Protein Engineering vol.14 no.8 pp.529–532, 2001

COMMUNICATION

Design of the linkers which effectively separate domains of a bifunctional fusion protein

Ryoichi Arai, Hiroshi Ueda, Atsushi Kitayama, Noriho Kamiya and Teruyuki Nagamune[1]

Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[1]To whom correspondence should be addressed.
E-mail: nagamune@bio.t.u-tokyo.ac.jp

With the aim of separating the domains of a bifunctional fusion protein, the ability of several lengths of helix-forming peptides to separate two weakly interacting β-can domains was compared with that of flexible linkers or of a three α-helices bundle domain. We introduced helix-forming peptide linkers A(EAAAK)$_n$A (n = 2–5) between two green fluorescent protein variants, EBFP and EGFP, and investigated their spectral properties. The fluorescence resonance energy transfer from EBFP to EGFP decreased as the length of the linkers increased. The circular dichroism spectra analysis suggested that the linkers form an α-helix and the α-helical contents increased as the length of the linkers increased. The results clearly suggested the ability of the helical linkers to control the distance and reduce the interference between the domains. This 'linker engineering' may open a way to the rational design of linkers which maximize the multiple functions of fusion proteins or *de novo* multi-domain proteins.

Keywords: α-helix/fluorescence resonance energy transfer/fusion protein/green fluorescent protein/linker engineering

Introduction

The gene fusion technique has become an indispensable tool in a variety of fields of biochemical research. The construction of recombinant fusion/chimeric proteins has been adopted as a means to increase the expression of soluble proteins and to facilitate protein purification. Furthermore, a wide range of applications of the gene fusion technique has been reported in the field of biotechnology. These include the immunoassays using chimeras between antibody fragments or antibody binding domains, and enzymes or green fluorescent protein variants (Maeda *et al*., 1996, 1997; Arai *et al*., 1998, 2000; Suzuki *et al*., 1999), the selection and production of antibodies (Bird *et al*., 1988) and the engineering of bifunctional enzymes (Bulow, 1987; Ljungcrantz *et al*., 1989).

The construction of a fusion protein involves the linking of two proteins or domains of proteins by a peptide linker. The selection of the linker sequence is particularly important for the construction of functional fusion proteins. Several studies have been made on the linker selection (Argos, 1990; Alfthan *et al*., 1995; Robinson and Sauer, 1998; Crasto and Feng, 2000). These studies suggested mainly that the flexibility and hydrophilicity of the linker were important not to disturb the functions of the domains. However, in our previous study on Streptococcal protein G-*Vargula* luciferase chimera, simple linking of the two moieties by a flexible linker did not retain the binding activity of protein G C1 domain (Maeda *et al*., 1997). The loss of the binding activity of protein G could be envisaged due to some sort of interaction and interference between the two moieties. When a three α-helices bundle domain of Staphylococcal protein A (PA) was introduced between protein G and luciferase as a linker, goat/sheep IgG binding activity derived from protein G was regained. We reasoned that PA could spatially separate protein G from luciferase and thus reduce the interference between them. These results indicate that spatial separation of the heterofunctional domains of a fusion protein by proper linker peptide might be so effective that the domains work independently. However, there have been very few reports on the linkers which effectively separate the domains. Although PA was effective as a linker in our case, the length of more than 50 amino-acid residues and the additional binding activity to IgG might hinder it from general use.

In this study, we designed shorter linkers than PA to effectively separate bifunctional domains of a fusion protein. The linkers, which were expected to form a monomeric hydrophilic α-helix, were designed according to the previous study on a short peptide forming a monomeric α-helix (Marqusee and Baldwin, 1987). In that study, the best helix-forming peptide, AEAAAKEAAAKEAAAKA [(i+4)E,K], which was stabilized by Glu^-–Lys^+ salt bridges, showed ~80% helicity. To our knowledge, there have been no reports that tried to introduce the helix-forming peptide to a bifunctional fusion protein as a linker, and it has been unknown whether it can actually separate the domains. To study these, we designed the linkers of different lengths A(EAAAK)$_n$A (n = 2–5) based on the helix-forming peptide (i+4)E,K, and introduced them between two domains of a bifunctional fusion protein. We also introduced flexible linkers and PA into the fusion protein for the comparison.

As a model fusion protein, we employed fusion proteins of two *Aequorea* green fluorescent protein (GFP) variants enhanced blue fluorescent protein (EBFP; F64L, S65T, Y66H, Y145F) (Heim and Tsien, 1996) and enhanced green fluorescent protein (EGFP; F64L, S65T) (Cormack *et al*., 1996). EBFP and EGFP are useful for fluorescence resonance energy transfer (FRET) studies (Miyawaki *et al*., 1997; Tsien, 1998; Arai *et al*., 2000). Because a FRET signal provides information about distances in the order of 10 to 100 Å, the technique is suitable for investigating spatial relationships of targeted molecules (Wu and Brand, 1994; Selvin, 1995). In this study, we examined the relative distance between EBFP and EGFP using FRET in order to evaluate designed linkers which effectively separate domains of a bifunctional fusion protein.

Materials and methods

Plasmid construction

For the expression of the fusion proteins, pET TRX Fusion System 32 (Novagen, Madison, WI, USA), a fusion expression system with *Escherichia coli* thioredoxin (Trx) was

employed to enhance the solubility of the highly expressed proteins in *E.coli* cytoplasm (LaVallie *et al*., 1993). The DNA fragment encoding EBFP was prepared from plasmid pEBFP-N1 (Clontech, Palo Alto, CA, USA) by standard PCR using *Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA) and the primers containing the cleavage sites for *Eco*RV and *Hin*dIII. The amplified fragments were digested and cloned into pET32/V_L–EGFP (Arai *et al*., 2000) between *Eco*RV and *Hin*dIII to give the plasmid pET32/EBFP–SL–EGFP. DNA sequencing was performed using the fluorescence DNA sequencer SQ-5500 (Hitachi, Tokyo, Japan). The DNA fragments encoding the designed linkers were prepared by annealing and extension using the appropriate oligo-nucleotide primers with the cleavage sites for *Hin*dIII and *Not*I. These were digested and cloned into pET32/EBFP–SL–EGFP between *Hin*dIII and *Not*I to give the plasmids pET32/EBFP–linker–EGFPs.

The amino acid sequences of the designed linkers were: short linker (SL), LAAA (4 aa) (derived from the cleavage sites for *Hin*dIII and *Not*I); flexible linker 3 (FL3), LGGGGSGGGGSGGGGSAAA (19 aa); flexible linker 4 (FL4), LSGGGGSGGGGSGGGGSGGGGSAAA (25 aa); helical linker 2 (HL2), LAEAAAKEAAAKAAA (15 aa); helical linker 3 (HL3), LAEAAAKEAAAKEAAAKAAA (20 aa); helical linker 4 (HL4), LAEAAAKEAAAKEAAAKEAAAKAAA (25 aa); helical linker 5 (HL5), LAEAAAKEAAAKEAAAKEAAAKEAAAKAAA (30 aa); B domain of Staphylococcal PA, LFNKEQQNAFYEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDAQAAA (55 aa).

Protein expression and purification

Escherichia coli AD494(DE3)pLysS (Novagen) was transformed with the plasmids, and selected on LB agar plates containing 50 μg/ml ampicillin, 34 μg/ml chloramphenicol and 15 μg/ml kanamycin. For all the cultivations thereafter, LB medium containing the same concentration of the three antibiotics was used. The medium (1.5 l) was inoculated with 5 ml overnight culture at 30°C of each strain harboring pET32/EBFP–linker–EGFPs containing the designed linkers, and cultured at 30°C for 7 h. At an OD_{600} of ≈0.5, IPTG was added to a final concentration of 1 mM to induce the expression of the fusion proteins, and cells were further cultured for 12 h at 16°C. Harvested cells were resuspended in sonication buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and then disrupted by freeze–thaws and sonication with a Sonifier 250 (Branson, Danbury, CT, USA). The supernatant was obtained after centrifugation twice at 20 000 *g* for 20 min at 4°C. The fusion proteins which have the $(His)_6$-tag were purified using Talon metal affinity resin (Clontech) according to the manufacturer's protocol. The fractions with sufficient fluorescence activity were collected. The proteins were specifically digested with thorombin (Novagen) between Trx and EBFP, and Trx fragments were removed by size exclusion chromatography with Superdex 75 (Amersham Pharmacia Biotech, Uppsala, Sweden). The protein concentration was determined by BCA assay kit (Pierce, Rockford, IL, USA) with bovine serum albumin as a standard.

Measurement of fluorescence intensity and circular dichroism (CD) spectra

The fluorescence intensity of the fusion proteins at 444 nm of the EBFP emission peak was measured using a fluorescence spectrophotometer F-2000 (Hitachi) with 380 nm excitation at 25°C. In order to cut the linkers between EBFP and EGFP domains, the fusion proteins (2 μM) were digested by trypsin (50 nM) at 20°C for 12 h in phosphate-buffered saline (pH 7.4). Specific digestion of the linkers was confirmed by SDS–PAGE. The CD spectra of the fusion proteins were measured using a CD spectropolarimeter J-725 (Jasco, Tokyo, Japan) at 25°C.

Table I. FRET of EBFP–linker–EGFPs

Linkers	F_{DA}/F_D	$E_T = 1–F_{DA}/F_D$	$r = R_0(1/E_T–1)^{1/6}$ (Å)
SL	0.739	0.261	50.0
FL3	0.735	0.265	49.8
FL4	0.752	0.248	50.5
HL2	0.766	0.234	51.2
HL3	0.811	0.189	53.5
HL4	0.864	0.136	57.1
HL5	0.910	0.090	61.8
PA	0.868	0.132	57.5

The fluorescence intensity of EBFP (donor) at 444 nm of its emission peak was measured with 380 nm excitation at 25°C. For the measurement, 2 μM of each fusion protein in PBS (pH 7.4) was used. F_{DA}, fluorescence intensity of donor in the presence of acceptor (fusion proteins not treated with trypsin); F_D, fluorescence intensity of donor in the absence of acceptor (fusion proteins treated with trypsin); E_T, efficiency of FRET; *r*, estimated distance between EBFP and EGFP assuming that the critical Förster distance R_0 of the pair EBFP–EGFP is 42 Å.

Results and discussion

Sufficient amounts of the fusion proteins having each linker were expressed and purified, with yields of 10–20 mg per 1.5 l culture. All of the fusion proteins retained sufficient fluorescence derived from both EBFP and EGFP (not shown). To estimate the FRET efficiency from EBFP to EGFP, the linkers between the two were digested by trypsin under the condition where both EBFP and EGFP domains remained intact, which was confirmed by SDS–PAGE (not shown). Based on this negative control where the fluorescence intensity of EBFP (donor) without FRET could be evaluated, FRET efficiency (E_T) was calculated from the ratio of EBFP fluorescence intensities before and after trypsin digestion (Table I). The higher FRET efficiency should mainly reflect the shorter spatial distance between EBFP and EGFP while it may be relevant to the orientation factor between EBFP and EGFP, because the FRET is more sensitive to the spatial distance than the orientation factor (Wu and Brand, 1994). In the case of the fusion proteins with the flexible linkers (FL3 and FL4), the FRET efficiency was highly comparable to that of EBFP–SL–EGFP although the flexible linkers were much longer than SL. Considering that GFP variants can form dimers (K_D = ~100 μM) (Yang *et al*., 1996; Phillips, 1997), FL3 and FL4 are so flexible and long that the two GFP variants may be interacting with each other to form a heterodimer. On the other hand, in the case of the fusion proteins with helical linkers (HL2, HL3, HL4 and HL5), the FRET efficiency decreased as the length of the helical linkers increased. When HL4 was compared with FL4 which has the same number of amino-acid residues as HL4, the FRET efficiency of EBFP–HL4-EGFP was much less than that of EBFP–FL4-EGFP. This means that the FRET efficiency was not simply relevant to the number of linker residues. Assuming that the critical Förster distance R_0 for an EBFP–EGFP pair is 42 Å (Tsien, 1998), the distances between the two domains can be calculated as shown in Table I. The estimated distances increased in

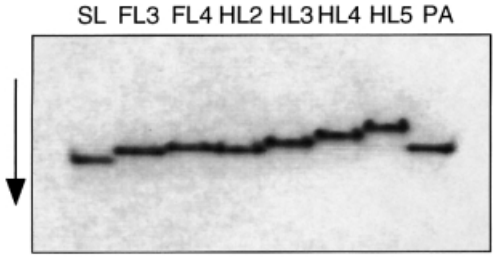

Fig. 1. Polyacrylamide gel (10%) electrophoresis of EBFP–linker–EGFPs in native conditions (native-PAGE). The purified fusion proteins (1.5 μg each) were loaded. The proteins were stained with Coomassie brilliant blue.

proportion to the length of the linkers. The results suggest that the helical linkers can effectively separate the neighboring domains of the fusion protein. It was noteworthy that the FRET efficiency of EBFP–HL4-EGFP was almost the same as that of EBFP–PA–EGFP. This suggests that HL4 and PA linkers can work similarly to separate EBFP and EGFP domains.

The result of 10% polyacrylamide gel electrophoresis of fusion proteins in native conditions (native-PAGE) is shown in Figure 1. The mobility of a protein in native-PAGE is relevant to the electrical charges, the molecular weight and the conformation of the molecules. Because the charges of Glu and Lys ion pairs in the helical linkers should cancel out, and SL, FL3 and FL4 do not have charged amino acids, the total electrical charges of the fusion proteins except EBFP–PA–EGFP are expected to be the same. Therefore, the fusion proteins except EBFP–PA–EGFP should be comparable with each other in respect to the molecular weight and the conformation of the molecules. It appeared that the mobility tended to decrease in the order SL, FL3 and FL4, and in the order HL2, HL3, HL4 and HL5, as the molecular weight increased. However, when HL4 was compared with FL4, irrespective of almost the same molecular weight, the relative mobility of EBFP–HL4-EGFP was less than that of EBFP–FL4-EGFP probably due to the difference of the molecular conformation. Presumably EBFP–HL4-EGFP may form an extended conformation like dumbbells, which have smaller mobility than a compact conformation. In contrast, EBFP–FL4-EGFP may take a more compact conformation because EBFP and EGFP can form a dimer. On the other hand, the relative mobility of EBFP–PA–EGFP was very large despite the molecular weight and the expected conformation. However, the result can be interpreted because the total electrical charges of PA, containing eight negative charged amino acids (Asp and Glu) and five positive charged amino acids (Lys and Arg), are more negative (–3) than the other linkers.

Figure 2A shows CD spectra of the fusion proteins. The spectral profiles show that they consist of mainly β-sheets due to β-can structure of EBFP and EGFP (Ormö *et al*., 1996; Yang *et al*., 1996). In the case of the fusion proteins with the helical linkers, the α-helical contents estimated by Bolotina's method (Bolotina *et al*., 1980; Deléage and Geourjon, 1993) increased as the number of the linker residues increased (HL2, 13%; HL3, 15%; HL4, 16%; HL5, 18%). Furthermore, to examine the secondary structure of the linkers, the differential CD spectra of EBFP–linker–EGFPs minus EBFP–SL–EGFP were drawn as shown in Figure 2B. The differential CD spectra should represent the secondary structure of the linkers, because the CD spectra of the parts other than the linkers should be canceled out. The profiles of the differential CD spectra of

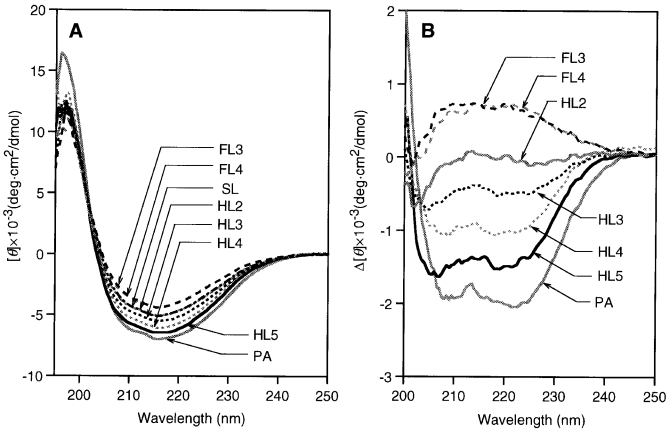

Fig. 2. CD measurement of the fusion proteins. (A) CD spectra of EBFP–linker–EGFPs. For the measurement, 7 μM of each sample in 10 mM sodium phosphate (pH 7.5) was used. (B) The differential CD spectra of EBFP–linker–EGFPs minus EBFP–SL–EGFP.

HL3, HL4, HL5 and PA show characteristics of α-helices with two minima at ~222 nm (n–π* transition) and ~208 nm (π–π transition). The results indicate that these linkers (HL3, HL4, HL5 and PA) form an α-helical structure in the fusion proteins. Considering the differential values of the mean residue ellipticity at 222 nm ($\Delta[\theta]_{222}$), which is an index of α-helical contents, the α-helical contents of the helical linkers (HL2, HL3, HL4 and HL5) increased in proportion to the number of linker residues. On the other hand, the $\Delta[\theta]_{222}$ of EBFP–FL3-EGFP and EBFP–FL4-EGFP were higher than the others. This was probably because the flexible linkers formed a random coiled conformation.

These results indicate that the helical linkers can effectively separate the domains of fusion proteins and the distance between the domains can be controlled by changing the repetitions of the EAAAK motif. According to the FRET measurements, the HL4 linker can equally work well as a PA linker. We think that HL4 and HL5 linkers are advisable candidates for the linker of bifunctional fusion proteins, because they can effectively separate the functional domains and avoid inadequate interaction between the domains. To our knowledge, this is the first comprehensive report of helical linkers with the aim of separating the domains of bifunctional fusion proteins and also controlling the spatial distance between them.

The design of linkers is more important if a multi-domain protein is designed *de novo*. The control of the distance and the orientation of the domains in order to maximize the desired functions is a demanding task in near-future protein engineering. Here we would like to propose a term 'linker engineering', whose objective is the rational control of the linker conformation, flexibility and the distance between protein domains to create a multi-functional fusion protein. This study suggests the potential of the helical linkers as first candidates for its building block.

Acknowledgements

We thank S. Yamaguchi, N. Shigi, T. Suzuki and K. Watanabe at the University of Tokyo for their help with CD measurements. R.A. was supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. This study was also supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (No. 296-10145107) and for Scientific Research (B09450301) from the Ministry of Education, Science, Sports and Culture of Japan, and funded in part by Biodesign Research Promotion Group of the Institute of Physical and Chemical Research (RIKEN), Japan.

References

Alfthan,K., Takkinen,K., Sizmann,D., Soderlund,H. and Teeri,T.T. (1995) *Protein Eng*., **8**, 725–731.
Arai,R., Ueda,H. and Nagamune,T. (1998) *J. Ferment. Bioeng.*, **86**, 440–445.
Arai,R., Ueda,H., Tsumoto,K., Mahoney,W.C., Kumagai,I. and Nagamune,T. (2000) *Protein Eng*., **13**, 369–376.
Argos,P. (1990) *J. Mol. Biol.*, **211**, 943–958.
Bird,R.E., Hardman,K.D., Jacobson,J.W., Johnson,S., Kaufman,B.M., Lee,S.M., Lee,T., Pope,S.H., Riordan,G.S. and Whitlow,M. (1988) *Science*, **242**, 423–426.
Bolotina,I.A., Chekhov,V.O., Lugauskas,V. and Ptitsyn,O.B. (1980) *Mol. Biol. (USSR)*, **14**, 902–909.
Bulow,L. (1987) *Eur. J. Biochem.*, **163**, 443–448.
Cormack,B.P., Valdivia,R. and Falkow,S. (1996) *Gene*, **173**, 33–38.
Crasto,J.C. and Feng,J. (2000) *Protein Eng*., **13**, 309–312.
Deléage,G. and Geourjon,C. (1993) *Comput. Appl. Biosci.*, **9**, 197–199.
Heim,R. and Tsien,R.Y. (1996) *Curr. Biol.*, **6**, 178–182.
LaVallie,E.R., DiBlasio,E.A., Kovacic,S., Grant,K.L., Schendel,P.F. and McCoy,J.M. (1993) *Biotechnology*, **11**, 187–193.
Ljungcrantz,P., Carlsson,H., Mansson,M.O., Buckel,P., Mosbach,K. and Bulow,L. (1989) *Biochemistry*, **28**, 8786–8792.
Maeda,Y., Ueda,H., Hara,T., Kazami,J., Kawano,G., Suzuki,E. and Nagamune,T. (1996) *Biotechniques*, **20**, 116–121.
Maeda,Y., Ueda,H., Kazami,J., Kawano,G., Suzuki,E. and Nagamune,T. (1997) *Anal. Biochem.*, **249**, 147–152.
Marqusee,S. and Baldwin,R.L. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8898–8902.
Miyawaki,A., Llopis,J., Heim,R., McCaffery,J.M., Adams,J.A., Ikura,M. and Tsien,R.Y. (1997) *Nature*, **388**, 882–887.
Morii,H., Uedaira,H., Ogata,K., Ishii,S. and Sarai,A. (1999) *J. Mol. Biol.*, **292**, 909–920.
Ormö,M., Cubitt,A.B., Kallio,K., Gross,L.A., Tsien,R.Y. and Remington,S.J. (1996) *Science*, **273**, 1392–1395.
Phillips,G.N.,Jr (1997) *Curr. Opin. Struct. Biol.*, **7**, 821–827.
Robinson,C.R. and Sauer,R.T. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5929–5934.
Selvin,P.R. (1995) *Methods Enzymol.*, **246**, 300–334.
Suzuki,C., Ueda,H., Tsumoto,K., Mahoney,W.C., Kumagai,I. and Nagamune,T. (1999) *J. Immunol. Methods*, **224**, 171–184.
Tsien,R.Y. (1998) *Annu. Rev. Biochem.*, **67**, 509–544.
Wu,P. and Brand,L. (1994) *Anal. Biochem.*, **218**, 1–13.
Yang,F., Moss,L.G. and Phillips,G.N.,Jr (1996) *Nature Biotechnol.*, **14**, 1246–1251.

Received February 20, 2001; accepted May 14, 2001