生物信息学

Bioinformatics

第八章 结构分析

生物信息学分析

- 根据研究对象及目的分类：

1. macromolecular sequences;
2. macromolecular structures;
3. expression profiles; (microarrays;RNA-seq;2D-PAGE)
4. biochemical network; (Interactions and reactions)
5. evolution history.

生物信息学分析

章节	源数据	结果知识	种类
四、序列分析 *	DNA序列	基因等特征序列	Seq.
	蛋白质序列	特征域、特性	
	EST	表达基因（mRNA）	Expr.
五、系统发育分析	DNA/蛋白质序列	进化历史	Evol.
六、基因组分析	基因组序列	基因位置、功能、物种进化历史	Seq. Evol.
（转录组分析）	Microarray	表达基因（mRNA）	Expr.
	RNA-seq		
七、蛋白质组分析	2D-Page	表达基因（蛋白质）	Expr.
	Y2-hybrid ...	蛋白质相互作用…	Net.
八、结构分析	蛋白质序列	蛋白质结构	Struct.
	RNA序列	RNA结构	

蛋白质结构分析

- Basic notes
- Structural alignment of proteins
- Structure-based protein classification
- Structure prediction of proteins (Comparative Modeling*)

序列通过折叠为结构而最终形成功能是使生命得以实现的重要自然原理。

生命的信息内涵

$Log_2 2 = 1$ bit

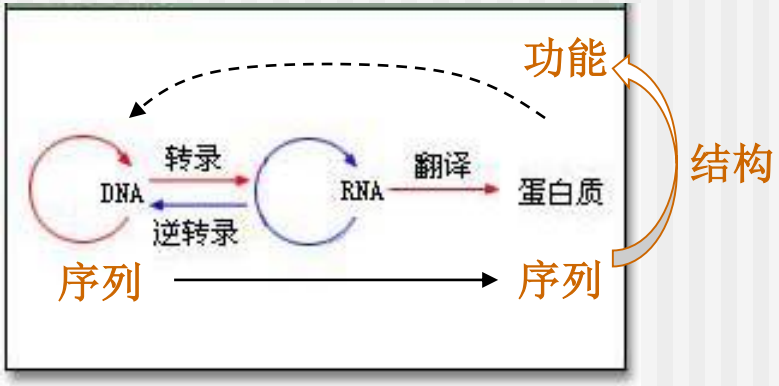
- 信息的产生：多种可能状态下进行选择。

——生物信息的产生源于自然选择。

ATCGC＝？ bits

- 信息的“行为方式”——传递：一种选择引起另一种选择；功能实现。

中心法则：



——生物学研究的就是生物信息的产生和传递方式。

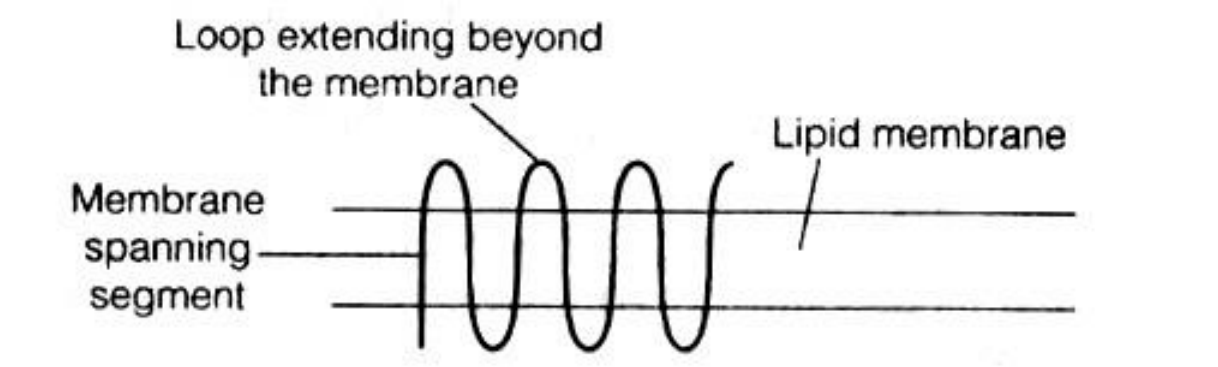
Structural types

- Globular proteins

  核心疏水，表面亲水。

- Integral membrane proteins

  膜内疏水，膜表面亲水。



结构的局部特征 —— Domain

- 序列特征域：

  A domain (or module) is a protein region that adopt a particular three-dimensional structure.

  A motif (or fingerprint, or pattern) is a short conserved protein region, typically 10~20 aa.

Find domains by Blastp

NCBI Conserved Domain Search - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/Structure/cdd/

NCBI Conserved Domain Search

NCBI

Conserved Domains

SH3 SH2

HOME | SEARCH | GUIDE | NewSearch | Structure Home | 3D Macromolecular Structures | Conserved Domains | Pubchem | BioSystem

Conserved domains on [gi|4826830|ref|NP_004983|]

View concise result

methyl-CpG-binding protein 2 isoform 1 [Homo sapiens]

Graphical summary

show options »

Query seq. 1 75 150 225 300 375 450 486

DNA binding site

Specific hits: MeCP2_MBD

Non-specific hits: MBD, MBD, MBD

Superfamilies: MBD superfamily

Search for similar d

cd01396

[Specific hit] cd01396, MeCP2, MBD1, MBD2, MBD3, and MBD4 are members of a protein family that share the methyl-CpG-binding domain (MBD). The MBD, consists of about 70 residues and is defined as the minimal region required for binding to methylated DNA by a methyl-CpG-binding protein which binds specifically to methylated DNA. The MBD can recognize a single symmetrically methylated CpG either as naked DNA or within chromatin. MeCP2, MBD1 and MBD2 (and likely MBD3) form complexes with histone deacetylase and are involved in histone deacetylase-dependent repression of transcription. MBD4 is an endonuclease that forms a

List of domain hits

Desc	PssmId	Multi-dom	E-value
[+]MeCP2_MBD[cd01396], MeCP2, MBD1, MBD2, MBD3, and MBD4 are m	29025	no	3.78e-17
[+]MBD[cd00122], MeCP2, MBD1, MBD2, MBD3, MBD4, CLLD8-like, and B	29023	no	1.79e-13
[+]MBD[smart00391], Methyl-CpG binding domain; Methyl-CpG binding do	128673	no	6.19e-23
[+]MBD[pfam01429], Methyl-CpG binding domain; The Methyl-CpG bindin ... etrically ...	144867	no	1.07e-20

Blast search parameters

Data Source: Live blast search RID = UYY6YB39015

User Options: Database: cdsearch/cdd v2.28 Low complexity filter: yes E-value threshold: 0.01 Maximum number of hits: 500

http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?ascbin=8&maxaln=10&seltype=2&uid=29025&querygi=48

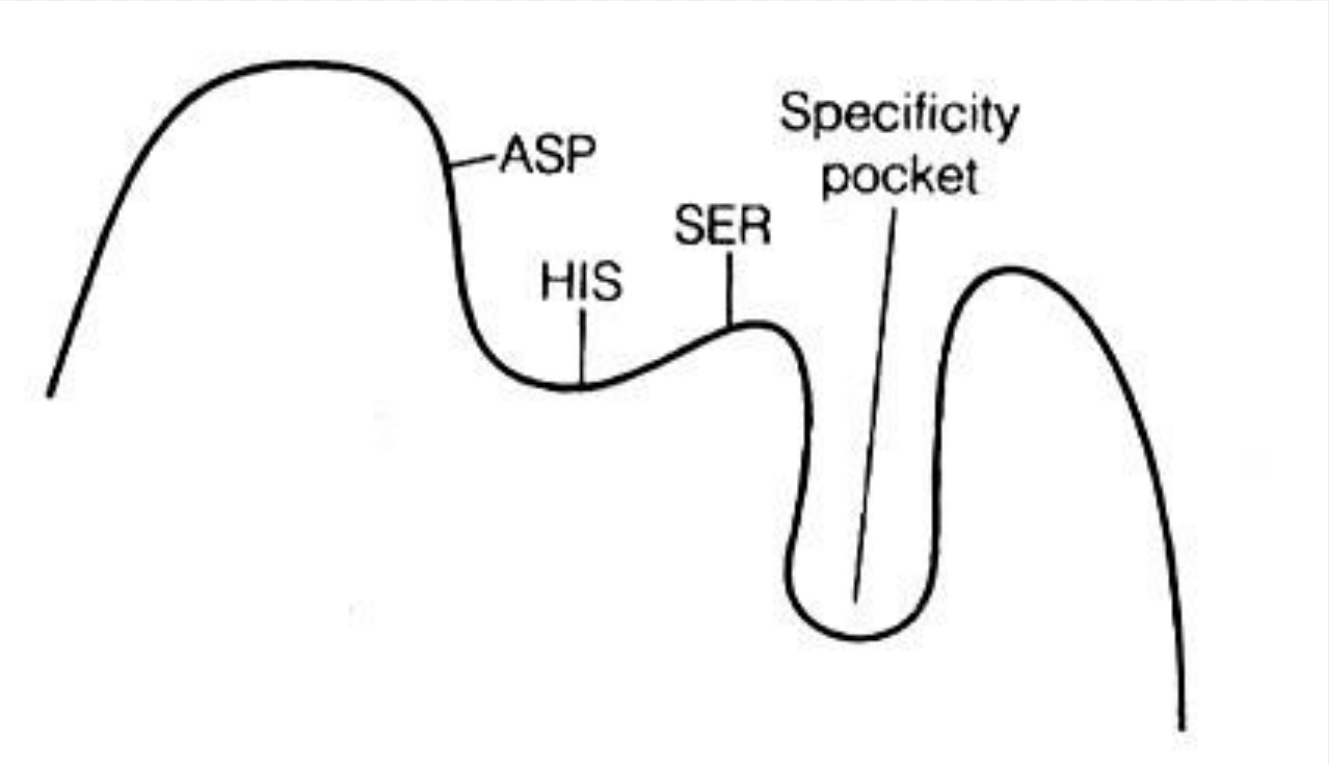
Internet 100%

开始 生物信息学2011 Microsoft Pow... NCBI Blast:re... NCBI Conserve... 8:27

What is a domain?

- 序列：A common subsequence observed to occur in many different proteins;
- 结构：A subsequence folding independently; a geometrically distinct substructure.
- 功能：A subsequence or substructure with a recognized function;

从结构到功能：



催化功能——形状＋特定位置的特定aa

蛋白质结构源数据库---PDB

http://www.rcsb.org/pdb/explore/explore.do?

收藏夹 | RCSB Protein Data Bank - RCSB PDB -...

RCSB PDB PROTEIN DATA BANK → RCSB PDB-101

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday May 29, 2012 at 5 PM PDT there are 81957 Structures | PDB Statistics

All Categories | Author | Macromolecule | Sequence | Ligand

Search | All Categories: e.g., PDB ID, molecule name, author

Browse | Advanced

MyPDB Hide

- Login to your Account
- Register a New Account
- Query Results (4)
- Query History (1)

Home Hide

- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- Sitemap
- New Website Features

Deposition Hide

- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamlines/Facilities
- Related Tools

Tools Hide

- Download Files
- Compare Structures
- File Formats
- Services: RESTful | SOAP
- Widgets

Summary | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Structure of Herring Type II Antifreeze Protein

2PY2

Display Files
Download Files
Share this Page

DOI:10.2210/pdb2py2/pdb

Primary Citation

Structure and evolutionary origin of Ca(2+)-dependent herring type II antifreeze protein.

Liu, Y., Li, Z., Lin, Q., Kosinski, J., Seetharaman, J., Bujnicki, J.M., Sivaraman, J., Hew, C.L.,

Journal: (2007) PLoS ONE **2:** e548-e548

PubMed: 17579720
PubMedCentral: PMC1891086
DOI: 10.1371/journal.pone.0000548
Search Related Articles in PubMed

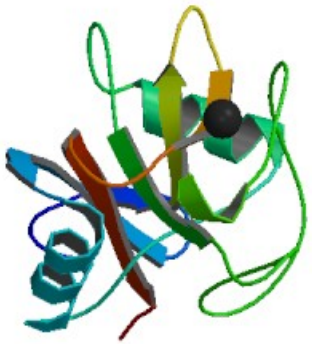
PubMed Abstract:

In order to survive under extremely cold environments, many organisms produce antifreeze proteins (AFPs). AFPs inhibit the growth of ice crystals and protect organisms from freezing damage. Fish AFPs can be classified into five distinct types based on their structures.... [Read More & Search PubMed Abstracts]

Molecular Description Hide

Classification: Antifreeze Protein
Structure Weight: 92991.48

Biological Assembly 1



More Images...

View in Jmol

Simple Viewer | Kiosk
Protein Workshop

Biological assembly 1 assigned by authors

Internet | 100%

PDB --- 结构数据

```
SITE     2 AC5  6 ASP E 114  HOH E 932
SITE     1 AC6  6 GLN F  92  ASP F  94
SITE     2 AC6  6 ASP F 114  HOH F 907
CRYST1   31.279  146.415  192.406  90.00  90.00  90.00 P 21 21 21    24
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.031970  0.000000  0.000000        0.00000
SCALE2      0.000000  0.006830  0.000000        0.00000
SCALE3      0.000000  0.000000  0.005197        0.00000
ATOM      1  N   CYS A   4      41.984  34.341  17.654  1.00 37.88           N
ATOM      2  CA  CYS A   4      41.522  34.320  19.073  1.00 38.27           C
ATOM      3  C   CYS A   4      40.007  34.287  19.172  1.00 38.11           C
ATOM      4  O   CYS A   4      39.306  34.731  18.264  1.00 38.43           O
ATOM      5  CB  CYS A   4      42.020  35.559  19.832  1.00 38.18           C
ATOM      6  SG  CYS A   4      43.805  35.582  20.183  1.00 40.36           S
ATOM      7  N   PRO A   5      39.480  33.755  20.284  1.00 37.81           N
ATOM      8  CA  PRO A   5      38.029  33.697  20.460  1.00 37.63           C
ATOM      9  C   PRO A   5      37.464  35.112  20.566  1.00 37.95           C
ATOM     10  O   PRO A   5      38.137  36.030  21.036  1.00 36.62           O
ATOM     11  CB  PRO A   5      37.868  32.890  21.749  1.00 37.89           C
ATOM     12  CG  PRO A   5      39.137  33.172  22.495  1.00 37.38           C
ATOM     13  CD  PRO A   5      40.172  33.094  21.404  1.00 37.61           C
ATOM     14  N   THR A   6      36.225  35.267  20.117  1.00 38.06           N
ATOM     15  CA  THR A   6      35.519  36.541  20.104  1.00 38.79           C
ATOM     16  C   THR A   6      36.154  37.767  20.777  1.00 38.89           C
ATOM     17  O   THR A   6      36.983  38.447  20.178  1.00 41.01           O
```

PDB --- 结构视图

http://www.rcsb.org/pdb/explore/jmol.do?structureId=2PY2&bior

收藏夹 RCSB PDB - Jmol Viewer for 2PY2

Summary | Sequence | Annotations | Seq. Similarity

MyPDB Hide

Login to your Account
Register a New Account
Query Results (4)
Query History (1)

Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools

Tools Hide

Download Files
Compare Structures
File Formats
Services: RESTful | SOAP
Widgets

PDB-101 Hide

Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources

Help Hide

Launch Help System
Display Settings

Structure of Herring Type II Antifreeze Protein

2PY2

Display Files
Download Files
Share this Page

Jmol Version 12.2.15

*Right-click Jmol to view additional options. Drag the bottom-right corner to resize.

Internet 100%

蛋白质结构分析

- Basic notes
- Structural alignment of proteins
- Structure-based protein classification
- Structure prediction of proteins (Comparative Modeling*)

结构直接支持功能

- 结构的保守性可能超出序列比较的可识别性 ——

  Protein structures tend to be conserved even when evolution has changed the sequence beyond recognition.

序列相似性分析的意义

- 进化关系推测

  序列相似一般是由于进化同源，也有例外。

- 结构推测
- 功能推测

序列及结构相似性比对的意义

- **序列比对：源序列与目标序列之间按残基位置相对排列。使序列之间的相似程度最大。**

  进化（同源）<==寻找序列相似物==>结构---------功能。

- **结构比对：源结构和目标结构之间按残基位置相对排列，使结构之间相似性程度最大——相应残基的α碳原子空间位置最接近。**

  进化（同源）<==========寻找结构相似物==>功能。

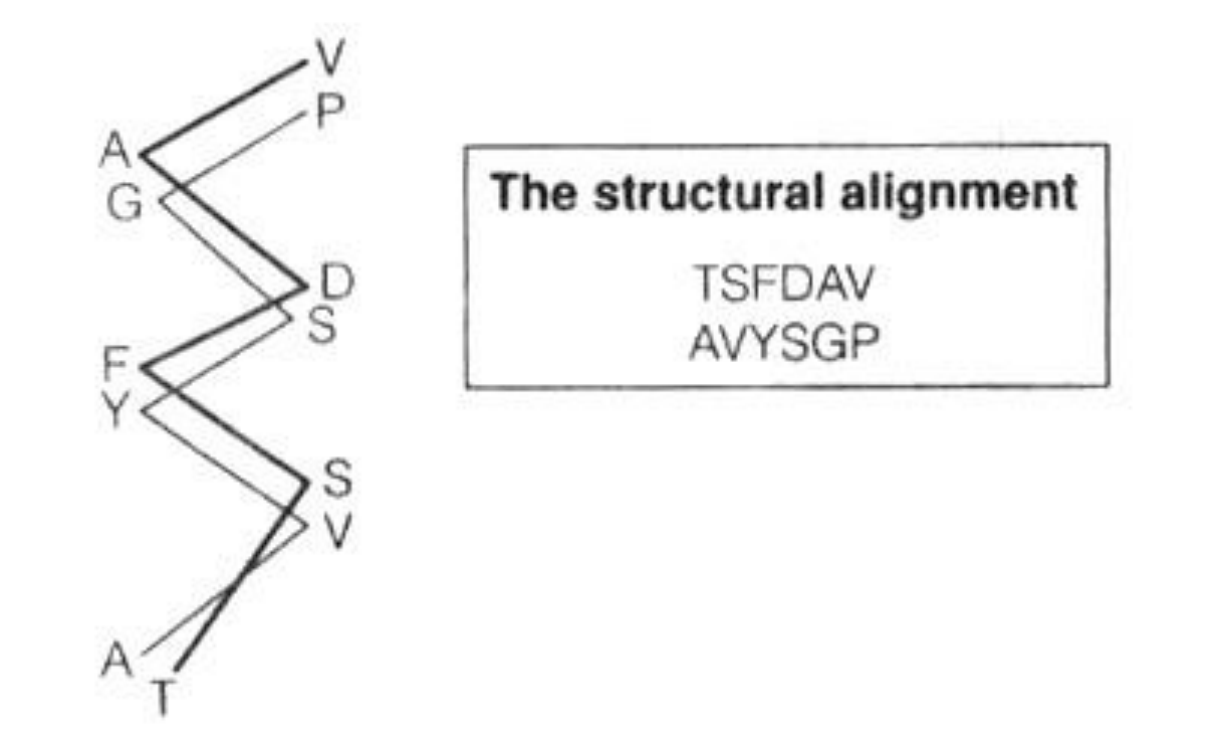
生物信息学分析

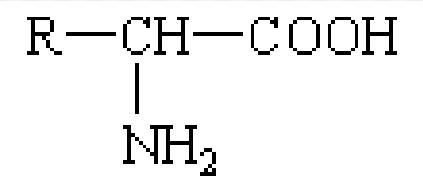
章节	源数据	结果知识	种类
四、序列分析 *	DNA序列	基因等特征序列	Seq.
	蛋白质序列	特征域、特性	
	EST	表达基因（mRNA）	Expr.
五、系统发育分析	DNA/蛋白质序列	进化历史	Evol.
六、基因组分析	基因组序列	基因位置、功能、物种进化历史	Seq. Evol.
（转录组分析）	Microarray	表达基因（mRNA）	Expr.
	RNA-seq		
七、蛋白质组分析	2D-Page	表达基因（蛋白质）	Expr.
	Y2-hybrid ...	蛋白质相互作用…	Net.
八、结构分析	蛋白质序列	蛋白质结构	Struct.
	RNA序列	RNA结构	

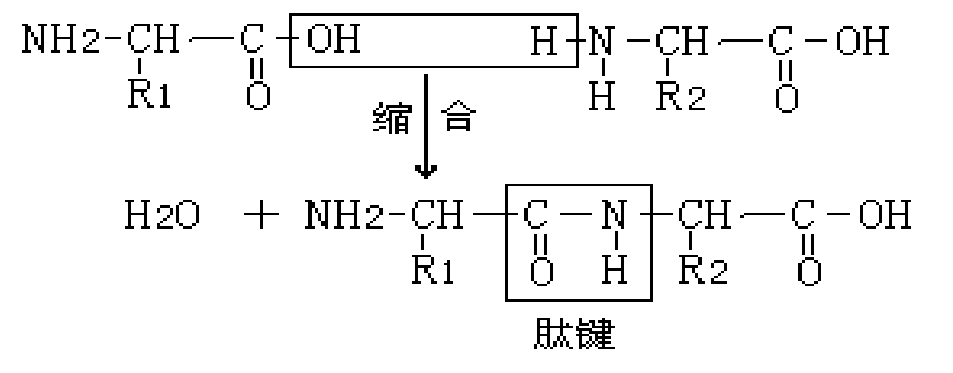
生物信息学分析

章节	源数据	结果知识	种类
四、序列分析 *	DNA序列	基因等特征序列	Seq.
	蛋白质序列	特征域、特性	
	EST	表达基因（mRNA）	Expr.
五、系统发育分析	DNA/RNA/ 蛋白质序列	进化历史	Evol.
	蛋白质结构	进化历史	….
			….
			….
	蛋白质结构	蛋白质功能	….
八、结构分析	蛋白质序列	蛋白质结构	Struct.
	RNA序列	RNA结构	

A conceptual view of structural alignment

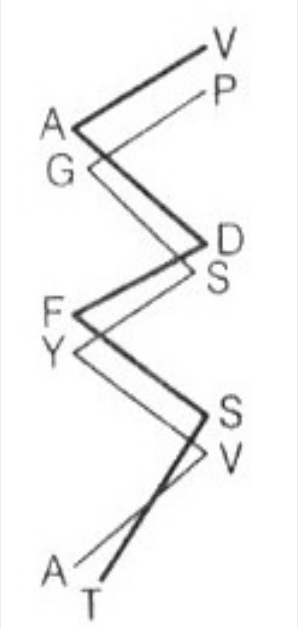






Root Mean Square Deviation (RMSD)

- 相应残基α碳原子平均空间距离的度量。



$$RMSD = \sqrt{\frac{1}{N}\sum_i d_i^2}$$

结构相似性衡量的重要指标

RMSD<1.5 埃 —— 非常相似

NCBI — Domain & Structure

http://www.ncbi.nlm.nih.gov/guide/domains-structures/#tools

Domains & Structures - Site Guide -...

NCBI Resources How To My NCBI Sign In

NCBI
National Center for Biotechnology Information

All Databases | Search

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis

Domains & Structures

All | Databases | Downloads | Tools | How To

Tools

CDTree
A stand-alone application for classifying protein sequences and investigating their evolutionary relationships. CDTree can import, analyze and update existing Conserved Domain (CDD) records and hierarchies, and also allows users to create their own. CDTree is tightly integrated with Entrez CDD and Cn3D, and allows users to create and update protein domain alignments.

Cn3D
A stand-alone application for viewing 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and UNIX and can be configured to receive data from most popular web browsers. Cn3D simultaneously displays structure, sequence, and alignment, and has powerful annotation and alignment editing features.

Conserved Domain Architecture Retrieval Tool (CDART)

Internet 150%

Conserved Domain Architecture Retrieval Tool (CDART)
Displays the functional domains that make up a given protein sequence. It lists proteins with similar domain architectures and can retrieve proteins that contain particular combinations of domains.

Conserved Domain Search Service (CD Search)
Identifies the conserved domains present in a protein sequence. CD-Search uses RPS-BLAST (Reverse Position-Specific BLAST) to compare a query sequence against position-specific score matrices that have been prepared from conserved domain alignments present in the Conserved Domain Database (CDD).

Related Structures
The Related Structures tool allows users to find 3D structures from the Molecular Modeling Database (MMDB) that are similar in sequence to a query protein. Although the query protein may not yet have a resolved structure, the 3D shape of a similar protein sequence can shed light on the putative shape and biological function of the query protein.

Vector Alignment Search Tool (VAST)
A computer algorithm that identifies similar protein 3-dimensional structures. Structure neighbors for every structure in MMDB are pre-computed and accessible via links on the MMDB Structure Summary pages. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone.

http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图书馆 网虫乐园 武大 本科教学系统

VAST: Vector Alignment Search Tool

VAST

NCBI

Structure

HOME | SEARCH | GUIDE | Structure Home | 3D Macromolecular Structures | Conserved Domains | PubChem | BioSystems

VAST: Vector Alignment Search Tool

ABOUT | SEARCH | HELP | nr-PDB | PUBLICATIONS | RESOURCES | NEWS

About VAST

VAST, short for **V**ector **A**lignment **S**earch **T**ool, is a computer algorithm developed at NCBI and used to identify similar protein 3-dimensional structures by purely geometric criteria, and to identify distant homologs that cannot be recognized by sequence comparison.

VAST is applied on every protein in the Molecular Modeling Database (MMDB) during MMDB data processing in order to identify similar 3D structures. The pre-computed results are accessible from a structure's summary page; to retrieve them, you can either:

1. view the "show annotation" graphic for any protein molecule of interest on a structure summary page, then click on the bar graphic for the overall protein molecule or for any 3D domain it contains in order to view a list of structures that are similar in shape to the molecule or 3D domain you selected. The VAST Help document provides additional details and illustrated examples.
2. follow the link for "Similar Structures: VAST" in the upper right corner of a structure summary page to open a tabular list of the protein molecules and 3D domains in the structure. Then select the protein or 3D domain of interest to view a list of structures that are similar in shape to the region you selected.

   Show "Similar Structures" for PDB ID or MMDB ID: 2TRX Go

If you have a newly determined protein structure that is not yet in MMDB, then you can use the **VAST Search** service

Internet 150%

http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图书馆 网虫乐园 武大 本科教学系统

MMDB Protein Structure Summary, 2TR...

VAST

Crystal structure of thioredoxin from Escherichia coli at 1.68 Angstroms Resolution

Citation:

Crystal structure of thioredoxin from escherichia coli at 1.68 a resolution.
Katti SK, Lemaster DM, Eklund H
J.Mol.Biol. (1990) **212** p.167

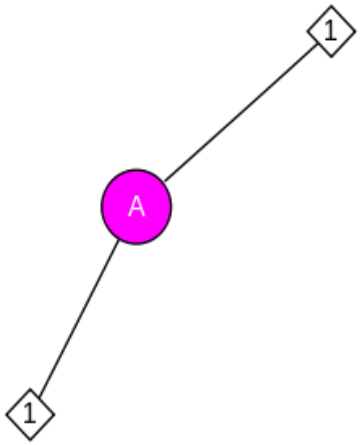
» All references (4)

PDB Deposition Date: 1990/3/19
Updated in MMDB: 05/2011
Experimental Method: X-Ray Diffraction
Resolution: 1.68 Å
Source Organism: Escherichia coli
Similar Structures: VAST

First Biological Unit | All Biological Units (2) | Asymmetric Unit

Biological Unit: monomeric; determined by: author

Interactions



Molecular Graphic

View or Save 3D Structure

File Format: Cn3D
Display As: 3D structure
Data Set: Single 3D structu

View structure

Download Cn3D

NOTICE
In order to view this biological unit properly, please upgrade to Cn3D 4.3.

Protein Chemical

Molecules and interactions

Label	Count	Molec
Protein and interactions		
		Thiore

VAST

VAST related structures have been calculated separately for individual protein chains and 3D domains present in this structure. To see the related structure list for each choose a chain or 3D domain from the table below.

Molecule	Domain Type	Alignment Range	# of Related Structures
[A]	Entire Chain	1-108	3598

Internet 100%

VAST 结果(1)

http://www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?sdid=7233

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐

Vast Neighbor Summary

View 3D Alignment of All Atoms with Cn3D Display Download Cn3D!

View Sequence Alignment using Hypertext for Selected VAST related structures

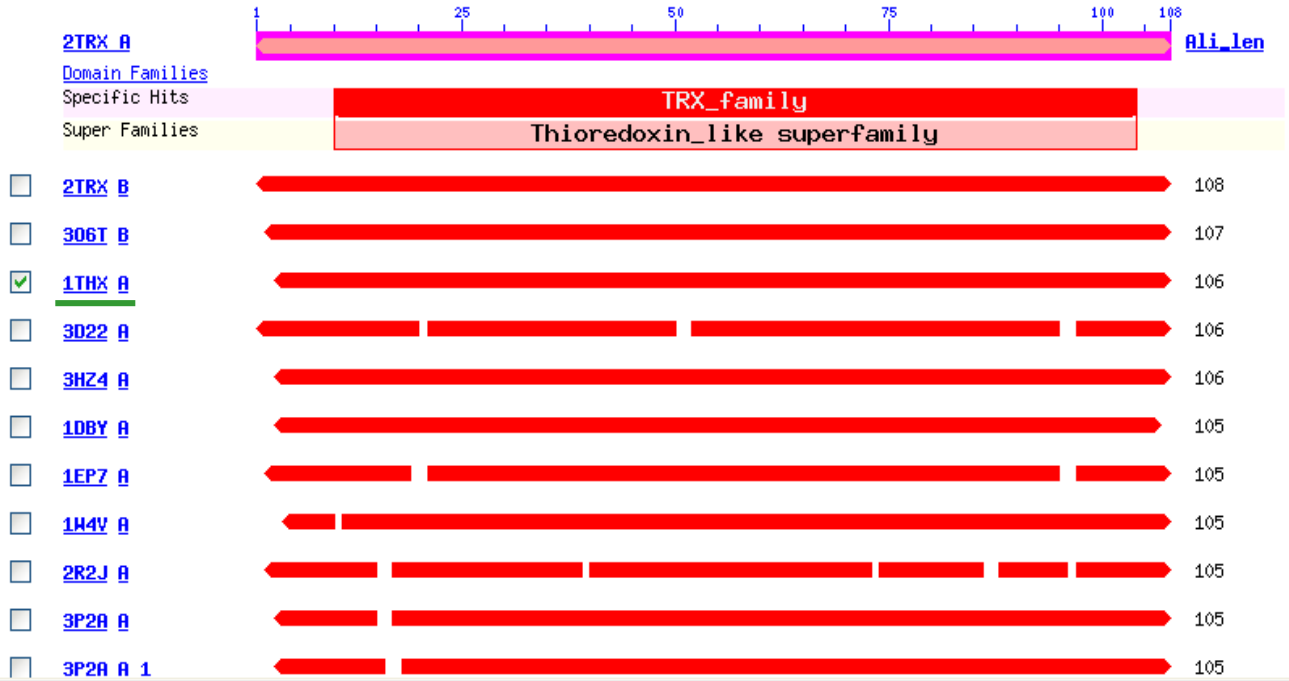
List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.

Total related structures: 3598; 1 - 60 of 566 representatives from the Medium redundancy subset displayed. Page: 1

Click to: Check All Uncheck All



已完毕，但网页上有错误。 Internet 100%

VAST 结果(2)

http://www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?sdid=7233

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图

Vast Neighbor Summary

View 3D Alignment of All Atoms with Cn3D Display Download C

View Sequence Alignment using Hypertext for Selected VAST related structures

List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-

Total related structures: 3598; 1 - 60 of 566 representatives from the Medium redundancy subset disp

Click to: Check All Uncheck All

1 25 50 75

2TRX A

Domain Families

Specific Hits TRX_family

Super Families Thioredoxin_like superfamily

2TRX B

3O6T B

1THX A

3D22 A

3HZ4 A

1DBY A

1EP7 A

1W4V A

2R2J A

3P2A A

3P2A A 1

2TRX neighbors - Cn3D 4.1

File View Show/Hide Style Window CDD Help

Cu

2TRX neighbors - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

2TRX_A ~~~sdKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFK

1THX_A tamskGVITITDAEFESEVLKAEQPVLVYFWASWCGPCQLMSPLINLAANTYSDRLKVVKLEIDPNPTTVKKYKVEGVPALRLVK

已完毕，但网页上有错误。 Internet 100%

VAST 结果(3)

http://www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?sdid=7233

Vast Neighbor Summary

View 3D Alignment of All Atoms with Cn3D Display Download C

View Sequence Alignment using Hypertext for Selected VAST related structures

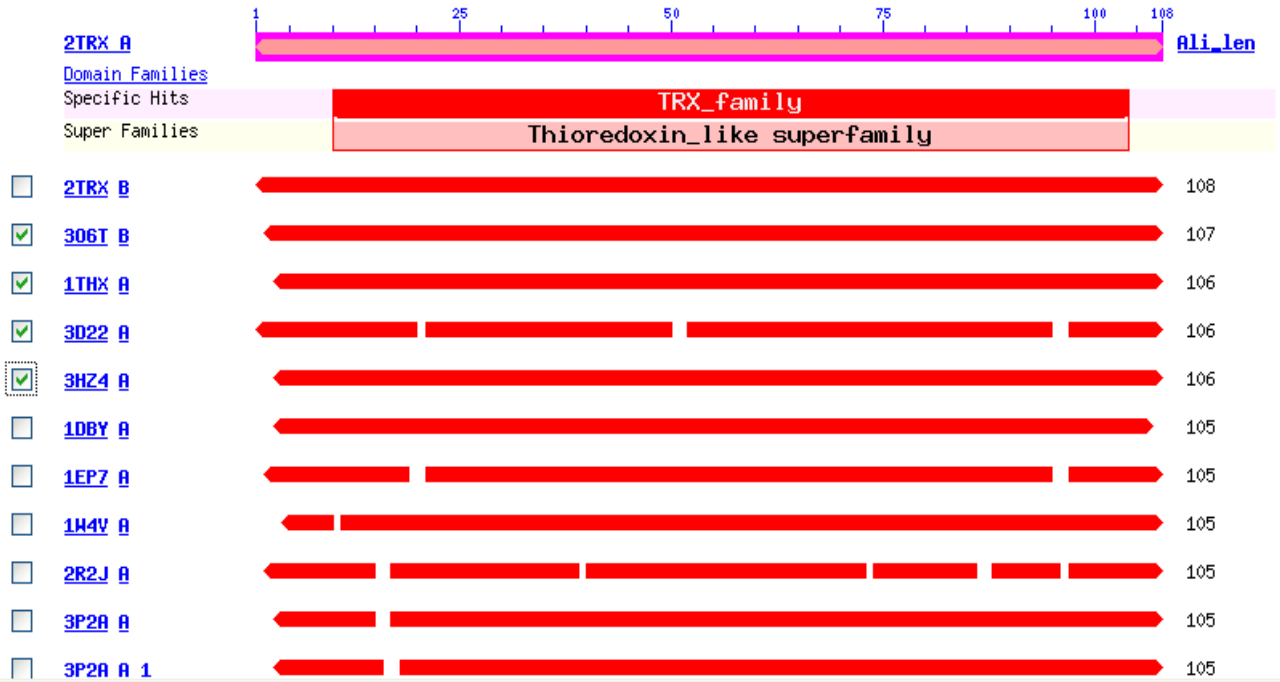
List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.

Total related structures: 3598; 1 - 60 of 566 representatives from the Medium redundancy subset displayed. Page: 1

Click to: Check All Uncheck All



VAST 结果(4)

http://www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?sdid=7233

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图

Vast Neighbor Summary

View 3D Alignment of All Atoms with Cn3D Display Download

View Sequence Alignment using Hypertext for Selected VAST related structures

List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure

Total related structures: 3598; 1 - 60 of 566 representatives from the Medium redundancy subset dis

Click to: Check All Uncheck All

1 25 50 75

2TRX A

Domain Families

Specific Hits TRX_family

Super Families Thioredoxin_like superfamily

2TRX B

3O6T B

1THX A

3D22 A

3HZ4 A

1DBY A

1EP7 A

1W4V A

2R2J A

3P2A A

3P2A A 1

2TRX neighbors - Cn3D 4.1

File View Show/Hide Style Window CDD Help

2TRX neighbors - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

```
2TRX_A   ~~~~sdKIIHLTDDSFDTDVLKAD~~~~GAILVDFWAEWCGPCKMIAPILDEIADEYQgKLTVAKLNIDQNPGTAPKYGIRGIPTL
3O6T_B   dseksATIKVTDASFATDVLSSN~~~~KPVLVDFWATWCGPSKMVAPVLEEIATERAtDLTVAKLDVDTNPETARNFQVVSIPTL
1THX_A   tamskGVITITDAEFESEVLKAE~~~~QPVLVYFWASWCGPCQLMSPLINLAANTYSdRLKVVKLEIDPNPTTVKKYKVEGVPAL
3D22_A   elaggNVHLITTKERWDQKLSEAsrdgKIVLANFSARWCGPSRQIAPYYIELSENYP~SLMFLVIDVDELSDFSASWEIKATPTF
3HZ4_A   slngsSIIEFEDXTWSQQVEDSK~~~~KPVVVXFYSPACPYCKAXEPYFEEYAKEYGsSAVFGRINIATNPWTAEKYGVQGTPTF
```

已完毕，但网页上有错误。 Internet 100%

蛋白质结构分析

- Basic notes
- Structural alignment of proteins
- Structure-based protein classification
- Structure prediction of proteins (Comparative Modeling*)

Structure-based protein classification

进化（同源）<==寻找序列相似物==>结构—功能。
进化（同源）<======== 寻找结构相似物==>功能。

---有些结构相似蛋白质的序列几乎看不出序列相似性。
Structure is more conserved, so structure-based classification is more powerful.

- 分类等级*:

Families —— 近缘同源（序列相似性较明显）;
Super-families —— 远缘同源;
Super-folds —— 有一定相似但不一定同源.

蛋白质结构分类数据库

- **SCOP (Structure Classification Of Proteins)**
  依结构相似性进行分类，等级细化顺序为——**Class**，**Folds**、 **Gene superfamily** 、**Gene family**。
- **CATH (classification by Class, Architecture, Topology, and Homology)**
  依结构相似性进行分类，等级细化顺序为——**Class**、**Architecture**、**Topology**、**Homology superfamiliy**、**Sequence family**。
- **FSSP (Fold classification based on Structure-Structure alignment of Proteins) or (Families of Structurally Similar Proteins)**
- **SARF (Spatial ARrangement of backbone Fragments)**

The legacy SCOP websites can be accessed at **SCOP 1.75** and **SCOP2 prototype**

SCOP 2

SCOP: Structural Classification of Proteins

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a
SCOP database, created by manual inspection and abetted by a battery of automated methods, aims
comprehensive description of the structural and evolutionary relationships between all proteins whose
provides a broad survey of all known protein folds, detailed information about the close relatives of ar
framework for future research and classification.

Latest update on **2022-06-29** includes **72,544** non-redundant domains representing **861,631** protein s
and families statistics **here**.

Keyword and ID search Sequence search

CATH / Gene3D v4.3

151 million protein domains classified into 5,841 superfa

Search by keywords, PDB code, GO term, etc Search

Core classification files for the latest version of CATH-Plus (v4.3) are now available to download. Daily updates of o

3D Structure

Find out what 3D structure your protein adopts

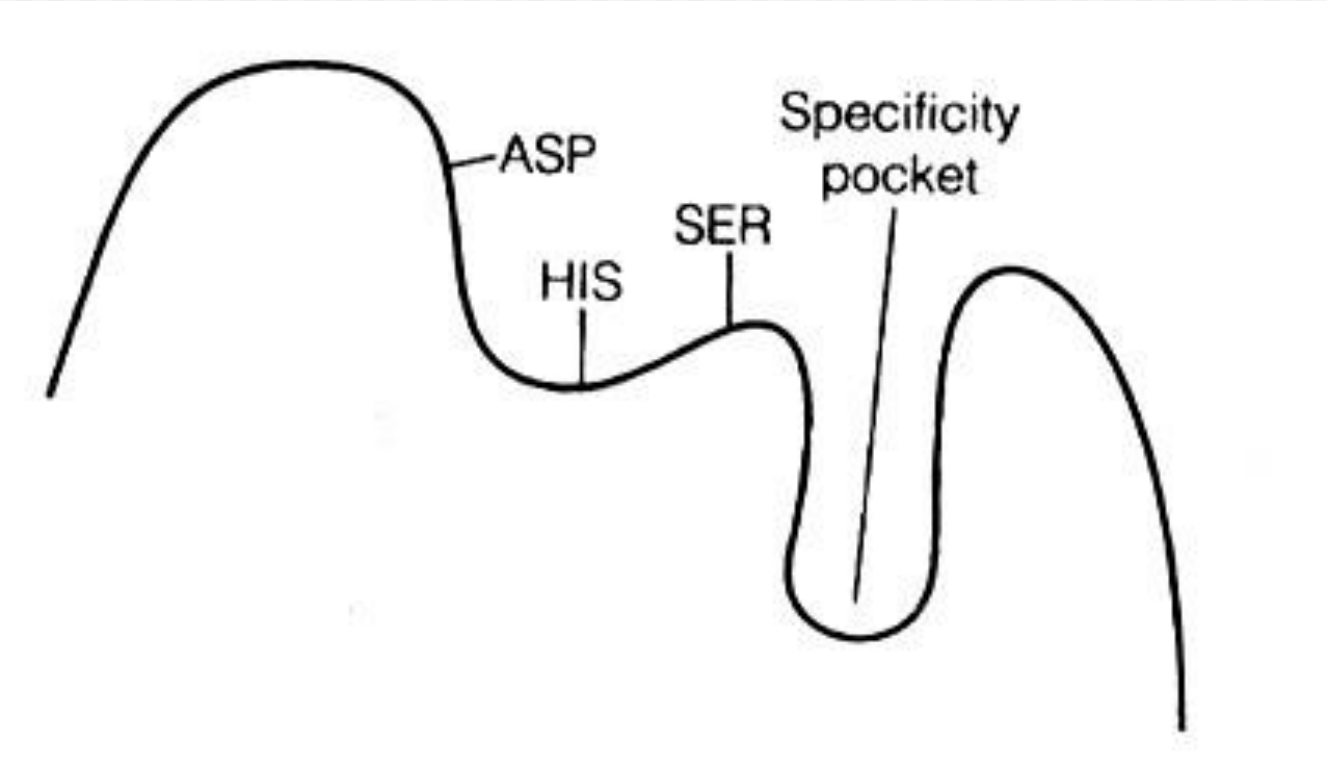
Protein Evolution

Learn about a particular protein family and how it evolved

蛋白质结构分析

- Basic notes
- Structural alignment of proteins
- Structure-based protein classification
- Structure prediction of proteins (Comparative Modeling*)

结构决定功能:

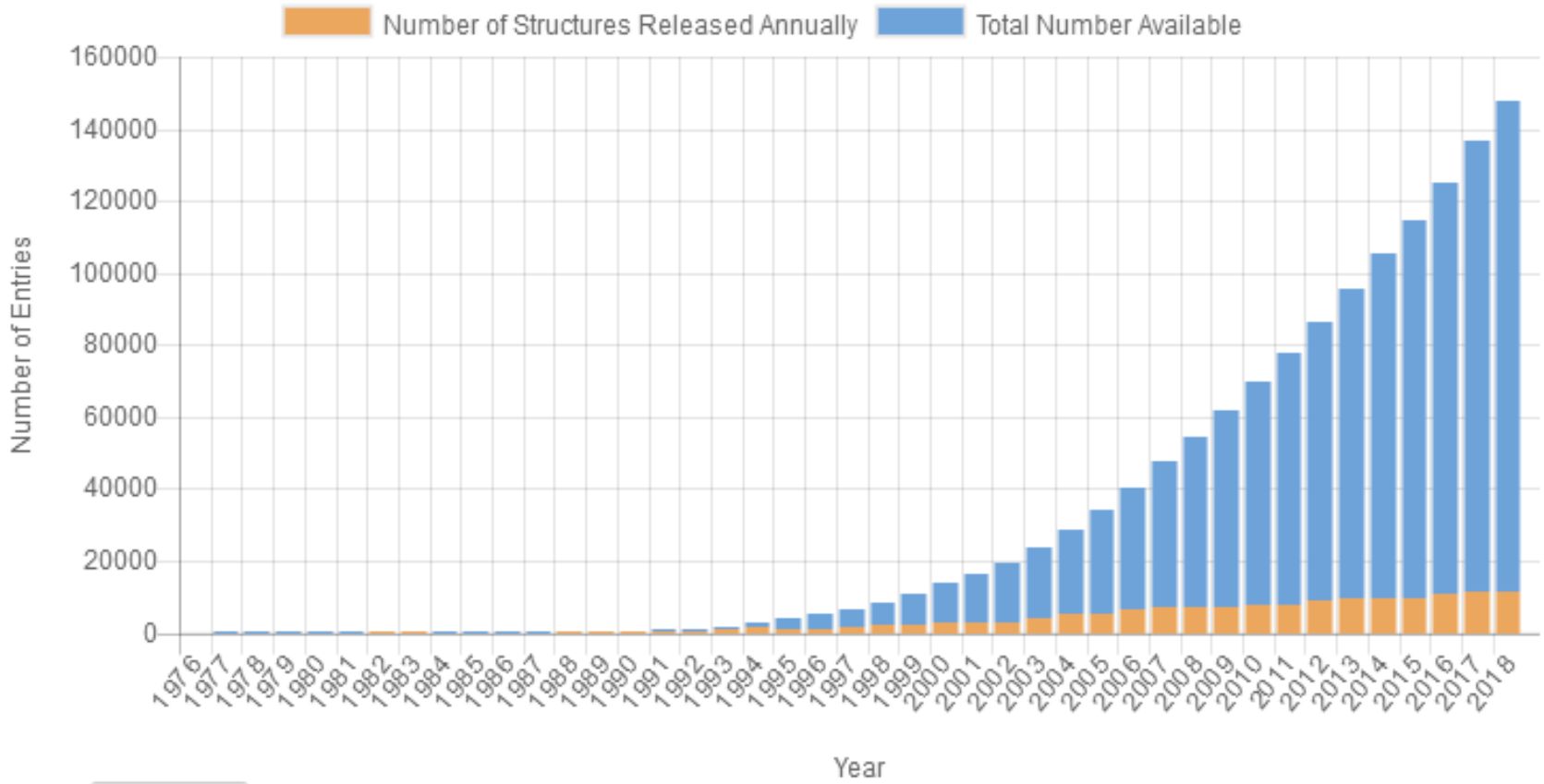


RCSB PDB Deposit Search Visualize Analyze Learn

PDB --- 统计

Other Statistics

PDB Statistics: Overall Growth of Released Structures Per Year



Contact Us

Show 10 entries

Protein structure prediction

序列数据 / 结构数据

June 2000, 86500 in SwissProt / 12500 in PDB;
Feb 2004, 144731 in SwissProt / 24358 in PDB;
Mar 2006, 208005 in SwissProt / 35343 in PDB.

.....

Jan 2017, 553231 in SwissProt / 116509 in PDB.

大量预测出的蛋白质序列.....

- 根据序列预测结构——序列决定结构。

History

- Many of the first bioinformatics programs were written in order to "solve the protein folding problem".
- Even though the field is more than 40 years old, protein structure prediction continues to be one of the most active areas in all of bioinformatics research.

CASP (Critical Assessment of Structure Prediction) competitions

Protein Structure Prediction Center

Menu

- Home
- PC Login
- PC Registration
- CASP Experiments
  - CASP15 (2022)
  - CASP14 (2020)
  - CASP13 (2018)
  - CASP12 (2016)
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)
- Initiatives
- Data Archive
- Proceedings
- CASP Measures
- Assessors
- People
- Community Resources

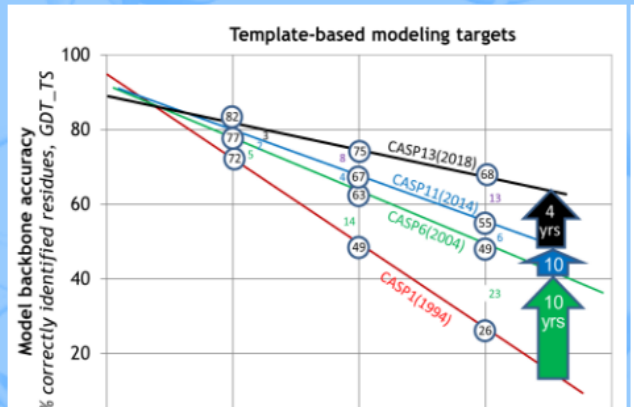
Success Stories From Recent CASPs

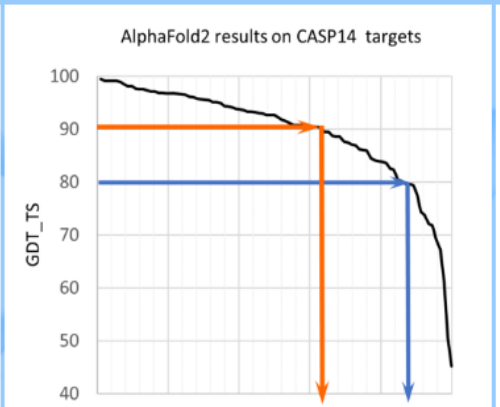
assembly modeling | **template-based modeling** | ab initio modeling | contact prediction | help structural biologists | refinement | data-assist modeling

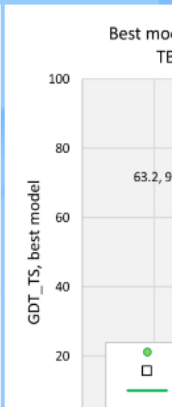
template-based modeling

Models based on templates identified by sequence similarity remain the most accurate. Over the course of there have been enormous improvements in this area. However, the overall accuracy improvements that w 10 years of CASP remained unmatched until CASP12 (2016), when a new burst of progress happened [Kry In two years from 2014 to 2016, the backbone accuracy of the submitted models improved more than in t The next CASP continued the trend [Croll et al, 2019], and the 2014-2018 model accuracy improvement d 2014 (see left plot). Several factors contributed to this, including more accurate alignment of the target se available templates, combining multiple templates, improved accuracy of regions not covered by template of models, and better selection of models from decoy sets due to improved methods for estimation of moc

CASP14 marked an extraordinary increase in the accuracy of the computed three-dimensional protein stru emergence of the advanced deep learning method AlphaFold2. Models built with this method proved to be experimental accuracy (GDT_TS>90) for ~2/3 of the targets and of high accuracy (GDT_TS>80) for almos (middle plot). The accuracy of CASP14 models for TBM targets significally superseeded accuracy of models simple transcription of information from templates, and reached the level of GDT_TS=92 on average, whic than the corresponding averages in previous two CASPs (right plot).







Prediction methods

- Ab initio:

  Ab initio prediction

- Knowledge-based:

  Comparative modeling （Homology modeling）*

  Fold recognition（Threading）

- Knowledge-trained:

  Secondary structure prediction

Ab initio prediction

- Method: it proceed from fundamental physical principles, involving quantum mechanics and statistical thermodynamics --- minimizing free energy.
- Difficult: proteins plus solvent molecules --- too large the system scale for calculation; adopting approximation to capture the essentials of the folding problem.
- Interesting: from an intellectual viewpoint; would be a huge scientific achievement; a challenge for bioinformatics.

Prediction methods

- Ab initio:

  Ab initio prediction

- Knowledge-based:

  Comparative modeling （Homology modeling）

  Fold recognition（Threading）

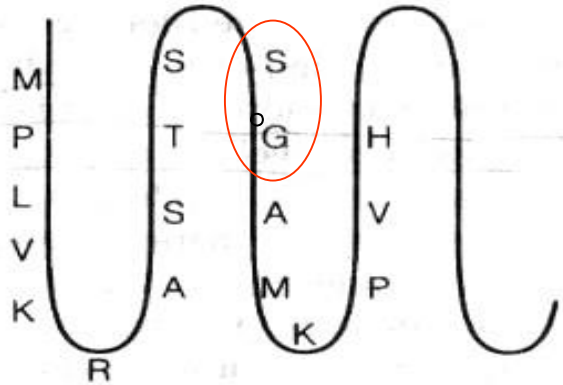
- Knowledge-trained:

  Secondary structure prediction

Comparative modeling (Homology modeling)

- **Theoretical basis**: Sequences with more than 25% identity over an alignment of 80 residues or more adopt the same basic structure.

序列比对：进化（同源）<==寻找序列相似物==>结构—功能
结构比对：进化（同源）<--------- 寻找结构相似物==>功能



Sequence of known structure:

MPLVKRASTSSGAMKPVH...

Schematic of known 3D structure (template)

Sequence of unknown structure (target):

MPILKRGTSTSYGAMRPIY...

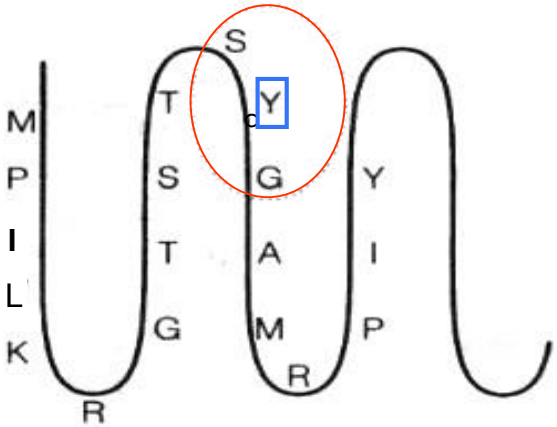
Aligned with sequence of known structure:

```
MPILKRGTSTSYGAMRPIY
MPLVKRASTSS_GAMKPVH
```

Predict the 3D structure of the target sequence by replacing the old residues on the known structure with those from the target sequence made equivalent by the alignment.

Need to position new side chain atoms.

Small structural changes where gaps are found in the alignment (loop modelling, dotted circle).



Predicted 3D structure

Fig. 1. Comparative modeling.

Preparing work

- Finding template structures:

  BLAST in the database of sequences whose structures are already known by experimental means. (--- PDB)

(Multiple alignment of the structure-template sequences with the target sequence.)

Modeling

- Backbone determination (α carbon)

  Placement of corresponding residues;

  ---- according to structure templates

  Loop modeling for insertions or deletions;

  ----spare parts algorithm; from a special loop library

- Side-chain positioning

  ---- conserved: according to structure templates,
  varied: using other sophisticated algorithms.

- Model refinement.

  ---- energy minimization……

Modeling

- Backbone determination (α-carbon)

  Placement of corresponding residues;

  ---- according to structure templates

  Loop modeling for insertions or deletions;

  ----spare parts algorithm; from a special loop library

- Side-chain positioning

  ---- conserved: according to structure templates,
  varied: using other sophisticated algorithms.

- Model refinement.

  ---- energy minimization……

Accuracy

- 序列相似性达70% ---自动预测质量较高---预测和实际结构相比，RMSD ~ 2-3埃。
- 序列相似性低于40％ --- 手工介入变得十分重要，否则预测 “can fail very badly”。
- 一般来说，30％以下的的序列相似性，用这种方法很不可靠。

Theoretical basis：Sequences with more than 25% identity over an alignment of 80 residues or more adopt the same basic structure.

Prediction methods

- Ab initio:

  Ab initio prediction

- Knowledge-based:

  Comparative modeling （Homology modeling）

  Fold recognition（Threading）

- Knowledge-trained:

  Secondary structure prediction

Fold recognition (Threading):

- A query sequence is tried to thread through a known structure to see how well it might fit.
- For an arrangement, observe how they match up with respect to properties that affect protein-folding such as whether it is hydrophobic or hydrophilic at particular points.
- The process is repeated for every other known structure in the database (fold-library) and finally, results are compared to determine which one is the most likely structure of the query protein.

-- 和同源模建类似，也限于有已知的蛋白质结构，但原则上并不要求序列十分相似，如可低于25%（同源模建的底限）。

Prediction methods

- Ab initio:

  Ab initio prediction

- Knowledge-based:

  Comparative modeling （Homology modeling）

  Fold recognition（Threading）

- Knowledge-trained:

  Secondary structure prediction

Secondary structure prediction

- Predicting the conformational state of each residue in three categories, helical, strand, and coil，usually based on ideas reflecting the preference of a residue for a particular secondary structure.
- Accuracy: (early) 60%, (+conserved domains) 66%, (+structural data + sophisticated algorithms) >70%.
- Not for integral membrane proteins（需要专门的算法）.

TABLE 10-4 Some Physical Properties of Proteins

Property	Classical Method	Example
Amino acid motifs	—	PDZ domain (e.g., nitric oxide synthase), coiled-coil domain (e.g., hemagglutinin, syntaxin, SNAP-25, myosin)
Isoelectric point (pI)	Derived from isoelectric focusing	—
Molecular weight	Derived from Stokes radius and sedimentation coefficient	—
Posttranslational modifications: phosphorylation	Enzymatic analyses	Synapsin
Posttranslational modifications: glycosylation	Enzymatic analyses	Nerve growth factor, neural cell adhesion molecule
Posttranslational modifications: isoprenylation	Biochemical analyses	Lamin B, G protein γ subunits, *rab3A*
Posttranslational modifications: palmitoylation	Biochemical analyses	β-Adrenergic receptor, *GAP-43*, insulin receptor, rhodopsin, nAChR
Posttranslational modifications: myristoylation	Biochemical analyses	PKA, $G_{i\alpha}$-subunit, MARCKS protein, calcineurin
Posttranslational modifications: GPI-anchored proteins	Enzymatic analyses	Alkaline phosphatase, *thy-1*, prion protein, 5′-nucleotidase, uromodulin
Sedimentation coefficient	Derived from sucrose density gradients	—
Stokes radius	Derived from gel filtration	—
Transmembrane domain	Derived from subcellular fractionation	—

Abbreviations: G protein, guanosine triphosphate-binding protein; GAP-43, growth-associated protein of 43 kDa; MARCKS, myristoylated alanine-rich C-kinase substrate; nAChR, nicotinic acetylcholine receptor; PDZ domain, post-synaptic density protein PSD-95, the *Drosophila* tumor suppressor discs-large, tight-junction protein ZO-1; PKA, protein kinase A; SNAP-25, synaptosomal-associated protein of 25 kDa; Rab3A, rat brain GTP-binding protein 3A; thy-1, thymocyte-1.

基于蛋白质序列的相关特性分析对后续其结构和功能分析可能十分有用。

特性 (Physical properties)

Strategy

0 → Preliminary sequence analysis

Prediction methods

- Ab initio:

3 → *Ab initio* prediction

- Knowledge-based:

1 → Comparative modeling （Homology modeling）

2 → Fold recognition（Threading）

- Knowledge-trained:

2 → Secondary structure prediction

AlphaFold: Using AI for scientific discovery

Today we´re excited to share DeepMind´s first significant milestone in demonstrating how artificial intelligence research can drive and accelerate new scientific discoveries. With a strongly interdisciplinary approach to our work, DeepMind has brought together experts from the fields of structural biology, physics, and machine learning to apply cutting-edge techniques to predict the 3D structure of a protein based solely on its genetic sequence.

Our system, **AlphaFold**, which we have been working on for the past two years, builds on years of prior research in using vast genomic data to predict protein structure. The 3D models of proteins that AlphaFold generates are far more accurate than any that have come before—making significant progress on one of the core challenges in biology.

ental techniques like cryo-electron microscopy, nuclear magnetic

ay crystallography, but each method depends on a lot of trial and error,

ars and cost tens of thousands of dollars per structure. This is why biologists are turning to AI methods as an alternative to this long and laborious process for difficult proteins.

Fortunately, the field of genomics is quite rich in data thanks to the rapid reduction in the

We're proud to be part of what the CASP organisers have called "unprecedented progress in the ability of computational methods to predict protein structure," placing **first** in rankings among the teams that entered (our entry is A7D).

Our team focused specifically on the hard problem of modelling target shapes from scratch, without using previously solved proteins as templates. We achieved a high degree of accuracy when predicting the physical properties of a protein structure, and then used two distinct methods to construct predictions of full protein structures.

Using neural networks to predict physical properties

CASP (Critical Assessment of Structure Prediction) competitions

Protein Structure Prediction Center

Menu

- Home
- PC Login
- PC Registration
- CASP Experiments
  - CASP14 (2020)
  - *CASP Commons (COVID-19, 2020)*
  - CASP13 (2018)
  - CASP12 (2016)
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)
- Initiatives
- Data Archive
- Proceedings

Success Stories From Recent CASPs

template-based modeling | *ab initio modeling* | ***contact prediction*** | *help structural biologists* | *refinement* | *data-assisted modeling* | ||

contact prediction

The most notable progress in recent CASPs (2014, 2016) resulted from sustained improvement in methods for predicting three-dimensional contacts between pairs of residues in structures. Average precision of the best CASP12 contact predictor almost doubled compared to that of the best CASP11 predictor (from 27% to 47% - see the plot). Advances in the field as a whole are not any less impressive: 26 methods in CASP12 showed better results than the best method in CASP11. [Schaarschmidt et al, 2018]

Theoretical advance in contact prediction lead to improved accuracy of 3D models, especially for the hardest template-free modeling cases (see models for CASP12 target T0915 below).

CASP13 (2018) registered yet another leap in accuracy of contact prediction, with the average precision of the best contact prediction group increasing by 23% (compared to CASP12) and reaching 70%.

Message Board

CASP14 Job Fair

Dear CASPers, Again, many thanks for your participation in th CASP14 meeting (and making it the most attended CASP meeting to date)! One thing that we were not able to make happen this tim around ...

CASP14 Program and updates

Dear CASPers, As you might have noticed, we had posted the CASP14 conference program online yesterday. It is available from the http://predictioncenter.org/casp1 web page. The CASP14 results w ...

Early registration deadline and instructions for poster uploads

Dear CASPers, If you plan to attend the CASP14 virtual meeting, please register. We will have limited capacity to process registrations during the conference and the access to sessions will not be ...

CASP-2018

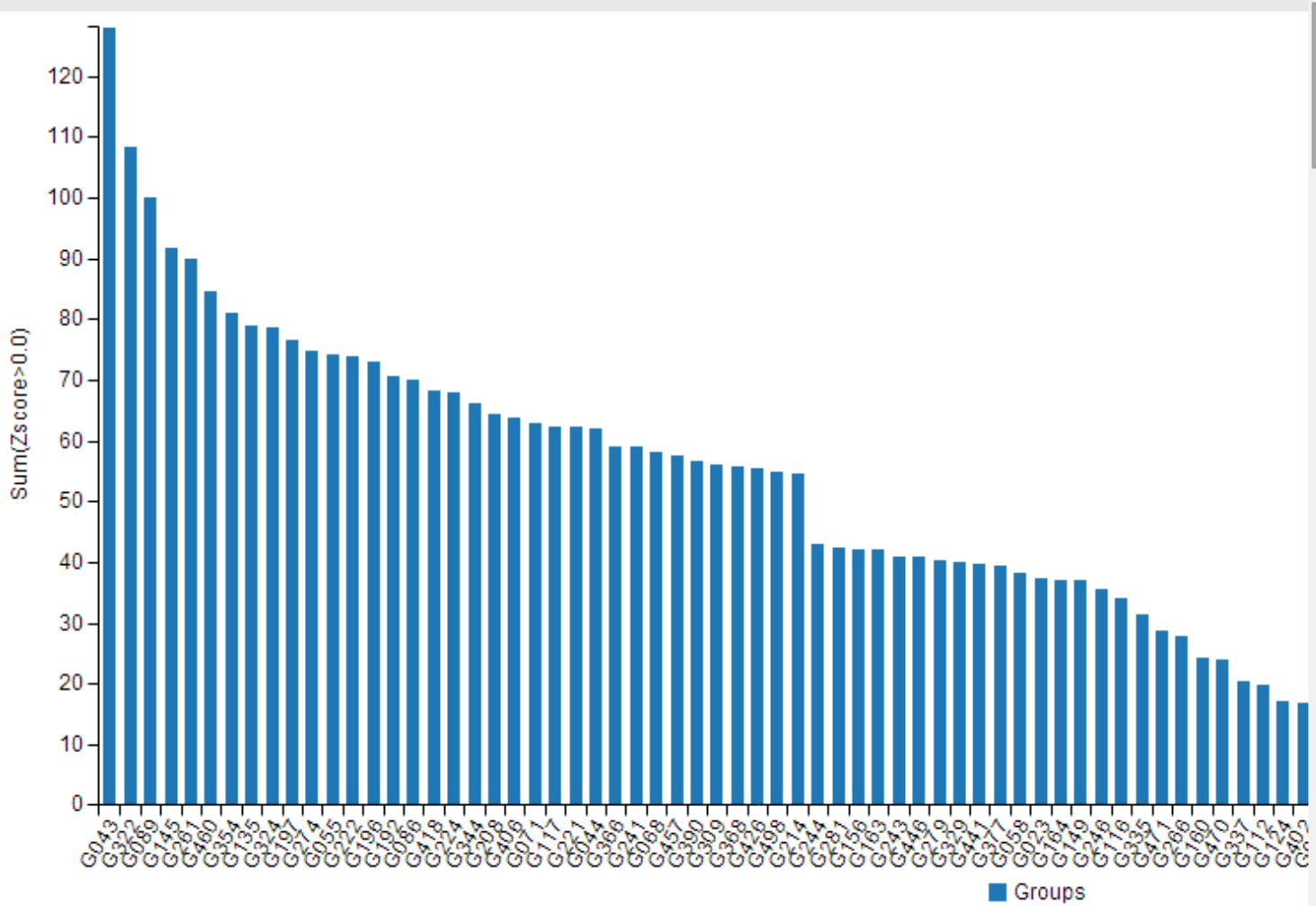
#	GR code	GR name	Domains Count	SUM Zscore (>-2.0)	Rank SUM Zscore (>-2.0)	AVG (>-2.
1	043	A7D	103	119.8429	1	1.1635
2	322	Zhang	103	106.7674	2	1.0366
3	089	MULTICOM	103	98.6232	3	0.9575
4	145	QUARK	103	90.1429	4	0.8752
5	261	Zhang-Server	103	88.1913	5	0.8562
6	460	McGuffin	103	80.7632	6	0.7841
7	354	wfAll-Cheng	103	76.8791	7	0.7464
8	135	SBROD	101	71.5034	9	0.7476
9	324	RaptorX-DeepModeller	103	75.4289	8	0.7323
10	197	MESHI	103	70.2323	10	0.6819
11	274	MUFold	103	66.9621	13	0.6501
12	222	Seok-refine	103	63.5832	15	0.6173
13	055	VoroMQA-select	103	67.2288	12	0.6527
14	196	Grudinin	103	68.8019	11	0.6680
15	192	Elofsson	102	62.8775	16	0.6361
16	086	BAKER	101	57.8510	18	0.6124
17	224	Destini	103	64.7773	14	0.6289
18	418	Seder3nc	103	40.3575	28	0.3918
19	344	Kiharalab	103	60.8535	17	0.5908
20	208	KIAS-Gdansk	99	49.0663	22	0.5764
21	406	Seder3mm	103	49.6085	21	0.4816
22	221	RaptorX-TBM	103	55.1204	19	0.5351

CASP13--2018

CASP3 (1998)
CASP2 (1996)
CASP1 (1994)
Initiatives
Data Archive
Proceedings
CASP Measures
Feedback
Assessors
People
Community Resources
Job Fair

FM
FM_sp (Combination of EUs: T0953s2-D23, T0984, T1000, T1002)

Show



CASP (Critical Assessment of Structure Prediction) competitions

Protein Structure Prediction Center

Menu

- Home
- PC Login
- PC Registration
- CASP Experiments
  - CASP15 (2022)
  - CASP14 (2020)
  - CASP13 (2018)
  - CASP12 (2016)
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)
- Initiatives
- Data Archive
- Proceedings
- CASP Measures
- Assessors
- People
- Community Resources

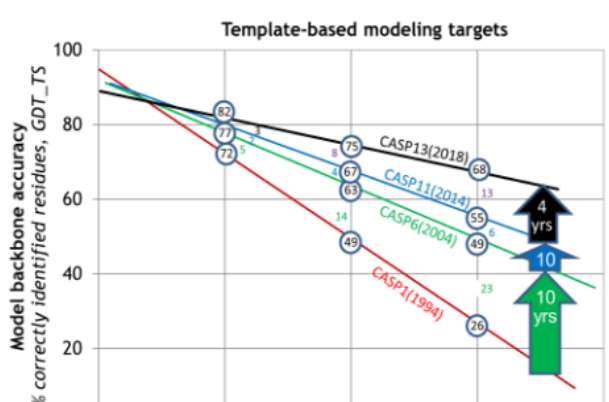
Success Stories From Recent CASPs

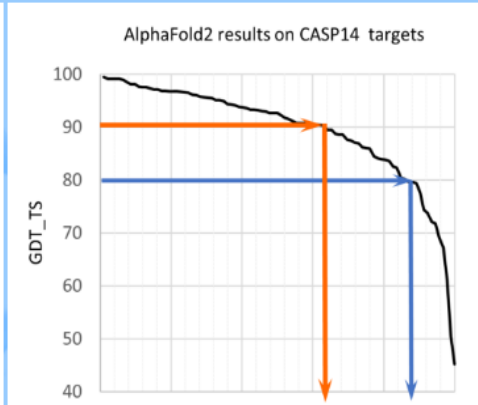
assembly modeling | **template-based modeling** | ab initio modeling | contact prediction | help structural biologists | refinement | data-assist modeling

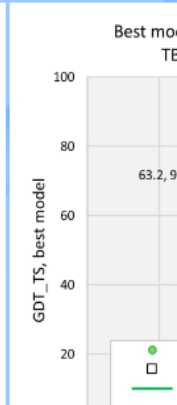
template-based modeling

Models based on templates identified by sequence similarity remain the most accurate. Over the course of there have been enormous improvements in this area. However, the overall accuracy improvements that w 10 years of CASP remained unmatched until CASP12 (2016), when a new burst of progress happened [Kry In two years from 2014 to 2016, the backbone accuracy of the submitted models improved more than in t The next CASP continued the trend [Croll et al, 2019], and the 2014-2018 model accuracy improvement d 2014 (see left plot). Several factors contributed to this, including more accurate alignment of the target se available templates, combining multiple templates, improved accuracy of regions not covered by template of models, and better selection of models from decoy sets due to improved methods for estimation of moc

CASP14 marked an extraordinary increase in the accuracy of the computed three-dimensional protein stru emergence of the advanced deep learning method AlphaFold2. Models built with this method proved to be experimental accuracy (GDT_TS>90) for ~2/3 of the targets and of high accuracy (GDT_TS>80) for almos (middle plot). The accuracy of CASP14 models for TBM targets significally superseeded accuracy of models simple transcription of information from templates, and reached the level of GDT_TS=92 on average, whic than the corresponding averages in previous two CASPs (right plot).







CASP-2020

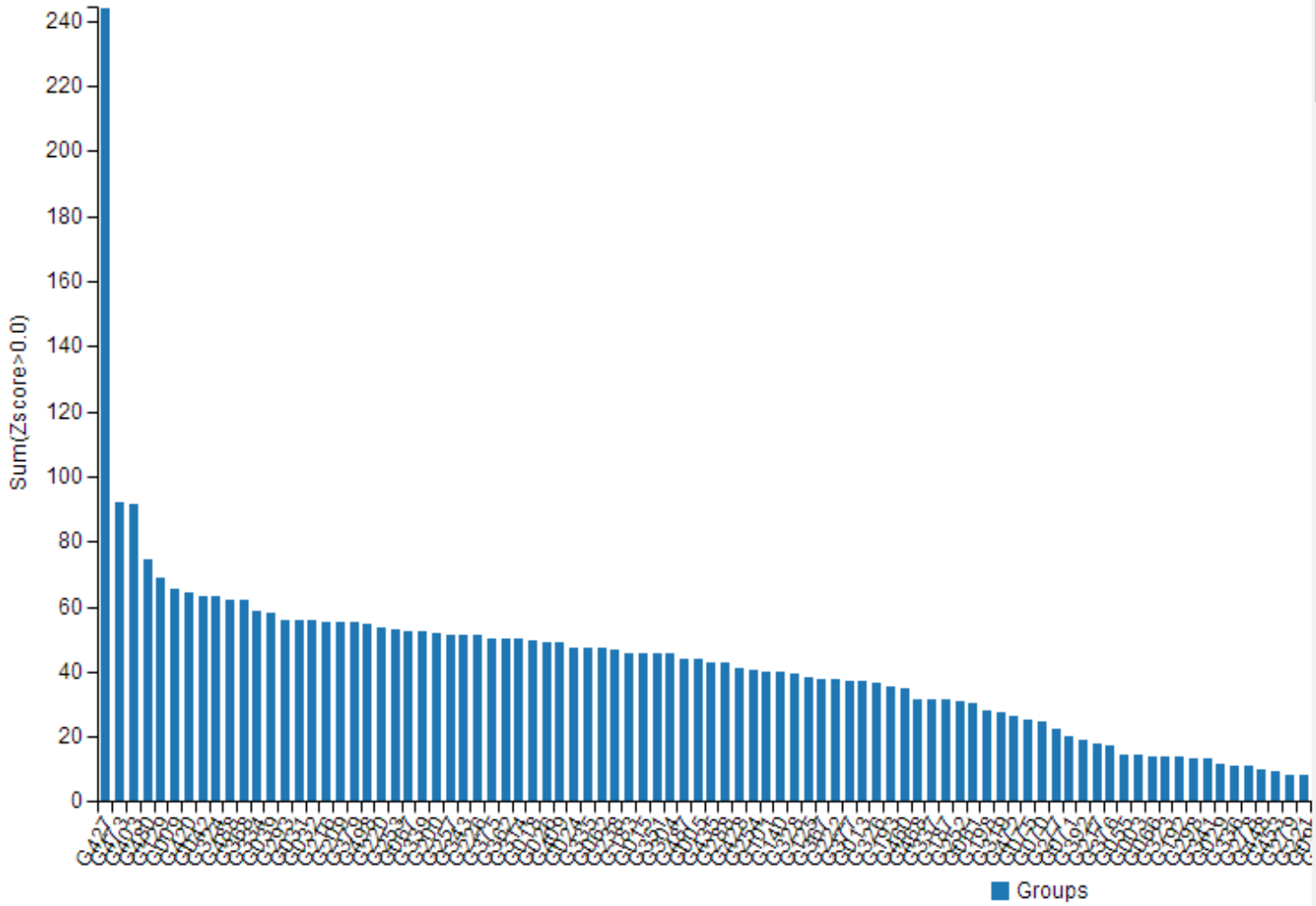
#	GR code	GR name	Domains Count	SUM Zscore (>-2.0)	Rank SUM Zscore (>-2.0)	AVG Zscore (>-2.0)	Rank AVG Zs (>-2.0)
1	427	AlphaFold2	92	244.0217	1	2.6524	1
2	473	BAKER	92	90.8241	2	0.9872	2
3	403	BAKER-experimental	92	88.9672	3	0.9670	3
4	480	FEIG-R2	92	72.5351	4	0.7884	4
5	129	Zhang	92	67.9065	5	0.7381	5
6	009	tFold_human	92	61.2858	7	0.6661	8
7	420	MULTICOM	92	63.2689	6	0.6877	7
8	042	QUARK	92	60.0226	10	0.6524	11
9	324	Zhang-Server	92	60.8875	8	0.6618	9
10	488	tFold-IDT_human	92	57.6435	11	0.6266	12
11	368	tFold-CaT_human	92	60.5423	9	0.6581	10
12	334	FEIG-R3	92	48.4424	20	0.5265	23
13	039	ropius0QA	92	55.7086	12	0.6055	13
14	293	MUFOLD_H	92	47.7806	21	0.5194	24
15	031	Zhang-CEthreader	92	49.5742	18	0.5389	21
16	032	MESHI	92	53.0953	14	0.5771	15
17	216	EMAP_CHAE	92	53.1597	13	0.5778	14
18	209	BAKER-ROSETTASERVER	92	46.1861	25	0.5020	28
19	379	Wallner	92	51.7365	15	0.5624	16
20	498	VoroMQA-select	92	51.4288	17	0.5590	18
21	220	McGuffin	92	49.5443	19	0.5385	22
22	253	Bhattacharya	92	51.5239	16	0.5600	17

CASP14--2020

CASP2 (1996)
CASP1 (1994)
Initiatives
Data Archive
Proceedings
CASP Measures
Feedback
Assessors
People
Community Resources
Job Fair

Multidom

Show



AlphaFold

蛋白质三维结构预测是生物学最严峻的挑战之一。继围棋、国际象棋等竞技项目之后，近日谷歌旗下DeepMind开发的人工智能程序AlphaFold在两年一次的蛋白质结构预测挑战赛CASP中再次大幅胜出。该程序在根据蛋白质氨基酸序列确定蛋白质三维结构方面取得巨大飞跃，准确性可与冷冻电子显微术（又称冷冻电镜）（Cryo-EM）和X-射线晶体学等实验技术相媲美。

第一代AlphaFold依托蛋白质数据库PDB作为训练数据集，构建神经网络，采用深度学习预测氨基酸残基间的方向和距离，混合传统算法Rosetta对蛋白质结构进行同源建模、结构优化；与此不同的是，第二代AlphaFold则将折叠蛋白质视为“空间图”，基于神经网络系统进行“端到端”的训练，使用了进化相关的氨基酸序列，多序列比对以及对氨基酸对的评估来优化结构预测。研究人员使用蛋白质数据库中接近17万个不同的蛋白质结构，通过不断地迭代，AlphaFold系统学习到了基于氨基酸序列精确预测蛋白结构的能力。这一基于原子坐标近乎“暴力”的算法是全新的途径，是全新算法与强大算力的强强联合。

AlphaFold

正如马里兰大学帕克分校计算生物学家，CASP共同创始人John Moult所言，从某种程度上而言，结构预测问题得到了解决。根据氨基酸序列准确预测蛋白质结构的能力将对生命科学和医学带来巨大的好处。这将极大地加快对细胞组成模块的理解，对于更快更先进的药物发现显然有很大帮助。Nature使用“它将会改变一切”来报道这一关键成果……

History

- Many of the first bioinformatics programs were written in order to "solve the protein folding problem".
- Even though the field is more than 40 years old, protein structure prediction continues to be one of the most active areas in all of bioinformatics research.

Article

Highly accurate protein structure prediction with AlphaFold

预测人类蛋白质组的结构

https://doi.org/10.1038/s41586-021-03819-2

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 15 July 2021

Open access

Check for updates

John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4]✉

2021, 7, 15
Nature

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1–4], the structures of around 100,000 unique proteins have been determined[5], but this represents a small fraction of the billions of known protein sequences[6,7]. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the 'protein folding problem'[8]—has been an important open research problem for more than 50 years[9]. Despite recent progress[10–14], existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method

蛋白质组（Proteome）

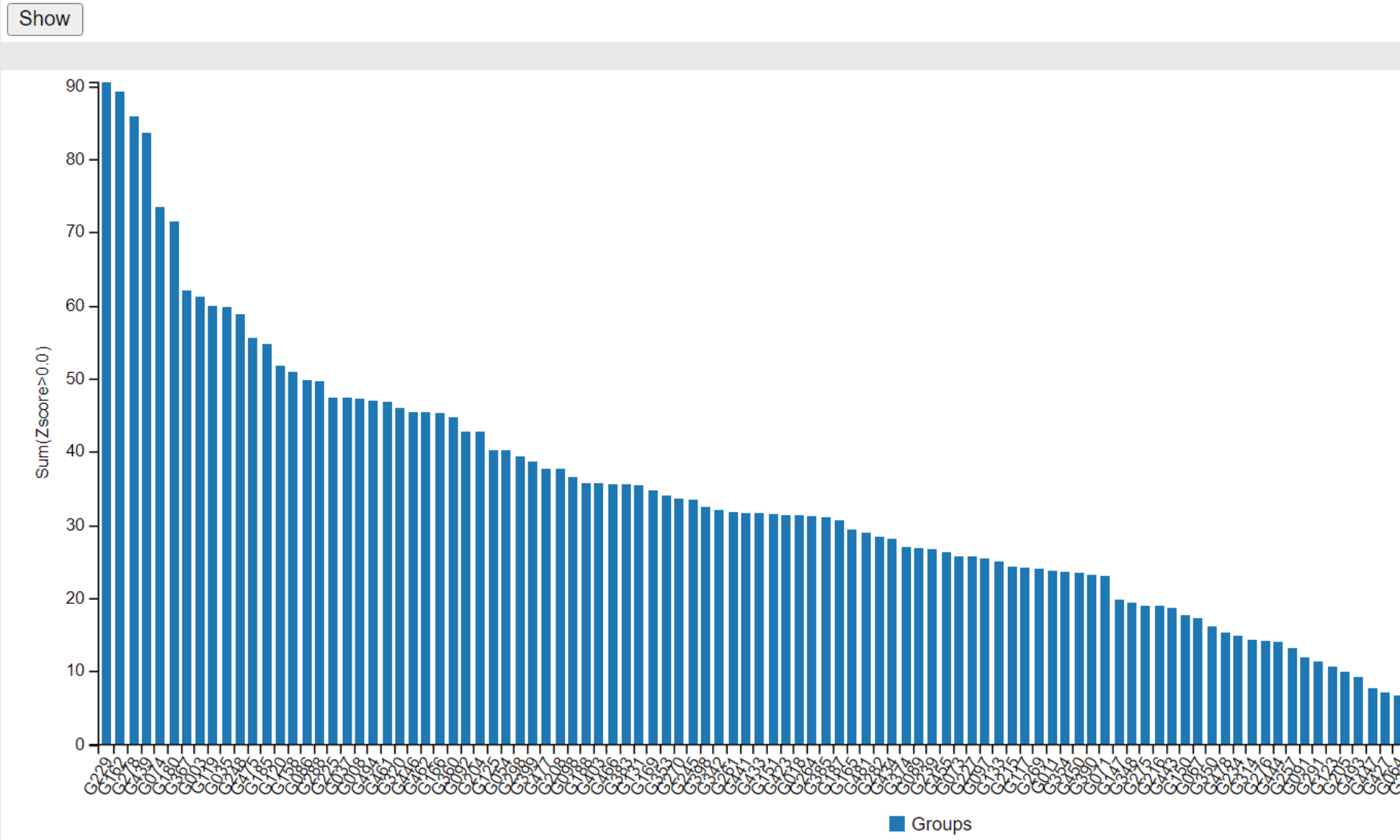
- 一个基因组表达产生的所有蛋白质的总体。（抽象） —— 意义不大。

  “the complete set of proteins expressed by an organism”

- 在某种内在和外在条件下，一个基因组表达产生的所有蛋白质的总体。（具体） —— 通常所指。

  “the complete set of proteins expressed by a cell or a tissue in a definitive situation”

CASP15--2022



Two years later, AlphaFold still dominates the competition.

Deepmind itself did not participate in this round, but AlphaFold has been open source since 2021 and the most successful participants have integrated Deepmind's AI system into their approaches.

In predicting the shape of individual proteins, participating teams achieved moderate improvements in accuracy. "The accuracy is already so high that it's hard to improve on it," …

Aside from AlphaFold's proven capabilities, several teams this year also demonstrated how the AI system can be used with modifications to predict protein interactions. Compared to CASP14, systems using such AlphaFold variants have made significant improvements and are slowly approaching the accuracy of experimental methods.

Strategy

0 → Preliminary sequence analysis

Prediction methods

- Ab initio:

3 → *Ab initio* prediction

- Knowledge-based:

1 → Comparative modeling （Homology modeling）

2 → Fold recognition（Threading）

- Knowledge-trained:

2 → Secondary structure prediction

http://us.expasy.org/structural_bioinformatics

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图书馆 网虫乐园 武大

ExPASy: SIB Bioinformatics Resource...

ExPASy

ExPASy
Bioinformatics Resource Portal

Home About Contact

Query all databases | search | help

Visual Guidance

Categories

- proteomics
- genomics
- **structural bioinformatics**
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- drug design

Resources A..Z

Links/Documentation

Databases

- SWISS-MODEL Repository • protein structure homology models • [more]
- Protein Model Portal • structural information for a protein • [more]
- SwissSidechain • non-natural amino-acid sidechains • [more]

Tools

- SWISS-MODEL Workspace • structure homology-modeling • [more]
- SwissDock • protein ligand docking server • [more]
- Click2Drug • drug design tools • [more]
- COILS • Prediction of Coiled Coil Regions in Proteins • [more]
- MARCOIL • coiled-coils prediction • [more]
- OpenStructure • molecular modelling and visualization • [more]
- Protein Model Portal • structural information for a protein • [more]
- QMEAN • estimate quality of protein models • [more]
- Swiss-PdbViewer • analyse protein 3D structures • [more]
- SwissParam • topology, parameters for small organic molecules • [more]

Internet 100%

Swiss-Model

http://swissmodel.expasy.org/workspace/

收藏夹 新浪竞技 建议网站 BLAST NCBI 华军 生科院 生信 搜狐 图书

SWISS-MODEL Workspace

BIOZENTRUM Universität Basel The Center for Molecular Life Sciences

SWISS-MODEL Workspace

Modelling Tools Repository Documentation

[myWorkspace] [login]

- myWorkspace
- Automated Mode
- Alignment Mode
- Project Mode

SWISS-MODEL Workspace

An Automated Comparative Protein Modelling Environment

SIB - Biozentrum Basel site provided by:

SIB Swiss Institute of Bioinformatics

SWISS-MODEL Version 8.05 released

We are pleased to announce a new release of Swiss-Model Workspace.

What's new?

- New automated modeling pipeline with improved hierarchical approach for template selection.
- New SWISS-MODEL template library (SMTL) HMM profiles
- Increased sensitivity of template detection (sequence to profile search using an adapted HHSearch protocol)
- New tools for model and structure quality assessment: Dfire and Qmean global scores; ProQres residue based assessment scores
- Additional hardware: As the new pipeline requires significantly more computational resources, we have added additional compute nodes to the cluster

http://swissmodel.expasy.org/workspace/index.php?func=show_workspace&userid=USERID&token=TOKEN Internet 100%

BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL

(2022)

Modelling Repository T

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Paste your target sequence(s) or UniProtKB AC here

Upload Target Sequence File... Validate

Project Title: Untitled Project

Email: Optional

Search For Templates Build Model

By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.

You are currently not logged in - to take advantage of the workspace, please log in or create an account.

(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projec

Modelling Projects in Session

BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL

(2022)

Modelling | Repository | Tools | Documer

All Projects

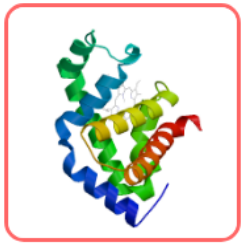
Untitled Project Created: today at 08:40

Summary | Templates 50 | Models 1 | Project Data

Model Results

Order by: GMQE

1 147



Model 01

Structure Assessment

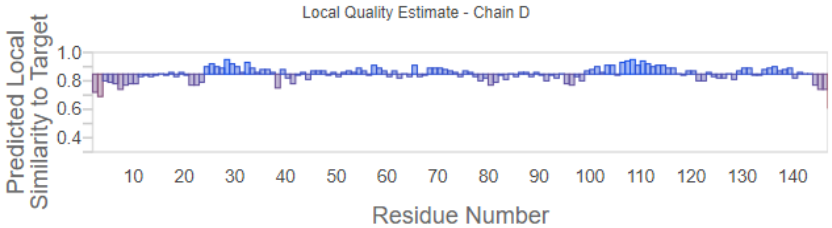
Oligo-State
Monomer

GMQE
0.93

QMEANDisCo Global:
0.85 ± 0.07

Ligands
1 x HEM

QMEANDisCo Local



QMEAN Z-Scores

Template