Protein molecules as computational elements in living cells

Dennis Bray

Many proteins in living cells appear to have as their primary function the transfer and processing of information, rather than the chemical transformation of metabolic intermediates or the building of cellular structures. Such proteins are functionally linked through allosteric or other mechanisms into biochemical 'circuits' that perform a variety of simple computational tasks including amplification, integration and information storage.

In unicellular organisms, protein-based circuits act in place of a nervous system to control behaviour; in the larger and more complicated cells of plants and animals, many thousands of proteins functionally connected to each other carry information from the plasma membrane to the genome. The imprint of the environment on the concentration and activity of many thousands of proteins in a cell is in effect a memory trace, like a 'random access memory' containing ever-changing information about the cell's surroundings. Because of their high degree of interconnection, systems of interacting proteins act as neural networks trained by evolution to respond appropriately to patterns of extracellular stimuli. The 'wiring' of these networks depends on diffusion-limited encounters between molecules, and for this and other reasons they have unique features not found in conventional computer-based neural networks.

Components

In principle, any protein that transforms an input signal into an output signal could act as a computational or information-carrying element. Thus an enzyme in a biochemical pathway 'reads' the concentration of its substrate and produces a corresponding level of product; a receptor on a cell surface reads the concentration of its ligand and produces a certain level of receptor–ligand complex (Fig. 1*a*). Simple enzymes and receptors generate a monotonic relationship between input and output, which saturates as the concentration of substrate or ligand rises. A more abrupt and switch-like performance is shown by many proteins, typically those composed of multiple subunits, where interactions between subunits cause the rate of enzyme reaction or ligand binding to rise steeply in sigmoidal fashion over a limited range of concentration (Fig. 1*b*).

These simple (although nonlinear) relationships are made more complex by allosteric mechanisms that allow proteins to be controlled by changes not only in their levels of substrate or ligand but also in the concentrations of regulator molecules. The enzyme aspartate transcarbamoylase, for example, is composed of multiple catalytic and regulatory subunits, and flips in an abrupt, cooperative fashion between an active and an inactive state. Binding of its substrates (aspartate and carbamoyl phosphate) drives the enzyme into an active conformation in which it makes *N*-carbamoylaspartate and hence begins the synthesis of the pyrimidine ring of C, U and T nucleotides. By contrast, binding of cytidine triphosphate (CTP), one of the end products of the pathway, converts the enzyme into an inactive conformation from which the substrates dissociate. The rate of reaction of this enzyme is consequently highly sensitive to variations in concentration of three input molecules[1] (Fig. 1*c*).

The activity of a protein can be altered not only by binding to a regulatory molecule but also by enzyme-catalysed modification. Many proteins are chemically modified after synthesis by the covalent addition of chemical groups such as methyl, nucleotidyl, fatty acyl, myristoyl and, particularly, phosphoryl groups, changes that, in general, alter the biological activity of the protein. In some cases protein modification changes the three-dimensional structure of the protein and, by altering the interactions with adjoining subunits, thereby causes a cooperative, or sigmoidal, response. The enzyme glycogen phosphorylase, for example, could be represented symbolically in a similar way to aspartate transcarbamoylase in Fig. 1*c* simply by replacing the inputs with (1) the concentration of its substrate, glycogen, (2) the activity of phosphorylase kinase, which adds a phosphate group to a site that is remote from the catalytic site, and (3) the activity of a protein phosphatase that removes this phosphate group[2].

Post-translational modification allows a single species (the modifying enzyme) to influence the activity of many others. Thus a change in activity of a protein kinase typically affects multiple target proteins. The action of a protein kinase may also be made more complex if it works back on itself to phosphorylate its own amino acids. Proteins that show this intramolecular feedback sometimes appears to act as an irreversible switch, as seen in the calcium-sensitive enzyme Cam II kinase (Fig. 1*d*). This protein, abundant in brain and concentrated in synaptic terminals, is stimulated by an increase in the level of cytosolic Ca^{2+} ions in the presence of calmodulin, leading to the phosphorylation of multiple target proteins. Among these target proteins is the Cam II kinase itself which, when it attains a sufficiently high level of phosphorylation, becomes irreversibly active regardless of the level of Ca^{2+}. This enzyme is indeed a sophisticated molecule, for its level of activity depends not only the size and number of Ca^{2+} spikes but also their frequency[3]. It is believed that this protein may be important for neuronal processes, such as synaptic plasticity, that depend on frequency-dependent modulation by Ca^{2+} spikes[4].

Analogous processing is seen in the membrane receptors for growth factors such as platelet-derived growth factor (PDGF) and insulin. When these receptors bind their ligand they associate into dimers, which activates a kinase associated with their cytoplasmic domains. Each kinase then phosphorylates, in reciprocal fashion, its sister domain on crucial tyrosine residues. These phosphorylated tyrosines are subsequently recognized by other proteins in the cytoplasm which bind to the sequence containing the modified tyrosine and hence change their own activity. In this way, information that the cell has encountered a specific hormone or growth factor is distributed along multiple divergent paths[5,6].

Phosphorylation also provides the basis of integration whereby a single protein combines multiple inputs to produce a single output in a manner analogous to the integration by a nerve cell of multiple synaptic inputs. This type of integration is seen in the product of the *c-src* gene, an intracellular signalling protein that is itself a kinase[7], and in glycogen synthase, the enzyme that synthesizes glycogen in a liver or muscle cell[8]. The latter protein has at least ten distinct sites at which phosphates can be added. Six or more different protein kinases are responsible for the addition of these phosphate groups, and one or more phosphat-

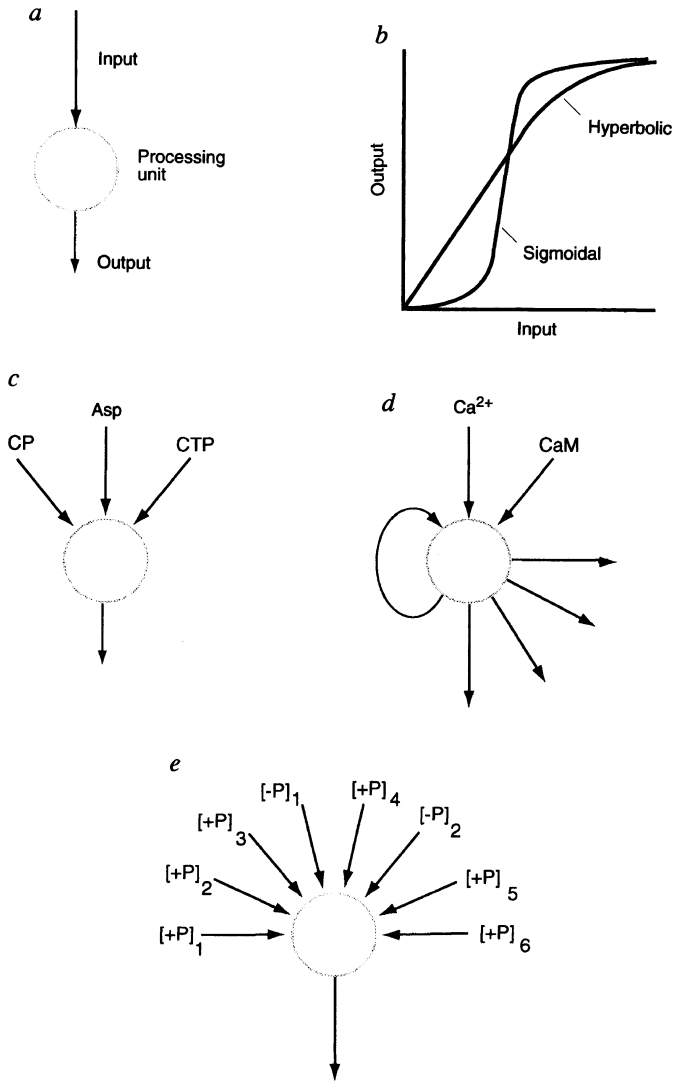

FIG. 1 Proteins as computational elements. *a*, A simple protein shown as a processing unit with the flow of information indicated by an arrow. The symbol used here is based on a perceptron, which is a device used to portray the processing of synaptic inputs by a nerve cell. For a simple enzyme, the input would be the concentration of its substrate and its output catalytic activity. For a membrane receptor, the input would be the concentration of ligand and the output the level of receptor occupancy. *b*, The input/output relationship of a simple enzyme or receptor is typically either hyperbolic or sigmoidal. *c*, Aspartate transcarbamoylase, a classic example of an allosteric enzyme. The activity of the enzyme is controlled by the concentration of its two substrates, carbamoyl phosphate (CP) and aspartate (Asp) in combination with the concentration of an end product, cytidine triphosphate (CTP), of the biosynthetic pathway initiated by the enzyme. *d*, Cam II kinase, a molecular switch, becomes active when both calcium ions (Ca^{2+}) and calmodulin (CaM) are present. The enzyme phosphorylates multiple target proteins including itself. When it reaches a sufficiently high level of phosphorylation, the enzyme switches to a permanently active state that is independent of the concentration of Ca^{2+}. *e*, Glycogen synthase, the enzyme that makes glycogen, is phosphorylated on multiple sites through the action of six protein kinases (+P) and several protein phosphatases (−P).

ases remove them. The target protein in this example serves to integrate the action of multiple modifying enzymes (Fig. 1*e*).

Living cells contain an enormous diversity of protein structures and the signals they carry are not limited to the binding of molecules or covalent modifications. Thus many proteins take as their 'input' contact with a macromolecule such as another protein or a molecule of DNA or RNA. Proteins are also known that respond specifically to light, to temperature, mechanical forces or voltage. The 'output' of a protein is just as diverse, and may be the formation of a macromolecular structure, the generation of a physical movement or production of light. Each input/output relationship is kinetically distinct, usually nonlinear, and often extremely rapid. Dynamic aspects of protein processing have been analysed in greatest detail in the behaviour of ion channels in the plasma membrane. Physiological recordings using patch-clamp electrodes show that a typical voltage-gated or ligand-gated ion channel—structures built from several protein molecules embedded in the plasma membrane—has multiple distinct states and changes from one to the next in less than 50 μ sec (ref. 9). In principle, such a channel could carry information at a substantial rate.

Building circuits with proteins

In an important theoretical paper[10] in 1943 on 'artificial intelligence', the neurophysiologists McCulloch and Pitts demonstrated that small numbers of idealized nerve cells, linked into circuits, can perform a variety of logical operations with no obvious upper limit to their complexity. Although it was recognized from the start that these devices ('McCulloch–Pitts neurons') were artificial and highly simplified, there was no reason, it was argued, why real nerve cells in the brain should not be capable of performing equally complex logical operations. Few today would disagree with this thesis.

As with neurons, so with proteins. Putting aside for the moment the question of whether it is useful or even sensible to view them in this manner, it is nevertheless true that protein molecules are in principle able to perform a variety of logical or computational operations. The symbols used in Fig. 1 to depict the operation of individual protein molecules are based on another, slightly more flexible, type of neuron-like unit, called a perceptron[11,12]. Thus protein molecules provide a cell with a toolkit of components from which virtually any circuit could be built. To take a simple example: a protein that is modified at two allosteric sites by two separate kinases (or by binding two independent allosteric regulators) could perform a variety of logical operations on these two 'inputs'. If this protein is active only when both sites are phosphorylated, it will act somewhat like a boolean AND gate: if it is active when either or both sites are active, it will be like an inclusive OR gate, and so on.

The crispness of the input/output response—how closely it resembles a digital switch—will depend on the kinetic parameters of the protein, such as its cooperativity[13]. Although each indi-

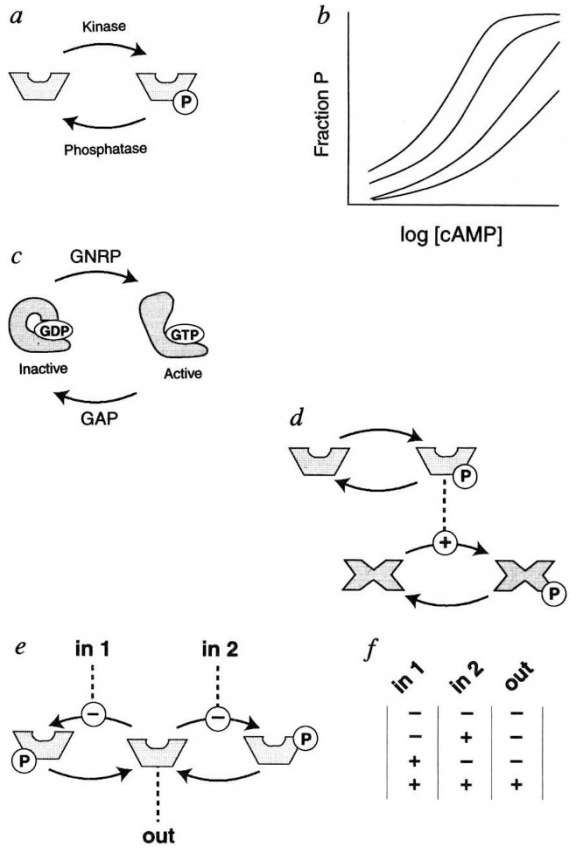

FIG. 2 Cyclic reactions as computational elements. *a*, Schematic representation of the cyclic phosphorylation and dephosphorylation of an enzyme, such as pyruvate dehydrogenase or glycogen phosphorylase. Phosphorylation is catalysed by a protein kinase and driven by the essentially irreversible hydrolysis of ATP; dephosphorylation is catalysed by a protein phosphatase driven by the release of inorganic phosphate (P_i). *b*, Experimentally measured performance of a cyclic reaction. The fractional phosphorylation of a target polypeptide in a system containing cAMP-dependent protein kinase and phosphoprotein phosphatase was measured over a range of concentrations of cAMP which stimulates the kinase. Each curve shows the effect of a range of cAMP concentrations (the 'input') on the fraction of target molecules that are phosphorylated (the 'output'). Measurements were repeated at different concentrations of P_i, which inhibits the phosphatase. Data redrawn from ref. 14. *c*, The cyclic interconversion of GTP-binding proteins forms the basis of many signal-transducing steps in living cells. The protein goes through a series of allosteric transitions driven by the irreversible hydrolysis of its bound GTP. The active form of the protein, associated with GTP, triggers one of a variety of different cell processes. Progression through the cycle, and hence the rate of formation of the active species, is controlled by regulatory proteins that either catalyse GTP hydrolysis (GTPase-activating protein (GAP)) or induce the protein to release GDP, thereby preparing it for association with GTP and another round of activation (guanine nucleotide releasing factor (GNRP)). These regulatory proteins thus perform a similar function to the kinase and phosphates enzymes in *a*. *d*, Two enzyme cycles coupled together. The phosphorylated enzyme produced in the first cycle acts as an enzyme to catalyse phosphorylation of the second enzyme. The potential for amplification achieved by the two cycles is a multiplicative function of the two individual cycles[22]. *e*, A reaction mechanism expected to function as an AND gate. Each of the two inputs ('in 1' and 'in 2') is assumed to inhibit one of the two protein kinases; when both are present, the concentration of A ('out') reaches its highest possible level. *f*, Given a suitable selection of rate constants and concentrations, a sharp ON response will be achieved only when both inputs are present. The addition of more enzyme cycles coupled to this one can create other logical devices, such as NOT and OR and XOR[13].

vidual protein molecule can exist in (usually) only two distinct conformations, the performance of a large assembly of such molecules depends on the probability that each molecule will occupy one or other of these conformations. For this reason, a different type of computational element in which the interconversion of two forms of a protein is energy driven is often able to give sharper and more easily controlled responses (Fig. 2*a*). Thousands of cyclic interconversions coupled to the hydrolysis of a molecule such as ATP occur in eukaryotic cells, each with the potential of acting as a molecular switch. In 1977, Stadtman and colleagues[14,15] saw that the cyclic interconversion of a protein between its phosphorylated and non-phosphorylated forms offers a flexible computational unit capable of very rapid responses. They demonstrated, first by theoretical analysis and later by actual experiment, that the input/output response of such cycles can be modulated over a wide range by changes in rate constants and metabolite levels (Fig. 2*b*)[14,15]. Sharp, digital-like responses can be attained if the target enzyme is highly cooperative or if the modifying enzymes are working near to saturation with respect to their substrates[16].

Some of the more subtle features of such switches have been recently explored by Arkin and Ross[13], who analysed the conversion of fructose-6-phosphate between two biphosphate forms by specific kinases and phosphatases. Control of the enzymes in this kinetic mechanism, through multiple allosteric regulators, serves to switch the pathway from glycolysis to gluconeogenesis according to the energy status of the cell. The operation of this switch, simulated either in isolation or in the context of a large model of glycolysis and the citric acid cycle, was substantially richer than that of a simple ON/OFF switch. Asymmetries and synergisms between the various inputs caused the gate to operate in some situations as a reasonably smooth AND gate, and in others as an OR gate, thereby conforming to a more general class of logic called 'fuzzy logic'[13].

Many proteins cycle between structural states that differ not in phosphorylation but in their association with a nucleoside triphosphate. Directed movements in living cells, for example, are driven by the cyclic hydrolysis of ATP catalysed by motor proteins such as myosin or kinesin. Many intracellular signals are carried by the cyclic, enzyme-catalysed hydrolysis of GTP (Fig. 2*c*). GTP-hydrolysing proteins are abundant in eukaryotic cells and used in many different processes, such as directing vesicular traffic through different membrane compartments of the cell and controlling the accuracy of protein synthesis[17,18]. The small GTP-binding protein, Ras, is an enabling switch for a cascade of protein kinases, and has a wide influence on many aspects of cell growth and differentiation[19,20]. The same fundamental cycle, in more complex form and involving more molecules, occurs in the heterotrimeric G proteins, which transduce sensory and hormonal stimuli across the plasma membrane in almost all eukaryotic cells[21].

It is possible to link two or more cycles together by making the output of one the input, or allosteric regulator, of the next. There is a question of matching here: the frequency and amplitude of an input cycle must be within the useful range of the next cycle in the series. If it is then systems of considerable complexity may be built. Sequential cascades can generate large amplifications, for example[22] (Fig. 2*d*), or create circuits that perform specific logical functions. The reaction mechanism shown in Fig. 2*e* would function as a boolean AND gate (Fig. 2*f*), and others can be devised that perform like OR, XOR and NOT gates[13]. Evidently this process may be continued indefinitely, so cyclic reactions could be put together to perform highly complex logical or computational functions. Indeed, it has been shown by formal analysis that cyclic chemical reactions can be used to construct a bistable system with similar properties to a McCulloch–Pitts neuron[23]. Sets of chemical neurons may then be linked through coupling reactions to build logic gates and a

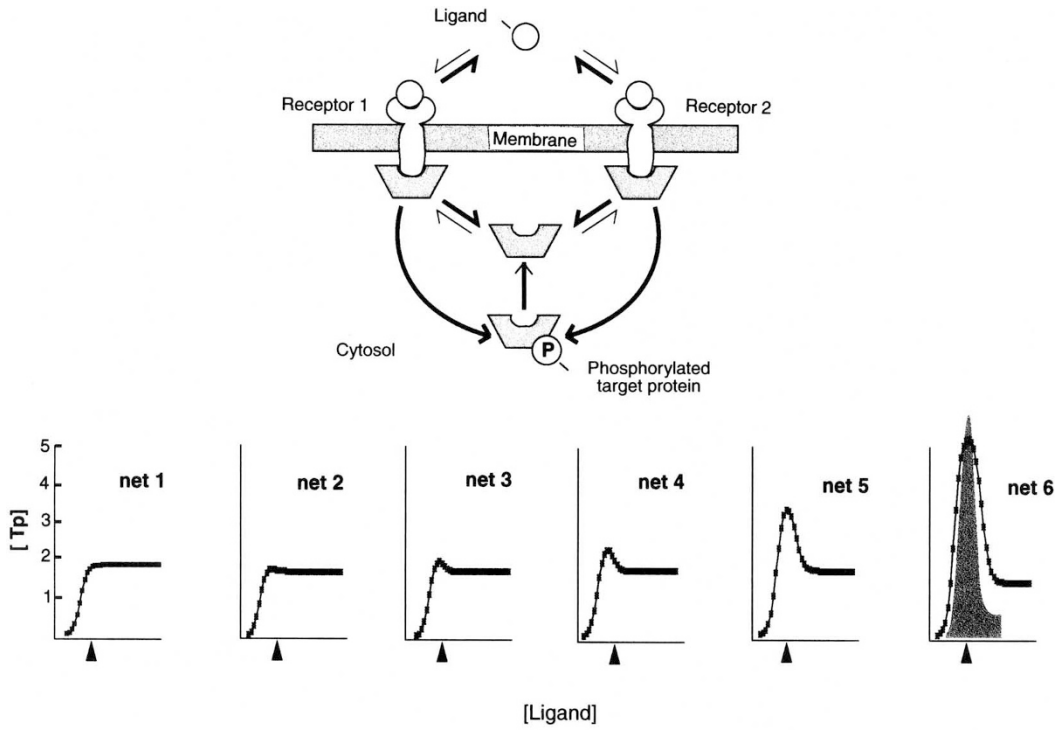

FIG. 3 Training a protein circuit by simulated evolution[25]. *a*, Model circuit built from two membrane receptors and an intracellular target protein. The two receptors, initially identical, have a binding site for the same extracellular ligand. When complexed with the ligand, the receptors bind to the target protein in the cytoplasm and phosphorylate it. The performance of the circuit, modelled by computer simulation, is specified by the seven rate constants shown as bold arrows. *b*, Optimization of the protein circuit to produce a specified input/output relationship. The circuit shown in *a* was assigned starting values of rate constants and the amount of phosphorylated target protein (output) measured for a range of ligand concentrations. This gave the dose–response curve net 1. Random changes in the rate constants were then made after which the response of the new circuit was again tested over a range of ligand concentrations. If the new circuit gave a dose–response curve closer to the desired one (in this case a response curve with a maximum at a specified concentration of ligand indicated by arrowheads) then this circuit was adopted in place of the preceding one. Outputs of a succession of new circuits, each closer to the target performances, are shown in net 2 to net 6. The target dose–response curve for the series is indicated by the shaded curve in net 6. This above simulation, run for a variety of different starting parameters, always achieved the same kind of solution, in which the two receptors had different affinities. The high-affinity receptor phosphorylates the target protein at a high rate, whereas the low-affinity receptor binds to the target protein but does not phosphorylate it. Some specific pairs of high- and low-affinity receptors in actual cells may have arisen by a similar process during evolution.

finite-state machine that generalizes to a universal Turing machine[24].

As the number of proteins in a multicyclic system grows larger, so the task of juggling its many independent rate constants and concentrations to achieve a desired input/output performance becomes increasingly difficult. Even a single cyclic interconversion may have ten or more independent variables[15]. A practical approach that involves biology is to use a form of optimization in which rate constants and binding constants are changed randomly until the system as a whole performs in a selectively advantageous way[25] (Fig. 3). It is not hard to imagine that the functional networks in living cells could have been shaped by a similar process taking place during evolution.

Protein circuits in living cells

Protein molecules can in some sense act as logical elements, and small sets of proteins can be artificially linked to perform simple computations, but whether it is useful from a biological standpoint to think of them in these terms is more debatable. Cells do not perform calculations for their own sake but as a means of monitoring and responding to their internal and external environments. Many proteins in a living cell are used to build macromolecular structures, produce movement, degrade unwanted molecules, or synthesize specific chemical species. To liken such proteins to computational devices is inappropriate, notwithstanding the fact that they are often allosteric and subject to regulation by the mechanisms discussed above. However, there also exist biochemical pathways, such as the bacterial chemotaxis pathway described below, the primary function of which is to transfer and process information. To describe the proteins in these pathways as 'chemical catalysts' or 'protein machines' is also inappropriate. The most useful analogy in this case is that of neuron-like elements that use a weighted sum of their inputs to produce one or more outputs. Of course, this analogy, like the others, is artificial and does not concern the cell. In a living cell the molecular interactions that regulate chemical catalysis and directed motion merge seamlessly with those involved in the transmission of information.

What kinds of protein-based circuit can be considered 'computational' or 'information processing'? An attempt to answer this question uncovers a plethora of physiological responses and biochemical pathways that for the most part are incompletely understood at the molecular level. Responses of living cells to their environment demonstrate their sophisticated processing of

environmental signals, and the multitude of intracellular signalling pathways offers abundant opportunities to construct computational circuits. A typical vertebrate cell has (at a guess) several hundred different channels and receptors on its surface, dozens of different G proteins and second messengers, and hundreds of different protein kinases. These elements control the concentration or activity of many thousands of target proteins in the cell (recent estimates suggest that as much as one-third of the proteins in a eukaryotic cell are phosphorylated, for example). In only a few instances are the molecular interactions of local circuits known in sufficient detail to make a quantitative prediction of performance.

Two of the most common responses of living cells to external stimuli are amplification and adaptation. Amplification enables a cell to transform faint and ephemeral stimuli into substantive biochemical changes; adaptation enables it to measure relative rather than absolute changes and hence respond over a wide range of input stimuli. The cyclic interconversion of proteins already mentioned can generate large amplifications both in the absolute number of output molecules produced for each input molecule (or input stimulus), and also in the rate of change of the output compared to the input[26]. Absolute amplification occurs, for example, when a single molecule of a hormone binds to a membrane receptor and, by triggering a protein kinase or activating a G protein, produces a change in many target molecules. Huge amplifications can in principle be achieved by cascades of cyclic interconversions coupled together: bleaching of a single molecule of rhodopsin in the membrane of a vertebrate photoreceptor, for example, can cause hydrolysis of 10^5 molecules of cyclic guanosine monophosphate (cGMP)[27,28]. In general, however, only a limited degree of amplification is desirable, and chains of signals commonly use both negative amplification (diminution) as well as positive amplification[26]. The other kind of amplification (in which small percentage changes in input produce larger percentage changes in output) can be obtained, as already mentioned, by the cooperativity of individual proteins, by coupling cyclic systems together or by enzymes that work close to saturation with respect to their substrates.

Adaptation lags behind the stimulus, following a time-course that approximates to the first differential of the input stimulus. Various mechanisms are used by cells to produce adaptation, including the reduction in number of membrane receptors, their desensitization, or the inhibition of one of the enzymes in the amplification cascade. A system in which both amplification and adaptation are sufficiently well understood at the molecular level to analyse it in terms of molecular circuitry[29] is that of the response of coliform bacteria to chemical attractants and repellents[30]. Here the amplification is due to an intracellular cycle of protein phosphorylation combined with a highly cooperative interaction with the flagellar motor. Transmission of the signal from the receptor to the flagellar motor is extremely rapid—less than 0.1 s—and is probably limited by diffusion. Adaptation, by contrast, is a slower process that requires several seconds to complete, and is controlled by the catalytic rate of two enzymes, one that methylates the membrane receptors and the other that removes these methyl groups. Methylation of the membrane receptors alters the strength of the signal they pass to the phosphorylation cycle (Fig. 4). This relatively simple circuit controls the swimming of the bacterium, enabling it to respond in an informed way to its chemical environment. It exemplifies how, in unicellular organisms, proteins act in place of nerve cells to control behaviour.

Bacterial chemotaxis illustrates other common features of protein circuits in living cells, such as their ability to integrate multiple inputs. Thus, in a complex chemical environment, bacteria add the influences of different attractants and repellents to determine their swimming behaviour[31]. Similarly, in the complex milieu of a developing embryo, the tip of a growth cone uses multiple cues in the extracellular matrix to select the path along which it grows[32]. The action of cytokines

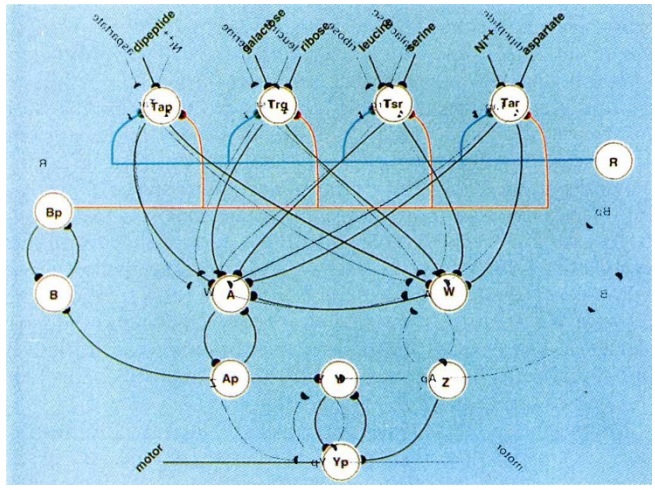

FIG. 4 Protein 'circuit' mediating the chemotactic response of coliform bacteria[30]. Tap, Trg, Tsr and Tar are membrane receptors, each of which binds to specific attractant and repellent molecules. Binding causes a signal to pass across the plasma membrane and to change the level of phosphorylation of an intracellular protein kinase, CheA (A). The transduction mechanism involves CheW (W) which, together with CheA, forms a complex associated with the cytoplasmic domain of the receptor. Phosphate groups are passed from CheA to CheY (Y), which is also believed to be part of the receptor complex, and the phosphorylated species CheYp (Yp) then diffuses to the flagellar motor. There it changes the direction of rotation and hence modulates the swimming of the cell so that it moves towards the attractant or away from the repellent. CheZ (Z) is a phosphatase that removes the phosphate group from CheYp, thereby allowing rapid changes in phosphorylation level. Two additional Che proteins add a capacity for adaptation that is essential for the chemotactic response. CheR (R) and the phosphorylated form of CheB (Bp) are enzymes that methylate (blue lines) and demethylate (red lines), respectively, the membrane-bound receptors. Methylation counteracts the effect of the attractant or repellent by changing the strength of the signal relayed by the receptor and works more slowly than the phosphorylation cascade. The level of methylation is continually adjusted according to the recent experience of the bacterium through phosphorylation of the methylesterase CheB by CheAp. The number of methyl groups per receptor (up to 4) changes slowly with the level of stimulation of the bacterium, producing a response that is proportional to the rate of change of the stimulus. This small network of proteins serves to measure the concentration of specific chemicals in the cell's environment, to integrate, amplify and determine the rate of change of these concentrations, and then to transmit the results of these computations to the flagellar motor. The system thus works in an analogous fashion to a group of nerve cells controlling a behavioural response in a multicellular organism. Note that at any instant of time the concentrations of the various phosphorylated proteins encode a representation, or 'memory', of the chemical environment in the past few seconds encountered by the bacterium.

on target haematopoietic cells is also characterized by pleiotropy and redundancy[33]. Integration in these and other systems rests ultimately on the convergence of signals into a single signalling pathway and the generation of an activity or concentration that is a graded function of the combination of inputs. Perhaps the most astonishing evidence of this combinatorial capacity is found on the boundary of our subject (as it involves DNA molecules) in the regulation of DNA transcription in eukaryotic cells. A typical gene in a multicellular organism requires the assembly of a transcriptional complex composed of enzymes, transcription factors and gene regulatory proteins[34]. Because these components are drawn from a very large pool of candidates, an extremely large number of different transcriptional complexes, each with a different 'blend' of proteins, is possible.

What other computational processes do protein circuits perform? A capacity for timing is evident at many levels: in the kinetic proof-reading that enhances the fidelity of protein trans-

lation (controlled by the slow hydrolysis of GTP by a protein); the detection of coincident signals[35]; and the orderly progression of a cell through its cycle of division, governed by cascades of protein phosphorylation[36]. Networks of proteins can also store information, not only as part of the dedicated apparatus that reinforces the effect of individual synapses, but also in a more general sense. The imprint of the environment on the concentration and activity of many thousands of molecules of the cell is in effect a memory trace. In contrast to the permanent information encoded in DNA, however, it is a 'random access memory' containing ever-changing information about the cell's surroundings. Developing organisms are shaped by protein circuits that carry signals not only from the plasma membrane to the nucleus of individual cells, but also act as diffusible inductive signals from one cell to the next.

Many aspects of cell behaviour display a capacity for information processing that is independent of the genome and hence, presumably, is controlled by sets of proteins[37]. The oriented growth of cilia on the surface of a *Paramecium*; crawling of cells over surfaces; replication of centrioles and subdivision of the cortex in a *Drosophila* blastoderm; these are all processes that take place independently of short-term control by DNA. The electrical response of excitable cells offers another entire repertoire of input/output responses, such as integration, amplification and the encoding of inputs into a series of signals of defined frequency[9]. Although these responses are usually measured by means of ionic currents (and represented symbolically as electrical circuits), we know that they are produced by sets of allosteric proteins embedded in the plasma membrane that respond to chemical and electrical stimuli. Following this line of thought, it must be admitted that the entire short-term behaviour of any organism depends on circuits of proteins which receive signals, transduce them and relay them to neighbouring cells, occasionally producing a permanent change in the synaptic connections.

Protein-based neural networks

This review of the many parallels between protein molecules and nerve cells is not yet complete. The representation of nerve cells as symbolic devices such as perceptrons presaged the development of neural networks: computer-based models used to simulate pattern recognition and other cognitive tasks[12]. Because proteins also integrate inputs and produce outputs it seems inescapable that the highly interconnected network of protein-based pathways in living cells will share some of the properties of neural nets[24,38]. Thus we expect this cell network to recognize combinations of environmental influences and to encode abstract features of its environment in the activity of particular proteins acting as 'hidden' units. Responses of this protein-based neural network may be expected to be robust and resistant to damage. The training of the network, which in the case of a computer-based network is achieved through the presentation of a training set of examples, would, in the case of living cells, be achieved by a darwinian process of random change and selection. Addition of new elements by gene duplication followed by independent variation would provide increasingly large networks able to respond appropriately to complex combinations of extracellular signals.

Curiously, in one respect the mathematical formalism of artificial neural networks is a more accurate approximation for networks of protein molecules than for networks of real neurons. The processing of electrical signals by a nerve cell is considerably more complex than the static sigmoidal nonlinearity of individual allosteric proteins. Many other features of biochemical signalling pathways, however, have no counterpart in most conventional computer-based neural nets[38]. Thus the architecture of the cytoplasmic network lacks well-defined boundaries, and individual units are not all equivalent in performance. The timescales of interactions between 'units' also varies enormously.

Arguably, the most important defining characteristic of protein-based neural networks is that they are governed by diffusive processes. Signals pass by means of physical contact between molecules, and their dispersion through the cytoplasm is limited by the random thermal motion of molecules. For small molecules and ions, diffusion through a cell is rapid: a molecule such as cyclic AMP can reach any part of a mammalian cell in a tenth of a second. Proteins are larger and diffuse more slowly, and are often impeded in their progress by associations with other components. Indeed the crowded conditions within a living cell force many proteins together in associations not seen *in vitro*[39]. Many steps in signal transduction consequently take place between protein molecules that are in physical contact, moving rapidly through the multimeric structure ('signalosome') by means of propagated allosteric changes or by internal catalytic changes. In the chemotactic relay system of coliform bacteria discussed previously, for example, a complex of proteins that includes CheA, CheW and CheY is associated with the cytoplasmic domain of individual membrane receptors[40]. Such complexes evidently work as computational cassettes that produce a stereotypical response to a specific input. Note that, because the links between individual protein molecules in such oligomeric assemblies can be modified by evolution, proteins are still the fundamental 'units' of computation from a neural network standpoint. It is just that these units are clustered spatially and limited in the extent of connections they make to other modules: signalling between signalosomes is presumably mediated by freely diffusing elements, especially second messengers of low molecular mass. □

Dennis Bray is in the Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK.

1. Stryer, L. *Biochemistry* (Freeman, San Francisco, 1995).
2. Johnson, L. N. & Barford, D. *A. Rev. Biophys. struct. Biol.* **22,** 199–232 (1994).
3. Hanson, P. I., Meyer, T., Stryer, L. & Schulman, H. *Neuron* **12,** 943–956 (1994).
4. Anwyl, R. *Curr. Biol.* **4,** 854–856 (1994).
5. Pawson, T. & Schlessinger, J. *Curr. Biol.* **3,** 434–442 (1993).
6. Saltiel, A. R. *FASEB J.* **8,** 1034–1040 (1994).
7. Morgan, D. O., Kaplan, J. M., Bishop, J. M. & Varmus, H. E. *Cell* **57,** 775–786 (1989).
8. Roach, P. J. *FASEB J.* **4,** 2961–2968 (1990).
9. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1992).
10. McCulloch, W. S. & Pitts, W. *Bull. math. Biophys.* **5,** 115–133 (1943).
11. Minsky, M. & Papert, S. *Perceptrons* (MIT Press, Cambridge, MA, 1969).
12. Rumelhart, D. E., McClelland, J. L. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition* (MIT Press, Boston, MA, 1986).
13. Arkin, A. & Ross, J. *Biophys. J.* **67,** 560–578 (1994).
14. Shacter, E., Chock, P. B. & Stadtman, E. R. *J. biol. Chem.* **259,** 12252–12259 (1984).
15. Stadtman, E. R. & Chock, P. B. *Proc. natn. Acad. Sci. U.S.A.* **74,** 2761–2765 (1977).
16. Goldbeter, A. & Koshland, D. E. *Proc. natn. Acad. Sci. U.S.A.* **78,** 6840–6844 (1981).
17. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348,** 125–132 (1990).
18. Hepler, J. R. & Gilman, A. G. *Trends biochem. Sci.* **17,** 383–387 (1992).
19. Boguski, M. S. & McCormick, F. *Nature* **366,** 643–654 (1993).
20. Hall, A. *Science* **264,** 1413–1414 (1994).
21. Gilman, A. G. *A. Rev. Biochem.* **56,** 615–649 (1987).
22. Chock, P. B. & Stadtman, E. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 2766–2770 (1977).
23. Okamoto, M., Sakai, T. & Hayashi, K. *Biol. Cybern.* **58,** 295–299 (1988).
24. Hjelmfelt, A., Schneider, F. W. & Ross, J. *Science* **260,** 335–337 (1993).
25. Bray, D. & Lay, S. *Biophys. J.* **66,** 972–977 (1994).
26. Koshland, D. E., Goldbeter, A. & Stock, J. B. *Science* **217,** 220–225 (1982).
27. Lamb, T. D. & Pugh, E. N. *J. Physiol. Lond.* **449,** 710–758 (1992).
28. Stryer, L. *Cold Spring Harb. Symp. quant. Biol.* **53,** 283–294 (1988).
29. Bray, D., Bourret, R. B. & Simon, M. I. *Molec. Biol. Cell* **4,** 469–482 (1993).
30. Parkinson, J. S. & Kofoid, E. C. *A. Rev. Genet.* **26,** 71–112 (1992).
31. Berg, H. C. & Tedesco, P. M. *Proc. natn. Acad. Sci. U.S.A.* **72,** 3235–3239 (1975).
32. Lin, D. M. & Goodman, C. S. *Neuron* **13,** 507–523 (1994).
33. Kishimoto, T., Taga, T. & Akira, S. *Cell* **76,** 253–262 (1994).
34. Tjian, R. & Maniatis, T. *Cell* **77,** 5–8 (1994).
35. Anholt, R. R. H. *Trends Neurosci.* **17,** 37–41 (1994).
36. Murray, A. & Hunt, T. *The Cell Cycle* (Oxford Univ. Press, Oxford, 1993).
37. Bray, D. *Cell Movements* (Garland Publishing, New York, 1992).
38. Bray, D. *J. theor. Biol.* **143,** 215–231 (1990).
39. Zimmerman, S. B. & Minton, A. P. *A. Rev. Biophys. struct. Biol.* **22,** 27–65 (1994).
40. Gegner, J. A., Graham, D. R., Roth, A. F. & Dahlquist, F. W. *Cell* **70,** 975–982 (1992).