Bayesian analysis of non-thermal structural changes induced by terahertz radiation in protein crystals

Gergely Katona[1], Ida Lundholm[1], Helena Rodilla[2], Maria-Jose, Garcia-Bonete[1], Annette Duelli[1], Weixiao Y. Wahlgren[1], Gleb Bourenkov[3], Josip Vukusic[2], Ran Friedman[4], Thomas Schneider[3] & Jan Stake[2]

1 Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden.

2 Department of Microtechnology and Nanoscience, Chalmers University of Technology, Gothenburg, Sweden.

3 European Molecular Biology Laboratory Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany.

4 Department of Chemistry and Biomedicinal Sciences and Centre for Biomaterials Chemistry, Linnaeus University, Kalmar, Sweden.

Abstract— **We have periodically (25ms on – 25ms off) illuminated lysozyme crystals with 0.4 THz radiation and simultaneously monitored their X-ray diffraction intensity in order to study non-thermal structural changes in the protein. In this work we analyze the X-ray scaled and unmerged diffraction intensity observations using a multivariate Bayesian model in order to improve the accuracy of the intensity estimates. The diffraction intensity pairs of the illuminated and non-illuminated state show a predominantly positive correlation. The correlation decreases with increasing resolution suggesting that finer slicing and faster sampling of the rocking curve may further improve the accuracy and effect size of structure factor amplitude differences, making the interpretation of structural changes more straightforward. The improved analysis retains the most important structural features described previously (in helix 3) and provide addition details about the B-factor changes close to the substrate binding site.**

I. INTRODUCTION

Historic evidence suggests that major technological advances precede rationalization of biological processes. For example the recognition of "genetic program" and "genetic code" occurred after the invention of computers. Clearly, our current technology places a strong bias on how we can understand living organisms. This line of reasoning suggests that fundamental biological discoveries will be made, at best, hand in hand with cutting edge technologies such as (terahertz) wireless communication, quantum computation and phononic transport/conversion technologies involving nanostructures. Protein crystals are not only essential for X-ray diffraction analysis, but also interesting high-density, ordered objects with emergent properties that influence the structure and dynamics of its constituent molecules. As a search for emergent dynamic mechanisms we tested Fröhlich's hypothetical condensation process [1] by incoherent terahertz excitation of protein crystals and monitor their non-thermal response by X-ray diffraction. We developed an interleaved data collection strategy that was expected to greatly minimize the effect the X-ray radiation damage and cancel out the temperature difference between the on and off measurements. In addition one may anticipate an improvement in the sensitivity of the Bragg intensity measurements. In this work we demonstrate that this is indeed the case through a multivariate Bayesian analysis of diffraction intensities.

II. RESULTS

Multivariate Bayesian analysis of diffraction intensities

We pumped protein crystals with terahertz radiation while recording x-ray diffraction images.[2] A new intensity merging strategy [3] applied to this data. In order to analyze a large number of a unique reflection with Markov chain Monte Carlo (MCMC) method an efficient parallelization approach was developed. In less than 1 % of the analyzed reflections the MCMC algorithm failed to converge on the first try, therefore an automated method was applied to detect failure through the autocorrelation and standard deviation of parameter traces and the Markov chains were recalculated if necessary. Figure 1 shows the THz_{on}/THz_{off} intensity measurement pairs of a unique reflection and the fitted posterior multivariate normal distribution (distribution of mean parameters and isodensity ellipses). An offset from the diagonal potentially indicate a change in crystal structure as a response to terahertz radiation. The positive correlation is apparent for positive and negative THz_{on}/THz_{off} measurement pairs and the posterior of multivariate normal distribution is affected by the prior requirement of positivity of intensity. As this example shows positive correlation narrows the posterior estimate of THz_{on}-THz_{off} intensity difference. Repeating the same type of analysis with 13622 unique reflections (Figure 2) shows that a vast majority of the measurement pairs are positively correlated. One expects negative correlation if the sampling of the rocking curves is not fine enough. Estimates from the not optimal sampling of THz_{on} and THz_{off} observations could spread over a wide range, but their pairwise sum would show less variation. At room temperature this is a real threat as protein crystals tend to have small mosaic spread (as small as 0.1°) and thus a narrow rocking curve especially at high resolution, however the lack of negatively correlated multivariate distributions indicate that at least up to 1.7 Å resolution 50% sampling of 0.01° oscillations is sufficient. From this analysis it is apparent that even modern crystallographic data reduction methods leave substantial amount of unaccounted variation/systematic errors behind. The potential sources of errors are many: crystal to crystal non-isoforms, inadequate models of the crystal shape (absorption and anisotropy correction), lack of good model for X-ray radiation damage, unmodeled (time-dependent) errors in the diffraction geometry and detector response. The interleaved sampling method in addition to achieving the primary experimental goal (eliminating the temperature difference between the observations) also trades these difficult to treat errors of X-ray crystallography to a manageable level of sampling errors.

Electron density and atomic model parameter changes after the reanalysis of primary diffraction data

X-ray crystallographic data analysis tools are not prepared for accepting probability distributions functions (PDF) for structure factor amplitudes therefore one has to simplify the description of the PDFs. PDFs are not necessary symmetric and their mean, median and mode do not coincide. The median

appear to combine the qualities of robustness and closeness to the value of highest posterior probability. Where necessary the standard deviation of the PDFs was used as substitute to the standard error. Despite of the inherent incompatibilities of the Bayesian and frequentist approach, the crystallographic R-factor only deteriorated by 0.5 % as compared to the traditional data treatment. The difference Fourier map still highlight positive electron density changes around helix 3 often paired with negative densities indicating a shift in atomic positions (Figure 3). The difference between the refined models are mainly in the temperature- (B-) factors which again dominated by decrease upon exposure to THz radiation. Figure 4 shows Asn-44 which is present at two alternative conformations. Instead of responding to a uniform increase or decrease or random changes in the B-factors of the atoms in this residue we see that one alternative conformation increases while the other decreases in B-factor. The conformer that is decreasing in B-factor is located closer to loop 45-50, which shows a general decrease in B-factor, whereas the other conformer is associated with Gln-57 with an increasing B-factor. Two mechanisms may act either exclusively or in concert: (1) the static occupancies of the two conformational minima are unchanged on the ms timescales and the vibrational distribution is differentially altered in the two conformer possibly through its environment, (2) there is a genuine shift between the conformers on the sub-ms time scales as a result of THz radiation and since the occupancies are fixed the refinement can only alter the free B-factor parameters. At room temperature both mechanisms are likely and crystallography cannot distinguish between them. In either way substrate binding Asn-44 residue undergoes a vibrationally induced change, which may play a regulatory role in enzyme catalysis.

III. Summary

The multivariate Bayesian analysis indicates that the interleaved data collection method successfully reduced the variation between the pairwise recorded Bragg reflection intensities. Terahertz pumping of protein crystals in conjunction with X-ray diffraction experiments facilitate the assignment of picosecond vibration modes to specific structural motifs.
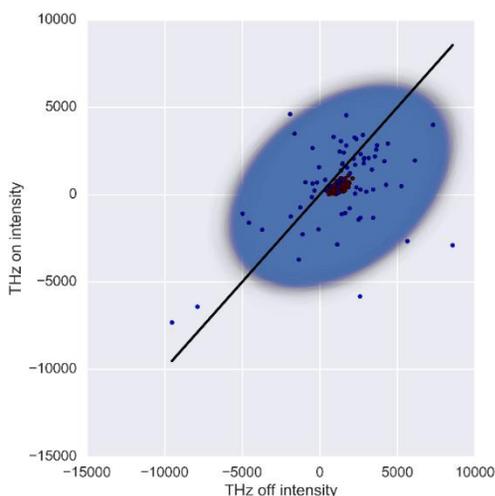

Figure 1 Light blue 90% isodensity ellipses illustrate samples from the multivariate normal distribution (Miller

index:(30,29,0)). Blue dots indicate intensity observation pairs (I_{THzoff}, I_{THzon}), red dots are samples from the posterior means of the multivariate normal distribution.
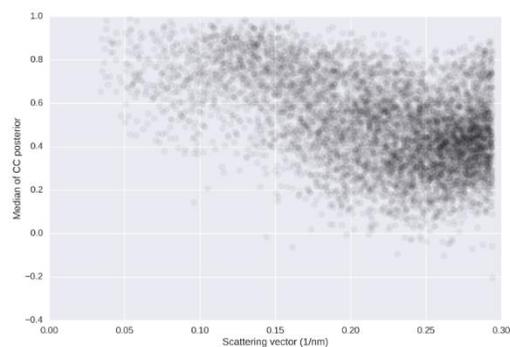

Figure 2 Posterior correlation coefficient between THz_{on} and THz_{off} measurements. Each transparent dot represent a unique reflection and plotted as a function of scattering vector length (resolution).

Figure 3 Electron density changes on helix 3 after the reanalyzed diffraction intensities. Green and red density indicate increase and decrease in electron density, respectively. (3.0σ)
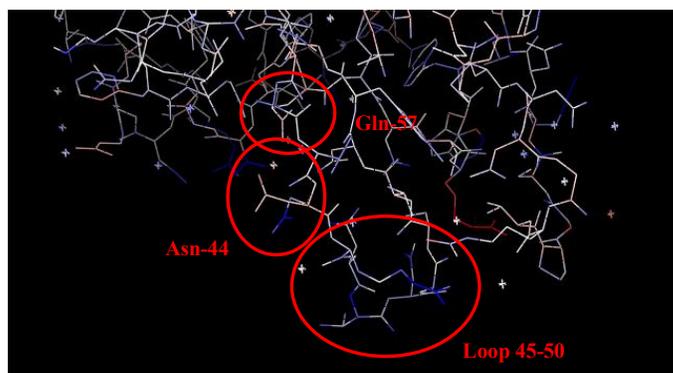

Figure 4 B-factor differences between THz_{on} and THz_{off} models. Red and blue colors indicate increase and decrease of B-factors in the illuminated state, respectively (-0.5 to 0.5 Å^2)

References

[1] H. Fröhlich, "Long-range coherence and energy storage in biological systems," *International Journal of Quantum Chemistry,* vol. 2, pp. 641-649, 1968.

[2] I. V. Lundholm, H. Rodilla, W. Y. Wahlgren, A. Duelli, G. Bourenkov, J. Vukusic, *et al.*, "Terahertz radiation induces non-thermal structural changes associated with Fröhlich condensation in a protein crystal," *Structural Dynamics,* vol. 2, p. 054702, 2015.

[3] G. Katona, M. J. Garcia-Bonete, and I. V. Lundholm, "Estimating the difference between structure-factor amplitudes using multivariate Bayesian inference," *Acta Crystallographica a-Foundation and Advances,* vol. 72, pp. 406-411, May 2016.