



【蛋白表达初级篇】 序列分析与表达片段选择

From Gene To Antibody

普健生物（武汉）科技有限公司

AtaGenix Laboratories





蛋白重组表达

XtenCHO高密度表达系统

昆虫杆状病毒表达系统

稳定细胞株构建

抗体定制服务

兔单克隆抗体制备

纳米抗体制备

抗体对开发

重组抗体表达

嵌合抗体生产

抗体片段生产

大规模重组抗体生产

抗体药物开发

人源化抗体

双特异性抗体

抗独特型抗体

普健生物（武汉）科技有限公司

始于2012，一站式抗体发现整体解决方案

3551光谷人才计划

——2013年

高新技术企业

——2017年

光谷瞪羚企业

——2018年

武汉国家生物产业基地
抗体发现与筛选公共服务平台

——2022年

目录

| CONTENTS |

-01-

蛋白序列分析



-02-

表达片段选择





BLAST: Basic Local Alignment Search Tool

<https://blast.ncbi.nlm.nih.gov>

网页 2023年1月9日 · The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to ...

Blast

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity ...

Protein BLAST

PSI-BLAST allows the user to build a PSSM (position-specific scoring matrix) using ...

Home

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity ...

Recent Results

This page provides links to your unexpired BLAST jobs. Enter a BLAST RID in the ...

Saved Strategies

clicking the 'Saved S ...
the top of a BLAST r

Help

Welcome to BLAST ...
Quick start guide. G

NCBI Homep

The .gov means it's ...
government website

PubMed

PubMed® comprises ...
citations for biomedic

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein

Enter Query Sequence

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

氨基酸序列

Query subrange [?](#)

From

To

Or, upload file [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases Non-redundant protein sequences (nr) [?](#)

- RefSeq Select proteins (refseq_select)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot (swissprot)
- Patented protein sequences (pataa)
- Protein Data Bank proteins (pdb)
- Metagenomic proteins (env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

数据库选择

Database [?](#)

Organism **物种选择** exclude

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

[Try experimental clustered nr](#)
For more info see [What is clustered nr](#)



UniProt

<https://www.uniprot.org>

网页 2023年2月8日 · At UniProt we are very interested in engaging with the machine learning community. Metrics, metrics, metrics: Measuring the Impact of UniProt. UniProt is brought ...

Align

Align - UniProt

BLAST

BLAST - UniProt

Peptide search

Find UniProt entries through parts of their peptide sequences, each more than two ...

ID mapping

ID mapping - UniProt

SPARQL

About. This SPARQL endpoint contains all UniProt data. It is free to access and ...

Tools Dashboard: BLAST

Cite UniProt; About & manual; Technical co... Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001). UniProt IDs

OR Tp53 in UniProt

Enter one or more sequences (20 max). You may also [load from a text file](#).

Protein or nucleotide sequence(s) in FASTA format.

氨基酸序列

Target database

UniProtKB reference proteomes + Swiss-Prot

Restrict by taxonomy

Enter taxon names or IDs to include

Name your BLAST job

"my job title"

Advanced parameters

Sequence type

Protein

Program

blastp

E-Threshold

10

Matrix

Auto - BLOSUM62

Filter

None

Gapped

yes

Hits

250

HSPs per hit

All

[Descriptions](#) | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

Sequences producing significant alignments

[Download](#) | [Select columns](#) | Show

select all *100 sequences selected*

[GenPept](#) | [Graphics](#) | [Distance tree of results](#) | [Multiple alignment](#) | [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 isoform X1 [Cucumis sativus]	Cucumis sativus	1320	1320	100%	0.0	100.00%	630	XP_004144109.2
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 [Cucumis sativus]	Cucumis sativus	1307	1307	100%	0.0	98.90%	635	NP_001295791.1
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 isoform X1 [Cucumis melo]	Cucumis melo	1281	1281	100%	0.0	97.30%	626	XP_008451093.2
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 [Cucumis melo var. makuwa]	Cucumis melo var. makuwa	1278	1278	100%	0.0	97.14%	626	KAA0055660.1
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 [Benincasa hispida]	Benincasa hispida	1179	1179	100%	0.0	94.44%	626	XP_038878611.1
<input checked="" type="checkbox"/>	protein ETHYLENE INSENSITIVE 3 [Cucumis melo]	Cucumis melo	1160	1160	97%	0.0	94.13%	693	NP_001315385.1

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	Score	Ident
<input type="checkbox"/> A0A0A0LXB4	A0A0A0LXB4_CUCSA	EIN3-like protein	Csa_1G601520	Cucumis sativus (Cucumber)	630 AA	100%	3390
<input type="checkbox"/> I3VKD3	I3VKD3_CUCSA	Ethylene-insensitive 3		Cucumis sativus (Cucumber)	635 AA	98.9%	3360
<input type="checkbox"/> A0A5D3CEG1	A0A5D3CEG1_CUCME	Protein ETHYLENE INSENSITIVE 3	E5676_scaffold16G00310	Cucumis melo var. makuwa	626 AA	97.3%	3286
<input type="checkbox"/> A0A5A7UKJ9	A0A5A7UKJ9_CUCME	Protein ETHYLENE INSENSITIVE 3	E6C27_scaffold181G00280	Cucumis melo var. makuwa	626 AA	97.1%	3278
<input type="checkbox"/> A0A1S3BQ52	A0A1S3BQ52_CUCME	protein ETHYLENE INSENSITIVE 3	LOC103492471	Cucumis melo (Muskmelon)	625 AA	97%	3276



Function

Names & Taxonomy

Subcellular Location

Phenotypes & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar Proteins



Proteinⁱ | EIN3-like protein

Statusⁱ | UniProtKB unreviewed (TrEMBL)

Organismⁱ |

Amino acids | 630

Protein existenceⁱ | Inferred from homology

Annotation scoreⁱ | 1/5

[Entry](#)

[Feature viewer](#)

[Publications](#)

[External links](#)

[History](#)

BLAST [Download](#) [Add](#) [Add a publication](#) [Entry feedback](#)

Functionⁱ

GO annotationsⁱ

Slimming set:

plant



一级结构分析

[ExpASy - ProtParam tool](#)

查看此网页的中文翻译, 请点击 [翻译此页](#)

[ProtParam tool](#) ProtParam(References/Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored i...

web.expasy.org/protparam/

Number of amino acids: 630

Molecular weight: 71859.94

Theoretical pI: 5.51

Amino acid composition:

Ala (A)	27	4.3%
Arg (R)	32	5.1%
Asn (N)	42	6.7%
Asp (D)	50	7.9%
Cys (C)	10	1.6%
Gln (Q)	34	5.4%
Glu (E)	39	6.2%
Gly (G)	35	5.6%
His (H)	15	2.4%
Ile (I)	28	4.4%
Leu (L)	48	7.6%
Lys (K)	41	6.5%
Met (M)	26	4.1%
Phe (F)	27	4.3%
Pro (P)	46	7.3%
Ser (S)	48	7.6%
Thr (T)	27	4.3%
Trp (W)	10	1.6%
Tyr (Y)	15	2.4%
Val (V)	30	4.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient 77975
 Abs 0.1% (=1 g/l) 1.085, assuming all pairs of Cys residues form cystines

Ext. coefficient 77350
 Abs 0.1% (=1 g/l) 1.076, assuming all Cys residues are reduced



一级结构分析

TMHMM - 2.0 - Services - DTU Health Tech

TMHMM - 2.0 Prediction of transmembrane helices in proteins NOTE: TMHMM-2.0 is a more recent and better transmembrane predictor, DeepTMHMM, has been released.

[services.healthtech.dtu.dk/ser...](https://services.healthtech.dtu.dk/services)

SignalP-5.0 - redirect - DTU

www.cbs.dtu.dk/services/SignalP

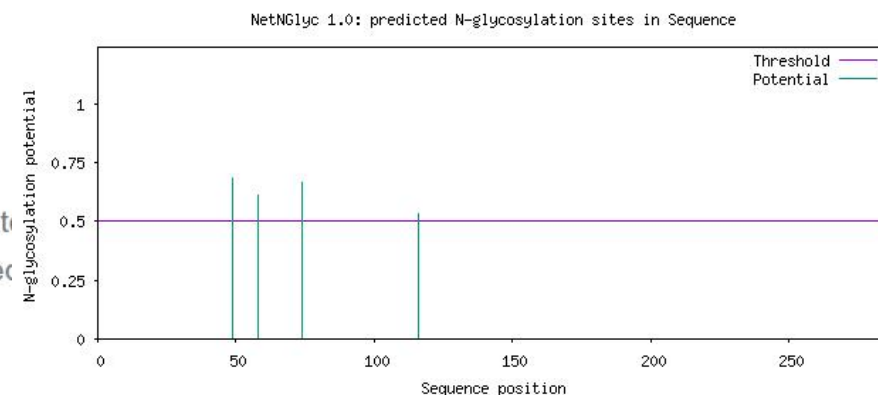
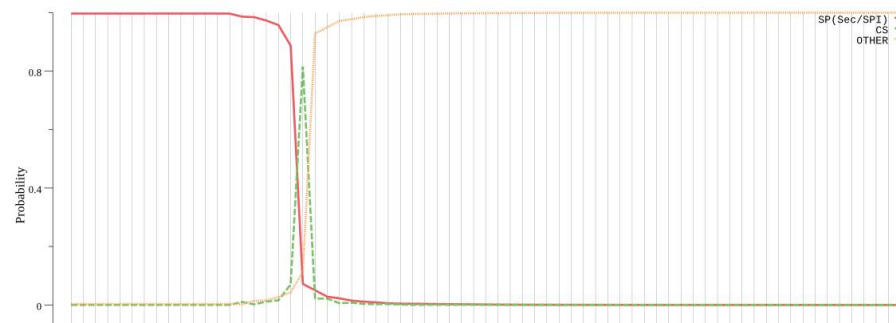
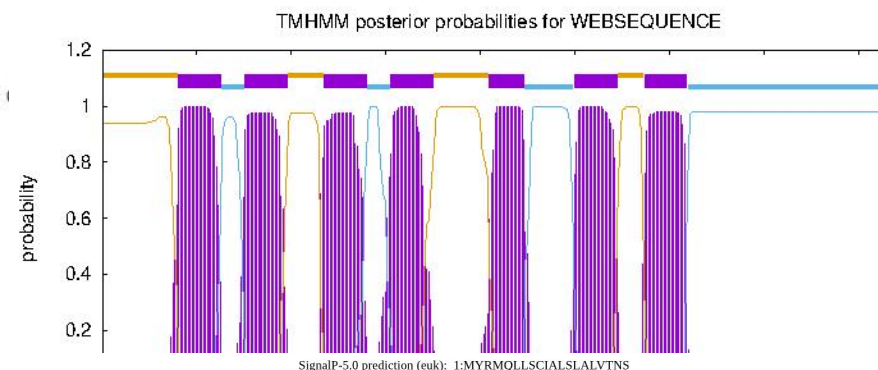
网页 2021年9月22日 · New home of SignalP-5.0 is:

<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>

NetNGlyc-1.0 - redirect - DTU

www.cbs.dtu.dk/services/NetNGlyc

网页 2021年9月22日 · N-glycosylation site prediction in human protein networks that examine the sequence context of Asn-Xaa-Ser/Thr sequences.





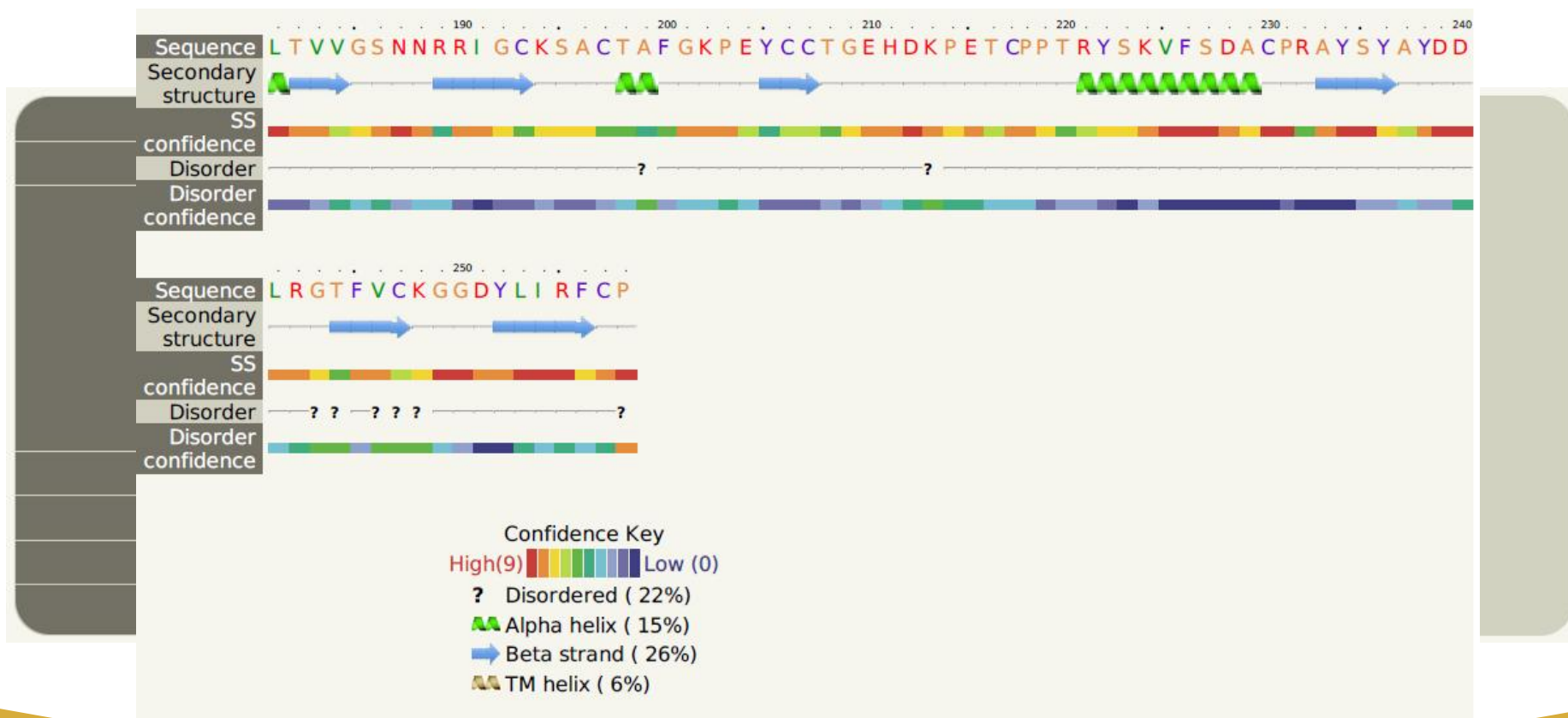
二级结构分析

PHYRE2 Protein Fold Recognition Server

查看此网页的中文翻译, 请点击 [翻译此页](#)



Current **Phyre2** server load =101%(expect slow processing) E-mail Address Optional Job description Amino Acid Sequence Or try the sequence finder Modelling ModeNormalIntensiveTest Plea...
www.sbg.bio.ic.ac.uk/~phyre2/h...





三级结构分析



<http://smart.embl.de/>



<http://www.ebi.ac.uk/interpro/search/sequence/>




<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

Conserved domains on [lcl|seqsig_MMMMF_7aa87fa235dc8ba3374d198280d9502f] View ?

Local query sequence

Graphical summary Zoom to residue level [show extra options »](#)



Query seq. Superfamilies

[Search for similar domain architectures](#) ? [Refine search](#) ?

List of domain hits

Name	Accession	Description	Interval	E-value
EIN3 super family	cl04813	Ethylene insensitive 3; Ethylene insensitive 3 (EIN3) proteins are a family of plant DNA-binding proteins that regulate transcription in response to the gaseous plant hormone ethylene, and are essential for ethylene-mediated responses including the triple response, cell growth inhibition, and accelerated senescence.	14-385	2.79e-178



三级结构分析

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein

Enter Query Sequence

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

From

To

Or, upload file [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases Standard databases (nr etc.): New Experimental databases [< Try experimental clustered nr](#)
For more info see [What is clustered nr](#)

Compare Select to compare standard and experimental database [?](#)

Standard

Database [?](#)

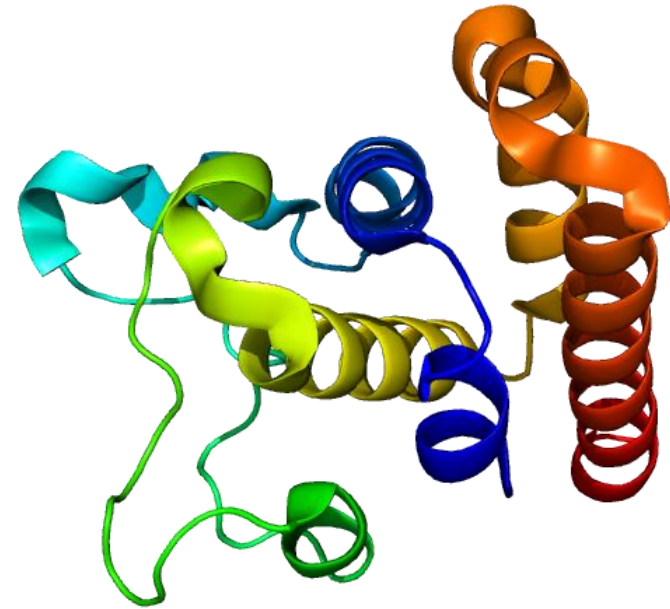
Organism Optional exclude

Exclude Optional taxa will be shown [?](#)

(WP) Uncultured/environmental sample sequences

Program Selection

- Non-redundant protein sequences (nr)
- RefSeq Select proteins (refseq_select)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot (swissprot)
- Patented protein sequences (pataa)
- Protein Data Bank proteins (pdb)**
- Metagenomic proteins (env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)



PDB: 4ZDS chain A

Sequence identity: 94%

实验测定得到的空间结构



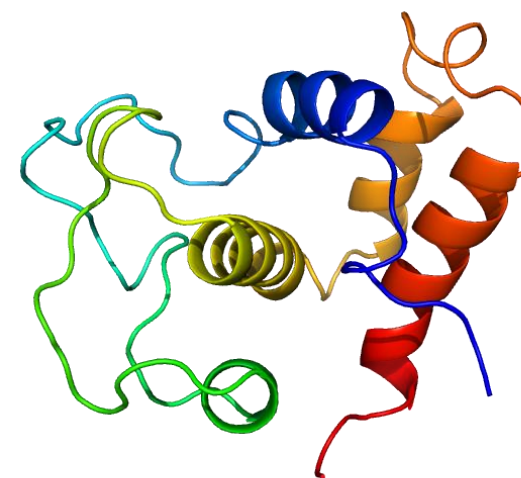
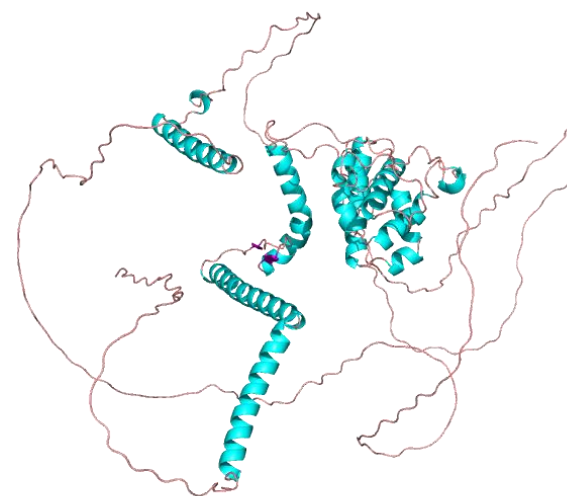
三级结构分析



<https://alphafold.ebi.ac.uk/>



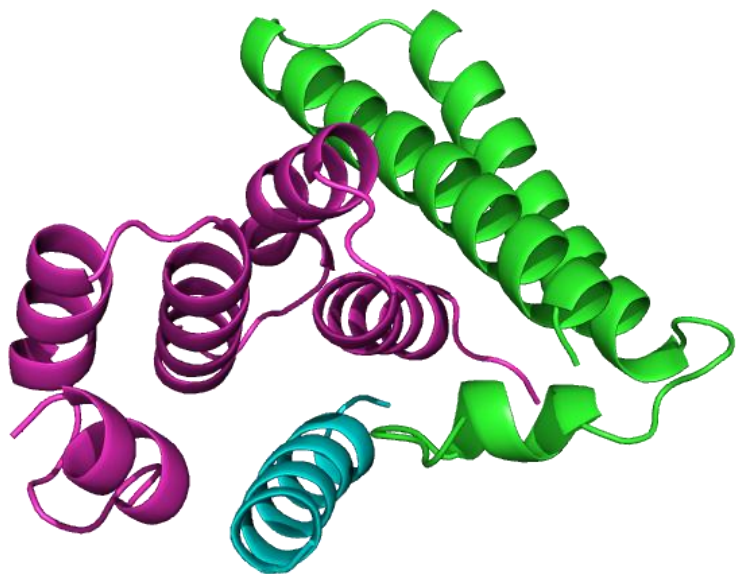
<https://swissmodel.expasy.org/interactive>



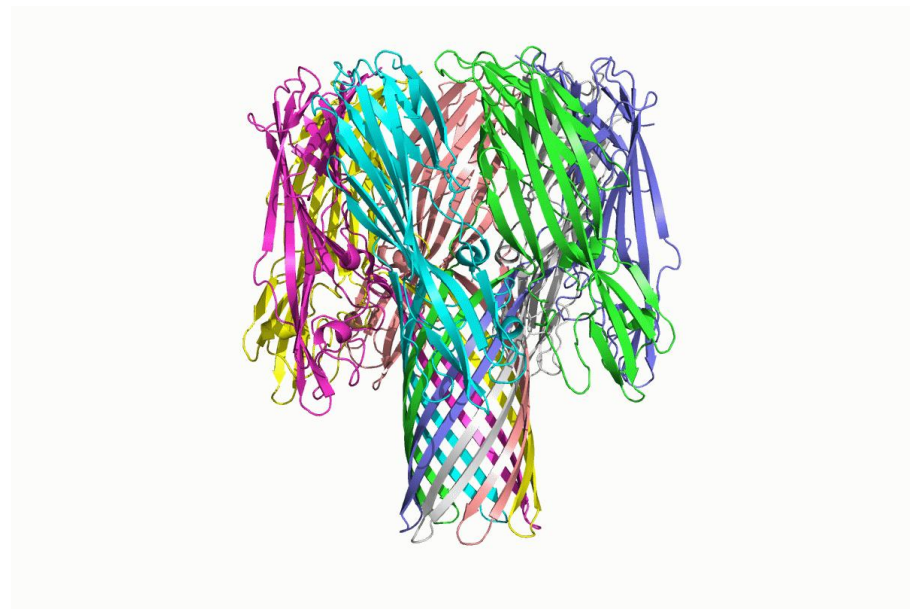


四级结构分析

蛋白质的四级结构是指蛋白质的多条肽链之间相互作用所形成的更为复杂聚合物的一种结构形式



异源三聚体



同源八聚体

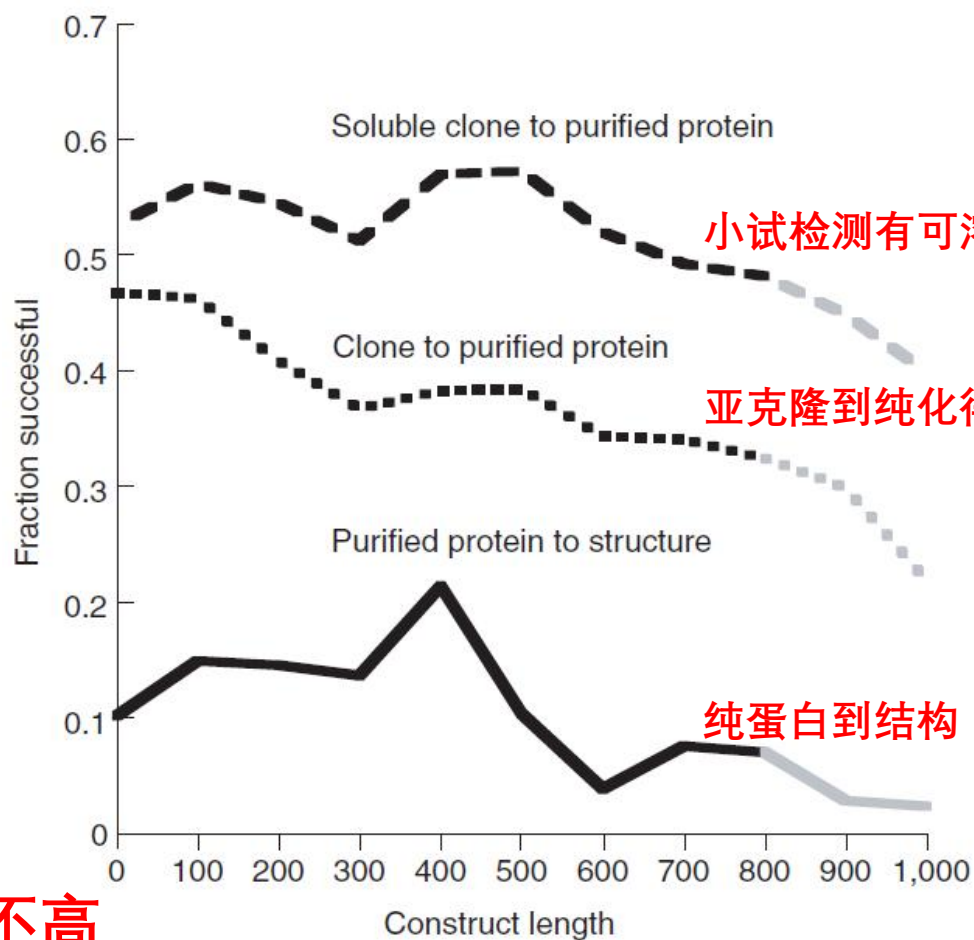


Protein production and purification

Structural Genomics Consortium¹⁻³, Architecture et Fonction des Macromolécules Biologiques⁴, Berkeley Structural Genomics Center⁵, China Structural Genomics Consortium^{6,7}, Integrated Center for Structure and Function Innovation⁸, Israel Structural Proteomics Center⁹, Joint Center for Structural Genomics^{10,11}, Midwest Center for Structural Genomics¹², New York Structural GenomiX Research Center for Structural Genomics¹³⁻¹⁷, Northeast Structural Genomics Consortium^{18,19}, Oxford Protein Production Facility²⁰, Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine²¹, RIKEN Structural Genomics/ Proteomics Initiative²² & SPINE2-Complexes^{23,25}



表达片段选择



小试检测有可溶蛋白到放大纯化得到

亚克隆到纯化得到蛋白

纯蛋白到结构

蛋白表达成功率不高

100-500个氨基酸成功率相对更高



表达片段选择

以蛋白应用为导向

- 抗原 兼顾可表达性和免疫原性
- 酶活 酶活结构域
- 结晶 有感兴趣活性的前提下越短越好
- 功能 结构决定功能，保留相关结构域
- ...



表达片段选择

全长蛋白表达注意事项：

去除原始信号肽或者更好适合表达系统的信号肽

截短蛋白表达注意事项：

- 去除原始信号肽或者更好适合表达系统的信号肽
- 不破坏保守结构域
- 不破坏蛋白球状结构
- 不影响二硫键形成
- 不影响四级结构形成
- 不包含长无序区
- 以参考文献为主，综合考虑预测结果

谢谢观看！

网址：www.atagenix.cn

邮箱：Sales@atagenix.com

武汉市东湖新技术开发区神墩四路666号C栋

