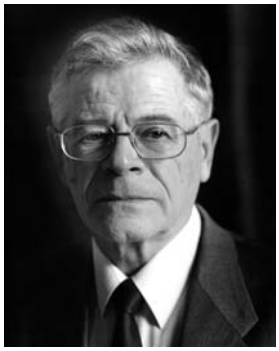
ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 12, pp. 1281-1283. © Pleiades Publishing, Ltd., 2007.
Published in Russian in Biokhimiya, 2007, Vol. 72, No. 12, pp. 1573-1575.



Ab Ovo Usque Ad Mala

A. S. Spirin[1,2]

[1]*Biological Faculty, Lomonosov Moscow State University, 119992 Moscow, Russia*
[2]*Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia; fax: (495) 632-7871; E-mail: spirin@vega.protres.ru*

Received July 25, 2007

DOI: 10.1134/S0006297907120012

Fifty years ago the Soviet journal *Biokhimiya* (*Biochemistry* (Moscow)) published a paper [1] that can be considered as the starting point for three important lines in the development in the field of science called molecular biology. First, analysis of the proportions of four types of nitrogenous bases in DNA—the so-called base composition—in bacteria of various taxonomic groups revealed its great variations: the ratio (G + C)/(A + T) in DNA of taxonomically remote species could differ more than sixfold, whereas relatives displayed similarity of their DNA base composition. For the sake of historical equity, it should be mentioned that in the international scientific literature the work under consideration was not the first where the comparative analysis of DNA base composition of different organisms was undertaken. The priority in the precise determination of proportions of the four nitrogenous bases— adenine (A), guanine (G), cytosine (C), and thymine (T)—in native (undegraded) DNA of various organisms and the discovery of species specificity of the chemical composition of nucleic acids undoubtedly belongs to Erwin Chargaff with associates (1950-1951) [2, 3]. Further, a year before the appearance of our paper in *Biokhimiya*, in 1956, French workers published two reports where the base composition of DNA of a number of bacterial species [4] and animals [5] were analyzed and the possibility of wide variations in DNAs among unrelated species was indicated. Unfortunately, these investigations were not continued in the above-mentioned groups and not completed with fixing of the correlations between the results on DNA base composition and evolutionary systematics of organisms. Only in the subsequent studies of A. N. Belozersky and his colleagues the problem was raised to the proper level, and the work in this direction led to the birth of a novel field of research, called **genosystematics** (see the first publications [6-9], as well as the articles of B. F. Vanyushin and A. S. Antonov in this issue). Later the methods of comparative analysis of nucleotide sequences in DNA and RNA were also used for genosystematics purposes, and this greatly promoted the identification of evolutionarily correct relations between species of all kingdoms of living organisms—animals, plants, fungi, protozoa, and bacteria. This approach was especially useful for detection and correction of errors in the classical systematics when evolutionarily remote species were found to be combined in one taxonomic group. The attraction of the methods of comparative analysis of nucleotide sequences of ribosomal RNA as the evolutionarily most conservative class of nucleic acids (see below) led C. Woese with associates to the construction of the modern genealogical tree of bacteria [10] and the epochal discovery—the revelation of a new super-kingdom (domain) of living beings, called Archaea, in addition to the previously known super-kingdoms (domains) of bacteria (now Eubacteria) and eukaryotes (now Eukarya) [11].

A principally novel result of the paper of 1957 [1], its essence, followed from the determination of base composition of total RNA, in parallel with that of DNA, in the same species of bacteria that were shown to display a great variety of their DNA composition. Earlier attempts to analyze RNA base composition were undertaken in some other laboratories, but they were not successful because of the extreme lability and vulnerability of isolated RNA, especially owing to omnipresent ribonucleases, so that RNA preparations usually contained damaged molecules with altered (corrupted) base composition. In the work under consideration we abandoned the isolation of RNA and used the technique of selective alkaline hydrolysis of RNA *in toto* and *in situ*, i.e. immediate incubation of ethanol-fixed and washed bacterial paste under alkaline conditions, during which the hydrolysis products, monoribonucleotides, were quantitatively released into solution and their mixture could be subjected to chromatographic separation and quantitative analysis. The

chromatographic separation of all four monoribonucleotides was also developed. All this had led us to success: true ("native") base composition of total RNA of a wide variety of bacteria was determined for the first time. The results were surprising. The expectation was that the bulk of cellular RNA, as the putative messenger in the transfer of genetic information from DNA to proteins, should copy the base composition of DNA of the same organism (just with the replacement of thymine for uracil). However, the base composition of the total cellular RNA was found to be dissimilar to DNA composition and demonstrated much higher evolutionary stability compared with DNA. The publication of this result in *Nature* [12] (and later in a chapter of the book *The Nucleic Acids* edited by E. Chargaff and J. N. Davidson [13]) was a sensation and evoked confusion. In 1959 F. Crick wrote: "The coding problem has so far passed through three phases. In the first, the vague phase, various suggestions were made, but none was sufficiently precise to admit disproof. The second phase, the optimistic phase, was initiated by Gamov in 1954, who was rash enough to suggest a fairly precise code. This stimulated a number of workers to show that his suggestions must be incorrect and, in doing so, increased somewhat the precision of thinking in this field. The third phase, the confused phase, was initiated by the paper of Belozersky and Spirin in 1958, although the experimental data had actually been published earlier, both by them and by Lee, Wahl, and Barbu. (*Author's note*: Here Crick made a mistake: Lee, Wahl, and Barbu analyzed only DNA composition, but did not deal with RNA!) The evidence presented there showed that our ideas were in some important respects too simple" (cited from [14], p. 35). F. Jacob and J. Monod described the situation as follows: "It has long been believed that structural information was transferred from the genes to stable templates, such as ribosomal RNA, copied along the genes and maintaining in the cytoplasm the information necessary for protein synthesis. Every gene was supposed to determine the production of a particular type of ribosomal particles, which in turn ensured the synthesis of a particular protein (see Crick, 1958). In recent years, however, this hypothesis has encountered several difficulties. 1. The diversity of base composition found in the DNA of different bacterial species is not reflected in the base composition of ribosomal RNA (Belozersky and Spirin, 1960) ..." (cited from [16], p. 195). Thus, the situation proved to be more complicated than it was first postulated by the "central dogma" of molecular biology in its original form: DNA $\rightarrow$ RNA $\rightarrow$ protein [15]. In any case, the results on the inconsistency (dissimilarity) of the base composition of total cellular RNA with DNA base composition indicated for the first time that the predominant fraction of total cellular RNA, which is the ribosomal RNA, is not a coding RNA. Therefore, it is the paper of 1957 [1] that began the history of the discovery of **non-coding RNAs** in living cells.

Since then, and especially most recently, various non-coding RNAs have occupied solid positions in molecular biology. Now their studies have become one of the most demanded and rapidly developing field of science. In addition to ribosomal RNA, a great variety of other classes of non-coding RNAs have been discovered. Adapter, or transfer RNAs (tRNAs), the second most abundant class of RNA after ribosomal RNA, were also discovered in 1957 [17, 18]; they do not serve as direct templates for protein synthesis either, although they are involved in the process of primary decoding of the genetic code via aminoacyl-tRNA synthetases that catalyze the specific attachment of amino acids to tRNA species. True non-coding RNAs, besides ribosomal RNA, were discovered much later, first of` all because of low abundance of each of their species in cells (see reviews in the book [19]). They include the following classes: RNAs serving as primers for the synthesis of DNA, and RNAs involved in the elongation of telomeres of chromosomes, both processes being strictly required for the reproduction of genes and cells; small cytoplasmic RNAs (scRNA) that participate in regulation of protein synthesis (translation) on ribosomes; the so-called 4.5S and 7S RNAs that play a role of structural scaffolds for assembly on themselves of special proteins and formation of functionally important ribonucleoprotein particles, such as SRP particles, responsible for the export of proteins through the cell membrane; and so on. Ribozymes, the catalytic RNAs discovered in 1982-1983, must be specially mentioned among non-coding RNAs. The most recent (2001) discovery of the class of the so-called microRNAs (miRNA) caused a sensation in science and attracted special interest. It was found that great varieties of the microRNAs (short RNA molecules of just 20-25 nucleotide residues in length, complementary to some sections of mRNAs) are typical minor components of the cells of higher organisms including humans, where they play a key role in the regulation of the syntheses of proteins that determine embryonic development, cell differentiation, and other important processes. Now there are all grounds to believe that the expression of at least one third of our genes is controlled by various microRNAs. On the whole, only about 2% of the human genomic DNA encodes proteins, whereas 80% of the genome is transcribed into numerous species of non-coding RNAs, the functions of which are not yet known. The analysis of the "transcriptome", i.e. the repertoire of RNA, in human and animal cells is one of the hotspots of modern molecular biology (see, e.g. [20]).

The third line of research initiated by the paper of 1957 under consideration [1] was the discovery and further studies of **informational**, or **messenger RNAs (mRNA)**, a special fraction of total cellular RNA that represents copies of structural genes encoding cellular proteins. The existence of a small fraction of RNA whose base composition resembles that of DNA of the same

organism was demonstrated in normal bacterial cells for the first time, whereas the base composition of the major part of the total RNA, presumably the RNA of protein-synthesizing particles, ribosomes, was found to be evolutionarily conservative [1]. These finding implied that ribosomal RNAs cannot be copies of structural genes, and that it is the mission of other RNAs to program the gene-unspecific ribosomal particles and serve as gene-specific templates for ribosome-synthesized proteins. In the scientific literature many authors often refer to the work of Volkin and Astrachan [21, 22] where it was shown a year before (1956) that the infection of *E. coli* cells with bacteriophage T2 induced the synthesis of a short-lived (presumably non-ribosomal) RNA with base composition similar to that of phage DNA. It should be mentioned, however, that this work attracted attention only later (see [16, 23]) when it became clear that the synthesis of phage-specific DNA-like RNA in infected cells is not a specific feature of phage infection, but may indicate the general phenomenon of programming protein synthesis by non-ribosomal RNA. The continuation of this work in brilliant experiments with phage-infected bacteria, the results of which were published almost simultaneously by three different groups of workers in 1961 [24-27], finally established the fact of existence, the position in the sequence of events, and the role of messenger RNA (mRNA) in the process of protein biosynthesis. The work of 1957 [1] was a necessary starting link in the establishment of general biological significance of the discovery of DNA-like RNAs in cells and their key role in the regulation of protein biosynthesis [16, 28].

REFERENCES

1. Spirin, A. S., Belozersky, A. N., Shugaeva, N. V., and Vanyushin, B. F. (1957) A study of species specificity of nucleic acids in bacteria. *Biokhimiya*, **22**, 744-754.
2. Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, **6**, 201-209.
3. Chatgaff, E. (1951) Some recent studies on the composition and structure of nucleic acids. *J. Cell. Comp. Physiol.*, **38**, Suppl. 1, 41-59.
4. Lee, K. Y., Wahl, E., et Barbu, E. (1956) Contenu en bases puriques et pyrimidiques des acides desoxyribonucleiques des bacteries. *Ann. Inst. Pasteur*, **91**, 212-224.
5. Lee, K. Y., et Barbu, E. (1956) Comparison du contenu en bases puriques et pyrimidiques des acides desoxyribonucleiques des quelques especes animals. *Comp. Rend. Soc. Biol.*, **150**, 865.
6. Belozersky, A. N., and Spirin, A. S. (1960) Nucleic acid composition and systematics. *Izvestiya Akad. Nauk SSSR, Ser. Biol.*, **1**, 64-81.
7. Vanyushin, B. F., and Belozersky, A. N. (1959) Base composition of deoxyribonucleic acids of higher plants. *Doklady Akad. Nauk SSSR*, **129**, 944-946.
8. Vanyushin, B. F., Belozersky, A. N., and Bogdanova, S. L. (1960) Comparative study of base compositions of ribonucleic and deoxyribonucleic acids in some fungi and myxomycetes. *Doklady Akad. Nauk SSSR*, **134**, 1222.
9. Antonov, A. S., and Belozersky, A. N. (1961) Comparative study of base composition of deoxyribonucleic acids in some vertebrate and invertebrate animals. *Doklady Akad. Nauk SSSR*, **138**, 1216-1219.
10. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA*, **87**, 4576-4579.
11. Woese, C. R., Stackenbrandt, E., Macke, T. J., and Fox, G. E. (1985) A phylogenetic definition of the major eubacterial taxa. *Syst. Appl. Microbiol.*, **6**, 143-151.
12. Belozersky, A. N., and Spirin, A. S. (1958) A correlation between the compositions of deoxyribonucleic and ribonucleic acids. *Nature*, **182**, 111-112.
13. Belozersky, A. N., and Spirin, A. S. (1960) Chemistry of the nucleic acids of microorganisms, in *The Nucleic Acids* (Chargaff, E., and Davidson, J. N., eds.) Vol. 3, Academic Press, New York, pp. 147-185.
14. Crick, F. H. C. (1959) The present position of the coding problem. *Brookhaven Symp. Biol.*, **12**, 35-39.
15. Crick, F. H. C. (1958) On protein synthesis. *Symp. Soc. Exp. Biol.*, **12**, 138-163.
16. Jacob, F., and Monod, J. (1961) On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.*, **26**, 193-209.
17. Hoagland, M. B., Zamecnik, P. C., and Stephenson, M. L. (1057) Intermediate reactions in protein biosynthesis. *Biochim. Biophys. Acta*, **24**, 215-216.
18. Ogata, K., and Nohara, H. (1957) The possible role of the ribonucleic acid (RNA) of the pH 5 enzyme in amino acid activation. *Biochim. Biophys. Acta*, **25**, 659-660.
19. Gesteland, R. F., Cech, T. R., and Atkins, J. F. (eds.) (2006) *The RNA World*, 3rd Edn., CSHOL Press, Cold Spring Harbor, N. Y.
20. Kapranov, P., Cheng, J., Dike, S., et al. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484-1488.
21. Volkin, E., and Astrachan, L. (1956) Phosphorus incorporation in *Escherichia coli* ribonucleic acid after infection with bacteriophage T2. *Virology*, **2**, 149-161.
22. Volkin, E., and Astrachan, L. (1956) Intracellular distribution of labeled ribonucleic acid after phage infection of *Escherichia coli*. *Virology*, **2**, 433-437.
23. Ycas, M. (1969) *The Biological Code*, North-Holland Publ. Co., Amsterdam.
24. Brenner, S., Jacob, F., and Meselson, M. (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, **190**, 576-581.
25. Gros, F., Gilbert, W., Hiatt, H., Kurland, C., Risebrough, R. W., and Watson, J. D. (1961) Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature*, **190**, 581-585.
26. Hall, B. D., and Spiegelman, S. (1961) Sequence complementarity of T2-DNA T2-specific RNA. *Proc. Natl. Acad. Sci. USA*, **47**, 137-146.
27. Hayashi, M., and Spiegelman, S. (1961) The selective synthesis of informational RNA in bacteria. *Proc. Natl. Acad. Sci. USA*, **47**, 1564-1580.
28. Spiegelman, S. (1961) The relation of informational RNA to DNA. *Cold Spring Harbor Symp. Quant. Biol.*, **26**, 75-90.