REVIEW ARTICLE

The structure and function of proline-rich regions in proteins

Michael P. WILLIAMSON

The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2UH, U.K.

INTRODUCTION

Proline-rich regions (PRRs) of proteins occur widely in both prokaryotes and eukaryotes. They are frequently found as multiple tandem repeats, often of considerable length. Despite this wide distribution, the functions of PRRs are often unclear. In this review, known structures and functions of PRRs are discussed, in an attempt to identify unifying properties.

The conformation of proline

Proline is a very unusual amino acid, in that the side-chain is cyclized back on to the backbone amide position. This has three important consequences. First, the backbone conformation of proline itself is very restricted. The available backbone ϕ dihedral angles are limited to a small range around $\phi = -65°$ [1,2] (Figure 1). Similar restrictions do not apply to ψ, which is able to populate either the α-helical region ($\psi \approx -40°$) or the β-sheet region (ψ approx. $+150°$). Surveys of prolines in crystal structures show that roughly 44% of prolines are in the α region and 56% are in the β region [4,5]. Second, the bulkiness of the N–CH_2 group places restrictions on the conformation of the residue preceding proline [6], disfavouring the α-helix conformation [4,5]. Third, because the amide proton is replaced by a CH_2 group, proline is unable to act as a hydrogen bond donor. This fact, plus the bulkiness of the side-chain, produces the well-known 'helix-breaker' (and β-sheet breaker) effect. However, it is of interest to note the relatively high proportion of prolines near the centre of transmembrane helices; it has been suggested that these residues play a role in signal transduction [7,8]. A statistical survey shows that Pro is often found one or two residues after the end of an α-helix [9]. However, there is an even stronger tendency to find Pro at the beginning of a helix. Presumably this is explained both by the positive benefit coming from not needing a hydrogen bond partner for Pro, and by the fact that the Pro ϕ angle is permanently constrained to an angle typically found in a helix.

These facts place restrictions on the conformation that is possible for the Xaa-Pro dipeptide. Xaa has a strong tendency to be in the β conformation, with less than 10% of Xaa being found in the α conformation [4], while the Pro ϕ angle is constrained close to $-65°$. The Xaa-Pro dipeptide therefore tends to be fairly rigid and extended.

The polyproline II helix

Naturally, a Pro-Pro dipeptide is even more restricted, and a considerable body of evidence [10–14] suggests that a sequence of four or more proline residues in a row adopts a single preferred conformation in solution, with $\phi = -78°$ and $\psi = +146°$, known as the polyproline II helix [15] (Figure 1). This is an extended structure with three residues per turn. It is found prominently in collagen, and also, but more rarely, in globular proteins. It has been identified as a major structural element in some pancreatic polypeptide hormones and neuropeptides [16,17]. In these polypeptides, the polyproline II conformation is stabilized by having proline as every third residue, in the motif $(PXX)_n$. Short sequences adopting the polyproline II conformation have been identified on the surface of proteins in a surprisingly large number of cases (96 occurrences in 80 non-homologous proteins) [18]. Although very few of these sequences consist entirely of proline residues, the majority contain at least one proline. The number of occurrences in crystal structures of polyproline II helices containing more than five C^{α} atoms is very low.

Nearly all the proline-rich sequences described here are repetitive and longer than five residues. Based on the remarks made above, it is therefore not surprising to find that they generally form extended structures and flexible regions that are hard to crystallize. For this reason, there are very few crystal structures of PRRs. Most structural information on such regions has come from solution-state n.m.r. and c.d. spectroscopy, and from modelling studies using secondary structure predictions.

The other well-known facet of proline is its unique ability to form *cis* peptide bonds, occurring to an extent of 5.7% in globular proteins [4], with Tyr-Pro as the most likely *cis*-bonded pair. In small peptides, a bulky hydrophobic residue before Pro increases the proportion of *cis* bonds [19], but this tendency is apparently not followed in proteins. A high proportion (41 out of 58) of Xaa-*cis* Pro occurrences in globular proteins are found

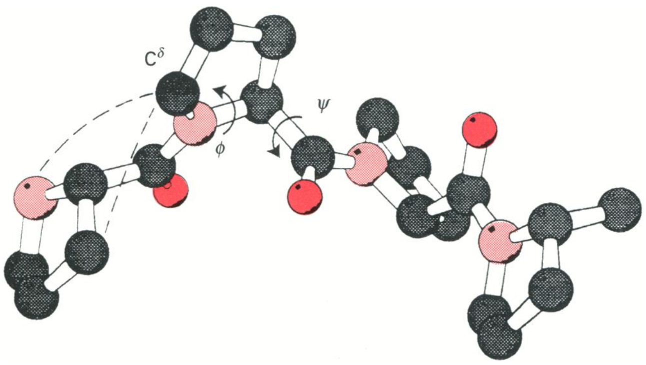

Figure 1 Part of a polyproline II helix

The helix is extended and repeats every three residues. The proline ϕ and ψ angles are indicated. The ϕ angle is constrained by the proline ring, while steric interactions between the proline δ carbon and the preceding residue limit the conformational freedom of the preceding residue: if the preceding residue is in the α-helix conformation, the interactions drawn as dashed lines are energetically unfavourable. Nitrogen atoms are shown in pink, and oxygen atoms in red. Figure prepared using the program MOLSCRIPT [3].

Abbreviations used: PRR, proline-rich region; PRP, proline-rich protein.

Table 1 Proteins with repetitive short proline-rich sequences

Name	Source	Sequence	Protein function	Comment	References
Light chain myosin kinase	Rabbit skeletal muscle	$(AP)_6$	Binds actin	PRR is at N-terminus	23
βB1 crystallin	Ox eye lens	$GP_3GPAPGSG(PA)_5Q(PA)_2$	Cytoskeletal binding?	PRR is at N-terminus	24,25
OmpA	*E. coli*	$(AP)_4$	Major outer membrane protein	Mediates F-dependent conjugation	26
Procyclin	*T. brucei*	$(DP)_2(EP)_{22–29}$	Membrane-bound coat protein	Developmentally regulated	27
TonB	Bacterial	$(EP)_5X_{13}(KP)_5$	Iron siderophore transport	Spans periplasmic space; binds FhuA	28,29
Group C M protein (equine)	*S. equi*	$(DPX)_{17}$ (15 are DPV)	Binds peptidoglycan	Next to membrane anchor; antigenic	30
Group B IgA receptor	*Streptococcus*	$(XPZ)_{30}$ (X = T, S, A, I, L, V; Z is alternately + and −)	Binds peptidoglycan?	195 residues from membrane anchor	31–33
p70 pertactin	*Bordetella parapertussis*	PQP nine times	Outer membrane protein. Cell adhesion?	PQP region not involved in adhesion	34
Amelogenin	Ox	$(QPX)_9$; 49 P in 170 residues	Tooth structure		35

as Type VI turns [20].^{13}C n.m.r. studies on proline-rich sequences in proteins show that *cis* Pro is rare in such sequences [21,22].

Following this brief review of proline conformation, we turn to consider proline-rich regions (PRRs) in proteins, concentrating on the large number of PRRs that contain repetitive proline-rich sequences, or multiple tandem repeats with minor variations between repeated sequences. In many of the examples discussed, the function of the PRR is uncertain. It is shown that the common element in almost all examples is that of binding, in a non-stoichiometric but functionally important way. However, in some cases, the PRR is largely used as a structural element; this function occurs most frequently in polypeptides containing hydroxyproline rather than proline. The division between the different sections below is intended to be by the type of sequence of the proline-rich section, although the differences occasionally become a little blurred.

A SURVEY OF PROLINE-RICH PROTEINS (PRPs)

Repetitive short proline-rich sequences

Many proline-rich sequences have a strikingly repetitive character, i.e. $(XP)_n$ or $(XPY)_n$, as illustrated by the examples in Table 1. Some of the sequences have unknown functions, but many have been demonstrated to be involved in binding processes. The $(AP)_n$ motif in myosin light chain kinase has been shown to bind to actin. N.m.r. studies have shown that $(AP)_n$ presents a rather stiff 'elbow-hinged' peptide chain, which would be consistent with our understanding of the effects of Pro on local conformation. In particular, the N-terminal proline-rich peptide of the myosin light chain kinase is likened to a tail 'wagging the dog' [22], giving the positively charged N-terminus a fairly defined spatial position with respect to the myosin core [36] and allowing the linker to operate as a functional linkage in the thin-filament-based actomyosin regulatory mechanism (Figure 2). The important binding interaction is explained as being strengthened by the smaller entropy loss that occurs upon binding of the more rigid peptide, a point that will be discussed at greater length later.

Similarly, the crystallins and ompA are thought to function by binding to cytoskeletal proteins. The function of procyclin is less certain. It contains a long $(XP)_n$ sequence, which presumably enables the protein to extend out from the cell membrane, and may also stabilize the protein outer coat by binding and non-covalently cross-linking other coat protein components.

TonB is a particularly interesting protein. It has an N-terminal membrane anchor embedding it in the cell membrane, followed by two charged $(XP)_n$ sequences, which apparently act to hold the protein pointing rigidly across the periplasmic space, with the $(XP)_5$ sequence being involved in interactions with the outer membrane siderophore receptor protein, FhuA. The protein, and particularly the $(XP)_n$ sequences, may therefore be a 'molecular trigger', passing extracellular signals to the inner membrane (Figure 3) [37].

Sequences of a repetitive nature such as those discussed in this section can be expected to adopt a polyproline II structure, as described above. This structure has three residues per turn, and it is therefore no surprise to find that many of the repetitive sequences have a periodicity of three. Thus the PRR in the group C equine streptococcal M protein, with the approximate sequence $(DPV)_{17}$, is expected to have a negatively charged face and a hydrophobic face. This PRR, which is adjacent to the membrane anchor, is thought to thread through the peptidoglycan layer, presumably interacting with it and strengthening it as it goes [38]. A similar structural function is particularly suggestive for the striking charge-alternating sequence of the streptococcal group B IgA receptor. Other extracellular PRPs, such as pertactin, appear to act by binding to proteins on the cell surface and influencing cell–cell recognition.

It is therefore likely that, in all proteins with repetitive $(XP)_n$ and $(XPY)_n$ sequences, the PRR functions as a stiff 'sticky arm', binding rapidly and reversibly to other proteins.

Tandemly repeated sequences

This group of proteins contains longer proline-rich sequences, typically 5–8 residues in length, which are repeated in tandem many times, often with slight variations (Table 2). In some cases, such as the salivary PRPs and the cereal storage proteins, the tandem repeats constitute almost the entire protein. The proteins of this group have better characterized functions than most of the proteins discussed in the previous section; nearly all of the functions involve protein–protein binding.

One of the best characterized groups is the salivary PRPs, which form 70 % of the protein in saliva. They appear to have several functions, but the most likely function of the proline-rich tandemly repeated section (which forms by far the largest part of the protein) is to bind plant polyphenols (tannins) present in the diet and to reduce their harmful effects by forming precipitates [51]. They do this by having long open extended structures which present a maximum surface area per residue, and achieve the precipitation of polyphenols by multivalent binding and non-covalent cross-linking [52], in much the same manner as multivalent antibodies bind, cross-link and precipitate antigens (Figure 4). The proline residues act not only to keep the structure

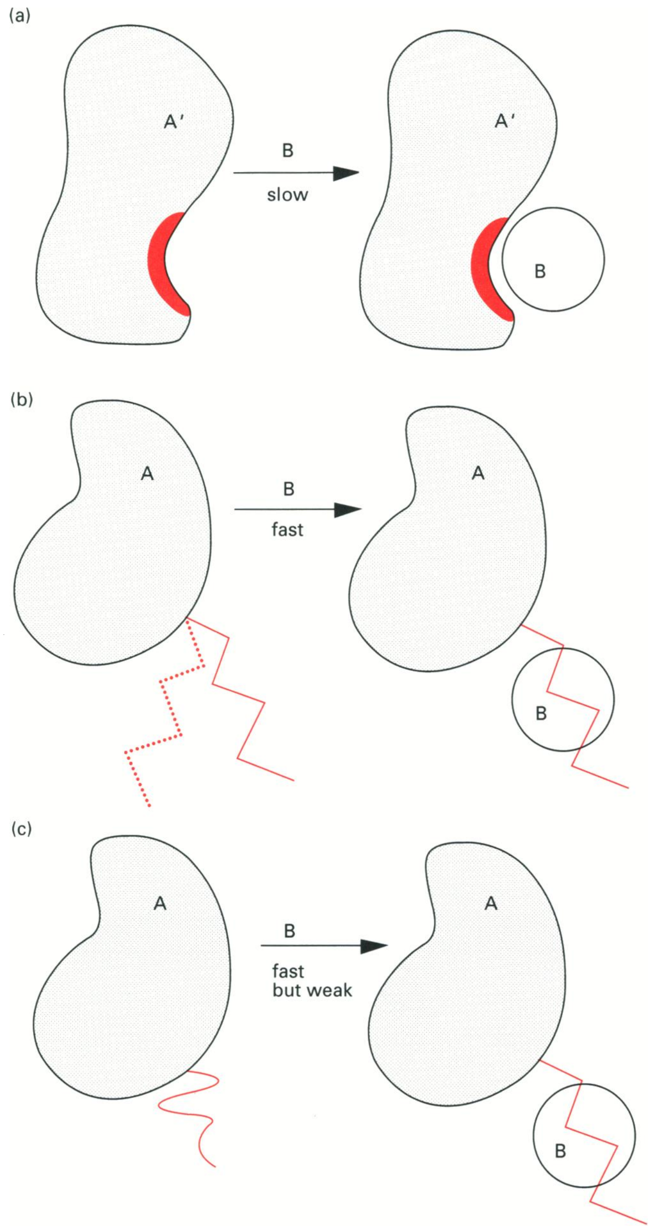

Figure 2 Protein–protein binding interactions

In all cases, the parts of protein A/A' that bind B are shown in red. (**a**) A conventional protein–protein binding interaction is usually very specific, and often slow. (**b**) By contrast, PRRs of proteins can bind rapidly and non-specifically to a range of other proteins, acting as 'sticky arms'. The stiffness of the PRR can give the interaction structural or mechanical significance. (**c**) The binding of a normal (non-proline-rich) hanging peptide arm requires conformational freezing of the arm, and therefore results in weaker binding.

extended, but also as binding sites for polyphenols; n.m.r. binding studies have shown that the proline residues are the primary sites for binding of the common polyphenol pentagalloyl glucose [53]. These tannin-binding PRPs are not confined to saliva; a fungal PRP has been identified that appears to be secreted specifically to bind to tannins, thereby allowing the fungus to grow on plant tissue that has a high polyphenol content [54].

The mammalian epithelial mucins contain proline-rich sequences broadly similar to those in the fungal proline-rich tannin-binding protein described above. They are large proteins, with a long tandemly repeated section. For example, the human tumour-associated polymorphic epithelial mucin has a 20-residue proline-rich sequence repeated between 21 and 125 times [41]. It is heavily glycosylated and is thought to function by creating an extensive network of interlocking extended chains anchored to the membrane, thus coating and lubricating the epithelial layer. However, its sequence similarity with the fungal protein suggests that it may have an additional function as a tannin-binding protein.

The parasitic circumsporozoite protein is of particular medical interest because its repeated proline-rich sequence makes it highly immunogenic. Its function is to form a tough interlocking network, as does the dec-1 eggshell protein (see below). The plant storage proteins play a vaguely analogous role, forming a tough extensible layer around the seed which is largely responsible for the texture of bread dough. The exact function of the plant storage proteins is unknown, but their situation in the periphery of the protein bodies [55] suggests that they may be involved in the support of these cellular organelles. This is presumably achieved by non-covalent interactions between protein chains (mediated in large part by the prolines), since there are few covalent cross-links. Elastin is probably somewhat different: its elasticity is thought to derive from the presence of repeated β-spirals, in which the regularly spaced proline residues play a key part by forming tight turns. The elastin structure is therefore one of the very few cases where the structural role of proline is to form turns, rather than to stabilize extended structure.

Several actin-binding proteins with highly repetitive sequences were described in the previous section. Others have longer tandemly repeated sequences, such as the actin-binding protein from *Dictyostelium discoideum*. Other tandem proline-rich repeats are thought to be involved in structural organization, such as the C-terminal extension of squid rhodopsin.

It is therefore clear that the longer tandemly repeated sequences discussed in this section play a qualitatively different role from that played by the $(XP)_n$ and $(XPY)_n$ sequences discussed in the previous section. Their greater length and flexibility allow them to form interlocking networks of high overall strength, suitable for external coats and irreversible precipitation of toxins. Nevertheless, the unique ability of PRRs to bind rapidly and tightly forms a common unifying motif.

Multi-PRR systems

Two systems have been described that both involve the binding of a PRR containing tandemly repeated proline-rich sequences to several other PRRs. The better characterized is RNA polymerase II, which contains 26 or 27 nearly identical copies of the sequence YSPTSPS, in a presumably extended C-terminal domain. There is good evidence that it interacts with the TFIID–IIA–IIB transcription factor complex [56], probably with the TATA-binding component of TFIID among others [49]. Many transcription factors have proline-rich termini. Careful studies on CTF/NF-1 (Table 3) show that the proline-rich segment, which is a 100-residue section at the C-terminus, is not required for DNA binding but is essential for transcriptional activation. It is presumed to bind to other factors involved in the initiation of transcription, such as RNA polymerase and possibly TFIID. Mermod et al. [57] list other PRRs known to be involved in transcriptional control. Many homeobox proteins, particularly the AntP-type homeodomains, have PRRs. For example, a chick homeobox protein has 44 prolines within residues 16–137, including a P_{10} stretch [78]. The combination of an extended proline-rich terminus on RNA polymerase II and proline-rich

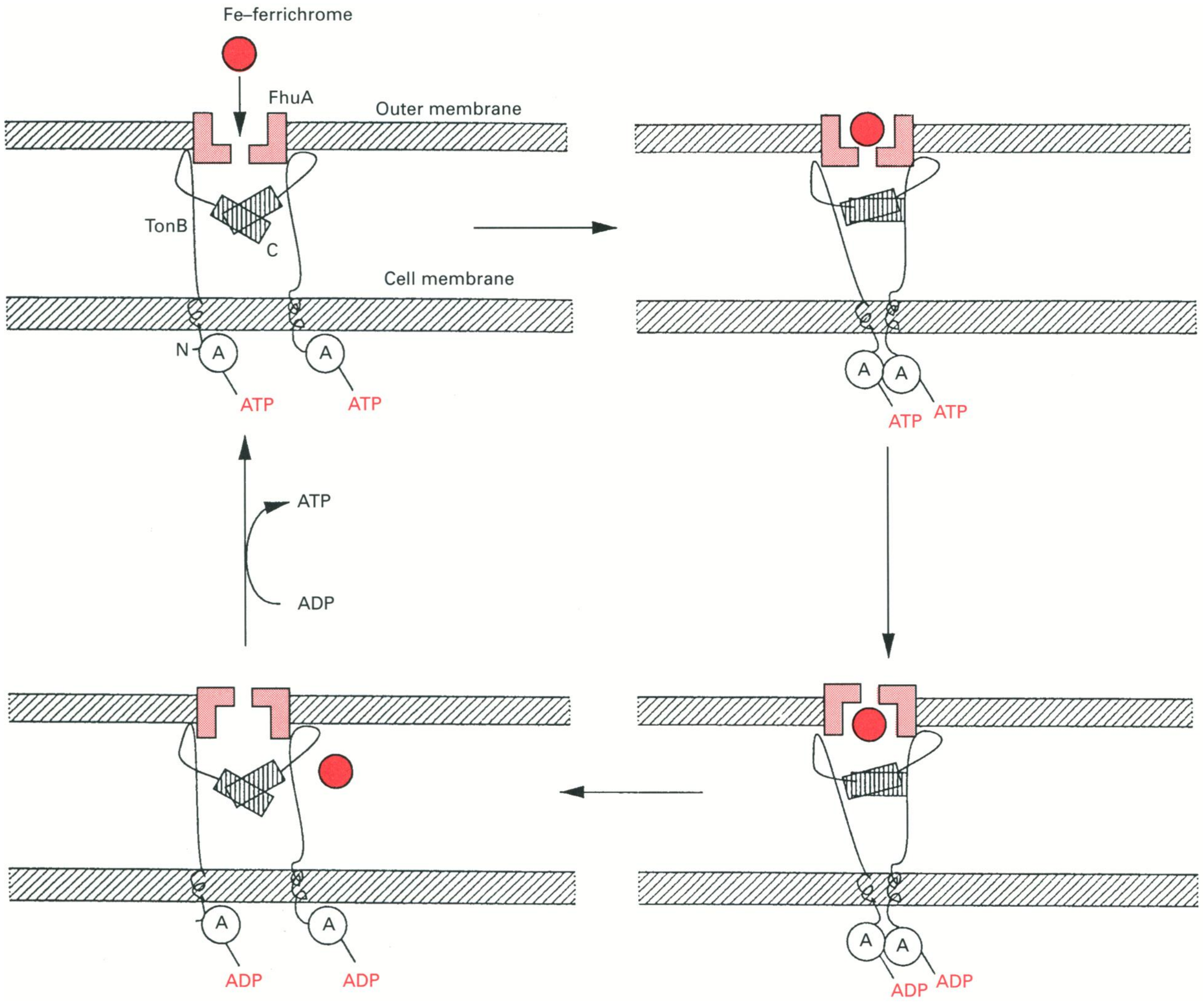

Figure 3 Hypothetical model of the mechanism of TonB

The proline-rich section extends across the periplasmic space and provides a mechanical link between an intracellular ATPase and the siderophore transporter FhuA. It is proposed that the binding of an iron–ferrichrome complex to FhuA triggers a conformational change in TonB that activates the ATPase. The hydrolysis of ATP produces a further conformational change in TonB that opens the siderophore channel, allowing transport of the iron complex. Exchange of ADP to ATP completes the cycle.

Table 2 Proteins with tandemly repeated proline-rich sequences

Name	Source	Sequence	Protein function	Comment	References
Salivary PRPs	Man, mouse	$(PQGPPQQGG)_n$	Polyphenol binding	Most of the protein is PRR	39,40
Mucins	Man	$(GSTAPPAHGVTSAPDTRPAP)_n$	Lubrication of epithelium	Glycosylated	41
Circumsporozoite protein	*Plasmodium berghei*	$(P_4NPND)_{13}PAPPQGN_3(PQ)_{17}$	Outer coat	Between two small domains	42
Gluten	Wheat	GYYPTSPQQ, PGQGQQ; many repeats	Cereal storage protein	Small N- and C-terminal domains	43
C hordein	Barley	PQQPFPQQ many times	Cereal storage protein	Small N- and C-terminal domains	44
Glutelin (zein)	Maize	VHLPPP eight times	Cereal storage protein	Small N- and C-terminal domains	45
Elastin	Man	$(VPGVG)_n$	Elastic connective tissue	β spiral?	46
Actin-binding protein	*Dictyostelium discoideum*	$[GYP(P)Q(P)]_5$	Actin assembly	Binds actin? at membrane?	47
Rhodopsin	Squid	$(PPQGY)_{10}$	Vision	Organizes microvillar structure?	48
RNA polymerase II	Man	YSPTSPS (26 times)	Transcription	Binds TFIID?	49
Synapsin I	Man	PQPAGPPAQQVPPPQQG (× 3)	Regulates vesicle release?	Binds vesicle and cytoskeleton?	50

Table 3 Non-repetitive PRRs

ψ indicates a hydrophobic amino acid.

Name	Source	Sequence	Protein function	Comment	References
CTF/NF-1 family	Man	PPHLNPQDPLKDLVSLACDPASQQPGPPTLRPTRPLQTVPLT	Transcription activator	Binds RNA polymerase II/TFIID(?)	57
Wilms tumour locus	Man	$PLPHFP_2SLP_2THSPTHP_3AP_3AP_9$	Transcription activator	Target unknown	58
VAMP-1	Man	PPSGPAPDAQGGAPGQPTGPPGAPP	Regulates vesicle release?	Binds vesicle and cytoskeleton?	59
Dynamin	Ox	PAVPPARPGSRGPAPGPPPAG	Mediates early stages of receptor-mediated endocytosis	PRR has critical regulatory role	60,61
shibire gene product	*Drosophila melanogaster*	PPLPPSTGRPAPAIPNRPGGGAPPLP	Endocytosis	Mechanochemical?	62
Consensus SH3-binding sequence	Mouse, rat	XPXXPPPψXP	Binds SH3	Signal transduction; cytoskeletal regulation	63
mSos1	Mouse	XPXXPPPψPPR	Binds SH3	Signal transduction	64
Vitelline	*Drosophila melanogaster*	$PYA_2(PA)_2YSAPA_2S_2GYPAP_2$ etc.	Eggshell structure	Membrane bound	65
Dec-1 eggshell protein	*Drosophila melanogaster*	PA 9 times; 27 P in 85 residues	Eggshell structure	Cleaved off? after covalent cross-linking	66
Colostrum PRP	Ovine whey	YVPLFP	Stimulates/suppresses immune response	Binds surface receptor(s)	67
Shaker family K^+ channel	Mammals	$PLPPALSP_3RP_3LSPVP$	Regulation of phosphorylation	Protein assembly and targetting	68
IgA_1	Man	PVPSTPPTPSPSTPPT	Immunoglobulin	Linker has binding function?	69
IgG Ike-N	Man	CPPCPAPE	Immunoglobulin	Linker has binding function?	70
Pyruvate dehydrogenase	*E. coli*	$GA_2PA_3PAKQEA_3PAPAAKAEAPA_3PA_2KA$	Dehydrogenase	Mobile linker	71
Cysteine proteinase	*T. brucei*	P_9	Divides two domains	C-terminal domain cleaved off on maturation	72
Proacrosin	Boar	P_{23} (42/127 residues are P)	Serine protease	May bind ovum; cleaved off on maturation	73
Orf E4	Human papillomavirus 8	$LPAP_4DH_2QDK(QT)_2P_3RP_5$	Associated with malignant conversion?		74
EBNA-2 nuclear protein	Epstein–Barr virus	$P_3LP_{26}SP_{11}$	Immortalization of B lymphocytes (?)		75
Huntington's disease gene product	Man	$Q_nP_{11}QLPQP_3$ (n=11–34 in normals; > 41 in disease states)	Determinant for Huntington's disease	Unknown function	76
Calcineurin A	Man	$MAAPEPARAAP_{11}GA$	Calmodulin-regulated phosphatase	N-terminus binds calmodulin along helix?	77

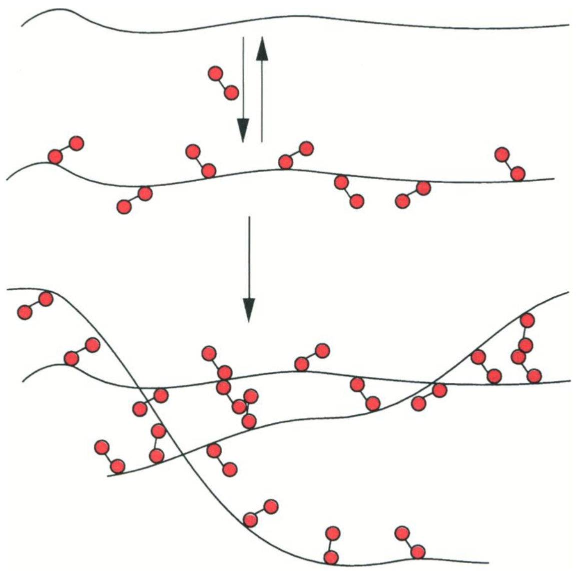

Figure 4 Model of the interaction between salivary PRPs and plant polyphenols

The polyphenols (shown in red) have several binding sites, and bind reversibly to the PRPs at prolines, which make up about 40% of the protein. Further intermolecular interactions lead to non-covalent cross-linking and then to precipitation of the complex. Eventually most of the polyphenols are precipitated.

termini on transcription factors has led to an attractive hypothesis supposing a preinitiation complex in which the proline-rich tails of several proteins interact to form a complex with indefinite stoichiometry but a limited range of spatial organization [79] (Figure 5). This hypothesis has received support from the finding that the C-terminal domain of RNA polymerase II can be phosphorylated, and that the phosphorylated form cannot bind to TFIID. This has been suggested to be the trigger to form an elongation-competent transcription complex [49].

The YSPTSPS repeat in RNA polymerase contains two copies of the SPXX motif. This motif is found in tandem repeats in a number of DNA-binding proteins and has been suggested to be itself a DNA-binding motif [80]. It is proposed that phosphorylation of the serine, which requires specific kinases, may regulate DNA binding [81]. The position of this hypothesis is still unclear. It may be that the SPXX motif is both a DNA-binding and a protein-binding motif; so far, the evidence for protein binding is more secure.

A protein designated PRP8 (pre-mRNA processing 8) has recently been characterized (J. Beggs, personal communication). This U5 snRNP protein is involved in stabilizing splicing complexes by binding to several components of the spliceosome [81a]. The N-terminus of PRP8 has four repeats of consensus sequence $LP_{5-8}G$ and appears to be required for cell viability, although whether it functions directly in the splicing process has not been determined. It may therefore play a similar role to the proline-rich C-terminal domain of RNA polymerase II.

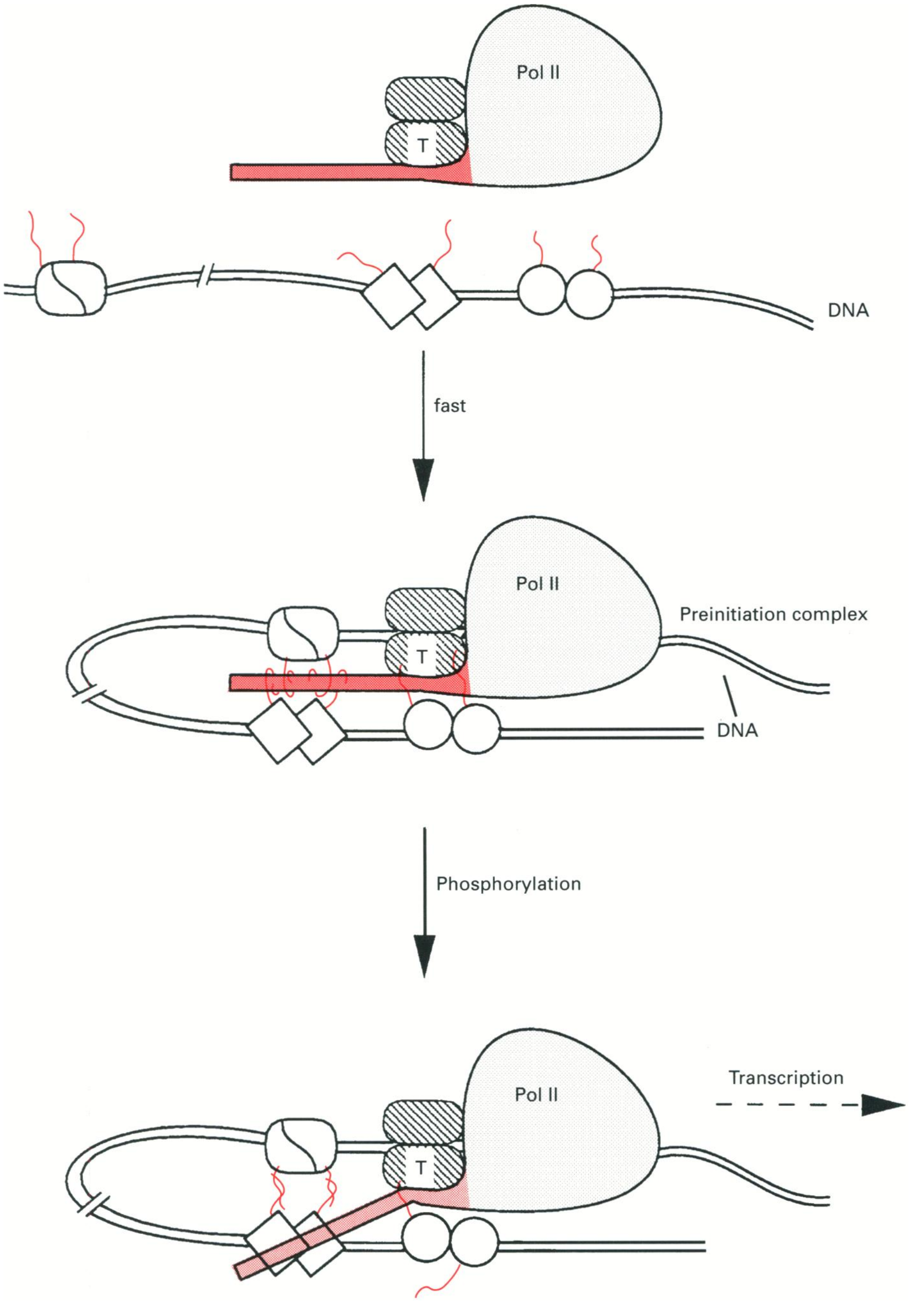

Figure 5 Hypothetical model of the preinitiation complex of RNA polymerase II

RNA polymerase II (Pol II) is shown with a globular domain and an extended proline-rich C-terminal domain (CTD), to which transcriptional activators can bind in a conformationally ill-defined manner. The protein marked T represents the class of specific polymerase II-associated proteins such as the TATA-binding element. The proline-rich C-terminal domain allows rapid binding of RNA polymerase II to the transcriptional activators, correct bending of the DNA, and the formation of a functional preinitiation complex. Phosphorylation of the C-terminal domain leads to its dissociation from the transcriptional activators and the start of transcription. Adapted from [56] and [79].

The second system that involves mutual interactions of several PRRs is the synaptic vesicle-associated neuronal proteins, of which the best characterized is synapsin I. Synapsin I is proline-rich throughout, containing in particular a 17-residue proline-rich sequence that occurs three times (residues 436–452, 460–476 and 620–636). The synapsins are soluble proteins that bind to the outside of synaptic vesicles and probably also to the cytoskeletal matrix [82]. Phosphorylation of serines in the PRR near the C-terminus of synapsin I (Table 2) leads to a reduction in its binding to an incompletely characterized vesicle-associated protein [83], implying a role for synapsin I in the phosphorylation-dependent transition of synaptic vesicles from a 'reserve pool' to a 'releasable pool' of vesicles [84]. At least two other synaptic proteins, vesicle-associated membrane protein 1 (VAMP-1) and synaptophysin [85], contain proline-rich segments (Table 3). These are probably intrinsic membrane proteins with proline-rich cytoplasmic regions, which function by interacting with synapsin I, in this case as part of the system for activating synaptic vesicles for release.

Other vesicle secretion and recycling systems appear to be regulated by homologous proteins [86]. Of these, the best understood is dynamin (Table 3), which binds to an SH3 (*src* homology 3) domain, and is therefore discussed in the next section with other SH3-binding domains.

Both the systems described in this section require the rapid and reversible association of several proteins into functional complexes, in which the prolines play a key part in the recognition and binding processes. Tight regulation of this association is necessary, which is achieved in both cases by phosphorylation of serines within the proline-rich sequence.

Non-repetitive PRRs

Several other proteins function in similar ways to the tandemly repeated PRRs described above (i.e. by facilitating protein–protein interactions), the only difference being that their PRRs are arranged in a non-repetitive manner. One such protein group that is currently of great interest is made up of the proteins that bind to SH3 domains (Table 3). SH3 domains are about 60 residues long, and have been found in association with catalytic domains, as in phospholipase Cγ, within structural proteins such as spectrin and myosin (in which they may regulate the cytoskeleton), and in small adaptor proteins such as Sem-5, Crk, Drk and Grb2. These adaptor proteins have received close attention because of their role in what now seems to be an evolutionarily conserved signalling pathway, leading from receptor binding to the stimulation of Ras and the start of a kinase cascade (Figure 6) [87,88]. The adaptor proteins consist of an SH2 domain, which typically binds to a phosphorylated receptor, and two SH3 domains, which bind to proline-rich sequences on the nucleotide-releasing factor Sos (similar in sequence to yeast CDC25 [90]). The binding is a somewhat atypical PRR-binding event. Although, like other PRR binding, more than one PRR is required [63,88], here the sequence requirements are rather stringent. Sos proteins from different organisms have fairly long and variable PRRs, but there is a consensus binding sequence, which is XPXXPPPψXPX (ψ indicates a hydrophobic residue), with prolines 2, 7 and 10 being essential [63]. Other residues, particularly 1 and 11, confer specificity on the binding [64]. A P_{10} sequence alone is incapable of binding. This means that each Sos binds with different affinities to SH3 domains from different sources [91].

In addition to their function in signal transduction, SH3 domain/PRR complexes also act as part of the vacuole sorting and receptor-mediated endocytosis pathways, which probably have many features in common with signal transduction. Thus it is now clear that one route of receptor-mediated endocytosis involves binding of the SH2 domains of phosphatidylinositol 3-kinase to autophosphorylated receptors [92]. PtdIns 3-kinase also has an SH3 domain, which binds to a PRR at the C-terminus of dynamin [61]. Dynamin is a GTP-binding protein, probably a GTPase, and shows sequence similarity to the *Drosophila shibire* gene product [62], mutants of which produce paralysis due to a defect in endocytosis. The N-terminal GTP-binding domain is similar in sequence to the yeast VPS1 (vacuolar protein sorting)/SPO15 gene product, which is involved both in vacuolar sorting and in meiotic chromosome segregation [93], while the C-terminal PRR is similar to the kinesin-related yeast KAR3 protein, which binds microtubules *in vivo* [94]. Dynamin is also thought to bind to microtubules. It is therefore tempting to postulate a pathway for endocytosis analogous to the signal transduction pathway shown in Figure 6, in which the chain of signal transduction proteins (phosphorylated receptor–Grb2–mSos1–Ras) is replaced by the chain phosphorylated receptor–PtdIns 3-kinase–dynamin–microtubule. However, both PtdIns 3-kinase and dynamin appear to have additional enzyme functions not possessed by Grb2 and mSos1.

Several SH3 domain structures are now available [61,95,96]. In all cases the PRR-binding site is a smooth hydrophobic surface, rich in conserved aromatic amino acids, with charged amino acids at the periphery. It has been suggested [61] that the hydrophobic surface provides a general platform for binding the

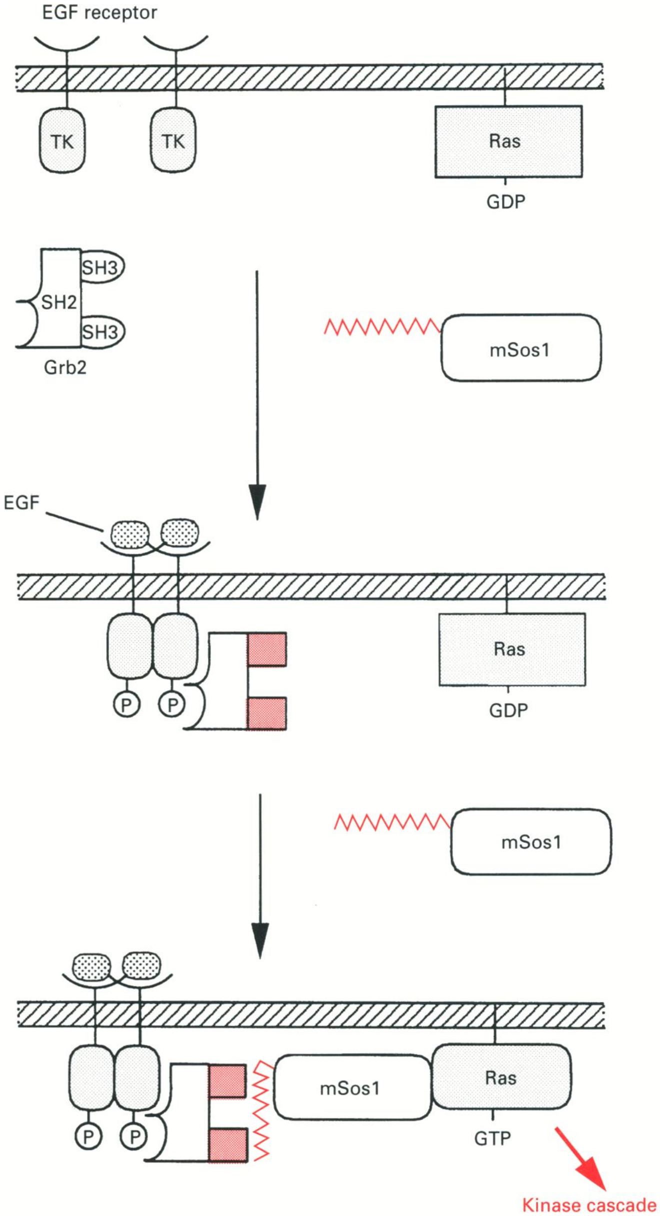

Figure 6 Mechanism of the activation of Ras by the epidermal growth factor (EGF) receptor in mammals

EGF binding leads to autophosphorylation of the tyrosine kinase (TK) domain of the receptor. The phosphorylated receptor then binds to the SH2 domain of the adaptor protein Grb2, which produces a conformational change in its SH3 domains [89], allowing Grb2 in turn to bind to mSos1 via two proline-rich regions with consensus sequence XPXXPPPψXP (shown in red). mSos1 is thought to act constitutively as a nucleotide exchanger, and its relocation to the plasma membrane activates Ras. The roles of EGF receptor/Grb2/mSos1 in mammals are taken respectively by Sevenless/Drk/Sos in *Drosophila*, and by Let-23/Sem-5/unknown protein in *Caenorhabditis elegans*. Adapted from [87].

PRR, with selectivity resulting from the charged amino acids. The conformation of the bound PRR has not been determined, but from the shape of the binding site it seems likely to be an extended polyproline II helix. These results are consistent with the PRR acting as a 'sticky arm', binding rapidly and reversibly to SH3 domains.

There are several proteins which are proline-rich throughout their entire sequence, notably the caseins and amelogenin. Caseins form about 80% of skim milk protein [97]. They have been divided by their electrophoretic mobility into α_s, β, κ and γ caseins, constituting respectively about 50, 10, 30 and 5% of skim milk protein. They all contain prolines spread throughout the sequence in a fairly regular (but not repetitive) manner, with a representative bovine α_s casein having proline as 17 out of 186 residues, and a β casein A^2 having 35 prolines out of 209 residues. The caseins assemble into micelles and clot by hydrolysis of a specific bond in κ casein, catalysed by proteases present in the stomach [98]. The caseins are phosphorylated and bind calcium. The structure of the micelle is not clearly understood [99], but is apparently produced by a semi-ordered aggregation of core polymers formed by the association of extended polypeptide chains. The regularly spaced proline residues are presumably important in maintaining an extended chain conformation and also in guiding associative processes.

Caseins, along with many other proteins, have sequences that are rich in proline, glutamic acid, serine and threonine and are flanked by positively charged residues. These sequences have been dubbed PEST sequences and have been suggested to be a signal for rapid degradation in eukaryotic cells, especially when phosphorylated [100,101]. The mechanism of degradation is as yet unknown, although it has been suggested that it may involve calcium-activated calpain proteolysis. The role of proline in the PEST sequence is unclear, but the similarities of this system to the phosphorylation-dependent RNA polymerase II and vesicle-associated protein systems may imply some evolutionary or functional similarities.

Amelogenin (Table 1) is the predominant constituent of developing teeth. Bovine amelogenin has 170 residues and contains 49 prolines, of which nine form a $(QPX)_9$ motif, with X = L, H or M [35]. Again, the protein functions by aggregation, and one can assume that the protein is largely extended, particularly the $(QPX)_9$ motif, which is presumably an approximate polyproline II helix.

Vitelline and the dec-1 eggshell protein are involved in the strengthening of eggshell structure and therefore have similar functions to the tandemly repeated circumsporozoite protein described above. The other proteins listed in Table 3 have more poorly defined functions. A protein isolated from ovine colostrum is proline-rich throughout much of its sequence, and regulates the immune response, by binding in some way to surface receptors. The interdomain linkers in immunoglobulins, which constitute the main difference between different IgG subtypes, are rich in prolines. The function of the proline residues may be simply to maintain an extended structure with limited mobility. However, since the Fc receptor binding site is located close to the interdomain 'hinge', the prolines may also be involved in interactions with Fc receptors.

There are a few other proline-rich sequences that appear to act solely as linkers, with no binding function. The most well studied is the approximately 30-residue linker that connects lipoyl domains in the dihydrolipoyl acetyltransferase component of 2-oxoacid dehydrogenase complexes [71]. This sequence contains essentially all-*trans* proline residues and is extended and mobile, even in the intact protein complex [102]. The function of the linker is to transfer acyl groups between different active sites in the complex. The proline residues limit the conformational freedom of the linkers and prevent adjacent lipoyl domains from interacting with each other, which would reduce the enzymic efficiency of the complex [103]. The mixed alanine/proline sequence has more mobility than either an all-alanine or an all-proline sequence [104].

Table 3 ends with some striking polyproline sequences. In both *Trypanosoma brucei* protease and proacrosin these sequences appear to have multiple functions: interacting with other proteins, separating two domains and acting as cleavage sites after the protein has attached to its target. The remarkable polyproline sequences in papillomavirus and Epstein–Barr virus have as yet no known function, but they are likely to involve protein–protein association in a manner similar to the CTF/NF-1 family of transcription factors. It is tempting to speculate that the Huntington's disease gene product is likewise involved in protein–protein association.

Calcineurin A, like the Wilms tumour protein, contains a long stretch of continuous prolines, 11 in this case. It is one of the few proteins for which a specific (but highly speculative) model for PRP–protein interaction has been proposed. It is suggested that the P_{11} sequence, if in the form of a polyproline II helix, could extend along the calmodulin central helix, thereby presenting the few residues N-terminal to the P_{11} stretch in a fixed position on one calmodulin domain, determined by the calcineurin-binding site on the other domain. As we have seen, this requires an uncharacteristically specific interaction mode for the proline residues.

In summary, all of the proteins presented in Tables 1–3 are very likely to have binding as a major function of the PRR, and in most cases binding is the only identifiable function. In a large number of cases the binding target is the cytoskeletal matrix, but many other ligands are also found. Proline-rich regions may therefore be taken to act as 'sticky arms' extending out from the rest of the protein (Figure 2). The binding is regulated by phosphorylation where required.

The next section describes proline-rich regions that have a structural role rather than a role in binding. This change in function is achieved by hydroxylation of some or all of the proline residues.

Hydroxyproline-rich proteins

The most well-known PRP, and probably the most abundant in the animal kingdom, is collagen (Table 4). It is a stiff high-tensile fibre found in connective tissue such as tendons and skin, and comprises up to a third of total body protein. Three chains are coiled around each other to give a triple-stranded helix, which is stabilized by hydrogen bonding of glycine between strands. Each strand forms a polyproline II helix. The regular sequence is crucial for maintaining the collagen structure; models of the collagen structure show that substitutions by other residues lead to steric clashes or unpaired hydrogen bonds.

The extended terminus of the blood complement protein C1q also seems to be collagen-like. It associates to form a triple helix, while a break in the $(GXX)_n$ sequence at around residue 39 forces the individual chains to bend, forming a 'bunch of tulips' structure.

The role of extensins is not clear. They form 5–10% of the plant cell wall, and probably strengthen it by covalent cross-linking of tyrosine residues. However, it is likely that the initial structure is produced by interwoven and non-covalently associated extensin chains. Extensins accumulate in plant cell walls upon wounding [113] and pathogen attack [114], indicating an

Table 4 Proteins rich in hydroxyproline

In this table, *P* denotes hydroxyproline

Name	Source	Sequence	Protein function	Comment	References
Collagen	Man	$(\mathrm{GP}P)_{350}$ (with variations on P and *P*)	Stiff connective fibre	Triple helix	105
C1q	Man	$(P\mathrm{GX})_3(\mathrm{ASXGX})_2(P\mathrm{GX})_2\mathrm{PGX}P$	Blood defence system	Kinked triple helix	106
Extensin P1	Tomato, carrot etc.	$[\mathrm{S}PPPP(\mathrm{VKPYHP})\mathrm{T}P\mathrm{VKY}]_n$	Cell wall constituent	Protects plant against damage?	107–109
Hydroxyproline-rich glycoprotein	Sorghum	$(\mathrm{PATKPPTPPVYTPSPKP})_n$	Cell wall constituent	Many are hydroxylated	110
Cucumber peel cupredoxin	Cucumber	*PPP*SSS*PP*SSVM*PPP*VMP*PP*S*P*S	Electron transfer	PRR is C-terminal extension (locates in matrix?)	111
Extracellular matrix protein	*Volvox carteri*	$P_2\mathrm{S}P_3\mathrm{S}P\mathrm{R}P_2\mathrm{SP}_4\mathrm{S}P\mathrm{S}P_{17}\mathrm{S}P_{18}\mathrm{S}P\mathrm{S}P_2$	Strengthens cell matrix	Also role in tannin binding?	112

additional role in defence, possibly due to agglutination of invading bacteria ([115] and refs. cited therein). This function is reminiscent of that carried out by the salivary PRPs, which also agglutinate bacteria.

Many microbial polysaccharide-digesting enzymes consist of two domains, a catalytic domain and a sugar-binding domain. The two domains are separated by a semi-rigid linker, whose function appears to be largely to hold the two domains apart, although it may play some role in stabilization of the domains against heat or chemical denaturation [116]. The linker can have a wide variety of sequences (reviewed for β-1,4-glycanases in [117]), which are generally rich in hydroxyamino acids (serine and threonine) or proline, or both. In the linkers of fungal proteins the hydroxyamino acids are heavily glycosylated, which both rigidifies and protects the linker [118].

The PRR of the extracellular matrix protein from *Volvox carteri* may play a similar role. It is possible that the PRR may also be involved in protection of the cell surface, by analogy with the epithelial mucins discussed above.

PROLINE IS INVOLVED IN BINDING

In this review I have sought to demonstrate that proline does not merely act as a spacer, but frequently has an important role in binding as well. This is true both for the $(\mathrm{XP})_n$ sequences and for the longer more varied tandem repeats. Clearly, the binding generated cannot be highly specific, but it can be both very rapid (because of the small surface area and flexibility involved) and remarkably strong. Less specific binding can be of positive advantage in some cases, allowing a wider range of ligands to be bound. This is of relevance to salivary PRPs, which have to bind a wide range of polyphenols and other substrates, and possibly also for the transcription factors, which probably need to bind to a range of different proteins involved in the initiation of transcription. Commenting on these systems, Sigler [79] writes: 'These systems ... share ... the need for a mechanism by which many and various proteins can interact with a common cellular element. These flexible and variable contact patterns depart from the traditional view of specific molecular interactions gained from studying assemblies of globular molecules that give crystalline images'. This comment has been fully borne out by the fuller and more recent data reported here.

The strength of the binding derives from the fact that proline-rich polypeptides have highly restricted mobility (and therefore relatively low entropy) even before binding. Thus binding leads to a smaller drop in entropy than it would do for a normal, more flexible, peptide, and hence a greater overall binding energy is achieved. To take an example, if we assume that each dipeptide Xaa-Pro has only two rather than the normal four degrees of rotational freedom around the backbone bonds, and we further assume that on binding all rotational freedom is lost, then an Xaa-Pro dipeptide loses two fewer degrees of freedom on binding. It has been estimated [119] that each degree of rotational freedom is worth 5–7 $\mathrm{kJ \cdot mol^{-1}}$ at 300 K; more recent estimates [120] place the figure somewhat lower, at around 3.5 $\mathrm{kJ \cdot mol^{-1}}$. Therefore the ΔG for the binding of an octapeptide (i.e. four dipeptides) could increase by 14 $\mathrm{kJ \cdot mol^{-1}}$, increasing the association constant from (for example) 10^3 to $2.7 \times 10^5\ \mathrm{M^{-1}}$, a value approaching a reasonable number for specific binding (cf. values of 10^5–$10^7\ \mathrm{M^{-1}}$ for peptides binding to major histocompatibility complex class I molecules of appropriate specificity [121]).

The multiple tandem repeats often found in PRRs appear to be another device for increasing weak binding, in much the same way as divalent antibodies bind to antigens much more strongly than monovalent ones [52]. Thus, for example, salivary PRPs act to bind to and precipitate dietary polyphenols. The precipitation reaction is mediated by cross-linking of one PRP to several polyphenols, and of one polyphenol to several PRPs. As discussed above, PRRs often have the additional function of a structural element, and multiple repeats are also necessary for extending the length of the protein. For example, they provide the coccal cell wall proteins with the length to span the peptidoglycan layer, and incidentally thereby take full advantage of it for binding purposes.

Similar observations have been made for so-called protein modules [122], which are single small protein domains that are repeated many times and which occur in a wide range of vertebrate blood and cell-surface receptor proteins. For example, fibronectin consists of 29 similar modules in a single polypeptide chain. The major function of these modular proteins appears to be in protein–protein binding, but they probably have an additional function of spacers, separating one functional part of the protein from another. Their functions therefore closely parallel those of the tandem proline-rich repeats discussed here.

The non-specific nature of the binding is supported by evidence showing that the exact number or sequence of the proline-rich repeats makes little difference to protein function. Thus deletion of almost half of the repeats in the C-terminal domain of RNA polymerase II still produces more or less functional proteins [123].

THE NATURE OF THE BINDING INTERACTION

Apart from the entropy advantage, proline has other features that make it a good ligand. It has a large flat hydrophobic surface and therefore binds well to other flat hydrophobic surfaces such as aromatic rings. Indeed, it is of more than passing interest that

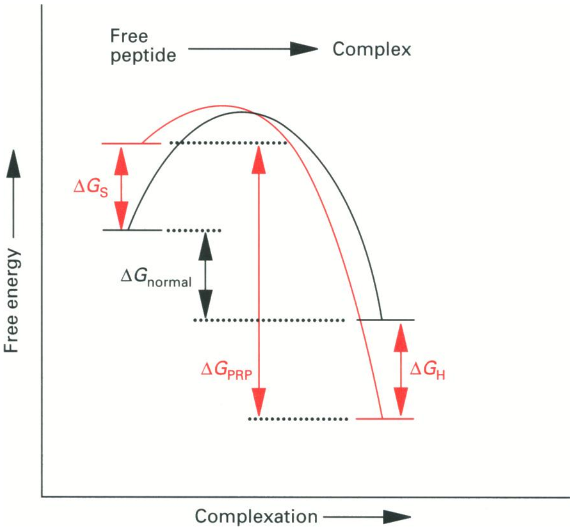

Figure 7 Schematic free energy diagram for the binding of a proline-rich peptide (red) or a normal peptide (black) to a globular protein

Because the free PRP is less flexible, it has less entropy and therefore a greater free energy than the normal peptide, by an amount ΔG_S. The bound peptides have similar entropy, but the PRP binds with more favourable enthalpy (heat of binding), by an amount ΔG_H, because of its more electron-rich amide bond. The overall binding energy of the PRP, ΔG_{PRP}, is therefore more favourable than that of the normal peptide, ΔG_{normal}, by $\Delta G_S + \Delta G_H$.

one of the very few proteins shown to be a receptor for a PRP, the SH3 domain, has a binding site lined with conserved aromatic residues [95]. The salivary PRPs have also been shown to interact with their principal physiological target, polyphenols, via proline residues [53]. The crystal structure of avian pancreatic polypeptide [16] shows that the polyproline structure is stabilized by interactions between the prolines in the N-terminal polyproline II helix and non-polar side-chains (many of them aromatic) in the C-terminal α-helix. Neuropeptide Y [17] is stabilized by similar interactions.

Although proline cannot act as a hydrogen bond donor, it is a very good hydrogen bond acceptor, possibly because the electron-donating potential of the methylene group attached to the amide nitrogen causes the amide carbonyl to be electron-rich [15,52,124]. It is presumably this property that causes proline-rich peptides to be highly soluble in water, and leads to confusion as to whether proline should be classed as a hydrophobic or a hydrophilic residue. Moreover, thermodynamic studies [125,126] have shown that tertiary amides such as the Xaa-Pro peptide bond are preferred to secondary amides as hydrogen bond donors, the enthalpy (heat) change for tertiary amides on forming a hydrogen bond being about 50% more favourable than for secondary amides. This observation may well be one aspect of a more general phenomenon, namely that if a solute is well solvated, its tendency to interact with other species will be reduced. Proline is more poorly solvated than other amino acids, in that it is hydrated by fewer water molecules around the amide bond, and so interacts more strongly with other solutes. In summary, proline may well be a preferred ligand enthalpically as well as entropically (Figure 7).

It may be of significance that many proline-rich sequences also contain large numbers of glutamine residues. Particularly striking are the nuclear protein SNF5, a transcription activator [127], which is proline-rich but also contains the sequence Q_7HQ_{37}; a *Drosophila* 20-hydroxyecdysone-inducible steroid nuclear hormone receptor [128], which has a P_{15} stretch immediately followed by a Q_6 stretch; and the Huntington's gene product (Table 3). It has been suggested [129] that 'Q-linkers' form an identifiable class of interdomain linkers in multidomain regulatory proteins. Q-linkers share a number of characteristics in common with proline-rich linkers, and many proline-rich linkers also contain high proportions of glutamine, for example the salivary PRPs and the cereal storage proteins. The significance of this is unclear, but it may be relevant to note that the second most likely residue to appear in a polyproline helix segment in globular proteins is glutamine (proline being the most likely) [18]. Indeed, glutamine is the only residue other than proline to have a Chou–Fasman conformational parameter that is higher for a polyproline II helix than for any other category of secondary structure. Thus it may be that glutamine is preferred as a linker component because, like proline, it preferentially forms polyproline II helices and makes an extended, conformationally restricted, polypeptide chain.

CONCLUDING REMARKS

Although proline often plays a purely structural role, particularly in hydroxyproline-rich proteins, the dominant role emphasized throughout this review is that of binding, in a non-stoichiometric but functionally important way. The binding ability has been rationalized as deriving from the restricted mobility of proline, which reduces the unfavourable entropy loss of peptides on binding, the flat hydrophobic surface of proline and the unique characteristics of the amide bond preceding proline, which make it a strong hydrogen bond acceptor.

Because of the rapid but non-specific nature of the interaction, PRRs are often involved in complex multiple protein association phenomena, as found for the RNA polymerase II pre-initiation complex, the vesicle-associated proteins and the SH3 domain-binding proteins. Studies of these associative processes are still in their infancy, and much work remains to be done to characterize their structure and kinetics. Many of these processes seem to be controlled or triggered by phosphorylation, and therefore one can expect that progress in understanding them will depend on characterization of the specific kinases involved.

The interactions seen for PRRs are qualitatively different from the archetypical enzyme–substrate interaction, since they rely on multiple weak binding sites rather than specific lock and key (or even induced fit) binding. The association will therefore not be amenable to site-directed mutagenesis, but will require larger-scale deletions, insertions and transpositions, as well as physicochemical studies, for example rapid kinetics and n.m.r. It is becoming clear that many key processes in the cell require the co-operative association of several different proteins into a functional complex. Understanding how such processes are regulated is one of the major challenges facing current enzymology, and it can only be tackled by the combined use of genetic and physicochemical techniques.

I thank Nicky Murray, John Guest and Jean Beggs for helpful discussions.

REFERENCES

1 Balasubramanian, R., Lakshiminarayanan, A. V., Sabesan, M. N., Tegoni, G., Venkatesan, K. and Ramachandran, G. N. (1971) Int. J. Protein Res. **3**, 25–33
2 Morris, A. L., MacArthur, M. W., Hutchinson, A. G. and Thornton, J. M. (1992) Proteins Struct. Funct. Genet. **12**, 345–364
3 Kraulis, P. J. (1991) J. Appl. Crystallogr. **24**, 946–950
4 MacArthur, M. W. and Thornton, J. M. (1991) J. Mol. Biol. **218**, 397–412
5 Nicholson, H., Tronrud, D. E., Becktel, W. J. and Matthews, B. W. (1992) Biopolymers **32**, 1431–1441

6 Hurley, J. H., Mason, D. A. and Matthews, B. W. (1992) Biopolymers **32**, 1443–1446

7 Williams, K. A. and Deber, C. M. (1991) Biochemistry **30**, 8919–8923

8 Sankararamakrishnan, R. and Vishveshwara, S. (1993) Proteins Struct. Funct. Genet. **15**, 26–41

9 Richardson, J. S. and Richardson, D. C. (1988) Science **240**, 1648–1652

10 Deber, C. M., Bovey, F. A., Carver, J. P. and Blout, E. R. (1970) J. Am. Chem. Soc. **92**, 6191–6198

11 Helbecque, N. and Loucheux-Lefebvre, M. H. (1982) Int. J. Peptide Protein Res. **19**, 94–101

12 Okabayashi, H., Isemura, T. and Sakakibara, S. (1968) Biopolymers **6**, 323–330

13 Dukor, R. K. and Kiederling, T. A. (1991) Biopolymers **31**, 1747–1761

14 Dukor, R. K., Kiederling, T. A. and Gut, V. (1991) Int. J. Peptide Protein Res. **38**, 198–203

15 Cowan, P. M. and McGavin, S. (1955) Nature (London) **176**, 501–503.

16 Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. and Wu, C.-W. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 4175–4179

17 Darbon, H., Bernassau, J.-M., Deleuze, C., Chenu, J., Roussel, A. and Cambillau, C. (1992) Eur. J. Biochem. **209**, 765–771

18 Adzhubei, A. A. and Sternberg, M. J. E. (1993) J. Mol. Biol. **229**, 472–493

19 Grathwohl, C. and Wüthrich, K. (1976) Biopolymers **15**, 2025–2041

20 Lewis, P. N., Momany, F. A. and Scheraga, H. A. (1973) Biochim. Biophys. Acta **303**, 211–229

21 Green, J. D. F., Perham, R. N., Ullrich, S. J. and Appella, E. (1992) J. Biol. Chem. **267**, 23484–23488

22 Bhandari, D. G., Levine, B. A. and Yeadon, M. E. (1986) Eur. J. Biochem. **160**, 349–356

23 Frank, G. and Weeds, A. G. (1974) Eur. J. Biochem. **44**, 317–334

24 Hejtmancik, J. F., Thompson, M. A., Wistow, G. and Piatigorsky, J. (1986) J. Biol. Chem. **261**, 982–987

25 Berbers, G. A. M., Hoekman, W. A., Bloemendal, H., de Jong, W. W., Kleinschmidt, T. and Braunitzer, G. (1983) FEBS Lett. **161**, 225–229

26 Chen, R., Schmidmayr, W., Krämer, C., Chen-Schmeisser, U. and Henning, U. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 4592–4596

27 Roditi, I., Schwarz, H., Pearson, T. W., Beecroft, R. P., Liu, M. K., Richardson, J. P., Bühring, H.-J., Pleiss, J., Bülow, R., Williams, R. O. and Overath, P. (1989) J. Cell Biol. **108**, 737–746

28 Gaisser, S. and Braun, V. (1991) Mol. Microbiol. **5**, 2777–2787

29 Brewer, S., Tolley, M., Trayer, I. P., Barr, G. C., Dorman, C. J., Hannavy, K., Higgins, C. F., Evans, J. S., Levine, B. A. and Wormald, M. R. (1990) J. Mol. Biol. **216**, 883–895

30 Timoney, J. F., Muktar, M. and Ding, J. (1991) in Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci (Dunny, G. M., Cleary, P. P. and McKay, L. L., eds.), pp. 160–164, American Society of Microbiology, Washington

31 Hedén, L.-O., Frithz, E. and Lindahl, G. (1991) Eur. J. Immunol. **21**, 1481–1490

32 Pancholi, V. and Fischetti, V. A. (1988) J. Bacteriol. **170**, 2618–2624

33 Fahnestock, S. R., Alexander, P., Nagle, J. and Filpula, D. (1986) J. Bacteriol. **167**, 870–880

34 Li, L. J., Dougan, G., Novotny, P. and Charles, I. G. (1991) Mol. Microbiol. **5**, 409–417

35 Takagi, T., Suzuki, M., Baba, T., Minegishi, K. and Sasaki, S. (1984) Biochem. Biophys. Res. Commun. **121**, 592–597

36 Abillon, E., Bremier, L. and Cardinaud, R. (1990) Biochim. Biophys. Acta **1037**, 394–400

37 Hannavy, K., Barr, G. C., Dorman, C. J., Adamson, J., Mazengera, L. R., Gallagher, M. P., Evans, J. S., Levine, B. A., Trayer, I. P. and Higgins, C. F. (1990) J. Mol. Biol. **216**, 897–910

38 Fischetti, V. A., Pancholi, V. and Schneewind, O. (1991) in Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci (Dunny, G. M., Cleary, P. P. and McKay, L. L., eds.), pp. 290–294, American Society of Microbiology, Washington

39 Bennick, A. (1982) Mol. Cell. Biochem. **45**, 83–99

40 Layfield, R., Bannister, A. J., Pierce, E. J. and McDonald, C. J. (1992) Eur. J. Biochem. **204**, 591–597

41 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. and Wilson, D. (1990) J. Biol. Chem. **265**, 15286–15293

42 Eichinger, D. J., Arnot, D. E., Tam, J. P., Nussenzweig, V. and Enea, V. (1986) Mol. Cell Biol. **6**, 3965–3972

43 Field, J. M., Tatham, A. S. and Shewry, P. R. (1987) Biochem. J. **247**, 215–221

44 Tatham, A. S., Drake, A. F. and Shewry, P. R. (1985) Biochem. J. **226**, 557–562

45 Pons, M., Feliz, M., Celma, C. and Giralt, M. (1987) Magn. Reson. Chem. **25**, 402–406

46 Venkatachalam, C. M. and Urry, D. W. (1981) Macromolecules **14**, 1225–1229

47 Noegel, A. A., Gerisch, G., Lottspeich, F. and Schleicher, M. (1990) FEBS Lett. **266**, 118–122

48 Hall, M. D., Hoon, M. A., Ryba, N. J. P., Pottinger, J. D. D., Keen, J. N., Saibil, H. R. and Findlay, J. B. C. (1991) Biochem. J. **274**, 35–40

49 Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. and Aloni, Y. (1992) Cell **69**, 871–881

50 McCaffery, C. A. and DeGennaro, L. J. (1986) EMBO J. **5**, 3167–3173

51 Mehansho, H., Butler, L. G. and Carlson, D. M. (1987) Annu. Rev. Nutr. **7**, 423–440

52 Hagerman, A. E. and Butler, L. G. (1981) J. Biol. Chem. **256**, 4494–4497

53 Murray, N. J., Williamson, M. P., Lilley, T. H. and Haslam, E. (1994) Eur. J. Biochem., in the press

54 Nicholson, R. L., Butler, L. G. and Asquith, T. N. (1986) Phytopathology **76**, 1316–1318

55 Ludevid, M. D., Torrent, M., Martínez-Izquierdo, J. A., Puigdomènech, P. and Palau, J. (1984) Plant Mol. Biol. **3**, 605–611

56 Koleske, A. J., Buratowski, S., Nonet, M. and Young, R. A. (1992) Cell **69**, 883–894

57 Mermod, N., O'Neill, E. A., Kelly, T. J. and Tjian, R. (1989) Cell **58**, 741–753

58 Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H. and Bruns, G. A. P. (1990) Nature (London) **343**, 774–778

59 Trimble, W. S., Cowan, D. M. and Scheller, R. H. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 4538–4542

60 van der Bliek, A. M., Redelmeier, T. E., Damke, H., Tisdale, E. J., Meyerowitz, E. M. and Schmid, S. L. (1993) J. Cell Biol. **122**, 553–563

61 Booker, G. W., Gout, I., Downing, A. K., Driscoll, P. C., Boyd, J., Waterfield, M. D. and Campbell, I. D. (1993) Cell **73**, 813–822.

62 van der Bliek, A. M. and Meyerowitz, E. M. (1991) Nature (London) **351**, 411–414

63 Ren, R., Mayer, B. J., Cicchetti, P. and Baltimore, D. (1993) Science **259**, 1157–1161

64 Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T. and Bowtell, D. (1993) Nature (London) **363**, 83–85

65 Gigliotti, S., Graziani, F., De Ponti, L., Rafti, F., Manzi, A., Lavorgna, G., Gargiulo, G. and Malva, C. (1989) Dev. Genet. **10**, 33–41

66 Waring, G. L., Hawley, R. J. and Schoenfeld, T. (1990) Dev. Biol. **142**, 1–12

67 Janusz, M., Wieczorek, Z., Spiegel, K., Kubik, A., Szewczuk, Z., Siemion, I. and Lisowski, J. (1987) Mol. Immunol. **24**, 1029–1031

68 McCormack, T., Vega-Saenz De Miera, E. C. and Rudy, B. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 5227–5231

69 Liu, Y. S. V., Low, T. L. K., Infante, A. and Putnam, F. W. (1976) Science **193**, 1017–1020

70 Endo, S. and Arata, Y. (1985) Biochemistry **24**, 1561–1568

71 Russell, G. C. and Guest, J. R. (1991) Biochim. Biophys. Acta **1076**, 225–232

72 Mottram, J. C., North, M. J., Barry, J. D. and Coombs, G. H. (1989) FEBS Lett. **258**, 211–215.

73 Adham, I. M., Klemm, U., Maier, W.-M., Hoyer-Fender, S., Tsaousidou, S. and Engel, W. (1989) Eur. J. Biochem. **182**, 563–568

74 Fuchs, P. G., Iftner, T., Weninger, J. and Pfister, H. (1986) J. Virol. **58**, 626–634

75 Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Séguin, C., Tuffnell, P. S. and Barrell, B. G. (1984) Nature (London) **310**, 207–211

76 MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. M., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P. and 42 others (1993) Cell **72**, 971–983

77 Guerini, D. and Klee, C. B. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 9183–9187

78 Sasaki, H., Yokoyama, E. and Kuroiwa, A. (1990) Nucleic Acids Res. **18**, 1739–1747

79 Sigler, P. B. (1988) Nature (London) **333**, 210–212

80 Suzuki, M. (1989) J. Mol. Biol. **207**, 61–84

81 Suzuki, M., Sohma, H., Yazawa, M., Yagi, K. and Ebashi, S. (1990) J. Biochem. (Tokyo) **108**, 356–364

81a Brown, J. D. and Beggs, J. D. (1992) EMBO J. **11**, 3721–3729

82 De Camilli, P. and Greengard, P. (1986) Biochem. Pharmacol. **24**, 4349–4357

83 Benfenati, F., Valtorta, F., Rubenstein, J. L., Gorelick, F. S., Greengard, P. and Czernik, A. J. (1992) Nature (London) **359**, 417–420

84 Benfenati, F., Valtorta, F. and Greengard, P. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 575–579

85 Trimble, W. S. and Scheller, R. H. (1988) Trends Neurochem. Sci. **11**, 241–242

86 Bennett, M. K. and Scheller, R. H. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 2559–2563

87 Schlessinger, J. (1993) Trends. Biochem. Sci. **18**, 273–275

88 Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B. and Schlessinger, J. (1993) Nature (London) **363**, 85–88

89 Buday, L. and Downward, J. (1993) Cell **73**, 611–620

90 Bowtell, D., Fu, P., Simon, M. and Senior, P. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 6511–6515

91 Cicchetti, P., Mayer, B. J., Thiel, G. and Baltimore, D. (1992) Science **257**, 803–806

92 Panayotou, G. and Waterfield, M. D. (1992) Trends Cell Biol. **2**, 358–360

93 Collins, C. A. (1991) Trends Cell Biol. **1**, 57–60

94 Meluh, P. B. and Rose, M. D. (1990) Cell **60**, 1029–1041

95 Yu, H., Rosen, M. K., Shin, T. B., Seidel-Dugan, C., Brugge, J. S. and Schreiber, S. L. (1992) Science **258**, 1665–1668

96 Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. and Saraste, M. (1992) Nature (London) **359**, 851–855

97 Whitney, R. M., Brunner, J. R., Ebner, K. E., Farrell, H. M., Josephson, R. V., Morr, C. V. and Swaisgood, H. E. (1976) J. Dairy Sci. **59**, 795–815

98 Dalgleish, D. G. (1982) in Developments in Dairy Chemistry, vol. 1 (Fox, R. F., ed.), pp.157–187, Applied Science, London

99 Farrell, H. M. (1973) J. Dairy Sci. **56**, 1195–1206

100 Rogers, S., Wells, R. and Rechsteiner, M. (1986) Science **234**, 364–368

101 Chevaillier, P. (1993) Int. J. Biochem. **25**, 479–482

102 Radford, S. E., Laue, E. D., Perham, R. N., Martin, S. R. and Appella, E. (1989) J. Biol. Chem. **264**, 767–775

103 Machado, R. S., Guest, J. R. and Williamson, M. P. (1993) FEBS Lett. **323**, 243–246

104 Turner, S. L., Russell, G. C., Williamson, M. P. and Guest, J. R. (1993) Protein Eng. **6**, 101–108

105 Eyre, D. R. (1980) Science **207**, 1315–1322

106 Reid, K. B. M. (1977) Biochem. J. **161**, 247–251

107 Smith, J. J., Muldoon, E. P., Willard, J. J. and Lamport, D. T. A. (1986) Phytochemistry **25**, 1021–1030

108 Chen, J. and Varner, J. E. (1985) EMBO J. **4**, 2145–2151

109 Li, X., Kieliszewski, M. and Lamport, D. T. A. (1990) Plant Physiol. **92**, 327–333

110 Raz, R., Crétin, C., Puigdomènech, P. and Martinez-Izquierdo, J. A. (1991) Plant Mol. Biol. **16**, 365–367

111 Mann, K., Schäfer, W., Thoenes, U., Messerschmidt, A., Mehrabian, Z. and Nalbandyan, R. (1992) FEBS Lett. **314**, 220–223

112 Ertl, H., Mengele, R., Wenzl, S., Engel, J. and Sumper, M. (1989) J. Cell Biol. **109**, 3493–3501

113 Chrispeels, M. J., Sadava, D. and Cho, Y. P. (1974) J. Exp. Bot. **25**, 1157–1166

114 Esquerré-Tugayé, M. T. and Lamport, D. T. A. (1979) Plant Physiol. **64**, 314–319

115 McNeil, M., Darvill, A. G., Fry, S. C. and Albersheim, P. (1984) Annu. Rev. Biochem. **53**, 625–663

116 Williamson, G., Belshaw, N. J. and Williamson, M. P. (1992) Biochem. J. **282**, 423–428

117 Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. and Warren, R. A. J. (1991) Microbiol. Rev. **55**, 303–315

118 Jentoft, N. (1990) Trends Biochem. Sci. **15**, 291–294

119 Page, M. I. and Jencks, W. P. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 1678–1683

120 Williams, D. H., Searle, M. S., Mackay, J. P., Gerhard, U. and Maplestone, R. A. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 1172–1178

121 Cerundolo, V., Elliott, T., Elvin, J., Bastin, J., Rammensee, H.-G. and Townsend, A. (1991) Eur. J. Immunol. **21**, 2069–2075

122 Baron, M., Norman, D. G. and Campbell, I. D. (1991) Trends Biochem. Sci. **16**, 13–17

123 Corden, J. L. (1990) Trends Biochem. Sci. **15**, 383–387

124 Veis, A. and Nawrot, C. F. (1970) J. Am. Chem. Soc. **92**, 3910–3914

125 Lilley, T. H. (1992) J. Chem. Soc. Chem. Commun. 1038–1039

126 Fernandez, J. and Lilley, T. H. (1992) J. Chem. Soc. Faraday Trans. **88**, 2503–2509

127 Laurent, B. C., Treitel, M. A. and Carlson, M. (1990) Mol. Cell Biol. **10**, 5616–5625

128 Feigl, G., Gram, M. and Pongs, O. (1989) Nucleic Acids Res. **17**, 7167–7178

129 Wootton, J. C. and Drummond, M. H. (1989) Protein Eng. **2**, 535–543