Rendiconti
Accademia Nazionale delle Scienze detta dei XL
Memorie di Scienze Fisiche e Naturali
124° (2006), Vol. XXX, P. II, pp. 291-300

ERNESTO DI MAURO [1] – LOREDANA VERDONE [2]

Epigenetic mechanisms and genome evolution

Summary – According to the Cartesian method, better dubbed in the variant we are more familiar with as "Structural" approach, evolution can be envisaged at four levels: molecular, genetic, epigenetic, behavioural/symbolic. The four levels are not necessarily linked by causal relationships, nor are they easily separated. Genetic evolution is partly superimposed to its epigenetic counterpart, behavioural evolution is grafted on these two, and both depend on molecular evolution. Deepest interconnection among all aspects is, therefore, intrinsic to the concept of evolution itself.

To talk about evolution is to talk about the very nature of the living matter. The thread of thought at this point is already irretrievably entangled: the Aristotelian logics (the logics we use to solve the simple day-by-day problems; the logics based on yes-or-not, tertium non datur; the logics by which our computer carries out its calculations) does not fit with the logics of the Living, to a logics that faces a reality that is so complex to be indefinable even in its essential components. Even in formulating the most basic questions we have to resort to a stringent approach, to the same indispensable consequentiality used to program, to carry on and to interpret any sensible experiment. The reality of the Living apparently escapes from this approach.

The definition of life

What is, then, that evolves in a living system? The quest for the definition of life is a hard and unsolved task. Whatever the proposed solution, exceptions can be raised: crystals, viruses, Gaia and the Planetary system to which it belongs, are they all definable as living systems?

[1] Dipartimento di Genetica e Biologia Molecolare, Università degli Studi "La Sapienza", Roma. E-mail: ernesto.dimauro@uniroma1.it
[2] Istituto di Biologia e Patologia Molecolari, CNR, Roma.
E-mail: loredana.verdone@uniroma1.it

The current definition of life is the one adopted by NASA: '*a self-sustaining chemical system capable of Darwinian evolution*' (Joyce *et al.*, 1994). It is obviously very interesting that the two concepts ("life" and "evolution") are here being linked in identificative terms. To live and to evolve are part of the same property. The third founding aspect of the system is "self-sustenance". Life must always and continuously sustain itself. In order to be able to carry out its functions and to replicate itself, a living structure must interact, be in dynamic equilibrium and co-evolve with the environment that hosts it. If the environment changes, only the organisms which change alongside will survive. For a living structure, evolution is means and aim at the same time.

Informational polymers

The origin of informational polymers (DNA, RNA, proteins, glyco- and lipo-structures) cannot find its justification outside the thermodynamic rules and the chemical scenario of reference. Polymeric information began its own self-organization, replicated itself and evolved starting from the physico-chemical structure of the first environment. To our knowledge, today on Earth organisms evolve on a nucleic acid-based information consisting in particular of a 'polyanionic ribbon' structure. Such structure is characterized by a linear matrix along which nucleotidic appendixes are arranged, kept open and available for replicative interactions by chemical groups of the same charge, easily undergoing ionization. Our informational polymers are essentially based on the linear repetition of phosphate groups. It was proposed, and accepted, that sequences of poly-electrolytes (poly-anionic or poly-cationic) are the diagnostic signature of life (Westheimer, 1987; Benner *et al.*, 2002, 2004). The synthesis of non-ionic linear structures has been the focus of intense research, from the pioneering observations by Pitha *et al.* (1970), to the recent synthesis of gaNA (Bean *et al.*, 2006). An important alternative polymer endowed of the characteristics of stability, autonomous replication and encoding capacity comparable to those of RNA and DNA is Peptide Nucleic Acid (PNA-Nielsen *et al.*, 1991), a polymer whose fundamental structure is a pseudo-peptidic skeleton consisting of N-(2-amino-ethyl)glicine to which the nucleobases are bound through methylencarbonilic linkers. Thus, PNA is a sort of natural DNA-protein hybrid, possibly a prebiotic molecule of great interest.

The first replicative molecule is not known. It is clear, however, that the first informational polymer was necessarily not what we know today: in the 'warm little pond' imagined by Darwin (1888) as the original cradle of evolution, many different chemical molecules were reasonably present.

It is the choice based on *the fittest will survive* that has determined the nature of chromosomes, exactly as we know them.

Chemical evolution

The chemical context to which we can refer is, plausibly, the formamide (H_2NCOH) chemistry. Formamide is formed by the hydrolysis of hydrogen cyanate (HCN) in the presence of water (H_2O). This molecule is stable in liquid form over a wide range of temperatures: 4-210°C; it reacts with several catalysts to produce all the precursors necessary for the synthesis of nucleic acids (Saladino *et al.*, 2004), including all the nucleoside species that favor the polymerization processes; it promotes their phosphorylation (Saladino *et al.*, 2006) and the protection of the generated polymers (Ciciriello *et al.*, 2006). The synthesis of nucleobases from formamide and their formamide-based phosphorylation would remain a biologically futile series of events if the newly formed polymers were not in the thermodynamic conditions allowing their survival as macromolecules.

The main interest of the origin of informational polymers is not simply focused towards understanding their chemical behavior. The main unresolved questions are: which were the physico-chemical conditions in which the first self-generated polymers appeared? Which are the chemical forces these ancestor molecules had to deal with?

The presence of formamide allows lipidic micelles formation. Probably, our simple heterogeneous molecular precursors have been protected by these self assembling structures.

The first molecules replicated themselves, they were genotype and phenotype at the same time. They solved the old bias of the egg and the chicken … *quod negant omnino posse reperiri, avesne ante an ova generate sint, cum et ovum sine ave et aves sine ovo gigni non possit* (Censorino, De die natali, IV, 3).

Living/Not-living

Where can the line between the Living and Not-living be drawn? In logical terms a defined punctuation does not exist. It is the Achilles-and-turtle paradox, unsolved aporia of systems whose properties are somehow interconnected but doomed to never mix. Every evolutionary tree is leading back to one central point, the common ancestor known to be a nucleic acid able to accumulate information, to protect itself (probably as proto-chromatin) from environment attacks, to allow evolution.

It is genetically sound to define as neo-Darwinian evolution the ability of the genetic information to adapt to environmental changes. More precisely, a genetic system cannot be defined as a living system: it can be reproduced in vitro, and in vitro it can replicate and evolve. A possible solution, notwithstanding the NASA definition accepted by most scientists, is that the Living starts to differentiate from the Not-living after the appearance and evolution of epigenetic properties.

What is epigenetics? And why it is important to discuss it in this context? The problem of the egg and the chicken is essentially due to the overlapping of two con-

cepts: the *phenotype*, how does the organism look, the working metabolic pathways, the muscle cell that contracts, the neuron that transports or elaborates the nervous impulse, the bone's saline and cellular structures. Our body is the phenotype.

The *genotype*: the DNA, the molecule containing all the information necessary to synthesize muscle proteins, to form neurons, to drive the metabolic pathways that organize the skeleton's mineral deposits. Moreover, DNA is not simply pure information. DNA must also be considered as a phenotype, it is a chemical molecule in which, like in any other molecule, atoms and chemical bonds are perfectly organized. DNA is phenotype and genotype at the same time.

If DNA is not pure information, if it is egg and chicken at the same time, if DNA derives from self-driven organization, if it is able to self-replicate according to chemical, thermodynamic and environmental rules, if all this is true, then where is the boundary line between Living and Not-living? If information derives from the combination of simple atoms (the most simple and abundant in the universe, hydrogen, oxygen, carbon, nitrogen) and if these were simply connected to become formamide, and if this latter simply gave adenine, tymine, uracil and all the components of nucleic acids. And if these compounds, easily phosphorylated onto ancient rocks, bound together to survive in thermodynamic niches set up by formamide, and slowly, nucleotide by nucleotide, bit by bit of new information, started to interact with the aminoacids produced by the same simple and spontaneous syntheses, creating chromatin and gaining a precise shape. If this, or one of the many other possible and similar paths were followed, it is then extremely difficult to draw the line where chemistry and thermodynamic end and the Living starts. Paradoxically, we maybe do not exist as living beings. Or, as paradoxically, we are not able to define our own living state. It is at this point that what we call epigenetics may provide an explanation and a meaning.

Molecular epigenetic mechanisms

Several definitions exist of epigenetics, a word born with embryology and genetics since the very beginning of both sciences. In terms of molecular meaning, epigenetics describes the heritable changes in genome function occurring by addition of specific chemical groups to DNA without modification of its nucleotide sequence. Epigenetics can be considered as the history of each individual's life engraved on DNA and on the proteins that regulate its function. When a cell has established a particular pattern of 'active' and 'non-active' genes this very pattern will be passed on to daughter cells during cell division. This process allows a marked increase of the informational capacity of the mechanisms governing genetic inheritance. Such variability of the individual rather than of the species is the most important ontological feature of the Living. Epigenetics also indicates the ensemble of regulatory mechanisms based on the interactions between chromosomal DNA and specific RNAs; a field now in active development.

The molecular mechanisms

The nucleotides of the DNA molecule may be chemically modified by different chemical mechanisms: methylation, acetylation, ADP-ribosylation, ubiquitination, sumoylation, phosphorylation.

The most important epigenetic modification of DNA is the methylation of cytosine. The addition of a CH_3 methyl group induces a modification of the cytosine ability to interact with the guanine and with the surrounding environment. In normal cells, DNA methylation occurs predominantly in repetitive genomic sequences, including satellite DNA and parasitic elements [such as long interspersed transposable elements (LINES), short interspersed transposable elements (SINES) and endogenous retroviruses] and, in particular, in specific tracts of CpG, dubbed CpG islands.

The mammalian DNA methylation machinery is composed of two elements, the DNA methyltransferases (DNMTs), which establish and maintain DNA methylation patterns, and the methyl-CpG binding proteins (MBDs), involved in 'reading' methylation marks. Clear evidence also exists that a DNA demethylase contributes the regulation of DNA methylation patterns during embryonic development, although the specifically responsible protein has not been identified (Mayer *et al*., 2000). Functionally, DNA methylation is a crucial epigenetic modification of the genome involved in the regulation of many cellular processes, including embryonic development, transcription, chromatin structure, X chromosome inactivation, genomic imprinting and chromosome stability. Consistent with these important roles, a growing number of human diseases have been found to be associated with aberrant DNA methylation. The study of these diseases has provided new and fundamental insights into the roles that DNA methylation and other epigenetic modifications play in development and in normal cellular homeostasis. A compilation of relevant articles on DNA methylation is in Nature Reviews, October 2005: "DNA methylation collection" (Nature Publishing Group).

Other epigenetic modifications are directly involved in the regulation and maintenance of gene expression. Reversible post-translational modification is a key mechanism for the regulation of protein function. Instead of single-site action, many eukaryotic proteins are dynamically modified at multiple sites by a diverse array of covalent modifications, including phosphorylation, lysine acetylation, methylation of both lysine and arginine residues, ubiquitination, ADP-ribosylation and sumoylation. All these reactions play important roles in a wide range of physiological and patho-physiological processes, including inter- and intracellular signaling, transcriptional regulation, DNA repair pathways and maintenance of genomic stability, telomere dynamics, cell differentiation and proliferation, necrosis and apoptosis.

The chromatin level

The DNA of a single eukaryotic cell is packed in the nucleus according to a hierarchical folding scheme. At the first level of organization, almost two tight superhelical turns of DNA (146 base pairs) are wrapped around an octamer of two copies each of the four histone proteins H2A, H2B, H3 and H4. This unit, dubbed the nucleosome core particle (NCP), is the basic repeating structure in chromatin and is invariant over the whole eukaryotic kingdom (Kornberg and Thomas, 1974; Luger *et al.*, 1997). Nucleosomes are connected to each other by 10-90 base pairs of linker DNA depending on cell type, organism and physiological status. This DNA tract interacts with the histone H1, also called the linker histone because it binds to the outer part of nucleosomes and stabilizes the highly condensed states of chromatin fibers. The simple arrays of nucleosomes along the DNA molecule represent the first level of chromatin structure (the 10 nm fiber, also called "the beads-on-a-string" structure). Most of the chromatin in the nucleus is even more tightly compacted.

Next stage of packaging involves folding the beaded structure into a 30 nm fiber. These fibers may be further folded on themselves into the thicker fibers visible in both metaphase chromosomes and nuclei of non-dividing (interphase) cells. The highly condensed regions of chromosomes are called heterochromatin, while the more open chromatin regions are defined euchromatin; it is in these relaxed chromatin domains that transcription may occur.

In addition to nucleosomes, the chromatin fiber contains a large variety of additional accessory proteins and numerous histone variants which are not randomly distributed but are, at the contrary, expressed in developmentally restricted and cell type specific patterns.

Histone variants were discovered on the basis of the usually limited differences in their amino-acid sequence relative to the major histone species (Sarma and Reinberg, 2005). The variants are usually present as single-copy genes whose expression is not restricted to the replicative S phase but are expressed throughout the cell cycle (Kamakaka and Biggins, 2005). This observation strongly suggests that their incorporation into chromatin might have considerable impact on both the generation and the epigenetic maintenance of regions with specialised chromatin structure (Jin *et al.*, 2005), playing major roles in gene silencing, gene expression and centromere function.

Among histone proteins, no H4 variant has been so far reported. All the other histones have variant counterparts which are found in different species with different abundance.

Histone proteins can be targets of epigenetic modification, carried on by modifying enzymes acting, in a substrate-specific manner, on both their unstructured N-terminals tails and their core regions (Peterson and Laniel, 2004). Combinations of these chemical alterations are thought to modify both the structure and the function of chromatin. Such chemical modification patterns are defined as epigenetic

tags. The specific combination of these epigenetic tags organizes different types of chromatin and is thought to represent, in analogy to the genetic code, a 'histone code'. Histone modifications include lysine acetylation, lysine and arginine methylation, serine phosphorylation, lysine ubiquitylation and sumoylation. All these modifications play major regulatory roles in many genetic events such as transcriptional activation and elongation, silencing and epigenetic cellular memory (Strahl and Allis, 2000; Berger, 2002; Turner, 2002). What appears to be safely established is that the various histone modifications have distinct functional effects and are mediated and recognized by conserved transcriptional regulatory protein modules (Verdone *et al*., 2006).

Acetylation

Among these modifications, acetylation is the one so far more thoroughly analysed (Kurdistani *et al*., 2003). The lysines present in the histones amino termini undergo acetylation-deacetylation switches depending on the different physiological conditions (Kornberg and Lorch, 1999). The balance between these modifications is achieved through the action of enzymes dubbed histone acetyltransferases (HATs) and histone deacetyltransferases (HDACs). These specific enzymes catalyze the transfer of an acetyl group from acetyl-CoA molecules to the lysine amino groups on the N-terminal histones tails (Yang X.J., 2004); analyses performed on large chromosomal domains indicate that the state of acetylation is in a continuous genome-wide flux (Waterborg J.H., 2001; Katan-Khaykovich and Struhl, 2002).

Acetylation of lysines neutralizes histones charges, therefore increasing chromatin accessibility. On the other hand acetylation, like the other covalent modifications, is also important as a signal for the binding of trans-acting factors. Numerous activating factors possess a region dubbed bromodomain specifically interacting with acetylated lysines (Dhalluin *et al*., 1999; Hassan *et al*., 2002; Mujtaba *et al*., 2004; Jacobson *et al*., 2000).

The existence of an informational code overlapped to the genetic sequence information suggests a solution to our initial query: the identification of the boundary between the Living and the Not-living. Before hinting to this solution let's take into consideration two recent studies:

Epigenetic differences arise during the lifetime of monozygotic twins. Fraga M.F., *et al*., Proc. Natl. Acad. Sci. USA (2005) 102:10604-10609.

Sex-specific, male-line transgenerational responses in humans. Pembrey M.E., *et al*., ALSPAC Study Team. Eur. J. Hum. Genet. (2006) 14:159-166.

The starting point of the first paper is that monozygous twins share a common genotype. However, it is easily realized that most monozygotic twin pairs are not identical; several types of phenotypic discordance may be observed, such as differences in susceptibilities to disease and a wide range of anthropometric features. Several possible explanations exist for these observations, one being the existence of epi-

genetic differences. Although twins are epigenetically indistinguishable during their early years of life, older monozygous twins exhibit remarkable differences in their overall content and genomic distribution of 5-methylcytosine DNA and of histone acetylation patterns, affecting their gene expression profile. These findings indicate the relevance of the fact that epigenetics is still largely overlooked in our understanding of how different phenotypes are originated from the same genotype.

The second study mentioned starts from the observation that the trans-generational effects of maternal nutrition or other environmental 'exposures' are well recognized determinants of future phenotypes, and that at the contrary the possibility that the exposure of the male influencing development and health in the next generation(s) is rarely considered. The authors show that male-line transgenerational responses exist in humans and that these transmissions are mediated by the sex chromosomes. Such responses open an entirely new dimension to the analysis of gene–environment interactions in development and health.

These and other recent studies, show that during their life genomes are modified and keep memory of the changes. The memory is maintained for a very short time (measured as fraction of seconds, like at the promoter of inducible genes transcription), for the entire human life (as for the twins in the Fraga's paper) or for a very long time, measured in millions of years (as in the case of the extreme gene silencing observed for the Phasmatoidea wings).

Ilya Prigogine said 'Life is time inscribed in matter'. These words, pronounced during a Conference on the New Alliance at the beginning of the '90s, may provide a solution to the definition of a border between Living and Not-living. From the first atoms formed at the origin of Universe to the DNA of our chromosomes there is no functional interruption, there is no separation between time and matter, there is only an increasing complexity, an increasing level of organization. The nature of the materials and the physico-chemical quality of the system do not change. Strictly, our chromosomes are objects as inanimate as the first informational polymers auto-assembled at the dawn of the world. Just a little more complex.

A functional genome, a genome that designs around itself a genotype to protect it, able to use and transform energy, survive and reproduce (essentially for keeping memory of itself), that very genome during this process writes on itself all the changes that it experiences, retains and uses them in different moments of its life, conveys them to the next generations. This mechanism makes it possible for the genome of the Species, almost an abstract entity and for the individual simply a starting point, to become in each organism a unique unrepeatable genome, resulting from different genetic combinations and from epigenetic modifications. Paraphrasing Prigogine, life is the time inscribed in DNA. It is not a great sin if this aphorism appears increasingly shaded of Lamarckism. Even Darwin would have not opposed to the evolution of his own thought.

REFERENCES

Benner, S.A., Ricardo, A., and Cardigan, M.A., Is there a common chemical model for life in the universe? *Curr. Opin. Chem. Biol.* (2004) 8, 672-689.

Benner, S.A., and Hutter. D., Phosphates, DNA, and the search for nonterrean life: a second generation model for genetic molecules. *Bioorg. Chem.* (2002) 30, 62-80.

Berger, S.L., Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* (2002) 12, 142-8.

F. Ciciriello, G. Costanzo, C. Crestini, R. Saladino, E. Di Mauro, Origin of informational polymers and the search for non-terran life: protection of the polymeric state of DNA by phosphate minerals. *Astrobiology* (2006) in press (accepted 14 oct. 2006).

Darwin, F., ed. (1888), *The Life and Letters of Charles Darwin*, London: John Murray, vol. 3, p. 18, letter to Joseph Hooker.

Dhalluin C., Carlson J.E., Zeng L., *et al.*, Structure and ligand of a histone acetyltransferase bromodomain. *Nature* (1999) 399, 491-496.

DNA by strand displacement with a thymine-substituted polyamide. *Science* (1991) 254, 1497-1500.

Hassan A.H., Prochasson P., Neely K.E., *et al.*, Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* (2002) 111, 369-379.

Jacobson R.H., Ladurner A.G., King D.S., *et al.*, Structure and function of a human TAFII250 double bromodomain module. *Science* (2000) 288, 1422-1425.

Jin J., Cai Y., Li B., Conaway R.C., Workman J.L., Conaway J.W., Kusch T., In and out: histone variant exchange in chromatin. *Trends Biochem. Sci.* (2005) 30, 680-687.

Joyce G.F., Young R. (Chair), Chang S., Clark B., Deamer D., De Vincenzi D., Ferris J., Irvine W., Kasting J., Kerridge J., Klein H., Knoll A., and Walker J. (1994), In *Origins of Life: The Central Concepts*. Deamer D.W., Fleischaker G.R., Eds. Boston, Jones and Bartlett.

Kamakaka R.T., Biggins S., Histone variants: deviants? *Genes Dev.* (2005) 19, 295-310.

Katan-Khaykovich Y., Struhl K., Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev.* (2002) 16, 743-752.

Kornberg R.D. and Thomas J.O., Chromatin structure; oligomers of the histones. *Science* (1974) 184, 865-868.

Kornberg R.D., Lorch Y., Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* (1999) 98, 285-94.

Kurdistani S.K., Grunstein M., Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell. Biol.* (2003) 4, 276-284.

Luger K. *et al.*, Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* (1997) 389, 251-260.

Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T., Demethylation of the zygotic paternal genome. *Nature* (2000) 403, 501-502.

Mujtaba S., He Y., Zeng L., *et al.*, Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell.* (2004) 13, 251-263.

Nielsen, P.E., Egholm, M., Berg, R.H., and Buchardt, O., Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* (1991) 254, 1497-1500.

Peterson, C.L., Laniel, M.A., Histones and histone modifications. *Curr. Biol.* (2004) 14, 546-551.

Pitha, J., Pitha, P.M., Ts'o, P.O., Poly(1-vinyluracil): the preparation and interactions with adenosine derivatives. *Biochim. Biophys. Acta.* (1970) 204, 39-48.

Saladino R., Crestini C., Neri V., F. Ciciriello, Costanzo G., Di Mauro E., Origin of informational polymers: the concurrent roles of formamide and phosphates, *Chem. Bio. Chem*. (2006) 7, 1707-1714.

Saladino R., Crestini C., Ciambecchini U., Ciciriello F., Costanzo G., and Di Mauro E., Synthesis and degradation of nucleobases and nucleic acids by formamide in the presence of montmorillonites. *Chem. Bio. Chem*. (2004) 5, 1558-1566.

Saladino R., Crestini C., Ciciriello F., Di Mauro E., Costanzo G., Origin of informational polymers: Differential stability of phosphoester bonds in ribo monomers and oligomers. *J. Biol. Chem.* (2006) 281, 5790-5796.

Sarma K., Reinberg D., Histone variants meet their match. *Nat. Rev. Mol. Cell. Biol.* (2005) 6, 139-149.

Strahl B.D., Allis C.D., The language of covalent histone modifications. *Nature* (2000) 403, 41-45.

Turner B.M., Cellular memory and the histone code. *Cell* (2002) 111, 285-291.

Verdone L., Agricola E., Caserta M. and Di Mauro E., Histone Acetylation In Gene Regulation, *Briefings in Functional Genomics and Proteomics* (2006) 5, 209-221.

Waterborg J.H., Dynamics of histone acetylation in Saccharomyces cerevisiae. *Biochemistry* (2001) 40, 2599-2605.

Westheimer F.H., Why Nature chose phosphates, *Nature* (1987) 235, 1173-1178.

Yang X.J., Lysine acetylation and bromodomain: a new partnership for signaling. *Bioessays* (2004) 26, 1076-1087.