Accommodating structurally diverse peptides in proteins

Anthony J Wilkinson

Many peptide-binding proteins must bind numerous ligands that differ in size, sequence and sometimes orientation. A variety of strategies for coping with structurally diverse peptide ligands have been revealed by biochemical and structural studies of proteins with roles in immunity, transport and signal transduction.

Address: Department of Chemistry, University of York, Heslington, York YO1 5DD, UK.

Chemistry & Biology July 1996, **3**:519–524

 ISSN 1074-5521

Peptides are widespread in nature and serve a variety of functions acting, for example, as a source of nutrients for microorganisms, as cell-signalling molecules, as recognition elements for the immune system and, of course, as segments of proteins. The interactions of peptides with their receptor proteins are central to the molecular biology of the cell and research in this area is consequently flourishing (reviewed in [1–3]). Generally the binding is sequence specific, but some proteins must tolerate peptide ligands that are structurally and chemically diverse. Different proteins have evolved different solutions to the problem of how to bind peptides efficiently without absolute sequence specificity. The varied nature of the protein–peptide interactions that give rise to this sequence tolerance is illustrated here by four examples: the major histocompatibility complex (MHC) molecules that display peptides as part of immune surveillance, the bacterial oligopeptide transporter protein OppA, and two peptide-binding molecules involved in signal transduction (calmodulin and SH3 domains).

MHC: sidechain pockets and the importance of ends

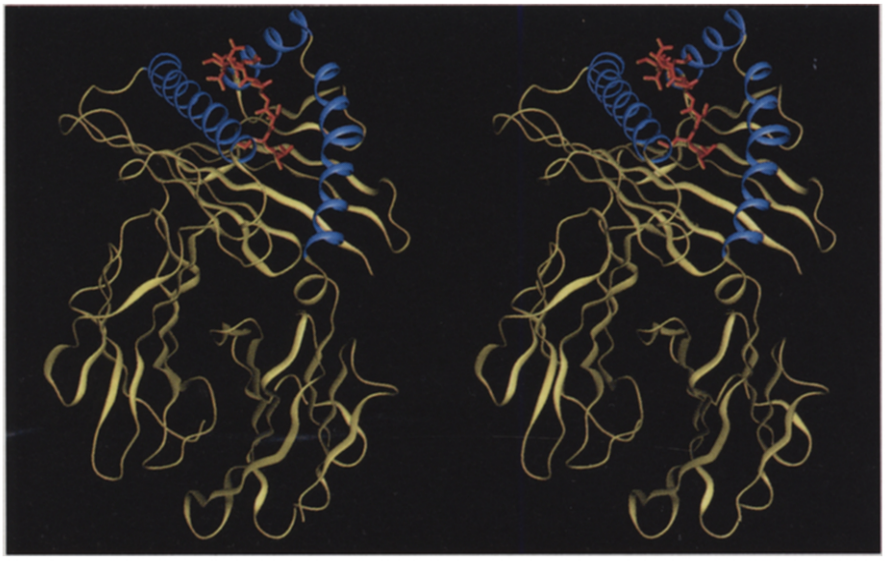
As part of a complex mechanism for combating intracellular pathogens such as viruses and mycobacteria, proteasomes in eukaryotic cells perpetually digest a proportion of the proteins in the cytosol, producing peptide fragments that are subsequently transported into the endoplasmic reticulum. Here, the peptides form complexes with class I MHC proteins. These complexes are then transported to the cell surface, where the membrane-associated MHC molecules display the bound peptides to the outside world. The peptides presented in this way are a representative sampling of the cellular environment; most will have originated from normal cellular 'self' proteins, but occasionally a cell will display 'foreign' peptides, providing a signal that it contains abnormal proteins, for instance those of an infecting virus. The class I MHC complexes are scrutinized by T-cell receptors (TCRs). These cell-surface receptors, like the immunoglobulins, have hypervariable antigen-recognition sequences. The recognition of a foreign peptide in complex with a class I MHC protein by TCRs on cytotoxic T cells leads to the activation of these cells and destruction of the infected cells. The activation of helper T cells, which coordinate other aspects of the immune response, is similar in principle, but in this case the TCRs recognize peptides derived from proteins taken up by endocytosis and displayed on class II MHC proteins.

Immune responses to the enormous array of potential pathogens rely on the capacity of the MHC proteins to bind peptides very tightly ($K_d < 10^{-9}$ M) but with low specificity. Although MHC proteins are quite variable in sequence (polymorphic) within a population, an individual expresses only a handful of MHC proteins of each class; to ensure that peptides can be presented from a vast array of potential pathogens, each MHC protein must be able to bind a huge number of peptides. Crystal structures of the MHC proteins show very clearly how this is achieved [4].

Class I and class II MHC proteins are closely related in structure. Each class of protein contains a peptide-binding groove whose floor is formed by an eight-stranded β-sheet (on a platform formed by a pair of immunoglobulin-like domains) and whose sides are formed by a concave arrangement of α-helices (Fig. 1). Polymorphic residues cluster in the peptide binding site, so that different MHC proteins bind to different subsets of available peptides. For class I MHC proteins, the preferred peptides are 8 to 11 amino acid residues long. In the complexes, the peptide is generally found in an extended conformation with the main-chain atoms of the amino- and carboxy-terminal residues interacting directly and extensively with the protein. In particular, the peptide's α-NH_3^+ and α-COO^- groups are buried and form interactions which contribute much of the overall binding energy.

Although MHC proteins can bind to a wide range of peptides, they are not completely non-selective. At some peptide positions there is a wide tolerance of different sidechains, but at a few positions the sidechains are buried in pockets which have strong preferences for residues of a given type. This restricts the range of peptides that can be bound to those possessing the necessary 'motif'. For example, one human class I MHC allotype, HLA-A2, prefers peptides with large aliphatic sidechains at residue-2 (P2) and small hydrophobic sidechains at the carboxy-terminal position, while in another, HLA-Aw68,

Figure 1



Peptides bind to MHC proteins in a peptide-binding groove that is partially exposed to solvent. Stereoview of a peptide complex of the class I MHC protein HLA-A2. The protein is represented as a yellow ribbon with the α-helical segments that frame the ligand coloured in blue. The peptide ligand is a nonamer (Thr-Leu-Thr-Ser-Cys-Asn-Thr-Ser-Val) derived from the envelope glycoprotein gp120 (residues 195–207) of human immunodeficiency virus (HIV), and is represented in red. The coordinates are from the Brookhaven Data Bank, file 1hhg.pdb [5].

the side chains at the corresponding positions are bound in cavities which are specific for valine/threonine and lysine/arginine respectively [5,6]. In the mouse class I MHC molecule H-2K^b the arrangement of specificity pockets is different and the sidechains of residues at positions 2 and 5 pack together in a shared binding pocket giving rise to preferred pairings of sidechains at these positions [7]. The interactions made by the peptide termini with class I MHC molecules are usually maintained in complexes with peptides of differing lengths. As a result, the central residues of peptides longer than nonamers bulge out from the protein [6]. These central residues exhibit the most variety in structure and conformation, and are among the most accessible for interactions with the TCR.

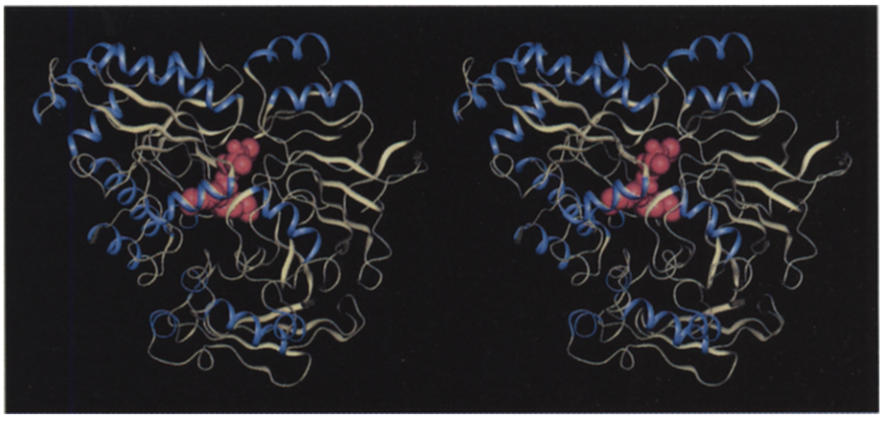
Class II MHC proteins have a very similar structure to the class I MHC proteins, with the notable difference that the peptide binding groove is open at both ends. Class II MHC proteins can thus bind peptides of arbitrary length, but, unlike class I MHC proteins, do not take advantage of the binding energy available from burying the α-NH_3^+ and α-COO^- groups at each end of the peptide. Peptides bound to HLA-DR1 (a human class II MHC allele) assume a polyproline type II helical conformation over much of the length of the 13 residues that reside in the binding site [8]. Three asparagine sidechains from the protein, each of which forms bidentate hydrogen bonds with main-chain C=O and N–H groups in adjacent peptide bonds of the ligand, appear to be important in determining the peptide conformation. In this arrangement, the ligand sidechains that are solvent exposed and the ligand sidechains that are buried in the protein are each regularly spaced along the length of the peptide.

OppA – a Venus flytrap

The oligopeptide transporter protein OppA must also bind peptides of diverse sequence, the more diverse the better. OppA is the initial receptor protein for, and defines the specificity of, the bacterial oligopeptide permease, which transports extracellular peptides to the cytoplasm. It is an abundant, soluble protein that captures peptides in the periplasm and delivers them to a complex of two membrane-spanning proteins (OppB and OppC), which form a channel through which the solute passes. Two cytoplasmic membrane-associated proteins (OppD and OppF) are responsible for coupling ATP hydrolysis to peptide transport. Although genetic studies indicate that the oligopeptide permease can transport peptides of two to five amino acid residues regardless of sequence, binding studies suggest that OppA has highest affinity for tripeptides and tetrapeptides (with a K_d of ~10^{-6} M). Dipeptides are transported predominantly by the closely related dipeptide permease, which has its own set of membrane components, and periplasmic binding protein (DppA).

The crystal structures of OppA and DppA reveal that their ligands are completely sequestered from bulk solvent and enclosed in the protein interior; the mechanism of binding has been likened to that of a Venus flytrap (Fig. 2; [9–12]). Such complete enclosure of ligands is a feature of the periplasmic substrate binding proteins, and is usually associated with complementary interactions and therefore with specificity. Indeed, the binding proteins associated with sugar and anion transporters are often exquisitely selective [13]. Complete removal of the ligand from bulk solvent in this way is presumably a prerequisite for specific transmembrane transport.

Figure 2



OppA completely surrounds its ligand, but makes few contacts with the peptide sidechains. Stereoview of the crystal structure of OppA, drawn as a yellow ribbon with α-helices in blue, in complex with the tetrapeptide Lys-Lys-Lys-Ala which is represented as red van der Waals spheres [10]. The coordinates are from the Brookhaven Data Bank, file 1olc.pdb.

The peptide in OppA is bound in an extended conformation so that the charged ligand termini and its polar main-chain hydrogen-bonding groups, common to all peptides, are available to make interactions with the protein. A salt bridge is formed between the acidic sidechain of Asp419 on the protein and the α-amino group of the peptide ligand, and there are extensive β-sheet-like interactions between main-chain hydrogen-bonding groups on the peptide and corresponding groups on the protein. Peptides of different length are accommodated not by bulging of the peptide as is seen for class I MHC, nor by extension out of the binding site as seen for class II MHC. Instead, a series of positively charged sidechains is located along the binding cavity in such a way that the α-carboxylate groups of peptide ligands of different lengths form ion pairs with different protein sidechains.

The sidechains of the peptide ligands, which can differ in size, shape, polarity and charge, project into spacious hydrated cavities. Few direct interactions are made between OppA and the ligand sidechains, presumably because such interactions would result in discrimination. Buried water molecules are important in adapting the pocket to accommodate the different ligands; water molecules occupy the volume not taken up by the peptide sidechains, fulfil the hydrogen-bonding potential of the ligand sidechains and the surrounding protein residues, and solvate any charged groups.

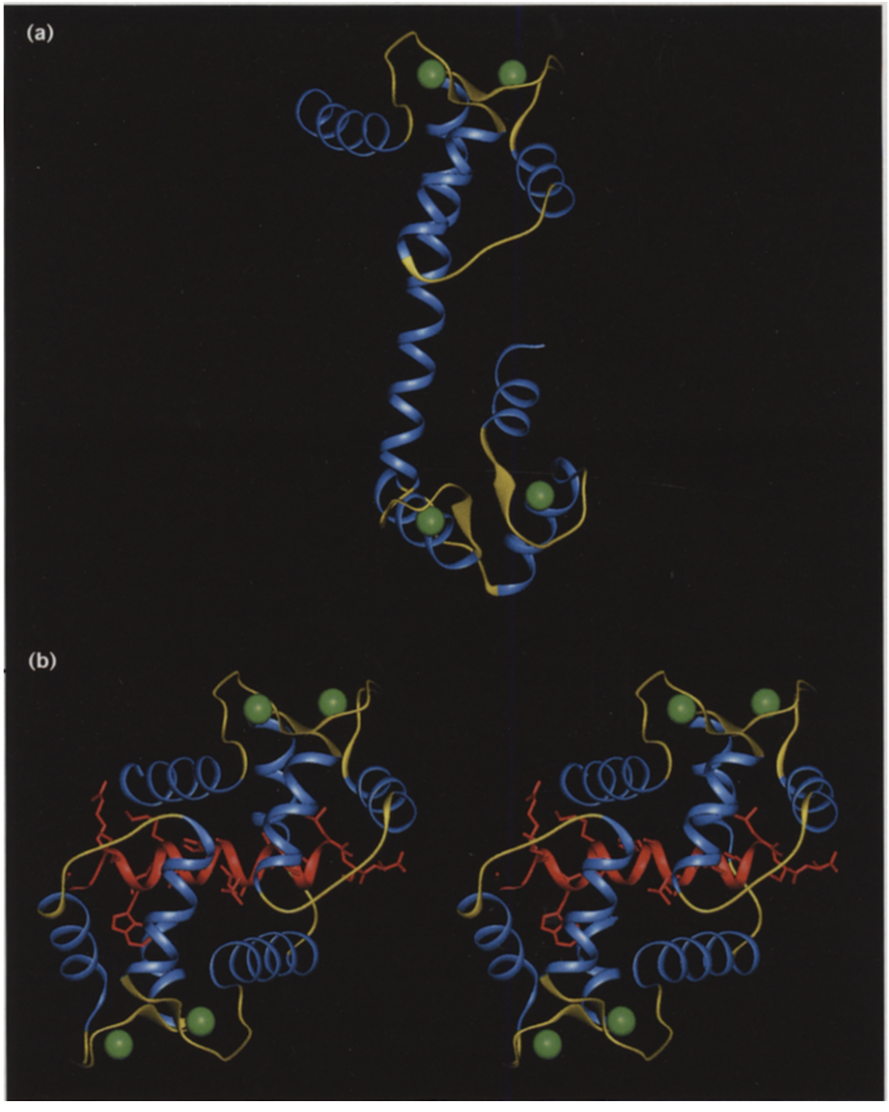
The different modes of peptide binding by MHC proteins and OppA are dictated by the receptors with which the complexes interact. MHC–peptide complexes are surveyed by a population of T cells, each with a different TCR, and the role of MHC is to advertise the peptides so that differences of sequence are readily apparent. This allows specific interactions to be made with those TCRs bearing complementary surfaces leading to directed immune responses. In marked contrast, OppA has a single receptor made up of the membrane components of its transport system. Efficient transport requires that this receptor be able to distinguish the unliganded from the liganded form of OppA, but interact with the liganded form of OppA regardless of the sequence of the bound peptide. A Venus fly-trap mechanism of ligand binding achieves this by masking differences among peptide sequences and by creating a protein surface that is unique to the liganded form.

Calmodulin: peptide wrapping

Permissive peptide binding is also observed in calmodulin (CaM), a calcium-binding protein that has a pivotal regulatory role in cellular metabolism. Calcium is a major intracellular signalling molecule in eukaryotic cells, its concentration transiently rising steeply and falling in response to extracellular stimuli. CaM is a subunit of a number of calcium-regulated enzymes, including protein kinases, protein phosphatases, cyclic nucleotide phosphodiesterases and nitric oxide synthetase, which themselves serve regulatory functions. The crystal structure of calcium–CaM reveals a bilobate molecule. Each lobe contains two calcium ions, each of which is bound in a helix-loop-helix motif that is referred to as an EF-hand (Fig. 3a). The two lobes are joined by a connecting segment of α-helix giving rise to an overall dumbbell-shaped molecule.

The interactions of CaM with its target enzymes are not understood in detail, although the CaM-binding regions within a number of these molecules have been narrowed down to ~20-residue segments of polypeptide. Synthetic peptides corresponding to these segments bind to calmodulin with affinities approaching those of the intact target proteins (K_d ~1 nM) implying that the major CaM-binding determinants reside in these short contiguous sequences. The paucity of sequence similarity among these peptides is striking, and it seems that CaM has low sequence specificity. Indeed, other studies have shown that CaM can bind a wide range of synthetic as well as

Figure 3



Calmodulin bends to wrap around a peptide ligand. The crystal structures of **(a)** calmodulin and **(b)** calmodulin in complex with a peptide (stereoview) are shown to illustrate the dramatic change in conformation that accompanies peptide binding [15]. The protein backbone is shown as a ribbon (yellow) with the α-helical regions highlighted in blue. The four calcium ions, two in each lobe, are represented as green spheres. In (a), the amino-terminal lobe is at the top. In (b), the 20-residue peptide (red) derived from smooth muscle myosin light chain kinase assumes an α-helical conformation with its amino terminus on the left. The view is such that the orientation of the amino-terminal lobe of the protein (right) is similar to that in (a). The coordinates are from the Brookhaven Data Bank, files 1cll.pdb and 1cdl.pdb.

natural peptides, including peptides composed of D-amino acids. These peptides share no obvious sequence similarity other than a high propensity for α-helix formation and a predominance of basic over acidic residues.

The three-dimensional structures of CaM–peptide complexes confirm that peptide ligands are bound in an α-helical conformation, and reveal a dramatic reorganization of the calmodulin domains. The α-helix connecting the two CaM lobes unravels to form two shorter α-helices connected by a short segment of coil. This unravelling allows relative movement of the two lobes so that they wrap around and substantially bury the peptide ligand in a manner that has been likened to two hands grasping a rope [14] (Fig. 3b). Different peptides can be accommodated through the capacity of the connecting helix to act as a variable 'expansion joint' that can be unravelled to varying extents, so allowing subtly different relative arrangements of the lobes according to ligand sequence [15]. The peptide-binding tunnel that results has a hydrophobic centre containing a number of apolar pockets, while its extremities feature acidic residues [16]. Modelling studies based on CaM–ligand structures led to the proposal that ideal calmodulin-binding peptides should contain a set of bulky aliphatic or aromatic sidechains to fill the apolar cavities, with flanking basic sidechains to give electrostatic interactions with the acidic residues on CaM [17]. However, analysis of a series of peptides derived from the CaM-recognition sequence of skeletal muscle myosin light chain kinase (sMLCK), corresponding to single point mutations of each residue to alanine, suggests that no single amino-acid residue is essential for tight binding [18]. Each of the alanine variants binds to CaM with greater affinity than the sMLCK peptide, emphasizing again the catholic character of the binding site and the difficulties in defining a CaM-binding motif.

The α-helical conformation of the peptide ligand provides no opportunity for CaM to form tight, specific interactions with the peptide backbone, as seen in OppA and the MHC complexes. Instead, CaM achieves high affinity binding by making extensive interactions with the sidechains on its peptide ligands. This is remarkable in view of its low sequence specificity. A further fascinating and unanswered question concerns how calmodulin grasps its physiological ligands, which are not free peptides but segments of a folded polypeptide that is usually part of a large enzyme assembly.

SH3 domains: backwards binding

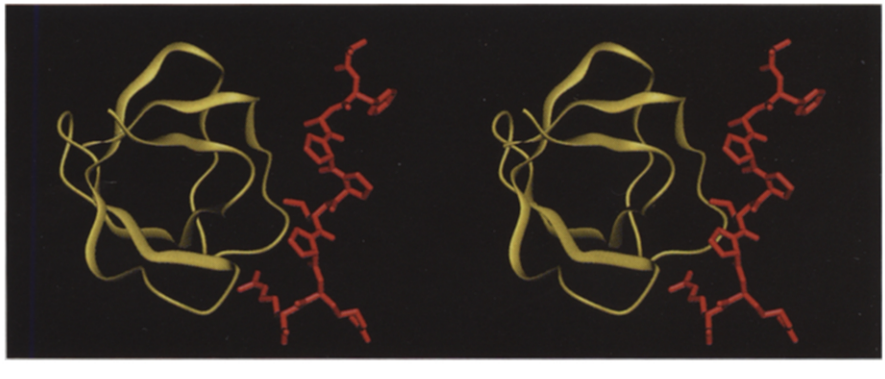
SH3 domains are also versatile in the organization of their complexes with peptides, although they are the most sequence-selective of the examples discussed here. SH, or Src-homology, domains are widespread in proteins involved in the transmission of signals from receptor protein-tyrosine kinases, and were initially discovered in cytoplasmic protein-tyrosine kinases related to c-Src, hence the name. They mediate protein–protein interactions among the components of the complex molecular circuitry that forms the signal transduction pathways of eukaryotic cells [19,20]. SH3 domains are composed of ~60 amino acid residues folded to form a pair of β-sheets that pack together in a barrel-like arrangement (Fig. 4). The amino- and carboxyl-termini are close together in space, consistent with the idea that these domains are independent folding units or modules that can be inserted at different locations within a protein. SH3 domains bind to short proline-rich sequences in their target proteins and in related synthetic peptides.

In the complexes, the peptides form a left-handed polyproline type II helix. This conformation results from the regular arrangement of prolines in the peptide, not from restrictions imposed by the binding surface (unlike the case of HLA-DR1–peptide complexes). The curious feature of SH3 domains is that, depending on the sequences that flank the proline-rich motif, the amino to carboxyl orientation of the ligand with respect to the protein can be reversed [21,22]. This is most vividly illustrated for the SH3 domain of Src, for which NMR structures of complexes with peptides bound in both orientations have been determined. Protein sidechains form hydrogen bonds with three main-chain carbonyl oxygens on the ligand, while three shallow pockets in the SH3 domain accommodate the sidechains of five peptide residues. The first two pockets each bind a proline sidechain and the sidechain of an adjacent, usually hydrophobic, residue. Due to pseudo-symmetry inherent in the polyproline type II helix, these interactions of the SH3 domain with the main-chain and sidechain groups on the ligand are largely preserved when the peptide binds in the opposite orientation. The third pocket binds an arginine sidechain, which forms an ion pair with a conserved acidic residue in the SH3 domain. It is the amino-terminal or carboxy-terminal location of this arginine with respect to the proline-rich motif that determines the orientation in which the peptide binds. Interestingly, in the crystal structure of the SH3 domain from phosphatidylinositol 3-kinase (PI3K), the three ligand-binding pockets are occupied by amino-acid residues from two different symmetry-related protein molecules [23]. The interactions of these sidechains with the SH3 domain are similar to those seen in the complexes with synthetic PI3K peptide ligands, though the main-chain conformation is clearly different. Although the peptide-complex structures provide models for how SH3 domains engage their intracellular ligands, structures of much larger complexes including SH3 domains and their target proteins are needed to illuminate the molecular mechanics of signal transduction properly.

Diverse routes to diversity

The studies of peptide complexes summarized above show that proteins can use a variety of strategies to bind peptides without imposing rigid sequence selectivity. Diversity in sequence recognition is most easily achieved by avoiding contact with some or all of the ligand sidechains; these sidechains can be exposed to the solvent, as seen for MHC complexes, or may be enclosed within a versatile ligand-binding envelope, as in OppA. A

Figure 4



Peptides bind to SH3 domains in a polyproline type II helix conformation. Stereoview of the crystal structure of the SH3 domain of c-Src in complex with a nine-residue proline-rich peptide (Ala-Phe-Ala-Pro-Pro-Leu-Pro-Arg-Arg; [21]). The protein backbone is represented as a yellow ribbon with the peptide residues in red. The peptide adopts a left-handed helical conformation that is characteristic of proline-rich sequences (a polyproline type II helix). The coordinates are from the Brookhaven Data Bank, file 1prm.pdb.

second source of diversity is structural malleability, either in the protein, as seen in calmodulin, or in the peptide, as with the bulging observed in complexes of class I MHC proteins with longer peptides. For the SH3 domains, the symmetry of the ligand's polyproline type II helix and the use of relatively non-specific hydrophobic sidechain binding pockets lead to a rather curious orientational ambivalence in peptide binding by these proteins.

Acknowledgements
I would like to thank Dr Leo Caves for helpful discussions.

References

1. Stanfield, R.L. & Wilson, I.A. (1995). Protein–peptide interactions. *Curr. Opin. Struct. Biol.* **5**, 103–113.
2. Zvelebil, M.J.J.M. & Thornton, J.M. (1993). Peptide–protein interactions: an overview. *Q. Rev. Biophys.* **26**, 333–363.
3. Marshall, G.R. (1992). Three-dimensional structure of peptide–protein complexes: implications for recognition. *Curr. Opin Struct. Biol.* **2**, 904–909.
4. Stern, L.J. & Wiley, D.C. (1994). Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure* **2**, 245–251.
5. Madden, D.R., Garboczi, D.N. & Wiley, D.C. (1993). The antigenic identity of peptide–MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* **75**, 693–708.
6. Guo, H-C., *et al.*, & Wiley, D.C. (1992). Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature* **360**, 364–366.
7. Fremont, D.H., Stura, E.A., Matsumura, M., Peterson, P.A. & Wilson, I.A. (1995). Crystal structure of an H-2K^b-ovalbumin peptide complex reveals the interplay of primary and secondary anchor positions in the major histocompatibility complex binding groove. *Proc. Natl. Acad. Sci. USA* **92**, 2479–2483.
8. Jardetzky, T.S., *et al.*, & Wiley, D.C. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline-II-like conformation for bound peptides. *Proc. Natl. Acad. Sci. USA* **93**, 734–738.
9. Tame, J.R.H., *et al.*, & Wilkinson, A.J. (1994). The structural basis of sequence-independent peptide binding by OppA protein. *Science* **264**, 1578–1581.
10. Tame, J.R.H., Dodson, E.J., Murshudov, G., Higgins, C.F. & Wilkinson, A.J. (1995). The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **3**, 1395–1406.
11. Dunten, P. & Mowbray, S.L. (1995). Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Protein Sci.* **4**, 2327–2334.
12. Nickitenko, A.V., Trakhanov, S. & Quiocho, F.A. (1995). 2 Å structure of DppA, a periplasmic dipeptide/transport chemosensory receptor. *Biochemistry* **34**, 16585–16595.
13. Quiocho, F.A. (1991). Atomic structures and function of periplasmic receptors for active transport and chemotaxis. *Curr. Opin. Struct. Biol.* **1**, 922–933.
14. Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B. & Bax, A. (1992). Solution structure of a calmodulin–target peptide complex by multidimensional NMR. *Science* **256**, 632–638.
15. Meador, W.E., Means, A.R. & Quiocho, F.A. (1993). Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science* **262**, 1718–1721.
16. Meador, W.E., George, S.E., Means, A.R. & Quiocho, F.A. (1995). X-ray analysis reveals conformational adaptation of the linker in functional calmodulin mutants. *Nat. Struct. Biol.* **2**, 943–945.
17. Afshar, M., *et al.*, & Haiech, J. (1994). Investigating the high affinity and low sequence specificity of calmodulin binding to its targets. *J. Mol. Biol.* **244**, 554–571.
18. Montigiani, S., Neri, G., Neri, P. & Neri, D. (1996). Alanine substitutions in calmodulin-binding peptides result in unexpected affinity enhancement. *J. Mol. Biol.* **258**, 6-13.
19. Pawson, T. & Schlessinger, J. (1993). SH2 and SH3 domains. *Curr. Biol.* **3**, 434–442.
20. Kuriyan, J. & Cowburn, D. (1993). Structures of SH2 and SH3 domains. *Curr. Opin. Struct. Biol.* **3**, 828–837.
21. Feng, S., Chen, J.K., Yu, H., Simon, J.A. & Schreiber, S.L. (1994). Two binding orientations to the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science* **266**, 1241–1246.
22. Lim, W.A., Richards, F.M. & Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372**, 375–379.
23. Liang, J., Chen, J.K., Schreiber, S.L. & Clardy, J. (1996). Crystal structure of PI3K SH3 domain at 2.0 Å resolution. *J. Mol. Biol.* **257**, 632–643.