Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution X-ray crystal structure of bovine H-protein at 0.88 Å resolution

Akifumi Higashiura,[a] Takeshi Kurakane,[a] Makoto Matsuda,[a] Mamoru Suzuki,[a] Koji Inaka,[b] Masaru Sato,[c] Tomoyuki Kobayashi,[c] Tetsuo Tanaka,[c] Hiroaki Tanaka,[d] Kazuko Fujiwara[e] and Atsushi Nakagawa[a]*

[a]Institute for Protein Research, Osaka University, Japan, [b]Maruwa Foods and Biosciences Inc., Japan, [c]Japan Aerospace Exploration Agency, Japan, [d]Confocal Science Inc., Japan, and [e]Institute for Enzyme Research, The University of Tokushima, Japan

Correspondence e-mail: atsushi@protein.osaka-u.ac.jp

Received 11 November 2009
Accepted 22 March 2010

PDB Reference: bovine H-protein, 3klr.

Recent technical improvements in macromolecular X-ray crystallography have significantly improved the resolution limit of protein structures. However, examples of high-resolution structure determination are still limited. In this study, the X-ray crystal structure of bovine H-protein, a component of the glycine cleavage system, was determined at 0.88 Å resolution. This is the first ultrahigh-resolution structure of an H-protein. The data were collected using synchrotron radiation. Because of limitations of the hardware, especially the dynamic range of the CCD detector, three data sets (high-, medium- and low-resolution data sets) were measured in order to obtain a complete set of data. To improve the quality of the merged data, the reference data set was optimized for merging and the merged data were assessed by comparing merging statistics and *R* factors against the final model and the number of visualized H atoms. In addition, the advantages of merging three data sets were evaluated. The omission of low-resolution reflections had an adverse effect on visualization of H atoms in hydrogen-omit maps. Visualization of hydrogen electron density is a good indicator for assessing the quality of high-resolution X-ray diffraction data.

1. Introduction

In recent years, macromolecular X-ray crystallography has been advanced significantly by the use of high-brilliance and low-divergence synchrotron beams, high-performance and high-precision large area detectors, cryocooling techniques and state-of-the-art data-reduction programs and mathematical improvement of the refinement software. Improvement of the methods and techniques of protein crystallography has increased the resolution limit and thus the quality of protein structures (Dauter *et al.*, 1997; Schmidt & Lamzin, 2002; Dauter, 2003; Vrielink & Sampson, 2003; Petrova & Podjarny, 2004). To date, the protein structures that have been refined against the highest resolution data include crambin (Jelsch *et al.*, 2000), hen egg-white lysozyme (Wang *et al.*, 2007) and human aldose reductase (Howard *et al.*, 2004) at resolutions of 0.54, 0.65 and 0.66 Å, respectively. As of February 2010, ∼250 structures at higher than 1.0 Å resolution have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000, 2002) out of a total of 63 000 structures. Examples of high-resolution X-ray crystallography are still not numerous: only 92 structures (<0.15% of the total) are beyond 0.9 Å resolution.

The number of measurements (namely, diffraction data) required for structure determination at 0.9 Å resolution is about 1.4 times that at 1.0 Å resolution. This increase in experimental data enables us to refine structural parameters

Table 1
Data-collection and scaling statistics.

Collection of the low-resolution data set was performed without a front beam stop in order to measure very low-resolution reflections. Values in parentheses are for the highest resolution shells.

	High-resolution	Medium-resolution	Low-resolution
Beamline	Photon Factory BL-5A		
Detector	ADSC Quantum315		
Wavelength (Å)	1.0		
Crystal-to-detector distance (mm)	59.9	90.8	284.2
Maximum resolution (Å)	0.88	0.99	1.96
Oscillation range (°)	1.0	1.0	1.0
No. of images	360	180	180
X-ray exposure time per frame (s)	10.0	3.0	1.0
Aluminium attenuator (mm)	0.0	0.0	0.3
Space group	*C*2		
Unit-cell parameters (Å, °)	a = 84.41, b = 41.25, c = 43.05, β = 91.18	a = 84.61, b = 41.34, c = 43.16, β = 91.10	a = 84.41, b = 41.30, c = 43.14, β = 91.14
Resolution range (Å)	14.9–0.88 (0.89–0.88)	20.4–1.00 (1.00–0.99)	43.1–1.96 (1.98–1.96)
No. of observed reflections	690629	277583	38945
No. of unique reflections	111864 (3714)	81155 (1903)	10767 (295)
Redundancy	6.2 (3.7)	3.4 (2.7)	3.6 (3.0)
Completeness (%)	95.6 (95.4)	97.9 (69.3)	99.5 (86.3)
$\langle I\rangle/\langle\sigma(I)\rangle$	47.4 (2.64)	34.3 (1.66)	49.2 (25.2)
$R_{\rm merge}$ (%)	4.7 (42.6)	3.2 (47.6)	4.4 (6.1)
No. of rejected reflections	12091	127	233
Wilson B factor (Å^2)	6.64	7.92	13.8

Table 2
Statistics of merged data sets with a reference data set.

Values in parentheses are for the highest resolution shell.

Reference data set	High-resolution	Medium-resolution	Low-resolution
Space group	*C*2		
Unit-cell parameters (Å, °)	a = 84.41, b = 41.25, c = 43.05, β = 91.18	a = 84.41, b = 41.34, c = 43.16, β = 91.18	a = 84.41, b = 41.31, c = 43.14, β = 91.14
Resolution range (Å)	43.1–0.88 (0.89–0.88)		
No. of observed reflections	1113257	1082860	1025526
No. of unique reflections	115668 (3712)	116414 (3683)	116058 (3690)
Redundancy	9.6 (4.7)	9.3 (3.4)	8.8 (3.6)
Completeness (%)	98.9 (95.4)	98.8 (95.0)	98.9 (95.8)
$\langle I\rangle/\langle\sigma(I)\rangle$	70.0 (3.72)	70.8 (2.45)	82.4 (2.60)
$R_{\rm merge}$ (%)	4.7 (38.6)	4.5 (44.5)	4.2 (42.7)
No. of rejected reflections	22060	25148	21843
Wilson B factor (Å^2)	7.50	8.65	10.7

more precisely and accurately. In high-resolution and well refined structures, we can visualize multiple conformations of main and/or side chains and accurate solvent structures and determine anisotropic temperature factors. Many H atoms can be visualized in hydrogen-omit maps generated using high-resolution data and the coordinates of H atoms can also be refined at ultrahigh resolution (<0.7 Å). H-atom positions are often important for understanding the function of enzymes (Vrielink & Sampson, 2003). As the amount of data increases, structural information also increases. Hence, it is clear that high-resolution data have an advantage in the structural refinement of proteins. However, it is difficult to collect, scale and merge high-resolution data and to carry out refinement of structures with a large number of parameters.

The glycine cleavage system is a mitochondrial multi-enzyme system that consists of four different proteins (P-, H-, T- and L-proteins); together, these proteins catalyze the oxidative cleavage of glycine. The glycine cleavage system is widely distributed in animals, plants and bacteria. The H-protein is a monomeric protein of molecular weight ∼14 kDa that plays a central role in glycine cleavage. The lipoic acid prosthetic group covalently bound to a specific lysine residue of the H-protein interacts with specific sites on the P-, T- and L-proteins (Kikuchi *et al.*, 2008). The gene for bovine H-protein was isolated from a bovine liver cDNA library (Fujiwara *et al.*, 1990) and the purified recombinant apo H-protein was lipoylated and activated *in vitro* by lipoyltransferase, using lipoyl-AMP as the lipoyl donor (Fujiwara *et al.*, 1992). The first structure of H-protein from pea leaves, with a reduced lipoic acid, was determined at 2.6 Å resolution (Pares *et al.*, 1994). The structures of H-protein from pea leaves with an oxidized lipoic acid and aminomethyllipoic acid were subsequently determined at 2.0 and 2.2 Å resolution, respectively (Pares *et al.*, 1995). Furthermore, the structure of *Thermus thermophilus* HB8 H-protein expressed in *Escherichia coli* was determined at 2.5 Å resolution and structural similarity to pea H-protein was reported (Nakai *et al.*, 2003). The structure of *Thermotoga maritima* H-protein determined at 1.65 Å resolution has been deposited in the Protein Data Bank (PDB code 1zko; Joint Center for Structural Genomics, unpublished work).

Here, we report the structure of bovine H-protein at 0.88 Å resolution. This is the first ultrahigh-resolution structure of an H-protein. In this report, we describe the features of the structure and investigate the treatment of high-resolution X-ray data. In addition, we investigate the importance of low-resolution data based on the visualization of H atoms. Visualization of hydrogen electron density is a good indicator for assessing the quality of high-resolution X-ray diffraction data.

2. Materials and methods

2.1. Expression, purification, crystallization, data collection and processing

The expression and purification of bovine H-protein were performed as described previously (Fujiwara *et al.*, 1992). The purified protein solution was concentrated to 15 mg ml^{-1}.

Crystallization was carried out by the hanging-drop vapour-diffusion method and the microseeding technique at 288 K. Cluster-like crystals for microseeding were grown in 2 µl drops

containing a 1:1(*v*:*v*) mixture of 15 mg ml^{-1} protein solution and 2.2–2.5 *M* ammonium sulfate in citrate buffer pH 2.8–3.2. Cluster-like crystals were crushed using a Seed Bead (Hampton Research) in the precipitation buffer. Single crystals for X-ray experiments were grown in 2 µl drops containing a 1:0.8:0.2(*v*:*v*) mixture of 15 mg ml^{-1} protein solution, 2.2–2.5 *M* ammonium sulfate in citrate buffer pH 2.8–3.2 and the diluted seed solution. Single crystals suitable for X-ray experiments, with dimensions of 0.2–0.3 mm, were obtained after a few days.

Data collections were performed using synchrotron radiation on Photon Factory beamline BL-5A equipped with an ADSC Quantum315 CCD detector under a nitrogen vapour stream at 90 K. Before freezing, each single crystal was transferred to reservoir solution including 30%(*v*/*v*) glycerol using a cryoloop. Three data sets, corresponding to high-, medium- and low-resolution reflections, were collected (Sevcik *et al.*, 1996). The high-resolution data set was collected with a long exposure time and a short crystal-to-detector distance. To collect the low-resolution data completely, the X-rays were attenuated by an aluminium attenuator in order to avoid saturation of high-intensity reflections. Additionally, the front beam stop was removed in order to measure the intensities of the lowest resolution reflections. The medium-resolution data were collected in order to connect the high- and low-resolution data. All data sets were measured from a single crystal in a fixed position and the sequence of data collection was high, medium and low. The details of data collection and the statistics of each data set are given in Table 1. All data sets were integrated, scaled and merged using the programs *DENZO* and *SCALEPACK* as implemented in the *HKL*-2000 program package (Otwinowski & Minor, 1997).

2.2. Phase determination and structure refinement

The high-resolution data set was used as the reference for scaling all three data sets (Table 2) and the resulting merged data were used for phase determination and refinement. The initial structure of H-protein was determined by molecular replacement with the program *MOLREP* (Vagin & Teplyakov, 1997) using the pea H-protein structure (PDB code 1hpc; Pares *et al.*, 1995) as a search model. Side-chain atoms beyond the C^{β} atom, water molecules and the lipoic acid were removed from the search model. In order to calculate the $R_{\rm free}$ factor, 5% of reflections were randomly selected and excluded from the following refinement. The program *ARP*/*wARP* (Perrakis *et al.*, 1999) was used to build missing side-chain atoms and add water molecules. The first step of an isotropic refinement was performed using the program *SHELXL* (Sheldrick, 2008) against the data from 10.0 to 1.5 Å resolution. After the isotropic refinement, previously unbuilt residues and multiple conformations were manually added using the program *Coot* (Emsley *et al.*, 2010). After manual adjustment, water molecules were located using *phenix.refine* (Adams *et al.*, 2002). Isotopic refinement was carried out and the maximum resolution was gradually extended to 1.0 Å. After further manual adjustments, we determined the occupancies of multiple conformations and flexible side chains. Anisotropic refinements were carried out using the program *SHELXL* against data from 10.0 to 1.0 Å resolution. After several rounds of anisotropic refinement using *SHELXL*, the resolution range was extended from 43.1 to 0.88 Å. H atoms were added to the model at the calculated positions without refinement of positions and *B* factors. No H atoms were added to waters or small molecules. The occupancies of the water molecules were refined. The quality of the final model was checked using the programs *WHAT IF* (Vriend, 1990) and *MolProbity* (Chen *et al.*, 2010). The statistics of the final model are summarized in Table 3.

Table 3
Statistics of final model.

Resolution range (Å)	43.1–0.88
R factor† (%)	11.3 (10.1)
Free *R* factor† (%)	13.2 (11.8)
No. of protein atoms	1135
No. of small molecules	22
No. of water atoms	274
R.m.s. deviations from ideal geometry	
Bond distances (Å)	0.017
Bond angles (°)	2.15
Standard deviation of ω value (°)	7.19
Ramachandran plot	
Residues in most favoured regions (%)	96.6
Residues in additional allowed regions (%)	3.4
Mean *B* factors (Å^2)	
Protein non-H atoms	
Overall	11.4
Main chain	9.28
Side chain	13.2
Small molecules	20.5
Water molecules	28.1

† *R* factors in parentheses are for reflections with $F_{\rm o} > 4\sigma(F_{\rm o})$.

2.3. Counting H atoms with significant electron density

In the refinement process, H atoms were added and subjected to refinement as a riding model. In order to avoid bias that might result from the riding-model treatment of hydrogen positions (Wang *et al.*, 2007), random coordinate noise with a maximum of 0.1 Å displacement was added to the final model using the program *PDBSET* (Collaborative Computational Project, Number 4, 1994). All anisotropic *B* factors were converted to isotropic *B* factors using *SHELXPRO*. The re-refinement was performed by *SHELXL*, using the input file derived from the coordinates with random errors. The first step was 30 cycles of isotropic refinement against the data to maximum resolution, followed by 30 cycles of anisotropic refinement against the same data. $\sigma_{\rm A}$-Weighted $F_{\rm o} - F_{\rm c}$ difference Fourier electron-density maps sampled at 864, 432 and 432 grid positions along the unit-cell edges were generated. Thus, the size of each grid was smaller than 0.1 × 0.1 × 0.1 Å. The number of electrons appearing at the calculated positions of atoms were determined using the program *MAPMAN* (Uppsala Software Factory, *RAVE* package) with the 'PEEK VALUE' command including an 'INTERPOLATE' option. H atoms bonded to protein atoms with partial occupancies, water molecules and small molecules were

Table 4
Maximum and mean differences and r.m.s. deviations between bovine and other H-proteins.

Tt and Tm represent *Thermus thermophilus* and *Thermotoga maritima*, respectively.

Reference range of bovine H-protein/ residue range	Species	PDB code (chain ID)/ residue range	Mean difference (Å)	R.m.s.d. (Å)	Maximum difference (Å)
3–125	Pea	1dxm(*A*)/7–129	0.729	0.864	3.245
3–125	Pea	1dxm(*B*)/7–129	0.700	0.828	2.833
3–125	Pea	1hpc(*A*)/7–129	0.697	0.824	3.176
3–125	Pea	1hpc(*B*)/7–129	0.658	0.796	2.817
3–125	Pea	1htp/7–129	0.623	0.730	2.797
3–125	Tt	1onl(*A*)/7–129	0.746	0.903	4.085
3–125	Tt	1onl(*B*)/7–129	0.887	1.142	5.826
3–125	Tt	1onl(*C*)/7–129	0.827	1.070	6.141
3–122	Tm	1zko(*A*)/4–123	0.466	0.591	4.493
3–122	Tm	1zko(*B*)/4–123	0.455	0.575	4.635

ignored. The threshold for determining H atoms was 0.16 e Å^{-3}. This threshold agrees with the results of manual counting of the H atoms.

3. Results and discussion

3.1. The structure of bovine H-protein

3.1.1. Overall structure. The overall structure of bovine H-protein is shown in Fig. 1. The structure mainly consists of two antiparallel β-sheets, with helices at the C-terminus joined to the main domain by a flexible linker. The two antiparallel β-sheets form a β-sandwich. One sheet is composed of four strands comprising residues 10–15, 18–23, 70–75 and 97–101 and the other is composed of three strands comprising residues 34–39, 51–57 and 61–65. There are also two short β-strands (4–5 and 45–46). A long disordered helix consisting of residues 104–109 (3_{10}-helix) and 113–125 (α-helix) is positioned at the C-terminus. Two other short α-helices (residues 25–31 and 85–88) and a 3_{10}-helix (residues 77–79) are positioned in the loops that join the β-sheets. Although Lys59 in one of these loops is the site of lipoylation, this residue has no lipoyl prosthetic group in this structure. The secondary structures were assigned by *DSSP* (Kabsch & Sander, 1983).

3.1.2. Comparison between bovine and other H-proteins. X-ray crystal structures of H-proteins derived from pea and bacteria have been deposited in the Protein Data Bank (PDB codes 1dxm, 1hpc, 1htp, 1onl and 1zko; Faure *et al.*, 2000; Pares *et al.*, 1995; Cohen-Addad *et al.*, 1995; Nakai *et al.*, 2003; Joint Center for Structural Genomics, unpublished work). The structural similarity between pea and *Thermus thermophilus* HB8 H-protein has previously been reported (Nakai *et al.*, 2003). Bovine H-protein shares 49% amino-acid sequence identity with the pea protein, 43% identity with the *T. thermophilus* HB8 protein and 50% identity with the *Thermotoga maritima* protein. As expected, the secondary structures of the bovine and other H-proteins are almost the same and the r.m.s. deviations for C$^{\alpha}$ atoms were 0.575–1.14 Å (Table 4). Differences are observed mainly in the C-terminal helices, flexible loops and around residue 59 of bovine H-protein. These results show that the secondary structures are highly conserved in the H-proteins of glycine cleavage systems.

3.1.3. Features of the high-resolution structure. The final electron-density map showed clear electron density for non-H atoms. The identities of atoms could be clearly assigned on the basis of electron density in many regions of the protein, as well as for solvent atoms. However, we could not model the N-terminus because of poor electron density. The final model consists of 1135 protein atoms, 274 water molecules, two sulfate ions and a glycerol molecule. Multiple conformations were observed for 29 residues. H atoms were observed as peaks in a hydrogen-omit electron-density map calculated at 0.88 Å resolution. ~40% of the H atoms could be visualized; H atoms bound to atoms with multiple conformations, waters and other solvent molecules were excluded from the count.

Crystals that diffract to atomic resolution tend to have lower solvent contents (35–40%) and lower symmetry (Schmidt & Lamzin, 2002; Bönisch *et al.*, 2005). However, the crystals of bovine H-protein had a solvent content of 54.8% (Matthews coefficient $V_{\rm M}$ = 2.72 Å^3 Da^{-1}). The high quality of the crystals despite their high solvent content could be rationalized in terms of crystal packing. The flexible loops that join the β-strands and the long helices at the C-terminus, which is separated from the core, are strongly held in place by other symmetry-related molecules (Fig. 2).

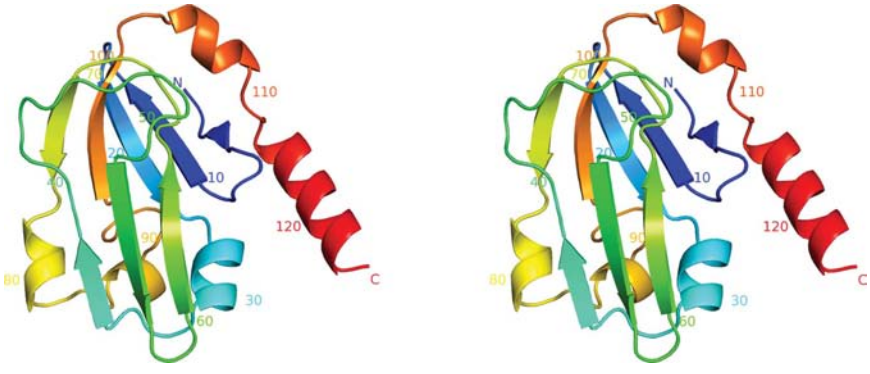

Figure 1
Stereoview of the overall structure of bovine H-protein. Schematic ribbon representation of bovine H-protein coloured from blue (N-terminus) to red (C-terminus). This figure was produced with the program *PyMOL* (DeLano, 2002).

3.2. Scaling and merging the high-, medium- and low-resolution data sets

3.2.1. Statistics of the high-, medium- and low-resolution data sets. Three data sets were collected at resolution cutoffs of 0.88 Å (high-resolution data set), 0.99 Å (medium-resolution data set) and 1.96 Å (low-resolution data set). The statistics of reduction for each data set are given in

Table 1. The overall R_{merge} values on intensities for the high-, medium- and low-resolution data were 4.7, 3.2 and 4.4% and their completeness values were 95.6, 97.9 and 99.5%, respectively. The resolution limit (0.88 Å) of the high-resolution data was determined based on the R_{merge} (∼40%) of the highest resolution shell; the $\langle I\rangle/\langle\sigma(I)\rangle$ values of the highest resolution shells were greater than 1.5 in all data sets. Fig. 3 shows the completeness as a function of resolution for the high-, medium- and low-resolution data sets. The high- and medium-resolution data sets were nearly complete at resolutions greater than ∼2.0 Å, but the completeness was not satisfactory at lower resolution (>∼15 Å), probably owing to saturation of intensities and the experimental conditions. The completeness of the low-resolution data set shows that we could measure the reflections of lowest resolution without a front beam stop between the crystal and the detector. To obtain a complete data set covering 43.1–0.88 Å resolution, these three data sets needed to be merged.

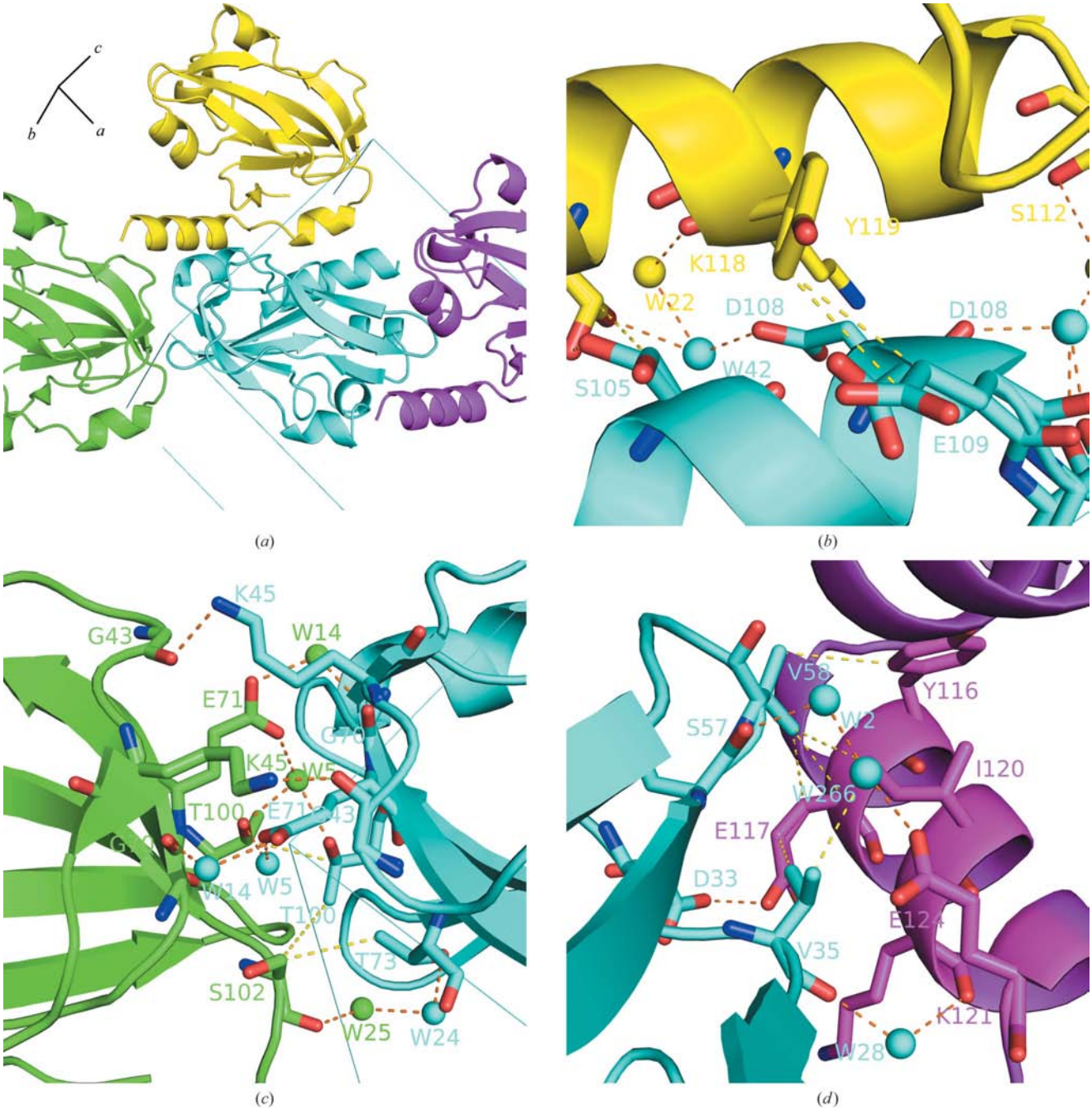

Figure 2
(*a*) Crystal packing of the bovine H-protein. Green, yellow and cyan molecules are at $(-x, y, -z)$, $(-x, y, -z + 1)$ and $(-x + 1/2, y + 1/2, -z + 1)$, respectively. (*b*), (*c*) and (*d*) show the regions of intermolecular interaction. Hydrogen bonds and hydrophobic interactions are represented by red and yellow dotted lines, respectively. These figures were produced with the program *PyMOL*.

3.2.2. Merging the high-, medium- and low-resolution data sets. Wilson plots for the three data sets are shown in Fig. 4. The large differences observed in the absolute scale and B factors of the three data sets reflect the conditions under which the data were recorded. Therefore, the three data sets had to be merged and scaled with caution.

To obtain the complete data set, the high-, medium- and low-resolution data sets were merged using *SCALEPACK* with the 'REFERENCE BATCH' option. The high-, medium- and low-resolution data sets were individually scaled and merged and used as the reference. The scaling and merging statistics are given in Table 2. The overall completeness was similar for the three data sets. The overall $R_{\rm merge}$ values on intensities were 4.7, 4.5 and 4.6% for the data obtained using the high-, medium- and low-resolution data as the reference data. When the medium-resolution data were used as the reference data in merging, the $R_{\rm merge}$ value was slightly better than for the others. On the other hand, the $R_{\rm merge}$ values in the highest resolution shells were 38.6, 44.5 and 54.0% for the high-, medium- and low-resolution reference data, respectively. The overall $\langle I\rangle/\langle\sigma(I)\rangle$ values were 70.0, 70.8 and 82.4, and the values in the highest resolution shell were 3.72, 2.45 and 2.60 for the high-, medium- and low-resolution reference data, respectively. The merged data set obtained using the high-resolution data as the reference showed the highest $\langle I\rangle/\langle\sigma(I)\rangle$ value in the highest resolution shell, but the overall value was the lowest. In contrast, the data set obtained using the low-resolution data as the reference showed a lower value in the highest resolution shell and the highest overall value. The data set obtained using the medium-resolution data as the reference had the lowest value in the highest resolution shell. The Wilson B values of the merged data obtained using high-, medium- and low-resolution data as the reference were significantly different (Table 2) and Wilson plots of the merged data sets showed significant differences in the high-resolution range (Fig. 5). We were not able to judge from these statistics which reference was the best with respect to the final merged data.

3.2.3. Counting visualized H atoms. In ultrahigh-resolution X-ray crystallography, significant electron density for a large number of H atoms can be observed in $\sigma_{\rm A}$-weighted $F_{\rm o} - F_{\rm c}$ hydrogen-omit maps. To assess the quality of the merged data sets obtained using the high-, medium- and low-resolution data sets as a reference, the numbers of visualized H atoms were counted in $\sigma_{\rm A}$-weighted $F_{\rm o} - F_{\rm c}$ hydrogen-omit maps.

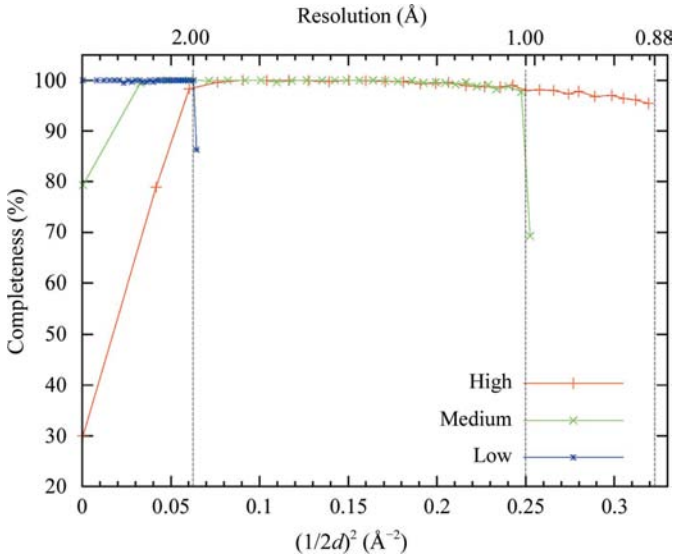

Figure 3
Completeness of the high-, medium- and low-resolution data sets. d represents the lattice-plane spacing (Å).

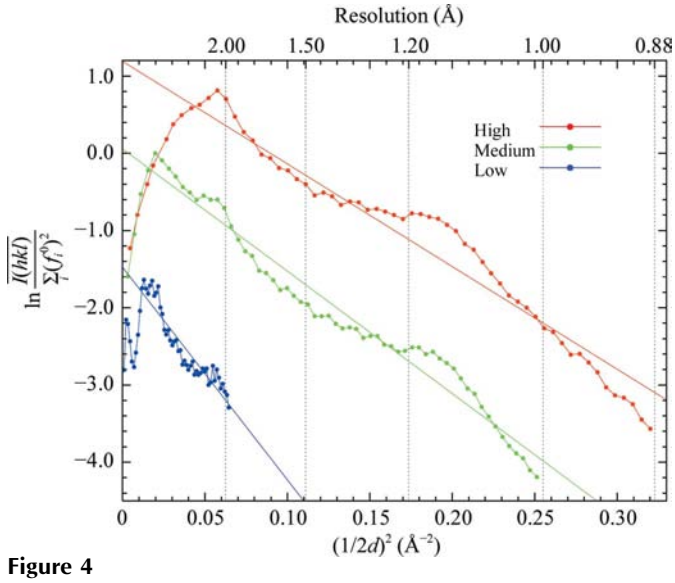

Figure 4
Wilson plots of the high-, medium- and low-resolution data set. The lines represent the regression curves: red, $y = 2(-6.64)x + 1.12$; green, $y = 2(-7.92)x + 0.0554$; blue, $y = 2(-13.8)x - 1.48$, where y and x are the vertical and horizontal axes, respectively. d represents the lattice-plane spacing (Å).

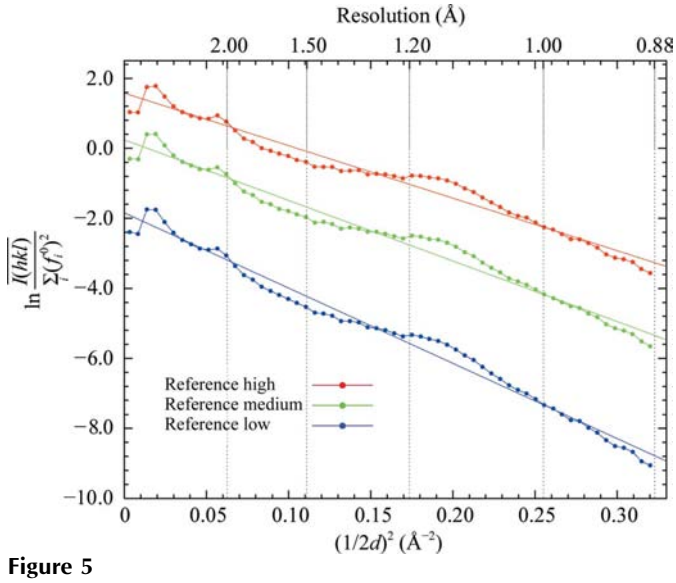

Figure 5
Wilson plots of the merged data set using the high-, medium- and low-resolution data sets as a reference. The lines represent the regression curves: red, $y = 2(-7.50)x + 1.58$; green, $y = 2(-8.65)x + 0.241$; blue, $y = 2(-10.7)x + 1.84$, where y and x are the vertical and horizontal axes, respectively. d represents the lattice-plane spacing (Å).

Table 5
R values against re-refinement models and the number of visualized H atoms for data sets with a reference.

Data set	R_{work}/R_{free} after re-refinement† (%)	Applied *B* factor (Å^2)	No. of visualized H atoms	R.m.s.d. for H-omit map (e Å^{-3})
Reference high	12.2 (11.2)/14.4 (13.0)	0.0	279 (41.5%)	0.0682
Reference medium	12.1 (11.1)/14.4 (13.2)	0.0	226 (33.6%)	0.0648
Reference low	12.4 (11.2)/14.5 (13.5)	0.0	170 (25.3%)	0.0615
Reference mid (−*B*)	—	−1.15	264 (39.3%)	0.0675
Reference low (−*B*)	—	−3.20	268 (39.9%)	0.0680

† *R* factors in parentheses are for reflections with $F_o > 4\sigma(F_o)$.

Results are given in Table 5 and parts of maps are shown in Fig. 6. The numbers of visualized H atoms were 279, 226 and 170 for the high-, medium- and low-resolution references, respectively; the number for the high-resolution reference is about 1.6 times larger than that for the low-resolution reference. The reliabilities of structures using each data set are validated by the *R* values in Table 5.

3.2.4. Applying negative *B* factors. Significant differences in the number of visualized H atoms and the Wilson *B* values were observed among data sets obtained using different data as the reference data for merging. To clarify the effect of Wilson *B* values in the visualization of H atoms, we modified the merged data obtained using the low- or medium-resolution data as the reference data as follows:

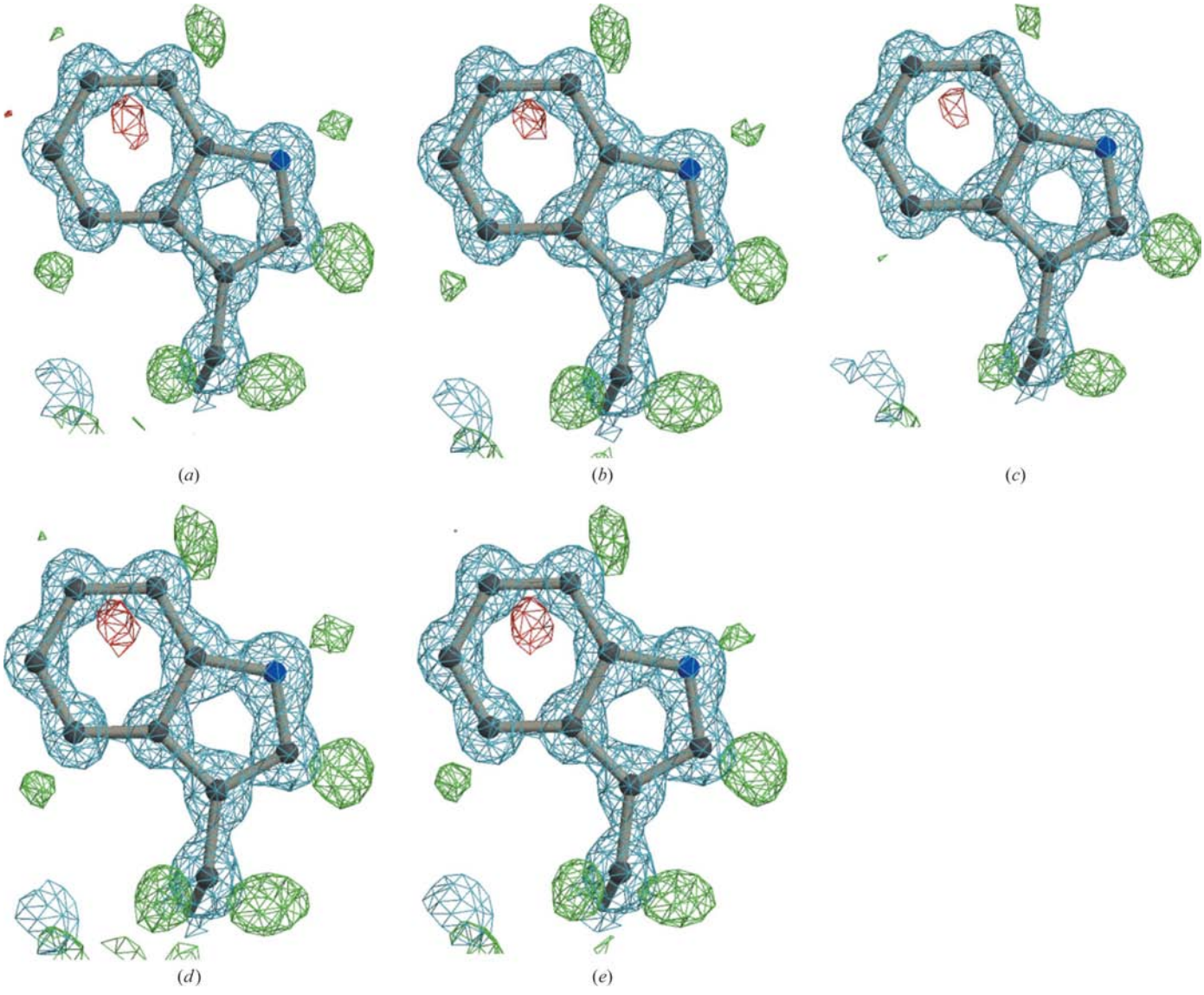

Figure 6
σ_A-Weighted $2F_o - F_c$ (blue) and σ_A-weighted $F_o - F_c$ (positive, green; negative, red) electron-density maps contoured at 1.00 and ±0.15 e Å^{-3}, respectively. All figures show the map around residue Trp11. (*a*) Reference high-resolution data set. (*b*) Reference medium-resolution data set. (*c*) Reference low-resolution data set. (*d*) Reference medium-resolution data set with application of $B = -1.15$ Å^2. (*e*) Reference low-resolution data set with application of $B = -3.20$ Å^2. The figures were produced with the program *POVScript+* (Fenn *et al.*, 2003).

$$F_{B\text{-map}} = \exp(-B\sin^2\theta/\lambda^2) \times (mF_o - DF_c), \qquad (1)$$

where $F_{B\text{-map}}$ is the structure factor for a σ_A-weighted $F_o - F_c$ map with a negative B factor and $mF_o - DF_c$ is the coefficient for calculation of the standard σ_A-weighted $F_o - F_c$ map. The negative B factors increased the number of visualized H atoms in the low- and medium-resolution reference data sets (Table 5). In addition, Figs. 6(*d*) and 6(*e*) show that the quality of the maps with negative B values is almost equal to that of the map from the high-resolution reference (Fig. 6*a*).

These results showed that the visualization of H atoms was sensitive to the B values in hydrogen-omit maps. In general, the overall B values were affected by radiation damage. In our data collection radiation damage was obvious, especially in the high-resolution data set, and the scaling B values were increased by 1.4 Å^2 in the final frame of the high-resolution data set relative to the initial frame. To consider the effects of radiation damage and the selection of the reference data set separately, the high-, medium- and low-resolution data sets were merged using typical images as references. The initial and the final batches of the high-resolution data set and the initial batches of the medium- and low-resolution data sets were chosen for the reference batch; the resulting Wilson B values of the merged data were 7.0, 8.4, 8.5 and 11.5 Å^2, respectively. These images were collected at the same position of the crystal. Compared with the total exposure time of the high- or medium-resolution data sets, the total exposure time of the low-resolution data set was very short; therefore, the increase in B factor must be smaller than that for the high- or medium-resolution data set. Nevertheless, the Wilson B factor of the merged data sets using the low-resolution data as a reference was higher than the others. In the merging process, the choice of a reference data set has to be completely irrelevant, so the higher Wilson B factor of the low-resolution reference data might be a consequence of a limitation of the scaling algorithm. We concluded that the choice of reference data set was important.

In addition, when we applied larger negative B values hydrogen densities were emphasized but noise also increased significantly (data not shown).

3.3. Combinations of data sets

3.3.1. Statistics of different combinations of data. To estimate the advantages of merging the high-, medium- and low-resolution data sets, we generated several merged data sets: the high-, medium- and low-resolution data set (HML; 43.1–0.88 Å resolution), the high- and medium-resolution data set (HM; 20.4–0.88 Å resolution) and the high- and low-resolution data set (HL; 43.1–0.88 Å resolution).

The statistics for HML are given in Table 2 (High-resolution column) and the statistics for HM and HL are given in Table 6. The statistics for the high-resolution data set (H; 13.5–0.88 Å resolution) are summarized in Table 1. The completeness values in the lowest resolution shell (43.1–2.73 Å) were 100.0% for HML, 85.0% for HM, 100.0% for HL and 30.0% for H data, respectively. The maximum resolution is the same in all data sets.

Table 6
Statistics of merged data sets.

Values in parentheses are for the highest resolution shell.

Data set	HM	HL
Space group	*C*2	
Unit-cell parameters (Å, °)	a = 84.61, b = 41.34, c = 43.05, β = 91.18	a = 84.41, b = 41.31, c = 43.14, β = 91.14
Resolution range (Å)	20.3–0.88 (0.89–0.88)	43.1–0.88 (0.89–0.88)
No. of observed reflections	1077943	839325
No. of unique reflections	115067 (3714)	115574 (3712)
Redundancy	9.4 (4.7)	7.3 (4.7)
Completeness (%)	98.3 (95.5)	98.8 (95.4)
$\langle I\rangle/\langle\sigma(I)\rangle$	60.8 (3.73)	60.8 (3.72)
R_{merge} (%)	4.4 (46.6)	4.4 (38.6)
No. of rejected reflections	17254	15246
Wilson B factor (Å^2)	7.39	7.49

3.3.2. Counting visualized H atoms. For each of the combined data sets, the H atoms were counted in σ_A-weighted $F_o - F_c$ hydrogen-omit maps. The numbers of visualized H atoms and the resulting values for the R and R_{free} factors of the re-refinements are summarized in Table 7. The number of visualized H atoms was largest for the HML data and second largest in the HL map. In addition, the number of visualized H atoms in the HML map was larger than that in the HM map and the number in the HL map was larger than that in the H map. These results, with and without the low-resolution data set, suggest that the low-resolution data are important for locating H atoms in σ_A-weighted $F_o - F_c$ hydrogen-omit maps (Figs. 6*a*, 7*a*, 7*b* and 7*c*). The completeness values for each combination data set, as a function of resolution, are given in Fig. 8. It is clear that the incompleteness of low-resolution reflections causes a reduction in the number of visualized H atoms.

To compare the effect of redundancy on the visualization of H atoms, the reflections that were not present in the H data set were excluded from the HML merged data. The quality of the reflections of the 'H from HML' data was higher than for the H data owing to the higher redundancy in the former. In addition, the quality of the 'H from HML' data was the same as for the HML data, whereas the completeness of the 'H from HML' data was lower than that of the HML data at low resolution. Similarly, we also generated 'HM from HML' and 'HL from HML' data. Using the HM, HL and H from HML data, the number of visualized H atoms was counted and the results are summarized in Table 7. The number of visualized H atoms in the high-redundancy data was slightly larger than or nearly equal to the number in the map of the low-redundancy data. Consequently, high redundancy of the merged data slightly improved the electron density of H atoms, but at low resolution the contribution was much smaller than that of the completeness.

3.3.3. Applying negative *B* factors. Figs. 7(*d*), 7(*e*) and 7(*f*) show σ_A-weighted maps that were calculated with negative B factors. The number of visualized H atoms was comparable to that for the HML data, but the map quality was poorer than those using HML. In Figs. 7(*g*), 7(*h*) and 7(*i*), σ_A-weighted $F_o - F_c$ maps contoured at the 2.1σ level were calculated for

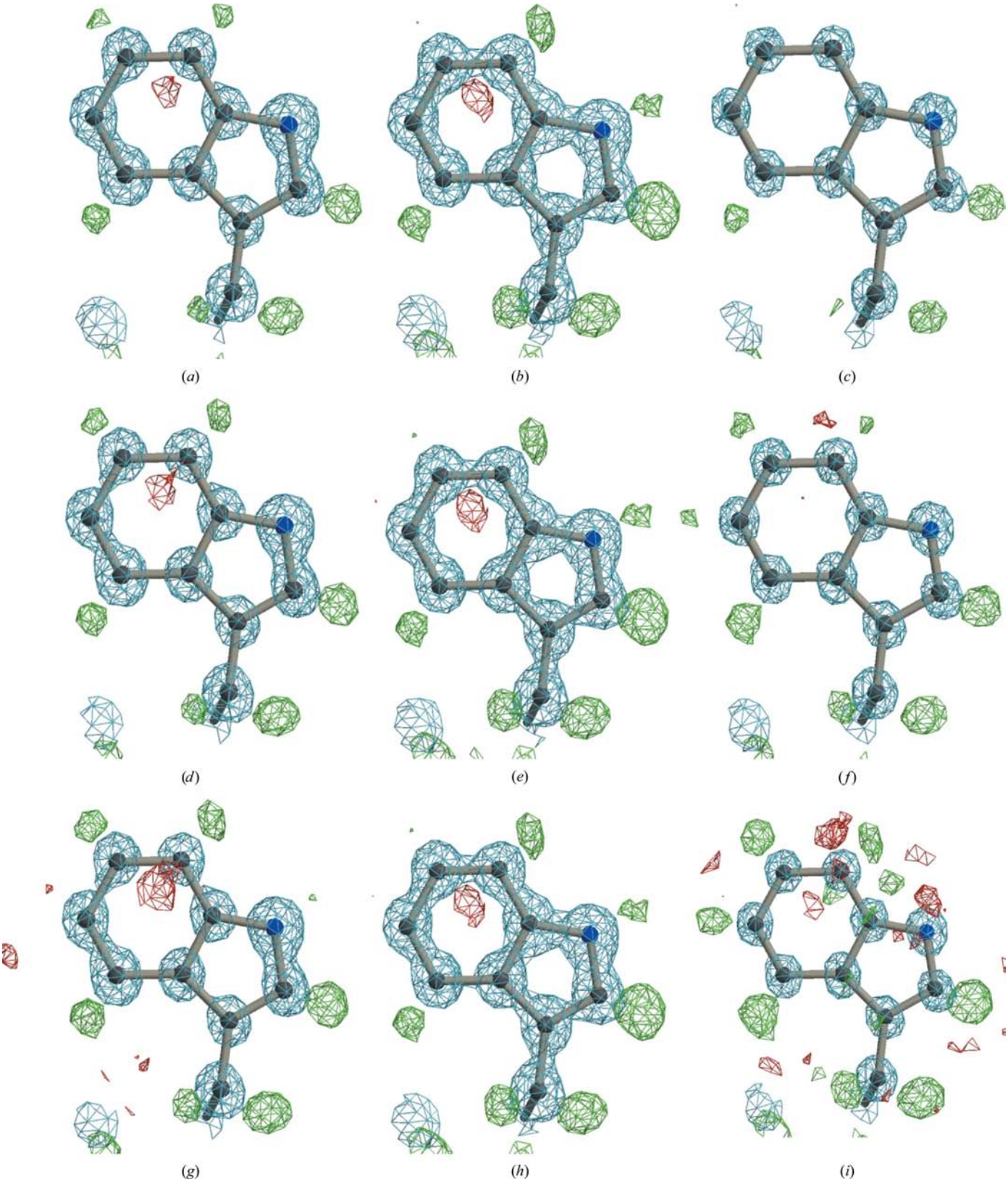

Figure 7
σ_A-Weighted $2F_o - F_c$ (blue) and σ_A-weighted $F_o - F_c$ (positive, green; negative, red) electron-density maps. All figures show the map around residue Trp11. The σ_A-weighted $2F_o - F_c$ (blue) maps are contoured at 1.00 e Å^{-3}. The σ_A-weighted $F_o - F_c$ maps of (*a*) HM, (*b*) HL and (*c*) H are contoured at $\pm$0.15 e Å^{-3}. The σ_A-weighted $F_o - F_c$ maps of (*d*) HM, (*e*) HL and (*f*) H were generated with an increasing B of -0.8, -0.1 and -2.4 Å^2 and are contoured at $\pm$0.15 e Å^{-3}. The σ_A-weighted $F_o - F_c$ maps of (*g*) HM, (*h*) HL and (*i*) H were contoured at 2.1σ with application of the same negative B factors. The figures were produced with the program *POVScript+* (Fenn *et al.*, 2003).

Table 7
R values against re-refinement models and the numbers of visualized H atoms for combinations of data sets.

Data set	R_{work}/R_{free} after re-refinement (%)	Applied *B* factor (Å^2)	No. of visualized H atoms (%)	R.m.s.d. for H-omit map (e Å^{-3})
HML	12.2 (11.2)/14.4 (13.0)	0.0	279 (41.5)	0.0682
HM	12.1 (11.0)/14.7 (13.3)	0.0	251 (37.4)	0.0581
HL	12.3 (11.2)/14.3 (13.0)	0.0	276 (41.1)	0.0683
H	11.9 (10.3)/14.0 (12.1)	0.0	133 (19.8)	0.0420
HM from HML	12.1 (11.0)/14.7 (13.3)	0.0	248 (36.9)	0.0584
HL from HML	12.2 (11.2)/14.3 (13.0)	0.0	278 (41.4)	0.0682
H from HML	12.0 (10.6)/14.1 (12.4)	0.0	137 (20.4)	0.0427
HM (−*B*)	—	−0.8	279 (41.5)	0.0605
HL (−*B*)	—	−0.1	279 (41.5)	0.0686
H (−*B*)	—	−2.4	278 (41.4)	0.0511

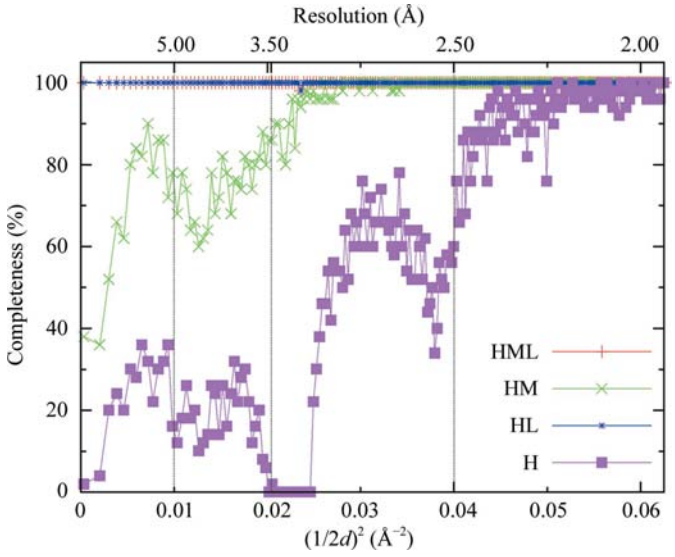

Figure 8
Completeness of each combination of data as a function of resolution at low resolution. *d* represents the lattice-spacing distance (Å).

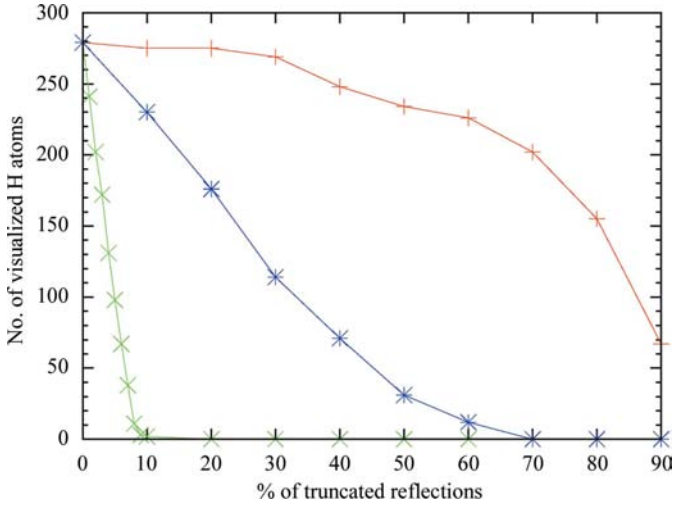

Figure 9
Truncation of reflections. Red and green curves show the truncation of high- and low-resolution reflections, respectively. The blue curve shows random truncation.

comparison with the HML maps. Although the number of visualized H atoms was almost the same, the negative peaks were more numerous in the σ_A-weighted $F_o - F_c$ maps of the HM, HL and H data.

3.4. Effects of removing selected reflections

To assess the contribution of low-resolution reflections to the visualization of H atoms, high- or low-resolution reflections, as well as randomly selected reflections, were removed from the HML data; σ_A-weighted $F_o - F_c$ hydrogen-omit maps were calculated using these truncated data with the phases and structure factors from the original data. The number of visualized H atoms is plotted as a function of the percentage of reflections removed in Fig. 9. Removing reflections from the high-resolution data reduced the number of visualized H atoms and only half the number were observed upon truncation of the reflections by 20%. Truncation of high-resolution reflections by 80, 50 and 10% resulted in resolution cutoffs of 1.5, 1.1 and 0.91 Å, respectively. The number of visualized H atoms in the 0.91 Å map was almost the same as the number in the 0.88 Å map. When reflections were removed randomly, the number of visualized H atoms was reduced in proportion to the percentage of reflections removed. The resolution range for the randomly truncated data was ∼43.1–0.88 Å. Removal of 10% of the reflections decreased the number of visualized H atoms by 13%. When low-resolution reflections were removed, the number of visualized H atoms dramatically reduced and no H atoms were observed when even as few as ∼10% of reflections were removed. Removal of ∼4% of the reflections halved the number of visualized H atoms. Removal of 4% of the low-resolution reflections is equivalent to removing data below 2.9 Å resolution and the removal of 10% of low-resolution reflections is equivalent to removing data below 1.9 Å resolution.

4. Conclusion

We collected X-ray diffraction data from a crystal of bovine H-protein using synchrotron radiation and determined the structure at 0.88 Å resolution. This is the highest resolution structure of an H-protein from any source. Comparisons of the atomic structure of bovine and other H-proteins demonstrate that the secondary structures of H-proteins are highly conserved.

To overcome the limitations of intensity measurements and the limited dynamic range of the CCD detector used, we measured three data sets in order to obtain a complete set of data under different experimental conditions. The high-, medium- and low-resolution data sets were all of very high quality. The three data sets could be merged using any of them as a reference. The statistics of the merged data sets were

similar, except for $\langle I\rangle/\langle\sigma(I)\rangle$ and the Wilson B values. To assess the quality of the merged data sets obtained using high-, medium- and low-resolution data as the reference, the number of visualized H atoms was counted in σ_A-weighted $F_o - F_c$ hydrogen-omit maps. The results showed that the merged data set using the high-resolution data as the reference was better than the other two sets. This effect is probably a result of radiation damage and the scaling algorithm. Thus, it is important to choose the high-resolution data set as the reference.

The advantage of merging three data sets was investigated using seven merged data sets (HML, HM, HL, H, HM from HML, HL from HML and H from HML). The qualities of the merged data sets were assessed by comparing R factors against the re-refinement models and by counting H atoms with significant electron density, and the results suggested that the low-resolution reflections contributed significantly to the visualization of H atoms. In addition, the medium-resolution reflections improved the redundancy and completeness, and the number of visualized H atoms slightly increased. In summary, merging of the three data sets was effective in determining a structure of high quality as well as for locating H atoms.

When high- and low-resolution reflections were removed from the list of reflections, the number of visualized H atoms was reduced. In particular, truncation of low-resolution reflections strongly affected the visualization of H atoms. These observations suggest that low-resolution reflections are critical in the visualization of H atoms even in high-resolution crystal structures.

This study was supported by the JAXA–GCF project 'High-quality Protein Crystallization Project on The Protein Structure and Function Analysis for Application' conducted by JAXA. The authors thank Professor Akira Sawaoka for his support of this project. The authors thank Professor Yoshiki Higuchi of the University of Hyogo, Professor Sam-Yong Park of Yokohama City University, Ms Sachiko Takahashi, Ms Mari Yamanaka, Mr Bin Yan and other members of Confocal Science Inc. and Mr Naoki Furubayashi of Maruwa Foods and Biosciences Inc. for their support throughout the project. The authors thank Dr Masahiko Hiraki and Dr Yusuke Yamada of the Structural Biology Research Center of the Photon Factory, Dr Eiki Yamashita of the Institute for Protein Research and Dr Masato Yoshimura of the National Synchrotron Radiation Research Center for their support and assistance with data collection on synchrotron-radiation beamlines.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Berman, H. M. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissing, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.
Chen, V. B., Arendall, W. B. III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cohen-Addad, C., Pares, S., Sieker, L., Neuburger, M. & Douce, R. (1995). *Nature Struct. Biol.* **2**, 63–68.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z. (2003). *Methods Enzymol.* **368**, 288–337.
Dauter, Z., Lamzin, V. S. & Willson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
DeLano, W. L. (2002). *PyMOL Molecular Viewer.* DeLano Scientific, San Carlos, California, USA.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Faure, M., Bourguignon, J., Neuburger, M., Macherel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C. & Douce, R. (2000). *Eur. J. Biochem.* **267**, 2890–2898.
Fenn, T. D., Ringe, D. & Petsko, G. A. (2003). *J. Appl. Cryst.* **36**, 944–947.
Fujiwara, K., Okamura-Ikeda, K. & Motokawa, Y. (1990). *J. Biol. Chem.* **265**, 17463–17467.
Fujiwara, K., Okamura-Ikeda, K. & Motokawa, Y. (1992). *J. Biol. Chem.* **267**, 20011–20016.
Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kikuchi, G., Motokawa, Y., Yoshida, T. & Hiraga, K. (2008). *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **84**, 246–263.
Nakai, T., Ishijima, J., Masui, R., Kuramitsu, S. & Kamiya, N. (2003). *Acta Cryst.* D**59**, 1610–1618.
Otwinowski, Z. & Minor, M. (1997). *Methods Enzymol.* **276**, 307–326.
Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. & Douce, R. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4850–4853.
Pares, S., Cohen-Addad, C., Sieker, L. C., Neuburger, M. & Douce, R. (1995). *Acta Cryst.* D**51**, 1041–1051.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Petrova, T. & Podjarny, A. (2004). *Rep. Prog. Phys.* **67**, 1565–1605.
Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Schmidt, A. & Lamzin, V. S. (2002). *Curr. Opin. Struct. Biol.* **12**, 698–703.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vrielink, A. & Sampson, N. (2003). *Curr. Opin. Struct. Biol.* **6**, 709–715.
Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Cryst.* D**63**, 1254–1268.