Cooperativity and symmetry at biological scales

A. Carbone ‡

Institut des Hautes Études Scientifiques, 35 route de Chartres, 91440 Bures-sur-Yvette, France

Abstract. Facts and ideas presented in this review paper have been written mostly as a guideline to orient the reader through some references in the field. Biological scales are intended in a broad sense and refer to bio-molecular structures as well as supra-molecular and cellular organisations.

1. Two examples.

The world of biology is a world of organisation: spatial and temporal. In general these two aspects are intricate and cannot be dissociated, with the consequent effect that the mechanisms underlying certain biological phenomena are hard to guess. In the following two examples, I still want to emphasize each aspect separately: temporal organisation is illustrated by protein networks in chemotaxis, and spatial organisation, by the structural regularity of protein coats of viruses.

Chemotaxis. Cells employ protein networks to control their behavior. In bacterial chemotaxis, for instance, the bacterium runs, tumbles, senses the average levels of nutrient. When the bacteria (too small in size to sense gradient differences in its receptors) senses *over time* that the attractant concentration is high, it keeps on swimming in the same direction, whereas when the time average concentration of the attractant decreases, it switches into its tumbling mode and changes direction, with the global result to swim towards high nutrient regions. The paradigm followed can be roughly described as a sequence of three actions: sense, compute, respond.

- *Sense:* peptide signals bind to receptors and they are read out in tiny amounts.

- *Compute:* the receptor, by various mechanisms, transduces this event into a change in the chemical state of a messenger molecule; this pathway involves thousands of bio-molecules, transport into the membrane and the cytoplasm, and polarization and migration of chemical gradients.

- *Respond:* molecules interact with others that affect the cell physiology, mechanics, or gene expression in the cell.

This rough schema is present in neutrophil chemotaxis as well, where the neutrophil (much bigger in size than a bacterium) locally senses the bacteria and by contracting its actin filaments accordingly, it chases the bacteria until phagocyting it [1, 2]. Neutrophil chemotaxis is a different and much more complicated process (at the molecular level) than bacterial chemotaxis, due to the feed-forward and feed-back coupling with the *mechanical* properties of the cytoskeleton (induced actin contraction). In both bacterial and neutrophil chemotaxis, the integration of multiple signals among different pathways is poorly understood.

‡ Electronic address: carbone@ihes.fr

Viruses: the example of bacteriophage HK97. The crystal structure of the mature empty capsid of the double-stranded DNA bacteriophage HK97 was recently determined at 3.6Å resolution [3]. It is an icosahedral capsid of 660Å in diameter formed by the assembly of 420 copies of a single 385-residue protein. This virus remarkably displays topologically linked proteins in its capsid, and subunits rotations and local refolding mechanisms in the mature capsid formation.

Each capsid subunit, a flat-folded protein, is joined to its neighbor by ligation on the side chain. This generates pentameric and hexamerics topologically linked rings of covalently joined subunits (recall that a covalent phosphate bond between nucleotides corresponds to $50kcal/mol$ while the energy of an hydrogen bond is $1-5kcal/mol$!). These catenanes have not been previously observed in viral capsid and they provide here a stabilisation mechanism for this very thin virus envelop [3].

Besides the beauty of the geometrical arrangements of proteins, an amazing discovery has been associated to bacteriophage HK97: it undergoes several capsid large-scale conformational changes which transform precursors into infectious virions [4]. Both the first and the last capsid of this series of conformations, have been experimentally identified (the first by cryoelectron microscopy and the last by crystallography), and the intermediate ones have been deduced by computational modeling, which turned out to be a crucial tool to understand the various steps of the transformation. It was detected that a rigid body rotation (≈ 40 degrees) caused switching to an entirely different set of interactions during folding and that in this occasion two motifs undergo refolding. These conformational changes stabilize the capsid by increasing the surface area buried at interfaces and bringing the cross-linking forming residues close together. The inner surface of one of the intermediate capsid is negatively charged and this suggests that the transition is triggered electrostatically by DNA packaging. These transitions are irreversible and their energetic basis lies in exothermic switching from one local minimum of the conformational free energy to another lower-energy state.

2. Genomes

The life of a cell is orchestrated by the information encoded in its *genome*, a double-stranded DNA molecule consisting of four nucleotides A, T, C, G. We shall point out some basic facts on cells and genomes of various organisms.

The bacterium *E. coli*, of volume $\approx 1\mu m^3$, is a rod-shaped cell with a thick cell wall composed of two concentric layers. Extending from the cell surface there are a few rotating flagella used by the cell as propellers. In the center of the cell, there is a region where the $4 \cdot 10^6$ base-pair DNA is packed. Between the inner cell wall and the region containing the DNA, the cell is filled in by the cytoplasm, a dense crowd of large and small molecules in thermal motion. There are about 300,000 (non-ribosomal) proteins of $\approx 10nm$ in diameter, 20,000 ribosomes (making the 25% of the cytoplasmic volume), 300,000 tRNAs, a couple of thousand mRNA molecules, 50,000,000 small organic molecules, including amino-acids, nucleotides, sugars, ATPs, and various ions. Yet, 70% of the cell volume is filled by $2 \cdot 10^{10}$ water molecules. Eukaryotic cells, which are typically 10 times bigger than bacteria (1,000 more voluminous), have also a cytoplasm but they are architecturally organized into several relatively well defined compartments and structural networks: nucleus, cytoskeleton, Golgi apparatus, endoplasmic reticulum, vacuole, other organelles, etc.

Molecules in the cytoplasm "communicate" with each other through *physical interactions*. This molecular communication represents the central interest of ongoing research: data gathered with different techniques are stored in databases and combined with DNA sequence information to infer, through intermediate conjectures and hypothesis, possible mechanisms regulating the cellular machinery.

What is coded in the genome? It is tempting to think that a genome encodes the full description of a given organism. But if the genome does encode the full description it is only by a series of indirect implications. The genome is expressed in the cell environment, and the cell carries a long history in its structure and molecular content. Initial conditions for a living cell are determined by the mother cell, and "the rest", such as the architecture, the consistency of the initial conditions, and the cell behavior, can be thought as being coded in its DNA. Still non-deterministic features are also omnipresent due to the importance of thermal agitation at sub-micrometric scales. For instance, the neuronal architecture is only roughly coded in DNA while many synaptic junctions between neurons can be seen as randomly distributed. More precisely, there are organized regions in the brain which are connected to each other in a pre-determined way and this is due to the cooperation of protein networks that induce neurons to grow in specific directions. Another basic observation which is important to keep in mind is that only a fraction of the genomic information is used at different stages of a cell life. There are genes whose role is crucial under only extreme conditions, and such events might never appear in a cell lifetime. Also, in multicellular organisms, differentiated cells employ distinct parts of the genome corresponding to their specialized function within the organism as a whole.

Genome sizes. There is a huge dispersion of genome sizes, going from $\approx 10^6$ basepairs (bps) for bacteria, to $\approx 10^{11}$bps for certain plants, like lily. *H. sapiens* has a genome of $\approx 10^9$bps, like sharks and frogs, or certain mollusks, reptiles, birds, plants, etc. In most multicellular eukaryotes, the actual DNA *coding* regions, where genomic information is stored, cover a small fraction of the genome sequence. *H.sapiens* has a 1% of coding DNA, the lungfish has 0.02% over its 140Gbps, the fruit-fly has 20%, etc... while the unicellular yeast has 70% and the bacteria *E.coli* nearly 100% of coding genome.

Genomes are highly constraint sequences. Genomes might be very long but they are far from being random; in fact, their emergence along evolution imposed constraints which are very limiting as can be seen easily with some simple observations. Take, for instance, the number 10^8 of existing viable genomes (that is, 10^8 different species on earth, including the large variety of living bacteria) against the 4^{10^6} possible genome sequences, where 4 is the number of bases A, T, C, G, and 10^6 is the average size of bacteria (since most existing species are bacterial species, it is fair to consider their genome size). Another example concerns proteins. If we assume that the "meaningful" part of a genome corresponds to proteins, then the space of all 4^{10^6} sequences is reducible to a space of 20^{300} potential proteins, where 300aa is the average length of a protein. Even if we would consider a larger size for proteins, as 2000aa for instance, the two numbers 4^{10^6} and 20^{2000} would still be far from being comparable. Of all potential protein sequences, only 5^{300} are (potentially) *functional* proteins (this rough estimate is easily computable from the secondary structure of a protein and the potential disposition of α-helices and β-sheets in the 3-dimensional space). If we look at proteins with *two functions*, this number decreases again to 2^{300}, and so on.

What are the factors that imposed these restrictions along Evolution? To give a definite answer is impossible, but two aspects concerning the structure of a cell can be noticed. The first is the presence of *symmetries* at the different levels of organisation in the cell. Syntactic symmetry (seen in genomic repetitions, and in the use of codons), spatial symmetry (seen in viral coats), temporal symmetry (seen in the cyclic behavior of bio-chemical processes), combinatorial symmetry (seen in the homology of protein sub-complexes), functional symmetry (seen in the function similarity of homologous complexes).

The second aspect concerns the presence of *universal machines* in the cell. These universal machines have the property to interact with a large class of molecules of a given type. For instance the ribosome which translates essentially any kind of RNA sequences into proteins, the RNA polymerase which transcribes essentially any DNA sequence into RNA sequences, the splicing machinery which cuts out introns from RNA sequences transforming them in mature messenger RNAs, the ubiquitin machinery that helps protein degradation, chaperons that help protein folding, etc.

What is the difference between organisms. One would like to understand where the difference between different living organisms lies. Remember that a man and a gorilla share more than the 99% of their coding sequences. To try to answer to the question we need to go down to a molecular level, but let us first fix some numbers:

organism	Number of genes	genome size
Homo sapiens (human)	$3-5\cdot 10^4$	$3.3\cdot 10^9$
Fugu rebripes (fish)	$3\cdot 10^4$	$4\cdot 10^8$
Drosophila melanogaster (fly)	$1.4\cdot 10^4$	$1.8\cdot 10^8$
Caenorhabditis elegans (worm)	$1.8\cdot 10^4$	$1\cdot 10^8$

The size of a genome is measured by its number of base pairs (remember that a stretched human DNA sequence is about 2 meters long). The table above shows that living organisms have genome sizes which are often very different, while the number of their genes is rather close. Actually, different organisms not only share the same number of genes, but those are almost identical (*homologous*), as shown by the recent sequencing of the mouse, rat, drosophila genomes, etc. In fact, less and less genes remain unknown as new genomes are sequenced: in 1996, with the completion of the *S. cerevisiae* genome sequence, a third of its genes did not match any previously detected sequence; in 1998, a sixth of the genes had no known homologues in *C. elegans*, and in 2000 only a tenth of the genes of *D. melanogaster* were defined as “unknown”. Even if genes are detected, it does not mean that their functions are known. The recently sequenced micro-organism *Plasmodium falciparum*, the malaria eukaryotic parasite, is $22\cdot 10^6$bps long and contains (an estimation of) 5,200 genes whose 60.9% have been identified as hypothetical proteins, that is coding sequences with unknown function [5].

Families of homologous genes. When we look closer to organisms with bilateral symmetry, like humans, flies, rats, sea urchins, we see that

- their genome contains at least one specific gene from each family of homologous genes, for each process of exchange of information in the cell, and that

- it is the number of genes coding for regulatory factors that varies from organism to organism (see Section 3).

This suggests that these organisms seem to be “equivalent” on the basis of genetic contents, and that we need to go down to the molecular level and study how their genes interact (through regulatory factors) to understand their actual differences. In practice, one looks for a model that is able to generate many different organisms from a combination of very few signals and very few genes! The proposed model is based on the idea of *cooperativity*: many different proteins interact in *concert* and with the DNA sequence to give a rich combinatorics of cellular behaviours.

3. Regulatory networks

A simplified model of transcription, translation and regulation. Before discussing the interactions between genes, let us briefly review basic mechanisms of gene expression. A gene along DNA is preceded by a so-called *promoter* region, where "regulatory" proteins interact with each other via direct or indirect binding to the DNA to "decide" whether the gene should be transcribed or not by an RNA-polymerase, the basic component of the transcriptional machinery in the cell. Hence, regulatory proteins *regulate the expression* of the gene.

Such proteins bind to promoter regions to either prevent transcription or enhance it. Therefore the level of gene expression decreases in presence of proteins inhibiting transcription, while it increases with a high concentration of proteins enhancing it. In between, the status of a gene is often tuned between on and off, by a relative concentration of inhibitors and enhancers. (A realistic picture should also take into account the continuous reduction in concentration due to protein degradation.)

If the combination of proteins on the promoter region enhances gene translation, the RNA-polymerase produces an RNA sequence, which is an exact copy of the gene. This sequence, after being often edited and spliced by the machinery of the cell (in case of eukaryotic cells), is ready to be translated into a protein by the ribosomal complex. The ribosome, which is the basic component of the translational machinery of the cell, gets the messenger RNA as input and synthesizes the corresponding protein from individual amino-acids.

Since regulatory proteins are themselves produced by the same transcription/translation process, their expression is also controlled by other regulatory proteins and even sometimes auto-regulated by their own concentration.

The complexity of regulation. The unfolding of development, from fertilization to the formation of a differentiated adult organism follows a blueprint encoded in its DNA both as transcriptional units and *cis*-regulatory elements. The DNA information is interpreted by transcription factors which respond to inputs from upstream stimuli and transfer these signals downstream to other regulatory genes and to structural genes controlling the proper development of the organism. Linear regulatory pathways have been studied from many years, but recent studies also point towards more complex transcriptional networks, reminiscent of electrical engineering diagrams. Such networks involve a vast number of genes and rely on positive and negative feed-back loops, amplification mechanisms, combinatorial interactions, cross-talks, signal thresholds and gradients of regulatory factor concentrations.

A logic behind regulation. Behind the combination of numerous regulatory proteins on promoter regions, there is a logic that biologists try to understand and which can be "rigorously" formalized. The regulation of a gene can be approximated by some boolean function of several variables describing the absence or presence of regulatory proteins. Suppose, for simplicity, to have just two sites of interaction in the promoter region and two proteins, interacting each one with one site. The presence or absence of these proteins at their respective sites, determines the on/off state of the gene. For two variables there are already 16 different possible boolean functions that can describe the gene expression. In general the story is much more complicated. Instead of 2 sites, several and organized subsets of sites are involved. Such regulatory modules in the promoter region are often called cis-*regulatory sequences*. Different parts of the control mechanism are under selection in different regimes. We shall see this in what follows.

Modules and the regulation network in early development: the case study of Sea Urchin. The Sea Urchin genome (complete sequence expected by 2004) is 800 Mega bases long, about $1/4$

of the human genome. Regulatory networks and *cis*-regulatory elements of some of the genes involved in endo-mesodermal specification during the early development of the Sea Urchin have been identified and thoroughly studied [6, 7, 8]. Let us outline some of the results and the complexity behind this regulation network by considering the *endo 16* gene, which plays a role in the development of the gastric apparatus.

The promoter region of *endo 16* is 2300bps long. A map of the regulatory proteins interacting with this region has been reconstructed: there are 7 modules consisting of 12 binding sites for *specific* proteins, as well as 23 other sites where *several different* proteins can bind. The regulatory circuit involving seven modules and 40 proteins was identified and can be compared to a computer program with 40 different variables. This logic model allows testable predictions to be emitted.

The data to reconstruct this circuit have been obtained experimentally. The protein associated to *endo 16* has been fluorescently labelled, and optical slices were recorded. Each slice corresponding to a different state of the "boolean circuit", corresponds to a different state of development of the Sea Urchin. Davidson and his group are about to reconstruct the *full* "logical" circuit of endo-mesodermal specification during early development (*endo 16* is just one player amongst more than 50 involved genes!). Their approach will remain based on careful analysis of expression data, perturbation data, *cis*-regulatory analysis of several genes, and other kinds of experiments [8].

Another example: the Drosophila development. By contrast with the previous example where a single mutation in one of the genes is usually sufficient to produce a new phenotype, John Reinitz showed how the "eve" gene of the Drosophila fly required a combination of mutations of its binding sites to display a phenotypic change [9]. Here the combinatorics becomes much more complicated. Reinitz developed a refined three-tiered model of transcriptional regulation. In particular, above direct DNA-binding by factors, the second tier involves adapter molecules binding to DNA-bound factors, while the third tier includes a physical model of how adapters initiate transcription.

4. A second level of organisation: network connectivity

The intertwined signaling cascades, superimposed to the transcriptional circuitry, together with the communication among structural genes producing the main building blocks of the organism, make matters even more complex. The most elementary kind of circuits that one can think of consists of two expression cascades. The first produces a protein that catalyzes a reaction providing some metabolite for the organism; the second leads to a regulatory protein with no catalytic function. The small molecular weight signaling molecule obtained from the first cascade, and the regulatory factor obtained from the second cascade, interact and mutually affect each other's production by interfering with the transcription of the two genes which are at the origin of the two cascades. This mutual connectivity can be thought as an effective wiring between the two cascades.

The connectivity of several signaling cascades easily generates a complicated network of interactions. Macromolecular networks come in three major flavours: *metabolic* networks, where interactions mostly involve enzymatic proteins and small molecules or metabolites, *protein-protein interaction* networks, and *genetic* networks, where regulatory proteins interact with promoter regions on DNA. These networks have been analyzed and are still studied from physiological, comparative and informational angles. Real progresses have been made in the study of metabolic flow charts, while methodologies for the other two kinds of networks are still rather poorly developed.

Genetic network homology does not imply gene homology: the example of circadian clocks. Different organisms are found to share the same networks. Circadian rhythms, for instance, are used by a wide range of organisms to provide an internal daily periodicity, and the associated clock networks share common features from cyanobacteria to mammals. Namely, all networks seem to include an interaction between two kinds of components: *positive* elements and *negative* elements [10]. In the simplest case, two proteins P and N, which play the role of the positive and negative elements, interact with each other as follows: P increases its own expression and that of N by binding to the promoter regions P_P of itself and P_N of N; on the other hand, a strong binding of N to P inhibits P activity, and represses the expression of both elements by preventing P to bind to the promoters P_P and P_N.

Circadian circuits show that the topology of a circuit is preserved from species to species but that this is not related to gene homology. For instance, the positive elements KaiA in *Synachococcus*, Clc and Cyc in *Drosophila* and Clock and Bmal in mice are not homologous, and this is also not the case for the negative elements KaiB and KaiC in *Synachococcus*, Tim and Per in *Drosophila*, Tim, Per1 and Per2 in mice.

An example of topologically equivalent genetic circuits that perform the same function and that display homologous genes across species, is bacterial chemotaxis [11].

Large networks: the protein-protein interaction networks. Protein-protein interaction networks are constructed from data on pairwise interactions between proteins (detected with two-hybrid systems, for instance) or on protein complexes (detected by mass spectrometry). The networks derived from these kinds of data are usually very different. This is due to two reasons: first, different experimental techniques cover different fractions of the complete set of proteins of a given genome, and second, data are strongly dependent on the experimental protocol.

A graph of pairwise protein interactions has been constructed for the yeast genome: nodes are proteins and edges are interactions. It contains a giant component of 466 proteins and 163 small disjoint subgraphs. (This graph has to be seen more as a statistical estimate of the actual protein-protein interaction in yeast because of the high number of false positive and false negative that the yeast-two-hybrid experiments might detect.) The mathematical analysis of its structure identifies it as a "small world" graph, which means that any two points in the graph can be connected by relatively short paths along existing links. In particular, these graphs are sparse graphs, cliquish, they are "scale-free", that is they satisfy a power law distribution of the node degrees $P(k) \approx k^{-\gamma}$, where $k \gg K$ for some threshold K, they contain few highly connected nodes, and they are insensitive to random errors, that is, if we take away a randomly chosen node, the average distance between nodes does not change [12, 13].

Visualization of protein complexes and a comparative study between complexes and pairwise protein interactions has been recently proposed [14].

Another example of large networks: metabolic networks. Metabolic networks represent the bio-chemistry of an organism, that is those processes that generate mass, energy, information transfert and cell-fate. Nodes are molecules and edges are reactions. As for protein-protein interaction networks, metabolic networks are scale-free graphs. A basic difference between the two types of graph is that protein-protein interaction networks have disjoint clusters while metabolic networks are connected. In particular, it has been noticed that metabolic networks of *different organisms* present the same topological scaling properties [15, 17, 16].

Networks and evolution. What is the dynamical process generating large networks? How metabolic and protein interaction networks have been shaped? Does their present structure provides robustness against mutations? What are the evolutionary models compatible with their growth? One way to *generate* a small world network is to take a regular graph and randomly reassign some of the connections. Another way is by adding new nodes to a graph, and preferentially connect them to existing nodes that have already large connection degrees. The first method displays a peaked distribution of connections, while the second displays a power law distribution for the connection degrees of the nodes. As mentioned above, large biological networks appear to be scale-free, and it was suggested that the most connected nodes in large networks might also be the most ancient [16].

By taking into consideration biologically relevant hypothesis in the network construction such as gene duplications, and both the possibility of adding and deleting nodes from the graph, Andreas Wagner [18] showed that theoretical models of network growth bore little relevance to evolutionary processes leading to real-life networks: protein connectivity does not reflect evolutionary constraints.

Networks and biological functions. What kind of biological information can be deduced from the structure of biological networks? Does the large scale structure of metabolic and protein-protein interaction networks reflect any biological function? Does it reflect chemistry or physics? Are the common combinatorial features of regulatory networks, as for instance circadian networks, consequences of some underlying "design principle"?

Network design. The ability to predict, control and design biological complex networks depends on having detailed quantitative descriptions of their dynamics, but absolute descriptions of mechanistic accuracy seem hopeless. Hence, one way to proceed might be to apply *physical/chemical* knowledge to develop abstract statistical models of these processes and use them to refine or disprove particular physical hypothesis.

Le us consider a concrete example. Regulatory networks that produce a regular oscillatory behavior over a 24-hour period and that are driven by light or temperature might be many. Biological circadian networks have, however, to satisfy some extra constraints such as a circadian period remaining relatively constant over a wide range of temperatures, and functioning reliably in the presence of internal noise, or under global changes in transcription and translation levels under certain nutrition or growth conditions. It seems plausible that a large scale design of "stable" *physical* components might induce a relatively "robust" behavior of such regulatory networks [10].

The study of cellular signal-processing pathways reveals recurring regulatory motifs. Small graph structures (where nodes are molecules and edges are *chemical* reactions) appear multiple times within an organism and among organisms. In this direction, Adam Arkin emphasizes the interest of breaking down some of the regulatory networks into recognizable engineering functions, irrespective of the identities of their constituent genes. Autoregulatory feedback loops in gene expression and cross-regulatory systems as well as amplifier functions, frequency-filtering functions and biphasic response motifs are identified in several concrete biological examples [19, 20].

A combinatorial approach to network design, has been taken in [21], where the design principles of transcriptional regulatory networks is delineated both through time-resolved gene expression on living bacteria and through a graph-theoretic approach to identify basic building blocks within the whole network. Relations between these basic blocks and their information processing function are established.

Conclusion. Networks should be viewed as a beginning, and not as an end to our understanding of the developmental, differentiation, growth, apoptosis programs of an organism. Equally complex mechanisms involving alterations in chromatin configurations, epigenetic events such as DNA methylation, associations of transcriptional activators and repressors that do not directly contact DNA, post-translational modifications of the transcriptional machinery impose additional layers of regulation on the hierarchy of mechanisms relating sequences to phenotypes.

Acknowledgment

We apologize for the many omitted and missed references. For more references and an introduction to some mathematical models in molecular biology the reader may consult [22].

References

[1] Ch.V. Rao, J.R. Kirby, A.P. Arkin. Design and diversity in bacterial chemotaxis, *Science*, 2002.
[2] "Neutrophil crawling", http://expmed.bwh.harvard.edu/projects/motility/neutrophil.html. Movie on a neutrophil chasing a bacteria which was turned in 1950 by David Rogers.
[3] W.R. Wikoff, L. Liljas, R.L. Duda, H. Tsuruta, R.W. Hendrix, J.Johnson, Topologically linked protein rings in the bacteriophage HK97 capsid, *Science*, 289:2129-2133, 2000.
[4] J.F. Conway, W.R. Wikoff, N. Chen, R.L. Duda, R.W. Hendrix, J.E. Johnson and A.C. Steven, Virus maturation involving large subunit rotations and local refolding, *Science*, 292:744-748, 2001.
[5] Plasmodium genomics, *Nature*, 419:489–542, 2002.
[6] C.H. Yuh, H. Bolouri, E.H. Davidson, Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 279: 1896-902, 1998.
[7] E.H. Davidson, *Genomic Regulatory Systems*, Academic Press, USA, 2001.
[8] E.N. Olson, Ed. *cis Network*, a special issue of *Developmental Biology*, Academic Press, USA, 2002. The issue collects papers on the work done in Davidson Laboratory around the deconstruction of the transcriptional circuitry of development.
[9] J. Reinitz, D.H. Sharp, Mechanism of eve stripe formation. *Mech. Dev.*, 49: 133-58, 1995.
[10] N. Barkai, S. Leibler, Circadian clocks limited by noise, *Nature*, 403:267-268, 1999.
[11] N. Barkai, S. Leibler, United we sense..., *Nature*, 393:18-21, 1998.
[12] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small world' networks, *Nature*, 393:440-442, 1998.
[13] R. Albert, H. Jeong, A.L. Barabási, Error and attacks tolerance of complex networks, *Nature*, 406:378-382, 2000.
[14] G.D. Bader, Ch.W.V. Hogue, Analyzing yeast protein-protein interaction data obtained from different sources, *Nature Biotechnology*, 20:991–997, 2002.
[15] D.A. Fell, A. Wagner, The small world of metabolism, *Nature Biotechnology*, 18:1121-1122, 2000.
[16] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabási, The large scale organisation of metabolic networks, *Nature*, 407:651-654, 2000.
[17] A.L. Barabási, R. Albert, Emergence of scaling in random networks *Science*, 286(5439):509-12, 1999.
[18] A. Wagner, The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, 18:1283-1292, 2002.
[19] A. Gilman, A.P. Arkin, Genetic "code": Representations and dynamical models of genetic components and networks, *Annu. Rev. Genomics Hum. Genet.*, 3:341-69, 2002.
[20] H.H. McAdams, A.P. Arkin, Towards a circuit engineering discipline. *Curr. Biol.*, 10: R318-20, 2000.
[21] S.S Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nature Genetics*, 2002.

[22] A. Carbone, M. Gromov. Mathematical slices of molecular biology, *La Gazette des Mathématiciens*, Société Mathématique de France, numéro spécial, 11–80, avril 2001.