The Disaster of Central Control

An Impressive Example from Nature

PETER SCHUSTER

Peter Schuster is Professor of Theoretical Chemistry at Vienna University. His publications, about 300 papers and nine books, deal with the analysis and prediction of molecular structures from small molecules to biopolymers and the dynamics of evolutionary processes. The books include "The Hydrogen Bond", three collective volumes together with Georg Zundel and Camille Sandorfy; "The Hypercycle—A Principle of Natural Self-Organization" together with Manfred Eigen; and "Evolutionary Dynamics—Exploring the Interplay of Selection, Accident, Neutrality, and Function", a collective volume together with James P. Crutchfield. His recent research focuses on optimization of RNA structures as a model for molecular evolution and the role of selectively neutral variants. Peter Schuster is at the Institüt für Theoretische Chemie und Molekulare Strukturbiologie der Universität Wien, Waeringerstraße 17, A-1090 Wien, Austria; E-mail: pks@thb@univie.ac.at.

It is commonplace to say that central control of large complex entities is doomed to fail. We have numerous examples from economies and societies of the past and the present demonstrating inefficiency when the systems exceed a certain size. In the old days, wise emperors were well aware of the problem and answered appropriately by the principle of *divide et impera*. Nature seems to have an elegant solution the problem too: Modular structure and partial autonomy of modules. The best example is the multicellular organism where the individual cell retains as much autonomy in metabolism as can be tolerated without endangering the whole system: A little more independence of somatic cells, for example, leads to tumor formation. Efficient division of labor is observed in bacterial cells too. No wonder I have thought that we can always learn from biology how to manage successfully the most sophisticated situations and how to handle and control complexity. It was a shocking experience therefore when I read a recent preprint and previous articles by John Mattick and his colleagues [1–3]. They present a plausible interpretation of the limitation in bacterial genome sizes based on DNA sequences: The number of genes in prokaryotes is bounded by an unaffordable regulatory overhead in genomes that are too large. This view is supported by an empirical fit of a power law to the data derived from some 90 fully sequenced prokaryotic genomes. The number of regulatory genes grows approximately with the square of the total number of genes, $n_r = 1.63 \times 10^{-5} \cdot n^{1.96} \approx 1.2 \times 10^{-5} \cdot n^2$, where n_r and n are the numbers of regulatory genes and all genes, respectively. Thus, complete regulation of ribosomal protein synthesis—when centrally organized on the DNA level in the spirit of the elegant *operon* mechanism as discovered by Jaques Monod, François Jacob, and André Lwow—falls into an inefficiency trap when genomes become too large. The guess is that the critical genome size for prokaryotes is in the range of 10,000 genes and indeed, this appears to be the size limit of bacterial genomes.

Eukaryotes—these are all higher organisms from yeast to man—make use of other control mechanism in addition to genetic control at the DNA level. Examples are alternative splicing of precursor messenger RNAs (for reviews see [4, 5]), small interfering RNAs [6], genomic imprinting [7, 8], and other forms of epigenetic regulation of gene expression and silencing. For the purpose of illustration we shall consider here an idealized—and perhaps not yet fully accepted—model view of alternative splicing: The translation product of the same DNA stretch yields slightly or substantially different proteins depending on the individual organism and the particular tissue in which it is expressed. Diverse proteins are obtained by cutting out different sections called *introns* from the precursor RNA sequence. A well-known example is the protein that controls sex

Certainly, we do not know yet the full story of the regulation of gene expression in multicellular organisms—transcription, maturation of messenger RNA, and translation—and we have to be prepared for further surprises.

in the fruit fly *Drosophila*. The amino acid sequence of a protein is thus laid down in part by factors that are determined by the parent organism—for example, the mother providing the egg cell—or through cell differentiation during development. The same piece of DNA sequence gives rise to different proteins, depending on the needs of the particular differentiated cell where it is expressed. What we observe is, therefore, a powerful interplay between central control that is represented by the regulation of DNA transcription and decentralized regulation at the RNA level prior to ribosomal translation into protein. Certainly, we do not know yet the full story of the regulation of gene expression in multicellular organisms—transcription, maturation of messenger RNA, and translation—and we have to be prepared for further surprises. What we know for sure, however, is that the eukaryotes have managed to go beyond the critical number of 10,000 genes by means of the initially mentioned *divide and impera* principle.

In order to illustrate my case for decentralized control I shall make a rough estimate that falls approximately within the range where John Mattick's empirical formula is strongly supported by data from sequenced prokaryotic genomes, $500 < N < 10000$, and compare a centralized genome of 10,000 genes with a virtual system of 10 locally regulated genomes of 1000 genes each, which are joined by a central unit. What we need in the first case, the centralized organization, is 1200 regulator genes according to Mattick's relation. For the decentralized system we assign to a small center, for example, 180 genes for controlling the activities of the 10 subunits—four for each of the 45 interactions in pairs—and 10×2 genes for regulating the communication between the center and the ten outposts at the periphery. The decentralized model would thus work with an overhead of only some 320 control genes, 200 in the center and, following the Mattick equation, 10×12 regulatory genes at the periphery.

To be fair to central organization, I shall mention also one clear disadvantage of decentralized regulation. A famous saying states "there is no free lunch" and by saving almost 900 "supernumerary" agents in the central organization of our example, we must have lost something as well. Because most control action in the decentralized system is done at the spot, little information needs to be exchanged with the center. The "headquarter" is managing as few messages as absolutely required, just enough to organize the interactions between the different units at the periphery. The lack of information in the center is a clear disadvantage for planning the future of the system as a whole. New challenges can hardly be met without a global view of the environment and consequences to be drawn from the global picture cannot be put in operation without central forces. The Darwinian mechanism of optimization through variation and selection need not care about the future because it operates with a "trial-and-error" mechanism on the population level. Hence solitary multicellular organisms do well with little central control and future planning. The situation becomes dramatically different in animal and human societies. Learning of the individual and education change the mechanism from Darwinian to Lamarckian: Information acquired by a single member of the society is readily transmitted to future generations and centralized forecast becomes important for the society as a whole. In such a scenario correct forecast of the future is of high value. Too much federalism and the lack of central power does often not allow for success in necessary planning. I mention two different examples: (i) the failure of worldwide reduction of carbon dioxide emissions despite clear evidence of its urgent necessity, and (ii) the difficulty of Switzerland to vote for joining the European Union despite clear advantages for her economy.

In closing I shall return to the inefficiency of central control that, I guess, we all can witness in everyday life. For me, being affiliated with a large university housing some eighty thousand students, the prokaryotic failure makes a strong case for the requirement of decentralized university organization rather than central control. One out of many reasons is to keep the costs of overheads small because expensive overheads reduce drastically the limited financial support from research grants: We should rather *do* research than keep our deans and rectors fully informed about our ideas and plans, which, moreover, will never materialize provided we proceed into unknown territories of knowledge as we should. In any case success is well documented in the scientific literature and provides the basis for evaluation organized by the central administration that in most other issues need not interfere and therefore could by as slim as possible.

The Darwinian mechanism of optimization through variation and selection need not care about the future because it operates with a "trial-and-error" mechanism on the population level.

REFERENCES

1. Croft, L.J.; Lercher, M.J.; Gagen, M.J.; Mattick, J.S. Is prokaryotic complexity limited by accelerated growth in regulatory overhead? Proc Natl Acad Sci USA 2004, Forthcoming.
2. Mattick, J.S.; Gagen, M.J. The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol 2001, 18, 1611–1630.
3. Mattick, J.S. Non-coding RNAs: The architects of eukaryotic complexity. EMBO Rep. 2001, 21, 986–991.
4. Black, D.L. Mechanisms of alternative pre-messenger splicing. Annu Rev Biochem 2003, 72, 291–336.
5. Graveley, B.R. Sex, Agility, and the regulation of alternative splicing. Cell 2002, 109, 409–412.
6. McManus, M.T.; Sharp, P.A. Gene silencing in mammals by small interfering RNAs. Nat Rev Genet 2002, 3, 739–747.
7. Reik, W.; Walter, J. Genomic imprinting: Parental influence on the genome. Nat Rev Genet 2001, 2, 21–32.
8. Wilkins, J.F.; Haig, D. What good is genomic imprinting: The function of parent-specific gene expression. Nat Rev Genet 2003, 4, 359–368.