PROTEIN

RCSB PDB-101

Introduction

Protein name is derived form a Greek word PROTOS which means “the first or the supreme’’.

Proteins are extremely complicated and nitrogenous molecule made up of variable number of amino acid residue joined to each other by a specific covalent bond called peptide bond.

20 amino acid which have been found to occur in all proteins, known as standard amino acid.

Why are proteins important to us?

- Proteins make up about 15% of the mass of the average person
- Enzymes act as a biological catalyst
- Storage and transport – Hemoglobin
- Defence -Antibodies
- Hormones – Insulin
- Ligaments and arteries (mainly formes by elastin Protein)
- Muscle – Proteins in the muscle respond to nerve impulses by changing the packing of their molecules (Actin and myosin)
- Hair, nails and skin: Protein keratin as main component

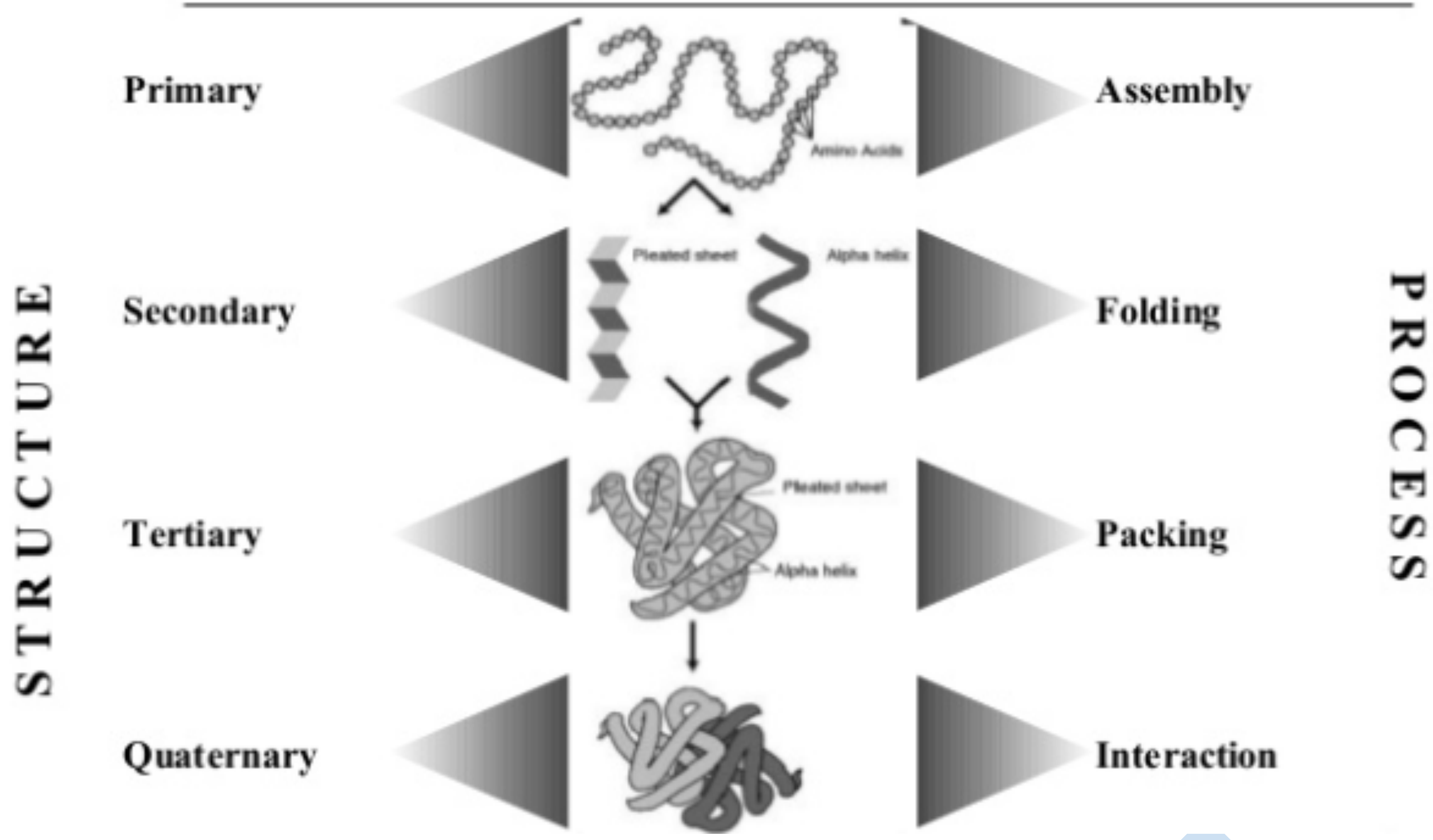
Levels in Protein structure

Majority of protein are compact and highly convoluted molecules.

Each polypeptide assumes at least three levels of structural organization termed as primary, secondary and tertiary structure.

Proteins which possess more than one polypeptide chain in their molecule also possess a fourth structure called quaternary structure.

Chemistry of Protein Structure



Primary structure

The sequence of amino acid residues along the peptide is called primary structure of the peptide.

It also include the determination of the number of amino acid residues in a peptide chain.

Shows whether the peptide chain is open, cyclic or branched.

Primary structure is linear, ordered and 1 dimensional.

Written from amino end to carboxyl end that is N to C.

primary structure of human insulin

CHAIN 1: GIVEQ CCTSI CSLYQ LENYC N

CHAIN 2: FVNQH LCGSH LVEAL YLVCG ERGFF YTPKT

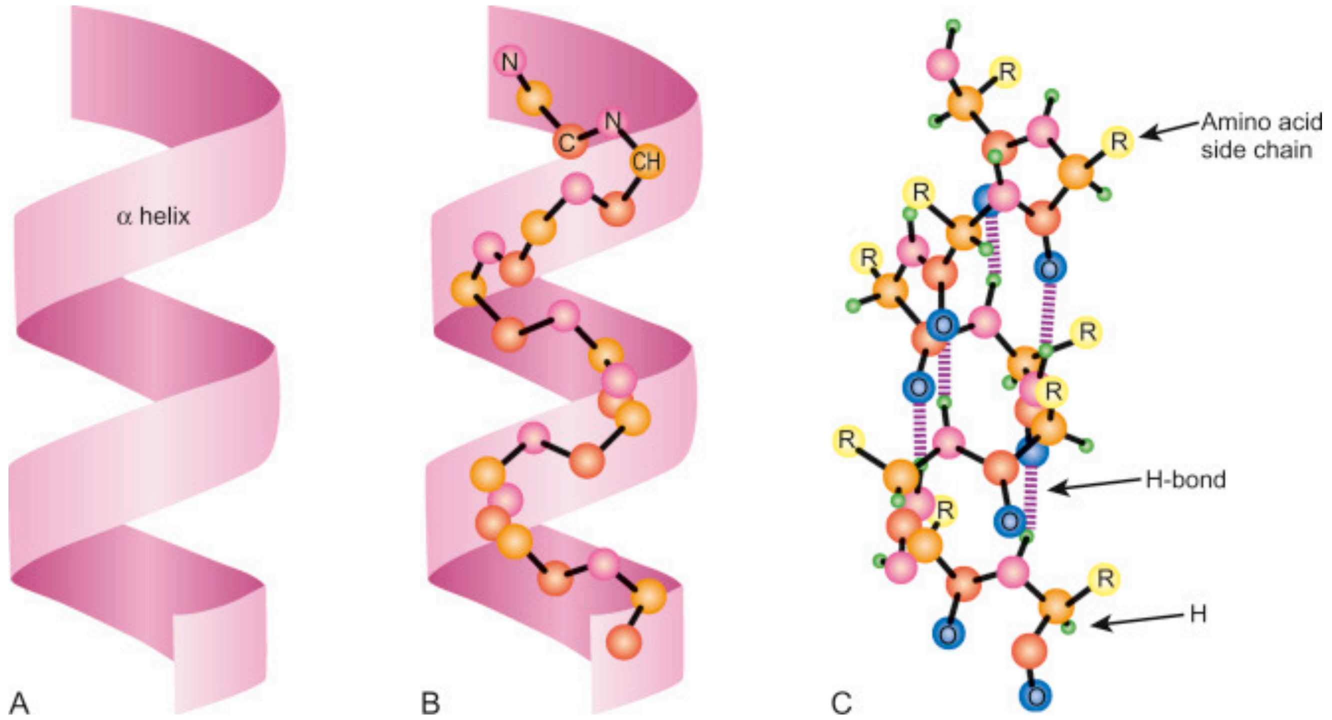
Secondary Structure

- Primary structure shows that peptide are quite straight and extended.
- X-rays diffraction on protein crystals shows that polypeptide chain tend to twist or coil upon themselves.
- The folding of the polypeptide chain into specific coiled structure held together by H bonds is called secondary structure of protein.
- Secondary structure may take one of the following form.

1. Alpha – Helix
2. Beta Pleated Sheet
3. Loop or Coil Conformation
4. Super secondary motifs

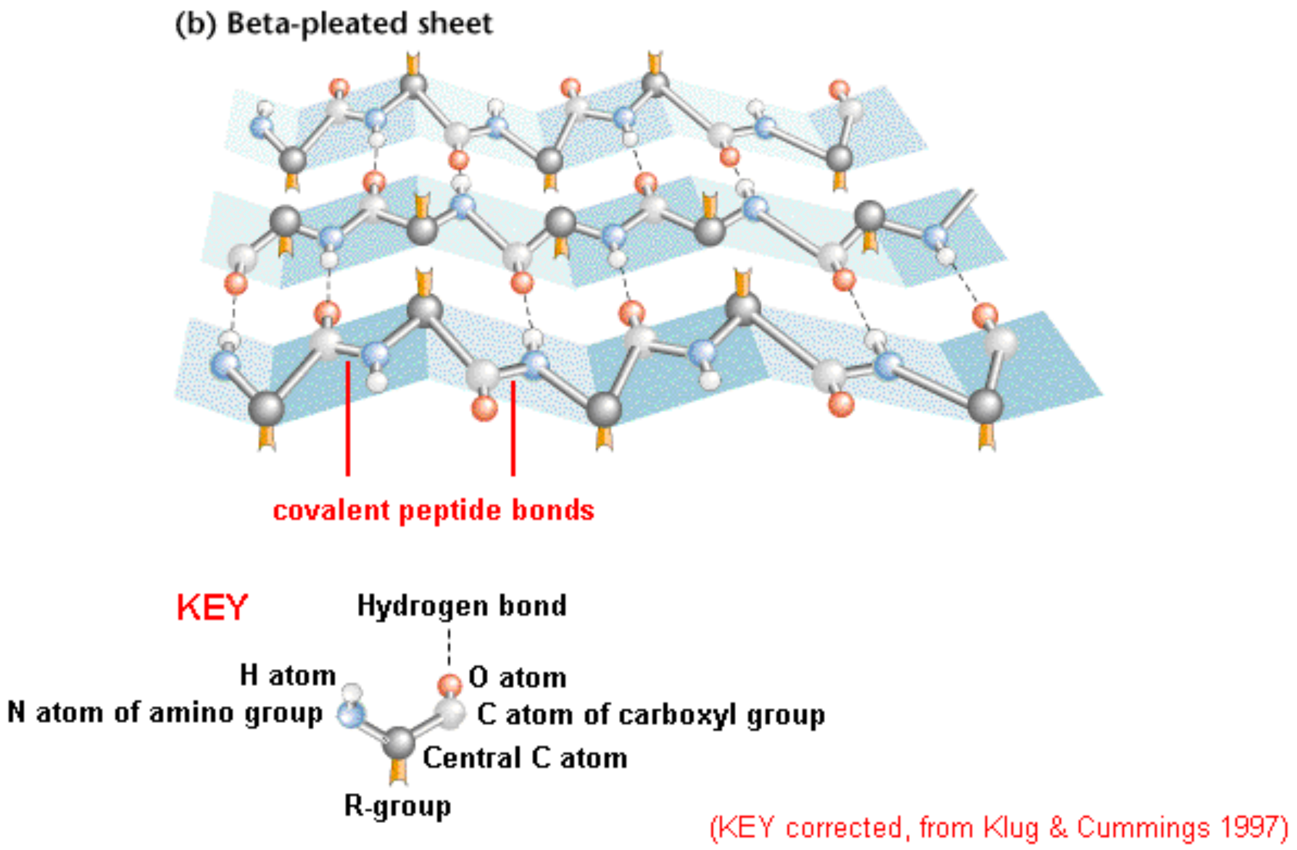
Alpha(α)- Helix

1. It is a clockwise rodlike spiral shape.
2. Formed by intrachain Hydrogen bonding between C=O group of each amino acid and NH_2 group that is present 4 residue ahead.
3. Proteins have great strength and elasticity.
4. Can easily be stretched due to tight coiling.



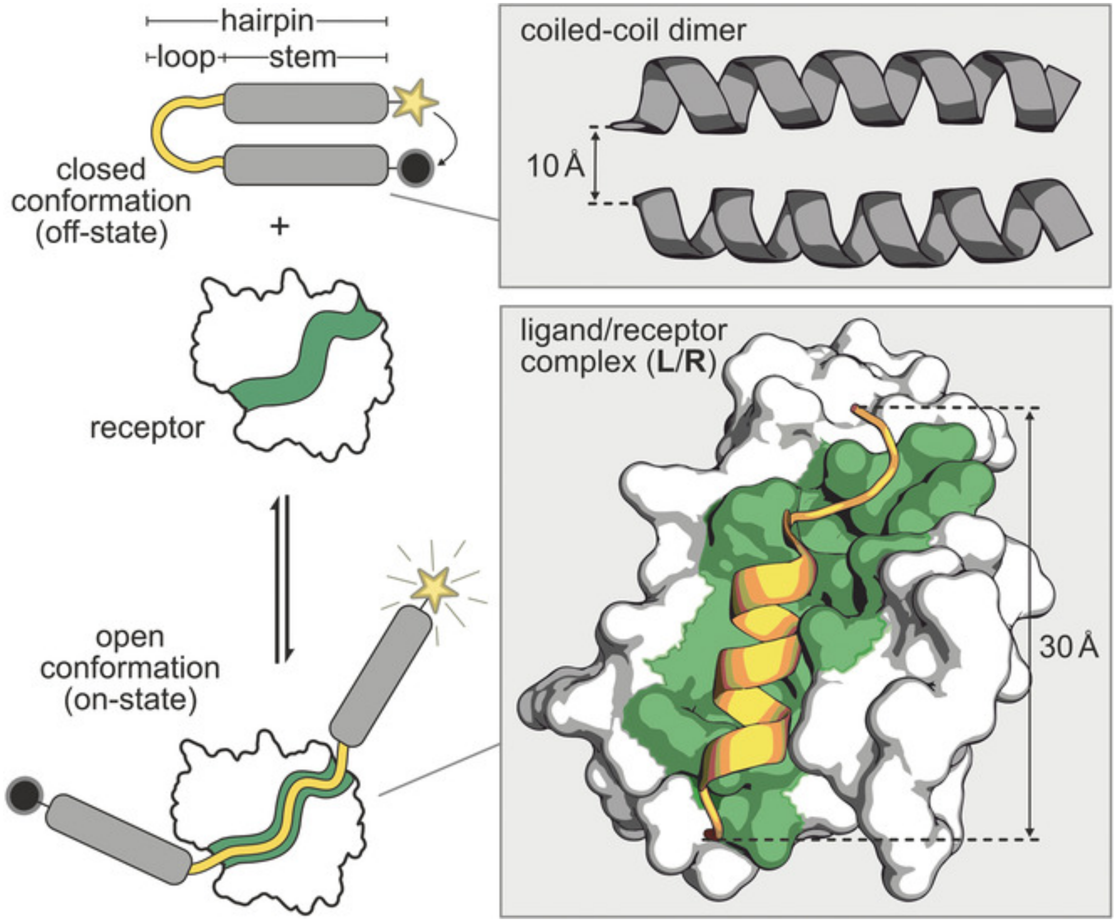
β- Pleated Sheath

1. 5 to 10 amino acid in this structure line up side by side just like a sheath of cloth can be folded again and again
2. Hydrogen bond present between the peptide strands that is interstrand.
3. This form is fully expended and can't be further stretched and they are inelastic



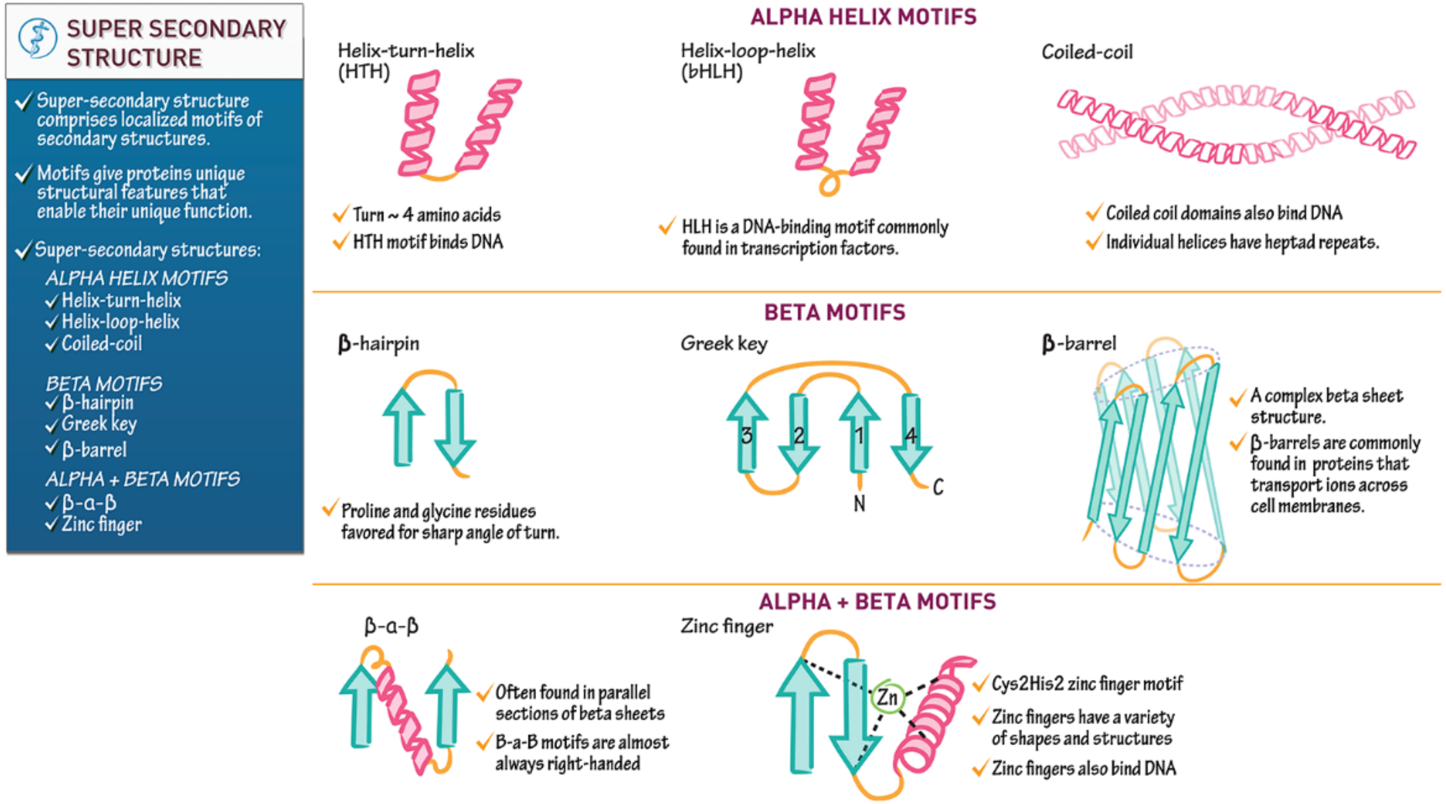
Loop or Coil Conformation

1. Present mainly in globular protein.
2. Connect two Alpha helix or Beta sheath.
3. Present in those area where bend is required.



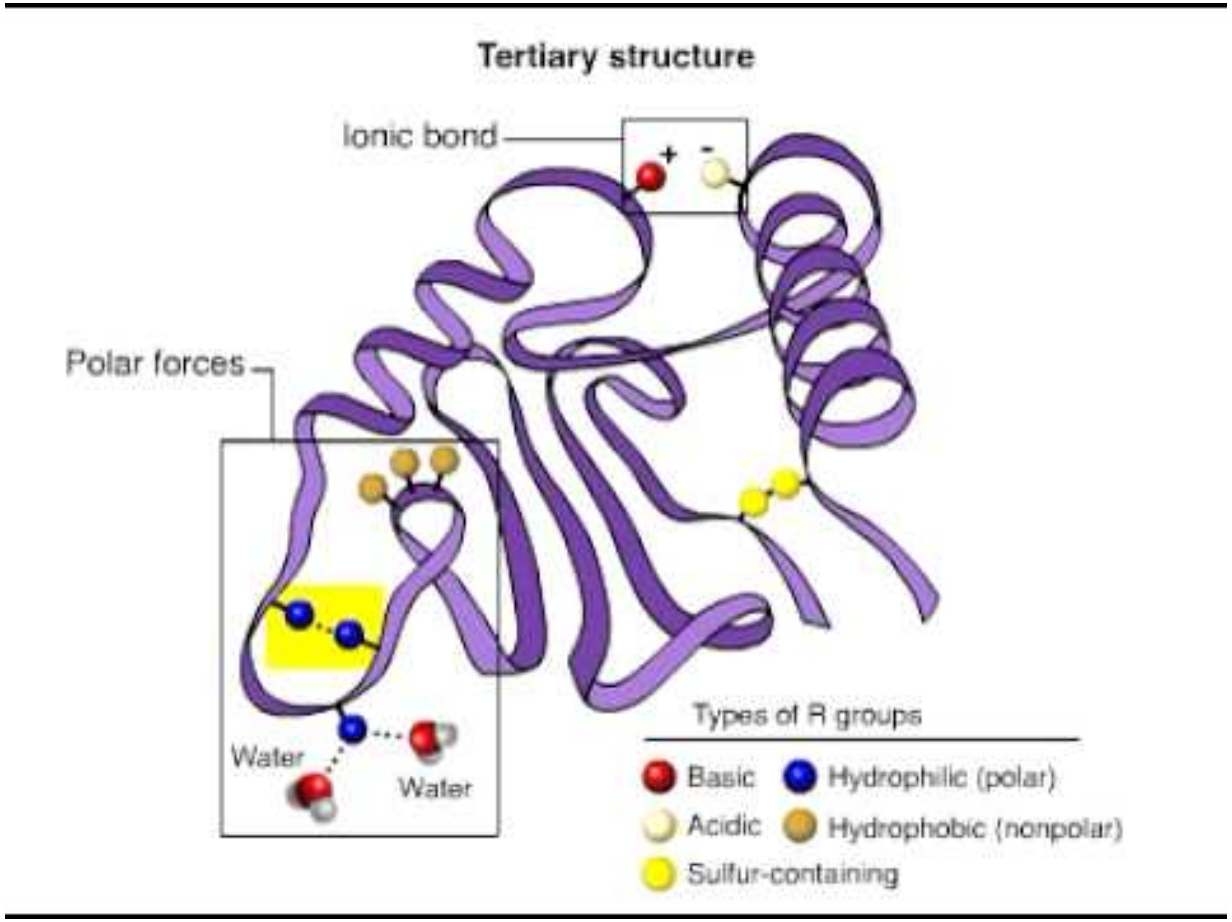
Super secondary Motifs

1. Present in Globular protein.
2. This structure form when two beta pleated sheath are connected to each other by an alpha helix.
3. For example β-α-β supersecondary motif



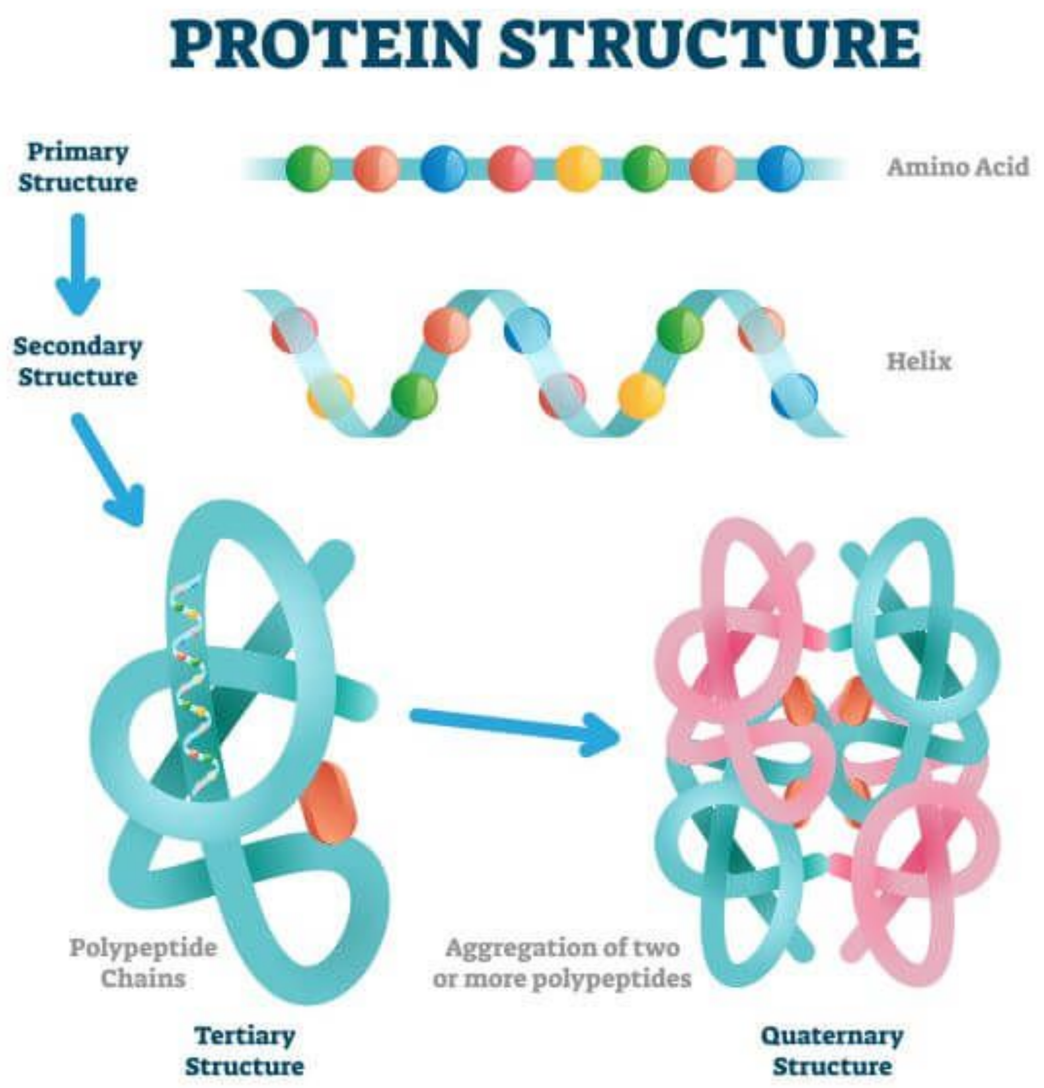
Tertiary structure

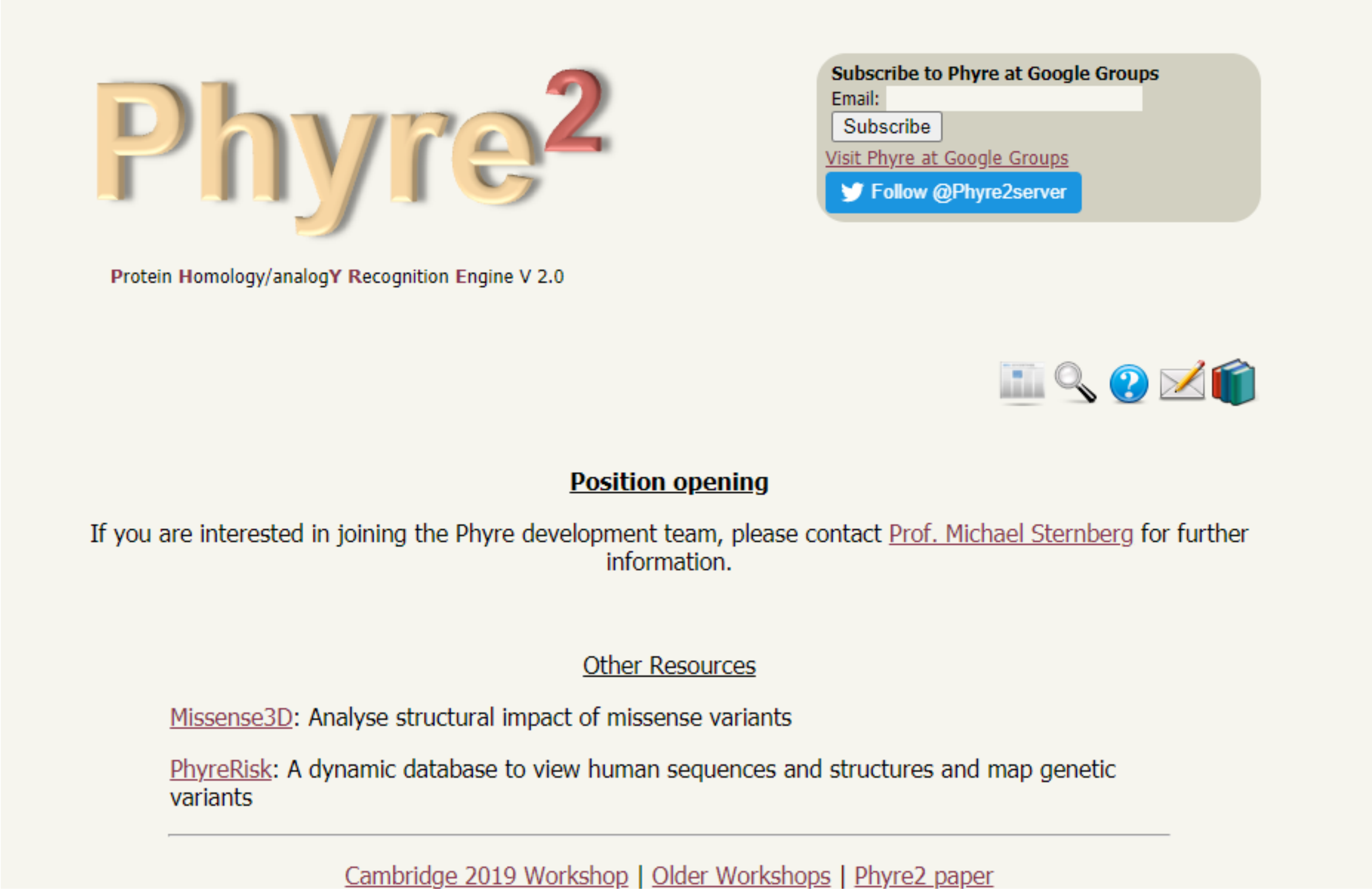
1. The tertiary structure mean the overall conformation of a polypeptide.
2. Myoglobin chain is when fully extended its length is 20 time than is width.
3. X-rays diffraction show that its structure is just like a foot ball i.e. globular.
4. The globular structure is due to folding and refolding



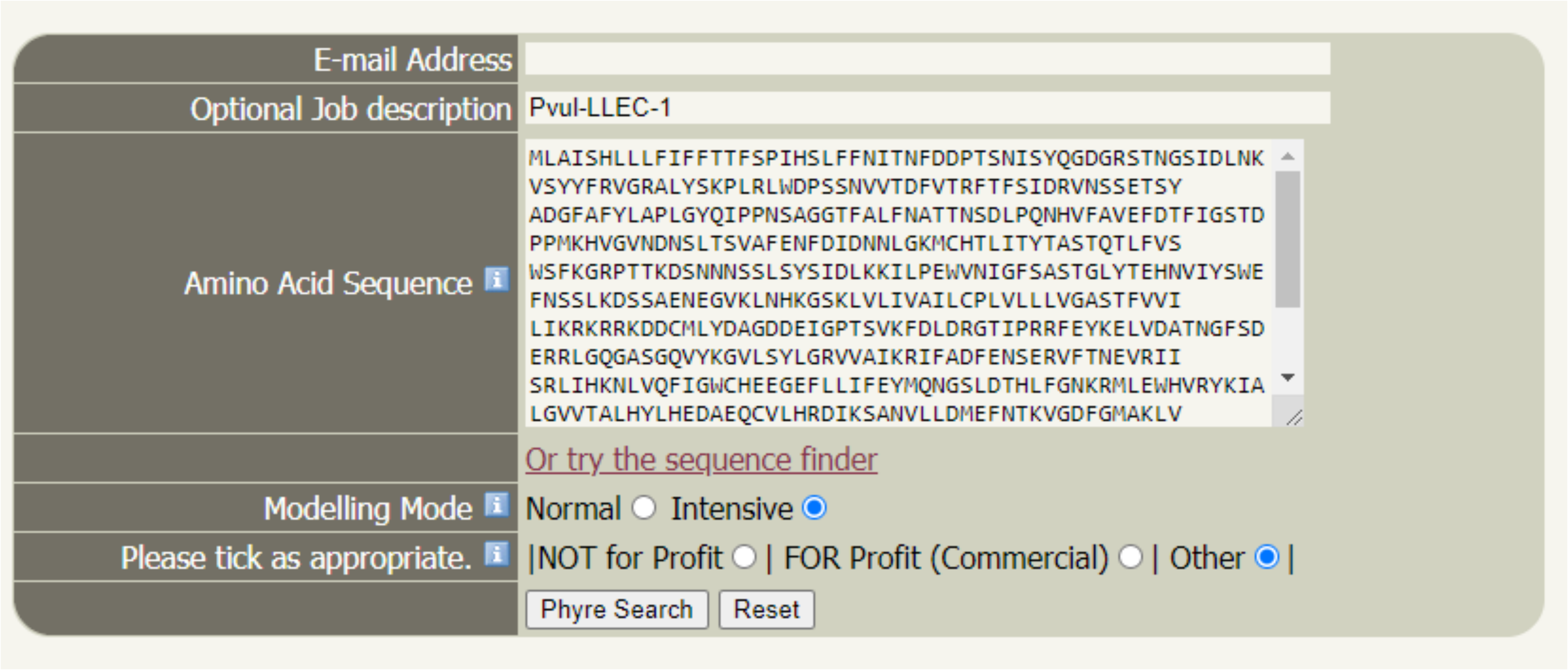
Quaternary Structure

1. Formed by those protein having more than one peptide chain subunit.
2. Each peptide have its own primary, secondary, and tertiary structure.
3. The number and arrangement of the over all structure of the peptide subunit is called quaternary structure.
4. For example structure of Hemoglobin.

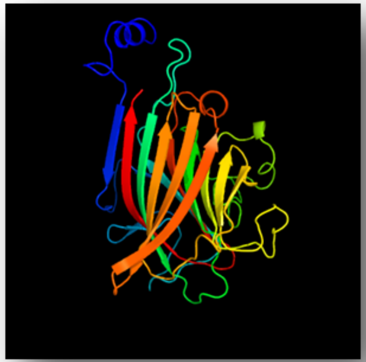

PROTEIN STRUCTURE
Primary Structure
Amino Acid
Secondary Structure
Helix
Polypeptide Chains
Tertiary Structure
Aggregation of two or more polypeptides
Quaternary Structure



- http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index



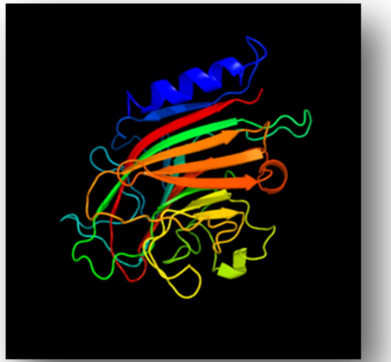
E-mail Address
Optional Job description
Pvul-LLEC-1
Amino Acid Sequence
MLAISHLLLFIFFTTFSPIHSLFFNITNFDDPTSNISYQGDGRSTNGSIDLNK
VSYYFRVGRALYSKPLRLWDPSSNVVTDFVTRFTFSIDRVNSSETSY
ADGFAFYLAPLGYQIPPNSAGGTFALFNATTNSDLPQNHVFAVEFDTFIGSTD
PPMKHVGVNDNSLTSVAFENFDIDNNLGKMCHTLITYTASTQTLFVS
WSFKGRPTTKDSNNNSSLSYSIDLKKILPEWVNIGFSASTGLYTEHNVIYSWE
FNSSLKDSSAENEGVKLNHKGSKLVLIVAILCPLVLLLVGASTFVVI
LIKRKRRKDDCMLYDAGDDEIGPTSVKFDLDRGTIPRRFEYKELVDATNGFSD
ERRLGQGASGQVYKGVLSYLGRVVAIKRIFADFENSERVFTNEVRII
SRLIHKNLVQFIGWCHEEGEFLLIFEYMQNGSLDTHLFGNKRMLEWHVRYKIA
LGVVTALHYLHEDAEQCVLHRDIKSANVLLDMEFNTKVGDFGMAKLV
Or try the sequence finder
Modelling Mode
Normal
Intensive
Please tick as appropriate.
|NOT for Profit | FOR Profit (Commercial) | Other |
Phyre Search
Reset



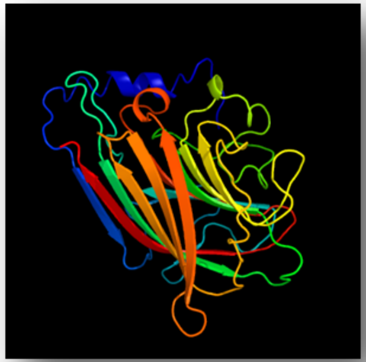
Pvul-BLEC-3

Pvul-BLEC-5

Pvul-BLEC-6

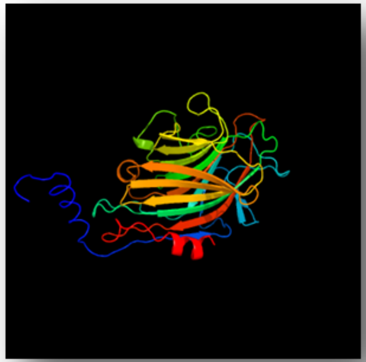


Pvul-BLEC-7

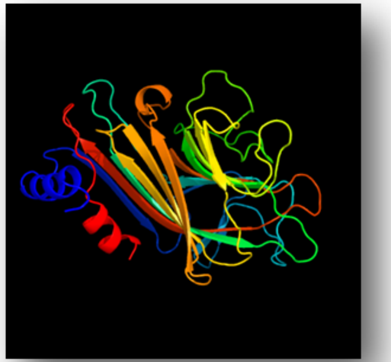


Pvul-BLEC-8

Pvul-BLEC-10

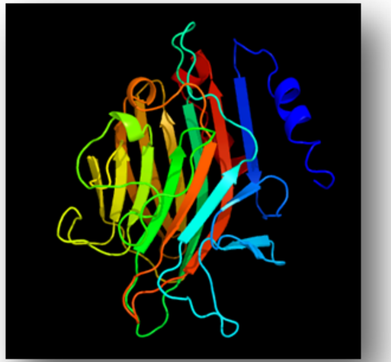


Pvul-BLEC-12



Pvul-BLEC-13

Pvul-BLEC-14



Pvul-BLEC-17

SECONDARY DATABASES

- **Secondary databases** make use of publicly available sequence data in primary **databases** to provide layers of information to DNA or protein sequence data.
- **Secondary databases** comprise data derived from analysing entries in primary **databases**.

- **ExPASy** (**Ex**pert **P**rotein **A**nalysis **Sy**stem)

https://web.expasy.org/protparam/

Expasy ProtParam

ProtParam tool

ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given p
user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic con
life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

RESET | Compute parameters

SNVVTDFVTRFTFSIDRVNSSETSY
IDNNLGKMCHTLITYTASTQTLFVS
KLVLIVAILCPLVLLLVGASTFVVI
VVAIKRIFADFENSERVFTNEVRII
IKSANVLLDMEFNTKVGDFGMAKLV
AADEKLRNEFDENEMRSLLVVGLWC

GFPLRFKNSSNGHVTDFSTRFSFTI
FFDIDSNIGNMGHVLITYNASAKLL
TVRNKLPVQVVAVAVSCSILFMLVV
VEDSERIFRNEVKIISGLVHRNLVQ
DFNTKISDFGIAKLVDPRLRTQKTK
YDVNEMTCLLTVGIWCSHPDHRQRP

LHLWDSSSSVVTDFTTRFTFSIEKG
ATTGKFDIDENLGKKCNALVTYNAS
NEKSGKGLSKVMIVVVGTCFMVFVA
LSYFGRVVAVKRIFTNFENSERVFI
NEVRIISRLIHRNLVQFVGWCHEQGEFLLVFDYMPNGSLDTHLFGDKKPLAWDIRYKVALGVALALRYLHEDAEQSVLHRDIKSANVLLDTDFSTKLGDF
GMAKLVDPRLKTQRTGVVGTYGYLAPEYMNGGRASKESDMYSFGVVALEIACGRRTYLDGEFHIPLMNWVWQQYVEGNVMDVVDERLNMEFDVDEMRSLI
IVGLWCTNPNDKERPKAAQVIKVLELEAPLPELPLDMHDRPPLSLSTYNHAQPTHNSLQSLPFTNSFVTIGR

>Pvul-LLEC-4
MLATSKNSHYFGSFLLLLILPRTIAQPFSFSITNFDDTENAGLIGYAGVAKILNGSIQLNSLIYSGIGRAIYGQPLHLKNSSNGKLTDFSTRFSFTIQSP
DTIYGDGFGFYVAPLSYQIPNTMIAGSGLGLYYENIPILAVEFDTFINDLDPPMQHVGINNGSVVSLNYTKFDIESNKGNMGHALITYNASAKLIAVSWF
FDGSSSASTPNAYLSYQIDLAELLPEWVAIGFSGSTGSSIEENVIHSWEFSSSLDLINFTHREANKEIVFTTEYKGREKVVAVAVIWSIIFALVVISITC
WMMKKRRNEDGFCFDREATPRRFGYNELVAATNGFADDRRLGEGGHGQVYKGFLSDLGRVVAVKWISSDVEDSERIFRNEVKIISRLIHKNLVQFIGWCQ
EEGKLFLVMDYLDNGSLETHLFGNRRSLTWGVRYSIALGVVRALGYLHEDVEQCVLHRDIKAGNVLLDRDFNAKLSDFGMAKLVDPRLRSEKTRVVGTYG
YLAPEYVKEGRASKESDMYGFGVLALEIACGRRTLRNWVWKHYVDGKILNAADEKLKKDFDVSEMTCLLTVGIWCTLEDHKERPTAEEVINVLKQNVSLP
ILSAKHP

>Pvul-LLEC-5
MRLLEKVTAFVLLLAFHPPFLKTVESLNFNITNFNDPESEKTMAYVGDGQATNGTIQLNIVDYLYRVGRALYAKPLHLWDASSSVLTDFTTRFTFTIDRA

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) (

Or you can paste your own amino acid sequence (in one-letter code) in the

```
MLAISHLLLFIFFTTFSPIHSLFFNITNFDDPTSNISYQGDGRSTNGSIDLNKVSYY
FRVGRALYSKPLRLWDPSSNVVTDFVTRFTFSIDRVNSSETSY
ADGFAFYLAPLGYQIPPNSAGGTFALFNATTNSDLPQNHVFAVEFDTFIGSTDPPMK
HVGVNDNSLTSVAFENFDIDNNLGKMCHTLITYTASTQTLFVS
WSFKGRPTTKDSNNNSSLSYSIDLKKILPEWVNIGFSASTGLYTEHNVIYSWEFNSS
LKDSSAENEGVKLNHKGSKLVLIVAILCPLVLLLVGASTFVVI
LIKRKRRKDDCMLYDAGDDEIGPTSVKFDLDRGTIPRRFEYKELVDATNGFSDERRL
GQGASGQVYKGVLSYLGRVVAIKRIEADFENSERVFTNEVRII
```

RESET | Compute parameters

Expasy

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MLAISHLLLF IFFTTFSPIH SLFFNITNFD DPTSNISYQG DGRSTNGSID LNKVSYYFRV

        70         80         90        100        110        120
GRALYSKPLR LWDPSSNVVT DFVTRFTFSI DRVNSSETSY ADGFAFYLAP LGYQIPPNSA

       130        140        150        160        170        180
GGTFALFNAT TNSDLPQNHV FAVEFDTFIG STDPPMKHVG VNDNSLTSVA FENFDIDNNL

       190        200        210        220        230        240
GKMCHTLITY TASTQTLFVS WSFKGRPTTK DSNNNSSLSY SIDLKKILPE WVNIGFSAST

       250        260        270        280        290        300
GLYTEHNVIY SWEFNSSLKD SSAENEGVKL NHKGSKLVLI VAILCPLVLL LVGASTFVVI

       310        320        330        340        350        360
LIKRKRRKDD CMLYDAGDDE IGPTSVKFDL DRGTIPRRFE YKELVDATNG FSDERRLGQG

       370        380        390        400        410        420
ASGQVYKGVL SYLGRVVAIK RIFADFENSE RVFTNEVRII SRLIHKNLVQ FIGWCHEEGE

       430        440        450        460        470        480
FLLIFEYMQN GSLDTHLFGN KRMLEWHVRY KIALGVVTAL HYLHEDAEQC VLHRDIKSAN

       490        500        510        520        530        540
VLLDMEFNTK VGDFGMAKLV DPRLRTQRTG VVGTYGYLAP EYVNGGRASR ESDMYSFGVV

       550        560        570        580        590        600
ALEIASGRRT YQDGEFHVCL MNWVWQLYVE GELLRAADEK LRNEFDENEM RSLLVVGLWC

       610        620        630        640        650        660
TNPNDKERPK AAQVMKVLQL EAPLPLLPLD MYERAPPMQL ITMPHHHSNP HSGPSQPITS
```

Dalton

Measure of molecular weight or molecular mass. One molecular hydrogen molecular atom has molecular mass of 1 Da, so 1 Da = 1 g/mol. Proteins and other molecular macromolecule molecular weights are usually measured in molecular kDa or kD (kilodaltons) - 1000 Da. molecular average molecular amino molecular acid = 110 Da.

Protein Isoelectric Point calculates the theoretical **pI** (isoelectric point) for the protein sequence you enter.

Number of amino acids: 667

Molecular weight: 75271.53

Theoretical pI: 6.04

Amino acid composition: CSV format

Ala	(A)	35	5.2%
Arg	(R)	36	5.4%
Asn	(N)	39	5.8%
Asp	(D)	39	5.8%
Cys	(C)	7	1.0%
Gln	(Q)	16	2.4%
Glu	(E)	36	5.4%
Gly	(G)	45	6.7%
His	(H)	19	2.8%
Ile	(I)	33	4.9%
Leu	(L)	69	10.3%
Lys	(K)	29	4.3%
Met	(M)	15	2.2%
Phe	(F)	41	6.1%
Pro	(P)	28	4.2%
Ser	(S)	56	8.4%
Thr	(T)	39	5.8%
Trp	(W)	9	1.3%
Tyr	(Y)	25	3.7%
Val	(V)	51	7.6%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%
	(B)	0	0.0%
	(Z)	0	0.0%
	(X)	0	0.0%

- The **Instability index** is a measure of proteins, used to determine whether it will be stable in a test tube. If the **index** is less than 40, then it is probably stable in the test tube. If it is greater (for example, enaptin) then it is probably not stable.
- The aliphatic index of a protein is a measure of the relative volume occupied by aliphatic side chain of the following amino acids: alanine, valine, leucine and isoleucine. An increase in the aliphatic index increases the thermostability of globular proteins. The index is calculated by the following formula.
  Aliphatic index = X(Ala) + a*X(Val) + b*X(Leu) + b*(X)Ile$ X(Ala), X(Val), X(Ile) and X(Leu) are the amino acid compositional fractions. The constants a and b are the relative volume of valine (a=2.9) and leucine/isoleucine (b=3.9) side chains compared to the side chain of alanine
- **Grand average of hydropathicity** index (**GRAVY**) is used to represent the hydrophobicity value of a peptide, which calculates the sum of the **hydropathy** values of all the amino acids divided by the sequence length.

Instability index:

The instability index (II) is computed to be 32.38
This classifies the protein as stable.

Aliphatic index: 87.06

Grand average of hydropathicity (GRAVY): -0.176

ID	*Phaseolus vulgaris* Genomic Database Identifier	Physical position on *P. vulgaris* genome			Protein length (aa)	pI	Molecular weight (Da)	Instability index	Aliphatic index	GRAVY	Stable or unstable	NCBI Accession No.
		Chr.	Start position (bp)	End Position (bp)								
Pvul-LLEC-1	Phvul.001G045400.1.p	1	3.677.580	3.680.026	667	6.04	75271.53	32.38	87.06	-0.176	stable	XP_007161136.1
Pvul-LLEC-2	Phvul.001G040800.1.p	1	4.327.634	4.329.496	620	5.68	68751.89	27.71	90.05	-0.085	stable	XP_007161076.1
Pvul-LLEC-3	Phvul.001G040700.1.p	1	4.336.106	4.338.606	672	5.77	74995.03	28.89	83.11	-0.189	stable	XP_007161075.1
Pvul-LLEC-4	Phvul.001G040600.1.p	1	4.339.156	4.340.979	607	6.01	67907.43	31.81	94.43	-0.064	stable	XP_007161074.1
Pvul-LLEC-5	Phvul.001G040500.1.p	1	4.359.201	4.361.618	661	5.73	73921.88	25.07	89.05	-0.140	stable	XP_007161073.1
Pvul-LLEC-6	Phvul.001G040400.1.p	1	4.370.066	4.372.118	636	5.62	70870.21	34.63	91.89	-0.142	stable	XP_007161072.1
Pvul-LLEC-7	Phvul.001G040300.1.p	1	4.379.952	4.382.583	664	5.63	73983.86	25.57	87.91	-0.161	stable	XP_007161071.1
Pvul-LLEC-8	Phvul.001G040100.1.p	1	4.394.207	4.396.266	636	5.62	71103.50	34.65	90.66	-0.154	stable	XP_007161068.1
Pvul-LLEC-9	Phvul.001G040000.1.p	1	4.399.520	4.401.949	666	5.83	74151.24	25.70	88.11	-0.129	stable	XP_007161067.1
Pvul-LLEC-10	Phvul.001G234200.1.p	1	4.880.986	48.811.980	664	8.68	73859.86	33.83	92.30	-0.104	stable	XP_007163432.1
Pvul-LLEC-11	Phvul.002G214900.1.p	2	38.345.248	38.347.335	695	5.85	77309.63	32.79	87.80	-0.128	stable	XP_007159169.1
Pvul-LLEC-12	Phvul.002G215200.1.p	2	38.371.310	38.373.253	647	7.31	72858.74	30.71	85.09	-0.229	stable	XP_007159173.1
Pvul-LLEC-13	Phvul.002G215300.1.p	2	38.383.003	38.384.994	663	7.32	74176.59	31.47	90.89	-0.201	stable	XP_007159174.1
Pvul-LLEC-14	Phvul.002G215400.1.p	2	38.393.780	38.395.786	668	6.75	75537.69	36.90	91.44	-0.147	stable	XP_007159175.1
Pvul-LLEC-15	Phvul.003G204500.1.p	3	43.061.296	43.063.308	670	5.11	73707.66	42.72	86.10	-0.074	unstable	XP_007155476.1
Pvul-LLEC-16	Phvul.005G103200.1.p	5	32.195.113	32.197.596	691	5.99	77701.63	41.04	86.43	-0.115	unstable	XP_007149844.1
Pvul-LLEC-17	Phvul.005G103300.1.p	5	32.198.665	32.200.731	688	6.34	77072.28	34.86	91.89	-0.078	stable	XP_007149845.1
Pvul-LLEC-18	Phvul.006G087700.1.p	6	19.974.245	19.976.281	678	5.75	72719.33	36.68	85.27	-0.040	stable	XP_007146994.1
Pvul-LLEC-19	Phvul.006G185000.1.p	6	28.613.191	28.616.354	692	5.51	78507.66	36.34	89.22	-0.143	stable	XP_007148157.1
Pvul-LLEC-20	Phvul.006G200800.1.p	6	29.750.251	29.752.428	639	6.96	72045.08	36.62	88.17	-0.155	stable	XP_007148350.1
Pvul-LLEC-21	Phvul.007G078200.1.p	7	7.489.146	7.491.182	678	6.24	75420.09	41.56	84.10	-0.087	unstable	XP_007143518.1
Pvul-LLEC-22	Phvul.007G260300.1.p	7	38.169.156	38.172.057	670	6.25	74866.04	36.79	84.51	-0.263	stable	XP_007145693.1
Pvul-LLEC-23	Phvul.007G260400.2.p	7	38.176.633	38.178.637	450	7.61	50666.97	39.48	84.60	-0.294	stable	XP_007145694.1
Pvul-LLEC-24	Phvul.007G260500.1.p	7	38.182.876	38.185.141	657	6.43	72940.88	37.51	88.23	-0.212	stable	XP_007145695.1
Pvul-LLEC-25	Phvul.008G117700.1.p	8	14.407.039	14.409.057	672	7.01	74202.12	34.28	99.33	-0.066	stable	XP_007140498.1
Pvul-LLEC-26	Phvul.008G117800.1.p	8	14.436.075	14.438.400	668	7.32	73975.61	33.83	94.03	-0.115	stable	XP_007140499.1
Pvul-LLEC-27	Phvul.008G239600.1.p	8	58.823.983	58.826.861	699	5.59	74937.86	33.42	87.78	-0.018	stable	XP_007141948.1
Pvul-LLEC-28	Phvul.008G279300.1.p	8	62.071.477	62.073.771	662	6.64	72808.09	38.34	86.56	-0.003	stable	XP_007142422.1
Pvul-LLEC-29	Phvul.010G015800.1.p	10	2.345.298	2.347.358	686	5.74	75220.67	36.40	93.76	-0.068	stable	XP_007134058.1
Pvul-LLEC-30	Phvul.011G119200.1.p	11	18.774.057	18.775.652	531	5.98	59066.39	34.97	94.11	-0.007	stable	XP_007132721.1

SEQUENCE MOTIF

1.

E-value: 1.0e-279 Site Count: 12 Width: 50

Log Likelihood Ratio: 1306 Information Content: 160.8 Relative Entropy: 157 Bayes Threshold: 10.373

	Name	Start	*p*-value	Sites
10.	Pvul-NIN-10	521	3.38e-60	AEKTISLPVL RQYFAGSLKDAAKSIGVCPTTLKRICRQHGITRWPSRKIKKVGHSLRKLQ LVIDSVQGAE
6.	Pvul-NIN-6	518	3.38e-60	AEKTISLPVL RQYFAGSLKDAAKSIGVCPTTLKRICRQHGITRWPSRKIKKVGHSLKKLQ LVIDSVQGAE
4.	Pvul-NIN-4	609	8.99e-58	SEKSVSLSVL QQYFSGSLKDAAKNIGVCPTTLKRICRQHGISRWPSRKINKVNRSLKKIQ TVLDSVQGVE
12.	Pvul-NIN-12	606	1.72e-57	VEKNVSLSVL QQYFSGSLKDAAKSIGVCPTTLKRICRQHGISRWPSRKINKVNRSLKKIQ TVLDSVQGVE
7.	Pvul-NIN-7	609	2.13e-57	AEKTITLQVL RQYFAGSLKDAAKNIGVCTTTLKRICRQHGIKRWPSRKIKKVGHSLQKLQ LVIDSVQGAS
3.	Pvul-NIN-3	584	5.27e-54	AEKSISLDVL QHYFTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSLSKLK CVIESVHGAE
5.	Pvul-NIN-5	593	7.24e-54	TEKSISLEVL QRYFAGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKINKVNRSLSKLK RVIESVQGAE
11.	Pvul-NIN-11	89	6.91e-44	SARMLSRKTV SQYFYMPISQAAKELNVGLTHLKKRCRELGIQRWPHRKLMSLQTLIRNMQ EQGQGEGPQN
2.	Pvul-NIN-2	143	3.83e-42	SSRMLSRKTI SQYFYMPITQAARELNVGLTLLKKRCRELGIRRWPHRKLMSLQTLINNVQ ELLKEEGPES
1.	Pvul-NIN-1	213	2.79e-40	LVAKISLSDL VQYFGMPIVEASRNLKVGLTVLKRKCREFGIPRWPHRKIKSLDSLIQDLQ EEAQNQELEN
8.	Pvul-NIN-8	166	1.16e-38	KPCALEFEEI KKHFDVPINEAAKQMNVGLTMLKRRCRELNIMRWPHRKLKSLQLLIDNVK ELGLAEEVSM
9.	Pvul-NIN-9	462	5.59e-28	RVPKMTMNDL SPFFHLTIRDAADKLDVSDSVVKKISRLGNLKRWPQRKLQSLAKDVRVLR KALNSPYEGT

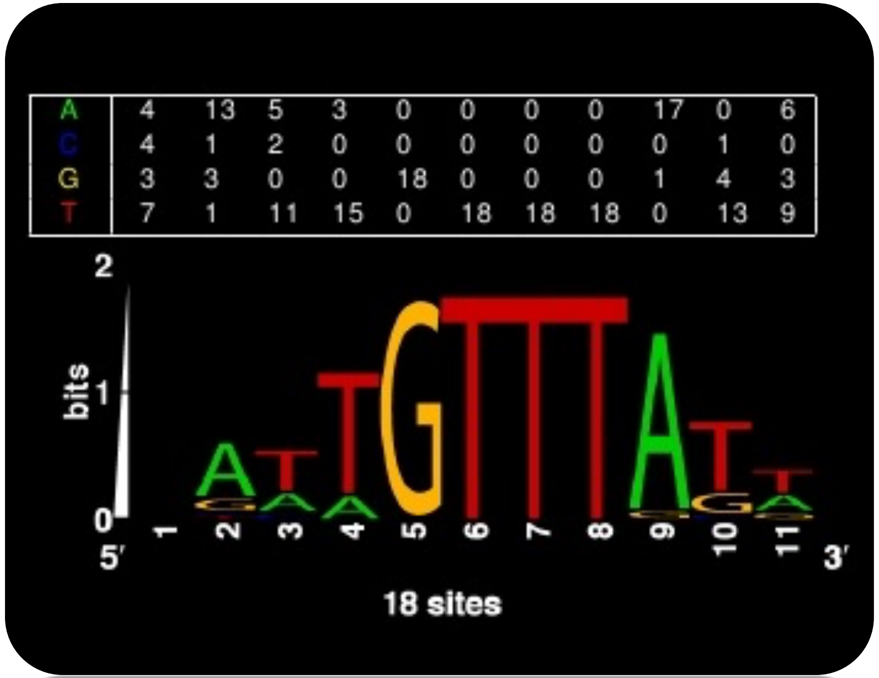
	Logo	E-value	Sites	Width	More	Submit/Download
2.		2.4e-123	7	50	↧	⇢

What is a protein motif?

- **Protein motifs** are small regions of **protein** three-dimensional structure or amino acid sequence shared among different **proteins**.
- They are recognizable regions of **protein** structure that may (or may not) be defined by a unique chemical or biological function
- Conserved string of amino acid residues

Sequence Logos

- A visual representation of the motif
- Each column of the matrix is represented as a stack of letters whose size is proportional to the corresponding residue frequency
- The total height of each column is proportional to its **information content.**



MEME Suite

- Suite of web based tools for motif discovery
- MEME - de-novo motif finding
- MAST - find matches to known motifs (MEME output)
- TOMTOM - Compare motifs to TRANSFAC and Jaspar
- http://meme-suite.org/tools/meme

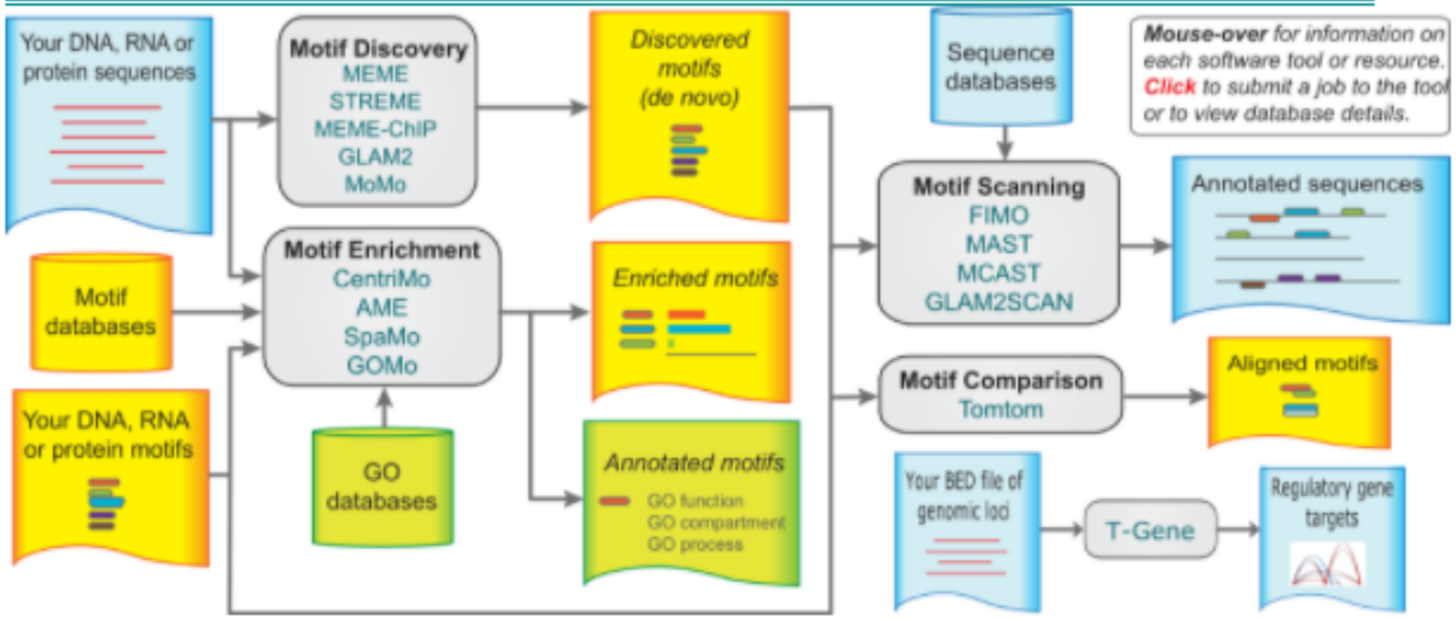
The MEME Suite

Motif-based sequence analysis tools

MEME Suite 5.3.3

- ►Motif Discovery
- ►Motif Enrichment
- ►Motif Scanning
- ►Motif Comparison
- ►Gene Regulation
- ►Manual
- ►Guides & Tutorials
- ►Sample Outputs
- ►File Format Reference
- ►Databases
- ►Download & Install
- ►Help
- ►Alternate Servers
- ►Authors & Citing
- ►Recent Jobs

↪ Previous version 5.3.2



MEME Multiple Em for Motif Elicitation

CentriMo Local Motif Enrichment Analysis

FIMO Find Individual Motif Occurrences

STREME Sensitive, Thorough, Rapid, Enriched Motif Elicitation

AME (New) Analysis of Motif Enrichment

MAST Motif Alignment & Search Tool

MEME-ChIP (Improved) Motif Analysis of Large Nucleotide Datasets

SpaMo Spaced Motif Analysis Tool

MCAST Motif Cluster Alignment and Search Tool

GLAM2 Gapped Local Alignment of Motifs

GOMo Gene Ontology for Motifs

GLAM2Scan Scanning with Gapped Motifs

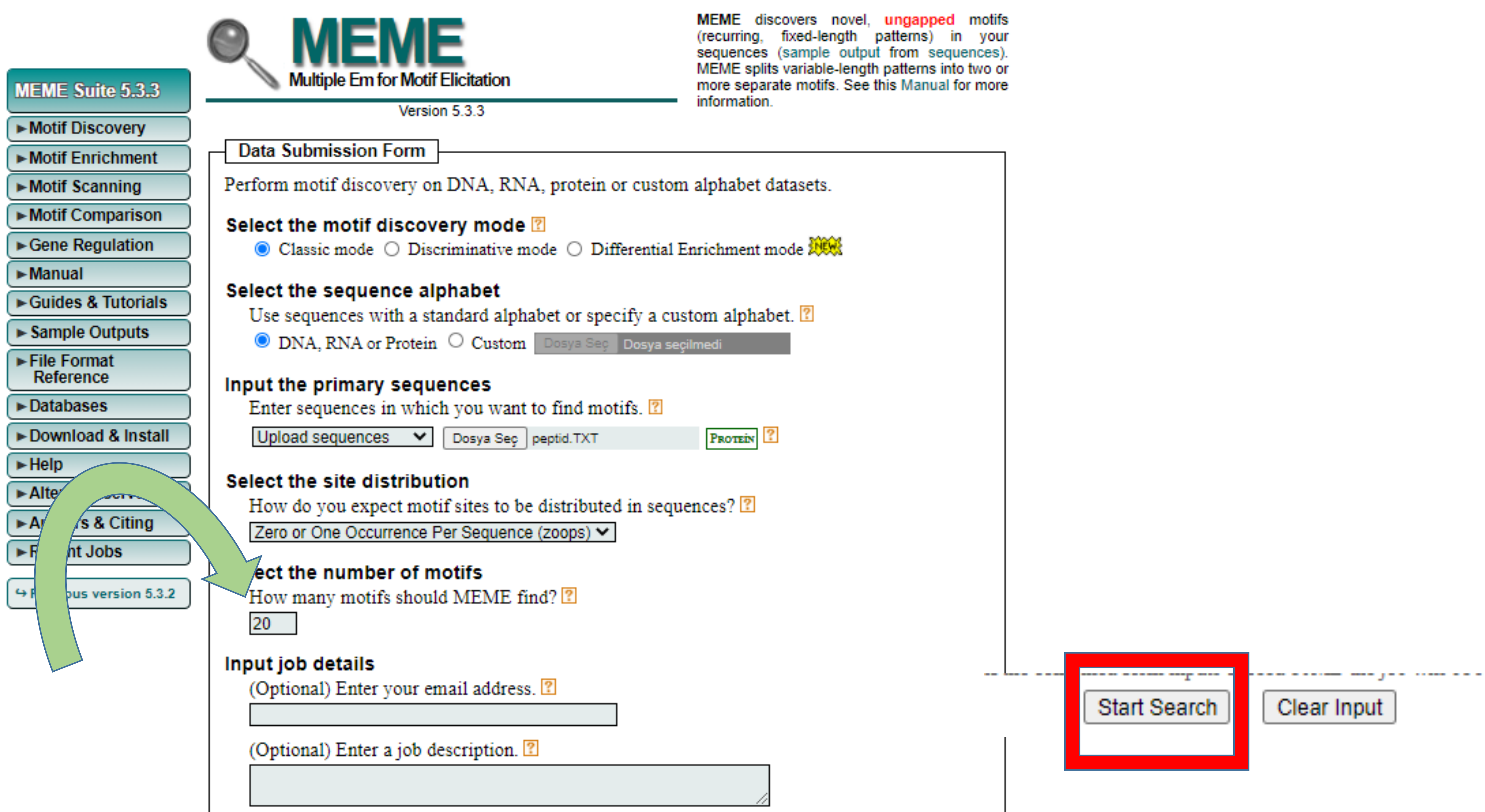
MoMo Modification Motifs

Tomtom Motif Comparison Tool

GT-Scan Identifying Unique Genomic Targets

T-Gene Predicting Target Genes

DREME Discriminative Regular Expression Motif Elicitation


MEME
Multiple Em for Motif Elicitation
Version 5.3.3
MEME discovers novel, ungapped motifs (recurring, fixed-length patterns) in your sequences (sample output from sequences). MEME splits variable-length patterns into two or more separate motifs. See this Manual for more information.
MEME Suite 5.3.3
Motif Discovery
Motif Enrichment
Motif Scanning
Motif Comparison
Gene Regulation
Manual
Guides & Tutorials
Sample Outputs
File Format Reference
Databases
Download & Install
Help
Data Submission Form
Perform motif discovery on DNA, RNA, protein or custom alphabet datasets.
Select the motif discovery mode
Classic mode
Discriminative mode
Differential Enrichment mode
NEW
Select the sequence alphabet
Use sequences with a standard alphabet or specify a custom alphabet.
DNA, RNA or Protein
Custom
Dosya Seç
Dosya seçilmedi
Input the primary sequences
Enter sequences in which you want to find motifs.
Upload sequences
Dosya Seç
peptid.TXT
PROTEIN
Select the site distribution
How do you expect motif sites to be distributed in sequences?
Zero or One Occurrence Per Sequence (zoops)
How many motifs should MEME find?
20
Input job details
(Optional) Enter your email address.
(Optional) Enter a job description.
Start Search
Clear Input

Discovered Motifs

	Logo	E-value	Sites	Width	More	Submit/Download
1.		1.0e-279	12	50	↧	⇢
2.		2.4e-123	7	50	↧	⇢
3.		3.4e-096	7	44	↧	⇢
4.		4.2e-077	7	39	↧	⇢
5.		1.2e-076	5	50	↧	⇢
6.		3.4e-070	7	41	↧	⇢
7.		4.3e-057	7	34	↧	⇢
8.		7.5e-026	7	21	↧	⇢
9.		1.3e-036	5	38	↧	⇢
10.		4.3e-026	7	21	↧	⇢
11.		8.2e-025	4	44	↧	⇢
12.		2.2e-023	7	20	↧	⇢

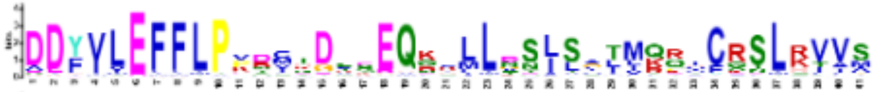
1.

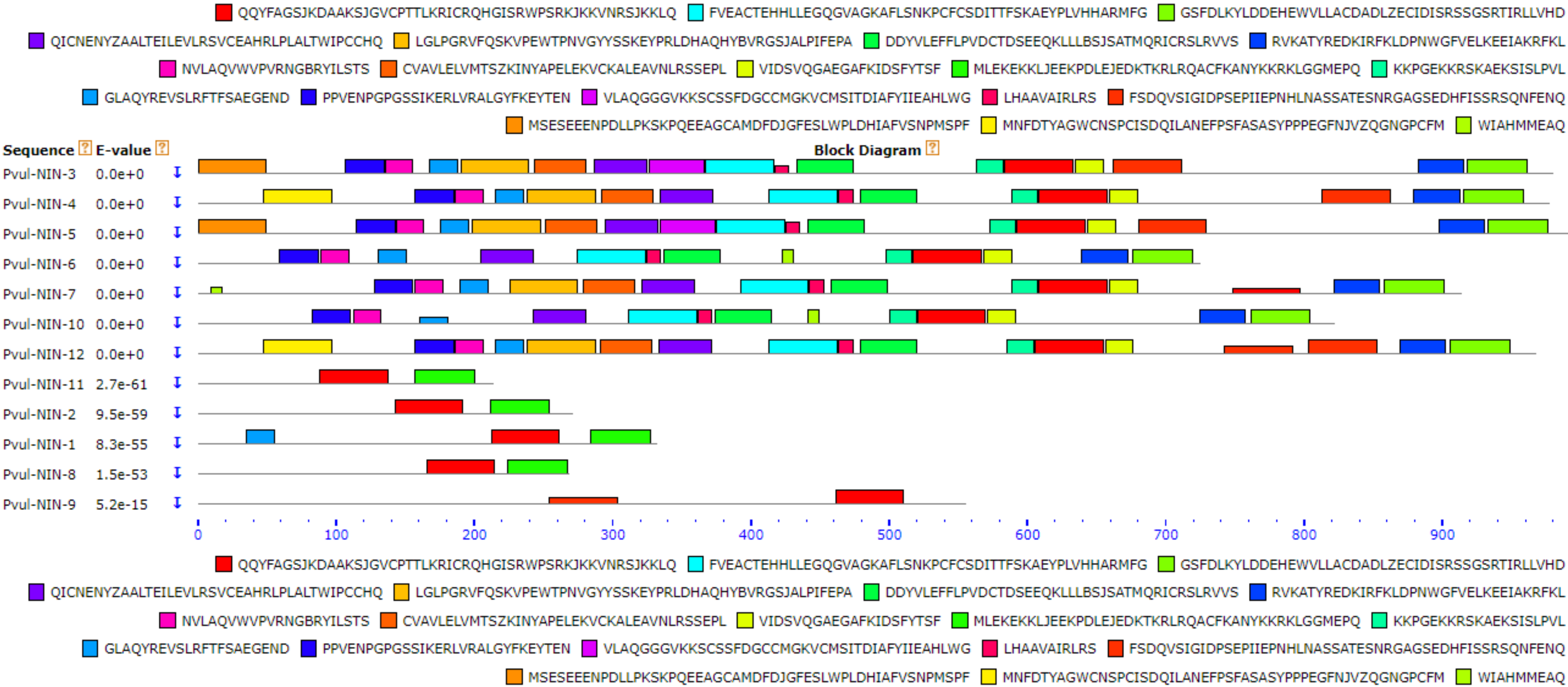
E-value: 1.0e-279 ? Site Count: 12 ? Width: 50 ?

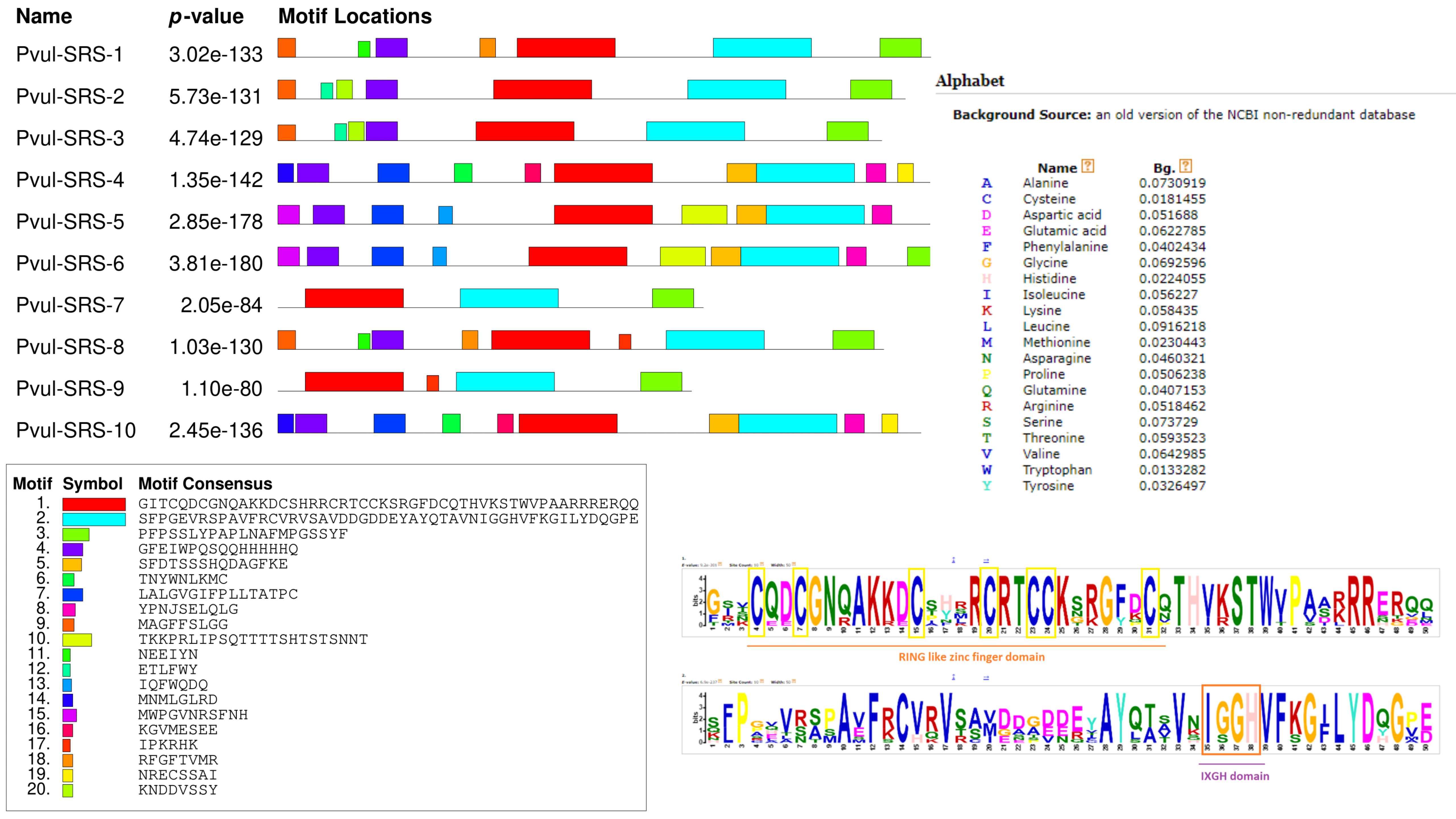
Log Likelihood Ratio: 1306 ? Information Content: 160.8 ? Relative Entropy: 157 ? Bayes Threshold: 10.373 ?

	Name ?	Start ?	*p*-value ?	Sites ?		
10.	Pvul-NIN-10	521	3.38e-60	AEKTISLPVL	RQYFAGSLKDAAKSIGVCPTTLKRICRQHGITRWPSRKIKKVGHSLRKLQ	LVIDSVQGAE
6.	Pvul-NIN-6	518	3.38e-60	AEKTISLPVL	RQYFAGSLKDAAKSIGVCPTTLKRICRQHGITRWPSRKIKKVGHSLKKLQ	LVIDSVQGAE
4.	Pvul-NIN-4	609	8.99e-58	SEKSVSLSVL	QQYFSGSLKDAAKNIGVCPTTLKRICRQHGISRWPSRKINKVNRSLKKIQ	TVLDSVQGVE
12.	Pvul-NIN-12	606	1.72e-57	VEKNVSLSVL	QQYFSGSLKDAAKSIGVCPTTLKRICRQHGISRWPSRKINKVNRSLKKIQ	TVLDSVQGVE
7.	Pvul-NIN-7	609	2.13e-57	AEKTITLQVL	RQYFAGSLKDAAKNIGVCTTTLKRICRQHGIKRWPSRKIKKVGHSLQKLQ	LVIDSVQGAS
3.	Pvul-NIN-3	584	5.27e-54	AEKSISLDVL	QHYFTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSLSKLK	CVIESVHGAE
5.	Pvul-NIN-5	593	7.24e-54	TEKSISLEVL	QRYFAGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKINKVNRSLSKLK	RVIESVQGAE
11.	Pvul-NIN-11	89	6.91e-44	SARMLSRKTV	SQYFYMPISQAAKELNVGLTHLKKRCRELGIQRWPHRKLMSLQTLIRNMQ	EQGQGEGPQN
2.	Pvul-NIN-2	143	3.83e-42	SSRMLSRKTI	SQYFYMPITQAARELNVGLTLLKKRCRELGIRRWPHRKLMSLQTLINNVQ	ELLKEEGPES
1.	Pvul-NIN-1	213	2.79e-40	LVAKISLSDL	VQYFGMPIVEASRNLKVGLTVLKRKCREFGIPRWPHRKIKSLDSLIQDLQ	EEAQNQELEN
8.	Pvul-NIN-8	166	1.16e-38	KPCALEFEEI	KKHFDVPINEAAKQMNVGLTMLKRRCRELNIMRWPHRKLKSLQLLIDNVK	ELGLAEEVSM
9.	Pvul-NIN-9	462	5.59e-28	RVPKMTMNDL	SPFFHLTIRDAADKLDVSDSVVKKISRLGNLKRWPQRKLQSLAKDVRVLR	KALNSPYEGT

	Logo	E-value ?	Sites ?	Width ?	More ?	Submit/Download ?
2.		2.4e-123	7	50		

	Logo	Name	Alt. Name	Width	1.	2.	3.	4.	5.	6.
1.		QQYFAGSJKDAAKSJGVCPTTLKRICRQHGISRWPSRKJKKVNRSJKKLQ	MEME-1	50	--	0.14	0.12	0.18	0.1	0.18
2.		FVEACTEHHLLEGQGVAGKAFLSNKPCFCSDITTFSKAEYPLVHHARMFG	MEME-2	50	0.14	--	0.14	0.18	0.33	0.17
3.		GSFDLKYLDDEHEWVLLACDADLZECIDISRSSGSRTIRLLVHD	MEME-3	44	0.12	0.14	--	0.18	0.16	0.17
4.		QICNENYZAALTEILEVLRSVCEAHRLPLALTWIPCCHQ	MEME-4	39	0.18	0.18	0.18	--	0.17	0.24
5.		LGLPGRVFQSKVPEWTPNVGYYSSKEYPRLDHAQHYBVRGSJALPIFEPA	MEME-5	50	0.1	0.33	0.16	0.17	--	0.16
6.		DDYVLEFFLPVDCTDSEEQKLLLBSJSATMQRICRSLRVVS	MEME-6	41	0.18	0.17	0.17	0.24	0.16	--
7.		RVKATYREDKIRFKLDPNWGFVELKEEIAKRFKL	MEME-7	34	0.18	0.16	0.19	0.16	0.18	0.24
8.		NVLAQVWVPVRNGBRYILSTS	MEME-8	21	0.21	0.19	0.19	0.35	0.25	0.26
9.		CVAVLELVMTSZKINYAPELEKVCKALEAVNLRSSEPL	MEME-9	38	0.15	0.22	0.19	0.21	0.16	0.2
10.		VIDSVQGAEGAFKIDSFYTSF	MEME-10	21	0.2	0.22	0.24	0.28	0.16	0.22
11.		MLEKEKKLJEEKPDLEJEDKTKRLRQACFKANYKKRKLGGMEPQ	MEME-11	44	0.2	0.15	0.16	0.18	0.16	0.15


QQYFAGSJKDAAKSJGVCPTTLKRICRQHGISRWPSRKJKKVNRSJKKLQ
FVEACTEHHLLEGQGVAGKAFLSNKPCFCSDITTFSKAEYPLVHHARMFG
GSFDLKYLDDEHEWVLLACDADLZECIDISRSSGSRTIRLLVHD
QICNENYZAALTEILEVLRSVCEAHRLPLALTWIPCCHQ
LGLPGRVFQSKVPEWTPNVGYYSSKEYPRLDHAQHYBVRGSJALPIFEPA
DDYVLEFFLPVDCTDSEEQKLLLBSJSATMQRICRSLRVVS
RVKATYREDKIRFKLDPNWGFVELKEEIAKRFKL
NVLAQVWVPVRNGBRYILSTS
CVAVLELVMTSZKINYAPELEKVCKALEAVNLRSSEPL
VIDSVQGAEGAFKIDSFYTSF
MLEKEKKLJEEKPDLEJEDKTKRLRQACFKANYKKRKLGGMEPQ
KKPGEKKRSKAEKSISLPVL
GLAQYREVSLRFTFSAEGEND
PPVENPGPGSSIKERLVRALGYFKEYTEN
VLAQGGGVKKSCSSFDGCCMGKVCMSITDIAFYIIEAHLWG
LHAAVAIRLRS
FSDQVSIGIDPSEPIIEPNHLNASSATESNRGAGSEDHFISSRSQNFENQ
MSESEEENPDLLPKSKPQEEAGCAMDFDJGFESLWPLDHIAFVSNPMSPF
MNFDTYAGWCNSPCISDQILANEFPSFASASYPPPEGFNJVZQGNGPCFM
WIAHMMEAQ
Sequence
E-value
Block Diagram
Pvul-NIN-3 0.0e+0
Pvul-NIN-4 0.0e+0
Pvul-NIN-5 0.0e+0
Pvul-NIN-6 0.0e+0
Pvul-NIN-7 0.0e+0
Pvul-NIN-10 0.0e+0
Pvul-NIN-12 0.0e+0
Pvul-NIN-11 2.7e-61
Pvul-NIN-2 9.5e-59
Pvul-NIN-1 8.3e-55
Pvul-NIN-8 1.5e-53
Pvul-NIN-9 5.2e-15
0
100
200
300
400
500
600
700
800
900



Name
p-value
Motif Locations
Pvul-SRS-1 3.02e-133
Pvul-SRS-2 5.73e-131
Pvul-SRS-3 4.74e-129
Pvul-SRS-4 1.35e-142
Pvul-SRS-5 2.85e-178
Pvul-SRS-6 3.81e-180
Pvul-SRS-7 2.05e-84
Pvul-SRS-8 1.03e-130
Pvul-SRS-9 1.10e-80
Pvul-SRS-10 2.45e-136
Alphabet
Background Source: an old version of the NCBI non-redundant database
Name Bg.
A Alanine 0.0730919
C Cysteine 0.0181455
D Aspartic acid 0.051688
E Glutamic acid 0.0622785
F Phenylalanine 0.0402434
G Glycine 0.0692596
H Histidine 0.0224055
I Isoleucine 0.056227
K Lysine 0.058435
L Leucine 0.0916218
M Methionine 0.0230443
N Asparagine 0.0460321
P Proline 0.0506238
Q Glutamine 0.0407153
R Arginine 0.0518462
S Serine 0.073729
T Threonine 0.0593523
V Valine 0.0642985
W Tryptophan 0.0133282
Y Tyrosine 0.0326497
Motif Symbol Motif Consensus
1. GITCQDCGNQAKKDCSHRRCRTCCKSRGFDCQTHVKSTWVPAARRRERQQ
2. SFPGEVRSPAVFRCVRVSAVDDGDDEYAYQTAVNIGGHVFKGILYDQGPE
3. PFPSSLYPAPLNAFMPGSSYF
4. GFEIWPQSQQHHHHHQ
5. SFDTSSSHQDAGFKE
6. TNYWNLKMC
7. LALGVGIFPLLTATPC
8. YPNJSELQLG
9. MAGFFSLGG
10. TKKPRLIPSQTTTTSHTSTSNNT
11. NEEIYN
12. ETLFWY
13. IQFWQDQ
14. MNMLGLRD
15. MWPGVNRSFNH
16. KGVMESEE
17. IPKRHK
18. RFGFTVMR
19. NRECSSAI
20. KNDDVSSY
RING like zinc finger domain
IXGH domain