

Fifty years of “Watson–Crick”*

Andrew M. MacMillan[‡]

*Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7,
Canada*

Abstract: Fifty years have passed since Watson and Crick proposed the molecular basis for the replication of nucleic acid and hence the transfer of genetic information. During this time, a model for the expression of this genetic information has been proposed and refined considerably. Coincident with these advances, the chemical synthesis of oligonucleotides combined with the power of molecular biology has facilitated dramatic advances in our understanding of the fundamental workings of the living cell.

STRUCTURE OF DNA

On 27 June 2000, the *New York Times* featured a dramatic banner headline announcing “Genetic Code of Human Life Is Cracked by Scientists”, a reference to the release of the first draft of the human genome. In fact, that code had been “cracked” many years before in an elegant piece of deduction [1] which had roots in one of the most celebrated papers of modern science, the proposal for the structure of DNA [2].

In the spring of 1953, three papers describing structural investigations of the DNA molecule were published in the journal *Nature* [2–4]. One of these, coauthored by James Watson and Francis Crick, put forward a model of DNA structure (Fig. 1A) which featured the now famous antiparallel double helix held together by hydrogen-bonding interactions between complementary base-pairs: adenine·thymine and guanine·cytidine (Fig. 1B). A high-resolution view of this structure was not obtained until Dickerson and coworkers reported the results of their crystallographic studies on the self-complementary DNA dodecamer 5′-CGCGAATTCGCG-3′ in 1980 (Fig. 1C) [5]. The Watson–Crick model represented, in one sense, the culmination of a spirited competition [6] to “solve” the mystery of the structure of DNA, but it also provided the key insight into the mechanism of hereditary information transfer and the means by which this information is communicated within the cell. That insight was communicated in one sentence in the Watson–Crick paper where the authors stated “it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.” This copying, based on complementary base-pairs, results in the replication of DNA and the synthesis of RNA, which then performs various functions in the cell.

One of the functions of RNA (mRNA) is to serve as a template for protein synthesis. Crick realized that the structure of DNA (and RNA) was too regular to directly specify incorporation of one of the 20 amino acids in a growing polypeptide chain. He therefore proposed “the adaptor hypothesis” whereby an intermediate molecule (now known to be tRNA) associated with a specific amino acid could interact with the coding RNA. He also formulated “the central dogma of molecular biology” whereby the transfer of sequence information occurs from DNA (or RNA) to protein but not from protein to nu-

*Adapted from remarks introducing the symposium “Chemistry of nucleic acids”, as part of the 39th IUPAC Congress and 86th Conference of the Canadian Society for Chemistry: Chemistry at the Interfaces, Ottawa, Canada, 10–15 August 2003. Other Congress presentations are published in this issue, pp. 1295–1603.

[‡]Corresponding author: E-mail: andrew.macmillan@ualberta.ca

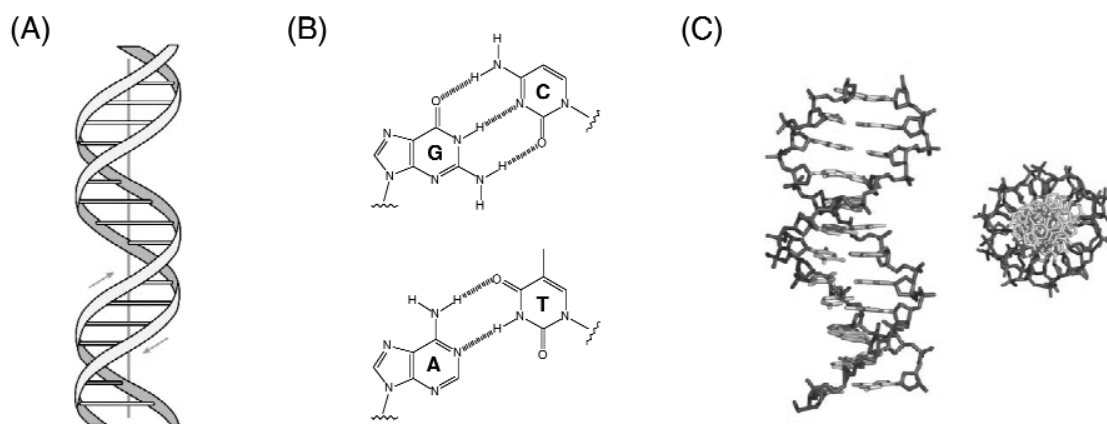


Fig. 1 The structure of DNA. (A) Proposed antiparallel duplex structure featuring external phosphate sugar backbone and internal base-pairs perpendicular to the helix axis (adapted from ref. [2]); (B) canonical Watson–Crick base-pairings; (C) high resolution model of B-form DNA showing side and top views based on crystals of 5′-CGCGAATTCGCG-3′ (nucleic acid database reference: BDL001) [5].

cleic acid [7]. This sometimes misunderstood statement is fundamental to our understanding of the expression of the gene in biological systems. Our description of the *mechanisms* of this expression has been elaborated in unforeseen ways as the complexity of biological systems has been further unveiled. Three examples illustrate this point. First, David Baltimore and Howard Temin showed, through their discovery of reverse transcriptase, an enzyme which synthesizes a DNA copy of RNA, that the flow of genetic information is not unidirectional [8,9]. Second, the split-gene intron/exon structure [10] found within eukaryotes means that coding mRNAs can be assembled in a combinatorial fashion by alternative splicing involving the inclusion or exclusion of different exons [11]. For example, the *Drosophila* gene *dscam* potentially encodes 38 000 distinct isoforms of an axon guidance receptor from a single precursor via this mechanism [12]. Finally, it is becoming increasingly clear that diversification of the immunoglobulin genes through somatic mutation, isotype switch recombination, and gene conversion depends on a complex of proteins, with a DNA mutating protein, activation-induced deaminase, at its heart [13].

DNA, RNA, PROTEINS....RNA

One paradox of the DNA \Rightarrow RNA \Rightarrow Protein model is that the mechanism of this information relay depends on proteins at all stages including the synthesis of DNA and RNA themselves (by DNA and RNA polymerases, respectively). A resolution of this chicken and egg problem would require the existence of a self-replicating biomolecule. Until the early 1980s, this was problematic since only proteins (non-coding according to the central dogma) were known to be capable of catalyzing biological reactions and only nucleic acids transmitted genetic information.

The paradox involved in the separation of coding and catalytic activity was resolved by studies on the post-transcriptional processing of two RNAs. Cech and coworkers were able to demonstrate that removal of Group I class introns from the *Tetrahymena* 26S pre-rRNA depends only on the presence of the pre-rRNA, Mg^{2+} , and a guanosine “cofactor” [14]. The two sequential transesterifications are autocatalytic, promoted by the folded intron itself. Separately, Altman and colleagues showed that the RNA components of the RNA-protein assembly RNase P were the catalytic subunits responsible for the maturation of pre-tRNA in *E. coli* and *B. subtilis* [15]. Thus, RNA, previously regarded as an information-transmitting or structural molecule, was shown to be capable of catalyzing biochemical transformations.

Subsequent to the work on the Group I intron and RNase P, other examples of autocatalysis by RNA have been documented and artificial RNA enzymes (ribozymes) have been created which catalyze key biochemical transformations including RNA synthesis and amino-acyltransfer [16,17]. The recent structure of the large subunit of the ribosome shows an active site exclusively composed of RNA, strongly suggesting that peptide bond formation during protein synthesis is RNA-catalyzed [18]. The involvement of RNA in key steps of gene expression from RNA processing through translation hints at an earlier “RNA World” as the origin of self-replicating biological systems [19].

CHEMICAL AND BIOCHEMICAL GENE SYNTHESIS

Parallel to the elucidation of the structure of DNA and the flood of work which sprang from this was an intense effort by chemists to chemically synthesize the gene. In 1955, Michelson and Todd reported the synthesis of dithymidyl nucleotide [20]. This was followed by somewhat more than two decades of effort on the chemical synthesis of longer oligonucleotides including complete genes. Key advances in this work included the development of a cyclic iterative approach to oligonucleotide synthesis and the use of the solid phase as well as the development of protecting group, phosphite triester and phosphoramidite chemistries with the landmark contributions being made by Khorana, Letsinger, and Carruthers [21–24]. Despite impressive achievements, the total chemical synthesis of the gene was not practical for general application much as the chemical synthesis of polypeptides had a limited utility for the study of specific systems. Instead, the combination of chemical synthesis of relatively short oligonucleotides with the techniques of molecular biology has provided some of the most powerful tools in modern biology. Three operations have been critical to the recent advances in biology based on the gene—these are isolation/synthesis, sequence characterization, and mutation—and all have made use of the chemical synthesis of DNA.

The most common methodology for the synthesis of genes now is the polymerase chain reaction (PCR) invented by Mullis and a rediscovery of the work of Khorana in the early 1970s [25,26]. In this protocol, short synthetic oligonucleotide primers flanking a sequence of interest are used to amplify that sequence through an iterative procedure of annealing, DNA polymerization, and denaturation. The applications of this technique have been myriad, ranging from cloning of genes and gene fragments to forensics and diagnostics. Synthetic oligonucleotides have also been key to the large-scale sequencing projects of the last decade via polymerization off of DNA primers with chain-terminating dideoxynucleotides [27]. As well, chemically synthesized pieces of DNA have facilitated the manipulation of the gene, either through the production of deletion mutants by PCR or by the introduction of point mutations in a particular sequence [28].

Finally, there has been great interest in the application of synthetic oligonucleotides and oligonucleotide analogs for modulation of gene expression including therapeutic purposes [29]. Most of the efforts in this regard have relied on an “antisense” strategy whereby hybridization to a sequence of interest would inhibit translation or trigger RNase H-mediated degradation of an RNA·DNA hybrid. Recently it was reported that double-stranded RNA triggers gene-specific silencing (RNA interference or RNAi), a response which appears to represent an antiviral mechanism as well as a system for regulation of endogenous gene expression [30]. The demonstration that short, synthetic RNA duplexes can induce RNAi in mammalian cells [31] has provided a very powerful tool for cell biology, one that should provide many more insights into the function of gene products in living cells.

ACKNOWLEDGMENTS

I would like to thank K. Chilibeck and O. Kent for help with figures. Research in the laboratory is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), and the Alberta Heritage Foundation for Medical Research (AHFMR).

REFERENCES

1. M. Nirenberg. "The Genetic Code" in Les Prix Nobel en 1968. Nobel Foundation, Stockholm, P. A. Norstedt and Söner, pp. 221–241 (1969).
2. J. D. Watson and F. H. C. Crick. *Nature* **171**, 737 (1953).
3. M. H. F. Wilkins, A. R. Stokes, H. R. Wilson. *Nature* **171**, 738 (1953).
4. R. Franklin and R. G. Gosling. *Nature* **171**, 740 (1953).
5. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson. *Proc. Natl. Acad. Sci. USA* **78**, 2179 (1981).
6. J. D. Watson. *The Double Helix : A Personal Account of the Discovery of the Structure of DNA*, Penguin, New York (1969).
7. F. H. C. Crick. *Symp. Soc. Exp. Biol. The Biological Replication of Macromolecules* **XII**, 138 (1958).
8. D. Baltimore. *Nature* **226**, 1209 (1970).
9. H. Temin and S. Mizutani. *Nature* **226**, 1211 (1970).
10. S. M. Berget, C. Moore, P. A. Sharp. *Proc. Natl. Acad. Sci. USA* **74**, 3171 (1977).
11. T. Maniatis and B. Tasic. *Nature* **418**, 236 (2002).
12. D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, S. L. Zipursky. *Cell* **101**, 671 (2000).
13. C.-A. Reynaud, S. Aoufouchi, A. Faili, J.-C. Weill. *Nature Immun.* **4**, 631 (2003).
14. K. Kruger, P. J. Grabowski, A. J. Zaugg, J. Sands, D. E. Gottschling, T. R. Cech. *Cell* **31**, 147 (1982).
15. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman. *Cell* **35**, 849 (1983).
16. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel. *Science* **292**, 1319 (2001).
17. N. Lee, Y. Bessho, K. Wei, J. W. Szostak, H. Suga. *Nat. Struct. Biol.* **7**, 28 (2000).
18. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz. *Science* **289**, 920 (2000).
19. R. F. Gesteland, T. R. Cech, J. F. Atkins (Eds.), *The RNA World*, 2nd ed., Cold Spring Harbor, Cold Spring Harbor Laboratory Press, NY (1999).
20. A. M. Michelson and A. R. Todd. *J. Chem. Soc.* 2632 (1955).
21. H. Schaller, G. Weimann, B. Lerch, H. G. Khorana. *J. Am. Chem. Soc.* **85**, 3821 (1963).
22. R. L. Letsinger and V. Mahadevan. *J. Am. Chem. Soc.* **87**, 3526 (1965).
23. R. L. Letsinger, J. L. Finnan, G. A. Heavener, W. B. Lunsford. *J. Am. Chem. Soc.* **97**, 3278 (1975).
24. L. J. McBride and M. H. Caruthers. *Tetrahedron Lett.* **24**, 245 (1983).
25. K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263 (1986).
26. K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, H. G. Khorana. *J. Mol. Biol.* **56**, 341 (1971).
27. F. Sanger, S. Nicklen, A. R. Coulson. *Proc. Natl. Acad. Sci. USA* **74**, 5463 (1977).
28. C. A. Hutchison 3rd., S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, M. Smith. *J. Biol. Chem.* **253**, 6551 (1978).
29. S. T. Crook. *Methods Enzymol.* **313**, 3 (1999).
30. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello. *Nature* **391**, 806 (1998).
31. S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, T. Tuschl. *Nature* **411**, 494 (2001).