

# Examining repeats with databases

Miguel Andrade  
Faculty of Biology,  
Johannes Gutenberg University  
Mainz, Germany  
[andrade@uni-mainz.de](mailto:andrade@uni-mainz.de)

# RepeatsDB



Home Browse Statistics About Help Curation API

Search



# RepeatsDB

Version: 4

Released: 14.9.2024

Regions: 49714

RepeatsDB is a repository for the annotation and classification of structural tandem repeat proteins (STRPs). Each entry has start and end positions of the repeat region, repeat units, classification into four levels (Class, Topology, Fold, Clan), and in-depth characterization of the repeat regions.

Search by UniProtKB, PDB or Pfam identifier

Search

Examples:

[AOA0H3JRU9](#)

[SEAM\\_A](#)

[PF01261](#) Pfam

## How to cite

**“RepeatsDB in 2025: expanding annotations of Structured Tandem Repeats proteins on AlphaFoldDB”**

Damiano Clementel, Paula Nazarena Arrías, Soroush Mozaffari, et al.

[Nucleic Acids Research](#)

[PubMed](#)

## Software

RepeatDB predictions are generated with **STRPsearch: fast detection of structured tandem repeat proteins**

Soroush Mozaffari, Paula Nazarena Arrías, Damiano Clementel, et al.

[bioRxiv](#)

[GitHub](#)

# RepeatsDB

Structure identifier  chain  source

Region classification

Pfam accession

---

#	Preview	Structure	Chain	Source	Regions, units	External features	Type
0		<a href="#">3vbp</a>	A	<a href="#">RCSB/PDB</a>	<a href="#">3.1 7 units</a>	<a href="#">0016740</a> 	<a href="#">Reviewed </a>
		Last updated the 1.8.2017					
1		<a href="#">3vbp</a>	C	<a href="#">RCSB/PDB</a>	<a href="#">3.1 7 units</a>	<a href="#">0016740</a> 	<a href="#">Reviewed </a>
		Last updated the 1.8.2017					
2		<a href="#">3vbp</a>	E	<a href="#">RCSB/PDB</a>	<a href="#">3.1 7 units</a>	<a href="#">0016740</a> 	<a href="#">Reviewed </a>
		Last updated the 1.8.2017					

# RepeatsDB

Structure 3vbp, chain A from RCSB/PDB type Review

Repeated regions: 15 → 152 Å 3.1 - Beta-solenoid

RepeatsDB entries: No RepeatsDB entry

UniProt sequences: No UniProt sequence

Pfam domains: No Pfam domain

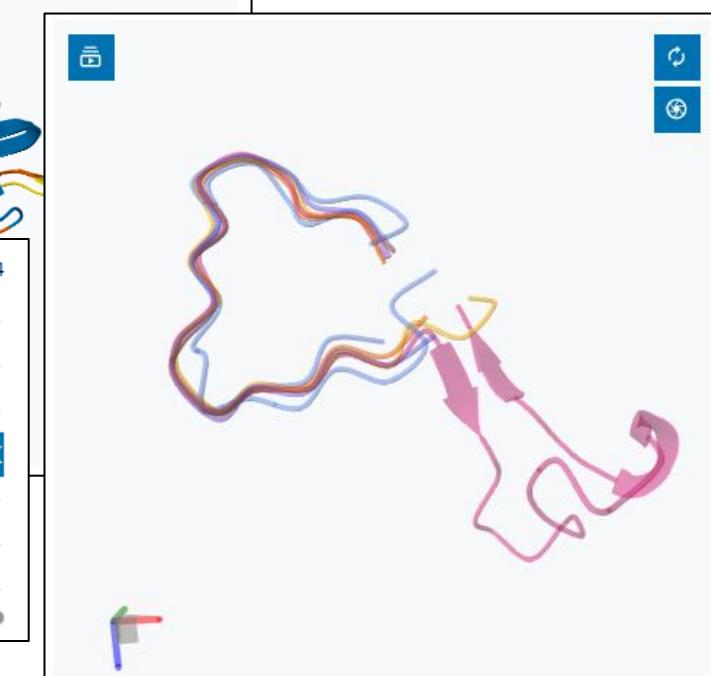
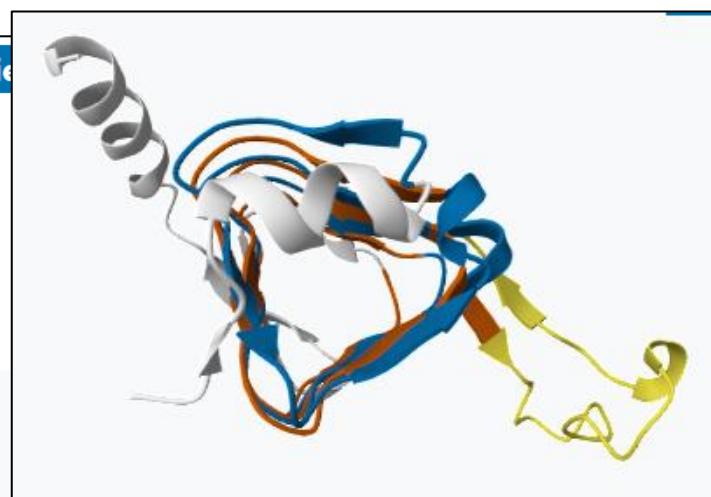
InterPro matches: No InterPro match

GO terms: GO:0016740 (IEA)

#	Start	End	Type	Parent region or Region's class
0	15	152	Region	3.1 - Beta-sc
1	15	35	Unit	#0 from 15 to 35
2	36	53	Unit	#0 from 15 to 53
3	54	74	Unit	#0 from 15 to 74
4	75	117	Unit	#0 from 15 to 117

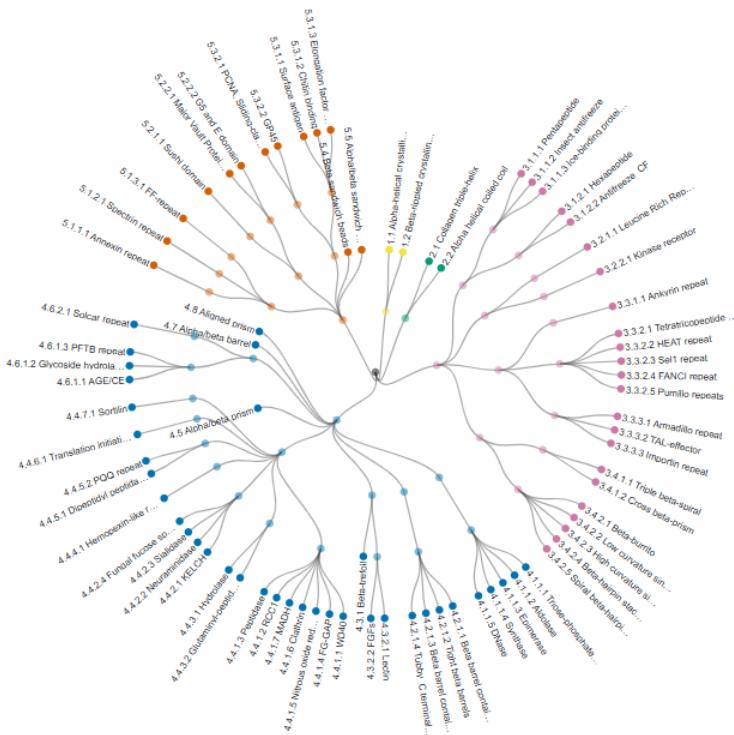
Index 5 10 15 20 25 30 35 40 44

	Consensus	-VSIG-	KNVLI	SAKTI	IYGP-	-	-	-	-
unit.1.fasta	F	LSVG	KNVLI	SKKAS	IYNP	G	-	-	-
unit.2.fasta	I	SIG	NNVRI	DDFCI	LSGK	-	-	-	V
unit.3.fasta	I	EMY	DFANI	SSRTI	VYAAI	NDFSG	NALMG	PTIPN	QYKNV
unit.4.fasta	V	I	KHVII	GAHSI	IFPN	-	-	-	KTGK
unit.5.fasta	V	TIG	SYSHI	AAYTA	LYGGE	-	-	-	VG
unit.6.fasta	V	VVIG	EGVAV	GAMSM	VKE	-	-	-	-



# Exercise 1. Evaluating annotations of repeat structures in RepeatsDB

- Go to RepeatsDB: <http://repeatsdb.bio.unipd.it/>
- Click on an outside node of the repeat map (only categories 3 “elongated repeats” or 4 “closed repeats”)



## 1 Crystalline aggregates

Crystalline aggregates formed by regions with 1 or 2 residue long repeats

## 2 Fibrous repeats

Fibrous structures stabilized by interchain interactions

## 3 Elongated repeats

Elongated structures whose repeat units require one another to maintain structure

## 4 Closed repeats

Closed structures whose repeat units need one another to maintain structure

## 5 Beads-on-a-string

Beads on a string structures whose repeat units are in tandem and large enough to fold independently

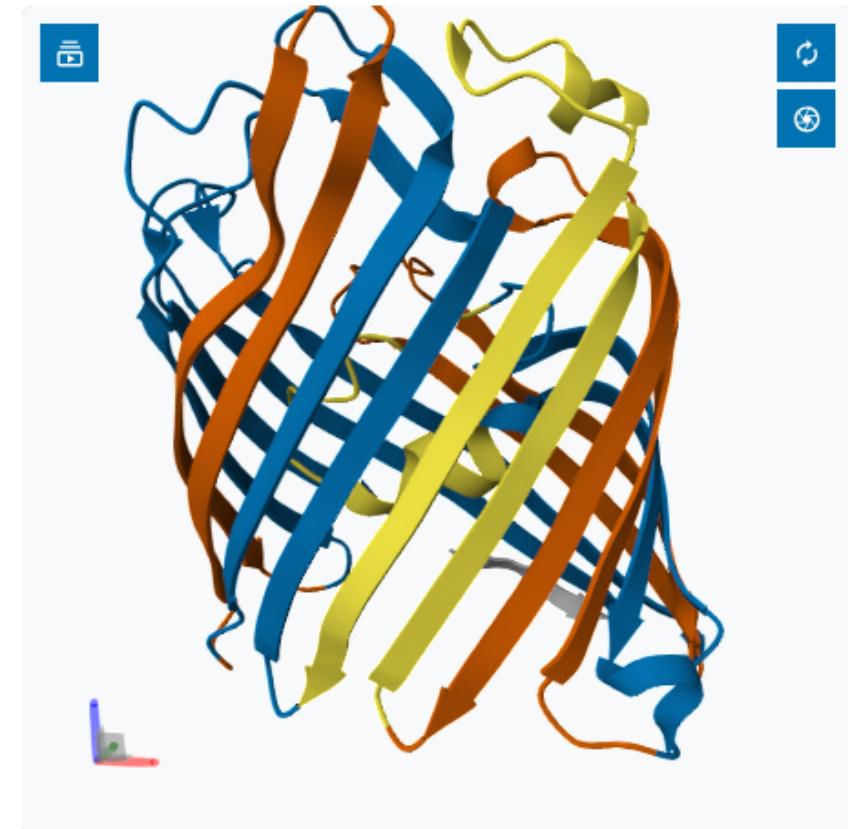
# Exercise 1. Evaluating annotations of repeat structures in RepeatsDB

- Go to RepeatsDB: <http://repeatsdb.bio.unipd.it/>
- Click on an outside node of the repeat map (only categories 3 “elongated repeats” or 4 “closed repeats”)
- Choose one example and write down the PDB name and the chain (e.g. PDB 6yc3, chain B).
- Check the assignment of the units. Is it correct? What about insertions? Are there mistakes? Write an evaluation: Looks good / one repeat wrong / many mistakes / total mess

# Example PDB 6yc3, chain B

#	Start	End	Type	Parent region or Region's class
0	9	339	Region	4.2.1.1 - Beta
1	9	53	Unit	#0 from 9 to
2	54	91	Unit	#0 from 9 to
3	92	147	Unit	#0 from 9 to
8	266	304	Unit	#0 from 9 to
9	305	339	Unit	#0 from 9 to

1. looks good
2. **one repeat wrong**
3. many mistakes
4. total mess



# Example PDB 1bt9, chain A

Structure **5cxb**, chain **B** from **RCSB/PDB** type **Reviewed** ★

Repeated regions: **452 → 801** & **4.4.1.1 - WD40**

RepeatsDB entries: **No RepeatsDB entry**

UniProt sequences: **GOSCK6** **433 → 801**

Pfam domains: **PF00400** **443 → 481** **759 → 801**

InterPro matches: **No InterPro match**

GO terms: **GO:0042254 (IEA)**

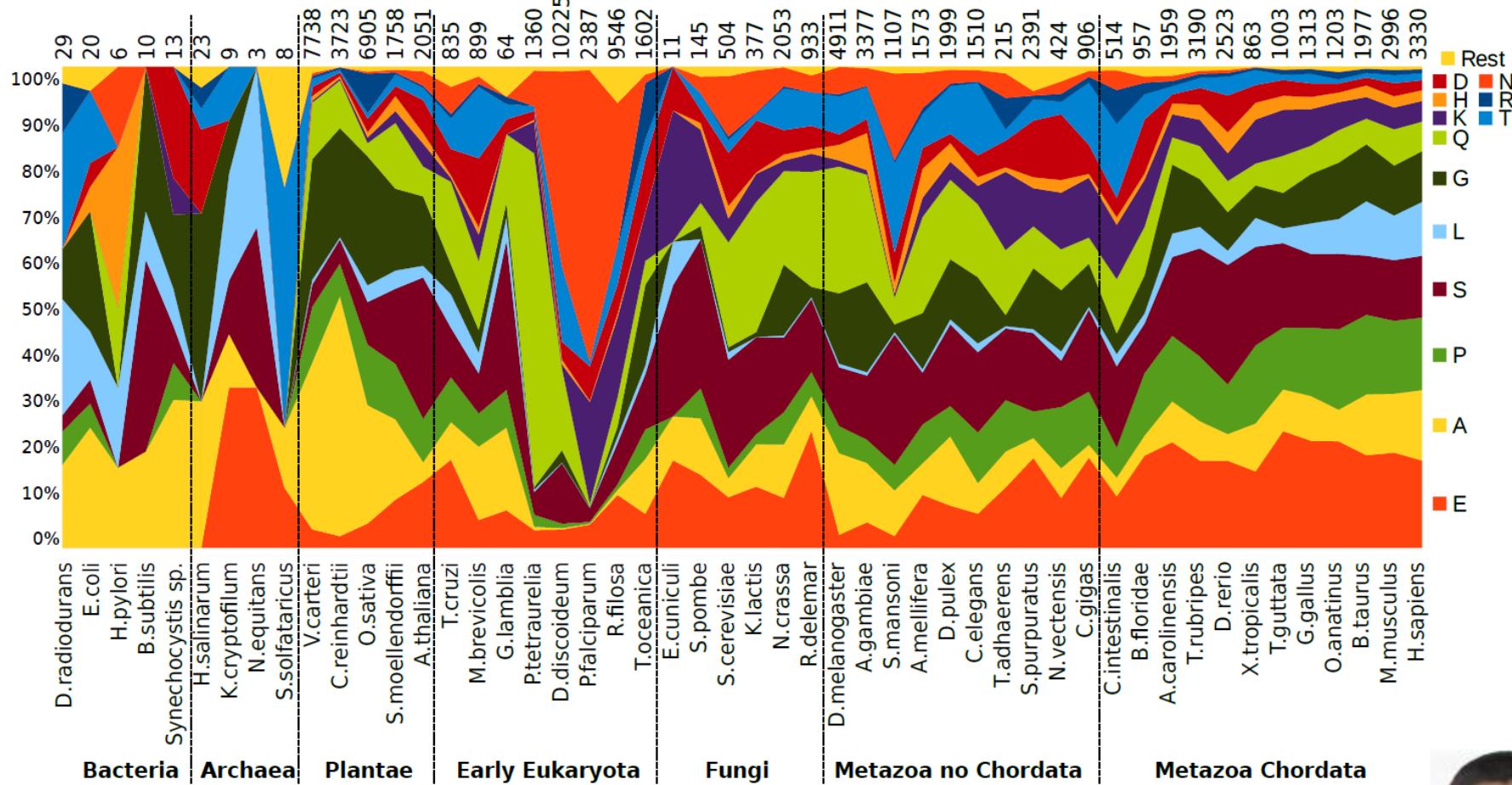
#	Start	End	Type	Parent region or Region's class
0	452	801	Region	4.4.1.1 - WD...

1. looks good
2. one repeat wrong
- 3. many mistakes**
4. total mess

5	674	718	Unit	#0 from 452
6	719	768	Unit	#0 from 452
7	769	801	Unit	#0 from 452



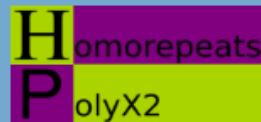
# Evolution of homorepeats in 50 species



Pablo  
Mier



# polyX2: Web tool to find homorepeats



PolyX2 is a web tool to search for homorepeats in a given protein dataset. It processes ~500 proteins per second. File upload is limited to 50 Mb (~2 minutes running time). For bigger datasets we recommend to use the standalone version of the script (see below).

Mier P and Andrade-Navarro MA. PolyX2: fast detection of homorepeats in large protein datasets. Genes 13(2022), 758. PMID:35627143.

## EXECUTION

Upload a file with one or more **protein sequence/s**, in **fasta format**  Keine ausgewählt

or paste the sequence/s here: [example1: [HD\\_HUMAN](#)] [example2: [SARS-CoV-2 complete proteome](#)]

Minimum number of  identical residues in a local window of  amino acids.

Will search homorepeats from amino acids:

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

# Exercise 2. Searching for homorepeats in a protein with polyX2

- Find a random human protein in **UniProt**. Write the Entry Name.

Entry	Entry Name	Protein Names	Gene Names	Organism
<input type="checkbox"/> P36888	FLT3_HUMAN	Receptor-type tyrosine-protein kinase FLT3 <a href="#">[...]</a>	FLT3, CD135, FLK2, STK1	Homo sapiens (Human)

## Exercise 2. Searching for homorepeats in a protein with polyX2

- Find a random human protein in **UniProt**.  
Write the Entry Name.
- Paste the sequence in polyX2
- Find: (1) The longest pure homorepeat(s) (length and amino acid type)? (2) The longest polyA region allowing 2 residues in a window of 3?

# Exercise 3. Find structures of short repeats

- Go to the PDB web page: <https://www.rcsb.org/>
- Go to Search > Advanced Search
- Type a short repeat in the Sequence Motif: as many repeats as possible; Use Mode PROSITE or RegExp (example: QQQQ / another example: [SP]SA[SP]SA[SP]SA )
- Click “count” to see if there are any cases
- Check the structure (3D View) to see if the repeat is in the structure (it could be disordered = absent)
- Write down the PDB name, the RegExp used and secondary structure type (absent if not part of the structure = flexible)  
Example 3NB9 QQQQ alpha-helix  
Example 3FRC NNNN absent