Acta Biochimica Polonica

Vol. 53 No. 1/2006, 1–9

on-line at: www.actabp.pl

Review

Regulatory mechanisms of gene expression: complexity with elements of deterministic chaos

Jolanta Jura[1], Paulina Węgrzyn[1], Jacek Jura[2] and Aleksander Koj[1]✉

[1]*Department of Cellular Biochemistry, Faculty of Biotechnology, Jagiellonian University, Krakow, Poland;*
[2]*Department of Animal Reproduction, National Research Institute of Animal Production, Balice, Poland;*
✉*e-mail: koj@mol.uj.edu.pl*

Received: 27 June, 2005; revised: 03 January, 2006; accepted: 05 January, 2006
available on-line: 23 February, 2006

Linear models based on proportionality between variables have been commonly applied in biology and medicine but in many cases they do not describe correctly the complex relationships of living organisms and now are being replaced by nonlinear theories of deterministic chaos. Recent advances in molecular biology and genome sequencing may lead to a simplistic view that all life processes in a cell, or in the whole organism, are strictly and in a linear fashion controlled by genes. In reality, the existing phenotype arises from a complex interaction of the genome and various environmental factors. Regulation of gene expression in the animal organism occurs at the level of epigenetic DNA modification, RNA transcription, mRNA translation, and many additional alterations of nascent proteins. The process of transcription is highly complicated and includes hundreds of transcription factors, enhancers and silencers, as well as various species of low molecular mass RNAs. In addition, alternative splicing or mRNA editing can generate a family of polypeptides from a single gene. Rearrangement of coding DNA sequences during somatic recombination is the source of great variability in the structure of immunoglobulins and some other proteins. The process of rearrangement of immunoglobulin genes, or such phenomena as parental imprinting of some genes, appear to occur in a random fashion. Therefore, it seems that the mechanism of genetic information flow from DNA to mature proteins does not fit the category of linear relationship based on simple reductionism or hard determinism but would be probably better described by nonlinear models, such as deterministic chaos.

Keywords: linear and nonlinear responses, alternative splicing, RNA editing, monoallelic expression, biallelic expression, somatic recombination, epigenetics

NONLINEAR DYNAMICS IN THE DESCRIPTION OF BIOLOGICAL PHENOMENA

There is no doubt that many spectacular achievements in molecular biology and medicine have come from applying linear theories based on proportionality between two variables. However, as pointed out by Higgins (2002), nonlinear behavior prevails within human systems due to their complex dynamic nature. For this reason nonlinear system theories are beginning to be applied in interpreting, explaining and predicting biological phenomena in categories of the theory of deterministic chaos. According to Higgins (2002) *"chaos theory describes elements manifesting behavior that is extremely sensitive to initial conditions, does not repeat itself and yet is deterministic. Complexity theory goes one step beyond chaos and is attempting to explain complex behavior that emerges within dynamic nonlinear systems".*

At present there are several examples of biological phenomena explained according to the theory of deterministic chaos or other nonlinear models: functioning of some neuronal networks (Korn & Faure, 2003), predictability of heart rhythm (Lefebvre *et al.*, 1993), pulsatile secretion of parathyroid hormone (Prank *et al.*, 1995), variability of cytokine receptors in cancer cells (Muc-Wierzgon *et al.*, 2004), functioning of RNA polymerase (Couzin, 2002). The non-linear patterns of gene expression have been extensively studied by Savageau (2001) and by Kauffman (Shmulevich *et al.*, 2005). In the following sections we review the complexity of the genetic information flow during phenotypic expression to conclude that nonlinear theories, such as deterministic

chaos, may better explain some biological phenomena without questioning of the current paradigm of molecular genetics (Chorąży, 2005).

THE CENTRAL DOGMA OF MOLECULAR BIOLOGY AND DETERMINATION OF HUMAN GENOME SEQUENCE

In April 1953, Watson and Crick (1953) published their *Letter to Nature* describing a structure for the salt of deoxyribonucleic acid – DNA. With the exception of some viruses, DNA is the genetic material of all organisms and genetic information is stored digitally, as defined by the order of the nucleotide bases: A,C,G,T. According to John Maynard Smith (2001) approximately 10^9 bits of information is needed for the formation of a complex living organism.

In each cell, DNA exists as very long chains packaged in the form of chromosomes. Humans have 22 pairs of autosomes and two sex-determining chromosomes, X and Y. The basic units of genetic information, the genes, are linearly arranged on chromosomes. According to "the central dogma of molecular biology" formulated by Crick the genetic information flows in principle in one direction: from DNA to RNA to proteins. The gene exerts its effect by having its DNA transcribed into messenger RNA, which is in turn translated into a protein. Every gene consists of several functional components; two main functional units are the promoter region and the coding region. In the promoter region there are specific structural elements that allow a gene to be expressed only in an appropriate cell, and at an appropriate time. These are *cis*-acting elements able to bind protein factors (*trans*-acting elements) that are physically responsible for transcription.

Each human body cell contains a complete set of genes (i.e., the full human genome), but only a fraction of these genes are used (or expressed) in any particular cell, at any given time. According to the current paradigm the genes carry the complete information on the structure and function of a living cell as well as a complex organism. Thus it was presumed that determination of the human genome sequence would allow us to comprehend how the organism functions, predict the molecular background of human disorders, and understand what causes the differences between individuals and between species. Although the completion of the Human Genome Project was celebrated in April 2003, exactly 50 years after the structure of DNA was described, the exact number of human genes encoded by the genome is still unknown (Ohta, 2005). The gene-prediction programs used by the International Human Genome Sequencing Consortium estimated the number of protein-coding genes at around 30 000–40 000, a figure much lower than previous estimates (around 100 000), and only 50–100% greater than the number possessed by the simple roundworm *Caenorhabditis elegans* (about 20 000 genes) (Claverie, 2001). In order to determine the exact number of genes and to locate them in the appropriate chromosome and locus, advanced molecular procedures have to be used. Moreover, these procedures should be based on parallel analysis of the transcript profile (***transcriptome***) and the corresponding set of proteins (***proteome***) of each type of tissue, at different stages of differentiation. One has to remember that all protein-coding sequences (exons) represent less than 2% of nuclear DNA, whereas gene-free DNA stretches are occupied by various repetitive sequences. These sequences comprise almost 45% of the human genome and are believed to play an important role in its stability and evolution (Jurka, 2004). It appears now that the popular belief in the omnipotence of individual genes cannot be upheld: it is the whole genome and its interaction with the environment that are responsible for the functioning of the cell and organism. Moreover, we still know very little on how the information encoded in a linear manner in DNA is converted into the three-dimensional morphological structures of the whole organism. Finally, as pointed out by Chorąży (2005), the current paradigm assuming that nuclear and mitochondrial DNA is the only genetic material completely neglects the contribution of other heritable material provided by the ovum.

WHY PROTEINS OUTNUMBER GENES

The real number and diversity of proteins encoded by the human genome is much higher than the number of genes. The previous estimation of the number of genes in the context of the Human Genome Project was based on the data obtained using computational programs to detect genes by determination of characteristic sequences, as the gene's beginnings and ends, or by comparing the sequence with known genes and proteins. Both strategies have disadvantages: small genes may be missed and not detected; a gene can code for several proteins but is recognized as encoding only one product; some genes can overlap, and there is a growing list of genes coding for different types of RNA only (such as tRNA, siRNA, microRNA), and not for proteins (Szymański & Barciszewski, 2003). So, depending on the computational methods and gene-finding programs used, the predicted number of all human genes is different, and, as we have already mentioned above, has to be verified by intensive work in the laboratory.

Even if we do not know the exact total number of genes, we already understand the reasons for the great difference between the number of genes and proteins. First of all, most eukaryotic genes are composed of coding exons and non-coding introns, and transcripts of many of these genes may undergo alternative splicing. Majority of genes have several splice forms in which specific exons can be excluded or included, and the length of the individual exons can be altered (Matlin *et al.,* 2005). The phenomenon of alternative splicing is quite a common process that affects the biological properties of a protein. According to Croft *et al.* (2000), around 50% of human genes have more than one alternative variant, and in most cases the functional significance of individual variants is poorly understood. The best known examples of alternative splicing include generation of tissue-specific isoforms, and variants with different cellular localization or altered function. For example, tropomyosin gene encodes two isoforms: one is expressed in smooth muscles and the other in nonmuscle cells (Cooper, 2002). Alternative splicing is responsible for altered intracellular localization of the product of Wilm's tumor gene (*WT1*), encoding a protein with four zinc finger motifs at the C terminus. This protein includes (or excludes) a sequence consisting of 17 amino acids in its central region; moreover, three amino acids (lysine, threonine and serine) are present (+KTS) or absent (–KTS) between the third and fourth zinc finger motifs (Fig. 1). Alternative splicing within the WT1 zinc finger region determines whether the protein has affinity for the essential splicing factors or for steroidogenic factor, SF1, in the nucleus: the +KTS isoform is localized in spliceosome sites whereas the –KTS isoform is localized in the nucleoplasm (Larsson *et al.*, 1995; Laity *et al.,* 2000). In many cases, alternatively spliced gene products fulfill different functions. Good examples of these are transcription factor isoforms which, according to the nature of domains, act as activators or repressors of transcription. Repressor activator protein 1 (Rap1p) in *Saccharomyces cerevisiae* is a model transcription factor with a silencing and putative activation domain playing an important role in the expression of glycolytic enzyme genes (Lopez, 1998).

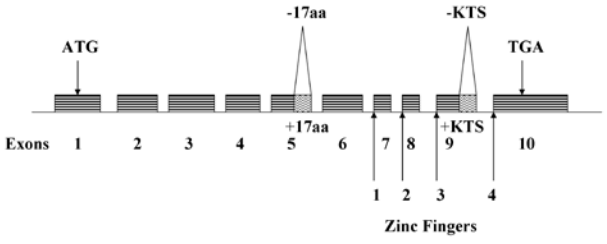

Figure 1. Diagram of the structure of *WT1* gene.
The boxes represent exons. In the C terminal region four zinc fingers motifs are indicated with numbered arrows. Alternatively spliced fragments (inclusion or exclusion of a sequence encoding 17 amino acids in exon 5, and 3 amino acids: lysine, threonine and serine in exon 9) give rise to four isoforms: +17aa, +KTS; –17aa, +KTS; +17aa, –KTS; –17aa, –KTS.

Another source of variation of a polypeptide encoded by one gene is the use of alternative promoters and activation of gene transcription at different sites, as well as the use of alternative polyadenylation sites. Both transcriptional processes contribute to the generation of variants that are tissue-specific, with expression in appropriate cellular organelles and at the proper developmental stage, or with expression associated with sex-specific regulation. An example of at least eight alternative promoters being used is the largest human gene, *DMD* at the Xp21 locus, responsible for Duchenne and Becker muscular dystrophy. Distinct promoters are utilized in lymphocytes, muscle and kidney cells, as well as in various cells of the central nervous system, making it possible to express cell-type specific proteins. The full length gene product consisting of 78 exons exists only in the cortex, muscles and Purkinje cells (Cox & Kunkel, 1997).

An additional mechanism increasing the number of proteins without the need to increase the number of genes is RNA editing. This is a very rare form of post-transcriptional processing involving base-specific alteration in the RNA after transcription but before translation. There are two distinct mechanisms of RNA editing: substitution catalyzed by enzymes that recognize a specific target sequence, and insertion/deletion mediated by guide RNA molecules. Insertion/deletion editing tends to occur in mitochondria and kinetoplastid protozoa and slime molds, while substitution editing is known to occur in human cells, although very rarely. The best documented example of substitution editing in humans is the *APO-B* gene, expressed in the liver and intestine (Driscoll *et al.,* 1989). The gene consists of 29 exons composed of 4564 codons. In the liver, a complete chain of 4563 amino acids (variant of apolipoprotein B-100) is expressed; the protein participates in the transport of cholesterol and other lipids in the blood. In the cells of intestine, chemical modification of the C nucleotide in codon 2153 (CAA) into a U (UAA) takes place and this results in glutamine codon changing to a STOP codon. The reaction is catalyzed by cytidine deaminase. Thus the intestine variant, apolipoprotein B-48, contains 2152 amino acids and takes part in the absorption of lipids from the intestine (Fig. 2). Other examples of substitution editing in human cells include subtle differences in the properties of some receptors of neurotransmitters and some voltage-gated ion channels. The modifications include A→I editing, where adenosine is deaminated to inosine, which normally is not present in mRNA, as is observed in the gluta-

mate receptor (Barbon *et al.,* 2003), and U→C editing in Wilm's tumor gene (*WT1*) (Sharma *et al.,* 1994). Presently, it is difficult to state what is the significance of RNA editing in human cells. Considering the fact that so far we know only a few examples of RNA editing, this phenomenon is not the major or the most important mechanism contributing to the increase in the number of different proteins. On the other hand, in the postgenomic era, we can expect the list of examples of RNA editing in humans to grow.

In addition to the processes already described, post-translational cleavage is another mechanism contributing to generation of a variety of gene products. Polypeptide cleavage is observed in the maturation of some plasma proteins (Brennan, 1989), hormones, neuropeptides (Hook *et al.,* 2004), growth factors (Lu, 2003), etc. Sometimes, cleavage includes only a signal peptide (leader sequence), but may also generate more than one functional polypeptide as in the case of preproinsulin. Also post-translational modifications, such as phosphorylation, methylation, hydroxylation, carboxylation, glycosylation, etc. may change the activity of the individual protein, may contribute to changes in protein–protein interaction or subcellular localization, and may also indicate the fate of the protein, e.g. its destiny for prompt degradation.

The synthesis of plasma glycoproteins in the liver may represent a model of limited determinism of certain biochemical processes in the cell. It is known that attachment of polysaccharides to a polypeptide chain requires the presence of certain amino acids, such as asparagine (Asn), which, moreover, must be spatially available to glycosyltransferases. Glycosylation occurs during migration of nascent polypeptides in the channels of endoplasmic reticulum. The efficiency of glycosylation, and thus the final form of a glycoprotein, depends on many factors: activity of glycosyltransferases, rate of polypeptide migration, concentration of active sugar pecursors used by glycosyltransferases, etc. We, and other authors, have demonstrated significant changes in the glycosylation pattern of liver-produced acute phase glycoproteins during a typical inflammatory response (Koj *et al.,* 1982; Van Dijk & Mackiewicz, 1993). Thus the existence of genetically controlled conditions, such as the presence of available Asn in the polypeptide, or an active specific glycosyltransferase in the endoplasmic reticulum are certainly necessary — but not sufficient - for the synthesis of "mature" plasma glycoproteins; their appearance depends also on variable metabolic conditions prevailing actually in a cell. This example may well illustrate the thesis stating that the expression of genetic information is better described by a model of deterministic chaos rather than a simple linear relationship.

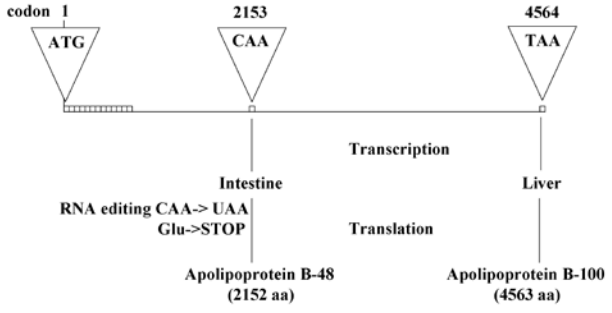

Figure 2. Substitution editing of human apolipoprotein B gene (based on the data of Driscoll *et al.,* 1989).

It appears that not only the number of proteins, but also the number of genes in the genome is in fact higher than the current estimates since some DNA regions can be used as a template for other genes, encoding functionally distinct proteins. Overlapping genes occur more often in simple genomes, such as those of phages and bacteria. Although in human cells only two cases of overlapping genes sharing a common sense strand and using different reading frames are known, there are examples where both strands, sense and antisense, are used as templates in the expression of distinct transcription units. The first case concerns genes for mitochondrial ATPase subunits 6 and 8. These two partially overlapping genes are transcribed in the heavy (H) strand and are translated in different reading frames. Other well-documented examples of overlapping genes have been described in *loci* for the neurofibromatosis type I gene (*NFI),* factor VIII gene (*F8C*) and retinoblastoma gene (*RB1*). Both strands, sense and antisense, are used for transcription. The antisense strand of intron 27 of the *NFI* gene contains three genes: *OGMP* — oligodendrocyte myelin glycoprotein, and *EVI2A* and *EVI2B,* which are homologs of murine genes involved in leukemogenesis (Cawthon *et al.,* 1991). Next, in intron 22 of the blood clotting factor VIII gene there are two genes, *F8A* and *F8B.* The latter is transcribed from the same strand as factor VIII gene. The generated transcript encoded by the *F8B* gene, besides the new exon spliced in intron 22, contains exons 23–26 of the factor VIII gene (Levinson *et al.,* 1992). In the case of the *RB1* gene, in intron 17, there is a coding sequence for a G-protein-coupled receptor gene (*U16*). Several overlapping genes exist in the class III region of the HLA complex in the 6p21.3 region. Also, small nucleolar RNA (snoRNA), siRNA and miRNA genes are located within other genes. It is likely that the continued study of human genome organization will show more examples of genes transcribed from the same stretch of DNA.

RESTRICTIONS IN GENE EXPRESSION

Considering the pattern of tissue-specific regulation, it must be noted that only some of the genes in the human genome are expressed in all types of cells. There are housekeeping genes and tissue-specific genes. The so-called housekeeping genes encode protein products responsible for general functions in all cells. These are, for example, genes encoding proteins engaged in protein synthesis and energy production. According to Hastie and Bishop (1976) and Jongeneel *et al.* (2003), only around 11500–12500 genes are expressed in a given cell type, and of these 9500–10500 are housekeeping genes. The rest are genes representing temporal as well spatial patterns of expression during growth, differentiation and development.

The so-called tissue-specific genes are involved in the functional and phenotypic characteristics of the cell. However, at this point it must be added that tissue-specific gene expression often show the phenomenon of "leakage" or "illegitimate transcription". Chelly and co-workers (1989) used PCR to amplify the cDNA of various tissue-specific genes (genes for anti-Mullerian hormone, β-globin, aldolase A, and factor VIIIc) in human fibroblasts, hepatoma cells, and lymphoblasts. Similarly, experiments performed in rats, where erythroid- and liver-type pyruvate kinase transcripts were detected in brain, lung, and muscle, confirmed that there was "illegitimate" transcription. The occurrence of these "illegitimate" transcripts is very low. For example, in the case of Duchenne muscular dystrophy gene transcripts, fibroblasts and lymphoblasts contain less than one molecule of specific RNA per 500–1000 cells (Chelly *et al.,* 1988). However, the existence of "illegitimate" transcripts provides a powerful tool for geneticists, who identify mutations in pathological transcripts and can use for this purpose any available cells.

In addition to restrictions on gene expression at the spatial and temporal levels, there is monoallelic *versus* biallelic expression: expression of only one of the two parental alleles, although studies on the developing embryo have shown that in mammals and some other animals there is an absolute requirement for a genetic contribution from the maternal and paternal genomes. McGrath and Solter (1984) and Surani *et al.* (1984; 1986) performed experiments with pronuclear transplantation in mice and showed that embryos containing only maternal genetic information develop minimal extraembryonic tissues (trophectoderm), whereas a poorly developed embryo is characteristic of embryos containing only the paternal genome. This experiment demonstrated the requirement for a genetic contribution from both sexes. Monoallelic *versus* biallelic expression concerns only dozens of genes and there are several mechanisms responsible for this phenomenon. One of these is genomic imprinting, where allelic exclusion occurs according to the parental origin (Brannan & Bartolomei, 1999). Elements that contribute to the functioning of imprinting centres and regional propagation of the imprints are CpG-rich differentially methylated regions (which, during development, retain germline-imposed methylation or demethylation), direct repeat clusters, and unusual RNAs (antisense, nontranslated, etc.) (Reik & Walter, 1998). Although numerous studies on genomic imprinting have been conducted in the past few years, our knowledge of imprinting is limited to the identification of imprinted genes and to several factors that contribute to the process.

In the mammalian genome, only a small number of genes are imprinted, and they show monoalleleic expression only in some cell types or certain stages of development. It appears that parental imprinting is a random, stochastic procedure. Examples of imprinting are found in Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS). Both diseases result from either a maternal or paternal deletion on chromosome 15 or from uniparental disomy — inheritance of both chromosomes as a pair from one parent (Ledbetter *et al.,* 1981). The mechanism resulting in monoallelic expression may also be independent of the parental origin. Examples of such expression include X-chromosome inactivation and allelic exclusion after programmed DNA rearrangement. In the first case, X-linked genes differ in dose between females (XX) and males (XY); therefore, in female mammalian embryos, in the late blastocyst stage inactivation of one of the X chromosomes occurs (Lyon, 1999). This process includes chromosomes of both maternal and paternal origin. Females become hemizygous, meaning that they have a single functional copy of each gene, exactly the same as in males. The inactive X acquires numerous features of silent chromatin, including the expression of a noncoding RNA, a switch to late replication, histone modifications, recruitment of the histone variant macroH2A, and DNA hypermethylation. The *XIST* gene plays a major role in X-chromosome inactivation, encoding quite a large RNA (17 kb), which is spliced and polyadenylated but not translated (Brown *et al.,* 1992; Chow & Brown, 2003). An example of monoallelic expression, or allelic exclusion independent of parental origin and following programmed DNA rearrangement, is also observed in the expression of immunoglobin genes in B lymhocytes, T-cell receptor genes in T lymphocytes (Skok *et al.,* 2001; Mostoslavsky *et al.,* 2001) and olfactory receptor genes (Chess *et al.,* 1994).

To control expression at different levels, eukaryotic organisms have developed many differ-

ent regulatory mechanisms. Knowledge about the regulation of all known human genes is far from being complete and further experimental analyses are required. However, we know that all nuclear processes, including gene expression, depend on an architectural framework. Thus, chromosomes in the nucleus are not randomly distributed, but occupy spatially defined subvolumes (Misteli, 2005). Despite the fact that chromosome territories exist, there is a tissue-specific arrangement of chromosomes (Boyle *et al.*, 2001; Parada *et al.*, 2004). It has been suggested that this positioning contributes to proper gene function (Ragoczy *et al.*, 2003). Moreover, bringing DNA and proteins together within a defined sub-region not only influences activation and repression of gene expression but may also be involved in the post-translational modification of proteins by sumoylation and ubiquitylation (Chambeyron & Bickmore, 2004). The best example of how nuclear architecture is important in cell functioning is that of laminopathies. Mutations of genes encoding these structural proteins contribute to weakening of the mechanical stability of nuclei, cell death or alteration in the gene expression pattern (Misteli, 2005).

Besides the importance of nuclear architecture, control at the transcriptional and translational levels seems to be of utmost importance in the regulation of gene expression. Transcriptional regulation occurs through the binding of *trans*-acting factors (transcription factors, hormones) to the *cis*-acting elements in the regulatory region of the gene. Modulation of the expression level may also be achieved by the binding of specific proteins to the regulatory regions of the gene (enhancers, silencers, boundary elements-insulators). The expression may also be regulated at the post-transcriptional level and includes different mechanisms of RNA processing. Some of these mechanisms, such as alternative splicing, alternative polyadenylation and RNA editing have been already described above. In recent years noncoding RNAs have been shown to constitute key elements implicated in a number of regulatory mechanisms in the cell of bacteria and eukaryotes. These types of RNA are involved in regulation of gene expression at both transcriptional and post-transcriptional levels, by mediating chromatin modifications, modulating transcription factor's activity and influencing mRNA stability, processing and translation (Szymanski & Barciszewski, 2003).

SOMATIC RECOMBINATION

The phenomenon of recombination is the source of genetic variations in germ cells, when during the early stages of cell division, in meiosis, two chromosomes of a homologous pair exchange DNA segments. Recombination is also important in somatic cells. Defects in recombination may be associated with an inability to repair damaged or broken chromosomes in somatic cells, resulting in cancer. Somatic recombination also refers to specialized immune cells — B and T cells. The immune system is remarkable in its ability to respond to the vast majority of foreign antigens. The antibodies produced by this system represent the best example of protein diversity. The explanation of the genetic basis of antibody diversity brought Susumu Tonegawa the Nobel prize in 1987 (Tonegawa, 1983).

B and T lymphocytes recognize a great variety of antigens. The immune response can be induced by different molecules, e.g. proteins, lipids, carbohydrates, DNA, etc. The specificity of antigen recognition is determined by the antigen receptors on B and T lymphocytes. An individual B or T lymphocyte is monospecific and produces a single type of immunoglobulin (Ig) and T-cell receptor (TCR). The molecular background of this diversity of proteins is the result of the unique organization of Ig and TCR genes.

The immunoglobulin molecule consists of four polypeptide chains: two heavy and two light ones. The variable part of the light chain of immunoglobulin is encoded by two regions: V (variable) and J (joining), and the heavy chain by three genes: V, D (diversity) and J. The C-terminal segment of the immunoglobulin molecule contains the constant region (C). The variable regions of both types of chains form a pocket located at the N-terminal segment of each chain and specifically bind the antigens. The numbers of V, J, and D genes in our genome are limited. They are organized in clusters on different chromosomes. The appearance of a new antigen in the body results in the replenishment of B- and T-cell clones expressing specific combinations of V, D and J genes and able to bind this antigen. Recombination of VDJ genes greatly enhances the versatility of the immune response and makes it possible to economize the genome size in comparison with a situation in which there were one gene for every antigen. It is obvious that this arrangement makes the notion "one gene – one protein" completely obsolete. Moreover, it points out to the importance of random processes (occurring in deterministic chaos) that are responsible for somatic recombinations.

The rearrangements of V, D, and J gene segments are mediated by RAG1 and RAG2, products of the recombination-activating genes, *RAG-1* and *RAG-2* (Fugmann *et al.*, 2000). Both factors have a long evolutionary history (Kapitonov & Jurka, 2005) and they act as a DNA recombinase (Schatz *et al.*, 1989; Oettinger *et al.*, 1990) that recognizes recombination signals, consisting of conserved nucleotide heptamers and nonamers separated by less con-

served strings of 12 ± 1 or 23 ± 1 nucleotides (Sakano *et al.*, 1979; Akira *et al.*, 1987).

Besides somatic recombination some additional mechanisms contribute to the diversity of Ig molecules. These include random formation of many different VJ_L and VDJ_H combinations, and alternative joining of D segments (V-D-D-J). The common phenomena additionally increasing the variability of immunoglobulins include imprecise joining of gene segments and addition of nucleotides to the DNA sequence at splice sites. Following the antigen-antibody contact frequent mutations occur in the recombined VDJ_H and VJ_L genes. Additionally, the heavy chain class is often changed during the cell lineage. This phenomenon is termed "class switching" or "isotype switching" and involves joining of the VDJ unit generated by somatic recombination to different segments of constant region (CH) genes. This results in production of antibodies with heavy chains of different classes, such as gamma, alpha, and epsilon.

The T-cell receptor (TCR) molecules are engaged in the cell-mediated immune response to foreign antigens. The molecule consists of two types of chains, and each chain has a variable and a constant region. The TCR heterodimer is usually composed of β and γ chains or, on a minority of T cells, α and δ chains. Both chains of the TCR are glycosylated at sites on their V and C regions. Genes encoding TCRs molecules are located on different chromosomes and are organized in clusters in a similar way as the Ig genes. The TCR diversity is mainly the result of somatic recombination, and the mechanism is the same as in the formation of Ig molecules. Individual gene segments for TCR are separated by the same recombination signal sequences as are found between the Ig gene segments, and the same RAG-1 and RAG-2 protein products (recombinases) are involved in somatic recombination. However, unlike for Ig molecules, somatic hypermutation does not seem to be an important diversity mechanism for TCR.

LIMITS OF DETERMINISM IN THE FLOW OF GENETIC INFORMATION

The "genocentric" approach to the functioning of the living organism based on the omnipotence of individual genes can no longer be upheld (Paszewski, 2005). A growing evidence suggests that DNA nucleotide sequences, although encoding the complete proteome, are unable to regulate directly all biological structures and functions of the cell or organism, as initially defined by the central dogma of molecular biology. We know now that the existing phenotype arises from a complex interaction of the whole genome and various environmental factors. To these factors important in the development and transmission of individual phenotype belong epigenetic instructions — changes of gene function not related to changes in DNA sequences. The most prominent examples of epigenetic mechanisms are: DNA methylation, histone acetylation and, changes in chromatin configuration, RNA interference, and altered protein conformation.

Silencing of genes by DNA methylation is a common mechanism of regulation of gene expression in the development and differentiation of an organism. However, sometimes methylation leads to pathogenic loss of function of a particular gene. For example methylation of CpG islands in promoter regions is associated with inactivation of genes and this type of undesirable effects on gene expression has been described for several tumor suppressor genes in many varieties of cancer (Jones & Laird, 1999). Also histone acetylation may have permissive or inhibitory effects on gene transcription. Certain transcription factors, for example p300/CBP, exhibit histone acetyltransferase activity. By binding to DNA they acetylate chromatin, relax the histone structure and permit the transcription to occur. How important chromatin configuration may be in the regulation of gene expression is shown in cases where endogenous and exogenous genes localized in regions with different level of transcription activity are inhibited or overexpressed. One of the best known examples is the *MYC* oncogene. Its translocation from chromosome 8 to a transcriptionally active immunoglobin region in chromosome 14 leads to overexpression and highly elevated level of the coded protein, and finally to the development of Burkitt's lymphoma.

In eukaryotes, including humans, there is a growing number of well described cases of influence of noncoding RNAs (ncRNAs) on gene expression modulation. The ncRNAs are engaged in chromatin modifications, modulation of transcription factor activity, mRNA processing and stability (Szymanski & Barciszewski, 2003). Discoveries in the field of epigenetics provide the evidence that studies at the transcriptome and proteome level are not sufficient to understand how a complex organism functions.

Conformational changes may alter the native structure of a protein's into a new form, with new properties. Such changes often lead to aggregation of proteins. The best known example are amyloid fibrils which are the feature of a group of late-onset degenerative diseases, such as prion diseases (Prusiner, 1998) and tauopathies characterized by aberrant intracellular aggregation of hyperphosphorylated tau protein (Vega *et al.*, 2005).

When evaluating the flow of genetic information in terms of determinism and reductionism the following constraints should be taken into account:

— DNA nucleotide sequences that occur in the genome and encode proteins, do not determine the

current phenotype that is dependent on the regulation of gene expression in response to challenges of the environment;

— Regulation of gene expression in animals is extremely complex due to the complicated structure and functions of gene promoter elements and additional modulation by hormones and some low-molecular forms of RNA;

— Thanks to the alternative splicing of mRNA, a gene can encode not only one specific peptide, but a whole family of polypeptide chains;

— Rearrangement of coding DNA segments during somatic recombination is a source of great variation in the structure of immunoglobulins that is necessary for antibody function;

— Some phenomena associated with the expression of genetic information are of a random nature: rearrangement of immunoglobulin genes, or parental imprinting of genes;

— Explanation of the processes of utilization of genetic information in the animal organism is further complicated by the phenomenon of emergence (Morowitz, 2002), in which new, unpredictable properties of a system emerge after it has exceeded a certain threshold of complexity (e.g., the emergence of awareness in animals);

— It seems that the mechanism of genetic information flow does not fit the category of linear models based on simple reductionism and hard determinism, but would be better described by nonlinear models such as deterministic chaos. The elements of deterministic chaos in genetic information might influence not only the phenotypic expression but also the rate of evolution. The proof of this conclusion must be provided by compatible mathematical models.

Acknowledgements

This work was partly supported by a grant (P05A01127) from the State Committee for Scientific Research (Poland). The authors are grateful to Professors M. Chorąży and S. Szala (Institute of Oncology, Gliwice, Poland) and to Dr J. Jurka (Genetic Information Research Institute, Mountain View, CA, USA) for helpful suggestions.

REFERENCES

Akira S, Okazaki K, Sakano H (1987) Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* **238:** 1134–1138.

Barbon A, Vallini I, La Via L, Marchina E, Barlati S (2003) Glutamate receptor RNA editing: a molecular analysis of GluR2, GluR5 and GluR6 in human brain tissues and in NT2 cells following *in vitro* neural differentiation. *Brain Res Mol Brain Res* **117:** 168–178.

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* **10:** 211–219.

Brannan CL, Bartolomei MS (1999) Mechanism of genomic imprinting. *Curr Opin Genet Dev* **9:** 164–170.

Brennan SO (1989) Propeptide cleavage: evidence from human proalbumins. *Mol Biol Med* **6:** 87–92.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71:** 527–542.

Cawthon RM, Andersen LB, Buchberg AM, Xu GF, O'Connell P, Viskochil D, Weiss RB, Wallace MR, Marchuk DA, Culver M, *et al.* (1991) cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene. *Genomics* **9:** 446–460.

Chambeyron S, Bickmore WA (2004) Does looping and clustering in the nucleus regulate gene expression? *Curr Opin Cell Biol* **16:** 256–262.

Chelly J, Kaplan JC, Maire P, Gautron S, Kahn A (1988) Transcription of the dystrophin gene in human muscle and non-muscle tissue. *Nature* **333:** 858–860.

Chelly J, Concordet JP, Kaplan JC, Kahn A (1989) Illegitimate transcription: transcription of any gene in any cell type. *Proc Natl Acad Sci USA* **86:** 2617–2621.

Chess A, Simon I, Cedar H, Axel R (1994) Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78:** 823–834.

Chorąży M (2005) Is gene concept facing dethronisation? *Folia Histochem Cytobiol* (Suppl 1) **43:** 9.

Chow JC, Brown CJ (2003) Forming facultative heterochromatin: silencing of an X chromosome in mammalian females. *Cell Mol Life Sci* **60:** 2586–2603.

Claverie JM (2001) Gene number. What if there are only 30,000 human genes? *Science* **291:** 1255–1257.

Cooper TA (2002) mRNA splicing: regulated and differential. In *Encyclopedia of Life Sciences*, www.els.net.

Couzin J (2002) Cell biology. Chaos reigns in RNA transcription. *Science* **298:** 1538.

Cox GF, Kunkel LM (1997) Dystrophies and heart disease. *Curr Opin Cardiol* **12:** 329–343.

Croft L, Schandorff S, Clark F, Burrage K, Arctander P, Mattick JS (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat Genet* **24:** 340–341.

Driscoll DM, Wynne JK, Wallis SC, Scott J (1989) An in vitro system for the editing of apolipoprotein B mRNA. *Cell* **58:** 519–525.

Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG (2000) The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* **18:** 495–527.

Hastie ND, Bishop JO (1976) The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9:** 761–774.

Higgins JP (2002) Nonlinear systems in medicine. *Yale J Biol Med* **75:** 247–260.

Hook V, Yasothornsrikul S, Greenbaum D, Medzihradszky KF, Troutner K, Toneff T, Bundey R, Logrinova A, Reinheckel T, Peters C, Bogyo M (2004) Cathepsin L and Arg/Lys aminopeptidase: a distinct prohormone processing pathway for the biosynthesis of peptide neurotransmitters and hormones. *Biol Chem* **385:** 473–480.

Jongeneel CV, Iseli C, Stevenson BJ, Riggins GJ, Lal A, Mackay A, Harris RA, O'Hare MJ, Neville AM, Simpson AJ, Strausberg RL (2003) Comprehensive sampling

of gene expression in human cell lines with massively parallel signature sequencing. *Proc Natl Acad Sci USA* **100:** 4702–4705.

Jones PA, Laird PW (1999) Cancer epigenetics comes of age. *Nat Genet* **21:** 163–167.

Jurka J (2004) Evolutionary impact of human Alu repetitive elements. *Curr Opin Genet Dev* **14:** 1–6.

Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *Plos Biol* doi: 10.1371

Koj A, Dubin A, Kasperczyk H, Bereta J, Gordon AH (1982) Changes in blood level and affinity to concanavalin A of rat plasma glycoproteins during acute inflammation and hepatoma growth. *Biochem J* **206:** 545–553.

Korn H, Faure P (2003) Is there chaos in the brain ? Experimental evidence and related models. *C R Biol* **326:** 787–840.

Laity JH, Dyson HJ, Wright PE (2000) Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc Natl Acad Sci USA* **97:** 11932–11935.

Larsson SH, Charlieu JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, Ross A, Cuzin F, van Heyningen V, Hastie ND (1995) Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* **81:** 391–401.

Ledbetter DH, Riccardi VM, Airhart SD, Strobel RJ, Keenan BS, Crawford JD (1981) Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. *N Engl J Med* **304:** 325–329.

Lefebvre JH, Goodings DA, Kamath MV, Fallen EL (1993) Predictability of normal heart rhythms and deterministic chaos. *Chaos* **3:** 267–276.

Levinson B, Kenwrick S, Gamel P, Fisher K, Gitschier J (1992) Evidence for a third transcript from the human factor VIII gene. *Genomics* **14:** 585–589.

Lopez AJ (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32:** 279–305.

Lu B (2003) Pro-region of neurotrophins: role in synaptic modulation. *Neuron* **39:** 735–738.

Lyon MF (1999) X-chromosome inactivation. *Curr Biol* **9:** R235-7.

Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: towards a cellular code. *Nature* **6:** 386–398.

Maynard Smith J (2001) Evolution and information. In *Images of the World – Science, Humanities, Art* (Koj A, Sztompka P, eds) pp 13–17, Uniwersytet Jagiellonski, Krakow.

McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37:** 179–183.

Misteli T (2005) Concepts in nuclear architecture. *Bioessays* **27:** 477–487.

Morowitz HJ (2002) *The Emergence of Everything*, Oxford University Press.

Mostoslavsky R, Singh N, Tenzen T, Goldmit M, Gabay C, Elizur S, Qi P, Reubinoff BE, Chess A, Cedar H, Bergman Y (2001) Asynchronous replication and allelic exclusion in the immune system. *Nature* **414:** 221–225.

Muc-Wierzgon M, Nowakowska-Zajdel E, Kokot T, Sosada K, Zubelewicz B, Wierzgon J, Cichocka M, Fatyga E, Brodziak A (2004) On the holistic approach in cancer biology: tumor necrosis factor, colon cancer cells, chaos theory and complexity. *J Biol Regul Homeost Agents* **18:** 261–267.

Oettinger MA, Schatz DG, Gorka C, Baltimore D (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248:** 1517–1523.

Ohta T (2005) Gene families, multigene families and superfamilies. *Nature Encyclopedia of the Human Genome,* http://www.ehgonline.net/contents.asp

Parada LA, McQueen PG, Misteli T (2004) Tissue-specific spatial organization of genomes. *Genome Biol* **5:** R44.

Paszewski A (2005) What is determined and and what random in biological systems — when does freedom begin? *Nauka* **1:** 53–66 (in Polish).

Prank K, Harms H, Brabant G, Hesch RD, Dammig M, Mitschke F (1995) Nonlinear dynamics in pulsatile secretion of parathyroid hormone in normal human subjects. *Chaos* **5:** 76–81.

Prusiner SB (1998) Prions. *Proc Natl Acad Sci USA* **95**: 13363–13383.

Ragoczy T, Telling A, Sawado T, Groudine M, Kosak ST (2003) A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res* **11:** 513–525.

Reik W, Walter J (1998) Imprinting mechanisms in mammals. *Curr Opin Genet Dev* **8:** 154–164.

Sakano H, Huppi K, Heinrich G, Tonegawa S (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280:** 288–294.

Savageau MA (2001) Design principles for elementary gene circuits: elements, methods and examples. *Chaos* **11:** 142–159.

Schatz DG, Oettinger MA, Baltimore D (1989) The V(D)J recombination activating gene, RAG-1. *Cell* **59:** 1035–1048.

Sharma PM, Bowman M, Madden SL, Rauscher FJ 3rd, Sukumar S (1994) RNA editing in the Wilms' tumor susceptibility gene, *WT1*. *Genes Dev* **8:** 720–731.

Shmulevich I, Kauffman SA, Aldana M (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc Natl Acad Sci USA* **102:** 13439–13444.

Skok JA, Brown KE, Azuara V, Caparros ML, Baxter J, Takacs K, Dillon N, Gray D, Perry RP, Merkenschlager M, Fisher AG (2001) Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat Immunol* **2:** 848–854.

Surani MAH, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308:** 548–550.

Surani MAH, Barton SC, Norris ML (1986). Nuclear transplantation in the mouse: heritable differences between parental genomes after activation of the embryonic genome. *Cell* **45:** 127–136.

Szymanski M, Barciszewski J (2003) Regulation by RNA. *Int Rev Cytol* **231:** 197–258.

Tonegawa S (1983) Somatic generation of antibody diversity. *Nature* **302:** 575–581.

Van Dijk W, Mackiewicz A (1993) Control of glycosylation alterations of acute phase glycoproteins. In *Acute Phase Proteins* (Mackiewicz A, Kushner I, Baumann H, eds) pp 559–580, CRC Press, Boca Raton, Ann Arbor, London, Tokyo.

Vega IE, Cui L, Propst JA, Hutton ML, Lee G, Yen SH (2005) Increase in tau tyrosine phosphorylation correlates with the formation of tau aggregates. *Brain Res Mol Brain Res* **138:** 135–144.

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. *Nature* **171:** 737–738.