

Computing and Chemistry

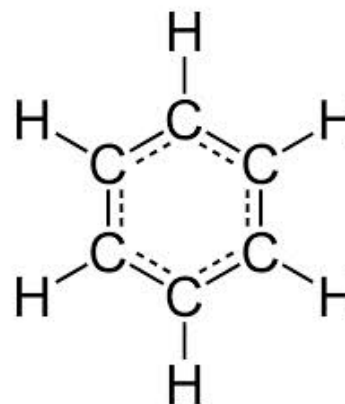
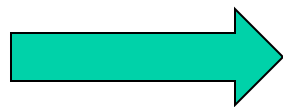
david.wishart@ualberta.ca

3-41 Athabasca Hall

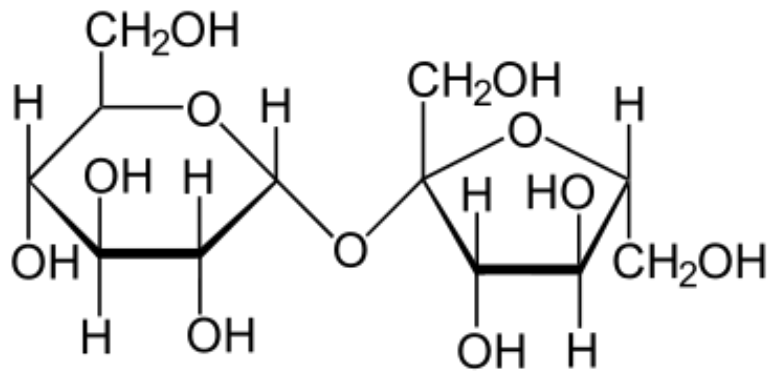
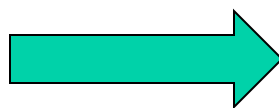
Sept. 16, 2013



How Do We Know?

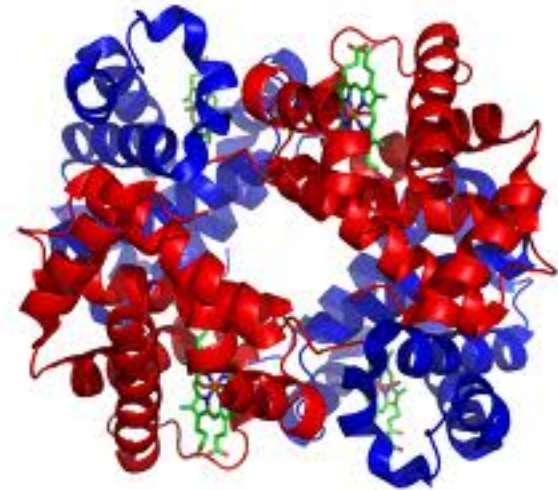


Benzene

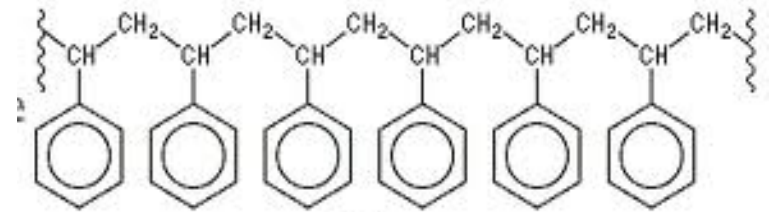


Sucrose

How Do We Know?



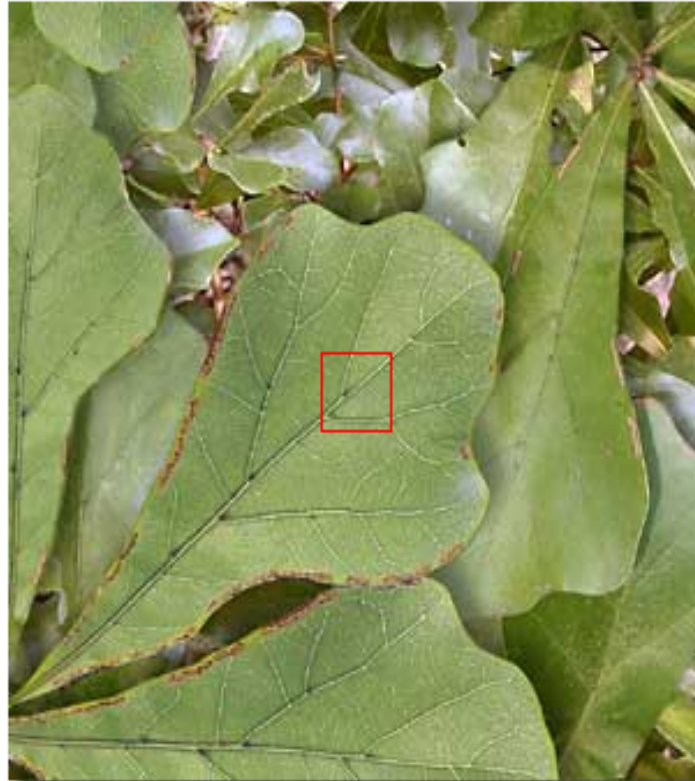
Hemoglobin



polystyrene

Powers of 10

Oak tree leaves at actual size.



10^{-1} meters

10 centimeters

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Surface of an Oak leaf magnified 10 times.



10^{-2} meters

1 centimeter

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Surface of an Oak leaf magnified 100 times.



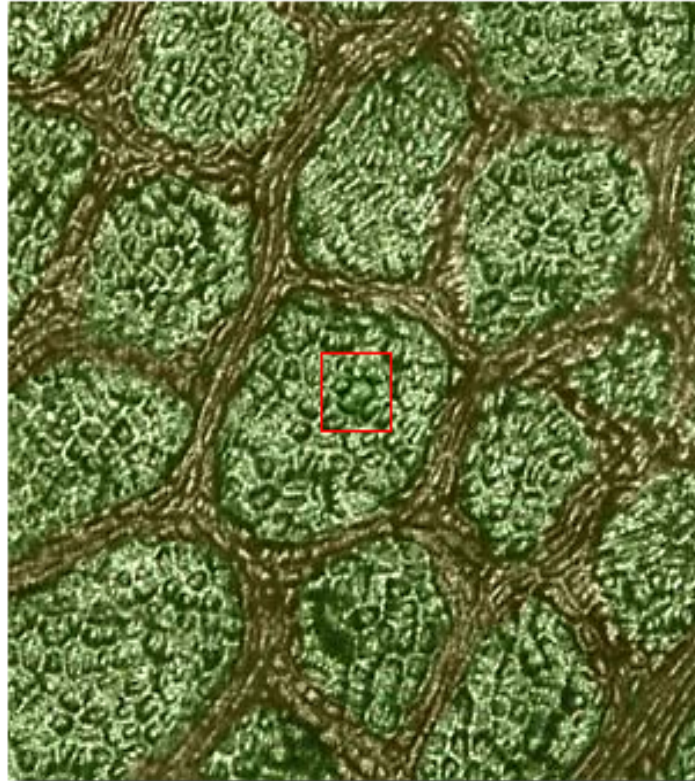
10^{-3} meters

1 millimeter

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Cells on the leaf surface.



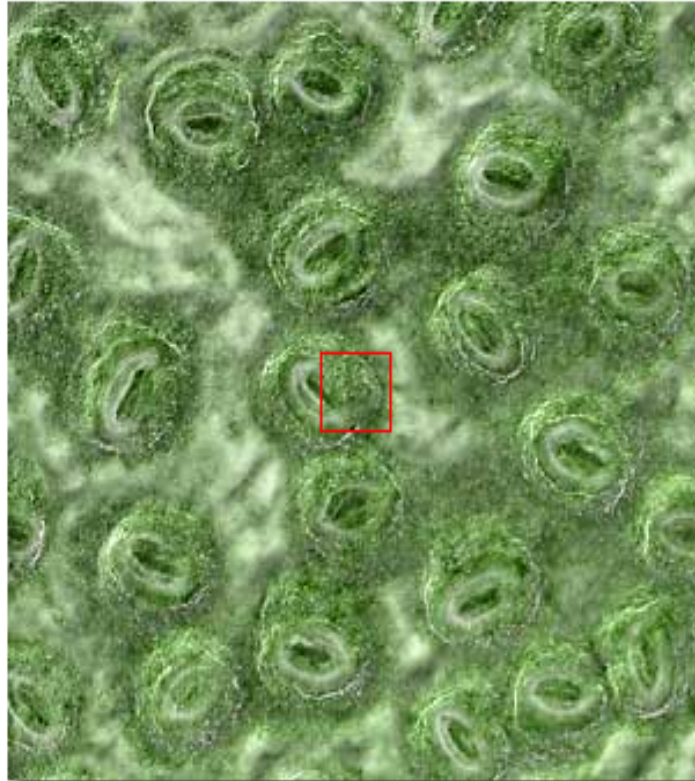
10^{-4} meters

100 micrometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Individual leaf cells.



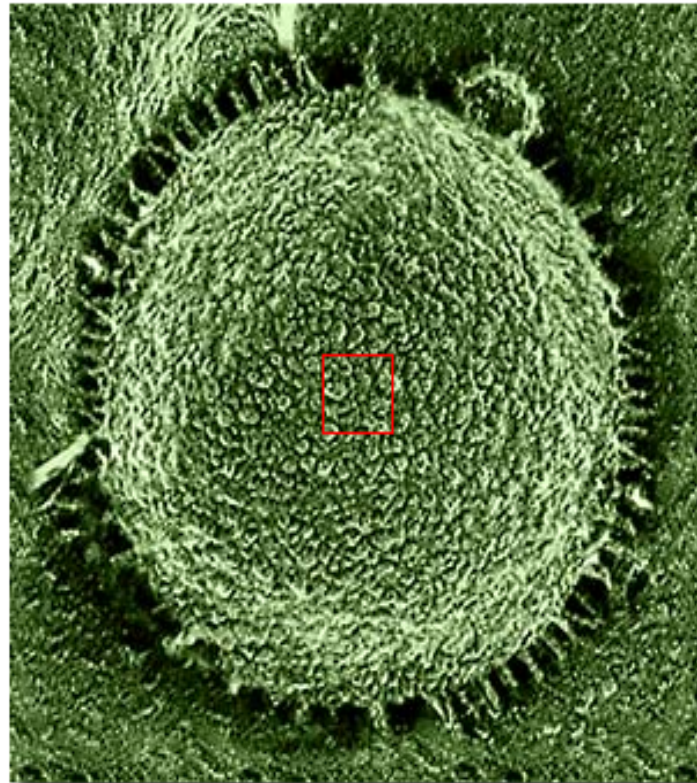
10^{-5} meters

10 micrometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

The nucleus of a leaf cell.



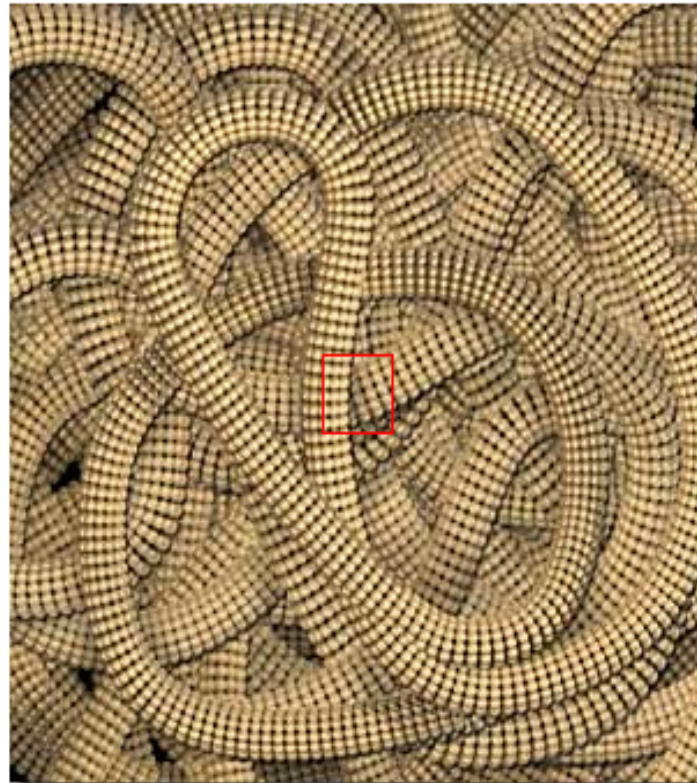
10^{-6} meters

1 micrometer

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Chromatin in the leaf cell nucleus.



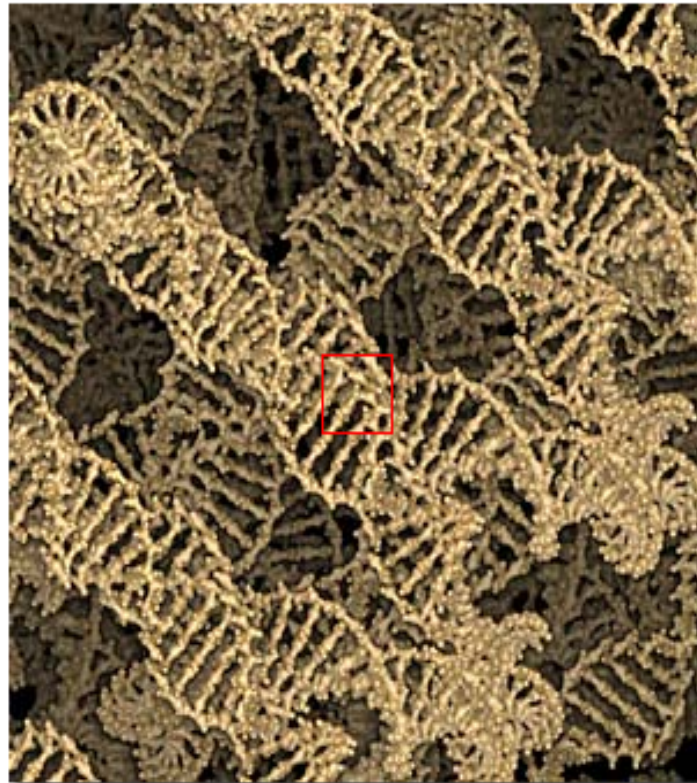
10^{-7} meters

100 nanometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Individual DNA strands.



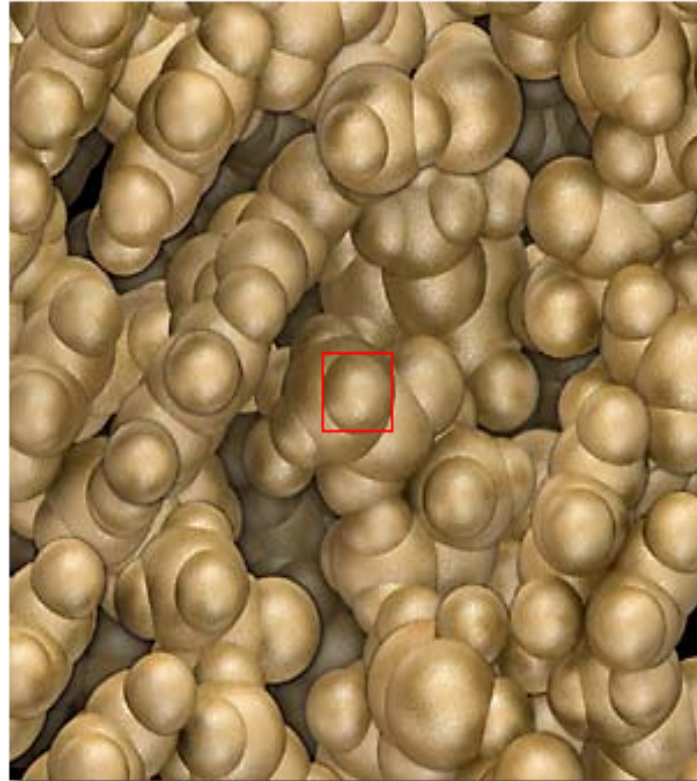
10^{-8} meters

10 nanometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

DNA nucleotide building blocks.



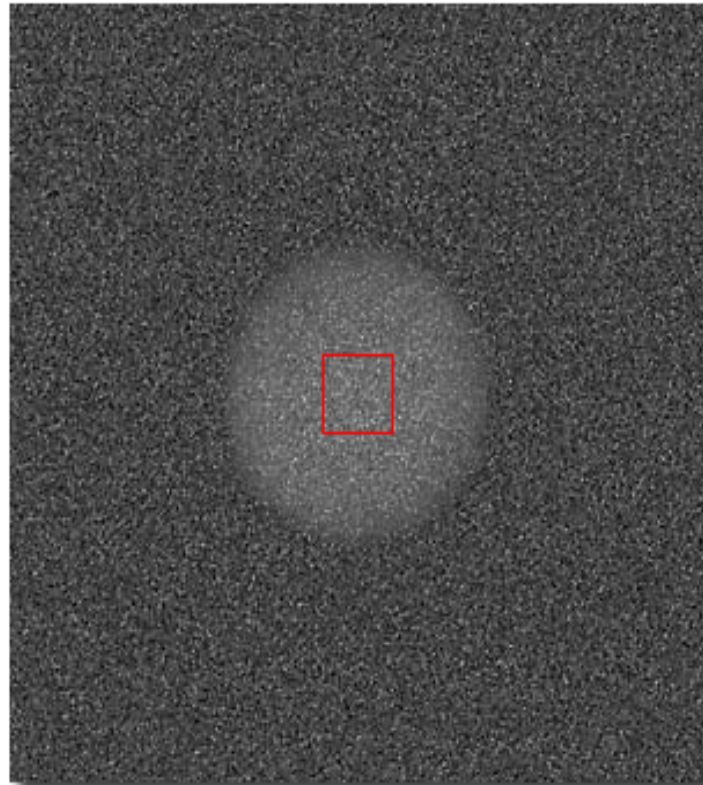
10^{-9} meters

1 nanometer

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Outer electron cloud of a carbon atom.



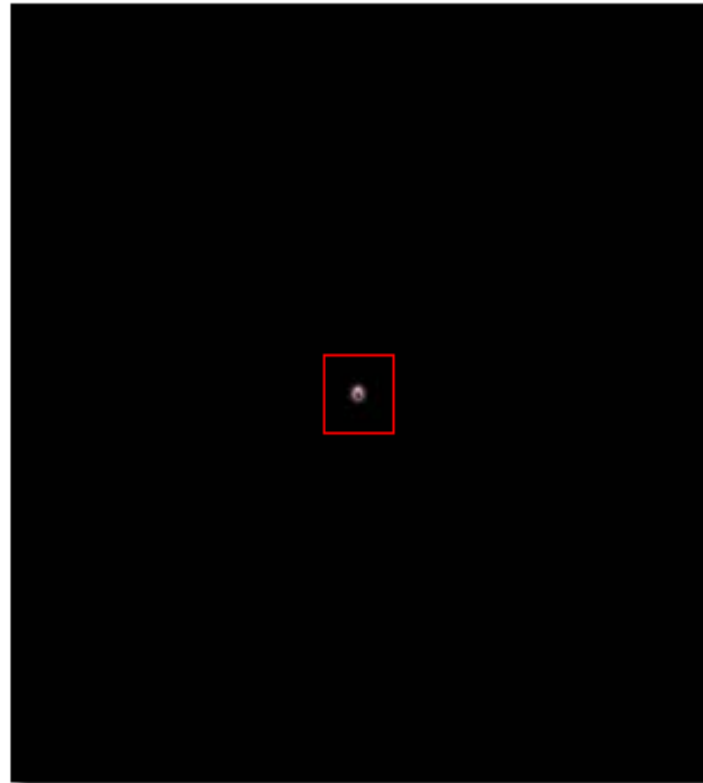
10^{-10} meters

100 picometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Empty space between inner shell and nucleus.



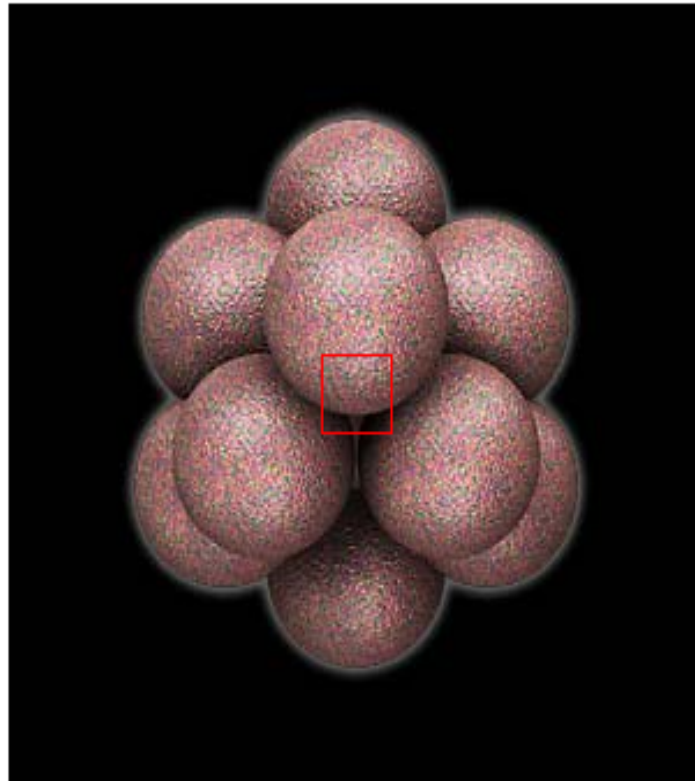
10^{-12} meters

1 picometer

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Powers of 10

Nucleus of the carbon atom.



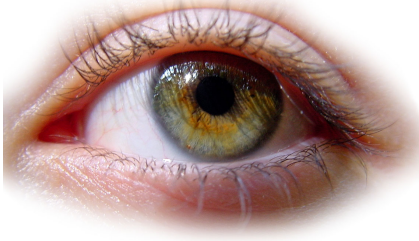
10^{-14} meters

10 femtometers

<http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/>

Our Seeing Limits (and Limitations)

Free



1 m

Live, moving

\$5



1×10^{-3} m

Live, moving

\$5000



1×10^{-6} m

Fixed, stained

\$500,000



1×10^{-9} m

Fixed, stained

Our Seeing Limits (and Limitations)

\$5,000,000



1×10^{-10} m

Extracted, crystallized

\$500,000,000



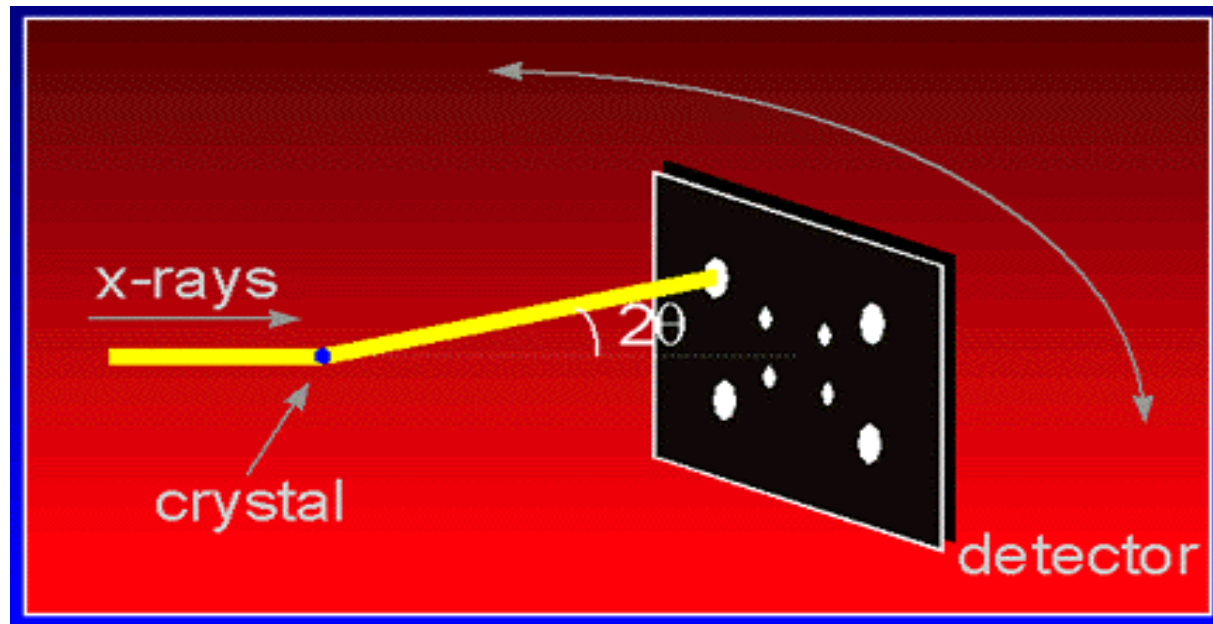
1×10^{-12} m

Atomized, vaporized

Seeing Molecules

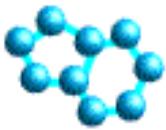
- **Can't use visible light**
- **Can't use electrons (EM)**
- **Have to use X-ray scattering**
- **Have to use Nuclear Magnetic Resonance (NMR) spectroscopy**
- **Have to use mass spectrometry**
- **All require computers & computing**

X-ray Crystallography

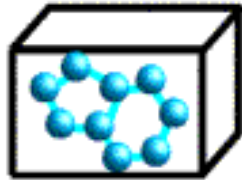


Crystallization

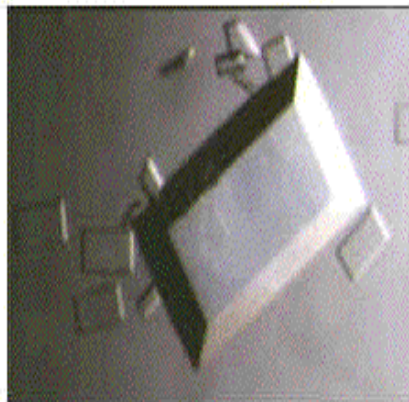
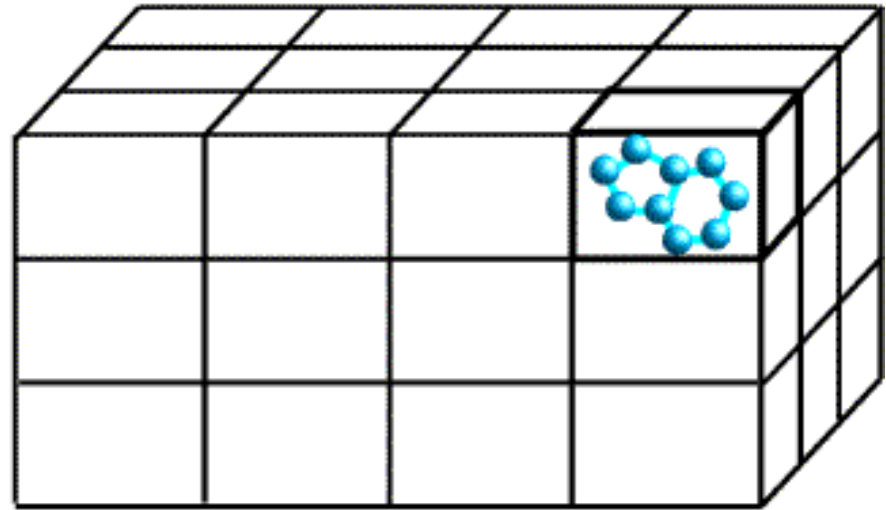
molecule



unit cell

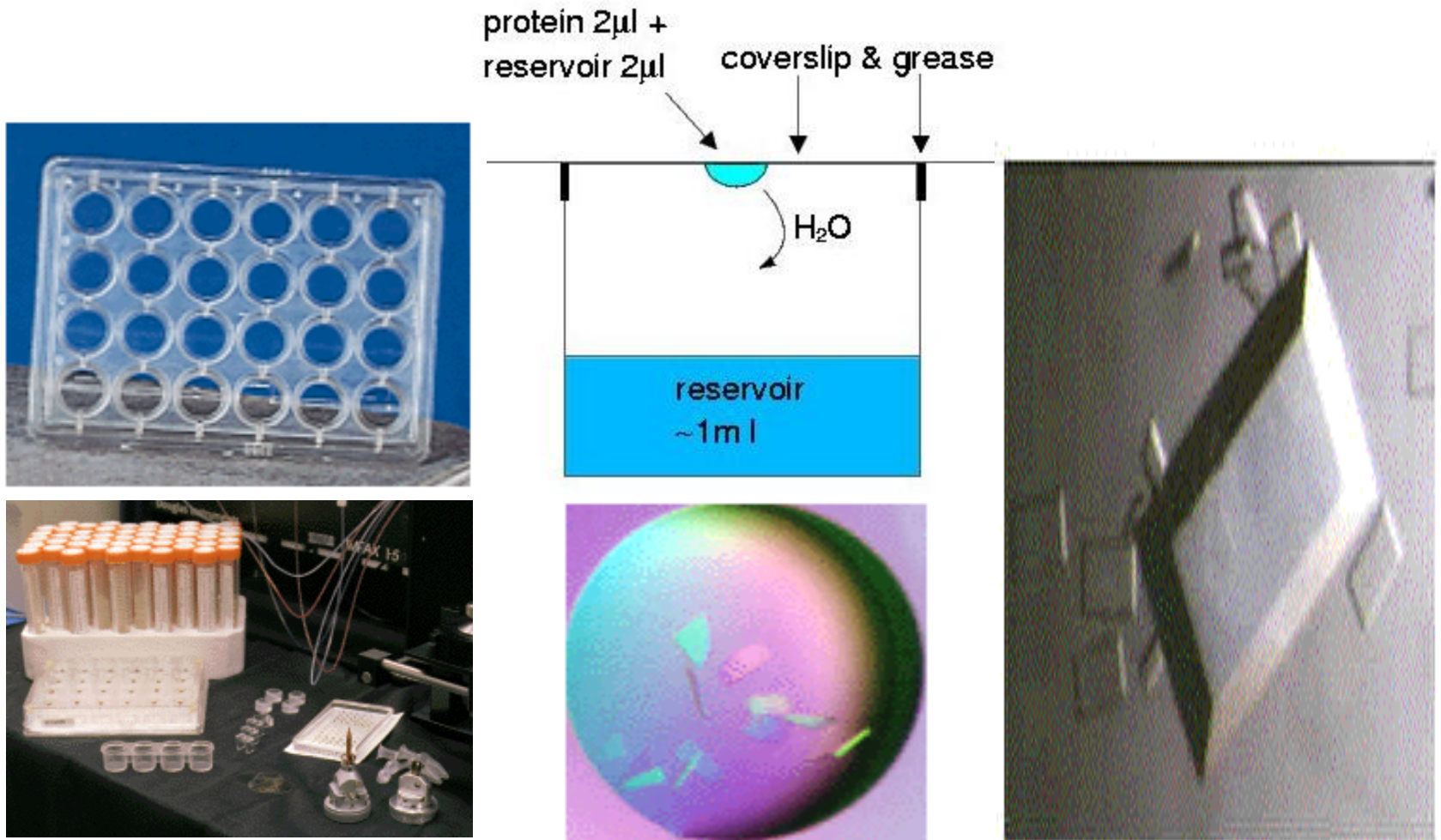


crystal



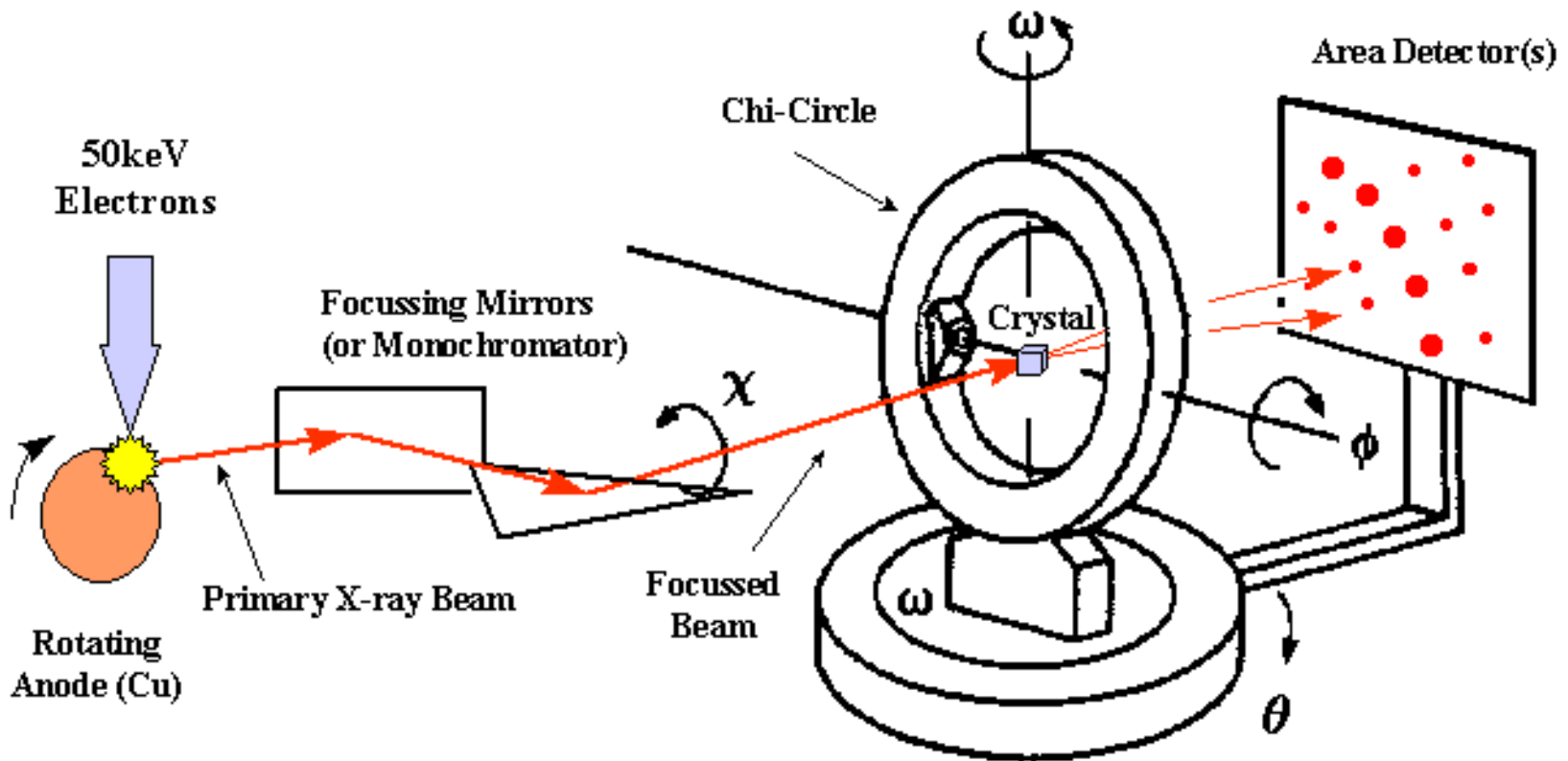
A Crystal

Crystallization



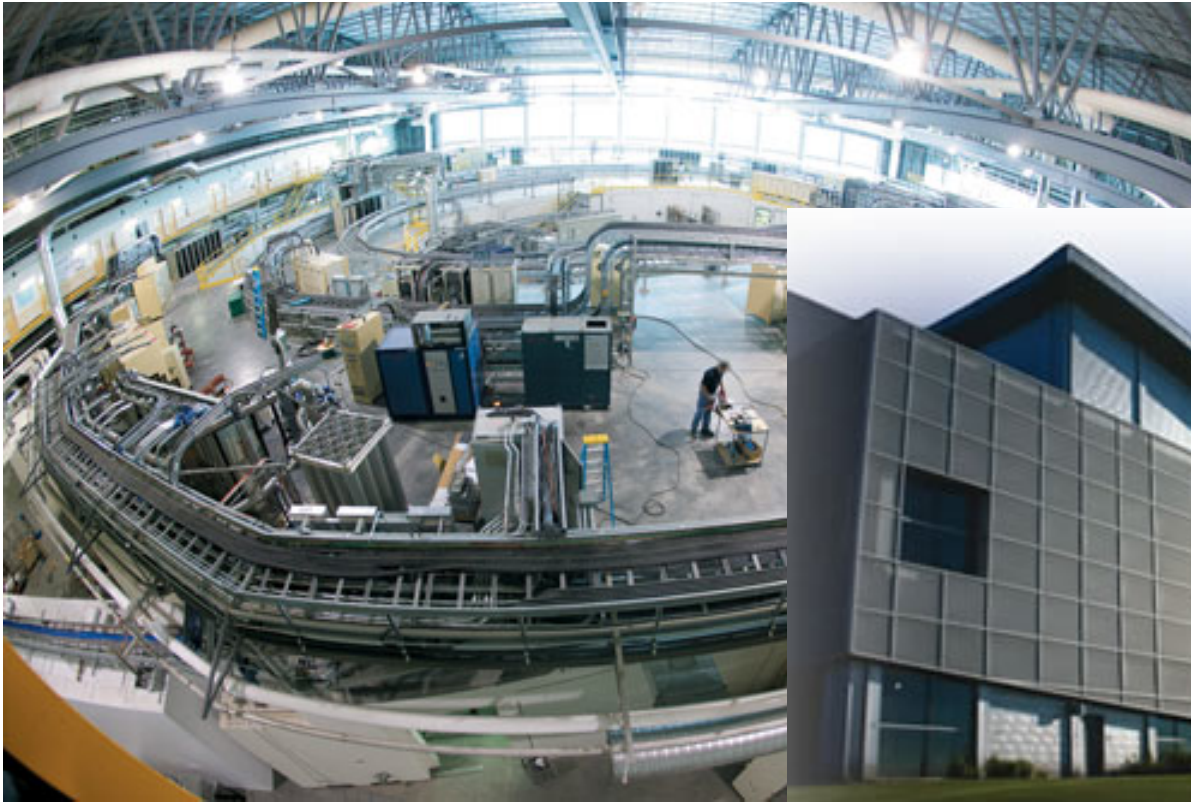
Hanging Drop Experiment for Crystallization

Diffraction Apparatus



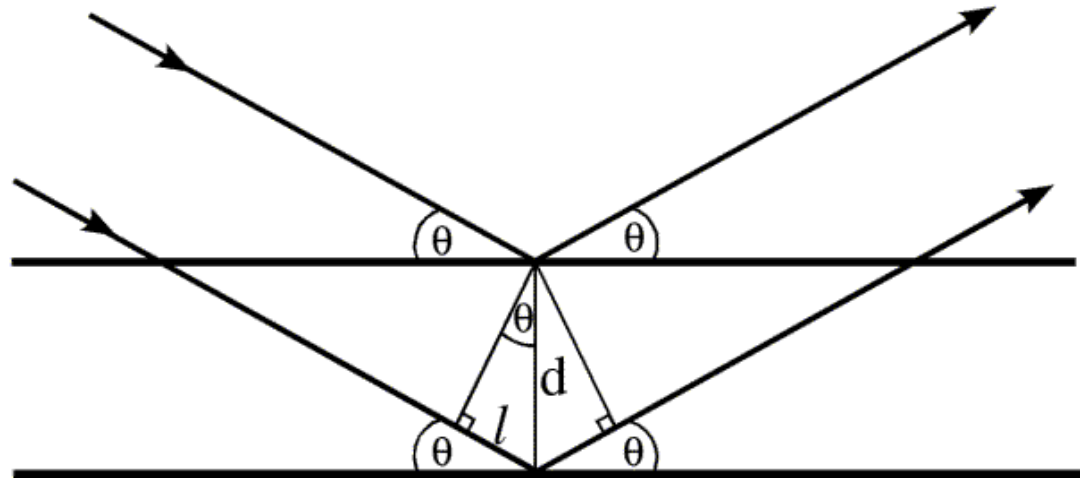
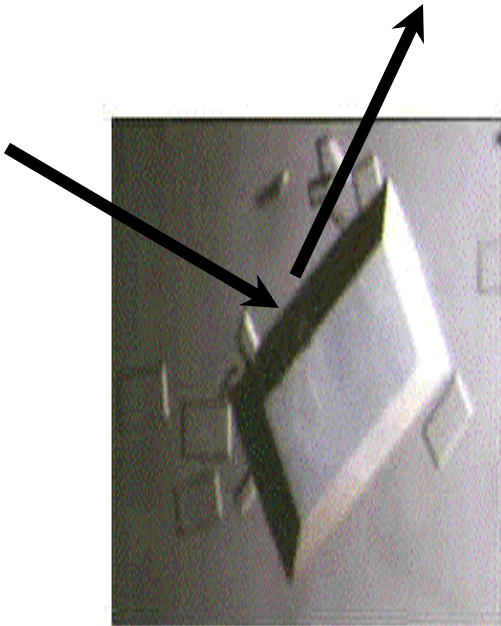
4-Circle Goniometer (Eulerian or Kappa Geometry)

A Bigger Diffraction Apparatus



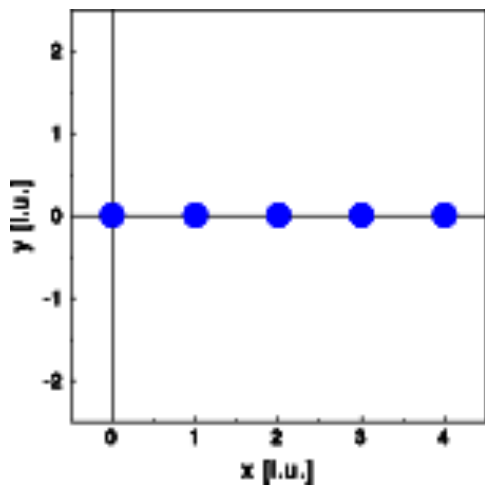
Synchrotron Light Source

Diffraction Principles***

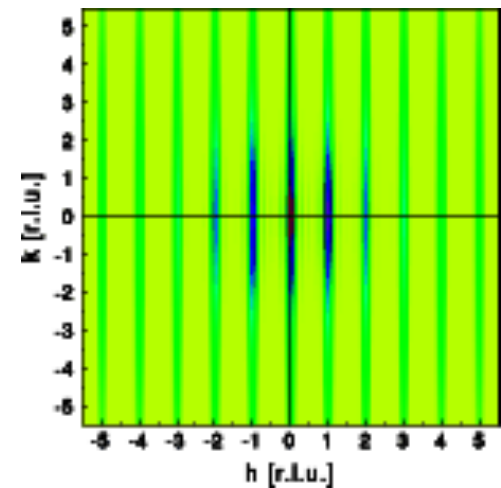
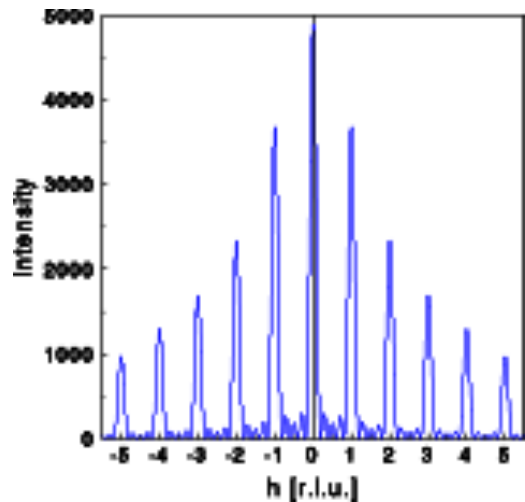


$$n\lambda = 2d\sin\theta$$

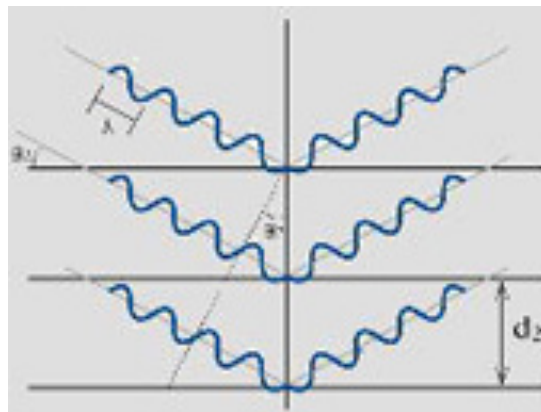
Diffraction Principles



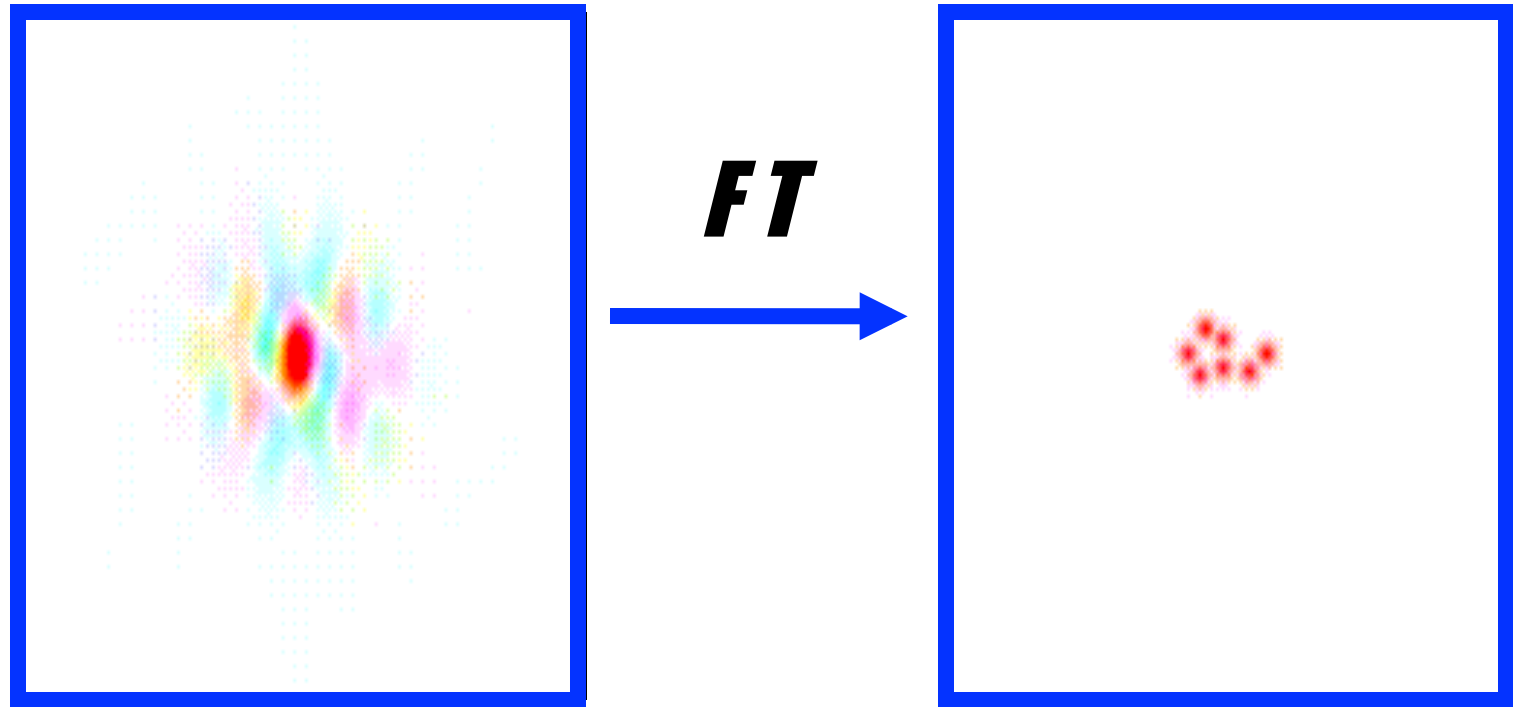
A string of atoms



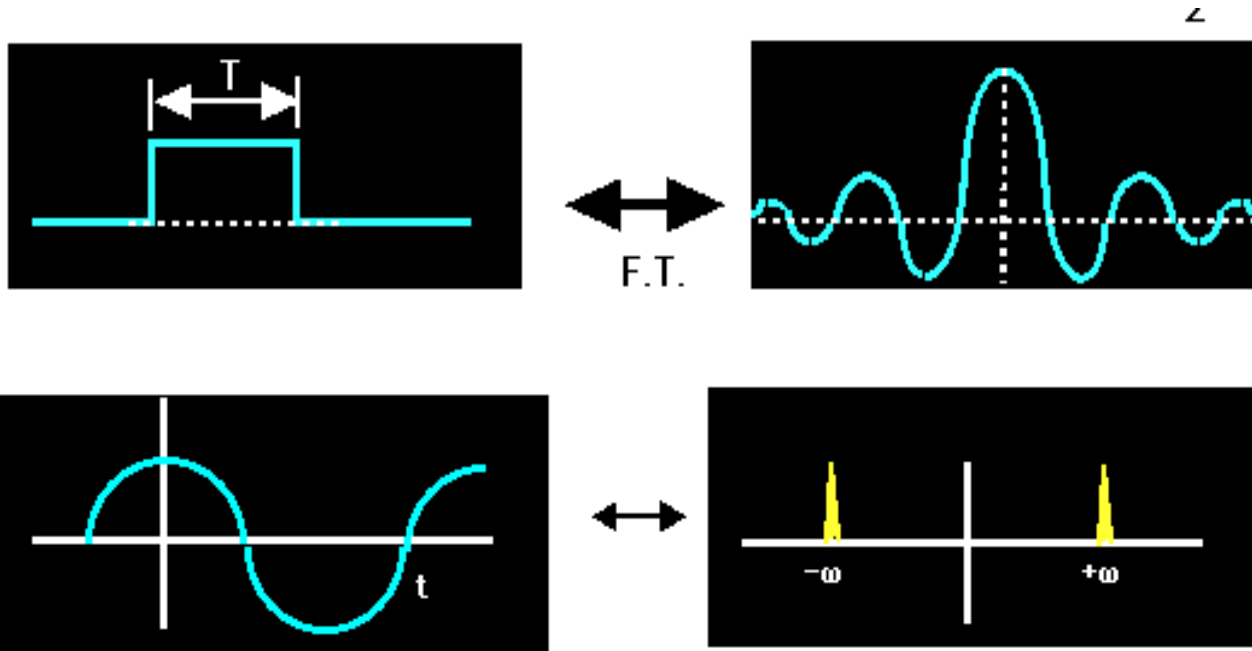
**Corresponding
Diffraction Pattern**



Converting Diffraction Data to Electron Density



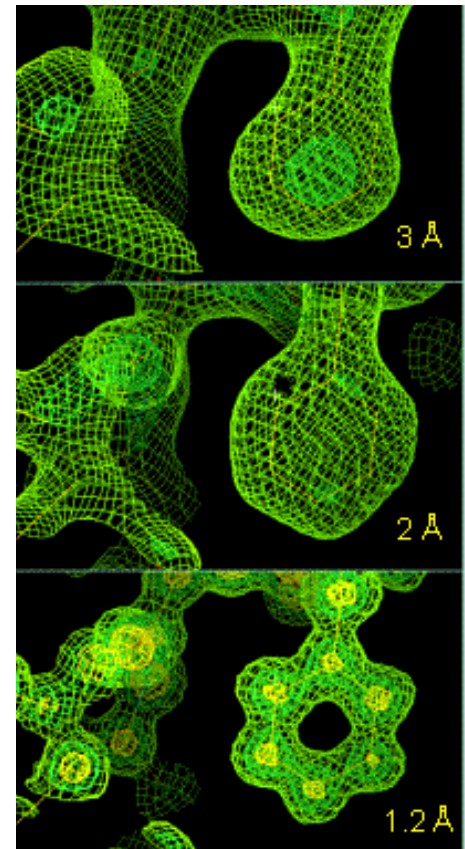
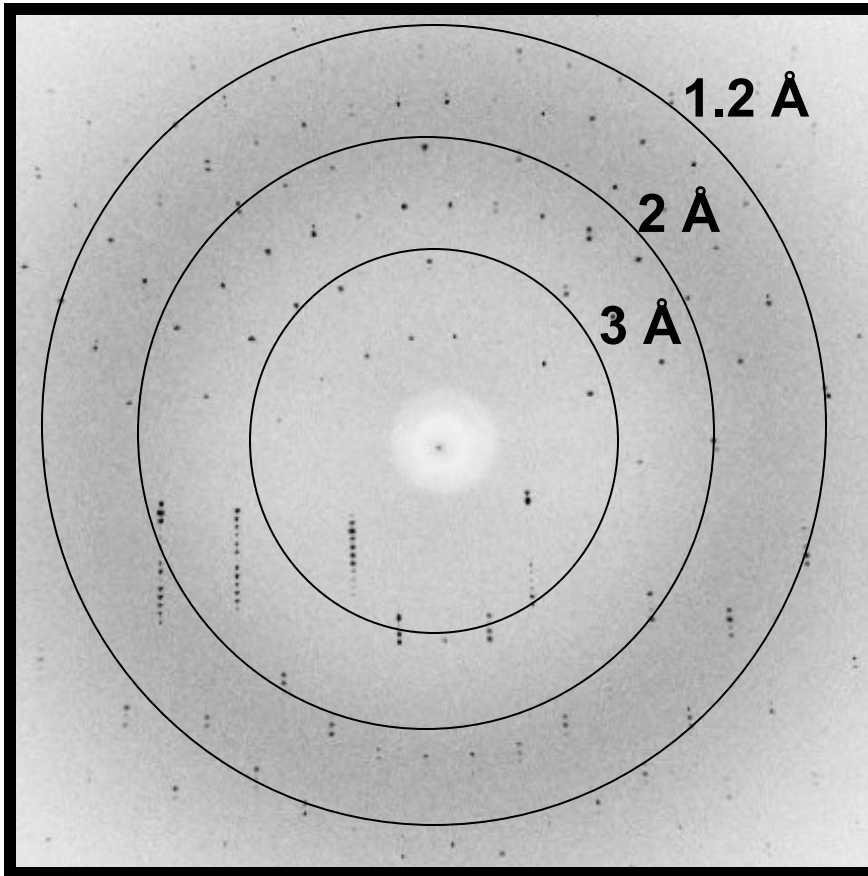
Fourier Transformation



$$F(x,y,z) = \int f(hkl) e^{i(xyz)(hkl)} d(hkl)$$

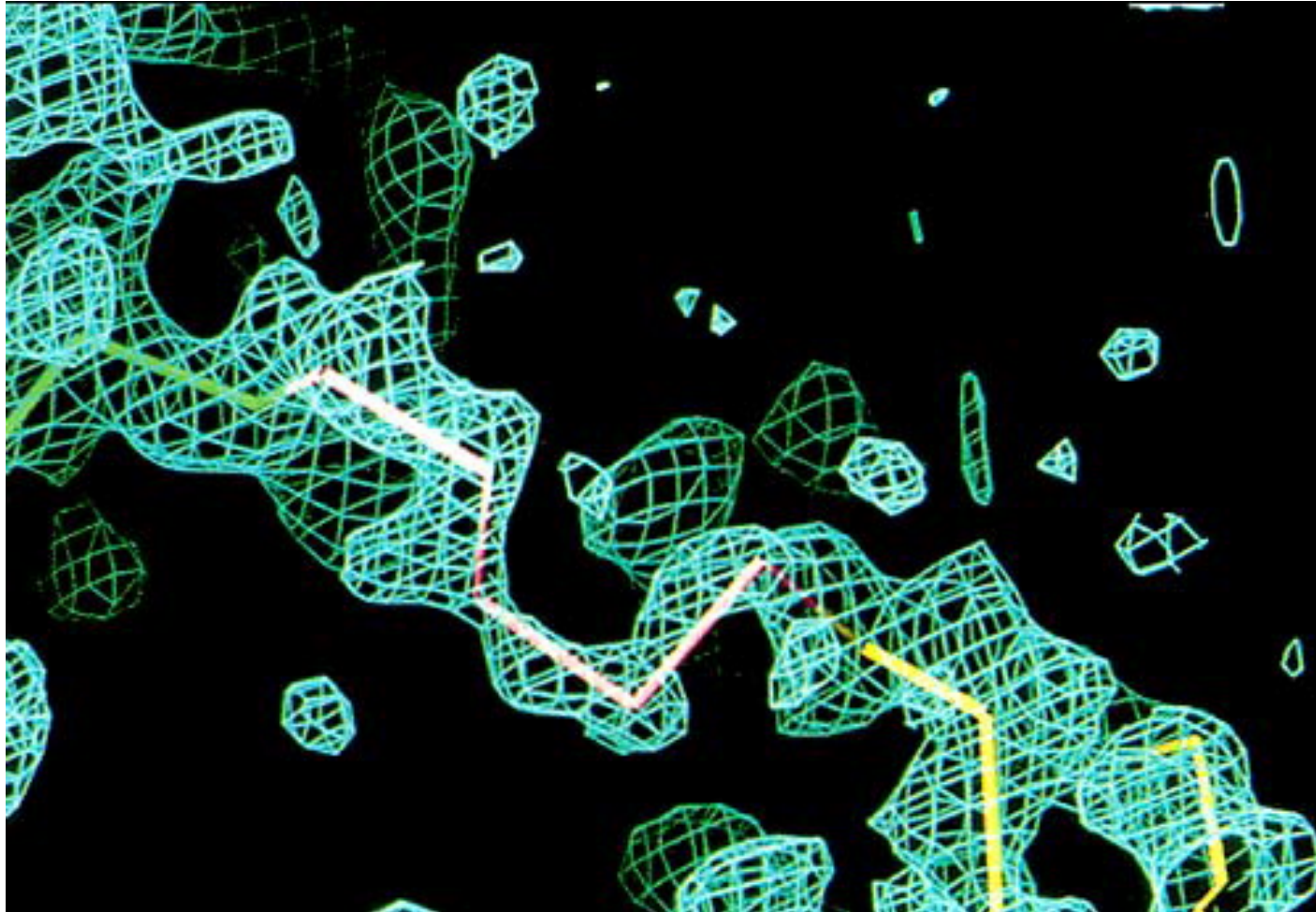
Converts from units of inverse space to cartesian coordinates

Resolution



Resolution describes the ability of an imaging system to resolve detail in the object that is being imaged.

Electron Density Tracing



Crystallography (Then & Now)



1959

2010

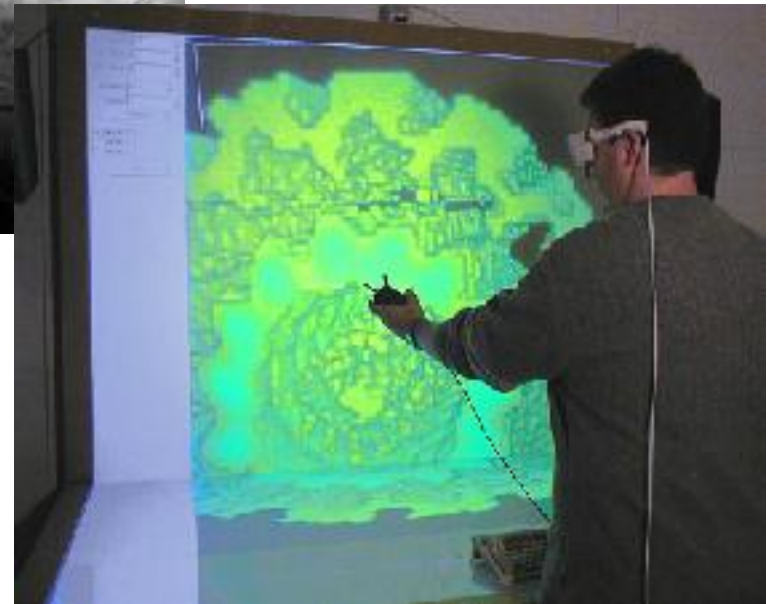


Crystallography (Then & Now)

1953



2010



X-ray Crystallography

- **Key is to measure both phase and amplitude of X-rays (unfortunately we can't measure phase)**
- **Trick is to guess phase, use a crutch (anomalous dispersion) or calculate the phase using pattern recognition (direct method)**
- **Direct method (purely computational) works for small molecules (<1000 atoms) but not for large**
- **Anyone who solves the “direct phasing problem” for all molecule sizes wins the Nobel Prize**

Computational Challenges in X-ray Crystallography

- Solving the direct phase problem
 - Algorithmics, Parallelism
- Developing robotic crystallography stations (doing what humans do)
 - Robotics
- Predicting and planning optimal crystallization conditions
 - Machine learning, Neural Nets
- Automated electron density tracing
 - AI, Machine learning

2 Main Methods to Solve Structures in Chemistry

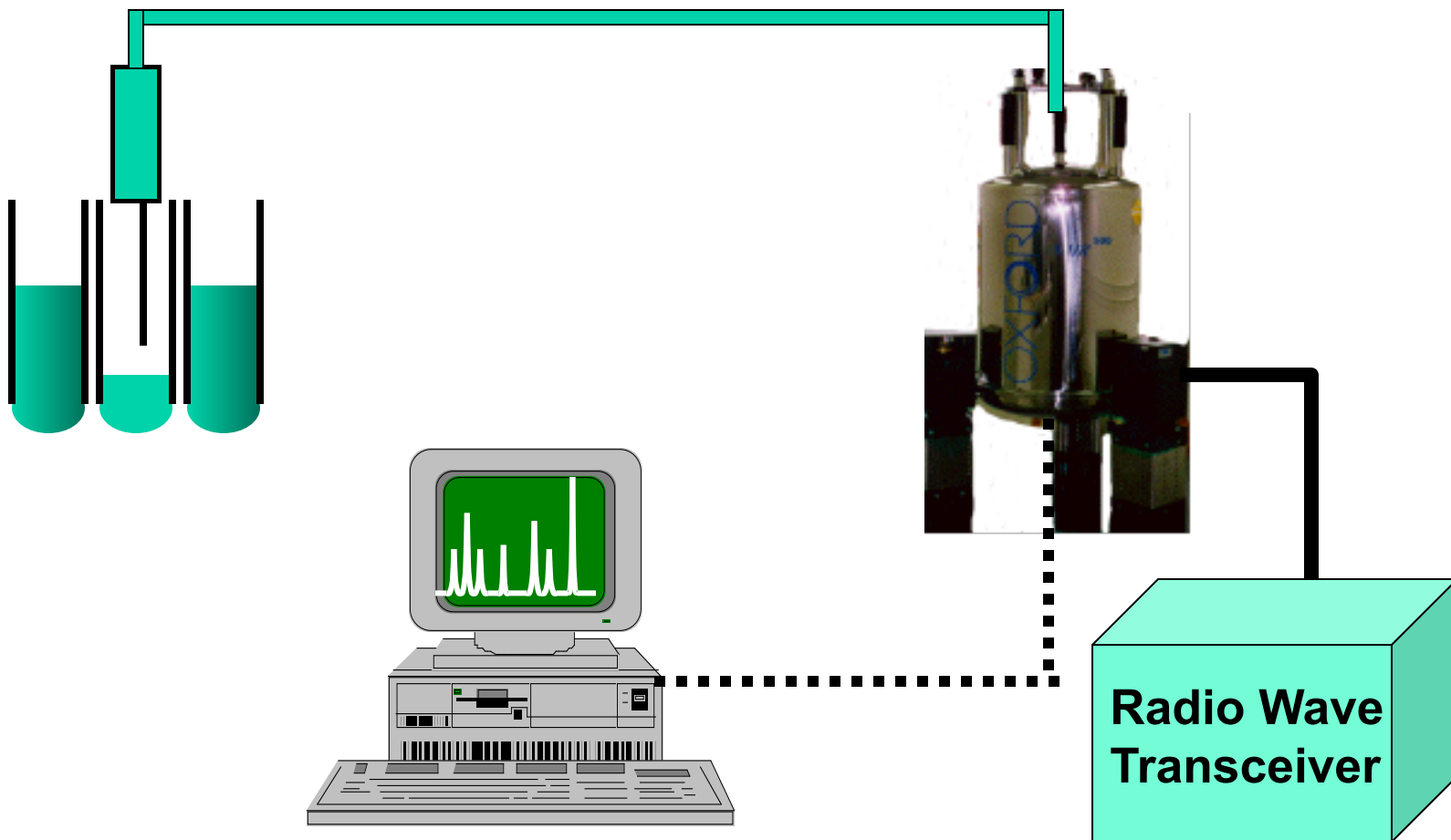


X-ray



NMR

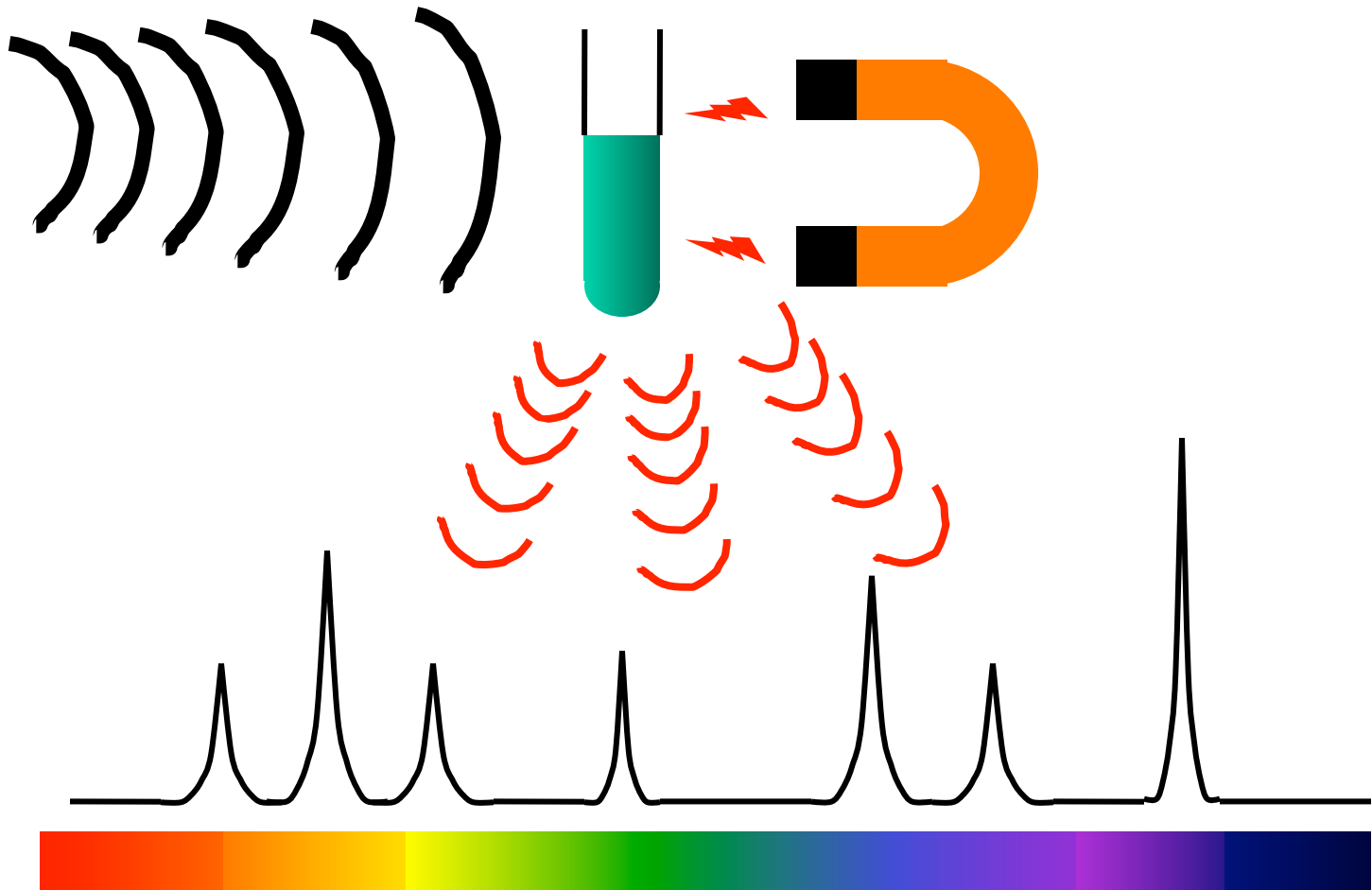
NMR Spectroscopy



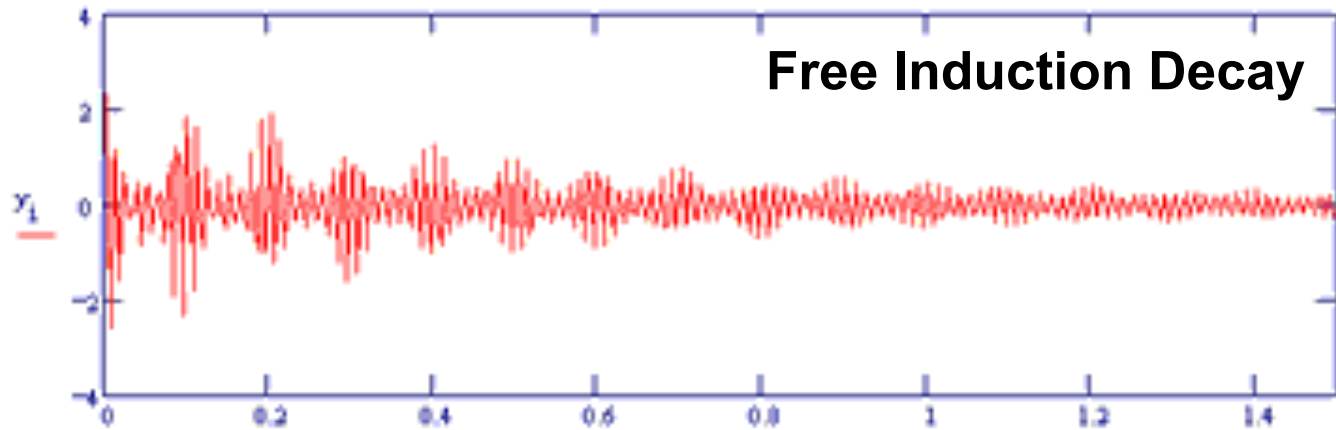
Principles of NMR

- Measures **nuclear** magnetism or changes in nuclear magnetism in a molecule
- NMR spectroscopy measures the absorption of light (radio waves) due to changes in nuclear spin orientation
- NMR only occurs when a sample is in a strong magnetic field
- Different nuclei absorb at different energies (frequencies)

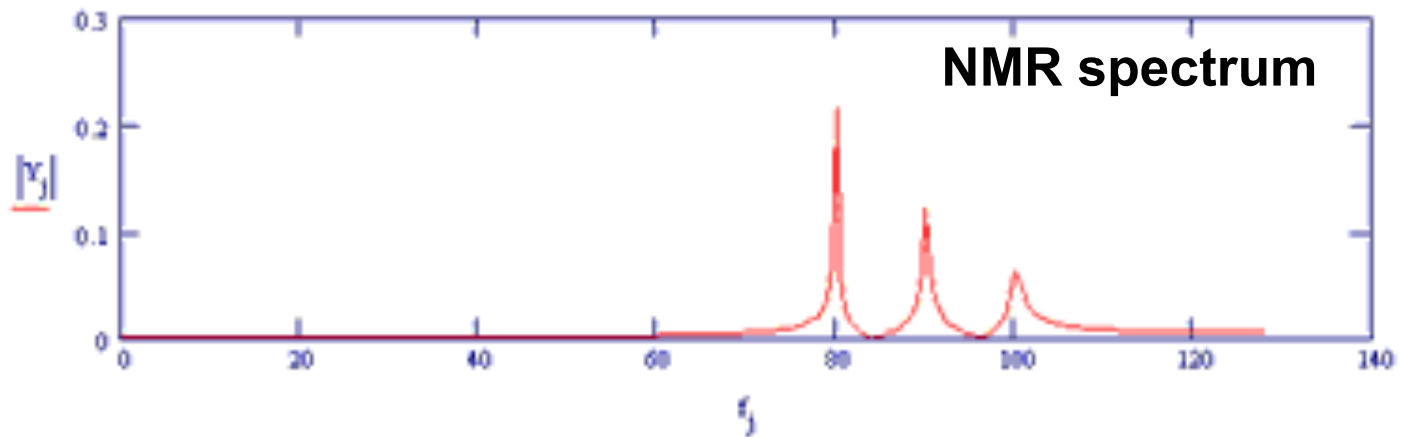
Principles of NMR



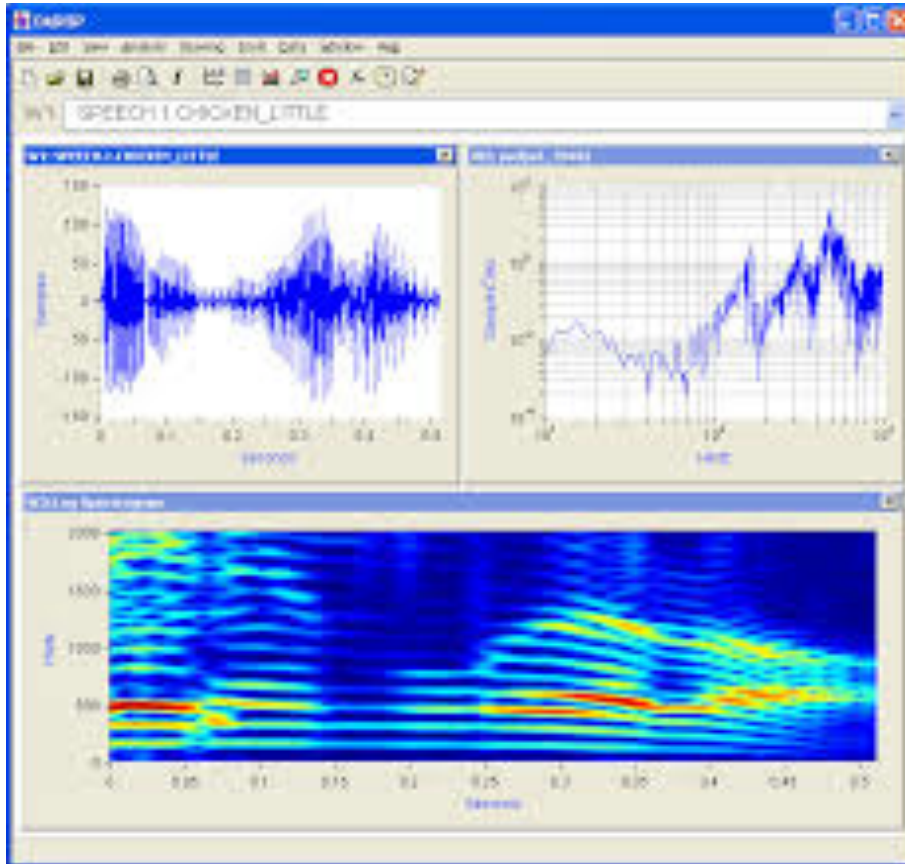
FT NMR



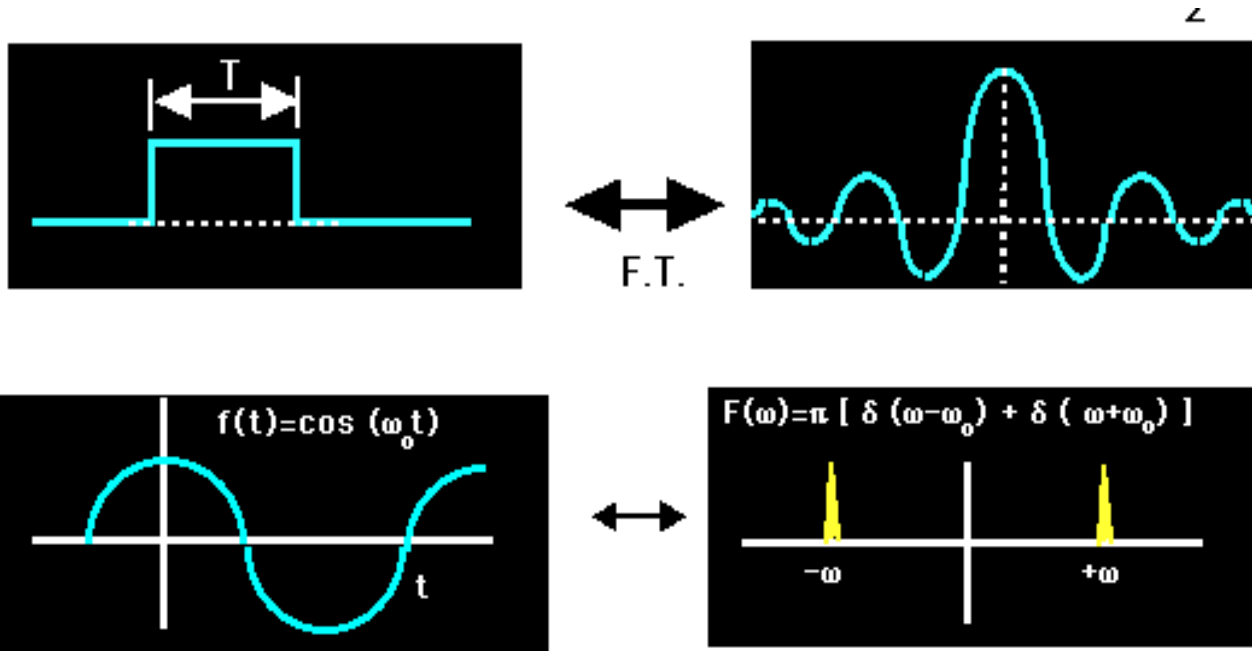
FT \downarrow



Signal Processing



Fourier Transformation

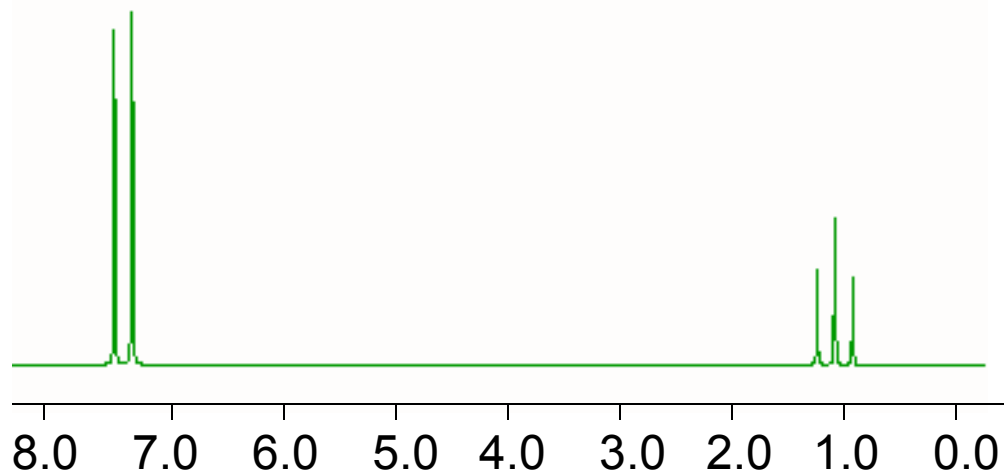


$$F(\omega) = \int f(t) e^{i\omega t} dt$$

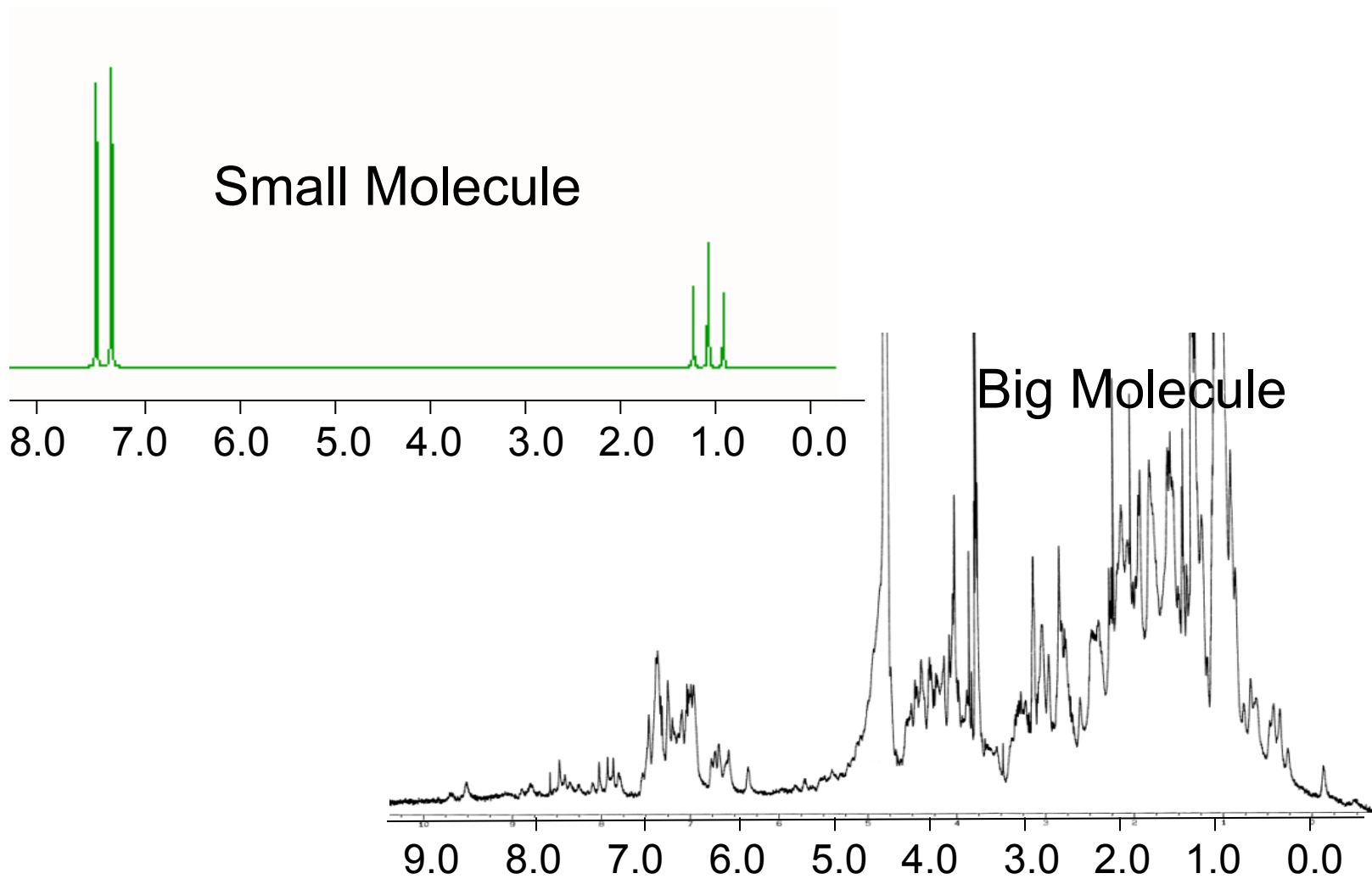
Converts from units of time to units of frequency

^1H NMR Spectra Exhibit...

- **Chemical Shifts (peaks at different frequencies or ppm values)**
- **Splitting Patterns (from spin coupling)**
- **Different Peak Intensities (# ^1H)**

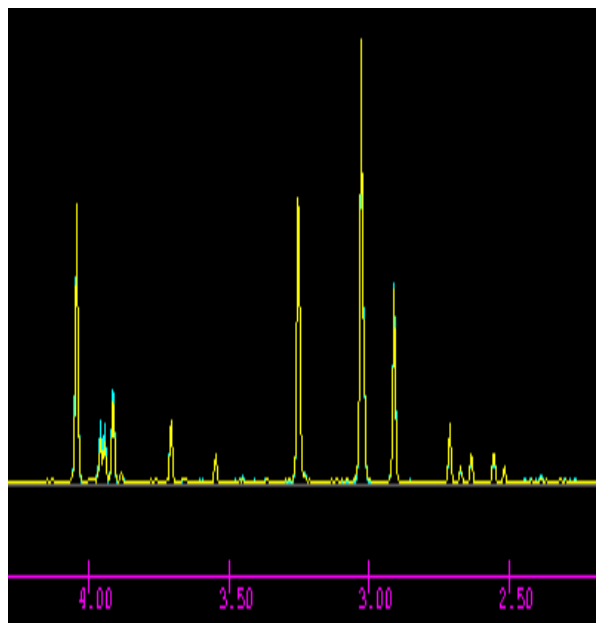


NMR Spectra



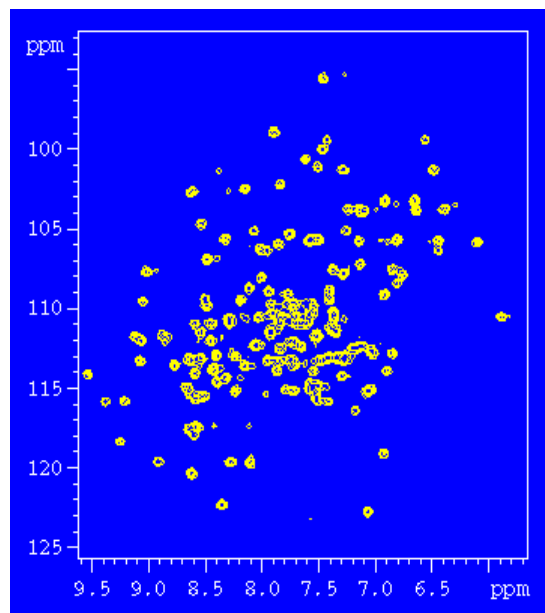
Multidimensional NMR

1D



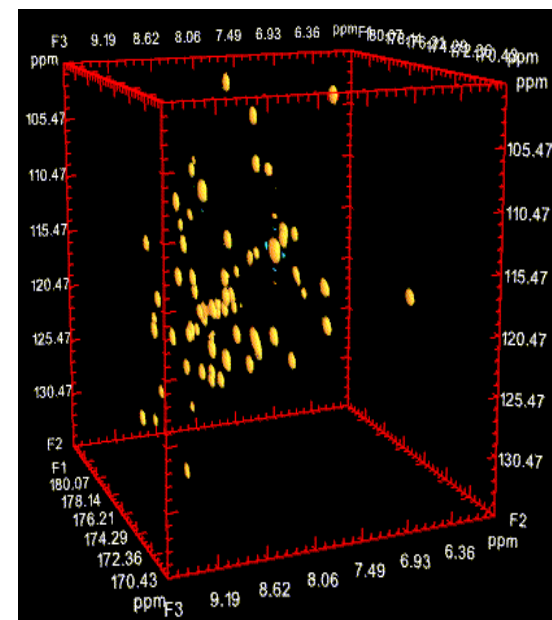
MW ~ 500

2D



MW ~ 10,000

3D

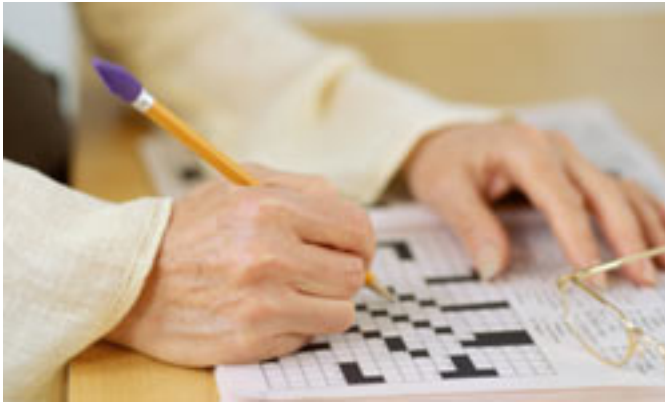


MW ~ 30,000

The NMR Challenge

- **Peak positions tell you atom types**
- **Peak clusters tells about atom type proximity or neighborhood**
- **Peak intensities tell you how many atoms**
- **How to interpret peak intensities, peak clusters and peak positions to generate a self-consistent structure?**

Solving a Crossword Puzzle



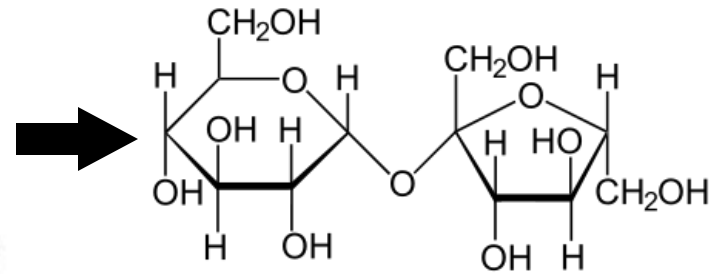
- **Dictionary of words and definitions (or your brain)**
- **Match word length**
- **Match overlapping or crossing words**
- **All words have to be consistent with geometry of puzzle**

NMR Spectroscopy (The Old Way)

Peak Positions

Peak Height

J-Couplings

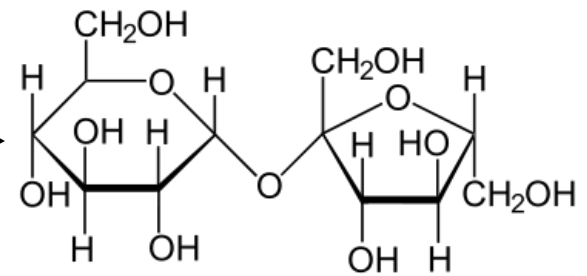
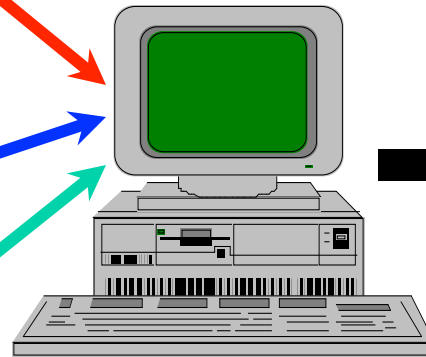


NMR Spectroscopy (The New Way)

Peak Positions

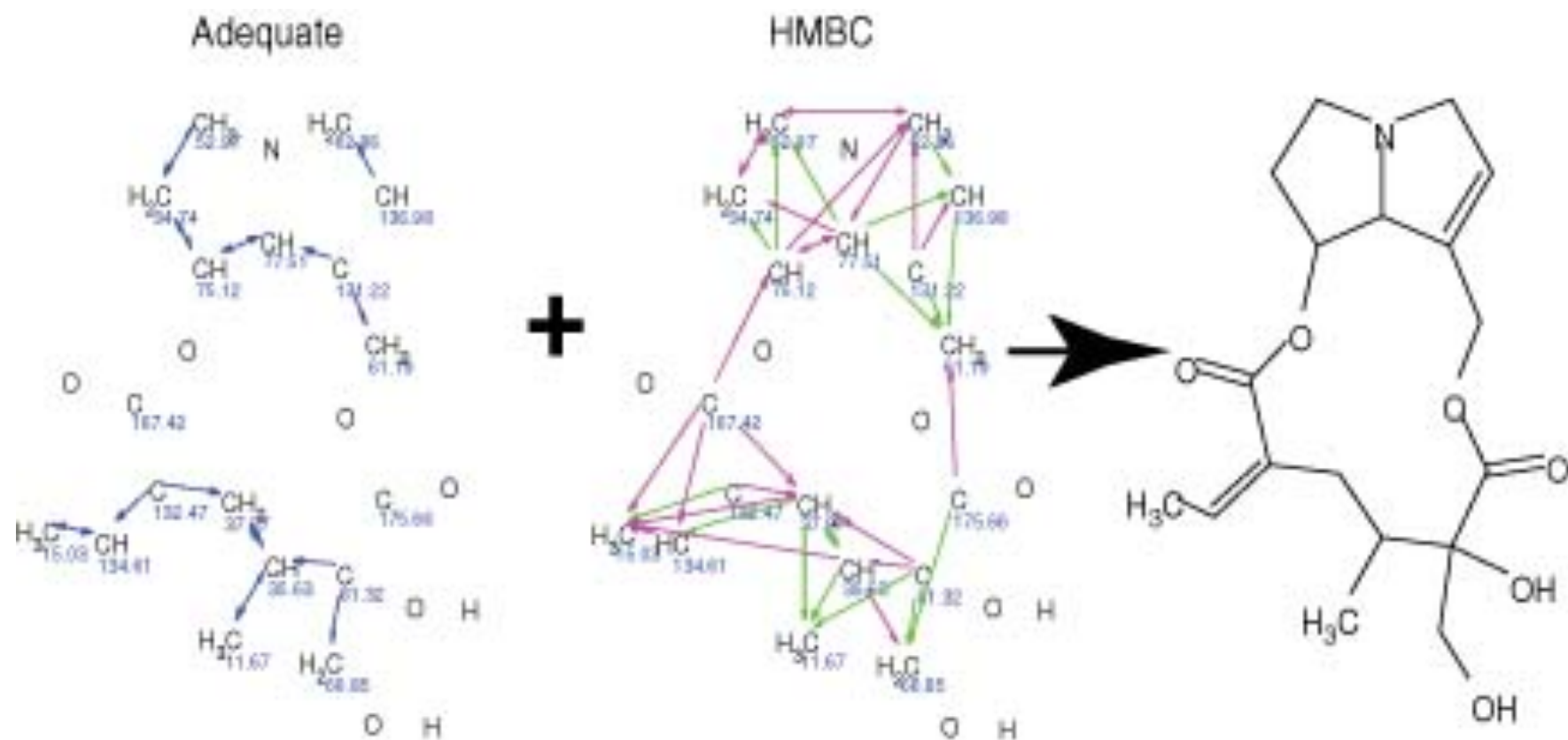
Peak Height

J-Couplings

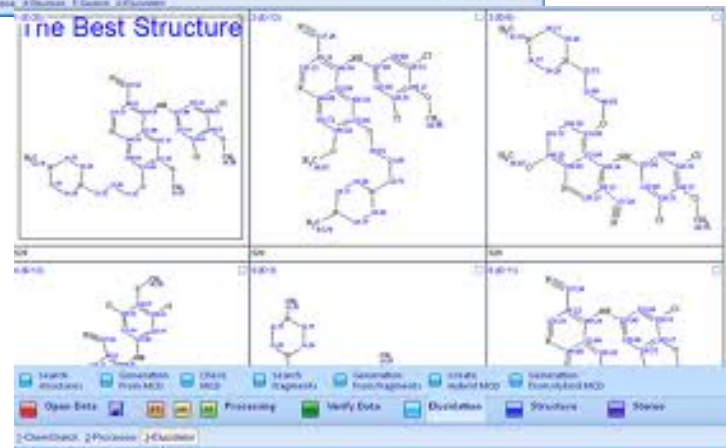
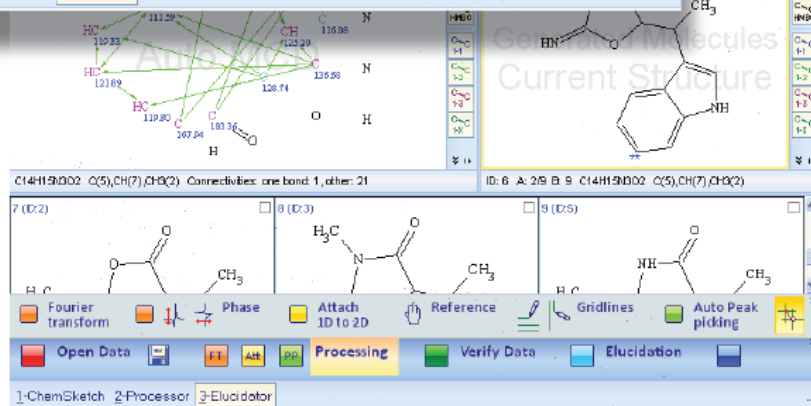
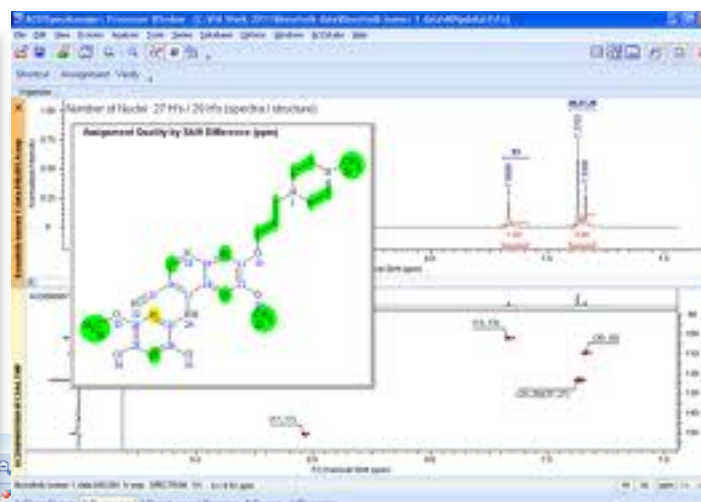
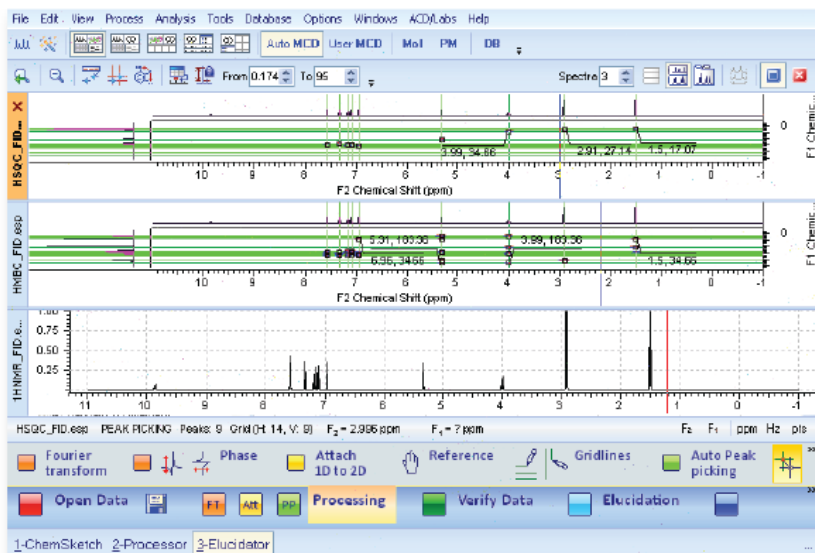


Computer
Aided
Structure
Elucidation

Computer-Aided Structure Elucidation



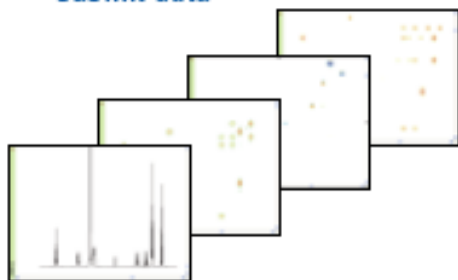
Structure Elucidator



Structure Elucidator

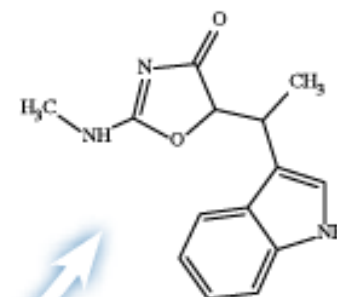
How it Works

Submit data

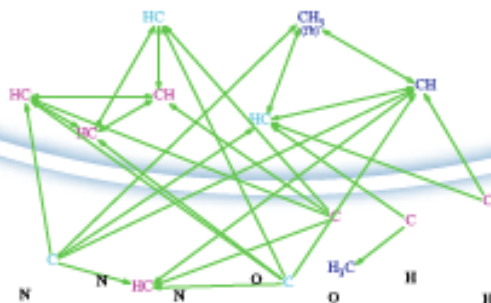


ACD/Structure Elucidator uses molecular formula, 1D, and 2D NMR data to generate a Molecular Connectivity Diagram (MCD). The information in this diagram is used to generate potential structures.

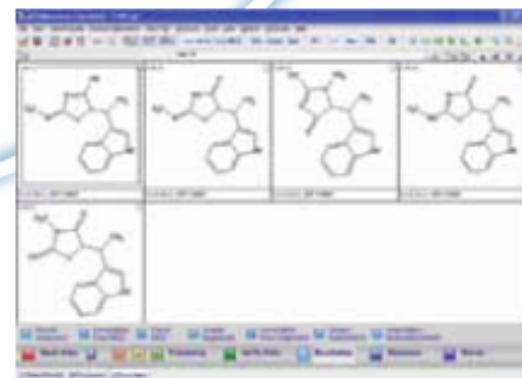
Identify best structure



MCD is automatically created



The MCD is also an interactive tool that a user can easily modify to add or remove information for consideration.



Beating Human Experts

Review

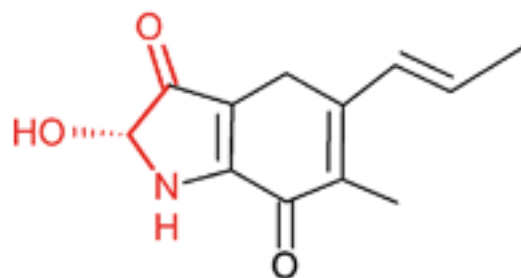
Article citation: Mikhail Elyashberg, *Nat. Prod. Rep.*, 2010, DOI: 10.1039/c002332a

Structural revisions of natural products by Computer-Assisted Structure Elucidation (CASE) systems

Mikhail Elyashberg, Antony J. Williams and Kirill Blinov

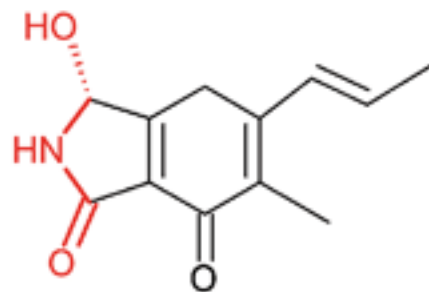
Covering: up to the end of 2009

Original structure



$d_A = 10.6$ ppm

Revised structure



$d_A = 2.02$ ppm

Key Computational Challenges in NMR

- **Solving structures for large molecules (i.e. proteins or RNA) using automated CASE methods**
 - **Monte Carlo Sampling, Neural Nets**
- **Extracting information about molecular motions from raw NMR data**
 - **Pattern recognition, Machine Learning**

Jobs in Computational Chemistry

- **Pharmaceutical and biotechnology companies**
- **Chemical products companies**
- **Universities and national labs**
- **Chemistry software development companies**
- **Cheminformatics – a rapidly growing field (not as large as bioinformatics)**

Questions?

david.wishart@ualberta.ca

3-41 Athabasca Hall