

The origin of protein interactions and allostery in colocalization

John Kuriyan^{1,2} & David Eisenberg³

Two fundamental principles can account for how regulated networks of interacting proteins originated in cells. These are the law of mass action, which holds that the binding of one molecule to another increases with concentration, and the fact that the colocalization of molecules vastly increases their local concentrations. It follows that colocalization can amplify the effect on one protein of random mutations in another protein and can therefore, through natural selection, lead to interactions between proteins and to a startling variety of complex allosteric controls. It also follows that allostery is common and that homologous proteins can have different allosteric mechanisms. Thus, the regulated protein networks of organisms seem to be the inevitable consequence of natural selection operating under physical laws.

One of the principal findings of molecular and cellular biology is that the metabolism and the homeostasis of cells are based on networks of interacting proteins¹. Every protein interacts with other proteins: in some cases, by binding directly to another protein; in other cases, by modifying another protein; and in still other cases, by acting on a substrate and converting it into a substrate for the next protein in a pathway. Proteins that have crucial cellular roles are usually 'switchable', with their activities being modulated by other molecules. When such switching in protein activity is regulated by communication between two sites in a protein — the active site and the site of modification or binding — we refer to this as allostery. Allostery was defined originally as the regulation of a protein by a small molecule that differs in shape from the substrate, and this definition was later modified to the regulation of a protein through a change in its quaternary structure induced by a small molecule². Our definition is broader than these and refers to a structural change — in the tertiary structure, the quaternary structure or both — induced by a small molecule or another protein. More generally, by our definition, the change induced by the modulator could be a change in the flexibility of the protein rather than simply a change in the structure³. In this broader sense, allostery accounts for the responsiveness of cells to external signals and for the regulation of metabolic pathways.

When confronting the intricacy of cellular networks and their exquisitely sensitive controls, scientists often wonder how such highly complex and regulated networks evolved. A few scientists go so far as to hold that "irreducibly complex" systems constitute a "powerful challenge to Darwinian evolution"⁴. The argument is that for a system that is "composed of several well-matched, interacting parts that contribute to the basic function, wherein the removal of any one of the parts causes the system to effectively cease functioning", these parts could not have evolved independently⁴. Protein networks with allosteric regulation are examples of such complex systems. Our view on the evolution of protein interactions and allostery is that natural processes of protein colocalization in cells, which effectively increase the local concentration

of neighbouring molecules, change what might have seemed to be improbable evolutionary events into probable ones. This view complements ideas found in many earlier articles on this topic^{2,5-8}.

Fundamental 'forces', such as compartmentalization and electrostatic or hydrophobic binding, target proteins to specific locations in the cell, where they are colocalized with other proteins. This natural process of colocalization is essential in metabolism, transcriptional control, and signalling⁹. We argue that colocalization, combined with other natural processes (such as genetic recombination), leads naturally to protein complexes, to networks of interacting proteins and, subsequently, to allosteric control. Every protein complex or allosteric system that develops in this way might seem 'irreducibly complex', but these assemblies form as a result of the accidental mutations that first led to the interactions or fixed the relative disposition of the interacting domains. As a consequence, homologous proteins often have different allosteric mechanisms. Thus, although allostery is expected to arise naturally and readily in molecular 'machines', the precise mechanism is usually specific to one molecule and its closest relatives, and is not present across a protein family. This review explains how the regulated complexes and pathways of cells might have emerged, step by step, through natural selection working on proteins that have been colocalized by natural processes. First, we discuss how fundamental thermodynamic principles led to the idea that protein interactions and allostery emerge in a random manner as a consequence of colocalization. Then, we illustrate this principle with specific examples of diversity in the allosteric control mechanisms that govern homologous proteins.

The effect of colocalization

Some of the ways in which two proteins can be brought together in a cell are illustrated in Fig. 1. One way is the fusion of the genes that encode the two proteins so that the gene product now consists of two domains linked by a short segment of polypeptide chain. Such covalent linkage boosts the effective concentration of the protein domains with respect to each other to values in the range of 0.05–3.6 mM¹⁰⁻¹³, greatly exceeding the usual

¹Howard Hughes Medical Institute, California Institute for Quantitative Biosciences, Department of Molecular and Cell Biology and Department of Chemistry, University of California, Berkeley, California 94720, USA. ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Howard Hughes Medical Institute, University of California at Los Angeles—Department of Energy Institute of Genomics and Proteomics, Institute of Molecular Biology, Department of Biological Chemistry, and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.

concentrations of proteins in cells, which tend to be in the nanomolar-to-micromolar range^{14,15}. For example, a single molecule in a cell of the bacterium *Escherichia coli* has a concentration of ~ 1 nM.

The increased concentration that results from colocalization can have profound consequences for protein–protein interactions. In the highly atypical case of haemoglobin packed into erythrocytes, the concentration is ~ 5 mM¹⁶; this approaches the concentration (~ 12 mM) in the crystals that Max Perutz and co-workers used to determine the structure of haemoglobin¹⁷. In physiological conditions, however, the solubility of haemoglobin in erythrocytes is greater than its concentration, so haemoglobin does not precipitate or crystallize. By contrast, in individuals with the mutation that causes sickle-cell anaemia, in whom the glutamic-acid residue at position 6 of the β -chain of haemoglobin is replaced with a valine residue, the solubility of this mutant haemoglobin is half that of wild-type haemoglobin. The concentration of the mutant protein therefore exceeds its solubility, and it forms fibrils. This fibril formation distorts the erythrocytes, which are then poor hosts for the parasites that cause malaria. This process would not occur so readily without the high concentrations that result from the sequestration of haemoglobin in erythrocytes. Thus, in selecting the glutamic-acid-to-valine mutation, natural selection operates on the high concentration of colocalized haemoglobin molecules in erythrocytes.

Colocalization gives evolutionary processes the opportunity to convert nonspecific binding interactions into interactions that have functional consequences. Good examples of nonspecific binding interactions between protein molecules are the molecular contacts that occur within crystals in regions that are not part of a functional interface. Studies of these nonspecific contacts have revealed that they usually cover a

small area, typically $200\text{--}1,200 \text{ \AA}^2$ (ref. 18), and consist of a few hydrogen bonds and limited hydrophobic interactions. Such nonspecific interactions often occur when protein concentrations approach 1 mM.

The effect of colocalization on binding is large, as illustrated in Fig. 2. The binding curve shows, as a function of the concentration of protein A, the fraction (ν) of another protein, B, that is bound to A. The fraction ν is given approximately by the equation $\nu = ([A]/K_d)/(1 + [A]/K_d) = K_a[A]/(1 + K_a[A])$, where $[A]$ is the concentration of A (strictly speaking, the activity of A), K_d is the dissociation constant for the binding of A to B, and K_a is the association constant for the binding of A to B (which equals $1/K_d$). Note that the binding of A to B is half complete ($\nu = 0.5$) when the concentration of A equals the dissociation constant. At a low concentration, such as $0.1K_d$, there is little binding, whereas at a high concentration, such as $10K_d$, nearly all of the binding sites on B are bound to A. Therefore, as the concentration of A varies from a low value typical of proteins in cells ($\sim 10\text{--}100$ nM) to a value that can be achieved by colocalization within a cell (~ 1 mM), the nonspecific interaction between A and B can increase from negligible to substantial.

In the absence of colocalization, no single mutation is likely to convert a pair of proteins with a nonspecific binding affinity ($K_d > 10$ mM) to a binding pair. The reason is that a single residue change introduces no more than a few hydrogen bonds, each of which reduces the standard free energy of binding (ΔG^0) by ~ 1 kcal per mol¹⁹. A mutation that increases the nonpolar contact area could reduce the standard free energy of binding by no more than ~ 3 kcal per mol (for replacement of glycine with tryptophan)²⁰. A reduction in the free energy of binding of ~ 3 kcal per mol reduces the dissociation constant by ~ 150 -fold to ~ 0.1 mM (as calculated by $\Delta G^0 = RT \times \ln K_d = -RT \times \ln K_a$, where R is the

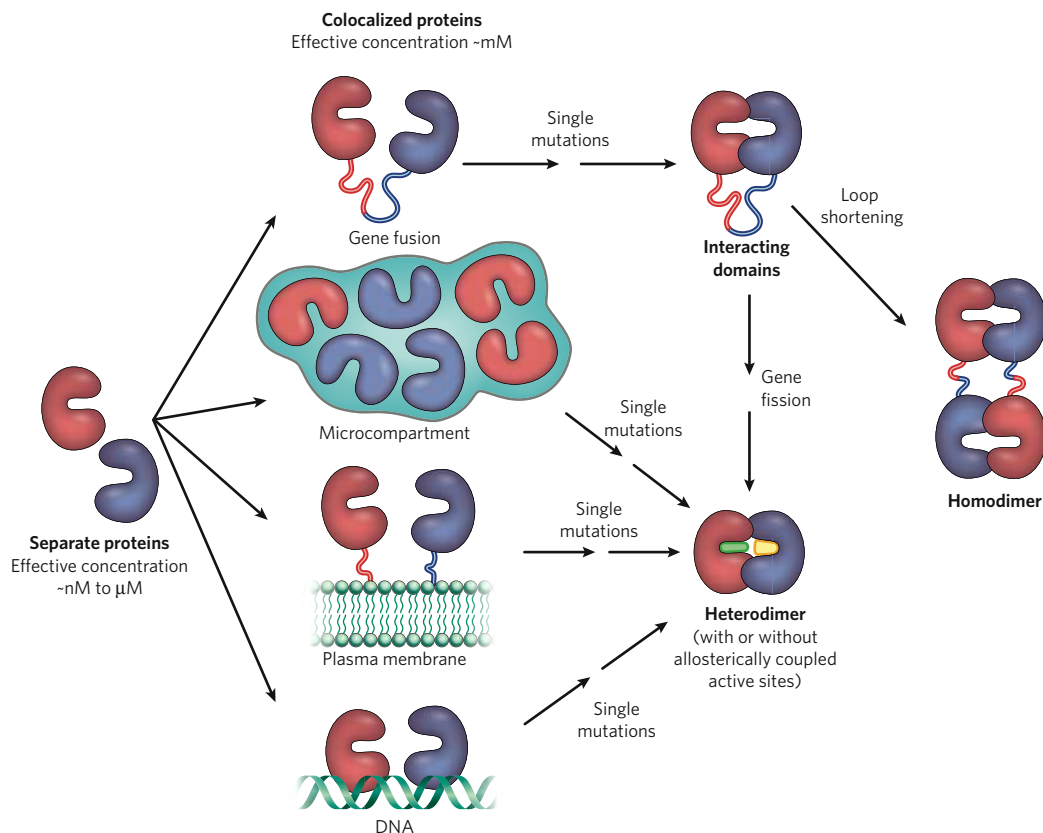


Figure 1 | The evolution of interacting proteins and allostery by single mutations. Two separate proteins in a cell are shown (left). Most cellular proteins are present at nanomolar-to-micromolar concentrations. A single random mutation in either protein is highly unlikely to result in binding or allostery. Interaction between these two proteins becomes probable when they are colocalized. Colocalization (second column) can occur by several mechanisms: by a gene fusion that results in both proteins being part of the same polypeptide chain, by concentration in a microcompartment, by association with the plasma membrane, or

by binding to DNA. This process boosts the effective concentration of the proteins with respect to each other. Now, a single point mutation can lower the dissociation constant enough for a selectable change in function to occur. Further single mutations that increase the affinity of the two domains for each other, or that introduce allostery, can be selected for, resulting in tight interactions between these sites or allosteric coupling. Additional single genetic events such as gene fission or loop shortening can result in a strongly interacting heterodimer or an oligomeric homodimer.

gas constant and T is absolute temperature). Because the concentration of proteins in cells is usually in the nanomolar-to-micromolar range, binding is still negligible in this case, so no new complexes would form as a result of the mutation.

By contrast, when proteins are colocalized, a single mutation can lead to the formation of a new complex. The potential decrease in the dissociation constant caused by the new mutation could bring its value below the effective concentration (~ 1 mM) of the colocalized binding partner. The result would then be substantial binding, and the protein pair would constitute a complex with a new function. If the mutation increases the fitness of the organism, natural selection will fix it in the population. A series of such random mutations can lead, step by step, to a tighter binding site or to an interdependent set of allosteric interactions between the two domains (Fig. 1).

We therefore conclude that when two proteins are tethered to produce high effective concentrations, their colocalization greatly increases the probability that a random mutation in one of the proteins will change their mutual affinity and thus opens up the possibility of a change in fitness. Because the same process of colocalization followed by random mutation and natural selection can operate on assemblies of any number of component proteins, there is no reason to suppose that there is an 'edge' to the power of darwinian evolution beyond which the formation of complex biological structures must be attributed to 'deliberate intelligent design', as has been postulated²¹. The examples that follow suggest that complex networks of interacting proteins could indeed have evolved through processes of colocalization and that allosteric controls emerge by chance within these networks. Rather than presenting a paradox, the step-by-step evolution of complex, regulated networks emerges naturally when the laws of chemistry are coupled with natural selection.

Genetic fusion and the evolution of interacting proteins

A possible pathway for the evolution of a strongly interacting protein pair starting from two non-interacting proteins²² is illustrated in the upper path of Fig. 1. When a random genetic event fuses the genes that encode two proteins, the expression product is a single polypeptide chain with domains corresponding to the initial proteins. Because these domains are colocalized on the same chain, their effective concentration increases from the nanomolar-to-micromolar concentration of the separated proteins to a millimolar concentration in the fused pair. Now, a single mutation in the gene can result in the replacement of an amino acid, possibly decreasing the free energy enough to create a tightly binding 'heterodimer' (still on the same chain) with an altered function, which is therefore selectable. Additional mutations can further stabilize the non-covalent bonds between the two domains or create allosteric interactions between them. At this point, another single genetic event can separate the gene that encodes the fused pair into two genes, each encoding one of the two proteins. Natural selection has therefore generated a heterodimer that can participate in signalling or metabolic processes.

Colocalization can also explain the evolution of multisubunit homooligomeric proteins²³. The process can begin as shown in the upper pathway in Fig. 1. After the tightly binding multidomain protein has evolved, only a single genetic event is needed to convert the single-chain multidomain protein to a homodimer. This event is a genetic deletion that shortens the loop tethering the two original proteins. Such loop shortening can prevent the two domains from binding to one another but allow each domain to bind to its complementary domain in a second molecule. Numerous examples of oligomer formation by loop shortening have been observed in studies of genetically engineered proteins²⁴, strongly suggesting that such processes occur in nature.

Comparative genomics studies have found thousands of examples, over evolutionary timescales, of gene fusion resulting in a larger protein and of gene fission resulting in the constituent domains of a protein becoming separate proteins. These findings are collected in the protein-domain databases Pfam²⁵ and ProDom²⁶, and several examples are shown in Fig. 3. Predictions of pairs of proteins that bind to each other or participate in the same metabolic pathway can be made by finding a

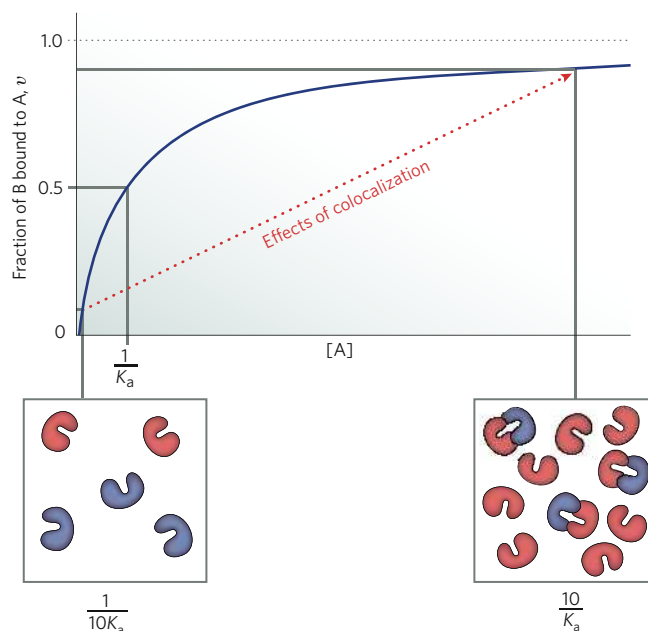


Figure 2 | The effect of colocalization on binding. Whether a protein, A (red), binds to another protein, B (blue), depends on the concentration of the first protein ($[A]$) and on the dissociation constant (K_d) of the complex (A–B). The fraction of B bound to A (v) is given approximately by the equation $v = ([A]/K_d)/(1 + [A]/K_d) = K_a[A]/(1 + K_a[A])$, where K_a is the association constant for the binding of A to B (which equals $1/K_d$). This describes a hyperbolic curve (blue). When $[A] = K_d = 1/K_a$, half of B is bound to A. A more exact relationship is needed, however, to distinguish between free A, plotted here on the x axis, and total A. This relationship gives a curve of a similar shape but with binding half-saturated at about $[A] = 0.4K_d$. Another approximation in this relationship is that activities are likely to differ from concentrations in the non-ideal environment of the crowded interior of a cell⁶⁷. Despite these approximations, it remains true that when $[A] < 0.1K_d = 1/10K_a$, little A is bound to B (left inset), and when $[A] > 10K_d = 10/K_a$, B is nearly saturated with A (right inset). This means that colocalization, which boosts greatly the effective concentration of A (red arrow), results in increased binding. The grey dashed line indicates the asymptote.

third protein that is homologous to both of the other proteins^{22,27}. For example, it can be inferred that the α -subunit and the β -subunit of carbamoyl-phosphate synthetase from *E. coli* bind to each other because both are homologous to the longer sequence of the same enzyme in humans (Fig. 3). Indeed, in *E. coli*, these two subunits form a complex, the structure of which has been determined²⁸. Systematic comparison of genomes shows that such gene-fusion and gene-fission events have been common in all three kingdoms of life^{29–31}.

In summary, commonly observed genetic events — recombinations, single-site mutations, and deletions — can account for the evolution of interacting proteins and of proteins in which multiple domains interact allosterically within a single chain. The allosteric mechanisms that emerge from colocalization and such genetic processes are likely to differ between homologous proteins, depending on which random mutations occurred in their evolutionary history. This principle is illustrated by the examples in the following three sections.

Allostery in DNA-binding proteins

Transcription factors recognize their target DNA sequences with high specificity by using cooperative binding: that is, the binding of one protein to DNA increases the affinity of another for an adjacent site^{32,33}. These cooperative interactions between multiple domains extend the effective length of the target DNA sequence. This cooperative binding is important because a typical DNA-binding domain does not make contact with enough bases for the interaction to be highly specific. A striking example is the structure of the interferon- β enhanceosome,

in which eight proteins are bound to DNA adjacent to each other and, collectively, recognize an extended DNA element³⁴.

Cooperativity in DNA binding often results from the colocalization of DNA-binding domains in oligomeric assemblies or in a single polypeptide chain. For example, many DNA-binding proteins, such as λ repressor, are dimeric even when they are not bound to DNA³². Others, such as zinc-finger-containing proteins, have multiple DNA-binding domains in the same polypeptide chain³². In all such cases, the DNA-binding domains do not necessarily need to make contact with each other for the binding to be cooperative. There are many transcription factors, however, that interact with each other only when bound to DNA³³. These interactions are allosteric, in the sense that contact with DNA results in the strengthening of protein–protein contacts. Although the mode of interaction with DNA is similar among evolutionarily related proteins that recognize DNA in this way, the nature of the protein–protein interactions can differ markedly. This principle is exemplified by the homeodomains, which are small DNA-binding domains found in many transcription factors that control development in animals.

The conserved core of homeodomains contains three α -helices, two of which form a helix–turn–helix motif³². By itself, a homeodomain binds weakly to DNA, but interactions with other DNA-binding domains, including other homeodomains, result in high-affinity and high-specificity DNA binding. In contrast to the highly conserved manner in which DNA is recognized by the homeodomain core, the interactions between the proteins bound to DNA are diverse, as shown for three homeodomain-containing proteins in Fig. 4a.

The homeodomain of *Drosophila melanogaster* Paired (PRD) proteins forms a highly cooperative homodimer on DNA, as a consequence of DNA deformations and reciprocal protein–protein contacts between the amino-terminal extension of one homeodomain and the second α -helix of the other³⁵ (Fig. 4a). The homeodomains of homeobox (HOX) proteins interact cooperatively with other homeodomains, such as those of *D. melanogaster* Extradenticle (EXD) and human pre-B-cell leukaemia homeobox 1 (PBX1). In contrast to the PRD dimer, in which the two homeodomains are located adjacent to each other on the DNA, the homeodomains of the HOX–protein–EXD and HOX–protein–PBX1 heterodimers are on opposite faces of the DNA, with the N-terminal linker of HOX reaching across the minor groove to engage the homeodomain of EXD or PBX1 (refs 36–38) (Fig. 4a).

Another distinct homeodomain interaction occurs for PIT1, which is a member of the POU family of transcription factors. These proteins contain a POU homeodomain and a POU-specific domain, the latter of which resembles the DNA-binding domains of bacterial transcription factors such as λ repressor³⁹. PIT1 binds cooperatively to DNA as a dimer; the protein–protein interactions involve the DNA-recognition helix of the POU homeodomain of one PIT1 protein and a surface element of the POU-specific domain of the other⁴⁰ (Fig. 4a). This is

in contrast to the situation for octamer-binding transcription factor 1 (OCT1), another member of the POU family, which recognizes DNA as a monomer³⁹. Yet another mechanism is used by the homeodomain protein Mata2, which controls mating type in yeast (*Saccharomyces cerevisiae*). When Mata2 interacts with either of two proteins, Mata1 or Mcm1, the affinity and the specificity of DNA binding increase. In the Mata2–Mata1 complex, the extra carboxy-terminal helix present in the homeodomain of Mata2 engages the Mata1 homeodomain⁴¹. By contrast, when the homeodomain of Mata2 interacts with Mcm1 (a MADS-box-domain-containing protein), the N-terminal region of the homeodomain mediates the contact⁴².

The unrelated nature of these DNA-dependent protein–protein contacts suggests that they evolved after primordial homeodomains first developed affinity for DNA, with the spacing between the binding sites on DNA determining which regions of the proteins can make contact with each other. That is, colocalization seems to have preceded the different random mutations that produced different binding relationships in each of these cases.

Allostery in haemoglobin

The pressure differential between oxygen in the lungs and in the tissues (roughly threefold in humans) is too small for oxygen to be effectively transported to the tissues by simple diffusion. In humans, haemoglobin therefore evolved into a highly cooperative tetramer (consisting of two α -globin– β -globin heterodimers; Fig. 4b), which switches from a structure with low affinity for oxygen to one with high affinity, in an ultrasensitive manner⁵. A key feature of Perutz's classic mechanism of allostery in haemoglobin⁵ is the coupling of the change in structure of an individual globin subunit after oxygen binding with the rotation of one α -globin– β -globin heterodimer with respect to the other. A crucial element in this coupling is the movement of the F helix, which is linked through a histidine residue to the iron in the haem group associated with each globin subunit¹⁷.

All animals with blood-based respiratory systems need to cope with the limited pressure differential between points of oxygen uptake and release, so it is not surprising that allosteric haemoglobin molecules are common in animals. Genomic studies have revealed the striking conservation of globin subunits throughout evolution. But, in contrast to the conserved structure of globin subunits, biochemical and structural studies have shown that the mechanism discovered by Perutz does not hold for all allosteric haemoglobin molecules⁴³. A haemoglobin of the clam *Scapharca inaequalvis*, for example, binds to oxygen cooperatively but is a dimer rather than a tetramer⁴⁴ (Fig. 4b). The mechanism discovered by Perutz, which relies on rotation of one dimer in the tetramer with respect to the other, cannot operate in a dimeric haemoglobin. Instead, the allostery in the *S. inaequalvis* haemoglobin results from a more direct transmission of the effects of oxygen binding at one haem group

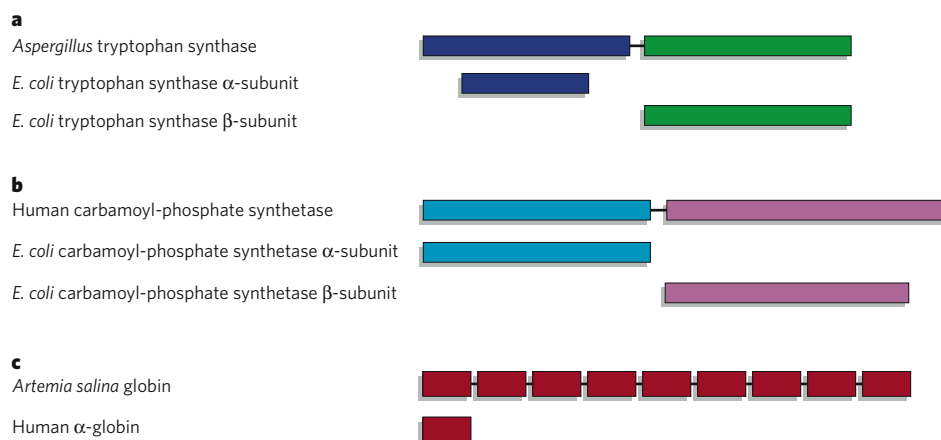


Figure 3 | Examples of fused protein domains in one organism that are homologous to separate domains in another organism. a, Tryptophan synthase. In *Escherichia coli*, the enzyme is a heterodimer formed from separate proteins that bind to one another. This pair of proteins is homologous

to a single protein in *Aspergillus*. **b**, Carbamoyl-phosphate synthetase. The situation is similar to that shown in part **a**. **c**, Globin. One of the subunits of human haemoglobin, α -globin, is homologous to a globin protein from the brine shrimp *Artemia salina* that consists of nine fused globin domains.

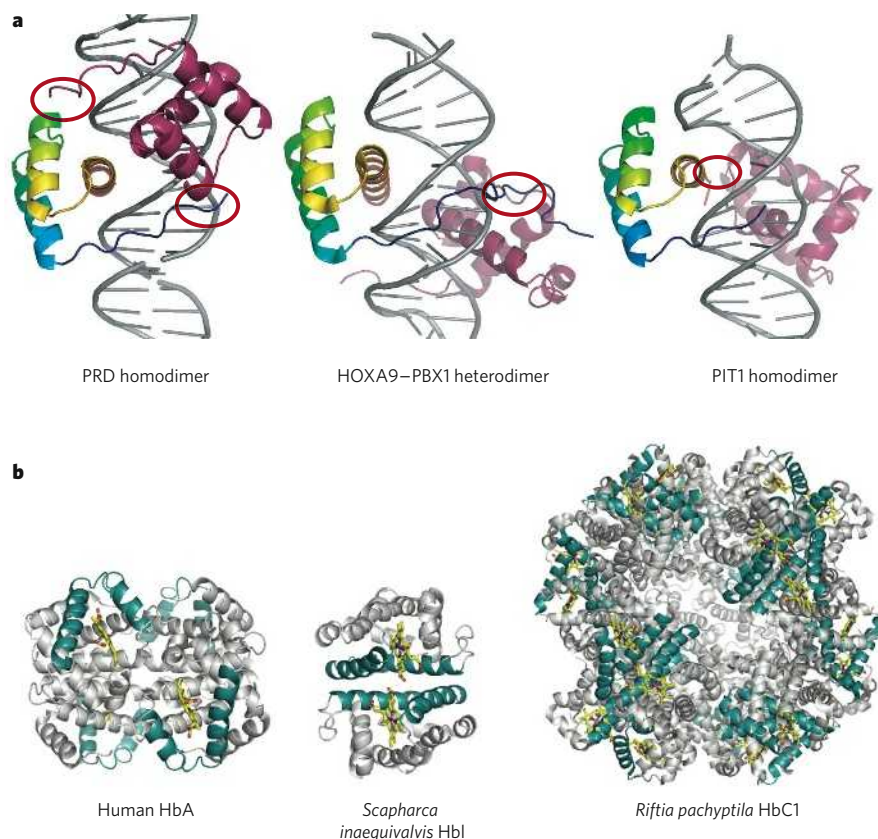


Figure 4 | Assemblies of homeodomain-containing proteins and haemoglobin. **a**, Three dimers of different proteins that contain homeodomains are shown bound to double-stranded DNA: a PRD homodimer, a HOXA9-PBX1 heterodimer, and a PIT1 homodimer. For clarity, the complete proteins are not shown. In each case, one homeodomain (left) is shown in the same orientation, with the colour varying from blue at the N terminus to pink at the C terminus. The domain that it interacts with is shown in pink, with the regions of contact indicated by red ovals. In the first two cases, the interaction occurs between homeodomains. For PIT1, the POU homeodomain in one molecule interacts with the POU-specific domain in the other molecule. Images generated from files from the Protein Data Bank (PDB), based on data from the following: ref. 35, file 1FJL (left); ref. 38, file 1PUF (centre); and ref. 40, file 1AU7 (right). **b**, Haemoglobin from three species is shown: humans (adult haemoglobin, HbA), the clam *Scapharca inaequivalvis* (HbI) and the tube worm *Riftia pachyptila* (HbC1). For each assembly, the two α -helices that bracket the haem group in each subunit are shown in blue. The haem group in each structure is shown in stick format, with carbon in yellow, oxygen in red, nitrogen in dark blue and iron in orange. Images generated from files from the PDB, based on data from the following: ref. 17, file 2HHB (left); ref. 68, file 4SDH (centre); and ref. 69, file 1YHU (right).

to the adjacent one, using conformational changes in the F helix that differ markedly from those in human haemoglobin.

A comparison of the quaternary structures of haemoglobin from various invertebrates reveals a striking diversity of assembly patterns⁴³. Some haemoglobin molecules are dimeric; some are tetrameric; and others are organized into higher-order oligomers (Fig. 4b). The diversity of interfacial packing between globin subunits indicates that the allosteric mechanisms differ in each case. The linked network of molecular interactions seen in human haemoglobin — in which the change in size of the iron atom is transmitted through the iron-linked histidine residue and the F helix, causing breakage of ion pairs at the interfaces between subunits⁵ — seems to be only one of several ways in which the globin fold can be adapted to yield a cooperative response to oxygen binding. Therefore, haemoglobin molecules in different organisms have acquired their allostery in several, apparently random, ways.

Mechanisms of control by phosphorylation

Phosphorylation is the most common covalent modification used to achieve allosteric control of proteins. In this section, we discuss two families of proteins — glycogen phosphorylases and a family of bacterial transcriptional activators — in which different mechanisms of control by phosphorylation have evolved within sets of homologous proteins. Then, we focus on protein tyrosine kinases, which carry out protein phosphorylation and are themselves regulated by phosphorylation, through extraordinarily diverse mechanisms.

Glycogen phosphorylases

Glycogen phosphorylase, which degrades glycogen to release glucose-1-phosphate, is activated by phosphorylation^{45–47}. Glycogen phosphorylase is a dimeric enzyme, and the structures of the subunits and the general features of the dimeric assembly are conserved in the yeast and mammalian enzymes (Fig. 5a), as well as in related non-allosteric bacterial proteins⁴⁸. The mechanism of control by phosphorylation of the yeast and mammalian enzymes is, however, different^{48,49}. In both proteins, the site of regulatory phosphorylation is in segments at the N terminus of

the polypeptide chain, but independent genetic-fusion events seem to have joined these unrelated segments to the conserved dimeric core of the enzyme^{50,51}. In the yeast enzyme, the N-terminal tail of the protein blocks the active site, preventing access to substrates. Phosphorylation of a serine residue in the N-terminal tail causes removal of the tail from the active site, thereby activating the enzyme. By contrast, allosteric control of the mammalian enzymes is much more complex. Inactivation of the unphosphorylated protein results from distributed conformational changes, rather than from physical occlusion of the active site. The N-terminal residue that is phosphorylated in the yeast enzyme is not present in the mammalian enzyme. Instead, phosphorylation occurs at another site in the N-terminal region, and the phosphorylated segment is docked differently (Fig. 5a). The structural changes induced by phosphorylation include a rotation of one subunit with respect to the other, and these changes correlate with responsiveness to ATP (an inhibitor) and AMP (an activator). Neither of these molecules has an effect on the activity of the yeast enzyme.

Bacterial transcriptional activators of the AAA+ superfamily

Phosphorylation also regulates the function of a family of bacterial transcriptional activators that belongs to a superfamily of proteins (known as AAA+ ATPases) with diverse ATP-dependent functions. These transcriptional activators contain a signal-receiver domain that is controlled by phosphorylation and an ATPase domain that couples to the σ^{54} -containing form of RNA polymerase. The basic switch that controls the activity of these proteins is transition from a monomeric or dimeric form, both of which are inactive, to a ring-shaped assembly, typically a hexamer. This structure helps the σ^{54} subunit of RNA polymerase to unwind duplex DNA. Although the signal-receiver and ATPase domains are highly conserved within this family of transcriptional activators, the mechanism by which phosphorylation regulates activity is not^{45–54} (Fig. 5b). In some members, such as nitrogen regulatory protein C (NtrC) from *Salmonella enterica* serovar Typhimurium, phosphorylation of the signal-receiver domain is required for the ATPase domain to oligomerize: that is, phosphorylation controls activity

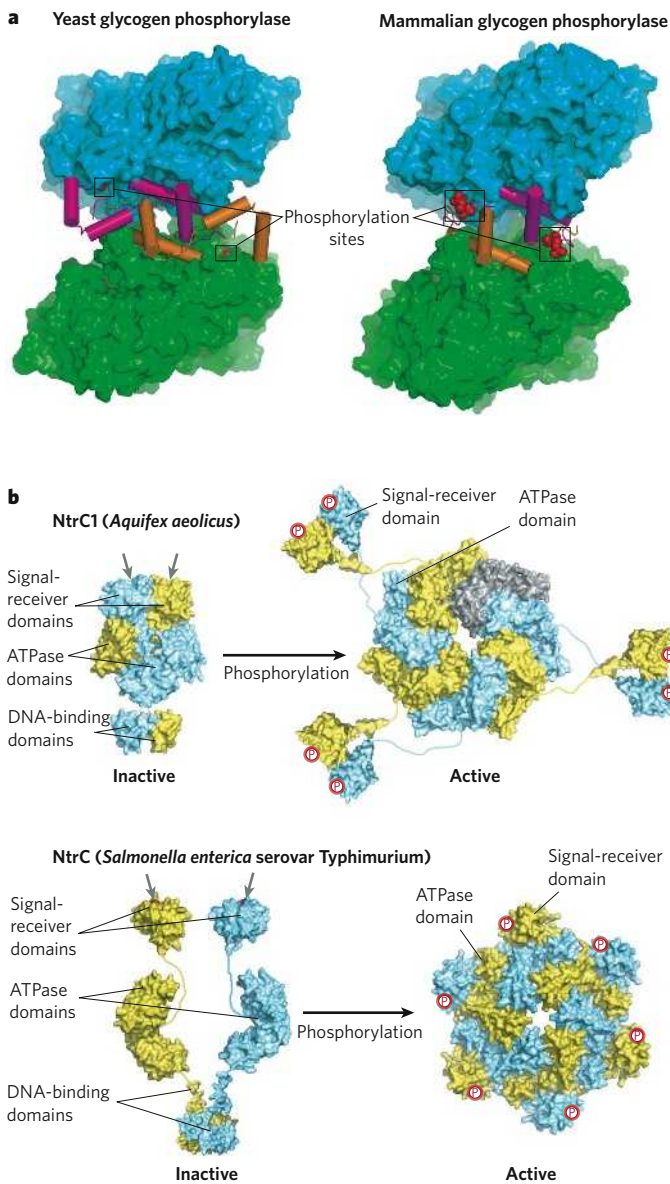


Figure 5 | Mechanisms of control by phosphorylation. **a**, Glycogen phosphorylase. The structures of the phosphorylated yeast enzyme (left) and mammalian enzyme (right) are shown, with one subunit in green and yellow and the other in blue and pink. Helices and loops in the N-terminal segments are shown as cylinders and coils, respectively. The phosphate groups are shown as red spheres (partly occluded in the yeast structure). This comparison shows that the structure at the N terminus (yellow and pink), which contains the sites of phosphorylation, differs between these proteins. Images generated from files from the PDB, based on data from the following: ref. 70, file 1YGP (left); and ref. 45, file 1GPA (right). **b**, Bacterial transcriptional activators of the AAA+ ATPase superfamily that associate with the σ^{54} form of RNA polymerase. Assembly of these proteins is controlled by phosphorylation, which results in a switch from the inactive (monomeric or dimeric) form (left) to the active form (right), in which the ATPase domains form an oligomeric ring. Two examples are shown, with one subunit in blue and the other in yellow. In NtrC1 from *Aquifex aeolicus* (upper panel), the signal-receiver domains hold the ATPase domains as dimers (which is an inactive conformation; left) until they are phosphorylated. After phosphorylation, a conformational change in the signal-receiver-domain dimer suppresses interaction with the ATPase domains, which are then free to assemble into the active oligomer (right). The crystal structure shows a heptamer rather than a hexamer, and the extra subunit is shown in grey. By contrast, in the homologue NtrC from *Salmonella enterica* serovar Typhimurium (lower panel), the signal-receiver domains and the ATPase domains do not interact in the absence of phosphorylation (left). After phosphorylation, the signal-receiver domains bind to a neighbouring ATPase domain, stabilizing the assembled active ring of the ATPase (right). The sites that undergo phosphorylation are indicated by grey arrows for the inactive molecules (left) and red circles for the active molecules (right). The DNA-binding domains are present at the bottom of the inactive forms of the proteins (an orientation that is required to maintain inactive NtrC in the dimeric state), and they are underneath the assembled oligomers (and therefore not visible in these diagrams).

positively (Fig. 5b). In other family members, such as NtrC1 from *Aquifex aeolicus* and dicarboxylate transport regulator D (DctD) from *Sinorhizobium meliloti*, phosphorylation is required to disrupt a dimeric state that prevents hexamerization of the ATPase domain (Fig. 5b). In this case, the signal-receiver domain is dispensable for activity. As is the case for glycogen phosphorylase, it is clear that phosphorylation-mediated control evolved after the basic mechanism of oligomerization had been set in place.

Protein tyrosine kinases

The clustering of receptor molecules at the plasma membrane is emerging as a key feature of intracellular signal transduction. Such clustering further increases the high local concentrations of membrane or receptor-associated signalling molecules, and it promotes a diverse range of protein–protein interactions⁵⁵. Allosteric is a common attribute of these proteins, with one domain modulating the activity of another domain in the same molecule. For proteins with homologous catalytic domains, these allosteric interactions follow no common pattern. This is exemplified by the protein tyrosine kinases, enzymes that are crucial for cell–cell communication in metazoans.

Cytoplasmic (non-receptor) protein tyrosine kinases have a conserved catalytic domain (known as the kinase domain), which is fused to targeting domains (also known as regulatory domains) that bind to other

proteins or to lipids⁵⁶. The primordial function of these targeting domains was probably to localize protein tyrosine kinases to sites of signalling, but they have evolved the ability to regulate the activity of the kinase domain. Here, we consider three cytoplasmic protein tyrosine kinases: Abl (the cellular homologue of the oncogene encoded by the Abelson leukaemia virus), ZAP70 (ζ -chain associated protein kinase of 70 kDa) and FAK (focal adhesion kinase; also known as PTK2). Each of these is activated by the phosphorylation of one or two tyrosine residues that are located between the targeting domains and the kinase domain, a process that releases the targeting domains from interaction with the kinase domains. In each case, however, the targeting domains suppress the activity of the kinase domain in a different manner.

Abl has two targeting domains — a Src-homology 2 (SH2) domain and an SH3 domain — fused to the kinase domain. These targeting domains clamp onto the distal surface of both lobes of the kinase domain, suppressing activity in a similar manner to that used by Src-family kinases, which are closely related^{57–59} (Fig. 6a). Phosphorylation of the linker between the SH2 domain and the kinase domain in Abl prevents the engagement of the SH3–SH2 unit with the kinase domain, thereby activating Abl⁶⁰. By contrast, ZAP70 has a tandem SH2 unit fused to the kinase domain, and this unit inhibits catalytic activity by interacting with the hinge region of the kinase domain of ZAP70, suppressing its flexibility⁶¹ (Fig. 6a). Phosphorylation of the linker between

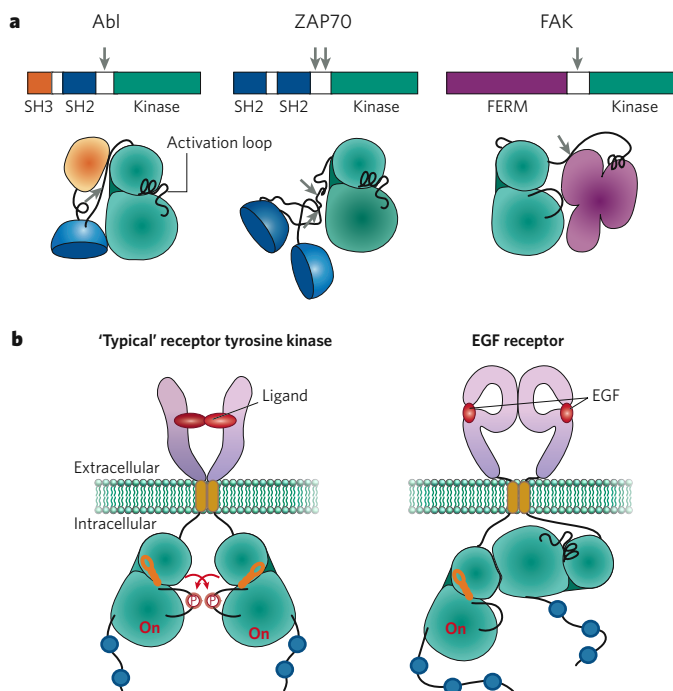


Figure 6 | Diverse regulatory mechanisms in tyrosine kinases. a, The domain organization and auto-inhibited structures of three cytoplasmic (non-receptor) protein tyrosine kinases (Abl, ZAP70 and FAK) are shown. These enzymes are activated by phosphorylation of tyrosine residues (indicated by grey arrows) located in the linker between the targeting domains and the kinase domain, but the molecular mechanism by which this phosphorylation causes activation differs for each enzyme. **b**, The activation mechanisms for a typical receptor tyrosine kinase and for the EGF receptor, an atypical receptor tyrosine kinase, are shown. Typical receptors undergo ligand-induced dimerization and can then be activated by the phosphorylation of each kinase domain by the other. For the EGF receptor, dimerization of the extracellular portion of the receptor by EGF results in the formation of an asymmetrical dimer by the kinase domains in the cytoplasm, which then activates one of these domains. This mechanism is unique to members of the EGF-receptor family. Signal transduction is propagated through the docking of SH2-containing molecules to phosphorylated residues (blue) adjacent to the kinase domains. This diagram is based on structures that were determined separately for the extracellular domains^{71–73} and the cytoplasmic domain⁶⁶.

the SH2 domain and the kinase domain activates ZAP70, by preventing the formation of interactions between aromatic amino acids that are crucial for assembly of the auto-inhibited ZAP70 (ref. 61). In FAK, there is a targeting domain known as a FERM domain, which is located N-terminal to the kinase domain (Fig. 6a). Unlike the interactions in the previous two examples, the FERM domain interacts with the 'front' of the kinase domain, where it directly blocks access to the active site⁶². Phosphorylation of the linker activates FAK by destabilizing the interaction between the FERM domain and the kinase domain. This comparison of three cytoplasmic protein tyrosine kinases is representative of a feature common to signalling pathways: they resemble the haphazard collection of assorted parts in the imagined devices of the cartoonist Rube Goldberg, with nature inventing multiple ways to regulate the activity of a conserved catalytic domain.

Receptor tyrosine kinases are transmembrane proteins in which an extracellular ligand-binding domain is separated from an intracellular kinase domain by the plasma membrane⁶³. The simplest mechanism (most probably the primordial mechanism) by which a receptor tyrosine kinase can be activated involves the phosphorylation of a centrally located activation loop in one subunit of a homodimer by the kinase domain in the other subunit, and vice versa, a reaction that is promoted by ligand-induced dimerization⁶⁴. Individual receptors, however, have evolved regulatory mechanisms that are layered on top of this simple mechanism. The insulin receptor, for example, is a covalently crosslinked

dimer, and *trans*-phosphorylation results from an insulin-induced conformational change rather than from a monomer–dimer transition⁶⁵. Unlike the insulin receptor and other typical receptor tyrosine kinases (Fig. 6b), the receptor for epidermal growth factor (EGF) does not require phosphorylation of an activation loop for catalytic activity. Instead, the activation mechanism involves an asymmetrical interaction between the large lobe of one kinase domain and the small lobe of the other, a process that stabilizes the active conformation of the latter⁶⁶ (Fig. 6b). This mechanism, which resembles the way in which protein kinases that control the cell cycle (CDKs) are activated by cyclins, does not seem to be used by other receptor tyrosine kinases. Within the EGF-receptor family, however, the ability of the kinase domains to function as both activators and transducers for each other leads to a powerful combinatorial response to a variety of ligands.

Conclusions

Genome sequencing has only begun to uncover the molecular details of the great puzzle of how complex and interacting molecular forms emerged from simpler ones. One of the findings that has emerged from genomic analysis is that the machinery of life is conserved across the evolutionary tree. Globin subunits, for example, have the same overall structure and the same chemical linkage to the haem iron in plants, invertebrates and mammals. Glycogen phosphorylase has the same dimeric structure in yeast and humans. Beginning with haemoglobin, researchers have come to appreciate that the regulated functioning of protein-based machines depends on allosteric interactions between one or more components in the assembly. Given the uniformity of the basic designs of protein modules, it could be expected that the allosteric mechanisms are also conserved. That they are not helps to resolve the otherwise insurmountable paradox of how such intricate mechanisms could have evolved from the constituent parts. The physical imperative for the allosteric control of oxygen binding to its transport protein has been solved by evolution in many organisms, but there is no combinatorial imperative that requires a particular interface or residue to be used in the mechanism. The variety of mechanisms that has been found seems to disclose the random nature of the events that gave rise to each.

Whether this 'rule of varied allosteric control' is generally applicable should emerge from further comparative studies of allosteric control in protein families. Similarly, genome sequencing of the organisms that diverged earliest might reveal ancestors of present-day proteins that gave rise to protein interactions. In particular, such analyses might definitively determine whether protein–protein interactions arise from the fusion of genes encoding protein domains followed by the fission of such fused genes. This hypothesis could be termed the 'rule of heterodimer evolution by protein fission'. It is now the turn of molecular scientists to uncover details of the process that Charles Darwin summarized famously in the final sentence of *On the Origin of Species*: "whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved".

- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).
- Beerink, P. T., Endrizzi, J. A., Alber, T. & Schachman, H. K. Assessment of the allosteric mechanism of aspartate transcarbamoylase based on the crystalline structure of the unregulated catalytic subunit. *Proc. Natl Acad. Sci. USA* **96**, 5388–5393 (1999).
- Behe, M. J. *Darwin's Black Box: The Biochemical Challenge to Evolution* (The Free Press, New York, 2003).
- Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726–739 (1970).
- Xu, D., Tsai, C. J. & Nussinov, R. Mechanism and evolution of protein dimerization. *Protein Sci.* **7**, 533–544 (1998).
- Ispolatov, I., Yuryev, A., Mazo, I. & Maslov, S. Binding properties and evolution of homodimers in protein–protein interaction networks. *Nucleic Acids Res.* **33**, 3629–3635 (2005).
- Liang, J., Kim, J. R., Boock, J. T., Mansell, T. J. & Ostermeier, M. Ligand binding and allostery can emerge simultaneously. *Protein Sci.* **16**, 929–937 (2007).
- Pawson, T. & Scott, J. D. Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278**, 2075–2080 (1997).

10. Klemm, J. D. & Pabo, C. O. Oct-1 POU domain–DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.* **10**, 27–36 (1996).
11. Robinson, C. R. & Sauer, R. T. Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA. *Biochemistry* **35**, 109–116 (1996).
12. Predki, P. F. & Regan, L. Redesigning the topology of a four-helix-bundle protein: monomeric Rop. *Biochemistry* **34**, 9834–9839 (1995).
13. Liang, H., Sandberg, W. S. & Terwilliger, T. C. Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc. Natl Acad. Sci. USA* **90**, 7010–7014 (1993).
14. Pedersen, S., Bloch, P. L., Reeh, S. & Neidhardt, F. C. Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179–190 (1978).
15. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnol.* **25**, 117–124 (2007).
16. Noguchi, C. T. & Schechter, A. N. Sickle hemoglobin polymerization in solution and in cells. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 239–263 (1985).
17. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **175**, 159–174 (1984).
18. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53**, 708–719 (2003).
19. Fersht, A. R. *Structure and Mechanism in Protein Science* (Freeman, New York, 1999).
20. Eisenberg, D., Wesson, M. & Yanashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.* **29A**, 217–221 (1989).
21. Behe, M. J. *The Edge of Evolution* (Free Press, New York, 2007).
22. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
23. Bennett, M. J., Choe, S. & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proc. Natl Acad. Sci. USA* **91**, 3127–3131 (1994).
24. Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv. Protein Chem.* **50**, 61–122 (1997).
25. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
26. Servant, F. *et al.* ProDom: automated clustering of homologous domains. *Brief. Bioinform.* **3**, 246–251 (2002).
27. Enright, A. J., Iliopoulos, I., Kyripides, N. C. & Ouzounis, C. A. Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**, 86–90 (1999).
28. Thoden, J. B., Raushel, F. M., Benning, M. M., Rayment, I. & Holden, H. M. The structure of carbamoyl phosphate synthetase determined to 2.1 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 8–24 (1999).
29. Snel, B., Bork, P. & Huynen, M. Genome evolution: gene fusion versus gene fission. *Trends Genet.* **16**, 9–11 (2000).
30. Kummerfeld, S. K. & Teichmann, S. A. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **21**, 25–30 (2005).
31. Fong, J. H., Geer, L. Y., Panchenko, A. R. & Bryant, S. H. Modeling the evolution of protein domain architectures using maximum parsimony. *J. Mol. Biol.* **366**, 307–315 (2007).
32. Pabo, C. O. & Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095 (1992).
33. Wolberger, C. Multiprotein–DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 29–56 (1999).
34. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-β enhanceosome. *Cell* **129**, 1111–1123 (2007).
35. Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709–719 (1995).
36. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S. & Aggarwal, A. K. Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. *Nature* **397**, 714–719 (1999).
37. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L. & Wolberger, C. Structure of a HoxB1–Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**, 587–597 (1999).
38. LaRonde-LeBlanc, N. A. & Wolberger, C. Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev.* **17**, 2060–2072 (2003).
39. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21–32 (1994).
40. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G. & Aggarwal, A. K. Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.* **11**, 198–212 (1997).
41. Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. Crystal structure of the MATA1/MATA2 homeodomain heterodimer bound to DNA. *Science* **270**, 262–269 (1995).
42. Tan, S. & Richmond, T. J. Crystal structure of the yeast MATA2/MCM1/DNA ternary complex. *Nature* **391**, 660–666 (1998).
43. Royer, W. E., Zhu, H., Gorr, T. A., Flores, J. F. & Knapp, J. E. Allosteric hemoglobin assembly: diversity and similarity. *J. Biol. Chem.* **280**, 27477–27480 (2005).
44. Royer, W. E., Hendrickson, W. A. & Chiancone, E. The 2.4 Å crystal structure of Scapharca dimeric hemoglobin. *J. Biol. Chem.* **264**, 21052–21061 (1989).
45. Barford, D., Hu, S. H. & Johnson, L. N. Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *J. Mol. Biol.* **218**, 233–260 (1991).
46. Barford, D. & Johnson, L. N. The allosteric transition of glycogen phosphorylase. *Nature* **340**, 609–616 (1989).
47. Sprang, S. R. *et al.* Structural changes in glycogen phosphorylase induced by phosphorylation. *Nature* **336**, 215–221 (1988).
48. Buchbinder, J. L., Rath, V. L. & Fletterick, R. J. Structural relationships among regulated and unregulated phosphorylases. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 191–209 (2001).
49. Rath, V. L. & Fletterick, R. J. Parallel evolution in two homologues of phosphorylase. *Nature Struct. Biol.* **1**, 681–690 (1994).
50. Palm, D., Goerl, R. & Burger, K. J. Evolution of catalytic and regulatory sites in phosphorylases. *Nature* **313**, 500–502 (1985).
51. Hwang, P. K. & Fletterick, R. J. Convergent and divergent evolution of regulatory sites in eukaryotic phosphorylases. *Nature* **324**, 80–84 (1986).
52. Lee, S. Y. *et al.* Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev.* **17**, 2552–2563 (2003).
53. Doucleff, M. *et al.* Negative regulation of AAA+ ATPase assembly by two component receiver domains: a transcription activation mechanism that is conserved in mesophilic and extremely hyperthermophilic bacteria. *J. Mol. Biol.* **353**, 242–255 (2005).
54. De Carlo, S. *et al.* The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev.* **20**, 1485–1495 (2006).
55. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452 (2003).
56. Neet, K. & Hunter, T. Vertebrate non-receptor protein-tyrosine kinase families. *Genes Cells* **1**, 147–169 (1996).
57. Nagar, B. *et al.* Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859–871 (2003).
58. Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src-family tyrosine kinase Hck. *Nature* **385**, 602–609 (1997).
59. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
60. Brasher, B. B. & Van Etten, R. A. c-Abl has high intrinsic tyrosine kinase activity that is stimulated by mutation of the Src homology 3 domain and by autophosphorylation at two distinct regulatory sites. *J. Biol. Chem.* **275**, 35631–35637 (2000).
61. Deindl, S. *et al.* Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. *Cell* **129**, 735–746 (2007).
62. Lietha, D. *et al.* Structural basis for the autoinhibition of focal adhesion kinase. *Cell* **129**, 1177–1187 (2007).
63. Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225 (2000).
64. Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669–672 (2002).
65. Hubbard, S. R. & Till, J. H. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **69**, 373–398 (2000).
66. Zhang, X., Gureasko, J., Shen, K., Cole, P. A. & Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125**, 1137–1149 (2006).
67. Deeds, E. J., Ashenberg, O., Gerardin, J. & Shakhnovich, E. I. Robust protein–protein interactions in crowded cellular environments. *Proc. Natl Acad. Sci. USA* **104**, 14952–14957 (2007).
68. Royer, W. E. High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. *J. Mol. Biol.* **235**, 657–681 (1994).
69. Flores, J. F. *et al.* Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc. Natl Acad. Sci. USA* **102**, 2713–2718 (2005).
70. Lin, K., Rath, V. L., Dai, S. C., Fletterick, R. J. & Hwang, P. K. A protein phosphorylation switch at the conserved allosteric site in GP. *Science* **273**, 1539–1542 (1996).
71. Cho, H. S. & Leahy, D. J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **297**, 1330–1333 (2002).
72. Garrett, T. P. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. *Cell* **110**, 763–773 (2002).
73. Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787 (2002).

Acknowledgements We are grateful to A. K. Aggarwal, R. E. Dickerson, S. C. Harrison, L. N. Johnson, E. M. Marcotte, M. Robertson, W. E. Royer, M. Seeliger, D. E. Wemmer, C. Wolberger, T. O. Yeates and many other colleagues for comments. We thank S. Deindl, L. Leighton, W. E. Royer, D. E. Wemmer and X. Zhang for assistance with the figures. Support from the Howard Hughes Medical Institute, the National Institutes of Health, the US Department of Energy Office of Biological & Environmental Research, and the National Science Foundation is gratefully acknowledged.

Author information Reprints and permissions information is available at npg.nature.com/reprints. Correspondence should be addressed to the authors (david@mbi.ucla.edu; kuriyan@berkeley.edu).