Part I

Systems Theory and Cooperative Bacteria
Chapter 1

Systems thinking in biology

Robert M. Seymour

1.1 Introduction

The aim of this volume is to bring those interested in the periodontal diseases up to date with advances in related areas of biology and physiology as they pertain to the cellular and molecular foundations of these complex conditions. A further aim is to introduce the reader to the major advance in biology in the 21st century: systems biology. Systems biology is a relatively new movement in the biological sciences that aims to reverse the decades-old paradigm of molecule-based research in which more and more detailed properties of smaller and smaller components of biological systems are isolated and dissected. While certainly not undervaluing the spectacular successes of this relentlessly reductionist approach, systems biology is pursuing a reverse, integrationist agenda, in which whole-system properties are ‘built-up’ from smaller component processes. These emergent, whole system, properties are considered the most significant biological phenomena, and the ones most likely to yield payoffs in areas such as clinical application and drug discovery/design. Systems biology is inevitably a multidisciplinary endeavour, involving the active participation not only of experimental biologists and medically orientated scientists but also physicists, mathematicians, engineers and computer scientists. The close integration of mathematical or computer model-building with hypothesis generation and experimentation is a hallmark.

Here we review some characteristic themes from this young, but already vast, research domain. It is the mainstream future of the biological sciences.

1.2 The molecular revolution

The molecular revolution that began in the 1950s has, by definition, revolutionised the biological sciences and still occupies the vigorous mainstream of research effort in almost all the major areas of the life sciences. This effort has resulted in the complete sequencing of the genomes of a growing number of organisms (see Chapters 13–15 for details on genome sequencing of oral bacteria), identification of proteins, their structure and functional properties, and their interactions in both intra- and inter-cellular signalling pathways, the mapping of metabolic processes, and the many knock-on effects of these developments for drug discovery, clinical applications and comparative methods in ecological and epidemiological studies.

This ongoing push, which continues to expand on the back of ever-improving technologies, such as DNA microchips, is generating an ever-increasing tidal wave of data. The major analytical response to this tsunami has been the development of the discipline of bioinformatics – in essence the systematic application of statistical and computational methodologies for data mining and data curation. Again, this is a thriving research area that continues to offer significant rewards to biologists (see Chapters 2 and 13–15). In addition, data mining of bacterial communities in the oral cavity is a rapidly growing area, which is described in detail in Chapter 4.

Nevertheless, the successes gained and momentum built up by the molecular revolution have tended to swamp some doubts about where, ultimately, all this reductionist effort will lead. Although there are still university departments named, for historical reasons, ‘Anatomy’ or ‘Physiology’, these have long since ceased to have any contact with classical, whole-organism biology, and nowadays are staffed almost exclusively by scientists trained in the reductionist methodologies and pursuing a reductionist research agenda. Doubts persist about what it is, in the end, we are
trying to understand about organisms. The answer must surely be something like: We want to know what makes them tick – how and why do they do what they do? This very complex question has many layers of meaning, and many avenues of response. Is the reductionist, molecular focus current in the biological sciences today in danger of not seeing the wood for the trees? Has the turnover from whole-organism biology to the total dominance of the molecular focus resulted in the loss of much potentially valuable expertise?

1.3 Mathematical models in biology

One doubt about the molecular focus has been its tendency to emphasise structure over function, and the piecemeal over the big picture. Methodologies, such as in vitro cell cultures, inevitably have tended to work with isolated biological components, often in artificial environments very remote from anything encountered in vivo, and used to identify specific molecular components involved in simple input–response structures. A picture of functional properties of cells is then inferred from the results of many such component-by-component analyses. This approach has led to, for example, an appreciation of many of the details, and the impressive level of complexity found in intra-cellular signalling pathways, as illustrated in Fig. 1.1.

Nevertheless, physiology has never died. Functional, dynamic properties of large-scale systems have been investigated successfully in parallel with the developing molecular revolution. Much of this work has been achieved with the judicious use of mathematical models. The systematic use of mathematical models in disparate areas of biology ranging from enzyme kinetics to ecology and population genetics has a long history (Murray, 2001). However, much of this is concerned with equilibrium properties of the modelled systems, rather than specific dynamic features, for example Lotka–Volterra-type models of ecosystem dynamics, and Fisher–Wright models in population genetics (Ewens, 1979). A classic and paradigmatic example of a non-equilibrium process model is the Hodgkin–Huxley equations describing the development and transmission of a membrane potential along the squid axon (Hodgkin and Huxley, 1952). Much successful mathematical modelling has built on this paradigm example, which concerns itself with electrophysiological phenomena. Similarly, the investigation of intra-cellular calcium dynamics is a significant area of research that has benefited from the extensive development of mathematical models (e.g. Sneyd et al., 1993; Höfer, 1999; Kummer et al., 2000). Extensive summaries of mathematical models in various areas of biology can be found in Keener and Sneyd (1998) and Murray (2001).

These modelling efforts, like the molecular revolution, have largely proceeded piecemeal, focusing on small, isolated systems, reflecting the prevailing implicit assumption that if we collect enough detail about individual structures and processes, a whole view of the ‘wood’ will emerge. However, over the past decade or so there has been a growing realisation amongst biologists and researchers in related disciplines that, in spite of the power and sophistication of bioinformatic tools, the huge and growing datasets now available are no longer amenable to researchers’ collective intuition, due to their scale and depth. Furthermore, this scale and depth makes the systematic inference of functional properties of whole biological systems of interest extremely problematic due to the presence of feedbacks and other non-linear interaction effects.

1.4 From structure to function: systems biology

Living systems are maintained by the continuous flow of matter and energy, and thus any biological system will inevitably be a subsystem of a larger one. The biologist therefore typically has to deal with an open, multi-level and multi-component system, the perceived nature of which evolves with our increasing understanding. A key feature of such a system is the interactions (or coupling in engineering terminology) among its components, in which a variety of spatial and temporal scales may exist. These interactions may be strong or weak, unidirectional or bidirectional, depending on the current state of the system, and often generate emergent properties through nonlinear interactions. Some of these multi-level, hierarchical features are illustrated in Fig. 1.2.

For example, to begin to represent a cell and its wide range of functions, we have to integrate individual models for relevant gene expression, metabolic and signalling pathways, as well as the
Levels of complexity

<table>
<thead>
<tr>
<th>Level</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene networks</td>
<td>Microarrays, gene expression modules, siRNA gene network models</td>
</tr>
<tr>
<td>Protein networks</td>
<td>Mass spectrometry, in vivo labelling, confocal microscopy, protein network models</td>
</tr>
<tr>
<td>Intracellular signalling and metabolomics</td>
<td>Intracellular location and concentration of signalling molecules, spatial analysis with CARS microscopy, models of interacting processes</td>
</tr>
<tr>
<td>Cells as units</td>
<td>Membrane receptors, protein interactions, signal transduction, cytoskeleton trafficking, models of individual processes</td>
</tr>
<tr>
<td>Multicellular systems – organs/organisms</td>
<td>Cell–cell communication, hormones, cytokines, integrated models of complete processes</td>
</tr>
</tbody>
</table>

Fig. 1.1 The complexity of the biological hierarchy revealed over the last half century by the molecular revolution. (a) A schematic view of a eukaryotic cell showing some intracellular pathways which transmit information from the cell membrane to the nucleus. (b) Levels of complexity: the major components of the biological hierarchy and their two-way information-and-control exchanges, together with some associated methodology used in their investigation.
associated biophysical processes for intracellular, extracellular and intercellular transport, etc. At the next scale up, a multi-cellular, tissue or organism level model has to be formed by integrating different kinds of cell functions and cell–cell communication in their intra- and extra-cellular environments. This is typical of the ‘bottom-up’ approach implicit in the reductionist, molecular revolution paradigm, and contrasts with the classical, physiological ‘top-down’ approach, which tends to start from the system as a whole.

In response to this now very apparent complexity, the last 10 years or so have seen the growth of a new paradigm in the biological sciences: systems biology. This is, broadly speaking, a move to bring a systems engineering perspective to the analysis and understanding of the integrated functioning of biological systems (Anand et al., 2000; Csete and Doyle, 2002; Alon, 2007; Boogerd et al., 2007). Thus, complex biological systems are conceived as being ‘engineered’ by natural selection in order to perform whatever functional roles they are perceived to have. Such ‘natural engineering’ is of course less systematic than the engineering that lies behind the design of complex human artefacts such as an automobile or jumbo jet. Nevertheless, there is anticipation that natural selection will have discovered and utilised some basic engineering design principles, for example the use of modular design structures, robustness through redundancy and the deployment of nested control systems. If these design principles can be uncovered in natural systems, then a deeper, more functionally based understanding can be achieved than is possible with the piecemeal methods of the molecular revolution.

The main ‘ideological’ goal of systems biology is to bring approaches from engineering disciplines into the problem-solving methodology of biologists. These approaches are not grouped around the peculiarities of any particular system (e.g. a cell type or organ such as the heart or the liver) but around individual biological problems whose treatment can be useful for other, possibly as yet unforeseen, problems. Biological systems on whatever hierarchical level of interest (be they organisms in populations, organs in organisms, cells, transcription factors, genes or single molecules) are using the common criteria of modularity, process and functionality. These can be formulated in the technical terminology of ‘modules’ and ‘protocols’ and find mathematical expression in control theory (Sontag, 1998), or conditional probability distributions and graphical models used to describe networks. Such concepts are critical if datasets are to be analysed under functional criteria. The emphasis then is on ‘generic’ properties – features that apply across a wide range of systems and circumstances. A schematic of this new analytical paradigm, and methods associated with it, is shown in Fig. 1.2. Paradigmatic issues that arise from this perspective are:

---

**Fig. 1.2** Schematic of components of a complex biological system from the point of view of ‘systems thinking’. This emphasises networks, dynamics, control structures, models and integrated behaviour of the whole system emerging from lower-level interactions. The whole model construct must be both informed by, and capable of, validation against data using methods relevant to the various hierarchical level (as indicated in Fig. 1.1b).
At what (conceptual) hierarchical level are biological systems predictable, and what functional features are robust enough to be predictable?

How does this robustness and predictability ‘emerge’ from possibly non-predictable lower level phenomena?

Are these predictable features generic properties of large classes of systems, or are they non-generic, highly optimised, and therefore highly evolved for specific roles in specific contexts?

The challenge posed by these objectives is considerable, and will require the efforts of many disciplines: experimental and theoretically minded biologists, engineers (especially control engineers), computer scientists (software engineers, bioinformaticians) and mathematical modellers, as well as the utilisation of the fast-developing in vivo imaging technologies coming out of the biophysics community. Some of the challenges to be faced in this endeavour are:

- To link the biological task of data collection and analysis with mathematical and computational modelling to enable the construction and analysis of integrated systems.
- To synthesise individual biological processes analysed with widely varying techniques across a range of scales into a constituent whole.
- To generate an experimental and theoretical scaffold that integrates the function of gene and protein networks with intra-cellular signalling and inter-cellular interactions mediated by direct cell–cell communication, surface membrane receptors and soluble factors such as cytokines and hormones.
- To elucidate the molecular, cellular and evolutionary bases for coordinated function of multicellular systems using data derived from genomics, proteomics, protein–protein interactions, transcriptional regulation, intracellular signalling mechanisms and metabolomics through to cell–environment and cell–cell interactions.

1.5 Systematic simplification

The ‘bottom-up’ approach to systems biology requires that biological modelling is scaled from relatively small components to the whole system level (Fig. 1.2). This can produce extremely complex models. The major success to date of this approach is the systems model of the heart (Noble, 2002), which has developed over several decades, building on a modification of the Hodgkin–Huxley model (Hodgkin and Huxley, 1952), and proceeding with gradual and well-supported increases in complexity. Currently, the model involves some hundreds of equations, largely representing electrophysiological and electromechanical processes, and is linked to sophisticated computer visualisations, particularly of solid geometries. It has more than proved its worth in aiding the understanding of cardiac arrhythmia, with consequences for drug design and testing (Noble and Colatsky, 2000).

Despite its complexity and the huge effort it has taken to build, the model covers only a part of the heart’s mechanical, electrical and chemical properties. This reveals the scale of the systems biology challenge. The model has been the seed for the Physiome Project (Hunter et al., 2002), which collects and catalogues physiological models and supports community access to these models. The project also provides web-accessible databases of biological data that potentially can be linked to models. In a similar vein is the UCL Beacon Project, which aims at systems modelling of liver physiology and glucose homeostasis.

Nevertheless, with few exceptions most mathematical models are stand-alone models of isolated biological processes, with limited ambition. Sometimes these models are provisional, and embed contested hypotheses concerning structure or function. One key reason for such small-scale modelling is the often insuperable difficulty of validation. However, some very successful models have been ambitiously large scale and also well validated. An important example is the model bacterium of Denis Bray and colleagues, which successfully simulates chemosensitivity and the operation of flagella (Bray et al., 1998; Levin et al., 1998). Thus, chemotactic signalling in Escherichia coli has the property that bacterial cells can detect chemical gradients ranging over five orders of magnitude in concentration, and Bray’s model showed how this low-threshold response and wide dynamic range can be achieved with a collection of single surface receptor molecules that communicate laterally with one molecule influencing its neighbours as a function of its own activity. The extent of this lateral activity spread can adapt to external circumstances.
Given the ambition of systems biology to capture functional properties at higher levels that emerge from molecular mechanisms, it will of necessity have to deal with large, composite and complex models, built systematically from smaller components spanning several hierarchical levels (Fig. 1.2). Such constructs may perform well as simulations of the real system, but to do so imposes an enormous burden on data gathering. Nevertheless, some successes are already apparent in this approach when coupled with large-scale, dedicated, high-throughput experimental work, for example in imaging (Nelson et al., 2002) and metabolomics (Maher et al., 2003).

Representing all aspects of a biological system in the smallest possible detail is clearly not feasible. In modelling, we cannot hope to recreate the world as an isomorphic in silico image of itself (Finkelstein et al., 2004) therefore judicious simplification is required in model construction. This is particularly true when attempting to link different processes at different spatio-temporal scales, such as gene and protein networks. For example, to represent biochemical networks involving many different proteins, we could model the interactions between proteins using simple stimulus–response functions, while retaining the complexity of the interaction network. Alternatively, we could focus on a few proteins and model the complex transformational processes involved in their interaction in great detail. Such simplification choices have at least two major facets:

- Choosing a modelling scheme that allows sufficient descriptive power at the relevant conceptual level(s) (Fig. 1.2), while having the flexibility to link to other component models and allowing contextualisation in terms of known or obtainable data.
- Choosing a level of detail in terms of the number and complexity of interactions to model explicitly within the descriptive scheme; whether and how to model space, and determining the relevant timescale(s).

Some work has been invested in analysing ‘systematic simplification’ procedures, in particular using a common simplification methodology on all parts of the system (e.g. Hetherington et al., 2006). This approach has had significant success in metabolic control analysis, in which useful (equilibrium) models have been constructed using only the stoichiometries of constituent reactions, interpreted using flux balance analysis or elementary mode analysis (Fell, 1997; Schuster et al., 1999; Edwards et al., 2001; Holzhutter, 2004; Hornberg et al., 2007).

One popular systematic simplification strategy is to replace smooth input–output response profiles with two-state (OFF–ON) functions. Sometimes three states are used (DOWN–0–UP). This is common in modelling gene networks in which a gene produces a transcription factor that binds to other genes (and possibly itself). The state of the receiving gene is then determined by ‘integrating’ all its arriving inputs (commonly, using a weighted summation with a threshold response). Examples of the use of this modelling strategy include investigating the evolution of the sex-determination gene network in Drosophila (MacCarthy et al., 2003), and the role of perturbation experiments in reverse engineering the structure of gene networks (MacCarthy et al., 2005). In a similar vein, such methods have been applied to biochemical networks (Glass and Kauffmann, 1973), and to network interactions in the innate immune response (Kaufman et al., 1985; Kaufman and Thomas, 1987), and also analysed more abstractly (Snoussi and Thomas, 1993; Thomas, 1998; Thomas and Kaufman, 2001).

A detailed statistical analysis of the sensitivity of the resilience of some features of the output of a complex model under similar systematic simplification procedures is given in Hetherington et al. (2006). They show that, although the detailed quantitative behaviour of the complex model is often not preserved under simplification, nevertheless, simplified models can perform well in predicting which parameters have significant control over various system properties.

1.6 Multiple stable states

Perhaps the central concept of systems biology – indeed, of complex systems more generally – is that of a network of interactions. The actors (or nodes) in these networks can be genes, proteins, cells or organisms (Fig. 1.1), and the mediating interactions can be of various kinds, including transcription factors, phosphorylation and dephosphorylation of proteins, binding of ligands to cell–membrane receptors, and changes of phenotypic response to environmental cues. The key point is that networks describe information
transfer (by whatever mechanism) between nodes (of whatever type), and this information induces changes in the properties or behaviour of the nodes, which in turn has knock-on consequences via network interactions. Most biological systems can be described in terms of one or more such networks. However, biological networks are dynamic structures, with behaviour that changes over time. They therefore exhibit adaptive behaviour in response to their external environment (Bhalla and Iyengar, 1999; Alon, 2007).

As indicated earlier, homeostatic mechanisms often allow biological systems – modelled as networks – to remain at or near equilibrium (or stationary) states. An equilibrium state is said to be (locally) stable if relevant homeostatic mechanisms will lead to a return to the equilibrium after a suitably ‘small’ perturbation away from it (a large enough perturbation can always destroy any state – indeed, the whole system). Features of stable equilibrium states are therefore of great interest in understanding system function. However, complex systems can have more than one possible stable equilibrium state, and in these circumstances it is often very unclear which state the system will assume under what conditions (Laurent and Kellershohn, 1999). Furthermore, it is possible that a system will remain in a particular stable state even after the stimulus that originally pushed it there is removed (Henderson and Seymour, 2003).

This issue of multistationarity is potentially important in relation to possible pathological states, in which one stable equilibrium may be regarded as ‘healthy’, but an alternative, less accessible, stable state is pathological, associated with disease symptoms. This point of view is developed by Henderson and colleagues in relation to the early-onset cytokine response of the innate immune system, for example in the context of rheumatoid arthritis and systemic inflammatory response syndrome (SIRS) (Henderson et al., 1998; Wilson et al., 1998; Seymour and Henderson, 2001; Jit et al., 2005). Similarly, for the adaptive immune response, the development of T-helper cell differentiation into Th-1 or Th-2 phenotypes in response to activation of T-cell receptors by cytokines – which in turn activate the transcription factors T-bet or Gata-3 – can be viewed as alternative, irreversible stable states of a network (Yates et al., 2004).

In Fig. 1.3 a simple network model of cytokine production and response by an immune system cell (e.g. monocyte/macrophage) is illustrated. This model features up to five (unspecified) cytokines that can bind to specific cell–membrane receptors. These cytokines are implicated in periodontal disease. Receptors are subject to (stimulated) internalisation, with consequent loss of bound ligand. Binding of the various cytokine ligands to their receptors results in a cell response that either increases or decreases the production of any given cytokine, depending on a threshold-type response to the (integrated) input of all the cytokines. In this model, these integrated threshold responses (which can be multiplicative as well as additive) are determined randomly, leading to random interaction networks. Different random choices of thresholds lead to different network instances. The model can result in unconstrained cytokine production – a clearly pathological outcome – or a constrained stable state, and sometimes also stable cycles. As is illustrated in Fig. 1.3, multiple stable states are possible in this model, but even with five cytokines, there are a small number of such states (no more than five), with most networks yielding only one. This suggests that multiple stable states may in fact be less frequently found than one might naively expect. Chronic pathologies arising from the immune response being stuck in the ‘wrong’ stable state are certainly possible, but perhaps to be expected only in individuals with untypical cytokine network structures.

### 1.7 Regulatory networks, modules, motifs and control structures

Networks are formally represented by nodes and edges, representing pathways of information flow between nodes. Edges may be directed, representing a specified direction of information flow, or undirected, in which case information flow is equally possible in either direction. The in-degree of a node is the number of incoming edges into the node, and the out-degree is the number of outgoing edges. In an undirected network, these concepts coincide. For a large network, the distribution of (in- or out-) degree \( d \) over the nodes of the network has been the subject of much study. For example, recent studies of transcription networks in *E. coli* and *Saccharomyces cerevisiae* have revealed a roughly exponential in-degree distribution, but a power law out-degree distribution (Guelzim et al., 2002; Regulon DB, see http://regulondb.ccg.unam.mx/).
**Fig. 1.3** Systems with multiple equilibria. (a) A model of production and linear decay of some product. The production occurs in response to some stimulus $p$ (horizontal axis). Decay is linear in $p$. Left panel: concave production response gives just one stable equilibrium (heavy point). Right panel: threshold production response function showing 0, 1 or 2 stable equilibria, depending on the slope of the decay line. (b) Dynamic trajectory of production model in response to external stimulus. Left panel: sustained stimulus leads system to equilibrium. Right panel: stimulus is removed after a time: the system does not return to its pre-stimulated state, but remains in chronic, low-level activation. (c) A schematic diagram of two cytokine production, $L_1$ and $L_2$, by a cell, showing interaction (+ = upregulation). (d) Statistical distribution of equilibria found in 100 000 randomly generated 2-cytokine networks. Constrained equilibria lead to sustained finite cytokine levels, unconstrained equilibria lead to runaway production of one or more of the cytokines. There are up to five constrained equilibria, but never more than two unconstrained equilibria.

That is, the probability that a randomly chosen node has out-degree $d$ is $P(d) \sim d^{-\gamma}$ for large $d$, where $\gamma$ is a positive constant.

Much has been claimed concerning the apparent ubiquity of power law degree distributions in biological networks – yielding so-called scale-free networks (even though this ubiquity may be somewhat illusory (Stumpf et al., 2005; May, 2006)). In particular, deep questions that naturally arise are:

- Does a network’s structure tell us anything about its function?
- How could natural selection have shaped a network’s global structure?
- Does a power law degree distribution in transcription, protein or metabolic networks confer any advantage on an organism?

For example, strong claims have been made that:

- (i) power law distributions endow a network with robustness against perturbations; (ii) networks with small average path lengths may adjust more rapidly to environmental perturbations, and (iii) this is likely to be particularly advantageous in metabolic networks in minimising transition times between metabolic states (Albert et al., 2000). However, recent work indicates that highly connected proteins in metabolic and protein interaction networks are not subject to more severe evolutionary constraints as might be expected. Also, genes whose protein products have many interactions do not have fewer duplicates in the genome (such genes would be expected to be deleterious) (Hahn et al., 2004). An alternative view, supported by this evidence, is that the evolution of protein interaction networks involves two processes: (i) gene duplications increasing the number of protein interactions, and (ii) addition and elimination of protein interactions. These processes are sufficient to generate a power law degree distribution provided: (i) the rate of interaction addition and deletion are nearly balanced; (ii) interaction turnover preferentially affects...
highly connected proteins, and (iii) some added interactions add new proteins to the network (Wagner, 2003). The important and complex issue of the evolution of ‘robust design’ in organisms and its relation to regulatory networks is analysed in detail in a recent book (Wagner, 2007).

Gene duplication and subsequent divergence is a common mechanism for the generation of diversity, even though this process does not appear to be important in all circumstances, for example in developmental mechanisms in multicellular organisms, which are largely controlled by highly conserved Hox genes acting within ‘modules’ such as arthropod limbs (Carroll, 2001). Thus, the great diversity of arthropod limb types is due to differences in the regulation of Hox genes along the anterior–posterior axis. More generally, the great variety of cell types within multicellular organisms is due ultimately to differences in gene expression, which can number in the hundreds of thousands. However, greater gene number itself is not a major driver of ‘complexity’, as, for example, measured by number of cell-types within an organism. Instead, the differences in gene expression are often controlled by a relatively small number of regulatory proteins. Thus, only about 3–5% of the proteins encoded by organisms’ genomes are transcription regulators. For example, in the bacterium Bacillus subtilis (∼4100 genes) a small number of regulators control the differential expression of several hundred genes during sporulation, while in yeast (S. cerevisiae, ∼6200 genes) a small set of transcription factors orchestrates the regulation of genes involved in cell-type differences. In metazoans, cell-type differences (e.g. between muscle and neural tissue) and body-region identity are typically regulated by a few proteins, whereas pattern formation within tissues is regulated by a larger set of proteins (Carroll, 2001). Regulatory evolution creates new combinations of gene expression and therefore enables increases in the information content of genomes, and the generative potential of development without expansion of gene number. However, we do not understand why the actual complexity realised in evolution is far less than appears to be possible genetically.

These and similar considerations have generated considerable interest in ‘modular design’ – either in structure or function or both – as a mechanism for generating diversity, robustness to component failure and adaptability (Csete and Doyle, 2002). A module is a structural/functional unit relatively separable from its surrounding structure. An important property bestowed on evolving organisms by modularity is the ability to dissociate developmental processes in one part of the body from another. Modularity facilitates change by conferring on organisms a greater ability to escape internal constraints, that is it confers robustness and evolvability (Wagner, 1996; Wagner and Altenberg, 1996). Several mechanisms can give rise to modularity: symbiosis (e.g. mitochondria in the eukaryotic cell), developmental segmentation, connectivity-sensitive growth and response to variation (Barabasi and Albert, 1999; Holland, 1999; Watson and Pollack, 2000; Lipson et al., 2002). Indeed, simple models show that modularity can arise spontaneously in evolutionary systems in response to environmental variation and selection (Lipson et al., 2002). Engineering design methods based on evolutionary simulation (genetic algorithms) could well benefit from evolving to variable, rather than stationary, fitness criteria, as a problem-independent method of inducing modularity.

Little is known about the design principles of transcriptional regulation networks that control gene expression in cells. For example, can such networks be broken down into basic building blocks? In taking a systems approach to understanding transcriptional regulatory networks, much emphasis has been placed on decomposing these networks into simpler subnetworks, often called motifs (Lee et al., 2002; Shen-Orr et al., 2002; Alon, 2007). Such motifs are sometimes construed as just small-scale patterns that occur more often than chance would predict, but also can sometimes be construed as representing specific ‘control modules’ with specific dynamical properties (Fig. 1.4). The motif structure then allows an easily interpretable view of the entire known transcriptional network of the organism. This approach may help to define the basic computational elements of many biological networks.

Figure 1.4 shows some of the proposed dynamic control motifs that are thought to be important in gene regulatory networks. Also shown is the output associated with three of these, obtained from a dynamic model of gene transcription which incorporates transcriptional delay. These are the auto-regulation loop, the two-component loop and the feed-forward loop. Each of these has very specific dynamical properties. Thus, the auto-regulation loop, in which the transcription factor produced
Fig. 1.4  Control modules and motifs in transcription networks derived from a dynamic model of transcription incorporating transcriptional delay. (a) Some control motifs which have been found in the transcriptional network of *E. coli*. Genes are squares, their protein products (transcription factors) are ovals. Dashed arrows: production of a protein by a gene. Solid arrows: binding of protein to gene leading to up- or down-regulation of the gene. (b) Dynamics associated with the auto-regulation motif. Left panels: external stimulus level of the gene. Right panels: protein-production response. Top pair: low-level stimulation produces no response. Bottom pair: high-level stimulation produces full response. (c) Similar stimulus–response graphs for the feed-forward loop. Two proteins are involved – one for each gene. A significant response occurs only when the stimulus is sustained for long enough. (d) Response graph for a two-component loop. The two protein products (one for each gene) oscillate once the system is activated (constant stimulus not shown). (e) Post-transcriptional protein modification such as that found in the NF-κB system. This can lead to cyclic behaviour: Horizontal straight line is the constant stimulus; two oscillating curves are the transcription factor and the protein product.

by a gene (in response to some background level of externally induced activation) binds to its parent gene and upregulates further production of itself, turns itself on only when background activation reaches some threshold level. In the feed-forward loop, an external protein factor activates a pair of linked genes; the transcription factor produced by the first of these binds to the second and upregulates (or downregulates) the protein produced by the second gene, with some time delay between the two responses. In the two-component loop, the transcription factors produced by a pair of genes bind to the non-parent gene, enhancing (or suppressing) its output. This leads to cyclic dynamics.

These and other regulated responses are thought to control the coordinated output of large gene networks composed of linked dynamic motifs.

Dynamic control structures of course can operate on many different levels in the biological hierarchy (Fig. 1.2) and indeed on more than one level simultaneously. Thus, in the examples of Fig. 1.4, both proteins and genes are active in control. However, this is construed as taking place in the nucleus over a small spatial range. Intracellular protein signalling pathways are more extensive in spatial scale, but generally operate on a fast timescale compared to that of gene transcription. One well-known interaction between the
time-scale of gene transcription and that of protein–protein signalling occurs in the NF-κB pathway. In this pathway, an initiating event, such as the binding of tumour necrosis factor (TNF) to its cell-surface receptor, leads to a chain of protein modifications culminating in the detachment of NF-κB from its inhibitor I-κB, which is degraded via ubiquitination. NF-κB is then free to migrate into the nucleus, where it initiates transcription of relevant genes producing, amongst other products, the inhibitor protein I-κB. This latter then migrates out of the nucleus, and proceeds to bind to free NF-κB, thereby inhibiting its activity (Hoffmann et al., 2002; Lipniacki et al., 2004; Nelson et al., 2004; Krishna et al., 2007). Although the details of all the steps in the NF-κB network are extremely complex, and not fully understood, nevertheless, this basic feedback loop is known to induce cycles of production and suppression of NF-κB. This is a now classic example of dynamic feedback control.

Control analysis of other signalling pathways is also an active area of research, e.g. the MAPK/ERK signalling pathway (Marshall, 1995; Hornberg et al., 2005a,b; Orton et al., 2005; Bluthgen, 2006; Santos et al., 2007).

1.8 Ecological models: quorum sensing and biofilms

Ecological communities have always been recognised as complex systems, in which there is a hierarchy of energy transfer from primary producers (photosynthesising plants) through herbivores, up to top predators, and down again to detritivores (Lindeman, 1942; Begon et al., 1990; Higashi and Burns, 1991). In this context, study of the dynamics of populations has a long history dating from the classic work of Lotka and Volterra in the 1920s and 1930s on pair-wise population dynamics. Interactions, either within or between species, in an ecological community can be conceptualised as one of several possible classic forms: competition (for a common resource), predator–prey; parasitism, mutualism/symbiosis, or commensalisms. The equilibrium structure of ecological systems can often be described by specifying what kinds of interactions are dominant (e.g. May, 1973).

Nevertheless, the properties and behaviour of ecological communities are still poorly understood. This is because there are so many factors relevant to a comprehensive understanding of the function of any community: the physical and chemical environment and its influence, succession and community development through time and space, mechanisms of communication (molecular or other forms of signalling) and interaction (competition, mutualism, etc.), and the role of disturbance in promoting diversity. Increasingly, the concept of a community as a network of trophic interactions, which has emergent, whole-system, properties, is coming to prominence (Patten, 1991; Kondoh, 2003; May, 2006). This is particularly true of communities of microorganisms, for example those forming gut and oral consortia in mammalian hosts, and the many examples of multi-species biofilms, which have only fairly recently become the targets of intensive research (Hooper et al., 1999; Horner et al., 2003; Handelsman, 2004; Rainey et al., 2005; IWA Task Group, 2006). The relevance for periodontal disease is obvious.

One particularly intriguing form of community interaction that has come to light through the study of microorganisms is quorum sensing. This was first demonstrated in the bacterium *Vibrio fischeri* and its host, the squid *Euprymna scolopes*, in which the bacteria colonise the squid light-organ and emit light at night, providing counter illumination that enables the host to avoid detection by its predators who attack from below in moonlit waters (Visick and McFall-Ngai, 2000; Visick et al., 2000). In quorum sensing, bacteria sense their own population density and express genes accordingly. Population density is detected by the accumulation of species-specific acyl homoserine lactones (AHLs). In *V. fischeri*, quorum sensing is used to regulate genes responsible for light emission. However, it is now known that quorum sensing is also widely used by pathogenic bacteria as a mechanism for evading the host immune response. Thus, invading bacteria switch on virulence genes only when their population density is sufficiently high, when they are in a position to overwhelm the host’s defences (Prichard et al., 2003). This kind of coordinated response – essentially through coalition formation – is certain to have significant implications for whole-system community dynamics in a wide range of contexts. For a more detailed discussion of the role of quorum sensing in oral bacterial biofilms the reader should refer to Chapter 12.

Not least of these is the interface between bacteriology and immunology through the interaction of animal – especially mammal – hosts with their
endogenous communities of ‘commensal’ microorganisms. This interaction is now thought to be highly dynamic, involving much molecular ‘cross-talk’, which, in particular, is important to normal gut development. Indeed, the gut of human neonates contains large numbers of facultative anaerobes, which decline in number during weaning as obligate anaerobes begin to take over. Over time, the gut community evolves to a stable climax community in which obligate anaerobes predominate. In parallel to this shift in microbial community composition is a marked morphological maturation of the host gut, along with changes in the host immune system (Hooper, 2005). The dynamics of microbial community development is even faster in the oral context, occurring after each toothbrushing session.

The possible role of pathogenic bacteria in the evolution of host characteristics has been hypothesised by Seymour et al. (2004). They considered the evolution of the human ABO blood-group system by constructing an epidemiological model involving a virus, which may pick up terminal glycans on its envelope during invasion and transmission between hosts (Preece et al., 2002), thereby biasing the next host immunological response, and opportunistic bacteria that adapt to the characteristics of host epithelial–mucosal surfaces. These forces generate differential responses between hosts which are able to maintain the commonly observed ABO frequencies in human populations. This model places the selective maintenance of the ABO polymorphism firmly within a larger community context.

Studies of interactions between the host immune system and microorganisms have generated some novel hypotheses. For example, one radical hypothesis has it that the evolution of the adaptive immune system in vertebrates occurred in parallel with, and in order to orchestrate, their strategy of hosting large consortia of resident microbes, presumably to exploit their nutrient-processing capacity (Ruby et al., 2004; McFall-Ngai, 2007). A system-level model of such adaptive orchestration of a commensal microorganism community has recently been developed (Seymour, 2005). Here, the host is modelled as a network of ‘sites’ in which bacteria can grow and reproduce, with variable configurations of possible migration between sites. These sites are favourable to different species of bacteria, so that, other things being equal, each site is dominated by its resident specialist and, if migration is allowed, the whole system is dominated by one specialist species. However, if a simple model of an adaptive immune response is imposed, which acts to suppress excessive growth in a site- and species-specific manner, then very diverse communities of microorganisms can be sustained, with diversity at a hierarchy of abundance scales.

Network interactions are one, rather schematic, way of representing the influence of space on community structure. Space, especially on sessile communities such as biofilms, allows the development of heterogeneous structures by preventing direct interaction between component parts of a system. This is an important mechanism for generating diversity and neutralising competitive exclusion (Johnson and Seinen, 2002). In the context of biofilm formation, there is now a huge modelling literature, both spatial and non-spatial (reviewed in IWA Task Group on Biofilm Modelling, 2006). The complex effects of spatial structure in a simple spatial, agent-based-model (ABM) of a growth–predation model is illustrated in Fig. 1.5.

Ecological communities exhibit all the system complexity of physiological systems (Fig. 1.2), even though they are not teleological in the same sense, that is physiological mechanisms are present for a purpose: as control mechanisms to maintain an organism within its domain of tolerance. For example, the glucose homeostasis system in mammals orchestrates the efficient delivery of energy to the various tissues in the face of the diverse demands that the organism encounters. In contrast, the mechanisms that sometimes maintain an ecological community in a (quasi) stable state are not selected for this high-level purpose: the stable state is an emergent property of these mechanisms and in itself has no function. Of course, the evolution of interaction mechanisms always takes place and is shaped within an ecological setting, and these mechanisms may therefore play a role in maintaining this context – the mechanisms themselves are adapted. But the whole-community system is not subject to natural selection, and hence is not adapted for any purpose.

This feature of ecological systems makes their analysis particularly difficult because our understanding cannot easily be directed by a ‘top-down’ view based on what the system is ‘for’, i.e. what its evolved function is. Instead, communities are
Fig. 1.5 A simple spatial community. (a, d) Light blue background is the substrate, represented on a $50 \times 50$ lattice. Red-coloured lattice squares indicate the presence of a base-level (pioneer) organism growing into the available space. (b, e) The red organism is a necessary substrate over which individuals on a higher trophic level can grow. These individuals move from square to square in search of the red organism. Density is shown as brown to purple on the light blue substrate, with darker colours representing higher density. The red organism is not shown. (c, f) Measure of the availability of red substrate to the higher trophic level over 100 time steps. Top row: growth of the red organism is too low for the higher-level individuals to thrive. Bottom row: vigorous growth of the red organism promotes a thriving community of the higher-level individuals.

1.9 Conclusions

The molecular revolution, beginning in the 1950s, which largely displaced classical physiology and whole-organism biology, has come full circle with the recent rise of the systems biology movement. This approach aims to reverse the trend toward ever more refined reductionism, and instead integrate the detailed knowledge that the molecular revolution has provided, to refocus on how whole system behaviour ‘emerges’ from detailed molecular information. This approach can be applied at many levels of the biological hierarchy (Fig. 1.1), although most current work tends to focus on the cell-as-system, with its major components conceived as networks of interacting components: gene regulatory networks, metabolic networks, protein signalling networks.

Systems thinking in biology is inherently multi-disciplinary, requiring the input of engineers and computer scientists as much as biologists. This is because whole system behaviour requires dynamic modelling, involving many temporal and spatial scales, stochastic effects, modular design and a hierarchy of orchestrating control structures (Fig. 1.2). The modular and object-oriented ‘design’ of many biological systems provides the flexibility for ‘plugability’, which facilitates the inter-operability of coupled components. Thus, mathematical/computational models of these systems must exhibit analogous properties:

- **multi-scaling**: temporal and spatial between different components
- **encapsulation**: information exchange determines a network of inputs and outputs between component models
- **heterogeneity**: different modelling paradigms and computational environments can be employed for different components.

Major issues for the systems modeller are: How do we decide what is the relevant scale on which to build models? What is the best strategy for handling the trade-off between quantitative models (lots of detailed data) versus qualitative models (capturing the phenomena)? These sorts of issues and their biological counterparts make the
traditional, hypothesis-driven paradigm of empirical biological research problematic. Indeed, some commentators believe that biology is in the throes of a ‘neo-Baconian crisis’, in which induction, and hypothesis generation from the ever-increasing mountain of high-throughput data, is becoming increasingly problematic.

It seems, therefore, that in the new systems biology era, the old hypothetico-deductive paradigm may be more of an obstacle to progress than an aid. Instead, a more ‘objectives’ driven approach is more appropriate, in the sense that the aim is to characterise the mechanisms underlying whole-system ‘emergent’ behaviour, and in so doing to understand the nature and scope of such behaviour. The premature formulation of often simplistic hypotheses does not help this process. In summary, systems biology should be not only a provider of models and simulations but also a source of understanding and a toolkit that complements and guides experimentation.

References


