

# SCIENTIFIC REPORTS



OPEN

## Application of profile fitting method to neutron time-of-flight protein single crystal diffraction data collected at the iBIX

Received: 02 August 2016

Accepted: 17 October 2016

Published: 01 December 2016

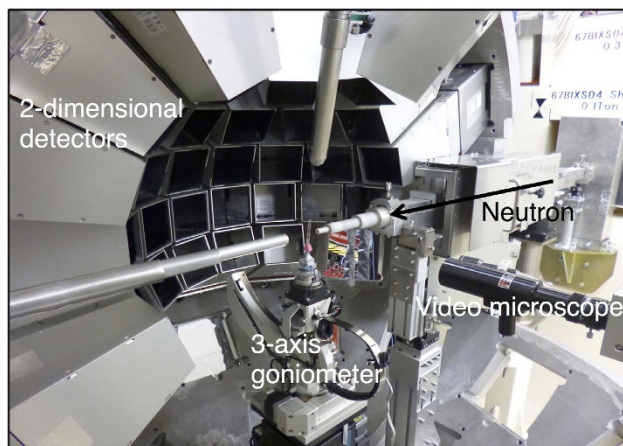
Naomine Yano<sup>1</sup>, Taro Yamada<sup>1</sup>, Takaaki Hosoya<sup>1,2</sup>, Takashi Ohhara<sup>3</sup>, Ichiro Tanaka<sup>1,2</sup> & Katsuhiko Kusaka<sup>1</sup>

We developed and employed a profile fitting method for the peak integration of neutron time-of-flight diffraction data collected by the IBARAKI Biological Crystal Diffractometer (iBIX) at the Japan Proton Accelerator Research Complex (J-PARC) for protein ribonuclease A and  $\alpha$ -thrombin single crystals. In order to determine proper fitting functions, four asymmetric functions were evaluated using strong intensity peaks. A Gaussian convolved with two back-to-back exponentials was selected as the most suitable fitting function, and a profile fitting algorithm for the integration method was developed. The intensity and structure refinement data statistics of the profile fitting method were compared to those of the summation integration method. It was clearly demonstrated that the profile fitting method provides more accurate integrated intensities and model structures than the summation integration method at higher resolution shells. The integration component with the profile fitting method has already been implemented in the iBIX data processing software STARGazer and its user manual has been prepared.

The neutron time-of-flight (TOF) method, which is similar to the X-ray Laue method, uses pulsed neutrons having continuous wavelengths. Because neutron velocities depend on the neutron wavelength, the flight times of the neutrons from their sources (the moderator) through the sample and to the detectors vary. Thus, we can obtain the neutron wavelength by measuring the flight times and separate diffraction peaks at different wavelengths using fixed detectors. In this regard, the TOF method is more efficient than the monochromatic method and can reduce the measurement time<sup>1</sup>. The IBARAKI Biological Crystal Diffractometer (iBIX)<sup>2</sup> at the Japan Proton Accelerator Research Complex (J-PARC)<sup>3,4</sup>, the Protein Crystallography Station (PCS)<sup>5</sup> at the Los Alamos Neutron Science Center (LANSCE)<sup>6</sup>, and the Macromolecular Neutron Diffractometer (MaNDi)<sup>7,8</sup> at the Spallation Neutron Source (SNS)<sup>9</sup> are TOF neutron diffractometers for protein single crystals. iBIX, PCS, and MaNDi have installed a H<sub>2</sub> coupled moderator (CM), partially coupled moderator, and H<sub>2</sub> decoupled moderator (DM), respectively.

The iBIX, which is installed on beam line BL03 at the Materials and Life Science Experimental Facility (MLF) of J-PARC, is a neutron TOF single-crystal diffractometer used mainly for elucidating the hydrogen, protonation, and hydration structures of biological macromolecules in various life processes. It is equipped with 30 time-resolved scintillator area detectors<sup>10</sup>, each with active areas of 133 × 133 mm (Fig. 1). The flight path lengths from the H<sub>2</sub> CM to the sample and from the sample to the detectors are 40 m and 490 mm, respectively<sup>2</sup>. At MLF BL03, fast neutrons are generated by the spallation of mercury by 3-GeV proton bombardment<sup>11</sup>, and become thermal neutrons after repeated collisions with hydrogen atoms in the CM. Although most of the moderated neutrons fly toward the sample, some neutrons return to the mercury target. Later, they fly back into the CM and are emitted to the sample. This causes significant broadening of the neutron pulse, leading to an asymmetrically shaped neutron pulse in the direction of the TOF axis<sup>1</sup>. However, the intensities of pulsed neutrons from the CM are stronger than those from DM and poisoned decoupled moderators (PM)<sup>11</sup>.

<sup>1</sup>Frontier Research Center for Applied Atomic Sciences, Ibaraki University, 162-1 Shirakata, Tokai, Ibaraki 319-1106, Japan. <sup>2</sup>College of Engineering, Ibaraki University, 4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan. <sup>3</sup>Neutron Science Section, J-PARC Center, Japan Atomic Energy Agency, 2-4 Shirakata-Shirane, Tokai, Ibaraki 319-1195, Japan. Correspondence and requests for materials should be addressed to N.Y. (email: naomine.yano.fron@vc.ibaraki.ac.jp)



**Figure 1.** The inside view of iBIX. 30 two-dimensional detectors are installed. The active area of each detector is  $133 \times 133$  mm and the distance from the sample to the detectors is 490 mm.

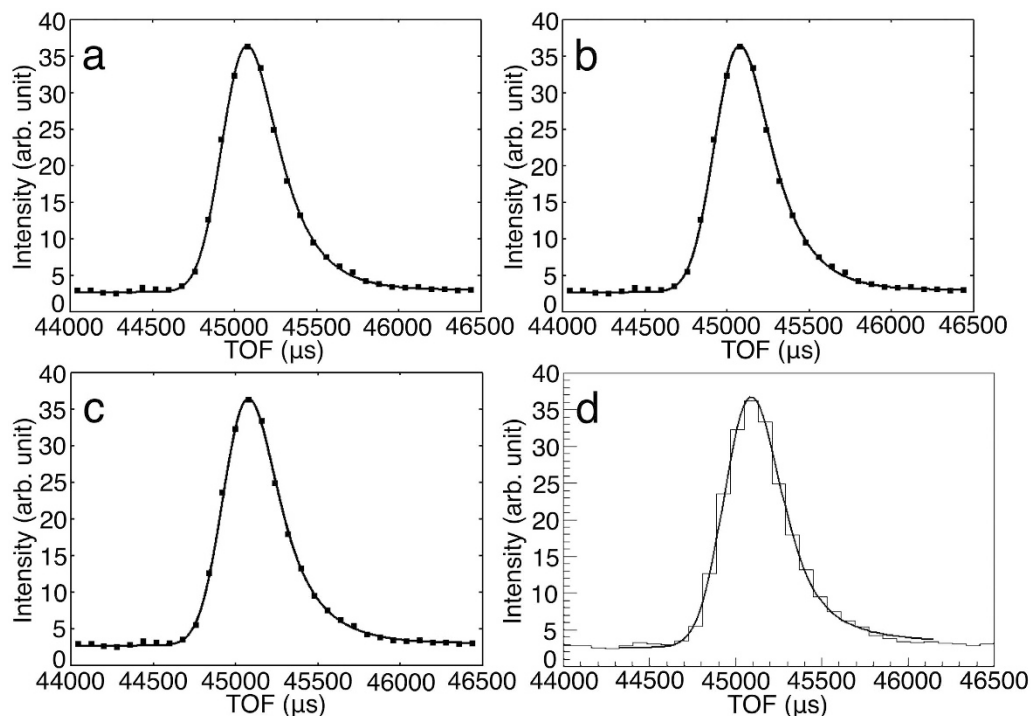
The peak intensities of neutron diffraction from protein single crystals are relatively weak because neutron beam intensity is lower than X-ray synchrotron intensity, and proteins have lower crystallinity than organic or inorganic compound crystals. Thus, some weak peaks are hidden under the error of strong background generated by strong incoherent scattering of hydrogen atoms in the protein crystals. TOF neutron protein crystallography (NPC) diffraction data processing software has been developed at each diffractometer facility independently. STARGazer<sup>12</sup>, d<sup>4</sup>TREK modified for wavelength-resolved Laue neutron crystallography<sup>13</sup>, and Mantid<sup>14</sup> are used at iBIX, PCS, and MaNDi, respectively. Overall, the data quality indices  $R_{\text{merge}}$  and  $R_{\text{pim}}$ <sup>15</sup> of TOF NPC diffraction data are higher than monochromatic synchrotron X-ray and nuclear reactor neutron diffraction data<sup>16–24</sup>. In comparison to monochromatic synchrotron X-ray and nuclear reactor neutron diffraction data, the data processing TOF NPC diffraction data can be improved, and the lowering of  $R_{\text{merge}}$  and  $R_{\text{pim}}$  is a common problem in all neutron TOF single-crystal diffractometers. Therefore, effective methods to determine accurate integrated peak intensities are essential for TOF NPC structural analysis. It has been demonstrated that the application of profile fitting methods to monochromatic X-ray diffraction data can provide more accurate integrated intensities than summation integration methods, especially for weak peaks<sup>25–27</sup>. The profile fitting method for the  $x$  and  $y$  detector positions has been implemented in monochromatic X-ray single crystal data processing software such as DENZO<sup>28</sup>, MOSFLM<sup>29</sup>, and XDS<sup>30</sup>. However, it has not yet been demonstrated that this method can be suitably applied to TOF NPC diffraction analysis at iBIX, PCS, and MaNDi.

To the best of our knowledge, one example of applying profile fitting methods in reciprocal space ( $Q$  space) to the TOF single crystal analysis involved TOPAZ<sup>31</sup>, which is installed on BL 12 at the SNS with a PM. In this example, a Gaussian convolved with two back-to-back exponentials<sup>32</sup> was used as the fitting function, and was applied to the TOF neutron diffraction data of two single crystals (sapphire and natrolite) and one cocrystal (betaine, imidazole, and picric acid (BIPa)). Because this case did not include protein single crystals measured using a diffractometer with a CM, the same fitting function could not be applied to neutron diffraction data collected from the iBIX. The beam intensity with a CM is more favourable to TOF NPC than a DM or PM. Thus, we attempted to find an appropriate fitting function and develop a profile fitting algorithm for the integration method applicable to the full set of TOF NPC diffraction data collected at the iBIX.

## Results and Discussion

**Evaluation of asymmetric fitting functions.** TOF NPC diffraction data collected at iBIX are processed in detector space. Protein diffraction data normally contains more observed reflections than organic or inorganic compounds because of the larger unit cell. Data processing time is an important factor when choosing a profile fitting algorithm. In order to avoid the processing time of transforming from detector space to  $Q$  space, detector space was adopted in our profile fitting method. First, TOF neutron diffraction data containing the spatial position and TOF for each neutron are converted to histogram data. The number of observed neutrons at  $x$ ,  $y$ , and  $t$  corresponding to the horizontal and vertical detector positions ( $x$ ,  $y$ ) and a time-of-flight channel ( $t$ ) are recorded. The procedure after this is similar to monochromatic methods and includes peak searching, determination of initial UB matrix, peak indexing, refinement of UB matrix, and peak integration<sup>12</sup>. Because the number of observed neutrons at each pixel of the peak position of three-dimensional (3D) data was insufficient to fit functions accurately, we used projections of the 3D peak in the direction of the TOF axis to improve the statistical precision. Because the pulsed neutron shape was asymmetric when using a CM, the projections of the 3D peaks in the direction of the TOF axis were also asymmetric. Thus, asymmetric functions had to be used in the profile fitting method.

On the other hand, although the procedure is different, the profile fitting method has also been used in TOF neutron powder diffraction studies (Rietveld method<sup>33</sup>), and many asymmetric fitting functions were proposed such as the pseudo-Voigt function convolved with two back-to-back exponentials<sup>34</sup>, and the Ikeda–Carpenter function<sup>35</sup>. A CM is not required to avoid peak overlapping, and fitting functions are utilized not to integrate Bragg intensities but rather to calculate the parameters of the fitting functions using the Rietveld method. The



**Figure 2. Profile fitting to the  $hkl = -8\ 1\ 0$  peak from ribonuclease A crystal using four asymmetric functions.** (a) Gaussian convolved with two back-to-back exponentials fit. (b) Pseudo-Voigt function convolved with two back-to-back exponentials fit. (c) Gaussian convolved with Ikeda–Carpenter function fit. (d) Gaussian convolved with Landau function fit. In panels (a–c), SciPy was used to fit the functions and results were plotted by Gnuplot. In panel (d), because Gaussian convolved with Landau function contains convolution part in the equation, ROOT was used to fit the function and plot the results. Both four points of the outside regions of the integration region were used as the background region.

purpose of using fitting functions and the modulator type are different between powder diffraction profile fitting and TOF NPC diffraction profile fitting in iBIX; however, there is a possibility that the proposed fitting functions can be used in both situations.

In order to determine proper fitting functions for TOF NPC diffraction data, four asymmetric functions were evaluated using strong intensity peaks, with no overlapping, of TOF NPC diffraction data from ribonuclease A collected at the iBIX. These were a Gaussian convolved with two back-to-back exponentials, a pseudo-Voigt function convolved with two back-to-back exponentials, a Gaussian convolved with the Ikeda–Carpenter function derived from TOF profile functions in the GSAS program suite<sup>36</sup>, and a Gaussian convolved with the Landau function<sup>37</sup>. The number of parameters for each fitting function was 5, 6, 6, and 4, respectively. It is important to predict the peak positions and estimate the background accurately to determine the integrated intensity, especially for weak peaks. We adopted a linear function as the background function. Non-linear least square minimization was performed using SciPy or ROOT<sup>38</sup> software. All four asymmetric functions fit well with strong intensity peaks and significant differences were not found (Fig. 2). It was shown that fitting functions proposed in powder diffraction studies are applicable to TOF NPC diffraction data collected at iBIX. Although the Gaussian convolved with Landau function has the least number of parameters, it is not an analytical function. In order to reduce the processing time and the number of parameters, the Gaussian convolved with two back-to-back exponentials was selected as the most suitable fitting function and used in profile fitting method. This function is analytical and has five parameters:  $A$ ,  $\alpha$ ,  $\beta$ ,  $\sigma$ , and  $T_{ph}$ .

**Implemented fitting function in profile fitting.** Further details on the fitting function and its parameters are provided as follows.

$$\begin{aligned} H(\Delta T) &= \int G(\Delta T - \tau)E(\tau)d\tau \\ &= AN[e^u \operatorname{erfc}(y) + e^v \operatorname{erfc}(z)] \end{aligned} \quad (1)$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \quad (2)$$

$\operatorname{erfc}(x)$  is an intrinsic function of the `scipy.special` package in SciPy.

Here,

$$E(\tau) = 2Ne^{\alpha\tau} \text{ for } \tau < 0 \quad (3)$$

and

$$E(\tau) = 2Ne^{-\beta\tau} \text{ for } \tau > 0, \quad (4)$$

where  $\alpha$  and  $\beta$  are the rise and decay coefficients for the exponentials.  $A$  is an intensity scale parameter. The original function without  $A$  is normalized, and the integrated function area is 1.  $A$  refers to the integrated function area. The normalization factor,  $N$ , is

$$N = \frac{\alpha\beta}{2(\alpha + \beta)} \quad (5)$$

The Gaussian function is

$$G(\Delta T - \tau) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\Delta T - \tau)^2}{2\sigma^2}\right], \quad (6)$$

and the Gaussian variance is the coefficient  $\sigma^2$ .

The coefficients  $u$ ,  $v$ ,  $y$ , and  $z$  are

$$u = \frac{\alpha}{2}(\alpha\sigma^2 + 2\Delta T) \quad (7)$$

$$v = \frac{\beta}{2}(\beta\sigma^2 - 2\Delta T), \quad (8)$$

$$y = \frac{\alpha\sigma^2 + \Delta T}{\sqrt{2\sigma^2}}, \text{ and} \quad (9)$$

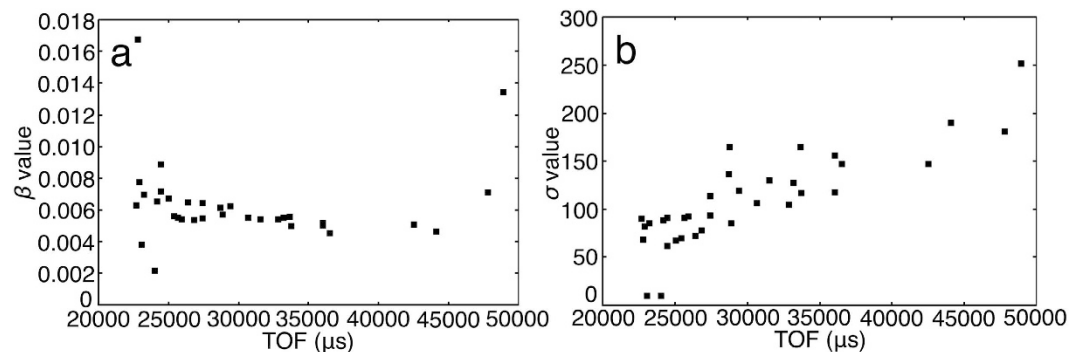
$$z = \frac{\beta\sigma^2 - \Delta T}{\sqrt{2\sigma^2}}. \quad (10)$$

$\Delta T$  is the difference in TOF between the peak position,  $T_{ph}$ , and the profile point,  $T$ .

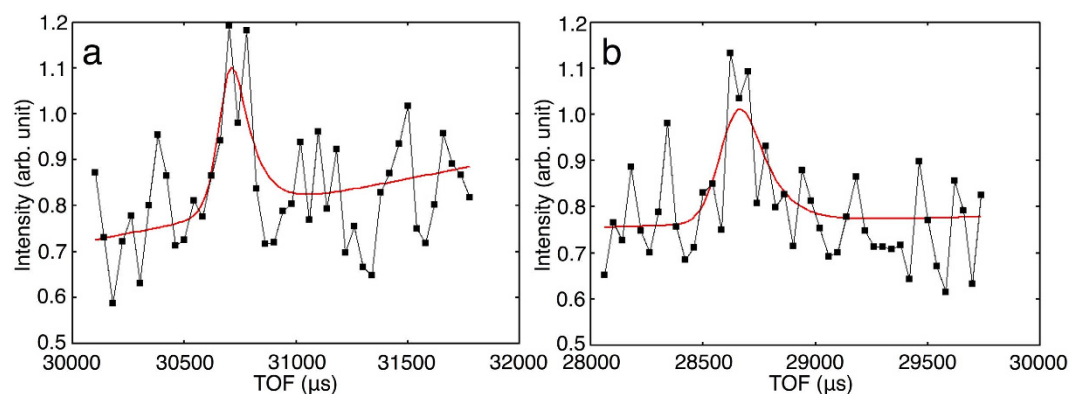
$$\Delta T = T - T_{ph} \quad (11)$$

The parameters  $\alpha$ ,  $\beta$ , and  $\sigma$  are related to the function profile and are not independent of each other.

**Correlation between fitting function parameters related to profile and TOF.** One of the assumptions in the profile fitting method is that peaks located within a region of reciprocal space or detector space have the same profile<sup>39</sup>. This is the key assumption when determining weak peak profiles. A projection profile in the direction of the TOF axis mainly depends on the TOF, scattering direction, and crystal orientation with respect to the beam. The active area of each detector is  $133 \times 133$  mm and the distance from the sample to the detectors is 490 mm. The maximum scattering angle difference among peaks within a detector is approximately  $15^\circ$ . iBIX contains 30 detectors that each measure different scattering angles. The scattering angle range of the lowest angle detector is from  $12.8^\circ$  to  $27.8^\circ$  and the  $Q$  range for a wavelength of  $4 \text{ \AA}$  is from  $0.35$  to  $0.75 \text{ \AA}^{-1}$  where  $Q = 4\pi(\sin\theta)/\lambda$  ( $\theta$  is half the scattering angle and  $\lambda$  is the wavelength). Additionally, the scattering angle range of the highest angle detector is from  $153.5^\circ$  to  $168.5^\circ$  and the  $Q$  range for a wavelength of  $4 \text{ \AA}$  is from  $3.06$  to  $3.12 \text{ \AA}^{-1}$ . Because the maximum scattering angle difference among peaks within a detector is approximately  $15^\circ$ , we assumed that peak profiles at one crystal orientation and one detector change depending only on TOF. In order to prove this assumption, we investigated whether there is a correlation between fitting function parameters related to profile and TOF. The parameters  $\alpha$ ,  $\beta$ , and  $\sigma$  are related to the profile. In order to reduce parameter error values,  $\alpha$  was fixed to a proper value and the  $A$ ,  $\beta$ ,  $\sigma$ , and  $T_{ph}$  parameters were calculated by non-linear least square minimization. Peaks whose  $I/\sigma(I)$  were over 5 and with lower parameter errors were selected and the  $\beta$  and  $\sigma$  parameter values were plotted against TOF at one crystal orientation and one detector. The result from the detector whose  $2\theta_{center}$  is  $51^\circ$  at one crystal orientation is shown in Fig. 3 as an example. The overall tendency regardless of detectors and crystal orientations is that  $\beta$  and  $\sigma$  are almost linearly dependent on TOF. Adjacent TOF peaks have similar  $\beta$  and  $\sigma$  values. Because  $\beta$  and  $\sigma$  are parameters related to profile shape, these plots show that adjacent TOF peaks have the same profile. Additionally, the maximum scattering angle difference among peaks within a detector is approximately  $15^\circ$ , so adjacent TOF peaks are located within a region of reciprocal space or detector space. We could confirm that the conditions of the profile fitting method are satisfied. The weaker peaks with adjacent TOF have larger parameter errors and the variances of  $\beta$  and  $\sigma$  in these peaks tend to be larger. In order to determine  $\beta$  and  $\sigma$  of weak peaks more accurately, adjacent TOF and accurate peaks were selected based on  $I/\sigma(I)$ , peak intensity, and parameter errors. Accurate peaks with larger TOF occasionally do not have similar  $\beta$  and  $\sigma$  values to Fig. 3a. Because there are few weak peaks adjacent to those peaks, we do not think it has a negative effect on determination of weak peak profiles.



**Figure 3. Plots of parameters related to peak profile against TOF in ribonuclease A neutron diffraction data.** The peaks whose  $I/\sigma(I)$  is over 5 and obtained by one detector located at  $51^\circ$  in  $2\theta_{\text{center}}$  and one crystal orientation were used. (a) Plot of parameter  $\beta$  against TOF. The peaks whose  $\beta$  errors are less than 1 were used. (b) Plot of parameter  $\sigma$ . The peaks whose  $\sigma$  errors are less than 10000 were used.



**Figure 4. Two examples of profile fitting to the weak peaks from ribonuclease A crystal using Gaussian convolved with two back-to-back exponentials.** Red solid line: fitting function. Black points and line: weak peak profile. These peaks were obtained by detector located at  $118^\circ$  in  $2\theta_{\text{center}}$  and one crystal orientation. Both four points of outside regions of the integration region were used as the background region. (a) Peak  $hkl$  is 11 3 16. (b) Peak  $hkl$  is 11 3 20.

**Application of profile fitting method to a full set of protein neutron diffraction data.** We developed a test program and applied it to a full set of ribonuclease A and  $\alpha$ -thrombin neutron diffraction data with non-overlapping peaks. The detailed procedure is described in the Methods section. The program could fit asymmetric functions to weak peaks as well as strong ones. The fitting results of two weak peaks obtained by the detector located at  $118^\circ$  in  $2\theta_{\text{center}}$  and one crystal orientation are shown as an example in Fig. 4. Intensity statistics, with peaks  $I/\sigma(I)$  larger than 0, were calculated using the iBIX data processing software STARGazer program<sup>12</sup>, and joint refinements of neutron and X-ray data were carried out using the PHENIX program<sup>40</sup>. In order to evaluate the algorithm utility, the intensity and refinement data statistics of the profile fitting method were compared to those of the summation integration method implemented in STARGazer by using the same integration regions, peaks and initial refinement model.

A summary of the intensity data and refinement results of ribonuclease A and  $\alpha$ -thrombin is provided in Tables 1 and 2. For both samples, more observed reflections could be used in the profile fitting method than in the summation integration method. This implies that the accuracy of a part of the negative intensity peaks on the summation integration method were improved by the profile fitting method. The higher the resolution, the lower the data quality indices  $R_{\text{merge}}$ ,  $R_{\text{pim}}$ ,  $R_{\text{work}}$ , and  $R_{\text{free}}$ <sup>41</sup> were on the profile fitting method than on the summation integration method. The greatest improvement in  $R_{\text{merge}}$ ,  $R_{\text{pim}}$ ,  $R_{\text{work}}$ , and  $R_{\text{free}}$  was observed in the highest resolution shell. The differences in ribonuclease A were 9.5%, 6.8%, 2.3%, and 1.7% and those in  $\alpha$ -thrombin were 7.2%, 6.2%, 3.8%, and 4.8%, respectively. Because higher resolution shells contain a higher percentage of weak peaks, it was shown that the profile fitting method is effective with weak peaks, as in previously reported results.  $R_{\text{merge}}$ ,  $R_{\text{pim}}$ ,  $R_{\text{work}}$ , and  $R_{\text{free}}$  in the overall resolution range were slightly improved over summation integration method. On the other hand, as peak intensities are stronger and background counts are relatively lower in the lower resolution shell, it seems that there is hardly any difference of  $R_{\text{merge}}$ ,  $R_{\text{pim}}$ ,  $R_{\text{work}}$ , and  $R_{\text{free}}$  between the profile fitting method and summation integration method.

Neutron data collection								
Space group	$P2_1$							
Cell dimensions of $a(\text{\AA}), b(\text{\AA}), c(\text{\AA}), \beta(^{\circ})$				30.29, 38.65, 53.36, 105.30				
Resolution ( $\text{\AA}$ )	Observed reflections		Independent reflections		$R_{\text{merge}}$		$R_{\text{pim}}$	
	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration
15.45–3.44	5,696	5,479	1,525	1,518	0.083	0.081	0.047	0.047
3.44–2.73	7,128	6,888	1,575	1,575	0.083	0.083	0.042	0.043
2.73–2.39	7,825	7,603	1,602	1,600	0.093	0.099	0.046	0.050
2.39–2.17	7,763	7,514	1,567	1,565	0.113	0.118	0.055	0.058
2.17–2.02	7,703	7,461	1,557	1,557	0.139	0.151	0.068	0.074
2.02–1.90	7,409	7,196	1,555	1,555	0.160	0.192	0.081	0.097
1.90–1.80	6,985	6,733	1,577	1,572	0.194	0.244	0.103	0.133
1.80–1.72	6,174	5,931	1,554	1,545	0.206	0.259	0.117	0.150
1.72–1.66	5,496	5,296	1,515	1,508	0.250	0.327	0.151	0.209
1.66–1.60	5,244	5,048	1,492	1,482	0.280	0.375	0.180	0.248
15.45–1.60	67,423	65,149	15,519	15,477	0.114	0.127	0.060	0.068
X-ray and neutron joint refinement								
Resolution ( $\text{\AA}$ )	No. of reflections (work/test)		$R_{\text{work}}$		$R_{\text{free}}$			
	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration		
15.37–2.88	2,478/130	2,472/128	0.180	0.179	0.210		0.211	
2.88–2.29	2,517/133	2,514/133	0.176	0.174	0.222		0.219	
2.29–2.00	2,503/141	2,502/141	0.147	0.152	0.174		0.183	
2.00–1.82	2,465/169	2,462/169	0.157	0.168	0.189		0.210	
1.82–1.69	2,442/119	2,431/118	0.167	0.180	0.209		0.222	
1.69–1.60	2,293/114	2,280/112	0.181	0.204	0.253		0.270	
15.37–1.60	14,698/806	14,661/801	0.170	0.175	0.207		0.213	

**Table 1. Statistics of neutron intensity data and X-ray and neutron joint refinement of ribonuclease A.**

<sup>†</sup>Cell constants were calculated from X-ray diffraction data.

**The reason for the success of the profile fitting method for iBIX data.** Two conditions for which the profile fitting method was accurate were proposed<sup>39</sup>. Firstly, peaks located within a region of reciprocal space or detector space have the same profile. Secondly, peak positions can be accurately predicted. It was shown that the first condition was satisfied (See Results and discussion: Correlation between fitting function parameters related to peak profile and TOF) and that weak peak parameters related to peak profile could be determined using more accurate peaks. Recently, detector parameters (distance between each detector and sample, detector position angles, etc.) and 3-axis goniometer offset angles were accurately calibrated and accuracy of peak position prediction was clearly improved. So, the second condition was satisfied. Additionally, an asymmetric fitting function that demonstrated good fitting of peak profiles was determined. In order to fit this function accurately to weak peaks by non-linear least square minimization, proper initial parameters corresponding to each peaks are essential. A procedure to determine proper initial parameter values was introduced (See Methods: Profile fitting procedure of test and implemented programs.) and most weak peaks whose  $I/\sigma(I)$  is less than 5 can be fitted using this function.

Because crystallinity of protein single crystals is relatively poor compared to that of inorganic and organic compounds, and iBIX has a CM installed, the projections of the 3D peaks in the direction of the TOF axis are wider and many fitting points can be prepared. It is advantageous for non-linear least square minimization to fit an asymmetric function. Thus, we think that the profile fitting method is more suitable for NPC data analysis obtained by diffractometers with a CM. For these reasons, it is concluded that the profile fitting method for iBIX data has been properly established and that the data statistics of the profile fitting method are an improvement over the summation integration method.

In this study, we definitively demonstrated for the first time that the profile fitting technique is also applicable to comparatively weak peaks and high-background TOF NPC diffraction data using a CM, and that the technique could improve data statistics. It seems that the established profile fitting method is applicable to NPC diffraction data collected from other diffractometers with CMs, and will be able to contribute to developing NPC.

The integration component with the profile fitting method has already been implemented in STARGazer and its user manual has been updated. The software and its manual are available for distribution to iBIX users.

**Future plan.** In the future, the accelerator power of J-PARC will be increased to 1000 kW, and we will be able to collect single crystal neutron diffraction data for larger unit cell crystals, including membrane proteins. Because iBIX was designed to measure samples with unit cells up to approximately  $135 \text{\AA}^3$ , problems caused by overlapping of adjacent peaks are expected. In order to solve those problems, we are attempting to apply the profile fitting technique to separate overlapped peaks in the TOF direction.

Neutron data collection								
Space group	C2							
Cell dimensions of $a(\text{\AA}), b(\text{\AA}), c(\text{\AA}), \beta(^{\circ})$				71.72, 72.72, 73.11, 100.51				
Resolution ( $\text{\AA}$ )	Observed reflections		Independent reflections		$R_{\text{merge}}$		$R_{\text{pim}}$	
	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration
20.64–4.51	3,394	3,259	1,738	1,712	0.053	0.053	0.041	0.042
4.51–3.59	4,901	4,661	1,978	1,954	0.058	0.060	0.040	0.042
3.59–3.13	5,391	5,132	2,046	2,023	0.081	0.087	0.054	0.058
3.13–2.85	5,937	5,607	2,033	2,012	0.124	0.137	0.078	0.089
2.85–2.65	5,492	5,122	2,008	1,971	0.154	0.167	0.101	0.112
2.65–2.49	4,464	4,170	1,892	1,851	0.178	0.202	0.125	0.145
2.49–2.36	3,922	3,659	1,814	1,784	0.203	0.234	0.149	0.177
2.36–2.26	3,058	2,904	1,631	1,586	0.226	0.272	0.174	0.215
2.26–2.18	3,072	2,870	1,610	1,557	0.248	0.289	0.192	0.224
2.18–2.10	2,788	2,646	1,514	1,496	0.250	0.322	0.196	0.258
20.64–2.10	42,419	40,033	18,264	17,946	0.103	0.110	0.072	0.079
X-ray and neutron joint refinement								
Resolution ( $\text{\AA}$ )	No. of reflections (work/test)		$R_{\text{work}}$		$R_{\text{free}}$			
	Profile fitting	Summation integration	Profile fitting	Summation integration	Profile fitting	Summation integration		
20.44–3.97	2,435/134	2,398/130	0.149	0.151	0.181	0.192		
3.97–3.15	2,756/130	2,724/132	0.146	0.149	0.186	0.197		
3.15–2.76	2,765/168	2,737/168	0.164	0.169	0.213	0.213		
2.76–2.50	2,619/146	2,557/144	0.160	0.166	0.221	0.221		
2.50–2.33	2,473/130	2,428/130	0.175	0.187	0.225	0.250		
2.33–2.19	2,185/99	2,111/96	0.192	0.216	0.274	0.281		
2.19–2.10	2,105/113	2,071/116	0.208	0.246	0.243	0.291		
20.44–2.10	17,338/920	17,026/916	0.163	0.171	0.209	0.220		

**Table 2.** Statistics of neutron intensity data and X-ray and neutron joint refinement of  $\alpha$ -thrombin.  $\ast$ Cell constants were calculated from X-ray diffraction data.

## Methods

**Preparation of ribonuclease A crystal and neutron and X-ray diffraction experiments.** Bovine pancreatic ribonuclease A was purchased from Sigma-Aldrich and crystallized as previously described<sup>2</sup>. A crystal with a volume of 6 mm<sup>3</sup> was soaked in D<sub>2</sub>O solution for 10 days in order to reduce background caused by incoherent scattering from hydrogen atoms. The neutron diffraction data were collected at room temperature over a wavelength range of 1.1 to 5.0  $\text{\AA}$ . Subsequently, X-ray diffraction data of the same crystal was collected at room temperature in the Photon Factory. The neutron diffraction data were processed using STARGazer<sup>12</sup> at a resolution of 1.6  $\text{\AA}$  and X-ray diffraction data were processed using HKL2000<sup>28</sup> at a resolution of 1.1  $\text{\AA}$ . Joint refinement of the neutron and X-ray data was performed in PHENIX<sup>40</sup>.

**Preparation of  $\alpha$ -thrombin crystals, and subsequent neutron and X-ray diffraction experiments.**  $\alpha$ -thrombin was purchased from Haematologic Technologies, Inc. (USA) and crystallized as previously described<sup>43</sup>. A crystal with a volume of 3.5 mm<sup>3</sup> was soaked in D<sub>2</sub>O solution for two weeks in order to reduce background caused by incoherent scattering from hydrogen atoms. The neutron diffraction data were collected at room temperature, using a wavelength range of 2.1 to 6.1  $\text{\AA}$ . Subsequently, the X-ray diffraction data of the same crystal was collected at room temperature in the Photon Factory. Neutron diffraction data were processed using STARGazer<sup>12</sup> at a resolution of 2.1  $\text{\AA}$  and the X-ray diffraction data were processed using HKL2000<sup>28</sup> at a resolution of 1.6  $\text{\AA}$ . Joint refinement of the neutron and X-ray data was performed in PHENIX<sup>40</sup>.

**Profile fitting procedure of test and implemented programs.** The profile fitting method was performed against peaks observed for every detector and every crystal orientation. A Gaussian convolved with two back-to-back exponentials and a linear function were adopted as fitting and background functions. The regions outside of the integration region in the TOF direction were defined as the background region. At first, the slope and the intercept of the background function were calculated using each peak by non-linear least square minimization. The parameters of the fitting function were calculated using background subtracted intensities by non-linear least square minimization. Proper initial parameter values are essential to obtain accurate results. If the initial parameter values are incorrect, accurate parameters cannot be calculated and incorrect integrated intensities will be obtained. The initial value of  $A$  is determined from the summation intensity of the projection profile in the direction of the TOF axis. If the obtained values are negative, then a positive value is given instead. Initial  $\beta$  and  $\sigma$  values are chosen based on previously successful examples. The predicted peak position TOF

is used as the initial  $T_{ph}$  value. The parameters  $\alpha$ ,  $\beta$ , and  $\sigma$  are related to the function profile and are not independent of each other. In order to reduce parameter error values,  $\alpha$  was fixed to a proper value and the  $A$ ,  $\beta$ ,  $\sigma$ , and  $T_{ph}$  parameters were calculated by non-linear least square minimization. The fitting function was integrated and  $I/\sigma(I)$  was determined at each peak. The strength of a peak was determined by  $I/\sigma(I)$ . For example, we can determine that the peaks whose  $I/\sigma(I)$  is over 5 are strong and those whose  $I/\sigma(I)$  is under 5 are weak. The main assumption in the profile fitting method is that peaks located within a region of reciprocal space or detector space have the same profile<sup>39</sup>. A projection profile in the direction of TOF axis mainly depends on the TOF, scattering direction, and crystal orientation. We assumed that peak profiles at one crystal orientation and one detector change depending only on TOF. In order to obtain more accurate  $\beta$  and  $\sigma$  values of weak peaks, the  $\beta$  and  $\sigma$  values of strong and adjacent TOF peaks selected based on  $I/\sigma(I)$ , peak intensity, and parameter error values were averaged, respectively, and were used as weak peak parameters.  $\alpha$ ,  $\beta$ , and  $\sigma$  were fixed, and  $A$  and  $T_{ph}$  were calculated by non-linear least square minimization. The fitting functions were integrated and the peak intensities corrected by the Lorentz factor were determined. These procedures were iterated for all detector and crystal orientations.

## References

- Niimura, N. & Podjarny, A. *Neutron protein crystallography. Hydrogen, protons, and hydration in bio-macromolecules*. 1–49 (Oxford University Press, 2011).
- Kusaka, K. *et al.* Evaluation of performance for IBARAKI biological crystal diffractometer iBIX with new detectors. *J. Synchrotron Rad.* **20**, 994–998 (2013).
- Ikeda, Y. Current status of 1 MW pulse spallation neutron source (JSNS) of J-PARC. *J. Nucl. Mater.* **343**, 7–13 (2005).
- Ikeda, Y. J-PARC status update. *Nucl. Instr. Meth. Phys. Res.* **A600**, 1–4 (2009).
- Langan, P., Greene, G. & Schoenborn, B. P. Protein crystallography with spallation neutrons: the user facility at Los Alamos Neutron Science Center. *J. Appl. Crystallogr.* **37**, 24–31 (2004).
- LANSCE into the Future. *Los Alamos Science*. **30** (2006).
- Coates, L., Stoica, A. D., Hoffmann, C., Richards, J. & Cooper, R. The macromolecular neutron diffractometer (MaNDi) at the Spallation Neutron Source, Oak Ridge: enhanced optics design, high-resolution neutron detectors and simulated diffraction. *J. Appl. Crystallogr.* **43**, 570–577 (2010).
- Coates, L. *et al.* The macromolecular neutron diffractometer MaNDi at the Spallation Neutron Source. *J. Appl. Crystallogr.* **48**, 1302–1306 (2015).
- Mason, T. E. *et al.* The Spallation Neutron Source: A powerful tool for materials research. arXiv:physics/0007068v1 (2000).
- Hosoya, T. *et al.* Development of a new detector and DAQ systems for iBIX. *Nucl. Instr. Meth. Phys. Res.* **A600**, 217–219 (2009).
- Maekawa, F. *et al.* First neutron production utilizing J-PARC pulsed spallation neutron source JSNS and neutronic performance demonstrated. *Nucl. Instr. Meth. Phys. Res.* **A620**, 159–165 (2010).
- Ohhara, T. *et al.* Development of data processing software for a new TOF single crystal neutron diffractometer at J-PARC. *Nucl. Instr. Meth. Phys. Res.* **A600**, 195–197 (2009).
- Langan, P. & Greene, G. Protein crystallography with spallation neutrons: collecting and processing wavelength-resolved Laue protein data. *J. Appl. Crystallogr.* **37**, 253–257 (2004).
- Arnold, O. *et al.* Mantid-Data analysis and visualization package for neutron scattering and  $\mu$  SR experiments. *Nucl. Instr. Meth. Phys. Res.* **A764**, 156–166 (2014).
- Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135 (2001).
- Unno, M. *et al.* Insights into the proton transfer mechanism of a bilin reductase PcyA following neutron crystallography. *J. Am. Chem. Soc.* **137**(16), 5452–5460 (2015).
- Yokoyama, T. *et al.* Protonation state and hydration of bisphosphonate bound to farnesyl pyrophosphate synthase. *J. Med. Chem.* **58**(18), 7549–7556 (2015).
- Nakamura, A. *et al.* “Newton’s cradle” proton relay with amide-imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Sci. Adv.* **1**(7), 1500263; 10.1126/sciadv.1500263 (2015).
- Chen, J. C. *et al.* Room-temperature ultrahigh-resolution time-of-flight neutron and X-ray diffraction studies of H/D-exchanged crambin. *Acta Crystallogr. F-Struct. Biol. Cryst. Commun.* **68**, 119–123 (2012).
- Chen, J. C., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. Direct observation of hydrogen atom dynamics and interactions by ultrahigh resolution neutron protein crystallography. *Proc. Natl. Acad. Sci. USA*. **109**(38), 15301–15306 (2012).
- Fisher, S. Z., Aggarwal, M., Kovalevsky, A. Y., Silverman, D. N. & McKenna, R. Neutron diffraction of acetazolamide-bound human carbonic anhydrase II reveals atomic details of drug binding. *J. Am. Chem. Soc.* **134**(36), 14726–14729 (2012).
- Vandavasi, V. G. *et al.* Exploring the mechanism of  $\beta$ -lactam ring protonation in the class A  $\beta$ -lactamase acylation mechanism using neutron and X-ray crystallography. *J. Med. Chem.* **59**(1), 474–479 (2016).
- McFeeters, H., Vandavasi, V. G., Weiss, K. L., Coates, L. & McFeeters, R. L. Neutron diffraction analysis of *Pseudomonas aeruginosa* peptidyl-tRNA hydrolase 1. *Acta Crystallogr. F-Struct. Biol. Cryst. Commun.* **72**, 220–223 (2016).
- Langan, P. *et al.* Evolution and characterization of a new reversibly photoswitching chromogenic protein, Dathail. *J. Mol. Biol.* **428**, 1776–1789 (2016).
- Diamond, R. Profile analysis in single-crystal diffractometry. *Acta Crystallogr.* **A25**, 43–55 (1969).
- Ford, G. C. Intensity determination by profile-fitting applied to precession photographs. *J. Appl. Crystallogr.* **7**, 555–564 (1974).
- Rossmann, M. G. Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *J. Appl. Crystallogr.* **12**, 225–238 (1979).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation Mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Powell, H. R., Johnson, O. & Leslie, A. G. Autoindexing diffraction images with *iMosflm*. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **69**, 1195–1203 (2013).
- Kabsch, W. XDS. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **66**, 125–132 (2010).
- Schultz, A. J. *et al.* Integration of neutron time-of-flight single-crystal Bragg peaks in reciprocal space. *J. Appl. Crystallogr.* **47**, 915–921 (2014).
- Von Dreele, R. B., Jorgensen, J. D. & Windsor, C. G. Rietveld refinement with spallation neutron powder diffraction data. *J. Appl. Crystallogr.* **15**, 581–589 (1982).
- Rietveld, H. M. A Profile refinement method for nuclear and magnetic structures. *J. Appl. Crystallogr.* **22**, 65–71 (1969).
- Thompson, P., Cox, D. E. & Hastings, J. B. Rietveld refinement of Debye-Scherrer synchrotron X-ray data from Al<sub>2</sub>O<sub>3</sub>. *J. Appl. Crystallogr.* **20**, 79–83 (1987).
- Ikeda, S. & Carpenter, J. M. Wide-energy-range, high-resolution measurements of neutron pulse shapes of polyethylene moderators. *Nucl. Instr. Meth. Phys. Res.* **A239**, 536–544 (1985).
- Larson, A. C. & Von Dreele, R. B. General Structure Analysis System (GSAS). *Los Alamos National Laboratory Report LAUR*. 86–748 (2004).



37. Tomoyori, K. *et al.* Profile functions to reproduce Bragg reflection shapes observed by a time-of-flight single-crystal diffractometer installed at a coupled moderator pulsed neutron source in J-PARC. *Nucl. Instr. Meth. Phys. Res. A* **723**, 128–135 (2013).
38. Bruna, R. & Rademakersh, F. ROOT - An object oriented data analysis framework. *Nucl. Instr. Meth. Phys. Res. A* **389**, 81–86 (1997).
39. Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1718–1725 (1999).
40. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **66**, 213–221 (2010).
41. Brünger, A. T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. **355**, 472–475 (1992).
42. Kusaka, K. *et al.* Peak overlapping and its de-convolution in TOF diffraction data from neutron biological diffractometer in J-PARC. *Physica B*. **385–386**, 1062–1065 (2006).
43. Skrzypczak-Jankun, E. *et al.* Structure of the hirugen and hirulog 1 complexes of alpha-thrombin. *J. Mol. Biol.* **221**, 1379–1393 (1991).

## Acknowledgements

This work was financially supported by the Ibaraki Prefectural Government. The authors wish to express their sincere thanks to Dr. Nobuo Niimura (Ibaraki University) for important comments regarding this research paper. The neutron data were collected in the iBIX at J-PARC (2012PX0006, 2013PX0008). The X-ray data were collected in the Photon Factory (2013G646). We would like to thank Editage ([www.editage.jp](http://www.editage.jp)) for English language editing.

## Author Contributions

All authors discussed and commented on the results and the manuscript. N.Y. determined the fitting function, developed the profile fitting algorithm for the iBIX data, applied it to the full set protein diffraction data, and wrote the manuscript. T.Y. and K.K. conducted protein crystallization and diffraction experiments. T.H. contributed to the maintenance of the neutron experimental instruments. T.O. and I.T. contributed to the discussion in this research paper.

## Additional Information

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

**How to cite this article:** Yano, N. *et al.* Application of profile fitting method to neutron time-of-flight protein single crystal diffraction data collected at the iBIX. *Sci. Rep.* **6**, 36628; doi: 10.1038/srep36628 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016