

Article

Open-Bundle Structure as the Unfolding Intermediate of Cytochrome *c'* Revealed by Small Angle Neutron Scattering

Takahide Yamaguchi ^{1,2}, Kouhei Akao ¹, Alexandros Koutsoubas ³, Henrich Frielinghaus ³ and Takamitsu Kohzuma ^{1,2,*}

¹ Institute of Quantum Beam Science, Graduate School of Science and Engineering, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan

² Frontier Research Center of Applied Atomic Sciences, Ibaraki University, 162-1 Shirakata, Tokai, Ibaraki 319-1106, Japan

³ Julich Centre for Neutron Science at Heinz Maier-Leibnitz Zentrum, Forschungszentrum Julich GmbH, Lichtenbergstrasse 1, Garching D-85747, German

* Correspondence: takamitsu.kohzuma.qbs@vc.ibaraki.ac.jp; Tel.: +81-29-228-8372

Abstract: The open-bundle structure of cytochrome *c'* as an unfolding intermediate was determined by small-angle neutron scattering experiment (SANS). The four- α -helix bundle structure of Cyt *c'* at neutral pH was transited to an open-bundle structure (at pD ~13), which is a joint-clubs consisting of four clubs (α -helices) connected by short loops. The compactly folded structure of Cyt *c'* (radius of gyration, R_g = 18 Å for the Cyt *c'* dimer) at neutral or mildly alkaline pD transitioned to a remarkably larger “open-bundle” structure at pD ~13 (R_g = 25 Å for the Cyt *c'* monomer). Cyt *c'* adopts an unstructured random coil structure at pD = 1.7 (R_g = 25 Å for the Cyt *c'* monomer). Numerical partial scattering function analysis (joint-clubs) and *ab initio* modelling gave structures similar to the “open-bundle”, which retains the α -helices but loses the bundle structure.

Keywords: Protein Unfolding; Cytochrome *c'*; Small-Angle Neutron Scattering; Open-Bundle Structure

1. Introduction

Protein molecule is folded into a distinctive structure to express the unique role in many biological systems. The mechanisms of protein folding/unfolding have been studied long time, because the knowledge of protein folding/unfolding is a key issue to predict the three dimensional protein structure from the amino acid sequence for the developments of medicinal drugs and functional cosmetics.

Anfinsen showed that most small proteins fold spontaneously into their specific functional structure [1,2]. Many protein folding/unfolding experiments have been performed to date to reveal protein self-assembly mechanisms, and several general mechanisms have been proposed. The “energy landscape” model is a widely accepted hypothesis for describing protein folding pathways. This model proposes that the protein folding reaction explores exothermic conformational arrangements of the polypeptide along the potential energy surface, leading to the native protein structure. Many non-native protein structures are trapped in local minima on the energy surface. The structurally unidentifiable non-native state (molten globule) is a presumed intermediate during the folding/unfolding reaction [3–7]. Numerous kinetic studies have probed the structures and properties of transient intermediate states [8,9], using far-UV CD, FT-IR, fluorescence, and NMR spectroscopy, as well as small-angle X-ray/neutron scattering and dynamic light scattering [7,10–16]. Englander et al. reported that the protein domains of cytochrome *c* and RNase are sequentially stabilized and folded in an orderly manner [17–19].

Together with experimental studies, molecular dynamics (MD) simulation is a useful technique to understand protein folding/unfolding mechanisms [20]. MD simulation of the unfolding of lysozyme showed that unfolding is triggered by the loss of hydrophobic

contacts at the inter-domain surface of the enzyme [21]. Lindorff-Larsen et al. reported that protein folding first occurs locally, followed by stabilization of the local structural elements accompanied by intermolecular interactions, which further promote the folding process [22].

Structural characterization of the folding/unfolding intermediate species is key to understanding the principles underlying protein structure formation. The application of small-angle X-ray (SAXS) and neutron scattering (SANS) techniques to protein molecules provides information on their flexibility, size, and structural morphology in solution [23,24]. The unfolding intermediate of lysozyme was previously characterized using SAXS [16]. The unfolding of lysozyme is triggered by extension of the β -domain, which preserves the folding of the α -domain. Synchrotron radiation SAXS rapidly provides data compared to SANS experiments. The combination of flow-cell experiments [25] and size exclusion chromatography [26] with SAXS reduces radiation damage to and aggregate formation by protein molecules.

SANS experiments are very useful for determining the solution structure of a protein without radiation damage. The energy of the neutron beam used for SANS experiments is on the order of meV, which is almost identical to the energy of the infrared region. Very recent progress in the *ab initio* structure analysis of small-angle scattering data has made possible “low-resolution structure models” that represent the overall shape of the solution protein structure [11,27,28].

Cytochrome c' (Cyt c') is a member of the c -type cytochrome family. Cyt c' comprises a four- α -helix bundle structure, in contrast to cytochrome c , with a heme prosthetic group embedded in the C-terminal region (Figure 1). Cyt c' undergoes a unique pH-dependent spin state transition. The spin state between pH 3 and 7 is likely a quantum mechanical admixture of an intermediate-spin (IS: $S = 3/2$) and a high-spin (HS) state. The spin state transition into the HS state at alkaline conditions ($8 < \text{pH} < 12$) is triggered by breakage of the inter-helix hydrogen-bonding linkage between helix C and helix D [29–32]. The structure of Cyt c' at pH \sim 13 was suggested to be a 6-coordinated low-spin (LS) state with imidazole (His120)/OH[−] or two OH[−] as the axial ligands, resulting in an “open-bundle” structure at pH \sim 13, as reported previously [33].

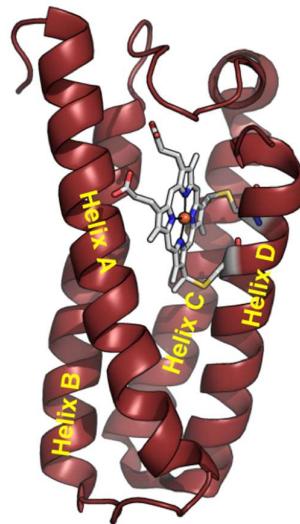


Figure 1. Structure of Cyt c' from *Alcaligenes xylosoxidans* NCIMB 11015 (PDB code: 4WGZ).

Very recently a program AlphaFold running on the Artificial Intelligence, Google DeepMind succeeded to provide 3D structure from the amino acid sequence [34]. The

experimental solution structure determination of unfolded protein is very important to verify such computationally predicted protein structure.

With these points in mind, we determined the SANS solution structures of Cyt *c'* under different pD conditions to better understand the mechanism underlying the alkaline structure transition of Cyt *c'*. This is the first report of the alkaline transition intermediate structure of Cyt *c'* at pD ~13. Determining the structures of Cyt *c'* in solution at pD 6.4, pD 9.4 and pD ~13 provides a general explanation of protein folding/unfolding mechanisms and the relevance of these mechanisms to the alkaline spin-state transition of Cyt *c'*.

2. Materials and Methods

2.1. Sample preparation for the SANS experiments

Cyt *c'* was extracted from *Alcaligenes xylosoxidans* and purified by cation exchange chromatography (CM-Sephadex C-50, GE Healthcare) and size-exclusion chromatography (Hiload Superdex 200 pg, GE Healthcare), followed by re-crystallization according to a previous method [33]. The purity of Cyt *c'* was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, Figure S1).

Samples for SANS experiments were prepared at pD 1.7, 6.4, 9.6, and ~13 with 20 mM phosphate buffer. The phosphate buffers used for pD 1.7 and ~13 samples were prepared by adding small amount of DCl and NaOD. The hydrogen atoms in potassium dihydrogen phosphate and dipotassium hydrogen phosphate were substituted with deuterium by repeated evaporation/addition of D₂O to the solution. The buffer solutions were prepared using D₂O, K₂DPO₄, K₂DPO₄, DCl, and NaOD. A small amount of concentrated Cyt *c'* D₂O solution was added to the D₂O buffer solution at the target pD (pD-jump method). The concentrations of Cyt *c'* at pD 1.7 and 6.4 were 5.5 mg/mL and at pD 9.6 and ~13 were 5.3 mg/mL. The electronic absorption spectrum of Cyt *c'* at each pD value was measured using a nanodrop-2000C spectrometer (Thermo Fisher) to check the electronic state of the heme in Cyt *c'* (Figure S2). The pD values of the buffers and samples prepared in D₂O were measured using a pH electrode and corrected to pD by adding 0.4, based on an earlier report.[35]

2.2. SANS measurements

The SANS experiments were performed at the KWS-1 beam line at the Julich Centre for Neutron Science (JCNS) at the Research Neutron Source Heinz Maier-Leibnitz Zentrum (FRM II) in Garching, Germany [36] using 2 mm path length quartz cells. The wavelength (λ) of the incident neutron beam was 5 Å, selected using the velocity selector, with a spread $\Delta\lambda/\lambda = 10\%$.

The collimator and detector were configured symmetrically at 8 m and 4 m for the low-Q region, respectively, and the configuration for the high-Q region was 4 m for collimation and 1.5 m for detection. Background scattering from the buffer solutions and empty cell were obtained in the same configurations and used to subtract unwanted background scattering. A secondary calibration standard of Plexiglas was used to determine the scattering intensity on an absolute scale and to correct detector sensitivity.

2.2. SANS data reduction and analysis

Data reduction from raw SANS profile images to 1D SANS curves, including all the above-mentioned corrections and calibrations, was conducted using qtiKWS. Radial averaging converted the 2D data to 1D intensity profiles ($I(Q)$), with the scattering vector Q being calibrated by the wavelength and scattering geometry. The data obtained for each sample with different sample-to-detector distances were merged. The background was removed by subtraction of the scattering intensity from individual buffer solutions.

The radius of gyration (R_g) of Cyt c' at each condition was estimated from the linear low- Q region of the Guinier plot and the pair distribution function, $P(r)$, using the PRIMUS program [37] in the ATSAS package. The gradual increase in the low Q intensity at $pD \sim 13$ could be due to a minor contribution of weak aggregation. The region corresponding to weak aggregation was not used for the Guinier and $P(r)$ analyses. The data ranges for Guinier analysis and the calculation of $P(r)$ were $0.001 \text{ \AA}^{-2} < Q^2 < 0.002 \text{ \AA}^{-2}$ and $0.05 \text{ \AA}^{-1} < Q < 0.25 \text{ \AA}^{-1}$, respectively. R_g at $pD 1.7$ was estimated from fitting with the Debye function ($D(Q)$) corresponding to the scattering curve from a polymer in random walk conformation using a numerical model described previously [38]. The R_g values were also calculated from the crystal structures at pH 6.0 (PDB ID: 4WGZ) and 11.4 (PDB ID: 4WGY) using CRYSON [39]. Further detailed analyses (joint-club model, molecular weight estimation, bead modelling, estimation of the aggregated particle fraction) were described in the Appendix A-D and Figure S3.

3. Results

SANS experiments with Cyt c' were performed at $pD 1.7$, 6.4 , 9.6 , and ~ 13 . Figure 2 shows plots of the SANS intensity of Cyt c' as a function of neutron scattering momentum, Q . The SANS curves at $pD 6.4$ and 9.6 have almost identical profiles, with a shoulder band at 0.2 \AA^{-1} which is absent at $pD 1.7$ and ~ 13 . The scattering intensities at $pD 1.7$ and ~ 13 were lower than at $pD 6.4$ and $pD 9.6$, likely due to dissociation of the dimer to the monomer. Furthermore the shoulder band at $pD 6.4$ and 9.6 observed at 0.2 \AA^{-1} is reproduced in the simulated SANS curve for the dimer structure (Figure 2B and 2C) and thus is characteristic of the dimer structure of Cyt c' .

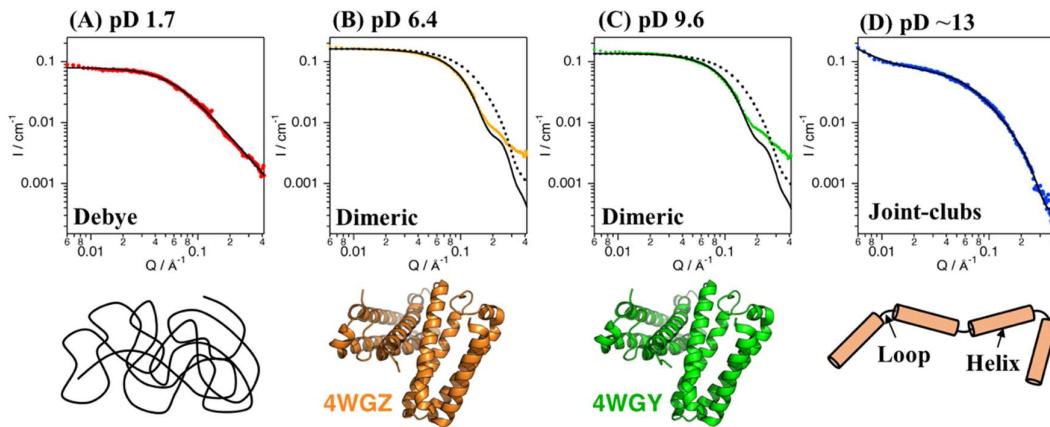


Figure 2. SANS data for Cyt c' at $pD 1.7$ (A, red), 6.4 (B, orange), 9.6 (C, green), and ~ 13 (D, blue). The black solid lines represent the simulated SANS profiles for Cyt c' . Fitting of the Debye function and joint-clubs model are shown with the SANS curves at $pD 1.7$ and ~ 13 , respectively. The simulated SANS curves obtained from the dimer (solid line) and monomer (broken line) crystal structures of Cyt c' at pH 6.0 and 10.4 are shown in panels (B) and (C), respectively. Schematic representation of the random coil (A), folded dimer (B, C), and open-bundle (D) structures of Cyt c' are provided under the respective SANS curves.

The experimentally obtained radius of gyration (R_g) values of Cyt c' at various pD were evaluated from the Guinier plot and $P(r)$ function (Figure 3) and are summarized in Table 1. The experimental and calculated R_g values from the crystal structures of Cyt c' at pH 6.0 and 10.4 are summarized in Table 2. The experimental R_g values of Cyt c' at $pD 6.4$ and 9.4 were $18\text{-}19 \text{ \AA}$ and $25\text{-}28 \text{ \AA}$ at $pD 1.7$. The larger R_g value at $pD 1.7$ is due to the expansion or oligomerization of Cyt c' . The SANS profile at $pD 1.7$ was evaluated using the Debye function (Figure 2A), used to analyze disordered polymer structures [40,41]. The SANS profile at $pD 1.7$ was fit well using the Debye function and gave $R_g = 25.7 \text{ \AA}$,

and clearly showed that the structure of Cyt *c'* transitioned to the unfolded random coil structure. The unfolded random coil structure of Cyt *c'* at pD 1.7 indicated by the SANS data is consistent with the structure previously proposed based on CD and ESI-MS spectrometry data [33].

Table 1. Estimated R_g values from experimental SANS curves

| Conditions | $R_g / \text{\AA}$ | | |
|--------------|--------------------|------------------|------------------|
| | Guinier | $P(r)$ | Debye |
| pD 1.7 | 23.02 \pm 0.57 | 25.03 \pm 0.03 | 25.65 \pm 0.15 |
| pD 6.4 | 18.93 \pm 0.45 | 18.24 \pm 0.08 | n/a |
| pD 9.6 | 19.24 \pm 0.49 | 18.10 \pm 0.05 | n/a |
| pD \sim 13 | 27.58 \pm 0.59 | 25.54 \pm 0.60 | n/a |

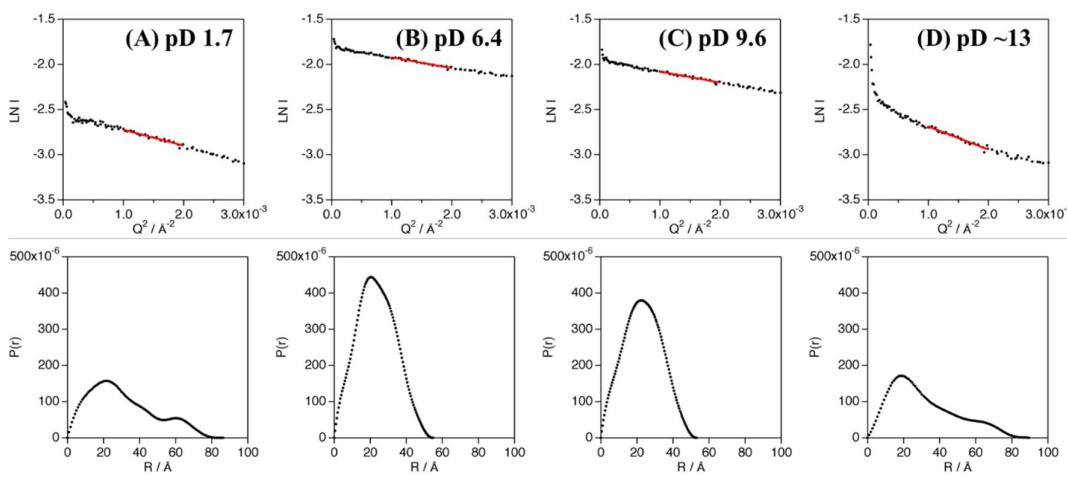


Figure 3. Guinier analysis (top panels) and $P(r)$ function (bottom panels) of Cyt *c'* at pD 1.7 (A), 6.4 (B), 9.6 (C), and \sim 13 (D). The linear regions of the Guinier plots used in the analyses are shown as red solid lines.

The crystallographic structure of Cyt *c'* from *Alcaligenes* was suggested to be a dimer, but the association state in solution was unknown [42]. The R_g values for the monomer and dimer structures of Cyt *c'* were calculated using the crystal structures of Cyt *c'* at pH 6.0 (PDB: 4WGZ) and pH 10.4 (PDB: 4WGY) and were 17-19 \AA (Table 2). Thus, the calculated R_g values are in good agreement with the experimentally obtained R_g values (Table 1), clearly demonstrating that the quaternary structure of Cyt *c'* is a dimer in solution at pD 6.4 and pD 9.6. The calculated SANS curves based on the crystal structure of dimeric Cyt *c'* at pH 6.0 and 10.4 are shown in Figure 2 and are quite similar to the experimental SANS curves at pD 6.4 and 9.6, especially for the range of $Q < 0.1 \text{ \AA}^{-1}$, which represents the global structure of Cyt *c'*. The similarity in the calculated and experimental SANS profiles strongly suggests that Cyt *c'* is a dimer in solution.

Table 2. Simulated R_g values calculated from crystal structures at pH 6.0 (PDB 4WGZ) and 11.4 (PDB 4WGY).

| Conditions | $R_g / \text{\AA}$ | |
|------------|--------------------|--------|
| | Guinier | $P(r)$ |
| pH 6.0 | Monomer | 14.0 |
| | Dimer | 17.5 |
| pH 10.4 | Monomer | 13.6 |
| | Dimer | 17.3 |

Fetler et al. reported that differences between experimental and simulated SAXS curves in the high-Q region of aspartate transcarbamoylase are due to quaternary structural differences in the solution and crystal structures [43–45]. In the present study, the discrepancy between the experimental and simulated curves of Cyt c' in the higher-Q region might similarly be due to quaternary structural differences in the dimeric structure in solution and in the crystal (Figure 2B, C). Kratky plots analysis of SANS profiles can distinguish flexible and rigid structures [46], with typical highly rigid folded states giving a bell-shaped curve in the low-Q region, and a disordered flexible structure giving a plateau shape in the high-Q region and lacking a bell-shaped curve in the low-Q region [47,48].

The Kratky plot of Cyt c' at pD 1.7 shows an intensified plateau region in the high-Q region, without a bell-shaped profile in the low-Q region (Figure 4A), strongly suggesting that Cyt c' becomes a very flexible random coil structure, in good agreement with the Debye function analysis and previous CD and ESI-MS experiments [33]. The Kratky plots of Cyt c' at pD 6.4 and 9.6 (Figure 4B and 4C) are clearly bell-shaped profiles centered at $Q = 0.1 \text{ \AA}^{-1}$ with a plateau shape in the higher-Q region ($Q > 0.2 \text{ \AA}^{-1}$), indicating a flexible moiety in the protein structure. The Kratky plot at pD \sim 13 shows a weaker, broader bell-shaped pattern in the low-Q region, suggesting a different size and/or shape of Cyt c' compared to at lower pD values.

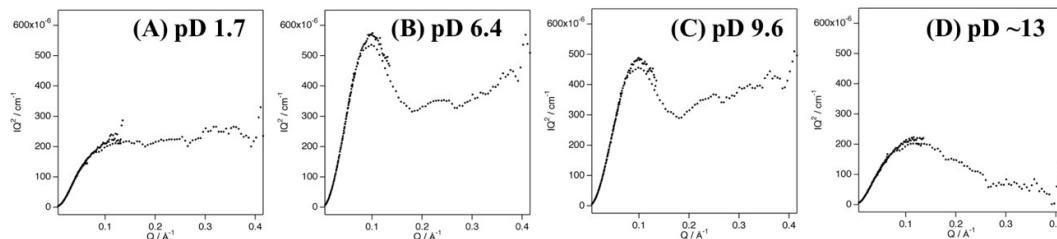


Figure 4. Kratky plots of the SANS curves of Cyt c' at pD 1.7 (A), 6.4 (B), 9.6 (C), and \sim 13 (D).

4. Discussion

In SANS curves, the extrapolated intercept intensity ($I(0)$) is proportional to the molecular weight (MW) and the weight concentration (C) of the scattering molecules.

The MW was determined according to equation 1:

$$MW = I(0)/C \frac{MW_{st}}{I(0)_{st}/C_{st}} \quad (1)$$

where MW_{st} , C_{st} , and $I(0)_{st}$ are the molecular weight, weight concentration, and extrapolated intensity of the standard sample, respectively [49]. The MW ratios at various pD values were calculated based on the $I(0)_{st}$ pD 1.7 value, used as the standard sample. The $I(0)$ values were obtained by extrapolation of the SANS $I(Q)$ curves (Fourier transformation of $P(r)$) and the Guinier plot (Figure S4). The calculated MW ratios of Cyt c' are summarized in Table 2. The MW ratios evaluated from Guinier analysis and $P(r)$ at pD 6.4 and 9.6 were twice that of the ratio at pD 1.7. Thus, the MW ratios also strongly support

the dimer structure of Cyt *c'* at pD 6.4 and 9.6. The $I(0)$ value estimated at pD \sim 13 is identical to the value at pD 1.7, further indicating that Cyt *c'* exists as a monomeric structure at pD 13.

Cyt *c'* has a four α -helix bundle structure. The structure at pH > 12 was suggested to be an “open-bundle” structure, which retains the α -helix structure [33]. The SANS curve at pD \sim 13 was numerically analyzed using a “joint-clubs model”, designed to describe the scattering patterns from four cylinder-shaped clubs connected by three short loops based on the zig-zag chain model described in earlier reports [50,51]. The “joint-clubs model” for analyzing the SANS curve at pD \sim 13 well reflected the “open-bundle” structure [33]. A schematic image of the joint-clubs model is shown in Figure 2D. Each α -helix of Cyt *c'* was modeled as a rigid club. Fitting using this joint-clubs model indicated that the length (L) of each club is 31.5 ± 3.1 Å. The Cyt *c'* structure consists of four α -helices: A(Ala3-Lys31), B(Asp37-Phe59), C(Ala76-Asp98), and D(Asp103-Arg124). The lengths of helices A, B, C, and D along each helix axis are 42.6, 35.4, 32.4, and 32.7 Å, respectively, as determined from the crystal structure (PDB ID: 4WGZ). The average length of the helices is 35.8 Å, which is in good agreement with the L value of 31.5 ± 3.1 Å calculated using the joint-clubs model. The average diameter of the clubs (R) was determined as 10.6 ± 0.9 Å by the fitting, in good agreement with an average diameter of the helices of \sim 10 Å in the crystal structure. Therefore, the joint-clubs model strongly supports the previously speculated “open-bundle” structure of Cyt *c'* at pD \sim 13.

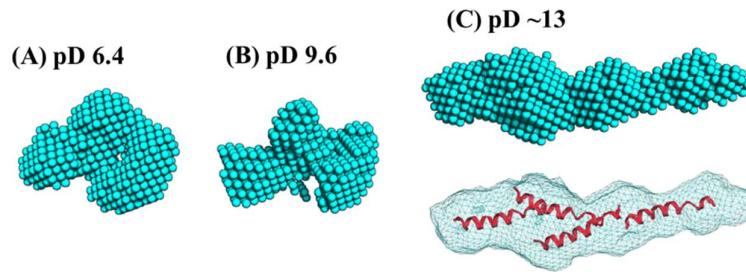


Figure 5. The *ab initio* bead models generated by the program DENFERT[27,28], and fitting of the crystal structure into each volumetric map (bottom panel) at pD 6.4 (A), 9.6 (B) and \sim 13 (C). The four B helices were docked in the volumetric map for (C) using the program Situs.[52] All models are drawn with the same scaling factor.

Ab initio analysis was conducted to clarify the low resolution solution structure of Cyt *c'*. Figure 5 shows the *ab initio* bead models drawn at the same scaling factor level to compare the size of the protein at different pD conditions. The bead models at pD 6.4 (Figure 5A) and 9.6 (Figure 5B) have essentially the same shape in the resolution of *ab initio* analysis. The bead model obtained at pD \sim 13 (Figure 5C) showed an extended structure. This elongation suggests two possibilities regarding the solution structure: the oligomerization of Cyt *c'*, or formation of the “open-bundle” structure. The first possibility can be excluded immediately because the molecular weight ratio (Table 3) reflects the monomeric state, as described above.

Table 3. Concentration (C) of Cyt *c'*, $I(0)$, and estimated molecular weight ratio from experimental SANS curves at each pD condition.

| Title 1 | C (mg/mL) | Guiner | | P(r) | |
|--------------|-----------|--------|----------|-------|----------|
| | | I(0) | MW ratio | I(0) | MW ratio |
| pD 1.7 | 5.5 | 0.078 | 1.0 | 0.080 | 1.0 |
| pD 6.4 | 5.5 | 0.165 | 2.1 | 0.161 | 2.0 |
| pD 9.6 | 5.3 | 0.141 | 1.9 | 0.139 | 1.8 |
| pD \sim 13 | 5.3 | 0.088 | 1.2 | 0.080 | 1.0 |

We also conducted docking simulation [52] of the helices to the elongated bead model structure at pD ~13 to clarify the suitability of the bead model to reflect an “open-bundle” structure of Cyt *c'* (Figure 5C, bottom). The volumetric map generated from the bead model at pD ~13 was fitted with four ideal helices. Docking with helices A, B, C, and D did not converge in the given space of the volumetric map and thus we chose helix B as a representative helix structure because its length is similar to the average length of the four helices in Cyt *c'*. The loop structures connecting each helix were omitted. Figure 5C shows a possible arrangement of the four helices in the “open-bundle” form of Cyt *c'* in the space of the volumetric map. This arrangement of the four helices is consistent with the joint-clubs model. The volumetric map at pD ~13 was also fitted to a monomer with the correlation coefficient 0.81 between the volumetric map and the volume calculated from the model (Appendix C “*Ab initio* bead modelling and analysis of bead model”). This volume is insufficient to accommodate the dimer structure. Therefore, the bead model for pD ~13 (Figure 5C) can reasonably be concluded to be the structure of the monomeric “open-bundle” structure with four α -helices, in contrast to the oligomerization of cytochrome *c* [53,54]. The oligomerization of cytochrome *c* is initiated by the displacement of the C-terminal helix domain to the corresponding position of other monomer [47]. The SANS experiment with Cyt *c'* at pD ~13 clearly demonstrated the monomeric “open-bundle” structure, and the fraction of aggregated particles was estimated to be only ~0.2% from the low-Q intensity (Appendix D “Estimation of aggregated particle fraction at pD ~13”). Therefore, Cyt *c'* does not undergo the intermolecular structural reassembly observed for cytochrome *c* [53].

5. Conclusions

The pH-induced structural transition mechanisms of Cyt *c'* were studied based on SANS solution structure determinations and were determined to be (i) random coil monomer at pD 1.7, (ii) folded dimer at pD 6.4, (iii) initial dimer dissociation at pD 9.6, and (iv) monomeric “open-bundle” structure at pD ~13. The comprehensive SANS analysis in the present study is consistent with previous spectroscopic studies (mass-spectrometry, CD/MCD spectroscopy) and precise crystal structure analyses. The relation between the pH-induced spin state and the structural transition of Cyt *c'* was described previously [33,55]. The present SANS study showed that the unfolding intermediate at pD ~13 is key for elucidating the structural transition mechanism. The alkaline transition structure at pD ~13 is the first folding/unfolding intermediate structure obtained for Cyt *c'* and has not been determined by crystal structure analysis. The “open-bundle” structure of Cyt *c'* as the unfolding/folding intermediate could provide insights into the initial or last step in the unfolding or folding process, respectively. Opening of the four α -helix bundle of Cyt *c'* into the intermediate “open-bundle” structure could be induced by the disappearance of inter-helix hydrogen-bonds at alkaline pH, reported previously [33].

Supplementary Materials: Figure S1: SDS-PAGE of purified Cyt *c'* on a 12.5% gel., Figure S2: Electronic absorption spectra of Cyt *c'* at pD 1.7, 6.4, 9.6, and ~13. The spectral patterns are annotated by charge-transfer 3 (CT3), Soret, CT2, Q and CT 1 at short wavelengths. The samples were prepared by diluting the samples following SANS experiments., Figure S3: Curves obtained by *ab initio* analyses (red lines) and data points used for each condition (squares: pD 6.4, circles: pD 9.6, triangles: pD ~13)., Figure S4: The estimation of $I(0)$ values by extrapolation of the $I(Q)$ SANS curves (top panels) and Guinier plots (bottom panels) at pD 1.7 (A, red), 6.4 (B, orange), 9.6 (C, green), and ~13 (D, blue). The black solid lines in the top panels are fitted $I(Q)$ SANS curves by Fourier transformation of $P(r)$ at each pD value, and the black dashed lines in the bottom panels were obtained from linear Guinier fitting. The $I(0)$ point at each pD value is marked by a diamond on the vertical axis.

Author Contributions: Conceptualization, T.K.; validation, T.Y., A.K., and H.F.; formal analysis, T.Y., A.K., and H.F.; investigation, T.Y. and H. F.; resources, T.Y., K.A., A.K., H.F. and T.K.; data curation, T.Y. and H.F.; writing—original draft preparation, T.Y.; writing—review and editing, T.Y., A.K., H. F. and T.K.; visualization, T.Y. and T.K. ; supervision, T.K.; project administration, T.K.;

funding acquisition, T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan under Contract Nos. 17-214, 18-190, and 19-198 to TK, and Grants-in-Aid for Scientific Research (Nos. 18550147 and 22550145 to TK) from the Japan Society for the Promotion of Science (JSPS).

Acknowledgments: Part of this study was performed under the approval of the Photon Factory Program Advisory Committee (Proposal Nos. 2018G018 and 2018G032). TY is grateful to the Early Researcher Encouragement Program of Ibaraki University. This work was performed using the KWS-1 instrument operated by the Jülich Centre for Neutron Science (JCNS) at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Joint-clubs model

The SANS curve at $pD \sim 13$ was analyzed using a numerical “joint-clubs” model, where the length of a club rod (L) and its diameter (R) correspond to the length and diameter of the α -helix, respectively. The other parameter, r , represents the radius of the blob scattering of the polymeric linkers within the Debye function $D(Q)$. The following functions (eq. A1-5) are involved in the model:

$$\Lambda(Q) = \frac{\text{Si}(QL)}{QL} \quad (\text{A1})$$

$$\Gamma_n(Q) = \frac{\sin(QL/n)}{\frac{QL}{n}} \quad (\text{A2})$$

$$B(Q) = 2 \frac{J_1(QR)}{QR} \quad (\text{A3})$$

$$D(Q) = \frac{2}{(Q^2 r^2)^2} (\exp(-Q^2 r^2) - 1 + Q^2 r^2) \quad (\text{A4})$$

$$I(Q) = I_0 \left[\left(\frac{1}{4} (2\Lambda(Q) - \Gamma_2^2(Q)) + \frac{2}{16} \frac{\Lambda^2(Q)}{1 - \Gamma_1(Q)} \left(4 - \frac{1 - \Gamma_1^4(Q)}{1 - \Gamma_1(Q)} \right) \right] B^2(Q) + I_1 D(Q) \quad (\text{A5})$$

Here, $\text{Si}(x)$ is the sine integral and $J_1(x)$ is the Bessel function of the first order. The prefactors I_0 and I_1 determine the scattering powers of all the helices and the linkers, respectively. In equation (eq. A5), the function in the large square brackets describes the zig-zag line scattering, while $B^2(Q)$ describes the spherical cross-section of the cylinder.[50,51] The total scattering intensity is given by the product of the single-particle scattering, $I(Q)$, and a simple structure factor [56], $S(Q)$ (eq. A6) emerging from single particles and aggregates, respectively:

$$S(Q) = \left(1 + \frac{S_{\text{agg}}(0)}{1 + \xi^2 Q^2} \right) \quad (\text{A6})$$

where $S_{\text{agg}}(0)$ is the dimensionless amplitude of the aggregates and ξ is a correlation length in the sense of van Hove [57], which represents the size of an aggregated particle. The contribution to scattering from aggregates was estimated to be $\sim 0.2\%$.

Appendix B. Molecular weight estimation

The estimated $I(0)$ values and the concentrations of Cyt c' were used to determine the ratio of the molecular weight, MW, to the molecular weight at pD 1.7 using equation S7:

$$Ratio = \frac{MW}{MW_{pD\ 1.7}} = \frac{I(0)/C}{I(0)_{pD\ 1.7}/C_{pD\ 1.7}} \quad (A7)$$

where C is the weight concentration. The $I(0)$ values at $Q = 0 \text{ \AA}^{-1}$ were estimated by extrapolation of the intercept intensities on the Guinier plots and $P(r)$ functions.

Appendix C. *Ab initio* bead modelling and analysis of the bead model

An indirect transform using the program GNOM [58] was performed to obtain the regularized scattering curves for the data where the protein retains its secondary structure (pD 6.4, 9.6, and ~ 13). The regularized data were used as input to the program DENFERT v.2 [27,28], which restores the low-resolution shape of the protein by considering the contribution of the hydration layer to the measured scattering. A 10% higher scattering length density for the hydration layer was applied. The protein scattering length density was set to $3.1 \times 10^{-6} \text{ \AA}^{-2}$ and the buffer scattering length density was set to $6.4 \times 10^{-6} \text{ \AA}^{-2}$. Ten independent runs for each pD condition were performed and compared using the program DAMAVER [59]. Those with the lowest normalized discrepancy (NSD; a measure of quantitative similarity among sets of three-dimensional points) were chosen as the typical model. The average NSD in all cases was in the range of 0.6-0.7, indicating solution stability, consistent with the experimental data. Figure S3 shows the curves resulting from the *ab initio* analyses for pD 6.4 and 9.6 and the joint-clubs model for pD ~ 13 with the experimental SANS curves. The square root values of the reduced χ^2 were 1.8, 2.9, and 1.9 for pD 6.4, 9.6, and ~ 13 , respectively. The bead radius in the *ab initio* analysis was set to 2 \AA . The models constructed for the pD ~ 13 condition was further converted to volumetric maps using the pdb2vol program, and the maps was used for docking analyses. The docking of models in map were performed using Powell's optimization program Collage in the Situs v.2.8 package [52]. The docking analysis at pH ~ 13 was carried out with four identical helices (Helix B), and the correlation coefficient for full multi-body docking was 0.81.

Appendix D. Estimation of the aggregated particle fraction at pD ~ 13

Forward scattering, $S(0)$, is generally expressed by equation A8:

$$S(Q \rightarrow 0) = (\Delta\varrho)^2 \cdot \phi \cdot V \quad (A8)$$

where $(\Delta\varrho)^2$ is the contrast of the scattering particle, ϕ is the volume fraction of the particles, and V is the volume of the particle. The ratio of forward scattering of the aggregate and a joint-club, $S_{agg}(0)/S_{jc}(0)$, is described by the volume fraction of Cyt c' in the beam, $\phi_{Cyt\ c'}$, the fraction of aggregated particles, ε , the volume of aggregated particles, V_{agg} , and the volume of the joint-clubs structure, V_{jc} (eq. A9).

$$\frac{S_{agg}(0)}{S_{jc}(0)} = \frac{(\Delta\varrho)^2 \cdot \phi_{Cyt\ c'} \cdot \varepsilon \cdot V_{agg}}{(\Delta\varrho)^2 \cdot \phi_{Cyt\ c'} \cdot (1 - \varepsilon) \cdot V_{jc}} \quad (A9)$$

The structure factor, $S(Q)$, is described by equation S6[56] The experimental data at pD ≈ 13 were fitted to the equation of the product of $S(Q)$ and the scattering function of the joint-clubs model defined by equations S1-5. Curve-fitting in Figure 2(D) gave the parameters $S_{agg}(0)/S_{jc}(0) = 9.2$ and $\xi = 432 \pm 64 \text{ \AA}$. The ratio of the volume between the joint-clubs structure and aggregated particles, V_{jc}/V_{agg} (eq. A10), was estimated:

$$\frac{V_{jc}}{V_{agg}} \approx \left(\frac{R_g}{\xi} \right)^3 \quad (A10)$$

where R_g is the radius of gyration at $pD \approx 13$ (ca. 25–28 Å, Table 1). The fraction of aggregated particles, ε , was estimated as ~0.2% by equation A9.

References

1. Anfinsen, C.B.; Haber, E.; Sela, M.; White, F.H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309–1314.
2. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230.
3. Ohgushi, M.; Wada, A. “Molten-globule state”: a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **1983**, *164*, 21–24, doi:10.1016/0014-5793(83)80010-6.
4. Kataoka, M.; Nishii, I.; Fujisawa, T.; Ueki, T.; Tokunaga, F.; Goto, Y. Structural characterization of the molten globule and native states of apomyoglobin by solution x-ray scattering. *J. Mol. Biol.* **1995**, *249*, 215–228, doi:10.1006/jmbi.1995.0290.
5. Bhuyan, A.K. Off-pathway status for the alkali molten globule of horse ferricytochrome c. *Biochemistry* **2010**, *49*, 7764–7773, doi:10.1021/bi100880d.
6. Arai, M.; Kuwajima, K. Role of the molten globule state in protein folding. *Adv. Protein Chem.* **2000**, *53*, 209–282.
7. Hamada, D.; Hoshino, M.; Kataoka, M.; Goto, Y.; Fink, A.L. Intermediate Conformational States of Apocytochrome c. *Biochemistry* **1993**, *32*, 10351–10358, doi:10.1021/bi00090a010.
8. Kieflhaber, T. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9029–9033, doi:10.1073/pnas.92.20.9029.
9. Elöve, G.A.; Bhuya, A.K.; Roder, H. Kinetic Mechanism of Cytochrome c Folding: Involvement of the Heme and Its Ligands. *Biochemistry* **1994**, *33*, 6925–6935, doi:10.1021/bi00188a023.
10. Kamatari, Y.O.; Konno, T.; Kataoka, M.; Akasaka, K. The methanol-induced transition and the expanded helical conformation in hen lysozyme. *Protein Sci.* **1998**, *7*, 681–688, doi:10.1002/pro.5560070317.
11. Hamill, A.C.; Wang, S.C.; Lee, C.T. Probing lysozyme conformation with light reveals a new folding intermediate. *Biochemistry* **2005**, *44*, 15139–15149, doi:10.1021/bi051646c.
12. Jamin, M.; Baldwin, R.L. Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin. *Nat. Struct. Biol.* **1996**, *3*, 613–618, doi:10.1038/nsb0796-613.
13. Kitahara, R.; Akasaka, K. Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3167–3172, doi:10.1073/pnas.0630309100.
14. Goto, Y.; Hagihara, Y.; Hamada, D.; Hoshino, M.; Nishii, I. Acid-Induced Unfolding and Refolding Transitions of Cytochrome c: A Three-State Mechanism in H₂O and D₂O. *Biochemistry* **1993**, *32*, 11878–11885, doi:10.1021/bi00095a017.
15. Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. Evidence for Identity between the Equilibrium Unfolding Intermediate and a Transient Folding Intermediate: A Comparative Study of the Folding Reactions of α -Lactalbumin and Lysozyme. *Biochemistry* **1986**, *25*, 6965–6972, doi:10.1021/bi00370a034.
16. Chen, L.; Hodgson, K.O.; Doniach, S. A lysozyme folding intermediate revealed by solution X-ray scattering. *J. Mol. Biol.* **1996**, *261*, 658–671, doi:10.1006/jmbi.1996.0491.
17. Krishna, M.M.G.; Maity, H.; Rumbley, J.N.; Lin, Y.; Englander, S.W. Order of Steps in the Cytochrome c Folding Pathway: Evidence for a Sequential Stabilization Mechanism. *J. Mol. Biol.* **2006**, *359*, 1410–1419, doi:10.1016/j.jmb.2006.04.035.
18. Hu, W.; Walters, B.T.; Kan, Z.Y.; Mayne, L.; Rosen, L.E.; Marqusee, S.; Englander, S.W. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 7684–7689, doi:10.1073/pnas.1305887110.
19. Hoang, L.; Bedard, S.; Krishna, M.M.G.; Lin, Y.; Englander, S.W. Cytochrome c folding pathway: Kinetic native-state hydrogen exchange. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12173–12178, doi:10.1073/pnas.152439199.
20. Freddolino, P.L.; Harrison, C.B.; Liu, Y.; Schulter, K. Challenges in protein-folding simulations. *Nat. Phys.* **2010**, *6*, 751–758, doi:10.1038/nphys1713.
21. Gilquin, B.; Guilbert, C.; Perahia, D. Unfolding of hen egg lysozyme by molecular dynamics simulations at 300K: Insight into the role of the interdomain interface. *Proteins Struct. Funct. Genet.* **2000**, *41*, 58–74, doi:10.1002/1097-0134(20001001)41:1<58::AID-PROT90>3.0.CO;2-3.
22. Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw, D.E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520, doi:10.1126/science.1208351.
23. Svergun, D.I.; Koch, M.H.J. Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol.* **2002**, *12*, 654–660.
24. Lattman, E.E. Small angle scattering studies of protein folding. *Curr. Opin. Struct. Biol.* **1994**, *4*, 87–92, doi:10.1016/S0959-440X(94)90064-7.
25. Akiyama, S.; Takahashi, S.; Kimura, T.; Ishimori, K.; Morishima, I.; Nishikawa, Y.; Fujisawa, T. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1329–34, doi:10.1073/pnas.012458999.
26. Watanabe, Y.; Inoko, Y. Size-exclusion chromatography combined with small-angle X-ray scattering optics. *J. Chromatogr. A* **2009**, *1216*, 7461–7465, doi:10.1016/j.chroma.2009.02.053.

27. Koutsoubas, A.; Pérez, J. Incorporation of a hydration layer in the 'dummy atom' ab initio structural modelling of biological macromolecules. *J. Appl. Crystallogr.* **2013**, *46*, 1884–1888, doi:10.1107/S0021889813025387.
28. Koutsoubas, A.; Jaksch, S.; Pérez, J. DENFERT version 2: extension of ab initio structural modelling of hydrated biomolecules to the case of small-angle neutron scattering data. *J. Appl. Crystallogr.* **2016**, *49*, 690–695, doi:10.1107/S1600576716003393.
29. Maltempo, M.M.; Moss, T.H. The spin 3/2 state and quantum spin mixtures in haem proteins. *Q. Rev. Biophys.* **1976**, *9*, 181–215, doi:10.1017/S0033583500002407.
30. Maltempo, M.M.; Moss, T.H.; Cusanovich, M.A. Magnetic studies on the changes in the iron environment in Chromatium ferricytochrome c'. *BBA - Protein Struct.* **1974**, *342*, 290–305, doi:10.1016/0005-2795(74)90084-1.
31. Yoshimura, T.; Suzuki, S.; Nakahara, A.; Iwasaki, H.; Masuko, M.; Matsubara, T. Spectral Properties of Nitric Oxide Complexes of Cytochrome c' from Alcaligenes sp. NCIB 11015. *Biochemistry* **1986**, *25*, 2436–2442, doi:10.1021/bi00357a021.
32. Weiss, R.; Gold, A.; Termer, J. Cytochromes c': Biological models for the S = 3/2,5/2 spin-state admixture? *Chem. Rev.* **2006**, *106*, 2550–2579.
33. Takashina, A.; Tiedemann, M.T.; Unno, M.; Yamaguchi, T.; Stillman, M.J.; Kohzuma, T. The pH Dependent Protein Structure Transitions and Related Spin-State Transition of Cytochrome c' from Alcaligenes xylosoxidans NCIMB 11015. *Bull. Chem. Soc. Jpn.* **2017**, *90*, 169–177, doi:10.1246/bcsj.20160316.
34. Alquraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35*, 4862–4865, doi:10.1093/bioinformatics/btz422.
35. Covington, A.K.; Paabo, M.; Robinson, R.A.; Bates, R.G. Use of the Glass Electrode in Deuterium Oxide and the Relation between the Standardized pD (paD) Scale and the Operational pH in Heavy Water. *Anal. Chem.* **1968**, *40*, 700–706, doi:10.1021/ac60260a013.
36. Feoktystov, A. V.; Frielingshaus, H.; Di, Z.; Jaksch, S.; Pipich, V.; Appavou, M.-S.; Babcock, E.; Hanslik, R.; Engels, R.; Kemmerling, G.; et al. KWS-1 high-resolution small-angle neutron scattering instrument at JCNS: current state. *J. Appl. Crystallogr.* **2015**, *48*, 61–70, doi:10.1107/S1600576714025977.
37. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M.H.J.; Svergun, D.I. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2003**, *36*, 1277–1282, doi:10.1107/S0021889803012779.
38. Debye, P. Molecular-Weight Determination by Light Scattering. *J. Phys. Colloid Chem.* **1946**, *51*, 18–32, doi:10.1021/j150451a002.
39. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503, doi:10.1107/S0021889892001663.
40. Mortensen, K.; Talmon, Y. Cryo-TEM and SANS Microstructural Study of Pluronic Polymer Solutions. *Macromolecules* **1995**, *28*, 8829–8834, doi:10.1021/ma00130a016.
41. Beaucage, G. Small-Angle Scattering from Polymeric Mass Fractals of Arbitrary Mass-Fractal Dimension. *J. Appl. Crystallogr.* **1996**, *29*, 134–146, doi:10.1107/S0021889895011605.
42. Dobbs, A.J.; Anderson, B.F.; Faber, H.R.; Baker, E.N. Three-dimensional structure of cytochrome c' from two Alcaligenes species and the implications for four-helix bundle structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1996**, *52*, 356–368, doi:10.1107/S0907444995008328.
43. Svergun, D.I.; Barberato, C.; Koch, M.H.J.; Fetler, L.; Vachette, P. Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins* **1997**, *27*, 110–117, doi:10.1002/(SICI)1097-0134(199701)27:1<110::AID-PROT11>3.0.CO;2-Q.
44. Fetler, L.; Vachette, P. The allosteric activator Mg-ATP modifies the quaternary structure of the R-state of Escherichia coli aspartate transcarbamylase without altering the T \leftrightarrow R equilibrium. *J. Mol. Biol.* **2001**, *309*, 817–832, doi:10.1006/jmbi.2001.4681.
45. Meisburger, S.P.; Thomas, W.C.; Watkins, M.B.; Ando, N. X-ray Scattering Studies of Protein Structural Dynamics. *Chem. Rev.* **2017**, *117*, 7615–7672, doi:10.1021/acs.chemrev.6b00790.
46. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577.
47. Doniach, S. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.* **2001**, *101*, 1763–1778.
48. Burger, V.M.; Arenas, D.J.; Stultz, C.M. A Structure-free Method for Quantifying Conformational Flexibility in proteins. *Sci. Rep.* **2016**, *6*, 29040.
49. Mylonas, E.; Svergun, D.I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **2007**, *40*, s245–s249, doi:10.1107/S002188980700252X.
50. Hermans, J.; Hermans, J.J. Light Scattering by Zig-zag and Worm-like Chains. *J. Phys. Chem.* **1958**, *62*, 1543–1546, doi:10.1021/j150570a018.
51. Lages, S.; Goerigk, G.; Huber, K. SAXS and ASAXS on Dilute Sodium Polyacrylate Chains Decorated with Lead Ions. *Macromolecules* **2013**, *46*, 3570–3580, doi:10.1021/ma400427d.
52. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 344–351, doi:10.1107/S0907444911049791.
53. Hirota, S.; Hattori, Y.; Nagao, S.; Taketa, M.; Komori, H.; Kamikubo, H.; Wang, Z.; Takahashi, I.; Negi, S.; Sugiura, Y.; et al. Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12854–12859, doi:10.1073/pnas.1001839107.
54. Margoliash, E.; Lustgarten, J. Interconversion of Horse Heart Cytochrome c Monomer and Polymers. *J. Biol. Chem.* **1962**, *237*, 3397–3405.

55. Takashina, A.; Unno, M.; Kohzuma, T. X-ray Crystallographic Elucidation for the Alkaline High-spin State Transition of Iron (III) Cytochrome c' from Alcaligenes xylosoxidans NCIMB 11015. *Chem. Lett.* **2015**, *44*, 268–270.
56. Xu, B.; Lynn, G.W.; Guo, J.; Melnichenko, Y.B.; Wignall, G.D.; McClain, J.B.; DeSimone, J.M.; Johnson, C.S. NMR and SANS Studies of Aggregation and Microemulsion Formation by Phosphorus Fluorosurfactants in Liquid and Supercritical Carbon Dioxide. *J. Phys. Chem. B* **2005**, *109*, 10261–10269, doi:10.1021/jp0580322.
57. Van Hove, L. Correlations in space and time and born approximation scattering in systems of interacting particles. *Phys. Rev.* **1954**, *95*, 249–262, doi:10.1103/PhysRev.95.249.
58. Svergun, D.I.; Richard, S.; Koch, M.H.J.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267–2272.
59. Volkov, V. V.; Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2003**, *36*, 860–864, doi:10.1107/S0021889803000268.