



ISSN 1399-0047

# Structure determination of an integral membrane protein at room temperature from crystals *in situ*

Danny Axford,<sup>a,‡</sup> James Foadi,<sup>b,c,‡</sup> Nien-Jen Hu,<sup>b,c,d,‡§</sup> Hassanul Ghani Choudhury,<sup>b,c,d,¶</sup> So Iwata,<sup>a,b,c,d,e</sup> Konstantinos Beis,<sup>b,c,d,‡‡</sup> Gwyndaf Evans<sup>a\*</sup> and Yilmaz Alguel<sup>b,c,d,\*</sup>

Received 31 January 2015

Accepted 1 March 2015

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

‡ These authors contributed equally to this work.

§ Current address: Institute of Biochemistry, National Chung Hsing University, Taichung 40227, Taiwan.

¶ Current address: Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia.

‡‡ Co-senior author.

**Keywords:** *in situ* data collection; membrane protein; multiple data sets; synchrotron beamline.

**PDB reference:** HiTehA, 4ycr

**Supporting information:** this article has supporting information at journals.iucr.org/d

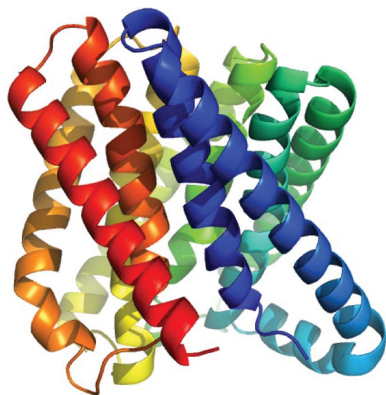
<sup>a</sup>Diamond Light Source, Harwell Science and Innovation Campus, Oxfordshire OX11 0DE, England, <sup>b</sup>Membrane Protein Laboratory, Diamond Light Source, Harwell Science and Innovation Campus, Oxfordshire OX11 0DE, England, <sup>c</sup>Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, England, <sup>d</sup>Research Complex at Harwell, Rutherford Appleton Laboratory, Oxfordshire OX11 0FA, England, and <sup>e</sup>Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan. \*Correspondence e-mail: gwyndaf.evans@diamond.ac.uk, y.alguel03@imperial.ac.uk

The structure determination of an integral membrane protein using synchrotron X-ray diffraction data collected at room temperature directly in vapour-diffusion crystallization plates (*in situ*) is demonstrated. Exposing the crystals *in situ* eliminates manual sample handling and, since it is performed at room temperature, removes the complication of cryoprotection and potential structural anomalies induced by sample cryocooling. Essential to the method is the ability to limit radiation damage by recording a small amount of data per sample from many samples and subsequently assembling the resulting data sets using specialized software. The validity of this procedure is established by the structure determination of *Haemophilus influenzae* TehA at 2.3 Å resolution. The method presented offers an effective protocol for the fast and efficient determination of membrane-protein structures at room temperature using third-generation synchrotron beamlines.

## 1. Introduction

Membrane-protein structure determination routinely uses X-ray diffraction data recorded at cryogenic temperatures from a single crystal, requiring a significant investment of effort to grow samples of sufficient size to allow a complete data set to be recorded. These two criteria have been driven by the typical nature of membrane-protein crystals: they are formed by limited crystal contacts, owing to a high solvent content and poor order, and are prone to non-isomorphism; these factors typically lead to weak diffraction (compared with most crystals of soluble proteins), requiring proportionally higher X-ray doses to allow measurement of high-resolution reflections. To compound the issue, phase transitions in any amphiphilic molecules in the crystal, such as detergents, can make the results of cryocooling less consistent and more likely to further compromise crystal order (Pebay-Peyroula, 2008).

It has been demonstrated that membrane-protein diffraction data can be recorded from micro/nanocrystal preparations injected into the intense pulsed beam of an X-ray free-electron laser (XFEL) at room temperature (Weierstall *et al.*, 2014). This significant step forward has been a consequence of the ‘diffraction before destruction’ experiment (Chapman *et al.*, 2011) made feasible by the very short, intense pulses from XFELs. Membrane-protein crystal structure determination has been beyond the reach of room-temperature crystal diffraction measurements at synchrotron-radiation sources, principally owing to the significant primary and secondary



OPEN ACCESS

radiation damage that occurs (Garman, 2010). *In situ* data-collection methodology (from crystals in crystallization plates) has matured to the point where the structure determination of viruses and other soluble proteins is now approaching routine (Axford *et al.*, 2012; Heidari Khajepour *et al.*, 2013; Wang *et al.*, 2012). *In situ* screening at synchrotrons (Axford *et al.*, 2012) has shown that membrane-protein crystals yield only a small number of images before losing their diffracting ability. High-resolution diffraction data can be recorded at room temperature for membrane-protein crystals. *In situ* data collection removes the need for cryoprotectant, a potential obstacle in membrane-protein crystallography, where the detergent composition can vary (Pellegrini *et al.*, 2011). Sufficient data for structure determination would require many isomorphous crystals. A recent development in data analysis of multiple crystals is the software *BLEND*, which has been shown to be applicable to the cases of soluble and membrane proteins and brings the benefit of accelerating the often time-consuming procedure of managing multiple data sets (Foadi *et al.*, 2013) and identifying isomorphous crystals. This, combined with high-frame-rate pixel-array detectors (Broennimann *et al.*, 2006) and the discovery of prolonged crystal lifetimes at room temperature for high dose and frame rates (Owen *et al.*, 2012, 2014) brings the possibility of room-temperature structure determination of membrane proteins using synchrotron radiation within the grasp of crystallographers.

Here, we describe the first *in situ* structure determination of a membrane protein, using *Haemophilus influenzae* TehA (HiTehA), which has previously been solved to 1.2 Å resolution from a single cooled crystal (Chen *et al.*, 2010). We present a method to collect data from multiple *in situ* crystals of membrane proteins and to form a sufficiently complete data set from many partial data sets. The validity of the approach is demonstrated both by the quality of the electron-density maps associated with the assembled data set and by a detailed comparison between the derived structure and the reference structure solved using data collected at 100 K from a single crystal.

## 2. Materials and methods

### 2.1. Protein expression, purification and crystallization

HiTehA was cloned into pWaldoGFPe and purified as described previously (Drew *et al.*, 2006), with the final buffer consisting of 20 mM Tris pH 7.5, 150 mM NaCl, 60 mM *n*-octyl- $\beta$ -D-glucopyranoside. The protein was screened for crystallization at 20 mg ml<sup>-1</sup> using the vapour-diffusion method. Crystals for the *in situ* data-collection experiment were grown by mixing 100 nl HiTehA solution with 100 nl reservoir solution in sitting drops using a Mosquito robot (TTP Labtech); drops were dispensed onto a hydrophobic-coated 96-well plate (CrystalQuick X). The best diffracting crystals grew over 7–10 d at 277 K from a reservoir solution consisting of 0.1 M NaCl, 120 mM Tris pH 9.4, 20% (v/v) PEG 400. The crystal plate was moved to ambient temperature

before mounting on a modified goniometer as described previously (Axford *et al.*, 2012).

### 2.2. *In situ* data collection

Data were collected on beamline I24 at Diamond Light Source using a dedicated goniometer for the mounting of SBS-format (now ANSI/SLAS standard; <http://www.slas.org>) crystallization plates and a Pilatus3 6M detector. We have previously shown that a 100  $\mu$ m offset must be added to the position of the rotation axis in the direction of the beam to account for the optical effect of viewing the crystals through the plate-base material, thereby ensuring that the crystals could be precisely located on the axis of rotation (Axford *et al.*, 2012). Centring was performed by positioning the crystals onto a cross-hair coincident with the beam position and then translating them along the beam axis into the focal plane of the on-axis microscope. Visible radiation damage to the crystals following data collection was clearly contained within the crystal volume rather than appearing as a vertical line, indicating that the crystals were indeed well centred using this method. The goniometer allows an angular movement of the plate of approximately  $\pm 20^\circ$  from the vertical.

A few crystals were initially used to optimize and fix the data-collection parameters. Based on the observed diffraction from these crystals,  $d_{\min}$  at the edge of the detector was set to 2.5 Å resolution (1.83 Å resolution in the detector corners). This was necessarily a best guess and could not be optimized on a per-crystal basis owing to the rapid onset of radiation damage at room temperature. Subsequent analysis has shown that several crystals diffracted to a higher resolution and into the corners of the detector. These effects are reflected in the completeness and multiplicity of the data in the highest resolution bin, as shown in §2.4. Thus, our initial estimate of 2.5 Å resolution turned out to be too conservative and the



**Figure 1**  
Example section of a diffraction image with spots extending to 2.1 Å resolution into the corners of the detector. The inset at the top right is an on-beam-axis view of two example crystals located at the edge of a crystallization drop captured during data collection. The red circle represents the full-width half-maximum of the beam profile and has been matched to the crystal size. The matching of the beam size to that of the crystal optimized the signal-to-noise ratio of the data.

structure was eventually refined using data to 2.3 Å resolution, with the initial electron-density maps and model building being aided by data to 2.1 Å resolution (Fig. 1). The final 2.3 Å resolution limit was selected in order to achieve an overall data completeness of greater than 90%.

Multiple wedges of data were measured consisting of 30–50 images of 0.2° rotation each at 25 frames s<sup>-1</sup> with 12% of the total beam flux, equating to ~2 × 10<sup>11</sup> photons s<sup>-1</sup>. Each wedge therefore consisted of 6–10° of data after X-ray exposure for a total of 1.2–2 s.

A total of 67 wedges of data were recorded from 56 separate crystals ranging in size from 10 to 75 µm in the largest dimension. The beam size on the sample was adjusted between 10 and 50 µm to best match the size of each crystal in order to optimize the signal-to-noise ratio of the measurements while distributing the X-ray dose through the whole crystal volume. For the larger crystals data could be recorded from up to three points on the sample using a beam size smaller than the crystal. The starting angle for each wedge was varied to cover a total sampled angular range of 24° with the intention of maximizing reciprocal-space coverage in the eventuality that the crystals were systematically orientated in the drops.

### 2.3. 100 K data collection

The reference cryocooled data set was recorded on beamline I24 from a single crystal grown using identical crystallization conditions to those described above. The crystal was flash-cooled in liquid nitrogen and maintained at 100 K in an open flow of cold N<sub>2</sub> gas for measurement.

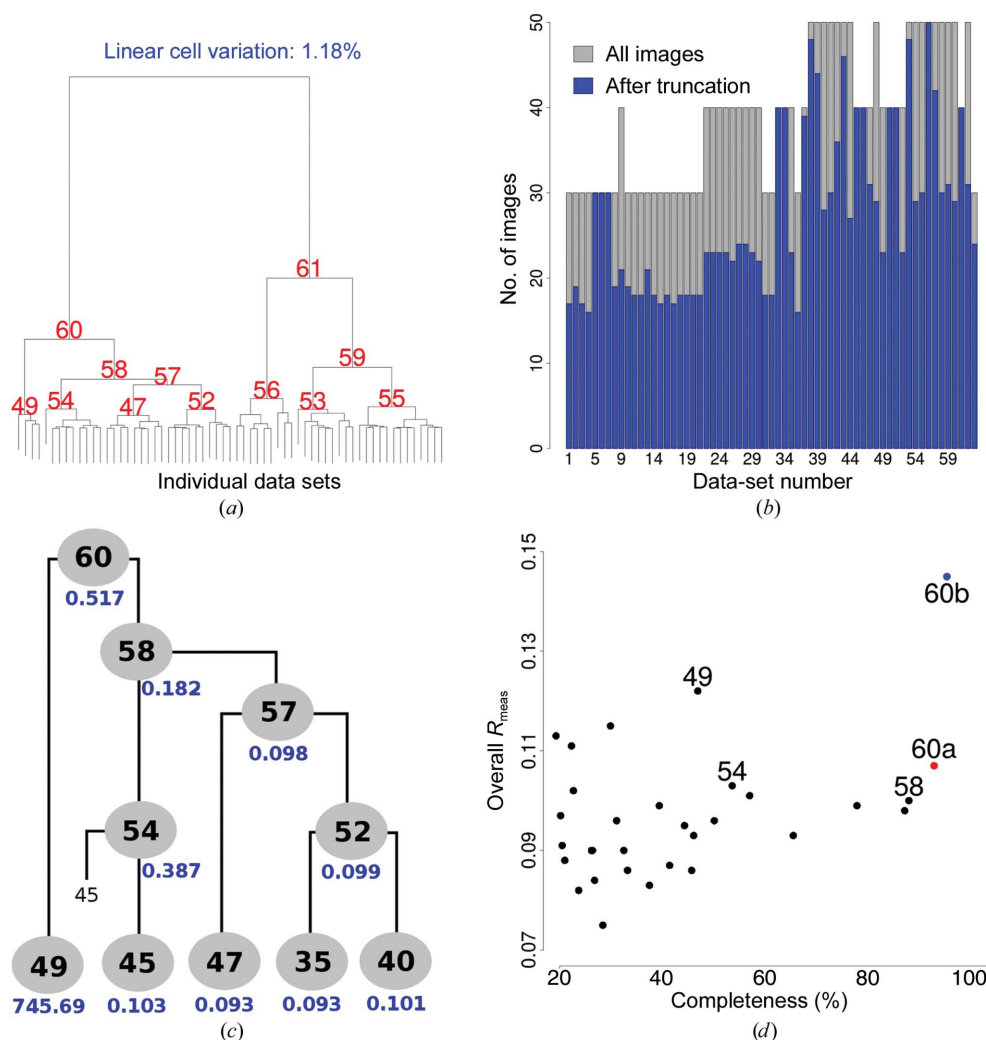
### 2.4. Data analysis, assessment of radiation damage and merging

Integration with *XDS* (Kabsch, 1993) proceeded smoothly for all but the last four data sets (64–67), for which *XDS* failed to integrate the data even when given the correct space group. These data sets were subsequently discarded from the analysis. A check of the diffraction images for the discarded data sets revealed split diffraction spots that were indicative of poor crystal integrity and were likely to be the reason that *XDS* failed to index the data.

The unit-cell parameters for all of the remaining wedges are displayed in Supplementary Table S1 along with the completeness up to 2.1 Å resolution.

*BLEND* was run in analysis mode on the remaining 63 data sets to produce a cluster dendrogram (Fig. 2*a*). The linear cell variation (LCV), which describes the maximum percentage change in the unit-cell face diagonals across all data sets, is 1.18%. Two major clusters emerged (Fig. 2*a*), cluster 60 and cluster 61, which showed a completeness of 89.7 and 70.7%, respectively, to 2.1 Å resolution. Cluster 60, being the most complete, was used for subsequent phasing by molecular replacement, model building and refinement.

Each wedge suffered to a varying extent from radiation damage. Rather than retaining a fixed number of images per wedge, a custom selection of data was made based on the procedure described in Appendix A. Briefly, a moving average



**Figure 2** Crystal selection and data processing carried out with *BLEND*. (a) Dendrogram showing all integrated data sets and their merging nodes, with two major clusters at nodes 60 and 61. (b) Graph showing the number of measured images in each wedge of data (grey bars) and the number of accepted images (blue bars) after radiation-damage assessment. (c) Final stage of data processing. The  $R_{\text{meas}}$  for each cluster (represented by a grey circle) is displayed in blue next to the node. (d) Plot of  $R_{\text{meas}}$  versus completeness for all subclusters of cluster 60, at 2.3 Å resolution, after the removal of data sets 45 and 46. Cluster 60a (red dot) includes the same data sets as cluster 60b (blue dot), but with some images removed, after correction for radiation damage. The reduction in  $R_{\text{meas}}$  is evident.



**Table 1**

Final merging statistics for cluster 60a.

All statistics were obtained after removing the outlier data sets. This data set was used to determine the room-temperature structure.

	Overall	Inner shell	Outer shell
Low-resolution limit (Å)	49.30	49.30	2.38
High-resolution limit (Å)	0.092	0.055	0.485
$R_{\text{merge}}$ (all $I^+$ and $I^-$ )	0.096	0.058	0.513
$R_{\text{meas}}$ (within $I^+/I^-$ )	0.107	0.064	0.585
$R_{\text{meas}}$ (all $I^+$ and $I^-$ )	0.107	0.063	0.590
$R_{\text{p.i.m.}}$ (within $I^+/I^-$ )	0.054	0.032	0.317
$R_{\text{p.i.m.}}$ (all $I^+$ and $I^-$ )	0.044	0.024	0.281
$R_{\text{merge}}$ in top intensity bin	0.050	—	—
Total No. of observations	99220	1770	7005
No. of unique reflections	20429	323	1945
Mean $I/\sigma(I)$	13.6	36.9	3.7
Mean $I$ half-set correlation $CC_{1/2}$	0.996	0.994	0.747
Completeness (%)	92.9	85.6	90.5
Multiplicity	4.9	5.5	3.6

intensity is determined as a function of diffraction image and resolution for each set and data are rejected when this intensity falls below a threshold, in this case 75% of the starting intensity. The number of images retained per wedge (ranging between 15 and 50) after application of this procedure is shown in Fig. 2(b). This approach is quite conservative, removing images only where it was statistically evident that global radiation damage had affected the data. Different approaches using the elimination of either a fixed number or a fixed fraction of images for all data sets have also been attempted, but in neither case were the merging statistics better than with this custom procedure.

Scaling and merging were performed using *AIMLESS* (Evans & Murshudov, 2013) at 2.3 Å resolution. Most data sets in cluster 60 merge well, with the exception of cluster 49 and data set 45 (Fig. 2c). Excluding data set 45 from cluster 58 reduced the overall  $R_{\text{meas}}$  from 0.182 to 0.100. Of the four data sets composing cluster 49 (52, 56, 41 and 46), data set 46 was found to be solely responsible for the poor merging and was therefore excluded. A new cluster, 60a, was therefore produced by discarding data sets 45 and 46. Fig. 2(d) shows a plot of  $R_{\text{meas}}$  versus completeness at 2.3 Å resolution for all of the clusters (nodes) in the left branch of the dendrogram after the removal of data sets 45 and 46. Structure factors were determined from scaled and merged intensities using *TRUNCATE* (French & Wilson, 1978).

### 2.5. Determination of high-resolution limits

Data to 2.1 Å resolution were initially used for phasing and model building as they resulted in a very clear and interpretable electron-density map. At a later stage, the resolution limit was cut to 2.3 Å resolution based on the application of an overall  $CC_{1/2} > 0.5$  criterion. This fairly stringent cutoff was used so that an overall completeness of greater than 90% was retained. This made comparison with the complete 100 K data set more meaningful.

The final overall  $R_{\text{meas}}$  was 0.107,  $R_{\text{p.i.m.}}$  was 0.044 and the completeness was 92.9%. The final statistical summary from *AIMLESS* for this data set is given in Table 1. As a note of

**Table 2**

Data-collection and refinement statistics (molecular replacement).

Values in parentheses are for the highest resolution shell. For the 100 K and room-temperature (RT) data sets, one and 56 crystals were used, respectively.

	100 K (one crystal)	RT (56 crystals)
Data collection		
Space group	<i>H3</i>	<i>H3</i>
Unit-cell parameters (Å, °)	$a = b = 97.03, c = 136.76, \alpha = \beta = 90, \gamma = 120$	$a = b = 98.60, c = 136.38, \alpha = \beta = 90, \gamma = 120$
Resolution (Å)	50.0–1.5 (1.54–1.50)	50.0–2.3 (2.38–2.30)
$R_{\text{merge}}$	0.035 (0.506)	0.096 (0.513)
$\langle I/\sigma(I) \rangle$	14.5 (2.4)	13.6 (3.7)
Completeness (%)	98.6 (98.3)	92.9 (90.5)
Multiplicity	3.3 (3.1)	4.9 (3.6)
$CC_{1/2}$	0.998 (0.843)	0.996 (0.747)
Refinement		
Resolution (Å)	50.0–1.5 (1.55–1.50)	50.0–2.3 (2.38–2.30)
No. of reflections	267005	99220
$R_{\text{work}}/R_{\text{free}}$ (%)	13.6/16.9 (21.7/22.7)	15.6/20.01 (23.3/28.2)
No. of atoms		
Protein	2461	2413
Ligand/ion	140	40
Water	158	71
$B$ factors (Å <sup>2</sup> )		
Protein	24.4	28.9
Ligand/ion	59.0	50.8
Water	42.0	34.7
R.m.s. deviations		
Bond lengths (Å)	0.006	0.01
Bond angles (°)	0.936	1.364

interest, equivalent analysis without the removal of radiation-damaged images gave an overall  $R_{\text{meas}}$  of 0.145, an  $R_{\text{p.i.m.}}$  of 0.050 and a completeness of 95.4%.

It is important to stress that structure determination was not complicated by the fragmented nature of the multiple data sets composing the final data. Molecular replacement, model building and final refinement were carried out in exactly the same way as for a complete data set from a single crystal.

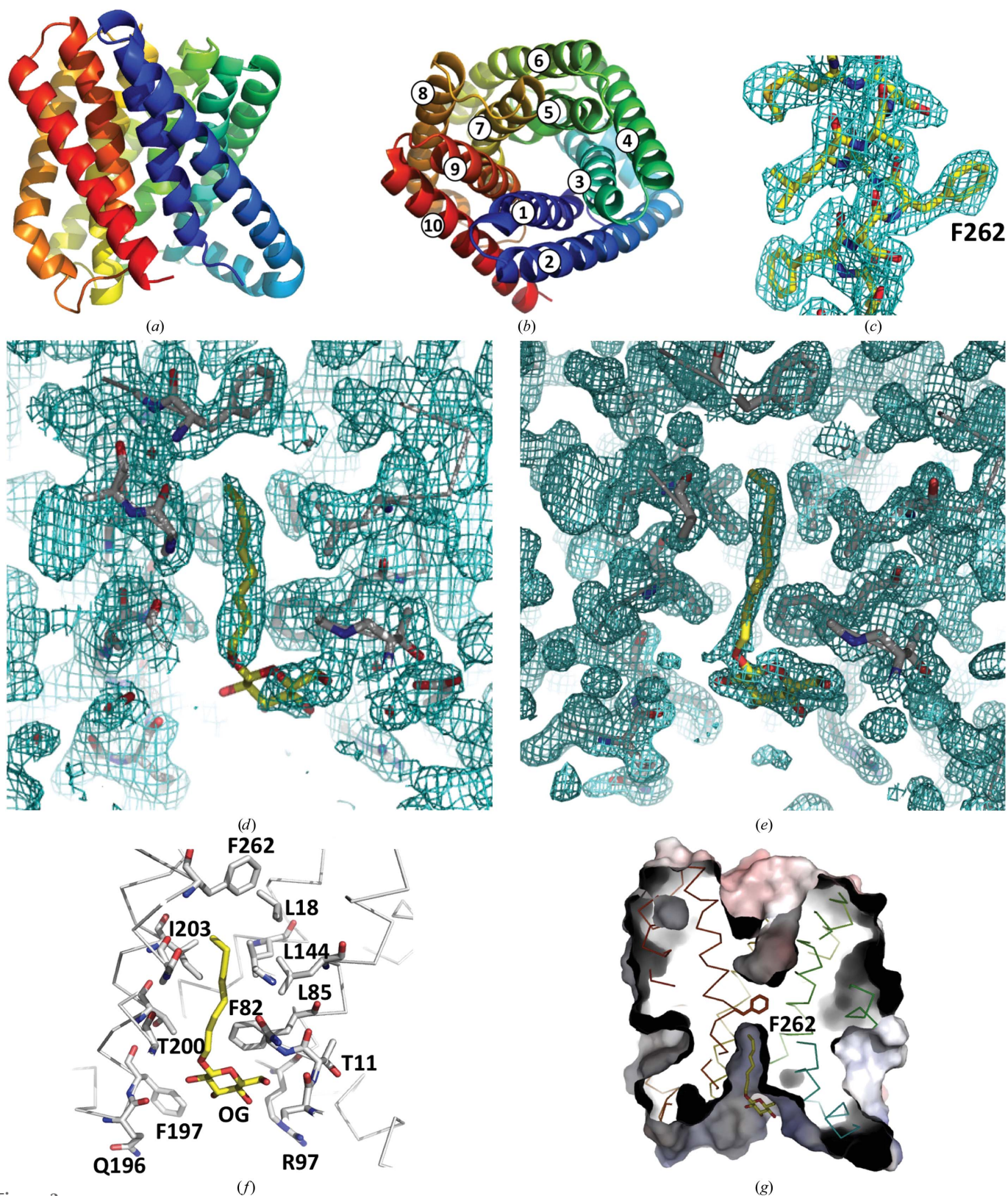
### 2.6. Structure determination and refinement

Phases related to the final data were obtained by molecular replacement in *Phaser* (McCoy *et al.*, 2007) using the deposited structure of HiTehA (PDB entry 3m71; Chen *et al.*, 2010) as a search model. The initial electron-density map was inspected and the model was built using *Coot* (Emsley & Cowtan, 2004). Model refinement was performed using *PHENIX* (Adams *et al.*, 2010). The structure was refined against the 2.3 Å resolution multi-crystal data set to an  $R_{\text{work}}$  of 15.6% and an  $R_{\text{free}}$  of 20.01%. The TehA structure has 97% of the residues in the favoured Ramachandran region and no outliers. The structure factors and coordinates have been deposited in the Protein Data Bank as PDB entry 4ycr. Detailed refinement statistics are given in Table 2.

## 3. Results and discussion

### 3.1. Data collection and analysis

A complete data set to 2.3 Å resolution was assembled from 63 partial data sets obtained by irradiating *in situ* 56 crystals of the membrane protein HiTehA distributed across a number

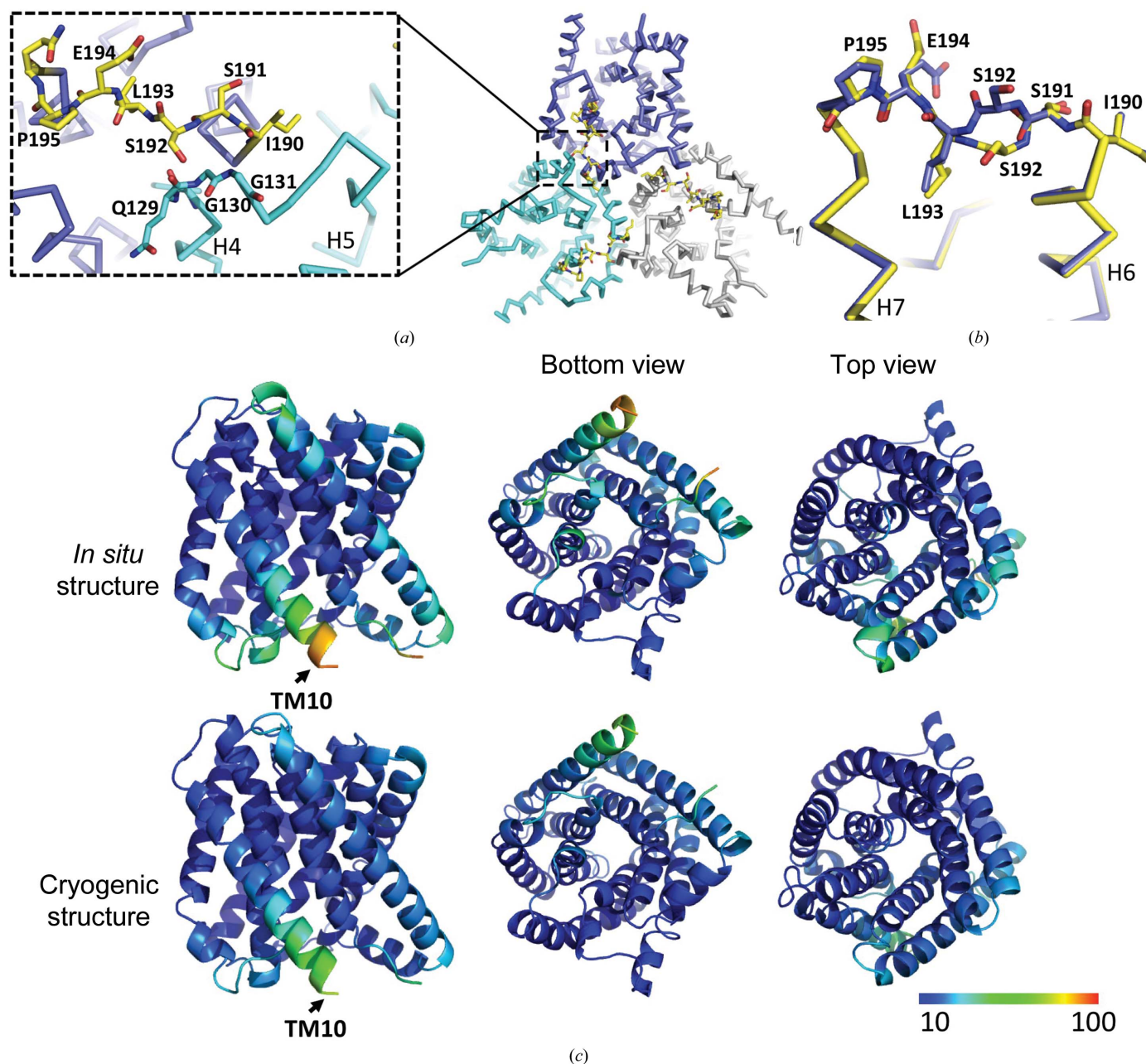


**Figure 3**  
 Crystal structure of HiTehA from *in situ* and cryogenic data. Cartoon representation of TehA (*a*) parallel to the membrane and (*b*) from the periplasmic face. The structure is coloured in a rainbow from the N-terminus (blue) to the C-terminus (red). (*c*)  $2F_o - F_c$  electron-density map section within TM9 contoured at  $1.0\sigma$  with the model in stick representation with carbon in yellow, nitrogen in blue and oxygen in red. Clear electron density is visible for the highly conserved gating residue Phe262. (*d*)  $2F_o - F_c$  electron-density map for OG with its proximate residues from the *in situ* 2.3 Å resolution data contoured at  $1\sigma$ . (*e*) The same representation as in (*d*) for the 1.5 Å resolution 100 K structure. (*f*) Ribbon representation with the OG detergent molecule and surrounding side chains shown as sticks. (*g*) A slice through the channel shows the path with the gating residue Phe262 (red sticks) on TM9. The OG detergent is bound to HiTehA on the cytoplasmic side and reaches deep into the hydrophobic channel.



of cells of a single 96-well crystallization plate mounted on a specialized goniometer (Axford *et al.*, 2012). Each crystal was exposed to  $\sim 2 \times 10^{11}$  photons  $s^{-1}$  for 1.2–2.0 s, during which 30–50  $0.2^\circ$  images were recorded at 25 frames  $s^{-1}$ . The total data collection for all crystals took less than 3 h. Data integration was carried out with *XDS* (Kabsch, 1993). Of the 67 wedges of data integrated with *XDS* only 63 indexed correctly in space group *H3*; the remaining four were associated with split crystals. The completeness of the individual data wedges

varied between about 12.3 and 22.6% at 2.1 Å resolution. These partial data sets were fed into *BLEND* (Foadi *et al.*, 2013) to carry out radiation-damage assessment, cluster analysis of unit-cell variation and to manage the subsequent collation, scaling and merging. Assessment and rejection of diffraction images overly affected by radiation damage was made by analysis of the average intensity reduction as a function of image and resolution (see §2.4). Diffraction images suffering from radiation damage were rejected from the



**Figure 4**

Comparison of the *in situ* and cryogenic structures. (a) The two models superimpose quite well in general. One notable exception is the loop connecting TM6 and TM7, as detailed in (b). This shifting loop is located towards the adjacent monomer and is proximate to the C-terminal region. (c) Colour representation of *B* factor across the chain for the two models. The blue to red spectrum indicates low to high *B* factors. The structure obtained with the *in situ* data exhibits higher *B* factors, especially at the ends of the helices C-terminal to TM10, than the cryogenic structure. The respective overall *B* factors are 26.4 Å<sup>2</sup> for the 295 K structure and 24.4 Å<sup>2</sup> for the 100 K structure. On average, the intracellular part of the models has a higher *B* factor than the extracellular part owing to the presence of larger loops.

analysis if their average intensity in the highest resolution shell dropped below 75% of the starting value. The final data set had an overall  $R_{\text{meas}}$  of 0.107, an  $R_{\text{p.i.m.}}$  of 0.044 and a completeness of 92.9% to 2.3 Å resolution (Table 1).

### 3.2. Structure of HiTehA

Initial structure determination was carried out *via* molecular replacement using the HiTehA structure (PDB entry 3m7b; Chen *et al.*, 2010) followed by refinement using the PHENIX platform (Adams *et al.*, 2010) to 2.3 Å resolution with final  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 15.6 and 20.01%, respectively (Table 2). The overall *in situ* structure is very similar to the published cryogenic structure, with an r.m.s.d. of 0.66 Å for all atoms. HiTehA is a trimeric membrane protein, with each monomer consisting of ten transmembrane (TM) helices linked by short loops (Fig. 3*a*). The HiTehA monomer consists of five two-transmembrane-helix hairpin repeats. TM1, TM3, TM5, TM7 and TM9 are part of the inner pore of the channel perpendicular to the membrane surrounded by TM2, TM4, TM6, TM8 and TM10 (Fig. 3*b*). The electron-density map after molecular replacement at 2.3 Å resolution was of high quality, allowing individual amino-acid side chains, water and detergent molecules to be fitted with accuracy. Residue Phe262, which was reported to be important for gating, was found to be in the same position and orientation (Fig. 3*c*) as observed by Chen *et al.* (2010).

### 3.3. Comparison of room-temperature and 100 K structures

In addition to the room-temperature data, a reference data set from a single crystal cryocooled to 100 K was collected and its structure was determined *via* molecular replacement in an identical way to the room-temperature structure (Figs. 3*d*–3*g*). These 100 K data were subsequently refined to 1.5 Å resolution with final  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 13.6 and 16.7%, respectively (Table 2).

The two structures superimpose very well with an r.m.s.d. of 0.55 Å for all atoms, but a clear shift in the loop connecting TM6 and TM7 is observed, with a maximum distance of 2.9 Å measured at residue Ser192 (Figs. 4*a* and 4*b*). In the case of the room-temperature structure the loop folds back towards the inside of HiTehA, whereas in the cryogenic model it folds outwards towards the cytoplasmic side. This loop is located on the interface with the next monomer of the trimeric HiTehA protein and interacts with the C-terminal end of TM helix 4. Ser192 interacts with the backbone of the adjacent monomer of the trimer in proximity to the backbone of the residues Gly130, Gly129 and Gln129. This loop shift in the monomeric interface does not impact the overall trimeric arrangement of HiTehA between the room-temperature and the 100 K model, as the trimeric superimposition involving a total of 912 C $\alpha$  atoms results in an r.m.s.d. of 0.279 Å. Furthermore, the loop shift neither alters the position of the gating Phe262, located on TM9, nor blocks the channel.

Analysis of the B-factor distribution reveals, as expected, regions of greater flexibility in the *in situ* structure compared

with the 100 K structure (Fig. 4*c*). However, the magnitude of this difference is small.

The electron density from the room-temperature and 100 K data both reveal one octylglucoside (OG) detergent molecule inside the channel cavity on the cytoplasmic site (Figs. 3*d* and 3*e*) that was not reported in the original structure. The hydrophobic alkyl tail of the OG detergent reaches deep into the channel and is surrounded by the hydrophobic residues Phe262, Ile203, Leu18, Leu144, Leu85 and Phe82. The polar glycoside head group of OG is proximate to the charged groups Arg97 and Gln196 and the backbone of HiTehA (Fig. 3*f*). As a note of interest, electron-density maps calculated using the structure factors from the structure of Chen and coworkers show OG-like density in the channel, but its interpretation was presumably hindered by discontinuity in this electron density.

### 3.4. Assessment of data quality using OMIT maps

In order to validate the multi-crystal data-set quality and to exclude model bias, an initial model of HiTehA with a C-terminal deletion ranging up to residue Val279, including the entire TM10 helix, was generated and used for refinement against the merged raw data set. Electron-density maps for the omitted region of the structure are shown in Figs. 5(*a*) and 5(*b*). The map shows continuously connected backbone and side-chain density for the omitted TM10 region.

### 3.5. Assessment of data quality by molecular replacement

There are no known structural homologues of HiTehA to provide a search model for molecular replacement. A feasible solution to this problem is suggested by the observation that  $\alpha$ -helical structures of membrane transporters and channels typically share common domains, motifs and repeats. These individual domains or motifs often serve as ensembles of plausible search models for molecular replacement (Pornillos & Chang, 2006; Sciara & Mancia, 2012). A potential match is represented by the backbone C $\alpha$ -atom superposition of helices TM1–TM4 onto helices TM7–TM10 (Fig. 5*c*); in this case the r.m.s.d. using secondary-structure mapping (SSM; Krissinel & Henrick, 2004) amounts to 2.6 Å calculated over 98 atoms. In order to simulate a *de novo* molecular replacement, TM1–TM4 were selected as the search model (Fig. 5*c*). The new truncated model consisted of 96 amino acids, making up 29% of the total HiTehA sequence. Molecular replacement using Phaser (McCoy *et al.*, 2007) indicated a prominent top solution with a rotation Z-score of 10.3, a translation Z-score of 16.8 and a log-likelihood gain (LLG) of 307.1. The calculated electron-density map at 2.3 Å resolution from the molecular-replacement solution clearly displayed the missing part of the model, and four additional TM helices were automatically traced by Buccaneer (Cowtan, 2006; Fig. 5*d*). The figure of merit associated with the resulting electron-density map was 0.599, a value indicating a high degree of map interpretability. Manual building of the model to completion was, at this stage, a straightforward procedure.



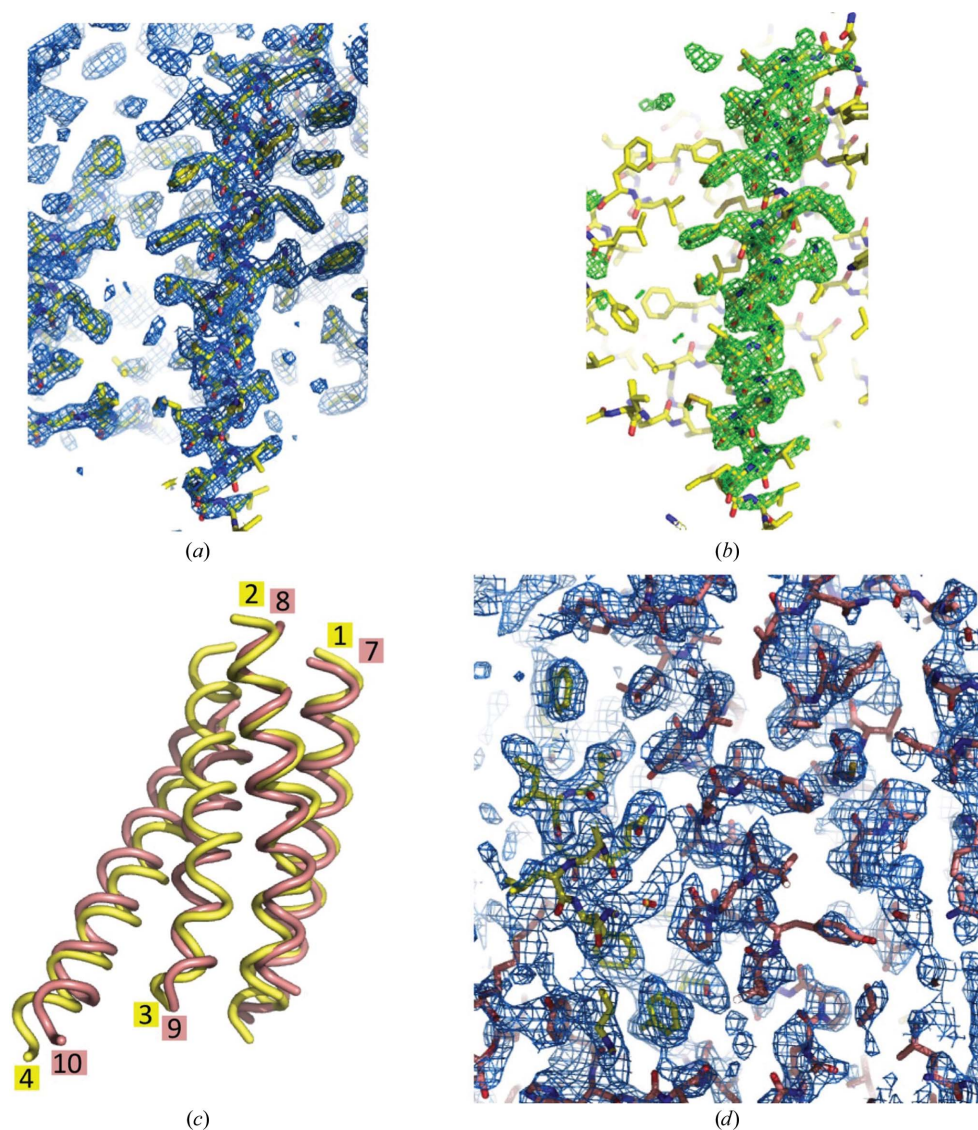
#### 4. Conclusion

The first structure of an integral membrane protein at room temperature determined by *in situ* data collection at a synchrotron has been presented. From a total of 56 measured crystals, a final scaled and merged data set reaching 2.3 Å resolution was obtained from 63 partial data sets. The results are of great value since, by their nature, membrane proteins struggle to form large, well ordered crystals that are amenable to cryocooling. The approach used here is fairly conservative regarding radiation-damage assessment and data rejection.

One could easily expect, however, that a more stringent application of the procedures outlined in the Supporting Information and the measurement of data from many more crystals could yield complete data for more challenging membrane-protein systems; for example, G-protein coupled receptors (GPCRs), which are typically grown in lipidic cubic phase and are known to require multiple data sets even under cryogenic conditions (Hanson *et al.*, 2008). Membrane-protein crystals grown in lipidic cubic phase are already screened routinely in plates for their initial diffraction on Diamond beamline I24 (Axford *et al.*, 2012). The collection of *in situ* data can decrease the number of crystals compromised through handling and increase the throughput, facilitating the acquisition of a full data set as produced by a suitable software package such as *BLEND*.

It is important to note that radiation damage in synchrotron X-ray diffraction data is inevitable. Recent free-electron laser (FEL) studies have shown that essentially radiation-damage-free membrane-protein diffraction data can be measured from crystals within a lipidic cubic phase 'jet' (Weierstall *et al.*, 2014). Currently, access to FELs is in heavy demand and the analysis of data obtained from serial femto-second crystallography is still in its infancy (Barends, 2014; White *et al.*, 2012) and is reliant on massive levels of averaging from tens of thousands of crystals to obtain data quality that approaches that attainable using a synchrotron. Whereas the radiation-damage-free nature of FEL diffraction data from biological macromolecules may be valued from the perspective of functional studies and biological interpretability (Neutze *et al.*, 2004), the practical problem of obtaining data that are of sufficient quality to determine *de novo* phase information and interpretable electron-density maps remains.

Using the approach presented in this paper, data collection from 56 crystals and the identification of 813 images of highest quality data sufficient for structure solution required around 150 min of beamtime. There is significant



**Figure 5**

The quality of the electron-density maps reflects the good quality of the *in situ* data. (a)  $2F_o - F_c$  electron-density map after an OMIT map related to a C-terminal HiTehA deletion including TM10. The map is calculated at 2.3 Å resolution and contoured at  $1.0\sigma$ . The fitted TM10 is shown for clarity. (b) Positive  $F_o - F_c$  electron-density map at 2.3 Å resolution contoured at  $3.0\sigma$  showing the missing TM10. TM10 is also shown here for clarity. (c) The four-transmembrane-helix search model. The TM1–TM4 helices (yellow) superimpose well onto the TM7–TM10 helices (salmon). (d)  $2F_o - F_c$  electron-density map calculated using the molecular-replacement phases at 2.3 Å resolution contoured at  $1.0\sigma$ . The missing part of the structure in the search model is revealed in the electron-density map and is well connected, with visible density for the side chains; the initial search model (yellow) and the built model (salmon red) are shown.



scope to increase the throughput of the data-acquisition procedure by automation, possibly by the use of image-recognition software to identify samples.

## APPENDIX A

### A procedure to assess and modify data sets affected by radiation damage

Intensity averages in data sets affected by radiation damage have relatively lower values than those in unaffected data sets. The effect is especially evident and is normally greater at increasing resolution (Garman, 2010). Several studies have ascertained that equivalent intensities vary monotonically with time once the crystal has been irradiated (Diederichs *et al.*, 2003), but the exact form of such behaviour is not easy to capture. If the average intensity is monitored in resolution shells during data collection, such a monotonic decrease should be quantitatively observable. In scaling programs a subdivision of data into resolution shells and time intervals (equivalent to a group of images) is always performed to implement any scaling algorithm, under the assumption of a constant irradiated dose. Here, we have adopted a similar approach for the determination of radiation damage. The goal of this procedure is to determine whether it is worth removing part of the data from the full data set and, if this is the case, which part should be removed. The main steps are as follows.

(i) Each data set is divided into resolution shells and groups of images. In the following, we will use  $s$  to indicate inverse resolution ( $s = 1/d$ ) and  $t$  (for time) to indicate image number.

(ii) Running averages are computed for each resolution shell. These can be represented as curves in an intensity–time plot. The average intensity in each shell has a behaviour characteristic of the specific data-collection experiment. In general, this average changes with time owing to factors such as beam-flux fluctuations, the exposure of different parts of the crystal during rotation, crystal absorption and radiation damage. For all data sets described in this paper the exposure was short (essentially constant beam flux), covered a small rotation range (a negligible change in absorption) and the crystals were quite small (fully bathed crystal); thus, it is relatively safe to consider radiation damage as the most prominent cause of dynamic behaviour for each data set. Under these conditions the average intensity is expected to follow an exponential decrease with time, with the decline being more rapid at higher resolutions.

(iii) An exponential regression is performed for individual curves in each resolution shell. The exponential coefficient, indicated as  $-\Omega$ , is calculated and stored. The regression model has the form

$$I(t) = I_0 \exp(-\Omega t). \quad (1)$$

(iv) A linear regression is carried out over all  $\Omega$  in relation to the resolution  $s$ . If the coefficient  $a$  of the regression is positive and different from zero by at least one standard deviation, then the decay is considered to be a genuine effect

of radiation damage. In such a case the decay coefficient ( $\Omega$ ) will be a function of resolution, with

$$\Omega = as + b. \quad (2)$$

(v) Radiation is supposed to affect decay from the start of data collection, as implied by the continuous nature of the model for  $I(t)$  (1). It is thus appropriate to limit the number of images used in further analysis to be compatible with a desired amount of decay. Let  $f$  indicate such an amount as a fraction between 0 and 1. A value of 1 means no decay and 0 means that the whole intensity average has been reduced to zero. A default value of  $f = 0.75$  (corresponding to data for which the average intensity has been reduced to 75% of its initial value) has been used for all data sets described in this paper. We are looking for the time at which the average intensity has decreased to  $f$  of its initial value. Using (1), it is found that

$$\frac{I(t)}{I_0} = \exp(-\Omega t) = f \rightarrow t = \frac{-\log f}{\Omega}. \quad (3)$$

(vi) If  $\Omega$  were constant with resolution, (3) would yield a single time (image) after which the intensity averages are decreased to more than  $f$  of their initial value. However, (2) tells us that  $\Omega$  typically increases with resolution. Thus, to retain data with average intensities greater than  $f$  of their initial value it is necessary to discard data from certain resolution shells in each image rather than whole images. The exact analytical value, obtained by substituting for (2) in (3), is

$$t = t(s) = \frac{-\log f}{as + b}. \quad (4)$$

(vii) Rather than the elimination of part of each image according to (4), we have preferred to use this formula to select an image after which all data are discarded. This is the average image among all reflections retained. The reason for this is related to the way that *AIMLESS* (Evans & Murshudov, 2013), and indeed most scaling programs, deals with radiation damage. This is implicitly included in scaling models as the temperature factor  $B$  and, as long as it is not too severe, the determined scale factors should correct for global radiation effects to a good approximation.

The regression model adopted in the procedure described here is an approximation to the actual decay. The linear increase of  $\Omega$  with resolution is also an approximation. They are, in essence, the simplest available models compatible with the observed phenomenon of radiation damage in crystals.

### Acknowledgements

We would like to thank Diamond Light Source for beam-time allocation and access. Sample preparation was carried out at the Research Complex at Harwell (RCaH) and the Membrane Protein Laboratory (MPL) at Diamond Light Source. We would also like to thank Professor David Stuart and Dr Martin Walsh for valuable feedback on the manuscript. We are grateful for financial support from the Wellcome Trust

(MPL: WT/099165/Z/12/Z to SI) and the Biotechnology and Biological Sciences Research Council (BB/G023425/1 to SI). NH, HC and YA prepared the samples. JF, DA, NH, HC, KB and YA performed the synchrotron *in situ* data collection. JF and YA performed the data analysis. All authors performed the research. GE, YA, DA, KB, JF wrote the manuscript, in discussion with the other authors. The authors declare no competing financial interests.

## References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Axford, D. *et al.* (2012). *Acta Cryst.* **D68**, 592–600.
- Barends, T. R. M., Foucar, L., Botha, S., Doak, R. B., Shoeman, R. L., Nass, K., Koglin, J. E., Williams, G. J., Boutet, S., Messerschmidt, M. & Schlichting, I. (2014). *Nature (London)*, **505**, 244–247.
- Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *J. Synchrotron Rad.* **13**, 120–130.
- Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
- Chen, Y.-H., Hu, L., Punta, M., Bruni, R., Hillerich, B., Kloss, B., Rost, B., Love, J., Siegelbaum, S. A. & Hendrickson, W. A. (2010). *Nature (London)*, **467**, 1074–1080.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Acta Cryst.* **D59**, 903–909.
- Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. (2006). *Nature Methods*, **3**, 303–313.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Foadi, J., Aller, P., Alguel, Y., Cameron, A., Axford, D., Owen, R. L., Armour, W., Waterman, D. G., Iwata, S. & Evans, G. (2013). *Acta Cryst.* **D69**, 1617–1632.
- French, S. & Wilson, K. (1978). *Acta Cryst.* **A34**, 517–525.
- Garman, E. F. (2010). *Acta Cryst.* **D66**, 339–351.
- Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V.-P., Chien, E. Y. T., Velasquez, J., Kuhn, P. & Stevens, R. C. (2008). *Structure*, **16**, 897–905.
- Heidari Khajepour, M. Y., Lebrette, H., Vernede, X., Rogues, P. & Ferrer, J.-L. (2013). *J. Appl. Cryst.* **46**, 740–745.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Neutze, R., Huldt, G., Hajdu, J. & van der Spoel, D. (2004). *Radiat. Phys. Chem.* **71**, 905–916.
- Owen, R. L., Axford, D., Nettleship, J. E., Owens, R. J., Robinson, J. I., Morgan, A. W., Doré, A. S., Lebon, G., Tate, C. G., Fry, E. E., Ren, J., Stuart, D. I. & Evans, G. (2012). *Acta Cryst.* **D68**, 810–818.
- Owen, R. L., Paterson, N., Axford, D., Aishima, J., Schulze-Briese, C., Ren, J., Fry, E. E., Stuart, D. I. & Evans, G. (2014). *Acta Cryst.* **D70**, 1248–1256.
- Pebay-Peyroula, E. (2008). *Biophysical Analysis of Membrane Proteins: Investigating Structure and Function*. Weinheim: Wiley-VCH.
- Pellegrini, E., Piano, D. & Bowler, M. W. (2011). *Acta Cryst.* **D67**, 902–906.
- Pornillos, O. & Chang, G. (2006). *FEBS Lett.* **580**, 358–362.
- Sciara, G. & Mancia, F. (2012). *Curr. Opin. Struct. Biol.* **22**, 476–481.
- Wang, X. *et al.* (2012). *Nature Struct. Mol. Biol.* **19**, 424–429.
- Weierstall, U. *et al.* (2014). *Nature Commun.* **5**, 3309.
- White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.