Search for organising principles: understanding in systems biology

M.D. Mesarovic, S.N. Sreenath and J.D. Keene

Abstract: Due in large measure to the explosive progress in molecular biology, biology has become arguably the most exciting scientific field. The first half of the 21st century is sometimes referred to as the 'era of biology', analogous to the first half of the 20th century, which was considered to be the 'era of physics'. Yet, biology is facing a crisis — or is it an opportunity — reminiscent of the state of biology in pre-double-helix time. The principal challenge facing systems biology is complexity. According to Hood, *'Systems biology defines and analyses the inter-relationships of **all** of the elements in a functioning system in order to understand how the system works.'* With 30 000+ genes in the human genome the study of all relationships simultaneously becomes a formidably complex problem. Hanahan and Weinberg raised the question as to whether progress will consist of *'adding further layers of complexity to a scientific literature that is already complex almost beyond measure'* or whether the progress will lead to a *'science with a conceptual structure and logical coherence that rivals that of chemistry or physics.'* At the core of the challenge is the need for a new approach, a shift from reductionism to a holistic perspective. However, more than just a pronouncement of a new approach is needed. We suggest that what is needed is to provide a conceptual framework for systems biology research. We propose that the concept of a *complex system*, i.e. *a system of systems* as defined in mathematical general systems theory (MGST), is central to provide such a framework. We further argue that for a deeper understanding in systems biology investigations should go beyond building numerical mathematical or computer models — important as they are. Biological phenomena cannot be predicted with the level of numerical precision as in classical physics. Explanations in terms of how the categories of systems are organised to function in ever changing conditions are more revealing. Non-numerical mathematical tools are appropriate for the task. Such a categorical perspective led us to propose that the *core of understanding in systems biology depends on the search for organising principles* rather than solely on construction of predictive descriptions (i.e. models) that exactly outline the evolution of systems in space and time. The search for organising principles requires an identification/discovery of new concepts and hypotheses. Some of them, such as coordination motifs for transcriptional regulatory networks and bounded autonomy of levccels in a hierarchy, are outlined in this article. Experimental designs are outlined to help verify the applicability of the *interaction balance principle* of coordination to transcriptional and post-transcriptional networks.

1 A unifying framework for systems biology: complex systems paradigm

The concept of systems biology can be traced back to 1968 [1] defined at the time as the use of systems theory for explaining biological phenomena in terms of information and decision-making/control concepts, i.e. the study of phenomena in terms of how the objects are related rather than what they are composed of. Progress was hampered by the fact that systems theory was not developed to address biological phenomena as such but rather for addressing problems in engineering, management, etc. Conversely, research in systems biology used only a fraction of the concepts and results that already existed in systems theory. The situation has hardly changed since. For example, in genetics only the most classical (and straightforward) concepts such as ordinary differential equations, feedback, and feed-forward control, perturbation analysis, etc. are used. Our reflections in this paper are aimed at bridging this gap between biology and systems science research.

It is well known that in the post-genome era systems biology was reborn out of necessity. It became apparent that a holistic rather than a reductionism approach for understanding in biology is imperative: not only how many genes there are or even how they are connected but how they interact to result in observed behaviour of the overall system. A new dimension had to be added to the core of systems biology research — beyond mathematical and computer modelling — namely: *complexity*. The principal obstacles to this goal are conceptual [2].

Beyond academic research, business has also discovered systems biology, again by necessity. Figure 1 compares the NCE approvals of new drugs with research and

Systems Biology online no. 20045010

doi: 10.1049/sb:20045010

Paper first received 16th April and in revised form 19th May 2004

M.D. Mesarovic and S.N. Sreenath are with the Systems Biology Center Initiative, Dept. of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106-7071, USA, email: mdm5@cwru.edu

J.D. Keene is with the Center for RNA Biology, Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

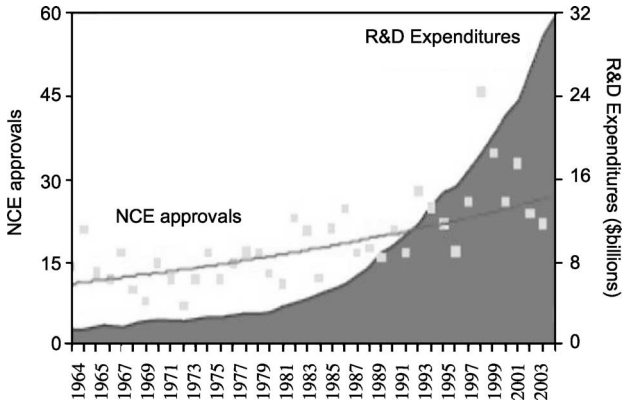


Fig. 1

development (R&D) expenditures [3]. Of particular note is a dramatic decline in the approvals, in spite of the increase in R&D expenditures from $12 billion to $32 billion over the last decade.

The starting point of our thesis is embodied in the contradistinction of the notion of a system and the notion of a complex system.

A system is a relation on items. Items within a system could be physical objects, indicators, variables, symbols, or any other conceptual category. For example, a body in motion viewed as a system is described in terms of relevant physical indicators: force, mass, velocity, acceleration; a genome is a set (sequence) of genes; a pathway is a set of kinematic reactions; a political or religious system is a set of interrelated dogmas, etc.

In mathematical general systems theory a system is defined as a relation in set theoretic terms

$$S \subset I_1 \otimes \ldots \otimes I_n$$

where I_i is the set of all ways in which the *i*th item can appear in an occurrence of the system. An element of *S* is then a multiple of the elements in each of the items, $s = (i_1, \ldots, i_n)$. The symbol $\otimes$ simply means that **all** items are viewed simultaneously. The symbol $\subset$ indicates that they are related, i.e. some of the multiples are missing; therefore there exists a relationship between the items since not all combinations of ways in which items appear simultaneously is observed. Hence, the items are mutually interdependent, constituting a relation.

A complex system is therefore defined by the following:

A complex system is a relation on systems such that:

- ***The 'overall' system has a distinct behaviour.***
- ***The subsystems preserve their own identities.***

By *distinct* it is meant that the behaviour of the complex overall system is described in terms of items that are different than the items of the subsystems.

A complex system is a 'system of systems.' Symbolically, a complex system is then defined by

$$S \subset S_1 \otimes \ldots \otimes S_n$$

The concept of a complex system can be viewed as the exemplification of the ancient dictum.

'*The whole is more than the sum of its parts.*'

The term 'sum,' of course, does not mean numerical addition but rather indicates the fact that the parts constitute the overall system by interacting. The parts (components, subsystems) have their own identities preserved within the interconnected whole. From that perspective, *systems biology constitutes much more than simply considering simultaneously all variables in biological observations*; i.e. observations on a systems level is not enough. *The distinct behaviour of the overall system has to be recognised.*

Items that form a complex system (i.e. subsystems) through interaction have their own recognisable boundary and existence while their behaviour (functioning) is conditioned by their being within the overall system. The human body is a prime example of a complex system. Human body parts (i.e. organs) are recognisable as such but their functioning (and even existence) is conditioned as being part of the total system (i.e. a body).

The term 'more' means the whole has an identity of its own, i.e. the behaviour of the whole is not recognisable in terms of indicators used to represent the parts.

A complex system, from this perspective, is not characterised solely by the number of items (i.e. being large), but by architecture of 'organised complexity' as a system of systems. It is more than just the items being interconnected: their organisation in a set of interconnected subsystems and resulting distinct behaviour of the overall system are defining characteristics of a complex system. A system with a very large number of items could be viewed as 'complicated' rather than the complex, with the latter term being reserved for organised complexity.

Mathematical and computer modelling are essential tools in systems biology but the explanations derived obtained their meanings in the context of the complex systems paradigm. There are at least two levels in a complex system: level of subsystems and level of overall system; they can be referred to as *functioning* and *behavioural* levels, respectively. In general, of course, there are many levels e.g. from molecular and cellular to organs, organisms, species, etc. Explicit recognition of multilevelness i.e. interdependence of the behavioural and functional levels in a living system — is central.

2 Principles of organised complexity

It is useful in the context of systems biology to make the distinction between *description*, *representation* and *understanding*. Description is a mathematical/computer model of a phenomenon under certain circumstances with numerically estimated parameters. Representation refers to a class/category of phenomena under a range of environmental conditions. A particular predictive model is then obtained by indicating a set of numerical parameters, each member of the set being appropriately specified. Understanding refers to ways the biological systems are organised, not necessarily in terms of mathematical/computer models but in terms of concepts. Multilevelness, concepts of feedbacks — negative, positive, multiple — feed-forward regulators, etc. are such principles. They are not mathematical but conceptual constructs. They are not 'engineering principles'; rather, they are concepts of how the real world functions from biology to psychology, economics to social sciences, etc. It is more appropriate to consider them as *systems science*, i.e. study of relationships within and among systems. Organising principles are on a higher level of perception than numerical representations; they could be based on observations but not restricted to numerical data. For example, Darwin's concept of evolution provided guidance and direction for data collection and numerical analyses.

A major challenge that has yet to be fully addressed in systems biology is the search for principles of organised complexity in biological systems. The organising principles play a role akin to the laws in physics in the sense that they

provide a starting framework to unravel the understanding of systems from observations and data. However, they are not numerical but 'relational', i.e. they indicate the functions of biological systems and/or their components. They provide the 'architecture' of the model — as an image of reality, i.e. projection of the reality on the focus/problem of interest. Some of the principles are presented below.

3 Multilevelness and bounded autonomy of levels

Multilevelness is the key organising complexity principle in complex systems. It appears in many forms of scaling in space and time. A fundamental question in systems biology is how the functioning on a higher level is related to the functioning on a lower level. In this respect, two alternatives are recognised.

Emergence. Distinct behaviour on the higher level is *solely* due to the way the components (subsystems) on the lower functioning level interact.
Coordination. Behaviour on the higher level depends not only on the interaction of the subsystems but also on the harmony of the subsystems on the lower level due to a *coordination process* or a *coordinator*.

Emergence is studied extensively as, for example, at the Santa Fe Institute, while coordination has received much less attention. In this paper we focus on coordination.

Consider two adjacent levels: two modes of inter-level relationships should be recognised. In the first mode the functioning/change on the lower level does not impact — as defined in an appropriate sense — the functioning on the higher level. In the second mode, the functioning on the lower level impacts the higher level. The first mode is referred to as *normal*, '*healthy*'. The second is referred to as *pathological*. We introduce the concept of a domain of autonomy of the levels to formalise the phrase 'in the appropriate sense'.

The **domain of autonomy** is a range of changes on the lower level and the corresponding range of normal behaviour on the higher level such that the two levels *do not interact* in the sense that the changes on the functional level are treated as 'background noise' on the higher level. Outside of the autonomy domain the influence of the lower level is treated as a *bona fide* signal by the higher level.

Making a distinction between the interaction and the interdependence is useful in this respect. Although the levels are interdependent in many ways — belonging to the same system — they can be viewed as non-interacting within domains of autonomy.

Identification of the autonomy domains is a major challenge in systems biology. Domains of normal behaviour are delineated by tolerances. The system becomes pathological either when the function on a lower level strays outside of the domain of autonomy or the tolerance on the higher level changes due to internal or external influences.

Bounded autonomy provides a 'division of labour' between the levels — a cross-level harmonisation — and illustrates that 'nature's design' is not optimisation of the behaviour over time, but rather optimisation of a system's organised complexity.

Bounded autonomy in multilevel biological systems is reminiscent of Herbert Simon's bounded rationality concept [3] and the contrast between satisfaction and optimisation [4] as studied in systems theory.

4 Coordination motifs

In Richard Young's Lab at the Whitehead Institute at MIT, the mapping of transcriptional regulatory networks for yeast has been developed identifying layers upon layers of regulators [5]. More recently, work has been reported using mammalian cells [6]. On the first level of these networks, the regulators are impacting on transcription promoters. On the higher level are 'regulators of the regulators' as they impact on lower level regulators.

In mapping regulatory networks, various feedback motifs are used as shown in Fig. 2. Feedback motifs are used to represent regulators regardless of their positions in the hierarchy. In a multilevel, hierarchical system, however, *the task of the higher-level regulators is not to control but to coordinate*, i.e. harmonise the functions of the first level regulators under changing conditions. For proper understanding of the functions of regulatory networks *coordination motifs that are distinct from lower level feedback motifs have to be discovered*. We will introduce some of them in the next Section.

There is a critical distinction between control and coordination. Control is 'dictating what is to be done.' Coordination is providing 'motivation' for the controllers

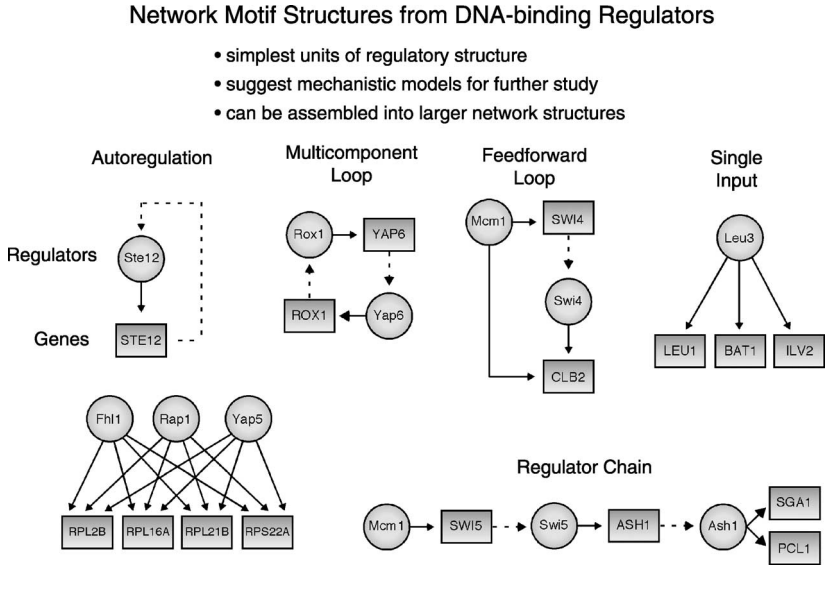
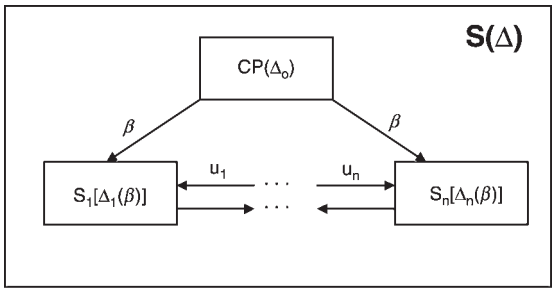
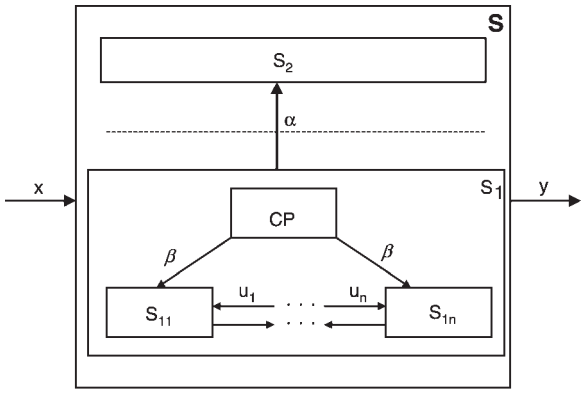


Fig. 2 *Motifs used in describing DNA binding regulators in Saccharomyces cerevisiae [5]*



S_i	First (Functional) Level Subsystem
CP	Coordination Process
S	Overall System
Δ_1	First Level Regulatory Function/Task
Δ_o	Coordination Task
Δ	Overall System Behavior
u_i	Interaction
β	Coordination Input

Fig. 3 *Coordination by a coordination process*



S_{1i}	First Level Subsystem	**S**	Overall System
CP	Coordination Process	u_1	First Level Interaction
S_1	First Level System	β	Coordination Input
S_2	Second Level System	α	Noise or Signal
x	External Change	y	Overall Systems Response

Fig. 4 *Autonomy domain of levels*

(regulators, modules, subsystems) to act so as to advance the overall system's objective while the subsystems are performing their own functions, modified by coordination.

Coordination theory has been developed within multi-level hierarchical system theory [7] (which is widely used in engineering [8], global issues [9, 10], organisation theory [11, 12] etc.). Figure 3 illustrates a coordinated system. On the first level there are n interacting subsystems, $S_1, \ldots, S_n$ forming the overall system by interaction, $u_1, \ldots, u_n$. The first level subsystems are homeostatic, i.e. they perform regulatory functions. Let $\Delta_1, \ldots \Delta_n$ represent the regulation objectives while $sat\Delta_i$ denotes that the regulatory function is satisfactory, i.e. the systems perform as required. Coordination has access to the first level regulators via an input β in order to modify the first level functioning in the direction of the overall system's objective. First level systems are then parameterised by β, i.e. $\Delta_1(\beta), \ldots, \Delta_n(\beta)$.

Each of the first level regulators focuses on a local subsystem to which it has direct access, and is 'ignorant' of what is happening in other parts of the overall system. To compensate for the deficiency due to the local focus, the first level regulators have to be harmonised. This is the task of coordination. Coordination function can be performed either by a coordination subsystem or by a properly designed *distributed coordination process*, CP.

Coordination is needed in order to harmonise the functioning of the regulators so that the overall integrated system functions properly. Let Δ represent the functioning of the system as a whole, while $sat\Delta$ indicates that the functioning of the overall system is satisfactory. A crucial step in understanding the concept of coordination is that *the task of coordination process*, Δ_o, *is not to explicitly pursue the overall objective*, Δ. Rather, the task of Δ_o is to influence all $\Delta_i(\beta)$ so that Δ is achieved while the first level systems perform their own, first level functions. The following coordination statement formally represents this relationship

$$[\boldsymbol{sat\Delta_1(\beta)\&\ldots\& sat\Delta_n(\beta)}]\ \boldsymbol{\&\ sat\Delta_o \Rightarrow sat\Delta}$$

which reads: *proper, satisfactory functioning of the first level systems*, $sat\Delta_i(\beta)$ *and simultaneous achievement of the coordination task*, $sat\Delta_o$, ***imply*** *the proper functioning of the second level*, $sat\Delta$. The word 'imply' is the key. Namely, Δ_o, and Δ are distinct

$$\Delta_o \neq \Delta$$

and are related to one another only by successful coordination.

Coordination is needed in order to harmonise the first level subsystem in external or internal changes. Let x denote a changing condition. The functioning of a first level system depends on x, i.e. it is a function of x, $sat\Delta_i(\beta, x)$. The task of coordination Δ_o is then the following: ***find*** $\hat{\beta}$ ***such that the coordination statement holds***.

$$[\boldsymbol{sat\Delta_1(x, \hat{\beta})\ \&\ldots\&\ sat\Delta_n(x, \hat{\beta})]\ \&\ sat\Delta_o \Rightarrow sat\Delta}$$

To explain what is meant by 'implication', we need to refer to the notion of the bounded autonomy of levels in a conceptual hierarchy, as introduced earlier. Conceptual levels refer to the focus of investigations. The first level could be the level of genes; the focus on the second level could be cellular or even higher. According to the domain of autonomy concept, the coordination of subsystems on the first level enables the behaviour on the higher level to stay within normal bounds. Within the domain of autonomy the cross-level interaction in effect remains on the noise level. The relationship is illustrated in Fig. 4.

A main obstacle to identifying coordination in biological systems is that the coordination function is distributed as a process and furthermore is 'fuzzy' because of the bounded autonomy property. This might very well be one of the reasons why a deeper understanding of the immune system has remained evasive.

So far we have considered what coordination should do — as defined by the logical coordination statement — but not how to do it. What is needed next is to identify the process of coordination by experimentation or by hypothesising its functioning and then to verify it experimentally in reference to the validity of the coordination statement. This will be outlined in the next Section.

5 Interaction balance principle of coordination

The starting point for understanding the concept of coordination by interaction balance is the observation that under 'healthy', perfectly normal, unperturbed conditions the interactions between the first level subsystems is harmonious in the sense that information transmitted by interactions between first level subsystems is as required for proper functioning of the overall system. There is no need for coordination! Behaviour/functioning follows the encoded script. Coordination is needed if, due to genetic or somatic mutations or perturbations, one or more of the subsystems is not sending the information (of the form or at

the magnitude) needed for proper functioning. A discrepancy that has to be restored provides a signal that coordination is needed.

Under normal conditions the *i*th subsystem receives the interaction $\mathbf{u}_i$ which it needs for its proper functioning. Let us assume that an external or internal change takes place denoted by x. In the new condition the *i*th subsystem needs different interaction which is a function of the change x, i.e. $\mathbf{u}_i(x)$. The actual interaction in the changed condition, $u_i(x)$, of course, is not necessarily the desired interaction, i.e.

$$\mathbf{u}_i(x) \neq u_i(x)$$

This results in uncoordinated behaviour. To bring harmony back the coordination process impacts the *i*th subsystem with the coordination signal, β.

The task of coordination based on the interaction balance principle is then to change the coordination signal β to such a value $\hat{\beta}$ that the interactions are balanced, i.e. desired and actual interactions are the same. The coordination task using the **interaction balance principle** (IBP) [6], is then: **find $\hat{\beta}$ such that the actual and desired interactions are the same for all i's, i.e.**

$$u_i(x, \hat{\beta}) = u_i(x, \hat{\beta})$$

The **coordination statement for IBP** is then

$$[\textit{sat}\Delta_1(x, \hat{\beta}) \,\&\, \ldots \,\&\textit{sat}\Delta_n(x, \hat{\beta})] = \,\&\, [u_i(\beta)\ u_i(\beta) \quad \text{for all i}] \Rightarrow \textit{sat}\Delta$$

The discrepancy between the desired and actual interactions is a trigger for coordination. A coordination process – acting across the entire overall system – sends the required coordination signals to all subsystems to restore their functioning so that the actual and desired interactions regain a balance, not necessarily the same as initially.

Two facts about coordination input are important:

- The coordination signal, β, has multiple components, i.e. it is a vector

$$\beta = (\beta_1, \ldots, \beta_n)$$

- Elements of the vector could be distinct for different subsystems, i.e. $(\beta_i \neq \beta_j)$. Yet, all elements of the vector are dependent on (are sensitive to) the discrepancy in any part of the overall system.

Discovery of a coordination principle is analogous to the discovery of a homeostatic feedback principle for the purpose of regulation. To illustrate this a block diagram of a system coordinated by IBP with two subsystems on the first level is shown in Fig. 5. The desired interactions $\mathbf{u}_1(\beta)$ and $\mathbf{u}_2(\beta)$ are contrasted with the actual interactions $u_1(\beta)$ and $u_2(\beta)$ and the resulting perceived discrepancies $\varepsilon_1(\beta)$ and $\varepsilon_2(\beta)$ are used as the basis by the coordination process to generate a new coordination signal, β.

Due to genetic errors or somatic impact on the system the balance is never fully achieved and an iterative process toward the balance evolves in time.

Robustness/resilience being ever present in biological phenomena, the coordinated behavior of the overall system does not require absolute balance. Rather, the system can tolerate imbalances within certain bounds, of which bounds, in turn, can also evolve over time.

Much like the feedback control principle, IBP is not a cure-all under all conditions. The usefulness of IBP and the motivation for the search for its application in systems biology depends on how broadly valid is the principle; i.e. whether it can, indeed, coordinate ever-evolving

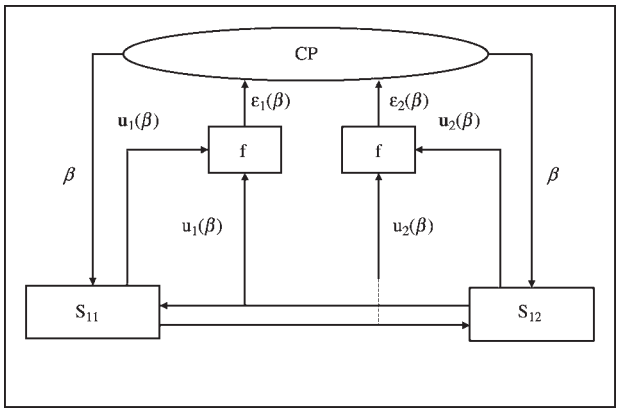


Fig. 5 *Interaction balance principle of coordination*

biological systems. Results in mathematical analysis in hierarchical systems theory shows that an *extremely broad category of systems is coordinable by IBP*. Therefore, IBP is a promising hypothesis for the study of coordination in complex biological systems. This is illustrated by the following theorem with the proof in [7].

To present the theorem we need the concept of the *inter-level function* in a hierarchical system and the property of *monotonicity* for a function.

Let v_i denote the state of the *i*th subsystem which is a function, g_i, of the change, x, and the coordination, β.

$$v_i = g_i(x, \beta)$$

For example, v_i could be the magnitude of transcription changed by x and modified by β. Let v denote the state of the overall system that corresponds to the given states on the first level $(v_1, \ldots, v_n)$.

If for any $(v_1, \ldots, v_n)$ there exists a unique v there exists an *inter-level function F* such that:

$$v = F(v_1, \ldots, v_n)$$

However, in general, for a given set of states $(v_1, \ldots, v_n)$ there might correspond different states of the overall system, v, at different points in time, e.g. due to the process of evolution.

Function F is monotone if for an increase of any argument the value of the function does not decline, i.e. either increases or at least stays the same. Monotonicity is defined, therefore, by

$$(v_i > v_i^*) \Rightarrow F(v_i) \geq F(v_i^*)$$

with all other arguments remaining unchanged. We then have the following:

Theorem: IBP successfully coordinates a complex system whenever:

- There exists an inter-level function F that for any given set of values on the first level $(v_1, \ldots, v_n)$ indicates a corresponding state, v, of the overall system.

$$v = F(v_1, \ldots, v_n)$$

- F is monotone, i.e. it is such a function that improvements in any of the first level states, v_i, will not negatively impact the second level performance, v.

The importance of IBP lies in the generality of the theorem, i.e. *if the conditions of the theorem are satisfied, IBP successfully coordinates a complex system regardless of any other properties of subsystems*. This is of particular

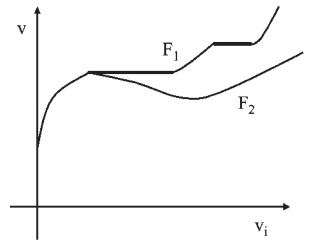


Fig. 6 *Monotonicity sufficient condition for IBP coordinability*

importance for biological phenomena for which analytical mathematical representations are at best approximations to various degrees.

For example, **the subsystems can be highly non-linear dynamical systems, or the subsystems do not have to be an analytical function at all**, such as in dynamics of the gene expression processes.

The theorem gives sufficient but not necessary conditions. Strict monotonocity guarantees (i.e. is sufficient for) coordinability by IBP.

Figure 6 helps illustrate the theorem. Whenever the inter-level function has the form F_1 it is guaranteed that IBP will successfully coordinate the system regardless of an other properties of the subsystems. If the form is as in F_2, IBP might still be valid, but might fail under certain circumstances.

IBP is presented here only as a case study. It is to be expected, of course, that other coordination principles are operative in biological systems.

6 Coordination motifs in regulated gene expression

We describe here J.D. Keene's experiments at Duke University Medical Center to search for coordination motif(s) in hierarchical regulatory networks starting with the IBP hypothesis. Applications to both transcriptional and post-transcriptional levels are considered. For the sake of completeness biological background for testing the IBP motif is outlined first.

The discovery of bacterial (prokaryotic) operons and their associated higher order regulons by Jacob and Monod in 1960 revealed one of the earliest examples of gene expression networks (Figs. 7*a* and 7*b*).

This discovery led to a frenzy of research for more than forty years into the regulation of gene expression because the proteins that are produced by bacterial operons are functionally related and efficiently coordinated [13, 14]. Therefore, much of our knowledge concerning protein function began with the elucidation of the prokaryotic operon. This research also led to rapid advances in our understanding of transcriptional mechanisms and illuminated many of the underlying principles of gene expression that occur in bacteria and bacteriophages [13]. However, in mammalian cells and other eukaryotic cells there is little evidence that operons of this type survived evolution. Yet it would seem important for gene expression to be coordinated in eukaryotes as well as in bacteria, while still accommodating greater genomic complexity.

It is logical to wonder whether there are gene expression networks in yeast and human cells that are analogous to operons. Indeed, it is known that transcriptional control is a major means of coordinated gene expression in eukaryotes. In fact, interaction motifs that are characteristic of many kinds of networks, including those in the worldwide web and power grids have been revealed by recent modelling of transcriptional control elements in sea urchin [15], yeast [16], and in human cells [17]. However, a new level of coordinate regulation has emerged at the post-transcriptional level that provides an operon concept and a gene expression network that complements transcriptional regulation, and appears to be synergistic with transcriptional networks [18]. Here, we first demonstrate the imperative to recognise coordination of transcriptional/post-transcriptional networks and then outline the design for experiments to test hypotheses of the applicability of IBP as a coordination motif.

Operons contain genes encoding bacterial proteins that function together in order to coordinate their expression. An example of a major operon is one that is responsible for forming the subunits of the ribosome, a large macromolecular machine that translates messenger RNA (mRNA) into protein. Other bacterial operons encode proteins and enzymes that function together in metabolic pathways that need to be coordinated efficiently. The coordination of gene expression by operons is straightforward because *transcription and translation are directly coupled* in bacteria (Figs. 7*a* and 7*b*).

Therefore, as mRNAs are being transcribed from the DNA in bacterial cells, ribosomes attach to the end of the nascent mRNA begin forming the proteins that are encoded

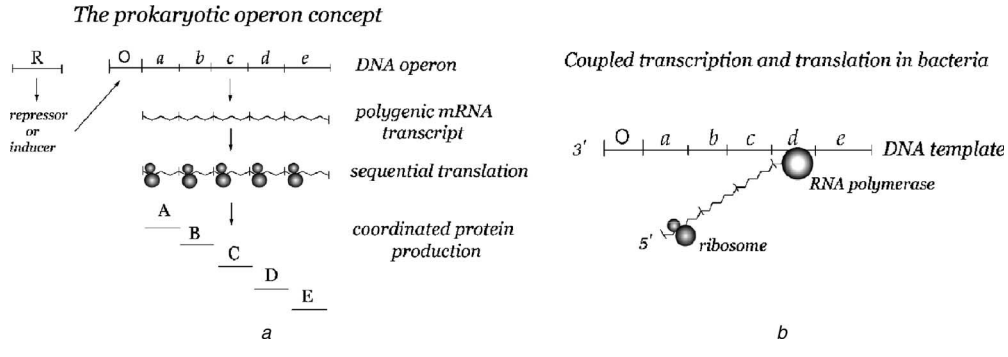


Fig. 7

a Illustration of a bacterial operon showing the genes tandemly linked on the DNA and a polycistronic (polygenic) mRNA transcript that is collinear with the DNA. Ribosomes can initiate protein synthesis at the start site of each open reading frame and coordinate the generation of equimolar amounts of the proteins A-E that are encoded by the genes a-e

b Transcription of the mRNA and its translation can be directly coupled in bacterial systems. Depiction of a mRNA transcript (zig-zag line) produced from an operon showing the RNA polymerase synthesising the mRNA while a ribosome begins protein synthesis at the 5′ end of the still nascent transcript

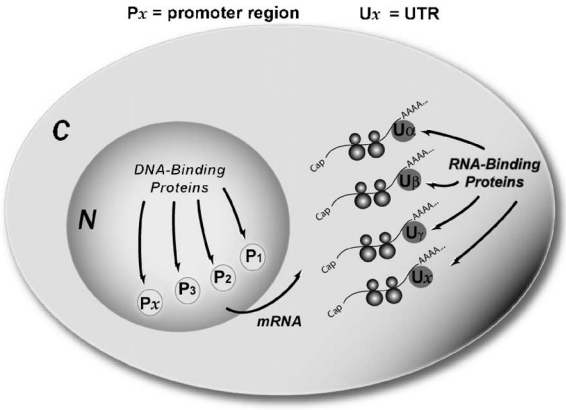


Fig. 8 *Depiction of a eukaryotic cell with both a transcriptional gene expression network in the nucleus (N) and a post-transcriptional gene expression network in the cytoplasm (C). DNA binding proteins such as transcription factors can activate or repress multiple promoter elements (P) on the DNA in order to coordinate the production of specific mRNAs that in turn are exported to the cytoplasm. Likewise, proteins that interact with mRNAs or mRNPs can organise the mRNAs using specific binding sequences (UTRs) to coordinate their expression at the level of translation or RNA stability. In addition, similar coordination processes can operate at the level of pre-mRNA splicing prior to export from the nucleus*

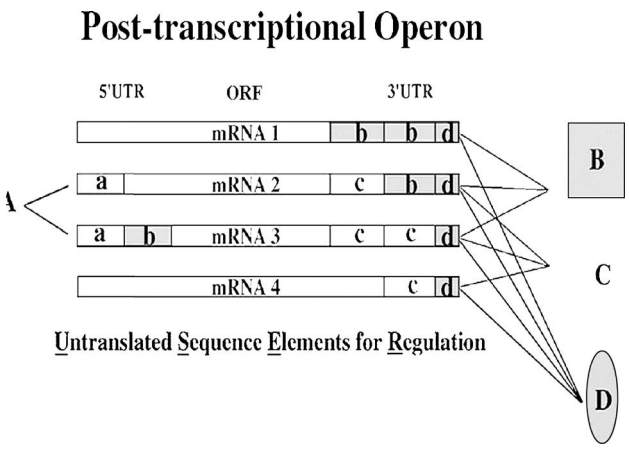


Fig. 9 *The concept of the post-transcriptional operon (or regulon) as proposed by Keene and Tenenbaum (2002) suggests that gene expression can be coordinated by factors (e.g. RNA-binding proteins or micro RNAs) that interact with multiple mRNAs. The activating or repressing factors are depicted as A-D and the RNA-binding sequences (UTR) are depicted as a-d. The essence of the theory is that multiple binding elements among related classes of mRNAs (USERs) allow them to be regulated together to coordinate their fates in concert. The significant advantage of the model is that mRNAs can be regulated in various combinations depending upon the needs of the cell, and this combinatorial power can generate genetic complexity at the post-transcriptional level. Likewise, steric occlusion of certain USER sites by binding factors could channel mRNAs into particular gene expression pathways, thus providing a means to evolve multi-functionality of the encoded proteins. During evolution the number of RNA-binding proteins and UTR sequences have expanded greatly concomitantly with increased multifunctionality. Reprinted by permission of Cell Press*

by it. This is best represented by the fact that many of the mRNAs produced from operons are polycistronic (polygenic), i.e. a single mRNA contains several tandem open reading frames and sequential start sites for protein production [14]. This architecture provides a very efficient and direct mechanism to assure that functionally related proteins are produced quickly and in the correct proportions by bacterial genes. However, in eukaryotic cells such as yeast and human, transcription and translation are not directly coupled to one another due to the fact that the nucleus compartmentalises the transcriptional apparatus on the DNA and the cytoplasm contains the ribosomes that are necessary for translation (Fig. 8).

Moreover, there is not a strict correlation between the amount of a protein that is made in a eukaryotic cell and the amount of steady-state mRNA that accumulates [18, 19]. Therefore, these synthetic processes of transcription and translation have become compartmentalised in different locations since evolution of the nuclear membrane [18]. Thus, eukaryotes lost the direct physical linkage between the functionally related genes on the DNA that formed the structural basis for bacterial operons, and a significant degree of posttranscriptional regulation evolved to govern protein production (Fig. 9). Therefore, the classical prokaryotic operons essentially disappeared as organisms evolved a defined nucleus.

Functionally related genes are not tandemly linked in eukaryotes, but there must be mechanisms to coordinate their expression. While transcription and translation are directly coupled in prokaryotes to allow collinear production of tandemly positioned genes, eukaryotes have their genes widely dispersed over multiple chromosomes without tandem linkage. Thus, it is not possible for functionally related proteins to be produced coordinately in eukaryotes by sequential transcription-translation. Therefore, it is important to ask how eukaryotic cells evolved a gene expression architecture that retained efficient protein production without coupled transcription and translation. Since the discovery of operons over forty years ago, it has been largely assumed that eukaryotes also regulate gene expression primarily at the level of transcription [13, 18]. This idea has been fixed in our thinking for many reasons. For example, viruses were used by molecular biologists as models of eukaryotic gene expression and it is believed that the gene expression is regulated chiefly at the level of transcription. More importantly, experimental methods to investigate coordinated gene expression at the level of transcription in all types of cells were easier than those needed to access post-transcriptional mechanisms. Thus, it is generally accepted that transcription factors, acting in various combinations at specific promoter sites on the DNA activate or repress gene expression [18]. Evidence indicates that transcription can be coordinated by these factors and that developmental processes in eukaryotes are orchestrated by transcriptional control. However, recent evidence indicates that gene expression can be orchestrated at the post-transcriptional level by RNA-binding proteins (Fig. 9) [18], and potentially by micro RNAs (mRNA).

In the search for network motifs to be tested as putative coordination models, application of IBP is under investigation through experiments in regulation of eukaryotic gene expression at both the transcriptional and the post-transcriptional levels. As noted, the steady-state levels of mRNA in a cell do not directly correspond with the steady-state levels of proteins present in that cell [18]. In part, this is because the biological fates of mRNAs are determined by the actions of RNA-binding proteins (RNA-BP) as the transcripts exit from the transcription machinery. Moreover, a significant filter between the transcription and translation networks involves selective export through the nuclear membrane by RNA-BPs. Therefore, in order to understand the regulatory linkages and coordination of global gene expression, one needs to reconcile the network systems that

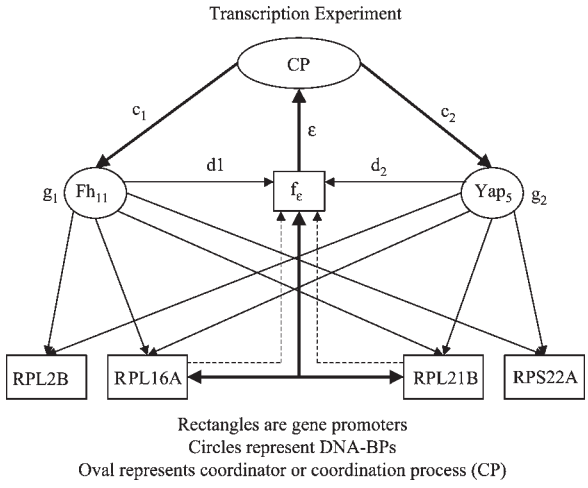


Fig. 10 *Example of interaction balance coordination motif involving transcriptional DNA-binding elements. The DNA and RNA binding factors are depicted as balls, while the proteins affected by their action at a DNA promoter or at a RNA UTR are represented by ribosomal proteins (RP)*

link the transcriptional apparatus with the post-transcription apparatus. Before this challenging task can be addressed experimentally, it is important to compare the regulatory motifs and IBP of each of these networks as they impact on the same regulated process.

These experiments are designed to evaluate the properties and parameters involved in coordinating these definitely linked, but poorly understood, gene expression networks (Figs 10 and 11). We compare transcriptional networks involving interactions of two DNA-binding proteins (DNA-BP), Fhl1 and Yap5, with promoter elements of a small group of ribosomal protein genes (RP) in parallel with interactions of two RNA-binding proteins (RNA-BP), Pub1c and Lhp1p, with the same group of RP messenger RNAs (mRNA).

In each case, a DNA-BP and a RNA-BP that do not interact with these DNA promoters and RNA USER elements will be used as controls. The experiments will involve chromatin immunoprecipitation (IP) of the DNA-BPs and ribonucleoprotein IP of the RNA-BPs using nuclear and/or cytoplasmic extracts. The recovery of the RP gene

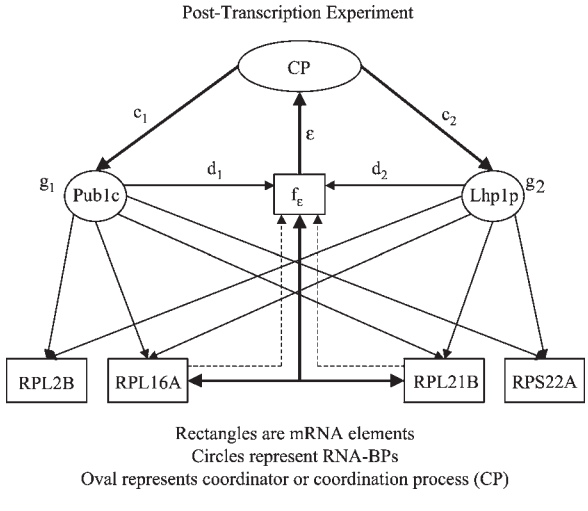


Fig. 11 *Example of interaction balance coordination motif for post-transcriptional RNA-binding elements. The DNA and RNA binding factors are depicted as red balls, while the proteins affected by their action at a DNA promoter or at a RNA UTR are represented by ribosomal proteins (RP)*

DNA containing the promoter and the RP mRNA containing the RNA USER element will be measured precisely using quantitative reverse transcription-polymerase chain reaction (RT-PCR).

Microarray analysis can be used also to expand the networks when appropriate. These experiments could be performed initially using logarithmic phase growing cultures of Saccharomyces cerevisiae, and will be compared in parallel with yeast cells that have been subjected to perturbations of the coordinating process (CP). These perturbations will include growth on galactose versus glucose, treatment with rapamycin to alter the TOR pathway. As many as twenty different perturbations can be used.

We hypothesise that the outcome of these calculations will show dual control at both the transcriptional and post-transcriptional levels, and synergy of the interactions leading to an asymmetric scaling of the network system that is characteristic of robustness. Therefore, based upon the transcription network analyses of [13] and the post-transcriptional Operon hypothesis of [18], we predict that the dual transcription/post-transcription network system will behave according to the interaction balance coordination principle.

7 Concluding remarks

Systems biology is a triad (Fig. 12). At the centre, of course, are the complex biological phenomena. Interpretation of observations and data is supported by computational algorithms, in particular, in genome sequencing, while nano-technology will play an equally important role in the future by providing access to observations inside the cells. Translation of interpretation to understanding is supported by systems science. Systems science provides a framework for understanding. It indicates hypotheses to be tested and modified in an iterative cycle of experimentation. Charles Darwin (1861) was quoted [20] as stating the following observation: *'How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service.'* The role of systems science is to provide such views.

Technology is considered as 'driving' systems biology, as an enabler of quantity and quality of data. A bottleneck in the advancement of understanding in systems biology is, however, conceptual: the gap between systems science and experimentation. We have indicated here some of the challenges. Many more have yet to be properly identified,

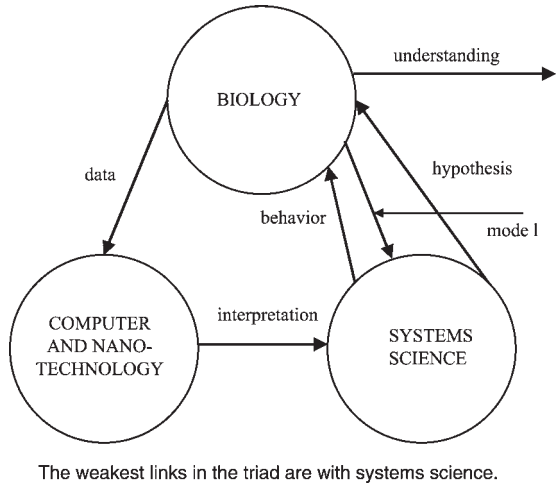


Fig. 12 *Systems biology triad*

for example, the question of teleology versus the classical input/output, 'Newtonian mechanics' view of the world. Teleology here simply means to indicate that many biological phenomena can be properly understood only if their behaviours are viewed as internally (goal-seeking) driven; for example, when bacterium follow a gradient toward food sources or the mutual attraction of tumor cells and endothelial cells residing in tumor vasculature.

Systems biology is a truly interdisciplinary science, the progress of which is seriously hampered by the lack of a language that provides a 'level playing field' for all participants. Drawing maps, networks, and analogous electronic circuits is not enough and can also be misleading. After all, biology is life — never static! Even the double-helix — the holy grail of molecular biology — has been found to turn around dynamically [21]. Categorical systems theory (aka mathematical general systems theory) could be a candidate for systems biology interdisciplinary language. To elaborate on that thesis would require much more space than is available in one article.

8 Acknowledgments

This work was supported by the Systems Biology Center of Excellence Initiative Provost Grant, Case Western Reserve University, 2004, and a gift from the Charles B. Wang Foundation to the Center for Computational Genomics and Systems Biology at Case. The authors benefited considerably from discussions with, and comments from, Drs. Joe Koonce, Patrick Lager, Ken Loparo, Leonardo Lozada, Joseph Nadeau, Yoh-Han Pao, Olaf Wolkenhauer, and many other colleagues through the years.

9 References

1 Mesarovic, M.D.: 'Systems theory and biology—view of a theoretician', in Mesarovic, M.D. (Ed.): 'Systems theory and biology' (Springer-Verlag, New York, 1968), vol. 351, pp. 59–87
2 Hanahan, D., and Weinberg, R.A.: 'The hallmarks of cancer', *Cell*, 2000, **100**, pp. 57–70
3 The Economist (April 2004)
4 Simon, H.: 'Science of the artificial' (MIT Press, 1984)
5 Lee, T.I., Rinaldi, N.J., Robert, R., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.D., Gordon, D.B., Ren, B., Wryrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., and Young, R.A.: 'Transcriptional regulatory networks in Saccharomyces cerevisiae', *Science*, 2002, **298**, pp. 799–804
6 Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A.: 'Control of pancreas and liver gene expression by HNF transcription factors', *Science*, 2004, **303**, pp. 1378–1381
7 Mesarovic, M.D., Macko, D., and Takahara, Y.: 'Theory of hierarchical, multilevel, systems' (Academic Press, New York, 1970)
8 Mesarovic, M.D., and Pestel, E.: 'Mankind at the turning point: the second report to the club of Rome' (Dutton, New York, 1974)
9 Jamshidi, M.: 'Large-scale systems: modeling, control, and fuzzy logic, North Holland series in system science and engineering' (Prentice-Hall, 1996, 1st edn.)
10 Mesarovic, M.D., and Pestel, E. (Eds.): 'Multi-level computer model of world development system'. Proc. Int. Symp. on Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 29 April–3 May 1974, vol. II
11 Takahara, Y., and Mesarovic, M.D.: 'Organization structure: cybernetic systems foundation'. Proc. IFSR Int. Series on Systems Science and Engineering, 2004, vol. V, p. 22
12 Takahashi, S., *et al.*: 'Applied general systems research on organisations' (Springer, New York, 2003)
13 Beckwith, J.: 'The Operon: an historical account', in Neidhart, F.C. (Ed.): 'Escherichia coli and salmonella cellular and molecular biology' (ASM Press, Washington, DC, 1996), pp. 1227–1231
14 Niedhardt, F.C., and Savageau, M.A.: 'Regulation beyond the operon' in Neidhart, F.C. (Ed.): 'Escherichia coli and Salmonella cellular and molecular biology' (ASM Press, Washington, DC, 1996), pp. 1310–1324
15 Davidson, E.H., Rast, J.P., Olivieri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., and Bolouri, H.: 'A genomic regulatory network for development', *Science*, 2002, **295**, pp. 1669–1678
16 Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R.: 'Correlation between protein and mRNA abundance in yeast', *Mol. Cell. Biol.*, 1999, **19**, pp. 1720–1730
17 Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L.: 'Integrated genomic and proteomic analyses of a systematically perturbed metabolic network', *Science*, 2001, **292**, pp. 929–934
18 Keene, J.D.: 'Ribonucleoprzotein infrastructure regulating the flow of genetic information between the genome and the proteome', *Proc. Natl. Acad. Sci. USA*, 2001, **98**, pp. 7018–7024
19 Keene, J.D., and Tenenbaum, S.A.: 'Eukaryotic mRNPs may represent post-transcriptional operons', *Mol. Cell*, 2002, **9**, pp. 1161–1167
20 Steven Weinberg, Article in *The New York Review of Books*, LETTERS
21 'Beyond the double-helix', *Nature*, 2003, **421**