

Paper for the Workshop on Evolution and Control of Complexity, Argonne, Oct 11-13, 2010

Use of correlated x-ray scattering (CXS) to track protein motions during enzyme catalysis by means of an x-ray free-electron laser.

Sebastian Doniach¹, Vijay Pande², John Spence³, Daniel Herschlag⁴, Sébastien Boutet⁵ and Jerome Hastings⁵

Departments of Applied Physics¹ and Physics¹, Chemistry², and Biochemistry⁴, Stanford University, Department of Physics, Arizona State University³, and Linear Coherent Light Source⁵, SLAC national Accelerator Lab.

Enzyme-driven catalysis controls most metabolic pathways in all living organisms, yet is poorly understood. We aim to develop an entirely new structural technique to measure enzyme-driven catalysis based on the world's first Xray free electron laser (xFEL), which will track enzyme conformational changes with unprecedented structural resolution for non-crystalline samples.

The xFEL, which has been developed at the SLAC National Accelerator lab as the Linac Coherent Light Source, or LCLS delivers pulse of x-rays which last a few tens of femtoseconds (10^{-15} seconds) and deliver about 2mJ of energy in a single pulse. The laser has already shown its ability to obtain terabytes of structural information on nano crystals of biomolecules. Our plan is to do measurements using Correlated X-ray Scattering (CXS) on droplets of solution of biomolecules in non-crystalline form in which enzyme molecules can go through their catalytic cycles under close to physiological conditions.

The measurement of correlated scattering events in which a given molecule in the ensemble can scatter 2 photons within the same x-ray pulse is made possible by the combination of x-ray flux in an FEL pulse and the recent advances in control of droplet formation and transport into an xFEL beam. The incidence of such CXS events is greatly amplified by the number of macromolecules in the droplet. Then by correlating scattering events at different pixels in an area detector, defining a pair of scattering vectors, q and q' , these 2-photon events define a 4-point correlation function C_4 which may be extracted by evaluating

$$C_4(q, q', \cos \theta) = \langle I(\vec{q})I(\vec{q}') \rangle - \langle I(q) \rangle \langle I(q') \rangle \quad (1)$$

where $I(q)$ measures the number of photon counts in the pixel defining a scattering vector q and the correlated average is over all pairs of pixels subtending a fixed relative angle θ . $\langle I(q) \rangle$ represents the usual 2-point SAXS scattering averaged over all orientations. Since the exposure to the x-ray pulse happens for times of order 10's of femtoseconds (fs), the molecules in the droplet do not have time to undergo significant brownian motion during the exposure. Hence two photon scattering from any molecule in the ensemble happens for a given molecular orientation and internal conformation. We estimate that some 10^7 correlated 2-photon CXS events could be expected from a single 1micrometer

(mm) sized droplet of protein solution at a concentration of 10mg/ml placed exposed to a single x-ray pulse in an xfel beam focused to an area of $\sim 1(\mu\text{m})^2$. An advantage of CXS relative to the usual single photon SAXS measurement is that background subtraction is self-generated by the act of correlating the data, with subtraction of the product of average scattering in the 2 wave vector

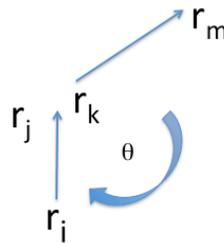
channels $|q|$ and $|q'|$ represented in equation (1). See Fig 1 for the relation of CXS data to molecular structure.

correlated x-ray scattering

Scattering events where two photons scatter off the same molecule

Scattering rate $C4(q_1, q_2, \cos(\theta))$ measures density of pairs r_{ij} within the molecule oriented at angle θ relative to pairs r_{km}

$C4$ is independent of the overall molecular orientation



In order to illustrate the kind of structural data available for non-crystalline samples from this kind of correlated measurement we show in Fig. 2 an estimate of changes in $C4$ for closed versus open states of the enzyme dihydrofolate reductase in the figure.

Fig 1.

The SAXS profile is similar to the $\cos(\theta)=0$ slice in $C4$, while changes at large $\cos(\theta)$ correspond to

changes in the internal structure of the enzyme not reflected in the SAXS profile.

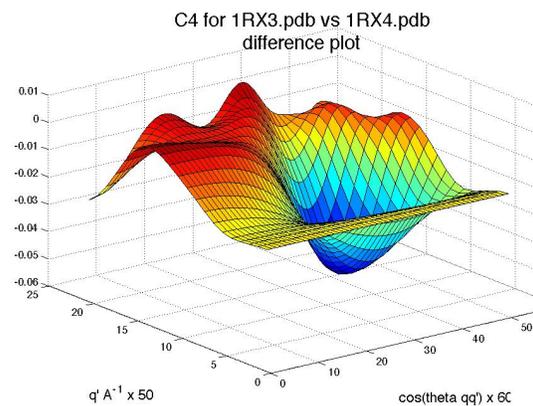
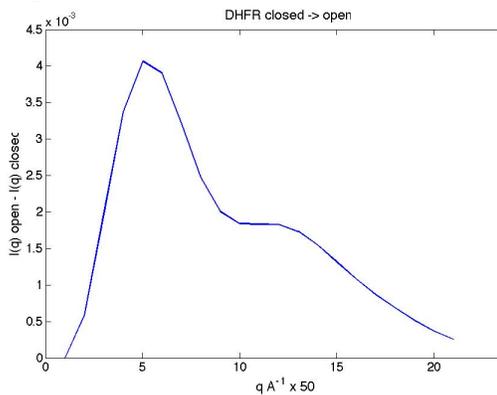


Figure 2: Difference in CXS for two states of dihydrofolate reductase (right) compared to the difference for one-photon scattering (SAXS) for the same pair of states (left).

Enzyme + substrate droplets will be injected into the beam at a range of times following mixing

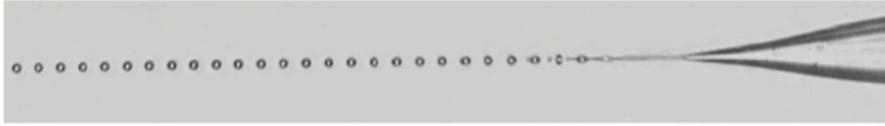


Figure 3. Rayleigh droplet source: Eight micron diameter droplets traveling at 20 m/sec in air, droplet breakup triggered by piezo at 20 kHz. Due to a necking instability, the continuous jet breaks into monodispersed droplets. Recently, aerodynamic focusing has been applied to achieve droplets of 1 micron and smaller, as used at FLASH.
(J. Spence and collaborators)

Each droplet will hold $\sim 10^8$ molecules

Scattering data will be collected in ~ 50 femtosec

Droplet will explode after $\sim 150 - 200$ femtosec

Correlated scattering: 2 photons scatter off the same molecule at an estimated yield of 10^7 events/pulse for 10^8 molecules per droplet

Droplets will hit the beam $\mu\text{sec} - \text{millisec}$ after mixing enzyme and substrate

The analysis of terabytes of data resulting from time resolved CXS measurements on enzyme systems, will be analyzed using an approach in which macromolecular kinetics are simulated using Markov state models (MSMs) built from molecular dynamics simulations.

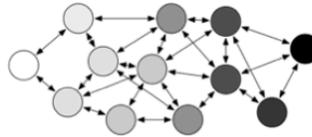
MSMs are essentially discrete-time master equations, where the macrostates and rates of transitioning between them are determined from molecular simulation. The Markov states are extracted by kinetic clustering of simulation data: conformations that can interconvert rapidly are grouped into the same state while conformations that can only interconvert slowly are grouped into separate

Markov State Models

Pande lab: build MSMs as a kinetic theory of a given long timescale, complex phenomenon

First step: build high resolution models from detailed molecular dynamics simulations on nanosec – μsec time scales to give a fully detailed master equation for the dynamics of interest

$$\frac{dp_{\alpha}}{dt} = \sum_{\beta} k_{\alpha\beta} p_{\beta}$$



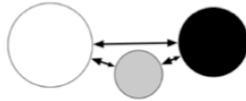
where the rate matrix is obtained from the short time simulations (quantitatively)

states. Such a kinetic clustering ensures that equilibration within a state, and therefore loss of memory of the previous state, occurs more rapidly than transitions between states. As a result, the model satisfies the Markov property—the

identity of the next state depends only on the identity of the current state and not any of the previous states.

Markov State Models

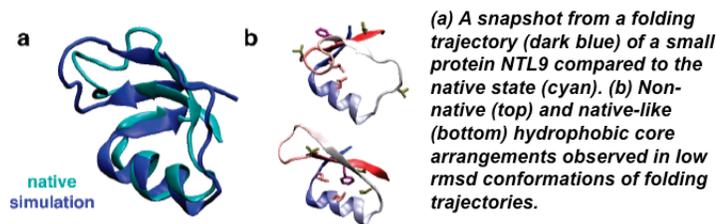
Second step: coarse grain in time to yield combinations of low resolution models $\{x_\alpha(t)\}$
(not cartoons – come straight from high resolution model/simulation data)



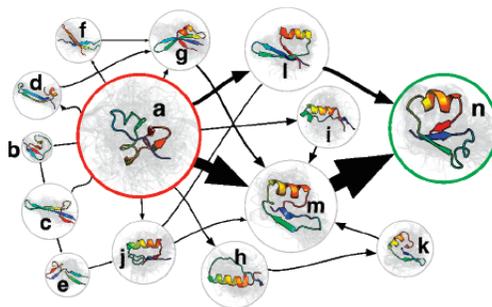
Markov State Models and CXS data

Third step: identify occupancies of coarse grained states at each time shot

Examples of the use of this approach include measurement of structural changes in Cyclophilin A. Human cyclophilin catalyzes *cis-trans* isomerization of a prolyl peptide bond of the HIV capsid needed to trigger a conformational change necessary for viral packaging. *cis-trans* Isomerization of prolyl peptide bonds is characterized by a very high activation energy barrier of around 16-22 kcal/mol, and the rate is in the order of tens to hundreds of seconds. Nature has provided the PPIase enzymes to circumvent this very slow kinetics by catalyzing the *cis-trans* isomerization and decreasing the time scale from seconds to the more biologically relevant millisecond time scale. In order to fully understand the catalytic mechanism of this speedup of more than 5 orders of magnitude, one



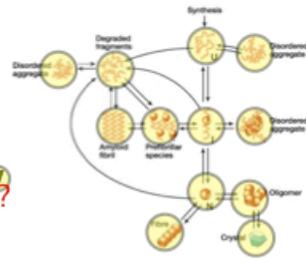
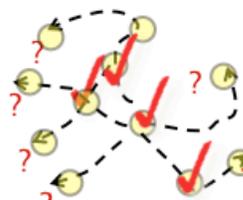
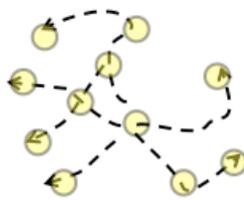
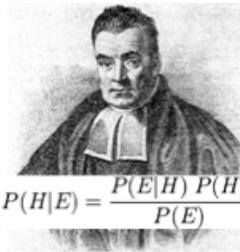
A 2000-state Markov State Model (MSM) was built using a lagtime of 12 ns. Shown is the superposition of the top 10 folding fluxes, which account for only ~25% of the total flux and transit only 14 of the 2000 macrostates.
V.A. Voelz, G.R. Bowman, K. Beauchamp, and V.S. Pande
J. AM. CHEM. SOC. 2010, 132, 1526–1528



has to be able to observe the *cis-trans* isomerization of the peptide bond at atomistic detail. CXS presents a revolutionary new method with which to directly observe the conformational changes essential to the catalytic function of Cyclophilin A.

Other possible examples of relevance to Bio-Energy and Bio-Remediation include modification of bacterial genes such as [Fe]-hydrogenases, involved in bio-fuel production and finding ways to avoid severe inhibition of Tetrachlorohydroquinone Dehalogenase made by bacteria used in environmental remediation.

Four key features of Markov State Models (MSMs)



- Statistical mechanics + Bayesian statistics lays a formal foundation for MSM theory.

- Many “short” trajectories can efficiently predict long timescale dynamics (as in TPS).

- Adaptive sampling methods can use simulation only where its needed most.

- Lumping and path analysis methods can yield insight directly

Singhal, Snow, Pande. *J. Chem. Phys.* (2004) 121, 415–425.
 Chodera, Singhal, Pande, Dill and Swope *J. Chem. Phys.* (2007), 126, 155101.
 Noé and Fischer *Curr. Opin. Struct. Biol.* (2008) 18 (2): 154-162
 Sarich, M., F. Noe, and C. Schütte. *SIAM Multiscale Model Simul.*, in press (2010).

Bowman, G. R.; Huang, X.; Pande, V. S. *Methods* (2009), 49, 197–201.
 Buchete and Hummer. *J. Phys. Chem. B* 2008, 112, 6057-6069
 Pan and Roux *J. Chem Phys.* (2008) 129(6): 064107.
 Noé, F., Schütte, C., Vanden-Eijnden, F., Reich, J., Weikl, TR. *PNAS* (2009) 106(45): 19011-6.
 Singhal, Snow, Pande. *J. Chem. Phys.* (2004) 121, 415–425.

Singhal and Pande. *J. Chem. Phys.* (2005) 123 204909
 Hinrichs and Pande *J. Chem. Phys.* (2007) 126, 244101
 Buchette and Hummer (2008)
 F. Noé *J. Chem. Phys.* (2008) 128, 244103
 Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. *PNAS* 2009, 106, 19765–19769.
 G. Bowman, V. Pande, *JCTC*, in press (2010)

Chodera, Singhal, Pande, Dill and Swope *J. Chem. Phys.* (2007), 126, 155101.
 Noé, Schütte, C., Vanden-Eijnden, Reich, Weikl. *PNAS* (2009) 106(45): 19011-6
 Berezhkovskii, A., Hummer, G., Szabo, A. *J. Chem Phys* 130:205102 (2009)
 Metzner, Schütte, Vanden-Eijnden *Multiscale Model Simul* (2009)