Highly brilliant and coherent XFEL beams for biological macromolecules

Arwen Pearson

Okay...ready? One, two, three,... THROW!

Time resolution in its infancy

By Betsy Streeter
What is the dream experiment?

• To observe a functioning system...
  • in real time
    • from fs to minutes
  • with high spatial resolution (Å)
    • but still the ability to see the whole thing!
  • and elemental specificity
  • and in situ
• There are many available tools, but they probe different time and length scales “windows” as well as different states (crystals, liquids, powders, organelles, cells...)

“We have to remember that what we observe is not nature in itself, but nature exposed to our method of questioning.”

Werner Heisenberg

Charles Maurice Stebbins & Mary H. Coolidge, Golden Treasury Readers: Primer
A well ordered ensemble

James Holton
“That’s no good. Crystals don’t wriggle and if it doesn’t wriggle, it’s not biology”

Commentary from Hill on Kendrew’s plans to study proteins in the crystalline form
A more realistic view of a biological ensemble
- Macromolecules are dynamic, flexible objects
- Any ensemble measurement sees all conformations at once
- The resulting ensemble structure is an average (over both space and time)

Dynamic information is lost and structural resolution is reduced

Subdividing the ensemble can reveal more detail, but at a cost of reduced signal

Can also be challenging to order the resulting structures along the reaction coordinate
Where do the photons go?

Protein
1A x-rays

- Elastic scattering (6%)
- Transmitted (98%)
- Inelastic scattering (7%)
  - Re-emitted (99%)
  - Absorbed (~0%)
- Photoelectric (87%)
  - Re-emitted (~0%)
  - Absorbed (99%)

James Holton
Increasing signal?

Dependent on both the source properties & the sample!

\[ I_{hkl} = I^0 (\lambda^3/\omega) (V_xLpA/V^2)|F_{hkl}|^2 \]

- Intensity of the incident beam
- Volume of the unit cell
- Volume of the crystal
- Scattering power of the sample
- Volume of the crystal
Increasing signal?

Dependent on both the source properties & the sample!

\[ I_{hkl} = I^0(\frac{\lambda^3}{\omega})(\frac{V_x L p A}{V^2})|F_{hkl}|^2 \]

- Intensity of the incident beam
- Volume of the unit cell
- Volume of the crystal
- Scattering power of the sample
- Properties of the sample that can’t be easily changed
Making and detecting X-rays

• All lab-based X-ray generators are fundamentally the same

• Use a *cathode* to generate a stream of electrons that impact a target metal *anode* to generate X-ray photons
Making and detecting X-rays

• All lab-based X-ray generators are fundamentally the same.

• Use a *cathode* to generate a stream of electrons that impact a target metal *anode* to generate X-ray photons.

• Data collection with early sealed tubes would take weeks.
Making and detecting X-rays

- Rotating anode generators work in the same way - but the anode is constantly turning

- Anode must be water cooled to carry away the excess heat

- For modern rotating anodes data collection takes hours
Synchrotrons

- Synchrotrons are particle accelerators that are able to deliver incredibly bright beams of light
Rotating anode
10 min exp.

Same crystal,
undulator, single
pulse of
100 ps exp.
Lysozyme 100 µs Exposure Time on P14 @ Petra III
XFELs deliver a huge increase in brightness
Levantino et al., 2015, Nat Comms
What sort of time-scales are we interested in for biology?

- **Allosteric transitions**
- **Helix/coil transitions**
- **Loop/hinge dynamics**
- **Enzyme catalysis (slowest steps)**
- **Water Structure reorganisation**
- **Side-chain rotations (surface)**
- **Chemistry**

**Time-scales**

- s (seconds)
- ms (milliseconds)
- µs (microseconds)
- ns (nanoseconds)
- ps (picoseconds)
- fs (femtoseconds)

**Methods**

- **Spectroscopy (electronic, vibrational, neutron, X-ray...)**
- **X-ray scattering/diffraction at synchrotrons**
- **X-ray scattering/diffraction at XFELs**
- **Magnetic Resonance (NMR & EPR)**
- **Single particle Cryo-EM**
• How can we access biochemical events on these different length scales?

  • Option 1: “Stop motion”

  • Need a way to arrest the reaction at a certain point
  • Need to be aware that off-pathway states can form
• Trapping methods are well established and have been used in structural enzymology since the 1960’s

• For slow reactions (> ms) can try cryo-trapping - plunge cool in liquid nitrogen

• Small drops in temperature can also be used to reduce reaction rates and bring specific intermediates within reach of cryo-trapping

• Mechanistic trapping can also be used, regardless of the rate of individual reaction steps
  • Alter reaction conditions to prevent full turnover
  • Use mutants to prevent full turnover
  • Use altered substrates to prevent full turnover
  • Drive the system into steady state
LCLS generating $\sim 10^{12}$ photons per pulse @ 9 keV

Electron bunch ($\sim 10^9$ electrons)

Accelerating modules ($\sim$ 1 km)

SASE undulator

$\sim 30$ fs

XPP beamline, 400 m from undulator

Si(111) monochromator

$\sim 10^{10}$ ph

Refractive lenses

Closed-loop circulation

Levantino et al., 2015, Nat Comms
• What about the sample size?

\[ I_{hkl} = I^0 \left( \frac{\lambda^3}{\omega} \right) \left( \frac{V_x L p A}{V^2} \right) |F_{hkl}|^2 \]

• On the face of it, it would seem that the bigger the crystal the better.

• But it is not so simple
Challenges for the pump-probe crystallographic experiment

**Limitations/drawbacks**
- Typically very short exposure times used so low signal to noise especially if using monochromatic beam
- Only one data point per cycle. So either...
  - a fully reversible reaction is needed, or
  - Lots of samples are required
- Also have the problem that in an XFEL experiment the sample is destroyed
- Very difficult for a non-reversible system
Serial experiments address the challenge of sample destruction and reaction irreversibility

- In a serial experiment each “shot” is taken from a new sample
  - Many flavours
  - “mesh and collect”
  - helical/grid scans
  - serial synchrotron crystallography (SSX)
  - serial femtosecond crystallography (SFX)
- Brings the new challenge of how to deliver the sample?
  - ideally sample should be delivered fast enough to
    - make best use of the available X-rays
    - allow the experiment to be done in a reasonable time
  - also puts a first practical limit on the sample size we can use
    - simply due to sample availability
• We can divide sample delivery methods into two classes
  • solid targets
  • jets
    • all jet experiments add some background to the diffraction pattern
• We can divide sample delivery methods into two classes
  • solid or fixed targets
  • jets

Oberthuer, Dominik http://dx.doi.org/10.1038/srep44628

• Sample delivery can be very fast, but is stochastic
• can use a LOT of sample
• need a way to stop the crystals settling
Viscous Jets

- First demonstrated with LCP
- Can also use “grease” and other polymers
- Sample delivery is slow - matches well to the rep rate of the LCLS and SACLA
- Also works well at synchrotrons
- Vital to test compatibility of media with YOUR sample

Uwe Weierstall Nature Comms (2014) doi:10.1038/ncomms4309

Kovascova et al., IUCrJ, 2017
Fixed/Solid Targets


- Samples can be presented randomly or in a defined array
  - if defined can achieve near 100% hit rates
  - useful for cases where sample is limited
  - Background can be minimised
Fixed Targets

Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography.
• [protein] in crystals ≈ [protein] in the cell

• many proteins retain catalytic activity in the crystal

• if there are no large conformational changes during catalysis, many proteins remain crystalline during turnover

David Goodsell, The machinery of life
• To really understand mechanism we need to be able to image the system “in action”

• “Movie-mode”

• Need a way to start the reaction at the same time for all molecules in the sample & to image faster than the reaction is occurring
• Sample delivery method and sample availability already put practical limits on sample size

• Additional constraints arise when we consider a time-resolved experiment that are associated with reaction initiation

• There are two basic ways to initiate a reaction
  • Mixing
  • Photoactivation

• These are associated with two concepts important for defining sample size
  • critical depth
  • laser penetration
• Critical depth
  • This defines the maximum distance a ligand has to diffuse for the diffusion rate to be faster than the process you’re interested in

• Depending on the reaction rate of the species you are looking at AND the buffer conditions this can be extremely variable

• There are a couple of cases to consider
  • diffusion and catalysis occur with similar rates
    • for a simple reaction
      \[
      E + S \overset{k_1}{\underset{k_-1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} E + P
      \]
    • we can estimate the critical depth as
      \[
      \lambda_c = (DK_M/k_2[E])^{1/2}
      \]

• Critical depth
  • This defines the maximum distance a ligand has to diffuse for the diffusion rate to be faster than the process you’re interested in
  • Depending on the reaction rate of the species you are looking at AND the buffer conditions this can be extremely variable

• There are a couple of cases to consider
  • diffusion and ligand binding occur with similar rates
    \[ E + S \overset{k_1}{\underset{k^{-1}}{\rightleftharpoons}} ES \]
  • we can estimate the critical depth as
    \[ \lambda_c = (D/k_1[E])^{1/2} \]

• Critical depth
  • This defines the maximum distance a ligand can diffuse for the diffusion rate to still be faster than the process you’re interested in

• Depending on the reaction rate of the species you are looking at AND the buffer conditions this can be extremely variable

• For extremely efficient enzymes the critical depth can be as short as a single unit cell!

• However for reaction steps with time-scales on the order of ms the critical depth is a few µm (assuming the crystallisation buffer is not too viscous)

• This means that if you are initiating a reaction using mixing and want to look at ms or shorter timescales you need a correspondingly small crystal (< 10-20 µm thickness)
Table 2 Comparison of reactivities of enzymes and proteins in crystal and solution states

<table>
<thead>
<tr>
<th>Enzyme or protein</th>
<th>Substrate or ligand</th>
<th>Relative activity</th>
<th>Concentration of enzyme in crystal (moles/liter)</th>
<th>$\lambda_c$ (µm)</th>
<th>$\lambda_{Experimental}$ (µm)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Chymotrypsin</td>
<td>acetyl-L-tyrosine hydrazide</td>
<td>1:0.18</td>
<td>0.0311</td>
<td>6.6</td>
<td>crystals of up to 0.2-0.8 mm dimensions used, thickness not stated</td>
<td>Kallos, BBA, 1964</td>
</tr>
<tr>
<td>Carboxypeptidase $A_\alpha$</td>
<td>carbobenzoxyglycyl-L-phenylalanine</td>
<td>1:0.33</td>
<td>0.0450</td>
<td>6.0</td>
<td>15-300</td>
<td>Quiocho &amp; Richards, Biochemistry, 1966</td>
</tr>
<tr>
<td>Carboxypeptidase $A_\gamma$</td>
<td>carbobenzoxyglycyl-L-phenylalanine</td>
<td>1:0.003</td>
<td>0.0450</td>
<td>6.0</td>
<td>not stated</td>
<td>Lipscomb, PNAS, 1973</td>
</tr>
<tr>
<td>Elastase</td>
<td>N-benzoyl-L-alanine methyl ester</td>
<td>1:0.46</td>
<td>0.0294</td>
<td>0.83</td>
<td>1.0</td>
<td>Shotton et al. Cold Spring Harbor Symp. Quant. Bio., 1971</td>
</tr>
<tr>
<td>Papain</td>
<td>acetylglycine ethyl ester</td>
<td>1:1.34</td>
<td>0.0279</td>
<td>8.0</td>
<td>1-6</td>
<td>Sluyterman &amp; Graaf, BBA, 1969</td>
</tr>
<tr>
<td>Phosphorylase b</td>
<td>maltotopease</td>
<td>1:0.09</td>
<td>0.0068</td>
<td>1.0</td>
<td>2.0</td>
<td>Kasvinsky &amp; Madsen, JBC 1976</td>
</tr>
<tr>
<td>Ribonuclease S</td>
<td>uridine-2', 3'-phosphate</td>
<td>1:0.09</td>
<td>0.0068</td>
<td>1.0</td>
<td>2.0</td>
<td>Doscher &amp; Richards, JBC, 1963</td>
</tr>
<tr>
<td>Ribonuclease A</td>
<td>cytidine-2', 3'-phosphate</td>
<td>1:1-1:10</td>
<td>0.0056</td>
<td>1.3</td>
<td>0.02</td>
<td>Bello &amp; Nowoswiat, BBA, 1965</td>
</tr>
<tr>
<td>Liver alcohol dehydrogenase</td>
<td>NADH</td>
<td>1:0.001</td>
<td>0.0080</td>
<td>0.16</td>
<td>&gt;1</td>
<td>Theorell et al., JMB, 1966</td>
</tr>
<tr>
<td>Sperm whale metmyoglobin</td>
<td>azide anion</td>
<td>1:0.048</td>
<td>0.0493</td>
<td>0.90</td>
<td>2-9</td>
<td>Chance et al., JMB, 1966</td>
</tr>
<tr>
<td>Horse methemoglobin</td>
<td>azide anion</td>
<td>1:0.4-0.045</td>
<td>0.0382</td>
<td>0.27</td>
<td>1</td>
<td>Chance et al., JMB, 1966</td>
</tr>
<tr>
<td>Fast reacting deoxyhemoglobin (horse Hb$^*$)</td>
<td>CO</td>
<td>1:1</td>
<td>0.0382</td>
<td>0.11</td>
<td>&lt;1, (but CO already in crystal after flash photolysis)</td>
<td>Parkhurst &amp; Gibson, 1967</td>
</tr>
</tbody>
</table>
Martin Trebbin & Diana Monteiro
(printed with support from PSCM ESRF)
• Laser penetration
  • This is related to the absorbance of your system at the wavelength you are exciting
  • Usefully this can be easily measured
    • Note you don’t need to excite at the maximum absorbance of the sample
    • Exciting off the maximum can increase your laser penetration
  • As a rule of thumb, for most samples ~ 10 µm should be OK
• Fast reactions (< ms) we need to trigger by light
  • T-jump via IR pulse (ns)
  • Photoisomerisation/direct photocleavage (few fs)

"Photoactive/Photoreceptor"
■ All other proteins

Results obtained by searching manually curated SwissProt entries for keywords “photoactive” or “photoreceptor.”

By approximation, only 0.25% of proteins are photoactive!!

Mike Thompson, UCSF
• Fast reactions (< ms) we need to trigger by light
  • T-jump via IR pulse (ns)
  • Photoisomerisation/direct photocleavage (few fs)
  • Photocaging (ns-ms)

[Diagram of photocaged inactive system and active system after photolysis]

Photocaged Inactive System

Photolysis

Active System

Photolysis by-product
What makes a good photocage?

(1) Clean and efficient photochemistry
(2) Good quantum yield
(3) Adequate absorption at wavelengths longer than 300 nm
(4) Good aqueous solubility
(5) The decaging rate must be much faster than the process of interest
(6) Synthetically tractable

What photocage scaffolds are available?

ortho-Nitrobenzyl
- Slowest (10 - 10^4 s\(^{-1}\))
- Short \(\lambda_{\text{max}}\) (254 - 320 nm)
- Solubility variable

Coumarinyl
- Faster (10^8 - 10^9 s\(^{-1}\))
- Longer \(\lambda_{\text{max}}\) (320 – 390 nm)
- Solubility mostly poor

para-hydroxyphenyl
- Faster (10^8 - 10^9 s\(^{-1}\))
- Short \(\lambda_{\text{max}}\) (280-304 nm)
- Solubility mostly good
Modifying photocage properties

- By modifying the substituents on the photocage moiety the photochemical properties can also be altered.
- But an improvement in one aspect is often offset by something else getting worse.

<table>
<thead>
<tr>
<th>λmax (nm)</th>
<th>ε (M⁻¹cm⁻¹)</th>
<th>φ</th>
<th>k (s⁻¹)</th>
<th>solubility (H₂O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>ca. 27000</td>
<td>0.1-0.2</td>
<td>10 - 200</td>
<td>Poor</td>
</tr>
<tr>
<td>254</td>
<td>ca. 27000</td>
<td>0.1-0.64</td>
<td>10 - 1000</td>
<td>Poor</td>
</tr>
<tr>
<td>262</td>
<td>ca. 5000</td>
<td>0.04-0.14</td>
<td>9×10³ - 3×10⁴</td>
<td>Good</td>
</tr>
<tr>
<td>345</td>
<td>ca. 6000</td>
<td>0.01</td>
<td>N/A</td>
<td>Poor</td>
</tr>
</tbody>
</table>
A current example

- $\text{o-nitrobenzyl with a methylenedioxy substituent}$
- attached to the alpha carboxylate of L-aspartate
- shows improved aqueous solubility compared to parent compound
- 49% yield over 7 steps
- shows increased extinction coefficient at longer wavelengths
A current example

- $o$-nitrobenzyl with a methylenedioxy substituent
- attached to the alpha carboxylate of L-aspartate
- shows improved aqueous solubility compared to parent compound
- 49% yield over 7 steps
- shows increased extinction coefficient at longer wavelengths

```
\[
\begin{align*}
\text{I} & \rightarrow \text{II} \rightarrow \text{III} \leftarrow \text{III}^+ \\
\text{II} & \rightarrow \text{IV} \rightarrow \text{V} \rightarrow \text{VI}
\end{align*}
\]
```
A current example

- o-nitrobenzyl with a methylenedioxy substituent
- attached to the alpha carboxylate of L-aspartate
- shows improved aqueous solubility compared to parent compound
- 49% yield over 7 steps
- shows increased extinction coefficient at longer wavelengths

Decay of the 1st intermediate

- $t_1 = 1 \, \mu s$, $t_2 = 10 \, \mu s$
- $t_1 = 0.21 \, \mu s$
An caging strategy that can be generalised to any system?

- protein crosslinking
A current example

- 3rd generation photocleavable crosslinker
- 30% yield over 4 steps
- good aqueous stability
- linked to protein via direct cysteine modification
- cleavage leaves a methyl carboxylate moiety
- extensible linker
- first photolysis tests using a mercury lamp show complete release of the photocage
Future Challenges
The coming Datapocalypse

- LCLS II will increase data throughout by three orders of magnitude by 2025
- A 1PB/day data firehose
The coming Datapocalypse

• Data rates and volumes will become untenable - need to be making decisions about what data we keep and what we don’t
• How do we avoid biasing our results by selective discarding?
• How do we know what to throw away?
  • even unmerged data cannot be trusted to represent the best that can be extracted from raw images
• All data collected have to be used. If the data have a change (resulting in bad merges) we should be modelling the change rather than throwing the data away
The coming Datapocalypse

- A few groups with too much data and lots of groups with no data
- Users aren’t interested in their own data, let alone anyone else’s!
The coming Datapocalypse

Problems facing humanity

- Storing serial data
- Everything else
Are we just fiddling while Rome burns?

- Ideally we should be doing better experiments, that don’t need so much post processing to fix all the problems (that can be fixed)
- Challenge of education and of providing sensible interfaces and real-time feedback to users