



InterPro



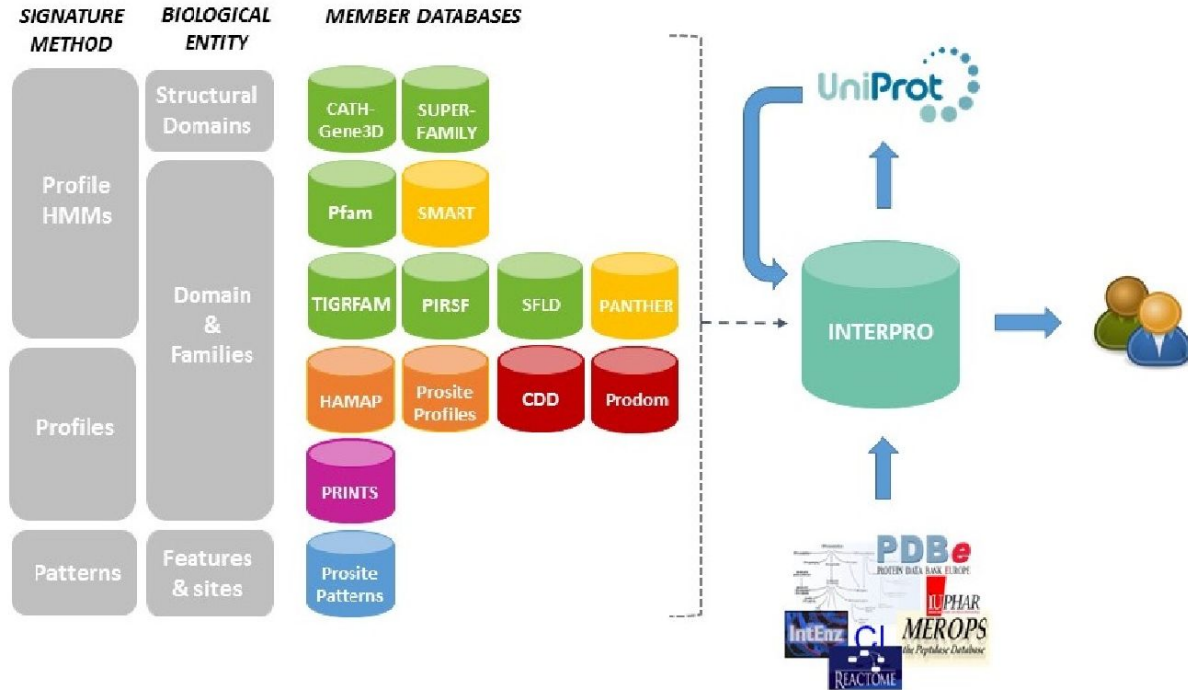
biopython

InterProFetcher

Biopython package for accessing data deposited in InterPro

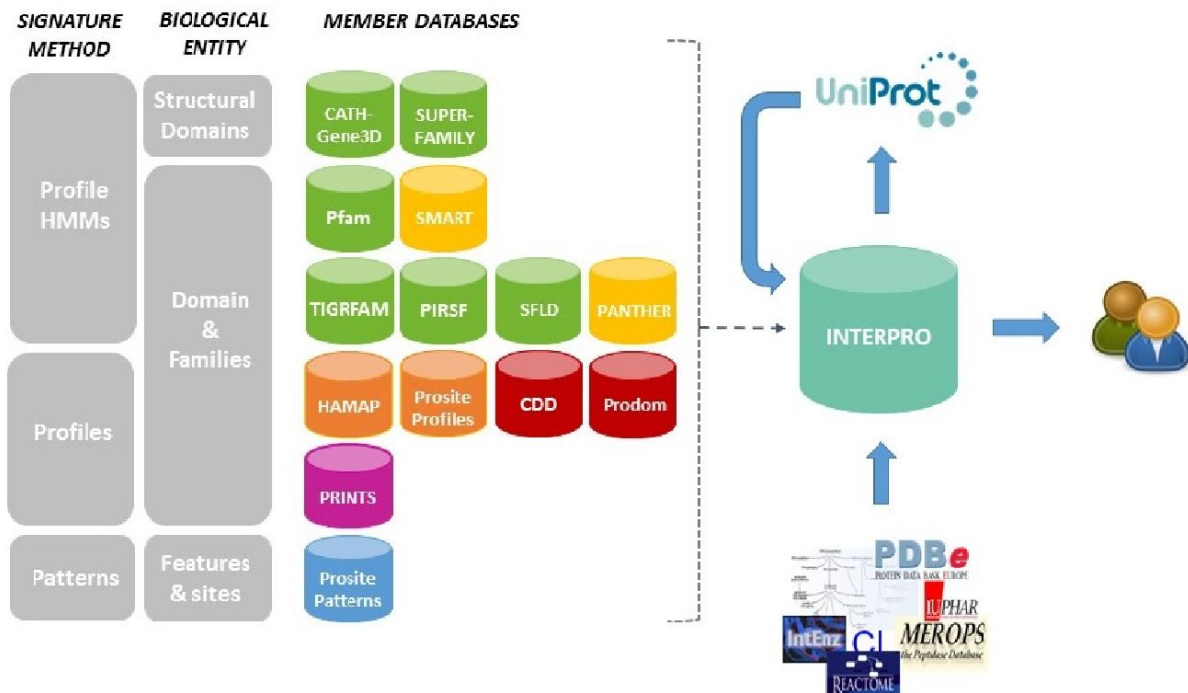
Motivations:

- InterPro is a comprehensive source as it has cross-references
- there is no such package in Biopython
- easier access to the data
- no need to manually download the data
- everything in one library :)



Types of data:

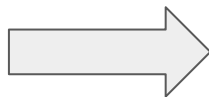
○ F	Family	21k
○ D	Domain	13k
○ H	Homologous Superfamily	3k
○ R	Repeat	340
○ S	Conserved Site	702
○ S	Active Site	132
○ S	Binding Site	75









Ideas:

- browsing protein families by a database and a keyword

Pfam,
transmembrane



ACCESSION	NAME	PFAM TYPE	DB	INTEGRATED INTO
PF00001	7 transmembrane receptor (rhodopsin family)	family		IPR000276
PF00002	7 transmembrane receptor (Secretin family)	family		IPR000832
PF00003	7 transmembrane sweet-taste receptor of 3 GCPR	domain		IPR017978
PF00664	ABC transporter transmembrane region	family		IPR011527
PF00939	Sodium:sulfate symporter transmembrane region	family		IPR001898
PF01007	Inward rectifier potassium channel transmembrane domain	domain		IPR040445

...

Ideas:

- browsing InterPro database by a keyword and type (e.g. domain)

domain, glucose



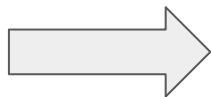
ACCESSION	NAME	INTEGRATED SIGNATURE(S)	GO TERMS BP MF CC
IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal	PS00624 PS00623 PF00732	<ul style="list-style-type: none">oxidoreductase activity, acting on CH-OH group of donorsflavin adenine dinucleotide binding
IPR001732	UDP-glucose/GDP-mannose dehydrogenase, N-terminal	PF03721	<ul style="list-style-type: none">oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptorNAD binding
IPR004719	Phosphotransferase system, maltose/glucose-specific subfamily IIC component	TIGR00852	<ul style="list-style-type: none">protein-N(P)-phosphohistidine-sugar phosphotransferase activityphosphoenolpyruvate-dependent sugar phosphotransferase systemmembrane
IPR007867	Glucose-methanol-choline oxidoreductase, C-terminal	PF05199	<ul style="list-style-type: none">oxidoreductase activity, acting on CH-OH group of donors
IPR010551	Glucose-6-phosphate isomerase, prokaryote	PF06560	<ul style="list-style-type: none">glucose-6-phosphate isomerase activitygluconeogenesisglycolytic processcytoplasm
IPR011299	PTS system glucose-specific IIBC component	TIGR02002	<ul style="list-style-type: none">glucose transmembrane transporter activityglucose transmembrane transportmembrane
IPR012938	Glucose/Sorbose dehydrogenase	PF07995	

...

Ideas:

- browsing proteins by database and an organism

InterPro,
homo sapiens



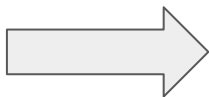
A0A023HHK9	Methylcytosine dioxygenase TET	Homo sapiens (Human)
A0A023HHL0	Methylcytosine dioxygenase TET	Homo sapiens (Human)
A0A023HJ61	RAB4A	Homo sapiens (Human)
A0A023I7F4	Cytochrome b	Homo sapiens (Human)
A0A023I7H2	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)
A0A023I7H5	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7J4	NADH-ubiquinone oxidoreductase chain 2	Homo sapiens (Human)
A0A023I7L8	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7N5	NADH-ubiquinone oxidoreductase chain 1	Homo sapiens (Human)
A0A023I7N7	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7N8	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)
A0A023I7R1	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)

...

Ideas:

- browsing protein structures by a database and a keyword + downloading them from PDB

InterPro,
lysozyme



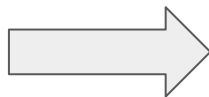
102I	HOW AMINO-ACID INSERTIONS ARE ALLOWED IN AN ALPHA-HELIX OF T4 LYSOZYME	X-ray	1.74 Å	
103I	HOW AMINO-ACID INSERTIONS ARE ALLOWED IN AN ALPHA-HELIX OF T4 LYSOZYME	X-ray	1.9 Å	
104I	HOW AMINO-ACID INSERTIONS ARE ALLOWED IN AN ALPHA-HELIX OF T4 LYSOZYME	X-ray	2.8 Å	
107I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.8 Å	
108I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.8 Å	
109I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.85 Å	
110I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.7 Å	
111I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.8 Å	
112I	STRUCTURAL BASIS OF ALPHA-HELIX PROPENSITY AT TWO SITES IN T4 LYSOZYME	X-ray	1.8 Å	











...

Ideas:

- browsing proteomes by organism

Escherichia coli

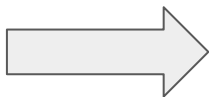


ACCESSION	NAME	ENTRY COUNT	ENTRY ACCESSIONS	PROTEIN COUNT	FASTA
UP000000296	Escherichia phage phiEB49	54	 Generate	34	 Generate
UP000000320	Escherichia phage D108 (Bacteriophage D108)	59	 Generate	38	 Generate
UP000000369	Escherichia phage 186 (Bacteriophage 186)	63	 Generate	37	 Generate
UP000000489	Escherichia phage K30	65	 Generate	32	 Generate
UP000000558	Escherichia coli O157:H7	7k	 Generate	5k	 Generate
...					

Ideas:

- downloading protein sequences by UniProt accession number (one or many proteins in an output file)

A0A023GYZ4



1 - 1 of 1 protein matching InterPro[®]

A0A023GYZ4

ACCESSION	NAME	SPECIES	LENGTH
A0A023GYZ4	Cytochrome c oxidase subunit 1	Schizonycha transvaalica	275

Cytochrome c oxidase subunit 1

A0A023GYZ4 **Fragment**

UniProtKB/TrEMBL protein

Overview
Entries 3
Sequence
Similar Proteins 2M
AlphaFold 1

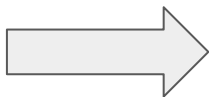
Short name A0A023GYZ4.9SCAR
Length 275 amino acids
Species Schizonycha transvaalica
Function Component of the cytochrome c oxidase, the last enzyme in the mitochondrial electron transport chain which drives oxidative phosphorylation. The respiratory chain contains 3 multisubunit complexes succinate dehydrogenase (complex II, CII), ubiquinol-...
[Show More](#)






External Links
[View A0A023GYZ4 in UniProtKB](#)
Export Matches [TSV]
[Generate](#)
[Search protein with HMMER](#)
[Search protein with InterProScan](#)
[Download Sequence](#)

Ideas:

- downloading proteomes by InterPro proteome IDs (sequences)

UP000000296



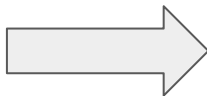
1 - 1 of 1 proteome matching InterPro [®]					UP000000296			 Export	
ACCESSION	NAME	ENTRY COUNT	ENTRY ACCESSIONS	PROTEIN COUNT	FASTA				
UP000000296	Escherichia phage phiEB49	54	 Download	34	 Download				

...

Ideas:

- downloading sequences for families/superfamilies and others by ID (e.g. family ID)

IPR036959



Download

Explanation

This FASTA file will contain **a list** of approximately 13k **UniProt proteins** which match with the **InterPro entry** with accession **IPR036959**.

We expect this file to contain 13k distinct proteins. If you encounter any problems during the creation of this file, please check the “Code snippet” section of this page for to see how to download the data directly onto your computer.

Please generate the file in order to download it.

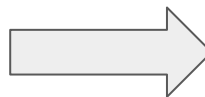
[Generate](#) [Download](#)

- online generation and downloading takes a long time and sometimes fails
- other option is to download a script

Ideas:

- adding option to make csv from results

A0A023HHK9	Methylcytosine dioxygenase TET	Homo sapiens (Human)
A0A023HHL0	Methylcytosine dioxygenase TET	Homo sapiens (Human)
A0A023HJ61	RAB4A	Homo sapiens (Human)
A0A023I7F4	Cytochrome b	Homo sapiens (Human)
A0A023I7H2	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)
A0A023I7H5	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7J4	NADH-ubiquinone oxidoreductase chain 2	Homo sapiens (Human)
A0A023I7L8	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7N5	NADH-ubiquinone oxidoreductase chain 1	Homo sapiens (Human)
A0A023I7N7	ATP synthase subunit a	Homo sapiens (Human)
A0A023I7N8	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)
A0A023I7R1	NADH-ubiquinone oxidoreductase chain 5	Homo sapiens (Human)



Technical stuff:

- tutorial and documentation
- learn how to test it with other BioPython packages
- exemplary usage in functions descriptions

Thank you for your attention!

Izabela Fedorczyk, Roksana Malinowska, Weronika Trawińska