Comprehensive application of XFEL micro crystallography for novel organic compounds

Kiyofumi Takaba[1, †], Saori Maki-Yonekura[1, †], Ichiro Inoue[1], Kensuke Tono[1, 2], Yasuhiro Fukuda[3], Yota Shiratori[3], Yiying Peng[3], Jumpei Morimoto[3], Satoru Inoue[4], Toshiki Higashino[5], Shinsuke Sando[3], Tatsuo Hasegawa[4], Makina Yabashi[1, 2] and Koji Yonekura[1, 6 *]

[1] RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

[2] Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo, 679-5198, Japan

[3] Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[4] Department of Applied Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

[5] Electronics and Photonics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, 305-8565, Japan

[6] Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

† These authors contributed equally.

* Corresponding author. E-mail: yone@spring8.or.jp. Phone: +81-791-58-2837.

Keywords:

Serial X-ray crystallography (SX); three-dimensional electron diffraction (3D ED); pharmaceutical; organic semiconductor; novel organic compounds

Abstract

24

25 There is a growing demand for structure determination from small crystals, and the three-

26 dimensional electron diffraction (3D ED) technique can be employed for this purpose.

27 However, 3D ED has certain limitations related to crystal thickness and data quality. We here

28 present the application of serial X-ray crystallography (SX) with X-ray free electron lasers

29 (XFELs) to tiny crystals of novel compounds dispersed on a substrate. For XFEL exposures,

30 two-dimensional (2D) scanning of the substrate, coupled with rotation, enables highly efficient

31 data collection. This approach is especially effective for challenging targets, including

32 pharmaceuticals and organic materials that form preferred-oriented flat crystals in low-

33 symmetry space groups. Some of these crystals have been difficult to solve or have yielded

34 incomplete solutions using 3D ED. Our extensive analyses confirmed the superior quality of

35 the SX data, regardless of crystal orientations. Additionally, 2D scanning with XFEL pulses

36 gives an overall distribution of the samples on the substrate, which can be useful for evaluating

37 the properties of crystal grains and the quality of layered crystals. Therefore, this study

38 demonstrates that XFEL crystallography has become a powerful tool for conducting structure

39 studies on small crystals of organic compounds.

40

Introduction

41

42 Structure determination from small crystals offers significant benefits for molecules that do

43 not readily form large single crystals. This is particularly relevant for organic compounds,

44 which often exhibit unique functional properties in their small crystal forms[1–4]. In the

45 pharmaceutical field, the morphology and grain size of crystals can influence crucial properties

46 such as solubility, safety, and efficacy[5]. Especially when considering bioavailability including

47 drug delivery, structure determination from small crystal grains could have significant

2

48 advantages across several applications such as the enhancement of active pharmaceutical

49 ingredients.

50

51 The three-dimensional electron diffraction (3D ED or simply ED in this manuscript)

52 technique has proven successful in studying such small crystals[6–8] thanks to the strong

53 scattering power of atoms for electrons[9]. Thin crystals often yield diffraction spots with

54 remarkably high resolution when the electron beam is directed perpendicular to the plate plane

55 of the crystal[10,11]. However, this property imposes severe limitations on the sample

56 thickness[12,13], particularly when the sample is highly tilted, further exacerbating these

57 limitations[14]. Certain regions in the reciprocal space suffer from a lack of measured data,

58 known as the 'missing wedge/cone'[7,15] and lower data quality as shown in this report. The

59 resulting incompleteness and/or lower quality of the data sometimes hinder correct phasing

60 using ab initio methods and inevitably lead to a degradation in the overall solution quality.

61 Higher discrepancies against atomic models (*R*-factors) are also generally observed in

62 structures solved by ED[16,17]. The underlying reasons for these higher errors are believed to

63 include effects of dynamical scattering[16,18], a suboptimal assignment of electron scattering

64 factors[17,19], and inelastic scattering[20,21]. Addressing and resolving these issues remain

65 significant challenges in the field.

66

67 Alternatively, recent studies have demonstrated that X-ray free electron lasers (XFELs)

68 can be utilized for structure determination from small crystals of inorganic-organic metal

69 hybrid materials[22] and a small organic compound[17]. The XFEL facility SACLA[23,24] generates

70 $\sim 10^{12}$ photons/pulse over 1 μm^2 [25,26], while a typical electron dose exposed for recording one

71 frame is calculated by multiplying ~ 0.01 electrons/Å^2 with the square of a few μm[27,28],

3

72   resulting in $\sim 10^{6\text{-}7}$ electrons/frame. A single XFEL pulse can compensate for the substantial

73   differences in the atomic cross-sections between X-rays and electrons, enabling the generation

74   of diffraction spots with sub Å resolutions from small crystals before the sample is

75   destroyed[29,30]. In our protocol, a custom-designed substrate covered with ~ mg sample crystals

76   is rapidly moved to expose each fresh area of the substrate plane to an XFEL pulse at a

77   repetition rate of 30 Hz. By collecting numerous diffraction images within a span of 1 to 2

78   hours, we can achieve ab initio structure determination for organic compounds consisting

79   solely of light atoms.

80

81   To extend the application of this serial X-ray crystallographic (SX) approach with

82   XFELs, we here conducted an investigation on novel compounds in the pharmaceutical and

83   organic material fields. These compounds presented unique challenges for structure

84   determination, as they formed small crystals belonging to low-symmetry space groups and/or

85   exhibiting heavy aggregation, and comprised plate-like crystals. In some cases, ED

86   encountered limitations due to an elongated electron path when dealing with thicker samples

87   and highly tilted ones. Our research shows the remarkable effectiveness of combining XFEL

88   scanning with tilting of the sample holder to overcome these challenges and acquire excellent

89   data from these difficult targets. Furthermore, we investigated the procedures and performance

90   of data processing and compared the resulting structures and data statistics of these samples

91   with those obtained through ED. This analysis and comparisons offer valuable insights into the

92   advantages of SX.

93

94   **Results**

95   **Strategy for data acquisition and analysis**

4

96
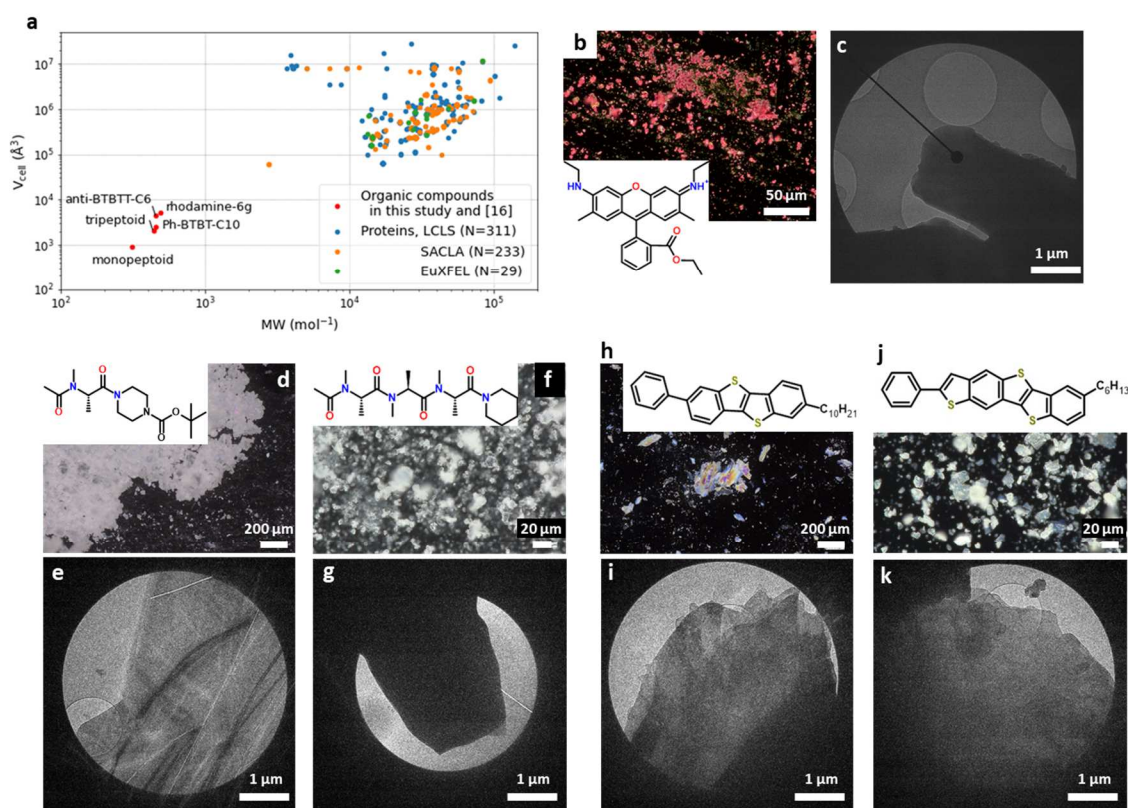
Fig. 1. **Information of sample crystals.**

(**a**) Plots molecular weight (MW) vs cell volume (V_{cell}) for crystals of organic compounds examined in this study (see below), rhodamine-6G in the previous study[17], and protein crystals in three XFEL facilities (derived from Protein Data Bank, www.rcsb.org). (**b, d, f, h, j**) Optical micrographs of the examined crystals overlaid with their molecular formula. (**c, e, g, i, k**) Typical electron micrographs of the corresponding crystals used for the acquisition of rotational ED patterns. (**b, c**) rhodamine-6G, (**d, e**) monopeptoid, (**f, g**) tripeptoid, (**h, i**) Ph-BTBT-C10, and (**j, k**) anti-BTBTT-C6. A shadow of a beam stop was included in (c).

5

105 Table 1. **Crystallographic parameters, and data and refinement statistics for the examined organic compounds by SX.**

	monopeptoid	tripeptoid	Ph-BTBT-C10	anti-BTBTT-C6	rhodamine 6g
Source/Beamline	SACLA/BL2	SACLA/BL3	SACLA/BL2	SACLA/BL3	SACLA/BL2
Photon energy (keV)	15	15	15	15	15
Pulse energy (μJ)	160	280	160	280	160
Camera distance (mm)[a]	100	90	100	90	100
Temperature (K)	Room temperature	Room temperature	Room temperature	Room temperature	Room temperature
Chemical Formula	$C_{15}H_{27}N_3O_4$	$C_{19}H_{34}N_4O_5$	$C_{30}H_{32}S_2$	$C_{28}H_{24}S_3$	$C_{28}H_{31}N_2O_{3.5}Cl$
Formula weight (mol^{-1})	313.4	441.6	456.7	456.5	487.0
Spacegroup	$P2_1$	$P2_1$	$P2_1/a$[g]	$P2_1/c$	$Pbca$
Unit Cell $a/b/c$ ($/\beta$)(Å, deg.)	6.5/11/13/93	5.8/6.8/53/90	6.0/7.8/53/93	13/7.8/47/97	15/15/23
V_{cell} (Å^3)	892	2079	2488	4548	5241
Z[b]	2	4	4	8	8
Total/hit/indexed frames (% of total frames)	141,929/47,693 /35,316 (100/33.6/24.9)	198,457/37,593 /8,850 (100/18.9/4.46)	188,516/107,114 /79,639 (100/56.8/42.2)	182,878/24,253 /6,308 (100/13.3/3.45)	265,624/112,781 /78,106 (100/42.5/29.4)
Average no. of spots per indexed frames	106.8	43.1	142.7	49.5	133.6
Average rate of indexed spots (%)[c]	37.5	40.7	49.2	44.1	51.9

6

Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Multiplicity	802.5 (630.9)	541.1 (530.6)	1,814 (413.5)	155.2 (158.4)	2,668 (993.7)
R_{split} (%)	7.53 (33.5)	7.40 (67.1)	6.42 (46.1)	20.36 (86.98)	8.97 (22.8)
$CC_{1/2}$ (%)[d]	99.2 (94.6)	99.5 (16.3)	99.4 (92.2)	96.6 (27.6)	98.4 (93.8)
$<I/\sigma(I)>$[e]	9.98 (4.10)	9.86 (1.79)	13.0 (2.08)	3.60 (1.36)	12.36 (3.81)
d_{min} (Å)	0.88	0.96	0.85	0.90	0.82
No. of parameters refined	243	506	292	561	322
R_1, wR_2	0.095, 0.296	0.129, 0.358	0.162, 0.450	0.189, 0.445	0.110, 0.301
Peak, hole (e$^-$/Å^3)	0.19, 0.34	0.27, -0.29	0.63, -0.95	0.60, -0.43	0.43, -0.59
<e.s.u.> for bond lengths (Å)[f]	0.014	0.025	0.022	0.024	0.003
<e.s.u.> for bond angles (deg.)	0.859	1.53	1.21	1.55	0.194
CCDC no.	2296549	2296550	2296551	2270804	2119567
Reference	This work	This work	This work	Higashino *et al.*, submitted, This work	Takaba & Maki-Yonekura *et al.*, 2023

[a] camera distances as nominal value [b] *Z*, formula units in unit cell [c] average ratio of the indexed spots per all detected spots for indexed frames [d] $CC_{1/2}$, the Pearson correlation coefficient between two half sets of intensities [e] *I*, measured diffraction intensity [f] e.s.u., estimated standard uncertainties calculated from full-matrix refinement [g] The notation of the crystal axes for Ph-BTBT-C10 is adopted from the reference[31].

111

112
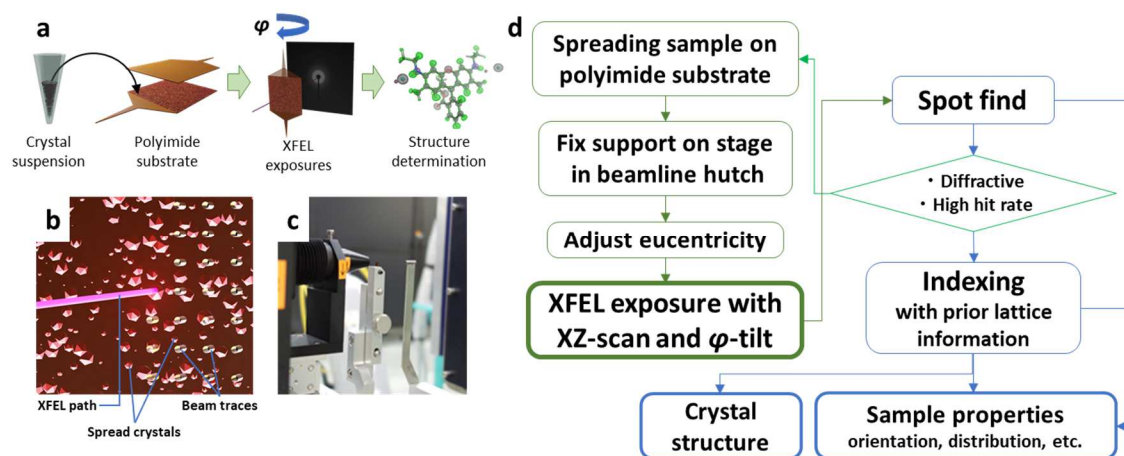
Fig. 2. **Illustration of data collection and flowchart of data processing.**

(**a**) Schematic diagram of structure determination for small crystals of compounds. (**b**) Two-dimensional (2D) scanning of crystals spread on polyimide substrate with XFEL pulses together with rotating the sample substrate around the vertical axis (where the rotation angle is denoted as φ). (**c**) Arrangement of a sample mounter placed between a beam collimator and a beam stopper. (**d**) Flowchart of data collection and processing.

119

In our previous work, we solved the crystal structure of rhodamine-6G at 0.82 Å resolution from tiny crystals by XFEL crystallography[17]. This structure corresponds to a new crystal form of a chloride compound in a high-symmetry space group (*Pbca*). The cell volume of this crystal is approximately 1/10 to 1/100 of typical protein crystals examined using XFELs (Fig. 1a). However, it still falls within the range of a middle-sized unit cell for small molecules. The crystals were well-dispersed without forming heavy aggregations and randomly oriented over the flat surface of the sample substrate. Therefore, this sample is particularly suitable for data acquisition from the flat substrate.

128

8

129   We applied this approach to investigate other types of organic molecular crystals. The

130   crystal images examined are summarized in Figs. 1b – k with their chemical formulas of the

131   molecules. The samples we studied presented specific challenges such as low-symmetry space

132   groups, smaller cell volumes, the presence of heavy aggregates, and preferred orientations. To

133   acquire data using XFEL, the crystal samples were attached to the surface of a polyimide

134   substrate. Subsequently, diffraction patterns were collected by performing a two-dimensional

135   (2D) scan of the substrate for exposures of XFEL pulses. In the case of crystals with preferred

136   orientations, the sample stage was initially brought to the eucentric position and then tilted

137   from 60° to 0° during XFEL exposures.

138

139   Data collection was automated using a SACLA-DAQ system[32], which involved

140   configuring a scheme for stage movement and tilt adjusted to each substrate. The automation

141   process encompassed the control and synchronization of various components, including the of

142   the beam line shutter, data readout from a Rayonix detector, and the movement and tilt of the

143   stage. Diffraction patterns were then converted to the HDF format and processed following

144   procedures similar to those used for protein crystals[33]. For the initial indexing process, we

145   employed pre-determined lattice parameters obtained from rotational ED patterns, which were

146   collected using a 300 kV electron microscope[14] (Supplementary Table 1). All the processes

147   are outlined in Fig. 2.

148

9

149   **Peptoids**
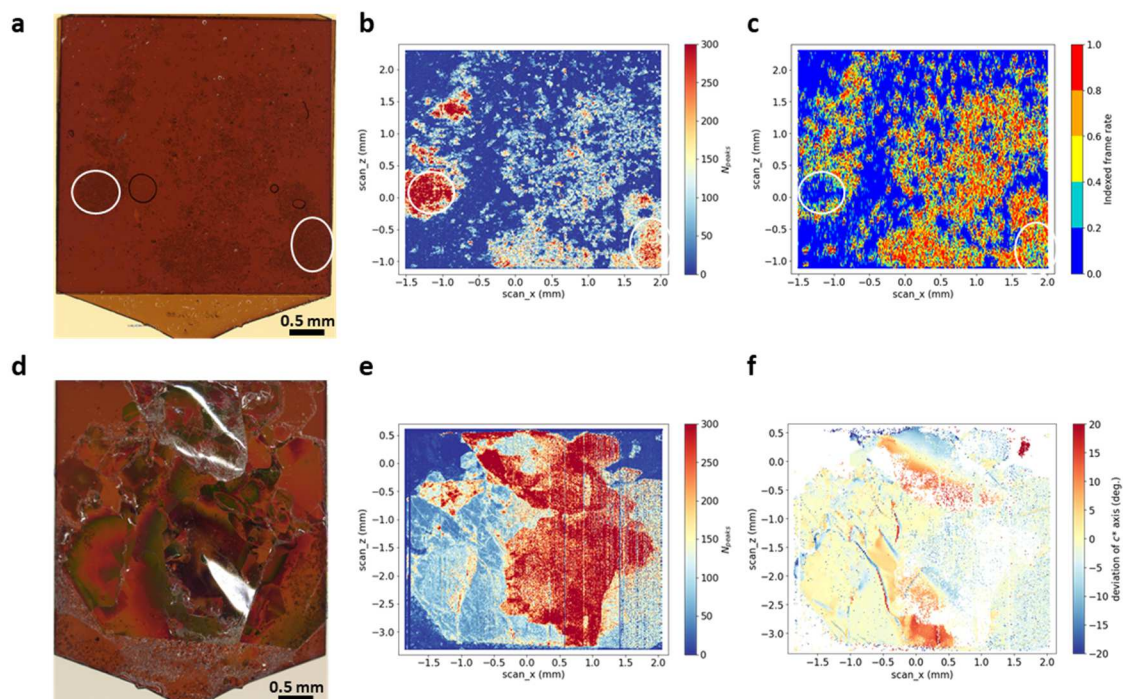
150

151   Fig. 3. **Distributions of crystals on the sample substrate.**

152   (**a**) A picture showing aggregated crystals of monopeptoid on a substrate after XFEL exposures.

153   (**b**) Reconstruction of the crystal distribution in (a) coloured by the number of detected spots

154   (N_{peaks}). (**c**) Distribution of the indexed crystals. Each crystal position is coloured by the number

155   of indexed frames in five consecutive frames divided by five. A white circle on the left or right

156   indicates an area where spot indexing was either hindered, likely due to excessive crystal

157   condensation, or possible to some extent, respectively. (**d**) A picture showing plate-like layered

158   crystals of Ph-BTBT-C10 on a substrate after XFEL exposure. (**e**) Reconstruction as in (b). (**f**)

159   Distribution of the orientation of crystals on the substrate in (d). Each dot indicates a position

160   yielding an indexed frame, which is coloured by the deviation in the $c*$ axis from the direction

161   perpendicular to the substrate plane.

10

162
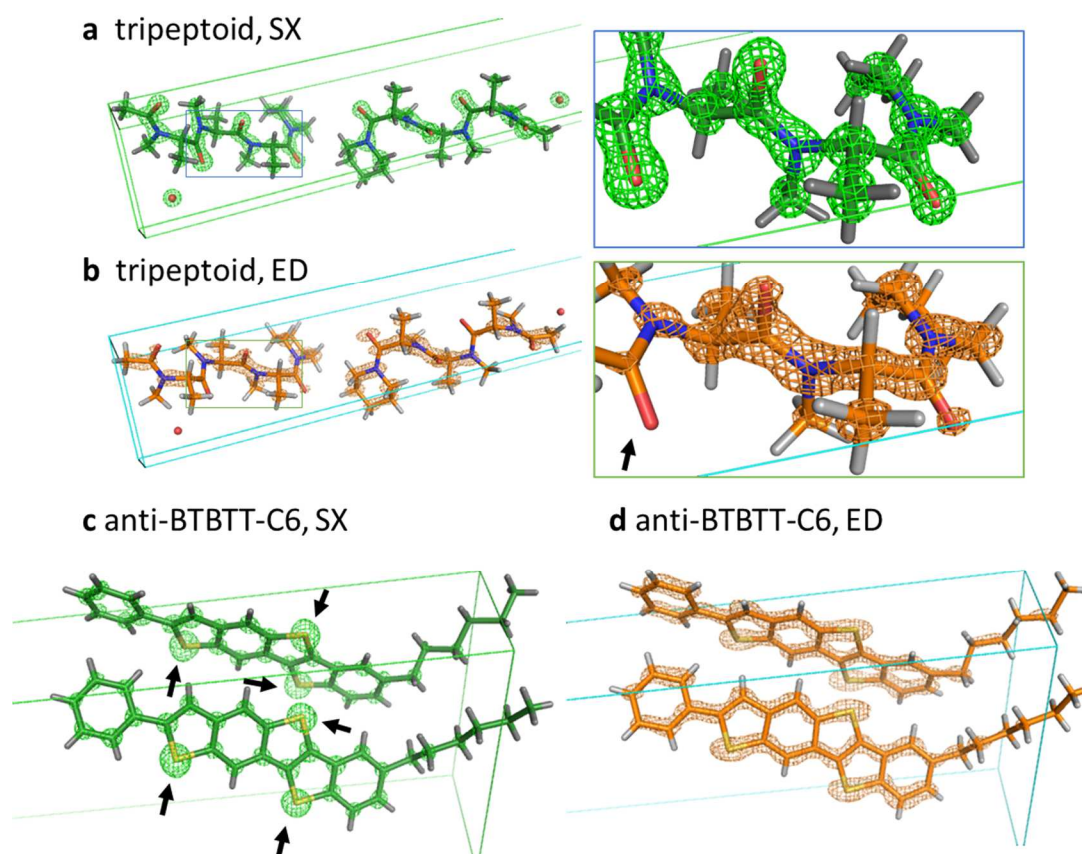
Fig. 4. **Crystal structures of tripeptoid and anti-BTBTT-C6 determined by SX or ED**.

(**a, b**) The tripeptoid structures determined from SX (a) and ED (b). (**c, d**) The anti-BTBTT-C6 structures from SX (c) and ED (d). Closeup views inside the boxes on the left side are shown on the right in (a) and (b). The atomic models are overlaid with $2F_{obs}$-F_{calc} maps contoured at a display level of 2.5σ. An arrow in (b) points to a missing density for some non-hydrogen atoms. Arrows in (c) indicate distinct spherical densities corresponding to sulfur atoms. The longest cell axis, c, of preferred-oriented tripeptoid crystals in (a) and (b) is perpendicular to the substrate or sample support plane, while that of anti-BTBTT-C6 in (c) and (d) are tilted by approximately 50° (see also Supplementary Figs. 1b, d and f).

11
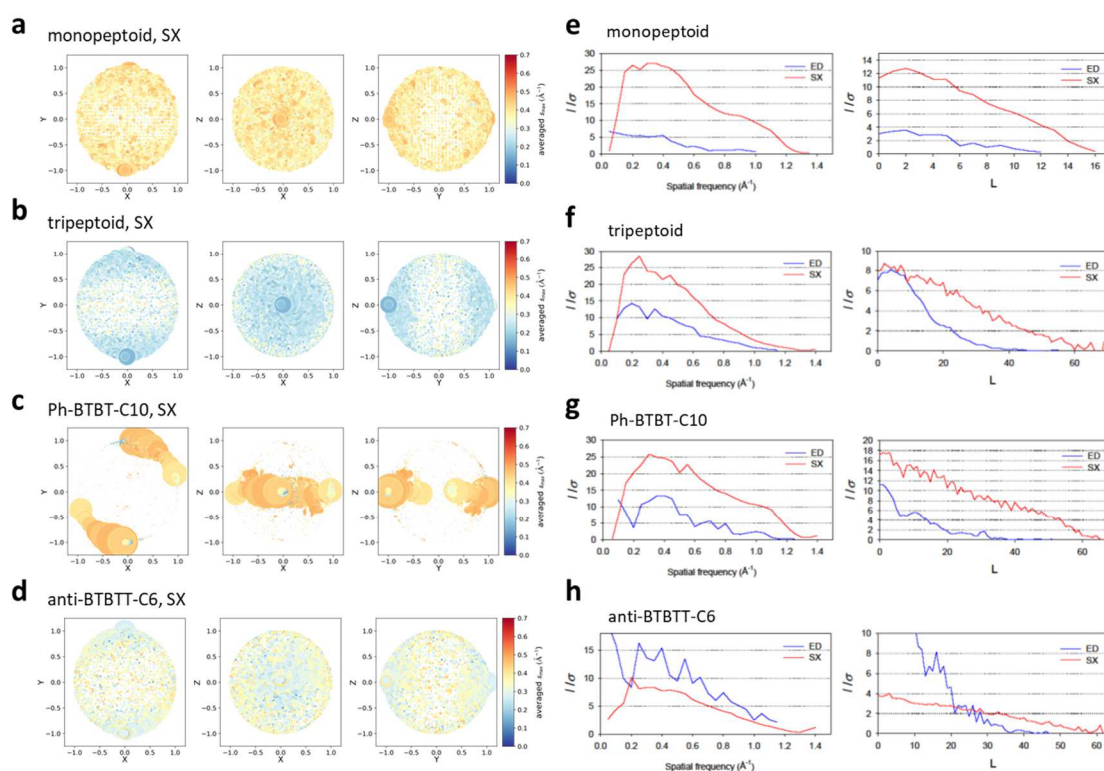
Fig. 5. **Distributions of crystal orientations for the SX data and plots of mean signal-to-noise-ratios, $<I/\sigma(I)>$, for the SX and ED data.**

(**a**) Orientation distributions of the monopeptoid crystals for SX data collection, as projections on the x–y (left), x–z (middle) and y–z (right) planes. (**b**) Distributions of the tripeptoid crystals. (**c**) Distributions of the Ph-BTBT-C10 crystals. (**d**) Distributions of the anti-BTBTT-C6 crystals. In (a - d), the incident X-rays are along the y axis. The data collection geometry for the SX data are shown in Supplementary Fig. 1a. The terminals of $c*$ (a unit vector of the reciprocal lattice) are shown with coloured points. The size and colour of each point represent the relative frequency of frames contributing to the corresponding orientation and the averaged $s_{max} = \frac{\sin\theta}{\lambda}$ (where θ is the maximum half the scattering angle and λ is the wavelength of X-rays), respectively. (**e**) Plots of $<I/\sigma(I)>$ in the SX and ED data of the monopeptoid crystals along the spatial frequency in the left, while those along a Miller index, L, along the $c*$ axis in

12

185 the right. (**f**) Plots of the tripeptoid crystal data. (**g**) Plot of the Ph-BTBT-C10 crystal data. (**h**)

186 Plots of the anti-BTBTT-C6 crystal data.

187

188 We examined the crystals of peptidomimetics called peptoids, which are peptide analogues

189 with functional groups on amide nitrogens and have demonstrated pharmacokinetic

190 properties[34,35]. These molecules exhibit high cohesion through intramolecular hydrophobic

191 interactions, making it challenging to obtain larger single crystals[36]. A peptoid monomer

192 named "monopeptoid" and an oligomer consisting of three residues named "tripeptoid" were

193 synthetized. The crystal structure of monopeptoid was recently reported[37], while the structure

194 of tripeptoid had not been determined yet. Powder crystals of these peptoids were first observed

195 using a cryogenic-electron microscope (cryo-EM). Monopeptoid formed aggregations of small

196 crystal grains (Fig. 1d), while tripeptoid crystal gains were more dispersed (Fig. 1f). However,

197 most crystals of both samples were too thick for rotational ED measurements. The crystals

198 yielded diffraction spots when untilted but exhibited fewer and lower-quality spots as the tilt

199 angles increased. We searched for thin crystals and collected their rotational ED patterns. The

200 crystals of both peptoids belong to a lower-symmetry space group, $P2_1$. Among all the samples

201 tested in this study (Fig. 1a), monopeptoid had the smallest cell volume, 891.5 $Å^3$, which is

202 even smaller than the volumes of the inorganic-organic hybrid material crystals analysed by

203 XFEL (1201-1313 $Å^3$)[22]. The structure of monopeptoid was solved using the ab initio method

204 by merging only 3 rotational datasets collected from 3 crystals, whereas diffraction data from

205 tripeptoid failed to be phased even from 30 data collections (Supplementary Table 1). With

206 the presence of preferred orientation (Supplementary Figs. 1b and d), the completeness of the

207 ED data for the tripeptoid crystals reached 94.6% (Supplementary Table 1). However, $<I/\sigma>$

208 values from high tilt angles was found to be poor (see Discussion).

13

209

210   We then conducted experiments using XFELs for both peptoid crystals. Crystals of

211 monopeptoid exhibited significant aggregation on the polyimide substrate (Fig. 3a) as seen in

212 Fig. 1d. However, we were able to process the SX patterns by supplying the lattice parameters

213 obtained from the rotational ED patterns. The reconstruction of the sample distribution on the

214 substrate revealed that SX patterns collected from the edge of the chunk were successfully

215 indexed (a white circle on the lower right in Fig. 3a-c). As for tripeptoid crystals, SX patterns

216 were collected by tilting the sample stage from 60º to 0º. Merging the data resulted in 100%

217 completeness (Table 1), and the crystal structure was determined from this dataset. The data

218 quality for both mono and tripeptoid crystals is superior to those by ED (cf. crystallographic

219 statistics in Table 1 and Supplementary Table 1). This reflects on the model refinement

220 statistics, and the R_1 values for mono and tripeptoid are 0.095 and 0.129 for SX and 0.172 and

221 0.242 for ED, respectively. The ED data of tripeptoid also produced the same structure using

222 the phase information derived from the determined SX structure (Fig. 4a, b). However, the

223 densities observed in the ED structure appeared broad perpendicular to the crystal plane (Fig.

224 4b)[7,14,38] Some densities are missing in the ED structure (an arrow in Fig. 4b), which had

225 probably impeded the structure determination from the ED data alone. The distributions of the

226 peptoid crystal orientations for both SX and ED data, which are illustrated in Supplementary

227 Figs. 1a and b, are shown in Figs. 5a and b and Supplementary Fig. 1c and d.

228

229 **Organic semiconductor materials**

230 Next we conducted experiments on organic semiconductor materials, 2-decyl-7-phenyl[1]-

231 benzothieno[3,2-*b*][1]benzothiophene (Ph-BTBT-C10)[31] and an anti-isomer of 7-hexyl-2-

232 phenyl-benzothieno[5,6-*b*]benzothieno[3,2-*b*]thiophene (anti-BTBTT-C6). Ph-BTBT-C10 is

14

233   known to form well-ordered plate-like layered crystals[31], although repeat distances between

234   the layers are approximately 7 to 9 times longer than those within the layer plane (Table 1) and

235   the crystallinity along the layers is worse. Ph-BTBT-C10 yields relatively thicker plate crystals

236   that are amenable to conventional X-ray diffraction (XRD). On the other hand, anti-BTBTT-

237   C6 formed only flakes of plate-like crystals (Fig. 1j) and its structure had not been determined

238   yet. A crystal structure of its syn-isomer (syn-BTBTT-C10) was recently reported[39]. For Ph-

239   BTBT-C10 and anti-BTBTT-C6, ED data were collected in the same way as for peptoids.

240

241   We applied XFEL crystallography to Ph-BTBT-C10 and anti-BTBTT-C6. The inherent

242   characteristics of these molecules resulted in preferred orientations on the substrate plane.

243   Consequently, diffraction patterns were collected by tilting the sample holder, as described for

244   the tripeptoid crystals, and processed using pre-determined lattice parameters obtained by ED.

245   This approach yielded datasets with 100% completeness for both Ph-BTBT-C10 and anti-

246   BTBTT-C6 (Table 1), even when the absolute maximum value of stage tilt angle is limited to

247   60°. The distributions of the crystal orientations are shown in Figs. 5c and d. This observation

248   suggests that the orientations of thicker crystals, which were not suited for ED measurements,

249   varied more in relation to the substrate plane and/or that the plane itself exhibited local bending

250   (see Fig. 3f). The crystal structures of both materials were successfully solved using the ab

251   initio method. The refinement with a twin option improved R-factors for both Ph-BTBT-C10

252   and anti-BTBTT-C6.

253

254   One sample of Ph-BTBT-C10 formed thin layered crystals over the substrate plane (Fig.

255   3d, e). From indexed frames of XFEL diffraction patterns, we were able to plot the orientations

256   over the crystal grains on the substrate (Fig. 3f), providing insights into the uniformity and

15

257   defects in the crystalline structure. Our XFEL scanning approach offers a spacing of 10 μm

258   between two exposed points, which is ~10^{4-5} times larger compared to electron beam scanning

259   techniques like 4D scanning transmission electron microscopy, where the scanning step was

260   typically 0.02-0.5 nm[40,41]. Nonetheless, XFEL scanning allows for ~10^7 times broader areas

261   and thicker samples to be studied.

262

263   The completeness of the ED data is 78.5% for Ph-BTBT-C10 [14] and 90.0% for anti-

264   BTBTT-C6 (Supplementary Table 1) due to the preferred orientations. Despite the presence of

265   a missing cone indicated in the ED data (Supplementary Fig. 1e and f), the ab initio method

266   gave the correct solutions for both molecules as for the SX data above. However, the data

267   quality of the ED data is inferior to those of the SX data (cf. crystallographic statistics in Table

268   1 and Supplementary Table 1; see also Discussion). The R_1 values for models of Ph-BTBT-

269   C10 and anti-BTBTT-C6 are 0.162 and 0.189 for SX and 0.242 and 0.250 for ED, respectively.

270   The SX map reveals distinct and isolated spherical electron densities for individual atoms, with

271   carbon and sulfur atoms distinguishable based on their sizes (Fig. 4c). In contrast, the ED map

272   appears more elongated nearly along the longitudinal axis of the molecule (Fig. 4d). This

273   elongated feature is consisted with the ED structure of tripeptoid shown in Fig. 4b.

274

275   **Discussion**

276   **Sample preparation, data collection, and processing**

277   Crystals of small organic compounds are often embedded in paraffin oil to facilitate the

278   collection of X-ray diffraction data. The oil embedding provides good adhesion of these

279   crystals including plate-like ones to the polyimide substrate plane. Utilizing the system

16

280   depicted in Fig. 2, we can acquire approximately 130,000 patterns from these samples on a

281   single substrate within approximately 1.5 hours.

282

283   The frames showing diffraction spots, referred to as "hitting" patterns, are identified in

284   over 13.3 – 56.8% of the total frames (total/hit/indexed frames (% of total frames) in Table 1),

285   which is up to approximately six times greater than the alternative method used for small

286   molecules, namely delivery through a liquid jet[22]. The sample consumption is 1 to a few mg

287   per sample substrate, which is 1/10 to 1/100 of the amount required by the liquid jet method.

288   This consumption can be further reduced to 1/5 to 1/10 [42] if the sample crystals are spread over

289   the substrate surface with minimal overlaps. Assuming a sample dimension of 10 μm and a

290   density of 1 mg/mm³, a 4×4 mm² plate can hold 0.16 mg of samples, without considering any

291   sample loss. The flat sample support can be suitable particularly for thin plate-like crystals that

292   grow only in a specific direction and may undergo deformation without the support (Fig. 3d).

293

294   In addition, our protocol involves a straightforward preparation process at the

295   experimental hall of SACLA. The procedure simply requires placing the sample substrate on

296   the goniometer stage and registering the stage positions for the areas that will be exposed to

297   XFELs.

298

299   Serial data collection covers the reciprocal space ideally through randomly-oriented

300   still diffraction frames. Permissible crystallographic symmetries efficiently fill space and

301   reduce the number of structure parameters to be determined. However, our method and the

302   results from the current and previous[17] studies have shown that higher symmetries are not

303   essential for structure determination using SX.

17

304

Indexing

306 In this and previous[17] studies we employed pre-determined lattice parameters derived from

307 rotational ED patterns as a reference for indexing diffraction spots in still SX frames. An

308 alternative approach involved creating a one dimensional (1D) profile from SX patterns to

309 directly estimate lattice parameters[22]. We also explored this method for our crystals

310 (Supplementary Fig. 2). For the monopeptoid crystal, which possesses the smallest cell volume

311 (Fig. 1a and Table 1), we confirmed that the correct lattice parameters were included in the

312 lattice candidate list derived from the 1D profile (Supplementary Table 2, Methods). However,

313 selecting the correct lattice from the list is not straightforward, even when considering the

314 figure of merit values (FOM in Supplementary Table 2). Testing these candidate lattices one

315 by one for all the SX patterns would incur a high computational cost.

316

317 The approach utilizing the 1D profile could still hold value in verifying whether other

318 possible solutions were overlooked, by cross-checking it with calculated 1D profiles from the

319 determined structure model. In this study, the average rates of indexed spots in the total

320 detected spots ranged from 37.5 to 49.2 % (Table 1). While a certain proportion of unindexed

321 spots remained after the indexing step, these spots did not contribute to additional unassigned

322 peaks in the experimental 1D profiles. Thus, the unindexed spots did not originate from other

323 overlooked polymorphic crystals in the measured specimens.

324

325   Longer cell parameters result in more complex 1D profiles with overlapping peaks,

326 making it difficult or even impossible to assign correct lattices. In contrast, our scheme utilizing

327 lattice parameters obtained by ED is less dependent on the cell lengths (Fig. 1a and Table 1).

328

329 **Comparison with ED data**

330 R-factors for ED structures are typically inferior to those for XRD structures, leading to greater

331 geometry errors in ED structures[16,17]. These trends were observed in all the crystals examined

332 in this study (Table 1 and Supplementary Table 1). To further investigate these observations,

333 we compared the completeness against spatial frequencies between the SX and ED datasets

334 (Supplementary Figs. 3). The values remain close to 100% up to the highest resolution shell

335 for all SX datasets except for that of anti-BTBTT-C6 with an I/σ cutoff < 2 (Supplementary

336 Figs. 3a - d). On the other hand, the values in the ED datasets exhibit gradual yet significant

337 decreases with the same I/σ cutoff (Supplementary Figs. 3e - h), confirming the superior

338 coverage of the reciprocal space with higher signal-to-noise reflections in the SX datasets.

339

340   We then plotted $<I/\sigma>$ values against spatial frequencies and Miller indices (Figs. 5e –

341 h and Supplementary Figs. 4c and 5) to explore the dependency on specific directions. The

342 plots reveal the superiority of the SX data over the ED data except for anti-BTBTT-C6.

343 Considering that the individual diffraction strength was sufficient for anti-BTBTT-C6 in the

344 SX (Fig. 5d), it is possible that the twinned component of the crystals was not completely

345 separated due to the limited number of spots, leading to reduced $<I/\sigma>$ values for integrated

346 intensity.

347

348   The crystals of tripeptoid and Ph-BTBT-C10 tended to orient their *c*-axes perpendicular

349   to the substrate plane (Fig. 5b and c). The crystals exhibited the same preferred orientations on

350   the support plane for ED measurement (Supplementary Fig. 1d and e). When examining the

351    $<I/\sigma>$ plots against the Miller index L, we observed rapid decreases in $<I/\sigma>$ values for higher

352   L indices in the ED data (Fig. 5f and g, and Supplementary Fig. 5). For these crystals,

353   reflections with higher Ls were measured only from higher tilt angles, indicating that tilting to

354   higher angles decreased the signal-to-noise ratio in the measured ED intensity. While the *c*-

355   axis of anti-BTBTT-C6 was leaned by ~50º from the direction perpendicular to the support

356   plane (Supplementary Fig. 1f), the same decrease in $<I/\sigma>$ were observed in the ED data of

357   anti-BTBTT-C6 (Fig. 5h).

358

359   Consequently, the structures determined by SX exhibit superior quality compared to

360   those obtained by ED (Table 1, Supplementary Table 1, and Fig. 4).

361

362   **Conclusion**

363   This report presents the successful application of XFEL crystallography to various organic

364   compounds dispersed on a substrate plane, resulting in the determination of their crystal

365   structures. Among the compounds studied were newly synthesized molecules, mono and

366   tripeptoids, and the organic semiconductor anti-BTBTT-C6.

367

368   These samples exhibited challenging characteristics including low-symmetry space

369   groups, smaller cell sizes, the presence of heavy aggregates, and preferred orientations. In

370   particular the preferred orientations and excessive crystal thickness often pose obstacles to

371   structure determination using ED.

372

373   Our approach combines 2D scanning and tilting of the sample holder and performs data

374   processing with lattice parameters obtained through ED. We demonstrated that these

375   challenges did not impede structure determination of these crystals, resulting in superior

376   structure data. Consequently, this study opens avenues for the extensive utilization of XFEL

377   crystallography. Furthermore, the presented approach will hold significant value for time-

378   resolved studies on organic molecules, such as exploring structural changes during chemical

379   reactions.

380

381   **Author contributions:** Y. F., Y. S., Y. P., J. M., S. I., and T. H. synthesized target compounds.

382   Ki. T., S. M-Y., I. I., Ke. T. and K. Y. conceived and designed the diffraction experiments. S.

383   M-Y. prepared target specimens for SX and ED experiments. I. I., Ke. T. and M. Y. set-up

384   XFEL beamline for the measurement. K. Y. set-up cryo-electron microscope for the

385   measurement. Ki. T., S. M-Y. and K. Y. collected SX data and S. M-Y collected ED data. Ki.

386   T. processed the raw-data from SX and ED, solved structures and analysed them. Ki. T. and K.

387   Y. discussed the results and wrote the manuscript. All authors approved the manuscript.

388

389   **Competing interests:** The authors declare no competing interests.

390

391   **Data availability**: Crystallographic data were deposited at the Cambridge Crystallographic

392   Data Centre, under deposition numbers CCDC 2296549 (monopeptoid), 2296550 (tripeptoid),

393   2296551 (Ph-BTBT-C10), 2270804 (anti-BTBTT-C6). SX image data was deposited at the

394   Coherent X-ray Imaging Database (CXIDB), under a deposition number XXX,

395   https://www.cxidb.org/id-***.html.

21

396

Acknowledgements

404

Supplementary information is available for this paper.

Supplementary Tables 1 and 2

Supplementary Figures 1 to 6

408

References

1. Cao, X., Tan, C., Sindoro, M. & Zhang, H. Hybrid micro-/nano-structures derived from metal-organic frameworks: Preparation and applications in energy storage and conversion. *Chem. Soc. Rev.* **46**, 2660–2677 (2017).
2. Fontana, F. *et al.* Production of pure drug nanocrystals and nano co-crystals by confinement methods. *Adv. Drug Deliv. Rev.* **131**, 3–21 (2018).
3. Andrusenko, I. *et al.* The Crystal Structure of Orthocetamol Solved by 3D Electron Diffraction. *Angew. Chem. Int. Ed. Engl.* **131**, 11035–11038 (2019).
4. Kato, K. *et al.* Double-Helix Supramolecular Nanofibers Assembled from Negatively Curved Nanographenes. *J. Am. Chem. Soc.* **143**, 5465–5469 (2021).
5. Blagden, N., de Matas, M., Gavan, P. T. & York, P. Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Adv. Drug Deliv. Rev.* **59**, 617–630 (2007).
6. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930 (2014).
7. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. Electron crystallography of ultrathin 3D protein crystals: Atomic model with charges. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3368–3373 (2015).

8. Van Genderen, E. *et al.* Ab initio structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector. *Acta Crystallogr. Sect. A Found. Adv.* **72**, 236–242 (2016).

9. Henderson, R. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).

10. Gim, S. *et al.* Supramolecular Assembly and Chirality of Synthetic Carbohydrate Materials. *Angew. Chemie - Int. Ed.* **59**, 22577–22583 (2020).

11. Inoue, S. *et al.* Emerging Disordered Layered-Herringbone Phase in Organic Semiconductors Unveiled by Electron Crystallography. *Chem. Mater.* **34**, 72–83 (2022).

12. Martynowycz, M. W., Clabbers, M. T. B., Unge, J., Hattne, J. & Gonen, T. Benchmarking the ideal sample thickness in cryo-EM. *RIKAGAKU KENKYUSHO Libr. Sept.* **118**, (2021).

13. Subramanian, G., Basu, S., Liu, H., Zuo, J. M. & Spence, J. C. H. Solving protein nanocrystals by cryo-EM diffraction: Multiple scattering artifacts. *Ultramicroscopy* **148**, 87–93 (2015).

14. Takaba, K., Maki-Yonekura, S., Inoue, S., Hasegawa, T. & Yonekura, K. Protein and Organic-Molecular Crystallography With 300kV Electrons on a Direct Electron Detector. *Front. Mol. Biosci.* **7**, 612226 (2021).

15. Chiu, M. Y. *et al.* Three-dimensional radiographic imaging with a restricted view angle. *J. Opt. Soc. Am.* **69**, 1323 (1979).

16. Clabbers, M. T. B., Gruene, T., van Genderen, E. & Abrahams, J. P. Reducing dynamical electron scattering reveals hydrogen atoms. *Acta Crystallogr. Sect. A Found. Adv.* **75**, 82–93 (2019).

17. Takaba, K. *et al.* Structural resolution of a small organic molecule by serial X-ray free-electron laser and electron crystallography. *Nat. Chem.* **15**, 491–497 (2023).

18. Klar, P. B. *et al.* Accurate structure models and absolute configuration determination using dynamical effects in continuous-rotation 3D electron diffraction data. *Nat. Chem.* **15**, (2023).

19. Yonekura, K. & Maki-Yonekura, S. Refinement of cryo-EM structures using scattering factors of charged atoms. *J. Appl. Crystallogr.* **49**, 1517–1523 (2016).

20. Tsuda, K. & Tanaka, M. Refinement of crystal structure parameters using convergent-beam electron diffraction: the low-temperature phase of SrTiO $_3$. *Acta Crystallogr. Sect. A Found. Crystallogr.* **51**, 7–19 (1995).

21. Yonekura, K., Maki-Yonekura, S. & Namba, K. Quantitative comparison of zero-loss and conventional electron diffraction from two-dimensional and thin three-dimensional protein crystals. *Biophys. J.* **82**, 2784–2797 (2002).

22. Schriber, E. A. *et al.* Chemical crystallography by serial femtosecond X-ray diffraction. *Nature* **601**, 360–365 (2022).

23. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region. *Nat. Photonics* **6**, 540–544 (2012).

24. Yabashi, M., Tanaka, H. & Ishikawa, T. Overview of the SACLA facility. *J. Synchrotron Radiat.* **22**, 477–484 (2015).

25. Tono, K. *et al.* Beamline, experimental stations and photon beam diagnostics for the hard x-ray free electron laser of SACLA. *New J. Phys.* **15**, (2013).

23

475   26.   Tono, K., Hara, T., Yabashi, M. & Tanaka, H. Multiple-beamline operation of
476   SACLA. *J. Synchrotron Radiat.* **26**, 595–602 (2019).
477   27.   Saha, A., Nia, S. S. & Rodríguez, J. A. Electron Diffraction of 3D Molecular Crystals.
478   *Chem. Rev.* **122**, 13883–13914 (2022).
479   28.   Gruene, T. & Mugnaioli, E. 3D Electron Diffraction for Chemical Analysis:
480   Instrumentation Developments and Innovative Applications. *Chem. Rev.* **121**, 11823–
481   11834 (2021).
482   29.   Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for
483   biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
484   30.   Barty, A. *et al.* Self-terminating diffraction gates femtosecond X-ray
485   nanocrystallography measurements. *Nat. Photonics* **6**, 35–40 (2012).
486   31.   Minemawari, H. *et al.* Crystal structure of asymmetric organic semiconductor 7-decyl-
487   2-phenyl[1]benzothieno[3,2- b][1]benzothiophene. *Appl. Phys. Express* **7**, 8–11
488   (2014).
489   32.   Joti, Y. *et al.* Data acquisition system for X-ray free-electron laser experiments at
490   SACLA. *J. Synchrotron Radiat.* **22**, 571–576 (2015).
491   33.   Nakane, T. *et al.* Data processing pipeline for serial femtosecond crystallography at
492   SACLA. *J. Appl. Crystallogr.* **49**, 1035–1041 (2016).
493   34.   Simon, R. J. *et al.* Peptoids: A modular approach to drug discovery. *Proc. Natl. Acad.
494   Sci. U. S. A.* **89**, 9367–9371 (1992).
495   35.   Morimoto, J. *et al.* A Peptoid with Extended Shape in Water. *J. Am. Chem. Soc.* **141**,
496   14612–14623 (2019).
497   36.   Darapaneni, C. M., Kaniraj, P. J. & Maayan, G. Water soluble hydrophobic peptoids
498   via a minor backbone modification. *Org. Biomol. Chem.* **16**, 1480–1488 (2018).
499   37.   Morimoto, J. *et al.* Bottom-up design of peptide nanoshapes in water using oligomers
500   of N-methyl-L/D-alanine. *ChemRxiv* (2023).
501   38.   Wennmacher, J. T. C. *et al.* 3D-structured supports create complete data sets for
502   electron crystallography. *Nat. Commun.* **10**, 3316 (2019).
503   39.   Higashino, T. *et al.* Architecting Layered Crystalline Organic Semiconductors Based
504   on Unsymmetric π-Extended Thienoacenes. *Chem. Mater.* **33**, 7379–7385 (2021).
505   40.   Gao, W. *et al.* Real-space charge-density imaging with sub-ångström resolution by
506   four-dimensional electron microscopy. *Nature* **575**, 480–484 (2019).
507   41.   Ophus, C. Four-Dimensional Scanning Transmission Electron Microscopy (4D-
508   STEM): From Scanning Nanodiffraction to Ptychography and Beyond. *Microsc.
509   Microanal.* (2019).
510   42.   Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an X-ray free
511   electron laser. *Sci. Rep.* **4**, 1–5 (2014).
512   43.   Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P. & Waterman, D. G.
513   Electron diffraction data processing with DIALS. *Acta Crystallogr. Sect. D Struct.
514   Biol.* **74**, 506–518 (2018).
515   44.   White, T. A. *et al.* Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–
516   689 (2016).
517   45.   Sheldrick, G. M. SHELXT - Integrated space-group and crystal-structure
518   determination. *Acta Crystallogr. Sect. A Found. Crystallogr.* **71**, 3–8 (2015).
519   46.   Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. Sect. C
520   Struct. Chem.* **71**, 3–8 (2015).

24

521 47. Toby, B. H. & Von Dreele, R. B. GSAS-II: The genesis of a modern open-source all
522   purpose crystallography software package. *J. Appl. Crystallogr.* **46**, 544–549 (2013).
523 48. Mastronarde, D. N. Automated electron microscope tomography using robust
524   prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
525 49. Yonekura, K., Ishikawa, T. & Maki-Yonekura, S. A new cryo-EM system for electron
526   3D crystallography by eEFD. *J. Struct. Biol.* **206**, 243–253 (2019).
527 50. Takaba, K., Maki-Yonekura, S. & Yonekura, K. Collecting large datasets of rotational
528   electron diffraction with ParallEM and SerialEM. *J. Struct. Biol.* **211**, 107549 (2020).
529 51. Yamashita, K., Hirata, K. & Yamamoto, M. KAMO: towards automated data
530   processing for microcrystals. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 441–449
531   (2018).
532 52. Evans, P. R. An introduction to data reduction: space-group determination, scaling and
533   intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 282–292 (2011).
534 53. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta
535   Crystallogr. Sect. D Biol. Crystallogr.* **66**, 133–144 (2010).
536 54. Foadi, J. *et al.* Clustering procedures for the optimal selection of data sets from
537   multiple crystals in macromolecular crystallography. *Acta Crystallogr. Sect. D Biol.
538   Crystallogr.* **69**, 1617–1632 (2013).
539 55. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing
540   with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485
541   (2010).
542
543
544

Methods

Sample preparation, data collection, and data processing were carried out as described in Takaba and Maki-Yonekura et al. (2023)[17] with some modifications.

Sample preparation for serial X-ray crystallography

In this study, we used a flat-faced polyimide plate with a size of 4×4 mm^2 and a thickness of 20 μm (Protein Wave Corporation). Microcrystal powder of rhodamine-6G was purchased from Tokyo Chemical Industry, and monopeptoid and tripeptoid were synthesized and crystallized as below. Crystals of organic semiconductor materials, Ph-BTBT-C10 and anti-BTBTT-C6 were prepared as described[31,39]. Monopeptoid and tripeptoid, which tended to aggregate and/or included larger grains ($\gtrsim$ 20 μm), were ground to fine grains between two glass slides. The raw or ground powder was suspended with low-viscosity liquid paraffin (Nacalai tesque) and spread over a polyimide plate. It was then sandwiched and held with another plate. The shape and distribution of the crystals were observed with an optical digital microscope, VHX-7000 (KEYENCE) on the plate or a glass slide (Fig. 2).

Data collection of serial X-ray crystallography

The prepared sample plate was fixed onto a sample mounter and vertically placed on a sample stage. The stage can move along the XYZ axes and rotate around the phi axis. The whole area of the sample plate was scanned with XFEL pulses by moving the stage in the XZ plane. For some compounds, the stage was also tilted during the data collection after the eucentric origin was adjusted. The tilt angle was changed stepwise by 5 or 10 degrees for the first trials. We then confirmed that the camera distance was robustly determined from these diffraction patterns. Thereby, continuous tilt was finally adopted.

26

569

570   The data collection described above was performed on beamlines BL2 and BL3 at the

571   SACLA XFEL facility[23,24]. The photon energy of XFEL and the beam size at the sample plane

572   were adjusted to 15.0 keV and ~ 1 µm, respectively. The pulse duration was ~7 fs and the

573   repetition rate was 30 Hz. The pulse energy in the beam time was ~ 160 µJ/pulse at BL2 and ~

574   260 µJ/pulse at BL3. Diffraction patterns were recorded on an MX300-HS CCD detector

575   (Rayonix) placed 90-100 mm downward from the sample plane. The stage movement, shutter

576   of the beamline, and readout of data from the detector were synchronized and controlled with

577   a python-based script. All data were collected at room temperature.

578

579   <u>Data processing and structure determination from serial X-ray diffraction images</u>

580   CCD frames showing Bragg spots were identified using a diffraction data processing program

581   DIALS version 3.5.0 [43]. Only frames with 10-300 identified spots were packed into the HDF

582   format and processed with the crystFEL suite version 0.9.1 for indexing and integration of the

583   intensities[44]. The indexing was performed with the lattice parameters obtained from rotation

584   ED patterns. The numbers of detected spots in the total frames and frames included in the

585   merge were associated with the geometry in the data collection recorded by the SACLA DAQ

586   system[32] and further used to evaluate the data collection efficiency. The integrated intensities

587   were used for ab initio phasing by SHELXT[45] and the obtained initial structures were refined

588   with SHELXL[46]. Hydrogen atoms were generated during the refinement as a riding model.

589

590   <u>Reconstruction of 1D profiles from SX diffraction patterns</u>

591   The 1D diffraction profiles represented in Supplementary Fig. 2 is reconstructed in the same

592   manner as described in Schriber et al. (2022)[22]. The position of Bragg spots found by DIALS

27

593 were converted to d spacing using the nominal XFEL photon energy (15.0 keV). The measured

594 d spacing were aligned into histogram that amounted to a sharpened powder diffraction pattern.

595 About 20 strong peaks in the pattern were selected as input indexing with GSASII[47]. For

596 monopeptoid, the candidate monoclinic lattices raised with a starting cell volume of 600 $Å^3$ are

597 listed in Supplementary Table 2 with the figure of merit defined as M^{*}[22]. The 4th lattice

598 highlighted in the table ($a/b/c/\beta$ [Å, deg.] = 6.47/10.60/13.09/94.59) exhibited a similar value

599 to the reference parameters used for indexing with crystFEL described above.

600

601 <u>Electron crystallography</u>

602 Microcrystals from the same sample batch as for SX were suspended in Novec7100 (3M) and

603 spread on a 200-mesh copper grid (Maxtaform) covered with holey carbon film (Quantifoil) or

604 simply attached electrostatically on the carbon film. The grids were immersed in liquid

605 nitrogen and transferred into a CRYO ARM 300 electron microscope (JEOL) operated at an

606 accelerating voltage of 300 kV under a specimen temperature of ~93 K. Semi-automated data

607 acquisition of rotational ED patterns was carried out by combined use of SerialEM[48] and

608 ParallEM[49,50] as previously described[14]. 33-156 rotation series were collected on an XF416

609 scintillator-coupled detector (TVIPS) for monopeptoid and a DE64 direct detection detector

610 (Direct Electron) for the other samples, and Debye-Scherrer patterns of gold sputtered on a

611 carbon film were measured to calibrate the camera distance at the end of each data collection

612 session. The camera distance and beam centre were determined from the spacing and centre

613 position of the measured gold rings. The diffraction spots were indexed and integrated by

614 DIALS. The reduced datasets were grouped and sorted by KAMO[51], which carried out scaling

615 and merging by using Pointless[52], XSCALE[53] and BLEND[54]. All the merged clusters were

28

616   subjected to phasing by SHELXT and SHELXD[55]. If this process gave initial structures, they

617   were refined with SHELXL.

618

619   The common lattice patterns derived from the ED datasets were used as initial parameters

620   for indexing the SX data as described above. The direct phasing failed for the ED patterns of

621   tripeptoid alone, whereas the structure of the same crystal was solved from the SX data indexed

622   with the ED lattice parameters. Then, the SX structure of tripeptoid was used as an initial model

623   for the ED data, and the obtained ED map was able to resolve the structure representing the

624   molecular formula of tripeptoid.

625

626   <u>Synthesis and crystallization of peptoids</u>

627   Acetyl-L-alanine piperazine, *N*-Boc was synthesized and crystallized as described in Morimoto

628   et al. (2023)[37], and referred to monopeptoid in this report. Tripeptoid was synthesized as

629   follows. *N*-Cbz-L-alanine (2.2 g, 10 mmol) was dissolved in 50 mL of tetrahydrofuran (THF)

630   and iodomethane (5.0 mL, 80 mmol, 8.0 equiv.) was added. After the solution was cooled on

631   ice, sodium hydride (60% in oil, 1.2 g, 30 mmol, 3.0 equiv.) was added. The solution was

632   allowed to reach room temperature and stirred overnight. Water was added and the solution

633   was adjusted to pH 2 using 2 M HCl. The aqueous solution was extracted with ethyl acetate

634   (EtOAc) three times and the organic phase was dried over Na_2SO_4. The solvent was removed

635   under reduced pressure. The crude product and DMT-MM (3.0 g, 11 mmol, 1.1 equiv.) were

636   dissolved in 10 mL of methanol and piperidine (1.5 mL, 15 mmol, 1.5 equiv.) was added. The

637   solution was stirred at room temperature overnight. After evaporation, water was added, and

638   the product was extracted with DCM three times. The organic phase was dried over Na_2SO_4.

639   The solvent was removed by evaporation. The crude product, palladium 10% on carbon (0.36

640　g), and methanol were added to a recovery flask. The flask was charged with H_2 and the mixture

641　was stirred for 3 days. The reaction mixture was filtered through celite. The solvent was

642　removed under reduced pressure. Fmoc-*N*-methylalanine (0.33 g, 0.48 mmol, 1.2 equiv.) and

643　triphosgene (48 mg, 0.16 mmol, 0.40 equiv.) were dissolved in 2 mL of THF, and *N,N*-

644　diisopropylethylamine (209 µL, 1.2 mmol, 3.0 equiv.) was added to the solution. After 1 min,

645　the product from the previous reaction (0.16 g, 0.40 mmol) in 3 mL THF was added to the

646　solution and stirred for 1 h. The reaction was quenched by adding saturated NH_4Cl solution.

647　The solution was extracted with EtOAc three times. After the solvent was evaporated, the

648　product was dissolved in 4 mL of THF and piperidine (0.36 mL, 3.6 mmol, 9.0 equiv.) was

649　added to the solution. The solution was stirred at room temperature for 30 min and the solvent

650　was removed under reduced pressure. The coupling reaction of Fmoc-*N*-methylalanine and

651　Fmoc deprotection were repeated once more. The product (58 mg, 0.17 mmol) was dissolved

652　in 2 mL of DCM, and acetic anhydride (80 mL, 0.85 mmol, 5.0 equiv.) and pyridine (68 mL,

653　0.85 mmol, 5.0 equiv.) were added to the solution. The solution was stirred for 30 min and the

654　solvent was removed under reduced pressure. The residue was purified by silica gel column

655　chromatography to give tripeptoid (56 mg, 0.15 mmol, 86%). HRMS (ESI-TOF MS) *m/z*:

656　[M+Na]$^+$ Calcd for $C_{19}H_{34}N_4O_4Na^+$ 405.2472; Found 405.2478. UPLC chromatogram is

657　shown in Supplementary Fig. 6. Tripeptoid dissolved in DCM and methanol was dried under

658　reduced pressure. Small crystals appeared in the container.

659

1
<p style="text-align:center">Supplementary Information for</p>
2
3
<p style="text-align:center">Comprehensive application of XFEL micro crystallography for novel organic compounds</p>
4
5
6
7 Kiyofumi Takaba, Saori Maki-Yonekura, Ichiro Inoue, Kensuke Tono, Yasuhiro Fukuda, Yota
8 Shiratori, Yiying Peng, Jumpei Morimoto, Satoru Inoue, Toshiki Higashino, Shinsuke Sando,
9 Tatsuo Hasegawa, Makina Yabashi and Koji Yonekura

10
11 Correspondence to: yone@spring8.or.jp
12
13
14 **This document file includes:**
15
16 Supplementary Tables 1 and 2
17 Supplementary Figures 1 to 6
18

19

20

21   Supplementary Table 1 **Crystallographic parameters, and data and refinement statistics for the examined organic compounds by**

22   **ED.**

	monopeptoid	tripeptoid	Ph-BTBT-C10	anti-BTBTT-C6	rhodamine 6g
Source	cryoARM300	cryoARM300	cryoARM300	cryoARM300	cryoARM300
High-tension voltage (kV)	300	300	300	300	300
Beam current (e^- Å^{-1} frame^{-1})	0.01	0.01	0.005	0.01	0.01
Calibrated camera distance (mm)	1,168	803	645	811	1,068
Temperature (K)	~93	~93	Room temperature	~93	Room temperature
Spacegroup	$P2_1$	$P2_1$	$P2_1/a$	$P2_1/c$	$Pbca$
Unit Cell $a/b/c(/\beta)$(Å, deg.)	6.5/11/13/94	5.8/6.9/53/90	5.9/7.5/51/93	13/7.5/47/97	15/15/23
V_{cell} (Å^3)	891	2118	2271	4439	4853
Z^a	2	4	4	8	8
No. of crystals measured/merged	48/3	156/30	33/7	96/38	30/17
Completeness (%)	97.3 (98.1)	94.6 (93.8)	78.5 (80.0)	90.0 (90.4)	100.0 (100.0)
Multiplicity	6.16 (6.26)	31.9 (31.6)	13.6 (13.7)	86.9 (90.2)	81.6 (84.5)
R_{merge} (%)	51.9 (96.9)	39.5 (200.6)	33.4 (93.3)	47.1 (69.6)	62.0 (372.4)
$CC_{1/2}$ (%)b	93.7 (18.2)	98.3 (12.4)	98.8 (88.6)	97.8 (92.1)	99.6 (56.0)
$<I/\sigma(I)>^c$	2.42 (1.04)	7.32 (1.33)	4.2 (0.5)	9.96 (4.66)	8.97 (0.83)
d_{min} (Å)	1.0	0.9	0.8	0.9	0.9
No. of parameters refined	243	506	290	560	310
R_1 ($F_o > 4\sigma$), wR_2 (all F_o)	0.172, 0.386	0.242, 0.534	0. 242, 0.617	0.250, 0.668	0.161, 0.378

Peak, hole (e⁻/Å)	0.20, -0.21	0.35, -0.30	0.58, -0.32	0.68, -0.36	0.22, -0.16
<e.s.u.> for bond lengths (Å)[d]	0.092	0.084	0.036	0.038	0.011
<e.s.u.> for bond angles (deg.)	6.709	5.619	2.060	2.384	0.792
Reference	This work	This work	Takaba et al., 2020	This work	Takaba & Maki-Yonekura et al., 2023

23   [a] Z, formula units in unit cell [b] $CC_{1/2}$, the Pearson correlation coefficient between two half sets of intensities [c] I, measured diffraction

24   intensity [d] e.s.u., estimated standard uncertainties calculated from full-matrix refinement

25

26   Supplementary Table 2 **Candidates of lattice parameters for monopeptoid.** The highlighted 4th cell is the correct one and it corresponds

27   to the cell derived from ED.

	M20	Bravais	a	b	c	alpha	beta	gamma	volume	FOM (M^*)
1	9.76	P2/m	8.5425	4.3302	19.6501	90.0	98.69	90.0	718.547	46.43
2	12.88	P2/m	17.2963	2.8573	22.3735	90.0	107.56	90.0	1054.169	27.36
3	14.17	P2/m	8.5025	8.5900	10.1590	90.0	104.56	90.0	718.162	26.29
4	15.70	P2/m	6.4683	10.5957	13.0782	90.0	94.59	90.0	893.459	25.04
5	17.77	P2/m	10.1834	13.9978	6.4325	90.0	92.91	90.0	915.747	24.94
6	17.77	P2/m	12.3180	13.9978	6.4325	90.0	124.35	90.0	915.747	24.92
7	17.92	P2/m	6.4890	10.7912	12.7547	90.0	95.67	90.0	888.763	22.35
8	14.99	P2/m	10.0744	13.4690	8.5375	90.0	105.70	90.0	1115.250	22.13
9	18.85	P2/m	12.8822	5.0899	10.4762	90.0	91.71	90.0	686.605	21.31

10	20.65	P2/m	12.9422	21.3752	3.8197	90.0	92.38	90.0	1055.782	21.20
11	18.19	P2/m	8.6067	4.9472	19.5974	90.0	99.27	90.0	823.559	20.35
12	18.19	P2/m	8.6067	4.9472	19.5974	90.0	99.27	90.0	823.559	20.35
13	18.24	P2/m	5.1117	13.4850	21.4089	90.0	104.41	90.0	1429.340	18.13
14	18.24	P2/m	5.1117	13.4850	20.7369	90.0	90.59	90.0	1429.340	18.12
15	17.45	P2/m	16.3375	4.2467	16.8612	90.0	101.95	90.0	1144.499	17.65
16	18.11	P2/m	8.6131	6.4484	21.8847	90.0	95.01	90.0	1210.846	17.63
17	19.47	P2/m	12.8876	3.3672	21.6095	90.0	90.77	90.0	937.663	17.06
18	23.21	P2/m	5.7615	6.4534	24.7010	90.0	91.75	90.0	917.980	16.67
19	28.42	P2/m	7.6269	24.6687	5.2645	90.0	98.39	90.0	979.894	16.57
20	30.32	P2/m	21.6443	2.9750	13.2662	90.0	92.25	90.0	853.579	12.24

28

29


30
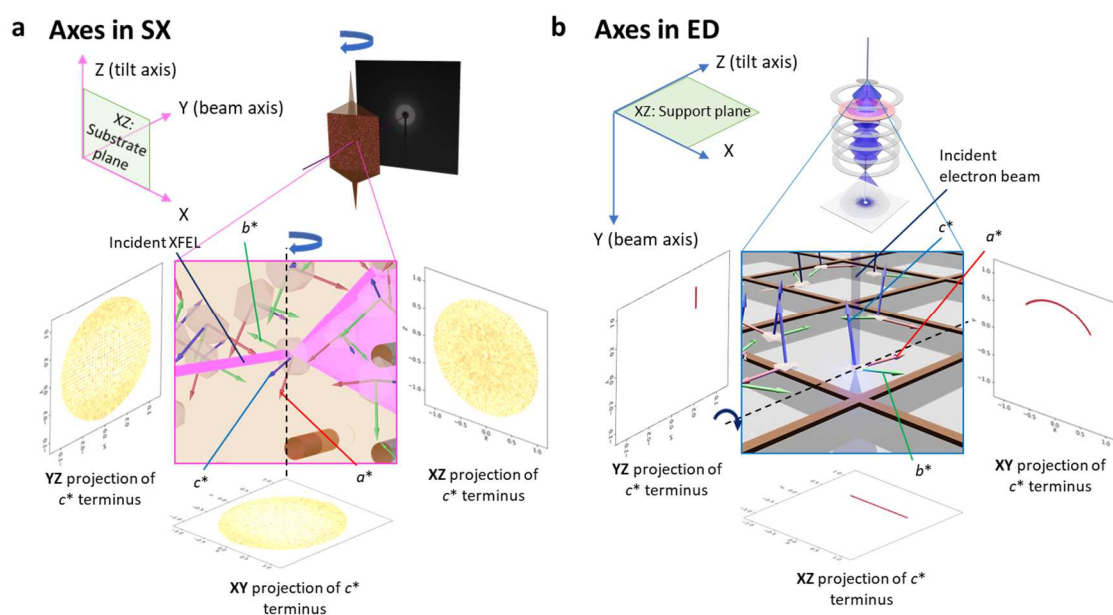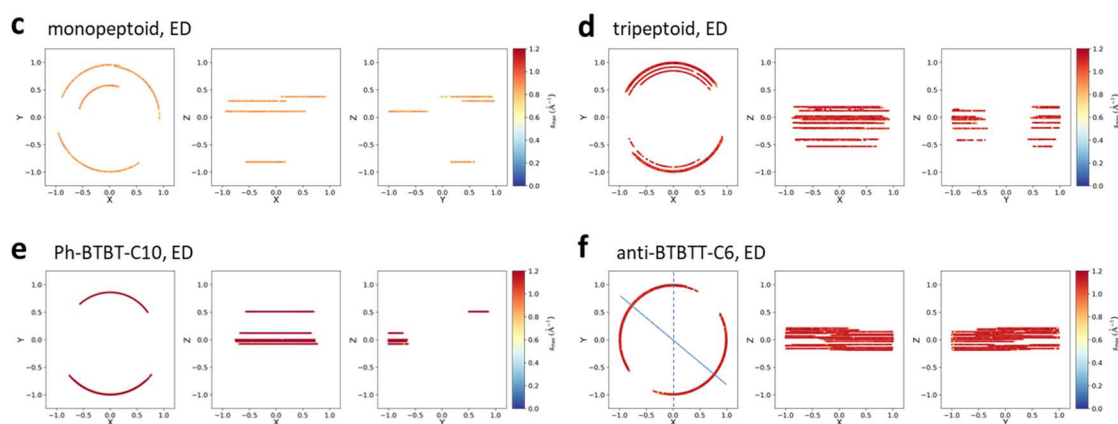
Supplementary Fig. 1. **Schematic drawings of the data collection geometry and distributions of crystal orientations for the ED data.**

(**a, b**) Schematic drawing showing the data collection geometry for SX (a) and ED (b) and the spatial relationships in projection diagrams in c-f, Figs. 5a-d, and Supplementary Figs. 4a and b. (**c**) Orientation distributions of the monopeptoid crystals on the x–y (left), x–z (middle) and y–z (right) planes in the ED data. (**d**) Orientation distributions of the tripeptoid crystals in the ED data. (**e**) Distributions of the Ph-BTBT-C10 crystals. (**f**) Distributions of the anti-BTBTT-C6 crystals. The beam is along the y axis. The colour scheme in (c – f) represents the s_{max} per

1

39   frame. The curves in the x-z projection indicates the stage rotation in a series of data collection.

40   The solid and broken lines, angled by 50 degree, in (f) corresponds to the electron beam and a

41   typical normal direction of the grid plane for the case of anti-BTBTT-C6.

42
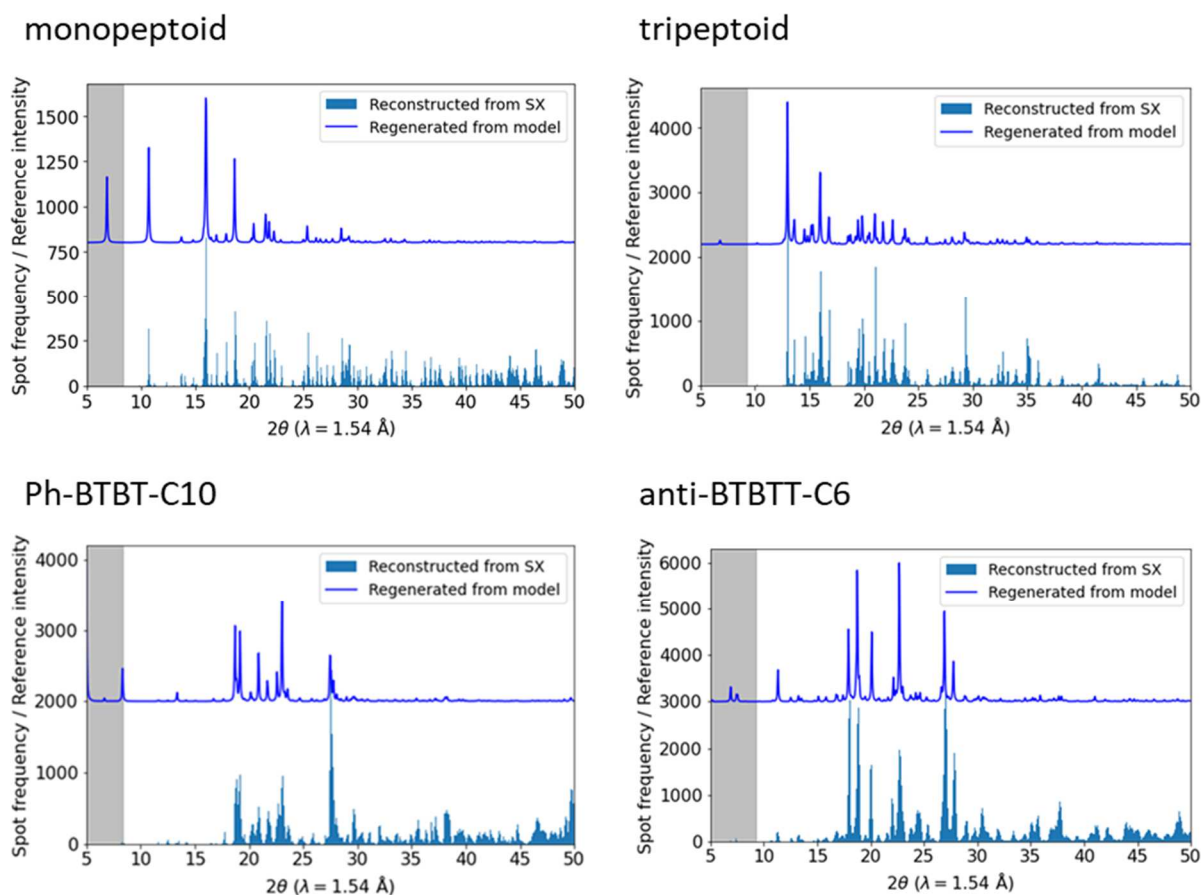
Supplementary Fig. 2. **Reconstructed 1D-pattern from the SX images.**

Histograms over all the recorded frames of individual detected spots, identified with the scatter angles (2θ). The calculated pattern from the determined model is overlaid in the upper part of each panel. The 2θ values are scaled to the wavelength of a general X-ray source (Cu Kα, λ=1.54 Å), according to Bragg's law (λ=2$d$sinθ). The undetectable region behind the beamstop is indicated as gray shadows in the histograms.

3

50
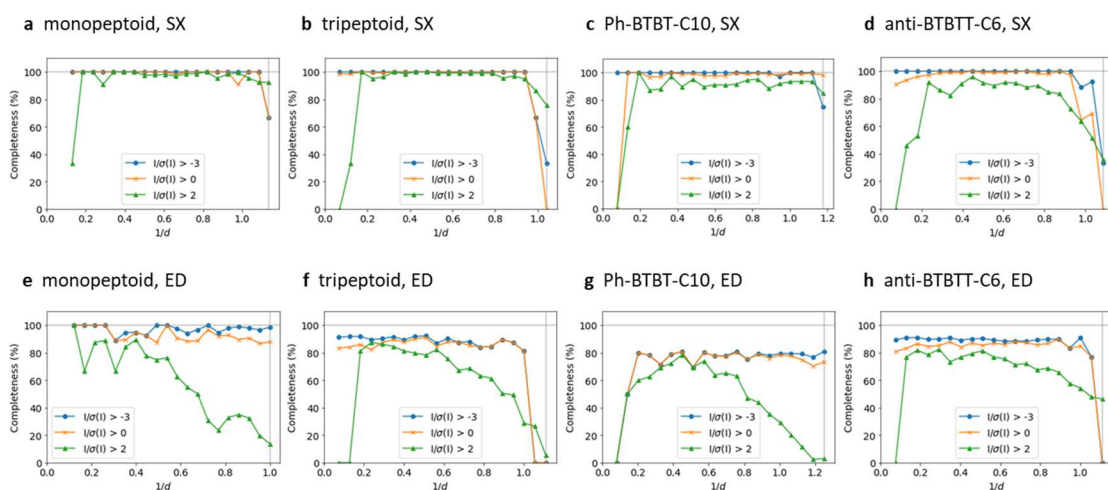
Supplementary Fig. 3. **Plots of the completeness in the SX and ED datasets.**

Plots of completeness in SX and ED data along the special frequency. The three series of plots

are respectively constructed from the limited diffraction datasets by the signal-to-noise ratio

($I/\sigma(I)$) of the individual reflections. (a-d) for SX and (e-h) for ED datasets. (a, e) monopeptoid.

(b, f) tripeptoid. (c, g) Ph-BTBT-C10. (d, h) anti-BTBTT-C6.

4
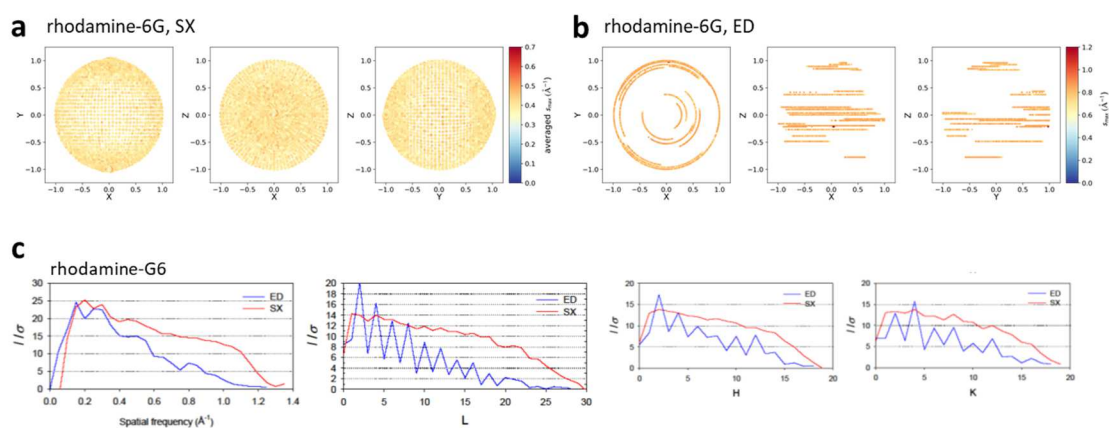
Supplementary Fig. 4. **Distributions of crystal orientations and plots of mean signal-to-noise-ratios, $<I/\sigma(I)>$ for the SX and ED data of rhodamine-6G.**

(**a**) Orientation distributions of the crystals in the SX data. (**b**) Distributions of the crystals in the ED data. The size and colour scheme in the SX data (a) is the same as in Figs. 5a - d and the colour in ED data (b) is the same as in Supplementary Fig. 1c – f. The traces represented as curves and lines in the projection in (b) indicates the sequential diffraction frames collected with the stage rotation. The data collection geometry for the SX and ED data are shown in Supplementary Figs. 1a and b, respectively. (**c**) Plots of $<I/\sigma(I)>$ in the SX and ED data of the rhodamine-6G crystals along the spatial frequency, while those along Miller indices, H, K and L axes.
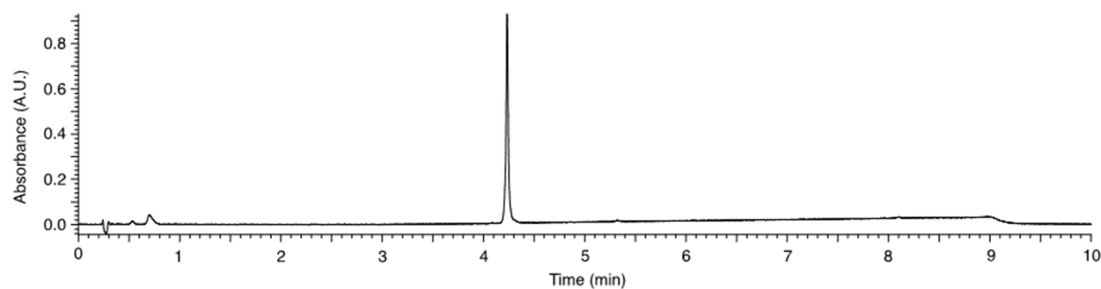
5
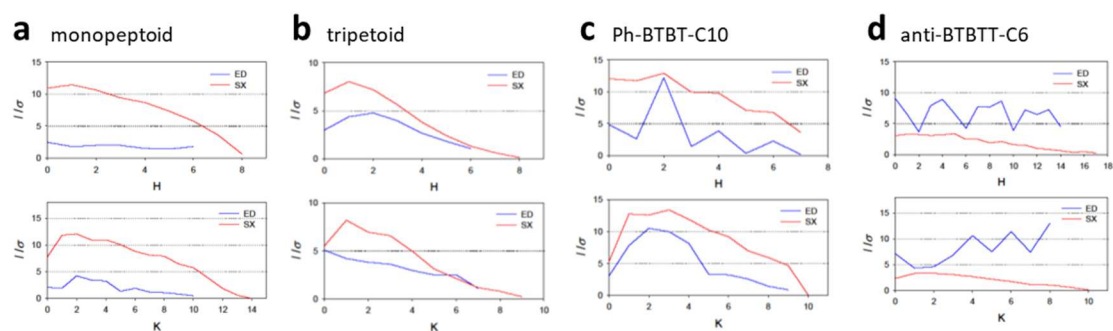
69

70 Supplementary Fig. 5. **Plots of mean signal-to-noise-ratios, $\langle I/\sigma(I)\rangle$, for the SX and ED**

71 **data along the Miller indices, H, K.**

72 (**a**) Plots of $\langle I/\sigma(I)\rangle$ in the SX and ED data of the monopeptoid crystals along Miller indices,

73 H and K axes. (**b**) Plots of the tripeptoid crystals. (**c**) Plots of the Ph-BTBT-C10 crystals. (**d**)

74 Plots of the anti-BTBTT-C6 crystals.

75

76

77 Supplementary Fig. 6. **UPLC chromatogram of tripeptoid.**

78 The product was analyzed on UPLC monitored at 220 nm. UPLC analysis was performed using

79 a linear gradient of solvent A (water containing 0.1% TFA) and solvent B (acetonitrile

80 containing 0.1% TFA).

6