生物信息学

蛋白质三维结构预测 2

蛋白质结构 --- 三级结构

PyMOL http://www.pymol.org

分子三维结构查看及分析软件**PyMOL**

PyMOL by Schrödinger

DOWNLOAD SCREENSHOTS PRODUCTS SUPPORT CONTACT

Download PyMOL 2.3

Version 2.3.0 - Updated February 11th 2019 (Installation instructions)

For previous versions and Python 2.7 bundles, see here.

These bundles include Python 3.7

Windows	Windows	macOS	Linux
EXE Installer	ZIP Archive	DMG Disk Image	TAR.BZ2 Archive

蛋白质结构 --- 三级结构

分子三维结构查看及分析软件VMD

下载：http://www.ks.uiuc.edu/Research/vmd/

从PDB数据库下载任一个.pdb文件，用写字板打开，对照VMD显示出来的东东，熟悉一下.pdb文件格式。

蛋白质结构 --- 三级结构

THEORETICAL *and* COMPUTATIONAL BIOPHYSICS GROUP

Home | Research | Publications | Software | Instruction | News | Galleries | Facilities | About Us

- Home
- Overview
- Publications
- Research
- Software
  - VMD Molecular Graphics Viewer
  - NAMD Molecular Dynamics Simulator
  - BioCoRE Collaboratory Environment
  - MD Service Suite
  - Structural Biology Software Database
  - Computational Facility
- Outreach

VMD Mailing List

Download VMD

VMD Tutorials

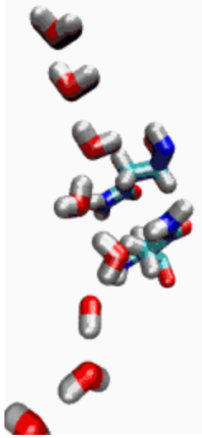
VMD Visual Molecular Dynamics

VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics and built-in scripting. VMD supports computers running MacOS X, Unix, or Windows, is distributed free of charge, and includes source code. (**more details...**)

Spotlight

VMD can now make movies easier than ever before, with the use of a **movie plugin** that takes care of the entire movie making process. The vmdmovie plugin generates one of several built-in movie types, according to user selectable options. Once preferences and selections are made, the movie generator takes control of VMD and takes care of the entire process, from the generation of individual movie frames using on-screen snapshots or ray tracers, image format conversion staging of the image data for compression, invocation of movie compressor programs, and final disk space cleanup and temporary file deletion. This makes the whole process of making movies much simpler for inexperienced users.

Other Spotlights



Overview

- **Molecular representations**
- **VMD plugin library**
- **Molecular file formats**
- **GPU-accelerated computing**
- **Interactive molecular dynamics**

News and Announcements

Immersive Molecular Visualization with Omnidirectional Stereoscopic Ray Tracing and Remote Rendering, HPDAV 2016 NEW

High Performance Molecular Visualization: In-Situ and Parallel Rendering with EGL, HPDAV 2016 NEW

Evaluation of Emerging Energy-Efficient Heterogeneous Computing Platforms for Biomolecular

蛋白质结构分类数据库

CATH是一个按等级分类PDB中蛋白质结构的数据库。只有分辨率在4埃以内的晶体结构以及NMR结构才被分类。蛋白质结构分类结合了自动与手动两个过程。总共有四个水平上的分类：

Class → **A**rchitecture → **T**opology → **H**omologous superfamily

Go to: http://www.cathdb.info/

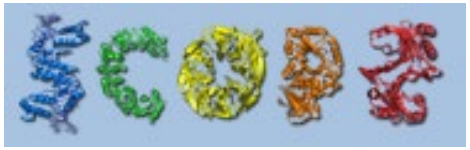
蛋白质结构分类数据库　SCOP

http://scop2.mrc-lmb.cam.ac.uk/

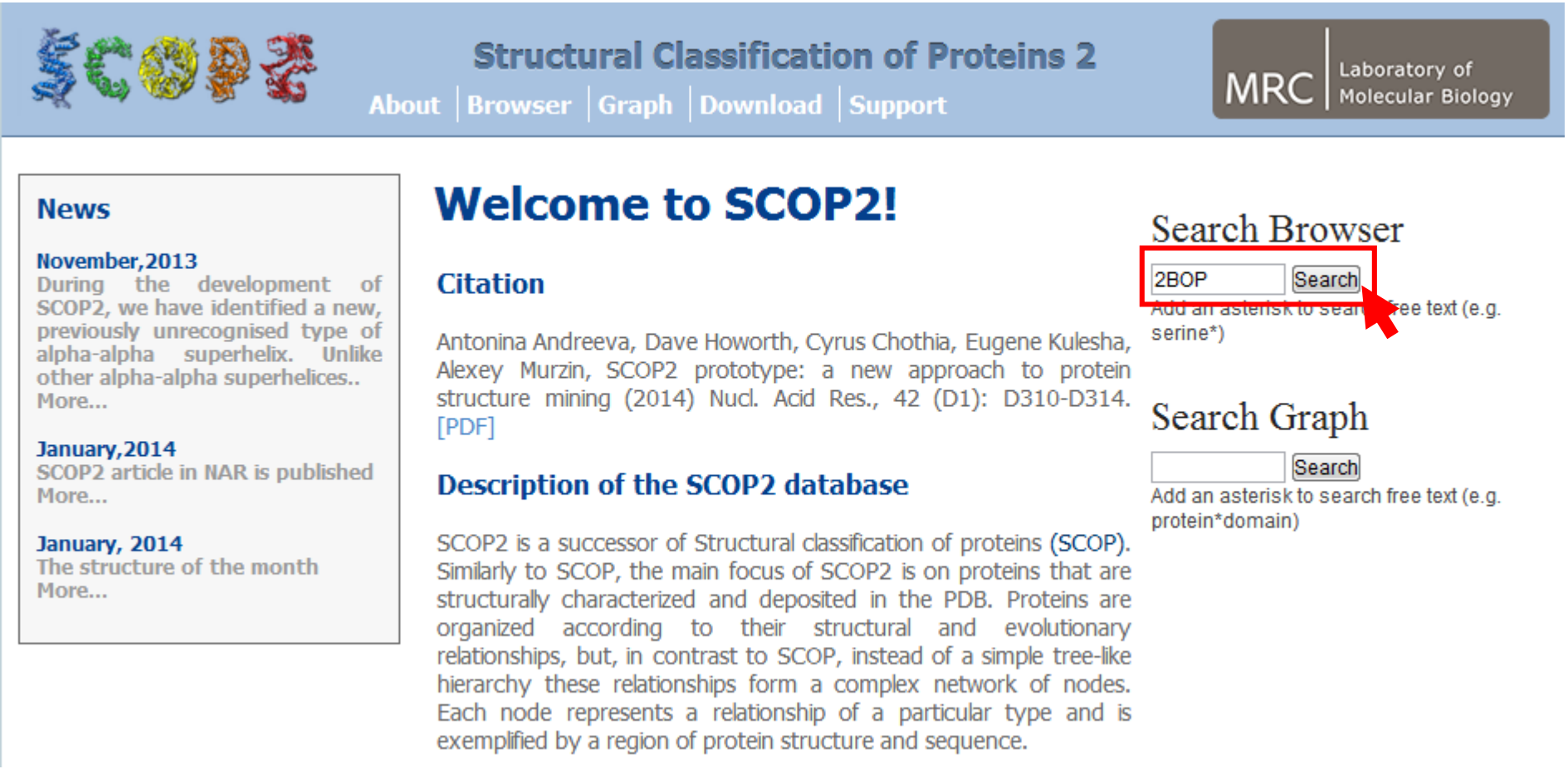
英国医学研究委员会（Medical Research Council, MRC）的分子生物学实验室和蛋白质工程研究中心于2014年2月正式发布了蛋白质结构分类数据库SCOP（structural classification of proteins）的全面升级版SCOP2。该数据库在搜集、整理、分析PDB数据中已知的蛋白质三维结构的基础上，详细描述了已知结构的蛋白质在结构、进化事件与功能类型三个方面的关系。 数据库的构建除了使用计算机程序外，主要依赖于人工验证。SCOP2把SCOP中仅基于蛋白质结构的树状等级分类系统发展成为单向非循环网状分类系统。

SCOP2分类基于四个层次，从顶部到底部分别为：
类（Class）、折叠（Fold）、超家族（Super family）、家族（Family）

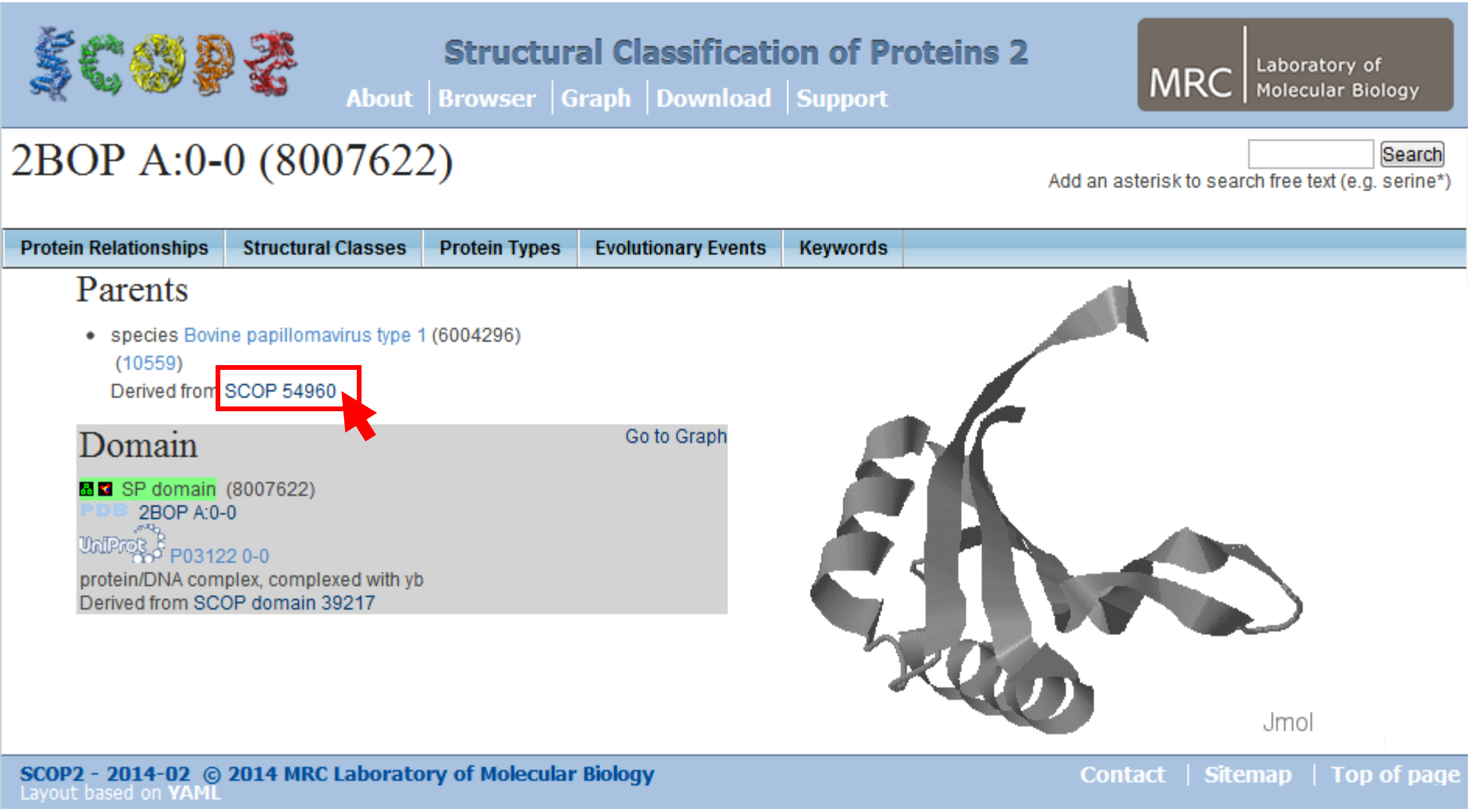
All alpha proteins
All beta proteins
Alpha and beta proteins (a+b)
Alpha and beta proteins (a/b)
Small proteins



蛋白质结构分类数据库　SCOP



蛋白质结构分类数据库　SCOP



蛋白质结构分类数据库 SCOP

Structural Classification of Proteins

Protein: Papillomavirus-1 E2 protein from Bovine papillomavirus type 1 [TaxId: 10559]

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a+b) [53931]
   Mainly antiparallel beta sheets (segregated alpha and beta regions)
3. Fold: Ferredoxin-like [54861]
   alpha+beta sandwich with antiparallel beta-sheet; (beta-alpha-beta)x2
4. Superfamily: Viral DNA-binding domain [54957]
   Superfamily
5. Family: Viral DNA-binding domain [54958]
6. Protein: Papillomavirus-1 E2 protein [54959]
   forms dimers with subunit beta-sheets making (8,12) barrel
7. Species: Bovine papillomavirus type 1 [TaxId: 10559] [54960]

PDB Entry Domains:

1. 2bop
   protein/DNA complex; complexed with yb
   1. chain a [39217] L

途径数据库

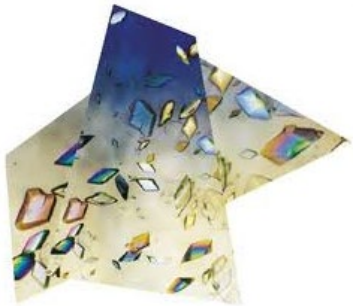
KEGG途径数据库是存储的是人工绘制的途径图谱，包括了目前已知的所有分子相互作用网络和生物反应网络：总途径图，代谢途径图，遗传信息加工图，环境信息加工图，细胞过程图，机体系统图，人类疾病相关途径图，药物开发图。

Go to: http://www.genome.jp/kegg/pathway.html

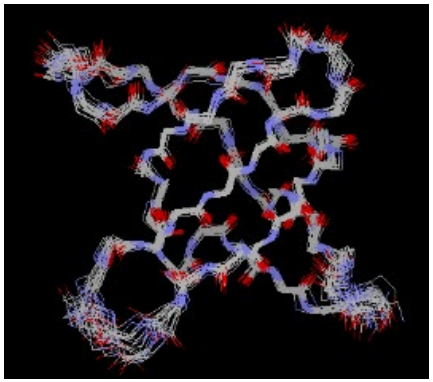
蛋白质结构 --- 三级结构测定

如何获得蛋白质的三级结构

1. 实验方法



X-ray Crystallography



Nuclear Magnetic Resonance (NMR) ~200AA

蛋白质结构 --- 三级结构测定

如何获得蛋白质的三级结构

1. 实验方法

SDU Experts in X-ray Crystallography

Prof. SUN Jinpeng Ph.D.

sunjinpeng@sdu.edu.cn

Institute of Biochemistry and Molecular Biology

School of Medicine, SDU

Prof. GU Lichuan Ph.D.

lcgu@sdu.edu.cn

State Key Laboratory of Microbial Technology

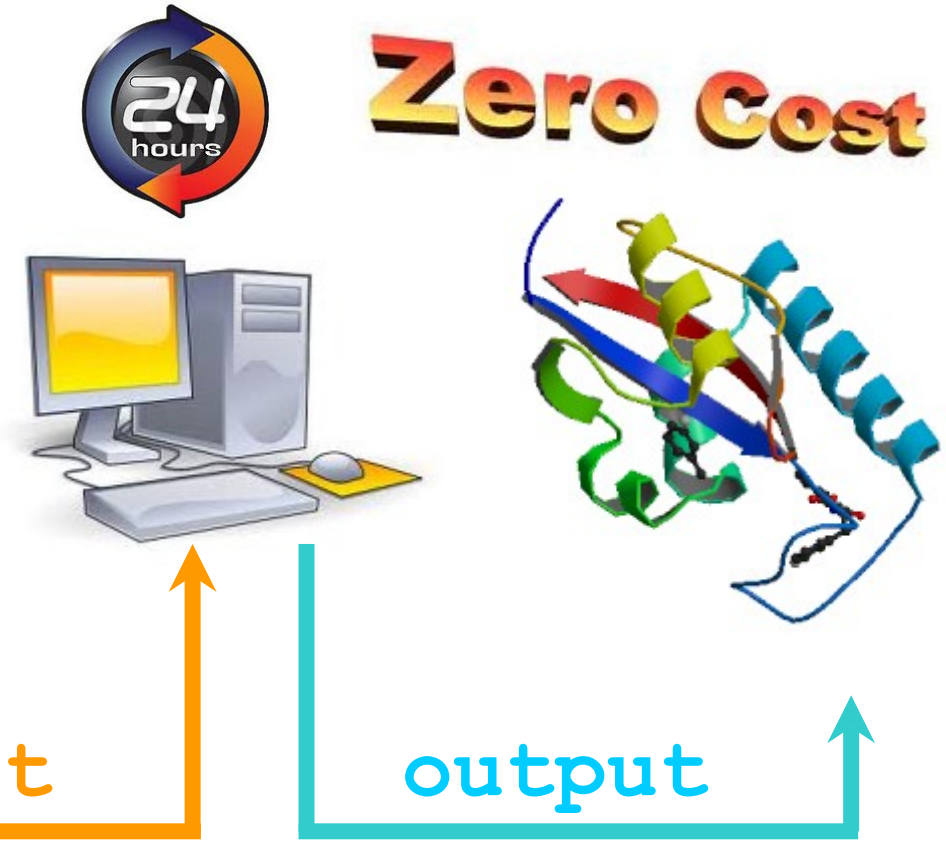
School of Life Sciences, SDU

蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法

MEAKIVKVLDSSRCEDGFGKKRKRAASYAAYVTGV
SCAKLQNVPPPNGQCQIPDKRRRLEGENKLSAYE
NRSGKALVRYYTYFKKTGIAKRVMMYENGEWNDL
PEHVICAIQNELEEKSAAIEFKLCGHSFILDFLHMQR
LDMETGAKTPLAWIDNAGKCFFPEIYESDERTNYC
HHKCVEDPKQNAPHDIKLRLEIDVNGGETPRLNLE
ECSDESGDNMMDDVPLAQRSSNEHYDEATEDSC
SRKLEAAVSKWDETDAIVVSGAKLTGSEVLDKDAV
KKMFAVGTASLGHVPVLDVGRFSSEIAEARLALFQ
KQVEITKKHRGDANVRYAWLPAKREVLSAVMMQG
LGVGGAFIRKSIYGVGIHLTAADCPYFSARYCDVDE
NGVRYMVLCRVIMGNMELLRGDKAQFFSGGEEYD
NGVDDIESPKNYIVWNINMNTHIFPEFVVRFKLSNL
PNAEGNLIAKRDNSGVTLEGPKDLPPQLESNQGAR
GSGSANSVGSSTTRPKSPWMPFPTLFAAISHKVAE
NDMLLINADYQQLRDKKMTRAEFVRKLRVIVGDDL
LRSTITTLQNQPKSKEIPGSIRDHEEGAGGL

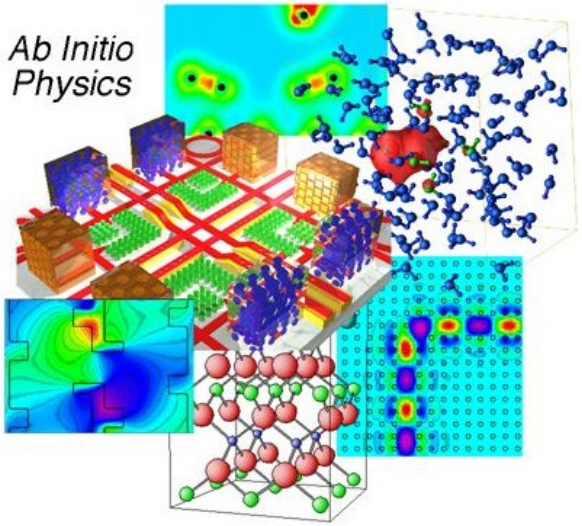


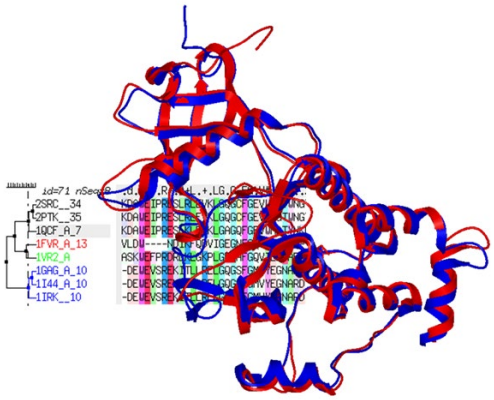
蛋白质结构 --- 三级结构预测

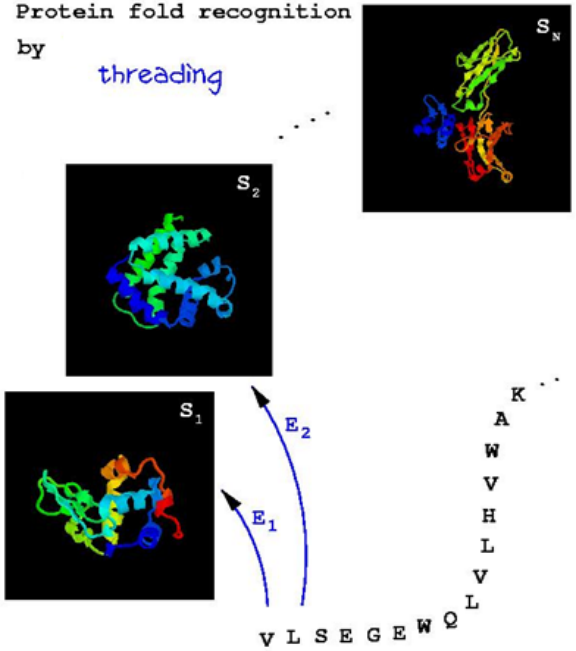
如何获得蛋白质的三级结构

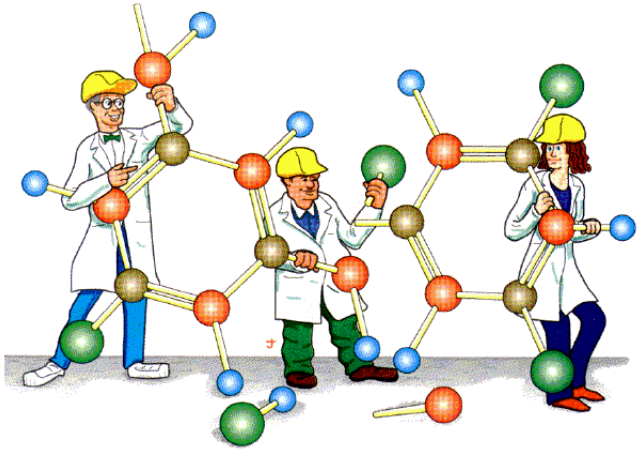
2. 计算方法

- 从头计算法
- 同源建模法
- 穿线法
- 综合法







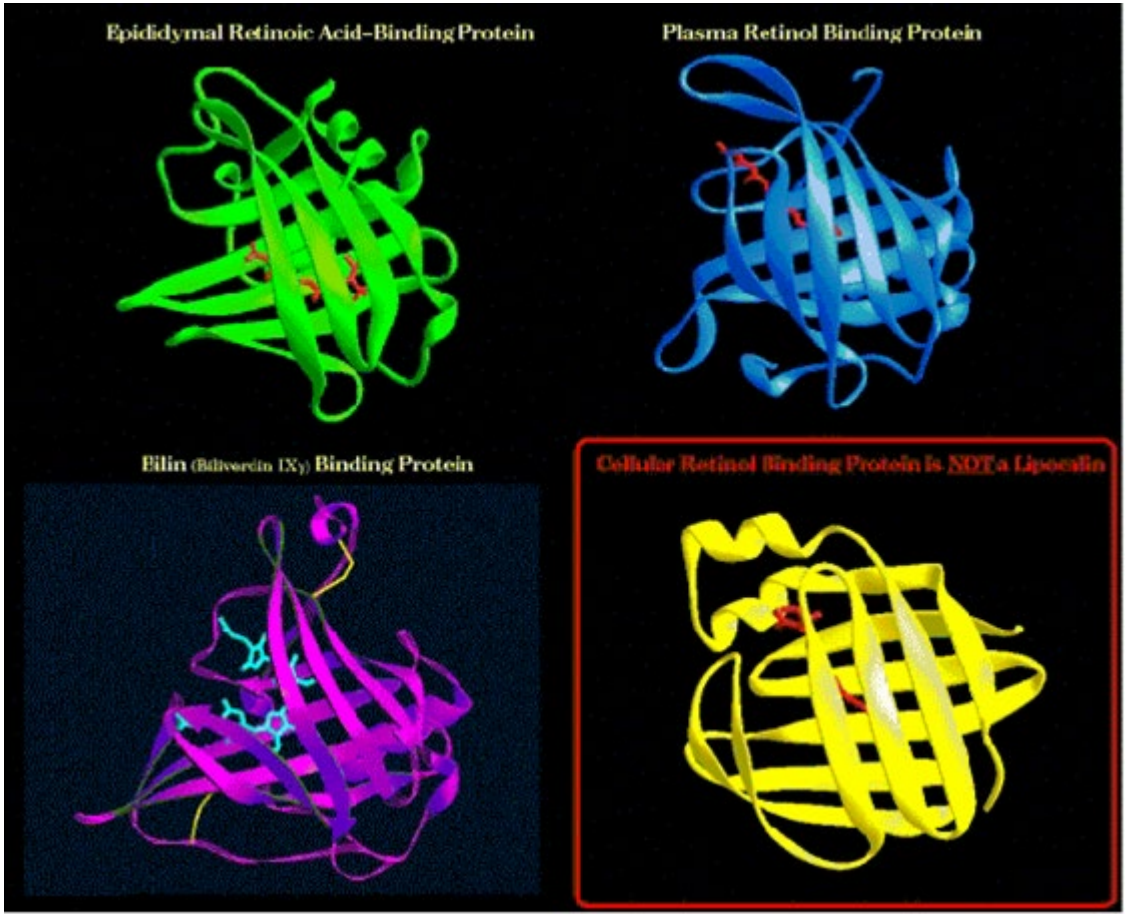


蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

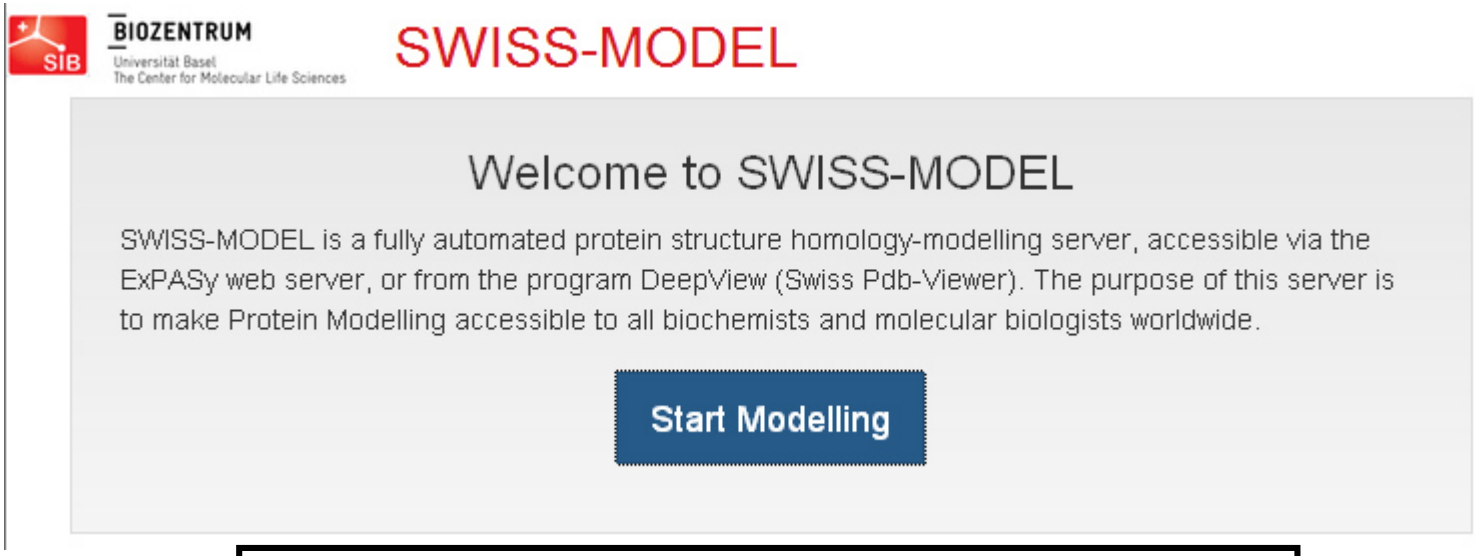
2. 计算方法 --- 同源建模法

同源建模法：相似的氨基酸序列对应着相似的蛋白质结构。



自动同源建模法: SWISS-MODEL

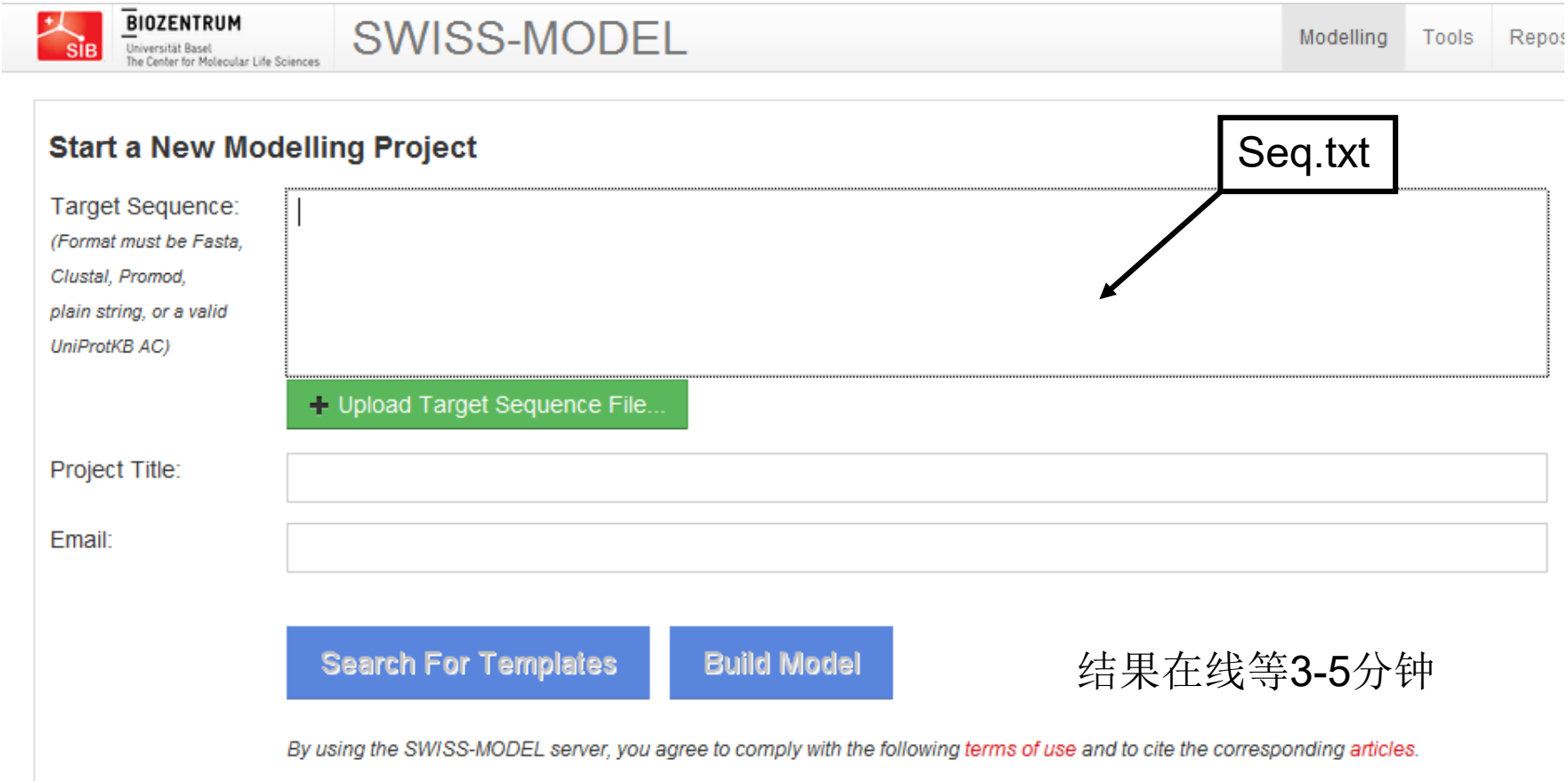
SWISS-MODEL http://swissmodel.expasy.org



SWISS-MODEL做出来的结果可以在SCI刊物论文上作为一项结果来发表，但前提条件是模板与目标的序列identity要足够高，至少高于30%，因为SWISS-MODEL是单纯的同源建模法，基本属于结构预测软件里的“傻瓜机”。

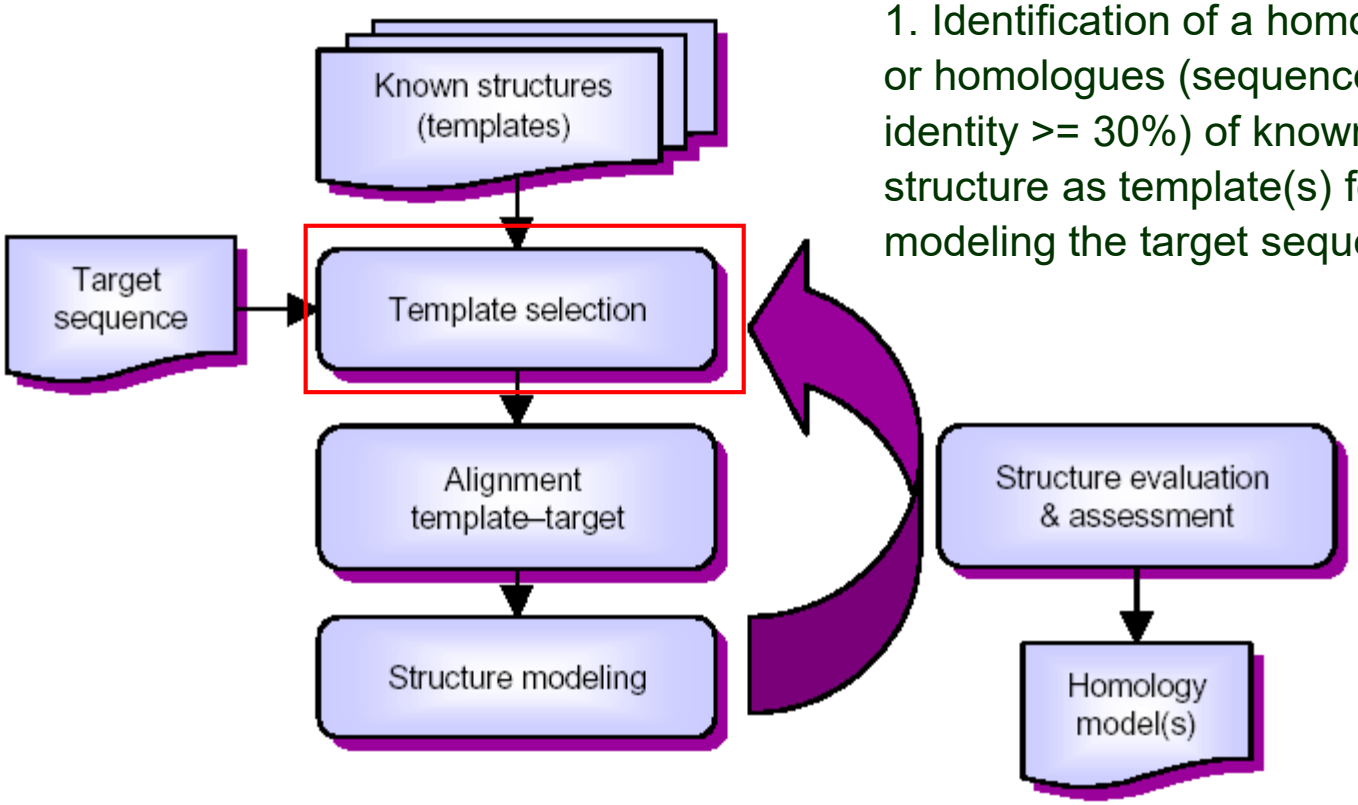
自动同源建模法: SWISS-MODEL

SWISS-MODEL http://swissmodel.expasy.org



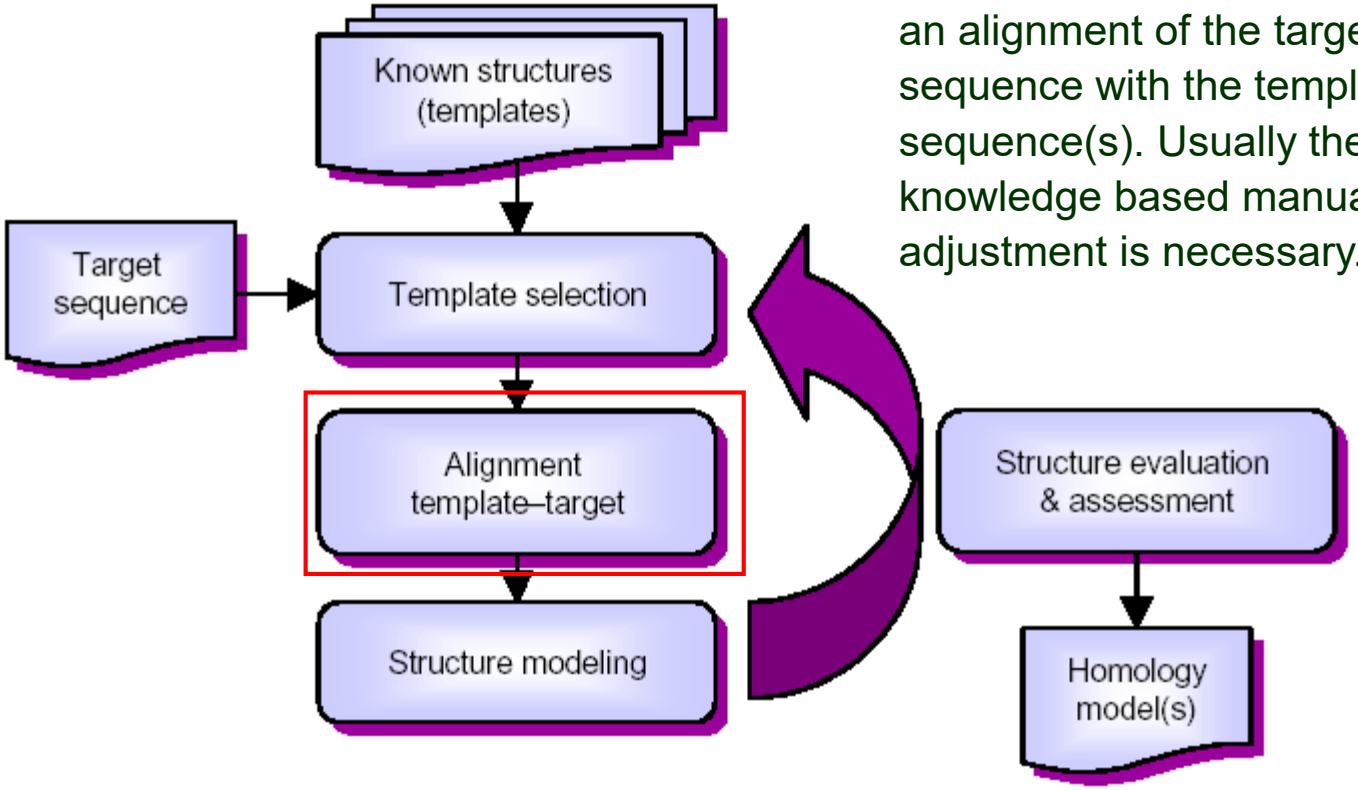
结果在线等3-5分钟

半手动同源建模法



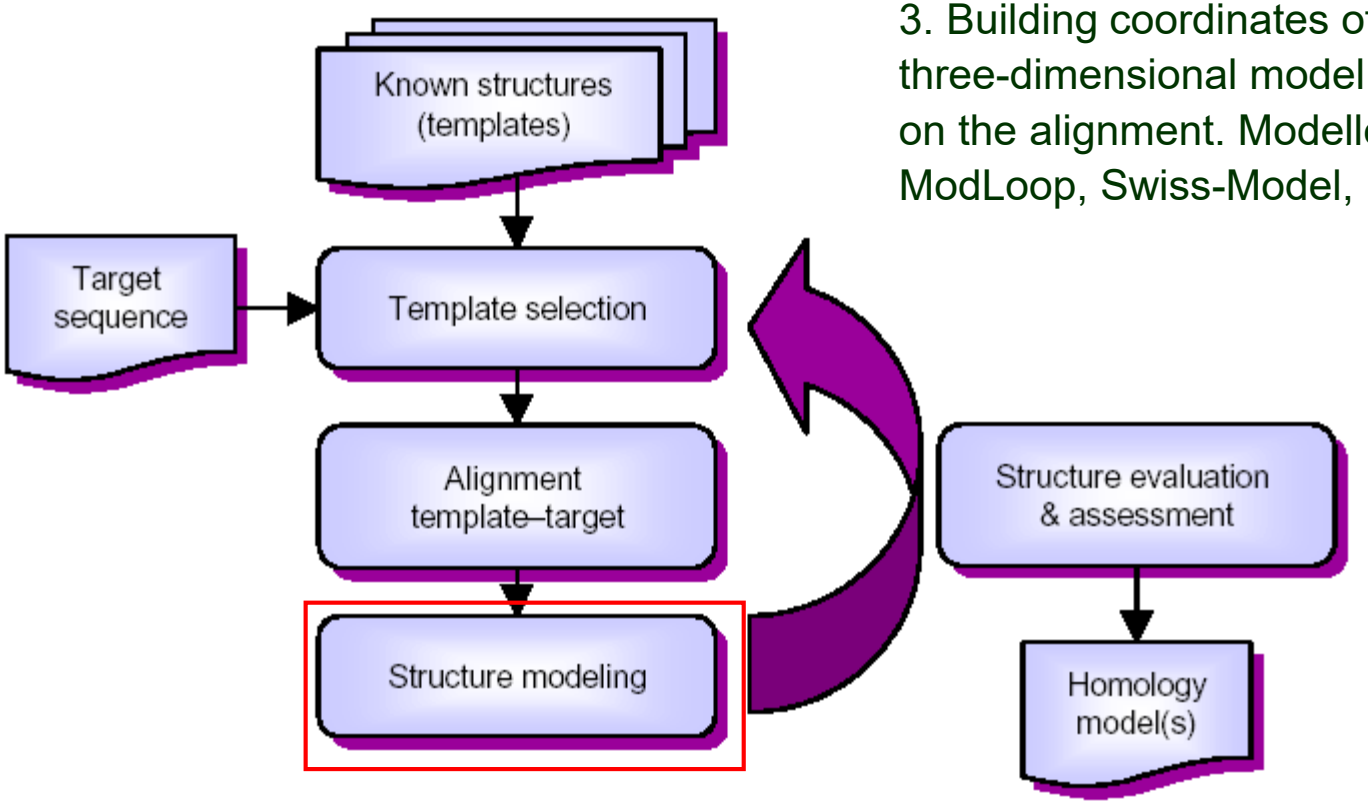
1. Identification of a homologue or homologues (sequence identity >= 30%) of known structure as template(s) for modeling the target sequence.

半手动同源建模法



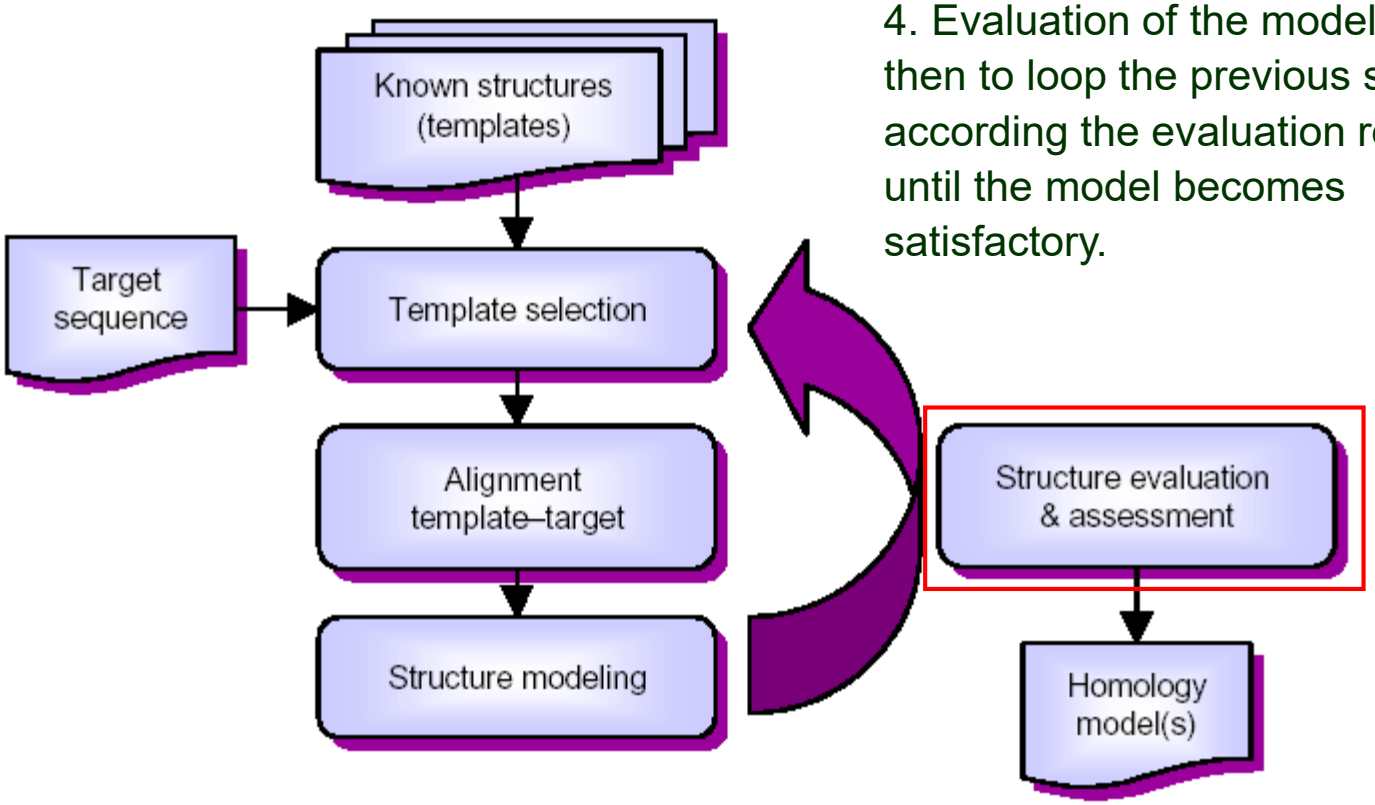
2. Generating and improving an alignment of the target sequence with the template sequence(s). Usually the knowledge based manual adjustment is necessary.

半手动同源建模法



3. Building coordinates of the three-dimensional model based on the alignment. Modeller, ModLoop, Swiss-Model, …

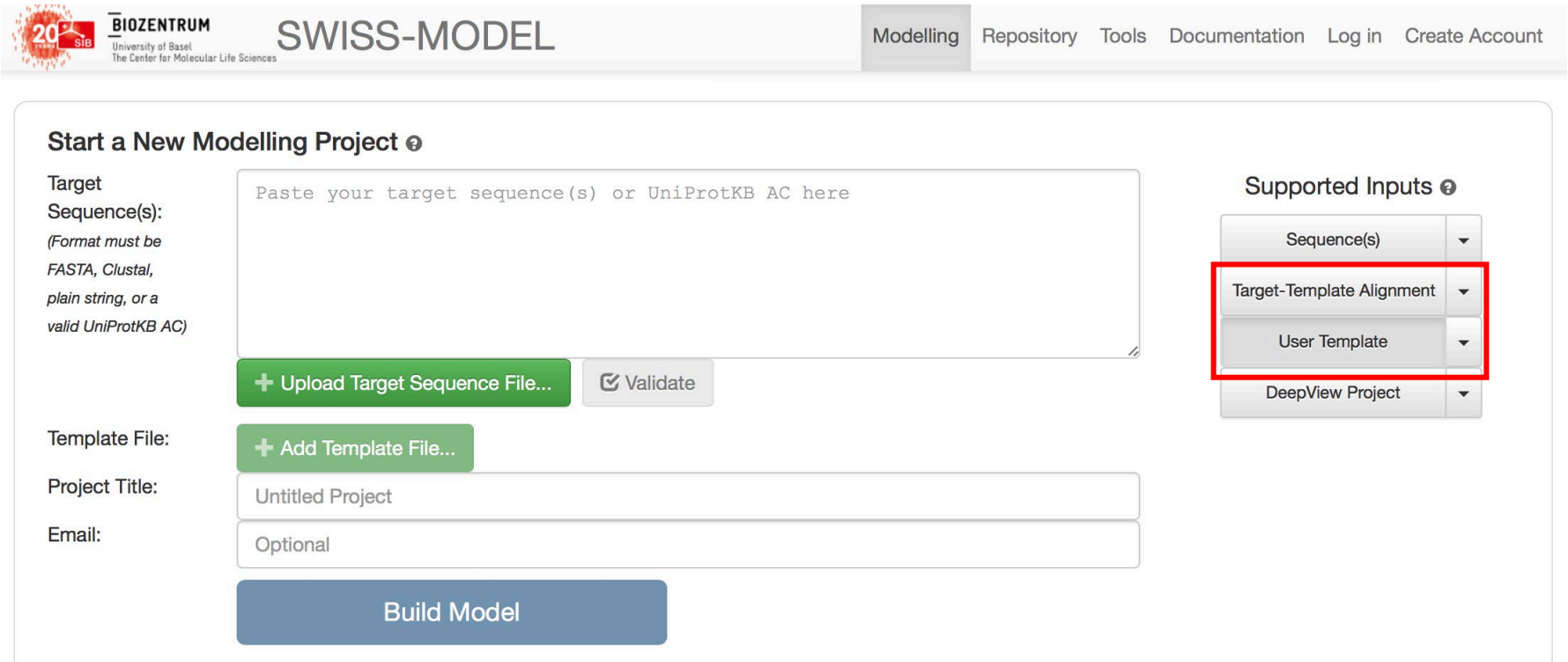
半手动同源建模法



4. Evaluation of the model and then to loop the previous steps according the evaluation results until the model becomes satisfactory.

半手动同源建模法: SWISS-MODEL

SWISS-MODEL http://swissmodel.expasy.org

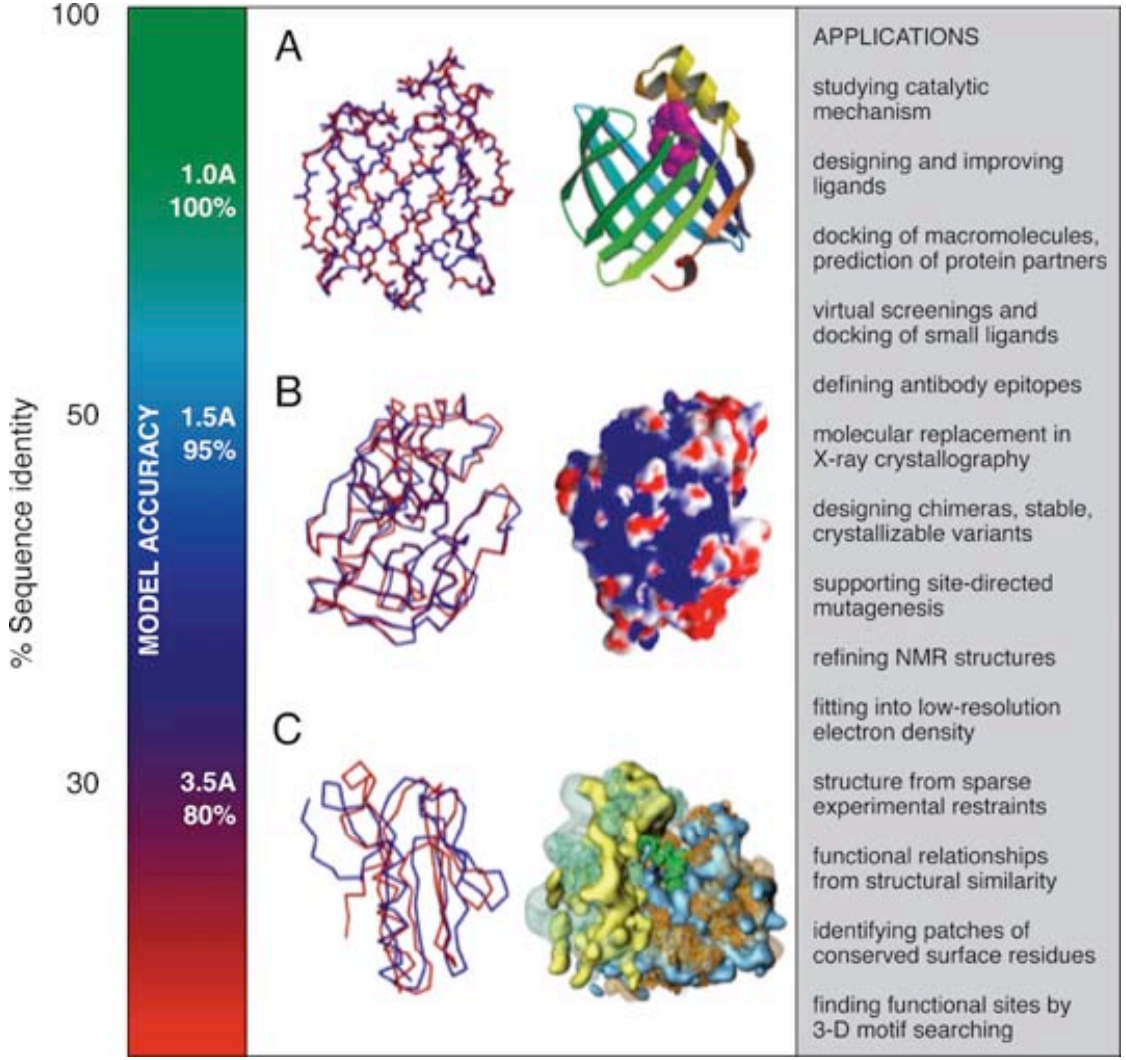


蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法 --- 同源建模法

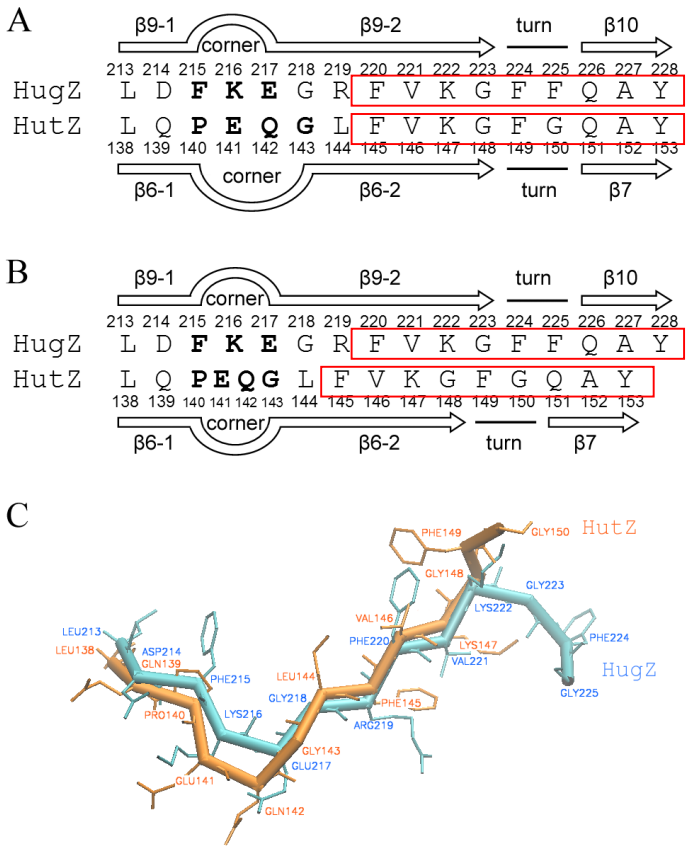
如果目标序列与模板序列相似度极高，那么同源建模法是最准确的方法。



蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法 --- 同源建模法



特例情况，虽然序列一致度达到很高水平，但是结构却并不相同。

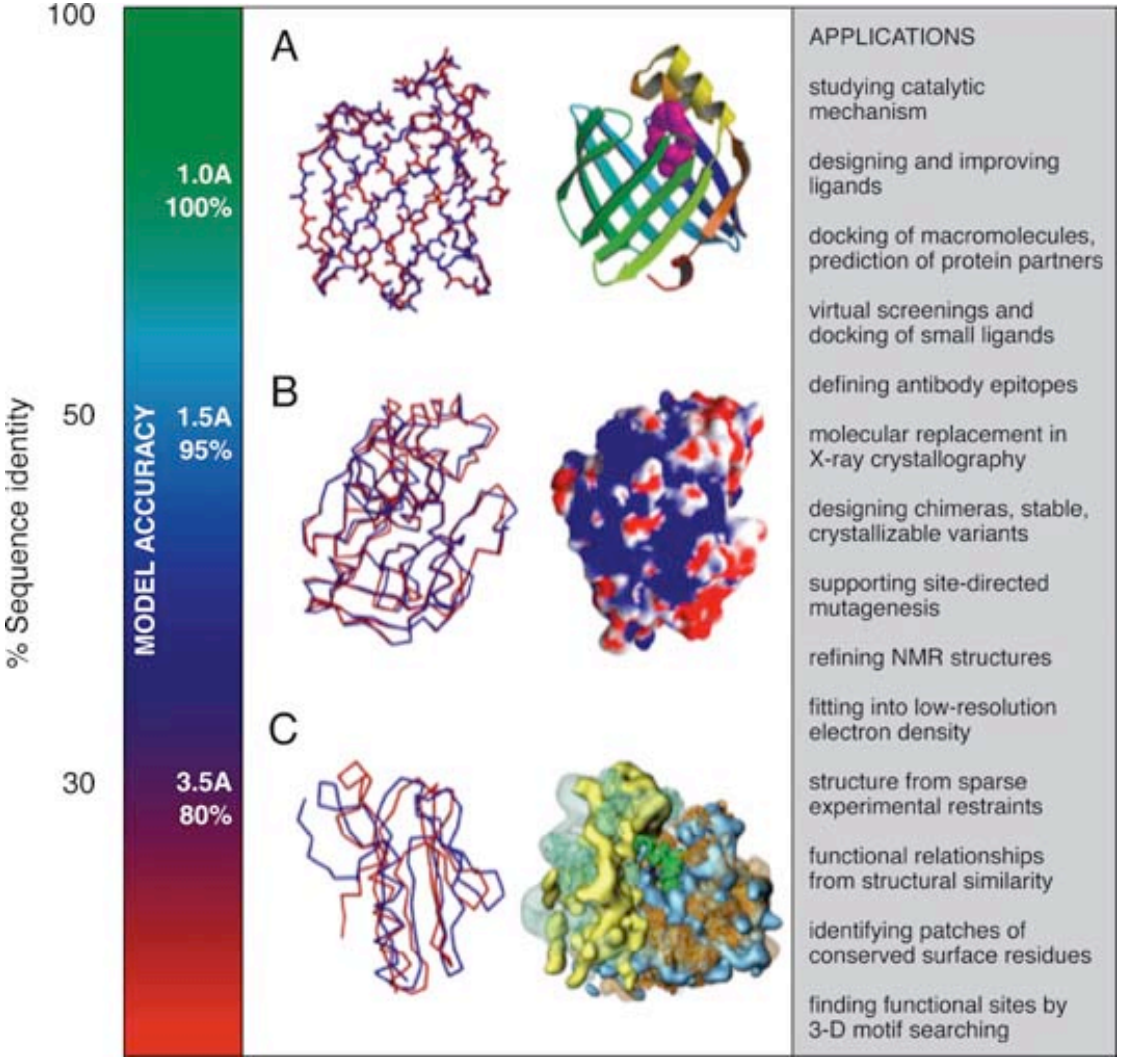
[*BMC Struct Biol, 2012, 12:23*]

蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法 --- 同源建模法

如果目标序列与模板序列之间的一致度 < 30%，那么同源建模法是不适用的。

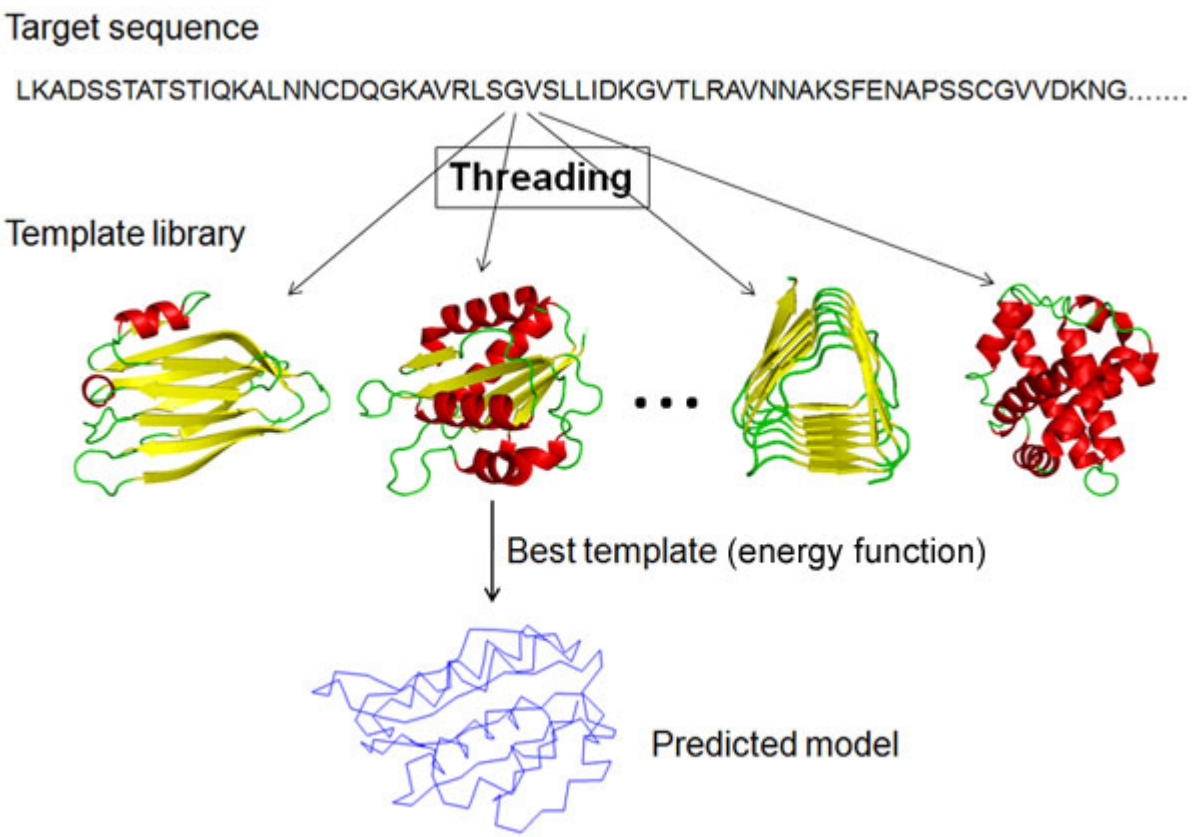


蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法 --- 穿线法

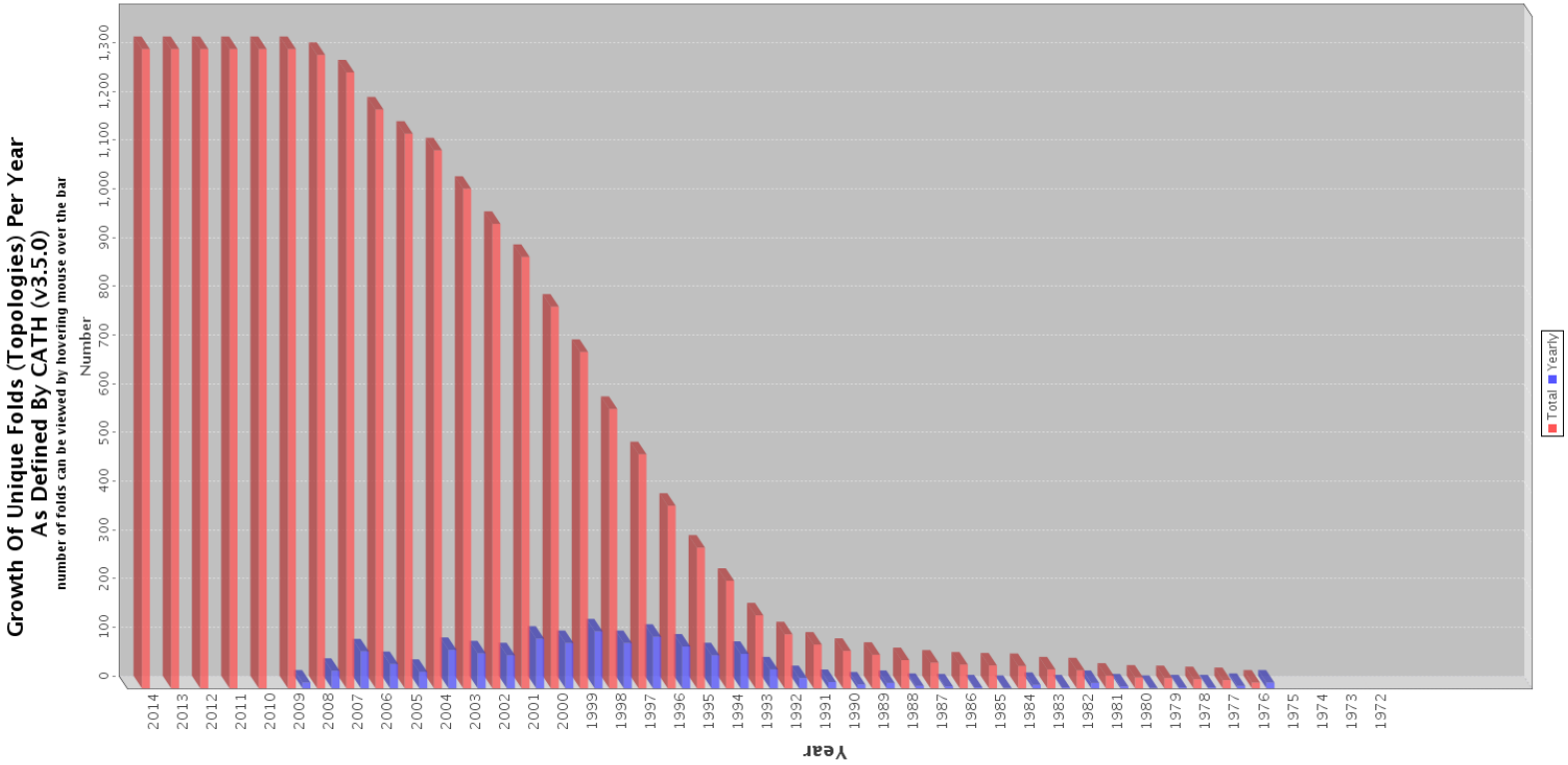
穿线法：不相似的氨基酸序列也可能对应着相似的蛋白质结构。



蛋白质结构 --- 三级结构预测

如何获得蛋白质的三级结构

2. 计算方法 --- 穿线法

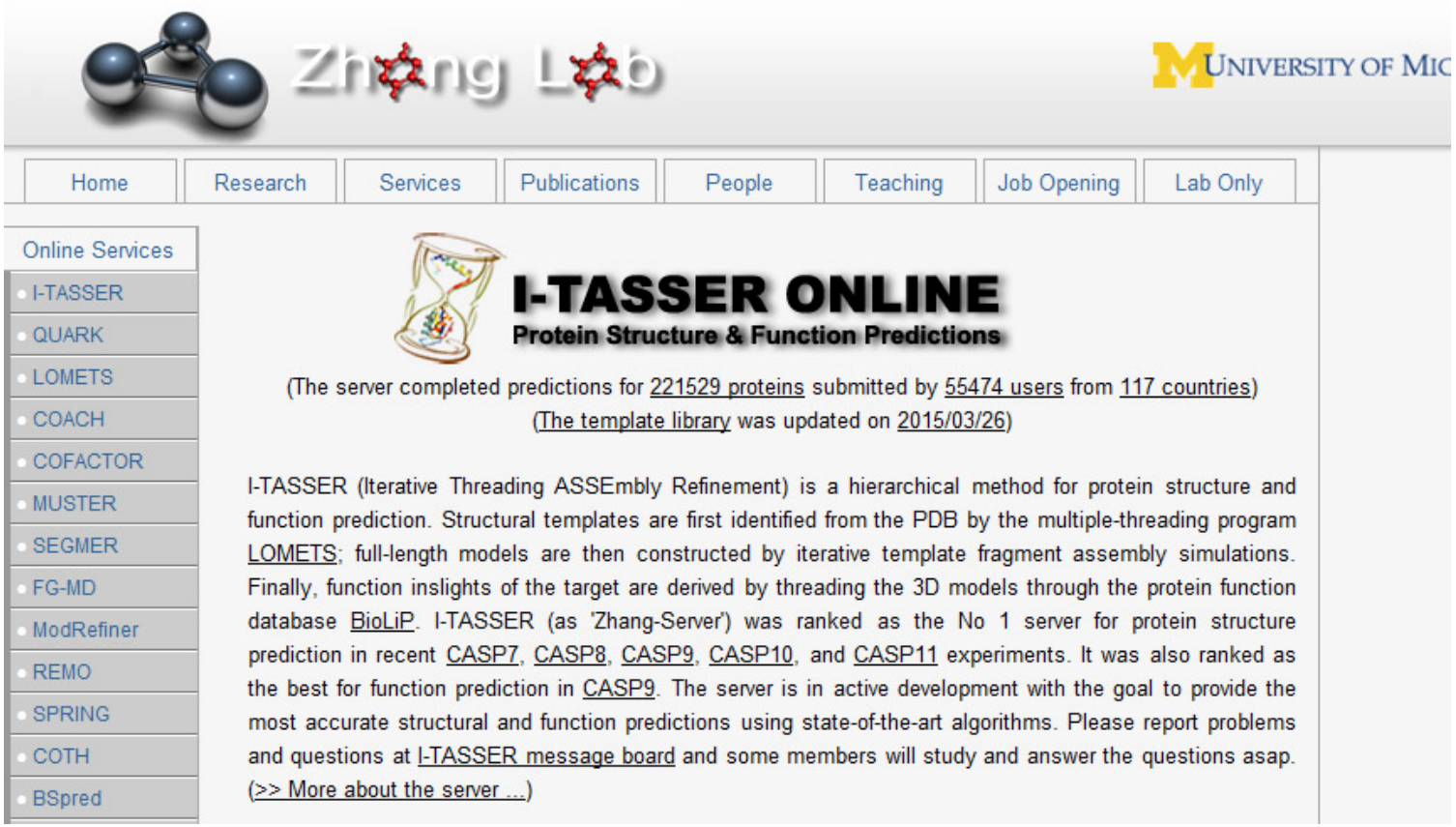


已知结构的蛋白质十几万，不同的结构拓扑 1313。

自动穿线法: I-TASSER

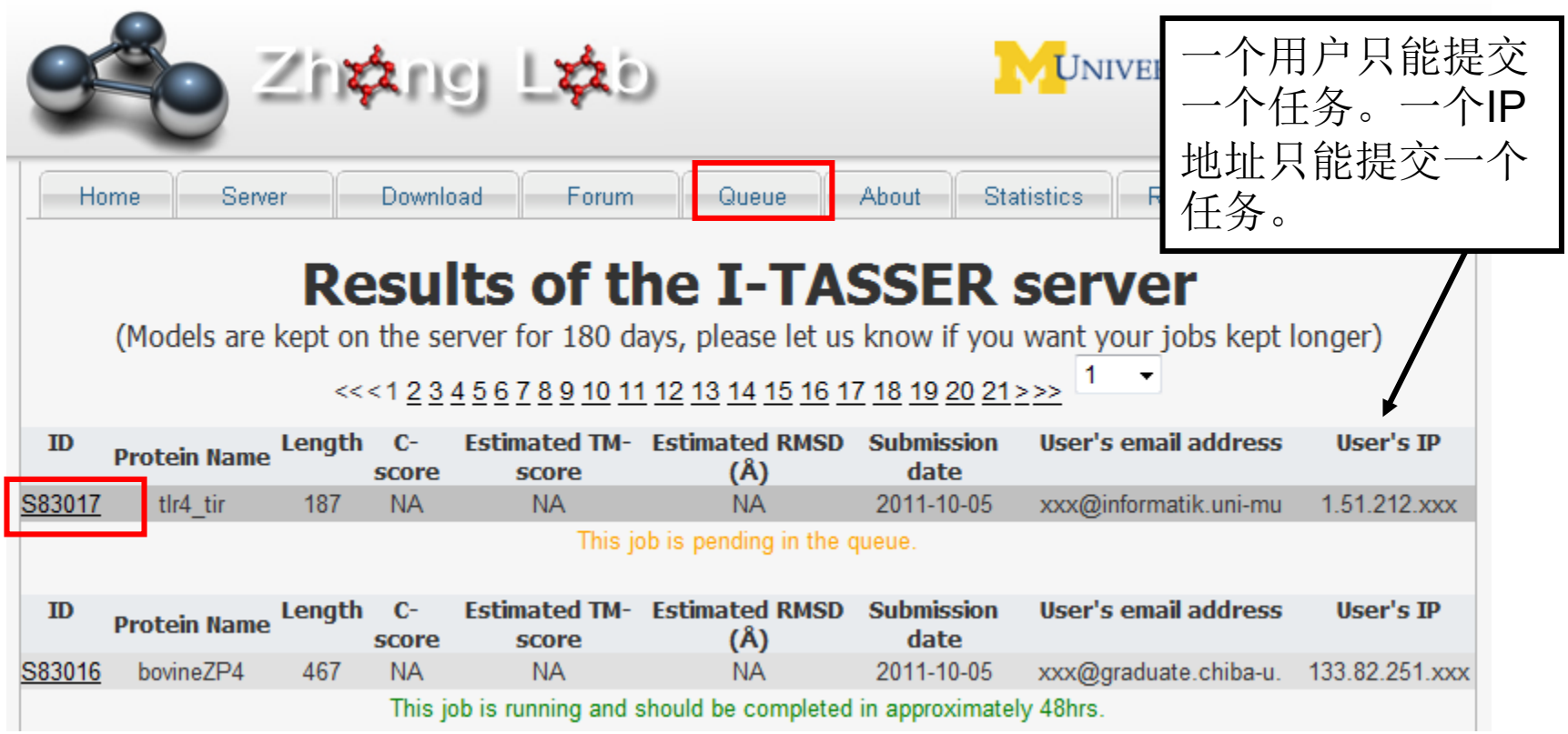
I-TASSER http://zhanglab.ccmb.med.umich.edu/I-TASSER

I-TASSER: 在线的蛋白质结构预测服务器，在近7届蛋白质结构预测比赛（CASP7/8/9/10/11/12/13）中皆排名第一。作者为美国密歇根大学的张阳教授。什么都不用自己操心，输入EMAIL和目标序列点RUN。



自动穿线法: I-TASSER

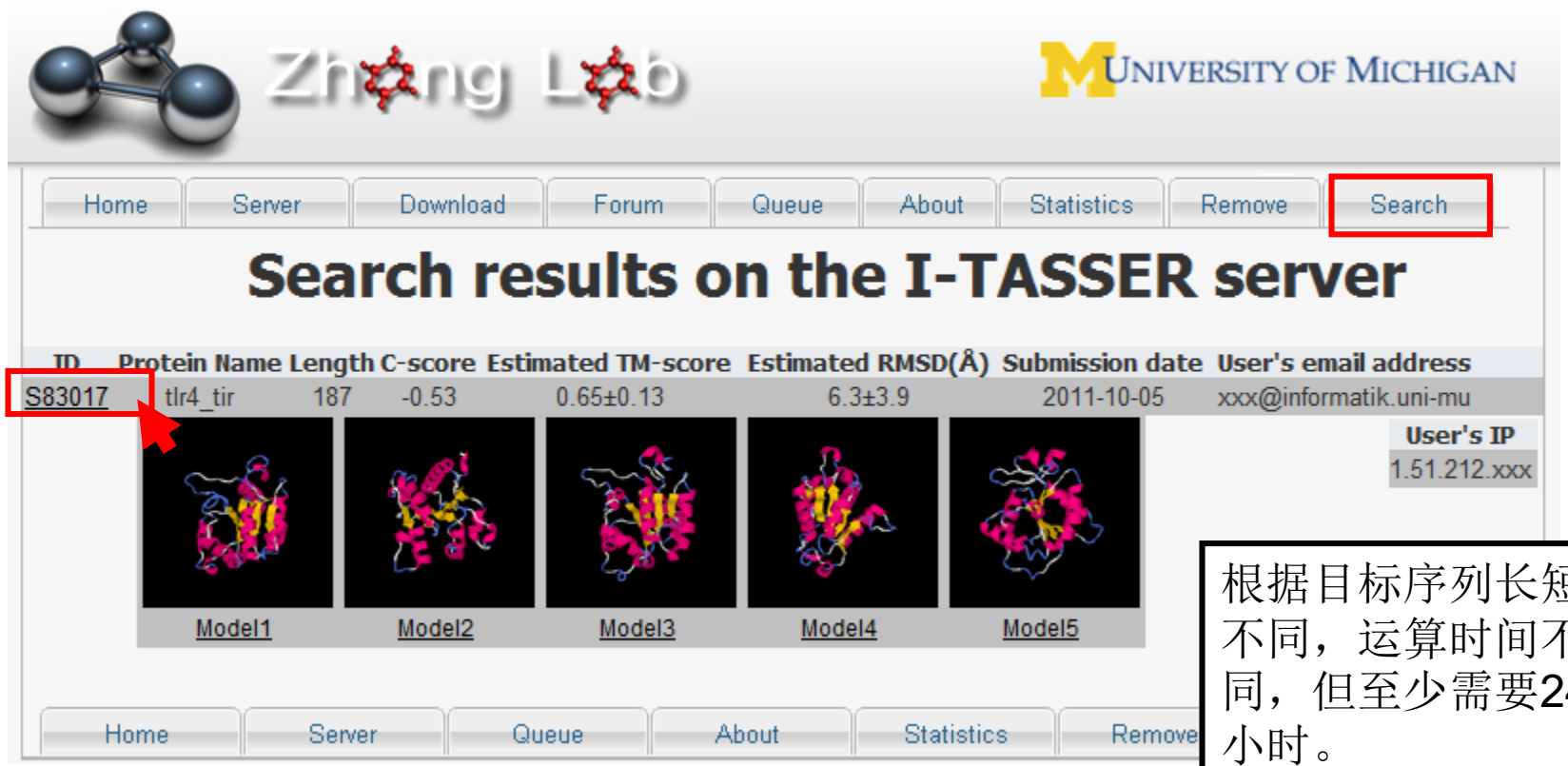
I-TASSER http://zhanglab.ccmb.med.umich.edu/I-TASSER



一个用户只能提交一个任务。一个IP地址只能提交一个任务。

自动穿线法: I-TASSER

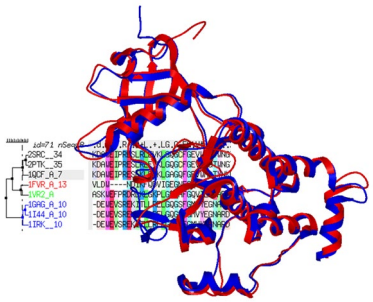
I-TASSER http://zhanglab.ccmb.med.umich.edu/I-TASSER

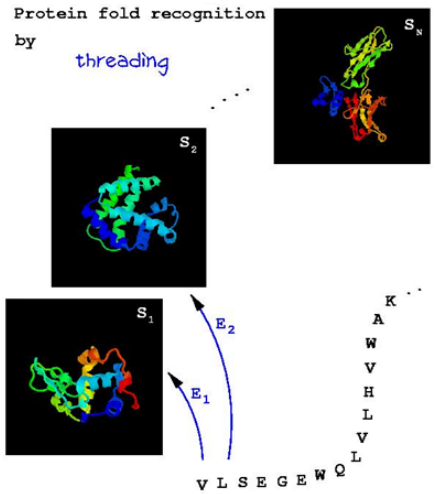


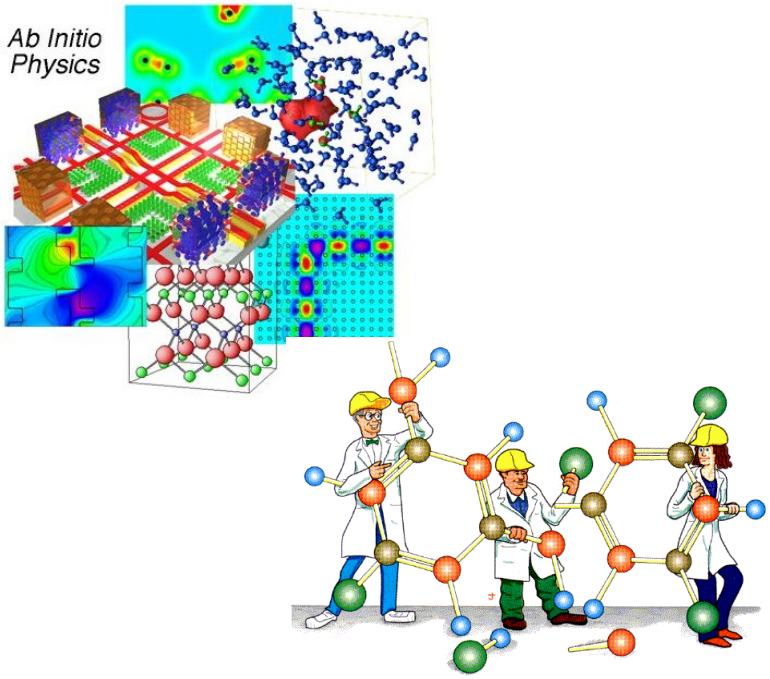
根据目标序列长短不同，运算时间不同，但至少需要24小时。

蛋白质结构预测

- 结构生物学中结合先验知识的计算方法
- **科学家们数十年的努力，只覆盖了人类蛋白质序列中17%的氨基酸残基。**







AlphaFold 2

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › news › article

NEWS | 22 July 2021

DeepMind's AI predicts structures for a vast trove of proteins

AlphaFold neural network produced a 'totally transformative' database of more than 350,000 structures from *Homo sapiens* and 20 model organisms.

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open Access | Published: 22 July 2021

Highly accurate protein structure prediction for the human proteome

Kathryn Tunyasuvunakool, Jonas Adler, [...]Demis Hassabis

Nature **596**, 590–596 (2021) | Cite this article

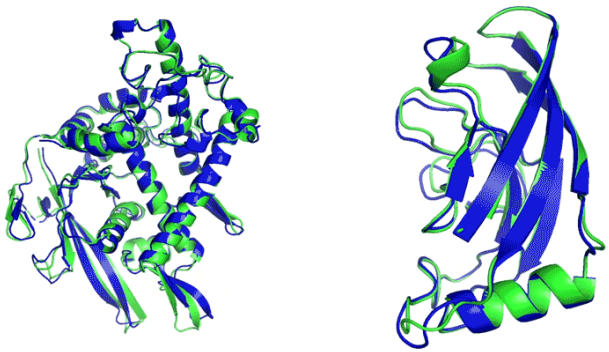
146k Accesses | **8** Citations | **1354** Altmetric | Metrics

2021年7月，98.5%的人类蛋白质结构被AlphaFold2预测出来。除了人类蛋白质组，数据集中还包括大肠杆菌、果蝇、小鼠等20个具有科研常用生物的蛋白质组数据，总计超过35万个蛋白质的结构。最重要的是，这些全都免费开放，交给欧洲生物信息学研究所托管。

DeepMind创始人哈撒比斯在官网发布题为《把AlphaFold的力量交到全世界手中》的文章，同时也在推特上表达了他抑制不住地兴奋：

这是我一生中梦寐以求的日子，也是创办Deepmind的初衷：用AI推进科学发展并造福人类。

AlphaFold 2



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDT
(adhesin tip)

Experimental result
Computational prediction

借鉴了AI研究中最近新兴起的Transformer架构。Transformer使用注意力机制兴起于NLP领域，用于处理一连串的文本序列。
而氨基酸序列正是和文本类似的数据结构，AlphaFold2利用多序列比对，把蛋白质的结构和生物信息整合到了深度学习算法中。

Alphafold2准确性：预测结构和蛋白质真实结构之间只差一个原子的宽度，真正解决了蛋白质折叠的问题。

AlphaFold 2

欧洲分子生物学实验室（EMBL）的负责人Edith Heard说：
“我们相信这对理解生命体是如何运作有着变革性的影响。”

哥伦比亚大学的计算生物学家Mohammed AlQuraishi表示，此前蛋白质结构预测领域总是要花费大量时间在一些基础工作上，浪费了学者的很多精力，现在他们可以更加专注于对蛋白质结构的研究了。

一些与DeepMind展开合作的研究团队，已经通过AlphaFold加速了研究进程：

DNDi（被忽视疾病药物开发组织）就表示，AlphaFold2推动了他们在热带疾病药物开发方面的研究。
朴茨茅斯大学酶创新中心（CEI）也表示，他们正在利用AlphaFold2开发一些新的酶，可以用来降解污染环境的一次性塑料。
科罗拉多大学波尔德分校的生化学家Marcelo Sousa则利用AlphaFold来制作蛋白质结构模型，开展一项关于抗生素的研究。
加州大学旧金山分校的一个团队则表示，AlphaFold2可以帮助他们更好理解SARS-CoV-2的生物学机制。

AlphaFold2论文地址：https://www.nature.com/articles/s41586-021-03828-1