High-Brightness Self-seeded X-ray Free Electron Laser to Precisely Map Macromolecular Structure

Heung-Sik Kang  
Pohang Accelerator Laboratory  
https://orcid.org/0000-0002-9861-6895

Chang-Ki Min  
Pohang Accelerator Laboratory

Inhyuk Nam  
Pohang Accelerator Laboratory  
https://orcid.org/0000-0002-6650-7854

Bonggi Oh  
Pohang Accelerator Laboratory  
https://orcid.org/0000-0001-7577-9563

Gyujin Kim  
Pohang Accelerator Laboratory

Donghyun Na  
Pohang Accelerator Laboratory

Young Jin Suh  
Pohang Accelerator Laboratory

Haeryong Yang  
Pohang Accelerator Laboratory

Myung-Hoon Cho  
Pohang Accelerator Laboratory

Changbum Kim  
Pohang Accelerator Laboratory

Min Jae Kim  
Pohang Accelerator Laboratory

Chihyun Shim  
Pohang Accelerator Laboratory

Jun Ho Ko  
Pohang Accelerator Laboratory

Hoon Heo  
Pohang Accelerator Laboratory

Jaehyun Park  
Pohang Accelerator Laboratory, POSTECH  
https://orcid.org/0000-0002-8321-026X

Jangwoo Kim  
Pohang Accelerator Laboratory

Sehan Park  
Pohang Accelerator Laboratory

Gisu Park  
Pohang Accelerator Laboratory

Seonghan Kim  
Pohang Accelerator Laboratory

Sae Hwan Chun  
Pohang Accelerator Laboratory

HyoJung Hyun  
Pohang Accelerator Laboratory

Jae Hyuk Lee  
Pohang Accelerator Laboratory  
https://orcid.org/0000-0002-8186-0226

Kyung Sook Kim
Keywords: hard-X-ray self-seeded (HXRSS) free-electron laser (FEL), femtosecond crystallography (SFX), macromolecular structure, lysozyme protein

DOI: https://doi.org/10.21203/rs.3.rs-82091/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

We demonstrate a hard-X-ray self-seeded (HXRSS) free-electron laser (FEL) at Pohang Accelerator Laboratory with an unprecedented peak brightness \(3.2 \times 10^{35}\) photons/(s·mm\(^2\)·mrad\(^2\)·0.1%BW). The self-seeded FEL generates hard X-ray pulses with improved spectral purity; the average pulse energy was 0.85 mJ at 9.7 keV, almost as high as in SASE mode; the bandwidth (0.19 eV) is about 1/70 as wide, the peak spectral brightness is 40 times higher than in self-amplified spontaneous emission (SASE) mode, and the stability is excellent with > 94% of shots exceeding the average SASE intensity. Using this self-seeded XFEL, we conducted serial femtosecond crystallography (SFX) experiments to map the structure of lysozyme protein; data-quality metrics such as \(R_{	ext{split}}\), multiplicity, and signal-to-noise ratio for the SFX were substantially increased. We precisely map out the structure of lysozyme protein with substantially better statistics for the diffraction data and significantly sharper electron density maps compared to maps obtained using SASE mode.

Introduction

The extreme peak brightness and ultrashort pulses provided by X-ray free-electron lasers (XFEL)\(^1\)–\(^5\) allow data collection from micrometer-sized protein crystals at room temperature (the functional temperature of their constituent molecules) while outrunning radiation damage. This 'diffraction-before-destruction approach' has been applied in serial femtosecond crystallography (SFX), which has revolutionized X-ray crystallography and has been considered an important tool to determine the structure of proteins that are difficult to crystallize\(^5\)–\(^9\).

XFELs have noisy and spiky spectra because the devices exploit the self-amplified spontaneous emission (SASE) that starts from the electron beam shot noise. The fluctuation of the noisy and spiky spectra of the XFEL can limit the data quality of SFX. Self-seeding is a promising approach to overcome the deficiencies of XFELs and to realize bright, fully-coherent FEL sources in the hard X-ray domain. The use of seeded FEL pulses with their higher reproducibility and 'cleaner' spectrum than SASE might accelerate convergence of the merged reflection intensities of the SFX data\(^10\). However, existing hard X-ray self-seeded (HXRSS) FELs have limited radiation pulse energy and spectral brightness, and inadequate stability. A previous study\(^11\) of the self-seeded XFEL for SFX did not show any improvement in the data-quality metrics of the SFX compared to SASE, in contrast to the expectation that the use of self-seeded pulses might result in SFX data of a superior quality to that collected using SASE pulses.

The idea of self-seeding\(^12\) has been proposed to overcome the limitation of SASE FEL, and later a self-seeding scheme with a four-crystal monochromator in Bragg reflection geometry was proposed\(^13\). A more-compact self-seeding scheme\(^14\) that uses a single-crystal monochromator was proposed for the hard X-ray region; this design exploits the phenomenon that the forward Bragg-diffracted monochromatic beam has a small delay time, which allows a very short (<5 m) chicane. The forward Bragg diffraction (FBD) through a thin diamond crystal produces a train of monochromatic wakes that trail the main X-ray pulse by a few tens of femtoseconds\(^15\)–\(^16\). By using a magnetic chicane to detour the electron bunch (bunch) so that it and the wake overlap in time, the monochromatic seed signal can be amplified in the downstream undulators.

The first successful demonstration of an HXRSS FEL using an FBD monochromator was performed out at the Linac Coherent Light Source (LCLS), produced 8.3-keV X-ray pulses with a bandwidth of 0.4–0.5 eV that was about 1/40–1/50 as wide as the SASE bandwidth\(^17\). The average pulse intensity of the HXRSS FEL pulses at the LCLS was 573 ± 290 µJ at 5.5 keV, and intensity fluctuation was ~50\%\(^18\). The average peak spectral intensity was 1.7 times larger than in SASE mode. In the self-seeding experiments at LCLS, including a soft X-ray self-seeding, the radiation spectrum often showed a pedestal-like distribution around the seeded frequency; this distribution limits spectral brightness\(^17\)–\(^19\)–\(^20\). The pedestals originate in longitudinal phase space modulations produced by microbunching instability (MBI) upstream of the undulators.

The problem of the large delays in Bragg-reflection monochromators\(^13\) was overcome, and self-seeding using Bragg reflections was demonstrated at the SPring-8 Angstrom Compact Free Electron Laser (SACLA); the design used a channel-cut Si crystal monochromator with a tiny gap of 90 µm\(^21\). The seed, with a bandwidth of 1.3 eV (full-width at half-maximum, FWHM) at 9.85 keV, was filtered from the SASE radiation by the 111 Bragg reflections from the channel-cut crystal. The average pulse energy of the self-seeded XFEL at 9.85 keV was 450 µJ, vs. 780 µJ in SASE mode. The peak spectral intensity of the self-seeded XFEL pulses was six times higher than in SASE mode. However, the X-ray pulses had a relatively large bandwidth of 3 eV (FWHM) because of the large bandwidth of the seed signal as well as an energy chirp in the e-bunch. Using the Si-220 Bragg reflections from the channel-cut crystal, the bandwidth was reduced to 0.6 eV (FWHM) at 9.0 keV with an average pulse energy of ~250 µJ\(^22\).

At PAL-XFEL we demonstrated an HXRSS FEL using an FBD monochromator (Fig. 1) that has favorable source features compared to SASE mode: a high average peak spectral intensity exceeded that of SASE by a factor of 12 in self-seeded mode, pulse energy of 0.85 mJ at 9.7 keV; substantial improvements in the stability of self-seeding; and substantially suppressed pedestal effects. We also demonstrated that
Improving The Spectral Brightness Of Self-seeded Xfel

To increase spectral brightness, we used e-bunches with a higher charge (180 pC) and a longer duration (42 fs FWHM) than in the previous study\textsuperscript{17,24}. The e-bunch energy is 8.538 GeV; the undulator parameter is 1.87; and the duration of the SASE FEL radiation pulse is 20 fs (FWHM), as measured using the cross-correlation method\textsuperscript{23,24}. A laser heater (LH) was used to suppress MBI (Methods Section).

One prerequisite to generate X-rays that have narrow bandwidth (narrowband) and high spectral brightness for the HXRSS FELs is a narrowband seed\textsuperscript{25}. For efficient seeding, the bandwidth of the monochromatic seed must match the FEL bandwidth, and the duration of the monochromatic wake should be comparable to or longer than that of the SASE signal for seeding. To generate such a seed, we use high-index Bragg reflections (ie., 33-3 or 115) with an FBD bandwidth of ~0.1 eV, instead of the 004 reflection with a ~0.3-eV bandwidth, as typically used in short-pulse modes\textsuperscript{17,24}. Unlike the self-seeding in reflection geometry\textsuperscript{21,22}, the path length delay of the seed pulse does not depend on the crystal index, so the high-index Bragg reflections, like 33-3 or 115, can be used to generate a seed bandwidth < 0.1 eV. For a better overlap with long e-bunch and a higher seed intensity, we use the 0\textsuperscript{th}-order wake of the FBD signal as the monochromatic seed instead of the 1\textsuperscript{st}-order wake that has been used in previous experiments (Fig. 1f) (details in ref. [25] and Supplementary Fig. 1). The duration of the 0\textsuperscript{th}-order wake, in this case, is sufficiently long (~65 fs) to allow for sufficient delay of the e-bunch with respect to the SASE pulse, and full separation of the seed signal from the SASE background (Supplementary Fig. 2).

Another prerequisite to increase spectral brightness is to suppress the pedestal-like distribution around the central seed frequency. This effect originates in the MBI induced by bunch compression, which creates detrimental sideband modulation of the e-bunch. Once the sidebands are generated, the electron oscillations are driven by the multiple-frequency ponderomotive potential. As a result, the efficiency of FEL generation at the carrier frequency is reduced, and the spectral quality is degraded by diversion of radiation power into sideband frequencies\textsuperscript{26,27}. A laser heater can efficiently suppress MBI in both SXRSS\textsuperscript{20} and SASE mode;\textsuperscript{28-31} however, for the HXRSS, the improvement of spectral brightness by using a laser heater has not been investigated experimentally.

The LH can suppress the MBI and increase the peak spectral intensity in self-seeded mode (Figure 2). The pedestal around the center peak is significantly reduced as slice energy spread increases, so the peak intensity increases (Fig. 2a). These results show that the sideband amplification due to the MBI is effectively suppressed, so the main peak is solely amplified. The fraction of FEL intensity enclosed within the bandwidth shows that spectral purity is significantly increased using the LH (Fig. 2b). As the slice energy spread increases, the peak intensity (solid red line) also increases (Fig. 2c); it reaches its maximum when the slice energy spread is ~27 keV. The optimal slice energy spread for self-seeded mode is about 5 keV higher than for SASE.\textsuperscript{32} To suppress pedestal effects due to microbunching instability substantially, a calm longitudinal-phase-space with energy modulation further suppression is required (Fig. 1e). However, the total sum of the spectrum (blue dotted line in Fig. 2c) remains almost constant until the LH = 27 keV; this result supports the hypothesis that unsuppressed MBIs channel the radiation power into the sidebands.

Single-shot spectra maps (Fig. 3a) were obtained using a 0.26-eV resolution Si(333) curved-crystal single-shot spectrometer\textsuperscript{35} (Supplementary Fig. 3) for 9.7-keV X-rays in SASE and self-seeded modes. The measured bandwidths of X-rays were 13.0 ± 0.1 eV in SASE and 0.35 ± 0.01 eV in self-seeded mode (Fig. 3b); the latter dropped to 0.24 eV after deconvolution from the spectrometer resolution. More-accurate measurements than these were obtained using a 0.09-eV-resolution Si(333) flat-crystal scanning spectrometer; they reveal a time-averaged bandwidth of 0.21 ± 0.01 eV in self-seeded mode, which drops to 0.19 eV after deconvolution (Fig. 3c). The FBD seed bandwidth in the 33-3 Bragg reflection from the diamond crystal is 0.06 eV (Supplementary Table 1), but the resultant bandwidth of the self-seeded XFEL increased to 0.19 eV because of the energy chirp of the e-bunch. Assuming the same pulse duration in the self-seeded mode as the SASE FEL radiation pulse (20 fs), and Gaussian pulse shape, the Fourier-transform-limited HXRSS FEL radiation bandwidth should be ~ 0.1 eV. The single-shot pulse bandwidth is definitely smaller than the 0.19-eV averaged bandwidth, so the PAL-XFEL HXRSS pulses are less than a factor of two larger than the Fourier-transform limit.

Self-seeded mode had 12 times higher average peak spectral intensity than SASE mode (Fig. 3b), but this number is limited by the spectral resolution of the single-shot spectrometer. The average pulse energy of the HXRSS at PAL-XFEL is 0.85 mJ, or 57% of the 1.5-mJ average pulse energy in SASE mode, as measured by the electron energy loss scan\textsuperscript{34}. Appropriate undulator tapering was applied for 20 undulators (Supplementary Fig. 4). The ratio of the integrated spectral area of the single-shot spectrometer for SASE to self-seeded mode is 1.64; this ratio is consistent with the XFEL intensity ratio of 1.76 (1.5 mJ to 0.85 mJ) that was measured by the electron energy loss scan. Overall, the
peak brightness of the PAL-XFEL HXRSS FEL is calculated to be $3.2 \times 10^{35}$ photons/(s·mm$^2$·mrad$^2$·0.1%BW), which is 40 times higher than that of SASE, the highest achieved to date.

The radiation pulse energy of the PAL-XFEL HXRSS is both high and very stable. The self-seeded mode has a consistently higher intensity than SASE mode (Fig. 3d). In self-seeded mode, > 94% of the shots have an intensity higher than the average SASE intensity (i.e., > 1 a.u.). Such seeding stability is mainly due to the stability of the PAL-XFEL, which has a very small shot-to-shot electron-energy jitter of 0.012% (r.m.s.)$^{24,31}$. The resultant shot-to-shot fluctuation of the central radiation wavelength of SASE FEL was measured to be 0.025% (r.m.s.), which is one-half the Pierce parameter $\rho = 5 \times 10^4$ (Relative SASE bandwidth, $\sim 5.6 \times 10^4$), so self-seeded pulses are almost always amplified.

Serial Femtosecond Crystallography With A Self-seeded Xfel

To solve a structure for the SFX, the necessary number of indexed snapshot patterns of crystals depends on the SNR of the individual patterns, the symmetry of the crystal, and the variability of parameters on which the diffraction depends from shot to shot (such as the chaotic spectrum of FEL pulses)$^6$. These factors influence the final accuracy of the merged data. To determine de novo the structure of a protein in which no homologous structures exist, the experimental phasing of SFX data, the data must have high resolution and a very high multiplicity of data sets for phase determination$^{7,10,35-38}$. The large shot-by-shot variations in X-ray intensity and photon energy may make experimental phasing of XFEL data very challenging.

We conducted test of self-seeded XFEL for SFX, because the self-seeded XFEL that we achieved performs extremely well. A previous did not show any difference in the data quality metrics of the SFX compared to SASE$^{11}$, but the peak spectral brightness of our XFEL is about ten times higher compared to the XFEL used previously and 40 times higher than SASE, with excellent stability. We expect that reduction in the relative bandwidth from $\Delta E/E = 1.3 \times 10^{-3}$ (SASE) to $\Delta E/E = 1.9 \times 10^{-5}$ (SS) will sharpen diffraction patterns, especially those collected at large scattering angles, which are responsible for increasing the resolution. Also, we expect an increase in filtration rate of raw data owing to the higher spectral intensity of the self-seeded XFEL compared to SASE.

We performed a demonstration experiment by mapping out the three-dimensional structure of the lysozyme from chicken eggwhite and performing a comparative analysis of the results obtained using the narrowband HXRSS FEL and the broadband SASE FEL (see Methods for the crystal preparation and experimental conditions).

We collected and processed three data sets that had different numbers of images for both self-seeded and SASE modes: SS1/SASE1 (111,467/101,443), SS2/SASE2 (38,510/38,686), and SS3/SASE3 (20,209/20,530). The indexing rates were substantial in all cases. For example, SS1; 70,656 crystal diffraction patterns (63.4%) were identified as crystal hits, and 33,663 of them were indexed (47.6%). The index rates of the self-seeding data sets were higher than those of the SASE data sets (Table 1). SFX data quality metrics such as SNR (or $I/\sigma$), multiplicity, $R_{\text{split}}$ (i.e., the consistency of merged intensity distributions between two half-datasets separated from the full dataset), and correlation coefficient $[C^{4}]$ strongly depend on the number of images, as is known (Fig. 4, Supplementary Table 2). However, the self-seeding data shows superior metrics than the SASE data at high resolutions, unlike a previous report$^{11}$. Remarkably, the self-seeding data sets had twice the multiplicity of the SASE data set at all resolutions (Fig. 4b), so the final accuracy of the merged data is improved, even with the same number of hit images (see Methods for SFX data processing).

Table 1: Statistics of data collection, phasing, and model refinement for three sets for self-seeded (SS1, SS2, and SS3) and SASE (SASE1, SASE2, and SASE3) modes. The models of self-seeded mode (SS1 and SS2) and SASE mode (SASE1 and SASE2) have been refined from 38.8 to 1.75 Å except for SS3 and SASE3 (38.8 to 1.85 Å). The higher-resolution shells (1.93-1.85 Å) for SS3 and SASE3 must be < 1.85 Å for validity. All models have one monomer in the asymmetric unit and adopt nearly identical structures, with r.m.s. deviations ~0.05 Å for 129 Ca atom pairs.
After refinement, when we compared the models with their structure maps (SS1 and SASE1), we found apparent improvements in 2
without bias or error, we performed structural determination using the same numbers of hit images for self-seeding and SASE data sets (see Methods for structure determination, refinement, and analysis). To compare and analyze the structures and their electron density maps,

| Data sets | SS1 | SASE1 | SS2 | SASE2 | SS3 | SASE3 |
|-----------|-----|-------|-----|-------|-----|-------|
| **A. Data collection** |     |       |     |       |     |       |
| Space group | P4₃2₁2 | P4₃2₁2 | P4₃2₁2 | P4₃2₁2 | P4₃2₁2 | P4₃2₁2 |
| Unit cell | a = 77.56, b = 77.56, c = 77.88, α = 90.0, β = 90.0, γ = 90.0 | a = 77.56, b = 77.88, c = 77.88, α = 90.0, β = 90.0, γ = 90.0 | a = 77.88, b = 77.88, c = 77.88, α = 90.0, β = 90.0, γ = 90.0 | a = 77.88, b = 77.88, c = 77.88, α = 90.0, β = 90.0, γ = 90.0 |
| Unit cell length (Å) | 37.32 | 37.32 | 37.32 | 37.32 | 37.32 | 37.32 |
| Unit cell angle (°) | α = 90.0, β = 90.0, γ = 90.0 | α = 90.0, β = 90.0, γ = 90.0 | α = 90.0, β = 90.0, γ = 90.0 | α = 90.0, β = 90.0, γ = 90.0 |
| X-ray wavelength (Å) | 1.2782 | 1.2782 | 1.2782 | 1.2782 | 1.2782 | 1.2782 |
| Number of collected images | 111,467 | 101,443 | 38,510 | 38,686 | 20,209 | 20,530 |
| Number of merged images | 70,656 | 70,656 | 27,926 | 27,926 | 12,377 | 12,377 |
| Number of indexed images | 33,663 | 28,301 | 14,256 | 11,809 | 7,091 | 5,686 |
| Indexing rate from hits (%) | 47.64 | 40.05 | 51.05 | 42.29 | 57.29 | 45.94 |
| Number of merged images | 33,663 | 28,301 | 27,926 | 27,926 | 12,377 | 12,377 |
| Resolution range (Å) | 38.8-1.75 | 38.8-1.75 | 38.8-1.75 | 38.8-1.75 | 38.8-1.85 | 38.8-1.85 |
| Total / unique reflections | 6,274,437 | 3,786,572 | 2,555,216 | 1,593,005 | 1,213,951 | 753,928 |
| Multiplicity | 269.2 (189.6) | 162.4 (108.9) | 108.0 (73.6) | 55.9 (37.5) | 49.9 (35.4) | 28.2 (19.7) |
| Completeness | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) |
| (%) | 94 / 30.1 | 65 / 38.6 | 62 / 36.4 | 89 / 31.4 | 95 / 31.4 | 88 / 37.5 |
| CC(a,b) | 0.992 (0.944) | 0.993 (0.855) | 0.979 (0.824) | 0.981 (0.353) | 0.971 (0.799) | 0.963 (0.523) |
| <I/Io>(a) | 6.2 (1.77) | 5.6 (1.29) | 4.0 (1.17) | 3.1 (0.22) | 2.9 (1.17) | 2.3 (0.33) |
| Rsplit (%) | 12.8 (52.5) | 12.7 (74.0) | 20.9 (79.2) | 21.3 (100) | 25.0 (68.9) | 30.5 (100) |
| Rwork / Rfree (%) | 20.7 / 23.2 | 20.9 / 24.5 | 22.4 / 24.4 | 21.4 / 25.9 | 22.1 / 24.7 | 21.6 / 26.4 |

B. Model refinement

| Protein | 1,001 / 19.8 | 1,001 / 29.6 | 1,001 / 21.3 | 1,001 / 30.2 | 1,001 / 20.0 | 1,001 / 27.2 |
| Water | 94 / 30.1 | 62 / 36.4 | 89 / 31.4 | 65 / 38.6 | 95 / 31.4 | 88 / 37.5 |
| Bond lengths (Å) | 0.009 / 0.927 | 0.008 / 0.917 | 0.010 / 1.137 | 0.007 / 0.875 | 0.011 / 1.293 | 0.007 / 0.887 |
| PDB code | 7BYO | 7BYP | 7D01 | 7D02 | 7D04 | 7D05 |
| Ramachandran plot (%) | Favoured | 99.2 / 0.0 | 99.2 / 0.0 | 96.85 / 0.0 | 98.4 / 0.0 | 98.4 / 0.0 | 99.21 / 0.0 |
| Outliers | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Rotamer outliers | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

* Values in parentheses refer to the highest-resolution shell (1.75-1.78 Å).
* CC* = √2CC.L/(1 + CC.L). ²
* Rsplit = 1 / √2. SSD | w getAll | 1 / √2.SSD | w only getAll |
* R = Σ | |Fo| - |Fc| | / Σ | |Fo|, where Rfree and Rwork are calculated for a randomly-chosen 5% of reflections that were not used for refinement or for the remaining reflections, respectively.

For quality assessment, we performed molecular replacement (MR) with the Phaser-MR in PHENIX, ⁴⁰ using a model of lysozyme (Protein Data Bank code 1VDS) as a search model, then conducted atomic model refinement using phenix.refine, then inspected of (mFo-DFc) omit maps, ⁴⁰ (see Methods for structure determination, refinement, and analysis). To compare and analyze the structures and their electron density maps without bias or error, we performed structural determination using the same numbers of hit images for self-seeding and SASE data sets (SS1/SASE1, SS2/SASE2, and SS3/SASE3).

After refinement, when we compared the models with their structure maps (SS1 and SASE1), we found apparent improvements in 2mFo-DFc maps of the self-seeded mode (Fig. 5a), even though lysozyme is a globular protein and has some buried residues that strongly interact with other residues. To get a much better view, we obtained bias-free mFo-DFc omit maps by sorting out the residues (Fig. 5b). Comparison of the
mFo-DFc omit maps at 1.75 Å resolution (Fig. 5b) clearly shows that the maps of the ten residues (Phe21/Ala28/Tyr41/Trp46/Phe52/Asn62/Tyr71/Trp81/Trp126/Trp141) are not blurred in self-seeded mode; the maps, including the side chains and the main chains (carboxyl groups, nitrogens on the peptide backbones, and α-carbons), are sharper than those obtained in SASE mode. For instance, in the Phe21 and Asn62 maps, β-carbons and side chains are revealed clearly only in self-seeded mode. Refined models without a specific residue were deleted from the original structure (Supplementary Table 3).

Comparative analysis of the mFo-DFc electron density maps of the ten residues reveals the superiority of the self-seeded data set over the SASE mode data sets (Table 2, Supplementary Fig. 5). For example, even though the data-quality metrics of the SS3 data are inferior to those of the SASE1 (SS3 dataset has one-fourth as many indexed images as the SASE1), the omit maps of the ten residues from the SS3 data are better than those from the SASE data. B-factors are crystallographic parameters to explain this big difference. The average B-factors of both protein and solvent waters models are relatively lower in the models from the self-seeded than in those from SASE mode, and the average B-factors are independent of the number of indexed images (Table 1: Model refinement). These traits indicate that the atomic displacement fluctuations are relatively weaker when a narrowband self-seeded FEL is used, than when a broadband SASE FEL is used. The reduced fluctuations might help increase the refinement of the model with sharpened electron density maps. The overall sharpening of the omit maps obtained from the self-seeding data resulted from phasing-quality data with fewer patterns. The high quality of data obtained in self-seeded mode is a result of the use of recurrent shots from a highly-stable self-seeded XFEL.

**Conclusion**

The PAL-XFEL HXRSS successfully demonstrated a forward Bragg diffraction self-seeded XFEL with unprecedented peak brightness \((3.2 \times 10^{35}\) photons/(s·mm\(^2\)·mrad\(^2\)·0.1%BW)) and stability; the average pulse energy is 0.85 mJ at 9.7 keV, the bandwidth (0.19 eV) is about 1/70 as wide, the peak spectral brightness is 40 times higher, and the stability is excellent with >94% of shots exceeding the average SASE intensity. We used high-index Bragg reflections (33 - 3 or 115) to exploit a narrow seed bandwidth < 0.1 eV and a long wake duration ~ 65 fs. A calm longitudinal-phase-space with energy modulation further suppressed is required to suppress pedestal effects due to microbunching instability.
substantially. We demonstrated that high-spectral-intensity and high stability self-seeded XFEL improves the data-quality metrics of SFX: it achieves outstanding quality in the SNR, multiplicity, CC*, and \( R_{\text{split}} \) compared to the large-bandwidth SASE. The high multiplicity of the self-seeding data sets yields phasing-quality data with fewer patterns than in SASE datasets, and improve the refinement of the model with sharpened electron density maps. The self-seeded data set achieves superior quality of electron-density map over the SASE mode data sets. Even with one-fourth of the indexed images of the SASE data set, the self-seeded data set shows a better or similar electron density maps for the residues. The improved structure map by the self-seeded XFEL indicates that high brightness narrowband XFEL increases the resolution of signal collection, and helps to solve three-dimensional macromolecular structures with high resolution, especially for a very small crystal.

**Methods**

**Laser heater to suppress MBI.** The laser heater adds a slice energy spread to the 150-MeV e-bunch. The IR laser beam size is comparable to that of the electron bunch, so the energy spread distribution assumes a super-Gaussian profile that can effectively suppress the MBI. The induced slice-energy spread of the e-bunch at LH as a function of the IR laser energy was measured using a transverse deflector and an energy spectrometer located after the first bunch compressor. The accelerating sections (L1 and XLIN) and the bunch compressor BC1 downstream of the laser heater were all turned off (Fig. 1). The longitudinal phase space of e-bunch was simulated for three cases of the laser heater condition, where no laser heater case (top), optimized cases for SASE (middle), and self-seeding (bottom) (Fig. 1e). The spectral purity of the self-seeded FEL is very sensitive to the energy modulation of the e-bunch, so the optimal condition of laser heater for self-seeding is different from that for SASE. The CSR (green dotted line in Fig. 2c) measured at the third bunch compressor using a visible CCD camera is due to the MBI; this result shows that MBI should be more suppressed by the laser heater for the self-seeding than for SASE.

**Crystal sample preparation.** As a reference sample, we purchased lysozyme from chicken eggwhite from Hampton Research (HR7-110). The lysozyme microcrystals were obtained by mixing 10 ml of protein solution (30 mg/ml protein concentration in 100 mM sodium acetate, pH 3.0) and 10 ml of crystallization solution (100 mM sodium acetate [pH 4.0], 6% [w/v] polyethylene glycol 8,000, and 3.5 M NaCl) at 291 K. Crystals were produced immediately; they were kept at 291 K before the experiment. The average size of lysozyme crystals ranged from 15 to 25 \( \mu \)m in one dimension. Crystal density was \( \approx 2.5 \times 10^7 \) per ml, as measured using a hemocytometer (Marienfeld). Before the sample was loaded into a lipidic cubic phase (LCP) injector, the lysozyme protein samples were mixed in a 1:1 volume ratio with monoolein by using a syringe mixer. For each loading, 40 \( \mu \)l of a sample (composed of 20 \( \mu \)l monoolein and 20 \( \mu \)l lysozyme crystal) was used to perform LCP injector sample delivery. The flow rate of the monoolein-mixed crystal sample was 0.032 ml/min. All chemical reagents were purchased from Hampton or Sigma-Aldrich.

**Experimental conditions.** SFX experiments using both self-seeded XFEL (0.85 mJ) and SASE (1.5 mJ) modes were performed at the NCI PAL-XFEL experimental station. The distances from the expected XFEL source position in the undulator to the centers of the vertical and horizontal focusing mirrors (VFM and HFM) were 147.1 m and 141.3 m, respectively. The focal lengths of the VFM and HFM are 5.99 m and 5.36 m, respectively. The predicted source size and divergence were \( \approx 37.9 \) mm (FWHM) and \( \approx 1.7 \) mrad (FWHM), respectively in both self-seeded FEL and SASE modes, the X-ray pulse was focused to a beam size of 2.5 \( \mu \)m (horizontal) \( \times 2.5 \mu \)m (vertical) (FWHM) using a Kirkpatrick–Baez mirror. The diffraction data were collected using an MX225-HS detector with a 4\( \times \)4 binning mode (pixel size: 156 \( \mu \)m\( \times \)156 \( \mu \)m) at room temperature and monitored by OnDa. The 4 \( \times \)4 binning mode was used to match the XFEL repetition rate of 30 Hz. The distance between the sample position and the detector was 111 mm and was validated by comparing the index rate of each data set. Aside from using the self-seeded FEL or SASE mode, all other conditions were identical in both experiments. We collected six data sets: three in self-seeded mode (SS1, SS2, and SS3) and three in SASE mode (SASE1, SASE2, and SASE3).

**SFX data processing.** After data collection, the hit images were filtered using Cheetah including min-snr of 4.0. The pre-processed images were further indexed, integrated, merged, and post-refined using CrystFEL (version 0.6.3). The experimental geometry was also refined for CrystFEL. Indexing was performed using DirAx (version 1.17) with peak integration parameters of int-radius = 3, 4, 5. The measured diffraction intensities were merged with process_hkl in the CrystFEL suite. To investigate the data statistics from the two modes (self-seeded and SASE) carefully, we processed data with the same methods and the same parameters for consistency.

**Structure determination, refinement, and analysis.** The structure of lysozyme from chicken eggwhite was determined (Table 1) by the molecular replacement method using the Phaser-MR in PHENIX (version 1.14-3260) using a model of lysozyme (Protein Data Bank code 1VDS) as a search model. During the calculation of molecular replacement, we excluded water molecules from the template model to avoid model bias. Water molecules were inspected and added manually using Coot (version 0.8.9) by reference to mFo–DFc maps. Water molecules were placed in correct positions depending on the density map where positive peaks higher than 1.5\( \sigma \) and 3.0\( \sigma \) occurred in the 2mFo-DFc map and mFo-DFc map, respectively. The molecular replacement model was first refined with a rigid-body protocol and Cartesian
simulated annealing (starting at 5,000 K) using phenix.refine to reduce model bias. After five cycles of restrained refinement, the model was evaluated by MolProbity (version 4.4)\textsuperscript{51}. The data of lysozyme crystals from both self-seeded and SASE modes belonged to the tetragonal space group \(P4_321\), with unit cell parameters of \(a = b = 77.56\text{ Å}, c = 37.32\text{ Å}, \alpha = \beta = \gamma = 90^\circ\). To inspect effects on map quality using the self-seeded mode, we made all of the omit maps on residues of the lysozyme model excluding glycine, which cannot present meaningful maps. Therefore, we manually deleted each residue from the lysozyme model and performed phenix.refine in PHENIX\textsuperscript{40} to generate an \(mFo-DFc\) map on each residue (Supplementary Table 2). Omit maps were generated to reduce the possible effect of model bias. Six models of the self-seeded and SASE modes were calculated in the same manner for a fair comparison.

Data availability

The coordinates and structural factors have been deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) under the accession codes 7BYO/7D01/7D04 (for lysozyme from self-seeded mode) and 7BYP/7D02/7D05 (for lysozyme from SASE mode).

Declarations

Acknowledgments

The authors gratefully acknowledge useful discussions with A. A. Lutman, F.-J. Decker, Z. Huang, and H. Loos. The XFEL experiments were performed at the NCI PAL-XFEL experimental station. This research has been supported by the Ministry of Science and ICT of Korea (Grant No. 2018R1A6B4023605) and in part by the Basic Science Research Program (Grant No. 2017R1A2B4007274; 2020R1F1A1075828; 2019R1A6B2A01081034; 2019R1C1C1003687) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT of Korea.

Author contributions

All authors designed, constructed, and tested the accelerator and X-ray systems, performed the experiments, and analyzed the data.

Additional information

The authors have no competing financial interests. Reprints and permission information are available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to H.-S. Kang.

References

1. Emma, P. \textit{et al.} First lasing and operation of an ångström-wavelength free-electron laser. \textit{Photon.} \textbf{4}, 641–647 (2010).
2. Tanaka, H. \textit{et al.} A compact X-ray free-electron laser emitting in the sub-ångström region. \textit{Photon.} \textbf{6}, 540–544 (2012).
3. Kang, H.-S. \textit{et al.} Hard X-ray free-electron laser with femtosecond-scale timing jitter. \textit{Photon.} \textbf{11}, 708–713 (2017).
4. Decking \textit{et al.} A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator, superconducting linear accelerator, \textit{Nature Photonics} \textbf{14}, 391–397 (2020)
5. Milne, C. \textit{et al.} SwissFEL: The Swiss X-ray Free Electron Laser. \textit{Sci.} \textbf{7}, 220 (2017).
6. Chapman, H.N. X-ray Free-Electron Lasers for the Structure and Dynamics of Macromolecules. Annu. Rev. Biochem. \textbf{88}, 35-58 (2019).
7. Fromme, P XFELs open a new era in structural chemical biology. \textit{Nature Chemical Biology} \textbf{11}, 895-899, doi:10.1038/nchembio.1968 (2015).
8. Boutet, S. \textit{et al.} High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. \textit{Science} \textbf{337}, 362-364, doi:10.1126/science.1217737 (2012).
9. Kern, J. \textit{et al.} Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. \textit{Science} \textbf{340}, 491–495 (2013).
10. Barends, R.M., Foucar, L., Botha, S., Doak, R.B., Shoeman, R.L., Nass, K., Koglin, J.E., Williams, G.J., Boutet, S., Messerschmidt, M. & Schlichting, I,\textit{De novo} protein crystal structure determination from X-ray free-electron laser data, \textit{Nature} \textbf{505}, 244–247 (2014).
11. Barends, T. White, T.A., Barty, A., Foucar, L., Messerschmidt, M., Alonso-Mori, R., \textit{et al.} Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data. \textit{Journal of Synchrotron Radiation} \textbf{22}(3), 644–652 (2015).
12. Feldhaus, J. \textit{et al.} Possible application of X-ray optical elements for reducing the spectral bandwidth of an X-ray SASE FEL. \textit{Comm.} \textbf{140}, 341–352 (1997).
13. Saldin, E. L., Schneidmiller, E. A., Shvyd’ko, Yu.V. & Yurkov, M. V. X-ray FEL with a meV bandwidth. *Instrum. Methods Phys. Res. A* **475**, 357–362 (2001).

14. Geloni, G., Kocharyan, V., & Saldin, E. A novel self-seeding scheme for hard X-ray FELs. *Mod. Opt.* **58**, 1391–1403 (2011).

15. Shvyd’ko, Yu. V. & Lindberg, R. R. Spatiotemporal response of crystals in X-ray Bragg diffraction. *Rev. ST Accel. Beams* **15**, 100702 (2012).

16. Lindberg, R. R. & Shvyd’ko, Yu. V. Time dependence of Bragg forward scattering and self-seeding of hard X-ray free-electron lasers. *Rev. Spec. Top. Accel. Beams* **15**, 050706 (2012).

17. Amann, J. *et al.* Demonstration of self-seeding in a hard-X-ray free-electron laser. *Photon.* **6**, 693–698 (2012).

18. Emma, C. *et al.* Experimental demonstration of fresh bunch self-seeding in an X-ray free electron laser. *Phys. Lett.* **110**, 154101 (2017).

19. Ratner, D. *et al.* Experimental demonstration of a soft X-ray self-seeded free-electron laser. *Rev. Lett.* **114**, 054801 (2015).

20. Marcus, G. *et al.*, Experimental observations of seed growth and accompanying pedestal, *Rev. Accel. Beams* **22**, 080702 (2019).

21. Inoue, I. *et al.* Generation of narrow-band X-ray free-electron laser via reflection self-seeding. *Photon.* **13**, 319–322 (2019).

22. Matsumura, S., Osaka, T., Inoue, I., Matsuyama, S., Yabashi, M., Yamauchi, K., and Sano, Y., High-resolution micro channel-cut crystal monochromator processed by plasma chemical vaporization machining for a reflection self-seeded X-ray free-electron laser, *Optics Express* **28**, No. 18, 25706 (2020).

23. Ding, Y. *et al.* Femtosecond X-ray pulse characterization in free-electron lasers using a cross-correlation technique. *Rev. Lett.* **109**, 254802 (2012).

24. Min, C.-K. *et al.* Hard X-ray self-seeding commissioning at PAL-XFEL. *Synchrotron Radiat.* **26**, 1101–1109 (2019).

25. Yang, X. & Shvyd’ko, Yu. V. Maximizing spectral flux from self-seeding hard X-ray free electron lasers. *Rev. ST Accel. Beams*, **16**, 120701 (2013).

26. Riyopoulos, S. & Tang, C. M. The structure of the sideband spectrum in free electron lasers, *Fluids* **31**, 1708 (1988).

27. Riyopoulos, S. & Tang, C. M. Chaotic electron motion caused by sidebands in free electron lasers, *Fluids* **31**, 3387 (1988).

28. Huang, Z. & Stupakov, G. Control and application of beam microbunching in high brightness linac-driven free electron lasers. *Instrum. Methods Phys. Res. A* **907**, 182–187 (2018).

29. Ratner, D. *et al.* Time-resolved imaging of the microbunching instability and energy spread at the Linac Coherent Light Source. *Rev. ST Accel. Beams* **18**, 030704 (2015).

30. Lee, J., Han, J., Lee, S., Hong, J., Kim, C. H., Min, C. K. & Ko, I. S. PAL-XFEL laser heater commissioning. *Instrum. Methods Phys. Res. A*, **843**, 39–45 (2017).

31. Kang, H.-S. *et al.* FEL performance achieved at PAL-XFEL using a three-chicane bunch compression scheme. *Synchrotron Radiat.* **26**, 1127–1138 (2019).

32. Nam, I. *et al.* Manuscript under preparation.

33. Zhu, D. *et al.* A single-shot transmissive spectrometer for hard X-ray free-electron lasers, *Phys. Lett.* **101**, 034103 (2012).

34. Loos, H. Operational experience at LCLS. In *2011 Free Electron Laser Conference* 166–172 (JACOW, 2011).

35. Spence, J.C.H., Kirian, R.A., Wang, X., Weierstall, U., Schmidt, K.E., White, T., Barty, A., Chapman, H.N., Marchesini, S., and Holton, J., Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals, *OPTICS EXPRESS* **19**, No. 4, 2866 (2011).

36. Nakane, T., Hanashima, S., Suzuki, M., Saiki, H., Hayashi, T., Kakinouchi, K., Sugiyama, S., Kawatake, S., Matsuoka, S., Matsumori, N., Nango, E., Kobayashi, J., Shimamura, T., Kimura, K., Mori, C., Kunishima, N., Sugahara, M., Takaku, Y., Inoue, S., Masuda, T., Hosaka, T., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Inoue, T., Nureki, O., Iwata, S., Murata, M., and Mizohata, E., Membrane protein structure determination by SAD, SIR, or SIRAS phasing in serial femtosecond crystallography using an iododetergent, *PNAS* **113** (46), 13039–13044 (2016).

37. Yamashita, K., Kuwabara, N., Nakane, T., Murai, T., Mizohata, E., Sugahara, M., Pan, D., Masuda, T., Suzuki, M., Sato, T., Kodan, A., Yamaguchi, T., Nango, E., Tanaka, T., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Manya, H., Endo, T., Kato, R., Senda, T., Kato, H., Iwata, S., Ago, H., Yamamoto, M., Yumoto, F. and Nakatsua, T., Experimental phase determination with selenomethionine or mercury-derivatization in serial femtosecond crystallography, *IUCrJ* **4**, 639–647 (2017).

38. Liu, H. and Spence, J.C.H., XFEL data analysis for structural biology, *Quantitative Biology* **4**(3): 159–176 (2016).

39. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. Science 336, 1030–1033 (2012).

40. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213–221 (2010).
41. Caldararu, O., Kumar, R., Oksanen, E., Logan, D.T. and Ryde, U. Are crystallographic B-factors suitable for calculating protein conformational entropy? Chem. Chem. Phys. 21, 18149 (2019).
42. Ko, J.H., Kim, G., Kim, C., Kang, H.-S., and Ko, I.S., Coherent synchrotron radiation monitor for microbunching instability in XFEL, REVIEW OF SCIENTIFIC INSTRUMENTS 89, 063302 (2018).
43. Park, J., Kim, S., Nam, K. H., Kim, B. & Ko, I. S. Current status of the CXI beamline at the PAL-XFEL. Korean. Phys. Soc. 69, 1089–1093 (2016).
44. Kim, J. et al. Focusing X-ray free-electron laser pulses using Kirkpatrick-Baez mirrors at the NCI hutch of the PAL-XFEL. Synchrotron. Radiat. 25, 289–292 (2018).
45. Mariani, V. et al. OnDA: Online data analysis and feedback for serial X-ray imaging. Appl. Cryst.49, 1073–1080 (2016).
46. Barty, A. et al. Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. Appl. Cryst.47, 1118–1131 (2014).
47. White, T. A. et al. CrystFEL: A software suite for snapshot serial crystallography. Appl. Cryst.45, 335–341 (2012).
48. White, T. A. et al. Recent developments in CrystFEL. Appl. Cryst.49, 680–689 (2016).
49. Duisenberg, A. J. M. Indexing in single-crystal diffractometry with an obstinate list of reflections. Appl. Cryst. 25, 92–96 (1992).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Cryst.D66, 486–501 (2010).
51. Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Cryst.D66, 12–21 (2010).

Figures

Figure 1

Schematic diagram of the self-seeding experiment at PAL-XFEL. a and b, Beam profiles of IR laser and electron beam at the end of the laser heater (LH). Initial and increased slice energy spread distribution with the laser heater turned OFF (c) and ON (d). e, Simulated longitudinal phase space of electron beam before the undulator from ELEGANT simulation using parameters of PAL-XFEL for the laser heater OFF (top) and ON for SASE (middle) and for self-seeding conditions (bottom). f, Time responses from the forward Bragg diffraction (black dotted line), measured seeded FEL intensity (solid red line), and SASE (solid blue line) without the crystal as a function of chicane delay and a seed pulse (green dotted line). Twenty undulators are used, 8 before and 12 after the self-seeding system. Two diamond crystal plates are mounted on one holder, with crystal surface parallel to (100) and (110) atomic planes and thickness of 100 μm and 30 μm, respectively.
Figure 2

Suppression of microbunching instability by a laser heater (LH) centered at 9.7 keV. a, Spectra of self-seeded XFEL as a function of the induced-energy spread by LH. b, Fraction of enclosed within the bandwidth for four different induced-energy spreads by LH. c, Peak intensity of self-seeded FEL (solid red line), total sum of spectrum in a (dotted blue line), and fraction of FEL intensity enclosed within ±0.5 eV (magenta dotted line) as a function of energy spread induced by LH. Green dotted line: coherent synchrotron radiation (CSR) due to the MBI, as measured at the third bunch compressor using a CCD visible-light camera. The optimized LH condition for self-seeding and SASE are different, where the induced energy spread is about 5 keV is higher for self-seeding than than for SASE.
Figure 3

Spectral intensity of self-seeded vs. SASE XFEL. a, Color maps of 1,000 SASE and self-seeded FEL spectra measured by a single-shot spectrometer. b, SASE and self-seeded FEL spectra averaged over 1,000 shots, with the peak value of the SASE spectrum set to 1, and Ec = 9.7 keV. c, Crystal angle-scanning spectrum measurement with Si-333 flat-crystal scanning spectrometer. Each data point is an average of 150 shots. Vertical axis represents the photo-diode current normalized by a quadratic beam position monitor (QBPM) for the FEL intensity measurement. The step of the crystal angle scan is 0.0001°, which corresponds to 0.022 eV. d, Histogram of radiation intensity (for 1-eV bandwidth around peak) expressed relative to average SASE intensity (1 a.u.) for SASE and self-seeded (SS) modes. Histogram data represent the 1000 single-shot spectra measurements in Fig. 3a. The 33-3 Bragg reflection is used in the diamond FBD monochromator. The e-bunch delay in the magnetic chicane is 30 fs.

Figure 4

Data quality indicators as a function of resolution. a, signal-to-noise ratio (SNR or I/σ), b, multiplicity, c, Rsplit, and d, correlation coefficient (CC*) derived from three HXRSS and three SASE data sets. The sets SS1/SASE1, SS2/SASE2, and SS3/SASE3 are calculated from 70,656, 27,926, and 12,377 total hit images, respectively. The resolution scale (x-axis) of each figure ranges from 1.75 Å to 3.0 Å to show differences between the self-seeded and SASE modes clearly (specific values, Supplementary Table 3). CC* represents a direct comparison of crystallographic model quality and data quality on the same scale, especially for multiplying measured data.
Overall structures of lysozyme models (self-seeded and SASE modes) from chicken eggwhite with 2mFo-DFc and mFo-DFc electron density maps at 1.75-Å resolution. a, Surface and ribbon diagram of the structure with 2mFo-DFc electron density maps on the ten residues (Phe21, Ala28, Tyr41, Trp46, Phe52, Asn62, Tyr71, Trp81, Trp126, and Trp141). Surface and model are colored yellow. The phases for diffraction data were determined using molecular replacement and refined (Materials section). The ten residues are drawn in balls and sticks and also shown in the black rectangles in detail with 2mFo-DFc electron density maps at right (black for SS1 and green for SASE1). The 2mFo-DFc maps are contoured to 1.5σ. b, Omit maps (mFo-DFc electron density maps) for the ten residues. Omit maps for self-seeded and SASE modes are colored in red and blue, respectively. mFo-DFc maps calculated in the same way and contoured to 3.0σ. (Omit maps (mFo-DFc electron density maps) for the ten residues that are calculated from the other self-seeding data sets (SS2/SS2 vs. SASE2/SASE3) are shown in the Supplementary Fig. 5.) The structure of an amino acid is shown in the black rectangular box in the middle of the figure, consisting of an NH group, carbonyl group, alpha-carbon atom (Cα), and an R-group. Comparative analysis on both the maps derived from self-seeded and SASE modes: Supplementary Table 4. Structures and maps were visualized using the program PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, Cambridge, MA, USA).

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- NPsupplementthskang.docx