UNIVERSITY OF PORTSMOUTH
CENTRE FOR ENZYME INNOVATION

The Future of Structural Biology (Protein X-ray crystallography)

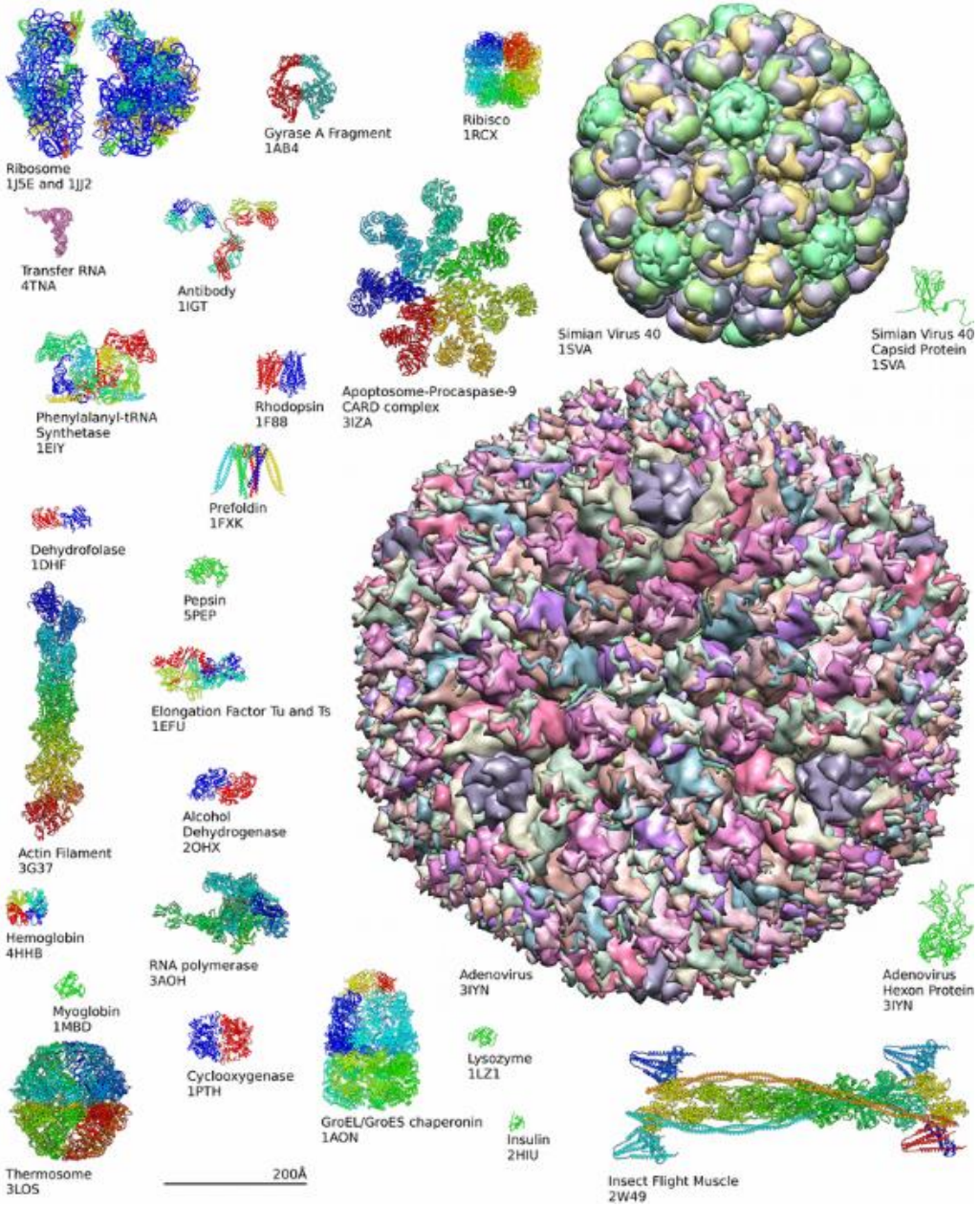
Dr Michael Zahn

06/12/2022

How does the Future of Structural Biology look like?

How can I be a part of it?

Structural Biology



➔ Structural biology = study of the structure of macromolecules and how they fold

➔ Primary purpose of structural biology is to better understand the proteins involved in biological processes and how they perform their functions

Structural Biology

X-ray crystallography

NMR

Cryo-EM

SAXS

Cryo-ET

AFM

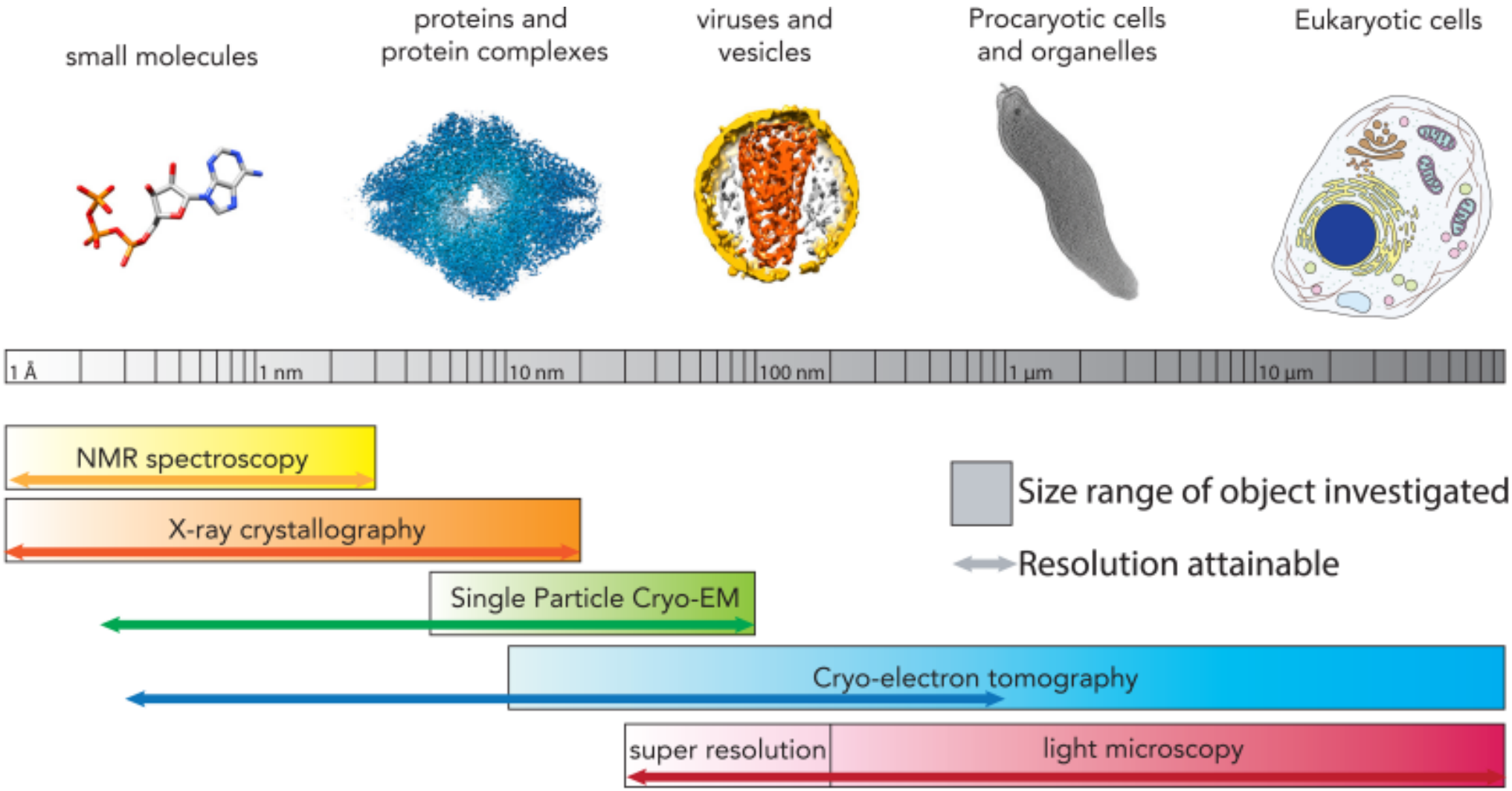
Molecular modelling

MicroED

EPR

MS

Structural Biology

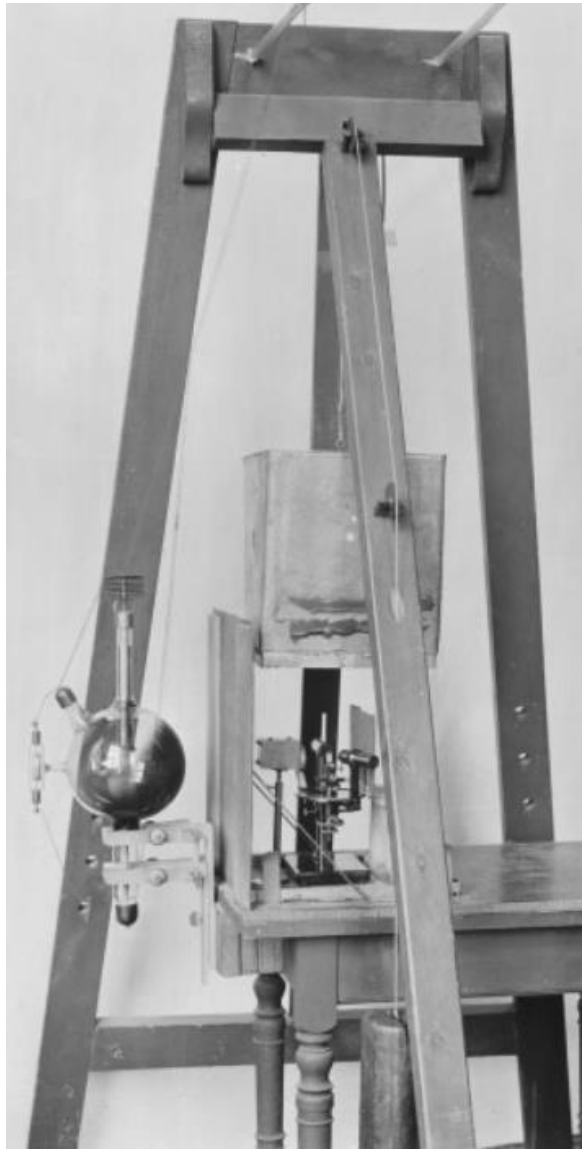


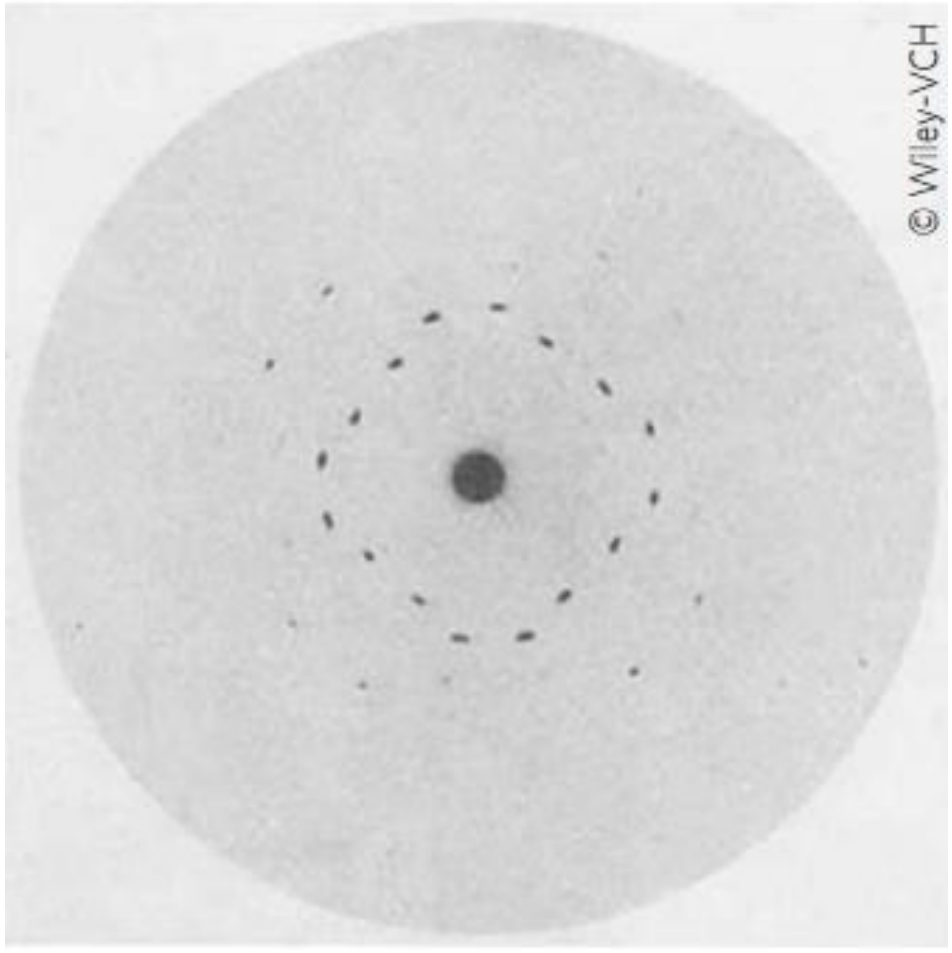
Biochem Soc Trans., 2018, 46(4):807-816

Milestones in Strutural Biology

1912 – Max von Laue, Walter Friedrich, Paul Knipping

➔ Discovery of X-ray diffraction = Birth of X-ray crystallography

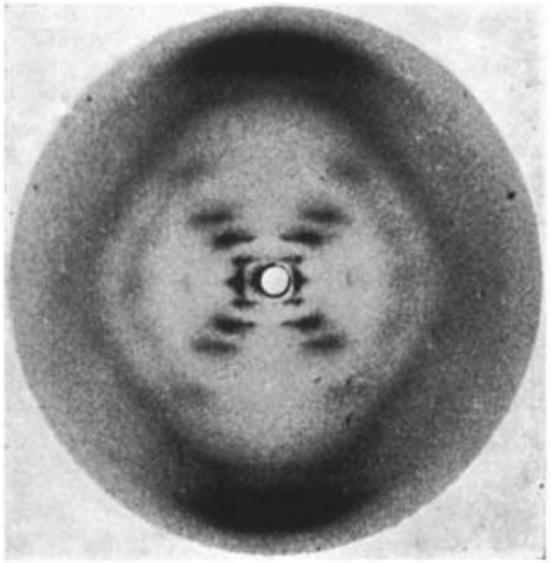




X-ray diffraction pattern from a zinc-blende (ZnS) crystal.

1953 – Rosalind Franklin & Raymond Gosling

➔ discovery of the DNA double helix structure with the help of fiber X-ray diffraction



Nature, 1953, 171(4356):740-1

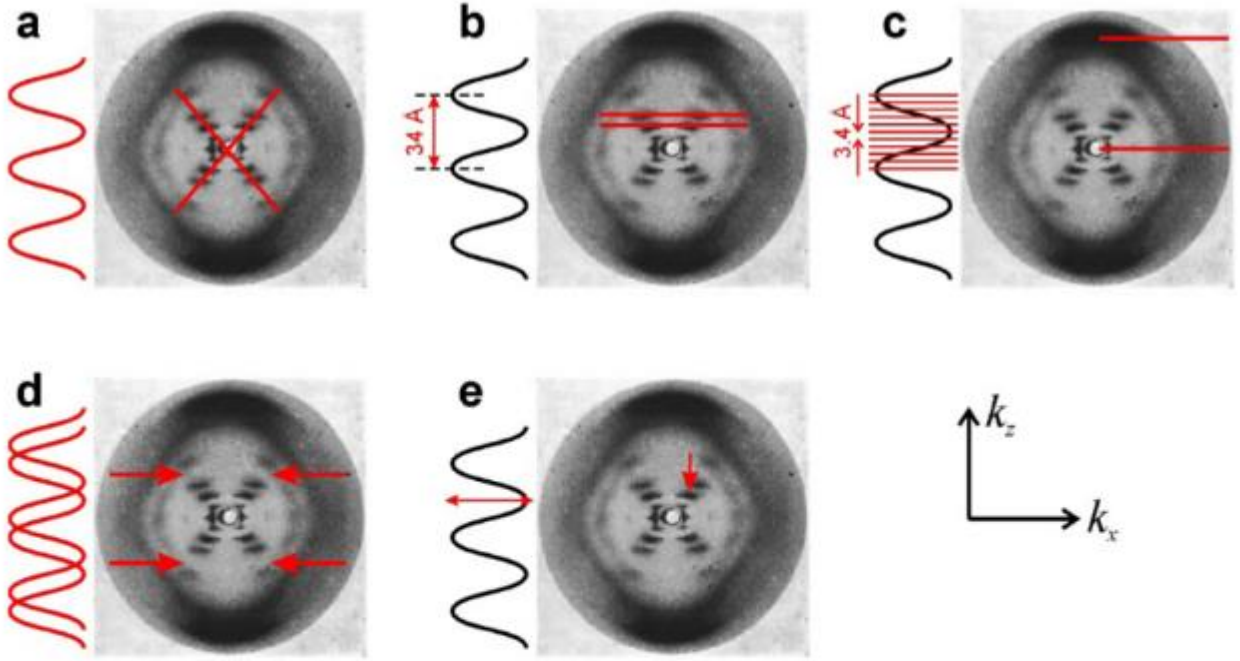


Fig. 11. Interpretation of Photo 51 B-form DNA diffraction pattern. (a) "X"-form distribution of the diffraction peaks (layer lines) is an indication of a helical structure. (b) The intensity exhibits non-zero values at $k_z^* = \frac{2\pi n}{h}$, where h is the period of the helical turn and n is the order of diffraction $n = 1, 2....$ (c) The broad extended peaks (layer lines) on the top and the bottom are formed by diffraction on small periodical features - the base pairs. (d) The missing diffraction spots is an indication of a double helix. (e) The position of the maxima of the diffraction spots is related to the radius of the DNA helix.

Opt Express, 2018, 26(23):30991-31017

1957 – John Kendrew

662 NATURE March 8, 1958 Vol. 181

A THREE-DIMENSIONAL MODEL OF THE MYOGLOBIN MOLECULE OBTAINED BY X-RAY ANALYSIS

By Drs. J. C. KENDREW, G. BODO, H. M. DINTZIS, R. G. PARRISH and H. WYCKOFF
Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge
AND
D. C. PHILLIPS
Davy Faraday Laboratory, The Royal Institution, London

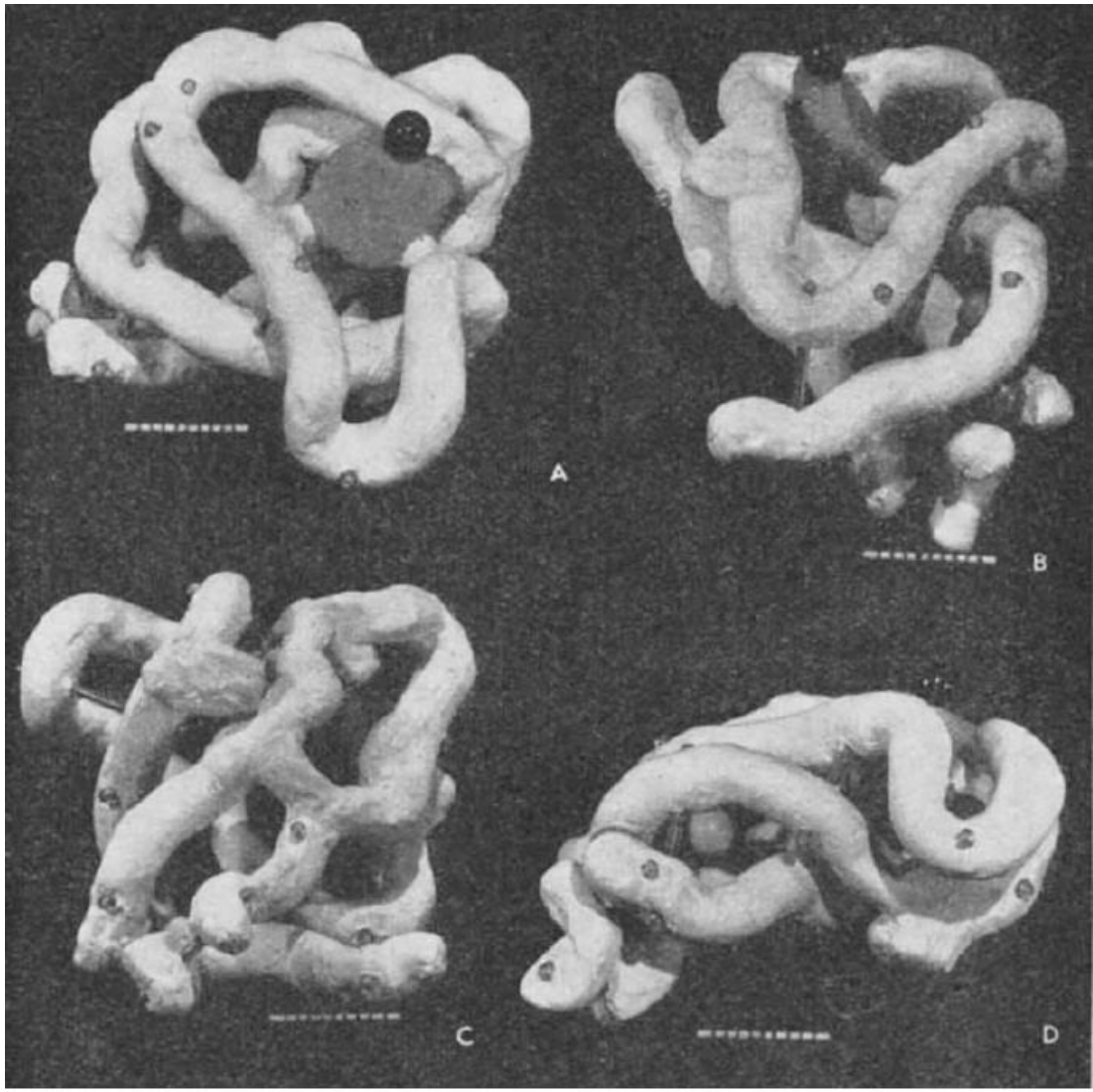


Fig. 2. Photographs of a model of the myoglobin molecule. Polypeptide chains are white; the grey disk is the hæm group. The three spheres show positions at which heavy atoms were attached to the molecule (black: Hg of *p*-chloro-mercuri-benzene-sulphonate; dark grey: Hg of mercury diammine; light grey: Au of auri-chloride). The marks on the scale are 1 A. apart

1971 – Protein Data Bank (PDB)

NATURE NEW BIOLOGY VOL. 233 OCTOBER 20 1971

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is received. The total holding will be announced annually in the organic bibliographic volumes of the reference series "Molecular Structures and Dimensions" published for the Crystallographic Data Centre and the International Union of Crystallography by Oosthoek's, Utrecht.

The success of the proposed system will depend on the response of the protein crystallographers supplying data. These will be accepted either "raw" or refined, in machine-readable form or as manuscripts. Laboratories intending to join the scheme should communicate with Mrs Olga Kennard or Dr D. G. Watson at the University Chemical Laboratories, Lensfield Road, Cambridge, who are responsible for the organization of the system. Data can be submitted to Cambridge or to Dr W. C. Hamilton at the Brookhaven National Laboratory, Upton, New York 11973, where the data will be computer processed.

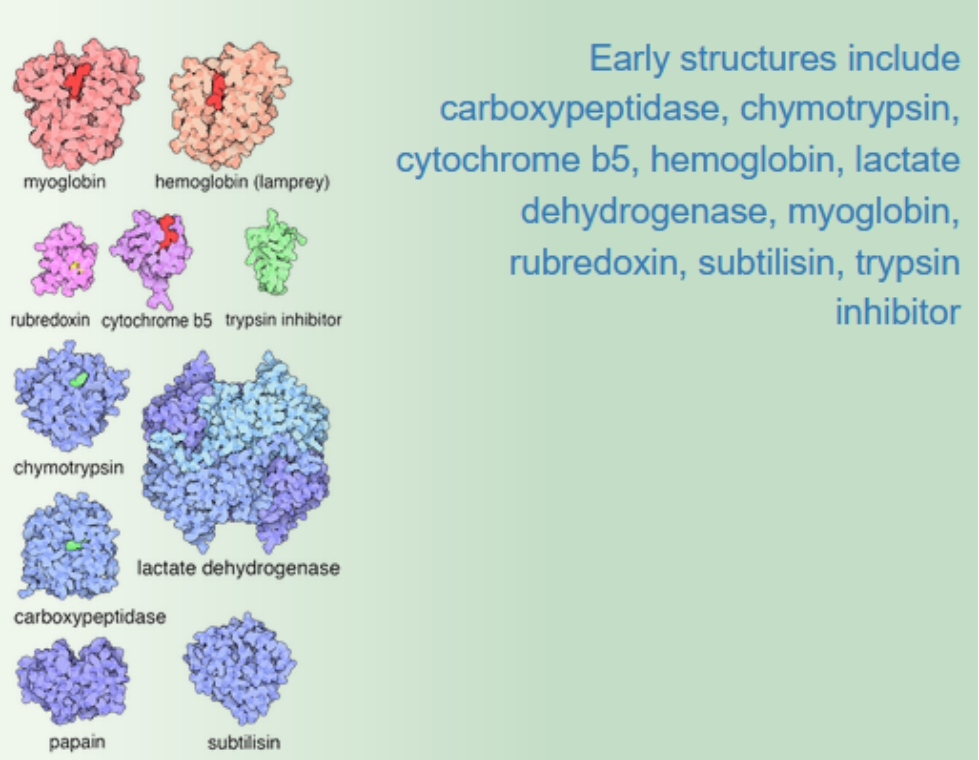
The two centres will maintain identical files and both will provide data services. The new data bank is intended to supplement existing publication media so that depositing material in this form is not a substitute for the publication of the results of structural investigations in a scientific journal.

➔established in 1971 at Brookhaven National Laboratory

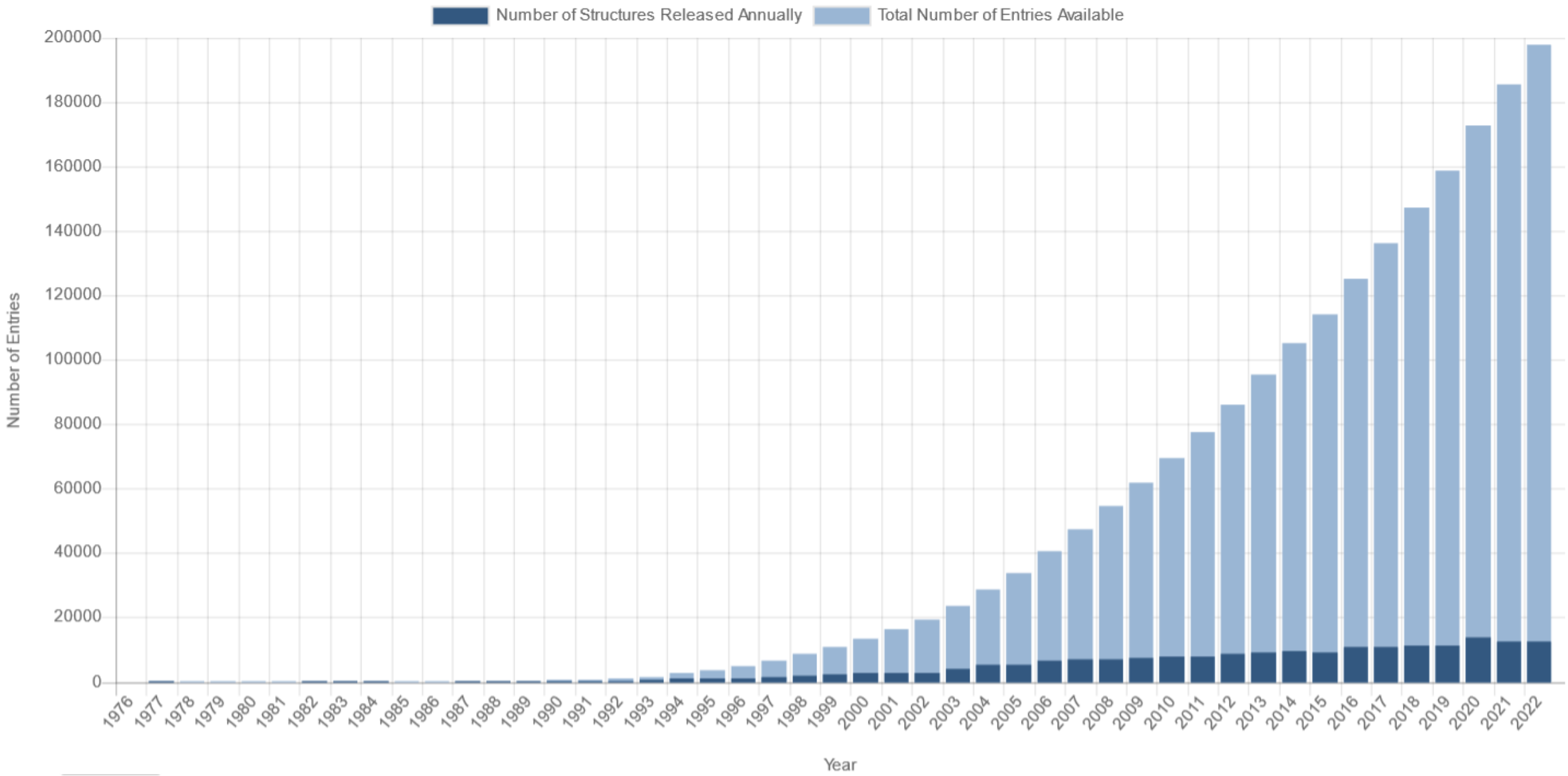
➔Member organisations: PDBe (Europe), PDBj (Japan), RCSB (US),

BMRB (NMR depositions), EMBD (EM depositions)

➔ PDB is overseen by Worldwide Protein Data Bank (wwPDB)



PDB

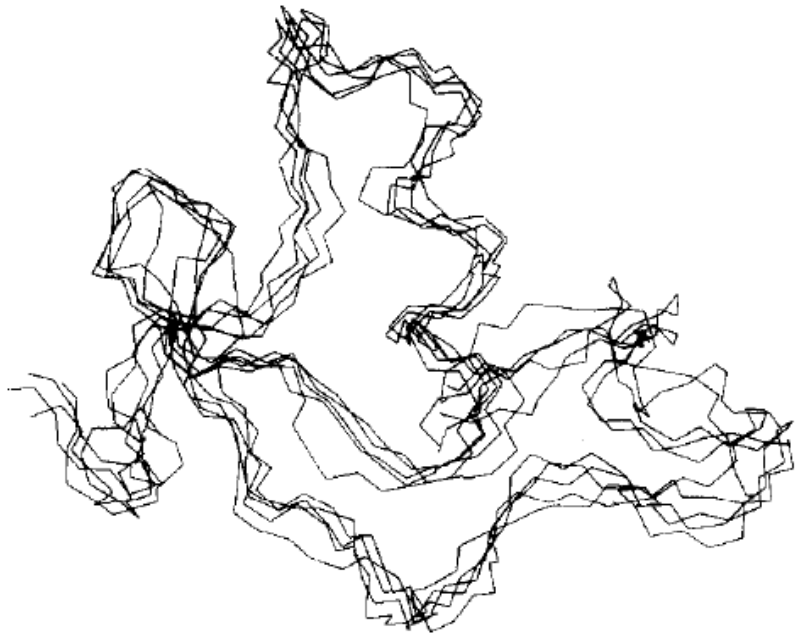

Number of Structures Released Annually
Total Number of Entries Available
Number of Entries
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year

1985 – First NMR protein structure

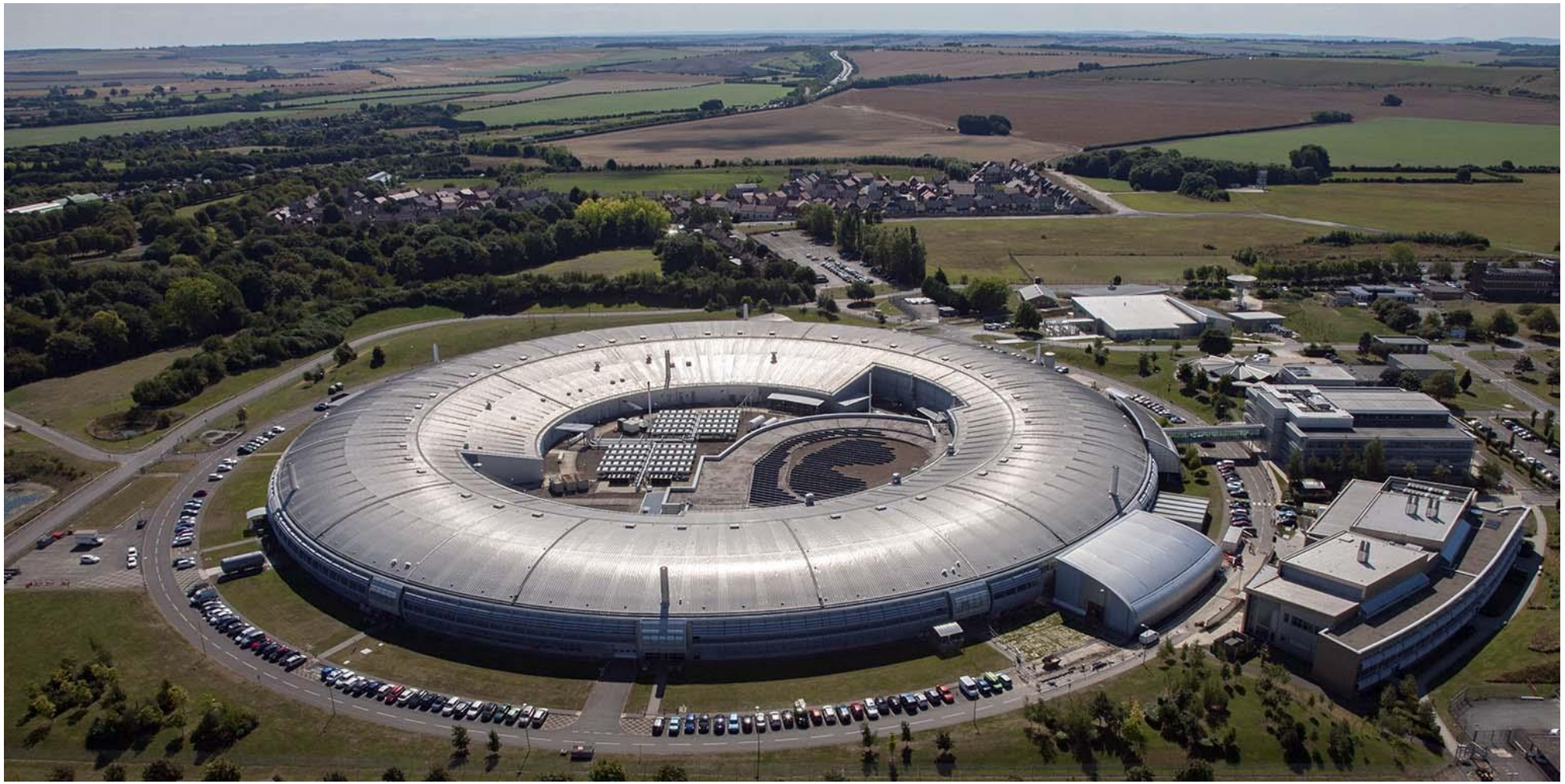
Solution Conformation of Proteinase Inhibitor IIA from Bull Seminal Plasma by ^{1}H Nuclear Magnetic Resonance and Distance Geometry

Michael P. Williamson, Timothy F. Havel and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule-Hönggerberg
CH-8093 Zürich. Switzerland



Synchrotrons revolutionized MX



➔ HT crystallography: High resolution structures
Automation of data collection and refinement
Suitable for ligand screening in drug discovery

2012 – 2014 Resolution Revolution in cryo-EM

Science, 2014, 343(6178):1443-4.

BIOCHEMISTRY

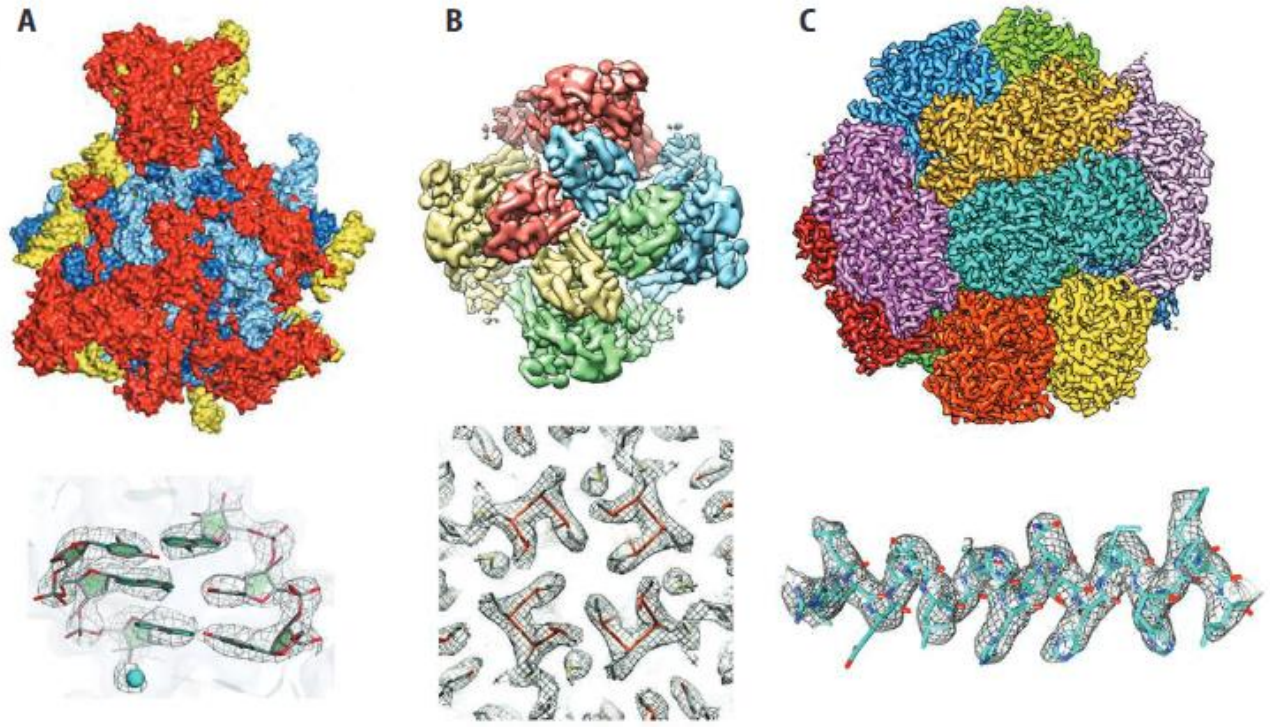
The Resolution Revolution

Werner Kühlbrandt

Advances in detector technology and image processing are yielding high-resolution electron cryo-microscopy structures of biomolecules.

Precise knowledge of the structure of macromolecules in the cell is essential for understanding how they function. Structures of large macromolecules can now be obtained at near-atomic resolution by averaging thousands of electron microscope images recorded before radiation damage accumulates. This is what Amunts *et al*. have done in their research article on page 1485 of this issue (*1*), reporting the structure of the large subunit of the mitochondrial ribosome at 3.2 Å resolution by electron cryo-microscopy (cryo-EM). Together with other recent high-resolution cryo-EM structures (*2*–*4*) (see the figure), this achievement heralds the beginning of a new era in molecular biology, where structures at near-atomic resolution are no longer the prerogative of x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

Ribosomes are ancient, massive protein-RNA complexes that translate the linear genetic code into three-dimensional proteins.



Near-atomic resolution with cryo-EM. **(A)** The large subunit of the yeast mitochondrial ribosome at 3.2 Å reported by Amunts *et al*. In the detailed view below, the base pairs of an RNA double helix and a magnesium ion (blue) are clearly resolved. **(B)** TRPV1 ion channel at 3.4 Å (*2*), with a detailed view of residues lining the ion pore on the four-fold axis of the tetrameric channel. **(C)** F_{420}-reducing [NiFe] hydrogenase at 3.36 Å (*3*). The detail shows an α helix in the FrhA subunit with resolved side chains. The maps are not drawn to scale.

Resolution Revolution in cryo-EM

THE REVOLUTION WILL NOT BE CRYSTALLIZED

MOVE OVER X-RAY CRYSTALLOGRAPHY. CRYO-ELECTRON MICROSCOPY IS KICKING UP A STORM IN STRUCTURAL BIOLOGY BY REVEALING THE HIDDEN MACHINERY OF THE CELL.

BY EWEN CALLAWAY

In a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

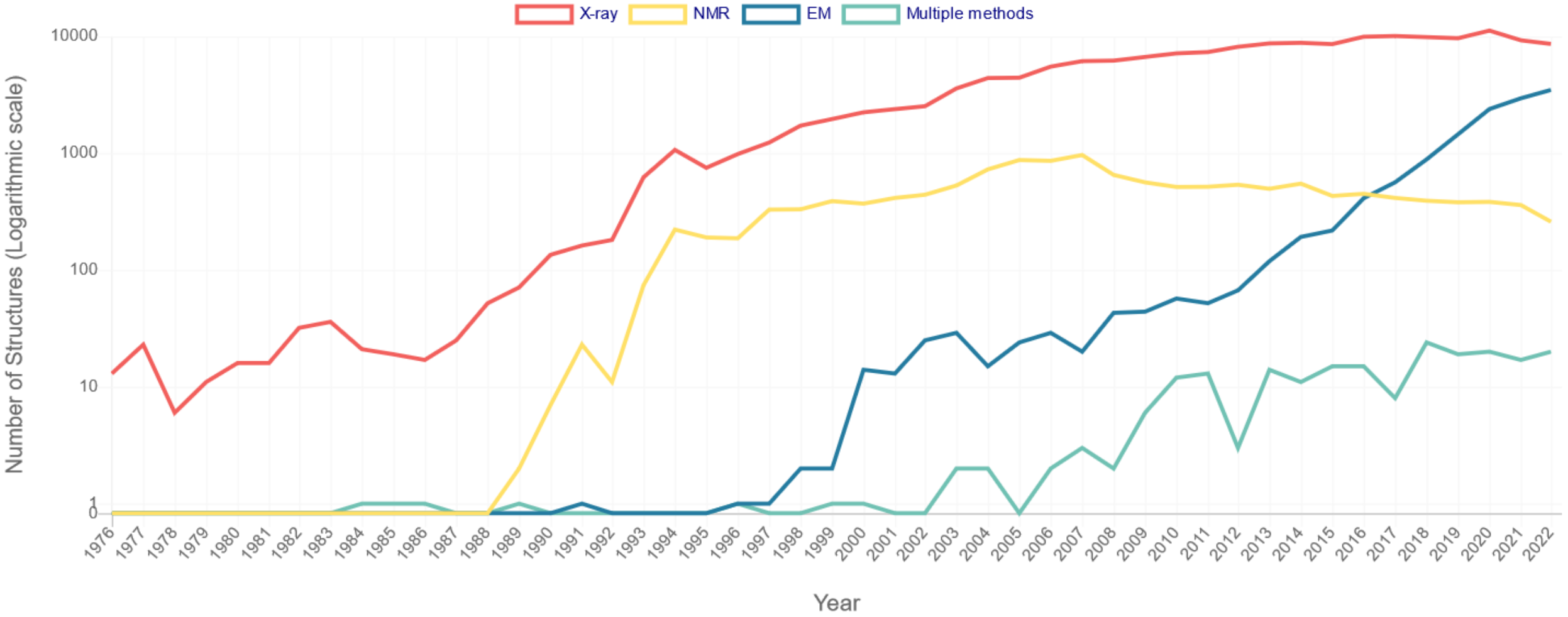
A hulking metal box, some three metres tall, is quietly beaming terabytes' worth of data through thick orange cables that disappear off through the ceiling. It is one of the world's most advanced cryo-electron microscopes: a device that uses electron beams to photograph frozen biological molecules and lay bare their molecular shapes. The microscope is so sensitive that a shout can ruin an experiment, says Sjors Scheres, a structural biologist at the UK Medical Research Council Laboratory of Molecular Biology (LMB), as he stands dwarfed beside the £5-million (US$7.7-million) piece of equipment. "The UK needs many more of these, because there's going to be a boom," he predicts.

In labs around the world, cryo-electron microscopes such as this one are sending tremors through the field of structural biology. In the past three years, they have revealed exquisite details of protein-making ribosomes, quivering membrane proteins and other key cell molecules,

ILLUSTRATION BY VICTOR HOEN

Nature, 2015, 525(7568):172-4

PDB


X-ray
NMR
EM
Multiple methods
Number of Structures (Logarithmic scale)
10000
1000
100
10
1
0
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year

AlphaFold2 from DeepMind – the next revolution

Article

Highly accurate protein structure prediction with AlphaFold

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

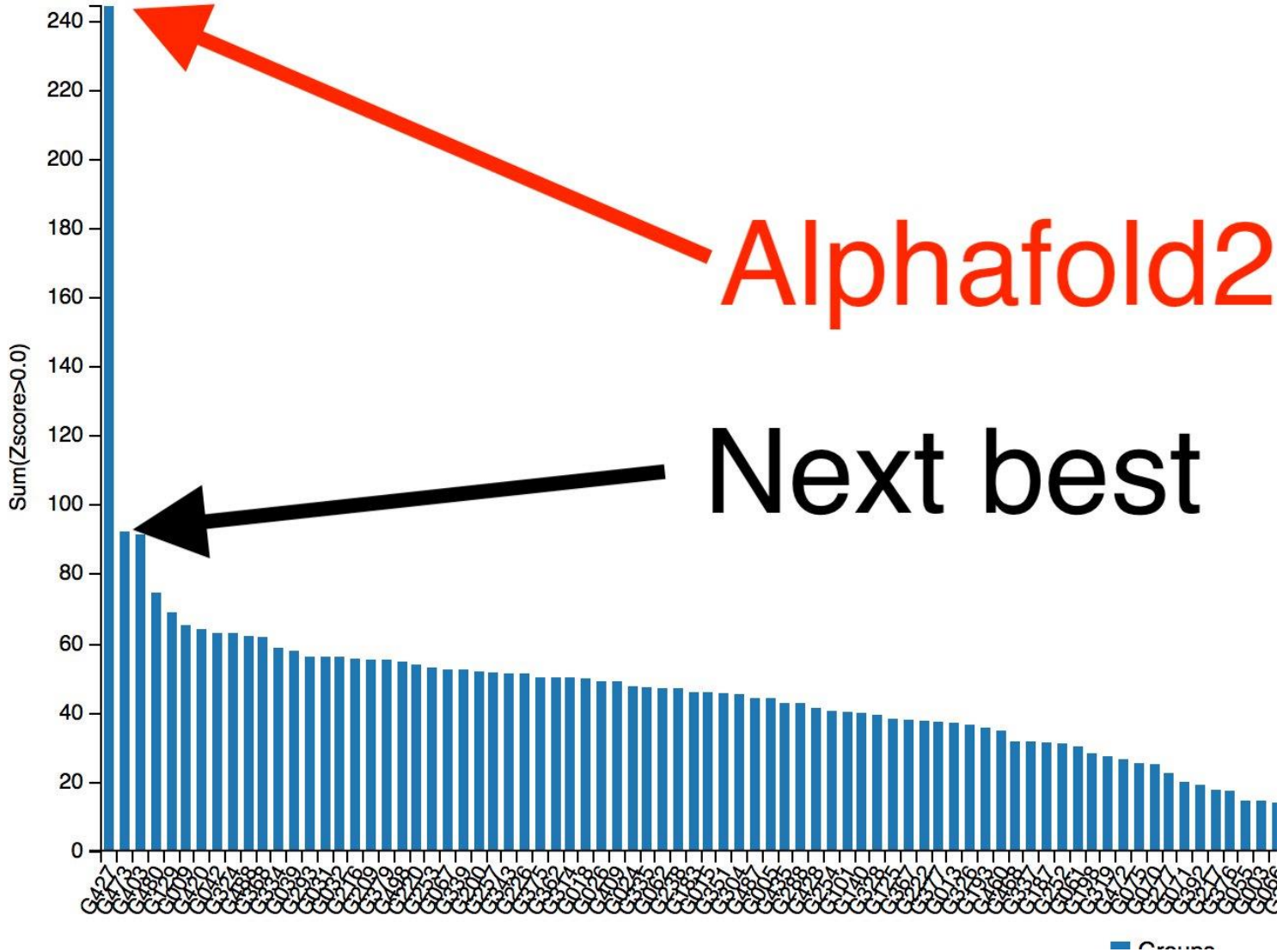
John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], 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]✉

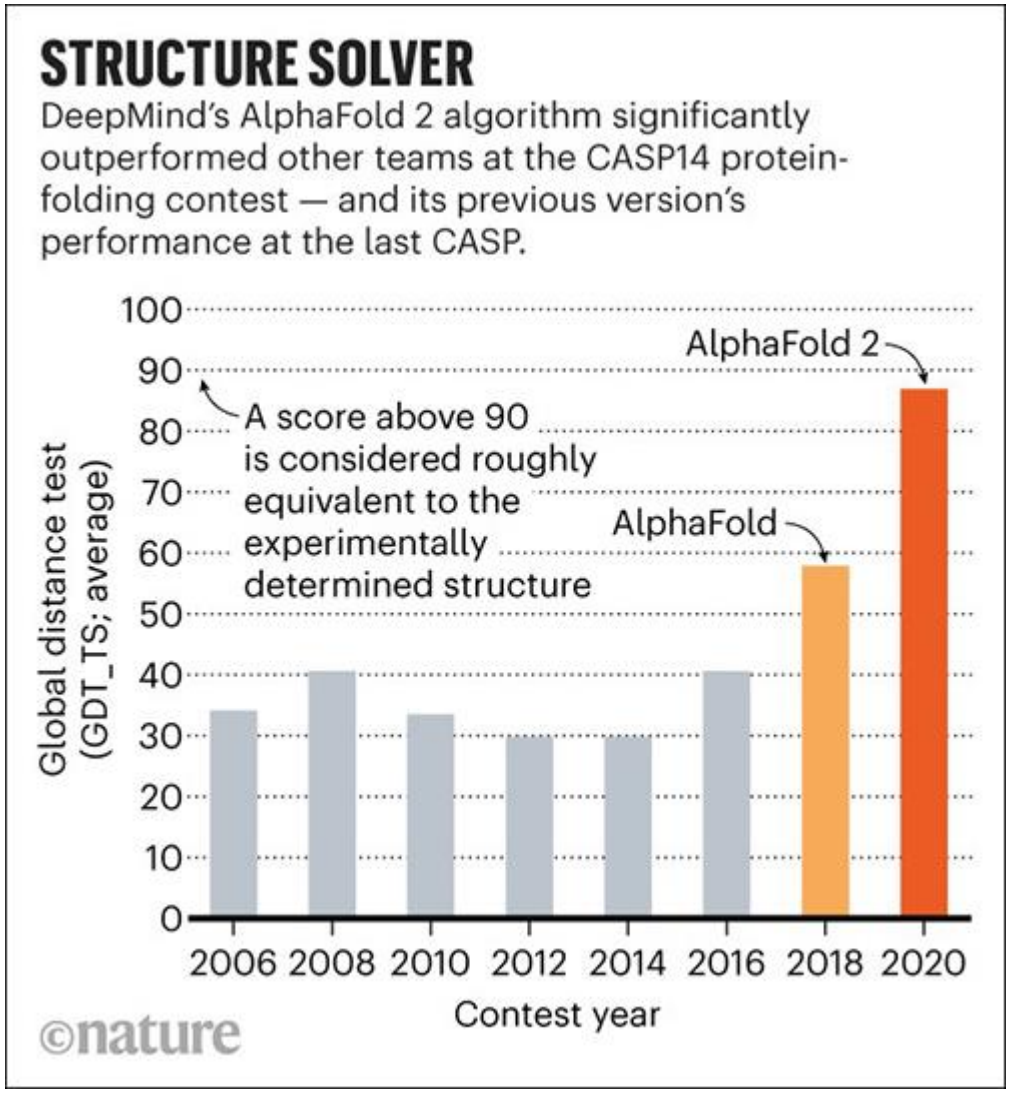
Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1–4], the structures of around 100,000 unique proteins have been determined[5], but this represents a small fraction of the billions 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 that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)[15], demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

Nature, 2021, 597(7873):583-589

CASP14

CASP = **C**ritical **A**ssessment of protein **S**tructure **P**rediction





AlphaFold2 - https://alphafold.ebi.ac.uk/

Information

Protein	Probable disease resistance protein At1g58602
Gene	At1g58602
Source organism	Arabidopsis thaliana (Mouse-ear cress) go to search
UniProt	Q8W3K0 go to UniProt
Experimental structures	None available in the PDB
Biological function	Potential disease resistance protein. go to UniProt

AlphaFold produces a per-residue confidence metric called predicted local distance difference test (pLDDT) on a scale from 0 to 100

3D viewer

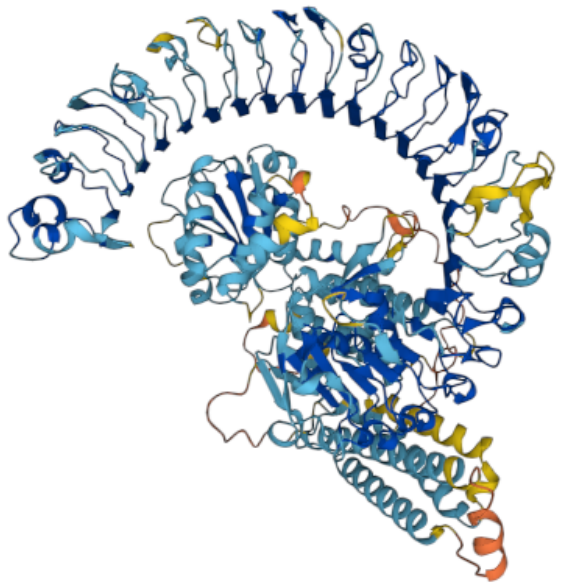
Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q8W3K0-F1 Chain 1: Probable dis... A

```
MAGELVSFAVNKLWDLLSHEYTLFQGVEDQVAELKSDLNLLKSFLKDADAKKHTSALVRYCVEEIKDIVYDAEDVLETFVQKEKLGTTSGIRKHIKRLTCIVPDRREIALYIGHVSKRITRVI
RDMQSFGVQQMIVDDYMHPLRNREREIRRTFPKDNESGFVALEENVKKLVGYFVEEDNYQVVSITGMGGLGKTILARQVFNHDMVTKKFDKLAWVSVSQDFTLKNVWQNILGDLKPKEEETKE
EEKKILEMTEYTLQRELYQLLEMSKSLIVLDDIWKKEDWEVIKPIFPPTKGWKLLLTSRNESIVAPTNTKYFNFKPECLKTDDSWKLFQRIAFPINDASEFEIDEEMEKLGEKMIEHCGGLPL
```



RoseTTAFold – Baker lab

➔ Developed while waiting for the release of the AlphaFold2 prediction

Science

RESEARCH ARTICLES

Cite as: M. Baek *et al.*, *Science*
10.1126/science.abj8754 (2021).

Accurate prediction of protein structures and interactions using a three-track neural network

Minkyung Baek[1,2], Frank DiMaio[1,2], Ivan Anishchenko[1,2], Justas Dauparas[1,2], Sergey Ovchinnikov[3,4], Gyu Rie Lee[1,2], Jue Wang[1,2], Qian Cong[5,6], Lisa N. Kinch[7], R. Dustin Schaeffer[6], Claudia Millán[8], Hahnbeom Park[1,2], Carson Adams[1,2], Caleb R. Glassman[9,10], Andy DeGiovanni[12], Jose H. Pereira[12], Andria V. Rodrigues[12], Alberdina A. van Dijk[13], Ana C. Ebrecht[13], Diederik J. Opperman[14], Theo Sagmeister[15], Christoph Buhlheller[15,16], Tea Pavkov-Keller[15,17], Manoj K. Rathinaswamy[18], Udit Dalwadi[19], Calvin K. Yip[19], John E. Burke[18], K. Christopher Garcia[9,10,11,20], Nick V. Grishin[6,21,7], Paul D. Adams[12,22], Randy J. Read[8], David Baker[1,2,23]*

Accurate Mutation Effect Prediction using RoseTTAFold

Sanaa Mansoor[1,2,3], Minkyung Baek[1,2,4], David Juergens[1,2,3], Joseph L. Watson[1,2], David Baker[1,2,5]

1. Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.
2. Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.
3. Molecular Engineering Graduate Program, University of Washington, WA 98195, USA.
4. School of Biological Sciences, Seoul National University, Seoul, 08826, Republic of Korea.
5. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

Meta

➔ 617 million structure predictions in two weeks from uncharacterized bacteria, viruses and microorganisms

Evolutionary-scale prediction of atomic level protein structure with a language model

Zeming Lin [1 2 *] Halil Akin [1 *] Roshan Rao [1 *] Brian Hie [1 3 *] Zhongkai Zhu [1] Wenting Lu [1] Nikita Smetanin [1]
Robert Verkuil [1] Ori Kabeli [1] Yaniv Shmueli [1] Allan dos Santos Costa [4] Maryam Fazel-Zarandi [1] Tom Sercu [1 †]
Salvatore Candido [1 †] Alexander Rives [1 † ‡]

How does the Future of Structural Biology look like?

How can I be a part of it?

Interpretation of structure prediction becomes important

➔ AlphaFold2 structure prediction will become the first step of any structural biology project

→ provides information on domain boundaries

→ provides model for Molecular Replacement

➔ solid background in structural biology and protein biochemistry remains essential for AlphaFold2 structure interpretations

➔ combination with molecular dynamics calculations could reduce the need of experimental structural biology to understand protein structure-function relationships

➔ Limitations:

→ AlphaFold2 provides a single prediction but many proteins have multiple conformations

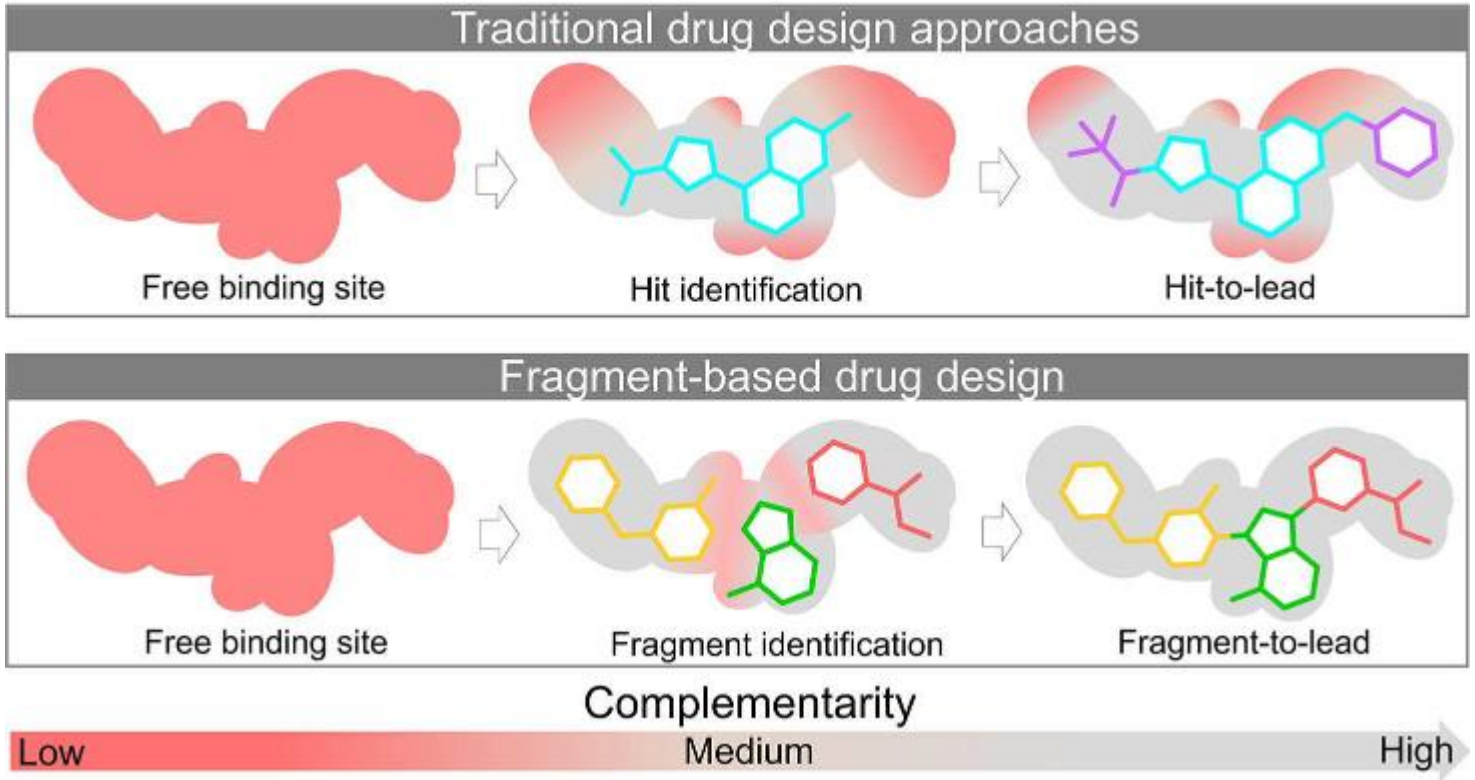
→ AlphaFold2 not suitable for point mutation structural alterations

X-ray crystallography vs. Cryo-EM

➔ X-Ray crystallography still remains the number one technique for 3D structure determination

➔ crystallographic fragment screening for fragment-based drug discovery

→ possible due to automation (XChem @ I04-1), detector speed and structure refinement pipelines



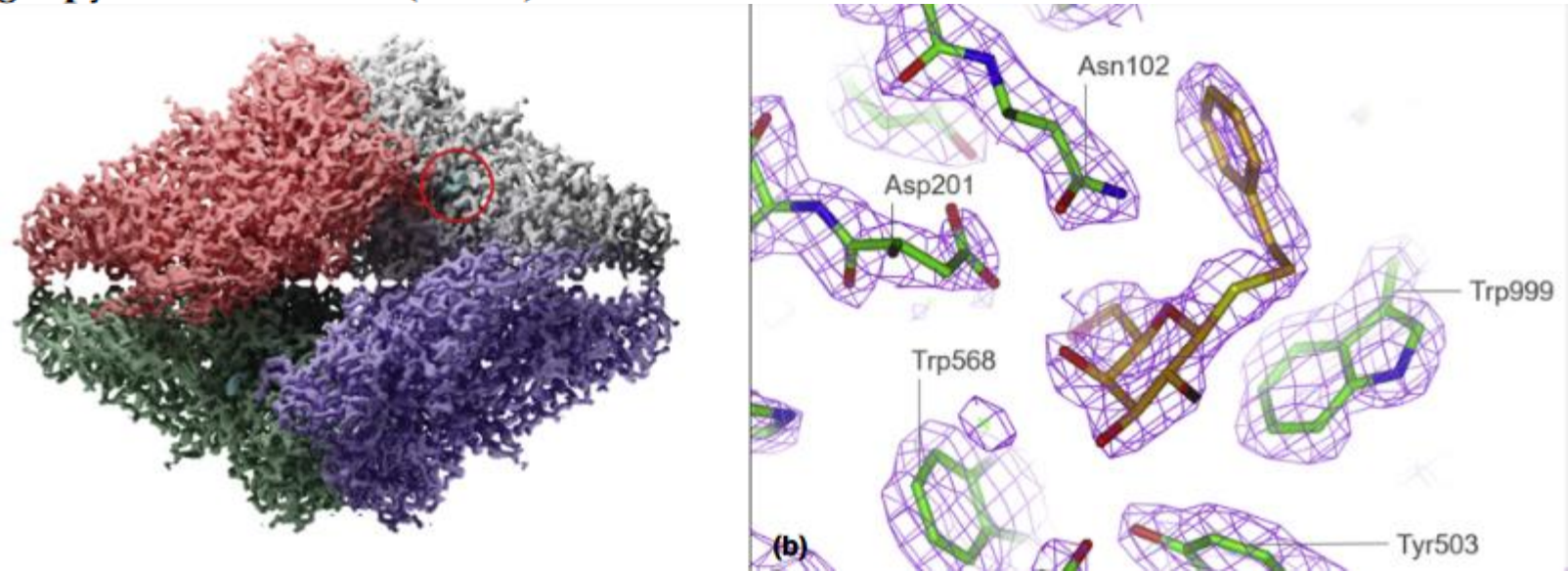
Front Chem., 2020, 8:93

X-ray crystallography vs. Cryo-EM

Fragment-based drug discovery using cryo-EM

Michael Saur[1,‡], Michael J. Hartshorn[2,‡], Jing Dong[1], Judith Reeks[1], Gabor Bunkoczi[1], Harren Jhoti[1], harren.jhoti@astx.com and Pamela A. Williams[1]

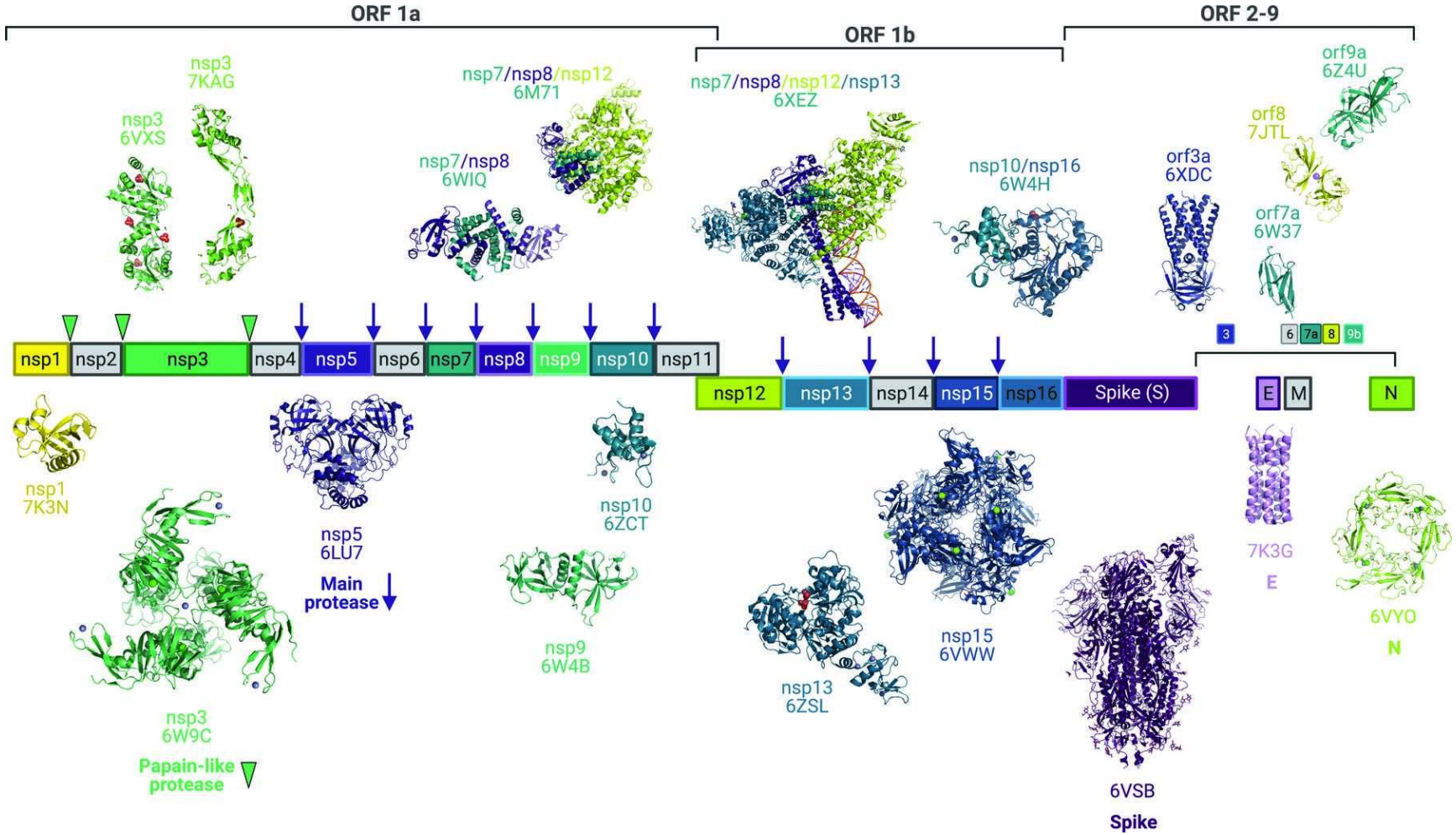
Recent advances in electron cryo-microscopy (cryo-EM) structure determination have pushed the resolutions obtainable by the method into the range widely considered to be of utility for drug discovery. Here, we review the use of cryo-EM in fragment-based drug discovery (FBDD) based on in-house method development. We demonstrate not only that cryo-EM can reveal details of the molecular interactions between fragments and a protein, but also that the current reproducibility, quality, and throughput are compatible with FBDD. We exemplify this using the test system β-galactosidase (Bgal) and the oncology target pyruvate kinase 2 (PKM2).



The Power of Structural Biology – example Covid19

IUCrJ

Structural biology in the time of COVID-19: perspectives on methods and milestones



- **Lynch *et al.***

- **Volume 8** | **Part 3** | May 2021 | Pages 335–341 | 10.11107/S2052252521003948

The Power of Structural Biology – example Covid19

RCSB.org/covid19:
PDB Structures (as of November 16, 2022)

Structure Determination Methodology

- [ ] experimental (2,784)

Experimental Method

- [ ] X-RAY DIFFRACTION (1,810)
- [ ] ELECTRON MICROSCOPY (959)
- [ ] SOLUTION NMR (13)
- [ ] NEUTRON DIFFRACTION (7)
- [ ] SOLID-STATE NMR (2)

Polymer Entity Type

- [ ] Protein (2,781)
- [ ] RNA (72)

Refinement Resolution (Å)

- [ ] 0.5 - 1.0 (25)
- [ ] 1.0 - 1.5 (495)
- [ ] 1.5 - 2.0 (575)
- [ ] 2.0 - 2.5 (471)
- [ ] 2.5 - 3.0 (368)
- [ ] 3.0 - 3.5 (451)
- [ ] 3.5 - 4.0 (279)
- [ ] 4.0 - 4.5 (74)
- [ ] > 4.5 (40)

PDBe-KB page	Data Summary
P0DTD1 - Replicase polyprotein 1ab The orf1ab polyprotein is a multifunctional protein involved in the transcription and replication of viral RNAs. It contains the proteinases responsible for the cleavages of the polyprotein.	1452 Structures · 1105 Ligands · 20 Interactions · 1 Functional Annotations · 11 Similar Proteins
PRO_0000449619 - Host translation Inhibitor nsp1 (nsp1) Inhibits host translation by interacting with the 40S ribosomal subunit. The nsp1-40S ribosome complex further induces an endonucleolytic cleavage near the 5'UTR of host mRNAs, targeting them for degradation. Viral mRNAs are not susceptible to nsp1-mediated endonucleolytic RNA cleavage thanks to the presence of a 5'-end leader sequence and are therefore protected from degradation. By suppressing host gene expression, nsp1 facilitates efficient viral gene expression in infected cells and evasion from host immune response.	21 Structures · 8 Ligands · 3 Interactions · 1 Functional Annotations · 1 Similar Proteins

Future directions in cryo-EM

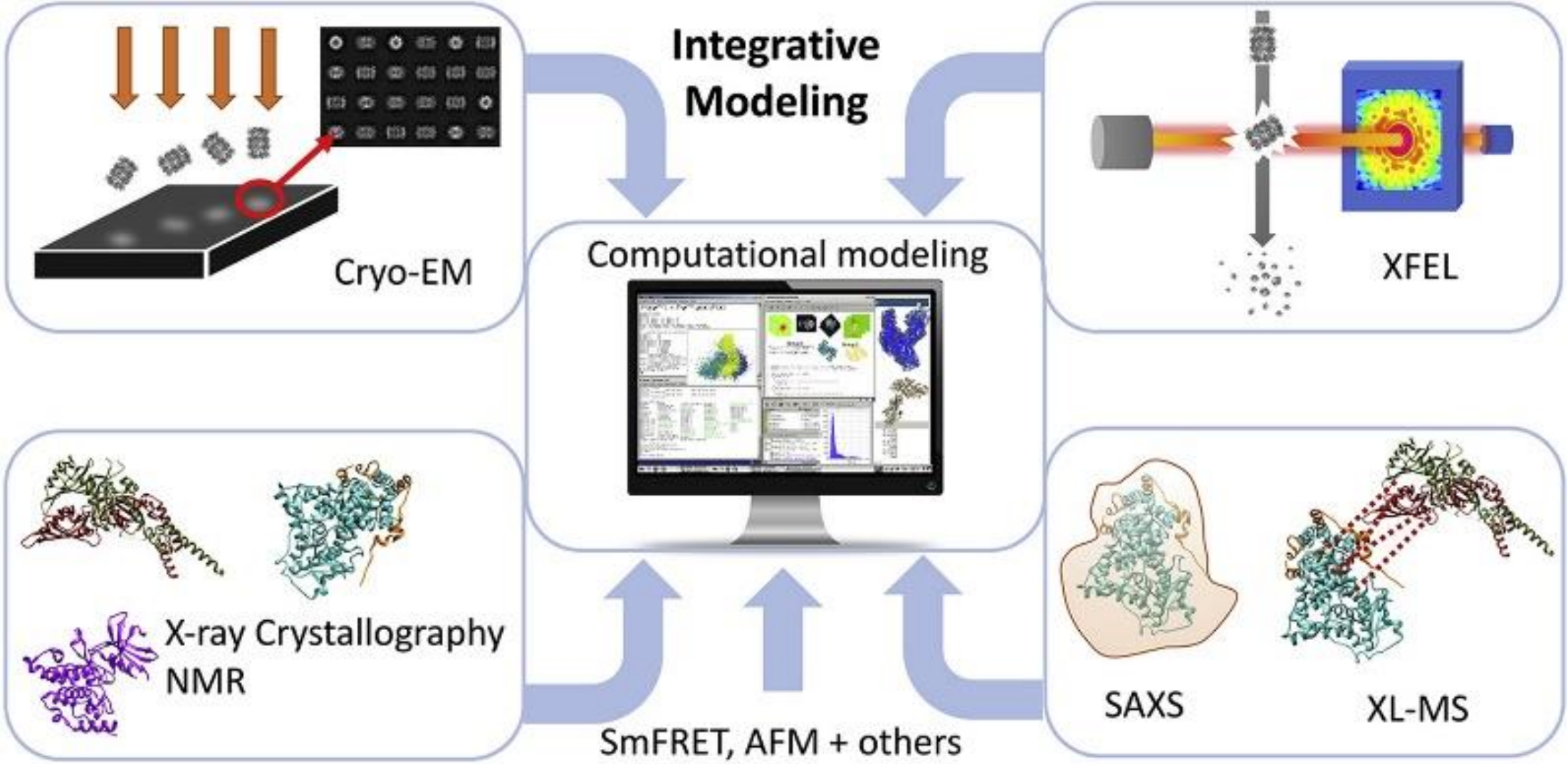
"We're now coming to the point where the easy samples have been done and people are looking at more complex problems," says Ardan Patwardhan, a structural biologist at the European Molecular Biology Laboratory-European Bioinformatics Institute in Hinxton, UK, who leads the team that runs the EMDB.

Henderson expects the boom in cryo-EM structures to slow at some point. One factor that could sap growth, he says, is the high cost of the most powerful microscopes, which can exceed £5 million (US$7 million). They also cost thousands of pounds each day to run, and require specialized laboratories that minimize vibrations. Henderson is campaigning to convince firms to develop cheaper, but still useful, microscopes that could spread the technique even further. "At the moment, you cannot go wrong by putting more investment into cryo-EM," he says.

Nature **578**, 201 (2020)

Integrative / Hybrid Modeling

➔ Combine data from multiple experiments to build a biomolecular complex model



J Mol Biol., 2020, 432(9):2846-2860

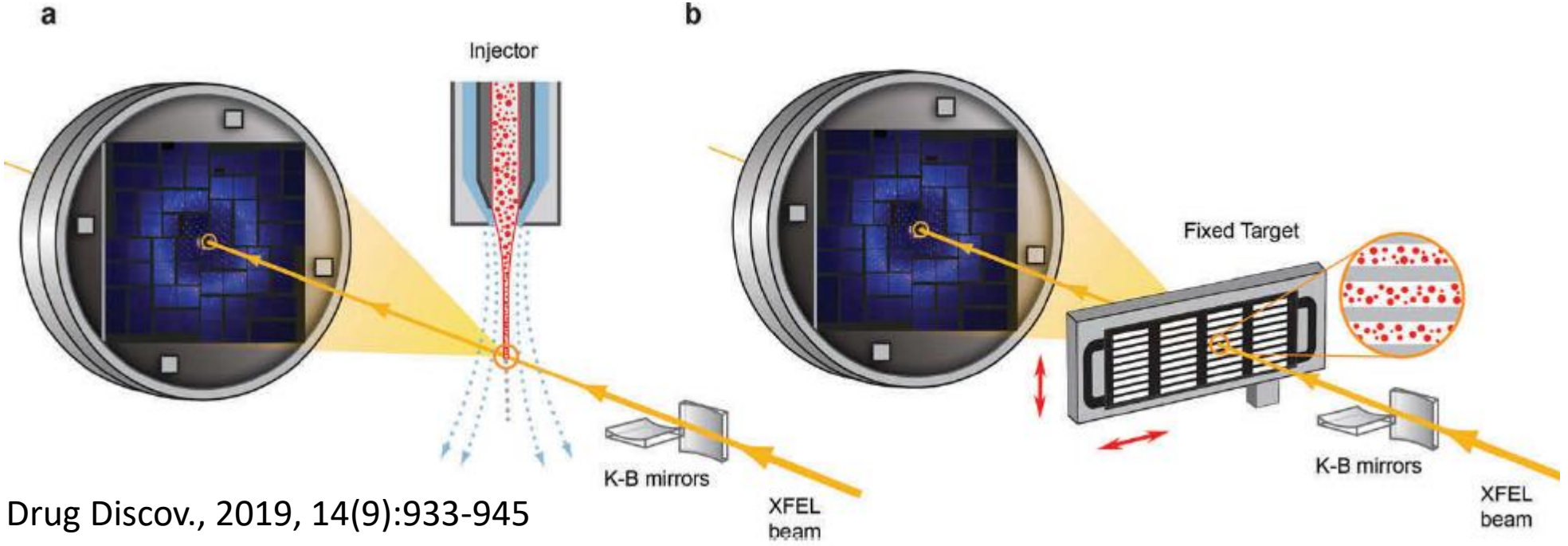
Serial Femtosecond Crystallography (SFX)

➔ "diffraction-before-destruction": femtosecond X-ray pulses from XFELs record diffraction image before sample is damaged

➔ Tiny crystals (micrometer- to nanometer) are used for data collection

➔ crystal structure determination at room temperature

➔ time-resolved measurements and observation of structural changes upon ligand binding

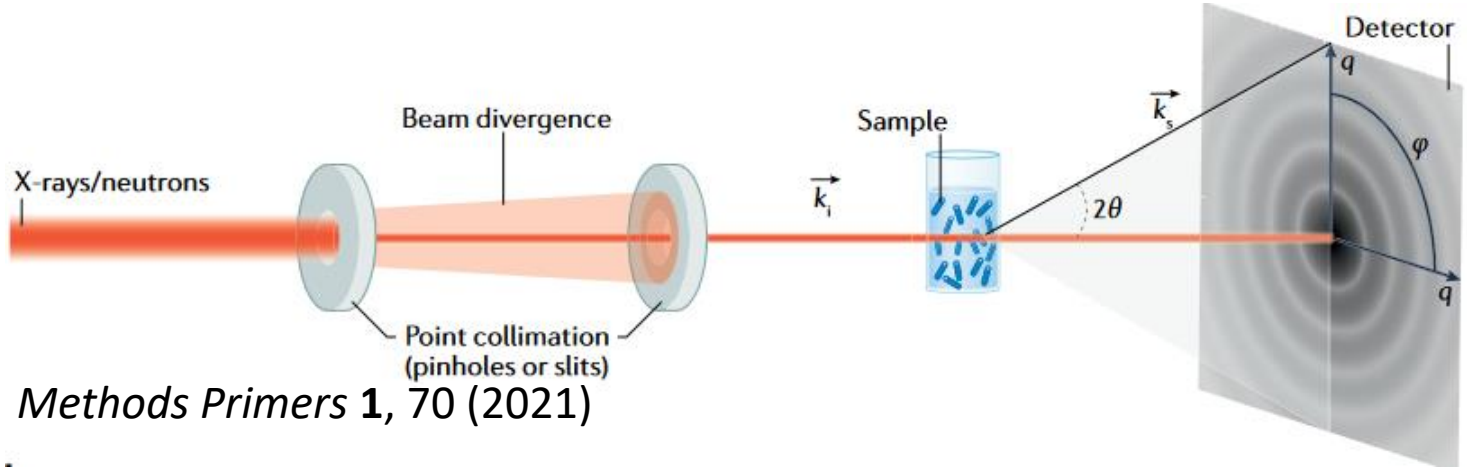


Expert Opin Drug Discov., 2019, 14(9):933-945

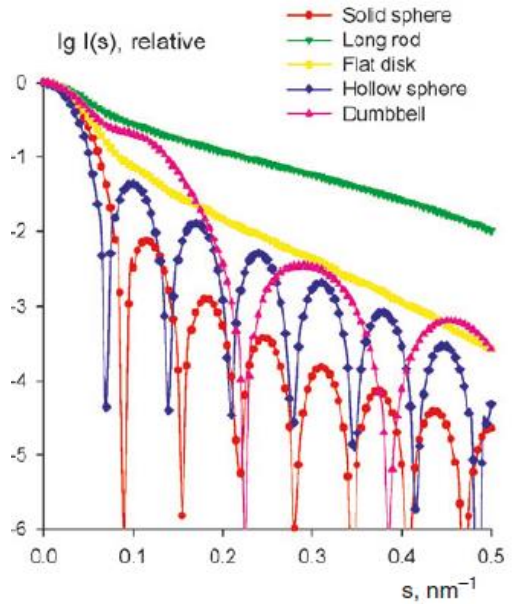
Small-Angle X-ray Scattering (SAXS)

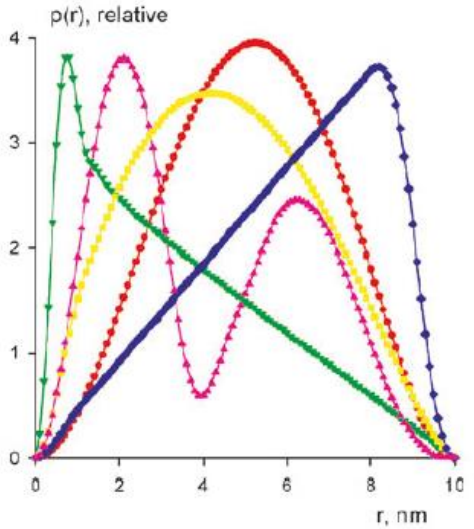
➔ information on size, shape and conformational flexibility

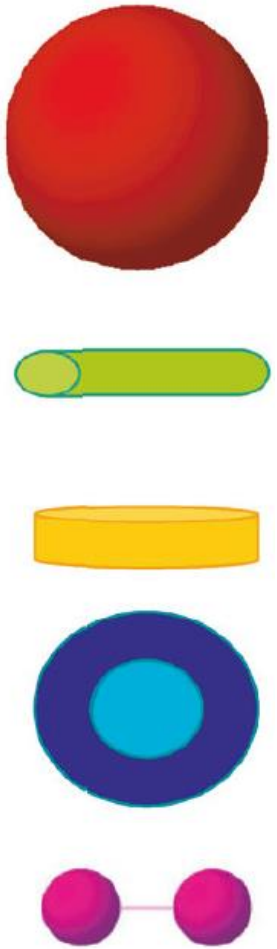
of biological macromolecules and complexes in solution



Methods Primers **1**, 70 (2021)

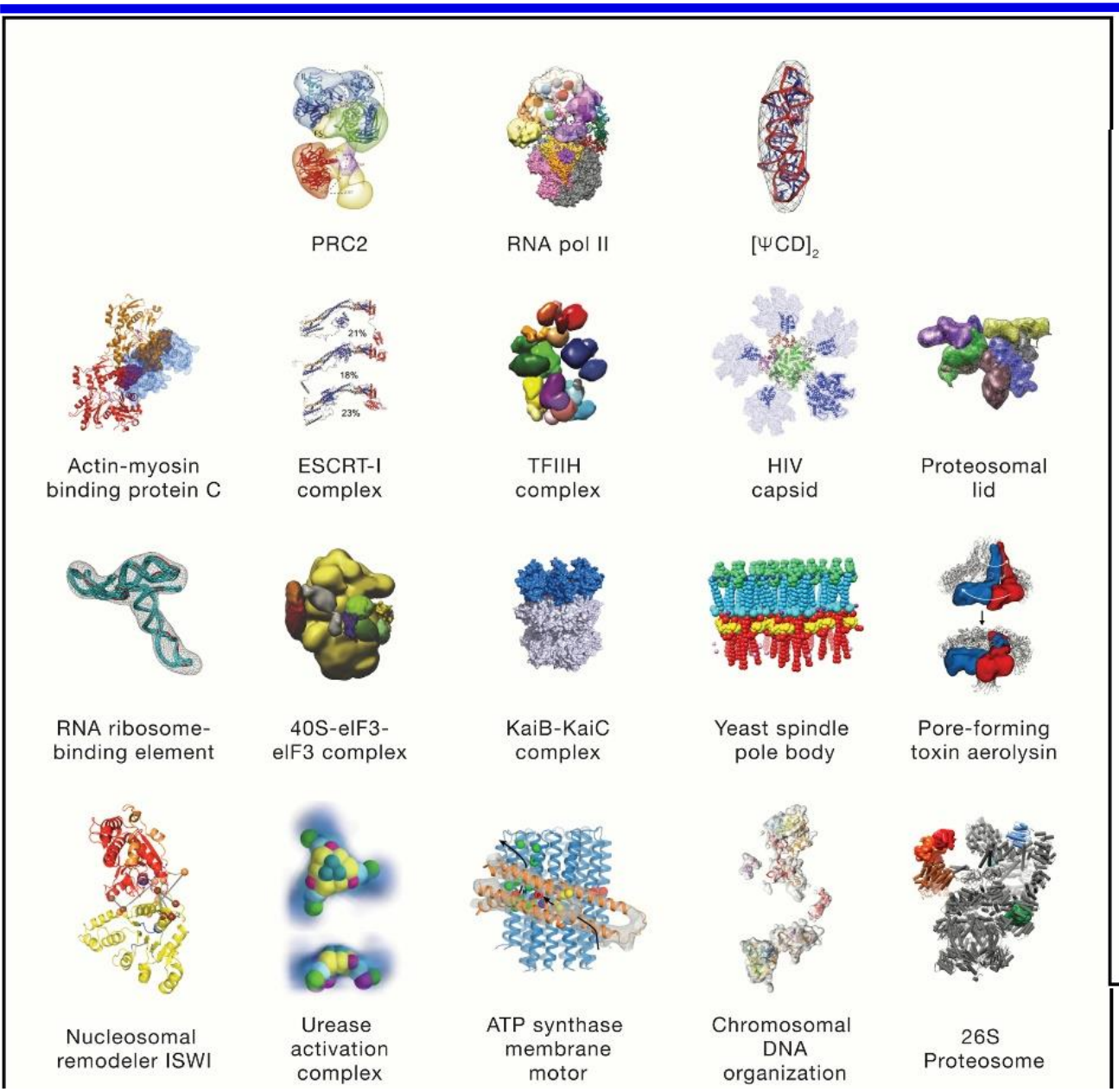


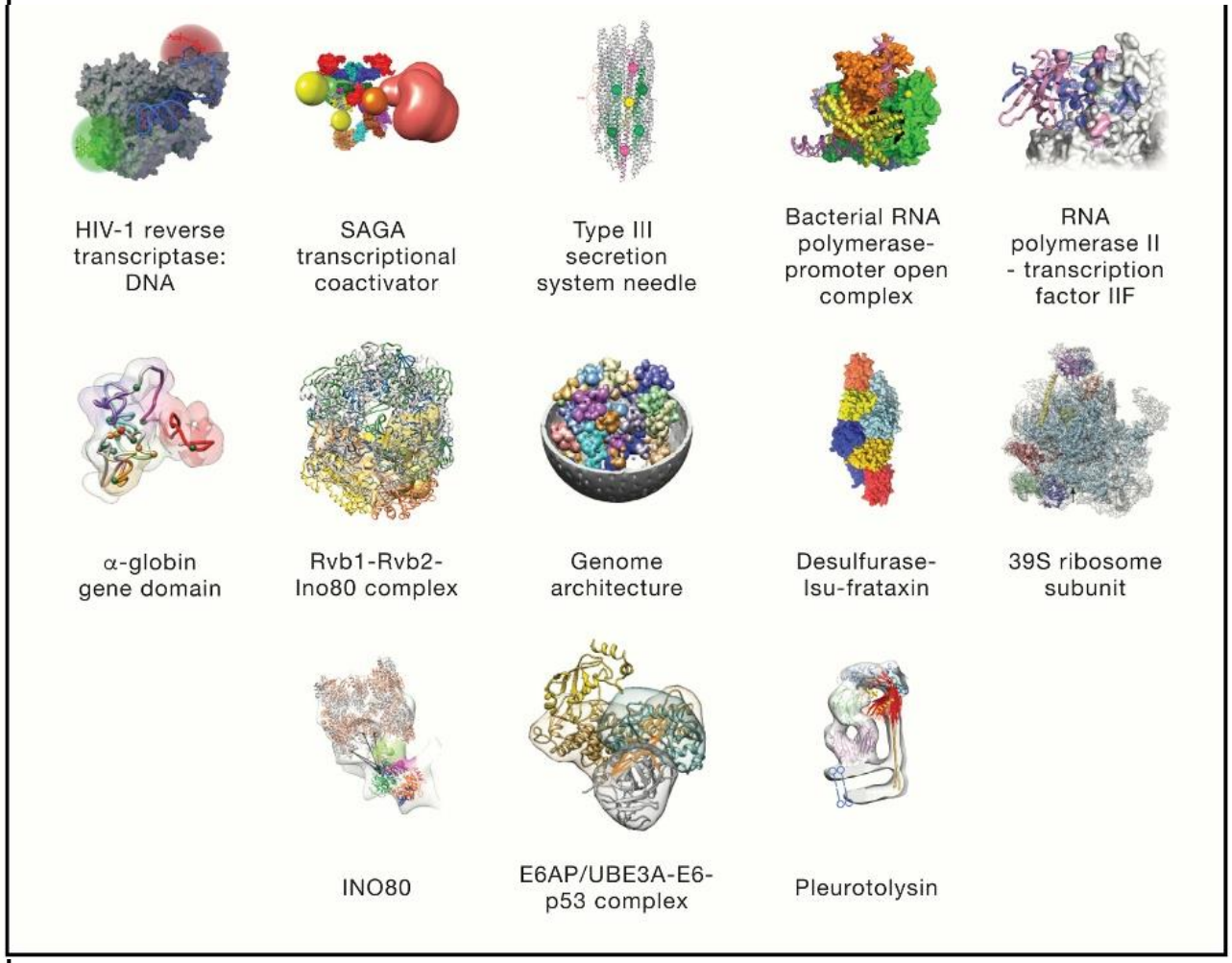




Rep. Prog. Phys. **66** (2003) 1735–1782

Integrative / Hybrid Modeling





Cell., 2019, 177(6):1384-1403

Integrative / Hybrid Modeling

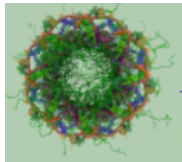
Table 1. Example Methods that Are Informative about a Variety of Structural Aspects of Biomolecular Systems

Structural information	Method
Stoichiometry	MS, quantitative fluorescence imaging
Atomic structures of parts of the studied system	X-ray and neutron crystallography, NMR spectroscopy, 3DEM, comparative modeling, and molecular docking
3D maps and 2D images	Electron microscopy and tomography
Atomic and protein distances	NMR, FRET, and other fluorescence techniques; DEER, EPR, and other spectroscopic techniques; and XL-MS and disulfide bonds detected by gel electrophoresis
Binding site mapping	NMR spectroscopy, mutagenesis, FRET, and XL-MS
Size, shape, and distributions of pairwise atomic distances	SAS
Shape and size	Atomic force microscopy, ion mobility mass spectrometry, fluorescence correlation spectroscopy, fluorescence anisotropy, and analytical ultracentrifugation
Component positions	Super-resolution optical microscopy, FRET imaging, and immuno-electron microscopy
Physical proximity	Co-purification, native mass spectrometry, XL-MS, molecular genetic methods, and gene/protein sequence covariance
Solvent accessibility	Footprinting methods, including HDex assessed by MS or NMR, and even functional consequences of point mutations
Proximity between different genome segments	chromosome conformation capture
Propensities for different interaction modes	Molecular mechanics force fields, potentials of mean force, statistical potentials, and sequence co-variation

Abbreviation are as follows: 3DEM, 3D electron microscopy; DEER, double electron-electron resonance; EPR, electron paramagnetic resonance; FRET, Foerster resonance energy transfer; HDex, hydrogen/deuterium exchange; NMR, nuclear magnetic resonance; SAS, small-angle scattering; XL-MS, cross-linking mass spectrometry.

Cell., 2019, 177(6):1384-1403

Integrative / Hybrid Modeling



PDB-Dev

Prototype Archiving System for Integrative Structures

Home | About | Deposit | Contact | FAQ | Browse Structures

Welcome to PDB-Dev

PDB-Dev is a prototype archiving system for structural models obtained using integrative or hybrid modeling and is funded by the NSF ABI Development Program. Structural characterization of many complex macromolecular assemblies is increasingly carried out using integrative modeling, where a combination of complementary experimental and computational techniques is used to determine the structure. The structural models obtained through integrative modeling are collected, archived and disseminated to the public through PDB-Dev. Once the mechanisms for processing integrative models are fully established through PDB-Dev, the key components will be integrated with the wwPDB OneDep system and the PDB-Dev holdings will be moved into the PDB.

PDB-Dev and Integrative Modeling

PDB-Dev is a prototype deposition and archiving system for structural models obtained using integrative or hybrid modeling. Structures of complex macromolecular assemblies are increasingly determined using integrative modeling, where a combination of complementary experimental and computational techniques is employed to model the structures. In addition to traditional structure determination methods such as X-ray crystallography (X-ray), NMR spectroscopy (NMR), and Electron Microscopy (3DEM), experimental techniques such as small angle scattering (SAS), atomic force microscopy (AFM), chemical cross-linking (CX), co-purification, Förster resonance energy transfer (FRET), electron paramagnetic resonance (EPR), mass spectrometry (MS), Hydrogen/Deuterium exchange (HDX), and various proteomics and bioinformatics approaches contribute to integrative modeling. Spatial restraints derived from the different kinds of experimental and computational methods are combined to derive the structure of the macromolecular assembly. Integrative modeling has been applied to determine the structures of complexes such as the nuclear pore complex and its sub-complexes, 16S rRNA complexed with methyltransferase A, human mitochondrial iron sulfur cluster core complex, the BBSome, ghrelin bound to its G protein-coupled receptor, complex of RNF168-RING domain and the nucleosome.

Integrative / Hybrid Modeling

First 100 structures in the PDB-Dev!

We are proud to announce an important milestone for PDB-Dev - the first 100 entries are released to the public. This achievement is a result of the continued collaboration among the PDB-Dev team, members of the wwPDB IHM task force, wwPDB leadership, RCSB PDB team members, and our pioneer depositors, who were the first to realize the importance of sharing the results of integrative structural biology investigations. The PDB-Dev project is funded by the United States National Science Foundation.

The first 100 entries also offer us a glimpse of the complex landscape of modern integrative modeling. While experimental models used with chemical cross-linking are a leading category, techniques such as three-dimensional electron microscopy (3DEM), small angle solution scattering (SAS), Forster resonance enery transfer (FRET), and others, together with powerful AI-based methods for the de novo structure prediction, give us more than 50(!) unique combinations of datasets used for integrative modeling (see figure below). Because integrative structures provide a unique opportunity to gain insights into how biomolecules function in conditions close to the native, many of these investigations are featured in journals such as Cell, Nature, Science, Proceedings of the National Academy of Sciences, and the Journal of Molecular Biology.

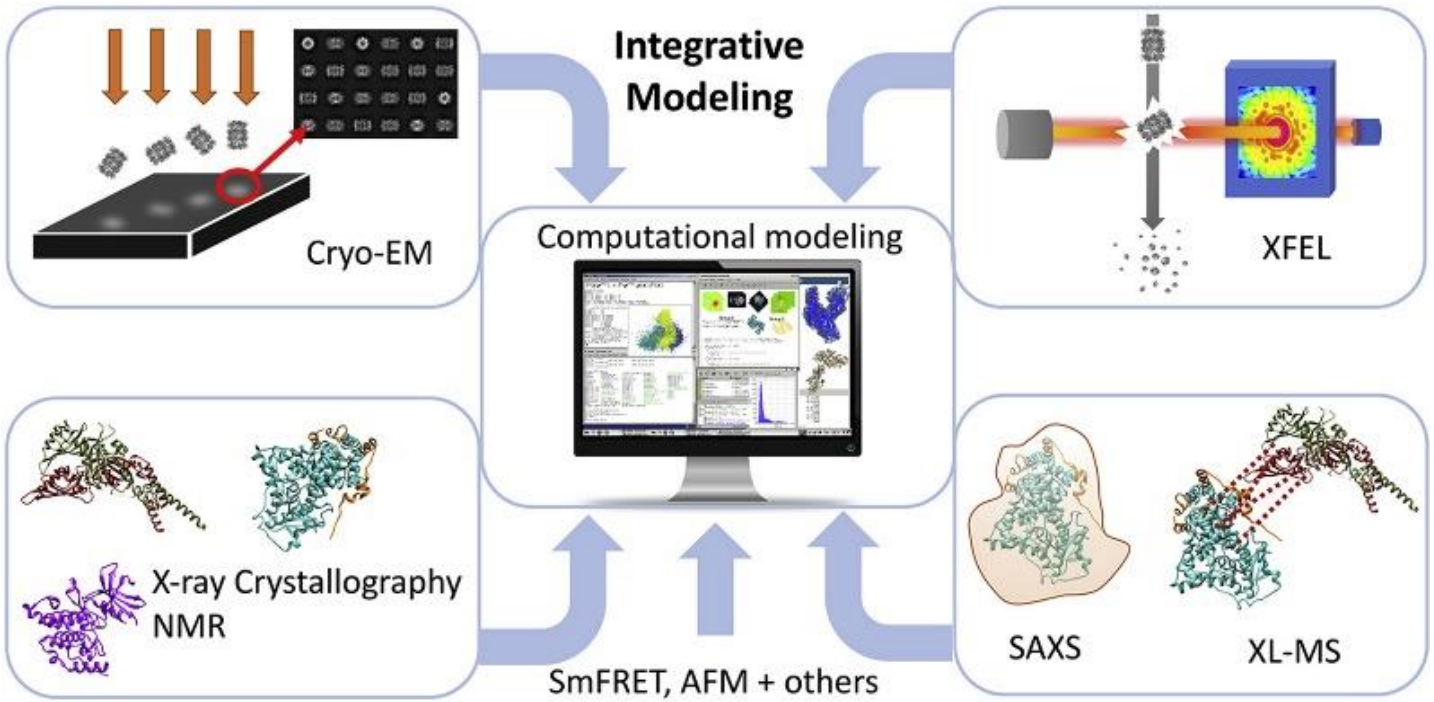
We designed PDB-Dev according to the FAIR guiding principles to make the integrative structures of biomolecules Findable, Accessible, Interoperable, and Reusable, and with the your help, we will continue to improve it and provide better archiving and structure validation services to our users.

First AlphaFold-based structure in PDB-Dev

For a long time, the primary sources of starting structures for integrative modeling were experimental structures from the Protein Data Bank or computational models generated using various comparative modeling approaches. The recent machine learning revolution has led to the development of modeling methods like AlphaFold2 and RoseTTAFold, and also brought a whole new universe of high-quality predicted structures of proteins. More than 200 million predicted structures of individual proteins and complexes are now available to researchers across multiple databases. Now these structures are finally making their way into integrative modeling.

In PDBDEV_00000141, Noone et. al., report a structure of pentraxin protein PTX3, a member of a family of soluble pattern recognition molecules that form an important part of innate immunity, where they facilitate the response to infections and damage by triggering processes such as inflammation. The complete structure of PTX3 was built by integrative modeling using three-dimensional electron microscopy map, mass spectrometry data, and AlphaFold-based starting models.

Questions?


Integrative
Modeling
Cryo-EM
Computational modeling
XFEL
X-ray Crystallography
NMR
SmFRET, AFM + others
SAXS
XL-MS