Nucleic Acids as Genetic Material

Burke H Judd, *Chapel Hill, North Carolina, USA*

Nucleic acids are long linear or circular macromolecules, either DNA or various types of RNA, that are composed of linked nucleotides. These molecules carry genetic information that directs all cellular functions.

Introductory article

Article Contents

Introduction

That the chromatin of the cell nucleus is the physical basis of inheritance was rather widely accepted by scientists in the late 1800s. Nägeli in 1884 had hypothesized that the stuff of inheritance resided in a cellular material he named idioplasm. There soon followed the proposal independently from several scientists, including Weismann, Strassburger and Hertwig, that Nägeli's hypothetical idioplasm was the nuclear chromatin detected by cytological observations. Support for this concept came from studies of sexual reproduction of plants and animals, where it was noted that both male and female parents contribute nuclei to the fertilized egg, while the cytoplasm is derived almost entirely from the female. Even more important, Roux had shown as early as 1883 that mitosis involves the accurate equal division of the nuclear material to the two daughter cells. In contrast, the mass fission of the cytoplasm is not necessarily an equal division of the individual parts.

The chemistry of the cellular contents at that time made it clear that there is a definite and constant difference between the molecular composition of the nucleus and cytoplasm. In 1871, Miescher subjected cells to digestion by gastric juices (essentially proteases) which left only nuclei, from which he obtained substantial quantities of nuclear substance. He called this material nuclein, on which he did accurate quantitative chemical analysis. His early results made it clear that nuclein was rich in phosphorus; however, determinations of its precise chemical composition from different cell types gave somewhat discordant results for amounts of other elements. Altmann, in 1889, opened the way to understanding the nature of nuclein by showing that it could be split into two substances: (1) an organic acid, rich in phosphorus, which he named nucleinic acid, later to be called nucleic acid; and (2) proteins, designated as a form of albumin. Subsequent analysis by a number of chemists, including Hoppe-Seyler, showed that the nucleic acid is completely free of sulfur, while the 'nucleoalbumins' contain no phosphorus. It was Kossel, from his own work, who drew the conclusion that what the cytologists designated as chromatin contains nucleic acid and various amounts of the nuclear proteins

Since nucleic acids were limited to the chromosomes of the nucleus, they were generally accepted as the material basis of inheritance. That correct view came into question in the 1920s, however, as more was learned about nucleic acid structure. From its chemical analysis, deoxyribonucleic acid (DNA) was interpreted to be a monotonous repeating polymer of only four nucleotides. That such a simple structure could encode sufficient information for directing cellular functions and development, therefore, seemed unlikely. Attention was then turned to proteins, which are diverse in type and complexity. The view that proteins are the genetic material temporarily supplanted the role earlier and correctly assigned to nucleic acids.

Proof of DNA as Genetic Material

The experiments of Griffith, in 1928, were one of the first steps toward proof that nucleic acids are the genetic material. He used different strains of the bacterium pneumococcus to demonstrate a genetic 'transformation' of one strain type into another. Different strains of pneumococcus can be distinguished by the type of polysaccharide found in the cell capsule. The capsule type is a constant, inherited characteristic of each strain. Occasionally some cells may lose the ability to form a capsule, but such unencapsulated cells do not change strain type. Encapsulated cells are deadly when injected into mice, while unencapsulated cells have lost their virulence.

Griffith carried out a series of experiments on mice injected with various combinations of pneumococcus strains. He discovered that mice injected with unencapsulated, avirulent cells mixed with an extract of heat-killed encapsulated cells generally died, just as if they had received live virulent cells. Clearly, the extract of heat-killed cells was somehow conferring virulence to the live unencapsulated cells. The crucial observation came from mice injected with live unencapsulated cells of one strain

mixed with the extract of heat-killed encapsulated cells of a different strain. Griffith noted that bacteria recovered from the dead animals proved to be encapsulated with the genetically determined polysaccharide characteristic of the heat-killed strain. Repeat experiments, with all of the necessary controls, indicated not only that association with an extract of dead encapsulated bacteria could in some way restore virulence to the unencapsulated cells, but that those cells were actually genetically transformed into the heat-killed encapsulated strain type. That this was indeed an inherited transformation, and not just the result of coating avirulent cells with the polysaccharide capsule material, was clear from culturing the cells to show that the capsule type was transmitted to daughter cells.

Purification of the transforming material

The experiments of Griffith were subsequently confirmed by other investigators, and in 1931 Dawson and Sia succeeded in inducing transformation *in vitro*. It was not until 1944, however, that Avery, MacLeod and McCarty identified DNA as the agent responsible for bacterial transformation. Even then there was some reluctance among scientists to assign to nucleic acids the role of the genetic material. The experiments of Avery and his colleagues marked a turning point, however, because the purification procedures and the control steps they employed to eliminate an active role for all but trace contaminants were quite convincing. Their experiments utilized encapsulated (S) type III pneumococci, from which they isolated a very highly purified DNA fraction. That fraction was capable of transforming unencapsulated (R) variants of type II cells into fully encapsulated cells of type III, from which the DNA was isolated. One key to eliminating any role of molecules other than DNA in the transformation process was the use of the enzyme deoxyribonuclease (DNAase), which depolymerizes DNA. Avery *et al.* showed that treating the active transforming fraction with the nuclease destroyed the structure of DNA and the ability of that fraction to transform pneumococcus cells. Treatments with enzymes that attack other cellular components had no effect on transformation activity.

The demonstration that DNA is the genetic material of a particular type of virus was provided by the elegant 'Waring blender' experiments of Hershey and Chase in 1952 (**Figure 1**). The virus T2 is a bacteriophage that infects *Escherichia coli* cells. It does so by attaching to the surface of the cell and injecting its core material into the bacterium. Many new viruses are produced inside the cell, causing it to burst (lysis), releasing new virus particles. Hershey and Chase used radioactive isotopes to label two different components of T2. ^{32}P, an isotope of phosphorus, was incorporated into the DNA, and the radioisotope of sulfur, ^{35}S, was used to label the protein. The double-labelled

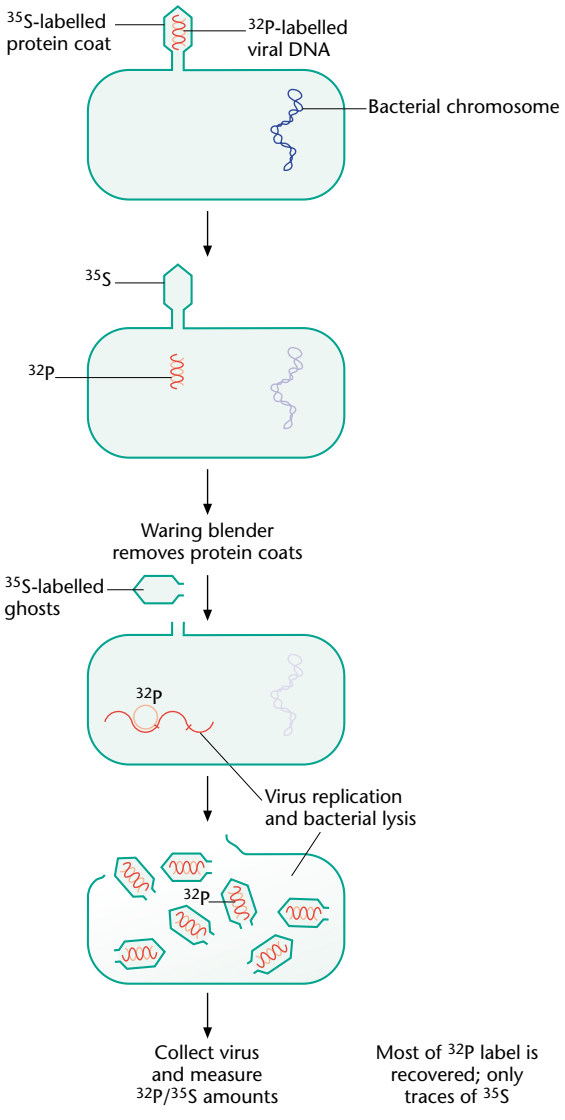

Figure 1 *Escherichia coli* cell infected with a T2 virus that is double-labelled with radioactive isotopes, ^{35}S in the protein coat and ^{32}P in the DNA. Results of infection and removal of protein coats shows that it is the DNA that carries the information necessary for viral replication.

viruses were then used to infect bacteria. After the viruses had attached and injected their core material into the cells, the mixture was placed in a Waring blender. The rapid mixing removed most of the T2 protein coats that remained on the outside of the cells. Inside the cells the infection continued to produce many new copies of T2 DNA that was packaged into new protein coats. After lysis, the new T2 viruses were separated from the cellular debris and analysed for their isotope composition. The results showed that almost all of the ^{32}P was recovered but almost none of the ^{35}S, demonstrating that it is the DNA that provides the

10.1038/npg.els.0000807, Downloaded from https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0000807, Wiley Online Library on [29/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

genetic information necessary to produce new virus particles.

RNA as the Genetic Material in Viruses

Many viruses have been shown to contain ribonucleic acid (RNA) as their genetic material. One of the early examples is the *Tobacco mosaic virus* (TMV), shown in the 1930s to be composed of protein and RNA. No DNA is found in the particle. The protein can be easily separated from the RNA by mild alkali treatment. That it is the RNA that acts as the genetic material in such particles was demonstrated directly by Gierer and Schramm, in 1956, who showed that tobacco plants could be infected by inoculation with the RNA alone.

Frankel-Conrat and his associates determined that infective virus particles could be reconstituted by mixing together the protein and RNA. Using this technique, he and Singer, in 1957, reconstituted viruses by mixing protein produced by a mutant strain together with the RNA from another strain, or vice versa. In either combination, the new virus particles produced by the infected host plant had protein of the type produced by the RNA parent.

Retroviruses are a family of RNA-containing viruses that replicate through a DNA intermediate. An important member of the retrovirus family is the *Human immunodeficiency virus* (HIV). The retrovirus replication process is accomplished through the use of reverse transcriptase, an enzyme discovered in 1970 by Temin and Mizutani and independently the same year by Baltimore. The enzyme catalyses the transcription of a complementary DNA strand from an RNA template. After a virus attaches to the surface of a cell, its RNA chromosome and reverse transcriptase protein are expelled into the cell cytoplasm. There, proviral DNA is transcribed using the viral RNA as a template. The DNA then enters the nucleus and integrates into the host chromosomal DNA. Synthesis of new viral RNA is made by transcription of the integrated DNA, and new viral particles are matured by encapsulating them with viral protein and host cell membrane as they are extruded from the host cell.

Retrotransposons found in eukaryotes are mobile elements that have structures very similar to retroviruses. They encode reverse transcriptase and move from one spot to another in the host genome by way of an RNA transcript that is used to make a DNA copy that can be inserted into another place in a chromosome. The retrotransposons do not form protein coats. They are transmitted vertically from one host to another through the germline. There are no proven cases of their horizontal transmission as infective particles.

Nonnucleic Acid Information Vectors

Infectious proteinaceous particles called prions have been characterized by Prusiner as the basis for some cases of central nervous system degenerative disorders, such as Creutzfeldt–Jacob disease (CJD) in humans. Bovine spongiform encephalopathy (BSE), known as mad cow disease, and scrapie in sheep are also diseases attributed to prions. CJD may occur as a sporadic, genetic or infectious illness, while kuru, a similar disease found in the Fore people in New Guinea, is known to be due to prion infection through ritual cannibalism of diseased brain tissue.

All forms of the degenerative disease involve aberrant metabolism and the resulting accumulation of prion protein. The basis of the disease was determined by Prusiner in the 1980s to be the conversion of a normal cellular protein designated PrP^{c} into an isoform protein called PrP^{Sc}. The normal PrP^{c} protein is encoded by the PrP gene and the genetic forms of CJD have been shown to be due to mutations in PrP.

Sporadic cases comprise the majority of CJD illness, however, and it is not known how the disease-causing prions originate in these cases. Three possible modes have been hypothesized: (1) horizontal transmission of prions from humans or other animals; (2) the spontaneous mutation of PrP in a somatic cell; (3) the spontaneous conversion of PrP^{c} to PrP^{Sc}. The important point for each possible mode is the introduction of PrP^{Sc} into cells where the normal PrP^{c} protein is produced. Experiments involving the inoculation of mice with different forms of prions show that it is likely that PrP^{Sc} acts as a template for the conversion of PrP^{c} to PrP^{Sc}. This model suggests that formation of the disease-inducing isoform involves a refolding of the normal protein into the less soluble, protease-resistant form. It is suggested by Prusiner that this conformational change may be assisted by a protein X that might act to chaperone PrP^{c} to the infectious PrP^{Sc} form.

It is important to distinguish between prions and nucleic acids as informational molecules. The transfer of information by prions is through the induction of conformation changes in other protein molecules. This is a very different type of replication and information transfer than that carried out by transcription and translation of nucleic acids. In the strictest sense, prions are not inherited; they are transferred only horizontally from one individual to another as an infectious agent. There is no known mechanism for the transfer of genetic information from proteins to nucleic acids.

10.1038/npg.els.0000807, Downloaded from https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0000807, Wiley Online Library on [29/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Organization of Nucleic Acids in Chromosomes: Overview

The chromatin of a cell is a highly organized aggregate of DNA, basic chromosomal proteins (histones) and nonhistone chromosomal proteins. In eukaryotes it is organized into some number (n) of individual chromosomes. Each chromosome in its simplest form, before replication, contains one long double-helical molecule of DNA plus the associated proteins.

Chromatin gets its name from its affinity for a variety of dyes, particularly in cells in the process of mitosis or meiosis, when all regions of the chromosomes are condensed. During interphase and early prophase, only some regions are highly condensed and stain readily. Those regions are called heterochromatin; the less condensed regions are known as euchromatin. Heterochromatin is usually subclassified into constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is highly condensed in all cell types, and generally positioned in a chromosome surrounding the centromere, at telomeres and nucleolar organizer regions. Facultative heterochromatin varies in its condensed state depending on cell type and developmental stage. It generally, though not always, represents euchromatin that is at times inactive in a particular cell type.

The chromosomal DNA of eukaryotes consists of unique and repetitive sequences of nucleotide pairs. This is measured by shearing purified chromosomal DNA into short segments, heating it to dissociate the strands and then lowering the temperature. The complementary strands begin to reassociate to produce double-stranded DNA again. Sequences that are highly repeated in the genome find pairing partners very rapidly compared to sequences that occur in low copy number. The kinetics of the reassociation can be used to measure genome size and complexity. For example, the total genome size of humans is 3.3×10^9 nucleotide pairs, of which 67% is nonrepetitive. Characterization of the various fractions shows that the repetitive sequences are distributed primarily at the centromeric and telomeric regions but they are dispersed throughout the unique sequence regions as well. The transcribed sequences of the genome are unique or in very low copy number.

In prokaryotes and many viruses the chromosome consists of a single DNA double helix in the form of a highly folded supercoiled circle, and associated proteins. The *E. coli* chromosome, for example, consists of 4.6×10^6 nucleotide pairs. As noted earlier, some classes of viruses have chromosomes consisting of RNA. Essentially all of the nucleic acid of bacteria and viruses is a unique sequence; however, a single bacterial cell may sometimes contain more than one copy of the circular chromosome, particularly in rapidly growing cultures.

Bacterial cells often also contain one or more minor DNA elements called plasmids, which generally may comprise as much as 2% of the DNA of the cell. Usually these small self-replicating circles separate from the bacterial chromosome, though some have the capacity to integrate into the main chromosome. An exceptional case is *Borrelia bergdorferi*, the causative agent of Lyme disease, which has a linear chromosome. It also has nine different circular plasmids and 12 different linear plasmids. Collectively, these plasmids contain approximately 40% of the genetic information of the cell.

There are several classes of plasmids: some carry antibiotic resistance genes; others encode colocins (proteins that can act to kill other bacterial cells) or immunity factors. The F-element is a plasmid that is important in bacterial conjugation and gene transfer.

Non-DNA Components of Chromosomes in Prokaryotes

The highly folded *E. coli* chromosomes can be prepared in their native state by lysing cells and centrifuging the lysate in a sucrose gradient. In such a form, the chromosomes sediment as particles that, in addition to the DNA, contain by weight about 30% RNA in the form of nascent RNA transcripts, and about 10% protein. The subunits of RNA polymerase account for more than 90% of the protein. The remaining proteins, part of which are contaminating membrane proteins, account for only about 1% of the total chromosomal particle weight. It is clear, however, that some of them are DNA-binding proteins, such as DNA polymerase. Also there are gyrases, forms of topoisomerase, which can catalyse transient breaks in the DNA to induce a supercoiled structure or to relax a supercoiled strand.

The ratio of DNA to protein in prokaryotes is very different than the approximate 1:1 ratio of DNA to protein in eukaryotic chromosomes. This suggests that, unlike the eukaryotes, bacteria do not depend on proteins other than the gyrases to maintain the compact conformation of the chromosome. There is, however, a chromosomal attachment site on the bacterial cell membrane that is important in maintaining conformation and facilitating chromosome replication.

Non-DNA Components of Chromosomes in Eukaryotes

In addition to DNA, eukaryotic chromatin contains a variety of proteins that are important to chromosome structure, its replication and transcription. The proteins are of two major classes, histones and the nonhistones.

10.1038/npg.els.0000807, Downloaded from https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0000807, Wiley Online Library on [29/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Histones are rich in basic amino acids lysine, arginine and histidine and form five separate classes, H1, H2a, H2b, H3 and H4. All classes are highly conserved, H4 probably the most highly and H1 the least, but in all cases they are much the same throughout the eukaryotic superkingdom.

The basic structural unit in the complex organization of chromatin is the nucleosome. It contains an octamer of histone molecules ($(H2a, H2b)_2$ $(H3_2, H4_2)$) plus one molecule of H1, around which is wrapped about 1.7 turns of DNA double helix (146 base pairs). This core unit, about 11 nm in diameter, is connected by DNA (about 27 base pairs) to the next adjacent unit. Strong surface interactions between the histones of the octamer are believed to be important in maintaining the integrity of the nucleosome core unit. The nucleosome structure, DNA plus histones, thus forms a single beaded fibre stretching the length of the chromosome. This fibre takes on a higher order of structure by winding into a fibre about 25–30-nm thick that is believed to be the basic form seen in the chromosomes of eukaryotes.

The nonhistone proteins of chromatin are very important in the replication, transcription and recombination of chromosomes. Prominent among those in this category are proteins that can bind to DNA and are important in effecting conformational changes necessary for action by polymerases and transcription factors. Two classes of topoisomerases are known that are involved in controlling the coiled or relaxed state of the DNA. Type I catalyses transient breaks in a single DNA strand, while type II catalyses transient double-strand breaks. The gyrases of bacteria belong to the type II class.

Many of the nonhistone chromosomal proteins vary in both type and amount in different cell types and at different stages of development. They are proteins, transiently associated with chromosomes, that are produced in response to cellular activities such as DNA repair, specific types of transcription or recombination, etc. Because of the complex array of cellular functions, there are numbers of nonhistone proteins in chromatin that are not well-characterized or are as yet unknown.

Special Chemistry of RNA and DNA in Chromosomes: Overview

The special aspects of the chemistry of nucleic acids dictate their structure, how they are replicated and how they, as the bearers of genetic code, transfer that information for use in cellular metabolism and development.

DNA and RNA are composed of two classes of nitrogen-containing bases: purines and pyrimidines. In DNA the most commonly occurring purines are adenine (A) and guanine (G); the pyrimidines are thymine (T) and cytosine (C). In RNA, thymine is replaced by uracil (U). A purine or pyrimidine base bonded to a pentose sugar (deoxyribose in DNA, ribose in RNA) forms a nucleoside. A nucleoside complexed with a phosphate group forms a nucleotide. Construction of a polynucleotide chain is accomplished by linking the phosphate group to the sugar of another nucleotide to form the 'backbone' of the chain (**Figure 2**). The covalent bonds of the phosphate linkages to the carbon in the sugar are formed in a manner that imparts a polarity to the chain. A double helix consists of two antiparallel polynucleotide chains wound around each other in such a way that the bases of the two chains pair and form hydrogen bonds between members of each pair. The bonds can form only between a particular purine in one chain with a pyrimidine in the other. A can pair only with T (or U in RNA), and G only with C. Our understanding this structure came from the analysis of the physical chemistry of nucleic acids worked out by a number of investigators and modelled by Watson and Crick in 1953.

That model opened the way for revolutionary advances in understanding the nature of genes and how they function. The genetic information is encoded in the sequence of the purine and pyrimidine bases along the chains. It is important to note that this information is carried in both chains. When a chromosome replicates, the two chains of the double helix unwind and separate, breaking the weak hydrogen bonds between the base pairs. Each single strand then serves as a template for the construction of a complementary antiparallel chain. The

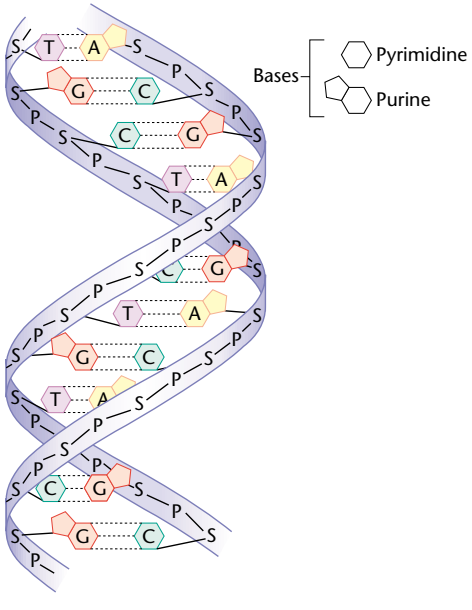

Figure 2 A DNA molecule showing two strands of opposite orientation arranged as a double helix. Pairing by virtue of hydrogen bonds (———) between purine bases on one strand with pyrimidine bases on the other, and their covalent attachments to the phosphate (P) and deoxyribose sugar (S) backbone, shows the three-dimensional relationships of these components.

10.1038/npg.els.0000807, Downloaded from https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0000807, Wiley Online Library on [29/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

end result is two double-helical molecules identical to the original one.

Considering that the average human chromosome consists of a single molecule of DNA about 3 cm in length, unwinding and copying such a molecule is a rather daunting prospect. The cell solves this problem by replicating small sections of the chromosome more or less simultaneously. Beginning at many points along the chromosome, replication bubbles form by the local unwinding of the DNA duplex. The unwinding is facilitated by the action of topoisomerases. Replication is then carried forth on both single chains through the action of DNA polymerases. The new chains are synthesized in a polar fashion, thus proceeding in opposite directions on the two antiparallel chains, resulting in bidirectional synthesis within the bubble. The chains grow until they meet and connect with the newly synthesized chains of the replication bubbles on either side. The whole process is much more complex than stated in this overview, involving a variety of steps and enzyme species.

Replication in most bacteria and viruses is initiated at a single origin of replication and proceeds around the circular chromosome by relaxing, unwinding and copying the duplex, either DNA or RNA, depending on organism. Similar uses of topoisomerases and polymerases as in eukaryotes facilitate the synthesis of two copies of the chromosome.

Utilization of the genetic information encoded in the nucleotide sequences of genes involves a series of steps and uses a variety of enzymes and auxiliary proteins. Though both strands of the duplex carry information, only one, the sense strand, is used to direct cellular functions. This is done first by unwinding the duplex and making an RNA copy, a transcript, of a gene. This involves the localized action of topoisomerases and RNA polymerase, along with a variety of transcription factors. The original transcript of a gene encoding a polypeptide is then processed into messenger RNA, which is transported to the cytoplasm for translation on ribosomes. Translation results in the synthesis of a polypeptide chain, the amino acid sequence of which has been specified by the nucleotide sequence of the particular gene. Other RNAs include ribosomal RNA, major structural components of ribosomes, transfer RNAs that transport amino acids to the ribosomes during translation; also mitochondrial and plastid RNAs encoded by genes in those cytoplasmic bodies.

Summary

It is evident that the proof that nucleic acids are the physical basis of inheritance grew from early cytological observations, followed by chemical analysis of the nuclear material. Actual proof was not established, however, until chemical analysis was coupled with genetic experimentation. The turning point was provided by the experiments of Avery and his colleagues and by Hershey and Chase. It was the model of the structure of the DNA double helix by Watson and Crick that opened the investigations about the genetic code residing in the nucleotide sequences. Our understanding of how that information is replicated and how it is expressed in cells has advanced dramatically since that time, resulting in greatly improved medical practices and more productive plant and animal varieties.

Further Reading

Avery OT, MacLeod CM and McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine* **79**: 137–158.

Hershey AD and Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* **36**: 39–56.

Prusiner SB (1997) Prion diseases and the BSE crisis. *Science* **278**: 245–251.

Wagner RP, Maguire MP and Stallings RL (1993) *Chromosomes: A Synthesis*. New York: Wiley-Liss.

Watson JD and Crick FHC (1953) A structure for deoxyribonucleic acid. *Nature* **171**: 737–738.

10.1038/npg.els.0000807, Downloaded from https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0000807, Wiley Online Library on [29/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License