NIH Public Access
Author Manuscript
Adv Drug Deliv Rev. Author manuscript; available in PMC 2014 October 15.

Published in final edited form as:
Adv Drug Deliv Rev. 2013 October 15; 65(10): 1357–1369. doi:10.1016/j.addr.2012.09.039.

Fusion Protein Linkers: Property, Design and Functionality

Xiaoying Chen[1], **Jennica Zaro**[1], and **Wei-Chiang Shen**[1,2]
[1]Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, California 90089, USA.

Abstract

As an indispensable component of recombinant fusion proteins, linkers have shown increasing importance in the construction of stable, bioactive fusion proteins. This review covers the current knowledge of fusion protein linkers and summarizes examples for their design and application. The general properties of linkers derived from naturally-occurring multi-domain proteins can be considered as the foundation in linker design. Empirical linkers designed by researchers are generally classified into 3 categories according to their structures: flexible linkers, rigid linkers, and *in vivo* cleavable linkers. Besides the basic role in linking the functional domains together (as in flexible and rigid linkers) or releasing free functional domain *in vivo* (as in *in vivo* cleavable linkers), linkers may offer many other advantages for the production of fusion proteins, such as improving biological activity, increasing expression yield, and achieving desirable pharmacokinetic profiles.

Keywords

bifunctional recombinant proteins; spacer; pharmacokinetics; pharmacodynamics

1. Introduction

As a product of recombinant DNA technology, fusion proteins have been developed as a class of novel biomolecules with multi-functional properties. By genetically fusing two or more protein domains together, the fusion protein product may obtain many distinct functions derived from each of their component moieties. Besides their wide applications in biological research such as protein purification [1] and imaging [2], recombinant fusion proteins have also become an important category of biopharmaceuticals (Figure 1) [3, 4]. For example, many protein drugs are fused to Fc domains of antibodies, such as Fc-immunoglobulin G1 (Fc-IgG1), or to carrier proteins such as human serum albumin (HSA) or transferrin (Tf) to extend their plasma half-lives and to achieve enhanced therapeutic effects [5-8]. They have also been widely applied for drug targeting, since proteins such as single chain antibodies or ligands for cell surface receptors can specifically target a linked functional protein (e.g. toxin or cytokine) to a specific type of cells [9, 10]. In drug delivery, the combination of protein drugs to carrier moieties such as cell penetrating peptides, antibodies or Tf can achieve efficient transport of the protein drugs across biological barriers such as cell membranes, the blood brain barrier or intestinal epithelium [11-13]. Several

© 2012 Elsevier B.V. All rights reserved.

[2]To whom correspondence should be addressed. (weishen@usc.edu) .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

fusion proteins drugs including Enbrel® (tumor necrosis factor/Fc-IgG1), Ontak® (Interleukin-2/diphtheria toxin), Orencia® (Cytotoxic T-Lymphocyte Antigen-4/Fc-IgG1), Amevive® (Leukocyte function antigen-3/Fc-IgG1), Arcalyst® (Interleukin-1 Receptor extracellular domain/ Fc-IgG1), and Nplate® (thrombopoietin/Fc-IgG1) have been approved by the FDA [14-16]. With the rapid advancement of biotechnology, it is foreseeable that fusion protein technology will have increasing importance in creating novel protein therapeutics and in improving the performance of current protein drugs.

The successful construction of a recombinant fusion protein requires two indispensable elements: the component proteins and the linkers. The choice of the component proteins is based on the desired functions of the fusion protein product and, in most cases, is relatively straightforward. On the other hand, the selection of a suitable linker to join the protein domains together can be complicated and is often neglected in the design of fusion proteins. Direct fusion of functional domains without a linker may lead to many undesirable outcomes, including misfolding of the fusion proteins [17], low yield in protein production [18], or impaired bioactivity [19, 20]. Therefore, the selection or rational design of a linker to join fusion protein domains is an important, yet underexplored, area in recombinant fusion protein technology.

This review will summarize the current knowledge of linker design in recombinant fusion proteins. First, an overview of the properties of linkers in naturally-occurring multi-domain proteins will be provided as a general reference for linker design. Next, the empirical linkers that have been applied to the successful construction of recombinant fusion proteins will be discussed along with examples. Lastly, various functions that can be achieved by utilizing linkers in recombinant fusion proteins will be presented, including improving folding and stability, facilitating protein expression, increasing the intrinsic biological activities, enabling targeting toward specific sites *in vivo*, and altering the pharmacokinetic (PK) profiles of fusion proteins.

2. General properties of linkers derived from naturally-occurring multi-domain proteins

Similar to recombinant fusion proteins, naturally-occurring multi-domain proteins are composed of two or more functional domains joined by linker peptides. These linker peptides serve to connect the protein moieties, and also provide many other functions, such as maintaining cooperative inter-domain interactions [21] or preserving biological activity [22]. Knowledge of natural linkers in multi-domain proteins is very helpful for the rational design of empirical linkers in recombinant fusion proteins. Two independent studies performed by Argos [23] and George and Heringa [24] examined the general properties of linkers in natural proteins. The differences in their study designs are worth noting. First, the size of the two databases of natural linkers was very different. While only 51 examples of linker peptides were examined in Argos's study, the latter database contains 1280 linkers. For delineation of linkers in the protein sequences, the Argos study visually inspected computer generated molecular models of the proteins. Due to the large sample size, George and Heringa developed an automated method to extract linker sequences from a database of proteins with known 3D structures. Therefore, the slight variation in results between these studies may be attributed to the differences in the databases, or in the methods used to assign secondary structures.

Several properties of the natural linkers, such as length, hydrophobicity, amino acid residues, and secondary structure were compared and the results are summarized in Table 1. Briefly, **t**he average length of linkers in natural multi-domain proteins was calculated to be 6.5 residues by Argos [23], and 10.0 ± 5.8 residues by George and Heringa [24]. In the latter

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

study, the linkers were grouped into small, medium, and large linkers with average length of 4.5 ± 0.7, 9.1 ± 2.4, and 21.0 ± 7.6 residues, respectively. To give insight into the structural environment of the linkers, the average normalized solvent accessibility and hydrophobicity were also calculated. The data showed higher solvent accessibility with increasing length of linkers [24], suggesting that longer linkers were more likely to be exposed to the solvent. Consistent with these data, the average hydrophobicity of the linkers decreased with the increase of length, indicating that longer linkers were more hydrophilic and therefore more exposed in the aqueous solvent than shorter linkers [24].

The preference of amino acid residues in natural linkers was also investigated by calculating the ratio of single amino acid occurrence in the linker and the full protein (Table 1), where values greater than 1 (shaded) indicate higher occurrences in the linker. By screening the database, threonine (Thr), serine (Ser), proline (Pro), glycine (Gly), aspartic acid (Asp), lysine (Lys), glutamine (Gln), asparagine (Asn), and alanine (Ala) were suggested to be preferable linker constituents by Argos [23], whereas Pro, arginine (Arg), phenylalanine (Phe), Thr, glutamic acid (Glu) and Gln were preferred in the study by George and Heringa [24]. Therefore, in general, preferable amino acids were polar uncharged or charged residues, which constitute approximately 50% of naturally encoded amino acids. Both studies suggested that Pro, Thr, and Gln were the preferable amino acids for natural linkers. Among them, Pro is a unique amino acid with a cyclic side chain which causes a very restricted conformation [25]. The lack of amide hydrogen on Pro may prevent the formation of hydrogen bonds with other amino acids, and therefore reduces the interaction between the linkers and the protein domains. As a result, the inclusion of Pro residues might increase the stiffness and structural independence of the linkers. Many natural multi-domain proteins contain Pro-rich sequences as interdomain linkers, including the linker between the lipoyl and E3 binding domain in pyruvate dehydrogenase ($GA_2PA_3PAKQEA_3PAPA_2KAEAPA_3PA_2KA$) [26] and between the central and C-terminal domains in cysteine proteinase (P_9) [27]. The small, polar amino acids, such as Thr (both studies), or Ser and Gly (Argos's study), were thought to be favorable because they might provide good flexibility due to their small sizes, and also help maintain stability of the linker structure in the aqueous solvent through formation of hydrogen bonds with water.

Natural linkers adopt various conformations in secondary structure, such as helical, β-strand, coil/bend and turns, to exert their functions. From George and Heringa's analysis, most linkers on average exhibited α-helix (38.3%) or coil/bend (37.6%) secondary structures (Table 1) [24]. The conformations were slightly changed when small vs. large linkers were compared, where the majority of linkers adopted coils. The study by Argos also showed that the majority of the linkers adopted coil structures (59%). Based on From George and Heringa's secondary structure analysis, linkers were grouped into two categories: helical and non-helical. The α-helix was a rigid and stable structure, with intra-segment hydrogen bonds and a closely packed backbone [28]. Some α-helical conformations form rapidly during folding [28], allowing the correct folding of connecting protein domains without non-native interactions with the linker. Linkers in an α-helix structure might also serve as rigid spacers to effectively separate protein domains, and to reduce their unfavorable interactions. Therefore, this conformation was commonly adopted by many natural and empirical linkers (to be discussed later). On the other hand, without an inherent rigid structure, the non-helical linkers tended to be rich in Pro, which could increase the stiffness of the linker as mentioned previously [25]. As a result, non-helical linkers with Pro-rich sequence could exhibit relatively rigid structures and serve to reduce inter-domain interference.

Overall, natural linkers mainly adopted extended conformations, and had independent structures that did not interact with the adjacent protein domains. Taken together, their length, composition, hydrophobicity, and secondary structure were all important to achieve

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the desirable functions. The natural linkers could serve as general reference for the rational design of empirical linkers in recombinant fusion proteins.

3. Empirical linkers in recombinant fusion proteins

The studies of linkers in natural multi-domain proteins have generated many candidates for the general purpose of protein fusion [23, 25]. In addition, researchers have designed many empirical linkers with various sequences and conformations for the construction of recombinant fusion proteins. In the following section, three types of empirical linkers (flexible linkers, rigid linkers, and cleavable linkers) will be discussed to illustrate their versatile applications for fusion protein construction (Table 3).

3.1 Flexible linkers

Flexible linkers are usually applied when the joined domains require a certain degree of movement or interaction. They are generally composed of small, non-polar (e.g. Gly) or polar (e.g. Ser or Thr) amino acids as suggested by Argos [23]. The small size of these amino acids provides flexibility, and allows for mobility of the connecting functional domains. The incorporation of Ser or Thr can maintain the stability of the linker in aqueous solutions by forming hydrogen bonds with the water molecules, and therefore reduces the unfavorable interaction between the linker and the protein moieties.

The most commonly used flexible linkers have sequences consisting primarily of stretches of Gly and Ser residues (“GS” linker). An example of the most widely used flexible linker has the sequence of $(Gly\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Ser)_n$. By adjusting the copy number “n”, the length of this GS linker can be optimized to achieve appropriate separation of the functional domains, or to maintain necessary inter-domain interactions. Besides the GS linkers, many other flexible linkers have been designed for recombinant fusion proteins. As suggested by Argos [23], these flexible linkers are also rich in small or polar amino acids such as Gly and Ser, but can contain additional amino acids such as Thr and Ala to maintain flexibility, as well as polar amino acids such as Lys and Glu to improve solubility.

Several other types of flexible linkers, including KESGSVSSEQLAQFRSLD and EGKSSGSGSESKST, have been applied for the construction of a bioactive scFv [29]. The Gly and Ser residues in the linker were designed to provide flexibility, whereas Glu and Lys were added to improve the solubility. These linkers were designed by computation methods and computational graphics. Specific amino acids, one near the carboxyl terminus of the antibody light-chain variable region (V_L) and one near the amino terminus of the heavy-chain variable region (V_H) domain, were first selected. Computation methods were then performed to search for suitable peptides from libraries of three-dimensional peptide structures derived from the Brookhaven Protein Data Bank (PDB) to span the distance in space between the selected amino acids. Alternatively, the linker peptide was built using computational design by adding single amino acids or short peptides extending from the carboxyl terminus of the V_L domain to the amino terminus of the V_H domain.

Another flexible linker, $(Gly)_8$, consisting of purely glycine was designed by Sabourin et al. to construct a Myc epitope-tagged Est2p fusion protein for easy protein detection [30]. The choice of the $(Gly)_8$ linker was based on the reports that flexible linkers can increase the accessibility of an epitope to antibodies or to improve protein folding. It was demonstrated that the $(Gly)_8$ linker is stable against proteolytic enzymes digestion during protein purification from the expression organism (yeast). This flexible linker improved the *in vivo* function of several epitope-tagged proteins involved in telomere maintenance. Similarly, a shorter $(Gly)_6$ linker was applied in the construction of human serum albumin-atrial natriuretic factor (ANF) fusion protein and maintained bioactivity of ANF [31].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Another Gly and Ser rich flexible linker, GSAGSAAGSGEF, was designed by Waldo et al. to express green fluorescent protein (GFP)-fusion proteins for rapid protein-folding assay [32]. The linker sequence avoided large hydrophobic residues to maintain good solubility in aqueous solutions. This linker provided similar performance of the GFP folding reporter as a longer $(GGGGS)_4$ linker. One advantage of this linker over the $(GGGGS)_4$ linker is that it did not have high homologous repeats in its DNA coding sequence. Therefore, it was less likely to be deleted by homologous recombination during the shuffling protocol for cloning.

In summary, flexible linkers are generally rich in small or polar amino acids such as Gly and Ser to provide good flexibility and solubility. They are suitable choices when certain movements or interactions (e.g. scFv) are required for fusion protein domains. In addition, although flexible linkers do not have rigid structures, they can serve as a passive linker to keep a distance between functional domains. The length of the flexible linkers can be adjusted to allow for proper folding or to achieve optimal biological activity of the fusion proteins.

3.2 Rigid linkers

While flexible linkers have the advantage to connect the functional domains passively and permitting certain degree of movements, the lack of rigidity of these linkers can be a limitation. There are several examples in the literature where the use of flexible linkers resulted in poor expression yields or loss of biological activity. For instance, a Tf-granulocyte colony stimulating factor (G-CSF) fusion protein failed to be expressed with a flexible $(GGGGS)_3$ linker [18]. In another report, the immunoglobulin binding ability of the protein G domain in a protein G-Vargula luciferase fusion protein was not recovered after inserting a flexible GGGGS linker [33]. The ineffectiveness of flexible linkers in these instances was attributed to an inefficient separation of the protein domains or insufficient reduction of their interference with each other. Under these situations, rigid linkers have been successfully applied to keep a fixed distance between the domains and to maintain their independent functions.

Alpha helix-forming linkers with the sequence of $(EAAAK)_n$ have been applied to the construction of many recombinant fusion proteins [18, 20]. As suggested by George and Heringa [24], many natural linkers exhibited α-helical structures. The α-helical structure was rigid and stable, with intra-segment hydrogen bonds and a closely packed backbone [28]. Therefore, the stiff α-helical linkers may act as rigid spacers between protein domains.

An empirical rigid linker with the sequence of $A(EAAAK)_nA$ (n = 2-5) was first designed by Arai et al. [34, 35]. The linker displayed α-helical conformation, which was stabilized by the Glu^- -Lys^+ salt bridges within segments. To test whether they could effectively separate the protein domains, these helical linkers were inserted between enhanced blue fluorescent protein (EBFP) and enhanced green fluorescent protein (EGFP), and the fluorescent resonance energy transfer (FRET) efficiency between EBFP and EGFP was measured [34]. The FRET efficiency decreased as the length of helical peptides increased, indicating that helical linkers can control the distance between domains by changing repetitions of the EAAAK motif. Compared to flexible linkers with the same length, the helical linkers induced much less FRET efficiency when inserted into EBFP-EGFP fusion proteins, suggesting that helical linkers can separate functional domains more effectively.

Another type of rigid linkers has a Pro-rich sequence, $(XP)_n$, with X designating any amino acid, preferably Ala, Lys, or Glu. As suggested by George and Heringa [24], the presence of Pro in non-helical linkers can increase the stiffness, and allows for effective separation of the protein domains. The structure of proline-rich sequences was extensively investigated by several groups [26, 36, 37]. For example, ^{1}H-NMR spectroscopy was conducted to elucidate

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the conformation of the $(Ala\text{-}Pro)_7$ dipeptide repeat in the N-terminal alkali light chain of skeletal muscle [36]. This sequence exhibited an extended and rigid conformation, probably due to the high frequency of Pro, which imposes strong conformational constrain [38]. In another study of 33-residue peptides containing repeating -Glu-Pro- or -Lys-Pro-also suggested that the X-Pro backbone displayed a relatively elongated and stiff conformation [37].

In conclusion, rigid linkers exhibit relatively stiff structures by adopting α-helical structures or by containing multiple Pro residues. Under many circumstances, they separate the functional domains more efficiently than the flexible linkers. The length of the linkers can be easily adjusted by changing the copy number to achieve an optimal distance between domains. As a result, rigid linkers are chosen when the spatial separation of the domains is critical to preserve the stability or bioactivity of the fusion proteins.

3.3 *In Vivo* cleavable linkers

The linkers discussed so far generally consist of stable peptide sequences that will not be preferentially cleaved *in vivo*. These stable linkers covalently join functional domains together to act as one molecule throughout the *in vivo* processes. The stable linkage between functional domains provides many advantages such as a prolonged plasma half-life (e.g. albumin or Fc-fusions). However, it also has several potential drawbacks including steric hindrance between functional domains, decreased bioactivity, and altered biodistribution and metabolism of the protein moieties due to the interference between domains [19, 39, 40]. Under these circumstances, cleavable linkers are introduced to release free functional domains *in vivo*. The design of *in vivo* cleavable linker in recombinant fusion proteins is quite challenging. Unlike the versatility of crosslinking agents available for chemical conjugation methods, linkers in recombinant fusion proteins are required to be oligopeptides. The linkers introduced in this section take advantage of the unique *in vivo* processes, and are cleaved under specific conditions such as the presence of reducing reagents or proteases. This type of linker may reduce steric hindrance, improve bioactivity, or achieve independent actions/metabolism of individual domains of recombinant fusion proteins after linker cleavage.

One well-studied *in vivo* process is the reduction of disulfide bonds, which has been widely applied in drug delivery by chemical conjugation methods [41]. Utilizing the reversible nature of the disulfide bond, an *in vivo* cleavable disulfide linker was designed for recombinant fusion proteins by Chen et al. [39], and offered the advantage of generating a precisely constructed, homogeneous product by recombinant methods. This disulfide linker (LEAGCKNFFPR↓SFTSCGSLE) was based on a dithiocyclopeptide containing an intramolecular disulfide bond formed between two cysteine (Cys) residues on the linker, as well as a thrombin-sensitive sequence (PRS) between the two Cys residues (Figure 2). This linker was inserted between G-CSF and Tf to construct a model fusion protein (designated as “G-C-T”). The *in vitro* thrombin treatment of G-C-T resulted in the cleavage of the thrombin-sensitive sequence, while the reversible disulfide linkage between the two domains of the fusion protein remained. The resultant disulfide-linked protein was designated as “G-S-S-T”. This disulfide-linked fusion protein was demonstrated to be cleavable *in vivo* following intravenous administration to CF1 mice. A rapid release of G-CSF from G-S-S-T in the blood was observed as early as 5 minutes, with a peak at ~15 minutes post injection. The released free G-CSF exhibited a quick elimination due to its short *in vivo* half-life. In contrast, no detectable amount of free G-CSF was released *in vivo* from G-C-T, which has a stable peptide linker. Therefore, this study demonstrated the construction of a disulfide-linked fusion protein for use in applications where the *in vivo* separation of protein domains is desired.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

More recently, a similar cyclopeptide linker was designed to create an *in vivo* cleavable disulfide linker in an interferon-α2b (IFN-α2b) and HSA fusion protein [42]. The dithiocyclopeptide sequence (CRRRRRREAEAC) contains an intramolecular disulfide bond between 2 Cys residues, as well as a peptide sequence sensitive to the secretion signal processing proteases resident in the yeast secretory pathway. During the protein expression, the linker was first cleaved by protease Kex2 at CRRRRRR↓EAEAC, followed by cleavage of proteases Kex1 and Ste13. As a result, the amino acids between two Cys residues in the linker were completely removed during secretion, and the disulfide linked fusion protein was directly expressed from *Pichia pastoris*.

Besides the reduction of disulfide bond, the *in vivo* cleavage of the linkers in recombinant fusion proteins may also be carried out by proteases that are expressed *in vivo* under pathological conditions (e.g. cancer or inflammation), in specific cells or tissues, or constrained within certain cellular compartments (Figure 2). Such *in vivo* cleavable linkers are designed to be liable to a particular protease by incorporating specific protease-sensitive sequences. Unlike the reduction of disulfide bond which happens rapidly in the blood circulation [39], the specificity of many proteases offers slower cleavage of the linker at constrained compartments. As will be further discussed in Section 5.4, this type of cleavable linker can therefore be applied to target activation of fusion protein bioactivity at specific sites *in vivo*.

3.4 Summary of empirical linkers

In summary, linkers can adopt various structures and exert diverse functions to fulfill the application of fusion proteins (Table 2). The flexible linkers are often rich in small or hydrophilic amino acids such as Gly or Ser to provide the structural flexibility and have been applied to connect functional domains that favor interdomain interactions or movements. In cases where sufficient separation of protein domains is required, rigid linkers may be preferable. By adopting α-helical structures or incorporating Pro, the rigid linkers can efficiently keep protein moieties at a distance. Both flexible and rigid linkers are stable *in vivo*, and do not allow the separation of joined proteins. Cleavable linkers, on the other hand, permit the release of free functional domain *in vivo* via reduction or proteolytic cleavage. They can be utilized to improve the bioactivity of chimeric proteins, or to specifically deliver prodrugs to target sites where the linkers are processed to activate bioactivity. The rational choice of linkers should be based on the properties of the linkers and the desired fusion proteins.

4. Linker designing tools and databases

The extensive studies about linkers in natural multi-domain proteins and recombinant fusion proteins fostered the idea of building databases and coming up with linker designing tools to aid the rational design of linkers based on the desired characteristics of fusion proteins.

An example of this type of tool is a program called LINKER, which searches its database of linker sequences with user-specified inputs (e.g., linker length, protease sensitive sequences to be avoided), and generates an output of several linker sequences that fit the criteria [43]. The linker database was built based on the assumption that the observed loop sequences in the X-ray crystal structures or the NMR solution structures are likely to adopt an extended conformation as linkers in the fusion protein. The final loop database contains 14,734 sequences that fit both the authors' records in the PDB files, and the searching criteria by the DSSP program which automatically identifies secondary and loop structures of proteins [44]. The basic input to the program is the desired length of the linker sequence, expressed as either the number of residues, or a distance in angstroms. Additional input parameters include potential cleavage sites for proteases or restriction endonucleases to avoid, so that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the selected linkers would be stable against the specified protease or be resistant to the restriction enzymes during cloning process. The users can also include amino acid composition preferences (e.g., eliminate bulky hydrophobic residues) to further select their linkers of interest. This program could serve as a convenient tool to generate linker sequences, which are observed loops in X-ray crystal or NMR solution structures. As a result, these loop sequences are likely to exhibit extended conformation, and serve as linkers to effectively construct fusion protein.

Another web-based program run by the Centre for Integrative Bioinformatics VU (IBIVU) at Vrije University of Amsterdam (http://www.ibi.vu.nl/programs/linkerdbwww/) also provides a database containing linkers with various confirmations and provides a search engine. The search algorithm accepts several query types (eg, PDB code, PDB header, linker length, C-alpha extent, solvent accessibility, secondary structure or sequence). The program can provide the linkers sequences meeting the searching criteria, and also provide other information such as the PDB code and a brief description of the source protein, linker's position within the source protein, linker length, solvent accessibility, and secondary structure. Users can search for sequences with desired properties, and obtain candidate sequences from natural multi-domain proteins.

5. Functionality of linkers in fusion proteins

The most basic function of linkers in recombinant fusion proteins is to covalently join the functional domains (e.g. flexible linkers or rigid linkers) or to release them under desired conditions (cleavable linkers). Linkers can also provide many derived functions in protein drug design such as improving biological activities, increasing production, achieving a controlled or targeted drug delivery, as well as achieving desirable PK profiles of the fusion proteins (Tables 1 and 2).

5.1 Linkers can improve folding and stability of fusion proteins

The flexible GS linker has been shown to improve folding and stability in several fusion protein examples. A first and very important application of the flexible GS linker is the construction of single-chain variable fragment (scFv), an antigen-binding fusion protein composed of antibody light-chain variable region (V_L) tethered to heavy chain variable region (V_H) via an oligopeptide linker [45]. A flexible linker $(GGGGS)_3$ was designed by Huston et al. to construct a scFv, since its flexible structure could allow for the correct orientation of the V_H and V_L domains, and would not interfere with the folding of the protein domains [46]. The length of the linker was adjusted according to the distance between the C-terminus of the V_H domain and the N-terminus of the V_L domain under its natural orientation (3.5 nm). The length of the $(GGGGS)_3$ linker was calculated to be about 5.7 nm, and was expected to bridge the V_H and V_L domains [46]. This $(GGGGS)_3$ linker was demonstrated to be suitable for constructing scFv due to its high flexibility, and has since been applied to many other scFvs [47-49].

Another type of flexible linker, $(Gly)_n$, has also been shown to improve folding of fusion proteins [30]. The insertion of the $(Gly)_8$ linker between a Myc epitope tag and the protein of interest (Est2p) greatly improved the functionality of epitope-tagged Est2p. The improvement was attributed to the correct folding of the epitope-tagged protein after linker insertion, as well as the reduced steric hindrance between functional domains.

In addition to flexible linkers, helical linkers can also improve fusion protein folding and stability. For instance, in the development of a virus coat protein-fusion protein, only peptides shorter than 20 amino acids could be expressed onto the coat protein of plant rod-shape viruses without preventing the assembly of functional virions [50]. Dramatically, with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the aid of a 15-amino acid linker $(EAAAK)_3$, a protein A fragment of 133 amino acids could be displayed onto the viral coat protein and the resultant fusion proteins could form functional virions [51]. By providing enough distance between protein A and the coat protein, the rigid helical linker permitted correct folding of the fusion protein, and allowed for the correct assembly of virus particles.

The length and structure of the linkers can also control the distance between functional domains and affect the stability of the fusion protein [34]. The α-helical linkers $(EAAAK)_n$ (n=1-3) as well as a flexible linker were applied to construct bifunctional fusions of β-glucanase and xylanase enzymes. The stability and catalytic activity of the resultant fusion proteins were significantly improved after linker insertion [52]. In addition, the thermal stability of β-glucanase in the fusion proteins was improved as the length of the helical linkers (controlled by copy number "n") increased. Compared to flexible linkers, helical linkers appeared to be more efficient in increasing the thermal stability. These improvements were likely due to the rigid structure of the α-helical linker that might provide enough space for protein domains to fold and function independently. These studies indicated that the insertion of a linker with proper structure and length could ultimately improve the stability and bioactivity of functional moieties.

5.2 Linkers can improve expression of fusion proteins

Besides impaired biological activity, the difficulty to express stable and high levels of recombinant fusion protein is often another hurdle during the application of fusion proteins for drug delivery. Due to the structural perturbation between protein domains, fusion proteins may be misfolded, unstable and appear as a heterogeneous product [17, 20], often resulting in a low expression yield. Although the expression of fusion proteins can sometimes be improved by simply switching the orientation of the component protein domains [53], the interference may not be effectively reduced due to the short distance between domains. Since many linkers can keep domains at proper distance and allow for their independent folding, they can serve as practical tools to enhance the expression yield of recombinant fusion proteins.

The effect of linker insertion on expression level of fusion proteins was observed in Tf-fusion proteins designed for Tf receptor-mediated protein drug oral delivery [18]. Fusion proteins consisting of Tf and human growth hormone (hGH) were constructed in two directions (hGH-Tf and Tf-hGH) to test the optimal orientation. A helical $(H4)_2$ linker $(A(EAAAK)_4ALEA(EAAAK)_4A)$ was then inserted into the two fusion proteins (designated as hGH-$(H4)_2$-Tf and Tf-$(H4)_2$-hGH), greatly improving their expression level in transiently-transfected HEK293 cells. The hGH-$(H4)_2$-Tf fusion proteins exhibited a 1.66-fold higher expression than hGH-Tf, while Tf-$(H4)_2$-hGH had an expression level 2.39-fold higher than that of Tf-hGH.

A more profound effect of the helical $(H4)_2$ linker insertion on protein expression level was observed in G-CSF-Tf-fusion proteins, G-CSF-Tf and Tf-G-CSF. With the aid of $(H4)_2$ linker, the production of G-CSF-$(H4)_2$-Tf was about 1.44-fold higher than that of G-CSF-Tf. On the other hand, expression of Tf-G-CSF, which had the reversed orientation, was almost undetectable, even with the insertion of a flexible linker $(GGGGS)_3$. With the $(H4)_2$ linker, high level of Tf-$(H4)_2$-G-CSF was expressed, with an 11.2-fold higher production compared to Tf-G-CSF. It was shown that the enhancement on the expression was sequence- and structure-specific, since inserting a linker encoded with the reversed oligonucleotides of $(H4)_2$ (which produces an identical peptide length without a helical structure) failed to facilitate the protein expression.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

These studies suggested that linker insertion may be a feasible approach to improve expression level of fusion proteins. The exact mechanism for the ability of the helical linker to improve expression of fusion proteins is still not very clear. The rigid, extended structure of the helical linker may effectively separate the functional domains and greatly reduce their interference. As a result, the fusion protein may be able to fold properly and obtain enhanced stability within the endoplasmic reticulum, resulting in an increased expression level. With further studies, more linkers may be identified for enhancing protein expression.

5.3 Linkers can improve bioactivity of fusion proteins

By fusing two or more protein domains, a fusion protein usually obtains the biological activities from each component. However, the direct fusion of proteins often results in impaired biological activity [20, 54, 55], probably because the functional domains are brought too close to properly interact with their corresponding binding proteins (i.e. receptors or ligands). Under these circumstances, linkers may be very effective tools to provide appropriate distance between domains to reduce their interference, restore or improve folding, or allow for the *in vivo* release of the free protein drug domain to ultimately improve bioactivity (Figure 3).

One example of applying linkers for bioactivity improvement is the development of Tf fusion proteins for the oral delivery of protein drugs via Tf receptor-mediated transcytosis across intestinal epithelium cells [13, 19, 20, 40, 56]. A direct fusion of G-CSF with Tf was constructed for the oral delivery of G-CSF [19]. The fusion protein exhibited suboptimal G-CSF bioactivity, retaining less than 10% of the bioactivity of parent G-CSF. To further improve the bioactivity, helical peptide linkers, $[A(EAAAK)_nA]_m$ (n = 2-4, m = 1 or 2), or a flexible linker, $(GGGGS)_3$, were inserted between the two domains of G-CSF-Tf fusion protein [20]. All of the resultant fusion proteins exhibited improved *in vitro* G-CSF biological activities. Particularly, the one with the helical $(H4)_2$ linker $(A(EAAAK)_4ALEA(EAAAK)_4A)$ exhibited the highest *in vitro* G-CSF bioactivity, with approximately 10-fold higher activity than the fusion protein without a linker. G-CSF-$(H4)_2$-Tf also exhibited much higher *in vivo* efficacy than G-CSF-Tf in animal models. A similar enhancement of the bioactivity was achieved with linker insertion in an hGH-Tf fusion protein, where the *in vitro* hGH bioactivity of hGH-Tf was improved with the $(H4)_2$ linker [40]. More remarkably, in animal experiments, orally administered hGH-$(H4)_2$-Tf fusion protein elicited a significant body weight gain in hypophysectomized rats (the biological response endpoint of hGH), while hGH-Tf failed to display a pharmacological effect [40].

In another study by Zhao et al., linker engineering greatly facilitated the development of bioactive and stable fusion proteins consisting of HSA and IFN-α2b. Albumin fusion proteins are widely used for prolonging the plasma half-life of protein drugs and enhancing *in vivo* efficacy [6, 57, 58]. Since the direct fusion of HSA and IFN-α2b caused diminished antiviral activity of IFN-α2b domain [53], three types of linkers including a flexible linker (GGGGS), a rigid Pro-rich linker (PAPAP) and a rigid helical linker (AEAAAKEAAAKA) [17] were investigated as spacers to reduce interdomain interference. With linker insertion, the anti-viral activities of 3 fusion proteins were dramatically increased by 39% (flexible linker), 68% (Pro-rich linker) and 115% (helical linker) compared to the fusion protein without any linker. The linker insertion also effectively facilitated the correct folding of the fusion proteins. Without a linker, a disulfide bond between two cysteine residues in IFN-α2b could not form due to the interference from HSA. With the aid of the 3 linkers, disulfide formation was effectively restored, and all fusion proteins appeared as a homogenous product on non-reducing SDS–PAGE. The enhanced bioactivity was likely attributed to the correct folding of the fusion protein and the proper separation of the protein domains after linker insertion.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The bioactivity of fusion proteins can also be improved by adjusting length of linkers to increase the space between fusion proteins. Bergeron et al. reported the construction of enzyme-chaperone chimeric protein as a new approach to stabilize an enzyme [55]. A fusion protein consisting of an enzyme (the model penicillin amidase, or PGA) and a chaperon (rTHS) was used as a model. Flexible linkers $(GGGGS)_n$ with different copy numbers (n = 1, 2, or 4) were inserted to test the optimal distance. The chimeric protein exhibited higher bioactivity as the length of the linker increased. The maximal activity was achieved with the longest linker, and was almost equivalent to that of the parent enzyme. By inserting flexible linkers with suitable length, the protein domains could be kept at a favorable distance, and the bioactivity of the fusion protein could be optimized.

The length of the rigid linkers can also have a major impact on protein bioactivity. Rigid peptide linkers $(Ala\text{-}Pro)_n$ (10 – 34 aa) were applied in an interferon-γ–gp120 fusion protein [54]. With a short 10-aa linker, the fusion protein possessed a relatively low biological activity of interferon-γ. By increasing the linker length, the bioactivity of the fusion protein was gradually improved, peaking at 88% activity of free interferon-γ with the longest 34-residue linker. The enhancement of bioactivity was sequence specific, since a $(Cys\text{-}Trp)_n$ linker with the same length was not able to improve the biological activity.

In some cases, even with the insertion of flexible or rigid linkers, the impaired bioactivity can still not be overcome due to steric hindrance between domains [19, 40]. With the aid of cleavable linkers, the steric hindrance may be effectively reduced after the release of free protein domains. GCSF-Tf fusion protein with a cleavable disulfide linker was reduced *in vitro* by dithiothreitol to mimic the expected *in vivo* released free G-CSF [39]. Following linker cleavage, the fusion protein exhibited a 2-fold increase of *in vitro* G-CSF bioactivity, and a much higher maximal response compared to the intact, covalently-linked fusion protein. The improved G-CSF bioactivity might be due to the reduced G-CSF receptor blockage after the release of G-CSF domain from Tf.

Recently, both the disulfide linkers (SS) and the protease-sensitive linkers (RKRR or RR) were applied in an IFN-α2b-HSA fusion protein for improving the *in vivo* efficacy [42]. The PK and pharmacodynamic (PD) properties of the resulting fusion proteins were compared against the IFN-α2b-HSA protein without a linker. The area under the concentration curve of IFN-α2b-HSA with cleavable linkers were generally lower than that of the nonreleasable fusion protein, consistent with the release of free IFN-α2b which has a short half-life. However, despite of the lower exposure, the PD properties were greatly improved, with an increase in the area under the anti-viral activity of 450%, 25% and 47% for IFN-RKRR-HSA, IFN-RR-HSA and IFN-SS-HSA, respectively. This result indicate that cleavable linkers may achieve higher *in vivo* efficacy compared to noncleavable linkers by alleviating the interference between domains and improving the intrinsic bioactivity of the fusion proteins.

These studies clearly demonstrate the capability of linker technology to improve suboptimal intrinsic activity of fusion proteins that are caused by insufficient separation or incorrect folding of functional domains. When the interference or steric hindrance between protein domains is greatly reduced, the bioactivity of fusion proteins may be greatly enhanced.

5.4. Linkers can target fusion proteins to specific sites *in vivo*

Linker insertion between fusion protein domains can also improve or enable targeting of fusion protein to specific sites *in vivo.* One way in which linkers can improve targeting is simply by increasing the binding affinity of the targeting protein domain for its receptor. This concept is very similar what has already been discussed in Section 5.3, where linkers can provide distance between domains, reduce their interference, and ultimately improve

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

their receptor binding affinity. A second approach for application of linkers to improve drug targeting involves introduction of a linker sequence that will enable specific activation of the fusion protein at the target site. In this approach, the intact fusion protein shows reduced or a lack of biological activity, but the cleavage of the linker at specific sites releases the free, biologically active protein drug domain at the target site (Figure 4). The focus of this section will be on the latter approach in using linkers for drug targeting.

The *in vivo* cleavable linkers sensitive to proteases that become active under certain physiological or pathological conditions are found to be ideal for drug targeting. For example, Schulte constructed a recombinant coagulation factor IX (FIX)-albumin fusion protein for the treatment of hemophilia B (Figure 4A) [59]. The fusion of albumin with FIX was intended to prolong the half-life of FIX and to improve its *in vivo* efficacy. With the insertion of a flexible GS linker, the fusion protein exhibited a poor FIX activity, probably because albumin blocked the interaction of FIX with other coagulation factors (e.g Factors VIII, Factor FX). To reduce the interference from albumin, a proteolytically-cleavable sequence (VSQTSKLT↓RAETVFPDV) derived from the N-terminal activation region of FIX was applied as a linker. Taking advantage of the activation process of FIX, the activation sites on FIX and the linker were simultaneously cleaved during clotting by either tissue factor/Factor VIIa or Factor XIa. Upon the cleavage of this linker, the clotting activity of the fusion protein was dramatically enhanced by 10- to 30-fold. At the meantime, since the linker was only cleaved during clotting, FIX was maintained in the inactive fusion protein form before linker cleavage, and exhibited prolonged half-life compared to FIX.

Similarly, *in vivo* cleavable linkers that are cleaved by proteases overexpressed at disease sites can also be utilized for drug targeting. For instance, matrix metalloproteinases (MMPs) are overexpressed during a variety of pathological conditions such as arthritic diseases [60], cancer [61] and inflammation [62, 63]. Linkers containing MMP cleavage sequences were investigated for the construction of several anti-inflammatory proteins, IFN-β, vasoactive intestinal peptide, and α- and γ-melanocyte-stimulating hormones, fused to latency-associated peptide (LAP) of transforming growth factor β1 [64-66]. The anti-inflammatory proteins are covered by the shell structure provided by LAP, and remain latent until cleavage of the linkers by MMPs at the disease site. For example, linkers consisting of two flexible GGGGS sequences flanking a MMP 1/9 cleavage sequence, PLGLWA, were inserted between IFN-β and LAP, allowing the release of free IFN-β at disease sites overexpressing MMPs (Figure 4B) [64]. Since the cleavage of the linker was limited, IFN-β could be masked by LAP and acted as a latent cytokine in the circulation. As a result, LAP-IFN-β fusion protein exhibited a 37-fold longer plasma half-life (55 hours) than native IFN-β due to the shielding of the latent cytokine from its cellular receptors until release from the LAP [67].

Another recent example of application of *in vivo* cleavable linkers involves targeting proteases that are specifically expressed by a pathogen. Park et al. engineered the MazE-MazF antitoxin-toxin system of *Escherichia coli* to fuse a C-terminal 41-residue fragment of antitoxin MazE to the N-terminal end of toxin MazF with a linker having a specific protease cleavage site for either HIV PR (HIV-1 protease), or NS3 protease (HCV protease) (Figure 4C). These fusion proteins are designed for the treatment of human immunodeficiency virus (HIV-1) and hepatitis C virus, which express corresponding proteases. The toxicity of the MazF toxin was not revealed until cleavage of the protease-sensitive linker by the virus-specific proteases, and thus provides a novel approach for designing fusion protein with improved safety profile [68]. The authors used the sequence RVL↓AEA as the HIV protease cleavage site, and EDVVCC↓SMSY as the HCV NS3 protease cleavage site for constructing the fusion proteins with cleavable linkers. Another cleavable linker GGIEGR↓GS containing Factor Xa cleavage site was also constructed for comparison. By incubating the respective

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

fusion protein with either HIV PR, HCV NS3 or Factor Xa protease *in vitro*, MazF was released from the MazE-MazF fusion protein, and exhibited its mRNA interferase activity.

Besides the extracellular spaces, the targeting/activation site for *in vivo* cleavable linkers can also be inside the cells. An example of targeting intracellular compartments is the *in vivo* cleavable linker that is specifically cleaved by furin, a cellular endoproteinase implicated in the proteolytic activation of diverse precursor proteins as well as toxins [69, 70]. Furin is mainly localized in the trans-golgi network (TGN) and also recycles between TGN, early endosomes, and the cell surface [69]. It has the consensus recognition sequence of -Arg-X-Arg/Lys-Arg↓-(↓ identifies the cleavage site, X represents any amino acid) [71]. *In vivo* cleavable linkers with furin-sensitive sequences have been applied in various recombinant fusion proteins such as immunotoxins or immunoproapoptotic proteins. Active toxins or apoptotic proteins fused to scFv can be released inside the target cells after cleavage of the linker by furin. Inclusion of furin-sensitive linkers in immunotoxins containing ribotoxin [72], caspase-3 [73], or granzyme B [73], human active truncated Bid [74] has shown significant improvements in cytotoxicity compared with constructs containing stable linkers, likely due to the regeneration of fully active toxic domains upon linker cleavage. In the example of immunotoxin containing ribotoxin, a linker sequence of TRHRQPR↓GWEQL was designed for furin cleavage, and it was found to be cleaved efficiently *in vitro* (Figure 4D). Compared to the immunotoxin without a linker, the ones with the proteolytically cleavable linker exhibited enhanced cell-killing activity by 2±30-fold on various target cell lines [72]. Wang et al. designed 3 furin cleavage sequences, including a synthetic polyarginine tract (RRRRRRR↓R↓R), and two furin cleavable sequences from PEA and diphtheria toxin for furin cleavage (TRHRQPR↓GWE, AGNRVRR↓SVG), and applied them in immunoproapoptotic proteins consisting of scFv and caspase-3/granzyme B. These fusion proteins were able to efficiently and selectively bind to the targeting tumor cells, being cleaved at the furin cleavage site within the endosome, translocate to cytosol to induce cell death, and ultimately reduce tumor size in nude mice.

Similarly, a protease present in the lysosome, cathepsin B, has been applied for targeted intracellular activation of cytotoxic proteins. Cathepsin B substrate peptides have previously been utilized as cleavable peptide linkers in many bioconjugates [75, 76]. For instance, a dipeptide of Phe-Lys was applied to serve as part of a cleavable linker in an albumin-binding prodrug of doxorubicin 1, for the *in vivo* release of doxorubicin after Cathepsin B cleavage in tumor. The cathepsin B-cleavable linkers have recently been applied to fusion proteins. Yuan et. al. used a cathepsin B sensitive peptide of GFLG together with a furin cleavage sequence of R_2KR_6, to link a tumor-targeting moiety (fragment of C. perfringens enterotoxin) and a toxin (recombinant gelonin) in order to release the toxin in the lysosome [77].

5.5 Linkers can affect the PK of fusion proteins

Fusion proteins obtain many advantages over the parent proteins, such as improved PK and PD properties as in albumin- and Fc-fusion proteins, as well as the drug targeting effects as in immunotoxins. Although several fusion proteins have been applied in the clinic, the mechanisms underlying PK of bifunctional fusion proteins are still largely unexplored, and a generalized PK model for fusion protein is not established. Target-mediated drug disposition (TMDD), which describes the process where drug-target binding significantly influences the PK and PD of the drug, has been established as a crucial mechanism for the elimination of many single domain protein and peptide drugs [78]. Generally, for many protein drugs, the disposition processes affecting their PK are relatively simple, e.g., binding to their cell surface receptor leads to endocytosis and lysosomal degradation. However, the disposition of bifunctional fusion proteins are affected by two different domains/binding sites, and therefore their PK/PD properties are much more complicated. Since linker insertion may

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

alter the receptor binding affinity of each protein domain, it can affect the *in vivo* disposition of fusion proteins and increase the complexity of PK studies.

A study by Chen et al. demonstrated that linkers could affect the PK of fusion proteins via their impact on receptor binding and subsequent intracellular processing [79]. Three linkers with different length and conformation were inserted into fusion proteins composed of Tf and hGH. The first linker is a short dipeptide, Leu-Glu (LE, designated as "dipeptide" in Table 4) [19]. The second linker is a cyclopeptide with a cyclic confirmation formed by the disulfide bond between 2 cysteine residues (LEAGCKNFFPRSFTSCGSLE, designated as "cyclo" in Table 4)[39]. This linker has a length of 20 amino acids and a rigid cyclic structure, and is originally designed to create an *in vivo* cleavable disulfide linker as discussed in Section 3.3 [39]. The third linker is an α-helix-forming linker $(A(EAAAK)_4ALEA(EAAAK)_4A$, designated as "$(H4)_2$" in Table 4) with the longest length of 50 amino acids and a rigid, extended structure [20, 40].

The receptor binding affinities of hGH-Tf fusion proteins were greatly altered with the linker insertion. The shortest dipeptide LE linker resulted in the lowest binding affinities for both hGH receptor (hGHR) and Tf receptor (TfR) (Table 4, as indicated by the highest IC_{50} values for inhibiting respective receptor binding), while the longer and more rigid cyclo and $(H4)_2$ linkers generated higher receptor binding affinities (Table 4). This result suggests that by adjusting the separation between domains via linker insertion, receptor binding affinity may be greatly altered due to the change in interdomain steric hindrance or interference.

Next, the impact of the two receptor binding sites (hGHR and TfR) on the plasma half-life of hGH-Tf was investigated. The binding to hGHR led to endocytosis and lysosomal degradation of the fusion proteins as evidenced by the prolonged plasma half-life of three hGH-Tf proteins after blocking their hGHR binding with excess hGH (Table 4), similar to previous reports for free hGH [80, 81]. The binding to TfR, on the other hand, led to recycling of the fusion proteins via the classic Tf-TfR recycling pathway [13, 81], where blockage of TfR binding by excess Tf significantly shortened the half-life of dipeptide-liked hGH-Tf (Table 4). Due to the presence of excess Tf in the blood, the binding to the protein drug domain (i.e. hGHR in this case) is considered the primary binding site, while binding to TfR is considered secondary binding under physiological conditions.

The impact of linkers on receptor binding affinity resulted in dramatic differences in the PK profiles of 3 hGH-Tf fusion proteins. Dipeptide-linked hGH-LE-Tf, which has the weakest binding for hGHR and hTfR, exhibited the longest plasma half-life due to less hGHR-mediated degradation (primary binding). In comparison, hGH-cyclo-Tf and hGH-$(H4)_2$-Tf, which exhibited stronger hGHR binding affinities, had shorter plasma half-lives because of more degradation via hGHR. The results indicated that linker insertion could greatly affect the receptor binding and subsequent intracellular processing, and ultimately alter the plasma half-lives of bifunctional fusion proteins.

Based on these findings, a mechanistic PK model was proposed for bifunctional fusion proteins composed of a protein drug domain (i.e. hGH) and a carrier protein domain that undergoes receptor-mediated recycling (i.e. Tf, albumin) (Figure 5). Since the endogenous concentration of the carrier protein is high, the fusion proteins likely bind first to the protein drug receptor on the target cells (primary binding), and get enriched to the cell surface. Subsequently, secondary binding to the recycling receptor (i.e. TfR, albumin-receptor) likely occurs on the cell membrane following the enrichment of fusion protein at the cell surface, or inside the acidic endosomes where fusion proteins may dissociate from the protein drug receptor and bind to recycling receptor [82]. The receptor-binding affinity inside the endosomes determines different subsequent intracellular processing (degradation or

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

recycling) for the fusion proteins, and ultimately affects the plasma half-life of the fusion proteins.

This study highlights the importance of linkers in designing and developing bifunctional therapeutic fusion proteins from a PK perspective. First, linker technology can serve an effective protein engineering tool to alter the receptor binding affinities of fusion proteins. Second, linker insertion can be applied to achieve the desirable PK profiles of fusion proteins. Third, since the different functional domains in fusion proteins may play distinct roles in determining the plasma half-life, the impact of linker insertion on each domain's receptor binding should be carefully evaluated and balanced to achieve the optimal PK profiles.

6. Summary and perspective

During the development of therapeutic recombinant fusion proteins, linker design has become a valuable means to achieve desired characteristics of the products. Linker sequences derived from natural multi-domain proteins may provide useful references for designing empirical linkers. Various empirical linkers such as flexible, rigid or cleavable linkers have been designed for various purposes, such as passively joining domains, spatially separating domains, or releasing free functional domains *in vivo*. Optimal linkers can provide many advantages for the fusion proteins production, including improving structural stability, enhancing bioactivity, increasing expression level, altering the PK profiles and enabling the *in vivo* targeting of the fusion proteins. Although many examples of various types of linkers have been developed in the past, the rational design of linkers for the construction of fusion proteins is still in its infancy. Systematic, strategic scientific endeavors are in demand to greatly advance the science of linker design and application. Many technology platforms may be investigated in more depth towards understanding the connection between linker composition and structure, and ultimately tie them to linker function.

The study of linker composition and structure, and the investigation of linker function should go hand in hand when designing a novel linker. An good example in the rational design of linkers is the rigid helical linkers ($A(EAAAK)_nA$) by Arai et al. [34, 35]. The idea of using these sequences as a linker started from the finding that they form an α-helical conformation in water as determined by circular dichroism [83]. It was then proposed to apply them to effectively separate protein domains in fusion proteins. To test this hypothesis, the linkers were investigated in a functional study, as well as conformational study. By determining the function of the fusion protein by measuring the FRET efficiency between two fluorescent protein domains [34], it was suggested that longer helical peptides increased the distance between domains. Analysis of variably linked chimeric proteins via synchrotron X-ray small-angle scattering [35] revealed that the chimeric protein with the helical linker assumed a more elongated conformation compared to the flexible linker, suggesting the helical linker was more effective for domain separation. These types of studies established a solid understanding of the linker, and greatly facilitated its further application. Further systematic studies about linker structures, including X-ray crystallography and NMR techniques [35, 36], would be greatly useful.

The establishment of more databases and searching programs for linkers would be another fruitful direction. As discussed earlier, only two studies have been performed to analyze the characteristics of the linkers in natural multi-domain proteins [24, 84]. With the rapid increase of the number of protein structures deposited in the PDB database, an updated study of natural linkers could be conducted. In addition to the properties analyzed in previous studies (e.g., amino acid composition, structure classification), it would be interesting to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

categorize the multi-domain proteins by their functions and structures, and identify the relationship between them and the linker properties. For example, is a more flexible linker confirmation preferred in interacting protein domains? Is there any preference of amino acids to be used in linkers in certain classes of proteins, e.g., transcription factors? A thorough study of the natural linkers could provide candidates for protein fusion, but more importantly, improve our understanding of linker properties. Similarly, building an empirical linker database could help summarize the knowledge and facilitate the future linker design. The extensive studies on the structures of empirical linkers have provided us with useful information for optimal linker design. Ultimately, more searching algorithms for linker databases could be developed, and provide more linker candidates for protein fusion based on user specifications.

With the rapid advancement of protein science and biotechnology, the design of linkers in fusion proteins has become more important than ever before. With a thorough understanding of their structures, conformations, and functions via future biomedical research, the incorporation of linkers will greatly facilitate the construction of stable and bioactive recombinant fusion proteins for drug delivery applications.

ABBREVIATIONS

aa	amino acid
EBFP	enhanced blue fluorescent protein
EGFP	enhanced green fluorescent protein
FIX	coagulation factor IX
FRET	fluorescent resonance energy transfer
G-CSF	granulocyte colony-stimulating factor
HSA	human serum albumin
hGH	human growth hormone
IFN	interferon
LAP	latency associated peptide
MMP	matrix metalloproteinase
PD	pharmacodynamic
PDB	Brookhaven Protein Data Bank
PK	pharmacokinetic
scFv	single-chain variable fragment
Tf	transferrin
TGN	trans-golgi network
$\mathbf{V_H}$	antibody heavy chain variable region
$\mathbf{V_L}$	antibody light-chain variable regions

References

[1]. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2003; 60:523–533. [PubMed: 12536251]

[2]. Yuste R. Fluorescence microscopy today. Nat Methods. 2005; 2:902–904. [PubMed: 16299474]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[3]. Schmidt SR. Fusion-proteins as biopharmaceuticals--applications and challenges. Curr Opin Drug Discov Devel. 2009; 12:284–295.

[4]. Chen X, Zaro JL, Shen WC. Pharmacokinetics of recombinant bifunctional fusion proteins. Expert opinion on drug metabolism & toxicology. 2012; 8:581–595. [PubMed: 22428984]

[5]. Kim B, Zhou J, Martin B, Carlson O, Maudsley S, Greig N, Mattson M, Ladenheim E, Wustner J, Turner A, Sadeghi H, Egan J. Transferrin fusion technology: a novel approach to prolong biological half-life of insulinotropic peptides. J Pharmacol Exp Ther. 2010; 334:682–692. [PubMed: 20498254]

[6]. Weimer T, Wormsbächer W, Kronthaler U, Lang W, Liebing U, Schulte S. Prolonged in-vivo half-life of factor VIIa by fusion to albumin. Thromb Haemost. 2008; 99:659–667. [PubMed: 18392323]

[7]. Dumont JA, Low SC, Peters RT, Bitonti AJ. Monomeric Fc fusions: impact on pharmacokinetic and biological activity of protein therapeutics. BioDrugs. 2006; 20:151–160. [PubMed: 16724863]

[8]. Schellenberger V, Wang C, Geething N, Spink B, Campbell A, To W, Scholle M, Yin Y, Yao Y, Bogin O, Cleland J, Silverman J, Stemmer W. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat Biotechnol. 2009; 27:1186–1190. [PubMed: 19915550]

[9]. Kreitman RJ. Immunotoxins for targeted cancer therapy. AAPS J. 2006; 8:E532–551. [PubMed: 17025272]

[10]. Manoukian G, Hagemeister F. Denileukin diftitox: a novel immunotoxin. Expert Opin Biol Ther. 2009; 9:1445–1451. [PubMed: 19817678]

[11]. Pardridge WM. Biopharmaceutical drug targeting to the brain. J Drug Target. 2010; 18:157–167. [PubMed: 20064077]

[12]. Patel LN, Zaro JL, Shen WC. Cell penetrating peptides: intracellular pathways and pharmaceutical perspectives. Pharm Res. 2007; 24:1977–1992. [PubMed: 17443399]

[13]. Widera A, Norouziyan F, Shen WC. Mechanisms of TfR-mediated transcytosis and sorting in epithelial cells and applications toward drug delivery. Adv Drug Deliv Rev. 2003; 55:1439–1466. [PubMed: 14597140]

[14]. Leader B, Baca Q, Golan D. Protein therapeutics: a summary and pharmacological classification. Nat Rev Drug Discov. 2008; 7:21–39. [PubMed: 18097458]

[15]. McDermott MF. Rilonacept in the treatment of chronic inflammatory disorders. Drugs Today (Barc). 2009; 45:423–430. [PubMed: 19649332]

[16]. Molineux G. The development of romiplostim for patients with immune thrombocytopenia. Ann N Y Acad Sci. 2011; 1222:55–63. [PubMed: 21434943]

[17]. Zhao H, Yao X, Xue C, Wang Y, Xiong X, Liu Z. Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. Protein Expr Purif. 2008; 61:73–77. [PubMed: 18541441]

[18]. Amet N, Lee HF, Shen WC. Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. Pharm Res. 2009; 26:523–528. [PubMed: 19002568]

[19]. Bai Y, Ann DK, Shen WC. Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. Proc Natl Acad Sci U S A. 2005; 102:7292–7296. [PubMed: 15870205]

[20]. Bai Y, Shen WC. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. Pharm Res. 2006; 23:2116–2121. [PubMed: 16952003]

[21]. Gokhale RS, Khosla C. Role of linkers in communication between protein modules. Curr Opin Chem Biol. 2000; 4:22–27. [PubMed: 10679375]

[22]. Ikebe M, Kambara T, Stafford WF, Sata M, Katayama E, Ikebe R. A hinge at the central helix of the regulatory light chain of myosin is critical for phosphorylation-dependent regulation of smooth muscle myosin motor activity. J Biol Chem. 1998; 273:17702–17707. [PubMed: 9651368]

[23]. 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. [PubMed: 2313701]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[24]. George R, Heringa J. An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng. 2002; 15:871–879. [PubMed: 12538906]

[25]. Williamson MP. The structure and function of proline-rich regions in proteins. Biochem J. 1994; 297(Pt 2):249–260. [PubMed: 8297327]

[26]. Turner S, Russell G, Williamson M, Guest J. Restructuring an interdomain linker in the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase complex of Escherichia coli. Protein Eng. 1993; 6:101–108. [PubMed: 8433963]

[27]. Mottram JC, North MJ, Barry JD, Coombs GH. A cysteine proteinase cDNA from Trypanosoma brucei predicts an enzyme with an unusual C-terminal extension. FEBS Lett. 1989; 258:211–215. [PubMed: 2599086]

[28]. Aurora R, Creamer T, Srinivasan R, Rose G. Local interactions in protein folding: lessons from the alpha-helix. J Biol Chem. 1997; 272:1413–1416. [PubMed: 9019474]

[29]. Bird R, Hardman K, Jacobson J, Johnson S, Kaufman B, Lee S, Lee T, Pope S, Riordan G, Whitlow M. Single-chain antigen-binding proteins. Science. 1988; 242:423–426. [PubMed: 3140379]

[30]. Sabourin M, Tuzon C, Fisher T, Zakian V. A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. Yeast. 2007; 24:39–45. [PubMed: 17192851]

[31]. de Bold MK, Sheffield WP, Martinuk A, Bhakta V, Eltringham-Smith L, de Bold AJ. Characterization of a long-acting recombinant human serum albumin-atrial natriuretic factor (ANF) expressed in Pichia pastoris. Regul Pept. 2012; 175:7–10. [PubMed: 22296859]

[32]. Waldo G, Standish B, Berendzen J, Terwilliger T. Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol. 1999; 17:691–695. [PubMed: 10404163]

[33]. 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. [PubMed: 9212866]

[34]. 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. [PubMed: 11579220]

[35]. Arai R, Wriggers W, Nishikawa Y, Nagamune T, Fujisawa T. Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. Proteins. 2004; 57:829–838. [PubMed: 15390267]

[36]. Bhandari D, Levine B, Trayer I, Yeadon M. 1H-NMR study of mobility and conformational constraints within the proline-rich N-terminal of the LC1 alkali light chain of skeletal myosin. Correlation with similar segments in other protein systems. Eur J Biochem. 1986; 160:349–356. [PubMed: 3769935]

[37]. Evans J, Levine B, Trayer I, Dorman C, Higgins C. Sequence-imposed structural constraints in the TonB protein of E. coli. FEBS Lett. 1986; 208:211–216. [PubMed: 3023135]

[38]. Morris A, MacArthur M, Hutchinson E, Thornton J. Stereochemical quality of protein structure coordinates. Proteins. 1992; 12:345–364. [PubMed: 1579569]

[39]. Chen X, Bai Y, Zaro J, Shen WC. Design of an in vivo cleavable disulfide linker in recombinant fusion proteins. Biotechniques. 2010; 49:513–518. [PubMed: 20615204]

[40]. Amet N, Wang W, Shen WC. Human growth hormone-transferrin fusion protein for oral delivery in hypophysectomized rats. J Control Release. 2010; 141:177–182. [PubMed: 19761807]

[41]. Thorpe PE, Wallace PM, Knowles PP, Relf MG, Brown AN, Watson GJ, Knyba RE, Wawrzynczak EJ, Blakey DC. New coupling agents for the synthesis of immunotoxins containing a hindered disulfide bond with improved stability in vivo. Cancer Res. 1987; 47:5924–5931. [PubMed: 3499221]

[42]. Zhao HL, Xue C, Du JL, Ren M, Xia S, Liu ZM. Balancing the pharmacokinetics and pharmacodynamics of interferon-alpha2b and human serum albumin fusion protein by proteolytic or reductive cleavage increases its in vivo therapeutic efficacy. Mol Pharm. 2012; 9:664–670. [PubMed: 22224503]

[43]. Crasto CJ, Feng JA. LINKER: a program to generate linker sequences for fusion proteins. Protein Eng. 2000; 13:309–312. [PubMed: 10835103]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[44]. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]

[45]. Hagemeyer CE, von Zur Muhlen C, von Elverfeldt D, Peter K. Single-chain antibodies as diagnostic tools and therapeutic agents. Thromb Haemost. 2009; 101:1012–1019. [PubMed: 19492141]

[46]. Huston J, Levinson D, Mudgett-Hunter M, Tai M, Novotný J, Margolies M, Ridge R, Bruccoleri R, Haber E, Crea R. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. Proc Natl Acad Sci U S A. 1988; 85:5879–5883. [PubMed: 3045807]

[47]. Trinh R, Gurbaxani B, Morrison SL, Seyfzadeh M. Optimization of codon pair use within the (GGGGS)3 linker sequence results in enhanced protein expression. Mol Immunol. 2004; 40:717–722. [PubMed: 14644097]

[48]. Reddy ST, Ge X, Miklos AE, Hughes RA, Kang SH, Hoi KH, Chrysostomou C, Hunicke-Smith SP, Iverson BL, Tucker PW, Ellington AD, Georgiou G. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat Biotechnol. 2010; 28:965–969. [PubMed: 20802495]

[49]. Alfthan K, Takkinen K, Sizmann D, Söderlund H, Teeri TT. Properties of a single-chain antibody containing different linker peptides. Protein Eng. 1995; 8:725–731. [PubMed: 8577701]

[50]. Takamatsu N, Watanabe Y, Yanagi H, Meshi T, Shiba T, Okada Y. Production of enkephalin in tobacco protoplasts using tobacco mosaic virus RNA vector. FEBS Lett. 1990; 269:73–76. [PubMed: 2387417]

[51]. Werner S, Marillonnet S, Hause G, Klimyuk V, Gleba Y. Immunoabsorbent nanoparticles based on a tobamovirus displaying protein A. Proc Natl Acad Sci U S A. 2006; 103:17678–17683. [PubMed: 17090664]

[52]. Lu P, Feng MG. Bifunctional enhancement of a beta-glucanase-xylanase fusion enzyme by optimization of peptide linkers. Appl Microbiol Biotechnol. 2008; 79:579–587. [PubMed: 18415095]

[53]. Zhao H, Xue C, Wang Y, Li X, Xiong X, Yao X, Liu Z. Circumventing the heterogeneity and instability of human serum albumin-interferon-alpha2b fusion protein by altering its orientation. J Biotechnol. 2007; 131:245–252. [PubMed: 17698234]

[54]. McCormick A, Thomas M, Heath A. Immunization with an interferon-gamma-gp120 fusion protein induces enhanced immune responses to human immunodeficiency virus gp120. J Infect Dis. 2001; 184:1423–1430. [PubMed: 11709784]

[55]. Bergeron LM, Gomez L, Whitehead TA, Clark DS. Self-renaturing enzymes: design of an enzyme-chaperone chimera as a new approach to enzyme stabilization. Biotechnol Bioeng. 2009; 102:1316–1322. [PubMed: 19170240]

[56]. Shah D, Shen WC. The establishment of polarity and enhanced transcytosis of transferrin receptors in enterocyte-like Caco-2 cells. J Drug Target. 1994; 2:93–99. [PubMed: 8069597]

[57]. Duttaroy A, Kanakaraj P, Osborn B, Schneider H, Pickeral O, Chen C, Zhang G, Kaithamana S, Singh M, Schulingkamp R, Crossan D, Bock J, Kaufman T, Reavey P, Carey-Barber M, Krishnan S, Garcia A, Murphy K, Siskind J, McLean M, Cheng S, Ruben S, Birse C, Blondel O. Development of a long-acting insulin analog using albumin fusion technology. Diabetes. 2005; 54:251–258. [PubMed: 15616036]

[58]. Halpern W, Riccobene T, Agostini H, Baker K, Stolow D, Gu M, Hirsch J, Mahoney A, Carrell J, Boyd E, Grzegorzewski K. Albugranin, a recombinant human granulocyte colony stimulating factor (G-CSF) genetically fused to recombinant human albumin induces prolonged myelopoietic effects in mice and monkeys. Pharm Res. 2002; 19:1720–1729. [PubMed: 12458679]

[59]. Schulte S. Half-life extension through albumin fusion technologies. Thromb Res. 2009; 124(Suppl 2):S6–8. [PubMed: 20109653]

[60]. Martel-Pelletier J, Welsch DJ, Pelletier JP. Metalloproteases and inhibitors in arthritic diseases. Best Pract Res Clin Rheumatol. 2001; 15:805–829. [PubMed: 11812023]

[61]. Lampert K, Machein U, Machein MR, Conca W, Peter HH, Volk B. Expression of matrix metalloproteinases and their tissue inhibitors in human brain tumors. Am J Pathol. 1998; 153:429–437. [PubMed: 9708803]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[62]. Yong VW, Krekoski CA, Forsyth PA, Bell R, Edwards DR. Matrix metalloproteinases and diseases of the CNS. Trends Neurosci. 1998; 21:75–80. [PubMed: 9498303]

[63]. Leppert D, Lindberg RL, Kappos L, Leib SL. Matrix metalloproteinases: multifunctional effectors of inflammation in multiple sclerosis and bacterial meningitis. Brain Res Brain Res Rev. 2001; 36:249–257. [PubMed: 11690622]

[64]. Vessillier S, Adams G, Chernajovsky Y. Latent cytokines: development of novel cleavage sites and kinetic analysis of their differential sensitivity to MMP-1 and MMP-3. Protein Eng Des Sel. 2004; 17:829–835. [PubMed: 15708865]

[65]. Wakefield LM, Winokur TS, Hollands RS, Christopherson K, Levinson AD, Sporn MB. Recombinant latent transforming growth factor beta 1 has a longer plasma half-life in rats than active transforming growth factor beta 1, and a different tissue distribution. J Clin Invest. 1990; 86:1976–1984. [PubMed: 2254455]

[66]. Vessillier S, Adams G, Montero-Melendez T, Jones R, Seed M, Perretti M, Chernajovsky Y. Molecular engineering of short half-life small peptides (VIP, alphaMSH and gammaMSH) fused to latency-associated peptide results in improved anti-inflammatory therapeutics. Ann Rheum Dis. 2012; 71:143–149. [PubMed: 21998117]

[67]. Adams G, Vessillier S, Dreja H, Chernajovsky Y. Targeting cytokines to inflammation sites. Nat Biotechnol. 2003; 21:1314–1320. [PubMed: 14528315]

[68]. Park JH, Yamaguchi Y, Inouye M. Intramolecular regulation of the sequence-specific mRNA interferase activity of MazF fused to a MazE fragment with a cleavable linker by specific proteases. Appl Environ Microbiol. 2012; 78:3794–3799. [PubMed: 22447587]

[69]. Thomas G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Nat Rev Mol Cell Biol. 2002; 3:753–766. [PubMed: 12360192]

[70]. Inocencio NM, Moehring JM, Moehring TJ. Furin activates Pseudomonas exotoxin A by specific cleavage in vivo and in vitro. J Biol Chem. 1994; 269:31831–31835. [PubMed: 7989356]

[71]. Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber R, Lindberg I, Bode W, Than ME. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. Nat Struct Biol. 2003; 10:520–526. [PubMed: 12794637]

[72]. Goyal A, Batra JK. Inclusion of a furin-sensitive spacer enhances the cytotoxicity of ribotoxin restrictocin containing recombinant single-chain immunotoxins. Biochem J. 2000; 345(Pt 2): 247–254. [PubMed: 10620501]

[73]. Wang T, Zhao J, Ren JL, Zhang L, Wen WH, Zhang R, Qin WW, Jia LT, Yao LB, Zhang YQ, Chen SY, Yang AG. Recombinant immunoproapoptotic proteins with furin site can translocate and kill HER2-positive cancer cells. Cancer Res. 2007; 67:11830–11839. [PubMed: 18089813]

[74]. Wang F, Ren J, Qiu XC, Wang LF, Zhu Q, Zhang YQ, Huan Y, Meng YL, Yao LB, Chen SY, Xu YM, Yang AG. Selective cytotoxicity to HER2-positive tumor cells by a recombinant e23sFv-TD-tBID protein containing a furin cleavage sequence. Clin Cancer Res. 2010; 16:2284–2294. [PubMed: 20371697]

[75]. Peterson JJ, Meares CF. Cathepsin substrates as cleavable peptide linkers in bioconjugates, selected from a fluorescence quench combinatorial library. Bioconjug Chem. 1998; 9:618–626. [PubMed: 9736496]

[76]. Abu Ajaj K, Graeser R, Fichtner I, Kratz F. In vitro and in vivo study of an albumin-binding prodrug of doxorubicin that is cleaved by cathepsin B. Cancer Chemother Pharmacol. 2009; 64:413–418. [PubMed: 19229536]

[77]. Yuan X, Lin X, Manorek G, Howell SB. Challenges associated with the targeted delivery of gelonin to claudin-expressing cancer cells with the use of activatable cell penetrating peptides to enhance potency. BMC Cancer. 2011; 11:61. [PubMed: 21303546]

[78]. Mager DE. Target-mediated drug disposition and dynamics. Biochem Pharmacol. 2006; 72:1–10. [PubMed: 16469301]

[79]. Chen X, Lee HF, Zaro JL, Shen WC. Effects of receptor binding on plasma half-life of bifunctional transferrin fusion proteins. Mol Pharm. 2011; 8:457–465. [PubMed: 21291258]

[80]. van Kerkhof P, Strous G. The ubiquitin-proteasome pathway regulates lysosomal degradation of the growth hormone receptor and its ligand. Biochem Soc Trans. 2001; 29:488–493. [PubMed: 11498015]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[81]. Grant B, Donaldson J. Pathways and mechanisms of endocytic recycling. Nat Rev Mol Cell Biol. 2009; 10:597–608. [PubMed: 19696797]

[82]. Muller-Eberhard U, Liem H, Grasso J, Giffhorn-Katz S, DeFalco M, Katz N. Increase in surface expression of transferrin receptors on cultured hepatocytes of adult rats in response to iron deficiency. J Biol Chem. 1988; 263:14753–14756. [PubMed: 3170563]

[83]. Marqusee S, Baldwin RL. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. Proc Natl Acad Sci U S A. 1987; 84:8898–8902. [PubMed: 3122208]

[84]. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. Journal of Molecular Biology. 1990; 211:943–958. [PubMed: 2313701]

[85]. Hu W, Li F, Yang X, Li Z, Xia H, Li G, Wang Y, Zhang Z. A flexible peptide linker enhances the immunoreactivity of two copies HBsAg preS1 (21-47) fusion protein. J Biotechnol. 2004; 107:83–90. [PubMed: 14687974]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

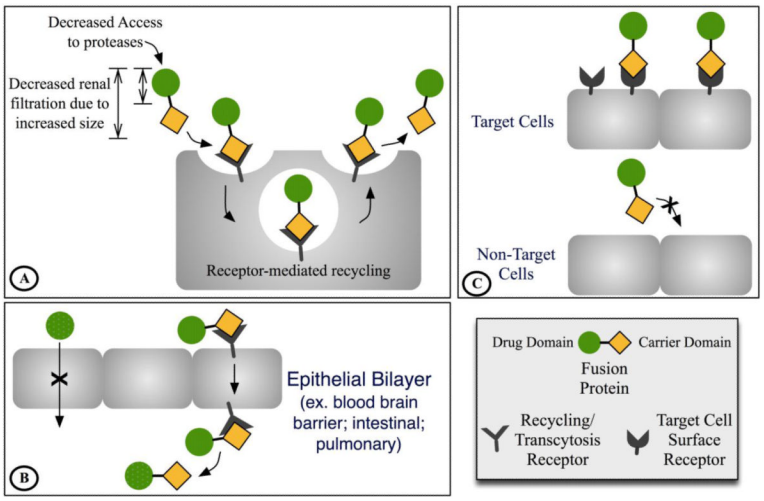


Figure 1. Applications of bifunctional fusion proteins in drug delivery

Bifunctional fusion proteins offer several advantages in drug delivery applications by (A) extending plasma half-life by decreasing access to proteases, decreasing renal filtration, or by altering the intracellular routing via receptor-mediated recycling; (B) enabling absorption across epithelial bilayers by binding to receptors that undergo transcytosis and (C) targeting *in vivo* sites that over-express or uniquely express specific receptors or antigens.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

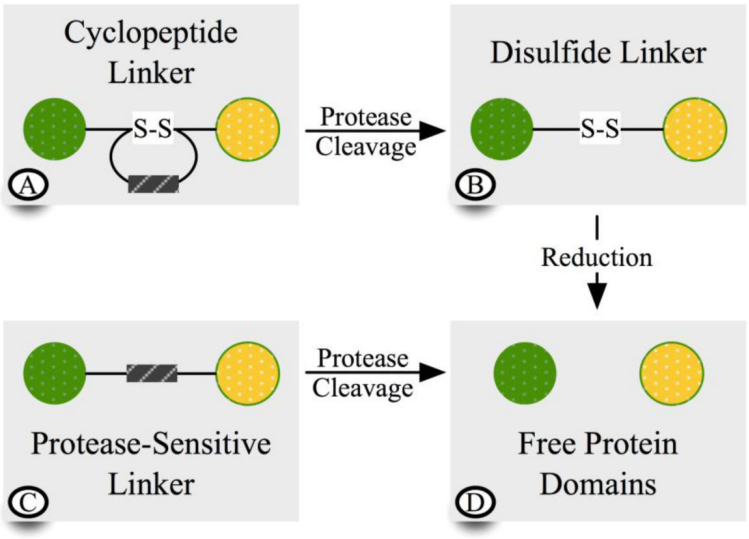


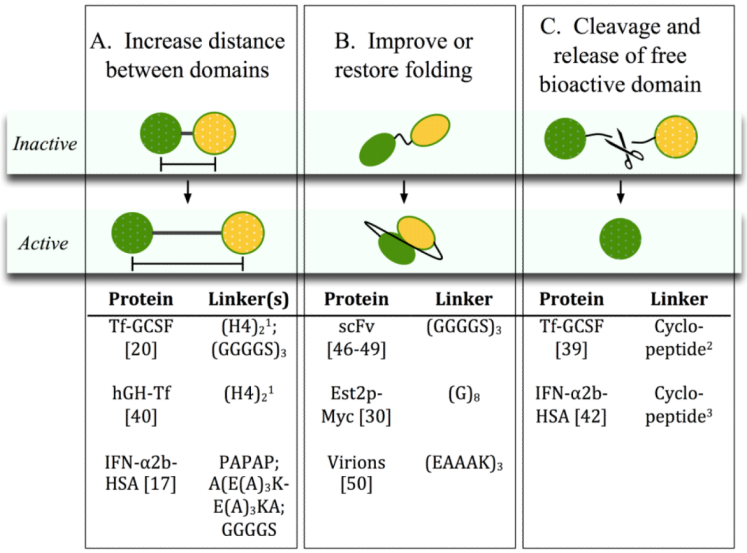
Figure 2. Illustration of in vivo cleavable linkers

The cyclopeptide linkers (A) contain a disulfide linkage between 2 Cys residues as well as a peptide loop containing a protease sensitive cleavage site. Cleavage of the protease-sensitive site generates a disulfide linked fusion protein (B) which, following reduction, releases the free protein domains (D). The protease sensitive linkers (C) generally contain a cleavage site sensitive to proteases present in specific tissues or intracellular compartments (eg. MMPs, furin, cathepsin B) to trigger the release of the free protein domains (D) at specific sites.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Protein	Linker(s)	Protein	Linker	Protein	Linker
Tf-GCSF [20]	$(H4)_2$[1]; $(GGGGS)_3$	scFv [46-49]	$(GGGGS)_3$	Tf-GCSF [39]	Cyclo-peptide[2]
hGH-Tf [40]	$(H4)_2$[1]	Est2p-Myc [30]	$(G)_8$	IFN-α2b-HSA [42]	Cyclo-peptide[3]
IFN-α2b-HSA [17]	PAPAP; $A(E(A)_3K-E(A)_3KA$; GGGGS	Virions [50]	$(EAAAK)_3$		

Figure 3. Improving bioactivity of fusion proteins using linkers

Insertion of linkers between fusion protein domains can increase bioactivity by (A) increasing distance between domains, (B) improving or restoring folding, or (C) releasing active domain using a cleavable linker.

[1]$(H4)_2$ sequence: $A(EAAAAK)_4ALEA(EAAAAK)_4A$

[2]Cyclopeptide sequence: LEAGCKNFFPR↓SFTSCGSLE containing a disulfide bridge between C residues and a thrombin-sensitive cleavage site

[3]Cyclopeptide sequence: CRRRRRR↓EAEAC containing a disulfide bridge between C residues and a furin-sensitive cleavage site.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

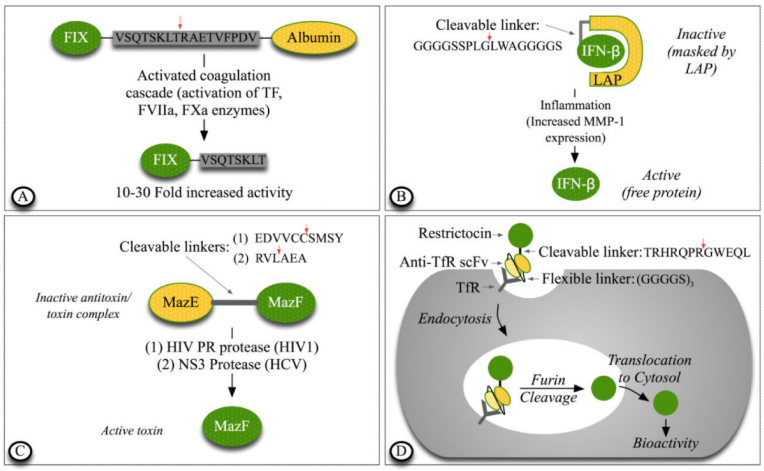


Figure 4. Use of linkers to target fusion proteins to specific sites

(A) Cleavable linkers sensitive to proteases that become active under certain physiological/ pathological conditions. For example, a protease sensitive linker has been inserted between Factor IX (FIX) and albumin protein domains to be selectively cleaved during activation of the coagulation cascade during clotting, thus releasing in the increased bioactivity of FIX [58]. (B) Linkers that are cleaved by proteases overexpressed at disease sites. For example, when fused to IFN-β, its bioactivity is masked by LAP. In areas of inflammation, the increased expression of MMPs (specifically MMP-1 in this example) results in the cleavage and un-masking of active IFN-β at the target site [66]. (C) Targeting proteases specifically expressed by a pathogen. For example, an inactive antitoxin (MazE): toxin (MaxF) complex was linked via a HIV or HCV-protease sensitive linkage. The toxicity of the MazF toxin was not revealed until cleavage of the protease-sensitive linker by the virus-specific proteases [67]. (D) Linkers that are cleaved at specific intracellular sites. In the example of immunotoxin (restrictocin) was linked to an anti-TfR scFV via a furin-sensitive linkage. Following Tf-receptor mediated endocytosis, the complex is exposed to furin proteases in the endosomes resulting in cleavage and cytosolic translocation of free restrictocin to the cytosol and subsequent bioactivity [71].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

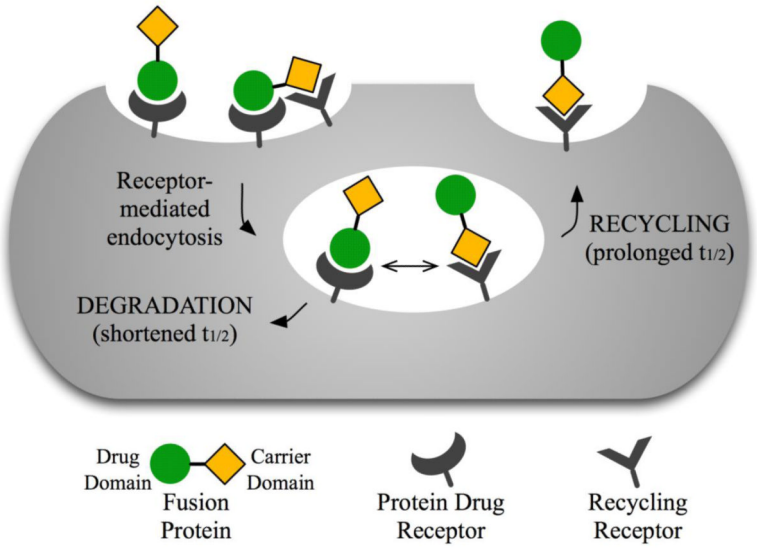


Figure 5. Relationship between receptor binding and pharmacokinetics of fusion proteins

In the presence of abundant endogenous Tf, the fusion proteins first bind to the protein drug receptor (e.g. GHR) on the target cell membrane. This binding is considered the primary binding, which enriches the fusion proteins onto the target cells and may lead to bivalent secondary binding of the Tf-domain to its recycling receptor (TfR). The fusion proteins are endocytosed into the early endosome, where fusion proteins that remain bound to GHR are degraded in the lysosome resulting in a shortened plasma half-life. By binding to the recycling receptor also present in the early endosomes, the fusion protein will be recycled back to the cell surface resulting in a prolonged plasma half-life. Therefore, the relative binding affinity of each protein domain for its receptor, which is altered by different linkers, affects the half-life of the fusion protein.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Properties of linkers derived from natural proteins.

Property		Argos[1]	George and Heringa[2]
Length (number of amino acid residues)		6.5	10.0 ± 5.8 (small: 4.5 ± 0.7; medium: 9.1 ± 2.4; large: 21 ± 7.6)
Hydrophobicity[3]		--	0.65 ± 0.09 (small: 0.69 ± 0.11; large: 0.62 ± 0.08)
Amino acid propensity[4]	Thr	1.55	1.017
	Ser	1.46	0.947
	Pro	1.35	1.299
	Gly	1.25	0.835
	Asp	1.25	0.916
	Lys	1.16	0.944
	Gln	1.13	1.047
	Asn	1.09	0.944
	Ala	1.05	0.964
	Val	1	0.955
	Glu	0.87	1.051
	Arg	0.84	1.143
	Ile	0.81	0.922
	Tyr	0.75	1
	Met	0.75	1.032
	Phe	0.69	1.119
	His	0.55	1.014
	Cys	0.35	0.778
	Trp	0.23	0.895
	Leu	N/A	1.085
Secondary Structure[5]	α-Helical	13%	38.3% (small: 21%; large: 31.4%)
	β-Strand	12%	13.6% (small: 33.6%; large: 10.4%)
	Coil	59%	37.6% (small: 36.9%; large: 45.4%)
	Turns	16%	8.4% (small: 8.5%; large: 12.8%)

[1] Data taken from study by Argos [23]

[2] Data taken from study by George and Heringa [24]

[3] Hydrophobicity values taken from Eisenberg's normalized consensus residue hydrophobicity scale, which ranges from 0 (hydrophilic) to 1 (hydrophobic)

[4] Calculated from the ratio of a single amino acid occurrence in the linker set compared to its occurrence in the full protein set, where values greater than 1 (shaded) indicate larger than average occurrences in linker sequences.

[5] Secondary structures were assigned using algorithms (Argos study [23]; George and Heringa study [24]), and the values are represented as a % of total linkers.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

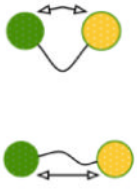
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Summary of empirical linkers

Linker		Advantages	Characteristics	Examples
Flexible		Allow for interaction between domains, or Increase spatial separation between domains	Rich in small or hydrophilic amino acids	$(GGGGS)_n$, $(G)_n$
Rigid		Maintain distance between domains	Helical structure or rich in Pro	$(EAAAK)_n$, $(XP)_n$
Cleavable		Allow for *in vivo* separation of domains	Reductive or enzymatic cleavage	Disulfide, protease sensitive sequences

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Examples of linkers and their functionalities.

Linker Function	Examples			
	Fusion Protein	Linker		Ref.
		Type	Sequence[a]	
Increase Stability/Folding	scFv	flexible	$(GGGGS)_3$	[46]
	G-CSF-Tf	flexible	$(GGGGS)_3$	[20]
	HBsAg preS1	flexible	$(GGGGS)_3$	[85]
	Myc- Est2p	flexible	$(Gly)_8$	[30]
	albumin-ANF	flexible	$(Gly)_6$	[31]
	virus coat protein	rigid	$(EAAAK)_3$	[50]
	beta-glucanase-xylanase	rigid	$(EAAAK)_n$ (n=1-3)	[52]
Increase expression	hGH-Tf and Tf-hGH	rigid	$A(EAAAK)_4ALEA(EAAAK)_4A$	[18]
	G-CSF-Tf and Tf-G-CSF	rigid	$A(EAAAK)_4ALEA(EAAAK)_4A$	[18]
Improve biological activity	G-CSF-Tf	flexible	$(GGGGS)_3$	[20]
	G-CSF-Tf	rigid	$A(EAAAK)_4ALEA(EAAAK)_4A$	[20]
	hGH-Tf	rigid	$A(EAAAK)_4ALEA(EAAAK)_4A$	[40]
	HSA-IFN-α2b	flexible	GGGGS	[17]
	HSA-IFN-α2b	rigid	PAPAP	[17]
	HSA-IFN-α2b	rigid	AEAAAKEAAAKA	[17]
	PGA-rTHS	flexible	$(GGGGS)_n$ (n=1, 2, 4)	[55]
	interferon-γ-gp120	rigid	$(Ala\text{-}Pro)_n$ (10 – 34 aa)	[54]
	GSF-S-S-Tf	cleavable	disulfide	[39]
	IFN-α2b-HSA	cleavable	disulfide	[42]
Enable targeting	FIX-albumin	cleavable	VSQTSKLTR↓AETVFPDV[b]	[59]
	LAP-IFN-β	cleavable	PLG ↓ LWA [c]	[64]
	MazE-MazF	cleavable	RVL↓AEA; EDVVCC↓SMSY; GGIEGR↓GS[c]	[68]
	Immunotoxins	cleavable	TRHRQPR↓GWE; AGNRVRR↓SVG; RRRRRRR↓R↓R[d]	[72]
	Immunotoxin	cleavable	GFLG↓[e]	[77]
Alter PK	G-CSF-Tf and hGH-Tf	dipeptide	LE	[79]
		rigid	$A(EAAAK)_4ALEA(EAAAK)_4A$	
		cleavable	Disulfide	

[a] Protease sensitive cleavage sites are indicated with "↓"

[b] Factor XIa/FVIIa sensitive cleavage

[c] Matrix metalloprotease-1 sensitive cleavage sequences, one example provided here

[d] HIV PR (HIV-1 protease); NS3 protease (HCV protease); Factor Xa sensitive cleavage, respectively

[e] Furin sensitive cleavage

[f] Cathepsin B sensitive cleavage

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Comparison of IC_{50} values and plasma-half lives of Tf fusion proteins (data taken from [79]).

Fusion Protein	Linker[a]	IC_{50} (nM)[b]		$t_{1/2}$ (h)[c]	$t_{1/2}$ with blockage (h)[c]	
		GHR/GCSFR	TfR		+GH	+Tf
GH-Tf	dipeptide	17.7	21.2	4.97 ± 0.34	5.95 ± 0.68	3.00 ± 0.94
	cyclo	8.2	4.2	1.76 ± 0.27	8.66 ± 2.98	2.14 ± 0.05
	$(H4)_2$	7.0	8.7	1.87 ± 0.44	6.73 ± 2.05	1.61 ± 0.75
GCSF-Tf	dipeptide	38.0	7.5	4.15 ± 0.75	--	--
	cyclo	39.5	0.9	5.69 ± 0.46	--	--
	$(H4)_2$	31.2	4.5	4.84 ± 1.18	--	--

[a] Linker sequences: Dipeptide (-LE-); Cyclo (-LEAGCKNFFPRSFTSCGSLE- with a disulfide bridge between two C); $(H4)_2$ (-LEA$(EAAAK)_4$ALEA$(EAAAK)_4$ALE-)

[b] IC_{50} values were obtained from competitive receptor binding assays. The lower IC_{50} value indicates higher receptor binding affinity for the fusion protein.

[c] Half-life ($t_{1/2}$) values represent mean ± SD.