Modular protein structures

John Kuriyan

Howard Hughes Medical Institute and Laboratories of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Many protein structures are intrinsically modular in that they are assembled from smaller structural units known as domains. In this article, the structures of several modular proteins are discussed. Comparison of the nucleotide-binding domains of two disulphide reductase enzymes, thioredoxin reductase and glutathione reductase, shows that they have diverged from a common ancestor. However, they appear to have arrived at their respective *active site architectures by independent (convergent)* means. The second example is the DNA clamping subunit (β subunit) of the bacterial DNA polymerase III. This molecule has quite unusual topology, as it dimerizes to form a torus that can encircle DNA. The structure is highly symmetric, with each molecule containing three topologically identical domains that nevertheless share very little sequence similarity. The final example is that of the SH2 domain, an eukaryotic module that binds to peptide sequences containing phosphorylated tyrosines. SH2 domains are found in many proteins that are key elements of cellular signal transduction.

THE complex shapes and intricate assemblies of protein molecules reflect the wide variety of functions performed by these molecular workhorses of the cell. The observed diversity in the amino-acid sequences and three-dimensional structures of proteins is quite bewildering, and extracting general principles from this complexity is a difficult task. Fortunately, some simplification is afforded by the fact that most proteins conform to a structural heirarchy, with larger protein molecules (200 to 1000 amino acid residues or even larger) being formed by the assembly of smaller protein units (50 to 150 amino acid residues)[1]. It is believed that the large assortment of unique protein molecules that are found in nature have evolved from combinations of a much smaller number of modular structural units[2,3].

Modular units that fold into compact three-dimensional shapes are referred to as domains. A protein domain has most of the characteristics of intact proteins: a well-defined hydrophobic core, a characteristic pattern of α-helices and/or β-strands and perhaps an active site with enzymatic or binding activity. In fact, some domains are best considered as independent protein molecules that just happen to be covalently linked to neighbouring molecules. But in many cases, and particularly in enzymes, each domain carries out specific tasks that are coordinated with events occurring at other sites in the intact protein molecule, and the isolated domains may not be able to function or even fold without the assistance of the rest of the protein.

The underlying modularity of protein architecture is perhaps easiest to recognize in the amino-acid sequences of eukaryotic proteins that are involved in controlling the transcription of DNA (transcription factors), or in mediating cellular signalling and association (for example, see refs. 4–6). Comparison of the sequences of these proteins reveals patterns (motifs) that are repeated in various combinations in other proteins. In many cases the sequence similarity is high enough to clearly define domain boundaries. These eukaryotic domains have been described as 'mobile' modules[5], as they are found inserted into many different protein molecules, with varying position and frequency.

Despite the current focus on these easily recognizable domains, most of our early understanding of the modularity of protein structures was based on the direct interpretation of three-dimensional structures, rather than the analysis of amino-acid sequences. As protein molecules evolve, their structural form is conserved long after the amino-acid sequences have mutated beyond recognition[7,8]. X-ray crystallographic studies over the last twenty years or so have revealed a level of modularity in protein architecture that was rarely hinted at by sequence analysis or previous biochemical experiments. The relatively obvious sequence signatures of 'mobile' eukaryotic modules are probably recognizable as such only because the proteins they are found in are modern in evolutionary terms (that is, the amino-acid sequences have not had sufficient time to diverge beyond recognition).

In this article I discuss the structures of three protein domains (Figure 1). The examples I have chosen are from work in my laboratory on quite different biological systems, but the emphasis here is on the underlying modular architecture rather than on the equally fascinating biochemistry represented by these proteins. The first, a nucleotide-binding domain, contains structural elements that were first identified by Michael Rossman and co-workers[9] in the dehydrogenase enzyme family[9]. Nucleotide-binding domains have now been found in numerous proteins, and I show how two disulphide reductase enzymes have probably evolved from a common ancestor containing two

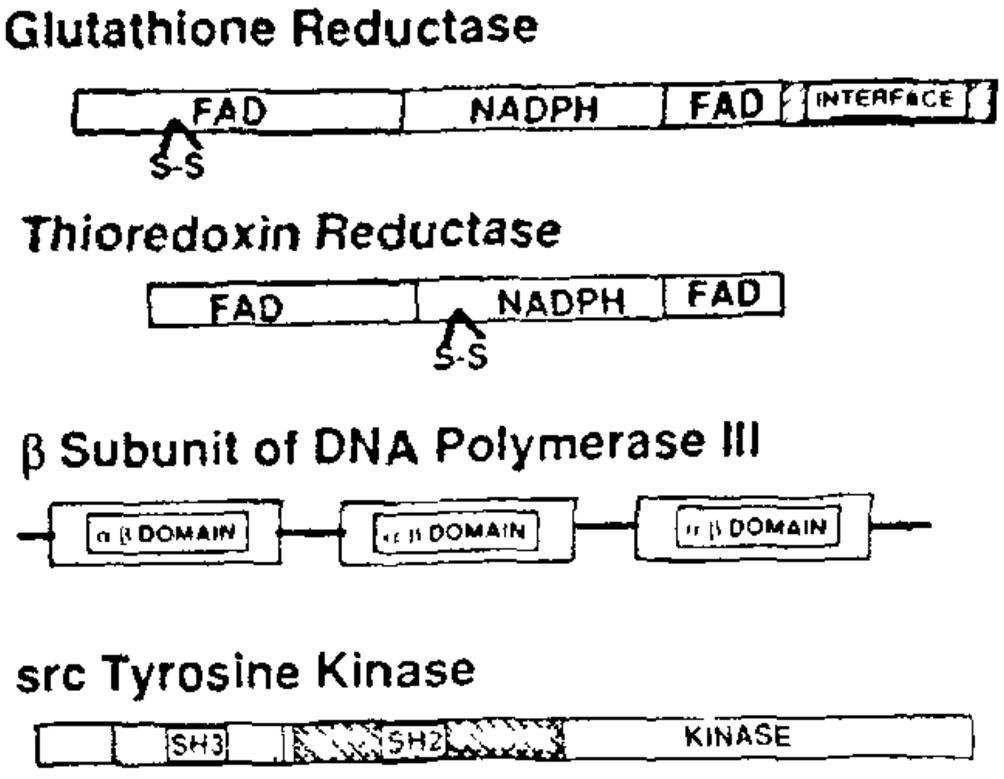

Figure 1. Schematic diagrams of the domain boundaries of the proteins discussed in the article. In these schematics, each protein molecule is shown as a long rectangle, which represents the linear amino-acid sequence. Each molecule is further subdivided into smaller rectangles, which are shaded variously to indicate the domain boundaries. The lengths of the rectangles are not drawn exactly to scale. The S-S symbols for the two reductases refer to the redox active disulphide bonds, which are present in different domains in the two molecules.

different, but related, nucleotide binding folds[10,11]. The second domain is a component of one of the subunits of the enzyme DNA polymerase III, which replicates the chromosomal DNA of bacteria during cell division[12]. Six of these domains together constitute the DNA clamping mechanism of this enzyme, and they provide a striking example of the evolution of complex protein structure from simpler component parts, and the conservation of three-dimensional structure despite the loss of amino-acid similarity.

The third module that I discuss, the SH2 domain, was discovered and defined by comparison of the primary sequences of a number of large proteins involved in cellular signal transduction[13]. Biochemical studies then elucidated the function of the isolated domain and demonstrated that it is, indeed, an independently functioning unit. The SH2 domain plays a central role in transmitting the effects of signals delivered by many polypeptide hormones, by binding to peptides that are phosphorylated on tyrosine. Proteins containing SH2 domains are implicated in the molecular origins of several forms of cancer, and there is great interest in understanding the effects of specifically blocking the actions of SH2 domains.

Nucleotide-binding domains in disulphide reductases

Disulphide bonds are bridges between cysteine residues and they often serve to provide additional stability by cross-linking the protein chain. Some proteins have disulphide bonds that play another important role: they can be reduced (to the dithiol form) and re-oxidized (back to the disulphide form) rather readily, and they serve as important redox agents in cellular biochemistry. The flavin-containing disulphide reductases are enzymes that utilize a redox active disulphide bond to mediate the transfer of electrons from nicotinamide adenine dinucleotide phosphate (NADPH) to various substrates, via an enzyme-bound flavin (flavin adenine dinucleotide, FAD)[10]. The primary electron donor, the nicotinamide ring of NADPH, can only participate in two-electron reductions (via hydride ion transfer) whereas the disulphide bond is reduced in two steps of one electron each. The flavin ring system is required to act as a broker between the otherwise incompatible reactive partners. In some members of the family the reaction runs in the opposite direction under physiological conditions, with the net oxidation of NAD^+ (ref. 10).

Schulz and his colleagues[14,15] have pioneered the structural analysis of this family of enzymes, by determining the structure of glutathione reductase (which maintains the redox balance of the cell by catalysing the reduction of glutathione disulphide, the oxidized form of the tripeptide glutathione). The modular architecture of this enzyme was immediately apparent. There are two Rossman-type nucleotide-binding domains that bind FAD and the NADPH, respectively. These domains are strikingly similar to one another in three-dimensional structure, and Schulz[16] has suggested that they probably arose by gene duplication followed by functional specialization. The enzyme is dimeric, and two sets of these binding domains are brought into juxtaposition by a third domain, known as the interface domain. Together, the three domains form the elements of the enzymatic machinery (Figure 2 *a*). The binding site for NADPH places the nicotinamide ring in close contact with the flavin ring system, allowing the stereospecific transfer of a hydride ion. Immediately adjacent on the other side of the flavin ring system is the enzyme's own disulphide bond, which is reduced by the flavin. This redox active disulphide bond is at the base of a deep crevice in the structure formed between the interface and FAD-binding domains, in which glutathione binds. The final step in the reaction, the disulphide/dithiol interchange, occurs at this site.

The overall picture that emerges from the structural studies on glutathione reductase is that of a large enzyme with well-defined binding sites for the various players in the reaction (Figure 2 *a*)[17]. The enzyme remains relatively static during the reaction (that is, it does not undergo any large-scale conformational changes), and provides a passive template within which the reactive groups transfer electrons to each other. Each of the domains plays distinct and important roles. The FAD and NADPH domains bind the requisite

Figure 2 *a*. For caption, see page 88.

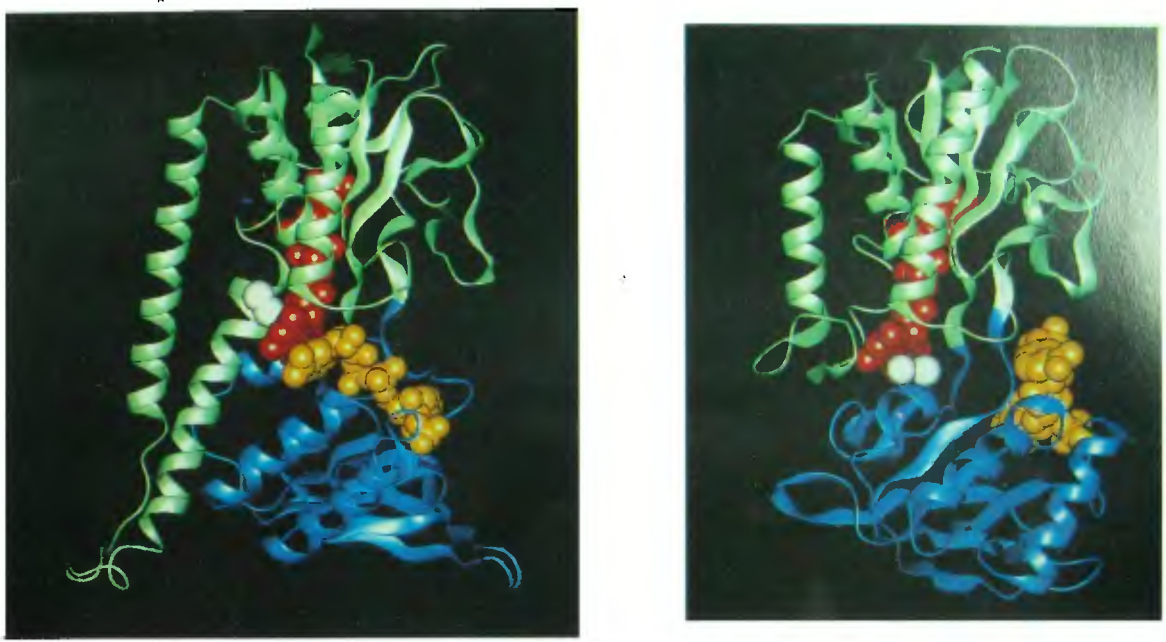

Figure 2. Comparison of the structures of glutathione reductase and thioredoxin reductase. ***a***, The active site of glutathione reductase, from the work of Schulz and colleagues[10]. The particular structure shown is from the high resolution refinement of Karplus and Schulz[15,17]. This view represents about half of the dimeric enzyme, which has two symmetrically related active sites. Bonds between atoms shown as lines that are coloured blue for basic residues, red for acidic residues and white for all others. The glutathione-binding site is the V-shaped crevice in the top portion of the molecule, in which glutathione is shown with thick green lines. The redox-active sulphur atoms of glutathione and the enzyme are shown as green spheres. Yellow and purple spheres represent the FAD and NADPH co-enzymes, respectively. Note that the nicotinamide ring of NADPH stacks against the ring system of FAD, and that the enzyme's redox-active disulphide bond is directly above the flavin. ***b***, Comparison of the FAD and NADPH domains of thioredoxin reductase[11] (on the right) and glutathione reductase[17], on the left. The polypeptide backbone of the protein structures is represented as a coloured ribbon, with the FAD and NADPH domains shown in green and blue, respectively. Note the similarities in the pattern of α-helices (shown as spirals) and strands of β-sheet (shown as flattened ribbons) between the two domains within one molecule, and between different molecules. One major difference between the two enzymes is the insertion of a long α-helix in glutathione reductase. This helix, along with the interface domain of glutathione reductase (not shown), serves to pin the NADPH domain of glutathione reductase into this particular orientation and rules out the large conformational change postulated for thioredoxin reductase. The atoms of FAD (red) and NADPH (yellow) are represented as solid spheres with the appropriate van der Waals radii. The sulphur atoms of the redox active disulphide bonds are shown as white spheres. Note that the disulphide bond is above the flavin ring system in glutathione reductase, which has the nicotinamide ring of NADPH stacking below the flavin ring. In thioredoxin reductase, the NADPH domain is rotated by approximately 70° with respect to the FAD domain. The position under the flavin ring is occupied by the disulphide bond, and the nicotinamide ring of NADPH is exposed to solvent. The structures of the NADPH cofactor in thioredoxin reductase is based on unpublished work (G. Waksman and J. Kuriyan).

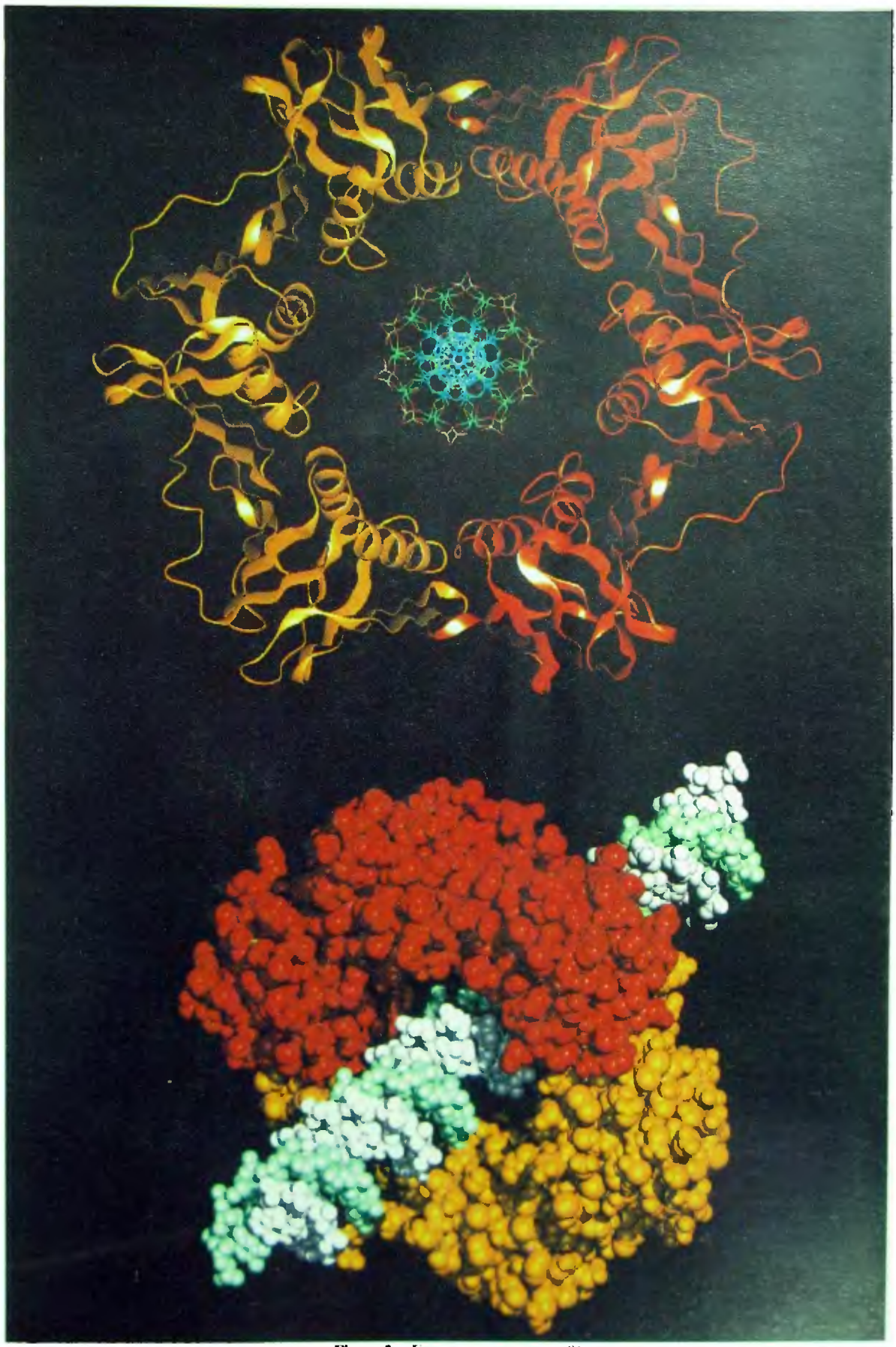

Figure 3. For caption, see page 91

Figure 4. The structure of the SH2 domain of the src tyrosine kinase, complexed with phosphopeptides. **a**, Ribbon diagram showing the β strands in yellow and the α helices in red. The bonds in the phosphopeptide are shown as rods, with all atoms of the peptide coloured white, except the phosphate group shown in green. The three positively charged sidechains interacting with the phosphotyrosine are shown in blue. **b**, Surface representation of the structure, in a similar orientation. The green dots represent the solvent accessible surface[39] of the SH2 domain, with the bonds in the structure represented by red lines. The three basic sidechains that coordinate the phosphotyrosine are shown in blue. The backbone atoms of the phosphopeptide are shown in yellow, with the sidechain atoms in white. The phosphate group is in red. The three purple spheres mark a groove on the surface that might bind longer peptides.

nucleotides. In addition, the FAD domain incorporates the enzyme's redox active disulphide bond, which is positioned near the flavin ring system, and the interface domain forms one wall of the glutathione-binding site and provides key catalytic groups. This enzyme architecture is seen to be essentially unchanged in several other members of this family that act on substrates different from glutathione: mercuric ion reductase[18], lipoamide dehydrogenase[19] and trypanothione reductase[20].

In light of the fact that the prototypic glutathione reductase mechanism involves such exquisite spatial relationships between the nucleotides, the enzyme's redox active disulphide and the substrates, the gene sequence of thioredoxin reductase came as a great surprise[21]. This enzyme is a well-known member of the family, and catalyses a reaction that is identical to that of glutathione reductase except that the substrate is a disulphide-containing protein, thioredoxin, instead of the small peptide glutathione[22]. The gene sequence revealed two unexpected features (Figure 1). One is that the interface domain of glutathione reductase is missing, which means that the dimeric enzyme and one wall of the active site cannot form as in glutathione reductase. The other is that the enzyme's redox active disulphide is present in the NADPH domain rather than the FAD domain. This is as if one of the players in a game of molecular tennis (with the flavin ring system as the net) has jumped over to the side of the other player.

The X-ray structure of thioredoxin reductase, determined in my laboratory, reveals that the molecule contains NADPH- and FAD-binding domains that are very similar to the domains of glutathione reductase (Figure 2*b*)[11]. However, this enzyme forms a dimer differently, utilizing the FAD domain instead of the interface domain, which is missing. The NADPH domain is rotated by about 70° relative to its orientation in glutathione reductase. This results in the enzyme's redox-active disulphide (located in this case in the NADPH rather than the FAD domain) to take up a position near the flavin ring system that corresponds to the location of the nicotinamide ring in glutathione reductase. Consequently, this conformation of the enzyme prevents the NADPH domain from binding in a productive mode and the nicotinamide ring is rotated away from the flavin and exposed to solvent (Figure 2*b*).

This structure of thioredoxin reductase presents two puzzles. Although the disulphide bond is located close to the flavin ring system, consistent with biochemical experiments, the protein substrate thioredoxin is too bulky to fit into the space near the disulphide. It is known that the nicotinamide ring of NADPH transfers a hydride ion to the flavin with the same stereospecificity as seen in glutathione reductase[23]. That is, at some point in the enzyme reaction, the NADPH group is likely to take up a conformation similar to that seen in glutathione reductase and excluded in this structure of thioredoxin reductase by the orientation of the NADPH domain. A simple hypothesis that provides a solution to both problems is that thioredoxin reductase, unlike glutathione reductase, undergoes a large conformational change during enzyme catalysis, involving a rotation of the NADPH domain relative to the rest of the structure. Such a large rotation is possible because in this enzyme the NADPH domain hangs relatively unimpeded from the bulk of the structure, a critical difference brought about by the absence of the interface domain of glutathione reductase. Thus, the enzyme may cycle between two states, one in which the disulphide bond is inacessible to thioredoxin but which is optimal for its reduction by the flavin. In the other state the disulphide bond is exposed and available for reaction with thioredoxin, and the NADPH binds productively so as to reduce the flavin. Further biochemical and structural studies are required to verify this proposed enzyme mechanism.

The comparison between glutathione reductase and thioredoxin reductase illustrates the process by which varied enzyme structures can arise from the functional adaptation of structural domains. Sufficient sequence similarity exists between the two enzymes that we can infer with confidence that they have diverged from a common ancestor. This ancestor is likely to have had two nucleotide-binding domains, since both enzymes preserve a curious feature: the NADPH domain is inserted into the sequence of the FAD domain (Figure 1). Although both domains catalyse the same reaction, that of pyridine nucleotide-linked disulphide reduction, they do so within entirely different structural templates. The enzyme reaction mechanisms therefore appear to have been arrived at independently, by a process of convergent evolution[11].

Figure 3. Two representations of the structure of the β subunit of DNA polymerase III. The X-ray crystallographic analysis was carried out on the isolated protein (without DNA). However, the dimensions of the molecular structure, as well as other properties, suggest that duplex DNA fits into the centre of the hole in the structure[12]. In these figures standard B-form Watson–Crick DNA is drawn in the geometric centre of the protein molecule. ***a***, The polypeptide backbone of the protein dimer (shown as ribbons), and all atoms of DNA (shown as a stick figure in the centre of the image). One molecule of the β subunit is coloured red, and the other is coloured yellow. The atoms of DNA are coloured according to their chemical type, oxygens are red, carbons green, nitrogens blue and phosphorus yellow. ***b***, Space-filling model of the β subunit and DNA. All atoms of DNA and the protein are represented as spheres with appropriate van der Waals radii. The two strands of DNA are coloured white and green, respectively.

The sliding clamp of DNA polymerase III

DNA polymerases are enzymes that copy the information stored in DNA, by catalysing the polymerization of nucleotides that are complementary to the DNA strand being replicated[24]. Several different kinds of DNA polymerase enzymes are known. Some of these are involved in replicating relatively short DNA sequences, as in the cellular processes that repair damage to the genome. The structure of one such polymerase, polymerase I of *Escherichia coli*, is known in atomic detail from the work of Thomas A. Steitz's laboratory[25]. Other DNA polymerases, such as polymerase III holoenzyme (Pol III) of *E. coli* are involved primarily in replicating the entire genome during cell division. These polymerases are much more complex assemblies, consisting of some 10 or more protein molecules that join together to carry out DNA replication[24]. One important property of these polymerases is that they are highly *processive*: once assembled onto DNA, the polymerase can replicate very long stretches of the genome without falling off the DNA.

In the case of Pol III, biochemical experiments have shown that the property of holding on tightly to DNA is conferred upon the polymerase by just one of its component protein molecules, the β subunit[26]. This protein does not normally associate with DNA. However, five other subunits of Pol III (known collectively as the γ complex) act to clamp the β subunit onto DNA in a process that expends ATP energy[27]. Once assembled upon DNA the β subunit can slide freely along duplex DNA, and can bind tightly to the subunits that carry out the actual chemical reactions of nucleotide polymerization and error checking (the α and ε subunits), and prevent them from falling off the DNA[26]. How does the β subunit manage to accomplish the task of clamping tightly to the DNA, and at the same time be free to slide along and allow replication to proceed rapidly?

The three-dimensional structure of the β subunit has been determined in my laboratory, by X-ray crystallography[12]. Although the structure has been determined in the absence of DNA, the mode of DNA binding can be immediately inferred. The X-ray structure reveals a molecule with unprecedented topology (Figure 3). Two molecules of the protein are tightly associated in a dimeric interaction to form a closed circle which can encircle duplex DNA. The clamping mechanism of the protein is now obvious: the protein is bound to DNA in a topological rather than a chemical sense, as originally proposed by Mike O'Donnell[26] on the basis of biochemical experiments. The hole in the middle of the donut-shaped structure is big enough (approximately 35 Å in diameter) for duplex DNA to slide through with no steric hindrance. The β subunit thus acts as a molecular 'curtain ring' through which the DNA slides like a curtain rod.

Although we do not as yet know the three-dimensional structures of the rest of the Pol III assembly, a general mechanism of how the sliding clamp works can be proposed on the basis of this structure and the extensive biochemical experiments on this system. The β subunit is a closed circular dimer in solution, and it has to be disassembled in someway before being snapped onto DNA. This process is carried out by the γ complex, which also serves to bind DNA at an initiation site for replication[27]. The γ-complex brings the β subunit to the vicinity of DNA, and then somehow opens the ring and re-closes it around DNA. The ATP energy that is known to be required[27] probably goes towards breaking the strong interactions at the dimer interface of the β subunit. Once assembled onto DNA the β subunit cannot easily fall off, and it serves to tether the rest of the polymerase machinery during replication.

A surprising feature of the structure is its high symmetry and modularity. Despite no detectable internal symmetry in the amino-acid sequence, the entire DNA-encircling duplex is constructed by repeating a simple β-α-β-β-β pattern 12 times around a circle (Figure 4)[12]. There are several levels of symmetry in the structure. First, each molecule is constructed of three easily recognizable domains that are topologically identical. Second, the topology of each domain is internally symmetric and contains 2 of the β-α-β-β-β motifs. Third, rotation of the dimeric molecule about a central axis reproduces the structure every 60°, that is the structure possesses the approximate symmetry of a circle.

It should be recognized that the symmetry noted here is not exact. Variations in the detailed structure come about because the amino-acid sequences of the individual domains have diverged considerably, with amino acid identities ranging from about 10% for two of the pairwise comparisons between the three domains, and 16% for the third[12]. This level of sequence identity is well below the threshold of about 20–25% identity that is required for structural similarity to be strongly predicated[8]. The β subunit structure presumably arose by gene duplication events starting from an ancestral module. The level of sequence conservation that is required to preserve a particular structural topology is clearly minimal, as known from studies of structural divergence in evolutionarily related proteins[7]. It may also be noted that some of the sequence variation is functional. The polymerase subunit (α) and the clamp-assembling proteins (γ complex) are unlikely to be symmetric, and the β subunit has evolved to make specific interactions with these proteins. Whatever the cause of the sequence drift, the fact that this level of topological symmetry is undetectable at the level of

sequence is a striking example of the structural degeneracy of protein sequences, and highlights one of the principal reasons why protein structure prediction is so difficult.

The SH2 phosphotyrosine recognition domain

The previous examples of modular protein structures presented domains that are intricately linked in the intact parent protein to form functional superstructures. In contrast, the SH2 domain is an example of a eukaryotic 'mobile module' that can exist in a functional state when isolated from the parent proteins. In fact, little is known about the three-dimensional forms of the intact proteins, such as the tyrosine kinases, that contain these modules. The SH2 domain takes its name from the Rous sarcoma virus gene *src*, which encodes a tyrosine kinase enzyme that phosphorylates tyrosine residues in certain target proteins[28,29]. The viral form of *src* is an oncogene, i.e. the protein encoded by the gene acts to transform normal cells to an abnormal growth state. In addition to a catalytic domain with kinase activity, the src protein also contains modular domains known as SH2 (src homology region 2) and SH3 (src homology region 3)[13]. These domains were identified from regions of sequence similarity between many different kinds of signal transduction molecules and are not known to have any enzymatic activity. Rather, they are domains that recognize and bind to specific structures in other proteins. The SH2 domain, for example, binds to sequences in certain proteins that contain phosphorylated tyrosine residues. The smaller SH3 domain is less well understood, and is thought to bind to signal transduction proteins involved in the formation of the cytoskeleton.

The phosphorylation (by kinases) and dephosphorylation (by phosphatases) of tyrosine residues is a central element in the signal transduction pathways that link growth hormone receptors to the modulation of DNA transcription in the nucleus[30]. When growth hormones bind to their receptors on the surface of the cell, they cause the receptor proteins to assemble into dimeric states. These receptors contain cytoplasmic domains with tyrosine kinase activity, and upon dimerization the kinase activity of the receptors is activated, resulting in autophosphorylation. Receptor tyrosine phosphorylation triggers the association of a variety of different cytoplasmic proteins with the receptors, many of which are known to encode enzymatic activities important in signal transduction. These include non-receptor tyrosine kinases, various phospholipases as well as GTPase activating proteins, all of which contain SH2 domains[30]. Recently it has become clear that the principal mechanism of this association with phosphorylated receptors is based on the recognition of specific phosphorylated tyrosine residues by the SH2 domains (for a recent review see ref. 31).

How do SH2 domains recognize phosphotyrosine residues? Although the large number of SH2 domains known are all clearly homologous (after all, that is how they were identified), they do not share sequence similarity with any protein of known structure. The mechanism of molecular recognition by SH2 domains has been revealed recently by the X-ray structure of the src SH2 domain, determined in my laboratory in complex with peptides containing phosphorylated tyrosine residues (Figure 4)[32]. At the same time, structures became available for the uncomplexed forms of the SH2 domains of two other proteins, the abl tyrosine kinase[33] and the p85 subunit of phosphatidyl inositol 3′-OH kinase[34], both determined in solution by NMR methods.

All the three structures show that the SH2 domain is shaped like a socket, with the N- and C-terminii of the module close together in space and away from the peptide-binding face. The domain is thus ideally modular, in the sense that it can be extracted from one protein and inserted somewhere else without affecting its ability to function properly. The architecture of the protein is particularly simple: a central anti-parallel β sheet is flanked by two α-helices. There is no well-defined groove or crevice into which the peptide binds. Rather, there is a relatively flat surface which runs perpendicular to the plane of the central β sheet and that binds phospho-peptides in extended conformations.

The phosphotyrosine sidechain is clamped between the tip of an α-helix and a loop connecting two strands of a β sheet (Figure 4). The mechanism of phosphotyrosine recognition has some unexpected elements. The two characteristic features of the phosphotyrosine sidechain are the aromatic ring system of the tyrosine, and the negatively charged phosphate group. One naturally expects positively charged residues such as arginine to form strong hydrogen bonding interactions with the phosphate, and mutagenesis experiments show that this is indeed the case[35]. The X-ray structure reveals a buried arginine forming a bidentate interaction with the phosphate group. With regard to the aromatic ring system, a naive expectation was that aromatic or hydrophobic sidechains would pack against the ring. Thus it was surprising to find that the two other positively charged groups at the active site, a lysine and an arginine, interact primarily with the ring system rather than the phosphate group. There are no non-polar or aromatic groups near the ring system. Instead, the lysine makes hydrophobic and amino-aromatic interactions with the ring. The arginine forms a hydrogen bond to a phosphate oxygen with its imino

hydrogen, and forms amino–aromatic with the ring system with one of the terminal nitrogens.

The amino–aromatic interactions observed here derive their stability from the fact that the tyrosine ring is polarized, with net negative charge on the carbons and net positive charge on the hydrogens. Hydrogen bond donors can thus interact favourably with the carbon atoms of the ring, with optimal positions being above and below the ring, as found in previous analyses of such interactions[36,37]. The stabilization of phosphotyrosine by amino–aromatic interactions makes logical sense: the basic residues at the active site would serve to set up a positive electric field that would attract the negatively charged phosphotyrosine. Furthermore, the net positive charge at the active site would inhibit tyrosine from binding, and the specific placement of the positive charges is incompatible with stabilization of phosphoserine or phosphothreonine, thus leading to specificity.

Although these SH2 structures have illuminated the mechanism by which phosphotyrosyl peptides are bound by the SH2 domain, they are silent about the nature of the multi-domain architecture of the much larger signal transduction molecules of which they are a part. These intact molecules have so far proven intractable to crystallographic analysis, but because of their importance they remain at the centre of intense efforts aimed at obtaining useable crystals.

Summary

The contemporary field of structural molecular biology has many similarities to descriptive botany and zoology in the latter half of the nineteenth century. Intense exploratory activity is revealing novel and strangely beautiful molecular structures every week, of which this article has described just three or four. As in pre-Darwinian botany and zoology, a central problem regarding the origin of these structures remains unsolved. In the case of structural biology this is the protein folding problem, which concerns the relationship between the linear amino-acid sequence that is encoded in the gene for a particular protein, and the specific three-dimensional form that the resulting amino-acid polymer takes up. We are currently unable to derive from the amino-acid sequence alone what the folded structure of a protein looks like[38]. However, increasingly powerful computational methods are being developed that allow the three-dimensional structure of a protein to be inferred if the structure of a protein with related amino-acid sequence is known. At the same time, the sustained efforts of crystallographers and NMR spectroscopists have resulted in the structures of several hundred different protein molecules being known today, with the rate of structure determination steadily accelerating. The accuracy with which protein structures can be predicted is therefore expected to improve considerably in the future. While we eagerly await the reduction of the protein folding problem to a computer algorithm capable of generating accurate three-dimensional structures from amino-acid sequences, we also recognize that the discovery of such an algorithm will take much of the fun away from experimental structure determination. After all, what is thought to be known is not eagerly pursued.

1. Branden, C. and Tooze, J., *Introduction to Protein Structure*, Garland, New York, 1991, pp. 1–302.
2. Chothia, C., *Nature*, 1992, **357**, 543–544.
3. Dorit, R. L., Schoenbach, L. and Gilbert, W., *Science*, 1990, **250**, 1377–1382.
4. Baron, M., Norman, D G. and Campbell, I. D., *Trends Biol. Sci.*, 1991, **16**, 13–17.
5. Bork, P., *Curr. Opinion Struct. Biol.*, 1992, **2**, 413–421.
6. Frankel, A D. and Kim, P. S., *Cell*, 1991, **65**, 717–719.
7. Chothia, C. and Lesk, A. M., *EMBO J.*, 1986, **5**, 823–826.
8. Sander, C. and Schneider, R., *Proteins*, 1991, **9**, 56–68.
9. Rossmann, M. G., Liljas, A , Brändén, C. I. and Benaszak, L. J., in *The Enzymes* (ed. Boyer, P. D), Academic Press, New York, 1975, pp. 61–102.
10. Schirmer, R. H. and Schulz, G. E., in *Pyridine Nucleotide Coenzymes Part B: Coenzymes and Cofactors; vol. 2* (eds. Dolphin, D., Poulson, R. and Avramovic, O.), John Wiley & Sons, New York, 1987, pp. 333–379.
11. Kuriyan, J , Krishna, T. S. R., Wong, L., Guenther, B., Pahler, A., Williams, C. H. J. and Model, P., *Nature*, 1991, **352**, 172–174.
12. Kong, X.-P., Onrust, R., O'Donnell, M. and Kuriyan, J., *Cell*, 1992, **69**, 425–437.
13. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. and Pawson, T., *Science*, 1991, **252**, 668–674.
14. Thieme, R., Pai, E. F., Schirmer, R. H. and Schulz, G. E., *J. Mol. Biol.*, 1981, **152**, 763–782.
15. Karplus, P. A. and Schulz, G. E., *J. Mol. Biol.*, 1987, **195**, 701–729.
16. Schulz, G. E., *J. Mol. Biol.*, 1980, **138**, 335–347.
17. Karplus, P. A and Schulz, G. E., *J. Mol. Biol.*, 1989, **210**, 163–180
18. Schiering, N , Kabsch, W., Moore, M. J., Distefano, M. D., Walsh, C. T. and Pai, E. F., *Nature*, 1991, **352**, 168–172.
19. Mattevi, A., Schierbeek, A. J. and Hol, W. G. J., *J. Mol. Biol.*, 1991, **220**, 975–994
20. Kuriyan, J , Kong, X-P., Krishna, T. S. R., Sweet, R. M., Murgolo, N J., Field, H., Cerami, A. and Henderson, G. B., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 8764–8768.
21. Russel, M. and Model, P., *J. Biol. Chem.*, 1988, **263**, 9015–9019.
22. Williams, C. H., Jr , *The Enzymes*, 3rd edn., 1976, **13**, pp. 89–173.
23. You, K., Arnold, L. J., Allison, W. S. and Kaplan, N. O., *Trends Biol. Sci* , 1978, **3**, 265–268.
24 Kornberg, A. and Baker, T. A., *DNA Replication*, W. H. Freeman, New York, 1991, pp. 1–931.
25. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. and Steitz, T. A., *Nature*, 1985, **313**, 762–766.
26. Stukenberg, P. T., Studwell-Vaughan, P. S. and O'Donnell, M., *J. Biol. Chem.*, 1991, **266**, 11328–11334.
27. Onrust, R., Stukenberg, P T. and O'Donnell, M , *J. Biol. Chem.*, 1991, **266**, 21681–21686.
28 Jove, R. and Hanafusa, H., *Annu. Rev. Cell Biol.*, 1987, **3**, 31–56.
29. Hanks, S. K., Quinn, A. M. and Hunter, T., *Science*, 1988, **241**, 42–52.

30. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R. and Soltoff, S., *Cell*, 1991, **64**, 281–302.
31. Pawson, T., *Curr. Opinion Struct. Biol.*, 1992, **2**, 432–437.
32. Waksman, G., Kominos, D., Robertson, S. R., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L. and Kuriyan, J., *Nature*, 1992, **358**, 646–653.
33. Overduin, M., Rios, C. B., Mayer, B. J., Baltimore, D. and Cowburn, D., *Cell*, 1992, **70**, 697–704.
34. Booker, G. W., Breeze, A. L., Downing, A. K., Panayotou, G., Gout, I., Waterfield, M. D. and Campbell, I. D., *Nature*, 1992, **358**, 684–687.
35. Mayer, B. J., Jackson, P. K., Van Etten, R. A. and Baltimore, D., *Mol. Cell. Biol.*, 1992, **12**, 609–618.
36. Levitt, M. and Perutz, M. F., *J. Mol. Biol.*, 1988, **201**, 751.
37. Burley, S. K. and Petsko, G. A., *Adv. Prot. Chem.*, 1988, **39**, 125–189.
38. Levitt, M., *Curr. Opinion Struct. Biol.*, 1992, **1**, 224–229.
39. Lee, B. K. and Richards, F. M., *J. Mol. Biol.*, 1971, **55**, 379–400.

ACKNOWLEDGEMENTS. I thank the members of my laboratory who have participated in the determination of the structures described here: Brian Guenther, Dorothea Kominos, Xiang-Peng Kong, Talleru S. R. Krishna, Jennifer L. Martin, Scott Robertson, Gabriel Waksman and Lim Wong. I also thank my colleagues and collaborators, David Baltimore, Stephen K. Burley, David Cowburn, Hidesaburo Hanafusa, Peter Model, Mike O'Donnell, Rene Onrust, Greg Petsko, Marilyn Resh, Robert M. Sweet and Charles Williams, Jr. This work was supported in part by the National Institutes of Health and the Pew Memorial Trust.

Received 21 December 1992; accepted 22 December 1992.

Recent developments in cell cycle regulation

G. Shanmugam

Cancer Biology Division, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India

Cell cycle of normal cells consists of precisely regulated series of events. Check points in each phase of the cycle enable the cells to complete specific functions before they transit to the next phase. Earlier studies emphasized the G0 to G1 transition as a crucial stage in cell proliferation; however, recent findings indicate decision-making points also at G1 and G2/M phases. Growth factors, oncogenes and cyclins are the major players regulating the cell cycle; they regulate the cycle by influencing phosphorylation and transcription processes. Entry into and exit from the cell cycle are determined by several processes such as the synthesis of new proteins, phosphorylation/dephosphorylation of pre-existing proteins and degradation of proteins like cyclins. Deranged regulation of cell cycle is the essence of neoplastic development.

A new-born cell has several options for its future depending on the program it has and on the signals it receives from the environment. When the signals are appropriate, the cell enters into the cycling process called cell cycle; if conditions are not favourable, the cell enters into a state of latency called quiescence. When the cell comes across differentiation factors, it starts to differentiate. Most normal cells have definite life span which is roughly equivalent to their division number *in vivo*; however, some of these cells overcome the rule of definite life span and get immortalized acquiring the capacity to divide infinitely. Programmed cell death (apoptosis) is another option in the life of cells. Since the cell cycle is a remarkably regulated phenomenon of normal cells and this type of regulation is often lost in the proliferation of cancer cells, the salient features of normal cell cycle and the various aspects of regulation of this cycle will be covered in this article. The objective of this review is to expose various aspects of recent developments in cell cycle research to an interdisciplinary readership. References of reviews on specific aspects of cell cycle are presented at appropriate places for those who seek detailed information on one or more aspects of cell cycle. Unless otherwise stated, the studies discussed in this review were from cultured mouse embryo fibroblasts or mouse fibroblastic cell lines (Balb/c 3T3).

The cell cycle of a growing cell is the period between the formation of the cell by the division of its mother cell and the time when the cell itself divides to form two daughters[1]. It is a fundamental unit of time at the cellular level since it defines the life cycle of a cell. A plethora of investigations into the temporal organization of cell division in a variety of cell types revealed that control of cell proliferation is principally determined in G1 phase of the cell cycle[2], and mitosis (M), cytokinesis and DNA synthesis (S) are other landmarks[3,4]. The temporal gaps separating cell division (mitosis) from S phase are designated as G1 from one side and G2 from the other side of the cycle. Modifications to this general concept were introduced to accommodate the G0—a state in which cells may exist for long periods of time in a quiescent stage[5,6].

Recent studies, however, show the presence of several stages which can be termed as sub-phases or check points[7], suggesting that the classical designation of four phases (G1, S, G2, M) may only serve as an organizing