

# Design and characterization of structured protein linkers with differing flexibilities

Joshua S. Klein<sup>1†</sup>, Siduo Jiang<sup>1†</sup>, Rachel P. Galimidi<sup>1†</sup>,  
Jennifer R. Keeffe<sup>1</sup> and Pamela J. Bjorkman<sup>1,2,3</sup>

<sup>1</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA and <sup>2</sup>Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125, USA

<sup>3</sup>To whom correspondence should be addressed.  
E-mail: bjorkman@caltech.edu

Received July 2, 2014; revised July 2, 2014;  
accepted August 22, 2014

Edited by Lynne Regan

**Engineered fusion proteins containing two or more functional polypeptides joined by a peptide or protein linker are important for many fields of biological research. The separation distance between functional units can impact epitope access and the ability to bind with avidity; thus the availability of a variety of linkers with different lengths and degrees of rigidity would be valuable for protein design efforts. Here, we report a series of designed structured protein linkers incorporating naturally occurring protein domains and compare their properties to commonly used Gly<sub>4</sub>Ser repeat linkers. When incorporated into the hinge region of an immunoglobulin G (IgG) molecule, flexible Gly<sub>4</sub>Ser repeats did not result in detectable extensions of the IgG antigen-binding domains, in contrast to linkers including more rigid domains such as β2-microglobulin, Zn-α2-glycoprotein and tetratricopeptide repeats. This study adds an additional set of linkers with varying lengths and rigidities to the available linker repertoire, which may be useful for the construction of antibodies with enhanced binding properties or other fusion proteins.**

**Keywords:** fusion proteins/hydrodynamic radius/linker design/size-exclusion chromatography

## Introduction

Fusion proteins are engineered biomolecules containing parts from two or more genes synthesized as a single multi-functional construct. These have been critical in many areas of biological research including affinity purification (Lichty *et al.*, 2005) and protein stabilization for structure determination (Zou *et al.*, 2012). Bi-specific fusion proteins have also been utilized as biopharmaceuticals, with an active drug domain fused to a carrier domain, allowing for the drug's proper transport (Chen *et al.*, 2013). Such proteins have been designed to penetrate epithelial membranes including the blood–brain barrier, as well as to target a specific cell population (Pardridge, 2010). Due to the modularity of protein domains in the generation of functional constructs, fusion

proteins will likely have increasing importance in research and drug design.

The successful construction of fusion proteins relies on the proper choice of a protein linker as direct fusion of two domains can lead to compromised biological activity (Bai *et al.*, 2005; Zhang *et al.*, 2009). Several studies have utilized existing databases to compile and characterize linkers in naturally occurring multi-domain proteins (Argos, 1990; George and Heringa, 2002). These studies have yielded amino acid sequence propensities for natural linkers of various sizes and lengths, as well as information on rigidity and secondary structure. This information has helped the empirical design of linkers that are customized for particular applications.

Linkers can be classified into three groups: flexible, rigid and cleavable (Chen *et al.*, 2013). Flexible linkers are generally composed of small, non-polar or polar residues such as Gly, Ser and Thr. The most common is the (Gly<sub>4</sub>Ser)<sub>*n*</sub> linker (Gly–Gly–Gly–Gly–Ser)<sub>*n*</sub>, where *n* indicates the number of repeats of the motif. Polyglycine linkers have also been evaluated, but the addition of a polar residue such as serine can reduce linker–protein interactions and preserve protein function. Due to their flexibility, these linkers are unstructured and thus provided limited domain separation in a previous study (Evers *et al.*, 2006). As a result, more rigid linkers including polyproline motifs (Schuler *et al.*, 2005) and an all α-helical linker A(EAAAK)<sub>*n*</sub>A (Arai *et al.*, 2001) have been developed.

We are interested in using relatively rigid protein linkers to separate anti-HIV binding proteins at distances that would permit bi- or multi-valent binding to HIV Env glycoproteins with the objective of creating reagents capable of cross-linking epitopes within a single Env trimer (intra-spike cross-linking). Such reagents would take advantage of avidity effects to minimize HIV's ability to evade neutralizing antibodies by rapidly mutating to lower the affinity between the HIV epitopes and the antigen recognition fragment (Fab) of the antibody (Klein *et al.*, 2009). Although the architecture of the HIV spike trimer does not permit intra-spike cross-linking by most natural antibodies (Zhu *et al.*, 2006; Klein and Bjorkman, 2010), it may be possible to create reagents capable of bivalent binding to an HIV Env trimer by fusing two identical reagents or two different reagents with an appropriate length linker. Here we report the design, construction and characterization of a series of structured protein linkers incorporating both rigid and flexible domains that can be used to achieve a variety of different desired separations. The linkers were incorporated into the hinge region of an intact immunoglobulin G (IgG) antibody and evaluated for their relative lengths and rigidities by dynamic light scattering (DLS).

## Methods

### Plasmid construction and protein purification

Genes encoding designed linkers were synthesized (Blue Heron Bio) with restriction sites for the enzymes *Nhe*I (5'-end) and either *Ngo*MIV or *Hind*III (3'-end). These sites

<sup>†</sup>These authors contributed equally to this work.

were also introduced into the gene encoding the heavy chain of the HIV-neutralizing antibody b12 (Roben *et al.*, 1994) such that the insert would be located between hinge region residues His235 and Thr236. Constructs encoding the b12 heavy chain gene with a linker inserted in the hinge region were subcloned into the pTT5 mammalian expression vector. The b12-linker IgGs were expressed transiently in HEK-6E cells by co-transfecting the b12-linker heavy chain genes with the b12 light chain gene as described (Diskin *et al.*, 2011).

IgG-linker fusion constructs were purified by protein A affinity chromatography (GE Healthcare) followed by purification and analysis by size-exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (GE Healthcare) in phosphate-buffered saline, 0.05% w/v sodium azide, pH 7.4.

### Dynamic light scattering

Fractions corresponding to the center of the SEC elution peak were concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore) with a molecular weight cutoff of 100 kDa to a volume of 80–400  $\mu$ l and concentrations of 0.5–1 mg/ml. Concentration differences within this range were not observed to affect the hydrodynamic radius values determined by DLS (data not shown). Sample sizes ranging from 80 to 350  $\mu$ l were loaded into a disposable cuvette, and measurements were performed on a DynaPro® NanoStar™ (Wyatt Technology) using manufacturer's suggested settings. A fit of the second-order autocorrelation function to a globular protein model was used to derive the hydrodynamic radius.

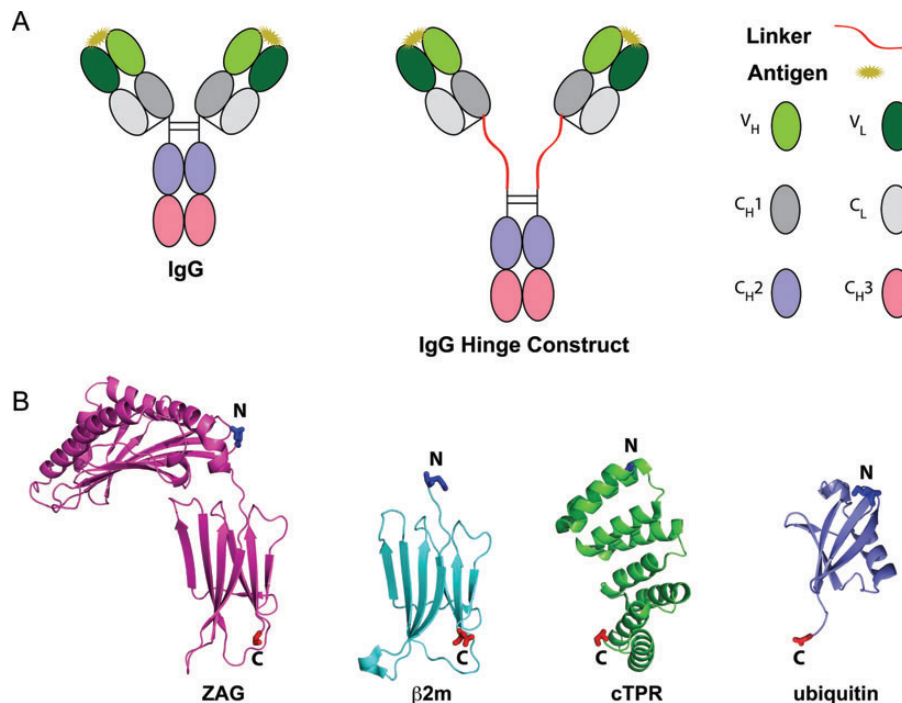
## Results and discussion

### Design and identity of designed linkers

In order to design potential structured linkers, we surveyed the Protein Data Bank (PDB) to find structures that were relatively

elongated and rigid, or represented small globular proteins. We chose Zn- $\alpha$ 2-glycoprotein (ZAG; PDB code: 1ZAG) as an example of a relatively elongated and rigid structure (Sanchez *et al.*, 1999), and  $\beta$ 2-microglobulin ( $\beta$ 2m; PDB code: 1LDS) and ubiquitin (Ub; PDB code: 1UBQ) as examples of small globular proteins (Fig. 1A). ZAG is a 31.5 kDa protein with a class I major histocompatibility complex heavy chain-like fold and a separation distance between the N- and C-termini of  $\sim$ 45 Å.  $\beta$ 2m is a stable 12 kDa protein with an immunoglobulin constant region-like fold that forms a rigid structure with a separation distance between the N- and C-terminus of  $\sim$ 35 Å (Trinh *et al.*, 2002). Likewise, Ub is a compact, stable 8.5 kDa protein with an N- and C-terminal separation distance of  $\sim$ 37 Å (Vijay-Kumar *et al.*, 1987). In addition to the structured linkers chosen from the PDB, proline-rich linkers were designed from the hinge sequence from IgA1 (polyPro and polyPro(Glyc)). This glycosylated region confers rotational flexibility of the Fab relative to the Fc in the context of wild-type dimeric IgA1 (Bonner *et al.*, 2008). In addition, glycosylation has been shown to potentially increase stability of polypeptide linkers (Imperiali and O'Connor, 1999). ZAG,  $\beta$ 2m and Ub proteins were joined in various combinations with short linker regions, either (Gly<sub>2</sub>Ser)<sub>n</sub> repeats, glycosylated proline-rich sequences (polypro(Glyc)), or unglycosylated proline-rich sequences (polypro), to create linkers L1–L12 (Table I).

We also created linkers using tetratricopeptide repeat domains (TPRs; PDB code: 2AVP; L13–L16; Table I; Fig. 1A) (Kajander *et al.*, 2007) that are found in natural proteins such as HSP70/90 (Scheufler *et al.*, 2000). These domains are optimal for use as potential structured linkers because the length of a set of tandem TPR domains corresponds predictably with the number of repeats. Each repeat consists of 34 amino acids with a defined sequence motif that forms two  $\alpha$ -helices (D'Andrea and Regan, 2003). Seven to eight TPRs



**Fig. 1.** (A) Schematic of wild-type IgG (left) and IgG with a designed linker in its hinge region (middle). IgG domains are color coded as shown in the right panel. (B) Ribbon diagrams for domains used in structured linkers shown to scale (pdb codes: ZAG (1ZAG),  $\beta$ 2m (1LDS), cTPR (2F07), ubiquitin (1UBQ)). The cTPR structure shown contains eight tandem repeats. N- and C-terminal residues are shown as sticks, color-coded blue for the N-terminus and red for the C-terminus.

**Table I.** Description of structured linker designs

Linker	Name	Description
L1	GPcPcPc	GlySer-polyPro(Glyc)-polyPro(Glyc)-polyPro(Glyc)
L2	GPPcP	GlySer-polyPro-polyPro(Glyc)-polyPro
L3	GPGcP	GlySer-polyPro-GlySer(Glyc)-polyPro
L4	GPPP	GlySer-polyPro-polyPro-polyPro
L5	GPbP	GlySer-polyPro- $\beta$ 2m-polyPro
L6	GPbG	GlySer-polyPro- $\beta$ 2m-GlySer
L7	PbGbG	polyPro- $\beta$ 2m-GlySer- $\beta$ 2m-GlySer
L8	GPbGbP	GlySer-polyPro- $\beta$ 2m-GlySer- $\beta$ 2m-polyPro
L9	GPUG	GlySer-polyPro-Ub-GlySer
L10	GPZP	GlySer-polyPro-ZAG-polyPro
L11	GGZGZP	GlySer-GlySer-ZAG-GlySer-ZAG-polyPro
L12	GcGcP	GlySer(Glyc)-GlySer(Glyc)-polyPro
L13	cTPR3	(G <sub>4</sub> S) <sub>3</sub> -cTPR3-(G <sub>4</sub> S) <sub>3</sub>
L14	cTPR6	(G <sub>4</sub> S) <sub>3</sub> -cTPR6-(G <sub>4</sub> S) <sub>3</sub>
L15	cTPR9	(G <sub>4</sub> S) <sub>3</sub> -cTPR9-(G <sub>4</sub> S) <sub>3</sub>
L16	cTPR12	(G <sub>4</sub> S) <sub>3</sub> -cTPR12-(G <sub>4</sub> S) <sub>3</sub>
L17	GS1	(G <sub>4</sub> S) <sub>1</sub>
L18	GS2	(G <sub>4</sub> S) <sub>2</sub>
L19	GS3	(G <sub>4</sub> S) <sub>3</sub>
L20	GS5	(G <sub>4</sub> S) <sub>5</sub>
L21	GS6	(G <sub>4</sub> S) <sub>6</sub>
L22	GS7	(G <sub>4</sub> S) <sub>7</sub>
L23	GS8	(G <sub>4</sub> S) <sub>8</sub>
L24	GS9	(G <sub>4</sub> S) <sub>9</sub>

(Gly<sub>4</sub>Ser)<sub>n</sub>, Gly–Gly–Gly–Gly–Ser sequence with *n* number of repeats; GlySer, (N-term: AGS(GGS)<sub>3</sub>; Middle: (GGS)<sub>4</sub>; C-term: (GGS)<sub>3</sub>GAS)<sub>2</sub>S); GlySer(Glyc), Gly–Gly–Ser sequence with an embedded potential N-linked glycosylation site (Asn–Ser–Ser); polyPro, proline-rich hinge sequence from IgA1; polyPro(Glyc), proline-rich hinge sequence from IgA1 with an embedded potential N-linked glycosylation site (Asn–Ser–Ser);  $\beta$ 2m,  $\beta$ 2-microglobulin; Ub, ubiquitin; ZAG, Zn- $\alpha$ 2-glycoprotein; cTPRX, consensus tetratricopeptide repeat sequence with X number of repeats.

form a complete superhelical turn with a pitch of  $\sim 72$  Å. For our TPR linkers, we used a consensus sequence defined by the amino acid of the greatest global propensity in the natural database of the TPR domains at each position, which was shown to form a stable superhelix and was therefore named the consensus TPR sequence or cTPR (Main *et al.*, 2003).

Finally, for comparison, we constructed a series of (Gly<sub>4</sub>Ser)<sub>n</sub> linkers (L17–L24; Table I) in order to determine the effect of increasing the number of flexible Gly<sub>4</sub>Ser repeats on the hydrodynamic radius of the IgG. The complete sequence of each linker is given in Table II.

As a scaffold for comparing the designed structured linkers, we inserted each into the hinge region of an intact IgG antibody (the anti-HIV antibody b12) (Roben *et al.*, 1994). We chose the hinge region of an IgG, which encompasses the amino acids between the C-terminus of the heavy chain portion of the antigen-binding fragment (Fab) and the N-terminus of the Fc, to insert the linkers because it can tolerate large protein insertions (Redpath *et al.*, 1998). In addition, extension in the hinge region could potentially increase the separation distance of the Fab arms (Fig. 1B).

### Characterization of the IgGs containing structured linkers

The b12 IgG proteins containing linkers L1–L24 were expressed by transient transfection in HEK 293-6E mammalian cells and purified by affinity and size exclusion chromatography. Visualization by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) for IgGs containing the L1–L8 linkers showed that all proteins were purified to  $>95\%$

homogeneity (Fig. 2). Under reducing conditions, two heavy chain bands were observed for b12-L1, which contained a linker containing three potential N-linked glycosylation sites, indicating the presence of multiple glycosylated isoforms. An overlay of the chromatograms derived from SEC showed that the IgGs containing the L1–L8 structured linkers all exhibited a decrease in retention volume relative to wild-type IgG, consistent with the expected increases in the radius of gyration ( $R_g$ ) of each of the constructs due to the addition of a structured linker (Fig. 3).

We next derived the hydrodynamic radii using DLS for wild-type b12 and the b12 proteins containing designed linkers. DLS measures fluctuations in the intensity of scattered light of a protein solution over time, which can be used to calculate an autocorrelation function of intensity (Nobmann *et al.*, 2007). Typical monodisperse samples (including our hinge-linked antibodies) generate an exponential decay in the autocorrelation. A least squares fit can be performed to calculate the decay constant, which directly relates to the diffusion coefficient. The diffusion coefficient is then inversely related to the characteristic hydrodynamic radius  $R_H$ , which reflects the radius of a hypothetical solid sphere that would diffuse at the same rate as the protein. The  $R_H$  value is not a direct measurement of the length that the linker contributes to the size of the IgG. However, comparative analysis can yield rank order differences for the relative lengths and rigidity of the various linkers. For example, if the separation between the IgG Fc and Fab domains were increased by the addition of a designed hinge linker, we would expect an observable increase in the  $R_H$  of the fusion construct compared with the parental b12 IgG due to increased size of the diffusion sphere.

The hydrodynamic radii were measured by DLS for each of the b12 IgG-linker fusion proteins and compared with an internal wild-type b12 IgG control (Fig. 4). By comparing constructs containing elongated or small protein domain linkers, cTPR repeat linkers and flexible (Gly<sub>4</sub>Ser)<sub>n</sub> linkers of various lengths (L17–L24), we could directly compare the effects of incorporating different lengths of flexible vs. structured proteins linkers.

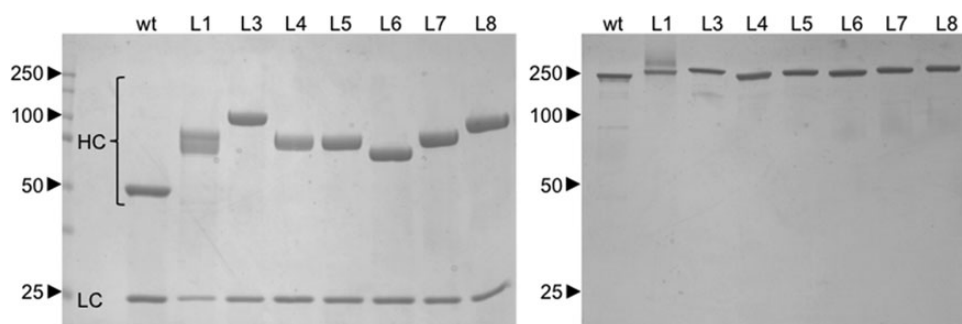
We observed a consistent trend for the  $R_H$  values between glycosylated and non-glycosylated linkers (L1, L2, L3 and L12 vs. L4). The incorporation of three potential N-linked glycosylation sites in proline-rich linkers derived from the hinge region of IgA1 (L1) appeared to increase the  $R_H$  relative to constructs containing similar linker sequences with only one (L2) or no (L4) N-linked glycosylation sites, possibly through stabilization of the folded state and leading the linker to adopt a more extended conformation (Shental-Bechor and Levy, 2008). While the addition of only a single potential N-linked glycosylation site did not seem to affect the diffusion rate of proline-rich linkers (compare L2 and L4), a single potential N-linked glycosylation in the GGSG–NSS–GSGG region of a combination proline-rich and Gly<sub>2</sub>Ser linker (L3) increased its  $R_H$  beyond the  $R_H$  of a proline-rich linker with three potential N-linked glycosylation sites (L1). These data are consistent with the observation that N-linked glycosylation confers rigidity in the backbone of a flexible linker (Liu *et al.*, 2000), suggesting these reagents contained linkers with a more extended conformation. Thus incorporating potential N-linked glycosylation sites within flexible linkers may be a general method to increase linker rigidity.

Adding a single  $\beta$ 2m domain to a linker increased the  $R_H$  of the b12-linker protein to a similar degree as a proline-rich repeat relative to IgG (compare L5 to L2, L4 and IgG),

**Table II.** Complete sequences of designed linkers

Linker	Name	Complete sequence
L1	GPcPcPc	AGSGGGSGGSGGSPVPSTPPTNSSTPPTPSPSPVPSTPPTNSSTPPTPSPSPVPSTPPTNSSTPPTPSPSAS
L2	GPPcP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSPVPSTPPTNSSTPPTPSPSPVPSTPPTPSPSTPPTPSPSAS
L3	GPcP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSGGSGNSSGGGSPVPSTPPTPSPSTPPTPSPSAS
L4	GPPP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSPVPSTPPTPSPSTPPTPSPSPVPSTPPTPSPSTPPTPSPSAS
L5	GPbP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSK DWSFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWDRDPVVPSTPPTPSPSTPPTPSPSAS
L6	GPbG	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFS KDW SFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWDRDGGSGGSGGSGSAS
L7	PbGbG	AGVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDW SFYLLYY TEFTPEKDEYACRVNHVTL SQPKIVKWDRDGGSGGSGGSGGSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNG ERIEKVEHSDLSFSKDW SFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWDRDGGSGGSGGSGSAS
L8	GPbGbP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSK DWSFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWDRDGGSGGSGGSGGSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSD IEVDLLKNGERIEKVEHSDLSFSKDW SFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWDRDPVVPSTPPTPSPSTPPTPSPSAS
L9	GPUG	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSK NIQKESTLHLVLRLLRGGGGSGGSGGSGSAS
L10	GPZP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSDGRYSLTYYTGLSKHVEDVPAFQALGSLNDLQFFRYNSKDRKSQPMGLWRQVE GMEDWKQDSQLQKAREDFMETLKDIVEYYNDSNGSHVLQGRFGCEIENNRSSGAFWKYYDGDYIEFNKEIPAWVPDPAQAQIT KQKWEAEPVYVQRAKAYLEEECPATLRKYLKYSKNILDRQDPPSVVVTSHQAPGEKKLKLCLAYDFYPGKIDVHWTRAGEVQE PELRGDVHLHNGNGTYQSWVVAVPPQDTAPYSCHVQHSSLAQPLVVPWEASVPVPSTPPTPSPSTPPTPSPSAS
L11	GGZGZP	AGSGGGSGGSGGSPVPSTPPTPSPSTPPTPSPSDGRYSLTYYTGLSKHVEDVPAFQALGSLNDLQFFRYNSKDRKSQPMGLWRQVE GMEDWKQDSQLQKAREDFMETLKDIVEYYNDSNGSHVLQGRFGCEIENNRSSGAFWKYYDGDYIEFNKEIPAWVPDPAQAQITKQKW EAEPVYVQRAKAYLEEECPATLRKYLKYSKNILDRQDPPSVVVTSHQAPGEKKLKLCLAYDFYPGKIDVHWTRAGEVQPELRGD VHLHNGNGTYQSWVVAVPPQDTAPYSCHVQHSSLAQPLVVPWEASGGSGGSGGSGGSDGRYSLTYYTGLSKHVEDVPAFQALG LNDLQFFRYNSKDRKSQPMGLWRQVEGMEDWKQDSQLQKAREDFMETLKDIVEYYNDSNGSHVLQGRFGCEIENNRSSGAFW KYYDGDYIEFNKEIPAWVPDPAQAQITKQKWEAEPVYVQRAKAYLEEECPATLRKYLKYSKNILDRQDPPSVVVTSHQAPGEK KKLKLCLAYDFYPGKIDVHWTRAGEVQPELRGDVHLHNGNGTYQSWVVAVPPQDTAPYSCHVQHSSLAQPLVVPWEASVPVPST PPTPSPSTPPTPSPSAS
L12	GcGcP	AGSGNSSGGSGGSGNSSGGGSPVPSTPPTPSPSTPPTPSPSAS
L13	cTPR3	KLGGGGSGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALEDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQRSAGGGGGSGGGGGGGGAS
L14	cTPR6	KLGGGGSGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALEDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAEAWKLNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNL GNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEDYQKALELDPNNRSRAGGGGGSGGGGGGGGAS
L15	cTPR9	KLGGGGSGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALEDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAEAWKLNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNL NAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEDYQKALELDPNNRSRAGGGGGSGGGGGGGGAS
L16	cTPR12	KLGGGGSGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALEDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNLQAEAWKLNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNL NAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEDYQKALELDPNNRSRAGGGGGSGGGGGGGGAS
L17	GS1	GGGGAS
L18	GS2	GGGGSGGGGAS
L19	GS3	GGGGSGGGGGSGGGGAS
L20	GS5	GGGGSGGGGGSGGGGGSGGGGAS
L21	GS6	GGGGSGGGGGSGGGGGSGGGGGSGGGGAS
L22	GS7	AGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGAS
L23	GS8	AGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGAS
L24	GS9	AGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGGSGGGGAS

Linkers L1–L21 were inserted into the hinge region of b12 IgG between residues His235 and Thr236. Linkers L22–L24 were inserted into the same hinge between residues Cys231 and Asp232.

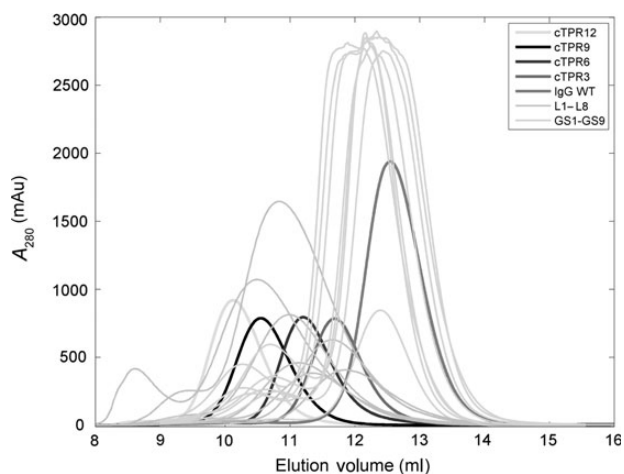


**Fig. 2.** SDS–PAGE analysis of b12 IgG-structured linker proteins run under reducing (left) and non-reducing (right) conditions.

suggesting that the structured  $\beta 2m$  domain provided similar bulk and separation to the polyPro repeat. However, adding a second tandem  $\beta 2m$  repeat separated from the first with a  $(Gly_2Ser)_4$  sequence did not increase the  $R_H$  appreciably (compare L8 and L5; L7 and L6). These results suggested that coupling a flexible Gly–Ser linker with the rigid  $\beta 2m$  domain partially diminished the separation between Fc and Fab regions provided by  $\beta 2m$  alone. A similar observation was made for hinge constructs containing ZAG (L10 and L11). A linker containing ZAG alone increased the hydrodynamic

radius of the b12-linker protein compared with IgG and more than the IgG–proline-rich linker (compare L10 to L2 and L4). However, replacement of the proline-rich domain by ZAG that was flanked at both termini by a  $(G_2S)_4$  peptide resulted in a decrease in hydrodynamic radius (compare L10 and L11).

We also investigated ubiquitin as a structured linker (L9). However, initial characterization by SDS–PAGE showed degradation at the linker site (data not shown). In addition, DLS measurements revealed that a purified sample of b12–L9 had a smaller  $R_H$  than IgG similar to a Fab or Fc region alone, further suggesting ubiquitin-specific degradation (Fig. 4).



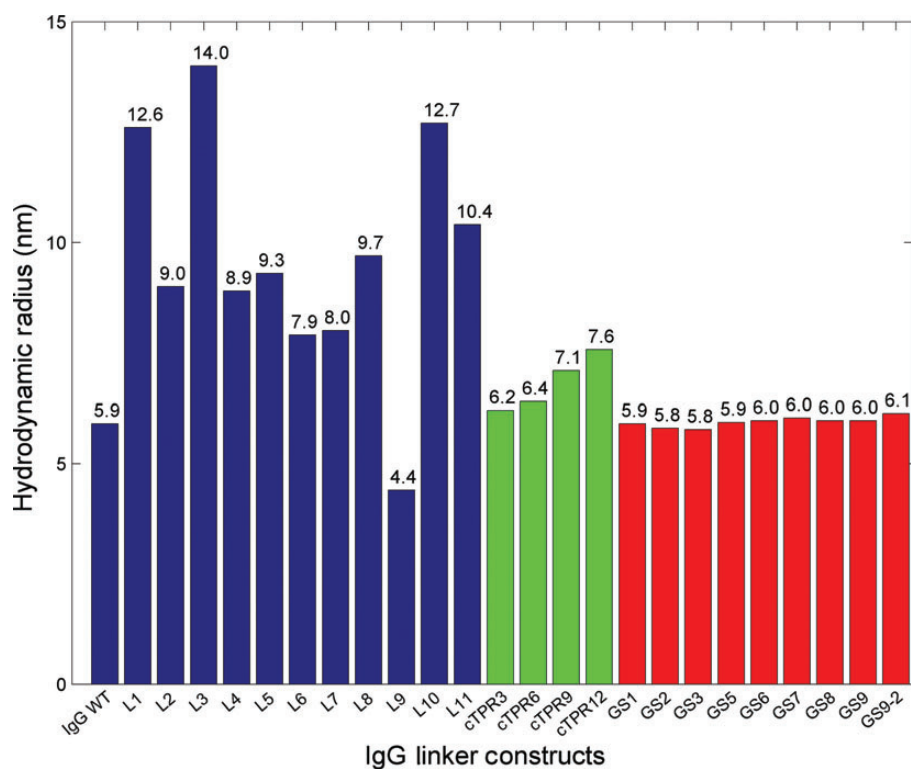
**Fig. 3.** Overlay of size-exclusion chromatograms for IgGs containing flexible and structured protein linkers. Structured linkers (L1–L8) exhibited larger decreases in retention volume with respect to wild-type compared with  $Gly_4Ser$  linkers, which exhibited little to no decrease depending on the number of repeats. Structured cTPR linkers also exhibited consistent decreases in retention volume as a function of the number of repeats.

### cTPR linker series

cTPR constructs were generated with 3, 6, 9 or 12 tandem repeats. All cTPR linkers were flanked by  $(Gly_4Ser)_3$  sequences (Table II). The constructs exhibited a consistent decrease in elution volume on SEC as a function of the repeat length (Fig. 3). These constructs also predictably increased the  $R_H$  of the linked IgG with increased number of tandem repeats (Fig. 4). The hydrodynamic radius of the cTPR12 construct corresponded to approximately the size of L4, which contained a proline-rich linker. These data suggested that, unlike with repeated domains of the structured linkers, the increase in separation between the Fab and Fc correlated predictably with the number of cTPR repeats despite the presence of  $Gly_4Ser$  peptides flanking the N- and C-termini.

### $(Gly_4ser)_n$ linker series

In order to compare our structured linkers to the typical unstructured Gly–Ser linkers commonly used in protein design and engineering, we constructed, expressed and purified eight IgG– $(G_4S)_n$  variants. In contrast to the SEC profiles for the structured linker constructs, there were only small differences in



**Fig. 4.** Comparative analysis by DLS of the hydrodynamic radii ( $R_H$ ) of designed linkers in the context of the b12 IgG.

elution volume for the IgGs including Gly<sub>4</sub>Ser linkers (L17–L24). These differences often did not correlate with molecular mass as IgG-GS9, the IgG with the largest linker, eluted at approximately the same volume as wild-type IgG, which eluted after some of the constructs with shorter linkers (Fig. 3).

Unlike proline-rich linkers and rigid linkers consisting of natural protein domains such as  $\beta$ 2m, Gly<sub>4</sub>Ser linkers that did not contain a potential N-linked glycosylation site did not detectably increase the hydrodynamic radius of the IgG, suggesting that these linkers did not provide increased separation between the Fab and Fc domains (Fig. 4). These results were consistent with the observation that Gly<sub>4</sub>Ser linkers did not provide significant separation between the joined domains in the context of other fusion proteins (Arai et al., 2001). Measurements of IgG-GS9 from two preparations showed only a slight difference in  $R_H$  (0.1 nm), indicating that these measurements were quite robust and relatively small differences in  $R_H$  may be significant.

Optimized linkers are important for the construction of multi-functional fusion proteins, in terms of both immunogenicity and conformational dynamics. Different linker compositions can alter their effective length and rigidity. In this study, we used SEC and DLS to characterize designed linkers in the context of an IgG to determine whether these linkers could increase the distance between the antigen-binding fragments. We found that flexible Gly<sub>4</sub>Ser linkers did not increase the  $R_H$  of fused reagents, suggesting these linkers did not provide increased separation between the Fab and Fc domains even with up to nine Gly<sub>4</sub>Ser repeats, in agreement with previous studies (Arai et al., 2001). By contrast, the structured helical cTPR linkers provided consistent increases in  $R_H$  and SEC elution volume as a function of repeat number, indicating that these repeats can be used to increase the separation distance between two proteins or domains. Our other designed linkers, including those containing naturally occurring proteins such as  $\beta$ 2m and ZAG, yielded increases in the observed  $R_H$  by as much as twice the  $R_H$  of a naturally occurring IgG. The systematic characterization of the lengths and rigidity properties of the structured protein linkers and a range of (Gly<sub>4</sub>Ser)<sub>n</sub> linkers reported here provide a new set of tools to the available linker repertoire for engineering fusion proteins.

## Acknowledgements

We thank Maria Politzer for assistance with sub-cloning and plasmid preparations and the Caltech Protein Expression Center for producing antibody-linker constructs.

## Funding

This work was supported by a grant from the Bill and Melinda Gates Foundation through the Grand Challenges in Global Health Initiative, the Director's Pioneer Award [1DP1OD006961-01 to P.J.B.] and the National Institutes of Health HIVRAD [P01AI100148 to P.J.B.]. Funding to pay the Open Access publication charges for this article was provided by NIDA/National Institutes of Health.

## References

Arai, R., Ueda, H., Kitayama, A., Kamiya, N. and Nagamune, T. (2001) *Protein Eng.*, **14**, 529–532.  
Argos, P. (1990) *J. Mol. Biol.*, **211**, 943–958.

Bai, Y., Ann, D.K. and Shen, W.-C. (2005) *Proc. Natl Acad. Sci. USA*, **102**, 7292–7296.  
Bonner, A., Furtado, P.B., Almogren, A., Kerr, M.A. and Perkins, S.J. (2008) *J. Immunol.*, **180**, 1008–1018.  
Chen, X., Zaro, J.L. and Shen, W.C. (2013) *Adv. Drug Deliv. Rev.*, **65**, 1357–1369.  
D'Andrea, L.D. and Regan, L. (2003) *Trends Biochem. Sci.*, **28**, 655–662.  
Diskin, R., Scheid, J.F., Marcovecchio, P.M., et al. (2011) *Science*, **334**, 1289–1293.  
Evers, T.H., van Dongen, E.M., Faesen, A.C., Meijer, E.W. and Merckx, M. (2006) *Biochemistry*, **45**, 13183–13192.  
George, R.A. and Heringa, J. (2002) *Protein Eng.*, **15**, 871–879.  
Imperiali, B. and O'Connor, S.E. (1999) *Curr. Opin. Chem. Biol.*, **3**, 643–649.  
Kajander, T., Cortajarena, A.L., Mochrie, S. and Regan, L. (2007) *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 800–811.  
Klein, J.S. and Bjorkman, P.J. (2010) *PLoS Pathog.*, **6**, e1000908.  
Klein, J.S., Gnanapragasam, P.N., Galimidi, R.P., Foglesong, C.P., West, A.P., Jr and Bjorkman, P.J. (2009) *Proc. Natl Acad. Sci. USA*, **106**, 7385–7390.  
Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J. and Tan, S. (2005) *Protein Expr. Purif.*, **41**, 98–105.  
Liu, H.L., Doleyres, Y., Coutinho, P.M., Ford, C. and Reilly, P.J. (2000) *Protein Eng.*, **13**, 655–659.  
Main, E.R., Jackson, S.E. and Regan, L. (2003) *Curr. Opin. Struct. Biol.*, **13**, 482–489.  
Nobmann, U., Connah, M., Fish, B., et al. (2007) *Biotechnol. Genet. Eng. Rev.*, **24**, 117–128.  
Pardridge, W.M. (2010) *J. Drug Target.*, **18**, 157–167.  
Redpath, S., Michaelsen, T.E., Sandle, I. and Clark, M.R. (1998) *Human Immunol.*, **59**, 720–727.  
Roben, P., Moore, J.P., Thali, M., Sodroski, J., Barbas, C.F., III and Burton, D.R. (1994) *J. Virol.*, **68**, 4821–4828.  
Sanchez, L.M., Chirino, A.J. and Bjorkman, P. (1999) *Science*, **283**, 1914–1919.  
Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U. and Moarefi, I. (2000) *Cell*, **101**, 199–210.  
Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M. and Eaton, W.A. (2005) *Proc. Natl Acad. Sci. USA*, **102**, 2754–2759.  
Shental-Bechor, D. and Levy, Y. (2008) *Proc. Natl Acad. Sci. USA*, **105**, 8256–8261.  
Trinh, C.H., Smith, D.P., Kalverda, A.P., Phillips, S.E. and Radford, S.E. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 9771–9776.  
Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.  
Zhang, J., Yun, J., Shang, Z., Zhang, X. and Pan, B. (2009) *Prog. Nat. Sci.*, **19**, 1197–1200.  
Zhu, P., Liu, J., Bess, J., Jr, Chertova, E., Lifson, J.D., Grise, H., Ofek, G.A., Taylor, K.A. and Roux, K.H. (2006) *Nature*, **441**, 847–852.  
Zou, Y., Weis, W.I. and Kobilka, B.K. (2012) *PLoS ONE*, **7**, e46039.