

The Discovery of mRNA

Part I (Crick's story): In Its Own Image?

The 1950's was a time when part of molecular biology was unknown, part was understood, and part was thought to be understood when in fact it was not. In this respect, the era was no different from the present. Examining the swirl of insight and confusion of fifty years ago may help us put into some perspective the conceptions (and hidden misconceptions) of our own day.

It was a dark and stormy room.^{1,2,3}

The room belonged to Sydney Brenner, who managed to find himself at many of the turning points in the history of molecular biology. On this day, in the spring of 1960, the point happened to turn in his cramped apartment in Cambridge, not well equipped to house his six guests who had repaired there for a quiet place to talk after a session of the Microbiological Society.

Brenner, a leprechaun with bushy eyebrows, quick wit, and an ear for mimicry, usually filled the air with ideas and rapid-fire stories on any subject. This afternoon he was uncharacteristically quiet, listening to Françoise Jacob go through once again the details of the experiment no one in the room except for him quite believed. Jacob had become known as a graduate student working with Elie Wollman and Andre Lwoff at the Institut Pasteur, when he developed conjugation, the passage of DNA from one bacterium to another, as a powerful tool to understand how the bacteriophage (virus) lambda represses the genes that would otherwise kill its *Escherichia coli* host. Now he had begun a collaboration with Jacques Monod and a visitor, Arthur Pardee, from Berkeley, applying the same techniques to the question of how *E. coli* regulates its own genes.

Jacob was being grilled by Francis Crick (Fig. 1). Crick had met Jacob through a talk he gave at the Pasteur during the missionary phase, the period immediately following the announcement of the proposed double helical structure of DNA. Those were heady times. The structure, with its series of complementary nucleotides offered a tantalizing answer to the ancient question of how the genetic material might replicate itself and pass faithfully from one generation to the next. However, there was a second ancient question whose answer was not so obvious: how does the genetic material dictate, as it must, the workings of the cell? This question occupied Crick and many others for the succeeding decade.



Fig. 1. Francis Crick (1950's).

It had long been clear that the workings of the cell were determined by the activity of its proteins, and Crick, along with many others, were enticed by what he called the Sequence Hypothesis, the proposition that the order of nucleotides in DNA somehow determined the order of amino acids in proteins. This in turn was postulated to be sufficient to determine the proteins' structures and activities and therefore the capabilities of the cell. But how could DNA nucleotides determine protein amino acids?

They couldn't, at least not directly. DNA lives in the nucleus of plants and animal cells, while proteins are synthesized in the cytoplasm. But RNA was known to live in both, and in the mid 1950's, Paul Zamecnik's group demonstrated that small cytoplasmic particles composed of

protein and RNA – now called ribosomes – were the sites of protein synthesis.⁴ Their results provided an explanation for the observed correlation between RNA levels and the amount of protein synthesized.⁵

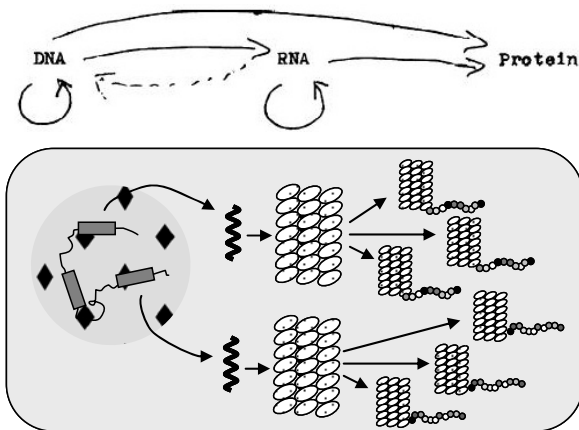


Figure 2. Models relating DNA and RNA to protein synthesis current in the 1950's. (A) Early version of Central Dogma, drawn by Jim Watson. Note that RNA is thought to replicate (arrow pointing to self). (B) One-gene-One-RNA-One-Ribosome-One-Protein hypothesis. In this hypothesis, each gene is thought to produce a single RNA that is central to a ribosome. The ribosome may replicate, but all replicates produce the same protein, different from proteins produced by ribosomes made from the RNA of different genes.

The capabilities of ribosomes stimulated the imagination, and Crick drew a powerful analogy between these particles and a class of plant viruses that were also composed of protein and RNA.⁶ "...I suspect [ribosomes] have a simple structure (akin to plant viruses)," he wrote in 1956 to Sydney Brenner.¹ Crick and Watson made the same suggestion in print a year later.⁷ Both viruses and ribosomes were small, spherical structures, relatively stable, turning over much less rapidly than the rate of protein synthesis. RNA was sufficient to determine the nature of the protein synthesized by the viruses,⁸ and the RNA in ribosomes was thought to be the cytoplasmic image of the DNA in the nucleus. It was easy to imagine ribosomes as viruses put in the employ of the cell -- and viruses as ribosomes gone wild. From this analogy came the unifying idea that one gene gives rise, through its corresponding RNA, to one ribosome, which (perhaps after replicating itself) determines one class of protein (Fig. 2).

To Crick and Jim Watson, further insight into protein synthesis would come from understanding the structure of the ribosome and in particular, the RNA that it contained. After the enunciation of the double helical structure of DNA, Watson went off to Cal Tech in 1953 to attempt to reproduce his earlier success, this time with RNA.

Some ideas are so beautiful that it is difficult to imagine how they could possibly be wrong. One beautiful idea, the replication of genetic information through two strands of DNA carrying complementary information, gained experimental support in the years following Watson and Crick's initial paper.⁹ A second beautiful idea, the connection of genetic information of one gene through one virus-like ribosome to produce one protein, suffered a different fate.

First, despite years of effort, Watson did not find a structure for RNA isolated from cells. Unlike DNA, RNA evidently did not possess a regular structure, certainly not the double helical structure predicted for it from the postulated ability of ribosomes to replicate. Furthermore, ribosomes were shown to be composed of two separable parts,¹⁰ nothing like the unitary capsule of a virus.



Figure 3. Larry Astrachan (left) and Ken Volkin (right). Latter courtesy of Oak Ridge National Laboratory.

Then, in 1956 Ken Volkin and Larry Astrachan (Fig. 3) reported¹¹ that when the bacteriophage T2 infects *E. coli*, the total amount of RNA in the cell does not change, but a small fraction of the total is made within the first few minutes. Since the cell changes over into a factory to make phage protein, you'd expect wholesale turnover of RNA, but this is not the case. The special fraction has a remarkable property that we will come back to later.



Figure 4. A. N. Belozersky (1968; left) and A.S. Spirin (1969; right). Courtesy of *Molecular Biology* (Moscow) and Cold Spring Harbor Laboratory, respectively.

Finally, Crick ran across a paper by Andrei Nikolaevich Belozersky and Alexander Sergeevich Spirin.¹² You can run across it too, by going to the cited reference. This is a highly unusual article, not only in what it says but in the way it says it. If you examine the opening paragraph of almost any random research article, you'll find the authors describing the context in which their work is set. In contrast, Belozerski and Spirin begin their article with a statement of their main result, as if it were old news. The reason is that it *was* old news! They had already published the work over the past two years in two Russian language journals. They were now reproducing it in English.

Belozersky and Spirin measured the base composition – the fraction of adenines (A), cytosines (C), guanosines (G), and either thymines (T) or uracils (U) – in the DNA and RNA from a great variety of bacteria.

Question 1: Given the model shown in Figure 2, what might you have expected to be the relationship between the DNA base composition and RNA base composition in a bacterium? What assumptions underlie that expectation?

Look at Table 1 of the article, showing the base composition of DNA. Scan down the columns showing the percentages of each of the four bases.

Question 2: Do all bacteria have more or less the same base composition?

This was the first shocker. The genetic code was not yet known, but the most popular model¹³ of the day predicted that each of the standard 20 amino acids would be encoded by one and only one triplet of nucleotides. If the genetic code were the same for all organisms – and it was difficult to believe otherwise – and if all organisms had proteins with amino acids in roughly the same proportions (also popularly believed), then it seemed to follow that the DNA base composition for all organisms should be the same. It wasn't.

The DNA base composition varies from organism to organism, but is it total chaos?

Question 3: Is there any general pattern that you can see in all DNA base compositions?

This was not a shocker. Erwin Chargaff had previously shown for bacteria and eukaryotes alike that there was a correspondence between the amount of A and the amount of T and, likewise between the amount of G and the amount of C.¹⁴ These pairing rules was one reason why the double helical structure of DNA with its complementary base pairs was instantly recognized as biologically plausible. Belozersky and Spirin confirmed Chargaff's finding that the fraction of G and A was almost the same as the fraction of C and T, respectively.

The last three columns relate ratios of different groups of bases, the first being the ratio of purines (*Pu*; A and G) to pyrimidines (*Py*; C and T).

Question 4: From Chargaff's rules, what value would you expect (and why) for:

- a. *Pu/Py*
- b. $(G+T) / (A+C)$
- c. $(G+C) / (A+T)$

So far, except for the lack of similarity amongst organisms, Belozersky and Spirin's results corresponded to expectation.

Now turn to Table 2, the RNA base compositions for the same bacteria.

Question 5: Do Chargaff's rules hold for RNA?

That's interesting! If RNA were double stranded, you'd expect the G-C and A-U pairings to lead to similar fractions. The fractions are very different, so RNA – or at least most of it – can't be double stranded. This was not much of a surprise. After all, the plant viruses that inspired the current view of ribosomes also contained single-stranded RNA. Surely they went through a double stranded stage for replication, but most of the RNA in the cell could well be single-stranded.

Now comes the big question. If RNA were the image of DNA, then you would expect to see more or less the same base compositions in both, so:

Question 6: Within a given bacterium, are the DNA and RNA base compositions approximately the same?

It seemed that RNA was *not* the working image of DNA... except perhaps for one known case. The special fraction of RNA Volkin and Astrochan observed to be made immediately after phage infection had a base composition unlike that of other RNA in the cell. It's composition was the same as that of the DNA of the infecting phage! No one knew what to make of this result.

In 1959, Crick summarized the state of affairs,¹⁵ describing the great variation in DNA base composition reported by Belozersky and Spirin as "very unexpected". He spent the remainder of his appraisal trying to rationalize the discrepancy between the DNA and RNA base compositions. Some of the possibilities he raised were:

- a. *The genetic code is not universal.* This was a most unappealing choice, as it would require mutations to have occurred over evolutionary time that would surely have raised immediate havoc within the organisms, as many proteins would have suffered changes.
- b. *Much of the DNA is nonsense, does not encode protein, and is not made into RNA.* We now know that this is partially true,... in eukaryotes. For example, about 98% of the human genome does not encode protein (it isn't nonsense, but that's another story). In bacteria, however, a typical figure is only 10% noncoding DNA, much too small to account for the results of Belozersky and Spirin.
- c. *The genetic code is degenerate,* i.e. there are multiple ways for DNA to encode the same amino acid. This turned out to be true: there are 61 triplet codons to encode 20 amino acids. The degeneracy of the genetic code and the tendency of organisms to prefer certain codons over others is sufficient to explain the variability in DNA base compositions but does not speak to the discrepancy between DNA and RNA.

Crick later spoke about this presentation, "...you can see there, we were completely lost... Didn't know where to turn. Nothing fitted."¹

This was Crick's gloomy state of mind as he listened to Jacob relate yet another experiment that suggested the world was not as it should be.

It was a dark and stormy room.

Question 7: In what way were the results of Belozersky and Spirin seemingly antithetical to the model of ribosomes that prevailed at the time? How can you reconcile their results?

To come: Part II (Jacob's story) Breaking the Mold

Scenario 1:

Two students of English were interested in the range of words in the language. One of them took to the street and wrote down each word as it was spoken. Then he went through the transcript, counting how many one-letter words there were, how many two-letter words, and so forth. When he finished the effort, he showed his finding to his colleague, marveling how English speakers tend towards small words (Fig. 5, solid line).

The second scholar was surprised at this, for she had done almost the same experiment but gotten quite different results. She had hired someone to go to the same street as his colleague, during the same period, and for the same length of time, and to return with an alphabetized list of the words he heard. Then the second scholar counted how many one-letter words there were in the list, etc. She concluded that English speakers tend towards middle-sized words (Fig. 5, dashed line).

How can you reconcile these two results?

Scenario 2:

You are an expert in the structure of automobiles. You have taken apart many automobiles of many different types. In fact, you've gone beyond this, running automobiles through a metal grinder, producing nuts, bolts, and bits of metal. You've done this enough times that you can sift through a bucket of car debris and make a very good guess from the ratio of nuts and bolts to sheet metal whether it came from a Ferrari or a Humvee.

Why are these ratios so different? You intend to find out. You go to a Ferrari factory and grind that up (discarding all the concrete, plastic, and perhaps stray organic material). Then you do the same to a Humvee factory. Comparing the two, you're astonished to find that the ratio of nuts and bolts to sheet metal is about the same! In fact, it doesn't matter which car factory you go to – Toyota, BMW,... – the ratio is pretty constant, even though that ratio varies significantly from car to car.

How can you reconcile your results?

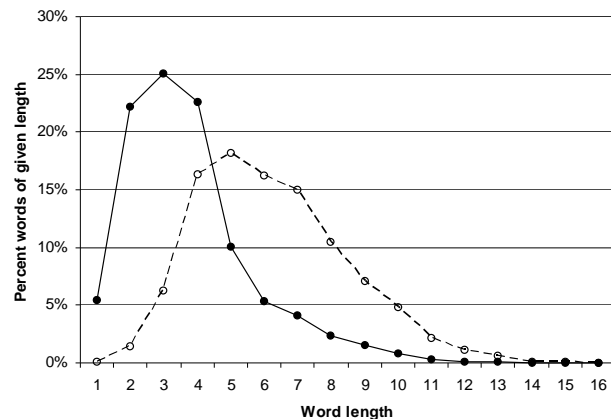


Fig. 5. Spoken English word usage. The frequency of words of a given length are shown, using either the number of times such words are used in speech (solid line) or the number of words of that length (dotted line). Data from reference 16.

References

1. The source for the historical ambience is H.F. Judson (1996), *The Eighth Day of Creation*, Cold Spring Harbor Laboratory Press. Woodbury, 2nd edition, and references 2 and 3.
2. Crick, F. (1988). *What Mad Pursuit*. Basic Books, Inc. New York
3. Jacob, F. (1988). *The Statue Within*. Basic Books, Inc. New York.
4. Littlefield, J. W., *et al.* Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *J. Biol. Chem.* **217**, 111-123 (1955).
5. Caspersson T. & Schultz, J. Pentose nucleotides in the cytoplasm of growing tissues. *Nature* **143**, 602-603 (1939).
6. Crick, F.H.C. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**,, 138-163 (1958).
7. Crick, F.H.C. & Watson, J.D. Virus structure: General principles. In: *Ciba Foundation Symposium on the Nature of Viruses*, G.E.W. Wolstenholme & E.C.P. Millar, eds. J. & A. Churchill, London. pp.5-13.
8. Fraenkel-Conrat, H. The role of nucleic acid in the reconstitution of active tobacco mosaic virus. *J. Amer. Chem. Soc.* **78**, 882-883 (1956).
9. Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
10. Tissières, A. & Watson, J.D. Ribonucleoprotein particles from *Escherichia coli*. *Nature* **182**, 778-780 (1958).
11. Volkin, E. & Astrachan, L. Phosphorus incorporation in *Escherichia coli* ribonucleic acid after infection with bacteriophage T2. *Viol.* **2**, 149-161 (1956).
12. Belozersky, A. N. & Spirin, A. S. A correlation between the compositions of deoxyribonucleic and ribonucleic acids. *Nature* **182**, 111-112 (1958).
13. Crick, F. H. C., *et al.* Codes without commas. *Proc. Natl. Acad. Sci. U.S.A.* **43**, 416-421 (1957).
14. Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**, 201-209 (1950).
15. Crick, F. H. C. The present position of the coding problem. *Brookhaven Symp.* **12**, 35-39 (1959).
16. Carnegie Mellon Pronouncing Dictionary (online). <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>. Accessed October, 2008.