	V	K	S	I	----	D	N	V	T	S	A	L	S	K	L	S	E	L	H	A	Y	V	L	----	R	V	D	P	V	N	F	K	F	L	S	H	C	L																																	
HBA3_PLEWA/7-107	EKAL	V	G	L	C	G	K	I	S	G	----	H	C	D	A	L	G	G	E	A	L	D	R	L	F	A	S	F	G	Q	T	R	T	F	S	H	F	----	D	L	S	----	P	G	S	A	D	V	K	R	H	G	K	V	L	S	A	I	G	E	A	A	K	H	I	----	D	S	M	D	Q	A	L	S	K	L	S	D	L	H	A	Y	N	L	----	R	V	D	P	G	N	F	Q	L	L	S	H	C	I									
HBA_CATCL/6-107	DKAD	V	K	I	A	W	A	K	I	S	P	----	R	A	E	I	G	A	E	A	L	G	R	M	L	T	V	Y	P	Q	T	K	T	F	A	H	W	----	A	D	L	----	P	G	S	G	P	V	K	H	G	K	V	I	M	G	A	I	G	D	A	V	T	K	F	----	D	L	L	G	G	L	S	L	S	E	L	H	A	S	K	L	----	R	V	D	P	S	N	F	K	I	L	A	N	C	I											
HBB_HETPO/7-106	ELHE	I	T	T	T	W	K	S	I	----	D	K	H	S	L	G	A	K	A	L	A	R	M	F	I	V	Y	P	W	T	T	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	V	E	S	F	K	L	L	A	K	C	F									
HBB_SQUAC/7-107	EKAL	V	N	A	V	W	T	K	T	----	D	H	Q	A	V	V	A	K	A	L	E	R	L	F	V	V	Y	P	W	T	K	T	Y	F	G	N	L	----	K	E	F	----	A	S	D	T	V	Q	T	H	A	G	K	V	S	A	L	T	V	A	N	H	I	----	D	D	V	K	P	H	F	V	E	L	S	K	H	Y	E	E	L	----	H	V	D	P	E	N	F	K	L	L	A	N	C	L												
HBB1_CYGMA/8-112	ELT	I	I	N	D	I	F	S	H	L	----	D	Y	D	D	I	G	P	K	A	L	S	R	C	L	I	V	Y	P	W	T	Q	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I								
HBB1_XENBO/7-111	DRQL	I	N	S	T	W	G	K	V	----	C	A	K	T	I	G	K	E	A	L	G	R	L	L	W	T	Y	P	W	T	Q	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I									
HBB_LITCT/1-105	GGSD	V	S	A	F	L	A	K	V	----	D	K	R	A	V	G	G	E	A	L	A	R	L	L	I	V	Y	P	W	T	Q	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I									
HBB_LEPPA/7-111	EKQY	I	V	S	V	F	S	K	I	----	D	V	D	H	V	G	A	N	T	L	E	R	V	L	I	V	Y	P	W	T	K	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I									
HBB2_XENLA/8-112	EKAA	I	T	S	V	W	Q	K	V	----	N	V	E	H	D	G	H	D	A	L	G	R	L	L	I	V	Y	P	W	T	Q	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	F	V	D	P	E	N	F	K	R	F	G	G	V	L									
HBB_ALLMI/7-111	ERKF	I	V	D	L	W	A	K	V	----	D	V	A	Q	C	G	A	D	A	L	S	R	M	L	I	V	Y	P	W	K	R	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I									
HBB0_MOUSE/8-112	EKAA	I	T	S	I	W	D	K	V	----	D	L	E	K	V	G	G	E	T	L	G	R	L	L	I	V	Y	P	W	T	Q	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I									
HBBN_AMMLE/2-106	BKAL	I	T	G	F	W	S	K	V	----	K	V	B	Z	V	G	A	Z	A	L	G	R	L	L	V	Y	P	W	T	Z	R	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	H	V	D	P	E	N	F	K	L	L	S	D	C	I										
HBA_LEPPA/7-108	DEV	L	I	K	E	A	W	G	L	L	H	----	Q	I	P	N	A	G	G	E	A	L	A	R	M	F	S	C	P	G	T	K	S	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	M	V	D	P	C	N	F	Q	L	I	E	A	I									
HBA1_TORMA/6-107	NKKA	I	K	N	L	L	Q	K	I	H	S	----	Q	T	E	V	L	G	A	E	A	L	A	R	L	F	E	C	H	P	Q	T	K	S	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	D	D	L	P	H	L	N	K	L	A	E	K	H	G	K	L	----	L	V	D	P	H	N	F	K	L	F	S	D	C	I								
HBA_SQUAC/6-107	DKT	A	I	K	H	L	T	G	S	L	R	T	----	N	A	E	A	W	A	G	E	S	L	A	R	M	F	A	T	T	S	T	K	T	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	D	N	V	A	G	H	L	D	P	L	A	V	L	H	G	T	T	L	----	C	V	D	P	H	N	F	P	L	L	T	Q	C	I						
HBA_HETPO/13-114	DRAE	L	A	A	L	S	K	V	L	A	Q	----	N	A	E	A	F	G	A	E	A	L	A	R	M	F	T	Y	A	A	K	S	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	D	D	L	K	T	H	L	K	L	A	T	F	H	G	S	E	L	----	K	V	D	P	A	N	F	Q	L	S	Y	L	C	I									
GLB1_TYLHE/7-110	QR	I	K	V	K	Q	Q	W	A	Q	----	Y	S	V	----	G	E	S	R	T	D	F	A	I	D	V	F	N	N	F	R	T	N	D	R	S	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	D	D	K	P	V	L	D	Q	A	L	A	H	A	F	H	K	Q	F	G	----	T	I	P	----	F	K	A	F	Q	T	M						
GLB4_LUMTE/11-120	DRRE	I	R	H	I	W	D	D	V	W	S	S	----	F	T	D	R	R	V	A	I	V	R	A	V	F	D	D	L	F	K	H	Y	P	T	S	K	A	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	G	V	T	K	E	Y	F	R	G	I	G	E	A	F			
GLB3_TYLHE/8-117	DRHE	V	L	D	N	W	K	G	I	W	S	A	E	----	F	T	G	R	R	V	A	I	G	Q	A	I	F	Q	E	L	F	A	L	D	N	A	K	G	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	D	D	P	K	A	L	Q	E	E	L	K	H	L	A	R	Q	H	R	E	R	S	----	G	V	K	A	Y	F	D	E	M	E	K	A	L
GLB4_TYLHE/8-117	DRRE	V	Q	A	L	W	R	S	I	W	S	A	E	----	D	T	G	R	R	T	L	I	G	R	L	L	F	E	E	L	F	E	I	D	G	A	T	K	G	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	G	D	V	K	S	Q	F	T	D	L	S	K	H	A	E	E	L	----	G	F	K	T	Y	F	K	E	F	G	K	A	L			
GLB2_TYLHE/9-115	QR	L	K	V	K	Q	Q	W	A	K	----	A	Y	G	V	----	G	H	E	R	V	E	L	G	I	A	L	W	K	S	M	F	A	Q	D	N	A	R	D	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	I	T	G	H	M	F	N	L	M	R	T	G	L																					
GLB2_LUMTE/8-114	EGL	K	V	K	S	E	W	G	R	----	A	Y	G	S	----	G	H	D	R	E	A	F	S	Q	A	I	W	R	A	T	F	A	Q	V	P	E	S	R	S	Y	F	G	N	L	----	K	E	F	----	A	C	S	Y	G	V	K	E	H	A	K	K	V	T	G	A	L	G	V	A	V	T	H	L	----	I	P	D	N	F	D	A	F	K	T	A	I																						
GLB_TUBTU/6-112	QR	F	V	K	H	Q	W	A	E	----	A	F	G	T	----	S	H	R	L	D	F	G	L	K	L	W	N	S	I	F	R	D	A	P	E	I	R	G																																																																						

# Family power

	10	20	30	40	50	60	70	80	90	100	110	120																																																																																																
HBAZ_CAPI/7-107	ERTI	LSLWSK	-IST	-----	QADV	IGTET	LERL	FS	CY	QAKT	EPHF	-DLH	-----	SGSA	QLRAHG	SKVVA	AVGD	AVKSI	-----	D-NVT	SALS	KLSL	EL	HAYV	VL	---	RVD	PVN	FKFL	SHCL																																																																														
HBA3_PLEWA/7-107	EKAL	VGLCGK	-ISG	-----	HCDA	LGGEAL	DRLE	FA	FG	QTRT	FSHF	-DLS	-----	PGSA	VDYKR	HGKVL	SAI	GEAA	KHI	-----	D-SMD	QALS	KLSL	SD	HAYN	NL	---	RVD	PGN	FQL	LSHC																																																																													
HBA_CATCL/6-107	DKAD	VKAIAW	K-ISP	-----	RADE	IGAEAL	GRML	TV	PQ	TKT	FAHW	-ADLS	-----	PGS	GPVK	HGKVV	IMGA	IGDA	VTKF	-----	D-DLL	GGLS	LS	EL	HASK	LL	---	RVD	PSN	FKI	LANCI																																																																													
HBB_HETPO/7-106	ELHE	ITTTWKS	-I	-----	DKHS	LGAKAL	ARMF	I	VY	PW	TTRY	FANL	KEFT	-----	ACSY	GVKE	HAKK	V	GAL	GV	AV	THL	-----	G-DVK	SQ	FT	DL	SKK	HAE	EL	---	HVD	VES	FKLL	LAKCF																																																																									
HBB_SQUAC/7-107	EKAL	VNAV	WT-K	-----	DHQAV	VAKAL	LERL	FV	VY	PW	TKT	FYK	FKF	-----	ASD	TVQ	HAGK	VVS	ALT	VAY	NHI	-----	D-DVK	PHF	V	EL	SKK	HY	EEL	---	HVD	PEN	FKLL	LANCL																																																																										
HBB1_CYGMA/8-112	ELT	IIND	I	F	SH-L	-----	DYDD	I	GP	KALS	RCL	I	VY	PW	TQR	FSGF	G	NLY	NAE	A	I	GNAN	VAAH	I	KVL	HGL	DR	GL	KNM	-----	D-NI	V	DAY	AEL	S	L	H	SEK	L	---	HVD	PN	FKLL	SDCI																																																																
HBB1_XENBO/7-111	DRQL	I	NST	WGK	-V	-----	CAKT	I	G	EAL	GR	L	L	W	T	Q	R	FS	F	GN	L	S	A	D	A	V	F	H	N	E	A	A	A	H	G	K	V	V	T	S	I	G	E	A	I	K	H	M	---	D-D	I	K	G	Y	A	Q	L	S	K	Y	H	S	E	T	L	---	HVD	P	C	N	F	K	R	F	G	G	C																															
HBB_LITCT/1-105	GGSD	VSAFLAK	-V	-----	DKRA	VGG	EAL	AR	L	L	I	V	Y	P	W	T	Q	R	FS	T	F	GN	L	S	A	D	A	I	S	H	N	S	K	V	L	A	H	G	R	V	L	D	S	I	E	E	G	L	K	H	P	---	Z	B	L	K	A	Y	A	K	L	S	E	R	H	S	G	E	L	---	HVD	P	A	N	F	R	L	G	N	V	L																											
HBB_LEPPA/7-111	EKQY	I	V	S	V	F	S	K	-I	-----	DVDH	V	G	A	N	T	L	E	R	V	L	I	V	Y	P	W	T	K	R	FS	F	GN	L	S	P	G	A	I	K	H	N	K	V	S	A	H	G	R	K	V	L	A	A	I	E	C	T	R	H	F	---	G-N	I	K	G	H	L	A	N	L	S	H	L	H	S	E	K	L	---	HVD	P	H	N	F	R	V	L	G	C	L																		
HBB2_XENLA/8-112	EKAA	I	T	S	V	W	Q	K	-V	-----	NVEH	D	G	H	D	A	L	G	R	L	L	I	V	Y	P	W	T	Q	R	FS	F	GN	L	S	N	S	A	A	V	A	G	N	A	V	Q	A	H	G	K	V	L	S	A	V	G	N	A	I	S	H	I	---	D-S	V	K	S	L	Q	L	S	K	I	H	A	T	E	L	---	F	V	D	P	E	N	F	K	R	F	G	V	L																	
HBB_ALLMI/7-111	ERKF	I	V	D	L	W	A	K	-V	-----	DVAQ	C	G	A	D	A	L	S	R	M	L	I	V	Y	P	W	K	R	R	FS	F	GN	K	M	C	N	A	H	D	I	L	H	N	S	K	V	Q	E	H	G	K	V	L	S	F	G	E	A	V	K	H	L	---	D-N	I	K	G	H	F	A	N	L	S	K	L	H	C	E	K	F	---	HVD	P	E	N	FK	L	L	G	D	I	I																
HBB0_MOUSE/8-112	EKAA	I	T	S	I	W	D	K	-V	-----	DLEK	V	G	G	E	T	L	G	R	L	L	I	V	Y	P	W	T	Q	R	FS	F	GN	L	S	S	A	Q	A	I	M	G	N	P	R	I	K	A	H	G	K	V	L	S	L	G	L	A	V	K	N	M	---	D-N	L	K	E	T	F	A	H	L	S	E	L	H	C	D	K	L	---	HAD	P	E	N	FK	L	L	G	N	M	L																	
HBBN_AMMLE/2-106	BKAL	I	T	G	F	W	S	K	-V	-----	KVBZ	V	G	Z	A	L	G	R	L	L	V	V	Y	P	W	T	Z	R	FS	F	GN	L	S	S	A	B	A	V	M	B	B	A	K	V	K	A	H	G	K	V	L	S	F	S	B	G	L	K	H	L	---	B	B	L	K	G	A	F	A	S	L	S	Z	L	H	C	B	K	L	---	HVB	P	Z	B	F	R	L	L	G	B	V	L																
HBA_LEPPA/7-108	DEV	L	I	K	E	A	W	G	L	-L	H	-----	QIPN	A	G	G	E	A	L	A	R	M	F	S	C	P	G	T	K	S	FS	F	GN	H	D	F	S	-----	ANNE	K	V	K	H	G	K	V	V	D	A	I	G	Q	G	V	Q	H	L	---	H-D	L	S	S	C	L	H	T	L	S	E	K	H	A	R	E	L	---	M	V	D	P	C	N	F	Q	L	I	E	A	I																			
HBA1_TORMA/6-107	NKKA	I	K	N	L	L	Q	K	-I	H	S	-----	QTEV	L	G	A	E	A	L	A	R	L	F	E	C	H	P	Q	T	K	S	FS	F	GN	S	G	F	S	-----	AND	K	R	V	K	H	G	A	L	V	L	K	A	L	V	D	T	N	K	H	L	---	D-D	L	P	H	L	N	K	L	A	E	K	H	G	K	L	---	L	V	D	P	H	N	FK	L	S	D	C	I																			
HBA_SQUAC/6-107	DKT	A	I	K	H	L	T	G	S	-L	R	T	-----	NAEA	W	G	A	E	S	L	A	R	M	F	A	T	P	S	T	K	T	FS	K	F	T	D	F	S	-----	ANG	K	R	V	K	A	H	G	K	V	L	N	A	V	A	D	A	T	D	H	L	---	D-N	V	A	G	H	L	D	P	L	A	V	L	H	G	T	T	L	---	C	V	D	P	H	N	F	P	L	L	T	Q	C	I															
HBA_HETPO/13-114	DRAE	L	A	A	L	S	K	V	-L	A	Q	-----	NAEA	F	G	A	E	A	L	A	R	M	F	T	V	Y	A	A	T	K	S	FS	F	GN	O	K	D	F	-----	AA	A	P	S	I	K	A	H	G	A	K	V	V	T	A	L	A	K	A	C	D	H	L	---	D-D	L	K	T	H	L	K	L	A	T	F	H	G	S	E	L	---	K	V	D	P	A	N	F	Q	L	S	Y	S	L	C	I													
GLB1_TYLHE/7-110	QR	I	K	V	K	Q	Q	W	A	-Y	V	S	V	-----	GES	R	T	D	F	A	I	D	V	F	N	N	F	R	T	N	D	R	S	-F	N	V	M	G	D	N	V	-----	Y	S	P	E	F	K	A	H	M	V	R	V	F	A	G	F	D	I	L	S	V	L	---	D	D	K	P	V	L	D	Q	A	L	A	H	Y	A	F	H	K	Q	F	G	---	T	I	P	---	F	K	A	F	G	T	M											
GLB4_LUMTE/11-120	DRRE	I	R	H	I	W	D	-V	W	S	S	-----	FTDR	R	V	A	I	V	R	A	V	F	D	L	F	K	H	Y	P	T	S	K	A	FS	F	GN	V	K	I	D	E	P	-----	E	S	G	E	F	K	S	H	L	V	R	V	A	N	G	L	D	L	I	N	L	---	D	D	T	L	V	L	Q	S	H	L	G	H	L	A	D	Q	H	I	Q	R	K	---	G	V	T	K	E	Y	F	R	G	I	G	E	A	F							
GLB3_TYLHE/8-117	DRHE	V	L	D	N	W	K	G	-I	W	S	A	E	-----	FTGR	R	V	A	I	G	Q	A	I	F	E	L	F	A	L	D	N	A	K	G	FS	F	GN	V	M	V	D	-K	-----	P	S	E	A	D	W	K	A	H	V	I	R	V	I	N	G	L	D	L	A	V	N	L	---	E	D	P	K	A	L	E	E	L	K	H	L	A	R	Q	H	R	S	---	G	V	K	A	Y	F	D	E	M	E	K	A	L									
GLB4_TYLHE/8-117	DRRE	V	Q	A	L	W	R	S	-I	W	S	A	E	-----	DTGR	R	T	L	I	G	R	L	F	E	E	L	F	E	I	D	G	A	T	K	G	FS	F	GN	V	M	V	D	T	-----	H	S	P	E	F	A	H	V	L	R	V	N	G	L	D	L	I	G	V	L	---	G	D	S	D	-T	L	N	S	L	I	D	H	L	A	E	Q	H	K	A	R	A	---	G	F	K	T	Y	F	K	E	F	G	K	A	L								
GLB2_TYLHE/9-115	QR	L	K	V	K	Q	Q	W	A	-A	Y	G	V	-----	GHER	V	E	L	G	I	A	L	W	K	S	M	F	A	Q	D	N	A	R	D	FS	F	GN	V	H	G	E	D	V	-----	H	S	P	A	F	E	A	H	M	A	R	V	F	N	G	L	D	R	V	I	S	S	L	---	T	D	E	P	V	L	N	A	Q	L	E	H	L	R	Q	H	I	K	L	G	---	I	T	G	H	M	F	N	L	M	R	T	G	L						
GLB2_LUMTE/8-114	EGL	K	V	K	S	E	W	G	R	-A	Y	G	S	-----	GHDR	E	A	F	S	Q	A	I	W	R	A	T	F	A	Q	V	P	E	S	R	S	FS	F	GN	V	H	G	D	T	-----	S	H	P	A	F	I	A	H	A	E	R	V	L	G	L	D	I	A	I	S	T	L	---	D	Q	P	A	T	L	K	E	E	L	D	H	L	Q	V	Q	H	E	G	R	K	---	I	P	D	N	F	D	A	F	K	T	A	I							
GLB_TUBTU/6-112	QR	F	V	K	H	Q	W	A	E	-A	F	G	T	-----	SHR	L	D	F	G	L	K	L	W	N	S	I	R	D	A	P	E	I	R	G	FS	F	GN	V	D	G	-N	-----	A	Y	S	A	E	F	E	A	H	A	E	R	V	L	G	L	D	M	T	I	S	L	---	D	D	Q	A	F	D	A	Q	L	A	H	L	K	S	Q	H	A	E	R	N	---	I	K	A	D	Y	G	V	F	N	E	L											
GLB3_LAMSP/7-113	QR	L	K	V	K	R	Q	W	A	E	-A	Y	G	S	-----	GND	R	E	E	F	G	H	I	W	H	V	F	K	D	A	P	S	A	R	D	FS	F	GN	V	R	G	D	N	-----	H	T	P	A	F	R	A	H	A	T	R	V	L	G	L	D	M	C	I	A	L	---	D	D	E	G	V	L	N	T	Q	L	A	H	L	A	S	Q	H	S	R	G	---	V	S	A	A	Q	V	D	V	E	H	S	V									
GLB_PAREP/8-117	QD	I	L	L	K	E	L	G	P	H	-V	D	T	-----	PAH	I	V	E	T	G	L	G	A	H	A	L	F	T	A	H	P	Q	Y	I	H	F	S	R	L	E	G	-H	T	I	E	N	V	M	Q	S	E	G	I	K	H	Y	A	R	T	L	T	E	A	I	V	H	M	L	K	E	I	---	S	N	D	A	E	V	K	K	I	A	A	Q	Y	G	K	D	H	T	S	R	K	---	V	T	K	D	E	F	M	S	G	E	P	I	F	
Q21978_CAEL/165-283	S	C	E	V	A	D	S	W	R	L	-V	E	S	R	S	A	A	E	T	S	A	C	F	G	L	F	W	F	Q	R	V	E	S	K	I	P	M	L	R	P	-L	S	E	S	D	D	V	D	L	P	D	N	H	P	V	R	R	H	A	L	F	T	S	I	L	H	S	V	K	N	V	---	D	E	L	E	A	Q	V	A	P	T	V	F	K	Y	G	E	R	H	Y	R	P	D	I	P	H	M	T	E	E	N	V	R	F	C	A	I

# Family power score(ab,i)

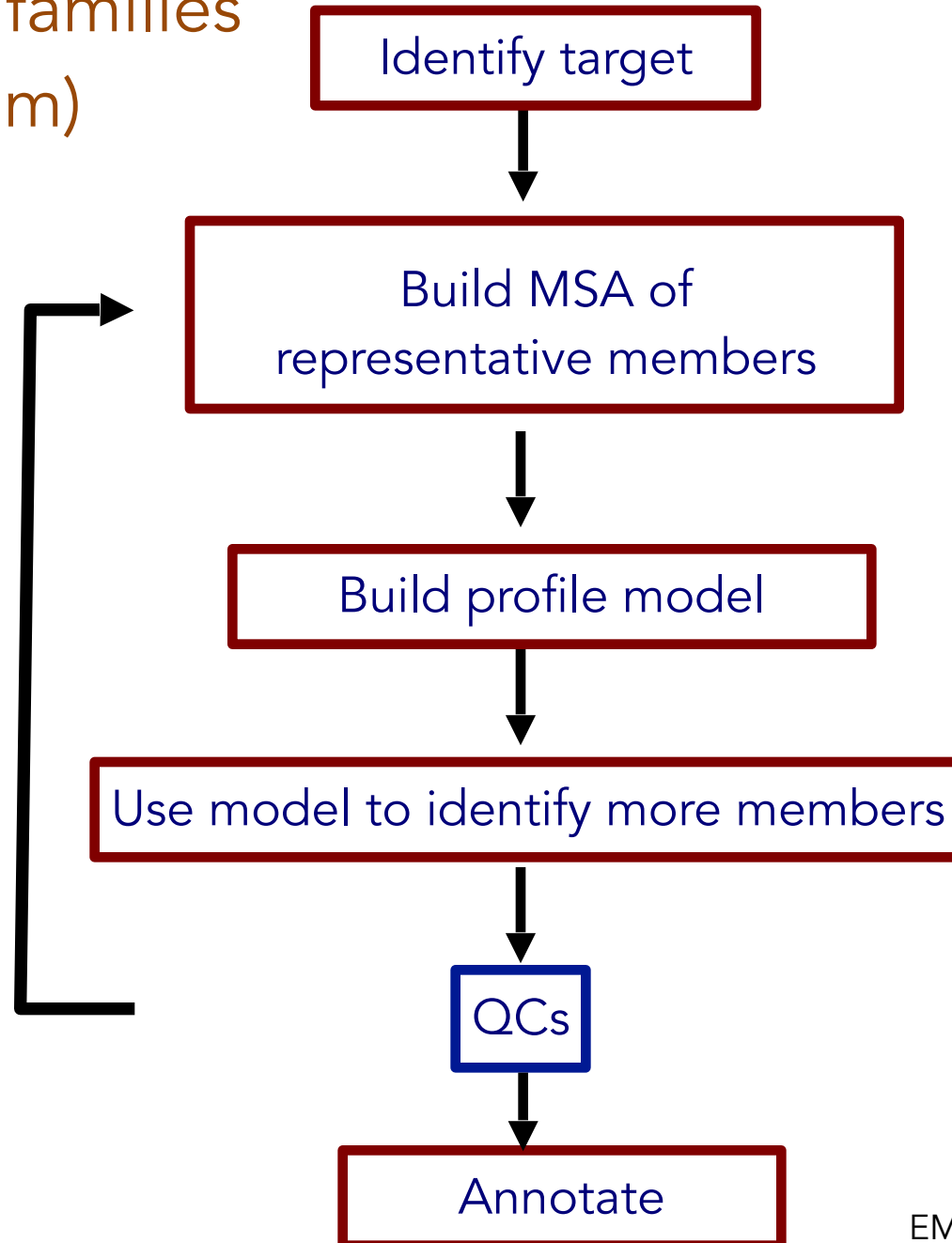
	10	20	30	40	50	60	70	80	90	100	110	120																																																																																															
HBAZ_CAPIH/7-107	ERTI	LSLWSK	IST	-----	QADVIGTETLERLFS	CYQA	KTY	EPHF	-DLH	-----	SGSAQLRAHGS	SKVVAAVGD	AVKSI	-----	D-NVTSALS	KLSEL	HAYVL	-----	RVDPVN	FKFL	SHCL																																																																																						
HBA3_PLEWA/7-107	EKAL	VGLCGK	ISG	-----	HCDA LGGEALDR	LFA	SFG	QTRT	FSHF	-DLS	-----	PGSADV	KRHG	GKVL	SAI	G	EAAKH	I	-----	D-SMD	QALS	KLSDL	HAYNL	-----	RVDPGN	FQL	LSHCI																																																																																
HBA_CATCL/6-107	DKAD	VIAWAK	ISP	-----	RADEIGEA	LGR	M	LTVY	PQTKT	FAHW	-ADLS	-----	PGSGPV	KHGK	VIM	G	ADAV	TKF	-----	D-DLL	GGLS	LSSEL	HASKL	-----	RVDP	SNFK	I	LANCI																																																																															
HBB_HETPO/7-106	ELHE	ITTTWKS	I	-----	DKHS	LGAKAL	ARM	FIVY	PWTRTY	FANL	-KEFT	-----	ACSY	GVKE	HAK	KVT	GAL	GVAV	THL	-----	G-DVK	SQFT	DL	SKKH	AEEL	-----	HVDV	ES	FKLL	LAKCF																																																																													
HBB_SQUAC/7-107	EKAL	VNAVWT	K	-----	DHQAV	VAKAL	LER	L	FVVY	PWTKT	FVYF	-GKF	-----	ASD	VTQ	HAG	KVVS	ALT	VAY	NHI	-----	D-DVK	PHF	V	ELSS	KKHY	EEL	-----	HVD	EN	FKLL	LANCL																																																																											
HBB1_CYGMA/8-112	ELT	IINDIFSH	L	-----	DYDD	IGPKALS	R	CLIVY	PWTRQR	FSGF	-GNLY	NAEAI	I	GNAN	VAAHG	I	KV	LHGL	DR	G	LKNM	-----	D-NI	V	DAY	AE	L	SH	SEKL	-----	HVD	PN	FKLL	LSDCI																																																																									
HBB1_XENBO/7-111	DRQL	IINSTW	GK	V	-----	CAKT	I	G	EAL	G	RLL	L	W	T	Y	P	W	T	Q	R	Y	FS	-	GN	L	S	A	D	A	V	F	H	N	E	A	A	H	G	K	V	V	T	S	I	G	E	A	I	K	H	M	-----	D-D	I	K	G	Y	A	L	S	K	Y	H	S	E	T	L	-----	HVD	PC	N	F	K	R	F	G	G	C																													
HBB_LITCT/1-105	GGSD	VSAFLAK	V	-----	DKRA	V	G	EAL	R	L	L	I	V	Y	P	W	T	Q	R	Y	FS	-	GN	L	S	A	D	A	I	S	H	N	S	K	V	L	A	H	G	R	V	L	S	I	E	E	G	L	K	H	P	-----	Z	B	L	K	A	Y	A	K	L	S	E	R	H	S	G	E	L	-----	HVD	P	A	N	F	R	L	G	N	V	L																										
HBB_LEPPA/7-111	EKQY	I	V	S	V	F	S	K	I	-----	D	V	D	H	V	G	A	N	T	L	E	R	V	L	I	V	Y	P	W	T	K	R	Y	FS	-	GN	L	S	P	G	A	I	K	H	N	K	V	S	A	H	G	R	K	V	L	A	A	I	E	C	T	R	H	F	-----	G	N	I	K	G	H	L	A	N	L	S	H	L	H	S	E	K	L	-----	HVD	P	H	N	F	R	V	L	G	C	L												
HBB2_XENLA/8-112	EKAA	I	T	S	V	W	Q	K	V	-----	N	V	E	H	D	H	D	A	L	G	R	L	L	I	V	Y	P	W	T	Q	R	Y	FS	-	GN	L	S	N	S	A	A	V	A	G	N	A	K	V	Q	A	H	G	K	V	L	S	A	V	G	N	A	I	S	H	I	-----	D-S	V	K	S	L	Q	L	S	K	I	H	A	T	E	L	-----	F	V	D	P	E	N	F	K	R	F	G	V	V	L											
HBB_ALLMI/7-111	ERKF	I	V	D	L	W	A	K	V	-----	D	V	A	C	G	A	D	A	L	S	R	L	I	V	Y	P	W	K	R	Y	FS	-	GN	K	M	C	N	A	H	I	L	H	N	S	K	V	Q	E	H	G	K	V	L	S	F	G	E	A	V	K	H	L	-----	D-N	I	K	G	H	F	A	N	L	S	K	L	H	C	E	K	F	-----	HVD	P	E	N	F	K	L	L	G	D	I	I														
HBB0_MOUSE/8-112	EKAA	I	T	S	I	W	D	K	V	-----	D	L	E	K	V	G	G	E	T	L	G	R	L	L	I	V	Y	P	W	T	Q	R	Y	FS	-	GN	L	S	S	A	Q	A	I	M	G	N	P	R	I	K	A	H	G	K	V	L	S	L	G	L	A	V	K	N	M	-----	D-N	L	K	E	T	F	A	H	L	S	E	L	H	C	D	K	L	-----	HAD	P	E	N	F	K	L	L	G	N	M	L											
HBBN_AMMLE/2-106	BKAL	I	T	G	F	W	S	K	V	-----	K	V	B	Z	V	G	A	Z	A	L	G	R	L	L	V	Y	P	W	T	Z	R	Y	FS	-	GN	L	S	S	A	B	A	V	M	B	B	A	K	V	K	A	H	G	K	V	L	S	F	S	B	G	L	K	H	L	-----	B	B	L	K	G	A	F	A	S	L	S	Z	L	H	C	B	K	L	-----	HVB	P	Z	B	F	R	L	L	G	B	V	L											
HBA_LEPPA/7-108	DEV	L	I	K	E	A	W	G	L	-L	H	-----	Q	I	P	N	A	G	E	A	L	A	R	M	F	S	C	Y	P	G	T	K	S	Y	FS	-	GN	L	S	S	P	G	A	I	K	H	N	K	V	S	A	H	G	R	K	V	L	A	A	I	E	C	T	R	H	F	-----	G	N	I	K	G	H	L	A	N	L	S	H	L	H	S	E	K	L	-----	MVD	P	C	N	F	Q	L	I	E	A	I										
HBA1_TORMA/6-107	NKKA	I	K	N	L	L	Q	K	I	H	S	-----	Q	T	E	V	L	G	A	E	A	L	A	R	L	F	E	C	H	P	Q	T	K	S	Y	FS	-	GN	L	S	G	F	S	-----	A	N	D	K	R	V	K	H	H	G	A	L	V	L	K	A	L	V	D	T	N	K	H	L	-----	D-D	L	P	H	L	N	K	L	A	E	K	H	G	K	L	-----	L	V	D	P	H	N	F	K	L	F	S	D	C	I								
HBA_SQUAC/6-107	DKT	A	I	K	H	L	T	G	S	-L	R	T	-----	N	A	E	A	W	A	G	E	S	L	A	R	M	F	A	T	P	S	T	K	T	Y	FS	-	GN	L	S	K	F	T	D	F	S	-----	A	N	G	K	R	V	K	A	H	G	K	V	L	N	A	V	A	D	A	T	D	H	L	-----	D-N	V	A	G	H	L	D	P	L	A	V	L	H	G	T	T	L	-----	C	V	D	P	H	N	F	P	L	L	T	Q	C	I				
HBA_HETPO/13-114	DRAE	L	A	A	L	S	K	V	-L	A	Q	-----	N	A	E	A	F	G	E	A	L	A	R	M	F	T	Y	A	A	T	K	S	Y	FS	-	GN	L	S	K	D	F	S	-----	A	A	A	P	S	I	K	A	H	G	A	K	V	V	T	A	L	A	K	A	C	D	H	L	-----	D-D	L	K	T	H	L	K	L	A	T	F	H	G	S	E	L	-----	K	V	D	P	A	N	F	Q	L	S	Y	S	L	C	I							
GLB1_TYLHE/7-110	QR	I	K	V	K	Q	Q	W	A	-Y	V	S	-----	G	E	S	R	T	D	F	A	I	D	V	F	N	N	F	R	T	N	D	R	S	Y	FS	-	GN	L	S	P	E	F	K	A	H	M	R	V	F	A	G	F	D	I	L	S	V	L	-----	D	D	K	P	V	L	D	Q	A	L	A	H	Y	A	F	H	K	Q	F	G	-----	T	I	P	-----	F	K	A	F	G	T	M															
GLB4_LUMTE/11-120	DRRE	I	R	H	I	W	D	-V	W	S	S	-----	F	T	D	R	R	V	A	I	V	R	A	V	F	D	D	L	F	K	H	Y	P	T	S	K	A	Y	FS	-	GN	L	S	E	G	F	K	S	H	L	V	R	V	A	N	G	L	D	L	I	N	L	-----	D	D	T	L	V	L	Q	S	H	L	G	H	L	A	D	Q	H	I	Q	R	-----	G	V	T	K	E	Y	F	R	G	I	G	E	A	F									
GLB3_TYLHE/8-117	DRHE	V	L	D	N	W	K	G	-I	W	S	A	E	-F	T	G	R	R	V	A	I	G	Q	A	I	F	E	L	F	A	L	D	N	A	K	G	Y	FS	-	GN	L	S	P	S	E	A	D	W	K	A	H	V	I	R	V	I	N	G	L	D	L	A	V	N	L	-----	E	D	P	K	A	L	E	E	L	K	H	L	A	R	Q	H	R	S	-----	G	V	K	A	Y	F	D	E	M	E	K	A	L									
GLB4_TYLHE/8-117	DRRE	V	Q	A	L	W	R	S	-I	W	S	A	E	-D	T	G	R	R	T	L	I	G	R	L	F	E	E	L	F	E	I	D	G	A	T	K	G	Y	FS	-	GN	L	S	P	S	E	E	F	A	H	V	L	R	V	N	G	L	D	T	L	I	G	V	L	-----	G	D	S	D	-T	L	N	S	L	I	D	H	L	A	E	Q	H	K	A	R	A	-----	G	F	K	T	Y	F	K	E	F	G	K	A	L							
GLB2_TYLHE/9-115	QR	L	K	V	K	Q	Q	W	A	-Y	G	V	-----	G	H	E	R	V	E	L	G	I	A	L	W	K	S	M	F	A	Q	D	N	A	R	D	Y	FS	-	GN	L	S	P	A	F	E	A	H	M	A	R	V	F	N	G	L	D	R	V	I	S	L	-----	T	D	E	P	V	L	N	A	Q	L	E	H	L	R	Q	H	I	K	L	G	-----	I	T	G	H	M	F	N	L	M	R	T	G	L										
GLB2_LUMTE/8-114	EGL	K	V	K	S	E	W	G	R	-A	Y	G	S	-----	G	H	D	R	E	A	F	S	Q	A	I	W	R	A	T	F	A	Q	V	P	E	S	R	S	Y	FS	-	GN	L	S	P	A	F	I	A	H	A	E	R	V	L	G	L	D	I	A	I	S	T	L	-----	D	Q	P	A	T	L	K	E	E	L	D	H	L	Q	V	Q	H	E	G	R	K	-----	I	P	D	N	F	D	A	F	K	T	A	I								
GLB_TUBTU/6-112	QR	F	V	K	H	Q	W	A	E	-A	F	G	T	-----	S	H	R	L	D	F	G	L	K	L	W	N	S	I	R	D	A	P	E	I	R	G	Y	FS	-	GN	L	S	A	E	F	E	A	H	A	E	R	V	L	G	L	D	M	T	I	S	L	-----	D	D	Q	A	F	D	A	Q	L	A	H	L	K	S	Q	H	A	E	R	N	-----	I	K	A	D	Y	G	V	F	N	E	L													
GLB3_LAMSP/7-113	QR	L	K	V	K	R	Q	W	A	E	-A	Y	G	S	-----	G	N	D	R	E	E	F	G	H	I	W	T	H	V	F	K	D	A	P	S	A	R	D	Y	FS	-	GN	L	S	H	T	P	A	F	R	A	H	A	T	R	V	L	G	L	D	M	C	I	A	L	-----	D	D	E	G	V	L	N	T	Q	L	A	H	L	A	S	Q	H	S	R	G	-----	V	S	A	A	Q	V	D	V	E	H	S	V								
GLB_PAREP/8-117	QD	I	L	L	K	E	L	G	P	-V	D	T	-----	P	A	H	I	V	E	T	G	L	G	A	I	H	A	L	F	T	A	H	P	Q	Y	I	H	S	R	L	EG	-H	T	I	E	N	V	M	Q	S	E	G	I	K	H	Y	A	R	T	L	T	E	A	I	V	H	M	L	K	E	I	-----	S	N	D	A	E	V	K	K	I	A	A	Q	Y	G	K	D	H	T	S	R	K	-----	V	T	K	D	E	F	M	S	G	E	P	I	F
Q21978_CAEL/165-283	S	C	E	V	A	D	S	W	R	L	-W	E	S	R	S	A	A	E	T	S	A	C	F	G	L	F	W	F	Q	R	V	E	S	K	I	P	M	L	R	P	-L	S	E	S	D	D	V	D	L	P	D	N	H	P	V	R	R	H	A	L	F	T	S	I	L	H	S	V	K	N	V	-----	D	E	L	E	A	Q	V	A	P	T	V	F	K	Y	G	E	R	H	Y	R	P	D	I	P	H	M	T	E	E	N	V	R			

# Sequence-profile alignments

- Position specific substitution matrices
- profile-hidden Markov models



# Building families (Pfam)



# Functions, organisms, structures

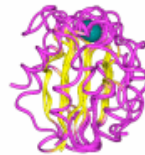
Marco Punta

Prokaryotes

The logo for TIGRFAMs, featuring the text "TIGRFAMs" in white, bold, sans-serif font centered on a solid blue rectangular background.

~4400 families

*Superfamily* 1.75  
HMM library and genome assignments server



Structural domains from SCOP

Signalling, extracellular and chromatin-associated proteins



~1000 domains

Gene3D  
Structural domains from CATH

# No limits, domains

Marco Punta

The logo for Pfam, consisting of the word "Pfam" in a bold, blue, sans-serif font.

~14000 families

The logo for ProDom, featuring the word "ProDom" in a bold, black, sans-serif font. The letters "o" and "o" are highlighted in green. Above the text are several horizontal lines in blue, green, and red. In the background, there is a semi-transparent box containing a protein sequence: "MSLQEEESIRTVQL" on the top line, "SIRTVQLQKNPLIS" on the bottom line, and "KLF...IS" in the middle.

# No limits, full-length proteins

Marco Punta

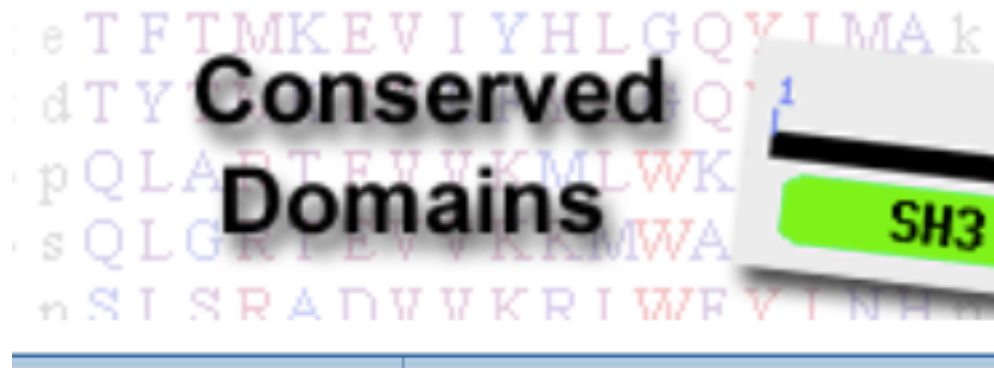


>7000 families, >50000 subfamilies



~2000 families

## CDD



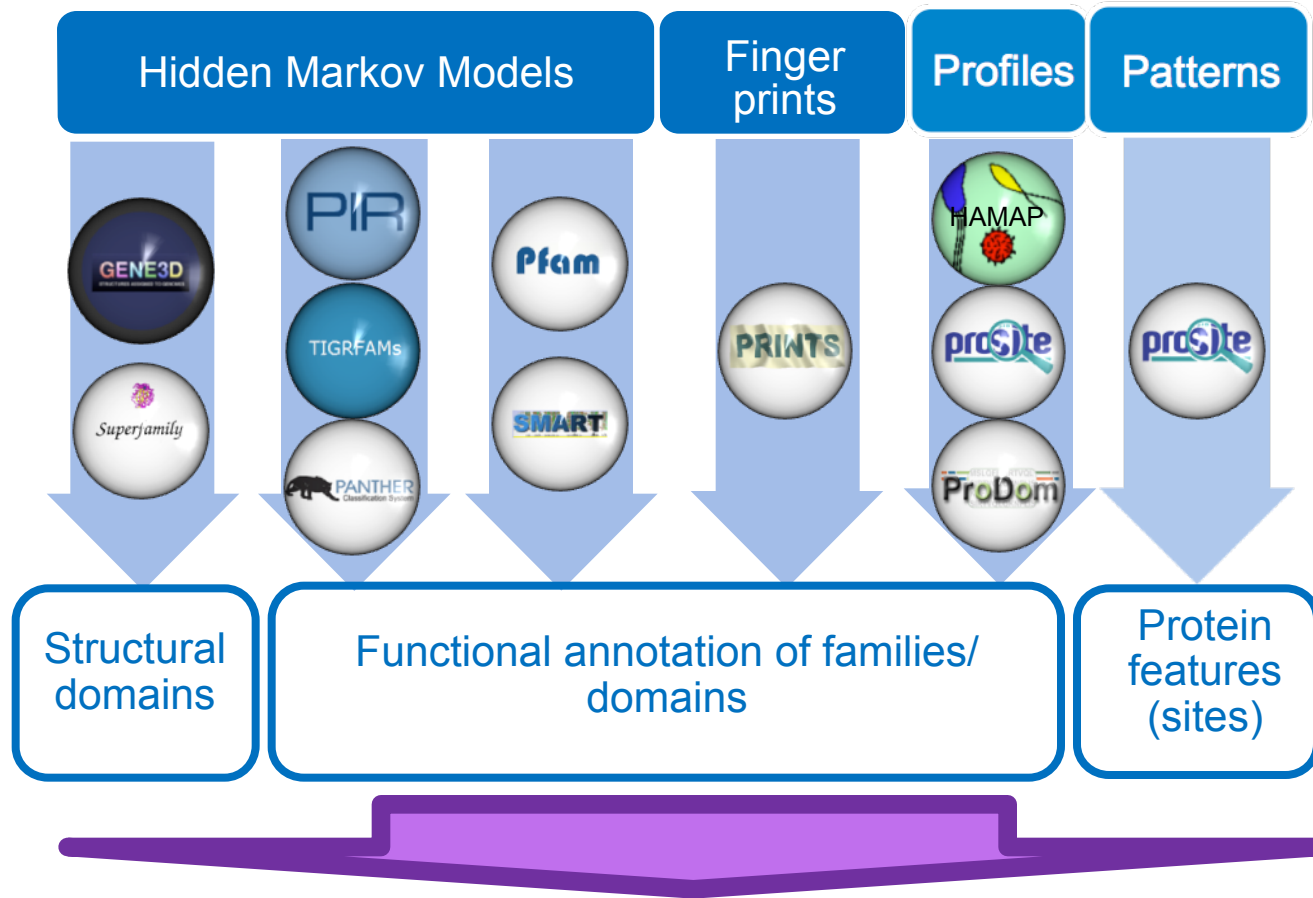
Uses RPS-BLAST

# Integration

Marco Punta



# Member databases



**Overview**

Similar proteins (2905)

Structures

**Filter view on**

**Entry type**

- Family
- Domains
- Repeats
- Site

**Status**

- Unintegrated

**Colour by**

- domain relationship
- source database

[help](#)

**P Protein**

**Phosphotransferase RcsD (P39838)**

**Accession** [P39838](#) (RCSL\_ECOLI)  
**Species** Escherichia coli (strain K12)  
**Length** 890 amino acids (complete)

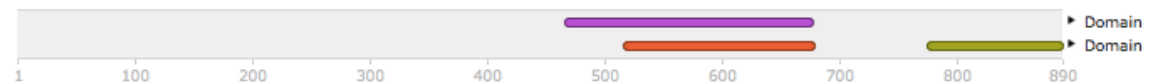
Export FASTA

Source: UniProtKB

**Protein family membership**

None predicted.

**Domains and repeats**



**Detailed signature matches**

- IPR005467** Signal transduction histidine kinase, core
  - ▶ PS50109 (HIS\_KIN)
- IPR003594** Histidine kinase-like ATPase, C-terminal domain
  - ▶ SSF55874
  - ▶ G3DSA:3.30.56...
  - ▶ SM00387 (HATPase\_c)
  - ▶ PF02518 (HATPase\_c)
- IPR008207** Signal transduction histidine kinase, phosphotransfer (Hpt) domain
  - ▶ SM00073 (HPT)
  - ▶ G3DSA:1.20.12...
  - ▶ SSF47226
  - ▶ PS50894 (HPT)
  - ▶ PF01627 (Hpt)
- no IPR** Unintegrated signatures
  - ▶ PB000390 (Pfam-B\_390)
  - ▶ PB001071 (Pfam-B\_1071)
  - ▶ PB002242 (Pfam-B\_2242)
  - ▶ PTHR26402
  - ▶ PTHR26402:SF483



**Overview**

Similar proteins (2905)

Structures

**Filter view on**

**Entry type**

- Family
- Domains
- Repeats
- Site

**Status**

- Unintegrated

**Colour by**

- domain relationship
- source database

[help](#)

**Protein**

**Phosphotransferase RcsD (P39838)**

**Accession** [P39838](#) (RCSL\_ECOLI)  
**Species** Escherichia coli (strain K12)  
**Length** 890 amino acids (complete)

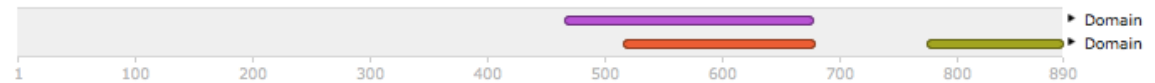
Export FASTA

Source: UniProtKB

**Protein family membership**

None predicted.

**Domains and repeats**



**Detailed signature matches**

- IPR005467** Signal transduction histidine kinase, core  
  - ▶ PS50109 (HIS\_KIN)
- IPR003594** Histidine kinase-like ATPase, C-terminal domain  
  - ▶ SSF55874
  - ▶ G3DSA:3.30.56...
  - ▶ SM00387 (HATPase\_c)
  - ▶ PF02518 (HATPase\_c)
- IPR008207** Signal transduction histidine kinase, phosphotransfer (Hpt) domain  
  - ▶ SM00073 (HPT)
  - ▶ G3DSA:1.20.12...
  - ▶ SSF47226
  - ▶ PS50894 (HPT)
  - ▶ PF01627 (Hpt)
- no IPR** Unintegrated signatures  
  - ▶ PB000390 (Pfam-B\_390)
  - ▶ PB001071 (Pfam-B\_1071)
  - ▶ PB002242 (Pfam-B\_2242)
  - ▶ PTHR26402
  - ▶ PTHR26402:SF483

**Overview**

Similar proteins (2905)

Structures

**Filter view on**

**Entry type**

- Family
- Domains
- Repeats
- Site

**Status**

- Unintegrated

**Colour by**

- domain relationship
- source database

[help](#)

**Protein**

**Phosphotransferase RcsD (P39838)**

**Accession** [P39838](#) (RCSL\_ECOLI)  
**Species** Escherichia coli (strain K12)  
**Length** 890 amino acids (complete)

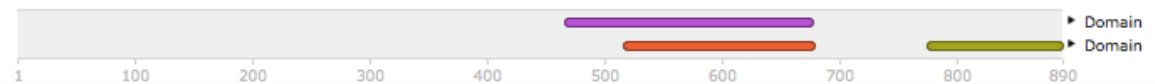
Export FASTA

Source: UniProtKB

**Protein family membership**

None predicted.

**Domains and repeats**



**Detailed signature matches**

<b>IPR005467</b>	Signal transduction histidine kinase, core		▶ PS50109 (HIS_KIN)
<b>IPR003594</b>	Histidine kinase-like ATPase, C-terminal domain		▶ SSF55874 ▶ G3DSA:3.30.56... ▶ SM00387 (HATPase_c) ▶ PF02518 (HATPase_c)
<b>IPR008207</b>	Signal transduction histidine kinase, phosphotransfer (Hpt) domain		▶ SM00073 (HPT) ▶ G3DSA:1.20.12... ▶ SSF47226 ▶ PS50894 (HPT) ▶ PF01627 (Hpt)

**no IPR** Unintegrated signatures

	▶ PB000390 (Pfam-B_390) ▶ PB001071 (Pfam-B_1071) ▶ PB002242 (Pfam-B_2242) ▶ PTHR26402 ▶ PTHR26402:SF483
--	---

**Overview**

Similar proteins (2905)

Structures

**Filter view on**

**Entry type**

- Family
- Domains
- Repeats
- Site

**Status**

- Unintegrated

**Colour by**

- domain relationship
- source database

[help](#)

**P Protein**

**Phosphotransferase RcsD (P39838)**

**Accession** [P39838](#) (RCSL\_ECOLI)  
**Species** Escherichia coli (strain K12)  
**Length** 890 amino acids (complete)

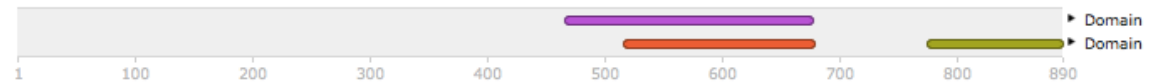
Export FASTA

Source: UniProtKB

**Protein family membership**

None predicted.

**Domains and repeats**



**Detailed signature matches**

- IPR005467** Signal transduction histidine kinase, core  
  - ▶ PS50109 (HIS\_KIN)
- IPR003594** Histidine kinase-like ATPase, C-terminal domain  
  - ▶ SSF55874
  - ▶ G3DSA:3.30.56...
  - ▶ SM00387 (HATPase\_c)
  - ▶ PF02518 (HATPase\_c)
- IPR008207** Signal transduction histidine kinase, phosphotransfer (Hpt) domain  
  - ▶ SM00073 (HPT)
  - ▶ G3DSA:1.20.12...
  - ▶ SSF47226
  - ▶ PS50894 (HPT)
  - ▶ PF01627 (Hpt)

- no IPR** Unintegrated signatures  
  - ▶ PB000390 (Pfam-B\_390)
  - ▶ PB001071 (Pfam-B\_1071)
  - ▶ PB002242 (Pfam-B\_2242)
  - ▶ PTHR26402
  - ▶ PTHR26402:SF483



P39838

Search

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

[Home](#) [Search](#) [Release notes](#) [Download](#) [About InterPro](#) [Help](#) [Contact](#)**Overview**[Proteins matched](#) (402223)[Domain organisations](#) (17400)[Pathways & Interactions](#)[Species](#)[Structures](#)[Literature](#) (21)[Cross-references](#) (2)**D Domain****Signal transduction histidine kinase, core (IPR005467)***Short name: Sig\_transdc\_His\_kinase\_core***Domain relationships**

None.

**Description**

Most prokaryotic signal-transduction systems and a few eukaryotic pathways use phosphotransfer schemes involving two conserved components, a histidine protein kinase (HK) and a response regulator protein (RR). The HK, which is regulated by environmental stimuli, autophosphorylates at a histidine residue, creating a high-energy phosphoryl group that is subsequently transferred to an aspartate residue in the RR domain. Phosphorylation induces a conformational change in RR that results in activation of an associated domain that effects the response.

Both prokaryotic and eukaryotic HKs contain the same basic signaling components, namely a diverse sensing domain and a highly conserved kinase core that has a unique fold, distinct from that of the Ser/Thr/Tyr kinase superfamily. The overall activity of the kinase is modulated by input signals to the sensing domain. HKs undergo an ATP-dependent autophosphorylation at a conserved His residue in the kinase core. Autophosphorylation is a bimolecular reaction between homodimers, in which one HK monomer catalyzes the phosphorylation of the conserved His residue in the second monomer.

The sensing domains are variable in sequence, reflective of the many different environmental signals to which HKs are responsive, whereas the about 250-residue kinase core is more conserved. The kinase core is composed of a dimerization domain and an ATP/ADP-binding phosphotransfer or catalytic domain and can be identified by five conserved primary sequence motifs present in both eukaryotic and prokaryotic HKs. These motifs have been termed the H, N, G1, F and G2 boxes. The conserved His substrate is the central feature in the H box, whereas the N, G1, F and G2 boxes define the nucleotide binding cleft. In most HKs, the H box is part of the dimerization domain. However, for some proteins, like CheA, the conserved His is located at the far N terminus of the protein in a separate HPT domain. The N, G1, F and G2 boxes are usually contiguous, but the spacing between these motifs is somewhat varied. The catalytic core forms an alpha-beta sandwich consisting of five antiparallel beta strands and three alpha helices [[PMID: 10966457](#)] [[PMID: 11406410](#)] [[PMID: 11369791](#)]

[Add your annotation](#)**Contributing signatures**

Signatures from InterPro member databases are used to construct an entry.

**PROSITE profiles**[PSS0109 \(HIS\\_KIN\)](#)

# Orthologous families, trees

**OMA**  
browser

**OrthoDB**

**COGs**  
Phylogenetic classification of proteins encoded in complete genomes

**eggNOG**  
version 3.0



Pros:

Better prediction of protein function (in principle, ortholog conjecture)

Gene history

Species trees

Caveats:

Lateral gene transfer difficult to model/recognise -> bacteria difficult


Gene loss difficult to account for, may lead to wrong ortho-para assignment

Large families difficult to model

# Team Exercise

## Building a new Pfam family

## QUICK SEARCH

Paste in your sequence or use the [example](#) 

search against

Reference Proteomes  UniProtKB  SwissProt  Pfam

**Submit**

Reset

[Alternative Search Options](#)

The [HMMER web server](#): fast and accurate  
This site has been designed to provide maximum flexibility  
coupled with **intuitive and** interactive

Quickstart



### Blog News

August, 2015

hmmmer.org is updating

hmmmer.org is moving off of Janelia tonight, into the great cloud. You may see some flakiness as DNS nameservers update.

### Download HMMER

Get the latest version

**v3.1b2**

### Recent

[HMMER web server](#)  
R.D. Finn, J. Hotelling,  
F. Schreiber,  
**Nucleic Acids Res**  
2015

[phmmer](#)[hmmscan](#)[hmmsearch](#)[jackhmmer](#)

## protein sequence vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#)

### ▼ Sequence Database

#### Frequently used databases

Reference Proteomes  UniProtKB  SwissProt  PDB

#### Representative Sets (UniProt)

rp75  rp55  rp35  rp15

#### Other databases

QfO  Pfamseq

▶ [Restrict by Taxonomy](#)





# protein sequence vs protein sequence database

Paste a Sequence | Upload a File | Accession Search

Paste in your sequence or use the [example](#)

```
HEAIGSGDLRLRSFRRRTSLAGAGRRTSDSHEDAGTLDFSSLLKKRD
SFRRDSKLEAPAEEDVWEILROAPPSEYERIAFOHGVTDLRGMLKRL
KGMKQDEKK
```

Submit Reset

## Sequence Database

Frequently used databases

Reference Proteomes  UniProtKB  SwissProt  PDB

Representative Sets (UniProt)

rp75  rp55  rp35  rp15

Other databases


QfO  Pfamseq

▶ Restrict by Taxonomy



## protein sequence vs profile-HMM database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Paste in your sequence or use the [example](#) 

```
HEAIGSGDLDLRSAFRRTSLAGAGRRTSDSHEDAGTLDFSSLLKCRD  
SFRRDSKLEAPAEEDYWEILROAPPSEYERIAFOHGVTDLRGMLKRL  
KGMKQDEKK
```

**Submit**

Reset

### ▼ HMM Database

#### Protein Families

- Pfam  TIGRFAM  Gene3D  Superfamily  PIRSF  
(select all)

## PHMMER Results

[Score](#) [Taxonomy](#) [Domain](#) [Download](#)
[Search Again](#)

## Sequence Matches and Features ?

Pfam  103hit coverage hit similarity 
 disorder
  coiled-coil
  tm & signal peptide ?
[Show hit details](#)

## Distribution of Significant Hits ?


 Bacteria
  Eukaryota
  Archaea
  Viruses
  Unclassified Sequences
  Other Sequences

[« First](#)
[« Previous](#)
**Page 1**
[of 4](#)
[Next »](#)
[Last »](#)

## Significant Query Matches (330) in uniprotrefprot (v.2015-06-24)

[Customize](#)

	Target	Species	E-value
>	<a href="#">Q3UIK0_MOUSE</a>	<a href="#">Mus musculus</a>	1.3e-60
>	<a href="#">E9Q9T8_MOUSE</a>	<a href="#">Mus musculus</a>	3.0e-58
>	<a href="#">Q3TF37_MOUSE</a>	<a href="#">Mus musculus</a>	1.0e-57
>	<a href="#">MYPC_RAT</a>	<a href="#">Rattus norvegicus</a>	2.0e-56
>	<a href="#">M3XYE3_MUSPF</a>	<a href="#">Mustela putorius furo</a>	9.1e-55

## UniProtKB - Q3UIK0 (Q3UIK0\_MOUSE)

**Protein** | Submitted name: **Myosin-binding protein C, cardiac-type****Gene** | **Mybpc3****Organism** | *Mus musculus (Mouse)***Status** | Unreviewed - Annotation score: - Experimental evidence at protein level<sup>i</sup>Display None

BLAST



Align



Format



Add to basket



History



Help video



Other tutorials and videos



Feedback

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

- identical protein binding Source: MGI
- myosin binding Source: MGI
- myosin heavy chain binding Source: MGI ▾
- structural constituent of cytoskeleton Source: MGI ▾

## Sequence Matches and Features ?

Pfam  103

disorder  coiled-coil  tm & signal peptide ?

No hits were found for your query.

### TIGRFAM Matches

[Advanced](#)

Family		Description	Start <span>▼</span>	End <span>▼</span>	Domain E-values	
Id <span>▼</span>	Accession				Ind. <span>▼</span>	Cond. <span>▼</span>

No hits were found for your query.

Your search took: 0.05 seconds

### Gene3D Matches

[Advanced](#)

Family		Description	Region	Start <span>▼</span>	End <span>▼</span>	Domain E-values	
Id <span>▼</span>	Accession					Ind. <span>▼</span>	Cond. <span>▼</span>

No hits were found for your query.

Your search took: 0.05 seconds

\* These hmmscan results have been modified by the Gene3D DomainFinder post processing program.

### Superfamily Matches

Superfamily			Family			Region	Model Match								
Accession	Description	E-value	Accession	Description	E-value		Start	End	Alignment		Model			Bit Score	Domain E-values
						Start			End	Start	End	Length	Ind		Cond
No hits were found for your query.															

Your search took: 0.05 seconds

\* These hmmscan results have been modified by Superfamily post-processing and family assignment code.

### PIRSF Matches

[Advanced](#)

Family			Subfamily			Region
Accession	Description	E-value	Accession	Description	E-value	
No hits were found for your query.						

Your search took: 0.04 seconds

\* These hmmscan results have been modified by PIRSF post-processing and family assignment code.

Search InterPro...

Search

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

[Home](#)
[Search](#)
[Release notes](#)
[Download](#)
[About InterPro](#)
[Help](#)
[Contact](#)
**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

**P** ProteinExport  Select format 

## Submitted

**Length** 103 amino acids

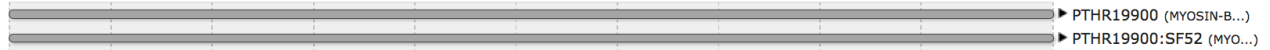
## Protein family membership

None predicted.

## Domains and repeats

None predicted.

## Detailed signature matches

**?** no IPR Unintegrated signatures

## GO term prediction

Biological Process

None predicted.

# UniProtKB - Q3UIK0 (Q3UIK0\_MOUSE)

**Protein** | Submitted name: **Myosin-binding protein C, cardiac-type**

**Gene** | **Mybpc3**

**Organism** | *Mus musculus (Mouse)*

**Status** | Unreviewed - Annotation score: - Experimental evidence at protein level<sup>i</sup>

Display None

BLAST
 Align
 Format
 Add to basket
 History

Help video
 Other tutorials and videos

Feedback

## Function<sup>i</sup>

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

- [identical protein binding](#) Source: MGI
- [myosin binding](#) Source: MGI
- [myosin heavy chain binding](#) Source: MGI ▾
- [structural constituent of cytoskeleton](#) Source: MGI ▾

Function

Names & Taxonomy

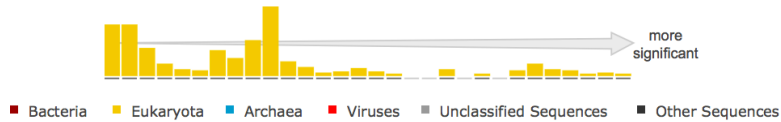
Subcell. location

Pathol./Biotech



[Show hit details](#)

**Distribution of Significant Hits** 



**Significant Query Matches (330) in uniprotrefprot (v.2015-06-24)**

[Customize](#)

		Target				Species					E-value	
v		<a href="#">Q3UIK0_MOUSE</a>				<a href="#">Mus musculus</a>					1.3e-60	
Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value	
start	end	start	end	start	end						Ind.	Cond.
1	103	263	365	263	365	1.99	1.00	100.0 (103)	100.0 (103)	209.4	3.1e-60	1.1e-64

```

.....*.....*.....*.....*.....*.....*.....*.....*.....*
Query 1  heaigsgdldlr safrrtslagagr rtsdshedag tldfssllkkrdsfrrdskleapaeedvweilrqappseyeriaf 80
Target 263 HEAIGSGDLDLRS AFRRTSLAGAGR RRTSDSHEDAG TLDFSSLLKKRDSFR RDSKLEAPAEEDVWEILRQAPPSEYERIAF 342
PP 9*****

```

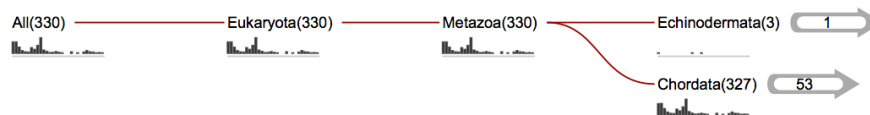
```

.....*.....*...
Query 81 qhgvtdlrgmlkrlkgmkqdekk 103
Target 343 QHGVTDLRGMLKRLKGMKQDEKK 365
PP *****98

```



## PHMMER Results

Taxonomic distribution of all search hits [?](#)

## Species Distribution

Species	Count	View
<a href="#">Takifugu rubripes</a>	22	<a href="#">Show</a>
<a href="#">Danio rerio</a>	19	<a href="#">Show</a>
<a href="#">Callithrix jacchus</a>	18	<a href="#">Show</a>
<a href="#">Mus musculus</a>	14	<a href="#">Show</a>
<a href="#">Oreochromis niloticus</a>	12	<a href="#">Show</a>
<a href="#">Homo sapiens</a>	11	<a href="#">Show</a>
<a href="#">Gasterosteus aculeatus</a>	11	<a href="#">Show</a>
<a href="#">Canis lupus familiaris</a>	10	<a href="#">Show</a>
<a href="#">Macaca mulatta</a>	10	<a href="#">Show</a>
<a href="#">Astyanax mexicanus</a>	9	<a href="#">Show</a>
<a href="#">Gorilla gorilla gorilla</a>	8	<a href="#">Show</a>
<a href="#">Tetraodon nigroviridis</a>	8	<a href="#">Show</a>



## PHMMER Results

[Jump to the exact match for your query architecture](#)

Domain Architectures [?](#)

« First « Previous **Page 1** of 2

**94**  
SEQUENCES

with domain architecture: **I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, *example:Q3TF37\_MOUSE*[↗](#)

[Show All](#)

Sequence Features  1113

**35**  
SEQUENCES

with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, *example:F6ZHP7\_HORSE*[↗](#)

[Show All](#)

Sequence Features  963

**34**  
SEQUENCES

with domain architecture: **I-set, I-set, I-set, I-set, I-set, I-set, fn3, fn3, I-set, fn3, I-set**, *example:Q3UIK0\_MOUSE*[↗](#)

[Show All](#)

Sequence Features  1278

**19**  
SEQUENCES

with domain architecture: **I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set**, *example:G1Q885\_MYOLU*[↗](#)

[Show All](#)

Sequence Features  1245

**12**  
SEQUENCES

with domain architecture: **I-set, I-set, I-set, I-set, fn3, fn3, fn3, I-set, fn3, I-set**, *example:W5MUP3\_LEPOC*[↗](#)

[Show All](#)

Sequence Features  1529

**10**

with domain architecture: **I-set, I-set, I-set**, *example:F7CWG3\_CAI\_1A*[↗](#)

## PHMMER Results

- **Job:** 5B650320-65F7-11E5-8E90-C8CCF69F8818.1
- **Started:** 2015-09-28 16:41:18
- **Algorithm:** phmmer
- **HMMER Options:** -E 1 --domE 1 --incE 0.01 --incdomE 0.03 --mx BLOSUM62 --pextend 0.4 --popen 0.02 --seqdb uniprotrefprot

### ▼ Format

#### Text

A plain text file containing the hit alignments and scores.



#### Tab Delimited

A tab delimited text file containing the hit information. No alignments.



#### XML

An XML file formatted for machine parsing of the data.



#### JSON

All the results information encoded as a single json string.



#### FASTA

Download the significant hits from your search as a gzipped FASTA file.



#### Full length FASTA

A gzipped file containing the full length sequences for significant search hits.



#### Aligned FASTA

A gzipped file containing aligned significant search hits in FASTA format.



#### STOCKHOLM

Download an alignment of significant hits as a gzipped STOCKHOLM file.



# 1. File -> Input alignment -> From file

File Edit Select View Annotations Format Colour Calculate Web Service

	10	20	30	40	50	60	70	80	90	100	110																		
Query/1-103	HEAIGS	GDLDLRS	AFRRT	S	L	A	G	AG	R	T	S	D	S	H	ED	A	G	T	L	D	F	S	S	L	L	K	K	R	N
H3B137_LATCH/266-360	-EASTS	-EEMDIR	AAFRRT	G			G				A	D	G	S	EE	A	G	E	L	D	F	S	A	L	L	K	K	R	N
F1QV58_DANRE/273-371	HEARTT	EGFDIRT	AFRRT	S							T	D	A	G	DD	S	G	E	L	D	F	S	A	L	L	K	K	R	N
F6ZHP7_HORSE/1-53																													
F1N8Z9_CHICK/247-343	-EAPVS	-GEMDIR	AAFRRT								T	E	G	L	EE	S	G	E	L	N	F	S	A	L	L	K	K	R	N
G3P589_GASAC/262-375	HEACAA	EGFDIRA	AFRRT	E	R	A	G	CA	K	h	g	l	f	s	N	S	N	D	G	K	ED	S	G	E	L	D	F	S	T
W5KF69_ASTMX/288-396	HEARAV	EGFDIRA	AFRRT	S	V	S	T	GG	K			K	r	m	s	g	l	s	T	D	G	G	DD	S	G	E	L	D	F
H3CVJ4_TETNG/260-370		EGLDIRA	AFRRT	S	E	A	k	TN	R	ie	t	k	Y	S	T	D	G	K	ED	S	G	E	L	D	F	S	T	L	L
V8PGT3_OPHHA/218-307	HEAPAT	ADLDIR	SAFRRT								V	D	G	Q	DE	G	R	E	L	D	F	T	T	L	L	R	K	R	
H2V4W4_TAKRU/111-197		EGLDIRA	AFRRT	S							T	D	G	K	ED	S	G	E	L	D	F	S	T	L	L	K	K	R	
H2TW89_TAKRU/93-190	--QGS	QNIDIR	SAFKRR	S	L	L	V	N	N	S	S	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K	H	R	
G3P586_GASAC/152-230					I	R	A	AF	R	R	T	N	D	G	K	ED	S	G	E	L	D	F	S	T	L	L	K	K	
G3VYD1_SARHA/186-279	HESTGT	PNIDIR	SAFKRS	N	N	S	G	EG				Q	ED	A	G	E	L	D	F	S	G	L	L	K	K	R	R		
F6PYQ5_ORNAN/158-246		PSIDIR	SAFKRS	K	N	S	G	EG				Q	ED	A	G	E	L	D	F	S	G	L	L	K	K	R	R		
F6PY53_ORNAN/195-283		PSIDIR	SAFKRS	K	N	S	G	EG				Q	ED	A	G	E	L	D	F	S	G	L	L	K	K	R	R		
H2V4W3_TAKRU/115-196		LDIRA	AFRRT	S							T	D	G	K	ED	S	G	E	L	D	F	S	T	L	L	K	K		
G3XAI7_HUMAN/193-275		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F1Q615_DANRE/273-362	HEARTT	EGFDIRT	AFRRT	S							T	D	A	G	DD	S	G	E	L	D	F	S	A	L	L	K	K		
M7B656_CHEMY/130-222	--A	PSIDIR	SAFKRR	F	D	L	F	F	I	N	N	S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L		
F6Y6C7_HORSE/69-151		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F6Y6F3_HORSE/69-151		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
G1R3C3_NOMLE/195-275				D	I	R	S	AF	K	R	S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
G3V1V7_HUMAN/69-151		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
MYPC1_HUMAN/168-250		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F8VZYD_HUMAN/181-263		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
G3RPA3_GORGO/168-250		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F7IK85_CAJJA/193-275		SIDIR	SAFK						R		S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
H2LNL0_ORYLA/260-380	HEARAP	GGLDIR	TAFRRT		H	t	n	L	i	v	m	e	k	k	p	c	v	P	F	A	V	C	S	T	D	G	N		
H2Q6Q1_PANTR/193-275		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F7CWG3_CAJJA/181-263		SIDIR	SAFK						R		S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
H0W0Z3_CAVPO/197-279		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
H0V1L5_CAVPO/197-279		NIDIR	SAFKR								S	G	E	G	Q	ED	A	G	E	L	D	F	S	G	L	L	K		
F1SR18_PIG/160-241		IDIRSA					F	K	R		S	G	E	G	Q	DD	A	G	E	L	D	F	S	G	L	L	K		
I3JTU4_ORENI/148-228				N	I	L	E	AF	K	R	S	G	E	D	A	D	ED	A	G	E	L	D	F	S	A	L	L		
F6W1U5_MACMU/142-224		NIDIR	SAFKR								S	G	E	G	Q	DD	A	G	E	L	D	F	S	G	L	L	K		
H2N1E5_PONAB/193-275		NIDIR	SAFKR								S	G	E	G	Q	DD	A	G	E	L	D	F	S	G	L	L	K		

Conservation



Quality



Consensus



HEARGS - ENIDIRSAFKRT - - N - - T - - - R - S - AF - K - - - R - M - - - I N - S - - - G - - - E G G - Q S E D F K A R K G - - E T P N L D F S G L L K K - - K - R L - - - S E

1. Edit -> Remove redundancy

2. Select 90% and Remove

1. Edit -> Remove redundancy

2. Select 90% and Remove

3. Edit -> Remove empty columns

1. Edit -> Remove redundancy

2. Select 90% and Remove

3. Edit -> Remove empty columns

4. Colour -> Clustalx

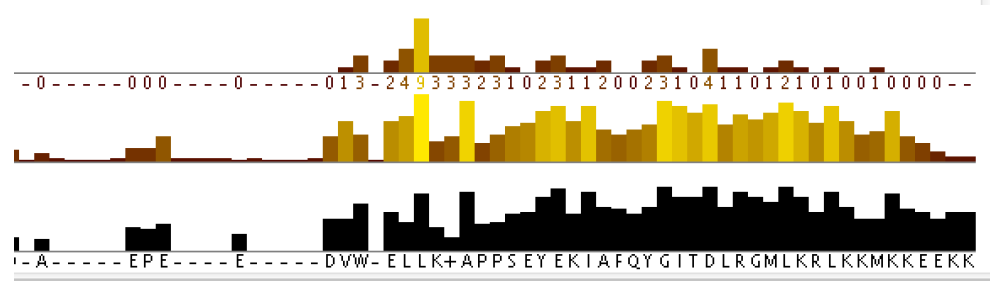




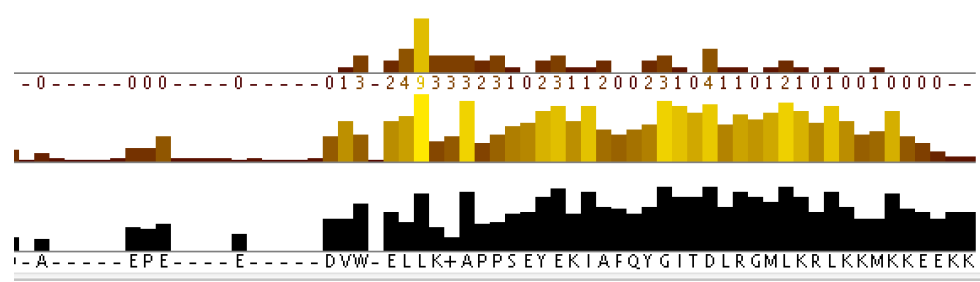
```

120      130      140      150      160      170
:-A-----PAE-----E-----DVW-EILRQAPPSEYERIAFQHGVTDLRGMLKRLKGMKDEKK
:-S-----EPD-----I-----DVW-DLLRQAPPSEYERIAFQYGITDLRGMLKRLKMKKKEKK
:-S-----QPD-----V-----DVW-EILRKAPPSEYERIAFQYGITDLRGMLKRLKRIKKEEKK
:-A-----EPD-----M-----DVW-NILSHAPSEYERIAFQHGVTDLRGMLKRLKMKKKEKK
:-S-----EPD-----V-----DVW-EILSHAPSEYERIAFQYGITDLRGMLKRLKMKKKEKK
:-Avhvs sEPE-----V-----DVW-EILSKAPSEYERIAFQHGVTDLRGMLKRLKMKKKEKK
:-V-----HSE-----Pd-----vDVW-SILSKAPSAFEKIAFQYGITDLRGMLKRLKMKKDEKK
:-V-----PAR-----K-----REK-DNITQRPT-----
:-Qd-----ePE-----I-----DVW-ELLKNANPNEYERIAFQYGITDLRGMLKRLKRMRRVEKK
:-T-----TAE-----K-----KKLI LKMGHAPPSEFERIAFQYGITDLRGMLKRLKMKKKEKK
:-D-----TPE-----V-----DVW-EILKKARPDEYERIAFTYGITDLRGMLLRRMKKIPEKK
:-C-----GIP-----P-----DVW-ELLKNAPSEDFERIAFEHGVTDLRGMLKRLKRVKKEVKK
:kE-----EPE-----I-----DVW-ELLKSAHPSEYERIAFQYGITDLRGMLKRLKMKKVE--
:-K-----DDD-----Dl g i p pEIW-ELLKGAKKSEYERIAFQYGITDLRGMLKRLKKAKEVKK
:-L-----GIP-----P-----EIW-ELLKGAKKSEYERIAFQYGITDLRGMLKRLKKAKEVKK
i-I-----PPE-----I-----IW-ELLKGAKKSEYERIAFQYGITDLRGMLKRLKKAKEVKK
:-M-----ETE-----E-----KVV-EILLSADKKDYERICAEYGITDFRGMLKRLNEMKKE--
:-F-----SAL-----L-----KAT-KKLKSAHPSEYERIAFQYGITDLRGMLKRLKMKKVE--
:-A-----DAK-----E-----DIW-ALLKSANPREYDRIAFYWGIKDLRKLKLANAKNNKK
|-A-----PEKkidIE-----QVW-QLLMTADRKDYEQICMKYGIVDYRGMLKRLQEMKKEQ--
:-K-----PI-----I-----DIM-ELLKNVDPKEYEKYARMYGITDFRGLLQAFELLKQSQ--
|-Q-----QKEgeidP-----KLL-ELLLSAPKKDYERICLEFGITDFRWFLLKQLKKE--
:-M-----EAD-----E-----KFF-EVLMSEAKKDYERICIQYGVTDYRGMLKRLNKKIE--
:-G-----EID-----P-----KFW-DVMLNAKSDYERICHEFGITDYRWMLKQLNKKKEK-
:-V-----IDE-----K-----EML-EILSKVPPKDFERVCMVYGFDFWGLLKKLEMKKVEEK
:-K-----E-----AIF-QLLLHADKKDYERICIKYGISDFRGMLRALQDLRKDT-
-----ESD-----E-----RFW-DVMLKADRNDYERICSEFGVKDLHSLKLLDEKKE--

```

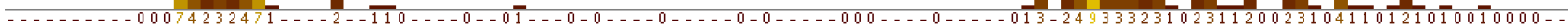


	120		130		140		150		160		170
-A-	PAE	-E-	DVW	EILRQAPPSEYER	IAFQHGVTDLRGMLKRLKGMKDEKK						
-S-	EPD	-I-	DVW	DLLRQAPPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-S-	QPD	-V-	DVW	EILRKAPPSEYER	IAFQYGITDLRGMLKRLKRIKKEEKK						
-A-	EPD	-M-	DVW	NILSHAPSEYER	IAFQHGVTDLRGMLKRLKMKKEEKK						
-S-	EPD	-V-	DVW	EILSHAPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-Avhvs	EPE	-V-	DVW	EILSKAPSEYER	IAFQHGVTDLRGMLKRLKMKKEEKK						
-V-	HPE	-Pd	DVW	SILSKAPPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-V-	PAR	-K-	REK	DNITQRPT							
-Qd-	ePE	-I-	DVW	ELLKNANPNEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-T-	TAE	-K-	KKL	LKMGHAPPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-D-	TPE	-V-	DVW	EILKKARPEYER	IAFTYGITDLRGMLKRLKMKKEEKK						
-C-	GIP	-P-	DVW	ELLKNAPSEYER	IAFEHGVTDLRGMLKRLKMKKEEKK						
kE-	EPE	-I-	DVW	ELLKSAPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-K-	DDD	-D i p	EIW	ELLKGAKKSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-L-	GIP	-P-	EIW	ELLKGAKKSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
i-I-	PPE	-I-	EIW	ELLKGAKKSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-M-	EIE	-E-	KVW	EILLSADKKDYER	ICAEYGITDFRGMLKRLKMKKEEKK						
-F-	SAL	-L-	KAT	KKLSAPSEYER	IAFQYGITDLRGMLKRLKMKKEEKK						
-A-	DAK	-E-	DIW	ALLKSANPREYDR	IAFYWGIKDLRKLKLANAKNNKK						
-A-	PEKkidIE	-E-	QVW	QLLMTADRKYEQ	ICMKYGIVDYRGMLKRLKMKKEEKK						
-K-	PI-	-E-	DIM	ELLKNVDPKEYER	YARMYGITDFRGLLAFELLLKSSQ						
-Q-	QKEgeidP	-E-	KLL	ELLLSAPKKDYER	ICLEFGITDFRWFLLKQLKKEER						
-M-	EAD	-E-	KFF	EVLMSEAKKDYESI	CIQYGVTDYRGMLKRLKMKKEEKK						
-G-	EID	-P-	KFW	DVMLNAKSDYER	ICHEFGITDYRWMLKQLNLKKEEKK						
-V-	IDE	-K-	EML	EILSKVPPKKDYER	VCMVYGFDFWGLKRLKMKKEEKK						
-K-	E-	-E-	AIF	QLLHADKKDYER	ICIKYGISDFRGMLRALDLRKT						
-E-	SD	-E-	RFW	DVMLKADRNDYER	ICSEFGVKDLHSLKLLDEKKEEKK						



File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service
Query/1-103	S	D	-	S	H	E	D	A
H3B137_LATCH/266-360	A	D	-	G	S	E	E	A
F1NBZ9_CHICK/247-343	T	E	-	G	L	E	E	S
G3P589_GASAC/262-375	N	D	-	G	K	E	D	S
W5KF69_ASTMX/288-396	T	D	-	G	G	D	D	S
H3CVJ4_TETNG/260-370	T	D	-	G	K	E	D	S
H2LNLO_ORYLA/260-380	T	D	-	G	N	D	D	S
G3UID8_LOXAF/265-336	S	D	-	S	H	E	D	A
G1PGG8_MYOLU/108-217	G	E	-	G	Q	E	D	A
MAA199_XIPMA/183-279	D	-	G	K	D	D	S	-
H2TW88_TAKRU/94-206	S	E	-	G	Q	E	D	A
F7DKM1_XENTR/161-262	-	-	-	D	E	D	A	-
H2UDR5_TAKRU/132-236	S	D	-	A	G	E	D	E
E2R072_CANFA/152-254	G	E	-	G	K	S	E	D
H0VYP7_CAVPO/151-251	E	G	-	K	S	E	D	A
F7EU62_CALLA/121-207	G	D	-	P	R	R	R	L
I3K8Q9_ORENI/154-222	-	-	-	S	A	-	-	-
H2UDR3_TAKRU/130-195	-	-	-	K	A	D	I	L
E4WXZ2_OKDI/125-189	-	-	-	F	R	A	A	L
F7FUF4_MACMU/92-170	-	-	-	P	Q	E	D	L
M7AMC7_CHEMY/10184-10245	-	-	-	D	R	A	K	L
MAAU85_XIPMA/120-195	G	-	G	-	S	N	-	E
H2RX04_TAKRU/156-219	-	-	-	E	D	F	K	K
E7FRW2_DANRE/153-222	-	-	-	D	P	E	D	F
G3U743_LOXAF/163-232	-	-	-	D	K	M	D	F
G1MT65_MELGA/145-206	-	-	-	Q	D	L	K	K
I3XJU9_ORENI/137-197	-	-	-	P	A	D	F	R

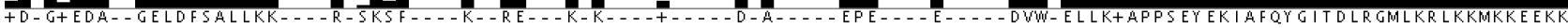
Conservation



Quality



Consensus



1. Edit -> Delete

2. Colour -> BLOSUM62 Score / Colour -> Percentage Identity

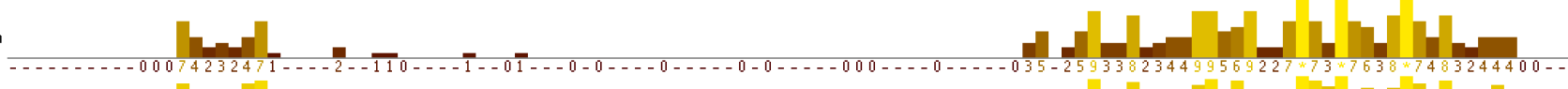
File Edit Select View Annotations Format Colour Calculate Web Service

```

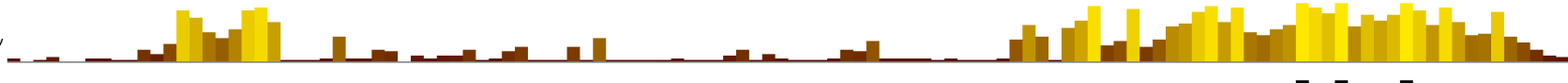
70      80      90      100     110     120     130     140     150     160     170
Query/1-103      SD-SHEDA--GTLDFSSLLKK---R--DSF---R--RD---S-K---L---E-A---PAE---E---DVW-EILRQAPPS EYERIAFQHGVTDLRGMLKRLKGMKQDEKK
H3B137_LATCH/266-360 AD-GSEEA--GELDF SALLKK---R--NNE--K--EV--K-Q---E---I-S---EPD---I---DVW-DILRQAPPS EYEKIAFQYGITDLRGMLKRLKMKKKEQKK
F1NBZ9_CHICK/247-343 TE-GLEES--GELNFSALLKK---R--NGFifcR--GD--G-K---S---D-S---QPD---V---DVW-EILRKAPPS EYEKIAFQYGITDLRGMLKRLKRIKKEEKK
G3P589_GASAC/262-375 ND-GKEDS--GELDFSTLLKK---R--TKV---L--IQvkyS-H--Mr--nK-A---EPD---M---DVW-NILSHAPAS EYEKIAFQHGITDLRGMLKRLKMKKKEEKK
W5KF69_ASTMX/288-396 TD-GGDDS--GELDF SALLKK---S--SGH---R--M--V-Q---V---S-S---EPD---V---DVW-EILSHAPAS EYEKIAFQYGITDLRGMLKRLKMKKKEEKK
H3CVJ4_TETNG/260-370 TD-GKEDS--GELDFSTLLKK---RmsSSF---L--KS---S;R---V---R-Avhvs sEPE---V---DVW-EILSKAPSS EYEKIAFQHGITDLRGMLKRLKMKKKEEKK
H2LNL0_ORYLA/260-380 TD-GNDDS--GELDF SALLKK;IkyH--ETF--N--RN--R-A---V---Q-V---HS E---Pd---vDVW-SILSKAPPS AF EKIAFQYGITDLRGMLKRLKMKKDEKK
G1PGG8_MYOLU/108-217 GE-GQEDA--GELDFSGLLKP---K--QRF---SpvRE--V-K---Q---Q-Qd---eEPE---I---DVW-ELLKNANPN EYEKIAFQYGITDLRGMLKRLKRMRRVEKK
M4A199_XIPMA/183-279 -D-GKDDS--GELDF SALLKK---rR--ESV---L--T-C--S-Q---K---K-T---TAE---K---KKLI LKMGHAPPS EF EKIAFQYGITDLRGMLKRLKMKKKEEKK
H2TW88_TAKRU/94-206 SE-GQEDA--GELDFSGLLKH---RIsDSV--K--TP--N-Kgmgq f---D-D---TPE---V---DVW-EILKKARPEY EYKIAFTYGITDLRGLLRMMKIPKEEKK
F7DKM1_XENTR/161-262 ---DEDA--GELDFSGLLKK---R--VED---K-qKE--Q-K---KkkdddC---GIP---P---DVW-ELLKNAKPS DFERIAFEHGITDLRGMLKRLKVKKEVKK
H2VDR5_TAKRU/132-236 SD-AGED E--GDLD FSALLKA---T--KNH; sItG--RN--K-K---P---QkE---EPE---I---DVW-ELLKSAPHS EYEKIAFQYGITDLRGMLKRLKMKVVE--
E2R072_CANFA/152-254 GEgKSEdG--GELDF SLLKK---R--EVV---E--EE--K-K---K---K---DDD---Dl g i pEIW-ELLKGAKKS EYEKIAFQYGITDLRGMLKRLKKAKEVVK
H0VY97_CAVPO/151-251 EG-KSEDA--GELDFSGLLKK---S--CLF--A--SP--Y-D---D---D-L---GIP---P---EIW-ELLKGAKKS EYEKIAFQYGITDLRGMLKRLKKAKEVVK
F7EY62_CAJA/121-207 GD-PRRRL--GQLTR E VVEE---K--KKK--K--DD--D-D---L---G-I---PPE---I---IW-ELLKGAKKS EYEKIAFQYGITDLRGMLKRLKKAKEVVK
I3KBQ9_ORENI/154-222 ---S A--ETIDFRKHLKK---R--NPD--gT--RE--H-K---T-M---ET E---E---KVW-EILLSADKKDY ERICAEYGITDFRGMLKRLNEMKKER--
H2VDR3_TAKRU/130-195 -----KADILSAFKR---A--DAG--E--DE--G-D---L---D-F---SAL--L---KAT-KLKSAPHS EYEKIAFQYGITDLRGMLKRLKMKVVE--
E4WX22_OIKDI/125-189 -----EFRAALRK---V--QKF--S--VG--K-K---V-A---DAK--E---DIW-ALLKSANPREYDRIFA FYWIKDLRKLKLLANAKNNKK
F7FUF4_MACMU/92-170 ---PQEDLrkELMDFRKLLKK---R--TTW--G--TR--A-R---A---p-A---PEKkidIE---QVW-QLLMTADRKDY EQICMKYGIVDYRGLRKLQEMKKEQ--
M7AMC7_CHEMY/10184-10245 -----DLRAKLKS---T--PTK--K--KE--E-E---E---E-K---PI---DIM-ELLKNVDPKEY EYARMYGITDFRGLLQAFELLLKQSQ--
M4AU85_XIPMA/120-195 -G-SN-EQ--ATQDFRSMLKK---T--TVA---T--RK--K-Q---L--P-Q---QKEg e i dP---KLL-ELLLSAPKKDY ERICLEFGITDFRWFLLKLLKQIKKER--
H2RX04_TAKRU/156-219 -----EDFKKALKN---K--IDI--D--AK--E-E---N---K-M---EAD--E---KFF-EVLSAEKKDY ESICIQYGVTFDRGMLKLLNEKKIE--
E7F8W2_DANRE/153-222 -----DPEDFRKMLKK---T--KIVk-krK--E--P-K---K---E-G---EID--P---KFW-DVMLNAPKS DY ERICHEFGITDYRWMLKQLNLKKE--EK-
G3U743_LOXAF/163-232 -----DKMDFKKMLKK---S--GPP--P--PE--K-K---Q---kK-V---IDE--K---EML-EILSKVPKKDF ERVCMVYGFDFWGLLKKLEMKKVEEK
G1MT65_MELGA/145-206 -----QDLKKT LKK---R--APL---P--KQ--K-E---V---D---K---E---AIF-QLLLHADKKDY ERICIKYGISDFRGLRALQDLRDKDT--
I3KJ09_ORENI/137-197 -----PADFRKLLKK---S--KVE---R--AD--G-----ESD---E---RFW-DVMLKADRNDY ERICSEFGVKDLHSLKLLDEKKKE--

```

Conservation



Quality



Consensus



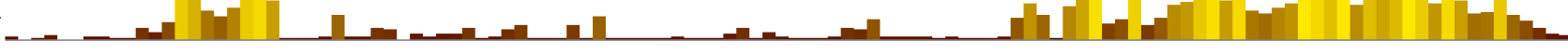
TD-G+EDA--GELDF SALLKK---R-SKS F---K--RE--K-K---+---D-A---EPE---E---DVW-ELLK+APPS EY EKIAFQYGITDLRGMLKRLKMKKKEEKK

	70	80	90	100	110	120	130	140	150	160	170
Query/1-103	SD-SHEDA	-GTLDFSSLLKK	---R-DSF	---R-DRD-S-K	---L-E-A	---PAE	---E---	DVW-EILRQAPPSEYERIAFHQGVTDLRGMLKRLKGMKQDEKK			
H3B137_LATCH/266-360	AD-GSEEA	-GELDFSAALLKK	---R-NNE	---K-EV--K-Q	---E-I-S	---EPD	---I---	DVW-DLLRQAPPSEYERIAFQYGITDLRGMLKRLKMKKEQKK			
F1N8Z9_CHICK/247-343	TE-GLAES	-GELNFSALLKK	---R-NGF	ilfcR-GD--G-K	---S-D-S	---QPD	---V---	DVW-EILRKAPPSEYERIAFQYGITDLRGMLKRLKRLKKEEKK			
G3P589_GASAC/262-375	ND-GKEDS	-GELDFSTLLKK	---R-TKV	---L-IQvkyS-H	---Mr--nK-A	---EPD	---M---	DVW-NILSHAPSEYERIAFHQGITDLRGMLKRLKMKKEEKK			
W5KF69_ASTMX/288-396	TD-GGDDS	-GELDFSAALLKK	---S-SGH	---R-M--V-Q	---V-S-S	---EPD	---V---	DVW-EILSHAPASEYERIAFQYGITDLRGMLKRLKMKKEEKK			
H3CVJ4_TETNG/260-370	TD-GKEDS	-GELDFSTLLKK	---RmsSSF	---L-KS--SsR	---V--R-Avhs	---EPE	---V---	DVW-EILSKAPSEYERIAFHQGITDLRGMLKRLKMKKEEKK			
H2LNLD_ORYLA/260-380	TD-GNDDS	-GELDFSAALLKK	siKyH-ETF	---N-RN--R-A	---V--Q-V	---HE	---Pd--v	DVW-SILSKAPPSEYERIAFQYGITDLRGMLKRLKMKKEEKK			
G1PGG8_MYOLU/108-217	GE-GQEDA	-GELDFSGLLKP	---K-QRF	---SvRE--V-K	---Q--Q-Qd--e	---EPE	---I---	DVW-EILKKNANSEYERIAFQYGITDLRGMLKRLKRRRVEEKK			
M4A199_XIPMA/183-279	-D-GKDDS	-GELDFSAALLKK	---rR--ESV	---L-T-C	---S-Q--K	---KT	---TAE	---K---	KKLI LKMGHAPPSEYERIAFQYGITDLRGMLKRLKMKKEEKK		
H2TW88_TAKRU/94-206	SE-GQEDA	-GELDFSGLLKH	---RIsDSV	---K-Tp--N-Kgmgqf	---D-D-TPE	---V---	DVW-EILKKARPSEYERIAFTYGITDLRGLRRMKKIPKEEKK				
F7DKM1_XENTR/161-262	---DEDA	-GELDFSGLLKK	---R-VED	---K-qKE--Q-K	---KkkdddC	---GIP	---P---	DVW-EILKNAKSPSEYERIAFHQGITDLRGMLKRLKMKKEEKK			
H2VDR5_TAKRU/132-236	SD-AGED	-GDLDFSAALLKA	---T-KNH	IsItG--RN--K-K	---P--QkE	---EPE	---I---	DVW-EILKSAHPSSEYERIAFQYGITDLRGMLKRLKMKVVE--			
E2R072_CANFA/152-254	GEGKSEDG	-GELDFSAALLKK	---R-EVW	---E-EE--K-K	---K--K-K	---DD	---Dliip	EIW-EILKGAKKSEYERIAFQYGITDLRGMLKRLKKAIVEEKK			
H0VVP7_CAVPO/151-251	EG-KSEDA	-GELDFSGLLKK	---S-CLF	---A-SP--Y-D	---D--D-L	---GIP	---P---	EIW-EILKGAKKSEYERIAFQYGITDLRGMLKRLKKAIVEEKK			
F7EU62_CALJA/121-207	GD-PRRRL	-GQLTREVVVEE	---K-KKK	---K-DD--D-D	---L--G-I	---PPE	---P---	IW-EILKGAKKSEYERIAFQYGITDLRGMLKRLKKAIVEEKK			
I3KBQ9_ORENI/154-222	---S-A--ETID	FRKHLKK	---R-NPD	---GT--RE--H-K	---T-M--ET	---ETE	---E---	KVW-EILLSADKKDYERICAEYGITDLRGMLKRLKMKKEEKK			
H2VDR3_TAKRU/130-195	---KAD	ILSAFKR	---A-DAG	---E-DE--G-D	---L--D-F	---SAL	---L---	KAT-KLKSAPSEYERIAFQYGITDLRGMLKRLKMKVVE--			
E4WZ22_OIKDI/125-189	---E	FRAALRK	---V-QKF	---S-VG--K-K	---V-A--DAK	---E---	DIW-ALLKSANPREYDRIAFYWGIKDLRKLKLANAKRNNKK				
F7FUFA_MACMU/92-170	---PQED	LrkELMDFRKLLK	---R-TTW	---G-TR--A-R	---A--pP-A	---PEK	---idIE	QVW-QLLMTADRKDYEQICMRYGIVDYRGMLRKLQEMKKEQ			
M7AMC7_CHEMY/10184-10245	---D	LRAKLKS	---T-BTK	---K-KE--E-E	---E--E-K	---PI	---	DIM-EILLKNVDKEYEYARMYGITDLRGMLKRLKQAFELKQSQ--			
M4AU85_XIPMA/120-195	-G-SN-EQ	-ATQDFRSMLK	---T-TVA	---T-RK--K-Q	---L--P-Q	---QKE	---eidP	KLL-EILLSAPKKDYERICLEFGITDFRWFLLKLLKQLKKEEKK			
H2RX04_TAKRU/156-219	---ED	FKKALN	---K-IDI	---D-AK--E-E	---N--K-M	---EAD	---E---	KFF-EVLSMAEKKDYESICIQYGVTDLRGMLKRLKNEKKEIE--			
E7F8W2_DANRE/153-222	---D	PEDFRKLK	---T-KI	vkkrK-EE--P-K	---K--K-E	---EG	---EID	---P---	KFW-DVMLNAKGDYERICHEFGITDYRWMLKQLNLKKEEKK		
G3V743_LOXAF/163-232	---DKMD	FKMKLKS	---S-GPP	---P-PK--K-K	---Q--kK-V	---IDE	---K---	EML-EILSKVPKDFERVCMVYGTDFWGLLKKLKEMKKVEEKK			
G1MT65_MELGA/145-206	---QD	LKTKLKK	---R-ABL	---P-KQ--K-E	---V--D--K	---E---	AIF-QLLHADKKDYERICIKYGISDFRGMLRALDRLKDT--				
I3KU9_ORENI/137-197	---PAD	FRKLLKK	---S-KVE	---R-AD--G	---	---ED	---E---	RFW-DVMLKADRNDYERICSEFGVKDLHSLKLLDEKKEE--			

Conservation



Quality



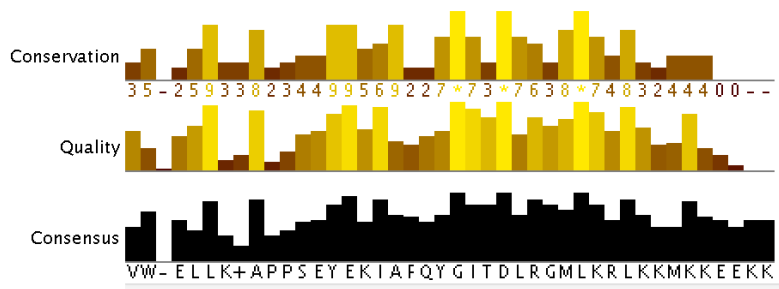
Consensus



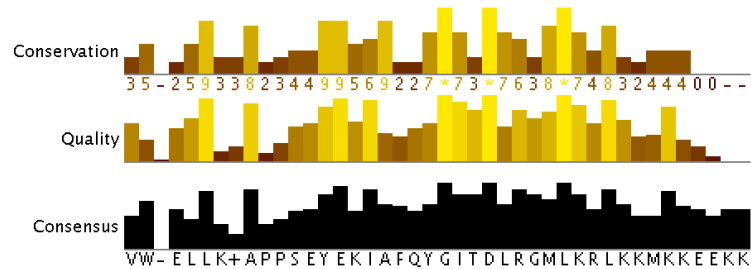
TD-G+EDA--GELDFSAALLKK---R-SKSF---K--RE--K-K---+---D-A---EPE---E---DVW-E LLK+APPS EY EK IAFQY GITDLRGMLKRLKMKKEEKK

1. Edit -> Remove left

		10	20	30	40
Query/63-103	VW-	EILRQAPPS	EYERIAFQHG	GVTDLRGML	KRLKKGMMKQDEKK
H3B137_LATCH/320-360	VW-	DLLRQAPPS	EYEKIAFQY	GITDLRGML	KRLKMKKKEQKK
F1NBZ9_CHICK/303-343	VW-	EILRKAPPS	EYEKIAFQY	GITDLRGML	KRLKRIKKEEKK
G3PS89_GASAC/335-375	VW-	NILSHAPPS	EYEKIAFQHG	GITDLRGML	KRLKMKKKEEKK
W5KF69_ASTMX/356-396	VW-	EILSHAPAS	EYEKIAFQY	GITDLRGML	KRLKMKKKEEKK
H3CVJ4_TETNG/330-370	VW-	EILSKAPSS	EYEKIAFQHG	GITDLRGML	KRLKMKKKEEKK
H2LNL0_ORYLA/340-380	VW-	SILSKAPPS	AFEKIAFQY	GITDLRGML	KRLKMKKKDEKK
G1PGG8_MYOLU/177-217	VW-	ELLKNANPN	EYEKIAFQY	GITDLRGML	KRLKRMRRVEKK
M4A199_XIPMA/238-279	KLI	LKMGHAPPS	EYEFKIAFQY	GITDLRGML	KRLKMKKKEEKK
H2TW88_TAKRU/166-206	VW-	EILKKARPDE	EYEKIAFTY	GITDLRGLLR	RMKKIKPEEKK
F7DKM1_XENTR/222-262	VW-	ELLKNAKPS	DFERIAFEH	GITDLRGML	KRLKMKKKEVKK
H2UDR5_TAKRU/198-236	VW-	ELLKSAHPS	EYEKIAFQY	GITDLRGML	KRLKMKKVVVE--
E2R072_CANFA/214-254	IW-	ELLKGAKKS	EYEKIAFQY	GITDLRGML	KRLKKAQVVEVKK
H0VYP7_CAVPO/211-251	IW-	ELLKGAKKS	EYEKIAFQY	GITDLRGML	KRLKKAQVVEVKK
F7EU62_CALJA/167-207	IW-	ELLKGAKKS	EYEKIAFQY	GITDLRGML	KRLKKAQVVEVKK
I3KBQ9_ORENI/184-222	VW-	EILLSADKKD	YERICAEY	GITDFRGML	KRLKLNEMKKER--
H2UDR3_TAKRU/157-195	AT-	KLLKSAHPS	EYEKIAFQY	GITDLRGML	KRLKMKKVVVE--
E4WXZ2_OIKDI/149-189	IW-	ALLKSANPRE	YDRIAFYWG	IKDLRKL	KLKLANAKKNNKK
F7FUF4_MACMU/132-170	VW-	QLLMTADR	KDYEQICMKY	GIVDYRGML	KRLKQEMKKEQ--
M7AMC7_CHEMY/10207-10245	IM-	ELLKNVDP	KEYKYARMY	GITDFRGLL	QAFELLLKQSQ--
M4A4U5_XIPMA/157-195	LL-	ELLLSAP	KKDYERICLE	FGITDFRFW	LKRLKQIKKER--
H2RX04_TAKRU/182-219	FF-	EVLMSAE	KKDYESICIQY	GVTFDFRGML	KRLKLNKKIE--
E7F8W2_DANRE/184-222	FW-	DVMLNAKK	SDYERICHE	FGITDYRWML	KQLNLKKEE-EK-
G3U743_LOXAF/192-232	ML-	EILSKVP	KKDFERVCM	YGFDFWGLL	KRLKEMKKEVEEK
G1MT65_MELGA/168-206	IF-	QLLLHAD	KKDYERICIKY	GISDFRGML	RALQDLRKDT--
I3KJU9_ORENI/160-197	FW-	DVMLKADR	NDYERICSE	FQWKLHSL	KLKLDKKEE---



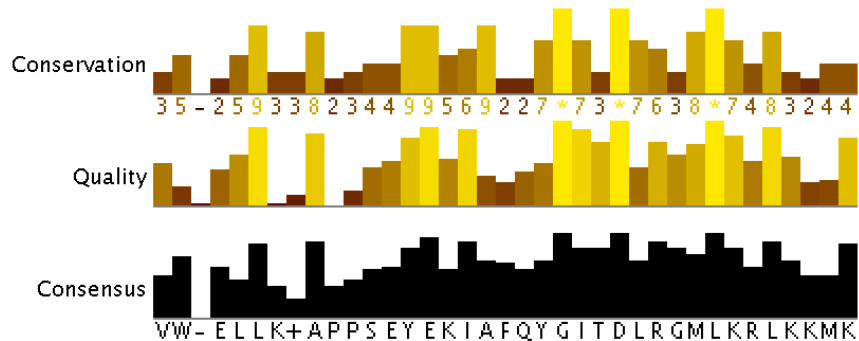
File	Edit	Select	View	Annotations	Format	Colour	Calculate	Web Service																																		
Query/63-103	VW-	E	I	L	R	Q	A	P	P	S	E	Y	E	K	I	A	F	Q	H	G	V	T	D	L	R	G	M	L	K	R	L	K	G	M	K	Q	D	E	K	K		
H3R137_LATCH/320-360	VW-	D	L	L	R	Q	A	P	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	E	Q	K	K		
F1N8Z9_CHICK/303-343	VW-	E	I	L	R	K	A	P	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	R	I	K	K	E	E	K	K		
G3P589_GASAC/335-375	VW-	N	I	L	S	H	A	P	S	E	Y	E	K	I	A	F	Q	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	E	E	K	K			
W5KF69_ASTMX/356-396	VW-	E	I	L	S	H	A	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	E	E	K	K			
H3CVJ4_TETNG/330-370	VW-	E	I	L	S	K	A	P	S	E	Y	E	K	I	A	F	Q	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	E	E	K	K			
H2LN10_ORVLA/340-380	VW-	S	I	L	S	K	A	P	S	A	F	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	D	E	K	K			
G1PGG8_MYOLU/177-217	VW-	E	L	L	K	N	A	N	P	N	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	R	M	R	R	V	E	K	K		
M4A199_XIPMA/238-279	KL	I	L	K	M	G	H	A	P	S	E	F	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	K	E	E	K	K		
H2TW88_TAKRU/166-206	VW-	E	I	L	K	K	A	R	P	D	E	Y	E	K	I	A	F	T	Y	G	I	T	D	L	R	G	L	L	R	R	M	K	I	P	K	E	E	K	K			
F7DKM1_XENTR/222-262	VW-	E	L	L	K	N	A	K	P	S	D	F	E	R	A	F	E	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	V	K	K	E	V	K	K			
H2VDR5_TAKRU/198-236	VW-	E	L	L	K	S	A	H	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	V	V	E	-	-		
E2R072_CANFA/214-254	IW-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	V	E	V	K	K			
H0VYP7_CAVPO/211-251	IW-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	V	E	V	K	K			
F7EU62_CAJJA/167-207	IW-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	V	E	V	K	K			
I3KBQ9_ORENI/184-222	VW-	E	I	L	S	A	D	K	K	D	Y	E	R	I	C	A	E	Y	G	I	T	D	F	R	G	M	L	K	L	N	E	M	K	K	E	R	-	-				
H2VDR3_TAKRU/157-195	A	T	-	K	L	K	S	A	H	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K	V	V	E	-	-	
E4WXZ2_OIKDI/149-189	IW-	A	L	L	K	S	A	N	P	R	E	Y	D	R	I	A	F	Y	W	G	I	K	D	L	R	K	L	L	K	L	A	N	A	K	K	N	N	K	K			
F7FUF4_MACMU/132-170	VW-	Q	L	L	M	T	A	D	R	K	D	Y	E	Q	I	C	M	K	Y	G	I	V	D	R	G	M	L	K	L	Q	E	M	K	K	E	Q	-	-				
M7FAM7_CHEMY/10207-10245	I	M	-	E	L	L	K	N	V	D	P	K	E	Y	E	K	Y	A	R	M	Y	G	I	T	D	F	R	G	L	L	Q	A	F	E	L	L	K	Q	S	Q	-	-
M4AU85_XIPMA/157-195	LL	-	E	L	L	S	A	P	K	K	D	Y	E	R	I	C	L	E	F	G	I	T	D	F	R	W	F	L	K	R	L	Q	I	K	K	E	R	-	-			
H2RX04_TAKRU/182-219	F	F	-	E	V	L	M	S	A	E	K	K	D	Y	E	S	I	C	I	Q	Y	G	V	T	D	F	R	G	M	L	K	L	N	E	K	K	I	E	-	-		
E7F8W2_DANRE/184-223	F	W	-	D	V	M	L	N	A	K	S	D	Y	E	R	I	C	H	E	F	G	I	T	D	Y	R	W	M	L	Q	L	N	L	K	K	-	E	K				
G3U743_LOXAF/192-232	M	L	-	E	I	L	S	K	V	P	K	D	F	E	R	V	C	M	V	Y	G	F	T	D	F	W	G	L	L	K	L	K	E	M	K	K	V	E	E	K		
G1MT65_MELGA/168-206	I	F	-	Q	L	L	H	A	D	K	K	D	Y	E	R	I	C	I	K	Y	G	I	S	D	F	R	G	M	L	R	A	L	Q	D	L	R	K	D	T	-	-	
I3KJU9_ORENI/160-197	F	W	-	D	V	M	L	K	A	D	R	N	D	Y	E	R	I	C	S	E	F	G	V	K	D	L	H	S	I	L	K	L	D	E	K	K	K	E	-	-		



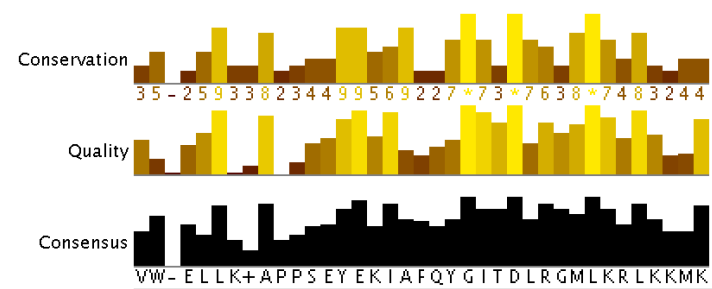
1. Edit -> Remove right



		10	20	30	
Query/63-98	VW-	EILRQA	PPSEY	ERIAFQ	HGVTDLRGMLKRLKGMK
H3B137_LATCH/320-355	VW-	DLLRQA	PPSEY	EKIAFQ	YGITDLRGMLKRLKKMK
F1NBZ9_CHICK/303-338	VW-	EILRKAP	PPSEY	EKIAFQ	YGITDLRGMLKRLKRIK
G3P589_GASAC/335-370	VW-	NILSHAP	PPSEY	EKIAFQ	HGVTDLRGMLKRLKKMK
W5KF69_ASTMX/356-391	VW-	EILSHAP	APSEY	EKIAFQ	YGITDLRGMLKRLKKMK
H3CVJ4_TETNG/330-365	VW-	EILSKAP	PPSEY	EKIAFQ	HGVTDLRGMLKRLKKMK
H2LNL0_ORYLA/340-375	VW-	SILSKAP	PPSAF	EKIAFQ	YGITDLRGMLKRLKKMK
G1PGG8_MYOLU/177-212	VW-	ELLKNA	NPN	EKIAFQ	YGITDLRGMLKRLKRMK
M4A199_XIPMA/238-274	KLIL	LKM	GHAPP	SEFEK	IAFQYGITDLRGMLKRLKKMK
H2TW88_TAKRU/166-201	VW-	EILKKAR	PDEY	EKIAFT	YGITDLRGLLRMMKKIP
F7DKM1_XENTR/222-257	VW-	ELLKNA	KPSDF	ERIAF	EHGITDLRGMLKRLKKVK
H2UDR5_TAKRU/198-233	VW-	ELLKSA	HPSEY	EKIAFQ	YGITDLRGMLKRLKKMK
E2R072_CANFA/214-249	IW-	ELLKGAK	KSEY	EKIAFQ	YGITDLRGMLKRLKKAK
H0VYP7_CAVPO/211-246	IW-	ELLKGAK	KSEY	EKIAFQ	YGITDLRGMLKRLKKAK
F7EU62_CAJA/167-202	IW-	ELLKGAK	KSEY	EKIAFQ	YGITDLRGMLKRLKKAK
I3KBQ9_ORENI/184-219	VW-	EILLSA	DKKDY	ERICAE	YGITDFRGMLKRLNEMK
H2UDR3_TAKRU/157-192	AT-	KKLKS	AHPSEY	EKIAFQ	YGITDLRGMLKRLKKMK
E4WXZ2_OIKDI/149-184	IW-	ALLKSA	NPREY	DRIAFY	WGIKDLRKLKLANAK
F7FUF4_MACMU/132-167	VW-	QLLMTA	DRKDY	EQICMK	YGIVDYRGMLRKLQEMK
M7AMC7_CHEMY/10207-10242	IM-	ELLKNV	DKPEY	EKYAR	MYGITDFRGLLQAFELK
M4AU85_XIPMA/157-192	LL-	ELLSA	PKKDY	ERICLE	FGITDFRWFLLKLLKQIK
H2RX04_TAKRU/182-217	FF-	EVLMSA	EKKDY	ESICIQ	YGVTDFRGMLKRLNEKK
E7F8W2_DANRE/184-219	FW-	DVMLNA	KKSDY	ERICHE	FGITDYRWMLKQLNLKK
G3U743_LOXAF/192-227	ML-	EILSKV	PKKDF	ERVCMV	YGFTDFWGLLKLLKEMK
G1MT65_MELGA/168-203	IF-	QLLLH	ADKKDY	ERICIK	YGISDFRGMLRALQDLR
I3KJU9_ORENI/160-195	FW-	DVMLKA	DRNDY	ERICSE	FVGVKDLHSILKLLDEKK



	File	Edit	Select	View	Annotations	Format	Colour	Calculate																													
					10	20	30																														
Query/63-98	VW-	E	L	R	Q	A	P	P	S	E	Y	E	K	I	A	F	Q	H	G	V	T	D	L	R	G	M	L	K	R	L	K	G	M	K			
H3B137_LATCH/320-355	VW-	D	L	R	Q	A	P	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K			
F1NBZ9_CHICK/303-338	VW-	E	L	R	K	A	P	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	R	I	K			
G3P589_GASAC/335-370	VW-	N	I	L	S	H	A	P	S	S	E	Y	E	K	I	A	F	Q	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K		
W5KF69_ASTMX/356-391	VW-	E	L	S	H	A	P	A	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K			
H3CVJ4_TETNG/330-365	VW-	E	L	S	K	A	P	S	S	E	Y	E	K	I	A	F	Q	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K			
H2LNL0_ORYLA/340-375	VW-	S	I	L	S	K	A	P	P	S	A	F	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K		
G1PGG8_MYOLU/177-212	VW-	E	L	L	K	N	A	N	P	N	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	R	M	R		
M4A199_XIPMA/238-274	K	L	I	L	K	M	G	H	A	P	P	S	E	F	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	
H2TW88_TAKRU/166-201	VW-	E	L	L	K	K	A	R	P	D	E	Y	E	K	I	A	F	T	Y	G	I	T	D	L	R	G	L	L	R	R	M	K	I	P			
F7DKM1_XENTR/222-257	VW-	E	L	L	K	N	A	K	P	S	D	F	E	R	I	A	F	E	H	G	I	T	D	L	R	G	M	L	K	R	L	K	K	V	K		
H2UDR5_TAKRU/198-233	VW-	E	L	L	K	S	A	H	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K		
E2R072_CANFA/214-249	I	W	-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	
H0VYP7_CAVPO/211-246	I	W	-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	
F7EU62_CAJJA/167-202	I	W	-	E	L	L	K	G	A	K	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	A	K	
I3KBQ9_ORENI/184-219	VW-	E	L	L	S	A	D	K	K	D	Y	E	R	I	C	A	E	Y	G	I	T	D	F	R	G	M	L	K	K	L	N	E	M	K			
H2UDR3_TAKRU/157-192	A	T	-	K	K	L	K	S	A	H	P	S	E	Y	E	K	I	A	F	Q	Y	G	I	T	D	L	R	G	M	L	K	R	L	K	K	M	K
E4WXZ2_OIKDI/149-184	I	W	-	A	L	L	K	S	A	N	P	R	E	Y	D	R	I	A	F	Y	W	G	I	K	D	L	R	K	L	L	K	L	A	N	A	K	
F7FUF4_MACMU/132-167	VW-	Q	L	L	M	T	A	D	R	K	D	Y	E	Q	I	C	M	K	Y	G	I	V	D	Y	R	G	M	L	R	K	L	Q	E	M	K		
M7AMC7_CHEMY/10207-10242	I	M	-	E	L	L	K	N	V	D	P	K	E	Y	K	A	R	M	Y	G	I	T	D	F	R	G	L	L	Q	A	F	E	L	L	K		
M4AU85_XIPMA/157-192	L	L	-	E	L	L	S	A	P	K	K	D	Y	E	R	I	C	L	E	F	G	I	T	D	F	R	W	F	L	K	K	L	K	Q	I	K	
H2RX04_TAKRU/182-217	F	F	-	E	V	L	M	S	A	E	K	K	D	Y	E	S	I	C	I	Q	Y	G	V	T	D	F	R	G	M	L	K	K	L	N	E	K	
E7F8W2_DANRE/184-219	F	W	-	D	V	M	L	N	A	K	K	S	D	Y	E	R	I	C	H	E	F	G	I	T	D	Y	R	W	M	L	K	Q	L	N	L	K	
G3U743_LOXAF/192-227	M	L	-	E	I	L	S	K	V	P	K	K	D	F	E	R	V	C	M	V	Y	G	F	T	D	F	W	G	L	L	K	L	K	E	M	K	
G1MT65_MELGA/168-203	I	F	-	Q	L	L	H	A	D	K	K	D	Y	E	R	I	C	I	K	Y	G	I	S	D	F	R	G	M	L	R	A	L	Q	D	L		
I3KJU9_ORENI/160-195	F	W	-	D	V	M	L	K	A	D	R	N	D	Y	E	R	I	C	S	E	F	G	V	K	D	L	H	S	I	L	K	K	L	D	E	K	



1. File -> Save as



# protein alignment/profile-HMM vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Upload a file  jalview.fasta

## ▼ Sequence Database ?

Frequently used databases

Reference Proteomes  UniProtKB  SwissProt  PDB

Representative Sets (UniProt)

rp75  rp55  rp35  rp15

Other databases

QfO  Pfamseq

► **Restrict by Taxonomy** ?

## ▼ Cut-Offs ?

E-value  Bit score

Significance E-values: Sequence  Hit



## HMMSEARCH Results

Score

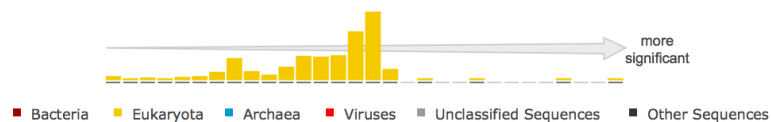
Taxonomy

Domain

Download

[Search Again](#)

## Distribution of Significant Hits


 < First < Previous **Page 1** of 4 Next > Last >
Significant Query Matches (362) in *uniprotrefprot* (v.2015-06-24)
[Customize](#)

	Target	Species	E-value
>	<a href="#">W4Z1N0_STRPU</a>	<a href="#">Strongylocentrotus purpuratus</a>	1.8e-34
>	<a href="#">W4Y2S3_STRPU</a>	<a href="#">Strongylocentrotus purpuratus</a>	3.1e-31
>	<a href="#">M7AMC7_CHEMY</a>	<a href="#">Chelonia mydas</a>	1.1e-25
>	<a href="#">W5LEB4_ASTMX</a>	<a href="#">Astyanax mexicanus</a>	3.5e-22
>	<a href="#">F7CWG3_CALJA</a>	<a href="#">Callithrix jacchus</a>	1.1e-20
>	<a href="#">H2UDR6_TAKRU</a>	<a href="#">Takifugu rubripes</a>	1.3e-20
>	<a href="#">H2UDR5_TAKRU</a>	<a href="#">Takifugu rubripes</a>	1.3e-20
>	<a href="#">F6W1U5_MACMU</a>	<a href="#">Macaca mulatta</a>	1.5e-20
>	<a href="#">F1QVS8_DANRE</a>	<a href="#">Danio rerio</a>	1.5e-20
>	<a href="#">H2UDR4_TAKRU</a>	<a href="#">Takifugu rubripes</a>	2.6e-20
>	<a href="#">H2UDR8_TAKRU</a>	<a href="#">Takifugu rubripes</a>	2.7e-20
>	<a href="#">H2UDR1_TAKRU</a>	<a href="#">Takifugu rubripes</a>	2.7e-20
>	<a href="#">H2UDR2_TAKRU</a>	<a href="#">Takifugu rubripes</a>	2.7e-20
>	<a href="#">H2UDR7_TAKRU</a>	<a href="#">Takifugu rubripes</a>	2.7e-20
>	<a href="#">M3ZYG5_XIPMA</a>	<a href="#">Xiphophorus maculatus</a>	2.7e-20
>	<a href="#">F1Q615_DANRE</a>	<a href="#">Danio rerio</a>	3.5e-20

# Annotation?

# UniProtKB - Q3UIK0 (Q3UIK0\_MOUSE)

**Protein** | Submitted name: **Myosin-binding protein C, cardiac-type**

**Gene** | **Mybpc3**

**Organism** | *Mus musculus (Mouse)*

**Status** | Unreviewed - Annotation score: - Experimental evidence at protein level<sup>i</sup>

Display None

BLAST
 Align
 Format
 Add to basket
 History

Help video
 Other tutorials and videos

Feedback

## Function<sup>i</sup>

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

- [identical protein binding](#) Source: MGI
- [myosin binding](#) Source: MGI
- [myosin heavy chain binding](#) Source: MGI ▾
- [structural constituent of cytoskeleton](#) Source: MGI ▾

 Function

 Names & Taxonomy

 Subcell. location

 Pathol./Biotech

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Proc

Expression

Interaction

Structure

Family & Domains

Sequence

Cross-references

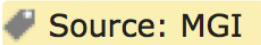
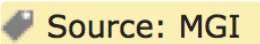
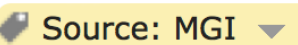
Publications

Entry information

Miscellaneous

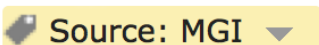
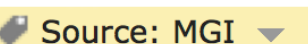
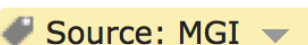
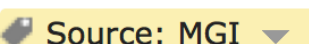
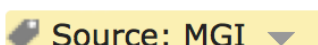
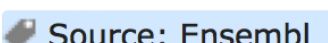
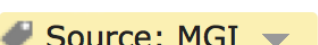
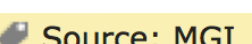
# Function<sup>i</sup>

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

- identical protein binding 
- myosin binding 
- myosin heavy chain binding 

Inferred from physical interaction<sup>i</sup>

PubMed 17075052

- heart morphogenesis 
- muscle contraction 
- myosin filament assembly 
- regulation of heart contraction 
- regulation of heart rate 
- regulation of striated muscle contraction 
- sarcomere organization 
- ventricular cardiac muscle tissue morphogenesis 

Complete GO annotation...

# Literature citation

Map to



Mapped (16)

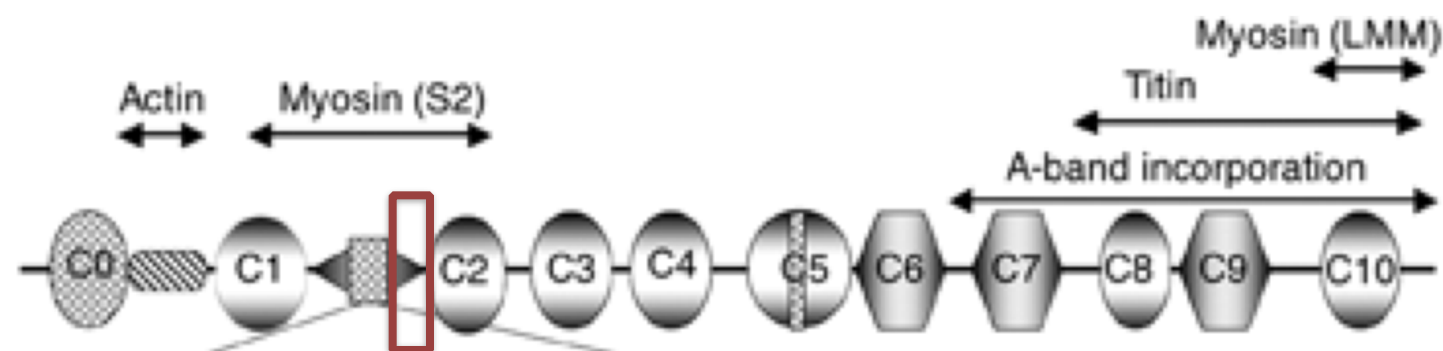
## Cardiac myosin binding protein C phosphorylation is cardioprotective.

Sadayappan S., Osinska H., Klevitsky R., Lorenz J.N., Sargent M., Molkentin J.D., Seidman C.E., Seidman J.G., Robbins J.






Cardiac myosin binding protein C (cMyBP-C) has three phosphorylatable serines at its N terminus (Ser-273, Ser-282, and Ser-302), and the residues' phosphorylation states may alter thick filament structure and function. To examine the effects of cMyBP-C phosphorylation, we generated transgenic mice with cardiac-specific expression of a cMyBP-C in which the three phosphorylation sites were mutated to aspartic acid, mimicking constitutive phosphorylation (cMyBP-C(AIIP+)). The allele was bred into a cMyBP-C null background (cMyBP-C((t/t))) to ensure the absence of endogenous dephosphorylated cMyBP-C. cMyBP-C(AIIP+) was incorporated normally into the cardiac sarcomere and restored normal cardiac function in the null background. However, subtle changes in sarcomere ultrastructure, characterized by increased distances between the thick filaments, indicated that phosphomimetic cMyBP-C affects thick-thin filament relationships, and yeast two-hybrid data and pull-down studies both showed that charged residues in these positions effectively prevented interaction with the myosin heavy chain. Confirming the physiological relevance of these data, the cMyBP-C(AIIP+:(t/t)) hearts were resistant to ischemia-reperfusion injury. These data demonstrate that cMyBP-C phosphorylation functions in maintaining thick filament spacing and structure and can help protect the myocardium from ischemic injury.



A



271-FRRTSLAGGG RRISDSHEDT GILDFSSLLK KRDSFRTPRK-Human  
 269-\*\*\*\*\*A\* \*\*T\*\*\*\*\*A \*TP\*\*\*\*\* \*\*\*\*\*RDS\*-Mouse  
 269-\*\*\*\*\*D\*\*\*A\* \*\*TD\*\*\*\*\*A \*TP\*\*\*\*\* \*\*\*D\*\*RDS\*-cMyBP-C<sup>Al1P+</sup>

 Cardiac-specific  IgG-like domains  Fibronectin type-III domains  cMyBP-C motif  Pro-Ala domain



# protein alignment/profile-HMM vs protein sequence database

[Paste a Sequence](#) | [Upload a File](#) | [Accession Search](#)

Upload a file  jalview.fasta

**Submit**

## ▼ Sequence Database ?

Frequently used databases

Reference Proteomes
  UniProtKB
  SwissProt
  PDB

Representative Sets (UniProt)

rp75
  rp55
  rp35
  rp15

Other databases

QfO
  Pfamseq

▶ **Restrict by Taxonomy** ?

## ▼ Cut-Offs ?

**E-value**
 Bit score



## HMMSEARCH Results

[Score](#) [Taxonomy](#) [Domain](#) [Download](#)
[Search Again](#)

## Distribution of Significant Hits



■ Bacteria
 ■ Eukaryota
 ■ Archaea
 ■ Viruses
 ■ Unclassified Sequences
 ■ Other Sequences

Significant Query Matches (1) in *pdb* (v.2015-06-24)
[Customize](#)

	Target	Species	E-value
>	<a href="#">2lhu_A</a>	<a href="#">Mus musculus</a>	5.8e-22

(show all) alignments

Your search took:0.01 secs  
showing rows 1 - 1 of 1

[Search Details](#)



■ Bacteria 
 ■ Eukaryota 
 ■ Archaea 
 ■ Viruses 
 ■ Unclassified Sequences 
 ■ Other Sequences

**Significant Query Matches (1) in *pdb* (v.2015-06-24)**

[Customize](#)

Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value	
start	end	start	end	start	end						Ind.	Cond.
23	31	56	69	58	66	0.06	0.85	33.3 (3)	77.8 (7)	0.5	7100	0.026

Query 23 tDlRgmLKr 31  
 D+ ++LK+  
 Target 58 LDFSSLLKK 66  
 PP 69\*\*\*\*\*8

Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value	
start	end	start	end	start	end						Ind.	Cond.
1	36	84	119	84	119	0.01	0.99	75.0 (27)	97.2 (35)	77.9	4.8e-21	1.8e-26

Query 1 vweLLkkAdkseYEkIafqYGI tDlRgmLKrLkkmK 36  
 vwe+L++A++seYE+Iafq+G+tDlRgmLKrLk mK  
 Target 84 VWEILRQAPPSEYERIAFOHGVTDLRGMLKRLKGMK 119  
 PP 8\*\*\*\*\*9

[\(show all\) alignments](#)

Your search took:0.01 secs  
showing rows 1 - 1 of 1

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

Go

[Advanced Search](#) | [Browse by Annotations](#)[Summary](#) [3D View](#) [Sequence](#) [Annotations](#) [Seq. Similarity](#) [3D Similarity](#) [Literature](#) [Biol. & Chem.](#) [Methods](#) [Links](#)

## Structural Insight into the Unique Cardiac Myosin Binding Protein-C Motif: A Partially Folded Domain

2LHU

Display Files ▾

Download Files ▾

Download Citation ▾

DOI:10.2210/pdb2lhu/pdb

### Primary Citation

#### Structural insight into unique cardiac myosin-binding protein-C motif: a partially folded domain.

Howarth, J.W. , Ramiseti, S. , Nolan, K. , Sadayappan, S. , Rosevear, P.R.

Journal: (2012) J.Biol.Chem. **287**: 8254-8262

PubMed: 22235120

PubMedCentral: PMC3318737

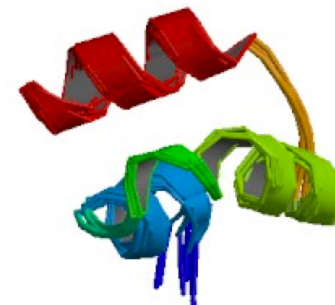
DOI: 10.1074/jbc.M111.309591

[Search Related Articles in PubMed](#)

#### PubMed Abstract:

The structural role of the unique myosin-binding motif (m-domain) of cardiac myosin-binding protein-C remains unclear. Functionally, the m-domain is thought to directly interact with myosin, whereas phosphorylation of the m-domain has been shown to

### Structure Image



**Chain A : Mybpc3 protein**

[FASTA](#) | [Sequence & DSSP](#) | [Image](#)

Polymer 1

Length: 124 residues

Chain Type: polypeptide(L)

Reference: [UniProtKB A9JR55](#)

**Annotations**

[Add Annotations](#)

Select

Secondary Structure: **DSSP** 18% helical (4 helices; 23 residues)

[\[hide\]](#) [\[reference\]](#)

DSSP

PDB **MGS S H H H H H S S G L V P R G S H M H E A I G S G D L D L R S A F R R T S L A G A G R R T S D S H E D A G T L D F**  
PDB

DSSP



PDB **S S L L K K R D S F R R D S K L E A P A E E D V W E I L R Q A P P S E Y E R I A F Q H G V T D L R G M L K R L K G M K Q**  
PDB

315 320 330 340

DSSP

PDB **DEKK**  
PDB

File Edit Select View Annotations Format Colour Calculate

		10	20	30	
Query/63-98	VW-	E I L R Q A P P S E Y E R I A F Q H G V T	D L R G M L K R L K G M K		
H3B137_LATCH/320-355	VW-	D L L R Q A P P S E Y E K I A F Q Y G I T	D L R G M L K R L K K M K		
F1NBZ9_CHICK/303-338	VW-	E I L R K A P P S E Y E K I A F Q Y G I T	D L R G M L K R L K R I K		
G3P589_GASAC/335-370	VW-	N I L S H A P S S E Y E K I A F Q H G I T	D L R G M L K R L K K M K		
W5KF69_ASTMX/356-391	VW-	E I L S H A P A S E Y E K I A F Q Y G I T	D L R G M L K R L K K M K		
H3CVJ4_TETNG/330-365	VW-	E I L S K A P S S E Y E K I A F Q H G I T	D L R G M L K R L K K M K		
H2LNL0_ORYLA/340-375	VW-	S I L S K A P P S A F E K I A F Q Y G I T	D L R G M L K R L K K M K		
G1PGG8_MYOLU/177-212	VW-	E L L K N A N P N E Y E K I A F Q Y G I T	D L R G M L K R L K R M R		
M4A199_XIPMA/238-274	KL	I L K M G H A P P S E F E K I A F Q Y G I T	D L R G M L K R L K K M K		
H2TW88_TAKRU/166-201	VW-	E I L K K A R P D E Y E K I A F T Y G I T	D L R G L L R R M K K I P		
F7DKM1_XENTR/222-257	VW-	E L L K N A K P S D F E R I A F E H G I T	D L R G M L K R L K K V K		
H2VDR5_TAKRU/198-233	VW-	E L L K S A H P S E Y E K I A F Q Y G I T	D L R G M L K R L K K M K		

**DSSP Legend**



T: turn



empty: no secondary structure assigned



G: 3/10-helix

S: bend