

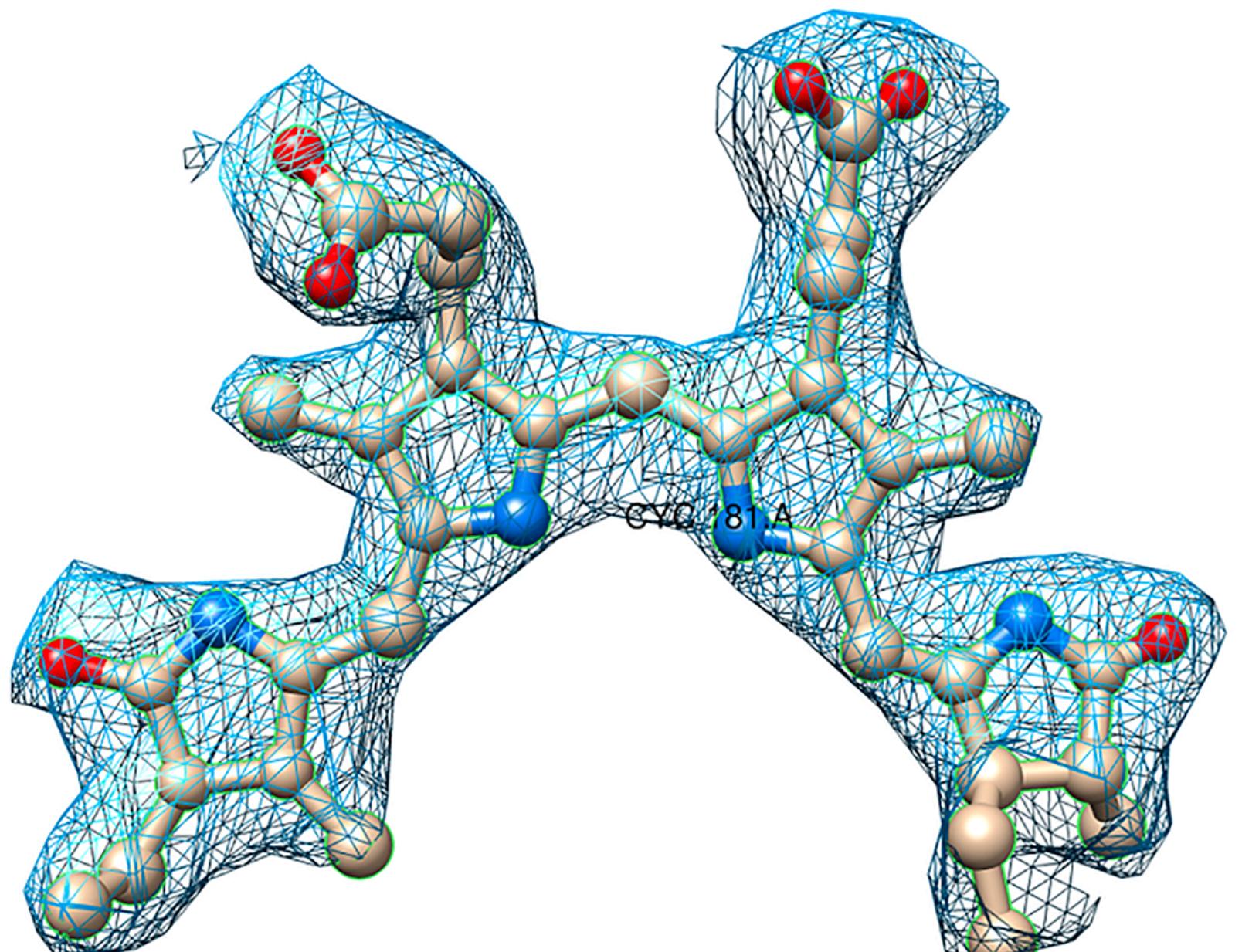
8/28/2024 Structure Determination

- This lecture is intended to help you achieve the following learning objective: Summarize the principles behind the main methods used to determine the structures of biological macromolecules. Compare their relative benefits and drawbacks.
- It will be about
 - General principles of structure determination
 - The Protein Data Bank
 - Experimental methods for structure determination
 - X-ray crystallography
 - Nuclear magnetic resonance (NMR)
 - Cryogenic Electron Microscopy (Cryo-EM)
- At the end of this lecture, you should be able to answer the following questions:
 - What are the major ways in which we obtain the structures used for molecular modeling?
 - What are the relative benefits and drawbacks of each technique?
- based on a lecture by Andy Howard (Summer 2021) adapted by Jennifer Sorescu (Spring 2022)

General structure determination principles

What is structure determination?

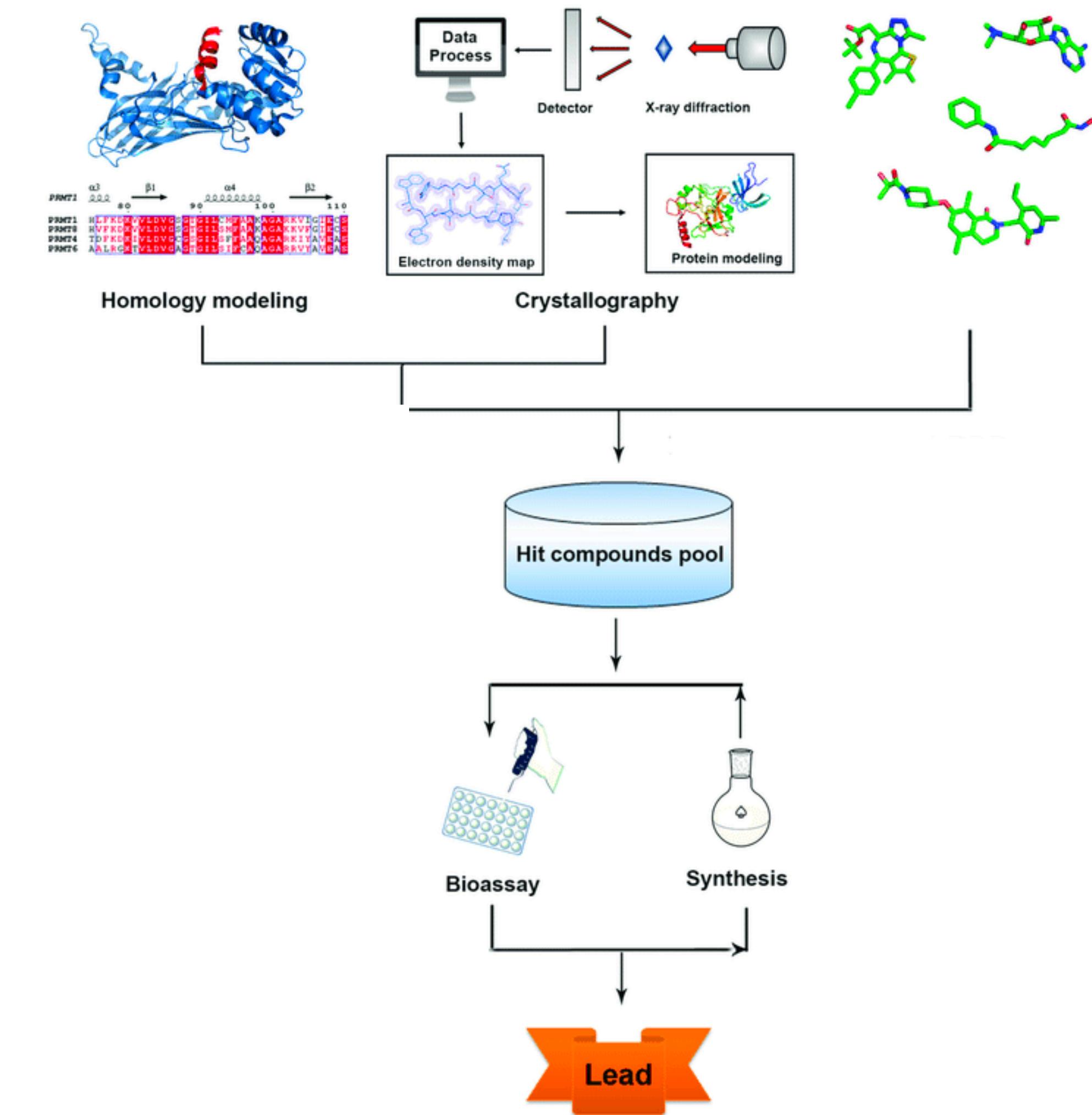
- Determining three-dimensional positions of atoms in a molecular system that are consistent with experimental measurements
- Structures do not come from experimental measurements, but are inferred. This means that “experimental” structures are models.
- If the positions of some atoms are not justified by the data, they may be omitted.
 - H atoms
 - Flexible regions in a protein
- Sometimes more than one model is a good fit to the data.



<https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0124580.g004>

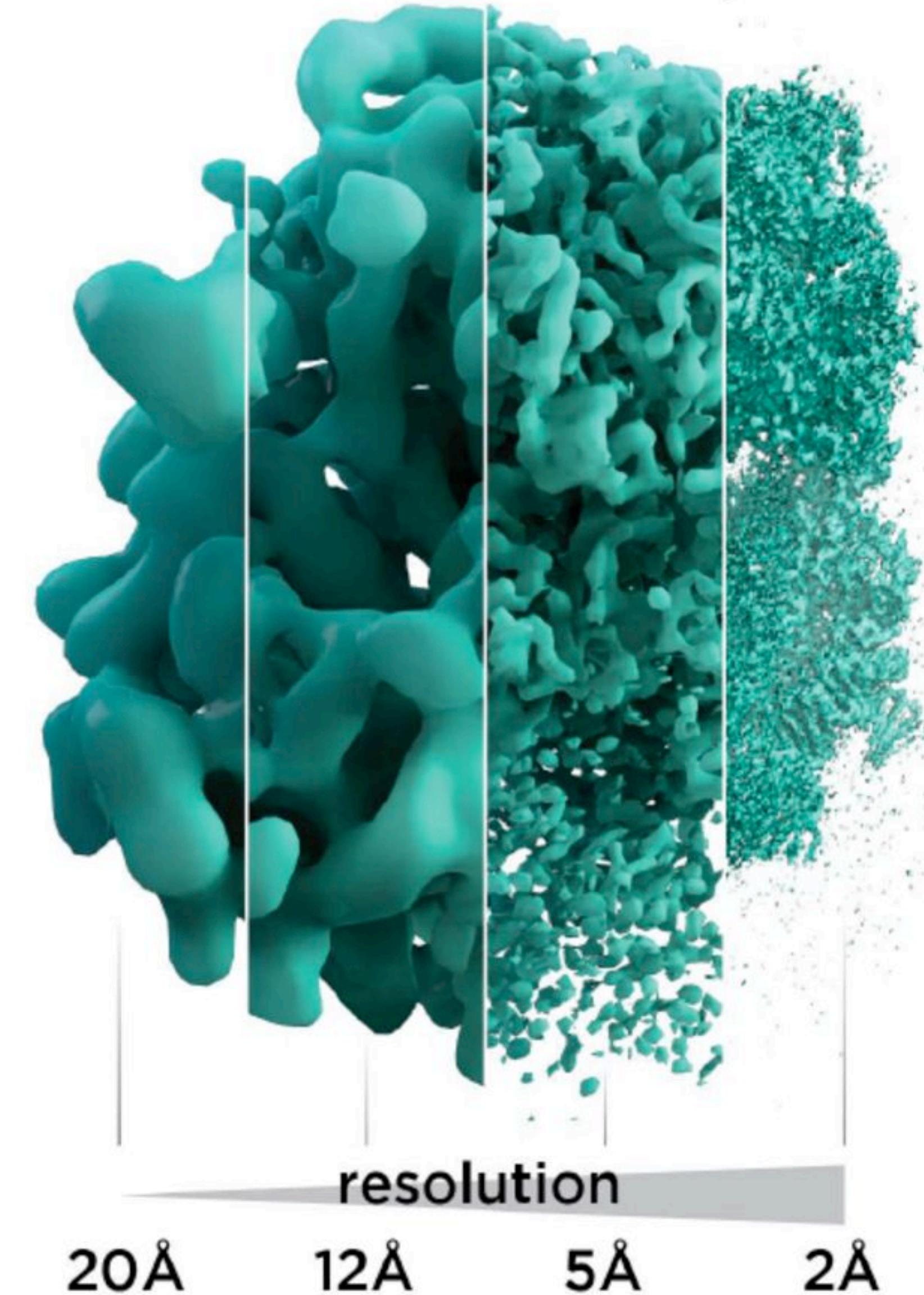
Why determine structure?

- Structure-function relationships - knowledge of a structure provides insight into how it works
- Identify critical residues for
 - binding to ligands and proteins
 - enzyme catalysis
 - connecting domains
- Modeling
 - Structure-based drug design
 - Starting point for molecular simulation
 - Structure prediction



Resolution

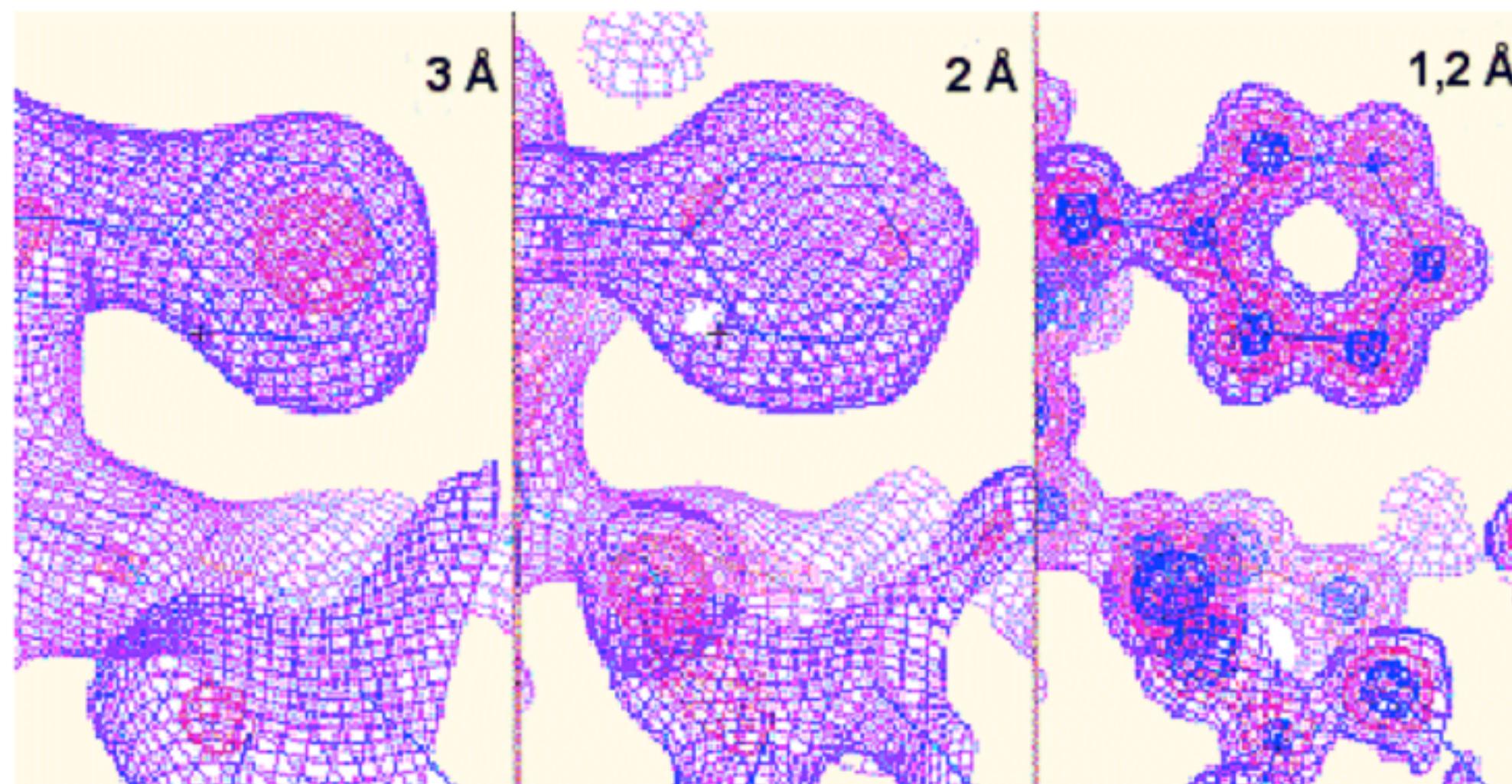
- Resolution = the size of the smallest observable detail that is discernible in a structural model
- High resolution = small #; low resolution = large #
- Summarizes the precision of atomic positions



<https://www.biorxiv.org/content/10.1101/2022.04.07.487522v1.full>

Needed Resolutions are Application-Dependent

- For drug design we generally need near-atomic ($\sim 1.8\text{\AA}$ resolution) so we can tell how the ligand is going to fit into the structure
- Alternate conformations or solvent dynamics might need better resolution
- For molecular dynamics, we need high resolution if we're trying to do QM in parts of the structure or if we want to begin the simulation with the atoms precisely located; but if we only want to model big changes, we can make do with 3 \AA or even 5 \AA resolution



The Protein Data Bank

The screenshot shows the main landing page of the RCSB PDB website. At the top, there is a dark blue header bar with the RCSB PDB logo and several navigation links: Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. To the right of these is a "MyPDB" dropdown menu. Below the header is a banner featuring the RCSB PDB logo and the text "194550 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education". To the right of the banner is a search bar with a magnifying glass icon and a "PDB Archive" dropdown. Below the search bar are links for "Advanced Search" and "Browse Annotations", along with a "Help" link. The main content area has a large background image of a protein structure. On the left, there is a sidebar with icons for Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area features a section titled "A Structural View of Biology" with text about the archive's purpose and the RCSB PDB's role as a member of the wwPDB. It also includes sections for "COVID-19 CORONAVIRUS Resources" and "Join the RCSB PDB Team". To the right, there is a "August Molecule of the Month" section showing a 3D model of secretory antibodies and a "Secretory Antibodies" section. At the bottom, there are three tabs: Latest Entries, Features & Highlights, News, and Publications.

<https://www.rcsb.org/>
8/27/2022

PDB Data Distribution by Experimental Method and Molecular Type

[Copy](#) [CSV](#)

Molecular Type	↑	X-ray↓	EM↑	NMR↓	Multiple methods↑	Neutron↓	Other↑	Total ↓
Protein (only)		166,100	14,873	12,470	208	77	32	193,760
Protein/Oligosaccharide		9,600	2,557	34	8	2	0	12,201
Protein/NA		8,668	4,531	286	7	0	0	13,492
Nucleic acid (only)		2,860	136	1,502	14	3	1	4,516
Other		167	10	33	0	0	0	210
Oligosaccharide (only)		11	0	6	1	0	4	22
Total		187,406	22,107	14,331	238	82	37	224,201

PDB Structures with Experimental Data

[177,482](#) have a structure factor file

[11,506](#) have an NMR restraint file

[6,038](#) have a chemical shifts file

[4,135](#) have NMR unified data files (NEF and/or NMR-STAR format files)

21,612 3DEM maps were used for the determination of [21,837](#) PDB structures

PDB Structures with Small Molecule Data

[170,164](#) with any non-polymer small molecules

[168,613](#) with non-polymer small molecules of 100-300 Da

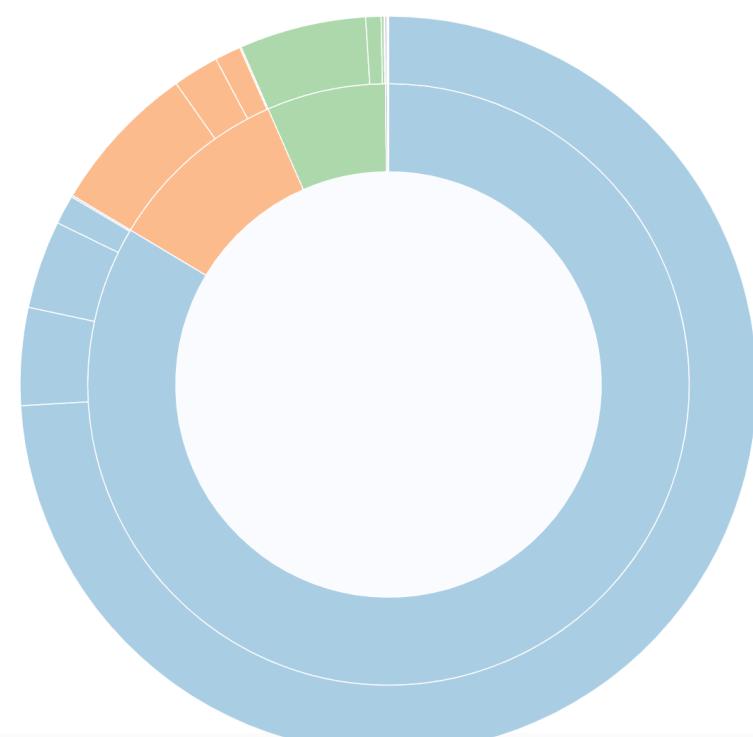
[58,582](#) with non-polymer small molecules of 300-500 Da

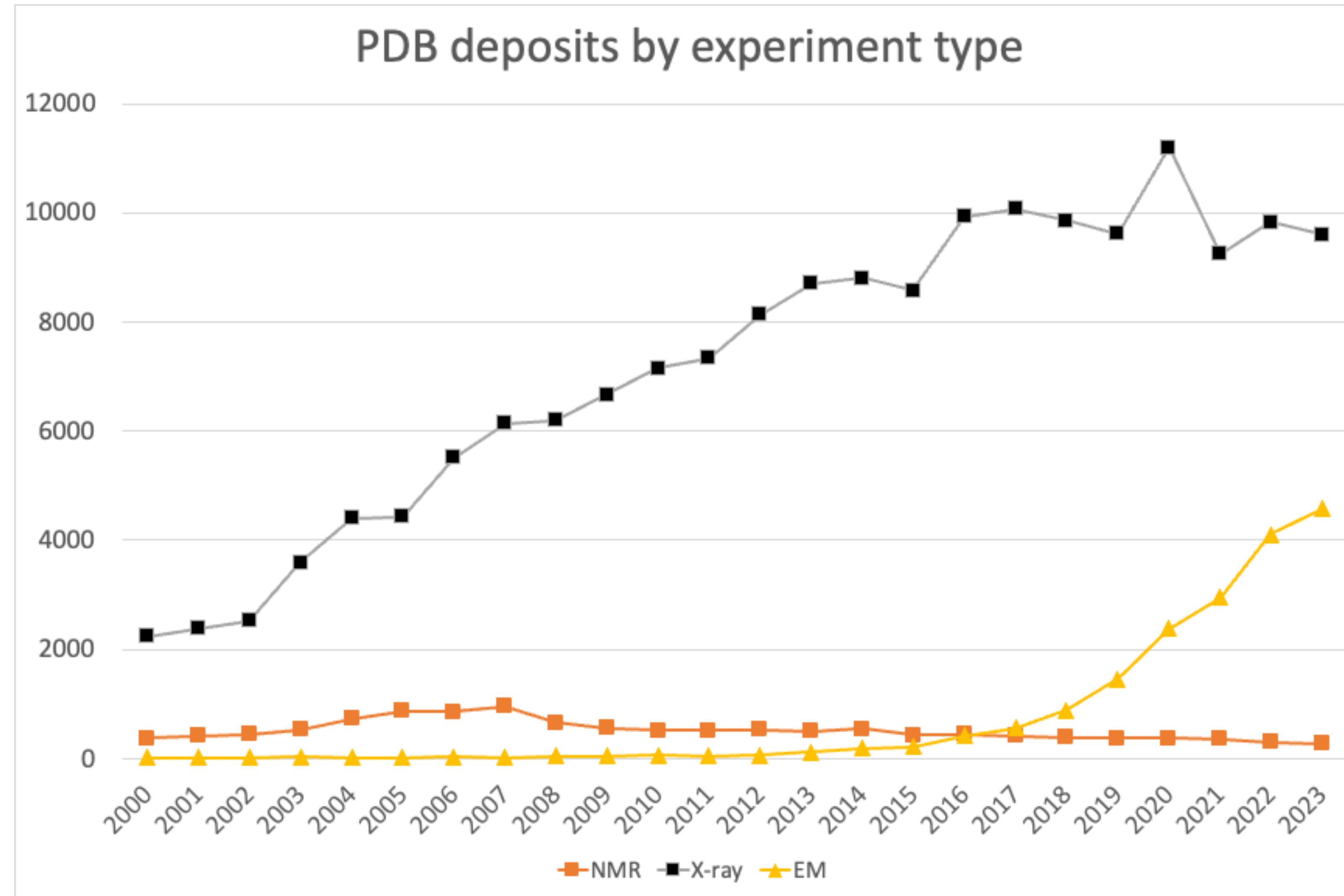
[40,153](#) with non-polymer small molecules of 500-1200 Da

[36,891](#) with author-designated Ligand/s of Interest

[4,503](#) with a biologically interesting short oligomer or oligomer-like BIRD molecules

[12,378](#) with branched entities of oligosaccharides





Data from

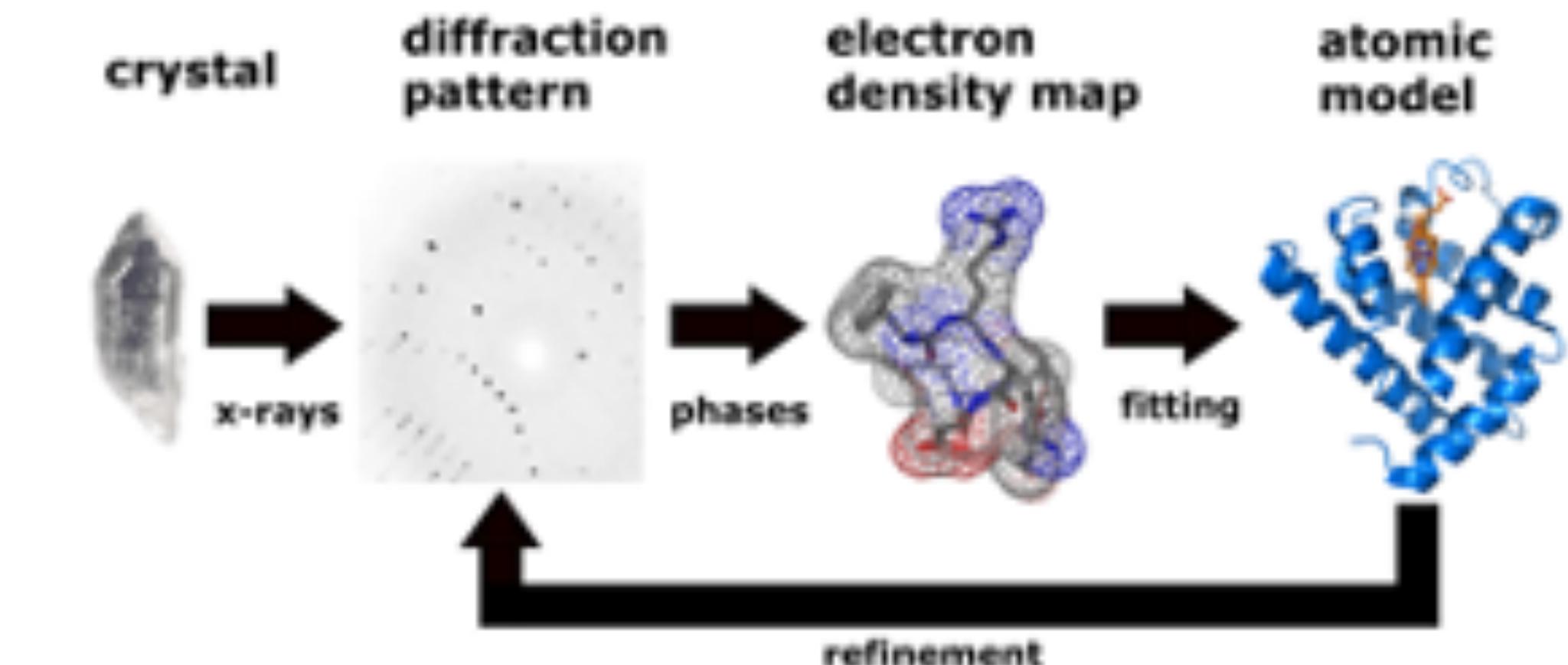
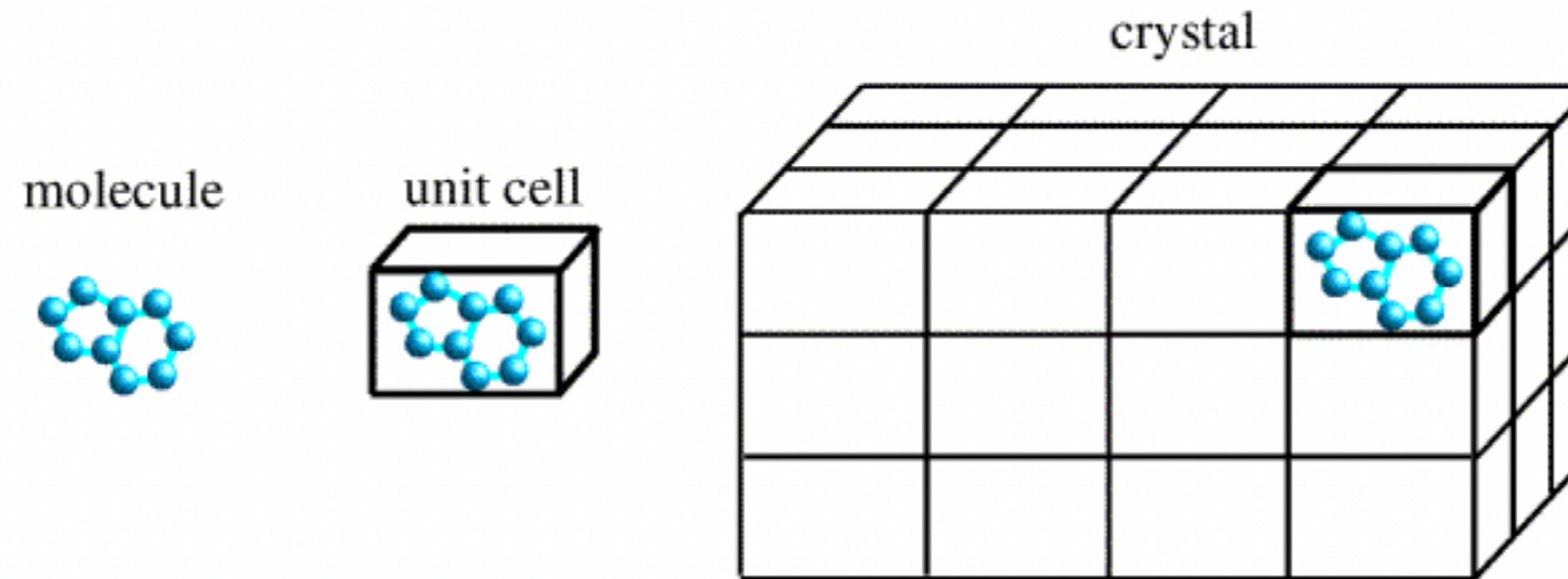
<https://www.rcsb.org/>

on 8/27/2024

Experimental methods for structure determination

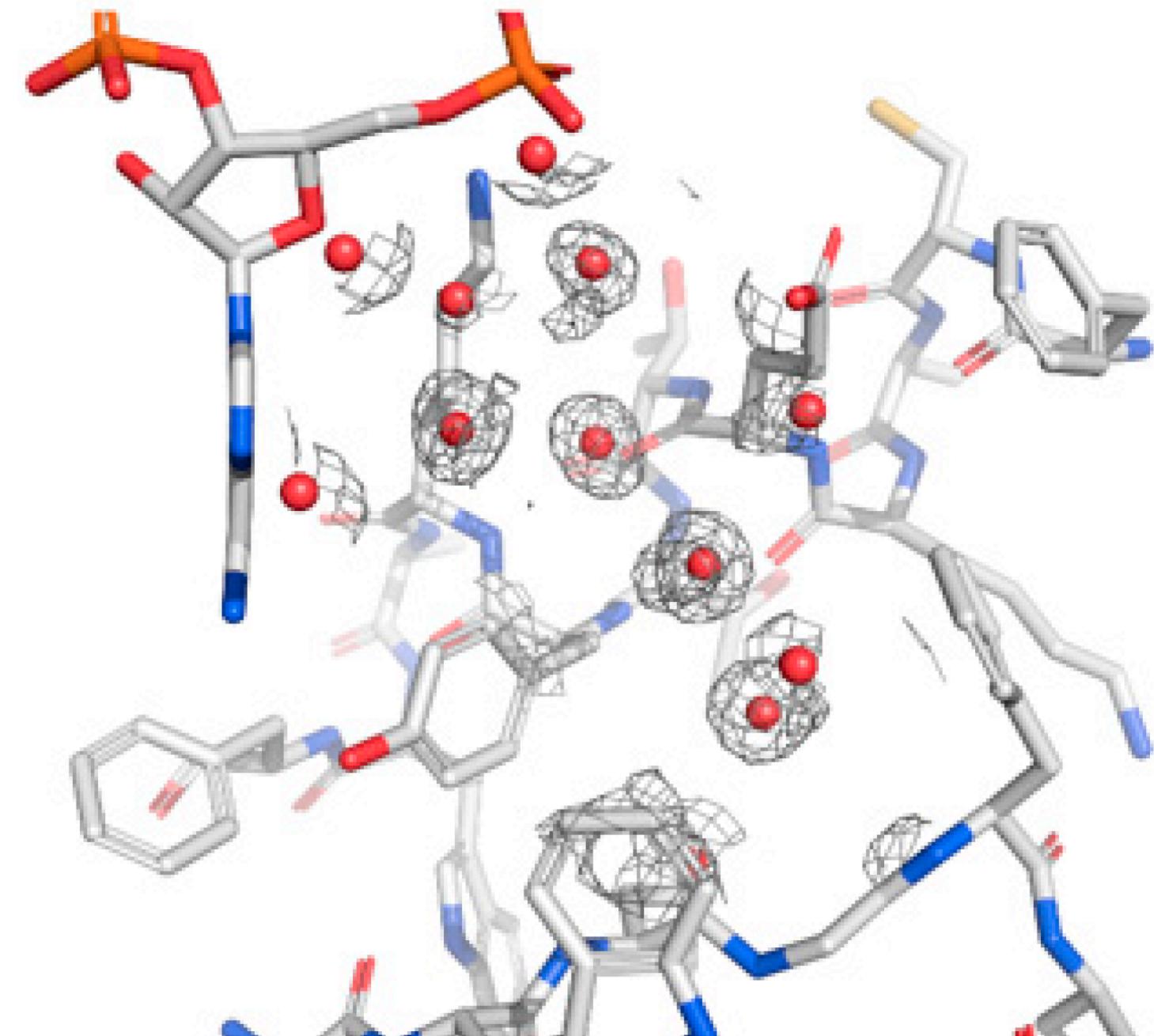
X-ray Crystallography

- Crystals
 - Translationally ordered arrays of molecules
 - Most solids are crystalline
 - Distances between atoms and orientations are similar for each unit cell
- X-rays are deflected by electrons
- A crystal causes interference between X-rays, leading to spots (Bragg reflections)
- Atomic model inferred from diffraction patterns at different angles

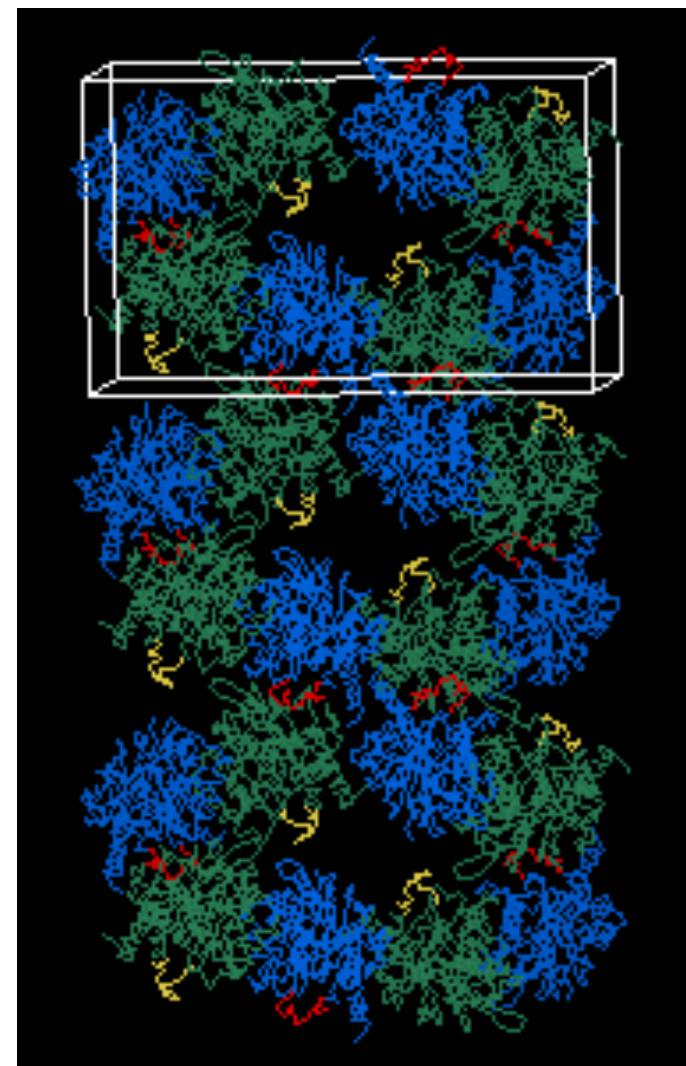


Benefits and Drawbacks

- Benefits
 - High resolution possible
 - High-throughput facilities, e.g. Advanced Light Source at Argonne National Laboratory
 - Ordered waters can be observed
- Drawbacks
 - Macromolecule(s) need to be crystallized
 - Can require specific conditions, e.g. >95% purity
 - May require special construct
 - Crystal contacts may impact structure, but
 - crystalline proteins still have biological activity
 - contacts have less influence on the interior than surface
 - Usually, but not always, collected at low temperature
 - Hydrogen, and thus protonation, usually unavailable



<https://www.mdpi.com/1420-3049/25/5/1030/htm#>



https://www.umass.edu/microbio/chime/pe_beta/pe/proteexpl/xtlcon.htm

Nuclear Magnetic Resonance

- uses a large magnet to probe spin properties of atomic nuclei
- like other spectroscopies, uses electromagnetic radiation (radio waves) to promote transitions between energy levels
- resonant frequency proportional to strength of magnetic field



https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance#/media/File:700_lab_fix.JPG

Nuclear Spin

- Like electrons, nuclei have spin
- Quantum numbers
 - I for spin
 - m for spin in a magnetic field
- If the number of protons and neutrons
 - is even, zero spin
 - is odd, non-zero spin
- If non-zero spin,
 - nucleus has magnetic moment $\mu = \gamma I$, where γ is a gyromagnetic ratio, that depends on the nucleus

Nuclei	Spin	Gyromagnetic Ratio (MHz/T)	Natural Abundance (%)
¹ H	1/2	42.576	99.9985
¹³ C	1/2	10.705	1.07
³¹ P	1/2	17.235	100
²⁷ Al	5/2	11.103	100
²³ Na	3/2	11.262	100
⁷ Li	3/2	16.546	92.41
²⁹ Si	1/2	-8.465	4.68
¹⁷ O	5/2	5.772	0.038
¹⁵ N	1/2	-4.361	0.368

The gyromagnetic ratios for several common nuclei

[https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_\(Physical_and_Theoretical_Chemistry\)/Spectroscopy/Magnetic_Resonance_Spectroscopies/Nuclear_Magnetic_Resonance/Nuclear_Magnetic_Resonance_II](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Magnetic_Resonance_Spectroscopies/Nuclear_Magnetic_Resonance/Nuclear_Magnetic_Resonance_II)

Nuclear spin in a magnet

- Nuclei behave as a tiny bar magnet
 - without magnetic field, randomly oriented
 - with strong magnetic field, aligns
 - with field or (low energy)
 - against field (high energy)

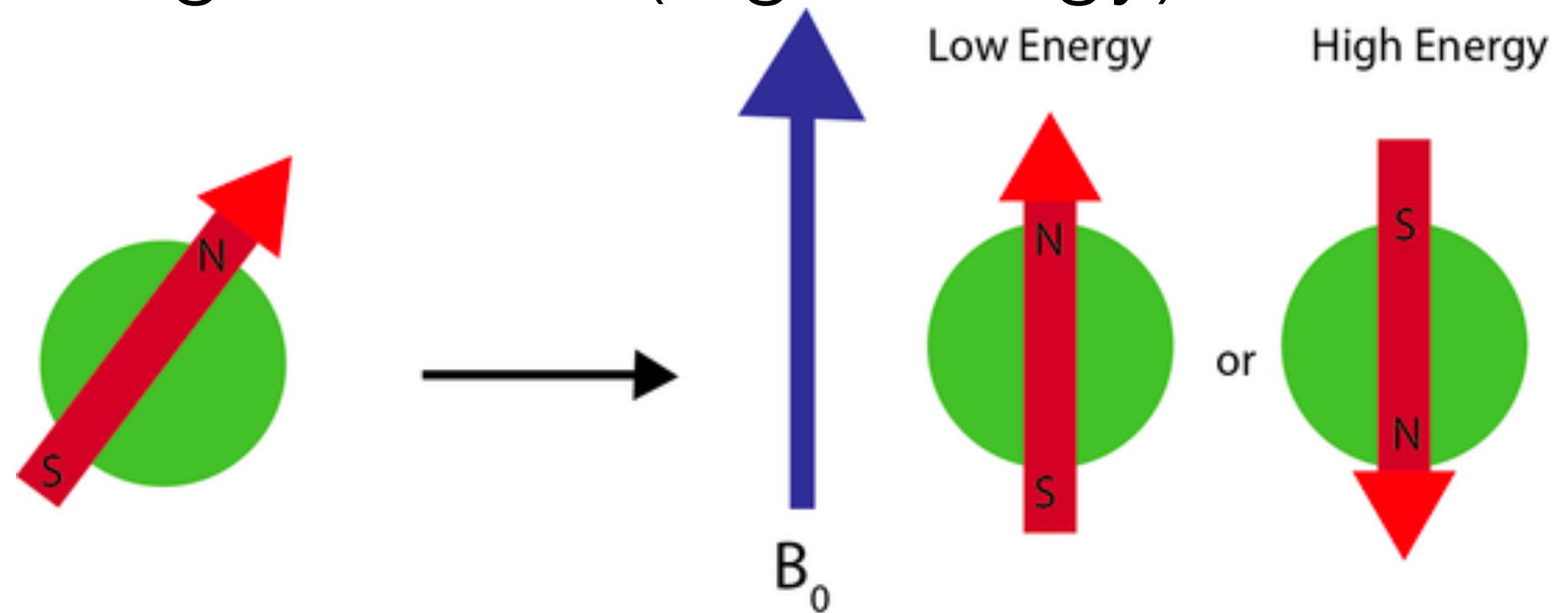


Figure NMR.1 from [https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_\(Physical_and_Theoretical_Chemistry\)/Spectroscopy/Magnetic_Resonance_Spectroscopies/Nuclear_Magnetic_Resonance/Nuclear_Magnetic_Resonance_II](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Magnetic_Resonance_Spectroscopies/Nuclear_Magnetic_Resonance/Nuclear_Magnetic_Resonance_II)

- Nuclear energy level is $E = - m\hbar\gamma B_0$
 - m is magnetic quantum number, e.g. $\pm \frac{1}{2}$
 - \hbar is Plank's constant over 2π
 - γ is magnetic moment

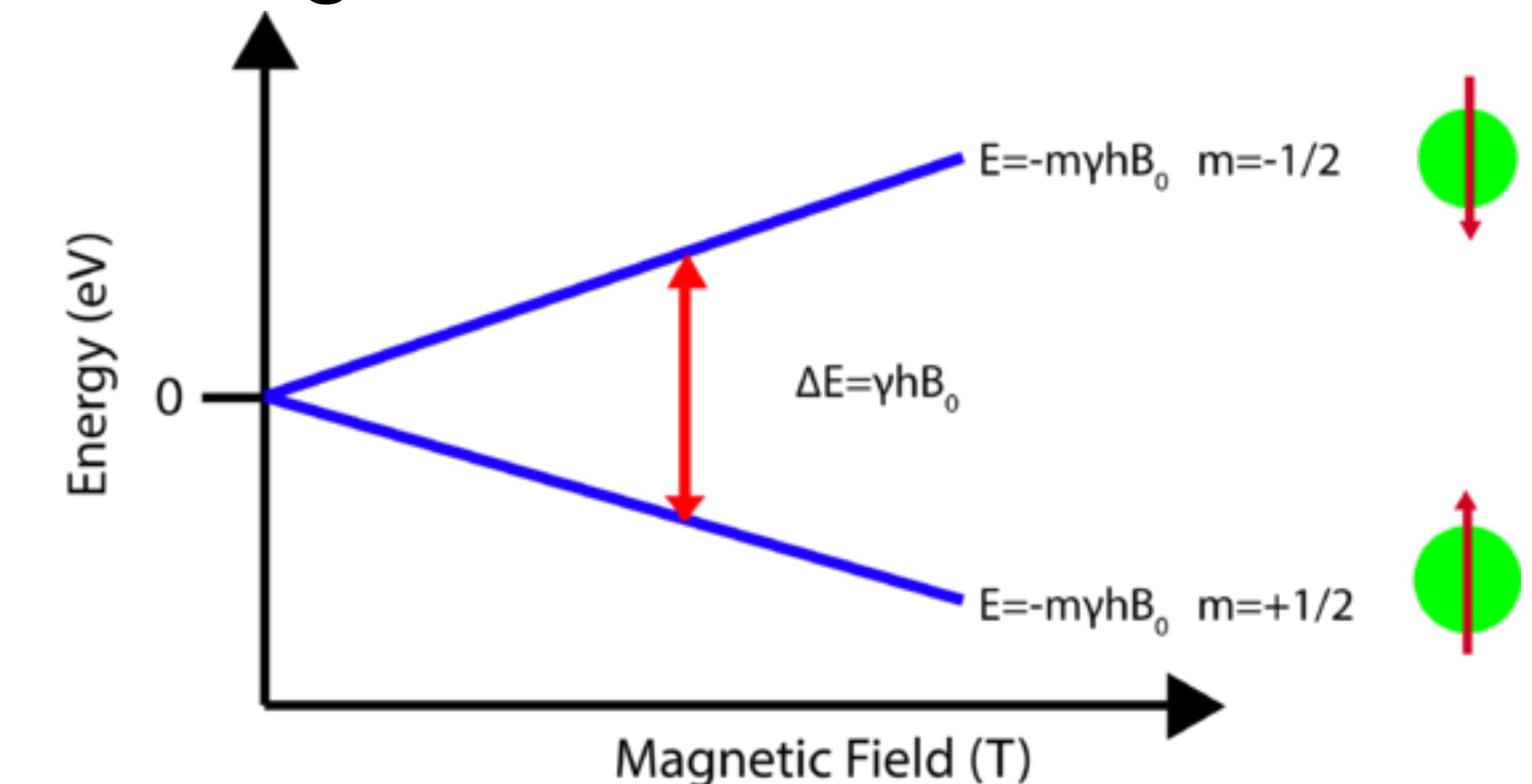


Figure NMR.2 The splitting of the degenerate nuclear energy levels under an applied magnetic field.

Spin flip

For a photon to be absorbed, its energy must match the energy gap, $\nu = \frac{\gamma B_o}{2\pi}$

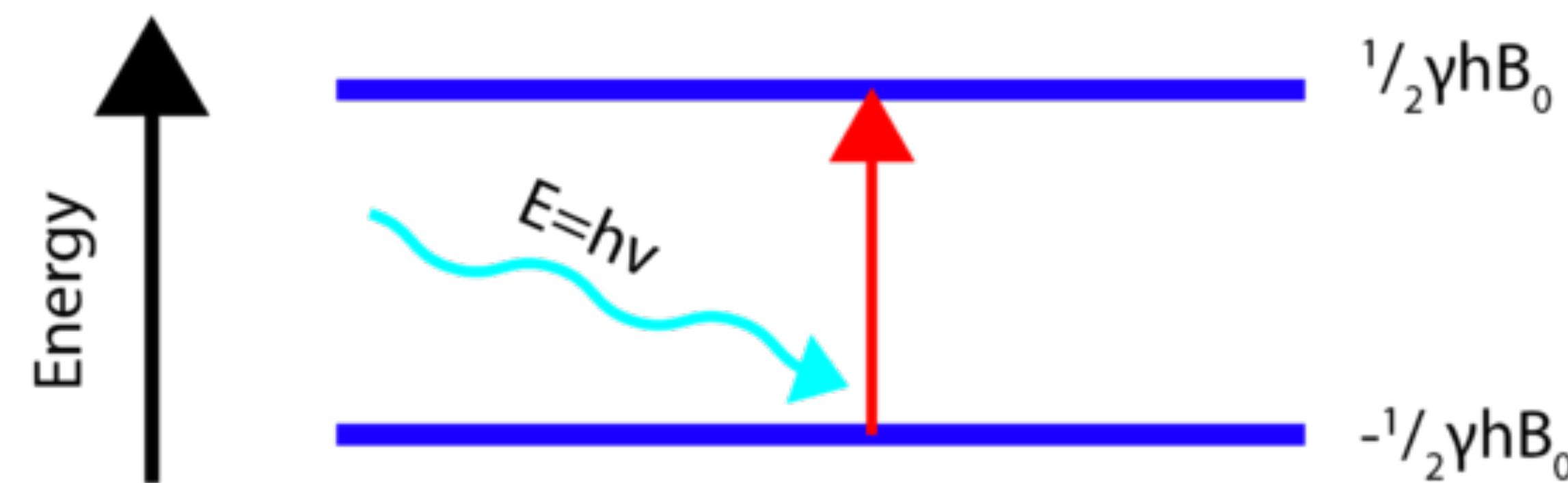
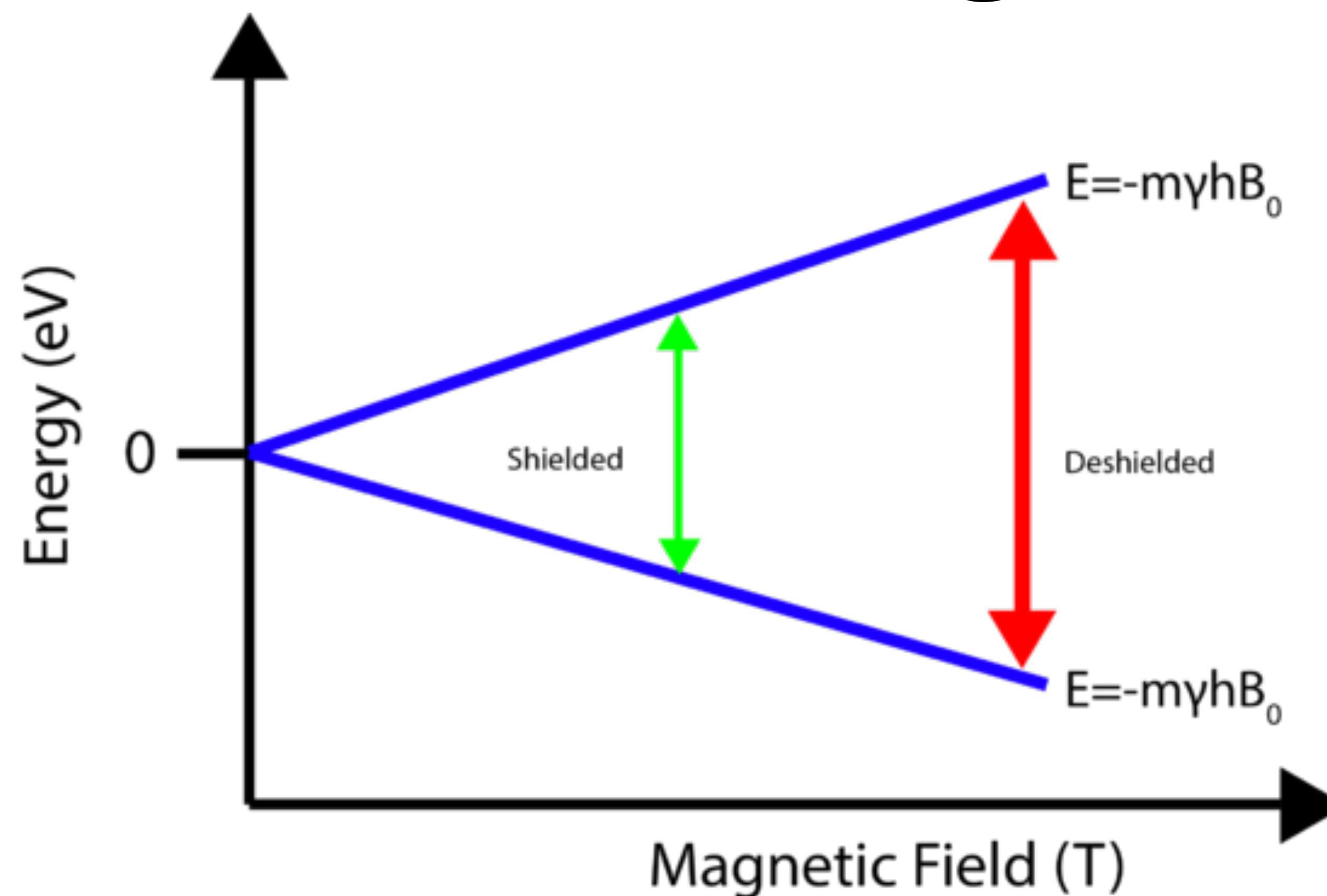


Figure NMR.3: Absorption of radio frequency radiation to promote a transition between nuclear energy levels, called a spin flip.

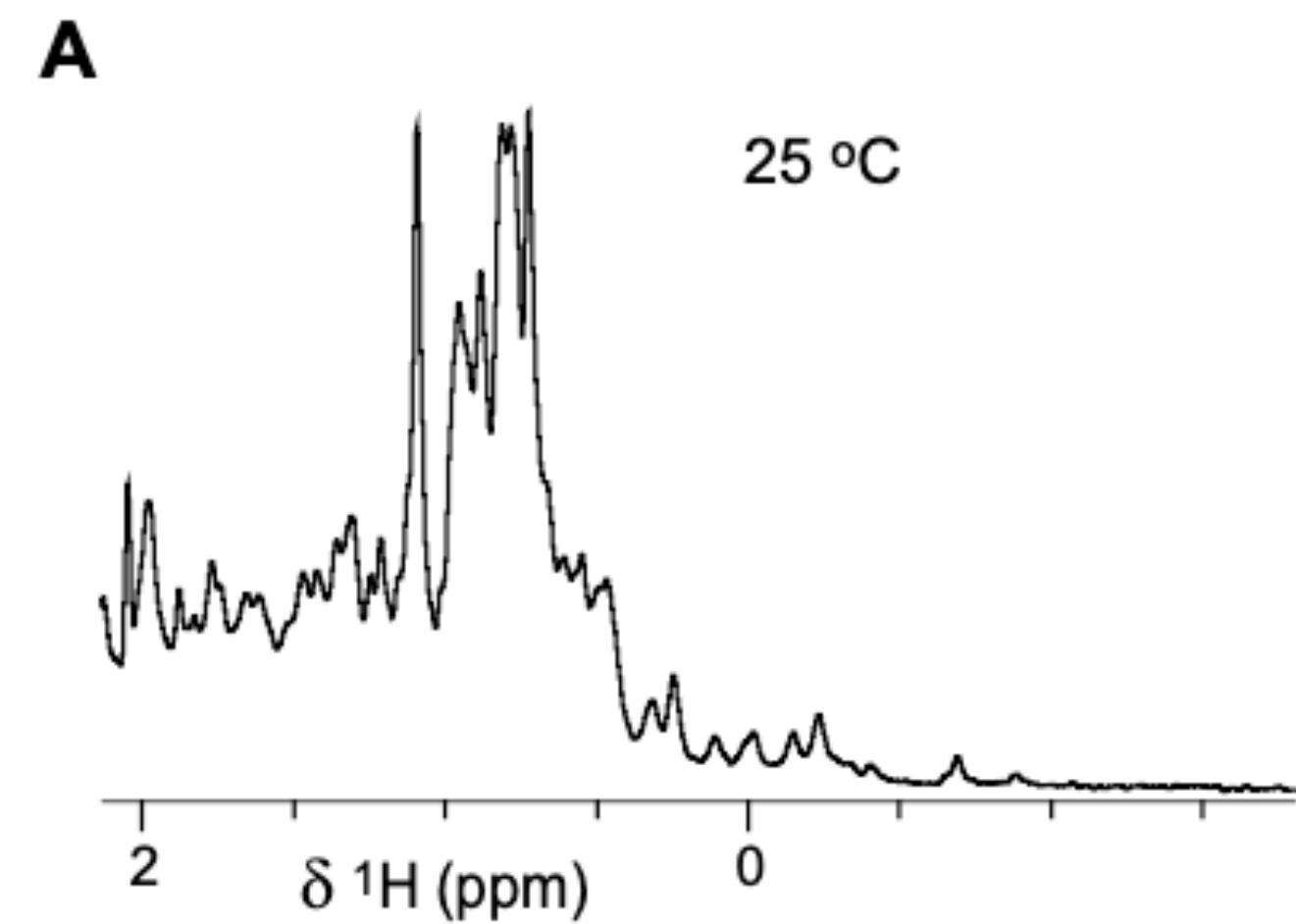
Shielding



The effect that shielding from electrons has on the splitting of the nuclear energy levels. Electrons impart their own magnetic field which shields the nucleus from the externally applied magnetic field. This effect is greatly exaggerated in this illustration.

Protein NMR

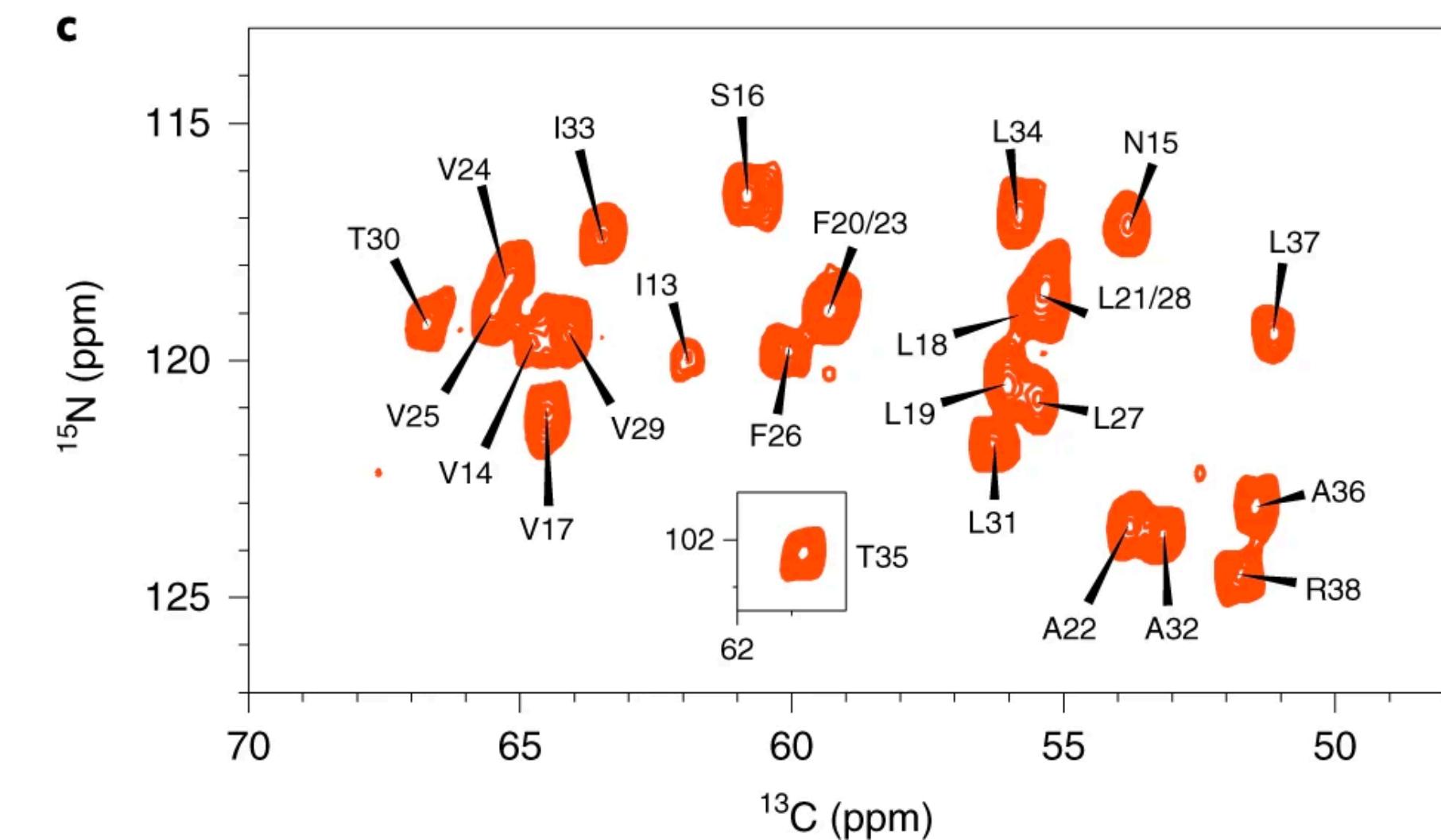
- Peak overlap is an even larger issue with proteins than small molecules!



Methyl region from ^1H NMR spectra of SARS-CoV-2 MPro.

Fig 1a of Kantsadi, A. L.; Cattermole, E.; Matsoukas, M.-T.; Spyroulias, G. A.; Vakonakis, I. A COVID Moonshot: Assessment of Ligand Binding to the SARS-CoV-2 Main Protease by Saturation Transfer Difference NMR Spectroscopy; preprint; Biochemistry, 2020. <https://doi.org/10.1101/2020.06.17.156679>.

- NMR is performed in *multiple dimensions*

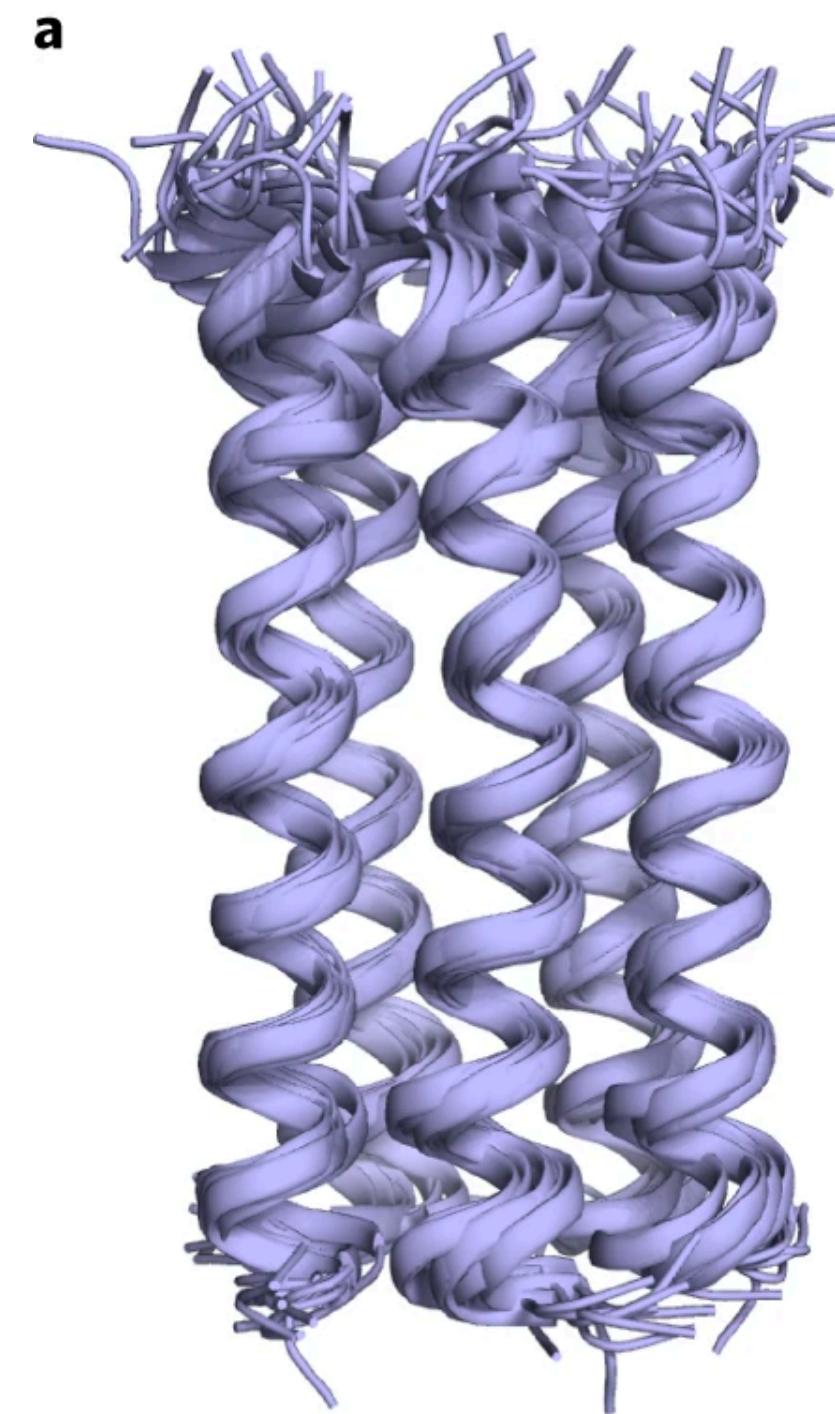


2D correlation spectrum of SARS-CoV-2 envelope transmembrane

Fig 1c of Mandala, V. S.; McKay, M. J.; Shcherbakov, A. A.; Dregni, A. J.; Kolocouris, A.; Hong, M. Structure and Drug Binding of the SARS-CoV-2 Envelope Protein Transmembrane Domain in Lipid Bilayers. *Nat Struct Mol Biol* 2020, 27(12), 1202–1208. <https://doi.org/10.1038/s41594-020-00536-8>.

Protein NMR structures

- Sample is irradiated in *pulse sequences*, which can probe magnetization transferred through
 - bonds - for assigning chemical shifts to nuclei
 - space - for distance restraints
- Distance restraints are used to build protein models



Ten lowest-energy structures of the SARS-CoV-2 envelope protein's transmembrane domain

Fig 3a of Mandala, V. S.; McKay, M. J.; Shcherbakov, A. A.; Dregni, A. J.; Kolocouris, A.; Hong, M. Structure and Drug Binding of the SARS-CoV-2 Envelope Protein Transmembrane Domain in Lipid Bilayers. *Nat Struct Mol Biol* **2020**, 27 (12), 1202–1208. <https://doi.org/10.1038/s41594-020-00536-8>.⁽¹⁾

Benefits and Drawbacks of NMR

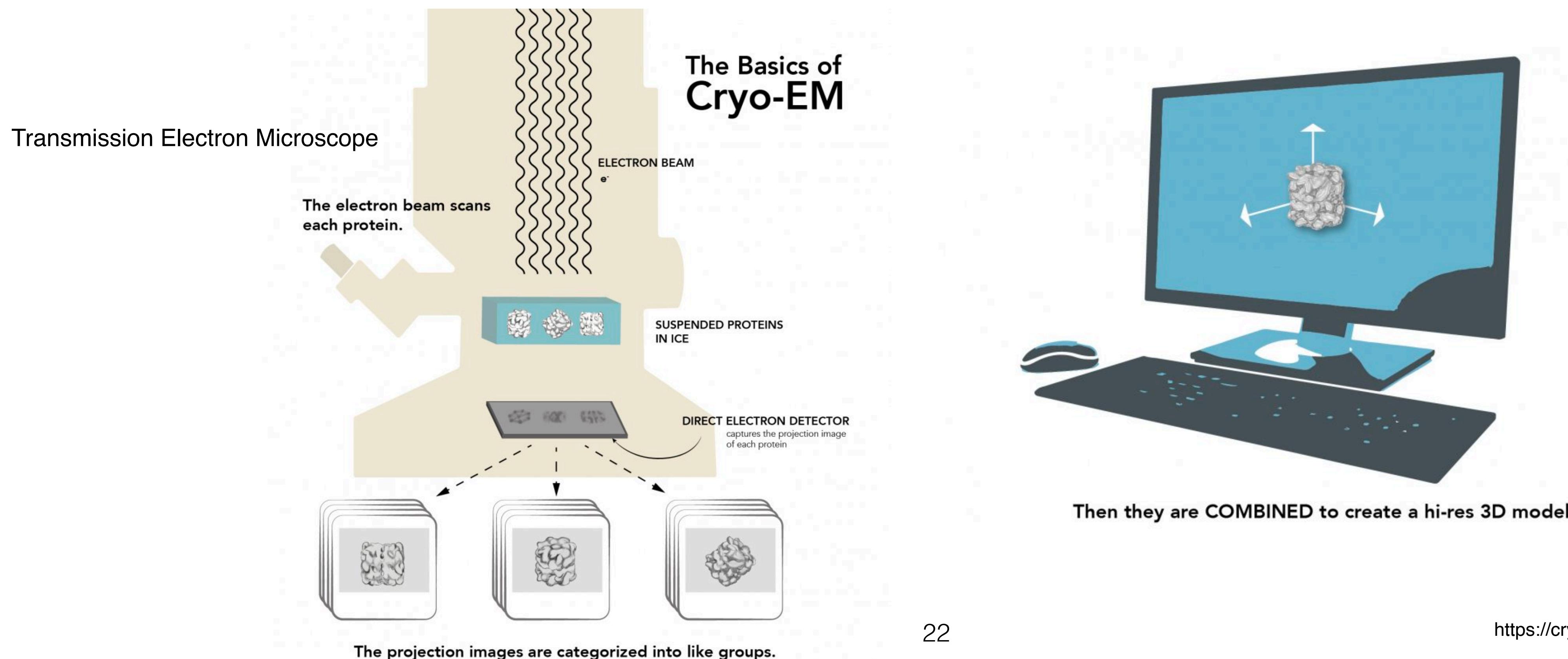
- Benefits
 - conditions mostly resemble natural surroundings of biological molecules
 - dynamics in various timescales spanning from picoseconds to days
 - hydrogen observable
- Drawbacks
 - peak overlap makes determination of larger structures untenable, but recent advances mean proteins (complexes) beyond 100 kDa are within our reach
 - instrumentation and maintenance expensive and no government-funded facilities
- Both: NMR provides ensembles, which may reflect heterogeneity or simply lack of knowledge
- Permi, P. Bioanalytical NMR Spectroscopy. In *Handbook of Spectroscopy*; John Wiley & Sons, Ltd, 2014; pp 1079–1114. <https://doi.org/10.1002/9783527654703.ch28>

Review Questions

- What is required for a nucleus to have an NMR signal?
- How is peak overlap addressed in protein NMR?

Cryogenic Electron Microscopy (Cryo-EM)

- illuminates samples with beams of electrons
- resolution limited by electron wavelength



Cryo-EM history

- EM long used for biological imaging at organelle scale (>10 nm)
- Before 1985, ~0.5 nm EM failed due to sample damage
- By 1990, ~0.4 nm was enabled by
 - cooling (< 100 K) to slow sample damage
 - use of software to merge images
- Now, high resolution (2 Å ideal, 3 Å in many cases, 4 Å in most) enabled by
 - more sensitive electron detectors
 - better sample preparation
 - better software
- 2017 Nobel Prize in Chemistry

The Nobel Prize in Chemistry 2017

Jacques Dubochet
Joachim Frank
Richard Henderson

Share this



The Nobel Prize in Chemistry 2017



© Nobel Media AB. Photo: A.
Mahmoud
Jacques Dubochet

Prize share: 1/3



© Nobel Media AB. Photo: A.
Mahmoud
Joachim Frank

Prize share: 1/3



© Nobel Media AB. Photo: A.
Mahmoud
Richard Henderson

Prize share: 1/3

The Nobel Prize in Chemistry 2017 was awarded jointly to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"

Benefits and Drawbacks

- Benefits
 - Method of choice for large multicomponent complexes
 - Lower purity requirements; starting substance must be ~85% pure
- Drawbacks
 - Frozen systems \neq physiological
 - Usually lower resolution
 - No hydrogen observed

Review Question

- 1. Cryo-EM**
 - 2. X-Ray Crystallography**
 - 3. NMR**
-
- 1. Which method uses electrons to form images of biological macromolecules at atomic resolution?**
 - 2. Which method analyzes non-zero nuclear spin interactions within a magnetic field?**
 - 3. Which method analyzes the angles and intensities of diffracted X-ray beams?**

Review Discussion Questions

What are some of the benefits (positive aspects) and drawbacks (negative aspects) of X-Ray Crystallography, NMR, and Cryo-EM?

Which structure determination method would be best for a large membrane-associated complex?

If you have a small, pure protein and you are interested its protonation states, which method for structural determination is most ideal?