"Seeing" Biological Polymers with X-Rays in Solution

Alexandra M. Deaconescu
Quantitative Biology Bootcamp
Brandeis University
January 2013

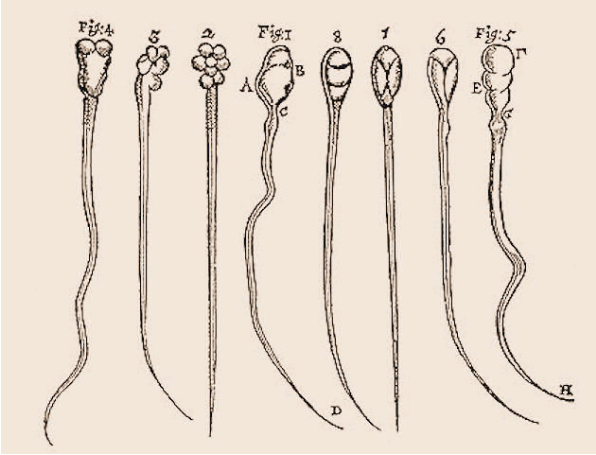
How can we “see” biological polymers?

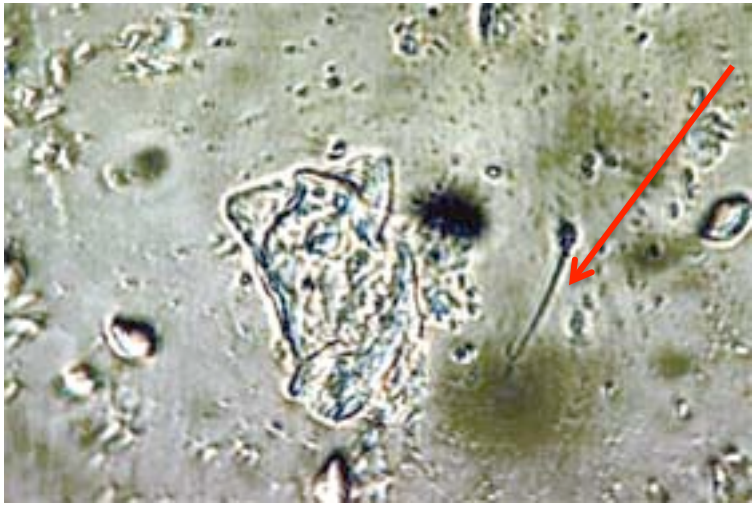
(ATOMIC or NEAR-ATOMIC) STRUCTURE DETERMINATION TECHNIQUES

- X-ray, neutron and electron crystallography
- Nuclear magnetic resonance (NMR)
- Cryo-electron microscopy (EM)
- Small-angle X-ray scattering (SAXS)

Why Is It Important to See Biological Assemblies and Polymers?

Structure Dictates Function





http://www.brianjford.com/wav-spf.htm

Antony van Leeuwenhoek, 1678

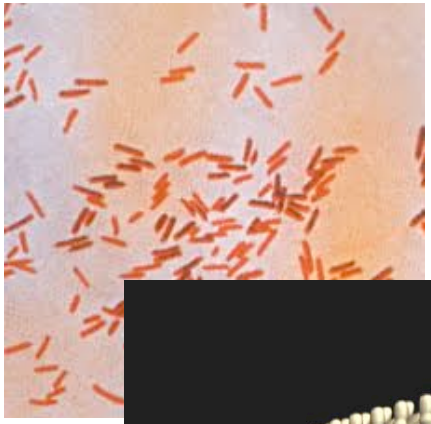
1674 The Infusoria - (Protist class in modern Zoology)

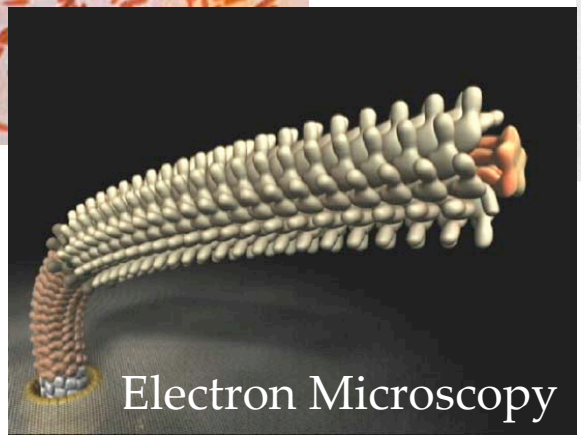
1676 The Bacteria (Genus Selenomonas - crescent shaped bacteria from human mouth)

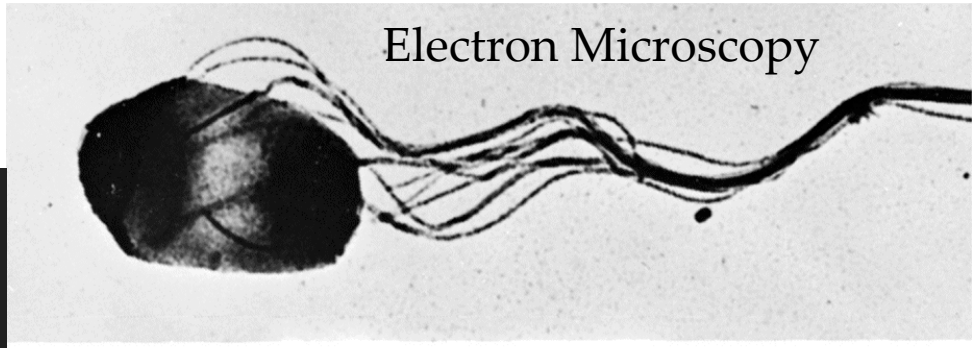
1677 The Spermatozoa

1682 The banded pattern of muscular fibers

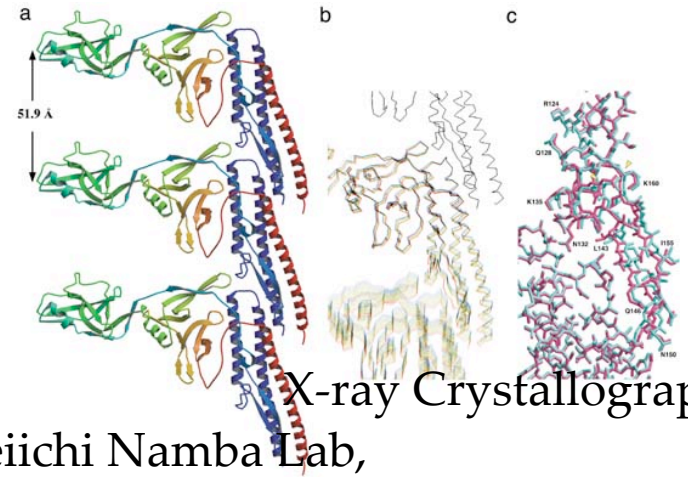
Structure Dictates Function: Bacterial Flagella





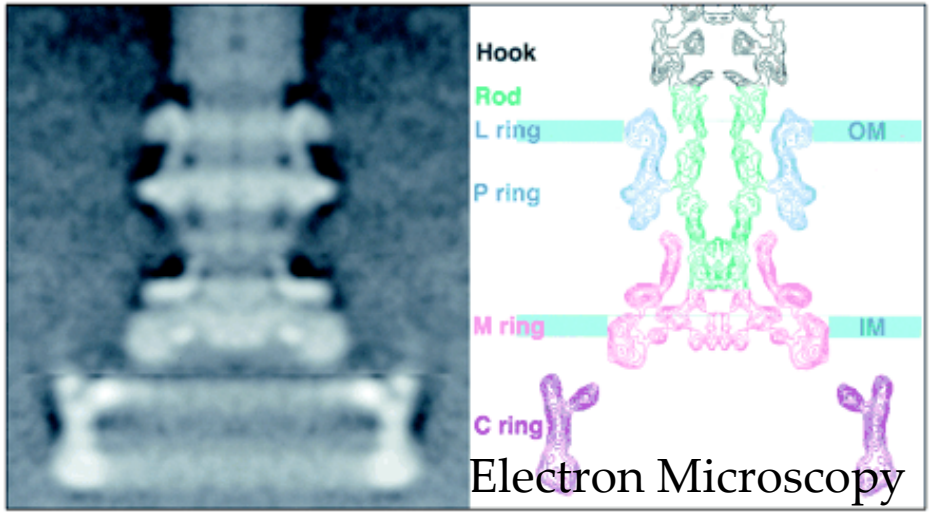


http://www.bmb.leeds.ac.uk/illingworth/6form/index.htm



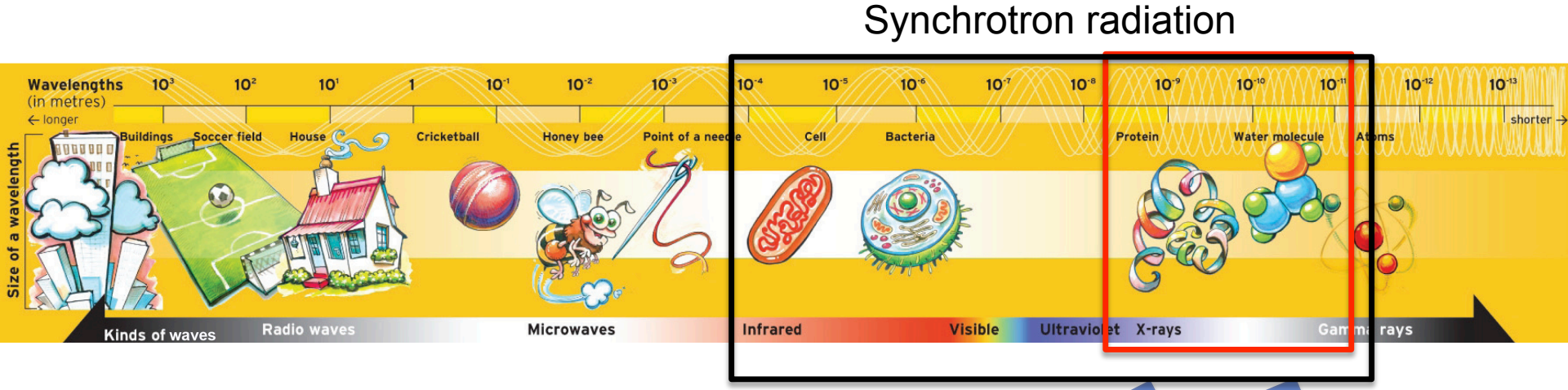
X-ray Crystallography

Keiichi Namba Lab,
Osaka University



Thomas et al., Mol Bio Cell (1999)

The Electromagnetic Spectrum



Soft X-Rays
(10 nm to 0.1n m)
Easily absorbed in air

Hard X-Rays
(<0.1nm)
Penetrant

- X-ray crystallography
- SAXS

(wavelength ~1Å)

http://aworldwideawake.com/2012/01/

Electrons, Neutrons or X-rays?

d ~ Wavelength

- **Electrons (**diffraction, EM) ➔ **Coulombic potential maps** (λ=pm at 200kV)

- **X-rays** wavelengths (diffraction and SAXS): typically 0.8-1.5Å
  - interact with electron cloud → **electron density maps**
  - can take advantage of anomalous scattering
  - many synchroton beamlines
  - (relatively) easy sample preparation

- **Neutrons** (diffraction and SANS) -> **neutron density maps (**λ=4-20Å @NIST
  - interact with atomic nuclei
  - generate fewer free radicals (minimal radiation damage)
  - very few beamlines (ESRF) and relatively weak and costly
  - low signal to noise ratio
  - more difficult sample preparation (deuteration)

What Can SAXS Do?

- Structure of metal alloys, synthetic polymers, emulsions, porous materials, nanoparticles, **biological macromolecules**
- **Works in solution under close-to-physiological conditions**
- **Measures shape and sizes**
- **Short response time**
- **Ideal for testing environmental parameters (pH, temperature, salt concentration, presence of ligands and cofactors)**

Why Choose SAXS (or not)?

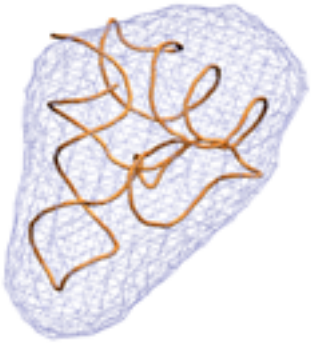
- No need for crystals (X-ray crystallography)
- No need to derivatize with heavy-atoms for phases (X-ray crystallography)
- No conformational selection (X-ray crystallography)
- In solution, under close-to-physiological conditions
- No grid-specimen interaction (EM)
- No staining artifacts (EM)
- Typically faster than X-ray crystallography, NMR or EM

Modest resolution (1-3nm)

How Big is Too Big (or Vice Versa)?

- **No size limitation (unlike in EM, NMR or crystallography)**
- **Suitable for molecules from kDa to megadaltons (nm to µm)**

P. furiosus protein
8.9 kDa
BID: 2HYPHP



SAM-1 Riboswitch
30.1 kDa
BID: 2SAMRR

30S ribosomal subunit
S. Solfataricus
~1MDa
BID: SS30SX

SAXS Development

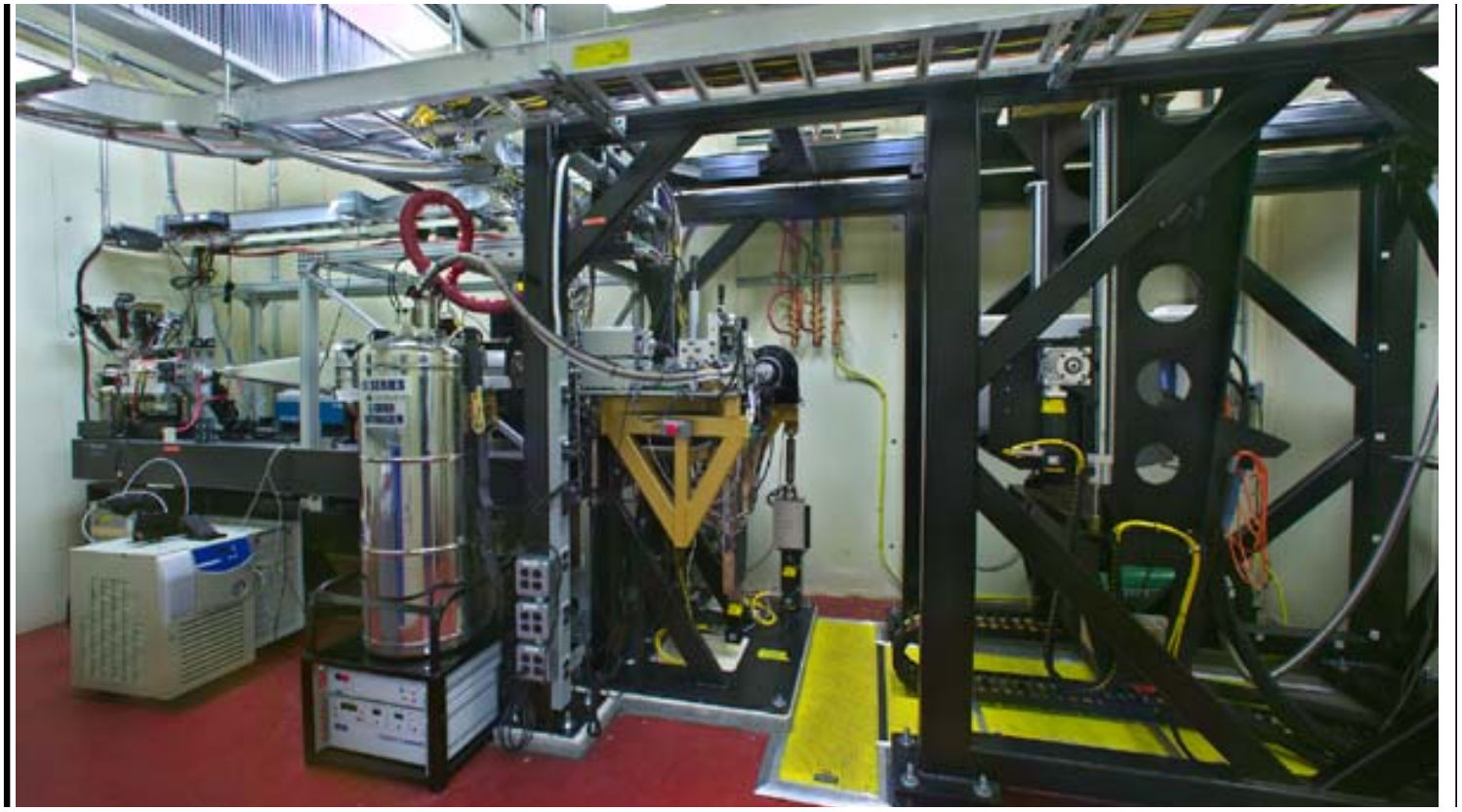
André Guinier (1911-2000)
(www.iucr.org)

With J. Friedel
http://www.lps.u-psud.fr/spip.php?article829&lang=en

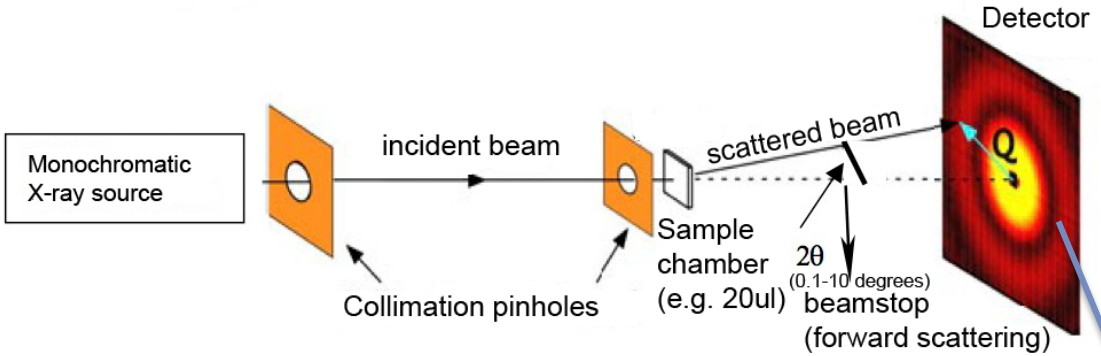
- 1930s-1950s polymers, porous materials (Guinier; Fournet; Kratky)
- 1960s and 1970s - biological SAXS (hardware development)
- 1990s – beginning of ab initio modeling for reconstruction of 3d-envelopes
- Software development: ATSAS by the group of Dmitri Svergun (EMBL)

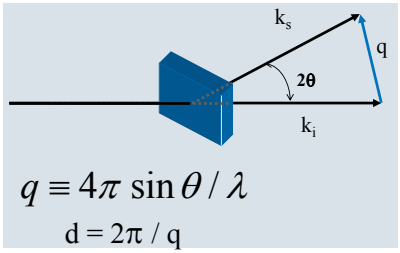
Experimental Setup

- Most hardware to generate, prepare and detect X-rays is shared with crystallography (dual purpose beamlines, e.g. Sibyls at ALS)



Experimental Setup





Adapted from Petoukhov, M., EMBO lecture
Bruker AXS

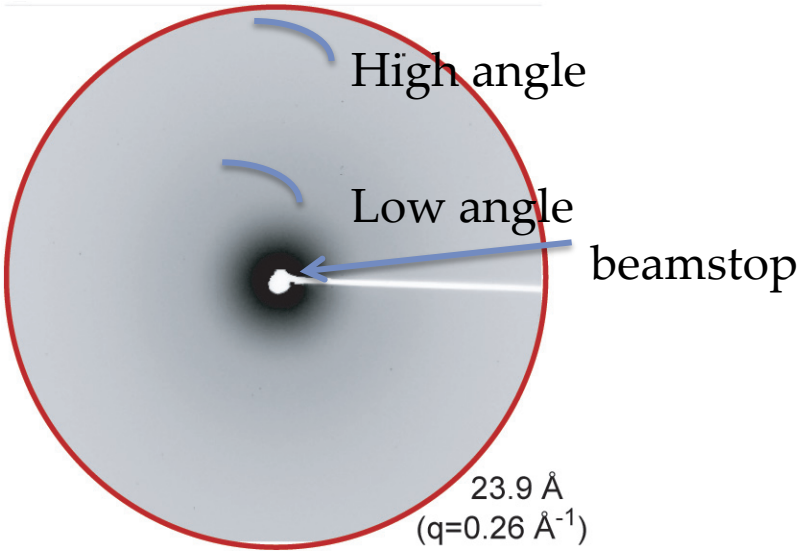
Sample: 1-2 mg (>0.5mg/ml)
Angles = 0-5 degrees
Q range: 0.001 to 0.45 Å-1 (d=μm to nm)

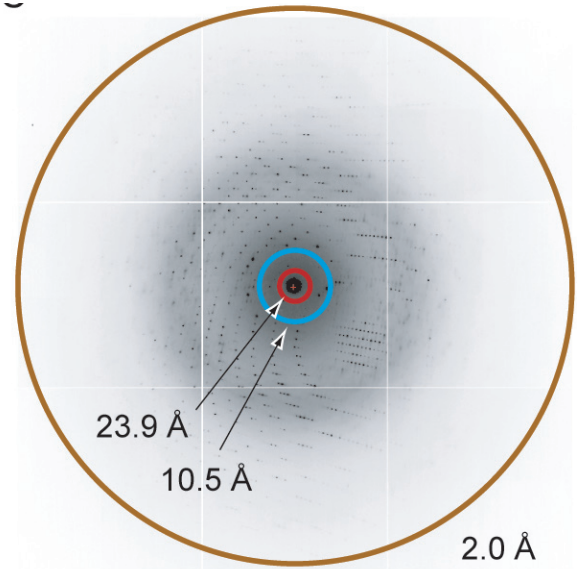
Thomson (elastic) scattering

Variations on the setup

- Flow cell (capillary) instead of a simple sample chambers
  minimize radiation damage
- Flow cell may be in-line with SEC

SAXS versus X-ray Crystallography





- Tumbling molecules
- Radially symmetric (isotropic)
- Low SNR
- Few observations/parameter
- (underdetermined)

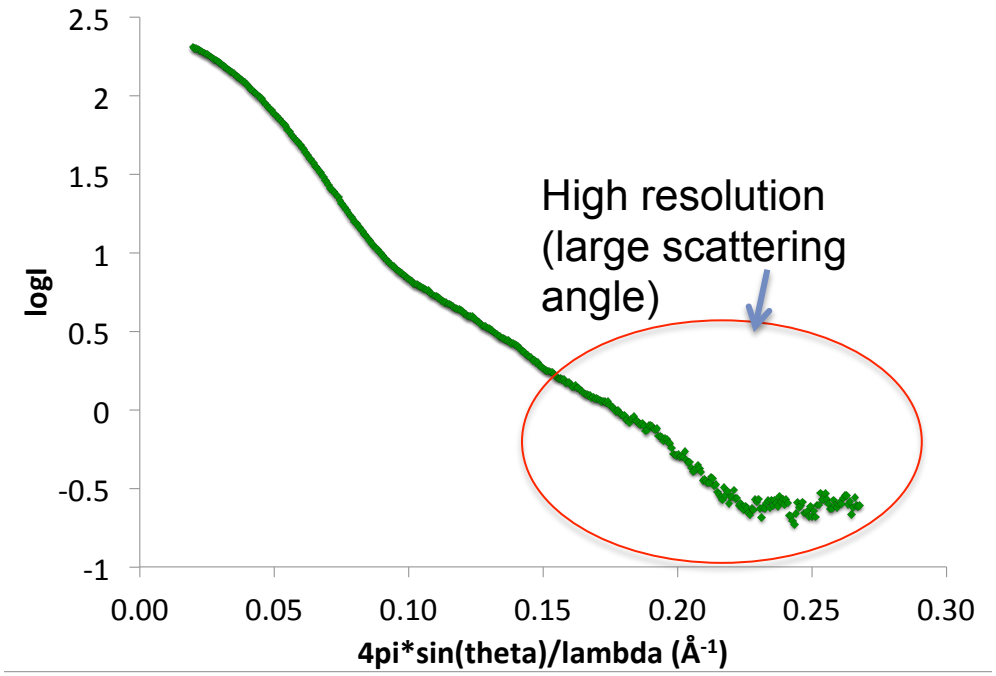
- Molecules “frozen” in lattice
- Non-isotropic ➔
- Convolution of the molecular transform with the lattice
- Discrete maxima
- High SNR
- Crystal needs to be rotated
- Many observations/parameter to be refined (at least at high resolution)

Putnam et al., Q Rev. Biophysics (2007)

Anatomy of a Scattering Intensity Curve

- radially-average intensity distribution to obtain 1-d curve, I(q)
- I is a function of momentum transfer $q=4\pi\sin\Theta/\lambda$ ($Å^{-1}$) or directional momentum change that photons undergo
- Normalization (against exposure time, transmitted sample intensity)

$q_{max}=2\pi/d$ 1/d reciprocal resolution (nominal)



After background subtraction
I~ scattering of single particle averaged over all orientations

What is Being Measured?

1. Scattering from sample of interest (protein)
2. Background scattering (buffer, water, quartz cell etc.)
3. Electronic noise, stray X-rays (not passing through samples)

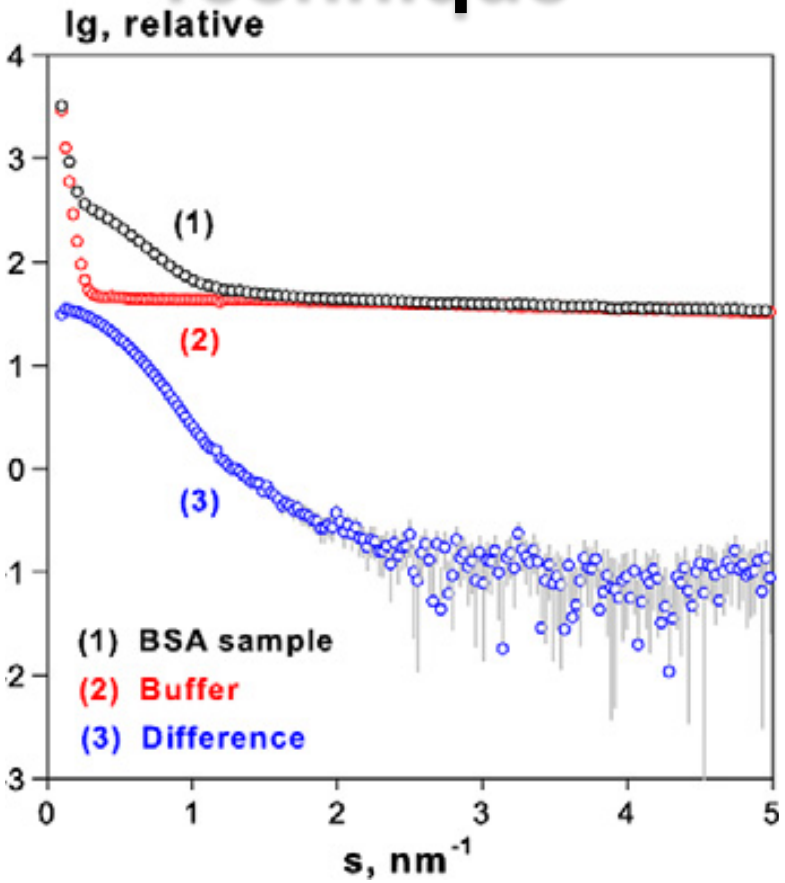
$$I(q) \sim (\rho_p - \rho_s)^2 P(q) S(q)$$

Contrast Factor
ρ=electron density

Form factor (SHAPE and SIZE)

Structure factor (1 for ideal, dilute solutions)

Small-Angle Scattering is a Contrast Technique



Mertens & Svergun (2010)
J. Struct. Biol

- **The contribution of bulk solvent to scattering is explicitly subtracted out**
- **Background subtraction is VERY important** (measure "sample" and "matching buffer" series)

SAXS is a Contrast Technique

$$I \sim \rho_{protein} - \rho_{solvent}$$

$\rho_{protein}$: electron density of protein

$\rho_{solvent}$: electron density of solvent

- **Proteins are made up of light atoms (low Z), which do not scatter very well (as opposed to DNA/RNA, which gives better contrast)**
- **typically 5% above background**

$\rho_{protein} = 0.44\ e^-/Å^3$
$\rho_{water} = 0.33\ e^-/Å^3$

- use relatively large protein concentrations (1-10mg/ml)

Scattering from an Ideal Solution

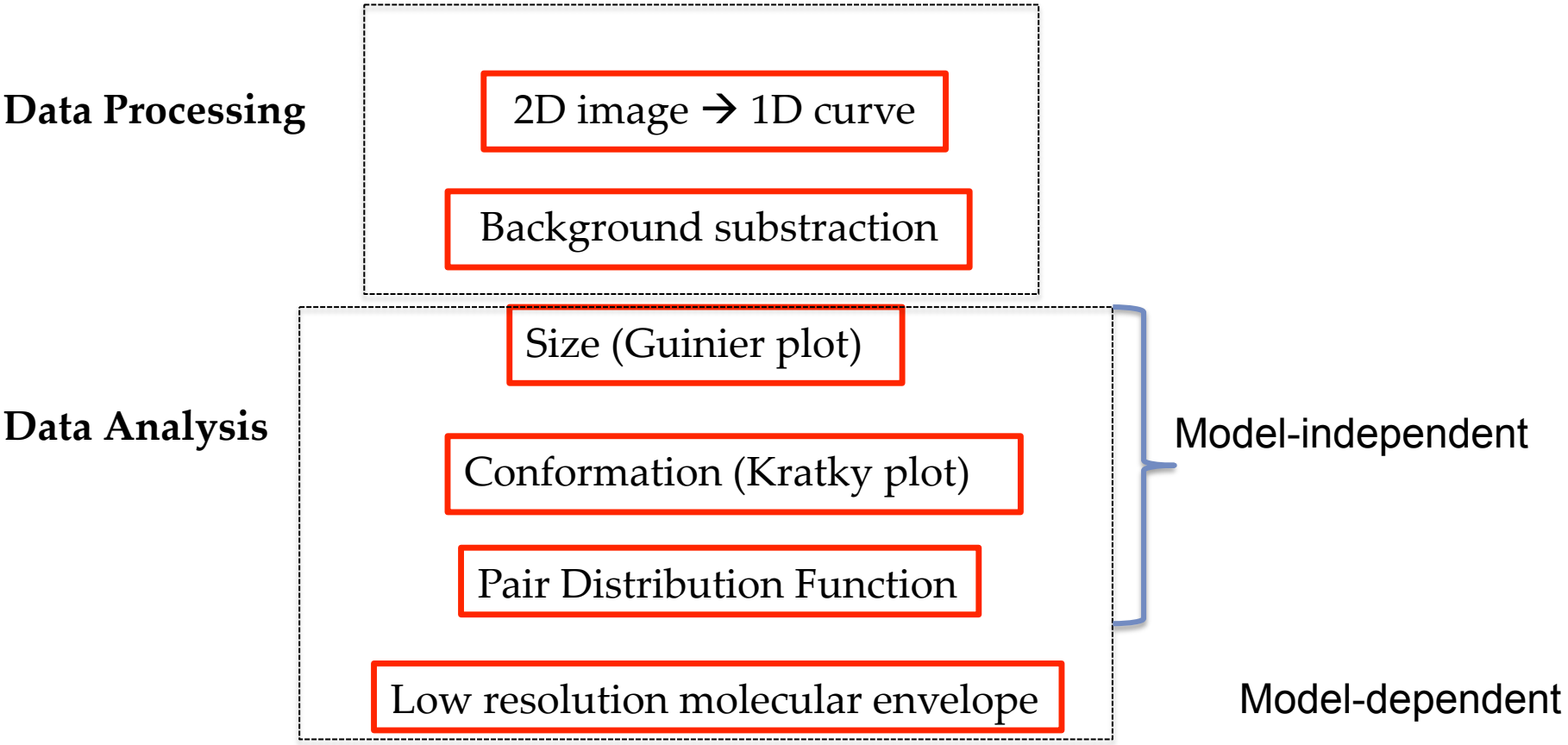
- No interaction between particles (no interparticle interference, e.g. aggregation or repulsion)
- Only one species (monodisperse)
- Particles are free to move (independent scatters)
- **$I(q) = (\rho_1-\rho_s)^2 P(q) S(q)$**

To a limited extent, interparticle interference can be dealt with.

But, for analysis, solution has to be monodisperse.

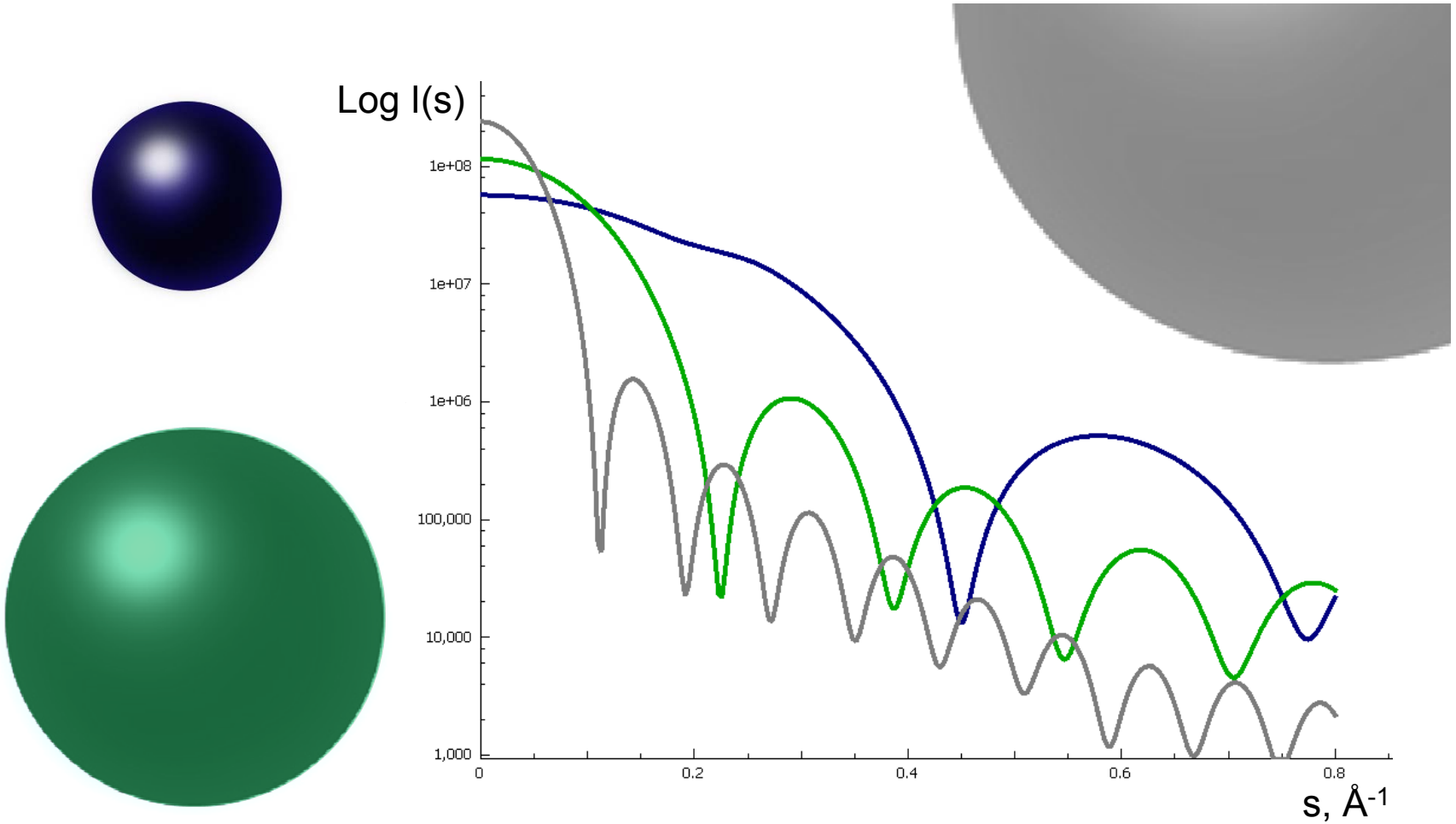
Best to use orthogonal methods (e.g. SEC, AU, maybe native PAGE, mass spectrometry, or best MALS-SEC) to ensure monodispersity

What Kind of Parameters Can We Extract from Scattering Curves?



A. Model independent analysis (directly from the scattering curve)

Sizes



Al Kikhney, BIOSAXS
http://www.embl-hamburg.de/biosaxs/courses/embo2012/

I. Forward Scattering I_0 and Molecular Masses

I_0 ~ (electrons in the particle)^2

I_0 ~ particle concentration

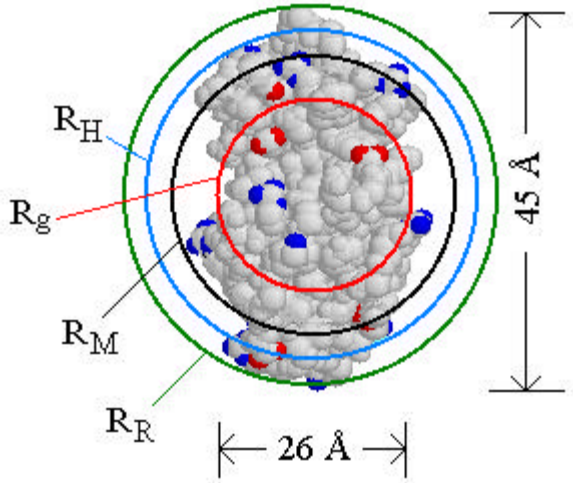
- **If the particle concentration is known, measurements can be calibrated with a known monodisperse protein (e.g. glucose isomerase NOT BSA), yielding the molecular mass of the solute of interest.**
- **An ensemble measurement (monodispersity again!)**

- Calculated by extrapolation (coincident with the direct beam)

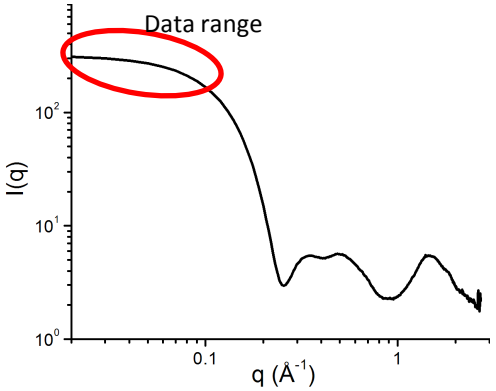
II. Radii of Gyration – the Guinier Plot

- R_g (root-mean-square distance of an object's part from the center of gravity), a function of a particle's mass distribution (size)

$$R_g^2 = \frac{\sum m_i r_i^2}{\sum m_i}$$



http://www.malverninstruments.com/

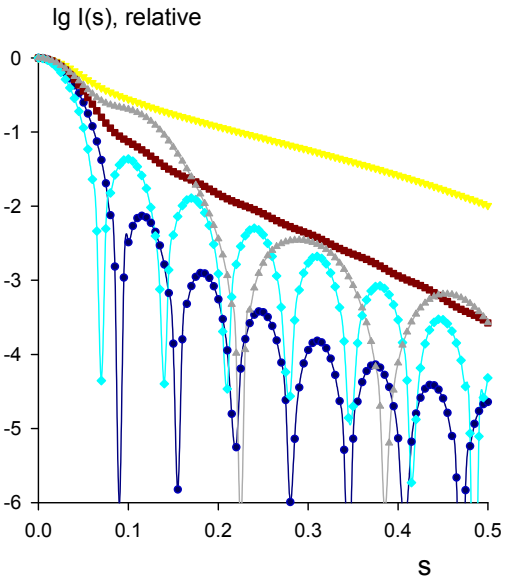


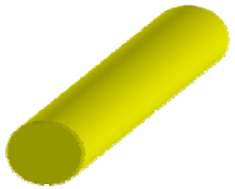
Guinier 1939
Guinier and Fournet 1955

- R_g can be calculated from the slope of the Guinier Plot (lnI versus q^2), but the limits of the Guinier regime is dependent on the type of shape (larger for globular objects, smaller for elongated shapes, $qR_g<0.8$)

Shapes
Shape

Solid sphere





Long rod

Flat disc

Hollow sphere

Dumbbell

Al Kikhney, BIOSAXS
http://www.embl-hamburg.de/biosaxs/courses/embo2012/

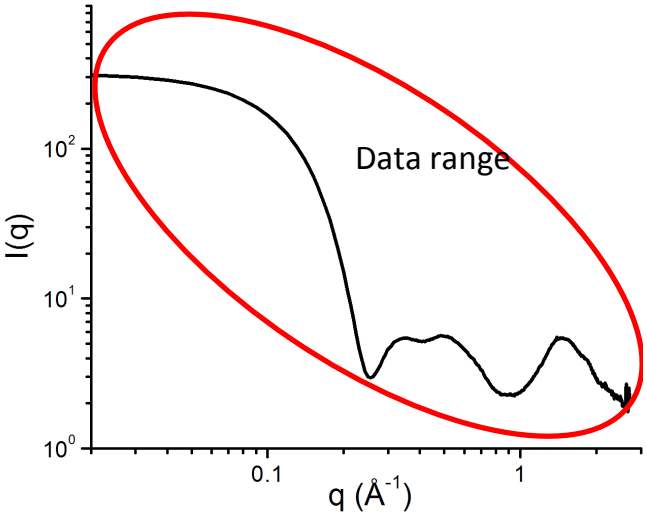
Persistence Length-Folded versus Unfolded

- **Kratky Plot:** $I(s) \cdot s^2$ versus s; generally **bell-shaped when folded**

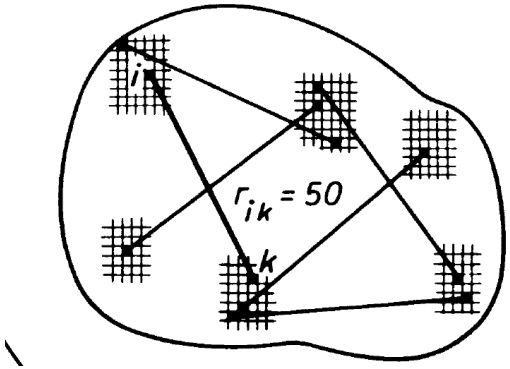
Mertens & Svergun (2010) *J. Struct. Biol.*

The Pair Distribution Function

Atom Pair Distance Histogram



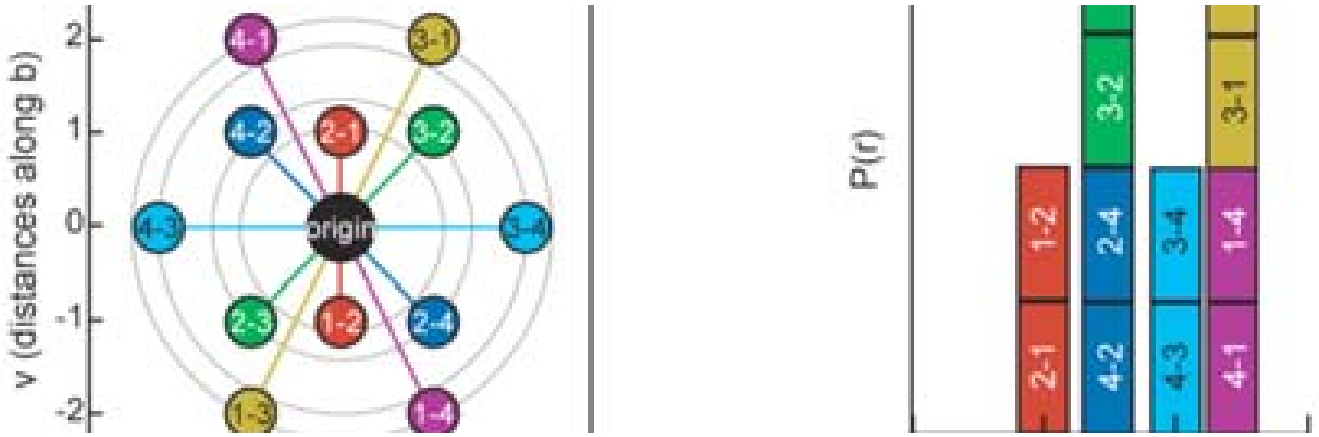
Fourier transform



Reciprocal (Fourier) space

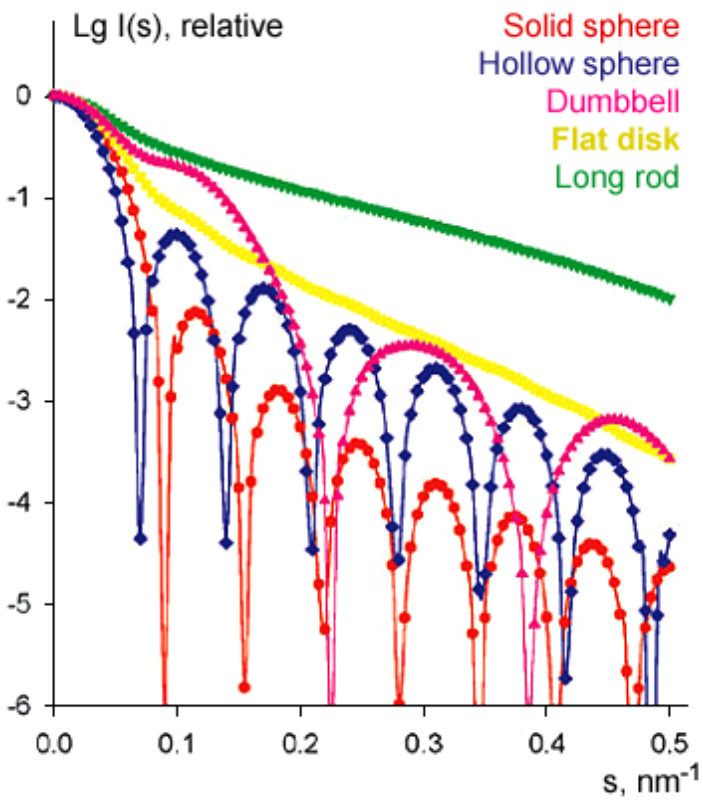
Real space

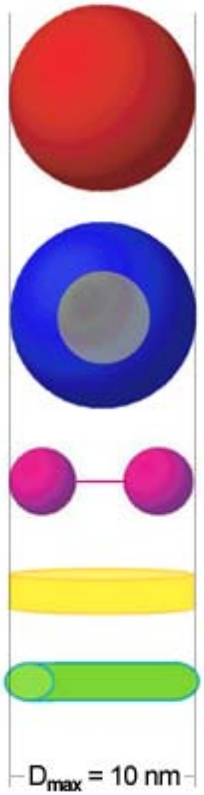
P(r) versus a Patterson Function

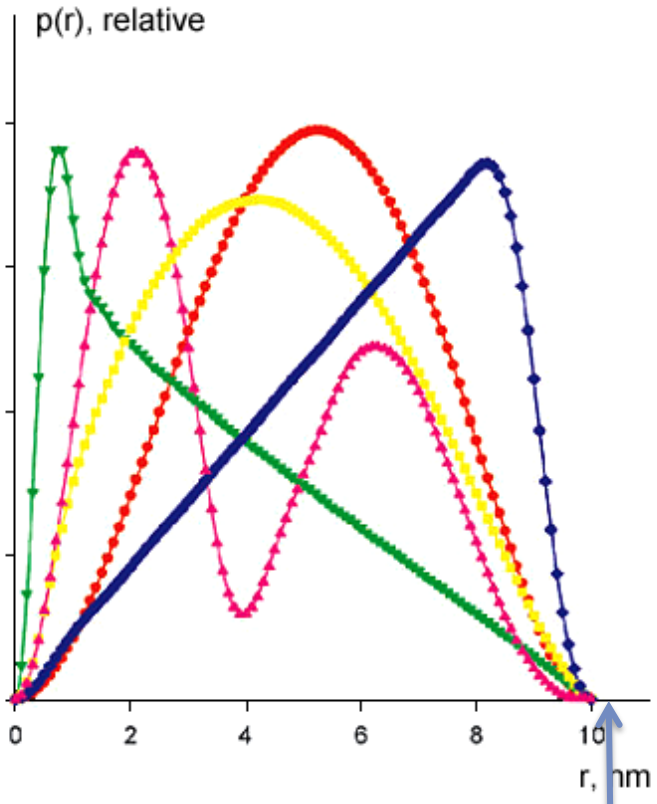


The Pair Distribution Function

(Similar to a "Patterson" Distance)





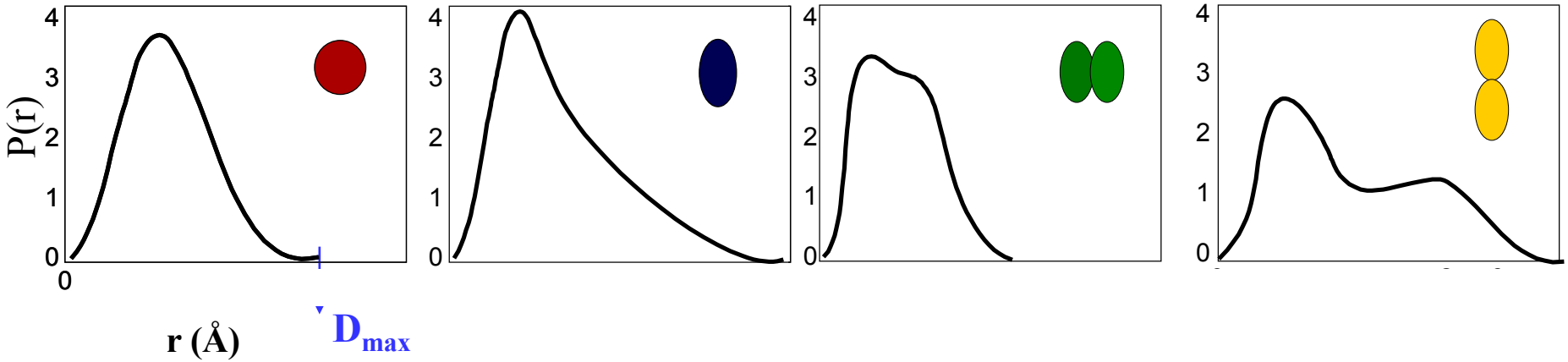


Svergun and Koch (2003)

@Dmax=0

Oligomerization Changes Dmax and p(r)

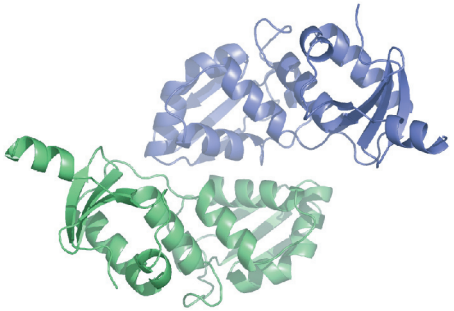
scattering particle.



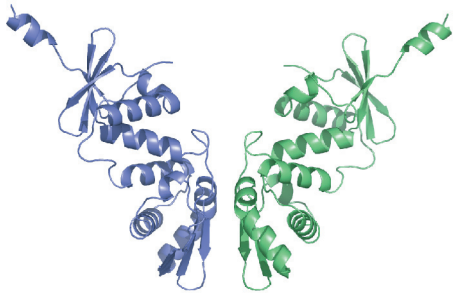
Shape : Modeled as a uniform density distribution that best

neutrons.ornl.gov/.../Small-Angle-Scattering_SAS_V-Urban-ORNL-...

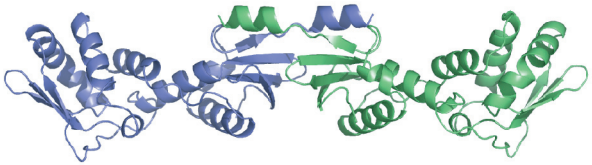
Differentiating between Crystal Packing and Oligomerization in Solution



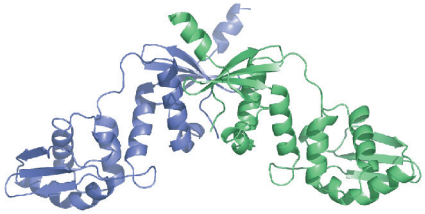
Dimer 1, (B/2 = 755 Å^2)
R_G = 26.08 Å D_{max} = 80 Å



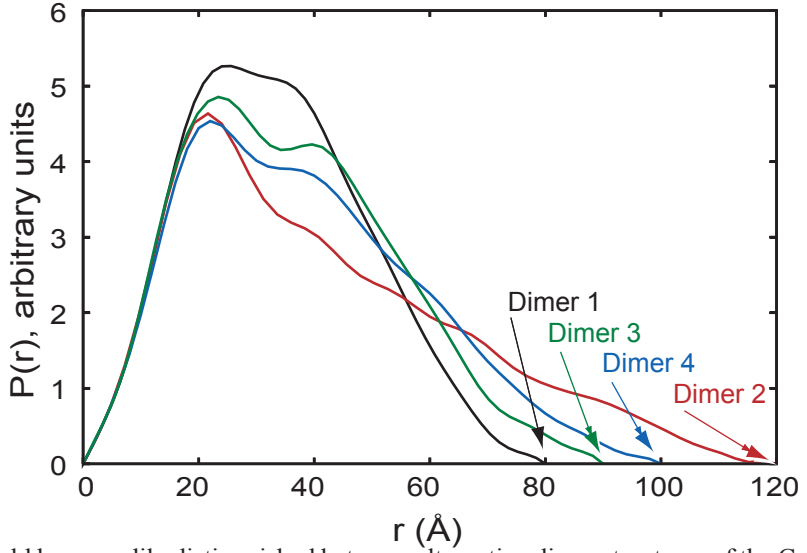
Dimer 3, (B/2 = 406 Å^2)
R_G = 28.3 Å D_{max} = 90 Å



Dimer 2, (B/2 = 923 Å^2)
R_G = 34.04 Å D_{max} = 120 Å



Dimer 4, (B/2 = 255 Å^2)
R_G = 30.4 Å D_{max} = 100 Å



SAXS could have readily distinguished between alternative dimer structures of the C-te

C-terminal domain of DNA repair protein MutL

Putnam et al., Q Rev. Biophysics (2007)

What kind of parameters can we extract from scattering curves?

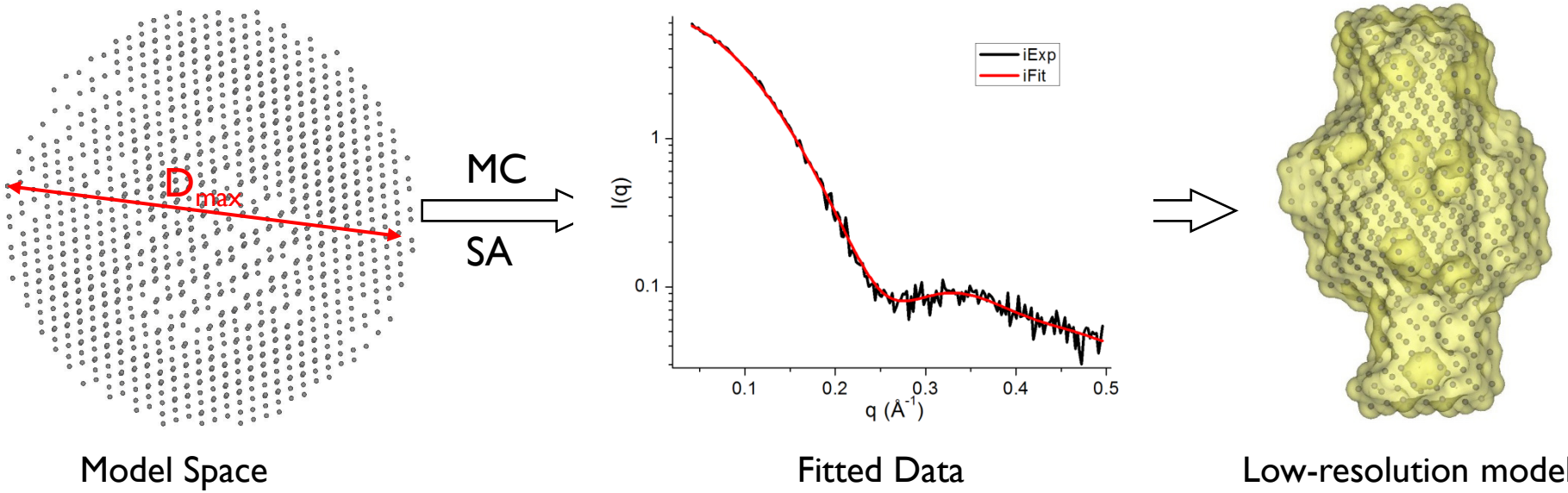
A. Model-dependent analysis (directly from the scattering curve)

B. Model-dependent (3D-reconstruction)

3D Reconstructions by Ab Initio Simulations: How?

- 3D search model -→ trial and error

→ fit against experimental data



Model Space | Fitted Data | Low-resolution model

Dummy Atoms/Residues Assigned to Either Solvent or Model
Simulated Annealing (to find "global" minimum)

3D Reconstructions by Ab Initio Simulations: How?

Constraints:

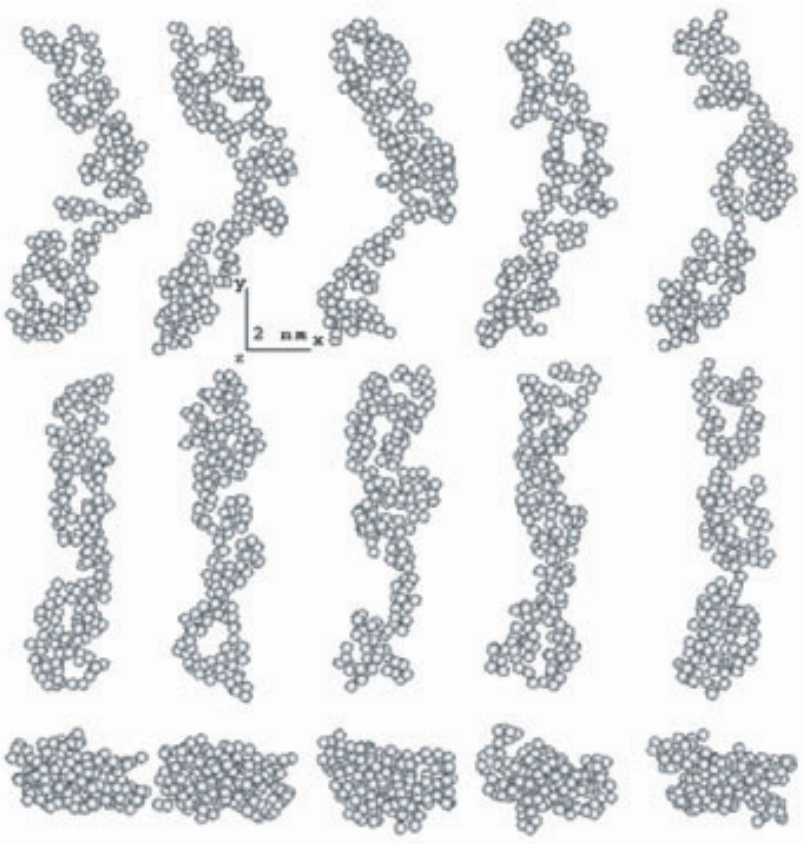
1. Packing and connectivity (3.8Å between scattering centers)
2. Symmetry (if present according to orthogonal method)

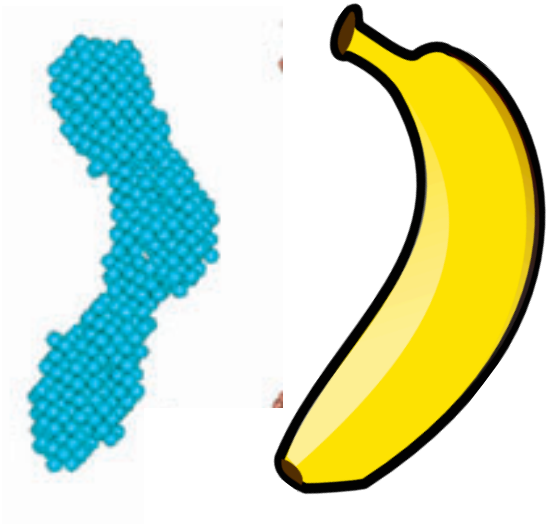
NIH SAXS Workshop
https://ccrod.cancer.gov/confluence/download/.../PartTwo.pdf

Multiple Simulations Need to Be Computed

- Reconstruction depends on initial conditions
- >10 independent simulations per sample
- Align models
- Analyze for convergence (NSD = normalized spatial discrepancy)
- Filter composite volume based on occupancy
- Find common features in your reconstructions

Solutions are Similar but Not Identical

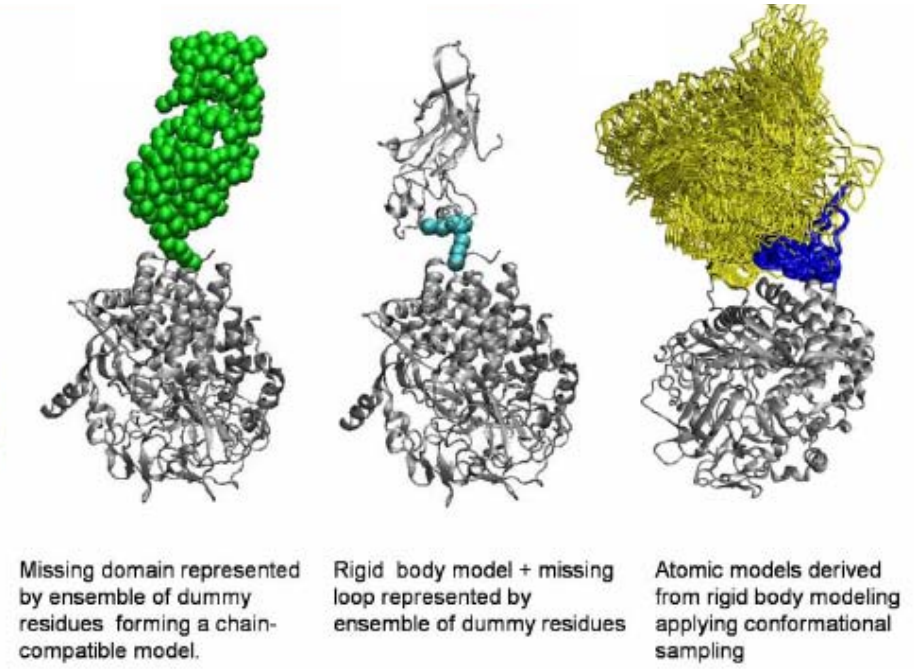




Z-disc domains of Titin (largest known protein, 35000 amino acids)
Svergun & Koch (2003)

So What Else Is It Good For?

- Validate crystal structures
- Help identify buffer conditions likely to produce crystals (non-aggregated protein)
- Locate domains and missing linkers (e.g. not visible in crystal structures)
- Look at dynamics of domains (ensemble of models, EOM and MES)



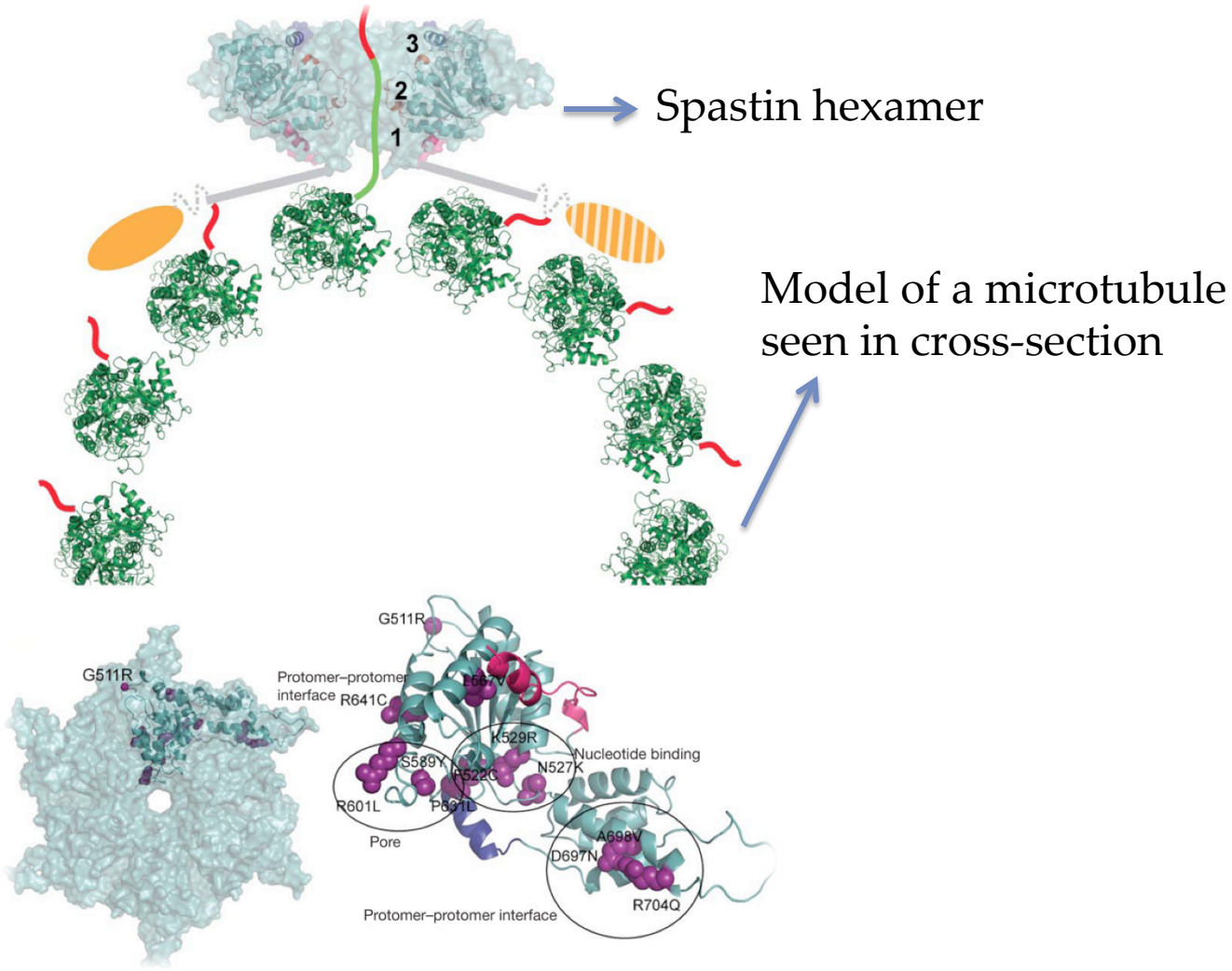
Missing domain represented by ensemble of dummy residues forming a chain-compatible model.

Rigid body model + missing loop represented by ensemble of dummy residues

Atomic models derived from rigid body modeling applying conformational sampling

nn, Grossman …)

Building Larger Assemblies from Known "Pieces"



Roll-Mecak & Vale, *Nature* (2008)

SAXS Is Versatile, Fast and Informative

TOMORROW:

1. Coupling transcription and DNA repair with a dsDNA-tracking motor

2. Post-translational modification of tubulin by tubulin tyrosine ligase (TTL)

Acknowledgements

The Grigorieff Laboratory

Dr. Niko Grigorieff

Dr. Jeff Gelles

Dr. Jané Kondev

Funding

Damon Runyon Cancer Research Foundation

HHMI HOWARD HUGHES MEDICAL INSTITUTE