UNIVERSITY of WISCONSIN

"Crystallography without Crystals"

Determining the Structure of Single Particles & Biological Molecules

Abbas Ourmazd, Dilano Saldin & Valentin Shneerson

Acknowledgments

Len Feldman

Paul Fuoss

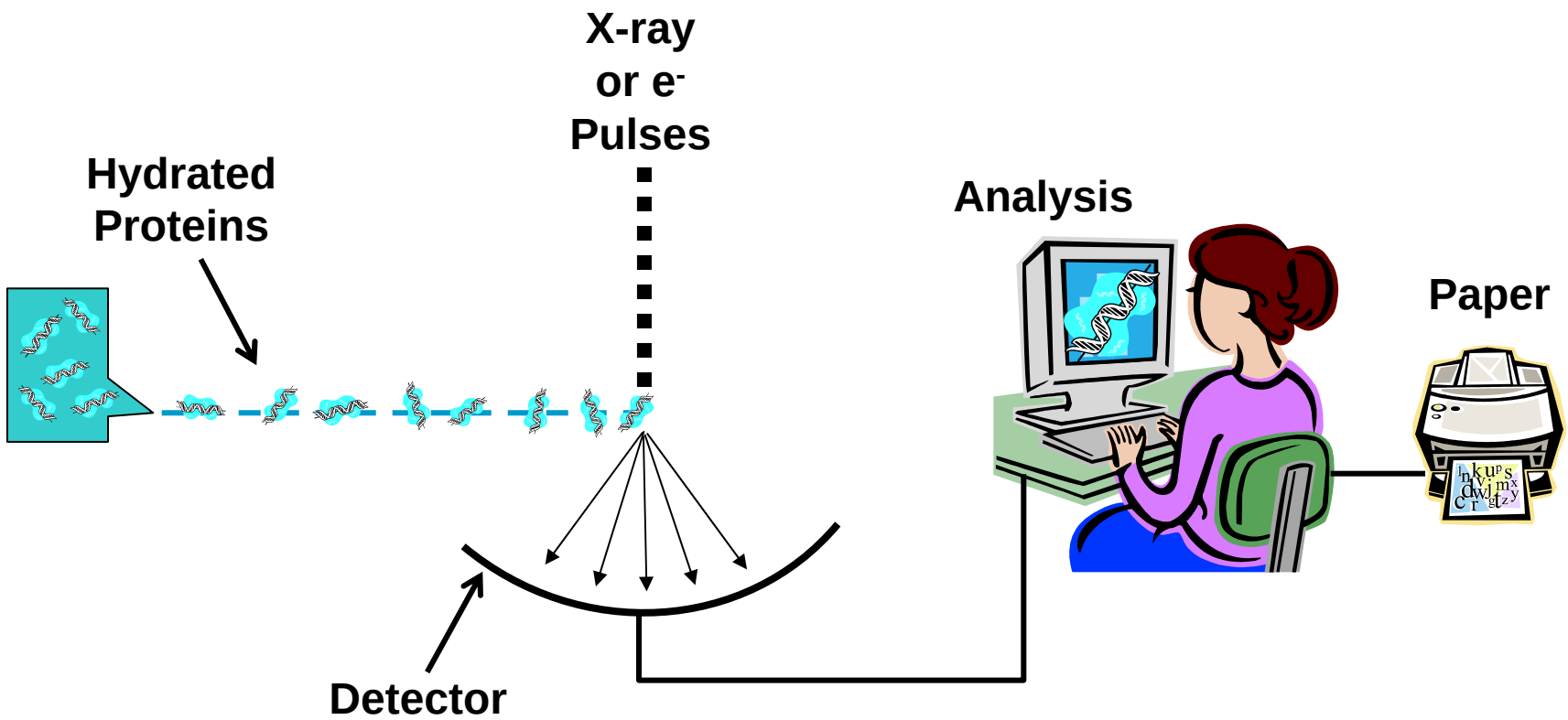
Eric Isaacs

Qun Shen

John Spence

Brian Stephenson

Experiment


X-ray
or e-
Pulses
Hydrated
Proteins
Analysis
Paper
Detector

Determining Structure of Individual Biomolecules

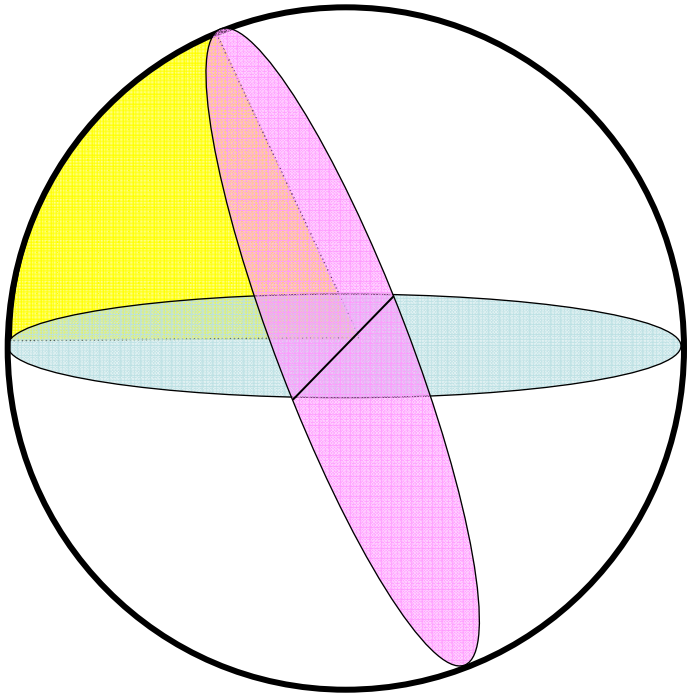
- **Expose beam of molecules to train of X-ray pulses**
  - **Unaligned molecules, "short pulses," many diffraction patterns**
  - **Each diffraction pattern from random, unknown molecular orientation**
- **Determine relative orientations of diffraction "snap-shots"**
  - **Each with very low signal/noise ratio**
- **Compile 3D diffracted intensity distribution**
- **Invert to obtain 3D electron density**

Key Messages

- **Robust algorithm for alignment key, not yet available**
  - **Only demonstration based on “common-line” approach**
  - **Works down to ~10 photons/pixel in presence of shot noise**
  - **Not good enough for atomic resolution experiments at 1Å**
  - **Might be close for lower resolution and/or longer wavelengths**
- **Best experimental regime to be determined**
  - **Are there alternatives to “scatter & destroy”?**
  - **Optics/detectors**
- **Need theory, key experimental parameters & equipment**
  - **Noise-robust, soup-to-nuts algorithm**
  - **Radiation damage threshold for single molecules**
  - **Focusing optics & detectors**

DPs from Random Orientations: Common-Line Approach

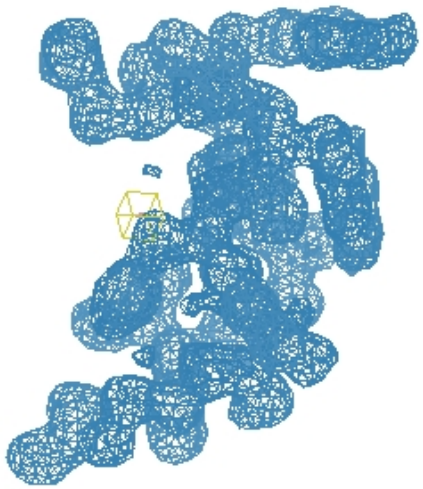
- **Diffraction patterns of same object share a "common line" of diffracted intensity**
  - **"Central Section Theorem"**
- **Three planes fix the relative orientations**
  - **Two with Ewald-sphere curvature**
- **No phase information available**
  - **"Friedel ambiguity"**
  - **Key difference with (cryo-) EM**
- **Developed means for resolving Friedel ambiguity**
  - **"Handedness" ambiguity remains**
  - **As in cryo-EM**



Electron Density Recovery

Model of protein 1UAO
(From atom coordinates in PDB)

Recovered Solution
(From DPs of random orientations)



- **Collection semi-angle ~ 32º; 12.4 keV photons; ~ 1 Å resolution**
- **Low-angle data excluded**
- **Correlation coefficient ~ 0.8**
- **Shneerson, Ourmazd & Saldin, Acta Cryst, submitted**

Noise

- **Can align dp's and recover structure in absence of noise**
  - **RMS alignment accuracy < 0.5°**
- **Anticipate severe noise (shot noise, background, detector)**
  - **Modeled as Poisson statistics for present**
- **CL method works with ≤ 10 photons/pixel + shot noise**
  - **Excluding large central region**
  - **Might be pushed to ≤ 1 photon/pixel (?)**
  - **Not close to photon counts anticipated for atomic resolution (at λ=1Å)**
- **Longer wavelengths would help substantially**
- **Need more sophisticated alignment algorithms**
  - **From <u>diffraction pattern alignment</u> to <u>photon assignment</u>**

Recovered vs. Actual Orientations

Sample of 20 Diffraction Patterns

	θ			ϕ			ψ		
Planes	*Recovered*	*Actual*	*Δ*	*Recovered*	*Actual*	*Δ*	*Recovered*	*Actual*	*Δ*
(0,1)	64	64.3	-0.3	144.3	144.6	-0.3	48	48.3	-0.3
(0,2)	19	20.8	-1.8	20.2	20.6	-0.4	180	177.6	2.4
(0,3)	49	48.7	0.3	82.9	81.8	1.1	90	90.4	-0.4
(0,4)	141	140.5	0.5	43.5	43.5	0.0	122	122.4	-0.4
(0,5)	16	13.9	2.1	16.7	16	0.7	108	110.3	-2.3
(0,6)	52	52.3	-0.3	15.7	15.6	0.1	61	60.4	0.6
(0,7)	174	174.2	-0.2	138.5	138.2	0.3	99	99.6	-0.6
(0,8)	90	90	0.0	89.6	87.9	1.7	170	169.7	0.3
(0,9)	86	85.2	0.8	177.3	177.8	-0.5	76	75.5	0.5
(0,10)	166	166.5	-0.5	33	33.7	-0.7	97	96.3	0.7
(0,11)	21	21	0.0	49.6	49.5	0.1	179	179.1	-0.1
(0,12)	58	58	0.0	41.9	42.7	-0.8	19	19.3	-0.3
		RMS Δ	**0.25**		**RMS Δ**	**0.21**		**RMS Δ**	**0.30**

Single-Molecule X-ray Scattering: Orders of Magnitude

- **Assumptions:**
  a. **Macromolecule with N atoms scatters as N carbon atoms**
  b. **Need ~ 1 pixel per atom spacing, i.e., atomic resolution**
  c. **Need 10^3 scattered photons per pixel**
  d. **Scattered amplitude: low-angle ~ N^2; high-angle ~ N**
  e. **Use 0.1nm radiation (12.4 keV)**
  f. **Take 500 kDa molecule**
     - **Yeast proteins: ~ 50kDa**
     - **Largest known proteins (titins) ~ 3000 kDa**

- **Consider current & next-generation X-ray sources**
  - **Current (3rd) generation running routinely (APS)**
  - **Next (4th) generation due to "start" in 2008 (XFEL)**

Single-Molecule Scattering: Orders of Magnitude

X-ray Beam		Flux per mm2	Counts per pulse per pixel		No. of Pulses for 1E3 photons/pixel		Time (sec) for 1E3 photons/pixel	
Source	Ø (μm)	per pulse	Small <	Large <	Small <	Large <	Small <	Large <
XFEL	0.1	3.E+20	2985	7.E-03	0.3	1.E+05	0.003	1.E+03
APS	0.01	1.E+18	11	3.E-05	87	4.E+07	0.9	4.E+05

XFEL could give diffraction pattern from single protein
Need ~ 10^5 shots; ~ 1 hour

Single-Molecule Scattering: Orders of Magnitude

X-ray Beam		Flux per mm2	Counts per pulse per pixel		No. of Pulses for 1E3 photons/pixel		Time (sec) for 1E3 photons/pixel	
Source	Ø (μm)	per pulse	Small <	Large <	Small <	Large <	Small <	Large <
XFEL	0.1	3.E+20	2985	7.E-03	0.3	1.E+05	0.003	1.E+03
APS	0.01	1.E+18	11	3.E-05	87	4.E+07	0.9	4.E+05

APS could give diffraction pattern from single protein
Need ~ 10^7 shots; ~ 100 hrs

Assumes every photon can be focused

Single-Molecule Scattering

- ~ 10^9 photons can be scattered in "reasonable time"
  - ~ hour on LCLS, days with APS
- Plenty to form high-quality diffraction pattern (~10^3 ph/pixel)
  - Also to deduce structure (Glaser "dose-fractionation")
- Problem: Do not know which photon belongs where
  - Each diffraction pattern from unknown random orientation of molecule
  - No technique demonstrated to align dp's with expected signal/noise
- "Diffraction pattern alignment" vs. "photon assignment"
  - Consider entire set of diffracted photons

DP Alignment vs Photon Assignment

- **CL method compares lines of diffracted intensity**
  - **"Line by line" comparison of diffraction pattern pairs/triplets**
  - **Poor utilization of available data**

- ***Improved strategy 1:* Natural coordinate system**
  - **Symmetry-based restrictions on number of possibilities**
  - **Increases "line" to "shell" or even entire diffraction pattern**

- ***Improved strategy 2:* Signal processing approaches**
  - **Hidden Markov models**
  - **Considers probabilities for entire data set (a la Viterbi)**

- ***Improved strategy 3:* Latent variable approaches**
  - **Considers internal correlations within entire data set**

Radiation Damage

- **Molecule explodes after single XFEL shot**
  - **Must collect data in first 20fs after pulse**
  - **"Molecule exploded, but not flown apart"**

- **Dose limit deduced from crystals**
  - **~ 10^{16} photons per mm^2**

- **Single-molecule damage threshold might be higher**
  - **Simulations indicate threshold significantly higher; reliable?**
  - **Energetic photoelectrons escape molecule, damage crystal**
  - **Single-molecule damage threshold might be *lower*!**

- **If higher, "sweet intensity spot" is ~ 10^{18} photons/mm^2/pulse**
  - **Approached by 3rd generation synchrotron with focusing optics**

Explode or Not Explode?

- Which operational regime?
  - "Scatter & destroy" (and collect within 20fs)
  - Not explode (and "expose" for longer)

- "Exposure-time" limited by molecular rotation: 10ps-10ns
  - Small (~10kDa, 3nm Ø) protein rotates ~2º in 10ps
  - Medium-size (3,000kDa, 30nm Ø) protein rotates ~2º in 2ns

- Regime within specifications of next-generation sources
  - Accessible w/ current sources w/ advanced optics?

- For atomic resolution, must integrate ~ 10^5 - 10^7 shots
  - To obtain reasonable signal/noise ratio
  - Must align to integrate, need to integrate to align

Conclusions

- **Single-particle/molecule crystallography "tantalizingly close"**
  - **Algorithms, molecular & X-ray sources under development**

- **Critical issues remain**
  - **Robust algorithms able to deal with noise**
  - **Radiation damage threshold, suitable operating regime**
  - **Reliable source of hydrated molecules**
  - **X-ray optics & detectors**

- **Success would have significant & broad impact**
  - **Access to all macromolecules "pure or not"**
  - **Possibility to investigate different molecular conformations**
  - **Implications for physics, materials, biochemistry, drug design**