

Conformations of Variably Linked Chimeric Proteins Evaluated by Synchrotron X-ray Small-Angle Scattering

Ryoichi Arai,^{1,2} Willy Wriggers,³ Yukihiro Nishikawa,² Teruyuki Nagamune,¹ and Tetsuro Fujisawa^{2*}

¹Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, Tokyo, Japan

²Laboratory for Structural Biochemistry, RIKEN Harima Institute/Spring-8, Hyogo, Japan

³School of Health Information Sciences and Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas

ABSTRACT We constructed chimeric proteins that consist of two green fluorescent protein variants, EBFP and EGFP, connected by flexible linkers, (GGGS)_n ($n = 3\sim 4$), and helical linkers, (EAAK)_n ($n = 2\sim 5$). The conformations of the chimeric proteins with the various linkers were evaluated using small-angle X-ray scattering (SAXS). The SAXS experiments showed that introducing the short helical linkers ($n = 2\sim 3$) causes multimerization, while the longer linkers ($n = 4\sim 5$) solvate monomeric chimeric proteins. With the moderate-length linkers ($n = 4$), the observed radius of gyration (R_g) and maximum dimension (D_{max}) were 38.8 Å and 120 Å with the flexible linker, and 40.2 Å and 130 Å with the helical linker, respectively. The chimeric protein with the helical linker assumed a more elongated conformation as compared to that with the flexible linker. When the length of the helical linker increased ($n = 5$), R_g and D_{max} increased to 43.2 Å and 140 Å, respectively. These results suggest that the longer helix effectively separates the two domains of the chimeric protein. Considering the connectivity of the backbone peptide of the protein, the helical linker seems to connect the two domains diagonally. Surprisingly, the chimeric proteins with the flexible linker exhibited an elongated conformation, rather than the most compact side-by-side conformation expected from the fluorescence resonance energy transfer (FRET) analysis. Furthermore, the SAXS analyses suggest that destabilization of the short helical linker causes multimerization of the chimeric proteins. Information about the global conformation of the chimeric protein is thus be necessary for optimization of the linker design. *Proteins* 2004;57:829–838.

© 2004 Wiley-Liss, Inc.

Key words: small-angle X-ray scattering (SAXS); green fluorescent protein; fusion protein; protein engineering; linker engineering; helical linker; *Situs*; modeling

INTRODUCTION

Gene fusion techniques have emerged as an indispensable tool in a variety of biochemical research areas. The construction of a recombinant chimeric/fusion protein is a standard method used to increase the expression of soluble

proteins and to facilitate protein purification. Furthermore, various applications of gene fusion techniques in the field of biotechnology have been reported. These include immunoassays using chimeras between antibody fragments and green fluorescent protein variants,^{1,2} selection and production of antibodies³ and engineering of bifunctional enzymes.⁴

Chimeric protein construction involves linking two proteins or domains of proteins by a peptide linker. The selection of the linker sequence is particularly important for the construction of functional chimeric proteins. The results of several linker selection studies^{5–8} have suggested that the flexibility and hydrophilicity of the linker are important factors in preventing the disturbance of the domain functions. However, a study on the streptococcal protein *G-Vargula* luciferase chimera suggested that the spatial separation of the hetero-functional domains of a

Abbreviations: a.a., amino acids; BCA, bicinchoninic acid; BFP, blue fluorescent protein; B-H2-G, EBFP-H2-EGFP; B-H3-G, EBFP-H3-EGFP; B-H4-G, EBFP-H4-EGFP; B-H5-G, EBFP-H5-EGFP; B-F3-G, EBFP-F3-EGFP; B-F4-G, EBFP-F4-EGFP; CCD, charge coupled device; CD, circular dichroism; D_{max} , maximum dimension of the particle; EBFP, enhanced blue fluorescent protein; EGFP, enhanced green fluorescent protein; F3, flexible linker 3; F4, flexible linker 4; FRET, fluorescence resonance energy transfer; GFP, green fluorescent protein; H2, helical linker 2; H3, helical linker 3; H4, helical linker 4; H5, helical linker 5; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; R_g , radius of gyration; SAXS, small-angle X-ray scattering

The Supplementary Materials Referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html>

Grant sponsors: NIH (1R01GM62968), Alfred P. Sloan Foundation (BR-4297) and Human Frontier Science Program (RGP0026/2003) to W.W.; Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government to T.F.; a grant-in aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan and Biodesign Research Promotion Group, RIKEN to T.N.; and Research fellowships of the Japan Society for the Promotion of Science for young scientists to R.A.

*Correspondence to: Tetsuro Fujisawa, Laboratory for Structural Biochemistry, RIKEN Harima Institute/Spring-8, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan. E-mail: fujisawa@spring8.or.jp

R. Arai's present address is RIKEN Genomic Sciences Center, Yokohama 230-0045 Japan.

Received 17 December 2003; Revised 25 April 2004, 19 May 2004, 20 May 2004

Published online 28 July 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20244

chimeric protein by an appropriate linker peptide is important for the domains to work independently.⁹

In our previous study, we designed linkers to effectively separate the two domains of a chimeric protein.¹⁰ We introduced helix-forming peptide linkers, (EAAAK)_n, between two green fluorescent protein variants, enhanced blue fluorescent protein (EBFP; F64L, S65T, Y66H, Y145F)¹¹ and enhanced green fluorescent protein (EGFP; F64L, S65T).¹² The circular dichroism (CD) spectroscopic analysis suggested that the introduced linkers form an α -helix, and that the α -helical contents increase as the lengths of the linkers increase. The fluorescence resonance energy transfer (FRET) analysis from EBFP to EGFP also suggested that the distance between the two domains increases as the lengths of the linkers increase. However, it is difficult to determine the distance and orientation of the domains and the linker of the chimeric proteins by FRET and CD analyses. Therefore, in this study, we used synchrotron X-ray small-angle scattering to analyze the conformations of chimeric proteins. The shapes and sizes of the chimeric proteins, consisting of EBFP and EGFP with the helical linkers (EAAAK)_n ($n = 4, 5$) and flexible linkers (GGGGS)_n ($n = 3, 4$), were deduced from the small-angle X-ray scattering (SAXS) pattern with an *ab initio* modeling procedure.

MATERIALS AND METHODS

Preparation of Chimeric Proteins

The chimeric proteins were prepared as has been previously described.¹⁰ In brief, we constructed chimeric proteins between EBFP and EGFP (Clontech, Palo Alto, CA) with designed linkers. Helical linkers were designed on the basis of a helix-forming peptide, [($i + 4$)E,K], which was described by Marqusee and Baldwin.¹³ The helix-forming peptide, A(EAAAK)₃A, forms a monomeric α -helix with the support of a Glu-Lys salt bridge. Therefore, we designed helical linkers consisting of the (EAAAK)_n motif ($n = 2\sim 5$). On the other hand, the flexible linkers consisted of the (GGGGS)_n motif ($n = 3\sim 4$). The amino acid sequences of the linkers are as follows: helical linker 2 (H2), LAEAAAKEAAAKAAA (15 a.a.); helical linker 3 (H3), LAEAAAKEAAAKEAAAKAAA (20 a.a.); helical linker 4 (H4), LAEAAAKEAAAKEAAAKEAAAKAAA (25 a.a.); helical linker 5 (H5), LAEAAAKEAAAKEAAAKEAAAKEAAAKAAA (30 a.a.); flexible linker 3 (F3), LGGGGSGGGGS-GGGGSAAA (19 a.a.); flexible linker 4 (F4), LSGGGGS-GGGSGGGSGGGGSAAA (25 a.a.) (L and AAA on the termini are derived from restriction enzyme sites). We expressed the chimeric proteins with thioredoxin, an S-tag and a His-tag, using the pET TRX Fusion System 32 (Novagen, Madison, WI) in *Escherichia coli* AD494(DE3) pLysS with the expression vectors pET32/B-H2-G, pET32/B-H3-G, pET32/B-H4-G, pET32/B-H5-G, pET32/B-F3-G and pET32/B-F4-G. Chimeric proteins with the His-tag were purified using Talon metal affinity resin (Clontech). The proteins were specifically digested with thrombin to remove the thioredoxin, and were further purified by size exclusion chromatography with Superdex 75 (Amersham Biosciences, Uppsala, Sweden). The protein concentration

was determined using a bicinchoninic acid (BCA) assay kit (Pierce, Rockford, IL) with bovine serum albumin (BSA) as the standard.

SAXS Measurements and Data Analyses

SAXS measurements were carried out using RIKEN structural biology beamline I (BL45XU),¹⁴ which employs a 1.0 Å wavelength X-ray from an undulator source of the electron storage ring at SPring-8. With a detector consisting of an X-ray image intensifier and a cooled charge coupled device (CCD) (XR-II+CCD),¹⁵ each scattering profile was collected at 25°C for 2 s. The sample-to-detector distance was 974 mm. Judging from the stability of intensity over time, the proteins suffered negligible radiation damage during the data collection. Preliminary data processing was performed using the program *iisgnapr*.¹⁵ Two-dimensional sample and buffer images were scaled with incident intensity, circular-averaged and then subtracted. The reciprocal parameter S , which is equal to $2 \sin \theta / \lambda$ (where 2θ is the scattering angle and λ is the X-ray wavelength), was calibrated by meridional reflections from chicken collagen. The radius of gyration, R_g , was determined by fitting the intensity profiles using the Guinier approximation: $I(S) = I(0) \exp(-4\pi^2 R_g^2 S^2 / 3)$, where $I(0)$ is the forward scattering intensity at a zero angle¹⁶ with a fitting region of S^2 (Å⁻²) from 10×10^{-6} to 30×10^{-6} . In order to eliminate inter-particle interference, measurements were taken at seven different protein concentrations, from 0.4 to 1.6 mg/mL, and these data points were extrapolated to a zero protein concentration. The pair distance distribution functions, $P(r)$, were calculated using the indirect transform package GNOM.¹⁷ The procedure for the determination of the maximum dimension of a particle, D_{\max} , is described elsewhere.¹⁸ The scattering region used was $0.00315 \text{ \AA}^{-1} < S < 0.05 \text{ \AA}^{-1}$. The accuracy of the fit R_f was evaluated using the following equation:

$$R_f = \frac{\sum |I_{\text{exp}} - I_{\text{cal}}|}{\sum |I_{\text{exp}}|}$$

where I_{exp} and I_{cal} are experimental intensity with homogeneous density correction¹⁹ and calculated intensity, respectively. The intensities were calculated using the Debye formula and crysol program²⁰ for bead models and high-resolution models, respectively.

Calculation of *Ab Initio* Models of Chimeric Proteins

Calculation of *ab initio* models was completed using the program DAMMIN.²¹ In dummy atom minimization (DAMMIN), a protein molecule is approximated by densely packed small spheres (dummy atoms). Minimization was performed using the simulated annealing method, starting from the dummy atoms placed at random coordinates within the search space, a sphere of diameter D_{\max} .¹⁷ More than ten independent bead models were calculated, aligned with each other, and then superimposed.²² The superimposed bead models were filtered against the volume of each run. The simulated scattering profiles, both from each run and from the filtered beads, usually fit the experimental data well²³ ($R_f = 0.00044$ of each run for B-H4-G). How-

ever, we observed a substantial deviation from the experimental data after the standard filtering procedure ($R_f = 0.013$ for B-H4-G). Therefore, we modified the filtering procedure as follows: First, we cut off the superimposed bead models with a larger volume than the average (approx. 10%). Then, the filtered beads were taken as an initial search model and were minimized again with simulated annealing. The resulting bead model exhibited good shape similarity, with a normalized standard deviation (NSD)²² lower than 0.4, and its scattering fit well to the experimental curve ($R_f = 0.0026$ for B-H4-G). Since EBFP and EGFP have 99% amino acid sequence identity and almost identical structures at low resolution, we calculated the bead models under the constraint of P2 symmetry.

High-Resolution Modeling of Chimeric Proteins with Helical Linkers (B-H4-G and B-H5-G)

Model volumetric structures were constructed from the bead models, using the Situs program package for the registration of the protein structures with low-resolution bead models from X-ray scattering, as described previously.^{24,25} The bead model atoms were each convoluted with a Gaussian kernel, with a half-max radius corresponding to the bead radius, using the program pdblur. The volumetric densities were later visualized at the half-max isocontouring threshold, which provided a molecular envelope of the bead model. To achieve the best fits into either lobe of the dumbbell-shaped density, we cropped the density so that it contained only one part of a lobe and the linker. In this way, the Colores fitting program distributed with Situs returned best fits specific to one lobe.²⁴ Using Colores, we performed the fitting of PDB entry 1BFP (a crystal structure of BFP)²⁶ to each lobe in the density, using standard volumetric cross correlation as a criterion. We assumed that both structures of EBFP and EGFP were coincident with that of 1BFP at low resolution, due to the approximately 99% sequence homology among EBFP, EGFP and BFP. We selected the best fit structure with the C- and N-terminal regions oriented towards the adjacent BFP, and used this to create our model. The models with the best three scores determined by the Colores fitting program coincided with each other: The root mean square deviation (RMSD) of the centers of these models was found to be less than 1.2 Å. The deviations in the orientations of the long and short axes of each lobe are within 6° and 30°, respectively.

We created α -helices as the initial models for the linker peptides, because the results of our previous CD analysis indicated that the linkers formed α -helical structures.¹⁰ For simplicity, the S-tag and the His-tag were omitted in the model. We then assembled everything into a model that starts at the EBFP of the first domain and ends at the EGFP of the second domain. While the two domains were fitted computationally, the position and orientation of the linker were fine-tuned manually with VMD²⁷ to ensure that it would originate at the C-terminus of the first-domain EBFP and end at the N-terminus of the second-domain EGFP. The model structures were converted into

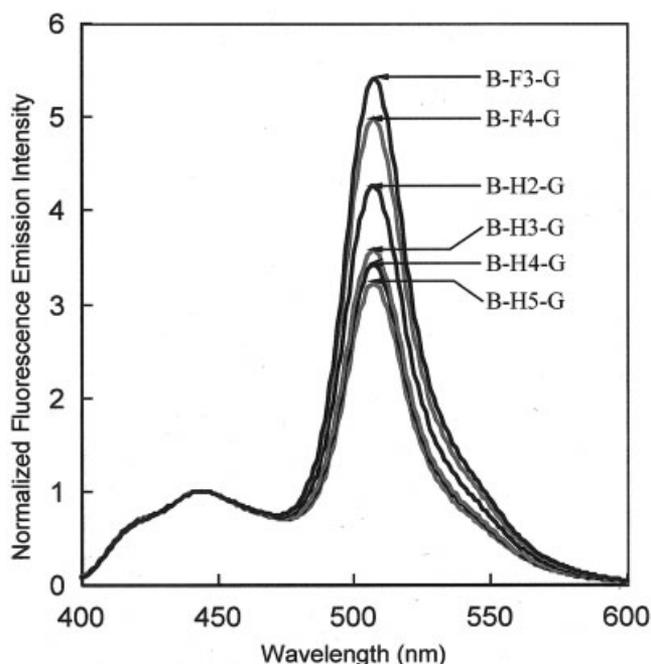


Fig. 1. Fluorescence emission spectra of the chimeric proteins. The fluorescence spectra were measured with 380 nm excitation at 25°C, using a fluorescence spectrophotometer F-2000 (Hitachi, Tokyo, Japan). The spectra were normalized at 444 nm to the emission peak of EBFP. The normalized peak value of EGFP emission at 507 nm, $I(507 \text{ nm})/I(444 \text{ nm})$, was taken as the index of FRET efficiency. For the measurements, a 2 μM concentration of each sample in phosphate-buffered saline (PBS) pH 7.4 was used.

the X-PLOR²⁸ format and were stereochemically optimized with a simple energy-minimization run. We used a force field (CHARMM19) with a distance-dependent dielectric constant, because this approach is suitable for molecular modeling in the absence of a solvent. The R_f values of these high-resolution models were 0.06 and 0.07 for B-H4-G and B-H5-G, respectively. They are far inferior to those of low-resolution models. The origin of this fitting discrepancy will be discussed later.

RESULTS

Fluorescence Spectra of Chimeric Proteins and FRET Analysis

The two GFP variants, EBFP and EGFP, have been widely used for FRET analyses.^{2,10,29,30} Figure 1 shows the fluorescence emission spectra of the chimeric proteins, normalized at 444 nm of the emission peak of EBFP. The chimeric proteins retained sufficient fluorescence activity derived from EBFP and EGFP. The observed emission peaks at 444 nm and 507 nm were found to be coincident with the emission peaks of the non-fusion EBFP and EGFP, respectively, indicating that the EBFP and EGFP domains of the chimeric proteins formed the proper structures. The peak intensity of EGFP emission at 507 nm [$I(507 \text{ nm})/I(444 \text{ nm})$] was used as the index of FRET efficiency. In the case of the chimeric proteins with helical linkers (H2, H3, H4, H5), the FRET indexes decreased as the lengths of the linkers increased. A higher FRET index

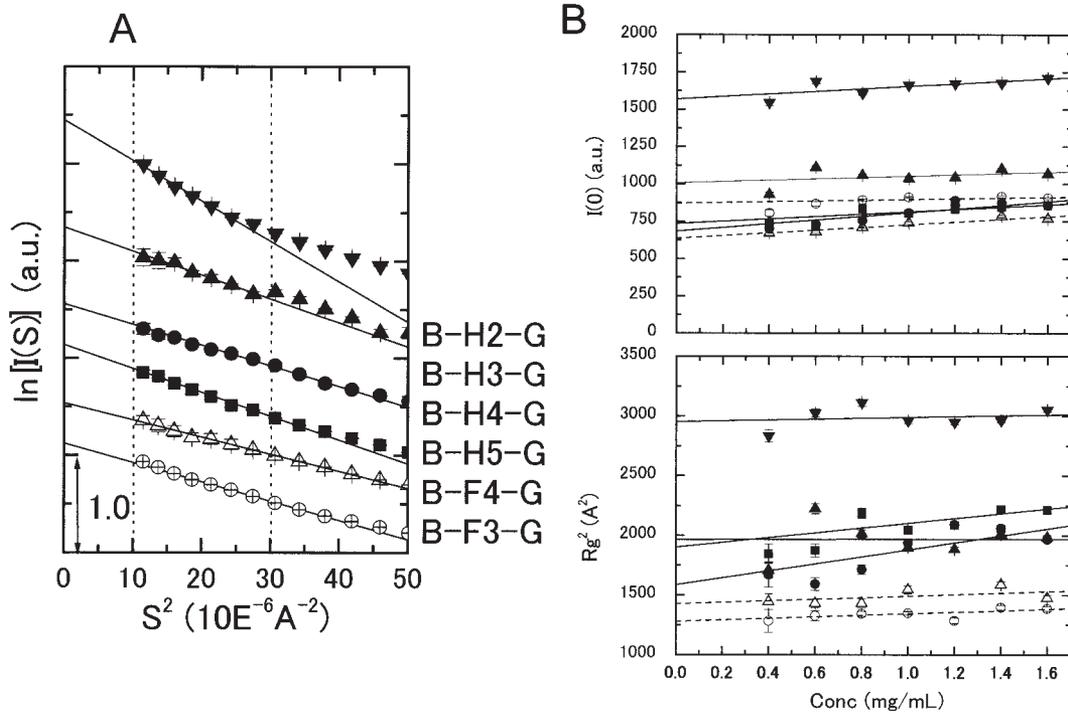


Fig. 2. (a) Guinier plots of X-ray scattering intensity from the chimeric proteins. The scattering intensity was extrapolated to zero protein concentration. The fitting region of S^2 (\AA^{-2}) was from $10 \times 10^{-6} \text{\AA}^{-2}$ to $30 \times 10^{-6} \text{\AA}^{-2}$, which is shown as the broken lines ($10 \times 10^{-6} \text{\AA}^{-2}$ to $25 \times 10^{-6} \text{\AA}^{-2}$ for B-H2-G). Data points are shifted equally for clarity, and symbols are as follows: B-H2-G (\blacktriangledown); B-H3-G (\blacktriangle); B-H4-G (\bullet); B-H5-G (\blacksquare); B-F4-G (\triangle); B-F3-G (\circ). (b) Protein concentration dependence on R_g and $I(0)/C$. The symbols are the same as those in panel A.

TABLE I. Summary of Structural Parameters Determined by SAXS

Chimeric proteins	Guinier approximation		$P(r)$ Function		
	R_g (\AA)	$I(0)$	R_g (\AA)	$I(0)$	D_{\max} (\AA)
B-H2-G	56.2 ± 0.5^a	1717 ± 22^a	52.2 ± 0.1	1402 ± 6	160 ± 10
B-H3-G	43.2 ± 2.2	941 ± 54	44.7 ± 0.5	912 ± 15	150 ± 10
B-H4-G	40.2 ± 1.5	705 ± 23	41.6 ± 0.1	724 ± 3	130 ± 10
B-H5-G	43.2 ± 0.7	764 ± 17	44.6 ± 0.1	738 ± 4	140 ± 10
B-F3-G	36.5 ± 2.1	888 ± 49	37.0 ± 0.1	862 ± 2	120 ± 10
B-F4-G	38.8 ± 0.3	753 ± 4	38.2 ± 0.04	710 ± 1	120 ± 10

^aThese values were determined with a fitting region of S^2 (\AA^{-2}) from 10×10^{-6} to 25×10^{-6} .

mainly reflects a shorter spatial distance between EBFP and EGFP, because FRET is more sensitive to the spatial distance than to the orientation factor.³¹ Thus, the FRET analysis suggests that the spatial distance between the two domains, EBFP and EGFP, increases as the lengths of the helical linkers increase. In the case of the chimeric proteins with flexible linkers, the FRET efficiency was relatively high compared to that of the helical linkers. For example, when H4 was compared to F4, which has the same number of amino-acid residues, the FRET efficiency of B-H4-G (EBFP-H4-EGFP) was found to be much less than that of B-F4-G (EBFP-F4-EGFP). This means that the FRET efficiency was not simply relevant to the number of linker residues. The helical linker is able to separate the two domains relatively.

Radius of Gyration and Maximal Dimension of Chimeric Proteins Determined by SAXS Analysis

Guinier plots of the scattering intensity profiles at high protein concentrations indicated an upward curvature in B-H2-G and B-H3-G (data not shown), but this was negligible for the chimeric proteins with other linkers. At a first glance, the tangent of the fitted curve (the radius of gyration) is steeper when the helical linkers are employed [Fig. 2(a)]. R_g and $I(0)$, determined by the Guinier plot, are summarized in Table I. $I(0)$ remained almost constant within the protein concentration range used for the SAXS measurements [Fig. 2(b)], except in B-H2-G and B-H3-G. The zero extrapolation $I(0)$ of the profiles is consistent with the molecular weights: the $I(0)$ values obtained from B-F3-G, B-F4-G, B-H4-G and B-H5-G ranged from $705 \pm$

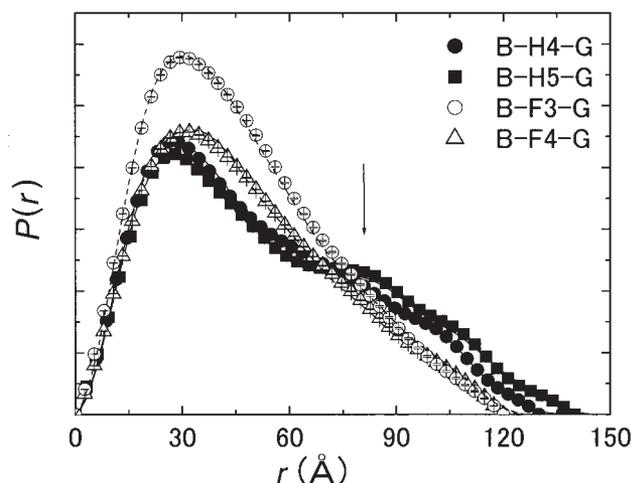


Fig. 3. Pair distance distribution functions $P(r)$ for the chimeric proteins. The ranges of S used for calculation were 0.00315 to 0.05 \AA^{-1} . Symbols indicate experimental $P(r)$ values and are the same as those used in Figure 2. Solid lines (B-H4-G and B-H5-G) and broken lines (B-F3-G and B-F4-G) show the $P(r)$ functions calculated from bead models, which were smoothed by Fourier transformation. The arrow reflects the inter-domain distance.

23 to 888 ± 49 (arbitrary units), whereas an $I(0)$ of 923 ± 1 was obtained for BSA [molecular weight (MW) of 68 kDa at 2 mg/mL] on the same beamline, giving rise to a MW of 52 to 65 kDa. The real MW values are 59.7 kDa (B-F3-G), 60.1 kDa (B-F4-G), 59.8 kDa (B-H2-G), 60.3 kDa (B-H3-G), 60.7 kDa (B-H4-G) and 61.2 kDa (B-H5-G). These results indicate that B-F3-G, B-F4-G, B-H4-G and B-H5-G are monomeric. The $I(0)$ value of B-H2-G is more than twice as high as those of the other linkers, indicating multimerization of B-H2-G in solution. The $I(0)$ values of B-H3-G are slightly larger than those of the other chimeras, which suggests a mixture of monomers and multimers.

The pair distribution function $P(r)$ (Fig. 3) and the maximum dimension D_{\max} are summarized in Table I. The increased D_{\max} values also support the idea that B-H2-G and B-H3-G are fully or partially aggregated. Apart from these two linkers, the R_g and D_{\max} values of the chimeric proteins, both with flexible linkers and with helical linkers, indicate that the chimeric proteins are elongated, because they are extraordinarily large as compared to the values for globular proteins with similar molecular masses (about 60 kDa). Comparing flexible linkers (F3, F4) to helical linkers (H4, H5), the smaller R_g and D_{\max} values suggest that the chimeric proteins with the flexible linkers assume more compact conformations than the chimeric proteins with the helical linkers. Comparing B-H4-G to B-H5-G, the R_g and D_{\max} values increased with the lengths of the helical linkers. On the other hand, comparing B-F3-G to B-F4-G, the changes in R_g and D_{\max} are relatively small: The lengths of the linkers contribute to separating the two domains with the helical linker, but not with the flexible linker.

All of the $P(r)$ values are typical of an elongated particle, and a particularly broad maximum was observed for the chimeric proteins with the helical linkers (B-H4-G, B-H5-G)

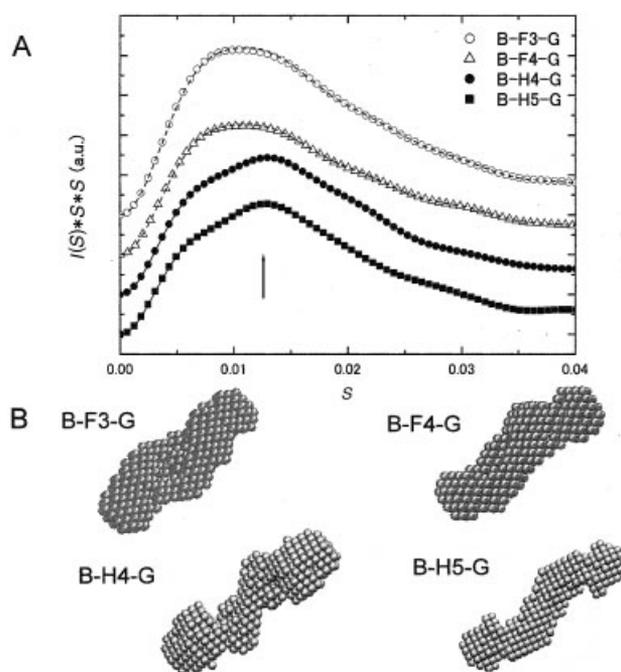


Fig. 4. Comparison of the scattering intensity data with fitted curves. (a) Data are expressed in an S vs. $S^*S^*I(S)$ plot and are shifted with respect to each other. Symbols are the same as those used in Figure 2. (b) The low-resolution bead models used for the best fits by the DAMMIN fitting are shown on the bottom of the graph. The arrow in the figure reflects the two-domain structure.

(indicated by an arrow in Fig. 3: $r \approx 80 \text{ \AA}$), although they were not observed in the case of flexible linkers. Together with the hump ($S \approx 0.0125 \text{ \AA}^{-1}$) observed in the plot of S versus $S^*S^*I(S)$, called the Kratky plot (see Fig. 4), they have the characteristics of a two-domain structure.³² The results suggest that the helical linkers separate the two domains more clearly than the flexible linkers. A refined image of the domain separation will be discussed in the following section.

Bead Models of Chimeric Proteins

To characterize the distance and relative orientation of the two domains of the chimeric proteins, the shapes of the chimeric proteins (B-H4-G, B-H5-G, B-F3-G, B-F4-G) were modeled using an *ab initio* modeling program, DAMMIN.²¹ The chimeric protein models are composed of small beads. The shapes were determined using non-linear least squares fitting to the experimental SAXS curves without any additional information. We examined various parameters, such as with/without symmetry constraints and different annealing protocols. The models possessed a common feature of an elongated shape with two cylindrical portions. We employed P2 symmetry, since the structures of EBFP and EGFP were almost the same.

Observations of the models revealed that they have the following characteristics (Fig. 4). First, the overall shapes of the four chimeric proteins are elongated particles ($30 \text{ \AA} \times 30 \text{ \AA} \times 120 \sim 140 \text{ \AA}$). Second, the chimeric proteins with flexible linkers were more compact than those with the

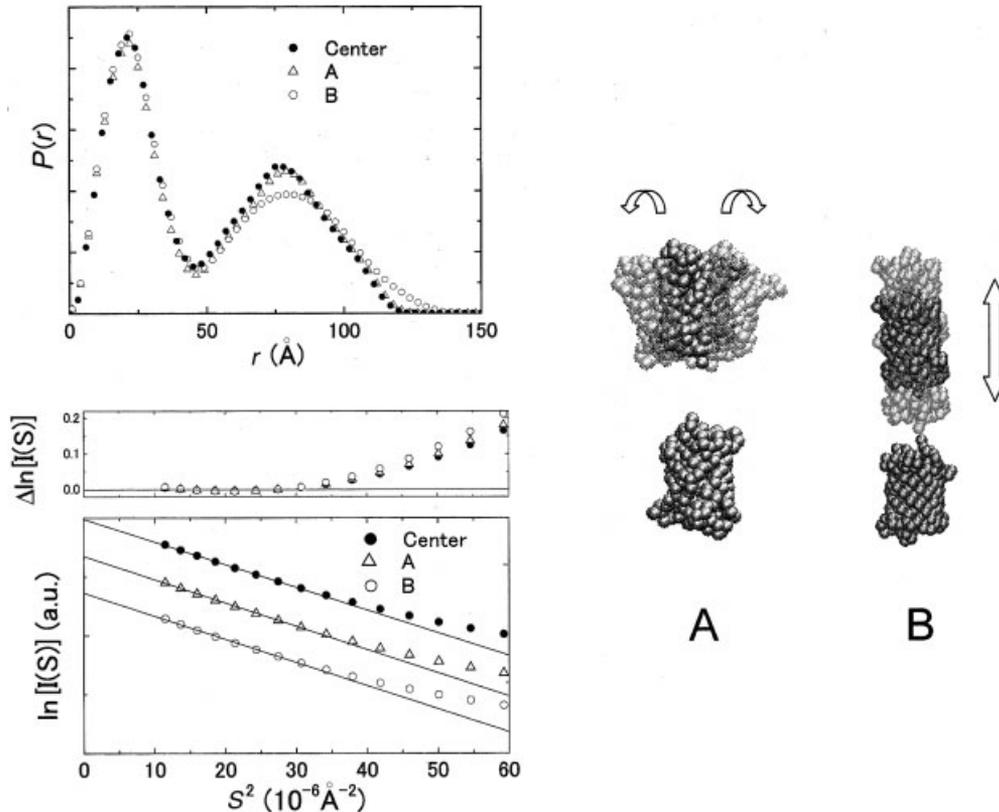


Fig. 5. Relationship between the ensemble-averaged and the center value of $P(r)$ and the Guinier plot. For simplicity, the linker parts, the S-tag, the His-tag and the C-terminal nine residues disordered in the crystal structure of EBFP/EGFP were omitted. One domain (PDB: 1BFP) was fixed, while the other was changed from the center value (\bullet), 20 Å apart from the other one. In order to enhance the effect of ensemble averaging, the population of each configuration was set to be unity. (a) Rotational movement (\triangle). The range of rotation was -20 to $+20^\circ$. (b) Translational movement (\circ). The range of movement was -10 to $+10$ Å. Both movements have the identical peak position $r \approx 80$ Å, which corresponds to the distance between the two domain centers. Since $P(r)$ depends highly on the distance between the two domains, the ensemble-averaged $P(r)$ is more sensitive to translational movement (b) than rotational movement (a). The Guinier plot and the deviation from the Guinier approximation are plotted. Symbols are the same as those used for the $P(r)$ functions.

helical linkers. Third, a dumbbell-like structure with a narrow part at the interface of the two domains is only observed in the chimeric proteins with the helical linkers (B-H4-G, B-H5-G), while the domains of the other two proteins with the flexible linkers do not look well separated. The domains of the chimeric proteins can be superimposed on each other. They had a similar shape, approximated well by a cylinder with diameter 30 Å and height 50 Å, which is consistent with the structures of EBFP and EGFP (a cylinder with diameter 24 Å and height 42 Å, according to the crystal structures of BFP and GFP^{26,33}) if we consider the hydration.³⁴ In addition, the lengths of the helical linkers in the bead models are apparently too short, as compared to the theoretical α -helix length: since the length of helix per residue corresponds to 1.5 Å,³⁵ the lengths of the helical linkers for H4 (25 a.a.) and H5 (30 a.a.) should be 37.5 Å and 45 Å, assuming that they completely adopt the α -helix conformation. At first sight, the domain orientations and the lengths of the helical linkers might contradict each other, if the two domains and the helical linker are

situated in a straight line. We constructed atomic model structures using the Situs program package^{24,25} to visualize the connectivity afforded by the helical linker, as discussed in the following section.

DISCUSSION

Structural Rigidity of Chimeric Proteins in Solution Suggested by SAXS Profiles

In a dumbbell like structure, the R_g values are mostly reflected by the two-domain conformation.³⁶ In other words, characterizing the distribution of the protein configuration that corresponds to the distribution of R_g will reflect the shape variations and thus the rigidity of the linker. The observed $R_{g \text{ average}}$ is the average of the various states R_{gn} :

$$R_{g \text{ average}}^2 = \frac{\sum n R_{gn}^2}{\sum n}$$

The variation of R_{gn} cannot be estimated with the observed value, $R_{g \text{ average}}$. The variation, however, might

be seen as a deviation from the Guinier approximation: the average intensity is the sum of the constituent intensities, and the Guinier approximation for R_g average is more limited for a wide variation of R_{gn} values. Figure 5 illustrates two different types of domain movements and their ensemble-averaged $P(r)$ and Guinier plots. For simplicity, the linkers and tag-sequences were omitted. Rotational movement of the domain (Fig. 5, mode A) results in less R_g deviation as compared to translational movement (Fig. 5, mode B). The ensemble-averaged $P(r)$ values from various conformations by rotational and translational movement are not very different from the center value, except in the region of D_{max} , especially in mode B. This suggests that the effect of the distribution should appear in the small-angle region at reciprocal space. In Figure 5, the deviation from the Guinier approximation as well as the Guinier plot itself is shown for the averaging modes A, B and the center configuration, respectively. In order to clarify the difference, the movements of mode A and mode B are set so that these R_g average values coincide with that of the center of the movement.

The linear region in the Guinier plot becomes narrower in mode B, as compared to those in mode A and the center. Our experimental data (Fig. 2) show that all Guinier plots exhibit a wide linear region, suggesting the narrow distribution of the domain conformations with translational movement (Fig. 5, mode B), rather than rotational movement (Fig. 5, mode A). In addition, Figure 4(b) shows that the bead models of B-H4-G and B-H5-G form dumbbell-like structures with narrow linkers. These SAXS analyses suggest that part of each helical linker seems to be rigid, supporting the α -helix structure of the helical linkers (H4, H5) in the chimeric proteins. The α -helix structures of H4 and H5 are consistent with the CD spectral analysis in our previous report.¹⁰

Comparison Between Chimeric Proteins with Helical Linker and Flexible Linker

As indicated in the Results section, the chimeric proteins with helical linkers are more elongated than those with flexible linkers. This was already suggested by the FRET analysis (Fig. 1), but SAXS provides more definitive information. In a previous study using FRET and CD,¹⁰ we speculated that the chimeric protein with the flexible linker forms the most compact conformation, with side-by-side contact [Fig. 6(a)], by means of the association of EBFP and EGFP. GFP variants reportedly formed dimers with weak binding constants ($K_{dimer} = \approx 100 \mu M$).³⁷ The dimer of GFP with the side-by-side domain contacts was also seen in the crystal structure.³⁸ However, both the R_g and the $P(r)$ values demonstrate that the chimeric proteins with flexible linkers assume the elongated conformation [Fig. 6(b)] rather than the most compact conformation [Fig. 6(a)]. Besides, all of the chimeric proteins have a narrow distribution of domain conformations, judging from the linearity of the Guinier plot [Fig. 2(a)]. The two domains, EBFP and EGFP, seem to have a tendency to maintain a distance from each other, indicating that

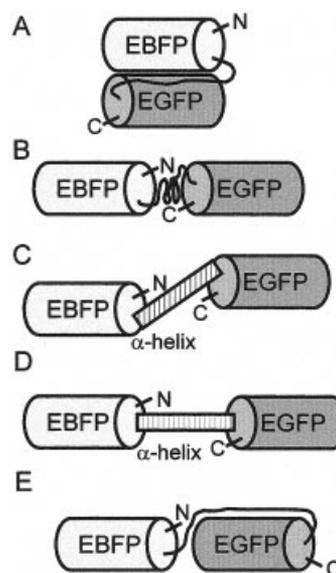


Fig. 6. Schematic diagrams of various conformations of the chimeric proteins with the linkers. (a) EBFP and EGFP side side by side, for the most compact conformation with the flexible linker. The expected R_g and D_{max} values of this side-by-side configuration were about 26 Å and 80 Å, respectively. (b) EBFP and EGFP are situated in a straight line, with the flexible linker between the two domains. The expected R_g and D_{max} values of this elongated configuration were about 30 Å and 100 Å, respectively. (c) The helical linker connects EBFP and EGFP diagonally. (d) The helical linker and the long axes of EBFP and EGFP are situated in a straight line. (e) The top surface of EBFP faces the bottom surface of EGFP with the flexible linker.

flexibility of the linker does not correspond to flexibility in the inter-domain distance. In the case of the flexible linker, the inter-domain distance was not regulated by the linker part. This is supported by the observation that the lengthening of the flexible linker did not change the D_{max} value. The difference between the helical linker and the flexible linker is the spatial separation of the two domains in the chimeric proteins. The flexible linker part is not conspicuous in B-F3-G and B-F4-G (Fig. 4). The structural independence of the two domains therefore seems to be very low. What is the origin of the loss of spatial separation of the two domains for the flexible linker shown in Figure 4? SAXS bead models are supposed to select one structure representing the highest population, as demonstrated by comparing the bead models and the nuclear magnetic resonance (NMR) structures of troponin-C.³⁹ Unlike the hinge bending in the linker helix of troponin-C, the present chimeric protein, especially with the flexible linker, may introduce a less-defined structure, which loosely occupies the space between the two domains [Fig. 6(b)]. It is possible for these disordered linkers to interact with the S-tag, the His-tag and some loops at the interface between the two domains, thereby disturbing the structure. In the interface space between the two domains, there are the linker sequences (19~25 a.a.), the S-tag (33 a.a.), the His-tag (8 a.a.) and nine residues at each C-terminus of EBFP and EGFP, which were disordered in the crystal structure of BFP.²⁶ Thus, 78~84 amino acid residues of

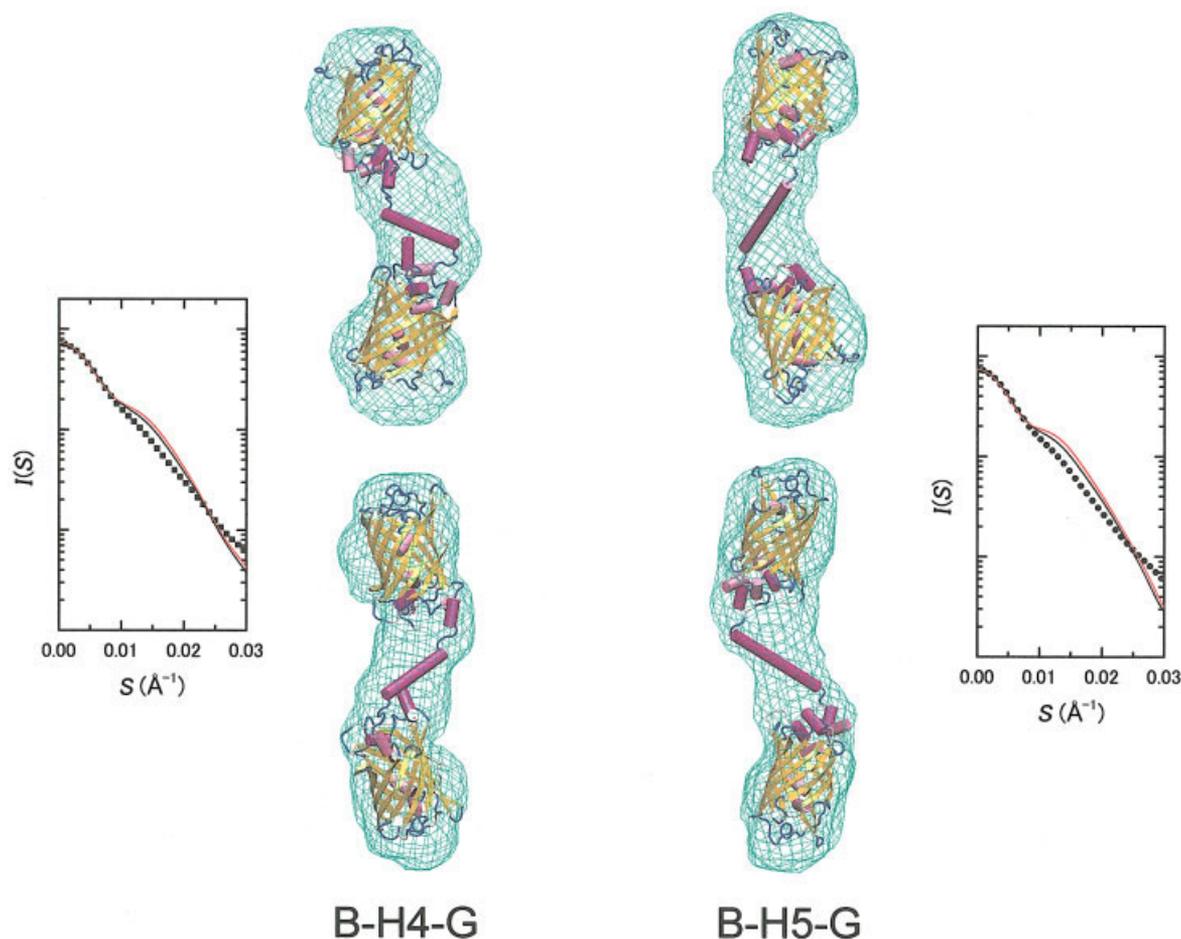


Fig. 7. High-resolution models (cartoon representation) of B-H4-G and B-H5-G, constructed using Situs²⁴ and the graphic program VMD.²⁷ Low-resolution models are shown as wire-frames. The linker and the C-terminal parts of the two domains are modeled. The S-tag and the His-tag were omitted. Two different views are shown. The intensity curves by the models are the experimental curve (dotted), and the simulations from the high-resolution model with (solid line) and without (red solid line) the linker. The calculated SAXS parameters are $R_g = 38.78 \text{ \AA}$ for B-H4-G with linker; $R_g = 39.44 \text{ \AA}$ for B-H4-G without linker; $R_g = 42.27 \text{ \AA}$ for B-H5-G with linker; $R_g = 42.84 \text{ \AA}$ for B-H5-G without linker, respectively. The accuracy of fit values are $R_f = 0.06$ for B-H4-G with linker; $R_f = 0.075$ for B-H4-G without linker; $R_f = 0.07$ for B-H5-G with linker; $R_f = 0.106$ for B-H5-G without linker, respectively.

the chimeric protein residues (538~544 a.a.) are present in this part. On the other hand, the helical linkers (H4, H5) probably form the rigid helix structure independently of the other portion, making it possible to regulate the spatial positions of the two domains of the chimeric proteins.

High-Resolution Modeling of B-H4-G and B-H5-G Based on SAXS Bead Models

Atomic models were created with the program Situs (Fig. 7) to investigate the connectivity of the helical linkers (H4, H5) evident in the low-resolution bead models. This is worthwhile because a low-resolution model often cannot explain atomic details. Sometimes the domain positions and orientations in a low-resolution envelope are too remote or bent too sharply to reveal the polypeptide connection. The high-resolution models provide a visualization of the results in a comprehensive manner. To arrive at a unique model, we assumed that: (1) the structures of

EBFP and EGFP are equivalent to that of BFP (PDB code: 1BFP),²⁶ due to the approximately 99% sequence homology among EBFP, EGFP and BFP; (2) the helical linkers (H4, H5) form an α -helix structure, as suggested by CD and SAXS analyses; (3) the S-tag (33 a.a.) at N-terminus and the His-tag (8 a.a.) at the C-terminus are omitted for simplicity; and (4) the top surfaces of EBFP and EGFP, including both the N-terminal and the C-terminal regions, face each other [Fig. 6(c and d)], and thus we excluded the conformation with the top surface of EBFP facing the bottom surface of EGFP [Fig. 6(e)].

As described in Materials and Methods, the calculated scattering curves from high-resolution models qualitatively simulate the experimental data. The calculated R_g values are also consistent with the experimental data (Fig. 7 legend). However, their accuracy of fit was much worse than that of the low-resolution model (Fig. 7). What is the origin of this discrepancy? One possible explanation utilizes the assumption (3), the omission of the tags, in

constructing high-resolution models. The good fit of the BFP structure onto each lobe of the envelope is also supported by the high scores of correlation determined by the Colores program. In the low-resolution models, the helical linkers (H4, H5) sharpen the interface of the two domains compared to the flexible linkers (F4, F5); however, the interface of B-H4-G and B-H5-G is not sharp enough to differentiate the linker helix clearly. There are still large densities in the interface that are not interpreted in the high-resolution model (Fig. 7). These residual densities are probably derived from the S-tag (33 a.a.) and the His-tag (8 a.a.), which are omitted in the high-resolution models. The contribution of the tags themselves is not as dominant in the small angle region, but its absence may affect the wider scattering region (see supplemental figure) because it changes the shape of the interface region.

The high-resolution models (Fig. 7) show that the helical linker connects the EBFP and EGFP domains diagonally [Fig. 6(c)] rather than longitudinally [Fig. 6(d)]. Even if we place the domain at the remotest distance in the bead model by hand, D_{\max} does not allow for the two domains and the helical linker to be situated longitudinally [Fig. 6(d)]. When the helical linker is situated diagonally, how will this affect the distance between the two domains? The inter-domain distances between the C_{α} of EBFP His66 and EGFP Tyr66 at the center of the chromophores for B-H4-G and B-H5-G were 73.4 Å and 83.5 Å, respectively. As the length of the helical linker increases, the calculated inter-domain distance increases. The increase in length of the helical linker (approx. $1.5 \text{ \AA} \times 5 \text{ residues} = 7.5 \text{ \AA}$) is roughly comparable with the increase in the calculated inter-domain distance (approx. 10.1 Å).

Features of Helical Linkers in Chimeric Proteins

The model of B-H4-G suggests that both the termini and the sides of the helical linkers coordinate the top surfaces of EBFP and EGFP. The helical linker $(EAAAK)_n$ is stabilized in solution by salt bridges (H-bonded ion pairs) of Glu-Lys.¹³ Since the linker contains the negative and positive charges of Glu and Lys, it can form ion pairs with the charged residues on the top surfaces of EBFP and EGFP. With the longer helical linkers (H4, H5), the linkers retained the α -helix structure, and some of the linker residues contacted the EBFP and EGFP domains. On the other hand, in the case of the shorter helical linkers, especially H2, most residues of the linkers are probably situated closer to the two domains. The charged residues, such as Glu and Lys, are likely to form ion pairs with the oppositely charged residues on the top surfaces of EBFP and EGFP, resulting in the destabilization of the short helix. Actually, the H2 linker did not form an α -helix structure, according to the CD analysis.¹⁰ The destabilized and melted short helix linkers (H2, H3) may act as attractants to attach neighboring molecules due to their charges and hydrophobicity, and consequently they cause multimerization of B-H2-G and B-H3-G.

CONCLUSIONS

In this work, synchrotron X-ray small-angle scattering revealed the average conformations of variably linked chimeric proteins with helical and flexible linkers in solution. Independent of the type of linker, the chimeric proteins exhibited elongated structures. The linkers, especially in the helical case, appear to be rather rigid. In the helical case, the separation of the two domains is well defined, while the spatial separation is not conspicuous if the linker is flexible. The helical linker can effectively separate the two domains of the chimeric proteins, without perturbing the domains. Superposition of the high-resolution models of B-H4-G and B-H5-G onto the low-resolution SAXS model suggests that the helical linker connects the two domains diagonally. This connection mode may destabilize shorter helical linkers.

The construction of chimeric/fusion proteins has become routine in biochemical and molecular biology research. The selection of the linker sequence is particularly important for the construction of multi-functional chimeric proteins. Since the length of the helical linker correlates with the distance between the two domains, the helical linker roughly controls the inter-domain distance. On the other hand, in the case of the flexible linkers, the flexible linker cannot change the inter-domain distance. Therefore, the helical linkers composed of $(EAAAK)_n$ motifs ($n = 4, 5$) seem to be good candidates for the linker of a multi-functional chimeric protein, because they form a rigid helix by themselves and can effectively separate the functional domains and keep them independent.

Linker engineering, to control the distance and orientation between two functional domains, will increase in importance if multi-domain proteins can be designed *de novo*. The low-resolution structures presented here exemplify the difficulty of this problem. It is important to consider the global conformational arrangements for linker design. SAXS is a very powerful tool for this purpose, and the present study encourages the use of routine SAXS analyses to provide feedback for linker design.

ACKNOWLEDGEMENTS

We thank Dr. H. Ueda at the University of Tokyo and Dr. N. Kamiya at Kyushu University for their helpful advice. We are also grateful to Dr. Y. Cong at the University of Texas Health Science Center for help in modeling. Dr. S. Akiyama kindly read this manuscript. Finally, we thank Dr. Y. Maeda for his continuous encouragement of T. F. in beamline management.

REFERENCES

1. Arai R, Ueda H, Nagamune T. Construction of chimeric proteins between protein G and fluorescence-enhanced green fluorescent protein, and their application to immunoassays. *J Ferment Bioeng* 1998;86:440–445.
2. Arai R, Ueda H, Tsumoto K, Mahoney WC, Kumagai I, Nagamune T. Fluorolabeling of antibody variable domains with green fluorescent protein variants: application to an energy transfer-based homogeneous immunoassay. *Protein Eng* 2000;13:369–376.
3. Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M. Single-chain antigen-binding proteins. *Science* 1988;242:423–426.

4. Bulow L. Characterization of an artificial bifunctional enzyme, beta-galactosidase/galactokinase, prepared by gene fusion. *Eur J Biochem* 1987;163:443–448.
5. Alfthan K, Takkinen K, Sizmann D, Soderlund H, Teeri TT. Properties of a single-chain antibody containing different linker peptides. *Protein Eng* 1995;8:725–731.
6. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol* 1990;211:943–958.
7. Crasto JC, Feng J. LINKER: a program to generate linker sequences for fusion proteins. *Protein Eng* 2000;13:309–312.
8. Robinson CR, Sauer RT. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc Natl Acad Sci USA* 1998;95:5929–5934.
9. Maeda Y, Ueda H, Kazami J, Kawano G, Suzuki E, Nagamune T. Engineering of functional chimeric protein G-Vargula luciferase. *Anal Biochem* 1997;249:147–152.
10. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 2001;14:529–532.
11. Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* 1996;6:178–182.
12. Cormack BP, Valdivia R, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 1996;173:33–38.
13. Marqusee S, Baldwin RL. Helix stabilization by Glu...Lys+ salt bridges in short peptides of de novo design. *Proc Natl Acad Sci USA* 1987;84:8898–8902.
14. Fujisawa T, Inoue K, Oka T, Iwamoto H, Uruga T, Kumasaka T, Inoko Y, Yagi N, Yamamoto M, Ueki T. Small-angle X-ray scattering station at the SPring-8 RIKEN beamline. *J Appl Crystallogr* 2000;33:797–800.
15. Fujisawa T, Inoko Y, Yagi N. The use of a Hamamatsu X-ray image intensifier with a cooled CCD as a solution X-ray scattering detector. *J Synchrotron Radiat* 1999;6:1106–1114.
16. Guinier A, Fournet G. Small-angle scattering of X-rays. New York: Wiley 1955.
17. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 1992;25:495–503.
18. Fujisawa T, Kostyukova A, Maeda Y. The shapes and sizes of two domains of tropomodulin, the P-end-capping protein of actin-tropomyosin. *FEBS Lett* 2001;498:67–71.
19. Porod G. General theory. In: Small-angle X-ray scattering. Glatter O, Kratky O, editors. London: Academic Press; 1982. p 17–51.
20. Svergun D, Barberato C, Koch MHJ. CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 1995;28:768–773.
21. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999;76:2879–2886.
22. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 2001;34:33–41.
23. Funari SS, Rapp G, Perbandt M, Dierks K, Vallazza M, Betzel C, Erdmann VA, Svergun DI. Structure of free *Thermus flavus* 5 S rRNA at 1.3 nm resolution from synchrotron X-ray solution scattering. *J Biol Chem* 2000;275:31283–31288.
24. Wriggers W, Milligan RA, McCammon JA. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 1999;125:185–195.
25. Wriggers W, Chacon P. Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *J Appl Crystallogr* 2001;34:773–776.
26. Wachter RM, King BA, Heim R, Kallio K, Tsien RY, Boxer SG, Remington SJ. Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. *Biochemistry* 1997;36:9759–9765.
27. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33.
28. Brunger AT. X-PLOR, version 3.1: A system for X-ray crystallography and NMR. New Haven: Yale University Press 1992.
29. Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY. Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin. *Nature* 1997;388:882–887.
30. Tsien RY. The green fluorescent protein. *Annu Rev Biochem* 1998;67:509–544.
31. Wu P, Brand L. Resonance energy transfer: methods and applications. *Anal Biochem* 1994;218:1–13.
32. 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.
33. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 1996;273:1392–1395.
34. Svergun DI, Richard S, Koch MH, Sayers Z, Kuprin S, Zaccai G. Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc Natl Acad Sci USA* 1998;95:2267–2272.
35. Branden C, Tooze J. Introduction to protein structure. New York: Garland Publishing 1991.
36. Fujisawa T, Ueki T, Iida S. Structural change of troponin C molecule and its domains upon Ca²⁺ binding in the presence of Mg²⁺ ions measured by a solution X-ray scattering technique. *J Biochem (Tokyo)* 1990;107:343–351.
37. Phillips Jr. GN. Structure and dynamics of green fluorescent protein. *Curr Opin Struct Biol* 1997;7:821–827.
38. Yang F, Moss LG, Phillips Jr. GN. The molecular structure of green fluorescent protein. *Nature Biotechnol* 1996;14:1246–1251.
39. Takahashi Y, Nishikawa Y, Fujisawa T. Evaluation of three algorithms for ab initio determination of three-dimensional shape from one-dimensional solution scattering profiles. *J Appl Crystallogr* 2003;36:549–552.