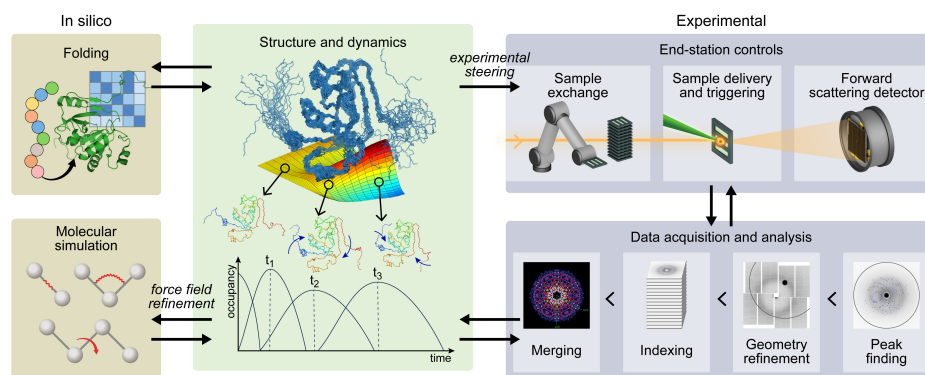# Graphical Abstract

Structural Biology in the age of X-ray Free Electron Lasers and Exascale Computing

Sandra Mous, Frédéric Poitevin, Mark S. Hunter, Dilipkumar N. Asthagiri, Thomas L. Beck

Highlights

Structural Biology in the age of X-ray Free Electron Lasers and Exascale Computing

Sandra Mous, Frédéric Poitevin, Mark S. Hunter, Dilipkumar N. Asthagiri, Thomas L. Beck

- X-ray free electron lasers and exascale computing provide synergistic insights into biomolecular dynamics.

- Integrated research infrastructures provide experimental data analysis and simulations on a minute timescale.

Structural Biology in the age of X-ray Free Electron Lasers and Exascale Computing

Sandra Mous[a], Frédéric Poitevin[a], Mark S. Hunter[a,*], Dilipkumar N. Asthagiri[b], Thomas L. Beck[b,**]

[a]*Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, 94025, CA, USA*
[b]*National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, 37830-6012, TN, USA*

Abstract

Serial femtosecond X-ray crystallography has emerged as a powerful method for investigating biomolecular structure and dynamics. With the new generation of X-ray free electron lasers, which generate ultrabright X-ray pulses at megahertz repetition rates, we can now rapidly probe ultrafast conformational changes and charge movement in biomolecules. Over the last year another innovation has been the deployment of Frontier, the world's first exascale supercomputer. Synergizing extremely high repetition rate X-ray light sources and exascale computing has the potential to accelerate discovery in biomolecular sciences. Here we outline our perspective on each of these remarkable innovations individually, and the opportunities and challenges in yoking them within an integrated research infrastructure.

Keywords: conformational dynamics, biomolecules, metalloproteins, nucleic acids, forcefields

---

[*]Corresponding author
[**]Corresponding author
Email addresses: `smous@slac.stanford.edu` (Sandra Mous), `fpoitevi@slac.stanford.edu` (Frédéric Poitevin), `mhunter2@slac.stanford.edu` (Mark S. Hunter), `asthagiridn@ornl.gov` (Dilipkumar N. Asthagiri), `becktl@ornl.gov` (Thomas L. Beck)

1. Introduction

The key to understanding proteins, nucleic acids, and other molecules of life depends on knowing the atomic arrangements in these molecules and how those arrangements change in response to interactions with the solvent or other molecules. To a great extent, our understanding of bio-macromolecular structure and function have relied on studying single, static structures, complemented by molecular simulations in favorable cases. Recent developments in time-resolved structural biology methods and molecular simulations promise to build on our knowledge of these static structures with a wealth of dynamic information, leading to a better understanding of the relationship between structure, dynamics, and function. In particular, the emergence of extremely bright X-ray light sources and exascale computing promises to usher in a new era of *dynamic* structural biology.

2. Studying the structural dynamics of biomolecules using time-resolved crystallography

X-ray Free Electron Lasers (XFELs), such as the Linac Coherent Light Source (LCLS)[1], have opened a new window to study the structural dynamics of macromolecules using serial femtosecond X-ray crystallography (SFX). XFELs produce X-ray radiation with exceptional properties, offering a billion-fold increase in peak brightness compared to synchrotron radiation. The femtosecond duration of the XFEL pulses ensures that atomic-level structural information is captured before the ionizing radiation can induce secondary electronic and atomic rearrangements (often referred to as "diffraction-before-destruction")[2]. As a result, performing measurements on samples at cryogenic temperatures – typically used for studying crystals at synchrotron light sources and in modern electron microscopy methods to mitigate radiation damage – is no longer necessary. Structures of macromolecules can therefore be obtained at near-physiological temperatures, including structures of particularly radiation-sensitive samples, including metalloenzymes[3–8] and protein radicals [9]. While the femtosecond pulse duration removes the need for cryogenic cooling of the samples, the high peak brightness of XFEL radiation enables X-ray diffraction data to be collected from crystals that are several micrometer or smaller in size. This capability has proven beneficial for the structural study of membrane proteins such as the medically relevant G-protein coupled receptors[10–13]. However,

the high peak brightness of the X-ray pulse also damages the sample shortly after the diffraction data are obtained, and therefore samples must be continually replenished in the X-ray interaction region (Fig. 1)[14, 15].
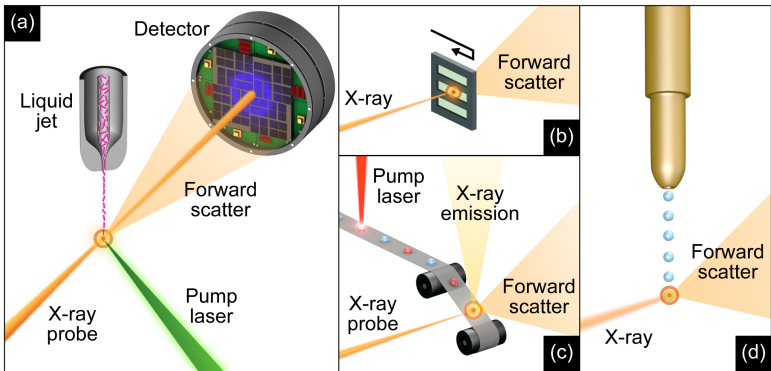

Figure 1: (a) Schematic illustrating the serial femtosecond X-ray crystallography (SFX) experiment. The liquid jet replenishes the crystalline sample at the interaction point, where the sample interacts with the X-ray pulse. Forward scattering data, which encodes the structural information, is collected by a detector placed closely to the interaction point. By pumping the sample with an optical laser pulse shortly before the X-ray pulse arrives, the dynamics of photosensitive systems can be studied. (b) Fixed-target sample delivery methods bring the crystalline sample into the interaction point using a chip or other sample holder[16]. (c) Tape-drive systems use piezo-acoustic injectors to deposit sample droplets on a conveyor belt system, from which X-ray diffraction and emission data can be collected simultaneously[17]. (d) Automated droplet-on-demand systems are another piezo-injector-based sample delivery system[18].

At XFELs, dynamics can be probed by expanding the experimental setup with a trigger, such as an optical pump laser (Fig. 1a)[19] or chemical pump, using a mixing device[20]. By varying the time-delay between the trigger and the X-ray pulse, macromolecular dynamics can be captured with femtosecond time resolution. Time-resolved methods at the XFEL have greatly increased our understanding of the ultrafast chemistry occurring within photosensitive systems (e.g., photoactive yellow protein[21], photosystem II[22], and the visual pigment rhodopsin[23]) and have captured the structural changes that occur in macromolecules upon ligand binding (including e.g., RNA[24, 25] and proteins that produce or provide resistance to antibiotics[3, 26]). Recently, the application area of time-resolved studies at the XFEL has been broadened through the development of new triggering methods, such as photocaged compounds[27], temperature jump excitation[28], and electric field stimulation[29].

Recent years have seen an increase in multimodal experiments, which help to validate X-ray diffraction studies through optical spectroscopy and computational simulations [30–33]. It is also possible to capture X-ray diffraction and X-ray spectroscopy data simultaneously (Fig. 1c), such that changes in both the atomic and electronic structure can be followed in time [34]. Among other examples, such multimodal experiments have recently enabled researchers to monitor the iron oxidation state while capturing the conformational changes of isopenicillin N synthase during catalysis [3]. New X-ray methods are still being developed, further expanding the toolkit for multimodal studies of macromolecular dynamics. For example, XES in the tender X-ray regime (∼1.5-5 keV) provides an exciting opportunity to capture the electronic state of elements common in biological systems, including sulfur, phosphorus, chlorine, and calcium [35]. The tender X-ray energy regime will also enable K-edge anomalous scattering methods [36] to be developed.

Further complimentary methods exist to probe biomolecular structure and dynamics (Fig. 2). While the SFX method is uniquely suited to probe ultrafast dynamics, serial synchrotron X-ray crystallography (SSX) can provide complimentary data in the millisecond time domain [32, 37, 38]. Both SFX and SSX rely on the cyrstallization of the macromolecule of interest, and alternative methods are therefore required to study single particles in solution. Among these methods, cryogenic electron microscopy (cryo-EM) has become an important tool in structural biology since advances in detector technology and image processing brought about the "resolution revolution" [39]. Particles in solution can also be studied using X-ray scattering methods, including time-resolved wide-angle X-ray scattering (TR-WAXS) [25] and fluctuation X-ray scattering (FXS) [40]), albeit at limited spatial resolution. Finally, single particle imaging (SPI) at XFELs [41] provides an opportunity to capture structural information from macromolecules resistant to crystallization. A unique advantage compared to other single particle methods such as cryo-EM is that the use of the ultrafast X-ray pulses removes the need for cryogenic cooling of the samples. While SPI was first demonstrated on large particles such as the giant mimivirus [42], the technique has progressed in the recent years, now allowing diffraction patterns to be collected from single proteins [43].
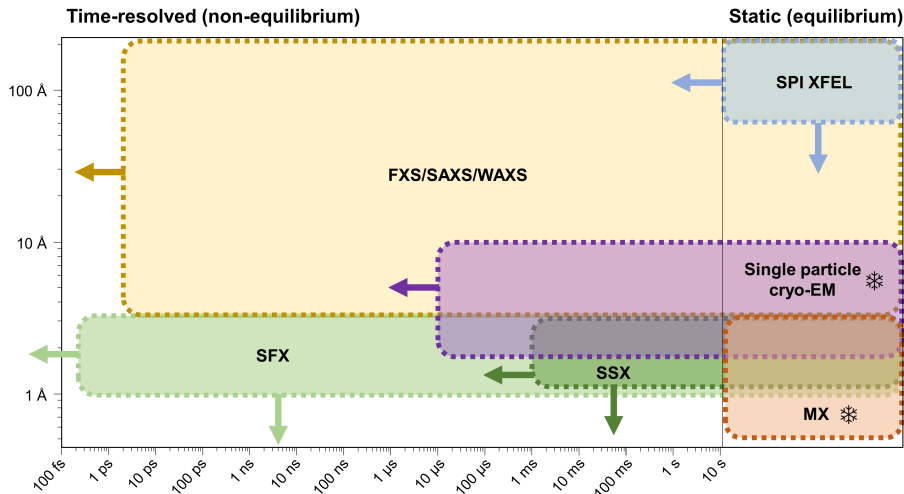
Figure 2: Resolution ranges achieved by the different diffraction and scattering techniques used for determining static structures and time-resolved dynamics: fluctuation X-ray scattering (FXS), small/wide-angle X-ray scattering (SAXS/WAXS), single particle imaging (SPI), cryogenic electron microscopy (cryo-EM), serial femtosecond X-ray crystallography (SFX), serial synchrotron X-ray crystallography (SSX), and macromolecular X-ray crystallography (MX). Methods that rely on cryogenic cooling of the sample (cryo-EM and MX) are marked with an ice crystal. Arrows indicate the expected improvement in resolution driven by further method development. It should be noted that the resolution range for SPI is expected to quickly expand in the following years as MHz XFELs become available.

3. Impact of high-repetition rate X-ray light sources

As discussed above, XFELs are able to capture the ultrafast chemistry and dynamics occurring within biological systems. However, the impact of these light sources has been limited due to the ∼100 Hz rate at which the X-ray pulses are delivered at most facilities. Recently, a new generation of XFELs (the European XFEL and LCLS-II) has come online, providing X-ray pulses at a megahertz rate. The European XFEL, in operation since 2017, recently demonstrated the use of megahertz repetition rates for studying biological structure and dynamics using time-resolved SFX [44, 45] and SPI [46]. Meanwhile, LCLS-II, which announced "first light" in 2023, will be upgraded to deliver hard X-rays (>5 keV) in the coming years (LCLS-II-HE)[47], making the facility suitable for structural biology experiments.

The ten-thousand-fold increase in the X-ray pulse rate of LCLS-II opens up new opportunities for studying biomolecular dynamics. For example, better sampling of the time axis in time-resolved pump-probe experiments allows for a more detailed mapping of the reaction coordinate. This will allow the rise and decay of reaction intermediates in heterogeneous samples to be captured in more detail (sometimes within a single crystal [48]), such that mixtures of intermediates can be modeled more accurately as a function of time. In addition, it is expected that a better temporal sampling will allow rare, short-lived events to be captured, without requiring prior information about reaction kinetics. Finally, we anticipate that the increased X-ray pulse rate will allow a wider parameter space and energy landscape of biological systems to be probed.

Large, comprehensive data sets will also benefit experiments with a limited signal to noise ratio by improving the precision of the measured structure factor amplitudes and scattering intensities through averaging. This will be particularly beneficial for experiments with a small differential signal (e.g., femtosecond pump-probe experiments) and experiments in which weak scattering signals are measured, including SPI, TR-WAXS, and FXS. In an SPI experiment, hundreds of thousands of snapshots have to be collected to reach near-atomic spatial resolution[49]. Hence, SPI, and other "data hungry" methods, will benefit significantly from an increased X-ray pulse repetition rate.

4. Studying the structural dynamics of biomolecules using exascale computing

From the early days of computer simulation of biomolecules, developments in the capability of computers and simulation algorithms have helped transform our understanding of biomolecular structure, function, and dynamics. To appreciate how exascale computing can complement XFEL experiments for the next stage of development in this area, it helps to consider just what exascale computing can deliver.

The world's first exascale supercomputer, Frontier, is able to deliver about 1.19×10^{18} floating point operations per second (flops) from its roughly 9,400 compute nodes packed into 74 cabinets. Each node comprises 4 GPUs and 1 CPU (with 64 cores). The Frontier supercomputer not only has a smaller footprint relative to its predecessor Summit (peak performance 0.2×10^{18} flops) but it is also 3.3 times more energy efficient (as measured

by flops/MW).

To better appreciate the capability of Frontier for biomolecular simulations, consider the simulation of Satellite Tobacco Mosaic Virus (STMV) in explicit water [50]. As Table 1 shows, the time to solution that required 256

Machine	hrs/ns
Ref. 50 (year 2006)	21.6
Workstation (with GPU)	16.8
Frontier 1 GPU	4.32
Frontier 2 GPU	2.40
Frontier 4 GPU	1.20

Table 1: Approximate times to obtain 1 ns of dynamics with a 1 fs timestep for the STMV model system comprising about 1.07×10^6 atoms. Throughout, we use 7 cores per GPU. Ref. [50] is the original study of the STMV model using 256 Altix nodes at the National Center for Supercomputing Applications (NCSA). The desktop workstation (built year 2022) is equipped with the NVIDIA RTX 6000 GPU. Frontier nodes have the AMD MI250X GPUs. Simulations were performed with the NAMD[51] (ver. 3.0), with the runs on Frontier in the so-called GPU-resident mode.

nodes of a supercomputer in year 2006 is now accessible with a single GPU on a standard workstation. The more capable GPUs in Frontier (or its predecessor, Summit) provide even better performance. Table 1 is meant only to provide a high-level guideline. The important point is that just *a fraction of one node* of Frontier is already a very capable machine for biomolecular simulations.

Ensemble computing for integrating with experiments
The Summit supercomputer enabled a record-breaking 10^9 atom simulation [52]. Preliminary testing shows that with 512 nodes of Frontier, we can credibly simulate 8×10^9 water molecules (Hagerty and Asthagiri, unpublished). But what is more interesting is the potential to study thousands of copies of a small ($\lesssim 10^6$ atom) system. The latter mode called *ensemble computing* not only enables enhanced sampling and scanning in parallel a large range of conditions of interest but is also of much interest in quantifying uncertainties [53]. The ensemble-mode will be of high interest in integrating with experiments at XFELs. We briefly review promising methodological developments that are well suited for ensemble computing.

Simulations to enhance the interpretation of XFEL experiments

XFELs can provide a detailed picture of conformational changes that are critical to biomolecular function. These could be intrinsic conformational changes that occur at thermal energies or changes driven by external stimuli (e.g. changes in solution condition, (un)binding of ligands, T-jump, or E-fields). These are also the questions that simulations have sought to address over the past decades, allowing molecular dynamics simulations to be directly integrated with insights from XFEL experiments. For example, simulated crystalline electron densities were directly compared to experimental electron densities obtained from light-activated photosystem II crystals, providing unique insights into the mobility of the water networks inside the protein [33]. The ongoing development of several important molecular dynamics algorithms, described below, is expected to further enhance the integration of simulation with experimental data.

When experiments indicate the direction of change, for example, isomerization about a specific bond or the (un)binding along a well-defined pathway, those directions serve as natural *order parameters*. In this case, biased sampling techniques could prove particularly helpful [54–56]. When sampling is desired without applying any external biases, as would be required for obtaining kinetic properties, the highly scalable weighted ensemble technique presents itself as a compelling choice [57, 58].

When experiments make available a collection of conformational states, but the directions of change are not immediately obvious or several concerted (allosteric) motions are indicated, it becomes important to classify the various conformations and elucidate the mechanism of transitions among them. Markov State Models (MSMs)[59–63] based on a time-lagged independent component analysis (TICA) [64, 65] are particularly compelling in this case. Based on a collection of vectors of user specified order parameters, the TICA method finds those linear combinations, the basis, that de-correlate slowly and are thus most useful in describing the slowest relaxation processes, such as large-scale conformational changes of interest (Fig. 3). Integrating adaptive sampling further enhances the efficiency and explanatory power of MSMs [66].

The comments above highlight the importance of order parameter(s). It is important to recognize that intuitively reasonable order parameters, such as the distance between a protein and a ligand, need not be the appropriate reaction coordinate for elucidating the kinetics [67]. To elucidate kinetics rigorously requires identifying the correct transition states, and this remains

a challenge in high-dimensions. An exciting recent development in this regard is machine-guided transition path sampling [68], wherein a neural network is trained to process a large collection of features, perhaps generate autonomously, to converge on a set of features that help locate the transition state. The final results can be interpreted using symbolic regression to derive chemically meaningful models.

Finally, developments in the molecular quasichemical theory (QCT) of solutions [69] now make it possible to obtain a comprehensive view of the thermodynamics of biomolecules in solution [70], treating the molecule holistically without relying on additivity assumptions that are often invoked in interpreting the thermodynamics of biomolecular transformations [71]. The QCT framework is rigorous, physically transparent, and designed to make good use of available data, be it from simulations or experiments. As such, the QCT framework is amenable to ensemble computing and can mesh with MSM or weighted sampling approaches to predict thermodynamics of conformational substates.

5. A vision for building an integrated research infrastructure coupling XFELs and exascale computing

The developments in XFEL and exascale computing are synergistic and can enable a comprehensive understanding of biomolecular dynamics. Figure 3 is our vision for integrating these two techniques. While experimental data is crucial for refining simulation parameters, integrating simulations into time-resolved data analysis can advance and improve the analysis and interpretation of experiments. We consider two specific possibilities, one in steering experiments using information from simulations and another in learning from experiments to improve simulations.

Simulations to steer the XFEL experiments

Machine learning methods are increasingly capable of characterizing and calibrating XFEL hardware [72] by integrating simulation and experimental data (Fig. 3). We envision that simulations, through providing a detailed kinetic model, can also predict the time delays at which we might find transition states and inform us about the triggering time delay (in applying external stimuli). A step in this direction is the use of a neural network to optimize experimental design[73]. In a similar vein, simulations could provide rich information about the macromolecule's response to varying physical/chemical
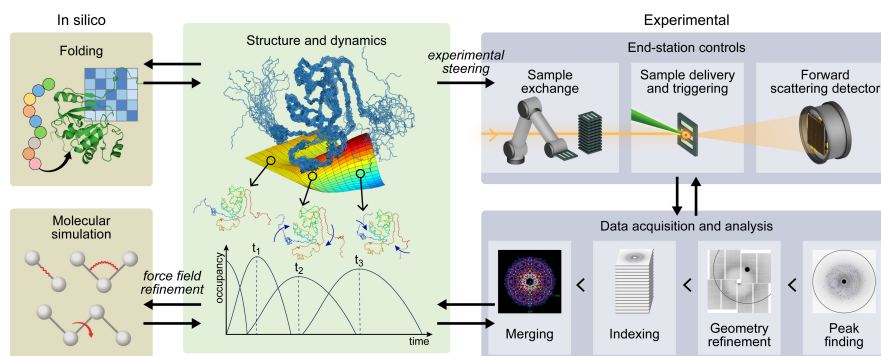
Figure 3: Molecular simulations (in silico) and time-resolved studies at XFELs (experimental) provide a comprehensive understanding of the structure and dynamics of macromolecules. Structural and dynamics information provided by simulations or experiments can be used to improve computational methods (e.g., through force field refinement) or help drive the experiment at the X-ray free electron laser (experimental steering). The higher X-ray pulse rate of the new generation of XFELs will allow for an increased sampling of the time axis in a time-resolved experiment. In this way, the rise and decay of reaction intermediates and short-lived states can be captured more accurately.

conditions (for example, ligand concentrations). This could provide a guide for identifying the most interesting samples to measure next, thereby informing the automated sample exchange.

Learning from experiments to improve simulations

Above we tacitly assumed simulations are well-crafted and the force fields are adequate. The former is user-dependent, and the latter may not always be the case. For systems involving metal cofactors we will need a quantum mechanical model for at least part of the metal-protein system. The physics of polarizability, often not included explicitly in standard force fields, and genuine many body effects could also play an important role. Modeling of nucleic acids is one area where some of these issues come to the fore [74]. However, the near one-to-one correspondence between XFEL experiments and simulations does open a way to use the knowledge from experiments to update the force fields for simulations.

Challenges

We have thus far sketched the potential for coupling XFELs and exascale computing, and the tools and techniques that would be of most interest. We

10

briefly highlight key challenges. While being US-centered, it is informative to consider the specifications for the plans of the US Department of Energy to build an integrated research infrastructure between its major experimental facilities, including XFELs, and the leadership computing facilities (Fig. 4): the total aggregate and peak compute performance should be on the order of exascale, filesystems must support high-throughput, concurrent data writing and reading from many sources at once, centralized data storage facilities should have capacities up to exabytes and networking bandwidths should be on the order of terabytes per second. The current standard for the DOE ESnet transmission rate is about 50 gigabytes per second, illustrating the importance of effective data compression and motivating the push to higher bandwidths across the network.

We should note that in addition to these infrastructure challenges, the many-fold increase in data generation capacity of LCLS-II also calls for algorithmic developments leveraging the state of the art in computer vision and structure generation from diffraction patterns arising from a large number of randomly oriented samples. The Exascale Computing Project within the DOE developed a wide range of such software tools and libraries, including the exaFEL suite of codes [75]. In addition, artifical intelligence and machine learning techniques are being developed to enable GPU parallel computing [76, 77] and tackle data compression[78, 79], so that actionable information can be extracted quickly and accurately from the deluge of data back to steer the experiment. Finally, to truly provide scientists with a seamless integration of data and algorithms which informs them on demand, it will be crucial to track data and processing workflows through delocalized registries across all facilities.

6. Conclusions

Being able to obtain time-resolved pictures of molecular transformations from experiments at the XFEL and simultaneously study the same in computer simulations opens broad possibilities for a deeper understanding of biomolecular structure, dynamics, and function at ambient temperatures. Technical challenges to integrate experimental and computational facilities separated by a continent are being addressed, and those developments will also favorably impact other areas of science and technology. We are at the threshold of a new age that fully integrates experiments and computing at a level that was not previously possible.
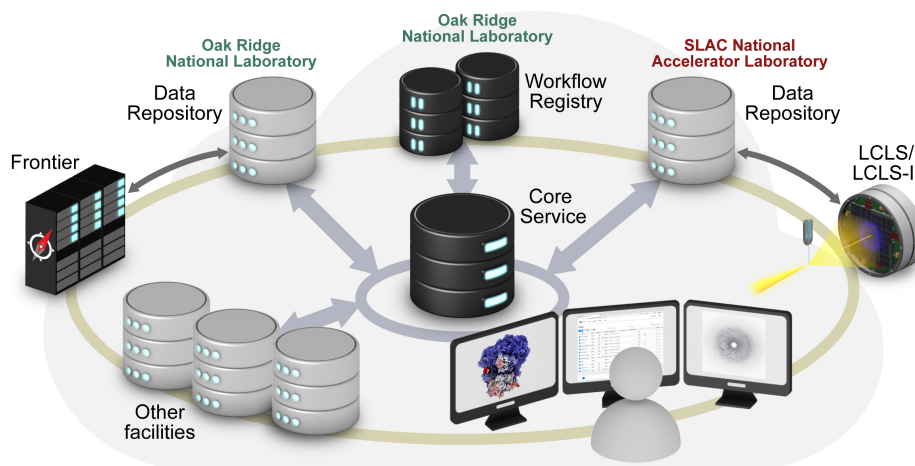
Figure 4: Data portal for high-throughput multimodal analysis. Structural and molecular dynamics data collected at the XFEL, such as LCLS or LCLS-II, will be registered in a central workflow registry to facilitate rapid collocation with data collected from other facilities and high-performance computing resources, such as the Frontier exascale supercomputer. These Integrated Research Infrastructures (IRI) will enable data processing, reprocessing, and large-scale multimodal studies.

7. Acknowledgements

References

[1] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, et. al., First lasing and operation of an Ångstrom-wavelength free-electron laser, Nature Photonics 4 (2010) 641–647.

[2] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, Potential for biomolecular imaging with femtosecond x-ray pulses, Nature 406 (6797), *Simulations that demonstrate that the ultrafast X-ray pulses from X-ray free electron lasers can be used to capture structural information before radiation damage destroys the sample. (2000).

[3] P. Rabe, J. J. A. G. Kamps, K. D. Sutherlin, J. D. S. Linyard, P. Aller, C. C. Pham, H. Makita, I. Clifton, M. A. McDonough, T. M. Leissing, et al., X-ray free-electron laser studies reveal correlated motion during isopenicillin N synthase catalysis, Science Advances 7 (2021) eabh0250.

[4] C. J. Ohmer, M. Dasgupta, A. Patwardhan, I. Bogacz, C. Kaminsky, M. D. Doyle, P. Y.-T. Chen, S. M. Keable, H. Makita, P. S. Simon, et al., XFEL serial crystallography reveals the room temperature structure of methyl-coenzyme M reductase, J. Inorg. Biochem. 230 (2022) 111768.

[5] J. John, O. Aurelius, V. Srinivas, P. Saura, I.-S. Kim, A. Bhowmick, P. S. Simon, M. Dasgupta, C. Pham, S. Gul, et al., Redox-controlled reorganization and flavin strain within the ribonucleotide reductase R2b–NrdI complex monitored by serial femtosecond crystallography, eLife 11 (2022) e79226.

[6] A. Bhowmick, R. Hussein, I. Bogacz, P. S. Simon, M. Ibrahim, R. Chatterjee, M. D. Doyle, M. H. Cheah, T. Fransson, P. Chernev, et al., Structural evidence for intermediates during O_2 formation in photosystem II, Nature 617 (2023) 629–636.

[7] I. Ishigami, R. G. Sierra, Z. Su, A. Peck, C. Wang, F. Poitevin, S. Lisova, B. Hayes, F. R. Moss, S. Boutet, et al., Structural insights into functional properties of the oxidized form of cytochrome c oxidase, Nature Communications 14 (2023) 5752.

[8] I. Ishigami, S. Carbajo, N. Zatsepin, M. Hikita, C. E. Conrad, G. Nelson, J. Coe, S. Basu, T. Grant, M. H. Seaberg, et al., Detection of a

geminate photoproduct of bovine cytochrome c oxidase by time-resolved serial femtosecond crystallography, J. Am. Chem. Soc. 145 (2023) 22305–22309.

[9] H. Lebrette, V. Srinivas, J. John, O. Aurelius, R. Kumar, D. Lundin, A. S. Brewster, A. Bhowmick, A. Sirohiwal, I.-S. Kim, et al., Structure of a ribonucleotide reductase R2 protein radical, Science 382 (2023) 109–113, **This study reports the first structure of an enzyme catalyzing radical chemistry, demonstrating the unique advantages of the diffraction-before-destruction technique made possible by X-ray free electron lasers.

[10] B. Stauch, L. C. Johansson, J. D. McCorvy, N. Patel, G. W. Han, X.-P. Huang, C. Gati, A. Batyuk, S. T. Slocum, A. Ishchenko, et al., Structural basis of ligand recognition at the human MT1 melatonin receptor, Nature 569 (2019) 284–288.

[11] L. C. Johansson, B. Stauch, J. D. McCorvy, G. W. Han, N. Patel, X.-P. Huang, A. Batyuk, C. Gati, S. T. Slocum, C. Li, et al., XFEL structures of the human MT2 melatonin receptor reveal the basis of subtype selectivity, Nature 569 (2019) 289–292.

[12] A. Luginina, A. Gusach, E. Marin, A. Mishin, R. Brouillette, P. Popov, A. Shiriaeva, Élie Besserer-Offroy, J.-M. Longpré, E. Lyapina, et al., Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs, Science Advances 5 (2019) eaax2518.

[13] H. Liu, R. N. V. K. Deepak, A. Shiriaeva, C. Gati, A. Batyuk, H. Hu, U. Weierstall, W. Liu, L. Wang, V. Cherezov, H. Fan, C. Zhang, Molecular basis for lipid recognition by the prostaglandin D_2 receptor CRTH2, Proc. Natl. Acad. Sci. USA 118 (2021) e2102813118.

[14] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, et al., Femtosecond x-ray protein nanocrystallography, Nature 470 (2011) 73–77.

[15] S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, et al., High-resolution protein structure determination by serial femtosecond crystallography, Science 337 (2012) 362–364.

[16] M. S. Hunter, B. Segelke, M. Messerschmidt, G. J. Williams, N. A. Zatsepin, A. Barty, W. H. Benner, D. B. Carlson, M. Coleman, A. Graf, et al., Fixed-target protein serial microcrystallography with an x-ray free electron laser, Sci. Rep. 4 (2014) 6026.

[17] F. D. Fuller, S. Gul, R. Chatterjee, E. S. Burgie, I. D. Young, H. Lebrette, V. Srinivas, A. S. Brewster, T. Michels-Clark, J. A. Clinger, et al., Drop-on-demand sample delivery for studying biocatalysts in action at x-ray free-electron lasers, Nat. Methods 14 (2017) 443–449.

[18] Z. Su, J. Cantlon, L. Douthit, M. Wiedorn, S. Boutet, J. Kern, C. H. Yoon, D. DePonte, Serial crystallography using automated drop dispensing, J. Synchrotron Rad. 28 (2021) 1386–1392.

[19] J. M. Glownia, J. Cryan, J. Andreasson, A. Belkacem, N. Berrah, C. I. Blaga, C. Bostedt, J. Bozek, L. F. DiMauro, L. Fang, J. Frisch, et al., Time-resolved pump-probe experiments at the LCLS, Opt. Express 18 (2010) 17620–17630.

[20] J. L. Olmos, Jr, S. Pandey, J. M. Martin-Garcia, G. Calvey, A. Katz, J. Knoska, C. Kupitz, M. S. Hunter, M. Liang, D. Oberthuer, et al., Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography, BMC Biol. 16 (2018) 59.

[21] A. Hosseinizadeh, N. Breckwoldt, R. Fung, R. Sepehr, M. Schmidt, P. Schwander, R. Santra, A. Ourmazd, Few-fs resolution of a photoactive protein traversing a conical intersection, Nature 599 (2021) 697–701.

[22] A. Bhowmick, P. S. Simon, I. Bogacz, R. Hussein, M. Zhang, H. Makita, M. Ibrahim, R. Chatterjee, M. D. Doyle, M. H. Cheah, et al., Going around the Kok cycle of the water oxidation reaction with femtosecond X-ray crystallography, IUCrJ 10 (6) (2023).

[23] T. Gruhl, T. Weinert, M. J. Rodrigues, C. J. Milne, G. Ortolani, K. Nass, E. Nango, S. Sen, P. J. M. Johnson, C. Cirelli, et al., Ultrafast structural changes direct the first molecular events of vision, Nature 615 (2023) 939–944.

[24] S. Ramakrishnan, J. R. Stagno, C. E. Conrad, J. Ding, P. Yu, Y. R. Bhandari, Y.-T. Lee, G. Pauly, O. Yefanov, M. O. Wiedorn, et al., Syn-

chronous RNA conformational changes trigger ordered phase transitions in crystals, Nat. Commun. 12 (2021) 1762.

[25] K. A. Zielinski, S. Sui, S. A. Pabit, D. A. Rivera, T. Wang, Q. Hu, M. M. Kashipathy, S. Lisova, C. B. Schaffer, V. Mariani, et al., RNA structures and dynamics with Å resolution revealed by x-ray free-electron lasers, Science Advances 9 (2023) eadj3509.

[26] S. Pandey, G. Calvey, A. M. Katz, T. N. Malla, F. H. M. Koua, J. M. Martin-Garcia, I. Poudyal, J.-H. Yang, M. Vakili, O. Yefanov, et al., Observation of substrate diffusion and ligand binding in enzyme crystals using high-repetition-rate mix-and-inject serial crystallography, IUCrJ 8 (2021) 878–895.

[27] M. Wranik, T. Weinert, C. Slavov, T. Masini, A. Furrer, N. Gaillard, D. Gioia, M. Ferrarotti, D. James, H. Glover, et al., Watching the release of a photopharmacological drug from tubulin using time-resolved serial crystallography, Nat. Commun. 14 (2023) 903, *The first demonstration of using photopharmacological compounds to study protein-ligand interactions at the XFEL.

[28] A. M. Wolff, E. Nango, I. D. Young, A. S. Brewster, M. Kubo, T. Nomura, M. Sugahara, S. Owada, B. A. Barad, K. Ito, et al., Mapping protein dynamics at high spatial resolution with temperature-jump x-ray crystallography, Nat. Chem. 15 (2023) 1549–1558, *This article demonstrates solvent-based temperature jumps as a universal method for perturbing molecular motion to study protein dynamics with time-resolved X-ray methods.

[29] D. R. Hekstra, K. I. White, M. A. Socolich, R. W. Henning, V. Šrajer, R. Ranganathan, Electric-field-stimulated protein mechanics, Nature 540 (2016) 400–405, * This work describes electric field–stimulated X-ray crystallography (EF-X), in which strong electric field pulses are used to apply of forces of controlled magnitude, direction and duration on crystals to drive conformational changes.

[30] P. Nogly, T. Weinert, D. James, S. Carbajo, D. Ozerov, A. Furrer, D. Gashi, V. Borin, P. Skopintsev, K. Jaeger, et al., Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser, Science 361 (2018) eaat0094.

[31] P. Skopintsev, D. Ehrenberg, T. Weinert, D. James, R. K. Kar, P. J. M. Johnson, D. Ozerov, A. Furrer, I. Martiel, F. Dworkowski, et al., Femtosecond-to-millisecond structural changes in a light-driven sodium pump, Nature 583 (2020) 314–318.

[32] S. Mous, G. Gotthard, D. Ehrenberg, S. Sen, T. Weinert, P. J. M. Johnson, D. James, K. Nass, A. Furrer, D. Kekilli, P. Ma, et al., Dynamics and mechanism of a light-driven chloride pump, Science 375 (2022) 845–851.

[33] M. D. Doyle, A. Bhowmick, D. C. Wych, L. Lassalle, P. S. Simon, J. Holton, N. K. Sauter, V. K. Yachandra, J. F. Kern, J. Yano, M. E. Wall, Water networks in Photosystem II using crystalline molecular dynamics simulations and room-temperature XFEL serial crystallography, J. Am. Chem. Soc. 145 (2023) 14621–14635.

[34] R. Alonso-Mori, J. Kern, R. J. Gildea, D. Sokaras, T.-C. Weng, B. Lassalle-Kaiser, R. Tran, J. Hattne, H. Laksmono, J. Hellmich, et al., Energy-dispersive x-ray emission spectroscopy using an X-ray free-electron laser in a shot-by-shot mode, Proc. Natl. Acad. Sci. USA 109 (2012) 19103–19107.

[35] B. Abraham, S. Nowak, C. Weninger, R. Armenta, J. Defever, D. Day, G. Carini, K. Nakahara, A. Gallo, S. Nelson, et al., A high-throughput energy-dispersive tender X-ray spectrometer for shot-to-shot sulfur measurements, J. Synchrotron Rad. 26 (2019) 629–634.

[36] K. Virk, K. Yonezawa, K. Choukate, L. Singh, N. Shimizu, B. Chaudhuri, K-edge anomalous SAXS for protein solution structure modeling, Acta Crystal. D 78 (2022) 204–211.

[37] T. Weinert, P. Skopintsev, D. James, F. Dworkowski, E. Panepucci, D. Kekilli, A. Furrer, S. Brünle, S. Mous, D. Ozerov, et al., Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography, Science 365 (6448) (2019) 61–65.

[38] F. Leonarski, J. Nan, Z. Matej, Q. Bertrand, A. Furrer, I. Gorgisyan, M. Bjelčić, M. Kepa, H. Glover, V. Hinger, et al., Kilohertz serial crystallography with the JUNGFRAU detector at a fourth-generation synchrotron source, IUCrJ 10 (6) (2023) 729–737.

[39] W. Kühlbrandt, The resolution revolution, Science 343 (6178) (2014) 1443–1444.

[40] K. Pande, J. J. Donatelli, E. Malmerberg, L. Foucar, C. Bostedt, I. Schlichting, P. H. Zwart, Ab initio structure determination from experimental fluctuation x-ray scattering data, Proc. Natl. Acad. Sci. USA 115 (2018) 11772–11777.

[41] J. Bielecki, F. R. N. C. Maia, A. P. Mancuso, Perspectives on single particle imaging with x rays at the advent of high repetition rate x-ray free electron laser sources, Structural Dynamics 7 (2020) 040901.

[42] M. M. Seibert, T. Ekeberg, F. R. N. C. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, et al., Single mimivirus particles intercepted and imaged with an x-ray laser, Nature 470 (7332) (2011) 78–81.

[43] T. Ekeberg, D. Assalauova, J. Bielecki, R. Boll, B. J. Daurer, L. A. Eichacker, L. E. Franken, D. E. Galli, L. Gelisio, L. Gumprecht, et al., Observation of a single protein by ultrafast x-ray diffraction, Light: Science & Applications 13 (1) (2024) 15, **This works reports the first X-ray diffraction pattern collected on a single protein, demonstrating the feasibility of single particle imaging at X-ray free electron lasers.

[44] M. O. Wiedorn, D. Oberthür, R. Bean, R. Schubert, N. Werner, B. Abbey, M. Aepfelbacher, L. Adriano, A. Allahgholi, N. Al-Qudami, J. Andreasson, et al., Megahertz serial crystallography, Nat. Commun. 9 (2018) 4025.

[45] S. Pandey, R. Bean, T. Sato, I. Poudyal, J. Bielecki, J. Cruz Villarreal, O. Yefanov, V. Mariani, T. A. White, C. Kupitz, et al., Time-resolved serial femtosecond crystallography at the european XFEL, Nat. Methods 17 (2020) 73–78.

[46] E. Sobolev, S. Zolotarev, K. Giewekemeyer, J. Bielecki, K. Okamoto, H. K. N. Reddy, J. Andreasson, K. Ayyer, I. Barak, S. Bari, et al., Megahertz single-particle imaging at the European XFEL, Communications Physics 3 (2020) 97.

[47] R. W. Schoenlein, S. Boutet, M. P. Minitti, A. Dunne, The Linac Coherent Light Source: Recent developments and future plans, Applied Sciences 7 (8) (2017).

[48] S. Holmes, H. J. Kirkwood, R. Bean, K. Giewekemeyer, A. V. Martin, M. Hadian-Jazi, M. O. Wiedorn, D. Oberthür, H. Marman, et al., Megahertz pulse trains enable multi-hit serial femtosecond crystallography experiments at x-ray free electron lasers, Nat. Commun. 13 (1) (2022) 4708.

[49] I. Poudyal, M. Schmidt, P. Schwander, Single-particle imaging by x-ray free-electron lasers — How many snapshots are needed?, Structural Dynamics 7 (2020) 024102.

[50] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, K. Schulten, Molecular dynamics simulations of the complete Satellite Tobacco Mosaic Virus, Structure 14 (2006) 437–449.

[51] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, et al., Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD, J. Chem. Phys. 153 (4) (2020) 044130. doi:10.1063/5.0014475.

[52] A. Dommer, L. Casalino, F. Kearns, M. Rosenfeld, N. Wauer, S.-H. Ahn, J. Russo, S. Oliveira, C. Morris, A. Bogetti, A. Trifan, A. Brace, et. al., COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol, International Journal of High Performance Computing Applications 37 (1) (2023) 28 – 44, *An important milestone in constructing and simulating a realistic model of an aerosolized virus.

[53] M. Vassaux, S. Wan, W. Edeling, P. V. Coveney, Ensembles are required to handle aleatoric and parametric uncertainty in molecular dynamics simulation, J. Chem. Theory Comput. 17 (2021) 5187–5197.

[54] C. Chipot, A. Pohorille (Eds.), Free Energy Calculations: Theory and Applications in Chemistry and Biology, Vol. 86 of Springer Series in Chemical Physics, Springer, Berlin, 2007, *Ideas presented in this monograph remain fresh and continue to be developed further.

[55] E. Darve, Thermodynamic integration using constrained and unconstrained dynamics, in: C. Chipot, A. Pohorille (Eds.), Free energy calculations: Theory and applications in chemistry and biology, Vol. 86 of Springer series in Chemical Physics, Springer, 2007, Ch. 4, pp. 119–170, **An accessible presentation of various sampling methods. Connections between methods such as umbrella sampling, adaptive bias forces, and metadynamics are shown.

[56] F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcare, M. Parrinello, V. Limongelli, Ligand binding to telomeric g-quadruplex DNA investigated by funnel-metadynamics simulations, Proc. Natl. Acad. Sci. USA 11 (2017) E2136–E2145.

[57] D. M. Zuckerman, L. T. Chong, Weighted ensemble simulation: Review of methodology, applications, and software, Annu. Rev. Biochem. 46 (2017) 43–57, *An accessible presentation of the weighted ensemble method. Impact of ideas going back to the early days of Monte Carlo sampling are revealed in a modern context.

[58] L. T. Chong, A. S. Saglam, D. M. Zuckerman, Path-sampling strategies for simulating rare events in biomolecular simulations 43 (2017) 88–94.

[59] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, A direct approach to conformational dynamics based on Hybrid Monte Carlo, J. Comput. Phys. 151 (1999) 146–168.

[60] V. S. Pande, K. Beauchamp, G. R. Bowman, Everything you wanted to know about Markov State Models but were afraid to ask, Methods 52 (2010) 99–105, *Very accessible account of MSMs.

[61] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation, J. Chem. Phys. 134 (2023) 174105, *Stresses the importance of validating MSMs.

[62] K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, V. S. Pande, MSMBuilder2: Modeling conformational dynamics on the picosecond to millisecond scale, J. Chem. Theory Comput. 7 (2011) 3412–3419, *Discussion of a software tool that has advanced the use of MSMs and the TICA procedure.

[63] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, P. Jan-Hendrik, F. Nóe, PyEMMA 2: A Software Package for Estimation, validation, and analysis of Markov models, J. Chem. Theory Comput. 11 (2015) 5525–5542, *Discussion of a software tool that has advanced the use of MSMs and the TICA procedure.

[64] C. R. Schwantes, V. S. Pande, Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9, J. Chem. Theory Comput. 9 (2013) 2000–2009, *Development of TICA for modeling macromolecular dynamics.

[65] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction, J. Chem. Phys. 139 (2013) 015102, *Development of TICA for modeling macromolecular dynamics.

[66] N. Plattner, S. Doerr, G. D. Fabritiis, F. Noé, Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling, Nature Chemistry 9 (2017) 1005–1011, **Efficiency of adaptive sampling and explanatory power of well-crafted MSMs come together to elucidate the kinetics and thermodynamics of association in the barnase-barstar system.

[67] P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark, Ann. Rev. Phys. Chem. 53 (2002) 291–318.

[68] H. Jung, R. Covino, A. Arjun, C. Leitold, C. Dellago, P. G. Bolhuis, G. Hummer, Machine-guided path sampling to discover mechnisms of molecular self-organization, Nature Comp. Sc. 3 (2023) 334–345, **Machine-guided approach to pick reaction coordinates for understanding molecular self-organization. Symbolic regression on the learned coordinate provides a chemically meaningful model. The authors treat a range of problems, including ion-pair association, formation of clathrate hydrates, folding of a polymer, and association of proteins in a membrane system.

[69] T. L. Beck, M. E. Paulaitis, L. R. Pratt, The Potential Distribution Theorem and Models of Molecular Solutions, Cambridge University Press,

Cambridge, UK, 2006, *Builds upon the partition function perspective of the potential distribution theorem. The work also provides a fresh perspective on earlier theories of solutions.

[70] D. N. Asthagiri, M. E. Paulaitis, L. R. Pratt, Thermodynamics of hydration from the perspective of the molecular quasichemical theory of solutions, J. Phys. Chem. B 125 (2021) 8294–8304.

[71] D. S. Tomar, M. E. Paulaitis, L. R. Pratt, D. N. Asthagiri, Hydrophilic Interactions Dominate the Inverse Temperature Dependence of Polypeptide Hydration Free Energies Attributed to Hydrophobicity, J. Phys. Chem. Lett. 11 (2020) 9965–9970, **Highlights the new insights enabled by molecular QCT, and reveals the conceptual limitations of additivity models which underpin the current understanding of dominant forces in protein folding and stability.

[72] N. Breckwoldt, S.-K. Son, T. Mazza, A. Rörig, R. Boll, M. Meyer, A. C. LaForge, D. Mishra, N. Berrah, R. Santra, Machine-learning calibration of intense x-ray free-electron-laser pulses using bayesian optimization, Phys. Rev. Res. 5 (2023) 023114.

[73] S. R. Chitturi, Z. Ji, A. N. Petsch, C. Peng, Z. Chen, R. Plumley, M. Dunne, S. Mardanya, S. Chowdhury, H. Chen, et al., Capturing dynamical correlations using implicit neural representations, Nature Communications 14 (2023) 5852, *Data-driven analysis tool that enables more efficient and informative experimental design.

[74] K. Liebl, M. Zacharias, The development of nucleic acids force fields: From an unchallenged past to a competitive future 122 (2023) 2841–2851.

[75] F. Alexander, A. Almgren, J. Bell, A. Bhattacharjee, J. Chen, P. Colella, D. Daniel, J. DeSlippe, L. Diachin, E. Draeger, et al., Exascale applications: skin in the game, Phil. Trans. Roy. Soc. A 378 (2166) (2020) 20190056.

[76] D. Mendez, J. M. Holton, A. Y. Lyubimov, S. Hollatz, I. I. Mathews, A. Cichosz, V. Martirosyan, T. Zeng, R. Stofer, R. Liu, J. Song, S. McPhillips, M. Soltis, A. E. Cohen, Deep residual networks for crystallography trained on synthetic data, Acta Crystallographica Section D 80 (1) (2024) 26–43. doi:10.1107/S2059798323010586.

[77] P. Gasparotto, L. Barba, H.-C. Stadler, G. Assmann, H. Mendonça, A. Ashton, M. Janousch, F. Leonarski, B. Béjar, TORO Indexer: A PyTorch-Based Indexing Algorithm for kHz Serial Crystallography, ChemRxiv (2023).

[78] J. B. Thayer, G. Carini, W. Kroeger, C. O'Grady, A. Perazzo, M. Shankar, M. Weaver, Building a data system for lcls-ii, in: 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE, 2017, pp. 1–4.

[79] M. Hadian-Jazi, A. Sadri, A. Barty, O. Yefanov, M. Galchenkova, D. Oberthuer, D. Komadina, W. Brehm, H. Kirkwood, G. Mills, et al., Data reduction for serial crystallography using a robust peak finder, Journal of Applied Crystallography 54 (5) (2021) 1360–1378.