REVIEWS

Linkers in the structural biology of protein–protein interactions

Vishnu Priyanka Reddy Chichili, Veerendra Kumar, and J. Sivaraman*

Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

Received 7 September 2012; Revised 8 November 2012; Accepted 13 November 2012
DOI: 10.1002/pro.2206
Published online 6 December 2012 proteinscience.org

Abstract: Linkers or spacers are short amino acid sequences created in nature to separate multiple domains in a single protein. Most of them are rigid and function to prohibit unwanted interactions between the discrete domains. However, Gly-rich linkers are flexible, connecting various domains in a single protein without interfering with the function of each domain. The advent of recombinant DNA technology made it possible to fuse two interacting partners with the introduction of artificial linkers. Often, independent proteins may not exist as stable or structured proteins until they interact with their binding partner, following which they gain stability and the essential structural elements. Gly-rich linkers have been proven useful for these types of unstable interactions, particularly where the interaction is weak and transient, by creating a covalent link between the proteins to form a stable protein–protein complex. Gly-rich linkers are also employed to form stable covalently linked dimers, and to connect two independent domains that create a ligand-binding site or recognition sequence. The lengths of linkers vary from 2 to 31 amino acids, optimized for each condition so that the linker does not impose any constraints on the conformation or interactions of the linked partners. Various structures of covalently linked protein complexes have been described using X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy techniques. In this review, we evaluate several structural studies where linkers have been used to improve protein quality, to produce stable protein–protein complexes, and to obtain protein dimers.

Keywords: protein–protein interactions; X-ray crystallography; linkers; weak binding

Introduction

Natural linkers act as spacers between the domains of multidomain proteins. Several interdomain linkers have been identified at the boundaries of functionally distinct domains.[1] These linkers adopt a coiled structure and show a preference for Gln, Arg, Glu, Ser, and Pro amino acids.[1] An automated method to extract interdomain linkers from a dataset of proteins with a known three-dimensional structure revealed that the majority of the interdomain linker sets constitutes Pro, Arg, Phe, Thr, Glu, and Gln residues that act as rigid spacers.[2] Proline is common to many naturally derived interdomain linkers, and structural studies indicate that proline-rich sequences form relatively rigid extended structures to prevent unfavorable interactions between the domains.[3] An independent analysis showed that Thr, Ser, Gly, and Ala are also preferred residues in natural linkers.[4] Besides, flexible Gly-rich regions have been observed as natural linkers in proteins, generating loops that connect domains in multidomain proteins.[5,6] The advancement in recombinant

*Correspondence to: J. Sivaraman, Department of Biological Sciences, 14 Science Drive 4, National University of Singapore, Singapore 117543, Singapore. E-mail: dbsjayar@nus.edu.sg.

DNA technology facilitated fusion of proteins or domains using these synthetic amino acid linkers. In addition to structural studies of protein–protein interactions, a wide range of applications in the field of biotechnology have employed these fused proteins to explore protein-based biochemistry, such as to create artificial bifunctional enzymes,[7] to produce antibodies[8] and proteins with specialized functions,[9] and as tools for FRET analysis.[10]

The characterization of protein–protein interactions is often required to gain an understanding of various biological processes. Yet, the study of protein–protein interactions for many complexes is hampered when one or more partners of the complex are unfolded or unstable. Traditionally, this problem has been addressed by the co-expression and/or co-purification of both proteins.[11] However, for weakly interacting or unstable complexes, the co-expression and/or co-purification often results in a single protein.[12] Protein engineering techniques were another option to address unfolded or unstable proteins, using a single polypeptide chain chimera to link the two binding partners via a flexible amino acid linker [13]. With these chimeric proteins, it was then possible to maintain both the intramolecular and intermolecular protein–protein interactions,[14] and chimeric proteins have been used to generate stable, soluble binary complexes for structural studies, as well as functional dimers.[15]

Linking binding partners using an artificial linker will increase the proximity between the interacting partners and preserve the natural interaction. In cases where the interacting partners are not linked, it is possible that the binding partners might dissociate due to their low affinity and/or due to the crystallization conditions. Most of the artificial linkers used to generate chimeric proteins for structural studies are rich in Gly residues (Table I). Interestingly, the naturally occurring, flexible linkers found in many proteins are also rich in Gly residues. In the next section, we discuss some of these naturally occurring Gly-rich linkers.

Gly-rich linkers in nature

Naturally occurring Gly-rich linkers exist in many proteins and, aside from linking domains, they are known to have a functional role in the protein. The transcription factor PAX6 consists of two DNA-binding domains, a paired domain (PD) and a homeodomain (HD), joined by a Gly-rich linker.[16] Crystal structure analysis of the human PAX6 PD-DNA complex revealed that the extended linker makes minor groove contacts with the DNA.[17] In transmembrane glycoproteins (TMs) of retroviruses, important functional roles are also carried out by the linkers, which mediate membrane fusion through an N-terminal fusion peptide. The fusion peptide is linked to the central coiled-coil core through Gly-rich linkers. The length and amino acid composition of linkers are conserved in many retroviruses and in HA2 of influenza viruses.[18] Alanine and proline scanning mutagenesis experiments showed that the Gly-rich linker of human T-cell leukemia virus Type 1 (HTLV-1) TM, gp21, is involved in membrane fusion function.[18] The N-terminus of the SR protein (sequence-specific RNA-binding factors), SRSF1, contains two RNA recognition motifs (RRM) connected through a Gly-rich linker, which is essential for the binding of exonic splicing enhancers (ESE).[19] In other cases, the role of the Gly-rich linker is still unknown. In filamentous Ff bacteriophages (M13, fd, and f1) minor coat gene 3 protein (g3p) consists of three domains (N1, N2, and CT) connected by a flexible Gly-rich linker;[6] a longer Gly-rich region exists in the g3p of the IKe filamentous phage, but this appears to confer no selective advantage.[20]Gly-rich linkers are also involved in disease conditions. TDP-43 (TAR DNA binding protein) consists of two RNA binding domains (RBD) and a Gly-rich C-terminus. Mutations in the Gly-rich domain of TDP-43 have been related to amyotrophic lateral sclerosis.[21] Gly-rich linkers are also employed for bioimaging studies. Novel fluorogen activating proteins (FAPs) consist of single-chain variable fragments (scFvs), which are made of human immunoglobulin variable heavy (VH) and variable light (VL) domains covalently attached via a Gly- and Ser-rich linker.[22] These FAPs produce fluorescence upon binding to a specific dye or fluorogen and are used in bio-imaging studies to track protein dynamics.

The roles of linkers

Poly-Gly and Gly-rich linkers can be considered as independent units and do not affect the function of the individual proteins to which they attach. The fused proteins behave independently, such that the single chained proteins can perform the combined function of fused partners.[23,24] For example, the native Type I' beta-turn loop in staphylococcal nuclease was replaced with a five-residue turn sequence from concanavalin A, without altering the activity of the nuclease. The crystal structure revealed that the hybrid protein was well folded and the introduced turn sequence retained the conformation of the parent concanavalin A structure.[25] In other systems, however, linker regions can affect the stability, solubility, oligomeric state, and proteolytic resistance of the fused proteins.[24] The repressor protein, Arc, is a dimer with identical subunits. A fused protein—Arc-linker-Arc—was tested for its stability, protein folding kinetics, and biological activity by varying the length and composition of the linker. While the poly-Gly linkers provided maximum conformational freedom, it failed to ensure optimal stability. In contrast, maximum stability was obtained with a linker containing 11 Ala and 5 Gly residues or 7 Ser and 9 Gly

residues, respectively, in the randomized region of the linker.[26] Flexible Gly linkers have been used to improve the folding and function of epitope-tagged proteins. Flexible polypeptide linkers of 5, 8, or 10 Gly residues have been placed between an epitope and a tagged protein to increase epitope sensitivity and accessibility without interfering in protein folding and function.[27,28] Thus, it is important that the length and amino acid composition of a potential linker is optimized in order to preserve the biological activity of the individual proteins in the fused complex.

The loop length created by the linker can have a profound effect on the action of the linker in the fused complex. A systematic study to investigate the role of loop length in protein folding and stability was carried out using the Rop protein as a model system.[24] Rop is a homodimer of helix-loop-helix monomers and, for this experiment, the two-residue loop linker between helix 1 and helix 2 was replaced by a series of non-natural $(Gly)_{1–10}$ linkers, with 1 to 10 Gly residues. As with the wild type Rop protein, all 10 mutants maintained a highly helical dimer structure and retained the same level of wild-type RNA-binding activity. Interestingly, however, the stability of the Rop protein toward thermal and chemical denaturation was inversely correlated with loop length; that is, a longer loop length resulted in a progressively decreased stability.[24] In cases where protein solubility and/or stability are the key problems for protein–protein interaction studies, tethering them to each other can improve the solubility and stability of the complex.[30,31] In some cases, the stability can be improved by altering the linker length and amino acid composition.[26] Thus, Gly-rich linkers are used between two interacting proteins either to improve the stability of one of the binding partners or, where the interaction is transient, to generate a stable protein–protein complex. Here, we present our analysis on various cases where artificial Gly-rich linkers have been used to fuse two proteins for structural studies. Table I provides details about the proteins that were fused and also about the length and composition of the linker used for fusion.

Linkers for unstable proteins and their binding partners

Some proteins exist in complex with their binding partner(s) in order to remain stable in cells.[23] They cannot exist as a single protein and require coexpression with their binding partner when expressed in a heterologous system. Overexpression of single proteins would result in unstable protein molecules or higher molecular weight aggregates, which hampers crystallization. In several cases, this is addressed via the use of flexible Gly-rich linkers. In the following sections, we briefly discuss the utilization of this strategy to generate stable major histocompatibility complex (MHC) Class II molecules, LIM (Lin-11/Islet-1/Mec-3) domains and Pregnane X receptor molecules for structural studies.

MHC class II-peptide complexes. The linkers used in MHC molecules are advantageous for two different scenarios: to stabilize proteins and to trap transient protein–protein interactions. Initially, it was shown that the covalent linkage of a peptide to an MHC Class II molecule could stabilize the MHC molecule.[13] Later, this strategy was applied to obtain the MHC/peptide/TCR (T-cell receptor) ternary complex,[75] which was revealed to be a fast dissociating complex (see Linkers to trap transient protein–protein interactions section).

MHC Class II molecules are heterodimeric proteins comprising noncovalently interacting α and β chains. The extracellular portion consists of two regions: a peptide binding site, consisting of an α1 or β1 domain, and an Ig-like domain, consisting of an α2 or β2 domain. When expressed as two different chains (α and β) and mixed together to form a αβ heterodimer, MHC Class II molecules resulted in high molecular weight aggregates, suggesting the possible heterogeneity of αβ heterodimers in the presence of free α and β chains. However, stable αβ heterodimers were produced when an antigenic peptide, recognized by a particular MHC Class II molecule, was covalently linked to the N-terminus of the MHC Class II β1 chain by a 16-amino acid Gly-rich linker (GGGGSLVPRGSGGGGS). The presence of linker had no effect on the recognition of peptide-MHC Class II complex by the specific T-cell hybridomas.[13] This particular linker has a thrombin cleavage site flanked by Gly-rich residues, so that the linker between the fused proteins can be cleaved. However, the crystal structure of the human class II MHC protein, HLA-DR1, complexed with an unlinked influenza virus peptide (PDB: 1DLH) revealed that a six-amino acid linker is sufficient to connect the C-terminus of the peptide to the N-terminus of the MHC class II molecule[76]. Further, tethering an ovalbumin (OVA) peptide to the N-terminus of the β1 chain of the $I\text{-}A^d$ MHC Class II molecule using a six-amino acid linker (Table I), was sufficient to enhance the stability of αβ heterodimers (Fig. 1).[78] Further crystallization trials found that many conditions could produce diffraction-quality crystals. This experimental approach led to the successful crystallization of other MHC Class II-peptide complexes using a polypeptide linker.[31–44]

LIM domain and related proteins. LIM (Lin-11/Islet-1/Mec-3) domains are a class of zinc binding domains that mediate protein–protein interactions to regulate cell fate.[23] LIM domain-containing proteins are divided into three groups: LIM-only

Table I. *Structurally Known, Covalently Linked Complexes*

N-terminus protein/ peptide	Linker	C-terminus protein/peptide	PDB	Method	Reference
Unstable proteins and their interacting partners:					
a. MHC Class II molecules and peptide complexes:					
HSP70 peptide	GGGGSLVPRGSGGGGS	MHC Class II I-E^k	1IEB	X-ray	31
Influenza virus hemagglutinin peptide	GSGSGS	MHC Class II I-A^d	1IAO	X-ray	32
Ovalbumin peptide	GSGSGS	MHC Class II I-A^d	2IAD	X-ray	32
Hen egg lysozyme	GGGGSLVPRGSGGGGS	MHC Class II I-A^k	1IAK	X-ray	33
Hen egg lysozyme	GGGGSLVPRGSGGGGS	MHC Class II I-Ag^7	1F3J	X-ray	34
pHB (64–76 of the d allele of mouse β globin)	GGGGSLVPRGSGGGGS	MHC Class II mut IE^k (E11Q, D66N on α chain)	1I3R	X-ray	35
Mylein Basic Protein (1-11)	eight residues of GS	MHC Class II I-Au	1K2D	X-ray	36
Pigeon Cytochrome c	GGGGSLVPRGSGGGGS	MHC Class II I-E^k	1KTD	X-ray	37
Moth Cytochrome c	GGGGSLVPRGSGGGGS	MHC Class II I-E^k	1KT2	X-ray	37
Eα 3K peptide	GGGGSLVPRGSGGGGS	MHC Class II I-A^b	1LNU	X-ray	38
class II invariant chain-associated peptide (CLIP)	GSGSGS	MHC Class II I-A^b	1MUJ	X-ray	39
hypocretin	GGGGSLVPRGSGGGG	MHC class II DQ0602	1UVQ	X-ray	40
Mylein Basic Protein (124-135)	GSGSGS	MHC Class II I-A^u	2P24	X-ray	41
Tu elongation factor	GGGGSLVPRGSGGGGS	DR52c	3C5J	X-ray	42
HLA-DR α-chain	GGGGSLVPRGSGGGGS	HLA-DP2	3LQZ	X-ray	43
Gliadin	GGGGSLVPRGSGGGGS	HLA-DQ2.3	4D8P	X-ray	44
b. LIM binding and related proteins:					
LIM only protein, LMO2	GGSGGHMGSGG	LIM domain-binding protein 1(Ldb1)	—	NMR	45
LIM only protein, LMO4, LMO2	GGSGGHMGSGG	LIM domain-binding protein 1(Ldb1)	1M3V, 1J2O	NMR	29
LIM only protein, LMO4	GGSGGSGGSGG	LIM domain-binding protein 1(Ldb1)	1RUT	X-ray	23
LIM only protein, LMO4	GGSGGSGGSGG	LIM domain-binding protein 1(Ldb1)	2DFY	X-ray	46
LIM homeobox protein 3, Lhx3	GGSGGHMGSGG	Isl1 (Islet-1)	2RGT	X-ray	47
LIM domain-binding protein 1(Ldb1)	GGSGGHMGSGG	LIM homeobox protein 3, Lhx3	2JTN	NMR	48
LIM only protein, LMO2-LIM2	GGSGGSGGSGG	LIM domain-binding protein 1(Ldb1)	—	NMR	48
LIM only protein, LMO2	GGSGGSGGSGG	LIM domain-binding protein 1(Ldb1)	2L6Y, 2L6Z	NMR	49
LIM homeobox protein 4, Lhx4	GGSGGHMGSGG	Isl2 (Islet-2)	3MMK	X-ray	30
LIM only protein, LMO4	GGSGGHMGSGG	CtlP (C-terminal binding protein interacting protein)	—	NMR	50
c. PXR-ligand binding domain and SRC-1:					
PXR-LBD	GGSGG	SRC-1	3CTB, 3CTC, 3HVL	X-ray	51
Transient protein–protein interactions:					
a. TCR-peptide-MHC molecule complexes:					
1. Peptide bound to MHC molecule:					
hen egg CA peptide (residues 134 to 146)	GGGGSLVPRGSGGGGS	MHC Class II I-A^k	1D9K	X-ray	52
Eα 3K peptide	GGGGSLVPRGSGGGGS	MHC Class II I-A^b	3C6L, 3C5Z, 3C60	X-ray	53
Mylein Basic Protein (114-126)	GGGGSLVPRGSGGGGS	HLA-DR4	3O6F	X-ray	54
2. Peptide bound to TCR:					
Hemagglutinin HA peptide	GGSGGGGG	TCR HA 1.7	1FYT	X-ray	55
Hemagglutinin HA1	GGSGGGGG	TCR HA1.7	1J8H	X-ray	56

Table 1 *(Continued)*

N-terminus protein/ peptide	Linker	C-terminus protein/peptide	PDB	Method	Reference
Mylein Basic Protein (89-101)	GGSGGGGG	TCR 3A6	1ZGL	X-ray	57
Mylein Basic Protein (85-99)	GGSGGGGG	TCR Hy.1B11	3PL6	X-ray	58
b. Others:					
Phosphoprotein (457–507) of paramyxoviral polymerase from measeles virus	GSGSGSGS	Nucleocapsid protein (486–505) of measeles virus	1T6O	X-ray NMR	14
FtsY	31 amino acids glycine-serine rich linker	Ffh	2XKV	Cryo-EM	59
Minor coat gene 3 protein (g3p)	GGGSEGGGSEGGGSE GGG	Integral membrane protein TolA (coreceptor)	1TOL	X-ray	6
Asf1 (anti-silencing function 1)	AAGAATAA	Histone H3	2IDC	X-ray	60
CaM	GGGGG	CBD of calcineurin	2F2P, 2F2O	X-ray	61
Covalently linked dimers:					
a. HIV-1 PR:					
HIV-1 protease	GGSSG	HIV-1 protease	1HPX	X-ray	62
C95M HIV-1 protease	GGSSG	HIV-1 protease	1G6L	X-ray	63
C95M HIV-1 protease	GGSSG	C1095A HIV-1 protease	1LV1	X-ray	64
HIV-1 protease	GGSSG	HIV-1 protease	2NPH	X-ray	65
HIV-1 protease	GGSSG	HIV-1 protease	3DOX	X-ray	66
HIV-1 protease	GGSSG	HIV-1 protease	2WHH	X-ray	67
G48V/C95F HIV-1 protease	GGSSG	G48V/C95F HIV-1 protease	3N3I	X-ray	68
HIV-1 protease	GGSSG	HIV-1 protease	3MIM	X-ray	69
HIV-1 protease	GGSSG	HIV-1 protease	4EP2, 4EPJ, 4EP3, 4EQJ, 4EQ0	X-ray	70
b. TTR dimer:					
Transthyretin	GSGGGTGGGSG	Transthyretin	1QWH	X-ray	71
GluR S1-S2 complex:					
GluR0 S1	GT	GluR0 S2	1II5, 1IIT, 1IIW	X-ray	72
GluR2 S1	GT	GluR2 S2	1FTK, 1FTL, 1FTO, 1FW0, 1FTJ, 1FTM	X-ray	73
GluR5/6 S1	GT	GluR5/6 S2	1S9T, 1SD3, 1S50, 1S7Y	X-ray	74
Recently determined structures:					
Fat domain of focal adhesion kinase	GSGSGSGSGGSG GSGGSGGSGGSGGS	Ld4 motif of paxillin	2L6H	NMR	—
Fat domain of focal adhesion kinase	GSGGSGSGGSGGSG	Ld2 motif of paxillin	2L6G	NMR	—
Fat domain of focal adhesion kinase linked to Ld2 motif of paxillin at its C-terminus via GSGGSGSGGSGG-SG linker	GSGSGSGSGGSG GSGGSGGSGGSGGS	Ld4 motif of paxillin	2L6F	NMR	—

(LMO), LIM homeodomain (LIM-HD) proteins, and LIM kinase families. Generally, LMO proteins are localized in the nucleus, but lack a nuclear localization sequence. Thus, to keep these proteins in the nucleus they must be bound to LIM domain binding protein-1 (ldb1). LMO1, 2, and 4 play a major role in normal and leukemic T-cell development, and the LMO-ldb1 complexes may modulate gene expression.[79] LMO proteins are insoluble and unstable when expressed by bacteria; but this can be rectified when they are fused with ldb1 using either a GGSGGHMGSGG or a GGSGGSGGSGG linker at

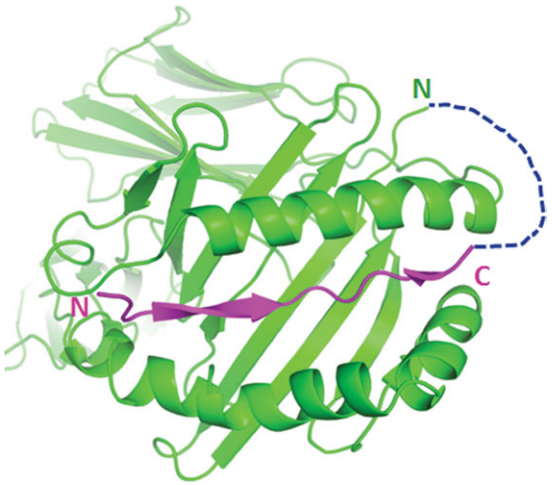

Figure 1. Crystal structure of MHC Class II-peptide complex. Ribbon representation of MHC Class II I-A^d (green) and ovalbumin peptide (magenta) complex (PDB code: 2IAD), where the C-terminus of ovalbumin peptide is linked to the N-terminus of the β chain of MHC Class II I-A^d molecule using a GSGSGS (6aa) linker. The electron density for the linker was not observed and the possible position was indicated as blue dotted lines. All structure-related figures of this review were prepared using the program PyMol.[77]

the C-terminus.[80] The LIM interaction domain (LID) of ldb1 (300–340) is unstructured in solution.[29] Upon binding to the N-terminal LIM domains of LMO2 and LMO4, the LID of ldb1 adopts an extended conformation, stretched along the face of the LIM domain [Fig. 2(A)] to facilitate binding.

Likewise, LIM-HD proteins tend to be insoluble and unstable. There are 12 mammalian LIM homeodomain proteins consisting of six pairs of paralogues, which often have overlapping expression patterns and functions. Two such paralogous pairs are LIM homeobox protein 3 and 4 (Lhx3 and Lhx4) and Islet-1 and -2 (Isl1 and Isl2). When expressed in bacteria, the LIM domains from Lhx3 and Lhx4 tend to aggregate and are largely insoluble. Linking an interaction partner, such as Isl1 or ldb1, to the LIM domains can prevent this aggregation and allow production of soluble tethered complexes[81] [Fig. 2(B)]. Since then, various LIM domain structures (Table I) have been solved by exploiting the same technique.[23,29,30,45–50] Figure 2(A,B) shows the interaction between lbd1 and the different LIM domains. These structures were solved using X-ray and NMR techniques. The backbone assignment by NMR experiments for the Gly-rich linker revealed that the linker exists as a random coil.[47]

PXR-LBD and SRC-1 complex. Pregnane X receptor (PXR) is a xenobiotic receptor and a ligand-inducible transcription factor involved in regulating drug metabolizing enzymes and transporters. PXR contains two domains: a highly conserved DNA-binding domain (DBD) at the N-terminus and a more divergent ligand-binding domain (LBD) at the C-terminus. The LBD of PXR is involved in ligand recognition and also interacts with co-activator proteins. PXR is an unstable protein and exists as a complex with its co-repressor. Ligand binding

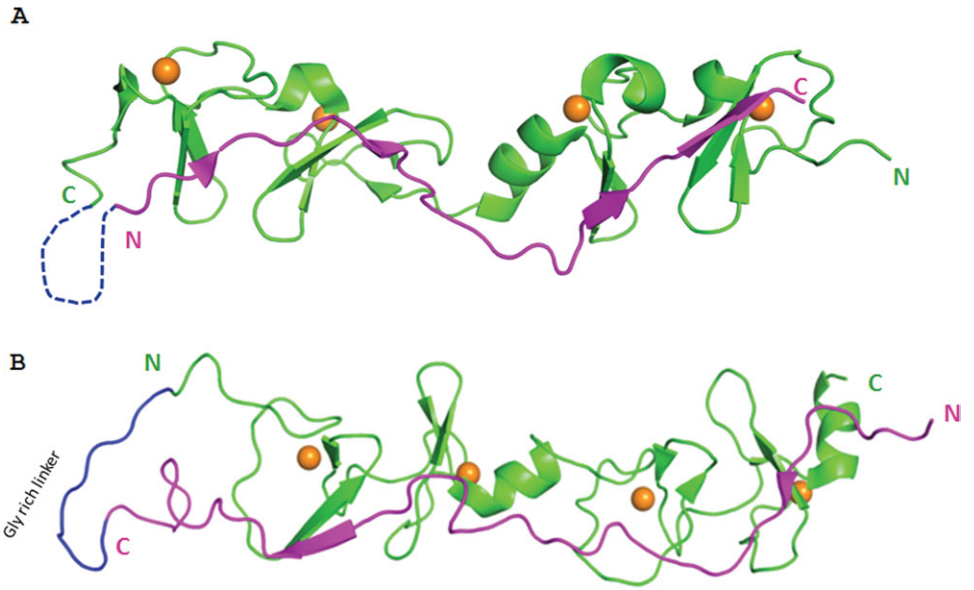

Figure 2. Structural details of LIM domains and their binding partner complexes. (A) Ribbon representation of the crystal structure of LMO4 (green)-ldb1 (magenta) complex (PDB code: 1RUT), where the C-terminus of LMO4 is linked to the N-terminus of Ldb1 using a GGSGGSGGSGG (11aa) linker. The blue dotted line represents the possible position of the missing amino acids of the linker. (B) NMR structure of Lhx3 (green)-ldb1 (magenta) complex (PDB code: 2JTN), where the C-terminus of ldb1 is linked to N-terminus of Lhx3 using a GGSGGHMGSGG (11aa) linker (dark blue). Zn^{2+} ions are shown as orange spheres.

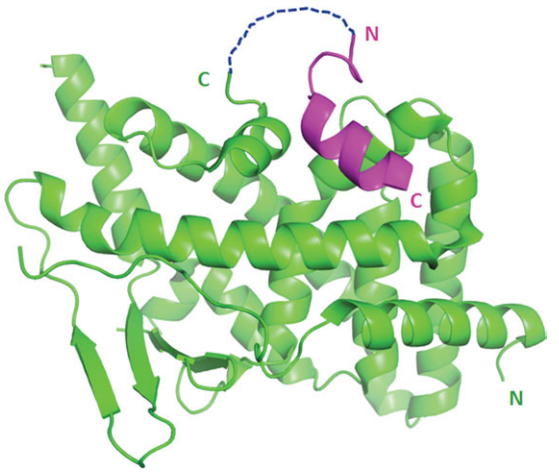

Figure 3. Crystal structure of the ligand binding domain of Pregnane X receptor (PXR-LBD) and steroid receptor activator 1 (SRC-1). Ribbon representation of PXR-LBD (green)—SRC-1 interacting peptide (magenta) complex, where the C-terminal of PXR-LBD is linked using a GGSGG (5aa) linker to the N-terminus of the SRC-1 interacting peptide (PDB code: 3CTB). The blue dotted line shows the possible positions of the missing amino acids of the linker.

displaces the co-repressor by a co-activator, the steroid receptor activator 1 (SRC-1). Studying the interaction between the PXR-LBD and SRC-1 is important in order to understand the mechanism of ligand-mediated transcriptional effects. A dual expression system was initially used to characterize this binding, but resulted in the production of unstable proteins and a stoichiometry of SRC-1 to PXR of less than 1:1. Watkins *et al.* purified PXR using the SRC-1 domain and then combined this PXR with a SRC-1 interacting peptide to form a complex and determine the crystal structure (PDB: 1NRL).[82] They found that the C-terminus of PXR lies near the N-terminus of SRC-1, at a distance of 18 Å, which would require at least five amino acids to link these proteins. Thus, the SRC-1 peptide was fused to the C-terminus of the PXR-LBD using various length linkers such as SGGGSSHS, SGGSGGSSHS, and SGGSGGSGGSSHS (underlined sequences are the added linker; the flanking regions are from the interacting proteins), with most crystals obtained with the 10-amino acid Gly-rich linker[51] (Fig. 3).

Linkers to trap transient protein–protein interactions

Several key biological events are dictated by protein–protein interactions, most of which are transient and weak. Trapping a weak or transient interaction for structural studies is always challenging, and flexible Gly-rich linkers have thus proven helpful in these instances; for example, trapping of fast dissociating reactions, such as the MHC molecule-bound peptide recognition by TCR; the interaction between a phosphoprotein and a nucleocapsid protein of the paramyxovirus; as well as trapping low-affinity binding interactions, such as in the case of g3p-TolA and Histone H3-Asf1 (Anti-silencing function 1). Further, in the case of calmodulin (CaM) and the calmodulin binding domain (CBD) of calcineurin, this strategy has been utilized to trap an unstructured protein that gains secondary structure upon binding with its partner. This wide usage of linkers is briefly discussed in the following sections.

T-cell receptor, peptide and MHC class II complexes. T-cell receptors (TCR) are found on the surface of T lymphocytes and recognize antigens bound to major histocompatibility complex (MHC) molecules. When the TCR engages with an antigen and MHC, it becomes activated to enhance the immune response. However, studying these interactions has proven difficult in the past due to a very low affinity between the TCR and the peptide/MHC molecule. Kinetic studies also indicated a slow association and fast dissociation, which made it difficult to trap this transient interaction for further studies.[75,83] Hence, to stabilize the TCR/peptide/MHC complex for crystallization, the antigen peptide was covalently linked using a Gly-rich linker at the N-terminus of the MHC molecule β chain (see also MHC class II-peptide complexes section); this linker helped to mediate the formation of a stable TCR-antigen peptide-MHC molecule ternary complex.[52–54] Another way to stabilize this complex was shown using an eight-amino acid polypeptide (GGSGGGGG) to link the antigen peptide and the N-terminus of the variable β chain of TCR.[55–58] The overall structure of the TCR-peptide-MHC and the MHC-peptide-TCR molecules did not show any differences in the binding regions of MHC and TCR molecules using both methods [Fig. 4(A,B)]; thus, the structure was able to be solved as a ternary complex. Table I outlines the various TCR-peptide-MHC ternary complexes that were solved using these methods.

Phosphoprotein-nucleocapsid protein complex. Within the virion of the paramyxovirus family of viruses, the nucleocapsid protein (N) packs the genomic RNA into a helical protein-RNA complex known as a nucleocapsid. It is used as a template for the transcription of mRNAs by a viral RNA polymerase, which consists of two components: the large protein and the phosphoprotein (P). While the large protein is involved in the catalytic activities of the polymerase, the phosphoprotein is involved in binding the polymerase to the nucleocapsid through its C-terminal region (459–507). The binding site for the phosphoprotein has been mapped to the C-terminus (486–505) of the nucleocapsid protein.[84] Kinetic studies showed that this is a fast-associating reaction, with a weak binding affinity. Hence, to stabilize

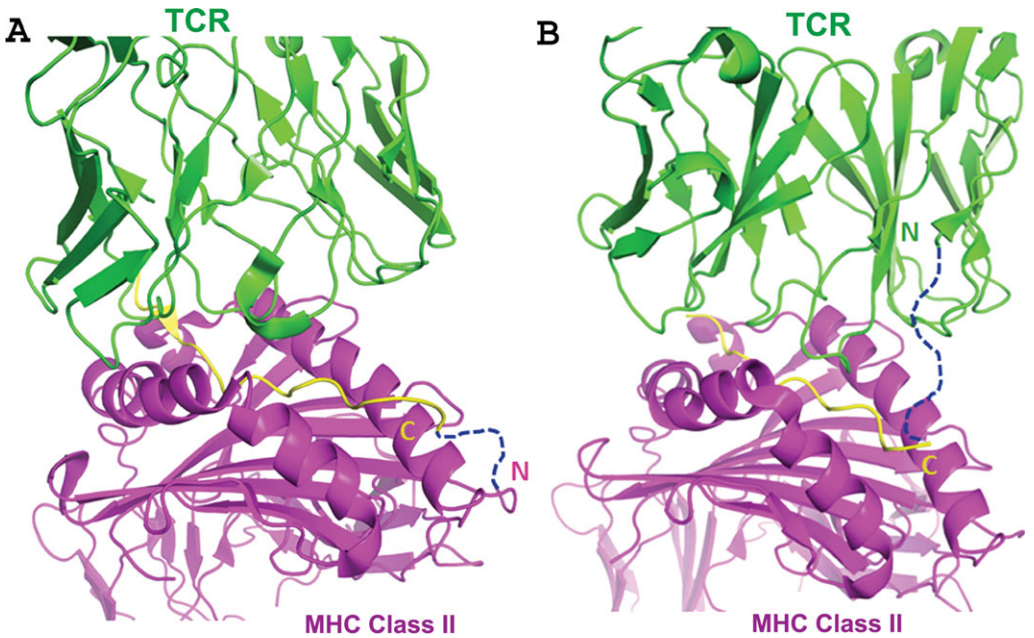

Figure 4. Crystal structures of different MHC molecule-peptide-T cell receptor (TCR) complexes. (A) Ribbon representation of TCR D10 (green)-hen egg CA peptide (yellow)-MHC Class II I-A^k (magenta) complex (PDB code: 1D9K), where the C-terminus of the CA peptide is linked to the N-terminus of the β chain of the MHC class II I-A^k using a GGGGSLVPRGSGGGGS (16aa) linker. An interactive view is available in the electronic version of the article. (B) Ribbon representation of TCR HA 1.7 (green)-Hemagglutinin (HA) peptide (yellow)-MHC molecule HLA-DR1 (magenta) complex (PDB code: 1FYT), where the C-terminus of the HA peptide is linked to the N-terminus of the variable β chain of the TCR HA 1.7 using a GGSGGGGG (8aa) linker. The blue dotted line shows the possible positions of the missing amino acids of the linker. An interactive view is available in the electronic version of the article.

this complex, the C-terminus of protein P was linked using a Gly- and Ser-rich linker (GSGSGSGS) to the N-terminus of protein N, and the linked complex was crystallized. This structure showed anti-parallel packing of the helix from the N protein [Fig. 5(A)]. However, the angle between the N protein helix and the final helix of the P protein suggests that it could be an intermolecular interaction between P and N proteins instead of intramolecular interaction.[14]

Ribosomes-SRP-FtsY Complex. *Escherichia coli* signal recognition particle (SRP) consists of a protein (Ffh) and 4.5S RNA. Ffh and FtsY (receptor of SRP) consist of an NG domain, a GTPase G domain, and an N domain. An association between SRP and FtsY via the NG domain is required for the delivery of ribosome-nascent chain complexes to the membrane, and a reciprocal GTPase activation coordinates the transfer of the signal sequence to the translocon.[59] Associations of SRP and FtsY involve a series of GTP-independent and -dependent steps, and form a relatively unstable complex that is visualized by cryo-EM. However, the interaction between Ffh and FtsY is stabilized by fusion of the C-terminus of FtsY to the N-terminus of Ffh via a 31-residue Gly- and Ser- rich linker, yielding a single-chain construct (scSRP).[59] The single chain scSRP behaves similar to an unlinked SRP and FtsY, and binds ribosomes efficiently, without altering the GTPase activity. The cryo-EM structure of scSRP bound to the ribosome was thus described.[59]

g3p-TolA complex. Ff phages are used as a selection tool by displaying proteins and peptides on the phage surface by genetic fusion with minor coat gene 3 proteins (g3p). g3p consists of three domains (N1, N2, and CT) connected to each other by natural, flexible Gly-rich linkers. Phage infection of *E. coli* cell progresses by the interaction between the N2 domain and the tip of an F pilus, with the pilus retracted by an unknown mechanism.[85] At this juncture, the N1 domain interacts with the C-terminal (D3) domain of the integral membrane protein TolA, an interaction which is not clearly understood, and leads to entry of the phage genome into the bacterial cytoplasm. To further understand the mechanism of host cell infection by phages, the interaction between the N1 domain of g3p and the D3 domain of TolA was required.[6] However, the affinity between these two domains is weak.[86] Hence, a Gly-rich linker ($(G_3SE)_3G_3$) was generated to fuse the C-terminus of the N1 domain of g3p and the N-terminus of the D3 domain of TolA. A long flexible linker was naturally present between the N1 and N2 domains

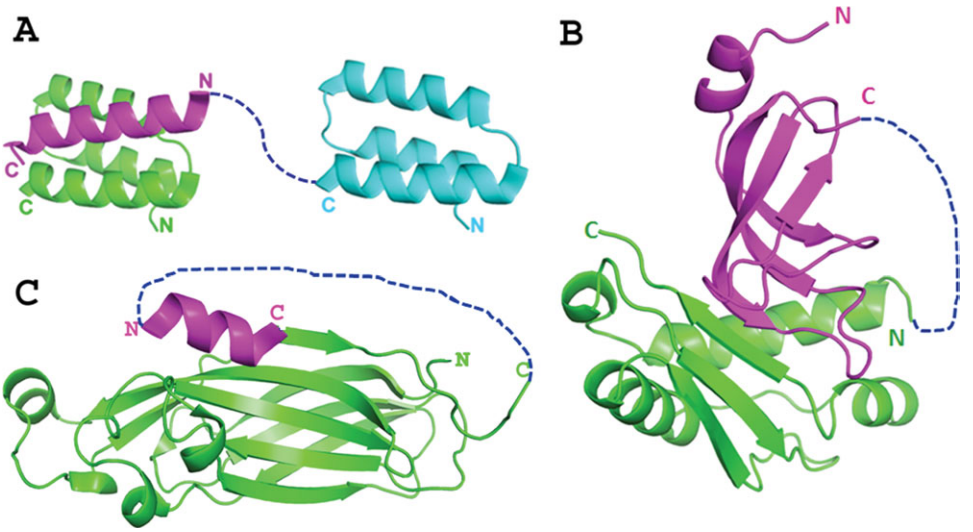

Figure 5. Structural details of transient protein–protein interactions using Gly-rich linkers. (A) Ribbon representation of the crystal structure of phosphoprotein ($P_{457–507}$) of paramyxoviral polymerase (green)–nucleocapsid protein ($N_{486–505}$) (magenta), where the C-terminus of $P_{457–507}$ is linked using a GSGSGSGS (8aa) linker to the N-terminus of the nucleocapsid protein ($N_{486–505}$) (PDB code: 1T6O). The interacting peptide (magenta) was from the adjacent symmetry-related molecule (cyan). (B) Ribbon representation of the N1 domain of g3p (magenta)—D3 domain of TolA (green) complex, where the C-terminus of g3p is linked using a GGGSEGGGSEGGGSEGGG (18aa) linker to the N-terminus of TolA (PDB code: 1TOL). (C) Ribbon representation of Anti-silencing function 1 (Asf1; green)–α3 helix of Histone H3 (magenta) complex (PDB code: 2IDC). The linker amino acids (AAGAATAA (8aa)) were not present in the pdb except for first two Ala residues. The predicted linker position is shown as a blue dotted line in all panels.

of g3p, and hence the use of a long linker between the N1 domain of g3p and the D3 domain of TolA did not impose any constraints [Fig. 5(B)].

Histone H3-Asf1 complex. Chromatin structure and organization is important during the various stages of the cell cycle. The initial steps of chromatin organization involve the deposition of a $(H3/H4)_2$ heterotetramer onto the DNA. Asf1 (anti-silencing function 1) is one of the several histone chaperones involved in this deposition. NMR and two-hybrid experiments have shown that the C-terminus of the α3 helix of yeast histone H3 interacts with Asf1,[87,88] but this interaction is weak.[87] Hence, to obtain a stable complex, the α3 helix of Histone H3 was linked using an alanine-rich linker (AAGAATAA) to Asf1[60] [Fig. 5(C)]. To our knowledge this is the only case where an alanine-rich linker was used.

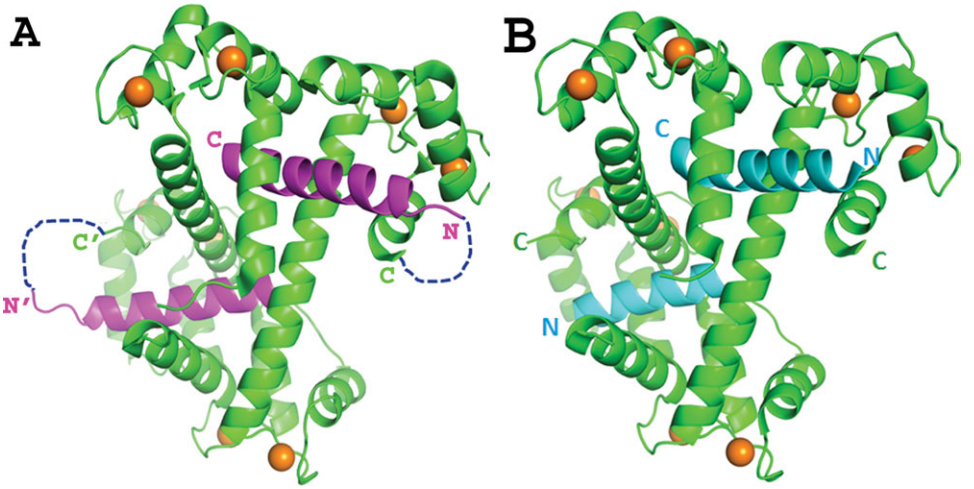

Figure 6. Structural similarities between linked and unlinked complexes of Calmodulin and the Calcineurin calmodulin binding domain (CBD). (A) Ribbon representation of Calmodulin (green)–CBD of Calcineurin (magenta) complex, where the C-terminal of Calmodulin is linked using a GGGGG (5aa) linker to the N-terminus of the Calcineurin CBD (PDB code: 2F2P). The predicted linker position is shown as a blue dotted line. (B) Ribbon representation of Calmodulin (green)–CBD of Calcineurin (cyan) complex, generated by co-crystallization (PDB code: 2R28). Both structures were found to be similar. Ca^{2+} ions are shown as orange spheres.

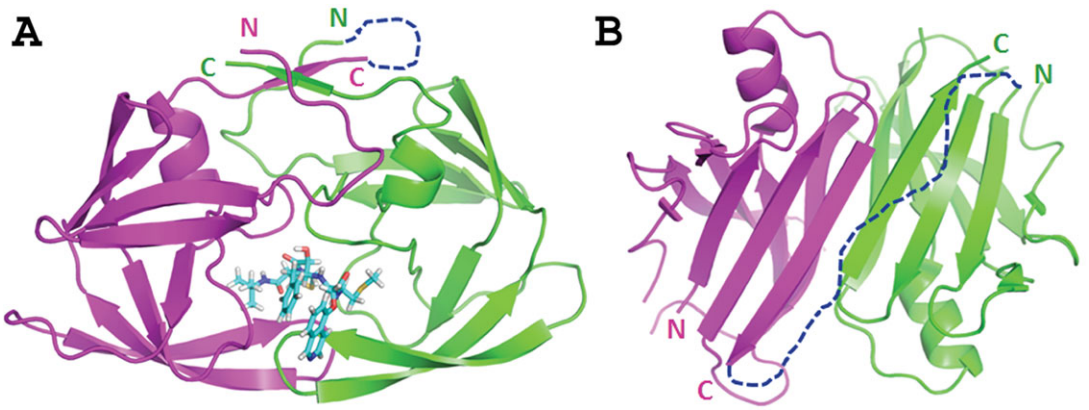

Figure 7. Crystal structure of covalently linked dimers. (A) Ribbon representation of covalently linked HIV-1 protease (PR) dimer (PDB code: 1HPX), where the C-terminus of one monomer (magenta) is linked to the N-terminus of another (green) using a GGSSG (5aa) linker (shown as a blue dotted line) in the presence of the protease inhibitor compound, KNI-272 (represented as light blue sticks). (B) Ribbon representation of covalently linked transthyretin dimer (PDB code: 1QWH), where the C-terminus of one monomer (magenta) is linked to the N-terminus of another (green) using a GSGGGTGGGSG (11aa) linker. The missing amino acids of the linker are shown as a blue dotted line.

CaM-calcineurin CBD complex. A similar approach was adopted to study the complex structure of calmodulin (CaM) and the calmodulin-binding domain (CBD) of calcineurin. Calcineurin is a CaM-dependent serine/threonine protein phosphatase involved in various signaling pathways.[89] The CBD of calcineurin is largely disordered in the absence of CaM and attains a secondary structure upon its interaction with CaM.[90] To facilitate the co-crystallization of the CaM-CBD peptide complex, the C-terminus of CaM was linked to the N-terminus of the CBD peptide using a five-amino acid Gly linker (GGGGG).[61] Later, the same group determined the crystal structure of CaM-CBD peptide by co-crystallization in the absence of a linker (PDB: 2R28) and found that the structures of CaM-CBD peptide with and without the linker were same (Fig. 6); this result proved that the presence of the flexible Gly-rich linker did not restrain the interactions between CaM and CBD peptide.[91]

Linkers to retain dimerization

Functional dimers are generated by fusing two monomers via a Gly-rich linker, and they are often involved in various enzymatic activities. The HIV-1 protease (HIV-PR) and other retroviral proteases are dimeric molecules that consist of two identical subunits. Crystal structures of HIV-PR have shown that the dimer interface of the protease contains the amino- and carboxyl-terminal strands of each monomeric subunit.[92] This tethered dimer provides a means to determine the effect of changes to one of

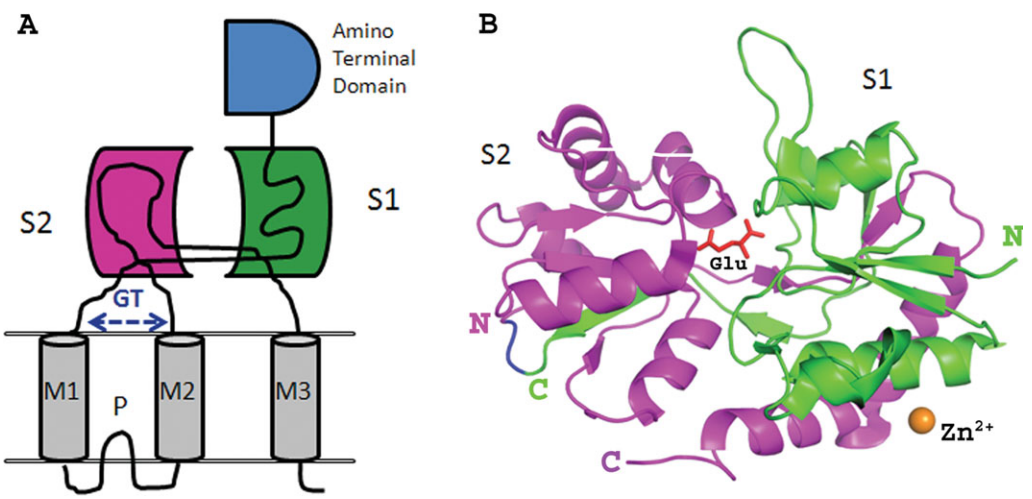

Figure 8. Structural details of the ligand binding core of Glutamate receptors (GluRs). (A) Cartoon representation of the domain structure of glutamate receptor ion channels (GluRs). The S1 and S2 domains are linked using a GT peptide to generate a ligand binding core for structural studies. (B) Ribbon representation of the ligand binding core of GluR2, where the C-terminal of S1 (green) is linked to S2 (magenta) using a GT di peptide (dark blue) in complex with glutamic acid (Red) (PDB code: 1FTJ). Zn^{2+} ion is shown as an orange sphere.

the two subunits on the activity of HIV-PR.[15] Thus, a covalently linked dimer of HIV-PR was constructed with a 5-amino acid (GGSSG) linker that connected the C-terminus of one monomer to the N-terminus of the other. This covalently linked dimer showed enzymatic properties very similar to those of the natural dimer and was found to be more stable than the natural HIV-PR dimer at pH 7.0, where the natural dimer dissociates. Figure 7(A) shows an example of covalently linked HIV-1 PR dimer. Later structures using various mutants of the HIV-PR tethered dimer and the structures of HIV-PR tethered dimer in the presence of substrate peptides were determined as shown in Table I.[62–70]

Similarly, transthyretin (TTR), a carrier of the thyroid hormone, thyroxine, and the holo-retinol binding protein, is known to form a tetramer, with two subunits each consisting of a dimer. TTR amyloid diseases are caused due to misfolding, aggregation and deposition of wild-type TTR in cardiac tissue and other destabilizing mutations that make TTR deposition prone in a variety of tissues. The rate-limiting dissociation of the tetramer into its two subunits is necessary for amyloid fibril formation.[93] Experiments were conducted to determine the effect of tethering these closely interacting subunits on the kinetic stabilization of the tetramer. A tethered construct (TTR-GSGGGTGGGSG-TTR)$_2$ was generated to stabilize the tetramer [Fig. 7(B)]. The linked complex was highly stable and also structurally and functionally similar to wild-type TTR. Besides, it was reported that urea was unable to denature the tethered dimer whereas guanidine hydrochloride treatment was able to denature the tethered dimer; the authors suggest that tethered complex is kinetically, rather than thermodynamically, stable.[71] This study also showed that tethering two subunits can dramatically reduce tetramer dissociation.

Linker to fuse domains within a protein

In the mammalian nervous system, fast synaptic transmission between nerve cells is carried out mainly by ionotropic glutamate receptors (iGluRs), a class of transmembrane proteins that form glutamate-gated ion channels. They are composed of distinct channel-forming and ligand binding domains, where many synthetic agonists and antagonists are known to bind.[72–74] The S1 and S2 domains form the core ligand binding region of these proteins. Although S1 and S2 form a ligand binding core, they are separated by three membrane spanning segments. Hence, to generate a ligand binding core, S1 and S2 were covalently linked (Fig. 8). For structural studies of the ligand binding domain, a Gly-Thr dipeptide linker was sufficient to covalently link the S1 and S2 domains of the ligand binding core of the glutamate receptors, and mimic its binding affinity for various agonists.[94]

Linker selection criteria

The selection of a linker sequence and length is dependent on the construction of functional chimeric proteins, and therefore, the optimal linker length will vary on a case by case basis. *De novo* linker design will be successful if the site of interaction between the two proteins is approximately known. More often, the existing knowledge of the known homologous complex structures is used to design linkers of an appropriate length that will mimic the natural interaction. The actual distance between the site of the interaction and the nearby N- or C-terminus to which the binding partner can be fused will provide clues for the *de novo* design of a linker or an appropriate length. Often, binding regions on the surface of proteins can be mapped using various biophysical and/or mutational analyses[95] and, based on these details, linker length and composition can be designed. In some cases, linking two proteins with a linker that is hypothesized to promote an intramolecular interaction of a chimeric protein may not be sufficient. Instead, it can form an intermolecular interaction, similar to the case of the phosphoprotein ($P_{459–507}$) and the nucleocapsid protein ($N_{486–505}$).[84] Similar binding was observed in our recent studies with CaM and the Neurogranin IQ motif peptide, where the binding partners are linked. We noticed that the peptide and protein from the adjacent molecules were interacting with each other (unpublished data). Our results and the results of others clearly show that if the linker length is not optimal, the linked partner will engage with an adjacent molecule to retain its specificity and natural interactions (intermolecular interaction). Thus, the linkers should be in an optimum length to promote either intramolecular or intermolecular interactions.

In general, the minimum and maximum lengths of linkers used in the studies reviewed here were between 2 and 31 amino acids (Table I), with a common linker lengths of 5 to 11 amino acids for structural studies. The length of the linker mainly depends on the system under study and a preunderstanding of the binding region may help to predict the approximate length of the linker. Apart from length of the linker, the composition should also be optimized. Not all amino acids make a perfect choice for linkers, since a certain degree of flexibility and hydrophilicity of the linker are important to retain the functions of the individual domains and allow the fused proteins to interact with each other.[4] Gly has low preference to form an α-helix; thus, the lack of a sidechain maximizes the freedom of the backbone conformation.[96] Furthermore, Ser and Thr are polar residues that prefer to interact with the solvent than with the fused proteins.[4] Thus, polypeptides rich in Gly, Ser, and Thr offer special advantages: (i) rotational freedom of the polypeptide backbone, so that the adjacent domains are free to

move relative to one another,[44] (ii) enhanced solubility,[26] and (iii) resistance to proteolysis.[97] The backbone dihedral angles of residues in linkers are highly variable; that is, in Ramachandran plots, they occupy a significantly larger area than either the α-helices or β-sheets.

Our literature analyses show that various compositions of linkers have been used for structural studies (Table I). In the case of MHC Class II-peptide complex, two different types of linkers (with and without thrombin cleavage site) were used, while in the case of MHC-peptide-TCR ternary complex, peptide linking was achieved through two different methods (peptide linked to either MHC or TCR molecule) (Table I). In other structural studies, linkers with almost equal composition of Gly and Ser were used, such as for the phosphoprotein-nucleocapsid protein complex, the ribosomes-SRP-FtsY complex and the Fat domain of focal adhesion kinase complexes. By comparison, for the linked complexes of the LIM domain, PXR-LBD, peptide bound to TCR-MHC, g3p-TolA, HIV-1 PR, and TTR covalently dimers, various lengths of Gly- and Ser-rich linkers were employed, for which over 60% of the linker amino acids were Gly. For structural studies of CaM-CBD of calcineurin complex, the study integrated a 5-amino acids linker with all Gly residues. Hence, a suitable linker might comprise a combination of small side chained amino acids, such as Gly and Ser, with an overall higher percentage of Gly.

Apart from designing linkers (length and amino acid composition), optimizing conditions to retain the interactions between the linked partners is important. In general, the flexible Gly-rich linkers do not alter the properties of the proteins to which they are bound, and permit the natural interaction to be retained. Optimized conditions for the interaction between the linked partners should be the same as that without the linker. This should be confirmed using unlinked interacting partners in various *in vitro* methods, such as immunoprecipitation or ITC/SPR experiments.

Concluding Remarks

Gly-rich linkers are naturally occurring separators, connecting domains within proteins while allow discrete functions of the domains. This system has now been widely exploited synthetically to optimize protein–protein interactions that are weak, transient, and otherwise complicated and have hitherto escaped being exposed using crystallography-based techniques. The recombinant fusion of interacting proteins through suitable linkers offer a unique advantage in retaining stable complex among the weak affinity or transiently interacting partners or binding with the unstructured partners. The studies explored here indicate that linkers are highly flexible and do not hamper the interactions between the linked proteins. This linker approach can be extended to other protein–protein interactions or to fuse different domains for structural studies. Various compositions of linkers have been used for structural studies, focusing on Gly and other small side-chained amino acids. However, in the future, other amino acids may be explored for their worth in the design of new artificial linkers.[98]

Acknowledgements

We are grateful to the Academic Research Funding (AcRF), National University of Singapore (NUS), Singapore, for the support of this study. VP is a graduate scholar in receipt of a research scholarship from the NUS.

References

1. Wootton JC, Drummond MH (1989) The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. Protein Eng 2:535–543.
2. George RA, Heringa J (2002) An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng 15:871–879.
3. Radford SE, Laue ED, Perham RN, Miles JS, Guest JR (1987) Segmental structure and protein domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Genetic reconstruction in vitro and 1H-n.m.r. spectroscopy. Biochem J 247:641–649.
4. Argos P (1990) An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol 211: 943–958.
5. Steinert PM, Mack JW, Korge BP, Gan SQ, Haynes SR, Steven AC (1991) Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. Int J Biol Macromol 13:130–139.
6. Lubkowski J, Hennecke F, Pluckthun A, Wlodawer A (1999) Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure 7:711–722.
7. Bulow L (1990) Preparation of artificial bifunctional enzymes by gene fusion. Biochem Soc Symp 57:123–133.
8. Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M (1988) Single-chain antigen-binding proteins. Science 242:423–426.
9. Tang L, Li J, Katz DS, Feng JA (2000) Determining the DNA bending angle induced by non-specific high mobility group-1 (HMG-1) proteins: a novel method. Biochemistry 39:3052–3060.
10. Biswas S, Deschenes I, Disilvestre D, Tian Y, Halperin VL, Tomaselli GF (2008) Calmodulin regulation of Nav1.4 current: role of binding to the carboxyl terminus. J Gen Physiol 131:197–209.
11. Tolia NH, Joshua-Tor L (2006) Strategies for protein coexpression in *Escherichia coli*. Nat Methods 3:55–64.
12. Patil A, Nakai K, Kinoshita K (2011) Assessing the utility of gene co-expression stability in combination with correlation in the analysis of protein–protein interaction networks. BMC Genomics 12:S19.
13. Kozono H, J White, J Clements, P Marrack, JW Kappler (1994) Functional soluble major histocompatibility proteins with covalently bound peptides. Nature 369: 151–154.

14. Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci USA 101:8301–8306.
15. Cheng YS, Yin FH, Foundling S, Blomstrom D, Kettner CA (1990) Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single-chain tethered dimer. Proc Natl Acad Sci USA 87:9660–9664.
16. Mishra R, Gorlov IP, Chao LY, Singh S, Saunders GF (2002) PAX6, paired domain influences sequence recognition by the homeodomain. J Biol Chem 277: 49488–49494.
17. Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. Genes Dev 13:1263–1275.
18. Wilson KA, Bar S, Maerz AL, Alizon M, Poumbourios P (2005) The conserved glycine-rich segment linking the N-terminal fusion peptide to the coiled coil of human T-cell leukemia virus type 1 transmembrane glycoprotein gp21 is a determinant of membrane fusion function. J Virol 79:4533–4539.
19. Cho S, Hoang A, Chakrabarti S, Huynh N, Huang DB, Ghosh G (2011) The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs. Nucleic Acids Res 39:9413–9421.
20. Bruno R, Bradbury A (1997) A natural longer glycine-rich region in IKe filamentous phage confers no selective advantage. Gene 184:121–123.
21. Pesiridis GS, Lee VM, Trojanowski JQ (2009) Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. Hum Mol Genet 18: R156–162.
22. Senutovitch N, Stanfield RL, Bhattacharyya S, Rule GS, Wilson IA, Armitage BA, Waggoner AS, Berget PB (2012) A variable light domain fluorogen activating protein homodimerizes to activate dimethylindole red. Biochemistry 51:2471–2485.
23. Deane JE, Ryan DP, Sunde M, Maher MJ, Guss JM, Visvader JE, Matthews JM (2004) Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex. EMBO J 23:3589–3598.
24. Nagi AD, Regan L (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des 2:67–75.
25. Hynes TR, Kautz RA, Goodman MA, Gill JF, Fox RO (1989) Transfer of a beta-turn structure to a new protein context. Nature 339:73–76.
26. Robinson CR, Sauer RT (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc Natl Acad Sci USA 95: 5929–5934.
27. Sabourin M, Tuzon CT, Fisher TS, Zakian VA (2007) A flexible protein linker improves the function of epitope-tagged proteins in *Saccharomyces cerevisiae*. Yeast 24: 39–45.
28. Honda S, Selker EU (2009) Tools for fungal proteomics: multifunctional neurospora vectors for gene replacement, protein expression and protein purification. Genetics 182:11–23.
29. Deane JE, Mackay JP, Kwan AH, Sum EY, Visvader JE, Matthews JM (2003) Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. EMBO J 22:2224–2233.
30. Gadd MS, Bhati M, Jeffries CM, Langley DB, Trewhella J, Guss JM, Matthews JM (2011) Structural basis for partial redundancy in a class of transcription factors, the LIM homeodomain proteins, in neural cell type specification. J Biol Chem 286:42971–42980.
31. Fremont DH, Hendrickson WA, Marrack P, Kappler J (1996) Structures of an MHC class II molecule with covalently bound single peptides. Science 272:1001–1004.
32. Scott CA, Peterson PA, Teyton L, Wilson IA (1998) Crystal structures of two I-Ad-peptide complexes reveal that high affinity can be achieved without large anchor residues. Immunity 8:319–329.
33. Fremont DH, Monnaie D, Nelson CA, Hendrickson WA, Unanue ER (1998) Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. Immunity 8: 305–317.
34. Latek RR, Suri A, Petzold SJ, Nelson CA, Kanagawa O, Unanue ER, Fremont DH (2000) Structural basis of peptide binding and presentation by the type I diabetes-associated MHC class II molecule of NOD mice. Immunity 12:699–710.
35. Wilson N, Fremont D, Marrack P, Kappler J (2001) Mutations changing the kinetics of class II MHC peptide exchange. Immunity 14:513–522.
36. He XL, Radu C, Sidney J, Sette A, Ward ES, Garcia KC (2002) Structural snapshot of aberrant antigen presentation linked to autoimmunity: the immunodominant epitope of MBP complexed with I-Au. Immunity 17:83–94.
37. Fremont DH, Dai S, Chiang H, Crawford F, Marrack P, Kappler J (2002) Structural basis of cytochrome c presentation by IE(k). J Exp Med 195:1043–1052.
38. Liu X, Dai S, Crawford F, Fruge R, Marrack P, Kappler J (2002) Alternate interactions define the binding of peptides to the MHC molecule IA(b). Proc Natl Acad Sci USA 99:8820–8825.
39. Zhu Y, Rudensky AY, Corper AL, Teyton L, Wilson IA (2003) Crystal structure of MHC class II I-Ab in complex with a human CLIP peptide: prediction of an I-Ab peptide-binding motif. J Mol Biol 326:1157–1174.
40. Siebold C HB, Wyer JR, Harlos K, Esnouf RE, Svejgaard A, Bell JI, Strominger JL, Jones EY, Fugger L (2004) Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proc Natl Acad Sci USA 101: 1999–2004.
41. McBeth C, Seamons A, Pizarro JC, Fleishman SJ, Baker D, Kortemme T, Goverman JM, Strong RK (2008) A new twist in TCR diversity revealed by a forbidden alphabeta TCR. J Mol Biol 375:1306–1319.
42. Dai S, Crawford F, Marrack P, Kappler JW (2008) The structure of HLA-DR52c: comparison to other HLA-DRB3 alleles. Proc Natl Acad Sci USA 105:11893–11897.
43. Dai S, Murphy GA, Crawford F, Mack DG, Falta MT, Marrack P, Kappler JW, Fontenot AP (2010) Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proc Natl Acad Sci USA 107: 7425–7430.
44. Tollefsen S, Hotta K, Chen X, Simonsen B, Swaminathan K, Mathews, II, Sollid LM, Kim CY (2012) Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. J Biol Chem 287:13611–13619.
45. Matthews JM, Visvader JE, Mackay JP (2001) 1H, 15N and 13C assignments of FLIN2, an intramolecular LMO2:ldb1 complex. J Biomol NMR 21:385–386.
46. Jeffries CM, Graham SC, Stokes PH, Collyer CA, Guss JM, Matthews JM (2006) Stabilization of a binary protein complex by intein-mediated cyclization. Protein Sci 15:2612–2618.
47. Bhati M, Lee C, Nancarrow AL, Lee M, Craig VJ, Bach I, Guss JM, Mackay JP, Matthews JM (2008)

Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. EMBO J 27:2018–2029.
48. Wilkinson-White LE, Dastmalchi S, Kwan AH, Ryan DP, Mackay JP, Matthews JM (2010) 1H, 15N and 13C assignments of an intramolecular Lmo2-LIM2/Ldb1-LID complex. Biomol NMR Assign 4:203–206.
49. Wilkinson-White L, Gamsjaeger R, Dastmalchi S, Wienert B, Stokes PH, Crossley M, Mackay JP, Matthews JM (2011) Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. Proc Natl Acad Sci USA 108:14443–14448.
50. Liew CW, Kwan AH, Stokes PH, Mackay JP, Matthews JM (2012) (1)H, (1)(5)N and (1)(3)C assignments of an intramolecular LMO4-LIM1/CtIP complex. Biomol NMR Assign 6:31–34.
51. Wang W, Prosise WW, Chen J, Taremi SS, Le HV, Madison V, Cui X, Thomas A, Cheng KC, Lesburg CA (2008) Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability. Protein Eng Des Sel 21:425–433.
52. Reinherz EL, Tan K, Tang L, Kern P, Liu J, Xiong Y, Hussey RE, Smolyar A, Hare B, Zhang R, Joachimiak A, Chang HC, Wagner G, Wang J (1999) The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science 286:1913–1921.
53. Dai S, Huseby ES, Rubtsova K, Scott-Browne J, Crawford F, Macdonald WA, Marrack P, Kappler JW (2008) Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. Immunity 28:324–334.
54. Yin Y, Li Y, Kerzic MC, Martin R, Mariuzza RA (2011) Structure of a TCR with high affinity for self-antigen reveals basis for escape from negative selection. EMBO J 30:1137–1148.
55. Hennecke J, Carfi A, Wiley DC (2000) Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. EMBO J 19:5611–5624.
56. Hennecke J, Wiley DC (2002) Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. J Exp Med 195:571–581.
57. Li Y, Huang Y, Lue J, Quandt JA, Martin R, Mariuzza RA (2005) Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. EMBO J 24:2968–2979.
58. Sethi DK, Schubert DA, Anders AK, Heroux A, Bonsor DA, Thomas CP, Sundberg EJ, Pyrdol J, Wucherpfennig KW (2011) A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. J Exp Med 208:91–102.
59. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C (2011) Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor. Nat Struct Mol Biol 18:88–90.
60. Antczak AJ, Tsubota T, Kaufman PD, Berger JM (2006) Structure of the yeast histone H3-ASF1 interaction: implications for chaperone mechanism, species-specific interactions, and epigenetics. BMC Struct Biol 6:26.
61. Ye Q, Li X, Wong A, Wei Q, Jia Z (2006) Structure of calmodulin bound to a calcineurin peptide: a new way of making an old binding mode. Biochemistry 45:738–745.
62. Baldwin ET, Bhat TN, Gulnik S, Liu B, Topol IA, Kiso Y, Mimoto T, Mitsuya H, Erickson JW (1995) Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. Structure 3:581–590.
63. Pillai B KK, Hosur MV (2001) 1.9 A X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. Proteins 43:57–64.
64. Kumar M, Kannan KK, Hosur MV, Bhavesh NS, Chatterjee A, Mittal R, Hosur RV (2002) Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. Biochem Biophys Res Commun 294: 395–401.
65. Das A PV, Mahale S, Serre L, Ferrer JL, Hosur MV (2006) Crystal structure of HIV-1 protease in situ product complex and observation of a low-barrier hydrogen bond between catalytic aspartates. Proc Natl Acad Sci USA 103:18464–18469.
66. Bihani S, Das A, Prashar V, Ferrer JL, Hosur MV (2009) X-ray structure of HIV-1 protease in situ product complex. Proteins 74:594–602.
67. Prashar V, Bihani S, Das A, Ferrer JL, Hosur M (2009) Catalytic water co-existing with a product peptide in the active site of HIV-1 protease revealed by X-ray structure analysis. PloS One 4:e7860.
68. Prashar V BS, Das A, Rao DR, Hosur MV (2010) Insights into the mechanism of drug resistance: X-ray structure analysis of G48V/C95F tethered HIV-1 protease dimer/saquinavir complex. Biochem Biophys Res Commun 396:1018–1023.
69. Das A, Mahale S, Prashar V, Bihani S, Ferrer JL, Hosur MV (2010) X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. J Am Chem Soc 132:6366–6373.
70. Alvizo O, Mittal S, Mayo SL, Schiffer CA (2012) Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease. Protein Sci 21:1029–1041.
71. Foss TR, Kelker MS, Wiseman RL, Wilson IA, Kelly JW (2005) Kinetic stabilization of the native state by protein engineering: implications for inhibition of transthyretin amyloidogenesis. J Mol Biol 347:841–854.
72. Mayer ML, Olson R, Gouaux E (2001) Mechanisms for ligand binding to GluR0 ion channels: crystal structures of the glutamate and serine complexes and a closed apo state. J Mol Biol 311:815–836.
73. Armstrong N, Gouaux E (2000) Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. Neuron 28:165–181.
74. Mayer ML (2005) Crystal structures of the GluR5 and GluR6 ligand binding cores: molecular mechanisms underlying kainate receptor selectivity. Neuron 45: 539–552.
75. Alam SM, Travers PJ, Wung JL, Nasholds W, Redpath S, Jameson SC, Gascoigne NR (1996) T-cell-receptor affinity and thymocyte positive selection. Nature 381:616–620.
76. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368:215–221.
77. DeLano WL (2005) PyMOL: A communications tool for computational models. Abstr Paper Am Chem Soc 230: U1371–U1372.
78. Scott CA, Garcia KC, Stura EA, Peterson PA, Wilson IA, Teyton L (1998) Engineering protein for X-ray crystallography: the murine Major Histocompatibility Complex class II molecule I-Ad. Protein Sci 7:413–418.
79. Deane JE, Sum E, Mackay JP, Lindeman GJ, Visvader JE, Matthews JM (2001) Design, production and characterization of FLIN2 and FLIN4: the engineering of

intramolecular ldb1:LMO complexes. Protein Eng 14: 493–499.
80. Deane JE, Maher MJ, Langley DB, Graham SC, Visvader JE, Guss JM, Matthews JM (2003) Crystallization of FLINC4, an intramolecular LMO4-ldb1 complex. Acta Crystallogr D Biol Crystallogr 59:1484–1486.
81. Bhati M, Lee M, Nancarrow AL, Bach I, Guss JM, Matthews JM (2008) Crystallization of an Lhx3-Isl1 complex. Acta Crystallogr F 64:297–299.
82. Watkins RE, Davis-Searles PR, Lambert MH, Redinbo MR (2003) Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. J Mol Biol 331:815–828.
83. Matsui K, Boniface JJ, Steffner P, Reay PA, Davis MM (1994) Kinetics of T-cell receptor binding to peptide/I-Ek complexes: correlation of the dissociation rate with T-cell responsiveness. Proc Natl Acad Sci USA 91:12862–12866.
84. Kingston RL, Baase WA, Gay LS (2004) Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. J Virol 78:8630–8640.
85. Jacobson A (1972) Role of F pili in the penetration of bacteriophage fl. J Virol 10:835–843.
86. Holliger P, Riechmann L (1997) A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5:265–275.
87. Mousson F, Lautrette A, Thuret JY, Agez M, Courbeyrette R, Amigues B, Becker E, Neumann JM, Guerois R, Mann C, Ochsenbein F (2005) Structural basis for the interaction of Asf1 with histone H3 and its functional implications. Proc Natl Acad Sci USA 102:5975–5980.
88. Munakata T, Adachi N, Yokoyama N, Kuzuhara T, Horikoshi M (2000) A human homologue of yeast antisilencing factor has histone chaperone activity. Genes Cells 5:221–233.
89. Bandyopadhyay J, Lee J, Lee JI, Yu JR, Jee C, Cho JH, Jung S, Lee MH, Zannoni S, Singson A, Kim DH, Koo HS, Ahnn J (2002) Calcineurin, a calcium/calmodulin-dependent protein phosphatase, is involved in movement, fertility, egg laying, and growth in Caenorhabditis elegans. Mol Biol Cell 13:3281–3293.
90. Romero, Obradovic, Dunker K (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform Wkshp Genome Inform 8:110–124.
91. Ye Q, Wang H, Zheng J, Wei Q, Jia Z (2008) The complex structure of calmodulin bound to a calcineurin peptide. Proteins 73:19–27.
92. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245:616–621.
93. Miroy GJ, Lai Z, Lashuel HA, Peterson SA, Strang C, Kelly JW (1996) Inhibiting transthyretin amyloid fibril formation via protein stabilization. Proc Natl Acad Sci USA 93:15051–15056.
94. Kuusinen A, Arvola M, Keinanen K (1995) Molecular dissection of the agonist binding site of an AMPA receptor. EMBO J 14:6327–6332.
95. Renzone G, Salzano AM, Arena S, D'Ambrosio C, Scaloni A (2007) Mass spectrometry-based approaches for structural studies on protein complexes at low-resolution. Curr Proteome 4:1–16.
96. Pace CN, Scholtz JM (1998) A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75:422–427.
97. Eldridge B, Cooley RN, Odegrip R, McGregor DP, Fitzgerald KJ, Ullman CG (2009) An in vitro selection strategy for conferring protease resistance to ligand binding peptides. Protein Eng Des Sel 22:691–698.
98. Matsushima N, Yoshida H, Kumaki Y, Kamiya M, Tanaka T, Izumi Y, Kretsinger RH (2008) Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. Curr Protein Pept Sci 9:591–610.