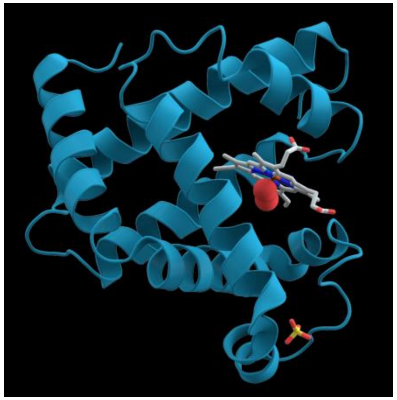
Al-Farabi Kazakh National University
Higher School of Medicine
Department of Fundamental Medicine



Proteomics and methods of proteome investigation.

Lecturer and creator: PhD Pinsky Ilya Vladimirovich

LEARNING OUTCOMES
As a result of the lesson you will be able to:

- 1. Give the definition to the terms “olygopeptides”, “polypeptides”, “proteins”, “proteome”, “proteomics”.
- 2. Explain and analyze the experimental methods of proteomics: MALDI-mass-spectrometry, ESI-mass-spectrometry, different types of chromatography, 2D- and 3D-PAGE, ELISA, nuclear magnetic resonance (NMR), X-ray diffraction and etc. What are the reasons for selecting any of these methods for specific proteins?
- 3. Describe the computational (bioinformatical) methods of protein research: amino acid sequence alignment, protein structure prediction, analysis of X-ray diffraction pattern and 3D-modelling of protein structure.
- 4. Characterize the Protein Data Bank (PDB), Uni-ProtKB/Swiss-Prot and other bioinformatical databases of protein information.

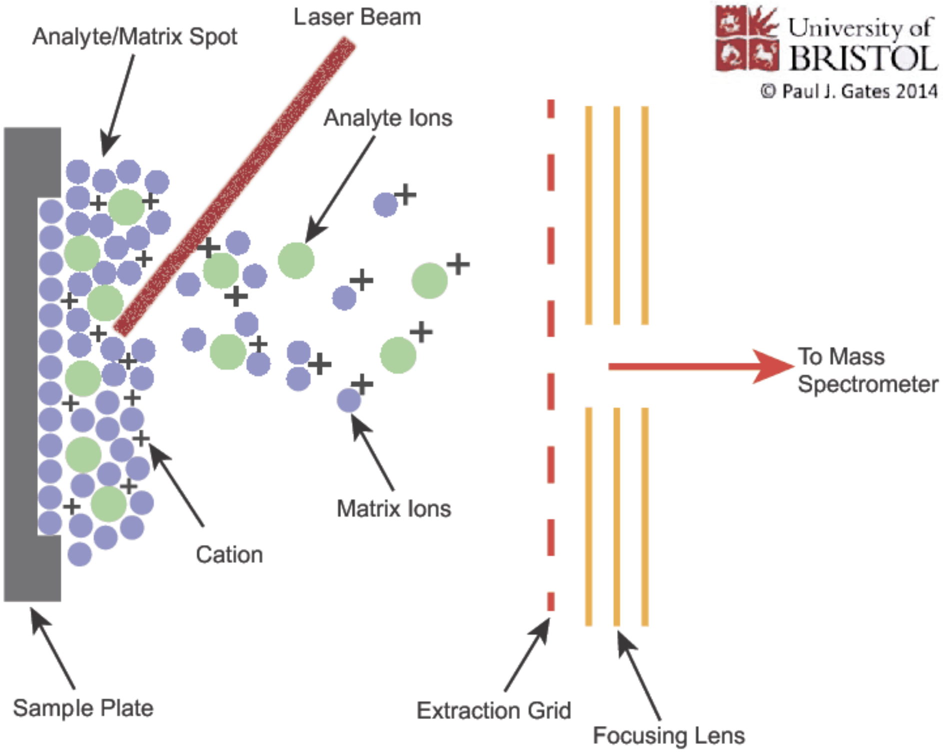
Definition

- **Proteomics** is the large-scale study of **proteins**.[1][2] Proteins are vital parts of living organisms, with many functions. The **proteome** is the entire set of proteins that is produced or modified by an organism or system. Proteomics has enabled the identification of ever increasing numbers of protein. This varies with time and distinct requirements, or stresses, that a cell or organism undergoes.[3] Proteomics is an interdisciplinary domain that has benefitted greatly from the genetic information of various genome projects, including the **Human Genome Project**.[4] It covers the exploration of proteomes from the overall level of protein composition, structure, and activity. It is an important component of functional genomics.

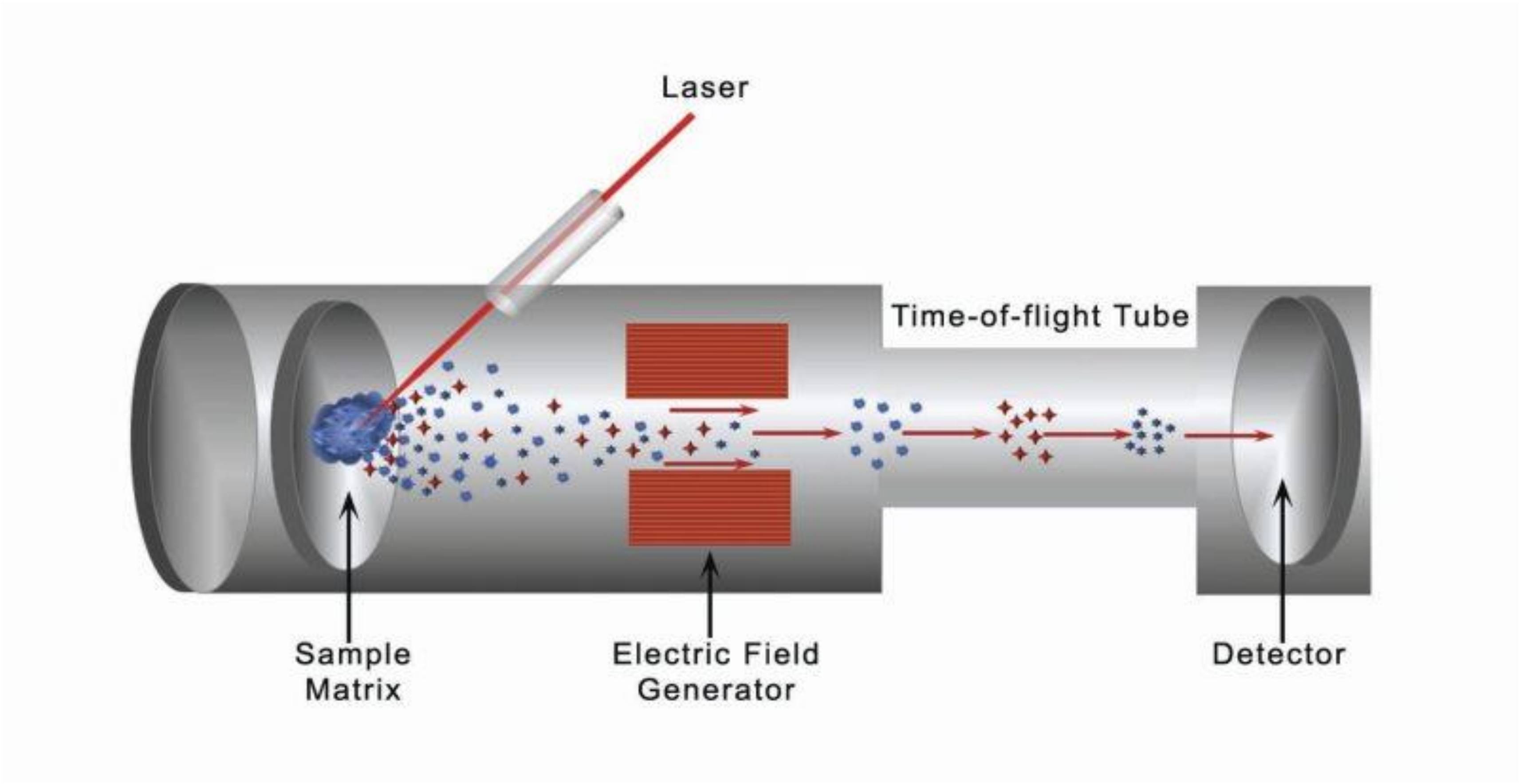
- Proteomics generally refers to the large-scale experimental analysis of proteins and proteomes, but often is used specifically to refer to **protein purification** and **mass spectrometry**.

Mass spectrometry

https://en.wikipedia.org/wiki/Proteomics#/media/File:Protein_pattern_analyzer.jpg

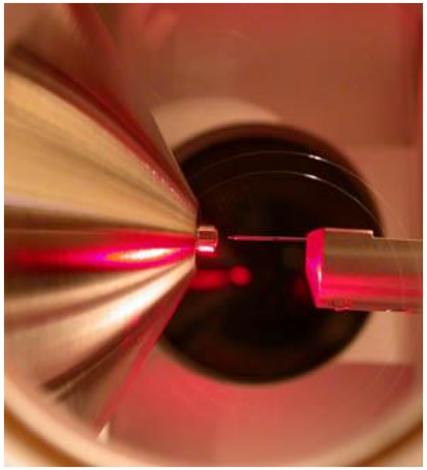


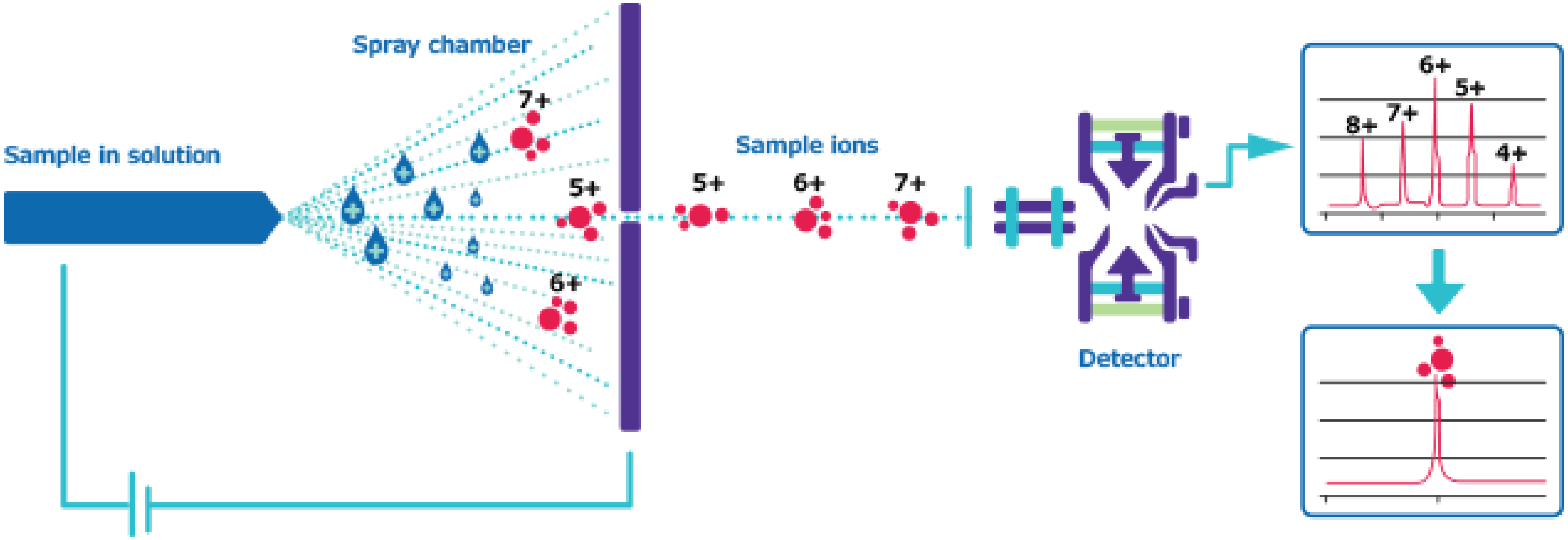
Matrix Assisted Laser Desorption/Ionization (MALDI)
newport.com



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.newport.com%2Fn%2Fmaldi&psig=AOvVaw2pBY9V8nhY1AulhJln3hjh&ust=1615810269525000&source=images&cd=vfe&ved=2ahUKEwj2rbG64K_vAhWymIsKHaHPAIEQr4kDegUIARDaAQ

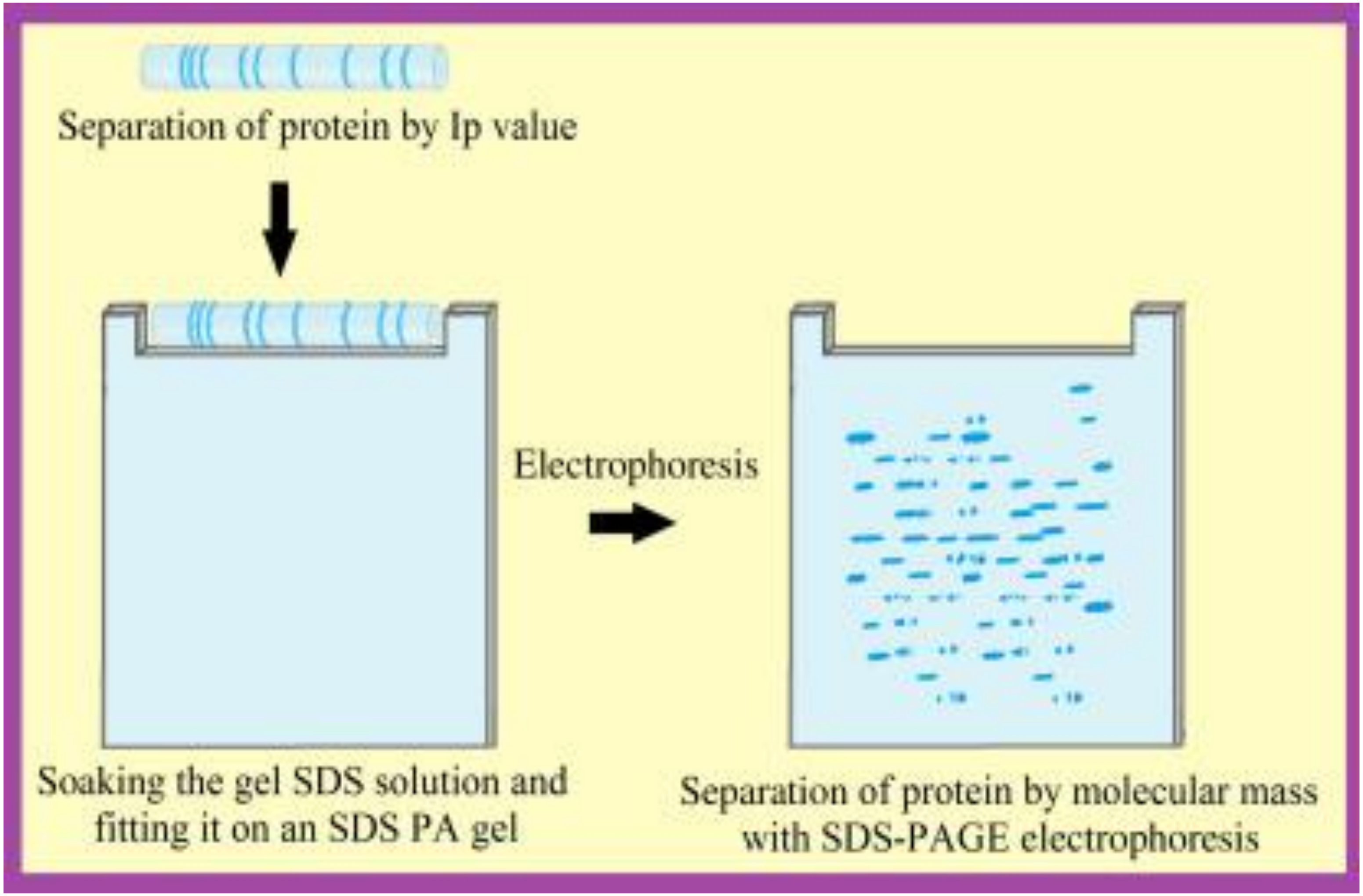
https://en.wikipedia.org/wiki/Electrospray_ionization#/media/File:NanoESIFT.jpg





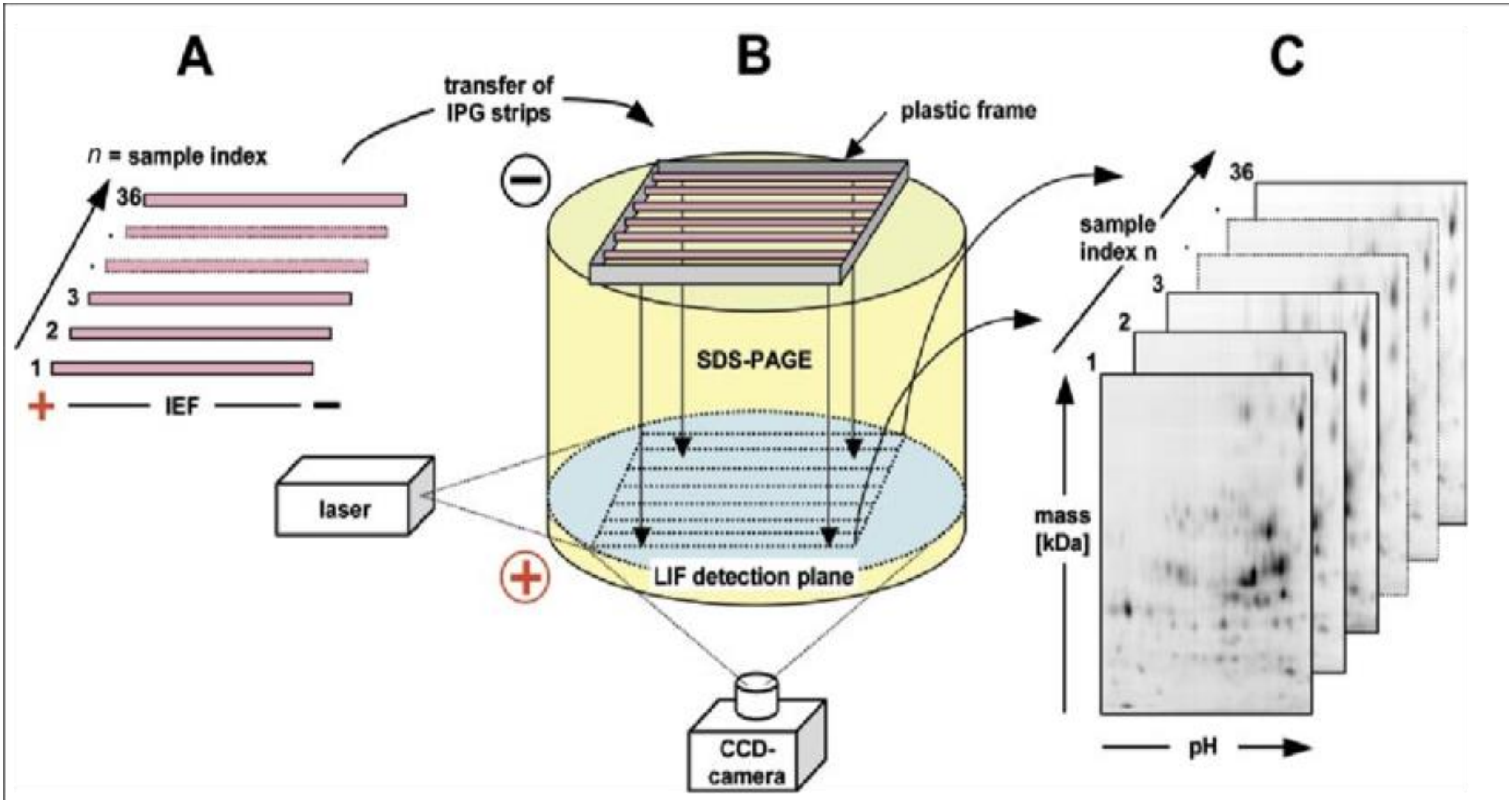
Electrospray Ionization Mass Spectrometry (ESI MS)
sigmaaldrich.com

2D PAGE electrophoresis

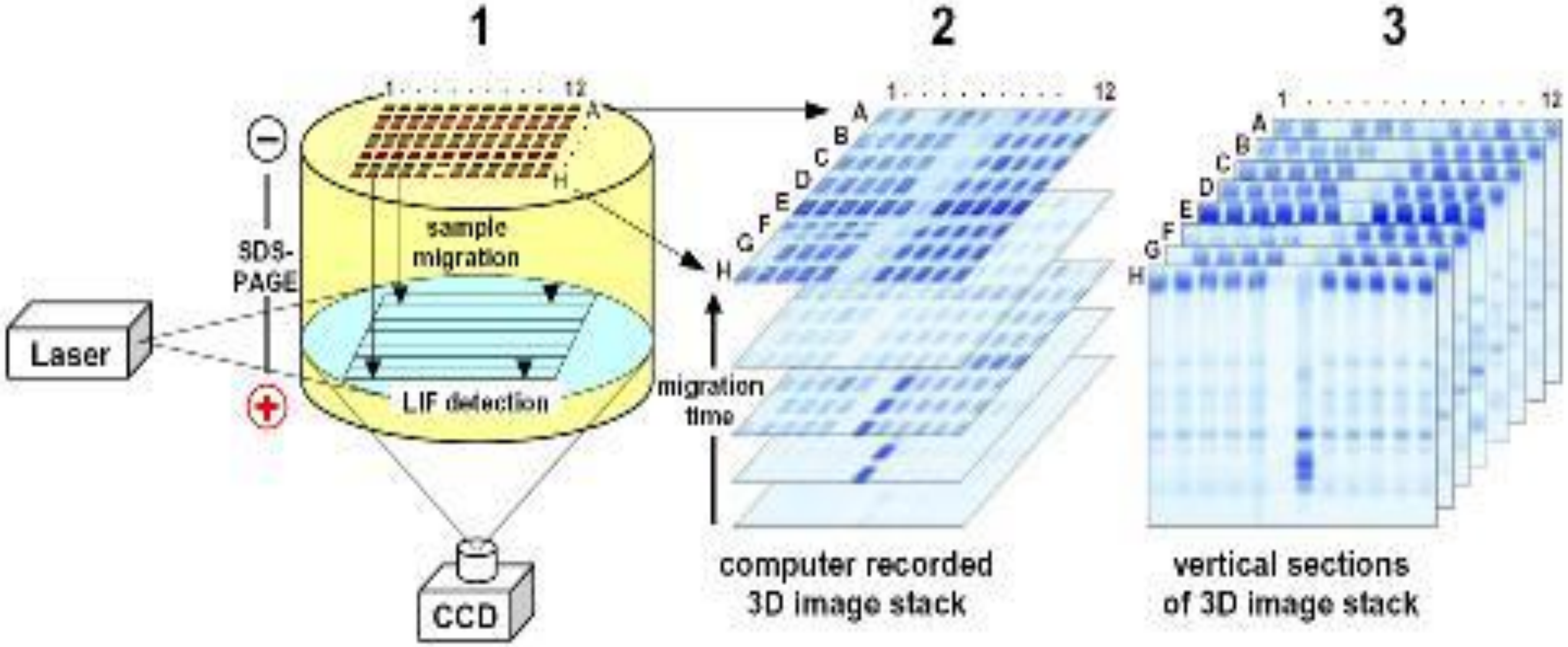


sciencedirect.com

3D PAGE electrophoresis

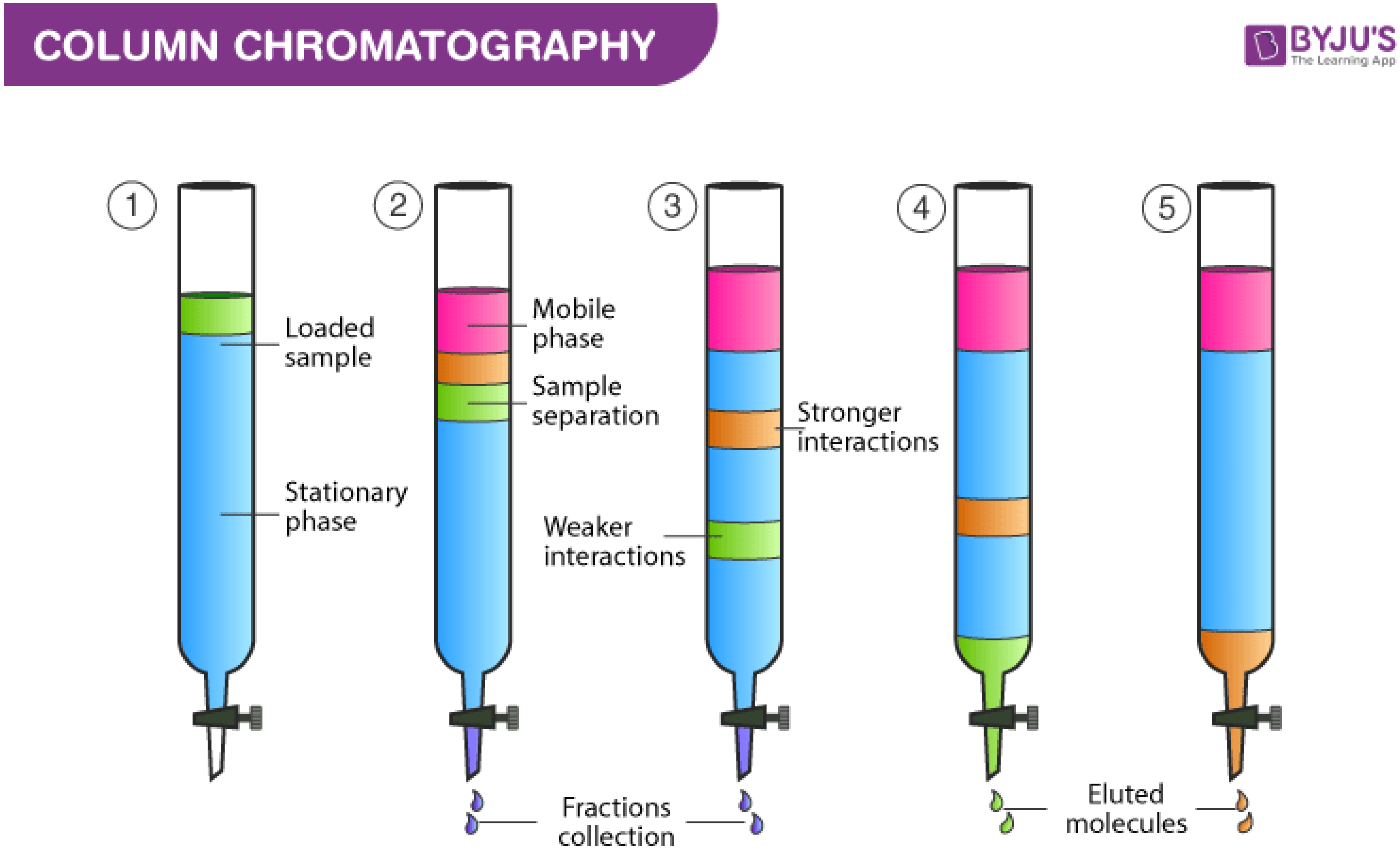


researchgate.net

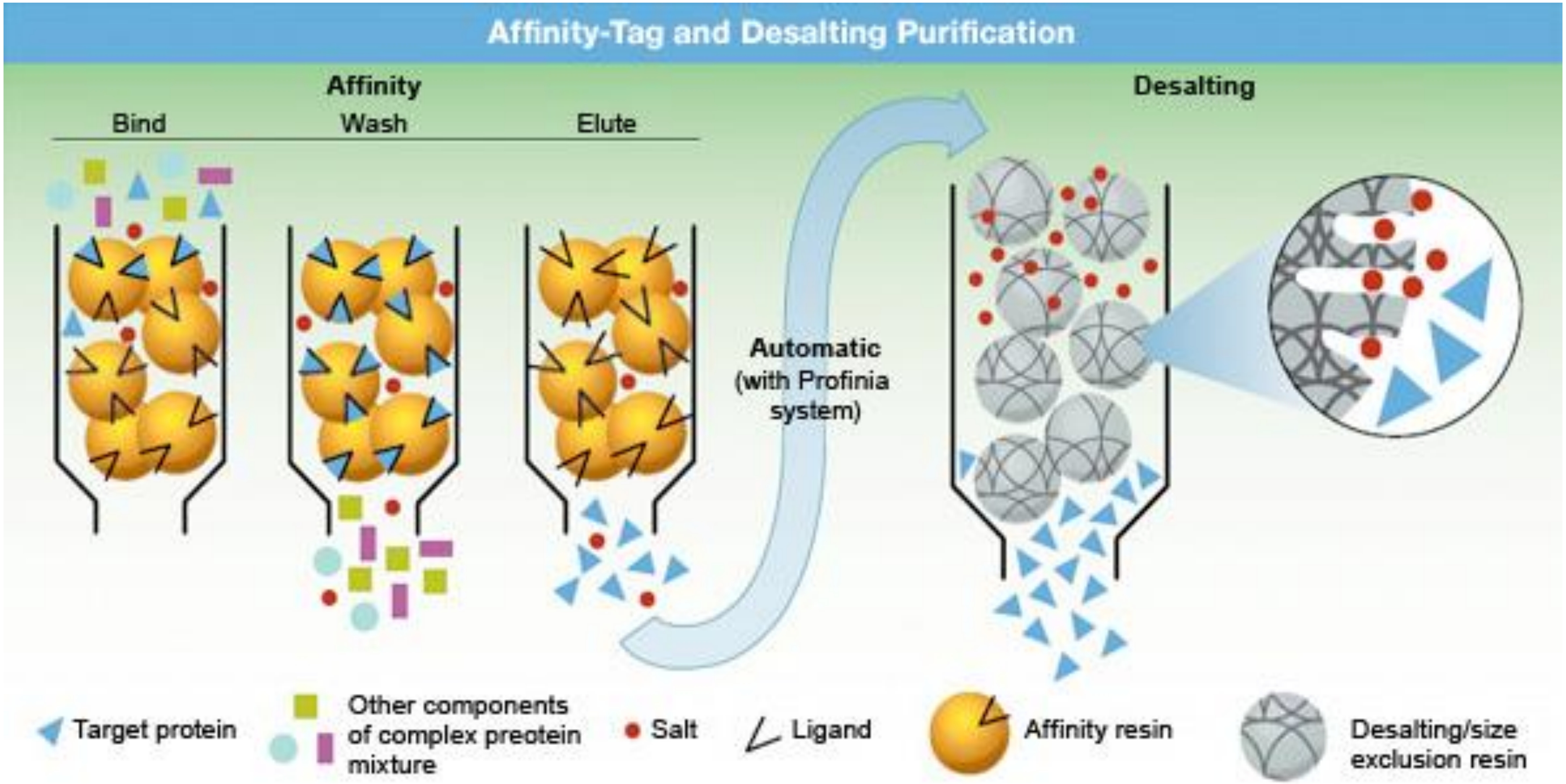


3D-Gel - Method & Workflow
3d-gel.de

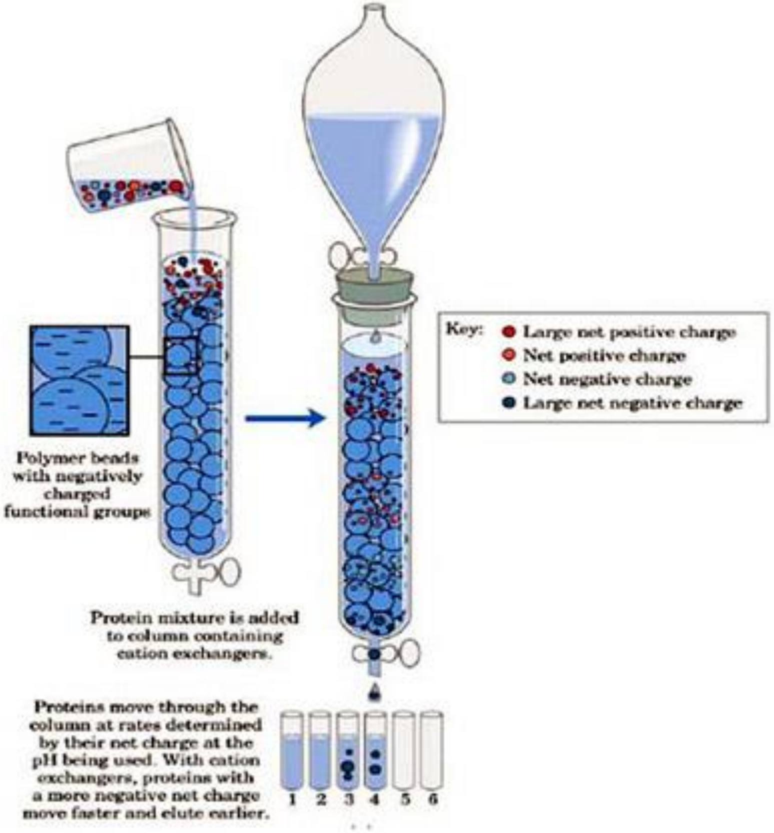
Chromatography

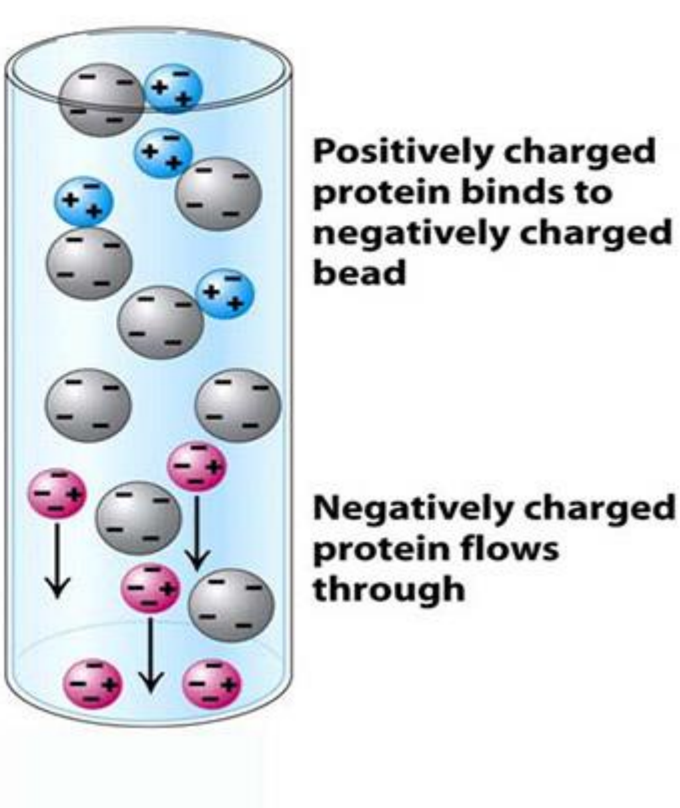


Column Chromatography
byjus.com



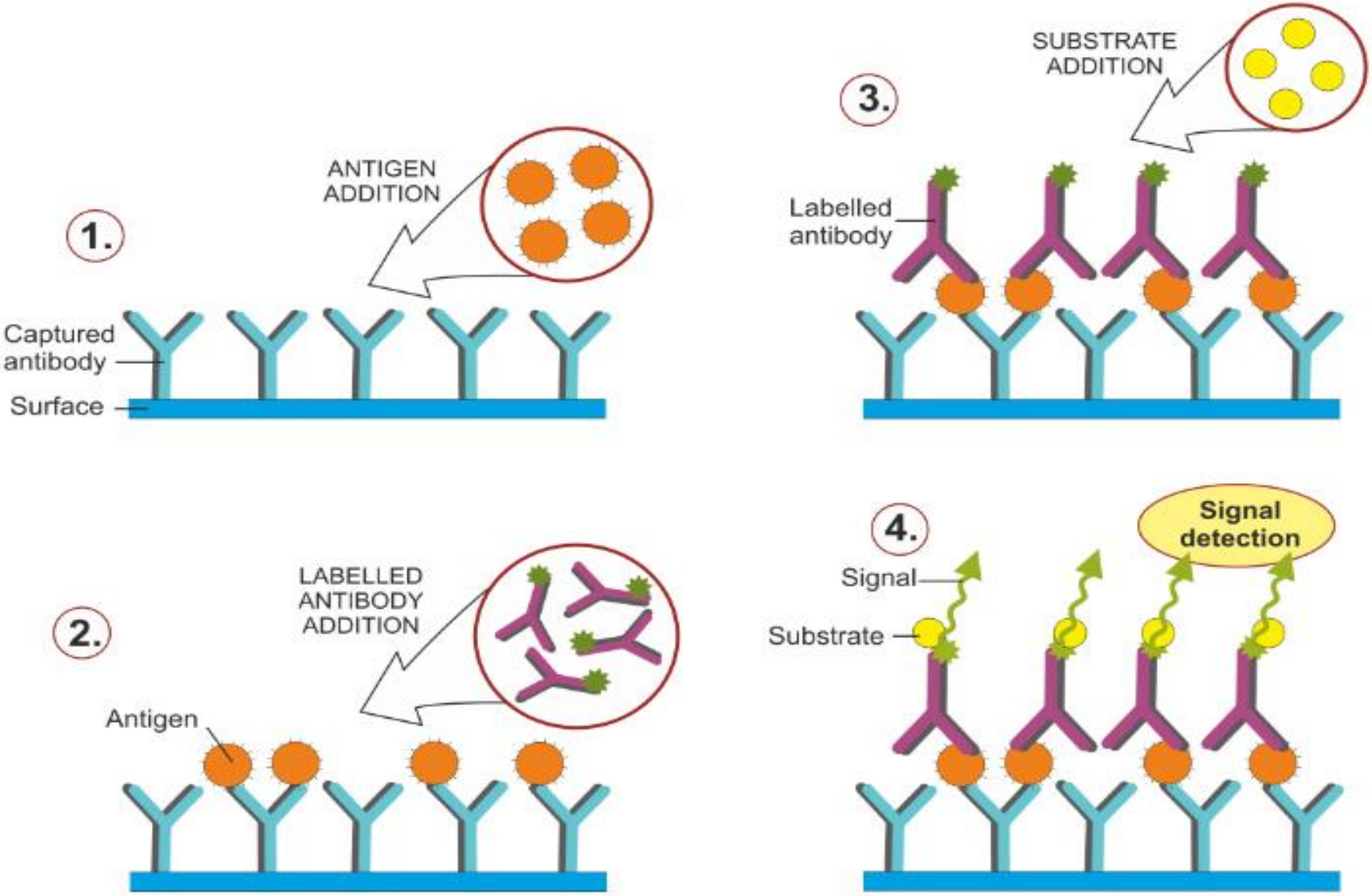
Introduction to Affinity Chromatography
bio-rad.com



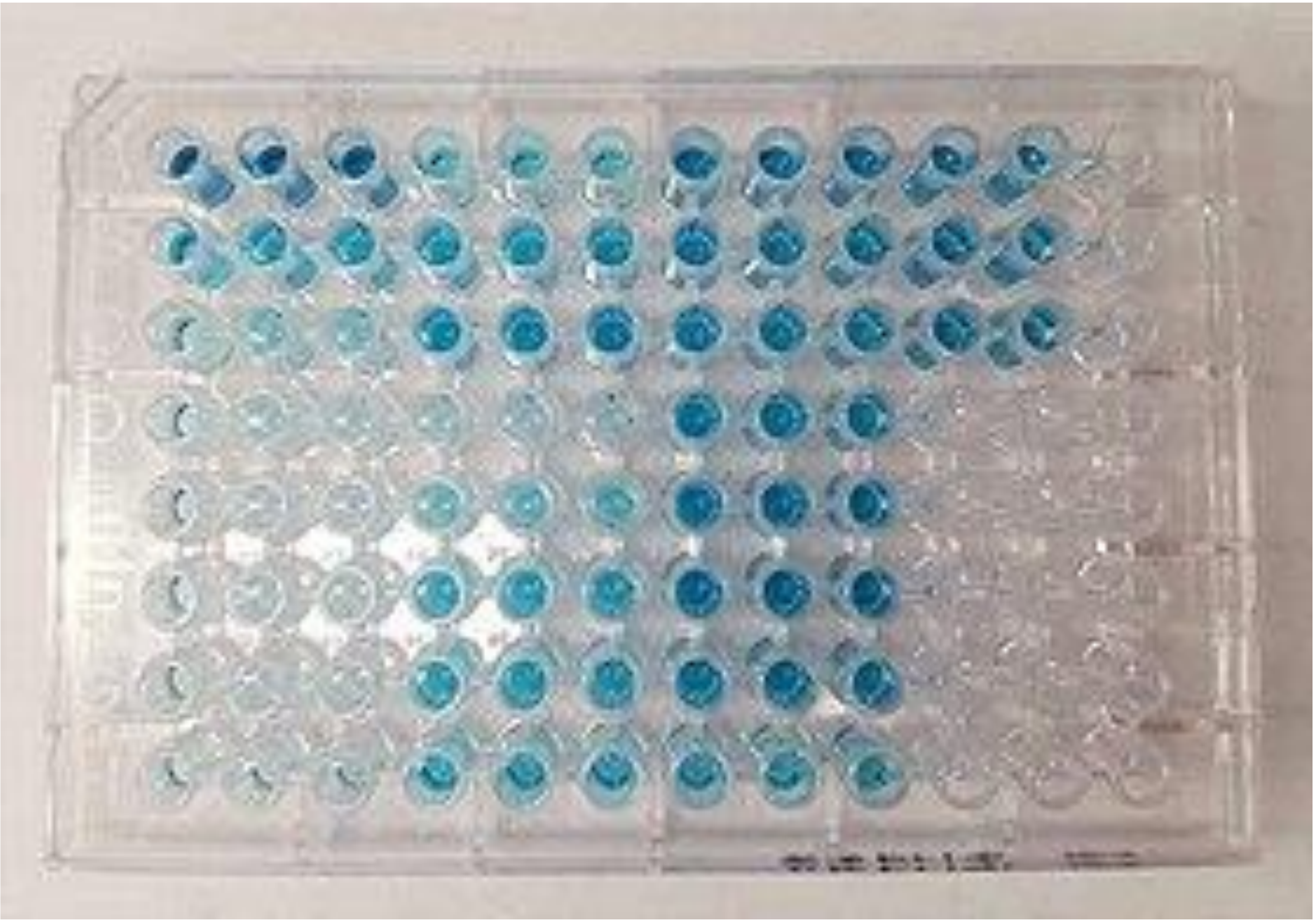


Ion Exchange Chromatography
microbenotes.com

ELISA (enzyme-linked immunosorbent assay)

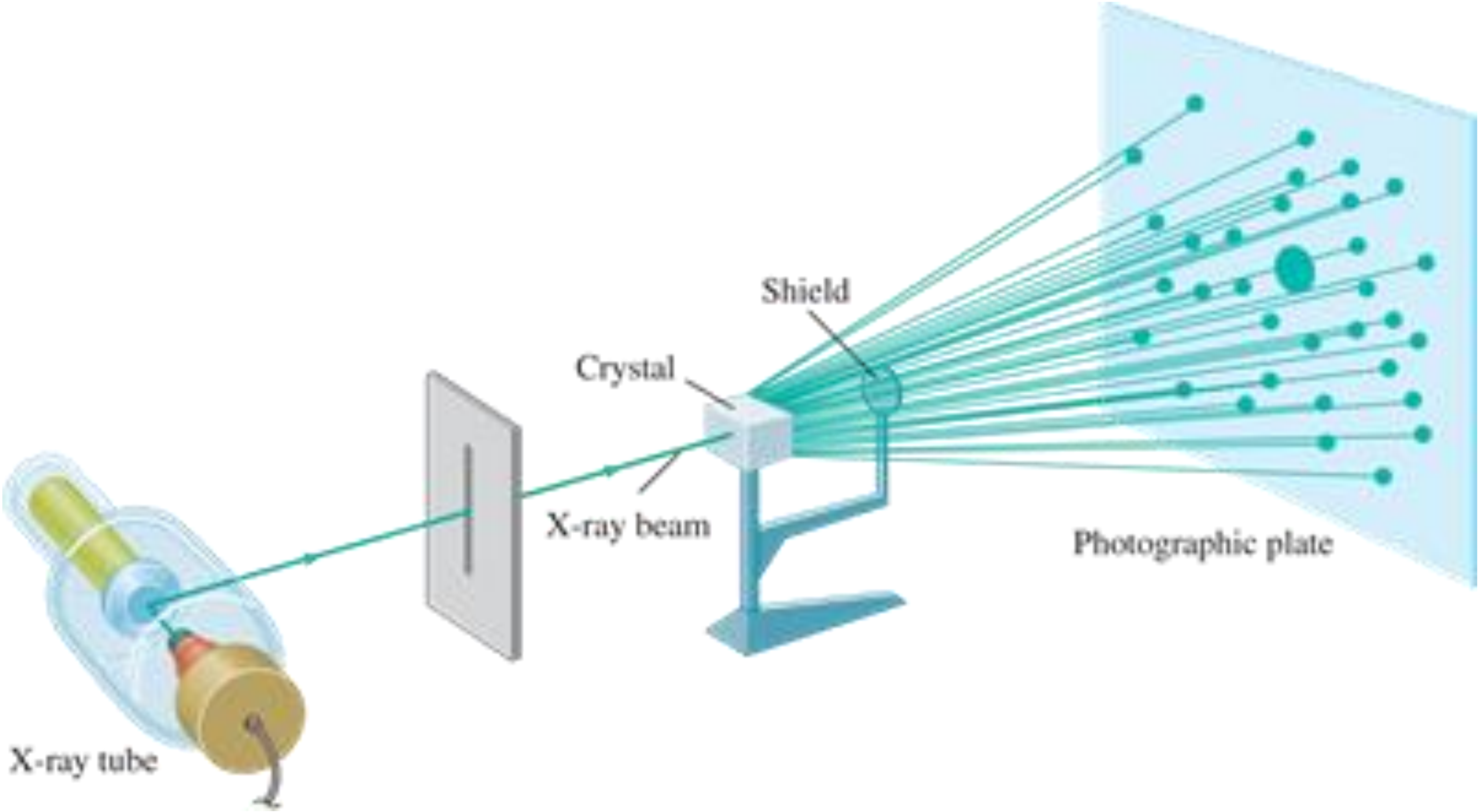


BioTek: ELISA and related immunoassays
biotek.com

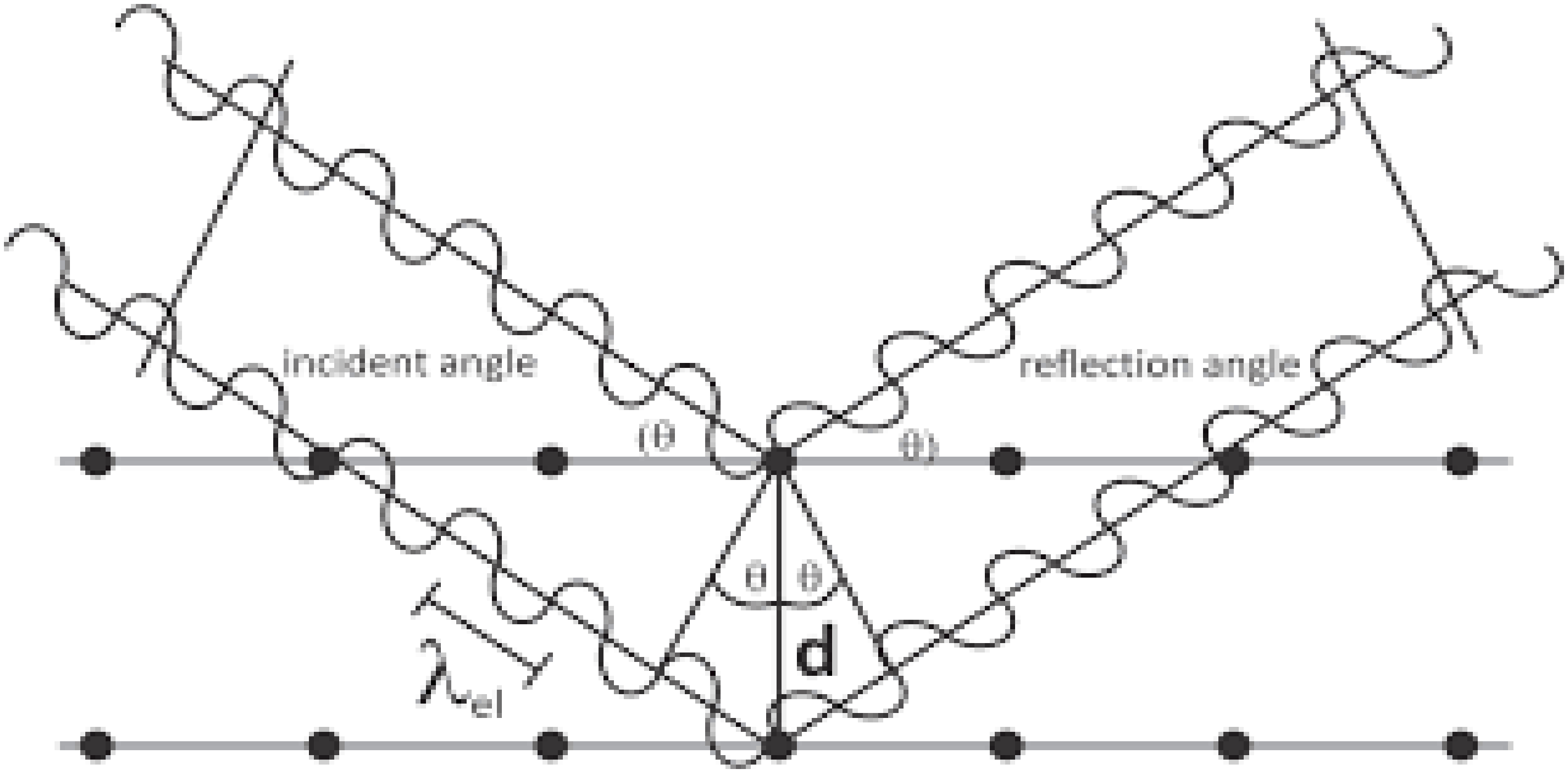


ELISA - Wikipedia
en.wikipedia.org

X-ray diffraction

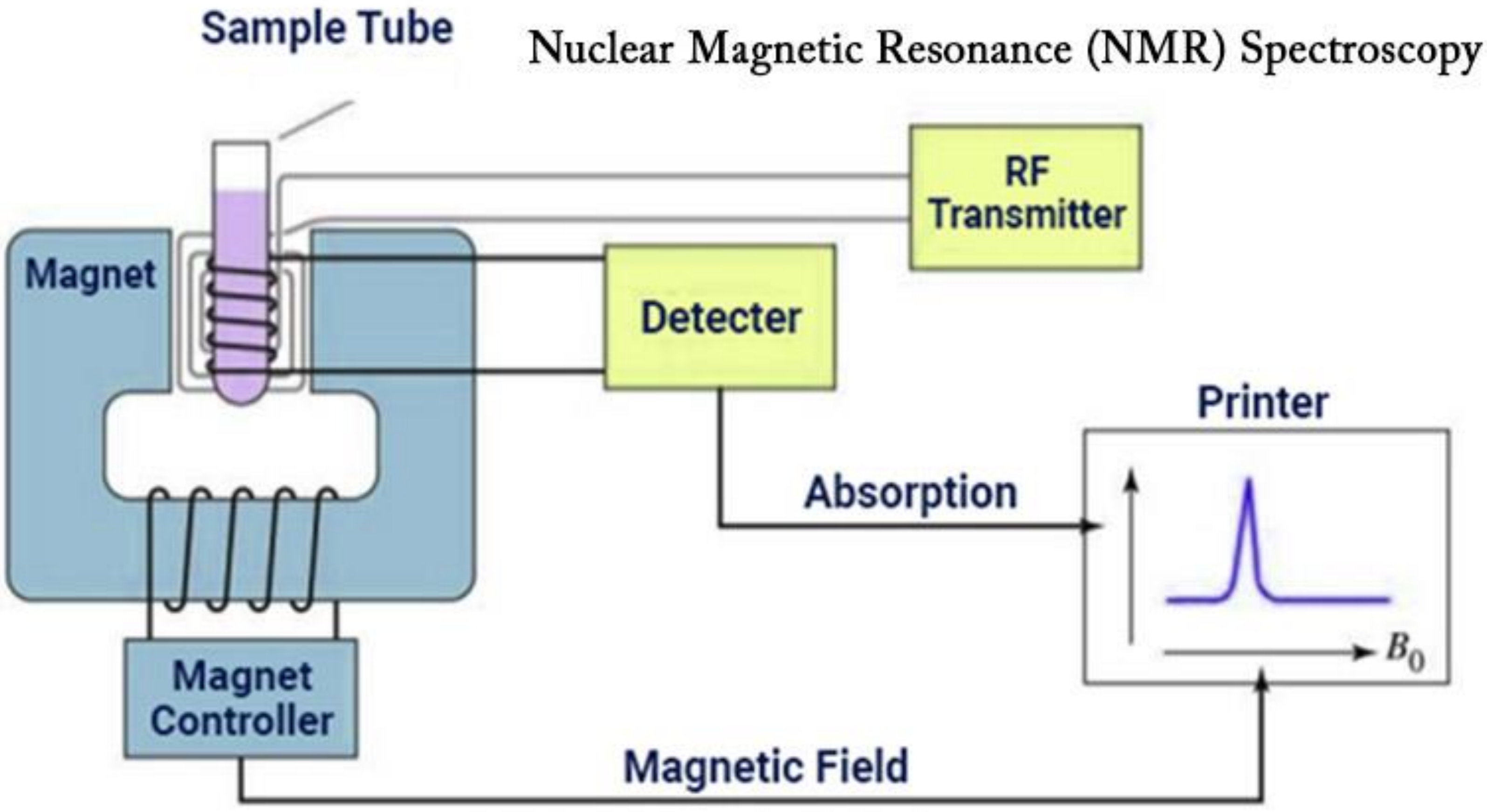


What is X-ray Crystallography?
medium.com

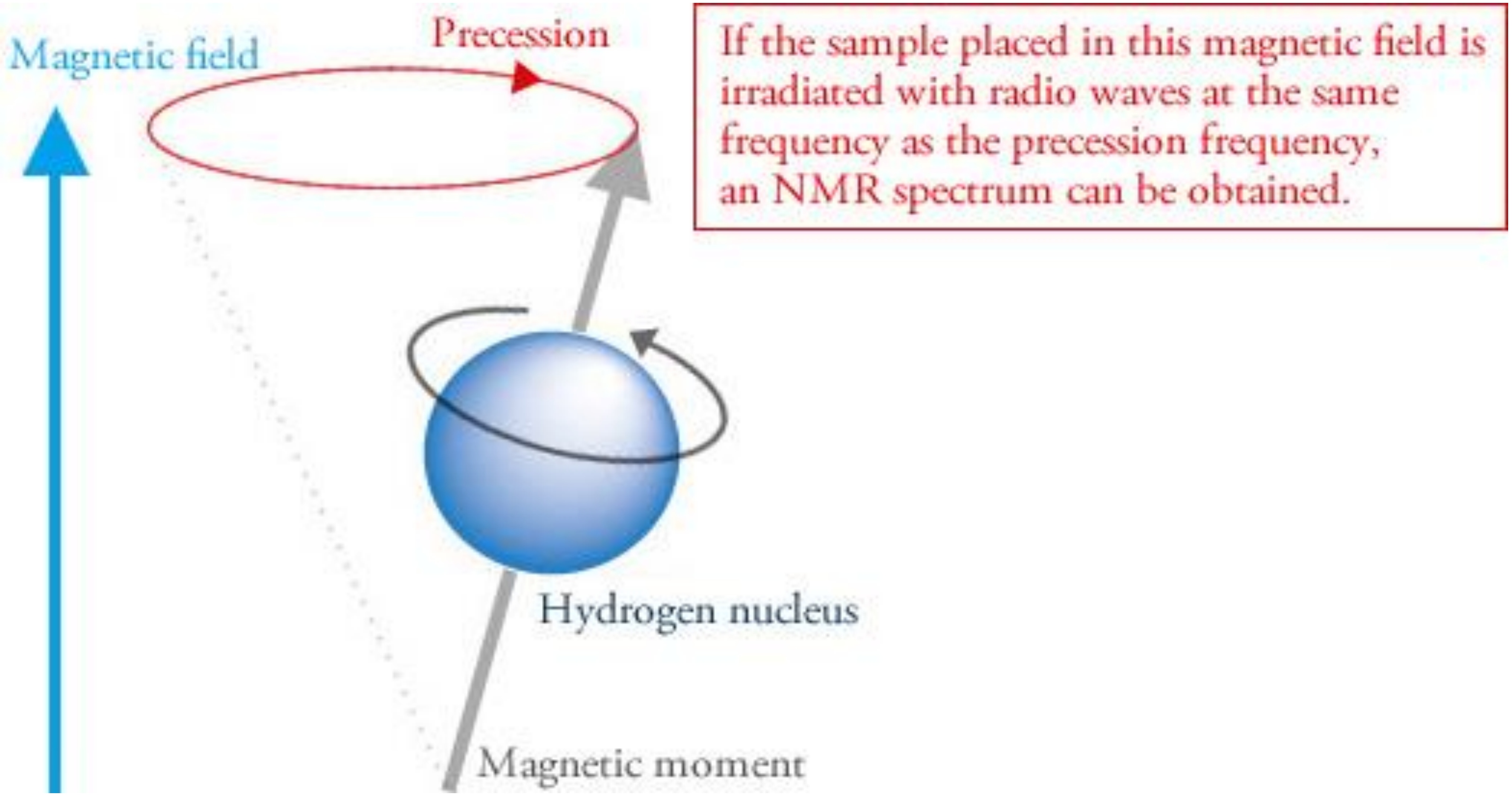


X-RAY DIFFRACTION
letstalkacademy.com

Nuclear magnetic resonance (NMR)



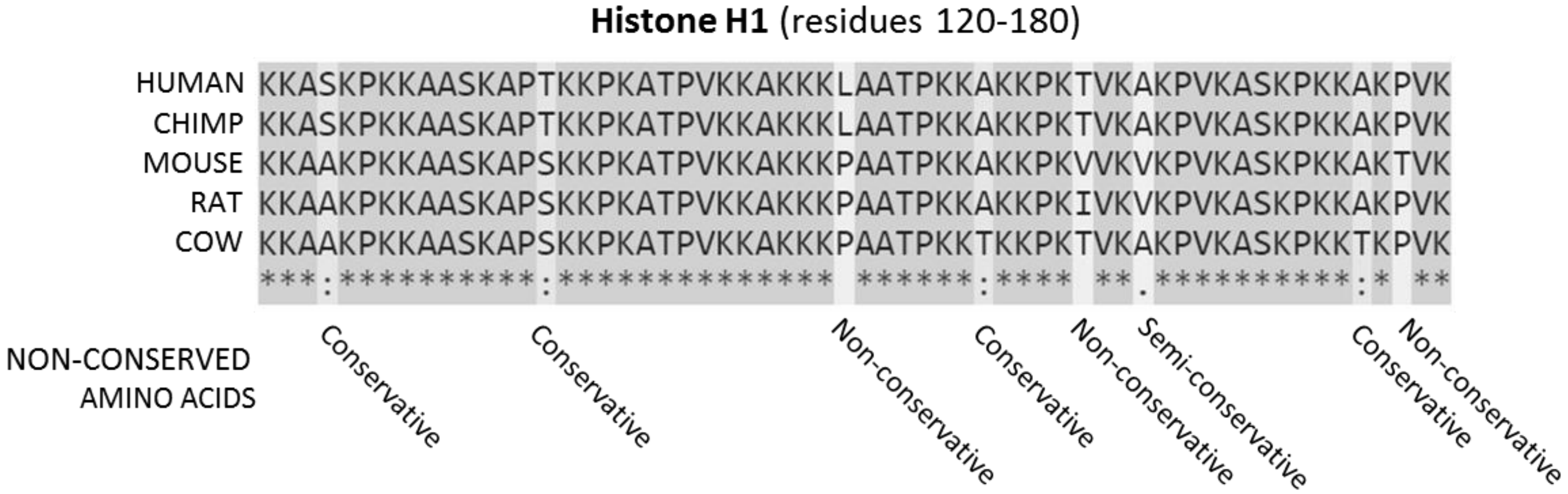
microbenotes.com



If the sample placed in this magnetic field is irradiated with radio waves at the same frequency as the precession frequency, an NMR spectrum can be obtained.

NMR basic knowledge
jeol.co.jp

Bioinformatic methods of protein research

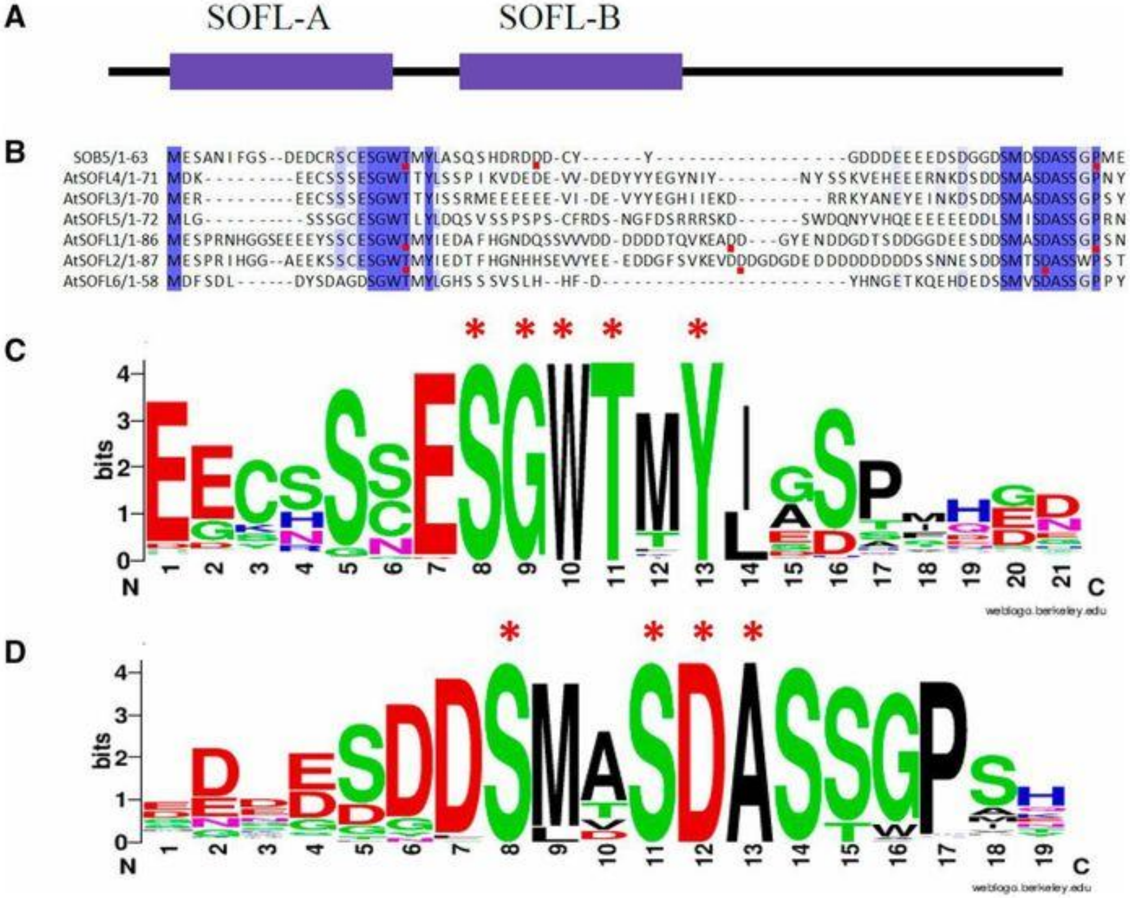


Histone H1 (residues 120-180)

```
HUMAN  KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
CHIMP  KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
MOUSE  KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKVVKVKPVKASKPKKAKTVK
RAT    KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKIVKVKPVKASKPKKAKPVK
COW    KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKTKKPKTVKAKPVKASKPKKTKPVK
       ***:**********:************** ******:**** **.**********:* **
```

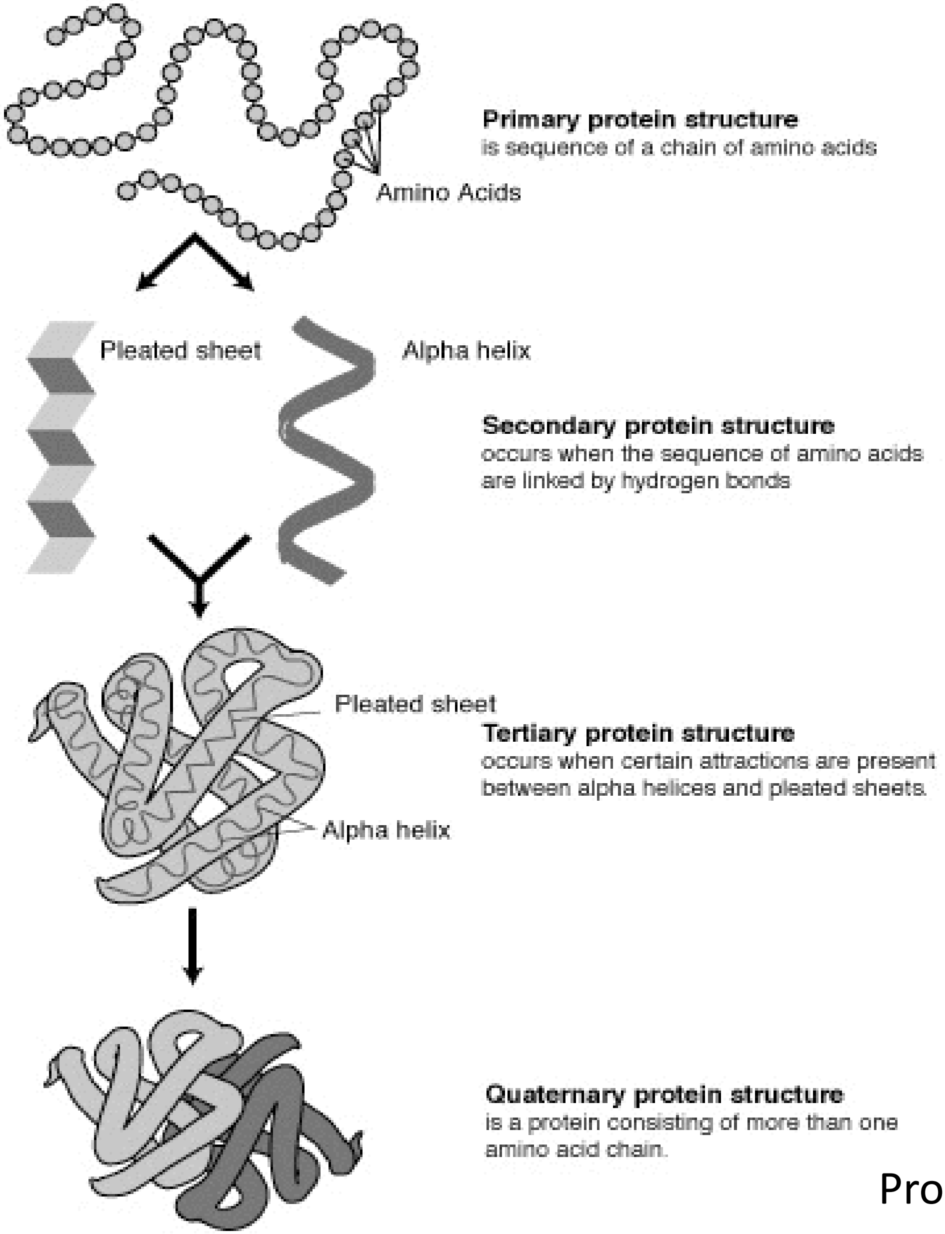
Sequence alignment - Wikipedia
en.wikipedia.org

Multiple sequence alignment and analysis of SOFL proteins.



Reuben Tayengwa et al. G3 2018;8:1281-1290

G3·Genes|Genomes|Genetics GSA



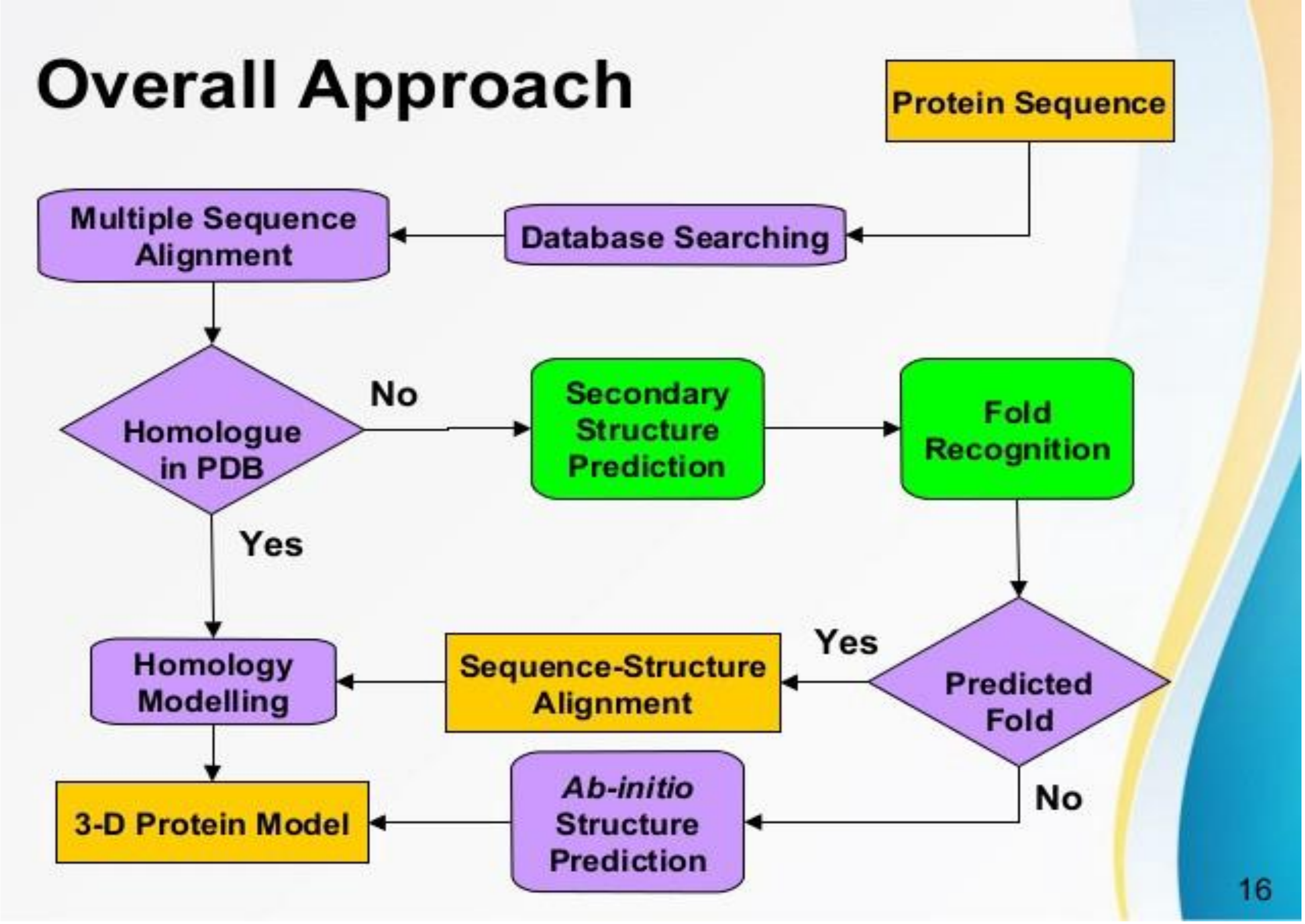
Primary protein structure
is sequence of a chain of amino acids

Secondary protein structure
occurs when the sequence of amino acids are linked by hydrogen bonds

Tertiary protein structure
occurs when certain attractions are present between alpha helices and pleated sheets.

Quaternary protein structure
is a protein consisting of more than one amino acid chain.

Protein structure prediction - Wikipedia
en.wikipedia.org



Overall Approach

Protein Sequence

Database Searching

Multiple Sequence Alignment

Homologue in PDB

No

Secondary Structure Prediction

Fold Recognition

Yes

Predicted Fold

Yes

Sequence-Structure Alignment

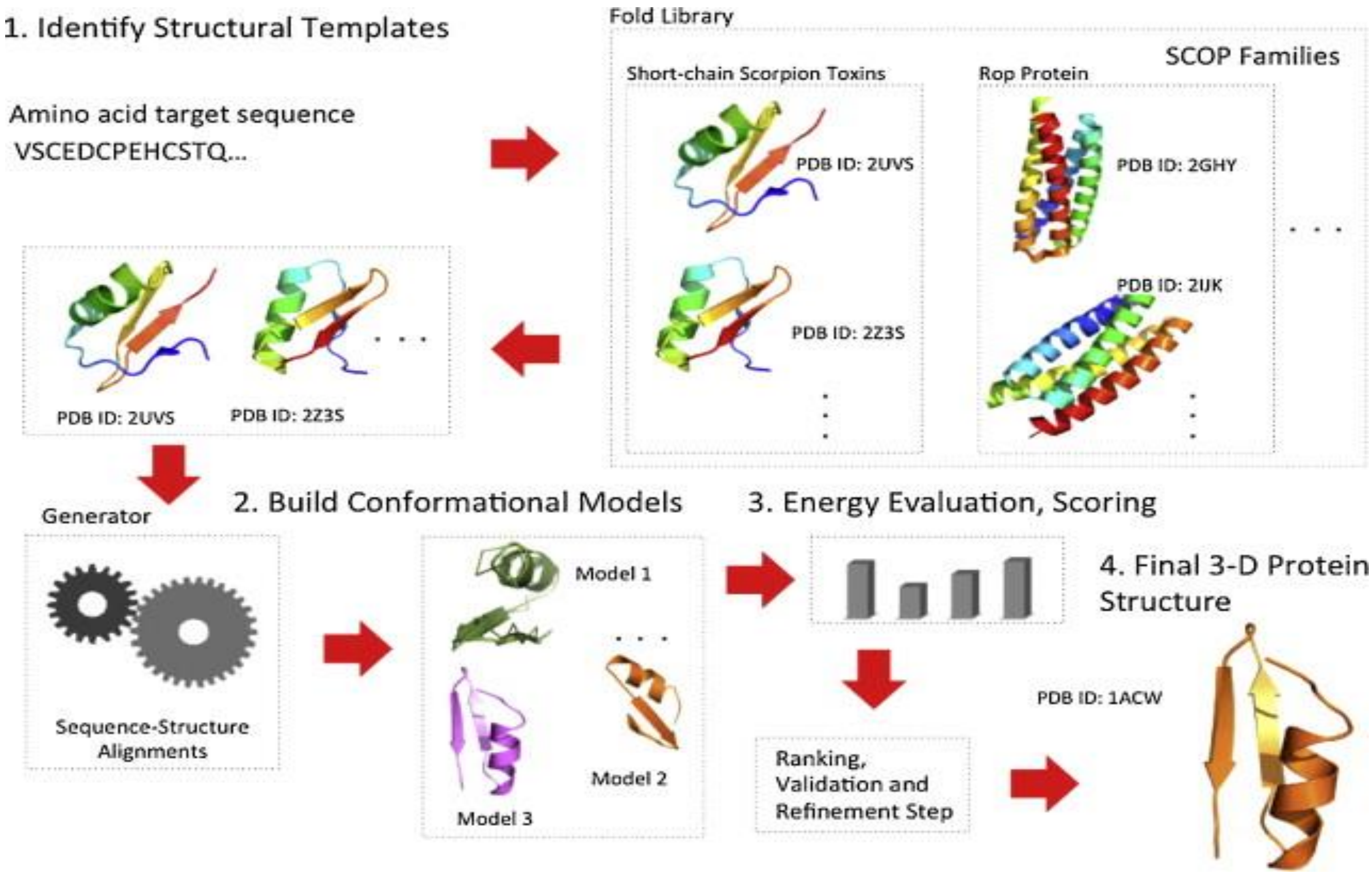
Homology Modelling

No

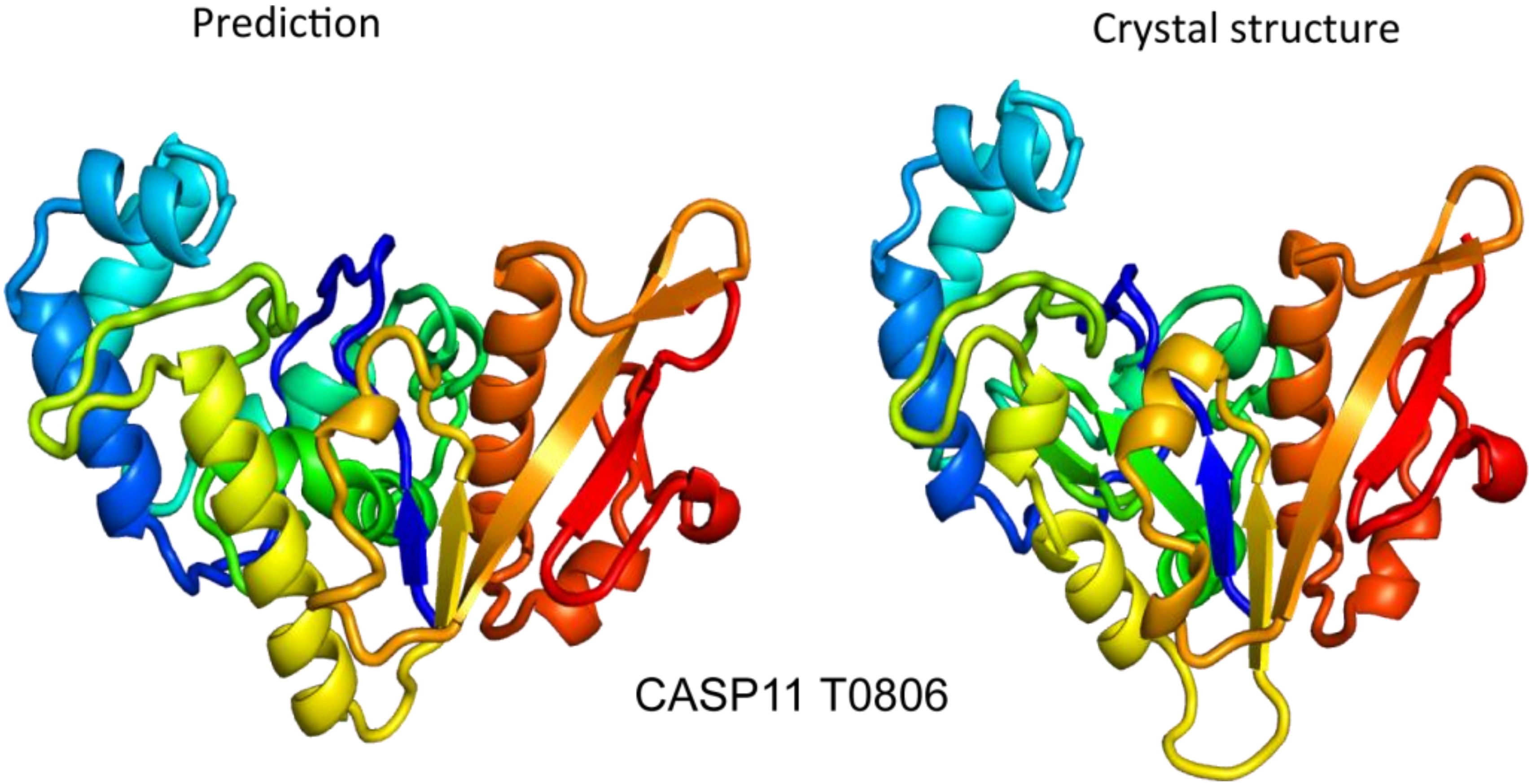
Ab-initio Structure Prediction

3-D Protein Model

A guide for protein structure prediction and 3D-modelling
medium.com



Three-dimensional protein structure prediction.
sciencedirect.com

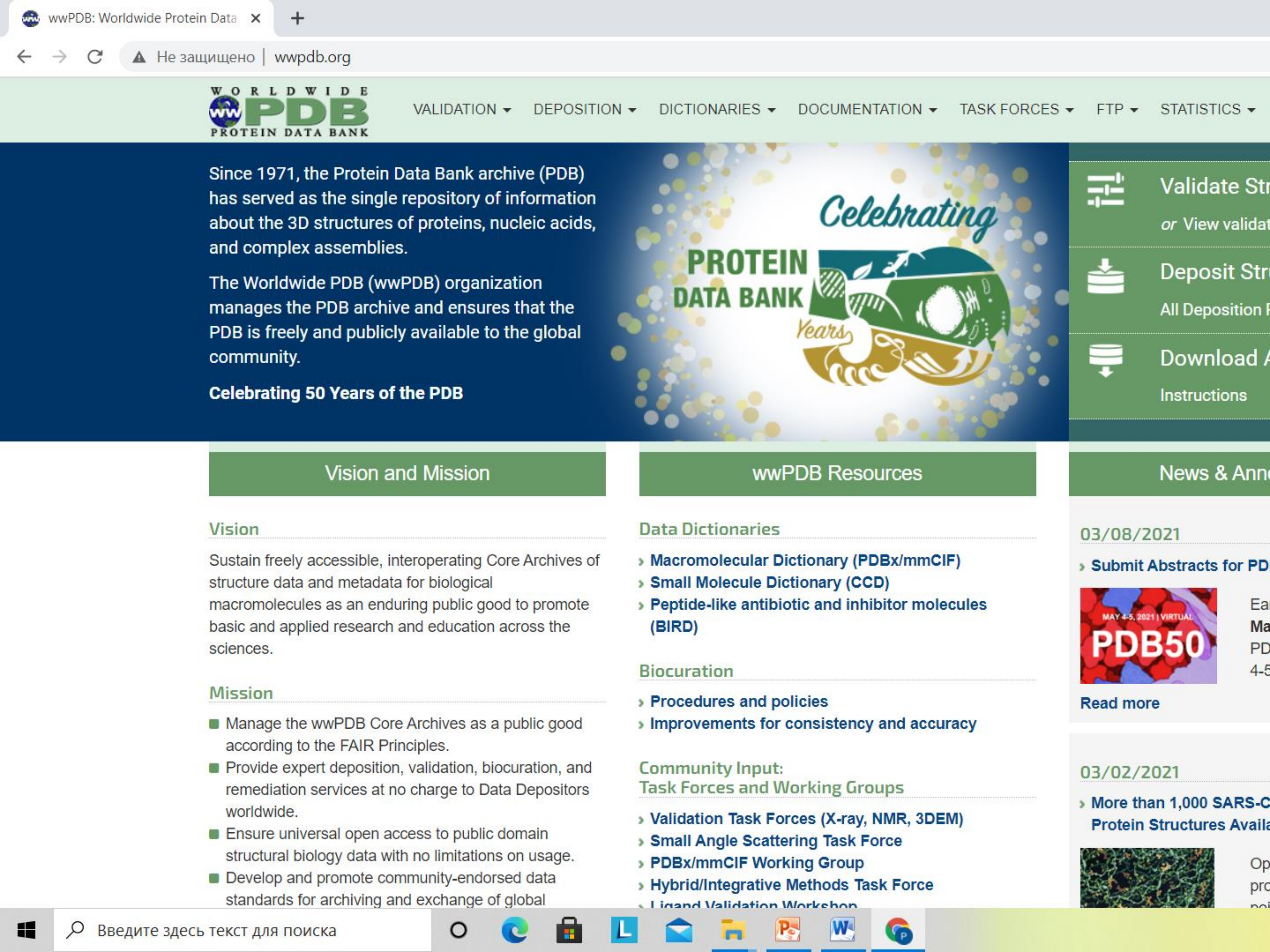


Prediction

Crystal structure

CASP11 T0806

November 2015 – Institute for Protein Design
ipd.uw.edu



Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Celebrating 50 Years of the PDB

Vision and Mission

Vision

Sustain freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences.

Mission

- Manage the wwPDB Core Archives as a public good according to the FAIR Principles.
- Provide expert deposition, validation, biocuration, and remediation services at no charge to Data Depositors worldwide.
- Ensure universal open access to public domain structural biology data with no limitations on usage.
- Develop and promote community-endorsed data standards for archiving and exchange of global

wwPDB Resources

Data Dictionaries

- **Macromolecular Dictionary (PDBx/mmCIF)**
- **Small Molecule Dictionary (CCD)**
- **Peptide-like antibiotic and inhibitor molecules (BIRD)**

Biocuration

- **Procedures and policies**
- **Improvements for consistency and accuracy**

Community Input: Task Forces and Working Groups

- **Validation Task Forces (X-ray, NMR, 3DEM)**
- **Small Angle Scattering Task Force**
- **PDBx/mmCIF Working Group**
- **Hybrid/Integrative Methods Task Force**
- **Ligand Validation Workshop**

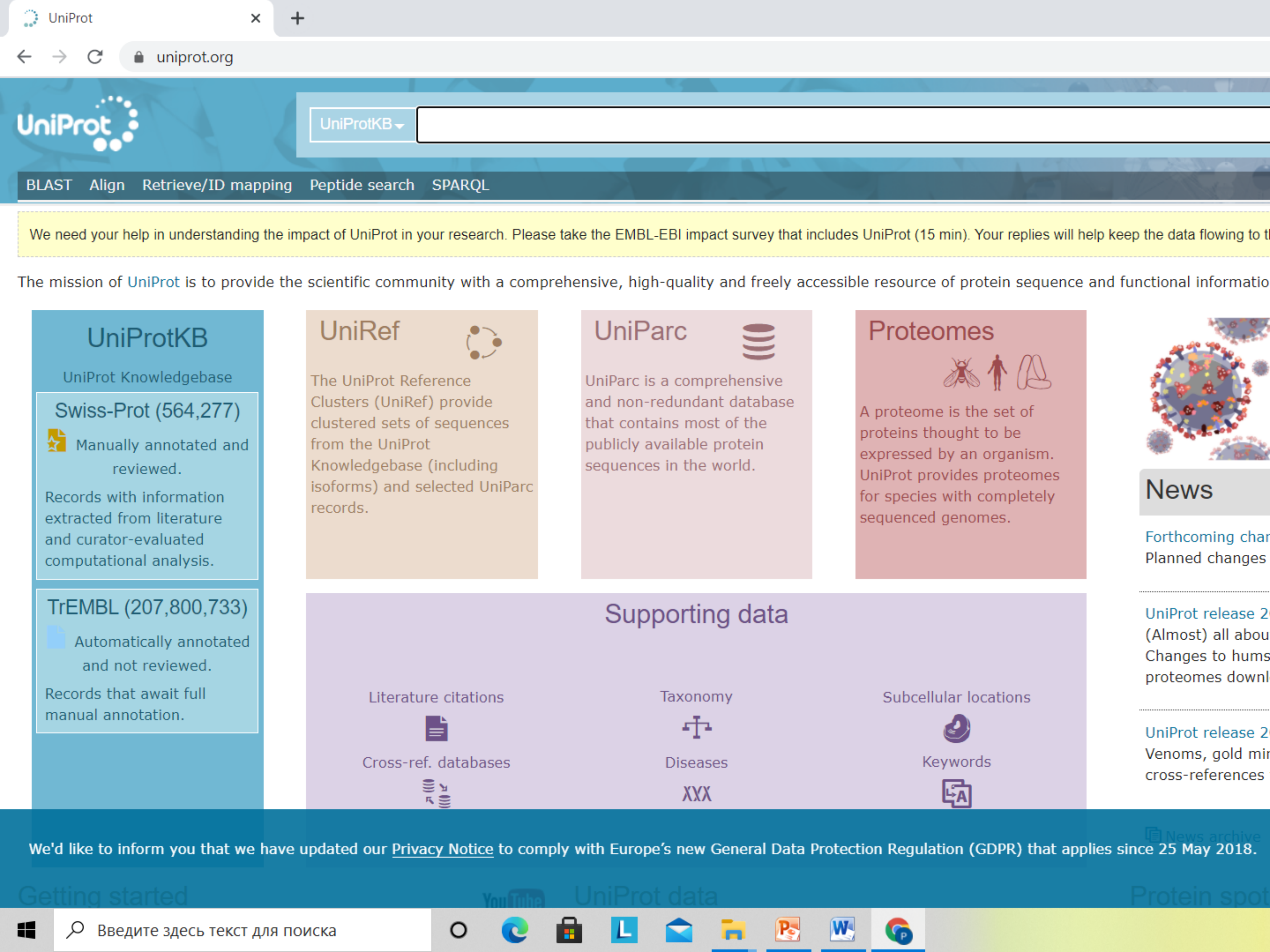
03/08/2021

Submit Abstracts for PD

Read more

03/02/2021

More than 1,000 SARS-C Protein Structures Availa



We need your help in understanding the impact of UniProt in your research. Please take the EMBL-EBI impact survey that includes UniProt (15 min). Your replies will help keep the data flowing to t

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional informatio

UniProtKB

UniProt Knowledgebase

Swiss-Prot (564,277)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (207,800,733)

Automatically annotated and not reviewed.

Records that await full manual annotation.

UniRef

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

UniParc

UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes

A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

Supporting data

Literature citations

Taxonomy

Subcellular locations

Cross-ref. databases

Diseases

Keywords

News

Forthcoming char

Planned changes

UniProt release 2

(Almost) all abou

Changes to hums

proteomes downl

UniProt release 2

Venoms, gold mir

cross-references

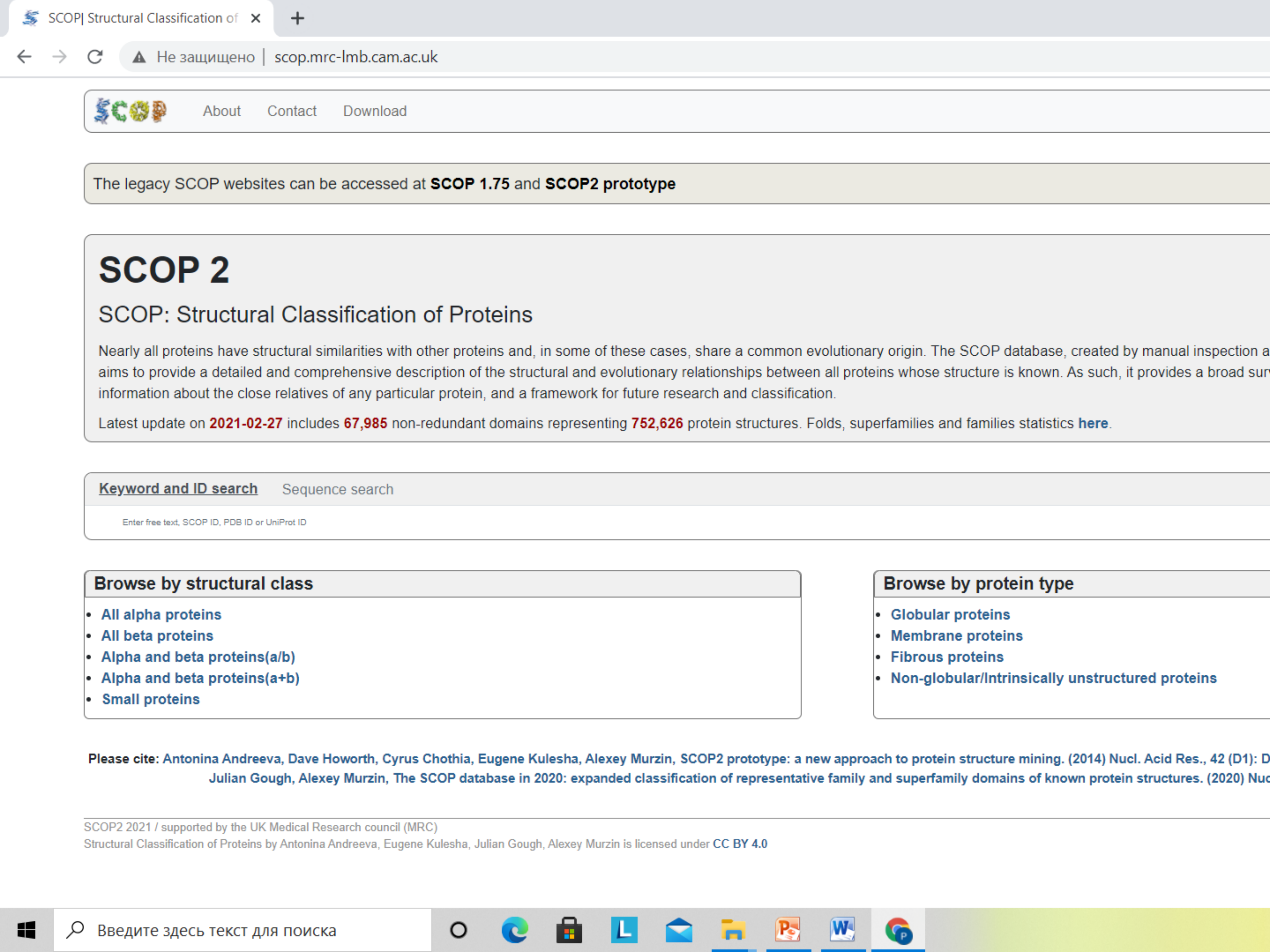
News archive

We'd like to inform you that we have updated our Privacy Notice to comply with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018.

Getting started

UniProt data

Protein spot



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 common evolutionary origin. The SCOP database, created by manual inspection a aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. As such, it provides a broad sur information about the close relatives of any particular protein, and a framework for future research and classification.

Latest update on **2021-02-27** includes **67,985** non-redundant domains representing **752,626** protein structures. Folds, superfamilies and families statistics **here**.

Keyword and ID search Sequence search

Enter free text, SCOP ID, PDB ID or UniProt ID

Browse by structural class

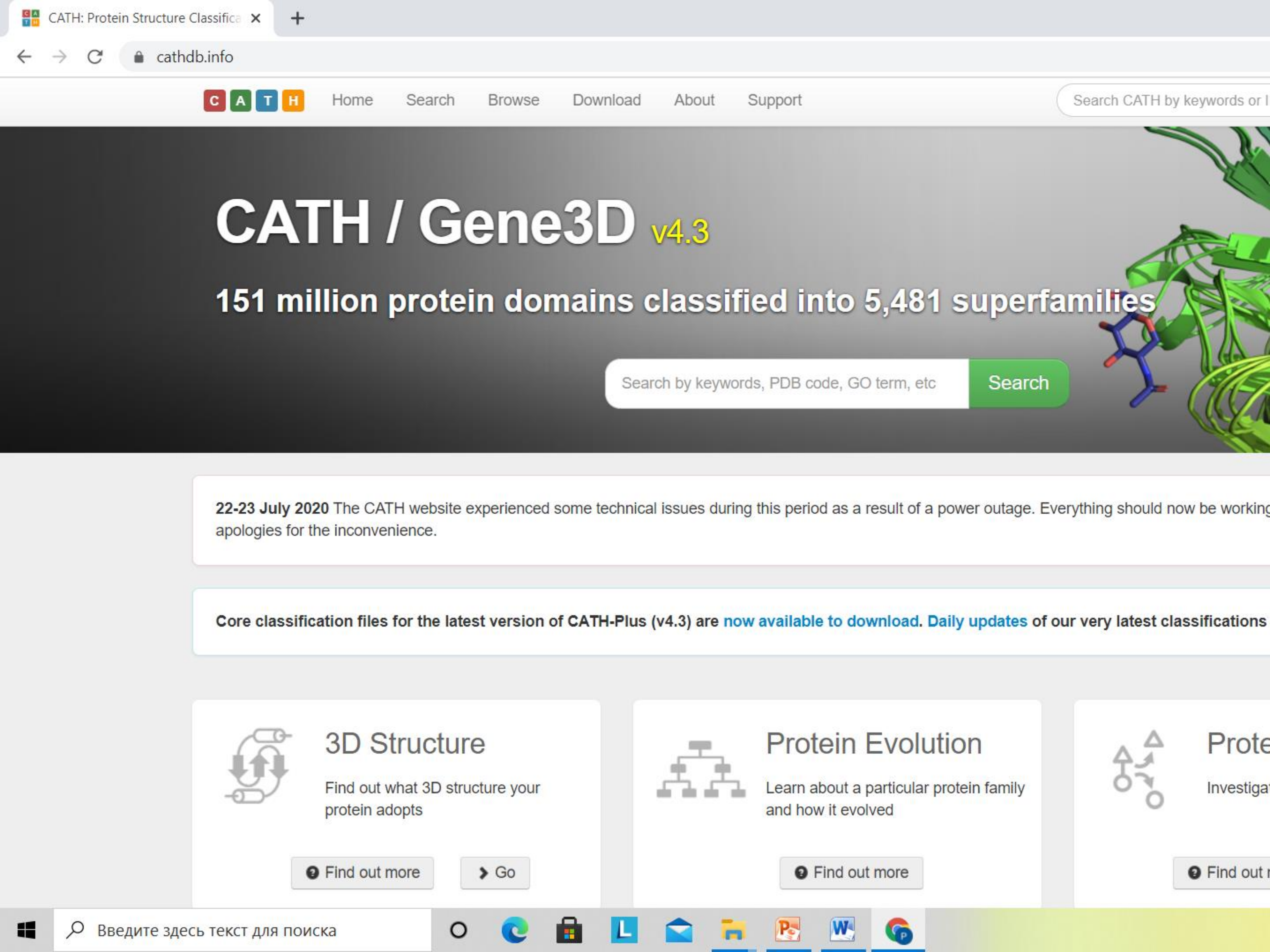
- **All alpha proteins**
- **All beta proteins**
- **Alpha and beta proteins(a/b)**
- **Alpha and beta proteins(a+b)**
- **Small proteins**

Browse by protein type

- **Globular proteins**
- **Membrane proteins**
- **Fibrous proteins**
- **Non-globular/Intrinsically unstructured proteins**

Please cite: Antonina Andreeva, Dave Howorth, Cyrus Chothia, Eugene Kulesha, Alexey Murzin, SCOP2 prototype: a new approach to protein structure mining. (2014) Nucl. Acid Res., 42 (D1): D
Julian Gough, Alexey Murzin, The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. (2020) Nuc

SCOP2 2021 / supported by the UK Medical Research council (MRC)
Structural Classification of Proteins by Antonina Andreeva, Eugene Kulesha, Julian Gough, Alexey Murzin is licensed under CC BY 4.0



CATH / Gene3D v4.3

151 million protein domains classified into 5,481 superfamilies

22-23 July 2020 The CATH website experienced some technical issues during this period as a result of a power outage. Everything should now be working
apologies for the inconvenience.

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

3D Structure

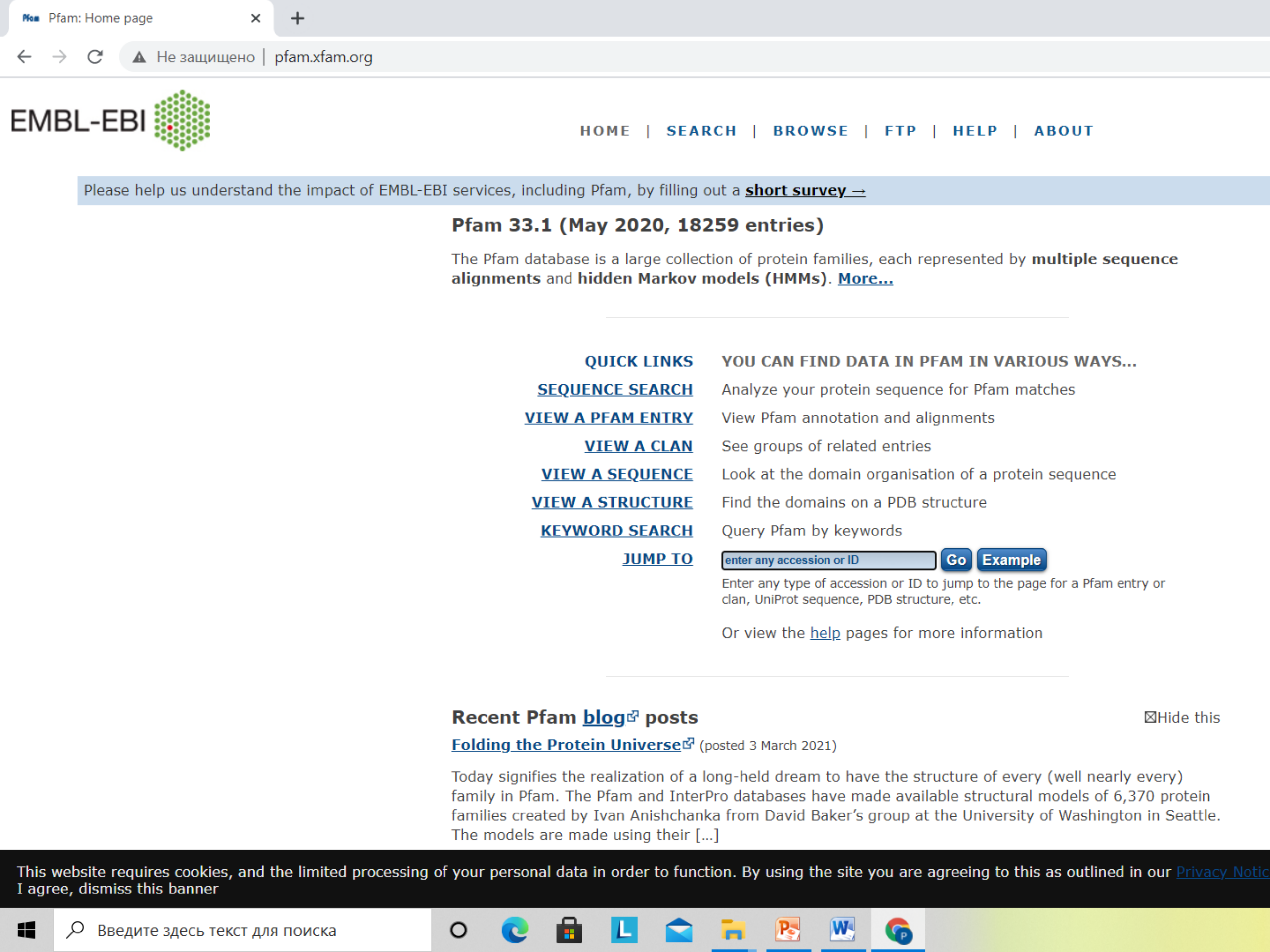
Find out what 3D structure your protein adopts

Protein Evolution

Learn about a particular protein family and how it evolved

Prote

Investiga



Please help us understand the impact of EMBL-EBI services, including Pfam, by filling out a **short survey →**

Pfam 33.1 (May 2020, 18259 entries)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. More...

QUICK LINKS	YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
SEQUENCE SEARCH	Analyze your protein sequence for Pfam matches
VIEW A PFAM ENTRY	View Pfam annotation and alignments
VIEW A CLAN	See groups of related entries
VIEW A SEQUENCE	Look at the domain organisation of a protein sequence
VIEW A STRUCTURE	Find the domains on a PDB structure
KEYWORD SEARCH	Query Pfam by keywords
JUMP TO	Enter any type of accession or ID to jump to the page for a Pfam entry or clan, UniProt sequence, PDB structure, etc.

Or view the help pages for more information

Recent Pfam blog posts

Folding the Protein Universe (posted 3 March 2021)

Today signifies the realization of a long-held dream to have the structure of every (well nearly every) family in Pfam. The Pfam and InterPro databases have made available structural models of 6,370 protein families created by Ivan Anishchanka from David Baker's group at the University of Washington in Seattle. The models are made using their [...]

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notic
I agree, dismiss this banner

References

1. Anderson NL, Anderson NG; Anderson (1998). "Proteome and proteomics: new technologies, new concepts, and new words". Electrophoresis. 19 (11): 1853–61. doi:10.1002/elps.1150191103. PMID 9740045. S2CID 28933890.
2. Blackstock WP, Weir MP; Weir (1999). "Proteomics: quantitative and physical mapping of cellular proteins". Trends Biotechnol. 17 (3): 121–7. doi:10.1016/S0167-7799(98)01245-1. PMID 10189717.
3. Anderson, Johnathon D.; Johansson, Henrik J.; Graham, Calvin S.; Vesterlund, Mattias; Pham, Missy T.; Bramlett, Charles S.; Montgomery, Elizabeth N.; Mellema, Matt S.; Bardini, Renee L. (2016-03-01). "Comprehensive Proteomic Analysis of Mesenchymal Stem Cell Exosomes Reveals Modulation of Angiogenesis via Nuclear Factor-KappaB Signaling". Stem Cells. 34 (3): 601–613. doi:10.1002/stem.2298. ISSN 1549-4918. PMC 5785927. PMID 26782178.
4. Hood, Leroy; Rowen, Lee (2013-09-13). "The human genome project: big science transforms biology and medicine". Genome Medicine. 5 (9): 79. doi:10.1186/gm483. PMC 4066586. PMID 24040834.