16 Functional States

Life is a program written in DNA – Craig Venter

Chemical reactions link cellular components together to form a network. Although we can specify the chemical properties of links in biological networks, it is the way in which a multitude of such links form networks that determines phenotypic functions. These integrated network functions are also called *functional states*, and they correspond to the observed biological functions or phenotypic states that networks create. A functional state may be viewed as the outcome of the execution of the genetic program written in the DNA. In this chapter we detail the concept of a functional state of a genome-scale network and how it represents a physiologically observable state. The following chapters then describe the framework for computing functional states using the constraint-based approach.

16.1 Components vs. Systems

Components come and go Biological components all have a finite turnover time. Most metabolites turn over within a minute in a cell, mRNA molecules typically have two-hour half-lives in human cells [463], 3% of the extracellular matrix in cardiac muscle is turned over daily, and so forth. So a cell that you observe today, compared with the same cell yesterday, may only contain a small fraction of the same molecules.

Similarly, cells have finite lifetimes. The cellularity of the human bone marrow turns over every two to three days. The renewal rate of skin is on the order of five days to a couple of weeks. The lining of the gut epithelium has a turnover time of about five to seven days. Slower tissues, like the liver, turnover their cellularity approximately once a year. So a mammal that you observe today may only contain a small fraction of the same cells as the same mammal observed a year ago. Thus, the components of a biological system come and go, and their turnover takes place on multiple time scales.

However, the system remains Most of the cells that are contained in an individual today were not there just a few years ago. However, we consider the individual to be

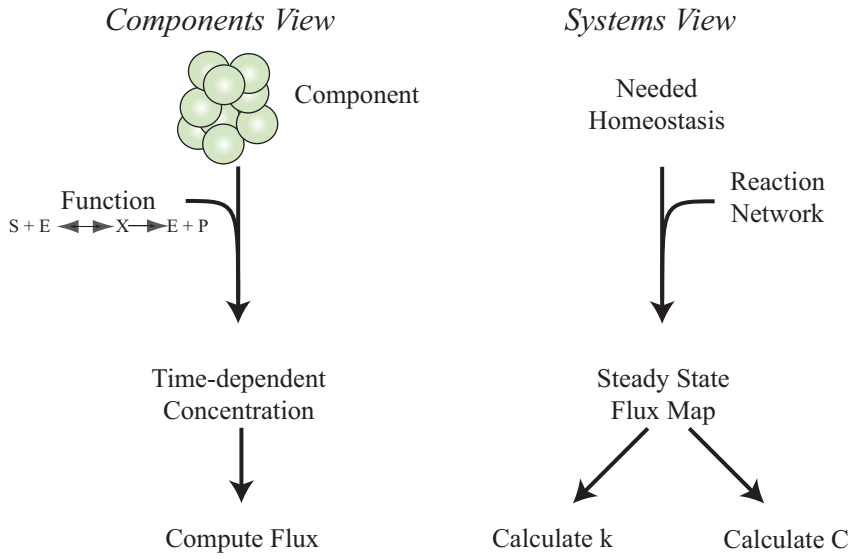

Figure 16.1 A contrast between the components view (on the left) and the systems view (on the right).

the same. Similarly, we consider one cell to be the same one a week later, even if most of its chemical components may have turned over. Thus, components come and go, but the system remains. Therefore, a key feature of living systems is how their components are connected. The interconnections between cells and cellular components define the essence of a living process.

Moving from viewing components in isolation to being a part of a system The difference between the familiar components view of a cell and its molecular biology is different from the less familiar systems view in many subtle ways. Here, we illustrate this difference in Figure 16.1.

On the left side we see the classical component point of view. When we are looking at one gene product, in this case an enzyme carrying out its function, we study this component by placing it in a beaker with its substrates and then observe the time-dependent disappearance of a substrate and the appearance of a product. The component that we are studying is the centerpiece of this experiment, and it is responsible for concentration changing in a time-dependent manner.

The right side illustrates a systems viewpoint of a biochemical network. Contrary to the components view, it is not the components that matter, but it is the state of the whole system that is important. Any biological network will have a nominal state that we recognize as a homeostatic state. Thus, the fluxes that reflect the interactions among the components to form the functional state of the network are dominant variables, and the concentrations of the individual components are 'subordinate quantities.' The concentrations of the network components are

determined to a first approximation by the flux map, or the state of the network, and then in detail by the kinetic properties of the links in the network.

Two key issues arise from the above considerations. The first deals with the nature of the links between components in a biological network, and the second deals with the functional states and the properties of a network that a set of links forms. At the genome-scale, the former is the process of network reconstruction discussed in Part I, and the latter is the subject of the current part of this book.

16.2 Properties of Links

Links between molecular components are given by chemical reactions or associations between chemical components. These links are therefore characterized and constrained by basic chemical rules. In tissue biology, the nature of links between cells is more complicated and often related to higher-order chemistry. We note that a T cell receptor, for instance, forms a complicated structure in the membrane of a cell. The properties of that structure, and how compatible it is with the complementary features of another cell, determine whether there is communication, or links, between these cells. Links between people in a social network have an even more complex basis. As we are focused on the characteristics of biochemical networks, we will discuss the chemical nature of links in molecular biology further.

Basic chemistry The prototypical transformations in living systems at the molecular level are bi-linear. This association involves two compounds coming together to either be transformed chemically through the breakage or formation of a new covalent bond, as is typical of metabolic reactions or macromolecular synthesis:

$$X + Y \rightleftharpoons X\text{–}Y \quad \text{covalent bonds}$$

or two molecules coming together to form a complex through hydrogen bonds and/or other physical association forces, a complex that has different functionality from individual components:

$$X + Y \rightleftharpoons X : Y \quad \text{association of molecules}$$

Such association, for instance, could designate the binding of a transcription factor to DNA to form an activated site to which an activated polymerase binds. Such bi-linear association between two molecules might also involve the binding of an allosteric regulator to an enzyme that induces a conformational change in the enzyme.

Chemical transformations have certain key properties.

Stoichiometry. The stoichiometry of chemical reactions is fixed, and is described by integers that count the molecules that react and that form as a consequence of the chemical reaction. Thus, stoichiometry represents 'digital information.' Chemical transformations obey elemental and charge balancing, as well as other features. Stoichiometry is invariant between organisms for the same reactions and

does not change with pressure, temperature, or other conditions. Stoichiometry gives the fundamental topological properties of a biochemical reaction network.

Relative rates. All reactions inside a cell are governed by thermodynamics. The relative rate of reactions, forward and reverse, are therefore fixed by basic thermodynamic properties. Unlike stoichiometry, thermodynamic properties do change with physico-chemical conditions such as pressure and temperature. The thermodynamics of transformation between small molecules in cells are fixed but condition-dependent. The thermodynamic properties of associations between macromolecules can be changed by altering the amino acid sequence of a protein.

Absolute rates. In contrast to stoichiometry and thermodynamics, the absolute rates of chemical reactions inside cells are adjustable. Highly evolved enzymes can be very specific in catalyzing particular chemical transformations. Cells can thus extensively manipulate the rates of reactions through changes in DNA sequence. Enzymes evolve to bring molecules into particular orientation to control the rate of appropriately oriented collisions between two molecules that lead to a chemical reaction (see Figure 7.1). It should be noted that much of the chemistry that takes place in cells occurs on the surfaces of protein. Surfaces are encoded in the DNA sequence and they determine the catalytic properties, such as rate constants, binding specificity, subunit association, protein binding to the DNA, and so forth.

The formation of links is restricted Links cannot just form between any two cellular components. The links that are formed are constrained by the nature of covalent bonds that are possible and by the thermodynamic nature of interacting macromolecular surfaces. The absolute rates are key biological design variables because they can evolve from a very low rate, as determined by the mass action kinetics based on collision frequencies, to a very high and specific reaction rate, as determined by appropriately evolved enzyme properties.

Information about links We do not have detailed information about the nature of all links between molecules inside a cell. In fact, there is a range of levels of knowledge that we have available to us (see Figure 16.2). Full information about the links allows for a full description of how a genetic property is mechanistically related to a phenotypic one. Most of this book is focused on this level of information, although additional auxiliary information can be included in the form of logistical relationships.

16.3 Links to Networks to Biological Functions

Reaction bi-linearity and network topology Most biochemical reactions are bi-linear. Bi-linearity gives the networks a hyper-graph property that is topologically non-linear. Consequently, biochemical reaction networks form a *tangle of cycles* [361] where different chemical properties and moieties are being transferred throughout the network from one carrier to the next. The coordinated movement of such *transferred*

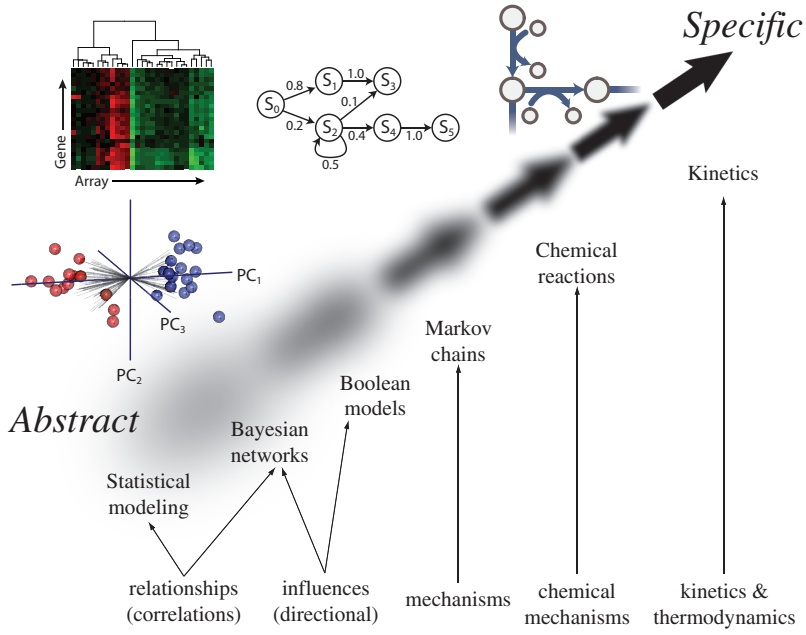

Figure 16.2 Illustration of different levels of knowledge that are available about a link in a network. Prepared by Nathan Lewis.

properties is determined by network topology and represents a key aspect of systems biology as they tie the whole system together.

We are familiar with the pathway maps that are used to describe cellular processes. We are less familiar with maps drawn around co-factors or carrier molecules that participate in multiple reactions. An example of the trafficking of redox equivalents in the core *E. coli* metabolic network is shown in Figure 16.3. This figure illustrates the two points of view, and shows how the carrier or co-factor molecules form a tangle of cycles that transmit the redox potential from one state to another. Another familiar example would be the movement of high-energy phosphate bonds between metabolites and proteins. ATP is the primary carrier of such high-energy bonds, and, for instance, a phosphate group is tied to glucose to form glucose-6-phosphate as the first step in glycolysis. The same feature is found in signaling networks whose components are in phosphorylated or dephosphorylated states. Other properties being transferred between molecules are one-carbon units, two-carbon units, ammonia groups, and so on.

Bi-linearity makes biochemical reaction networks highly interwoven and confers on them certain stoichiometric texture that affects their steady and dynamics states.

One network, many functional states One interesting feature of biochemical networks as they grow in size is that due to combinatorics, the number of possible functional states that they can take can grow faster than the number of components in a network. Therefore, the number of phenotypic functions derivable from a genome

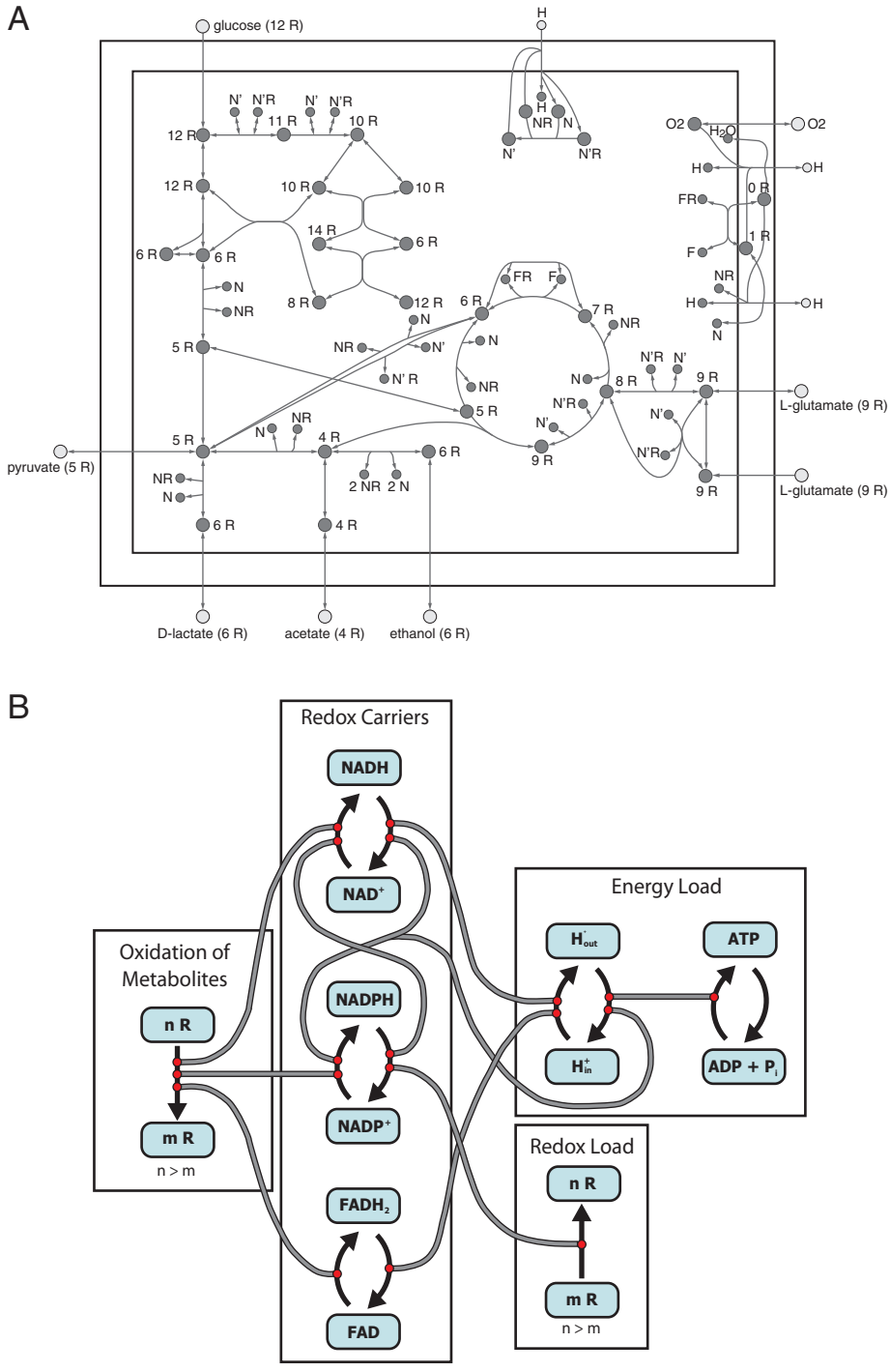

Figure 16.3 The tangle of cycles in trafficking of redox potential (R) in *E. coli* core metabolic pathways showing the redox equivalents (R) of the metabolites and carriers in the core *E. coli* metabolic model under aerobic conditions. (A) A reaction map organized around the core pathways. (B) A series of node maps organized around the molecules that carry redox potential. This map looks like a tangle of cycles. Taken from [317].

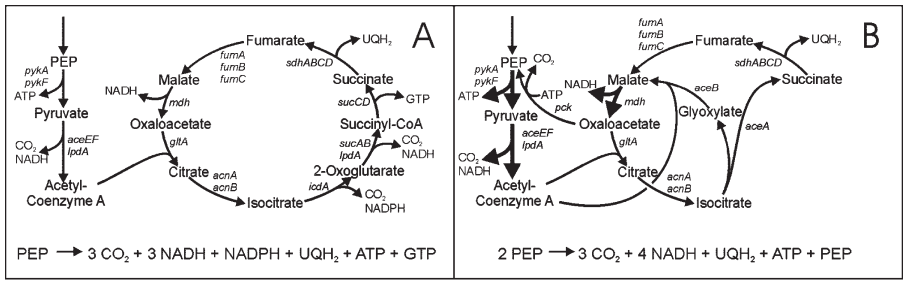

Figure 16.4 Stoichiometries of two alternative cycles for complete oxidation of PEP. The tricarboxylic acid cycle (A) and the PEP–glyoxylate cycle (B). Large solid arrows indicate reactions that are used twice per turn of the cycle. Gene names are shown in italics. Taken from [122].

does not linearly scale with the number of genes. For instance, the human genome may only have 50% more genes than the genome of *Caenorhabditis elegans*, a small worm, but nevertheless, human beings display much more complicated phenotypes and in greater variety. Thus, in general, it is hard to correlate organism complexity and functions to the number of genes its genome contains.

The fundamental property of biochemical networks having many possible functional states leads to the possibility of having the same network display many different phenotypic behaviors. A specific example in Figure 16.4 shows two experimentally determined alternative functional states of some of the metabolic pathways in the core pathways of *E. coli*. An organism does not fully exploit or use all possible functional states.

Many possible states will be useless to the organism in its struggle for survival. Therefore, a limited subset of these functional states needs to be selected and expressed by cells by imposing regulatory constraints. As we will discuss in Chapter 20, complex biological reaction networks can also have *equivalent* functional states, that is, there are identical overall functional states that differ in the ways in which they use the underlying links in the network.

Changing properties through evolution: distal causation Some of the key features of biological networks that distinguish them from other networks need to be accounted for in the analysis of their systemic properties. The first basic feature of biological networks is that they evolve; they change with time. They are *time-variant*. Principally, such changes occur through the kinetic properties of the links in the network and the changing of the available or active links in the network at any given point in time. The number of available links can be manipulated by regulation of gene expression, by horizontal gene transfer, and by other mechanisms.

The second feature that has to be taken into account is the fact that they have a sense of *purpose*. The fundamental purpose is survival. However, in complicated organisms that are fundamentally composed of many networks, some will have goals that are subtasks to the overall goal of survival. For instance, the goal of adipocytes would be to collect and store fat if there is an abundance of energy resources in its environment. A goal of the mitochondrion, being the powerhouse of the cell, seems to be to maximize ATP production from available resources. Therefore, the study of

objectives, i.e., purpose, of biochemical reaction networks becomes a relevant and central issue (Chapter 21).

Temporal–spatial organization Thus, linking many biological components together forms a network. This network can have many functional states from which a subset is selected. Links, network topology, and functional states can all change with time or environmental conditions. It is important to be cognizant of the fact that biochemical reaction networks have to operate in the crowded interior of a cell (see Figures 15.3A and 17.1). Thus, the network view of the biological process has to be considered in the context of the three-dimensional physical arrangement of such networks. These considerations may limit the usefulness of analogies with other man-made networks such as electrical circuits.

16.4 Constraining Allowable Functional States

Disciplines differ in their approach to model-building The above considerations of the nature of links, how they form networks, and how networks form functional states, make it likely that *in silico* modeling and simulation of genome-scale biological systems is going to be different from that practiced in the physico-chemical sciences. First is the notion that a network can fundamentally have many different states or many different solutions. Which states (or solutions) are picked is up to the cell and such choices can change over time based on the selection pressure experienced. This difference from the physico-chemical sciences is illustrated in Figure 16.5.

Theory- versus constraint-based thinking All theory-based considerations in engineering and physics leads one to attempt to seek an 'exact' solution, typically computed based on the laws of physics and chemistry. However, in biology it appears that not only can a network have many different behaviors that are picked based on the evolutionary history of the organism, but also, as we shall see, these networks can carry out the same function in many different and equivalent ways. This leads to an interesting distinction in mathematical modeling philosophy between the key disciplines (Table 16.1).

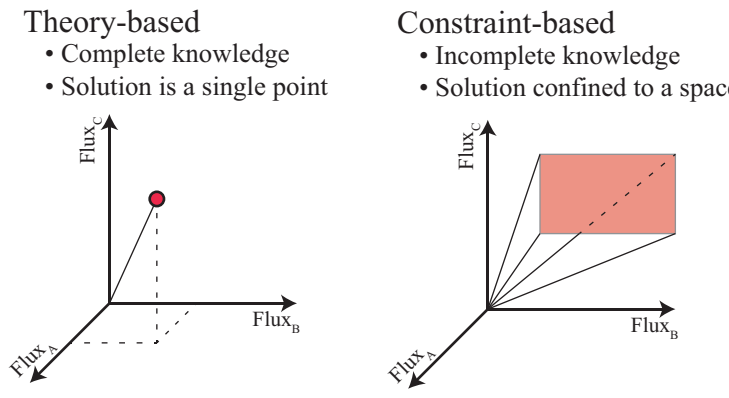

Figure 16.5 Theory- vs. constraint-based analysis. Illustration of finding an exact solution (a point) versus finding a range of allowable solutions (a solution space).

Table 16.1 Disciplinary differences in modeling philosophy.

	Equations	Boundary conditions	Nature of solutions
Physics	+++	+	Unique
Engineering	++	++	Design
Biology	+	+++	Multiple and changing

In physics, the emphasis has always been on deriving theory. Quantum mechanics developed about 100 years ago. Boltzman derived his famous equation prior to that. Theory, as expressed by mathematical equations representing our understanding of fundamental physical mechanisms, has been central to physics. If one wants to obtain particular solutions to these equations, one imposes boundary conditions that typically lead to the calculation of a unique solution.

Engineering takes a bit of a departure from this philosophy. The equations used in engineering do not need to be correct mechanistically, in a fundamental theoretic sense, as long as they describe the process at hand phenomenologically. Furthermore, the boundary conditions that need to be stated are very important and are often very specific to what an engineer is designing. In engineering, though, one is used to the fact that a problem can have multiple solutions, and that often comes down to the use of design variables to try to optimize a design.

In biology, based on the above consideration, we find that the equations needed to describe the physics of the intracellular environment may never be well-known, and furthermore, network functionalities evolve and change over time. Therefore, the fundamental equations describing biological functions may be hard to formulate and fully define. On the other hand, the boundary conditions or the constraints under which cells operate and evolve against are easier to identify, state, and use.

Constraint-based analysis methods These considerations give a general conceptual background for functional states and that there are constraints on what functional states a cell can take on. There are many methods that have been developed under the constraint-based modeling approach. They can be used to address many network properties, functional states, and biological questions; some are summarized in a later chapter (Table 18.1). To complete this chapter, we will discuss the general types of constraints that biology operates under before we proceed to formalize and mathematically deploy them.

16.5 Biological Consequences of Constraints

The constraints under which a cell operates Cells operate under myriad constraints. There are different ways to classify these constraints, and many authors have discussed them from different points of view. A few will be mentioned here.

- A statement of two very general categories of constraints imposed by natural selection have been described by F. Jacob [179]. They are basically (i) the requirement for reproduction and the genetic mechanisms required to produce offspring with non-identical genetic composition of the parent(s), and (ii) the permanent interaction with the environment that imposes thermodynamic constraints of constant flux of matter, energy, and information. The latter constraints are easier to describe in the language of the basic physical laws while the former describe distal causation.
- A. Danchin [93], in his insightful book about genomes, divides the cellular processes and their associated constraints into four general categories: (i) compartmentalization to segregate function in space and to differentiate the 'inside' from the 'outside;' (ii) metabolism that determines the flow of matter, energy, and redox potential within cells, and its relationship with the outside world; (iii) the transfer of memory to physico-chemical processes (i.e., 'actuating' inherited information); and (iv) memory transmitted from one generation to the next. This classification is similar to that of Jacob, with the first two describing the physico-chemical constraints that a cell deals with while the latter two are related to biological causation.
- In Chapter 15 we defined four categories of constraints that can be used to analyze the capabilities of reconstructed biochemical reaction networks: (i) physico-chemical constraints, (ii) spatial and topological constraints, (iii) environmental constraints, and (iv) regulatory constraints. This classification is operational and these constraints can be described mathematically and used to assess the capabilities of networks.

Picking candidate states: the role of regulation Cells are subject to inviolable constraints such as those associated with mass and energy balances. Their underlying biochemical networks must obey these, and other spatial constraints. These constraints have been called *hard constraints* and, as illustrated by the pentagon in Figure 16.6, give a range of all allowable states of the network. One or more states may be deemed suitable by the cell, based on its evolutionary history and current challenges (i.e., the prevailing environmental constraints). A way to exclude all the unwanted states (i.e., those that are unsuitable, or selected against) is to implement a regulatory network that eliminates a large portion of the solution space (the pentagon), and by default forces the expression of the 'desired' phenotype.

If a state or phenotype is not the best one under given conditions, the solution can move within the allowable range. This change in the selection of a functional state can be accomplished by regulating the expression of the genes or by regulating the activity of the corresponding gene products. Such regulation has a relatively short time profile. Over longer times, of course, the components of the network can evolve and the properties change slightly, allowing a drift in the phenotypic function of the cell.

Hierarchical organization in biology Many facets of cellular function and properties are organized hierarchically. The spatial organization of DNA is shown in Figure 16.7A. The linear dimension of the *E. coli* cell is about 1 mm while the length

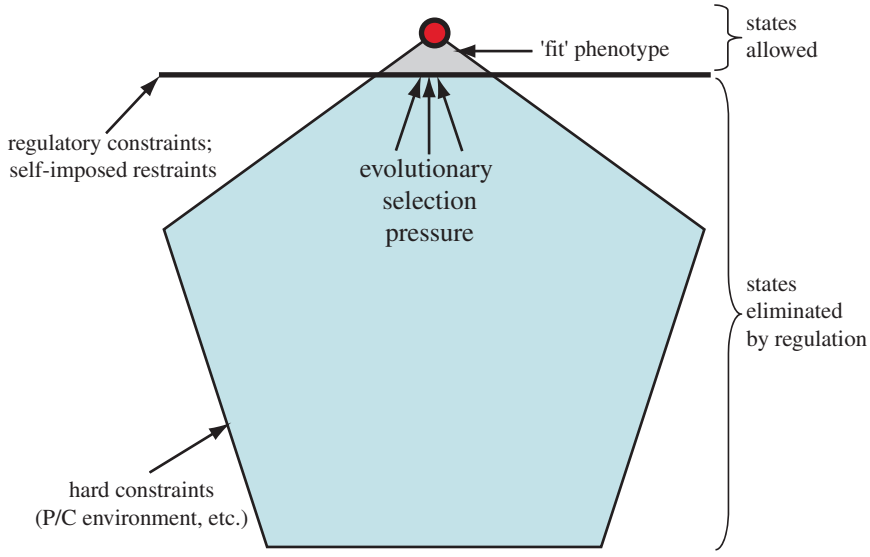

Figure 16.6 Illustration of the constraints on network functions. The pentagon illustrates the range of allowable functions based on hard physico-chemical and environmental constraints. The solid horizontal line illustrates self-imposed constraints (restraints) produced by regulatory networks, i.e., all the states below the line are ruled out by regulatory mechanisms (the blue segment). The red dot denotes the desired functional state, that is found among the admissible states (gray segment) after regulatory constraints have been imposed.

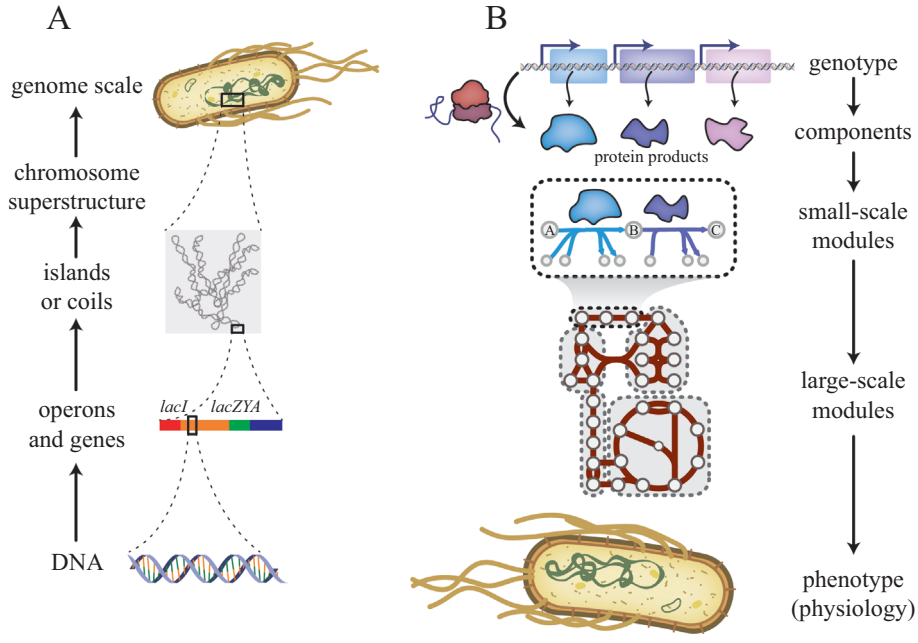

Figure 16.7 Illustration of hierarchical organization in biology: (A) of the DNA, (B) in network function. Prepared by Nathan Lewis.

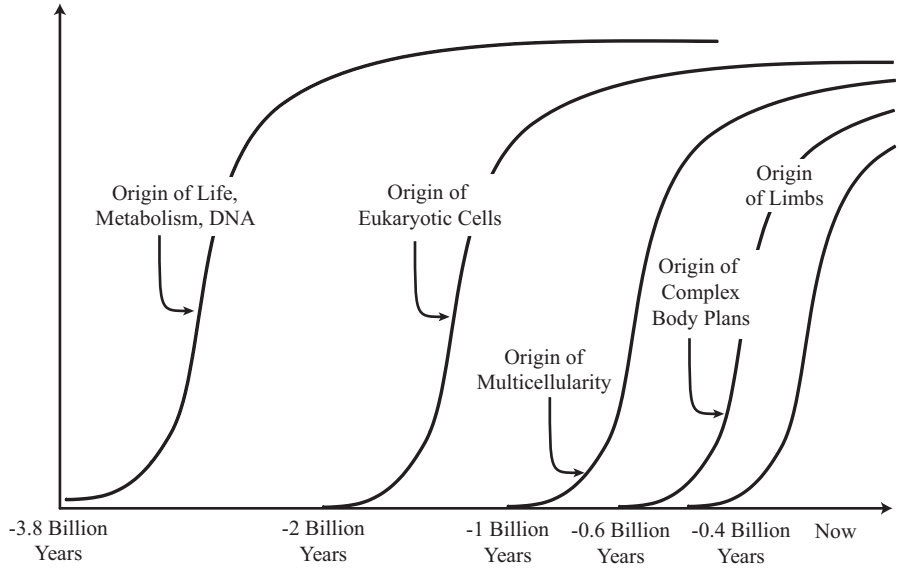

Figure 16.8 Timeline of the development and biological 'fraction' of major cellular and developmental processes. Inspired by Marc Kirschner.

of the cell is on the order of 1 μm, a 1000-fold difference. The bacterial genome is thus 'folded' a thousand times in a hierarchically organized fashion. Biochemical reaction networks can be similarly decomposed (Figure 16.7B). Reactions group together into coordinated units that may be co-localized in space, or even compartmentalized. Many such coordinated units can form a larger organized unit.

The constraints that apply to the lower levels of organization by necessity will constrain the subsequent higher-level functions. This upward application of constraints necessitates a *bottom-up* approach to the analysis of complex biological phenomena. Gödel's completeness theorem in mathematics that showed an axiomatic approach to proving mathematical theorems could not prove all properties of a system may in a general sense apply to biology. By analogy, we would expect that we cannot construct all higher-level functions from the elementary operations alone. Thus, observations and analyses of system-level functions will be needed to complement the bottom-up approach. Therefore, bottom-up and top-down approaches are complementary to the analysis of the hierarchical nature of complex biological phenomena.

Evolutionary adoption of constraints and formation of hierarchy The successive adoption of cellular functions over evolutionary times are illustrated in Figure 16.8. The basic biochemistry of cellular processes and the maintenance and expression of the information on the DNA molecule evolved early. This basic set of processes is found in all organisms today. The genetic code is essentially universal and most proteins are made up of about 20 amino acids. These are basic constraints under which all subsequent cellular processes must operate. The genetic code cannot be predicted from basic theory or physics [91], but is consistent with the basic laws of physics and chemistry. Once picked, it is essentially fixed over evolution. Similarly, most modern

proteins are made up of a limited number of motifs, and the basic circuits that lay out the body plan are remarkably conserved. Thus, the constraints set at a lower level of biological hierarchy confine higher levels of organization, but may not explain or predict the more complex functions. Evolution is a tinkerer that combines the elements at hand together in new and unpredictable ways. The first 'wave' in Figure 16.8 is close to the underlying chemical principles and represents the focus of this text.

16.6 Summary

- Biological systems are defined by the interactions between their components.
- The links between molecular components are constrained by the basic laws of chemistry.
- Multiple links between components form a network, and the network can have functional states.
- Functional states of networks are constrained by various factors that are physico-chemical, environmental, and biological in nature.
- The number of possible functional states of networks typically grows much faster than the number of components in the network.
- The number of candidate functional states of a biological network far exceed the number of biologically useful states to an organism.
- Cells select useful functional states by elaborate regulatory mechanisms.
- One may view hierarchical organization and evolutionary change as biological consequences of dealing with constraints.