A NEW APPROACH TO THE STUDY OF NUCLEOTIDE SEQUENCES IN DNA

by

GIORGIO BERNARDI

With 34 Figures

AKADEMIE DER WISSENSCHAFTEN UND DER LITERATUR · MAINZ
IN KOMMISSION BEI FRANZ STEINER VERLAG GMBH · WIESBADEN

A new approach to the study of nucleotide sequences in DNA

G. BERNARDI

Contents

1. A brief introduction on the general issue, the eukaryotic genome

A comparison of prokaryotic and eukaryotic cells reveals that the major existing differences concern the *genome size* and the *genome organization*. These two points will be briefly presented here together with models which have been put forward in order to explain some of the problems raised.

a) Genome size

The amount of DNA per eukaryotic (haploid) cell is constant in each species and covers an extremely wide range of values (1,2). Some fungi have DNA contents almost equal to those of bacteria, whereas some animals and plants have DNA contents per cell 10,000 times as large (Table I). Such a wide range of values raises a number of problems. In view of the very pirimtive state of our understanding in this extremely complex area and of the introductory nature of this section it seems justified to discuss here only two of them.

The first problem is that wide variations of genome sizes are found within single phyla, within single genera and even within interfetile species. To quote some examples, unicellular algae span a 2000-fold range in genome size (4); among animals, amphibians span a 100-fold range (fig. 1); an 80-fold variation is seen among various species of Ranunculaceae possessing a (nearly) constant karyotype (13); of the two interfertile species *Chironomus thummi thummi* and *Chironomus thummi piger*, the first one has 27% more DNA than the second one (14). Whatever the reason for such differences, it is clear that they do not correspond to equivalent differences in the amount of genetic information. Letting aside, for this reason, this first paradox, we can proceed further, but only to be confronted with a second paradox.

If we consider the *minimum genome size* for some major steps in evolution we can notice a systematic upward trend (fig. 2), corresponding to the increasing grade of organization. The 1000-fold increase in genome size shown by mammals compared to bacteria is, however, not accom-

Table I

DNA content per haploid cell *

	Picograms (10^{-12} g)
Bacteria	
Haemophilus influenzae	0.002–0.0027
Escherichia coli	0.0047
Ascomycetes	
Neurospora crassa	0.042
Saccharomyces cerevisiae	0.0245
Protozoa	
Euglena gracilis, Z	3.33
Tetrahymena pyriformis, GL	15.7
Porifera, Coelenterata, Echinodermata	
Tube sponge	0.06
Jellyfish, *Cassiopeia*	0.33
Sea urchin, *Arbacia*	0.67
Annelida, Mollusca	
Nereid worm	1.45
Snail, Tectarius muraticus	0.67
Arthropoda	
Cliff crab, Plagusia depressa	1.49
Drosophila melanogaster	0.18
Fish	
Dipnoan, African lungfish	50
Shark, Charcarias obscurus	2.73
Carp	1.64
Trout	2.45
Amphibians, Reptiles	
Rana pipiens	6.48
Xenopus laevis	4.2
Bufo viridis	5.4
Triturus viridescens	4.9
Amphiuma	84
Green turtle	2.6
Water snake	2.5
Birds	
Chicken	0.73
Duck	1.25
Mammals	
Rat	3.9
Rabbit	3.25
Ox	3.2
Man	3.12
Plants	
Vicia faba	19.2
Vicia sativa	3.6
Trade scantia paludosa	29.7

* Data mostly from ref. 3.

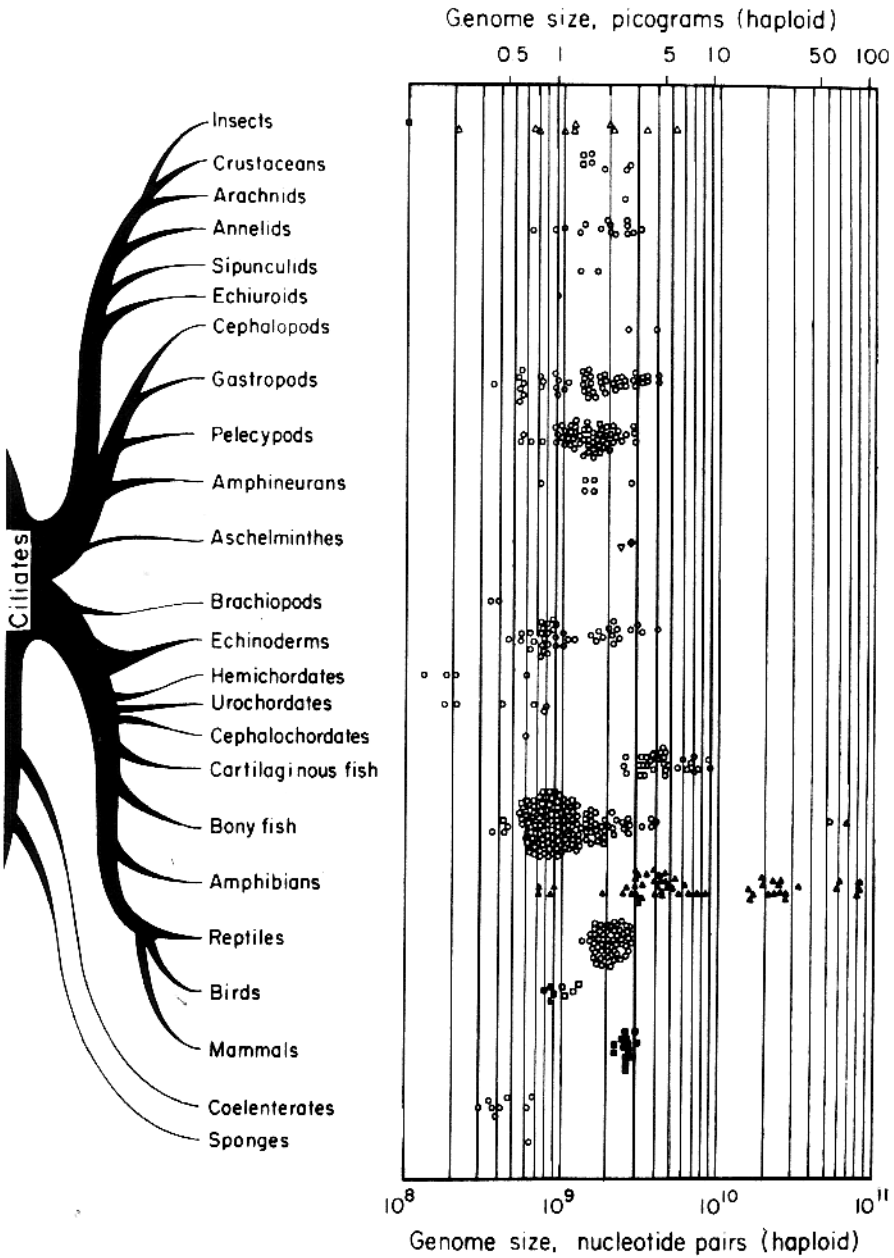

Fig. 1. Distribution of genome size in animals (5). The data have been assembled from many sources: ○ ref. 6, □ ref. 2, ◆ ref. 7, ▽ ref. 8, ▲ ref. 9, ■ ref. 3, △ ref. 10 and ref. 11, ● ref. 12.

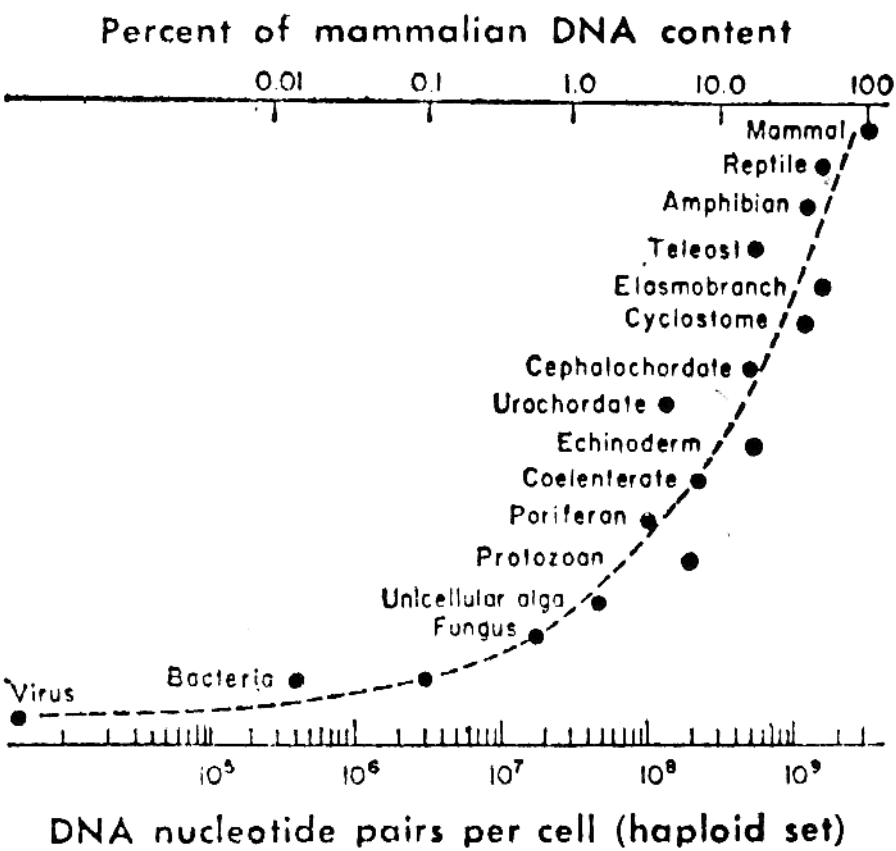

Fig. 2. The minimum amount of DNA has been observed for species at various grades of organization (15). The ordinate is not a numerical scale and the exact shape of the curve has little significance. Data refer (from top to bottom) to the following species: *Bos bos* (16); *Chelonia mydas* (17); *Scaphiopus couchi* (9); *Tetraodon fluviatilis* (18); *Carcharias obscurus* (2); *Lampetra planeri*, *Amphioxus lanceolatus* (19); *Acidea atra*, *Paracentrotus lividus* (20); *Cassiopea*, *Dysidea crawshagi* (2); *Amoeba hystolytica* (21); *Saccharomyces* (22); *E. coli* (23); *Mycoplasma* (24); Simian virus (25).

panied, by far, by a corresponding increase in the number of different proteins encoded by the mammalian genome. For example, 93% of the enzyme activities listed in the International Enzyme Commission tables are known to be present in both prokaryotes and mammals (5). These facts raise the problem of the role of the "excess DNA" found in eukaryotes.

This problem can be explained in two different ways, namely by postulating that in eukaryotes structural genes are reiterated, or by postulating that the "excess DNA" is essentially formed by regulatory genes. Accordingly, two types of general models have been proposed: 1) Models based on the simple reiteration of a gene or a group of genes; 2) "Super-operon" models where regulatory genes are predominant over structural genes.

CALLAN's model (26–28; fig. 3) suggests that eukaryotic genes are "serially repeated" 100 to 10,000 times along the DNA duplex. The nucleotide sequence of each of the repeated or "slave genes" is brought into accord with that of a "master-gene" at least once per generation, possibly during the meiotic prophase, by an unspecified rectification mechanism. Because of rectification, only mutational alterations in the master gene are ever detected.

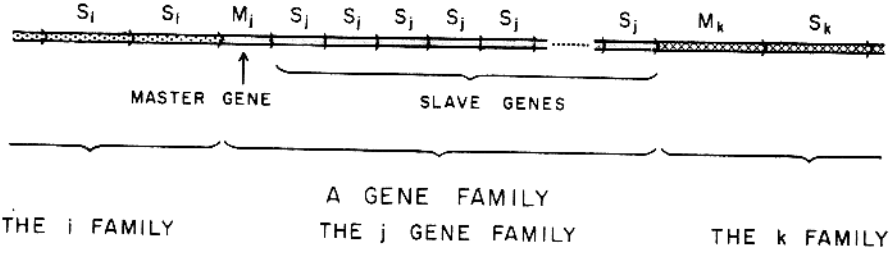

Fig. 3. CALLAN's model of the eukaryotic chromosome: masters and slaves. The chromatid is pictured to be a single DNA molecule containing gene families, each made up of a defined (large) number of identical genes. One gene, called the master, is depicted as the first of a series of identical slaves. The genes of one family can differ in sequence from those in another, although they may be related sequences (28).

BRITTEN and DAVIDSON (15; fig. 4) assume that in eukaryotic genomes a relatively small number of individual "producer genes" (cistrons carrying structural information) are under the control of many "receptor genes", structurally linked to them. Transcription of a producer gene can only occur if at least one of its receptor genes is activated by forming a sequence-specific complex with "activator RNA". The latter RNA would be synthetized by specific "integrator genes" which, in turn, are thought to be under the control of "sensor genes", the latter being sensitive to signals of various types.

GEORGIEV (29; fig. 5) assumes that the operon of higher organisms is divided into a promoter-proximal acceptor (or non-informative) zone

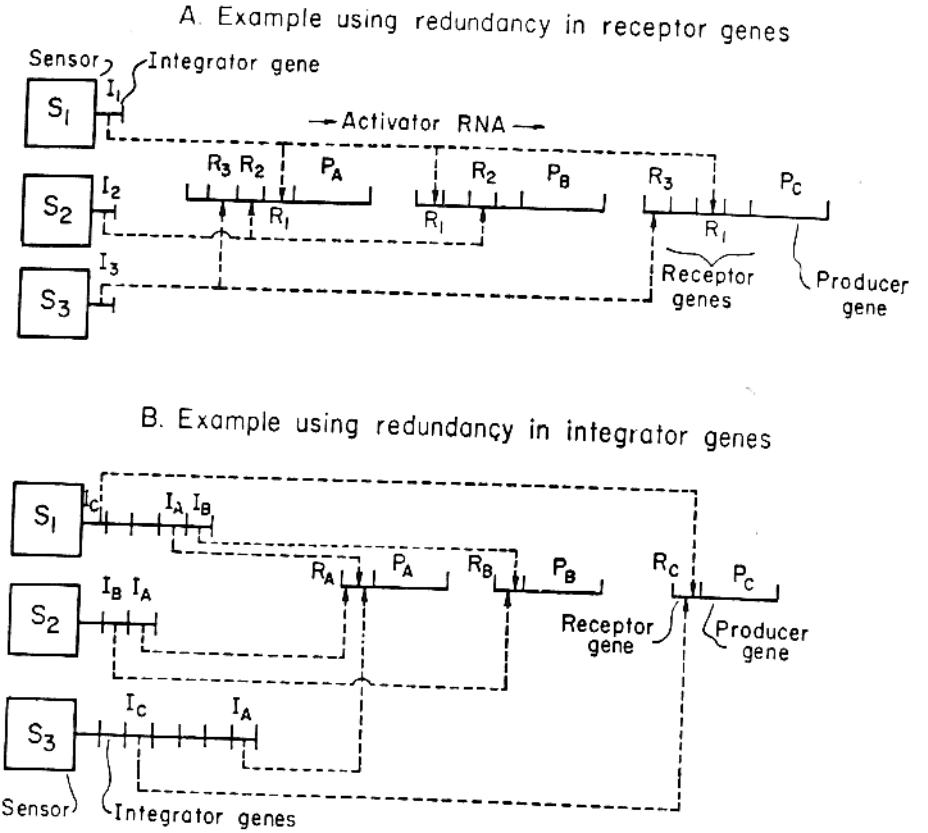

Fig. 4. Types of integrative system within the BRITTEN-DAVIDSON model. (A) Integrative system depending on redundancy among the regulator genes (B) Integrative system depending on redundancy among the integrator genes. These diagrams schematize the events that occur after the three sensor genes have initiated transcription of their integrator genes. Activator RNAs diffuse (symbolized by dotted line) from their sites of synthesis – the integrator genes – to receptor genes. The formation of a complex between them leads to active transcription of the producer genes P_A, P_B, and P_C (15).

and a promoter-distal structural (or informative) zone. Numerous "acceptor loci" in the non-informative zone are believed to interact specifically with regulatory proteins in the manner of operators, or promoters. The informative zone not only contains cistrons for structural or enzyme proteins but also for regulatory proteins. The entire unit is

believed to be transcribed in one piece with the acceptor portion being degraded before the messenger leaves the nucleus.

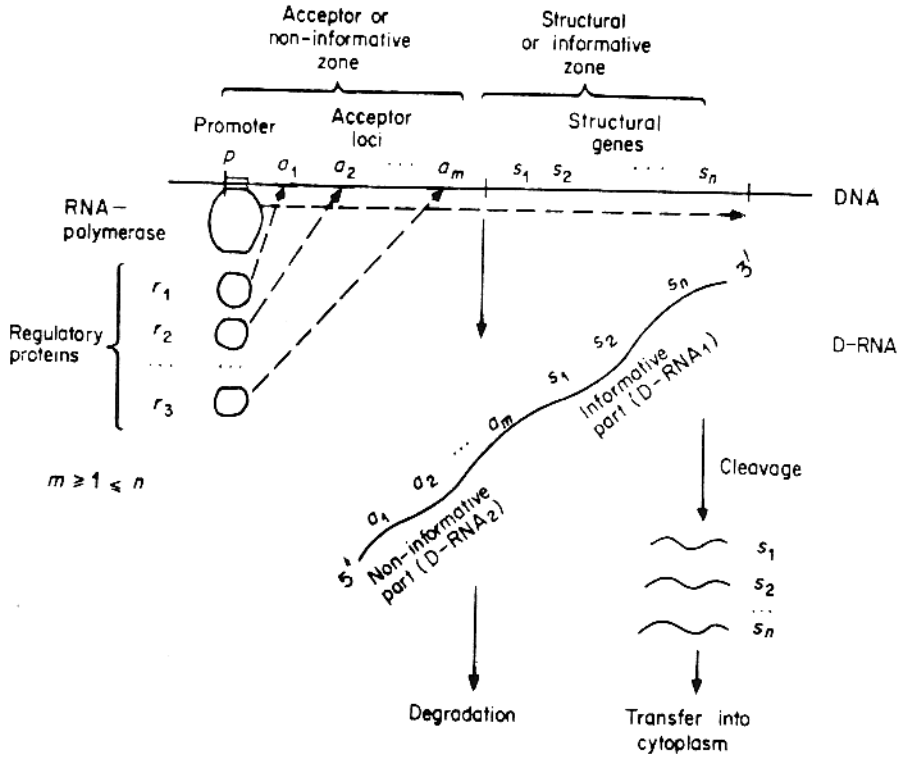

Fig. 5. Scheme of the eukaryote operon according to GEORGIEV (29).

b) Genome organization

The genetic material of eukaryotes is organized in a much more complex way than that of prokaryotes. DNA is combined with basic and acidic proteins to form chromatin fibers which are folded each one into a chromatid. Before the chromatid is replicated, there is only one chromatid (unit chromatid) per chromosome (at least according to the more generally accepted mononeme theory); at mitotic metaphase the chromosome is made up of two (sister) chromatids. Polytene chromosomes, which are formed by hundreds of chromatids having undergone internal replication, display the chromomeric structure of chromatids in the form of

characteristic bands (fig. 6). Structurally, chromomeres correspond to local packing of chromatin; functionally, they are units of mutation, replication and transcription, and correspond to Mendelian genes (30). Cromomeres are the seat of the paradoxical high amount of DNA of eukaryotic genomes since they contain much more DNA than required by a gene coding for an average protein. In the case of the two *Chironomus*

Fig. 6. A comparison of the DNA contents in homologous bands of giant chromosomes of *Chironomus thummi thummi* and *C. thummi piger* (14). The *thummi* bands (right-hand side) have 2, 4, 8 or 16 times as much DNA as the homologous *piger* bands (left-hand side).

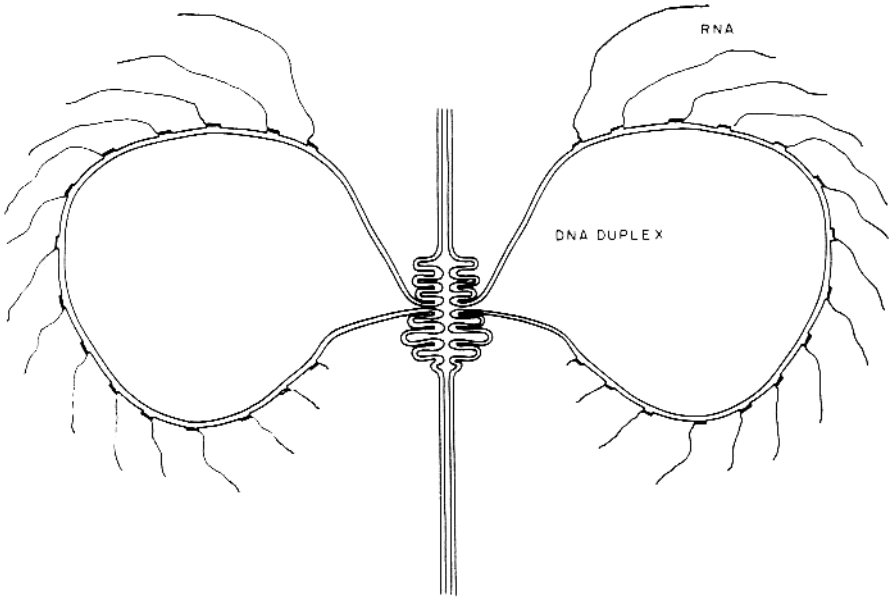

Fig. 7. Schematic diagram of how the DNA in the loops of a lampbrush chromosome is thought to be "expressed"-transcribed. It is not yet clear whether the RNA is translated into protein as a general rule, although some protein synthesis is thought to occur in the oocyte nucleus (28).

species mentioned above, the extra DNA of *thummi* is localized in 30 (or more) of the 5000 or more bands present in the chromosome sets. Very interestingly, the *thummi* bands have 2, 4, 8 or 16 times as much DNA as the homologous piger bands (14; fig. 6).

It is of interest to see how the two types of models proposed to account for the "excess DNA" can be fitted into the chromomeric structure of chromosomes.

In the "master and slaves" model of CALLAN, which was originally developed in connection with the lampbrush chromosomes of amphibian oocytes, the genes are visualized as serially repeated along the chromomere loops (fig. 7), which undergo an unwinding and rewinding process.

As far as "super-operon" models are concerned, the first suggestion came from CRICK (31), who placed the structural cistrons exclusively in the interbands and left the DNA of the bands entirely to the task of regulation (fig. 8), implying the requirement for very long regulatory

segments. Subsequently, another model was proposed by PAUL (32; figs. 9, 10) who assumed that chromomeres originate through an initial reiteration of conventional genetic units comprising an "address locus", a promoter, a regulator, an initiator, protein coding and terminator sequences; all but the first of the reiterated units are thought to gradually degenerate during further evolution, both with respect to base sequence and to their original functions, with the exception of the "address loci" which continue to be needed as recognition sites for a postulated unwind-

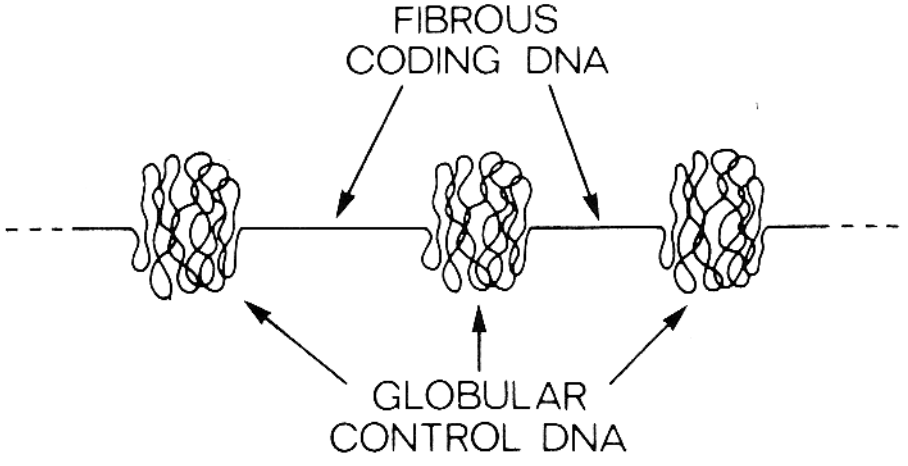

Fig. 8. An extremely schematic drawing of the general structure of the DNA of the chromatid proposed by CRICK. The line represents part of the continuous DNA molecule in the mononemic chromatid. The straight portions correspond to the interband regions of the giant polytene chromosomes of the *Diptera*, which are postulated to be similar in their general character to the corresponding interphase chromatids, which are the active form. The mitotic chromosome is relatively intert. The DNA sequences coding for protein are postulated to be mainly, if not entirely, in these extended regions. For convenience this DNA is referred to as fibrous DNA. The intricately folded regions correspond to the bands seen in the polytene chromosome. Na attempt has been made to represent their detailed structure. They are postulated to be the sites of the control regions. The model implies that a genetic complementation group is usually contained in either an interband plus a band or an interband plus part of the bands on either side. When a gene is active the bands are probably at least partly unfolded. The globular DNA is certainly complexed with chromosomal proteins, the fibrous DNA probably so. Thus both should be more strictly referred to as nucleo-protein (31).

ing mechanism, plus the terminator of the last unit: a fully evolved chromomere would thus have only one "sensible" gene at its beginning, this initial segment being followed by an appendix of nonsense DNA sequences which, although being transcribed, possess only auxiliary functions.

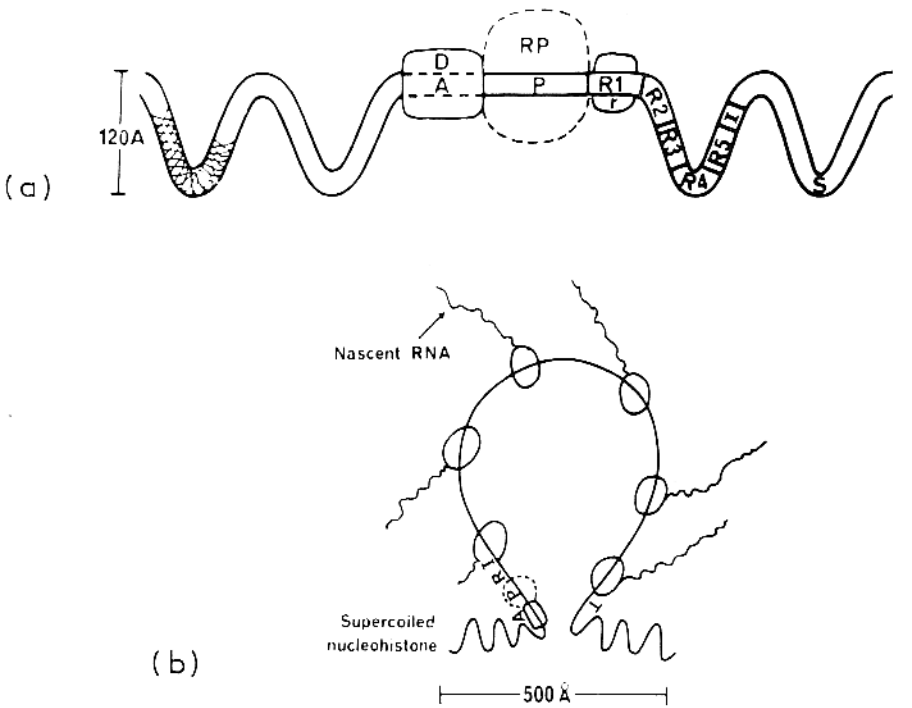

Fig. 9a) Postulated structure of a transcribable region of DNA in chromatin according to Paul. Nucleohistone is assumed to form a supercoiled structure with a diameter of approximately 120 Å. A transcribable region of DNA is postulated to contain an address site (A) closely linked to a promoter site (P), regulator sites (R1, R2, etc.) and an initiator site (I). A destabilizing molecule (D) binds to A to cause local relaxation of supercoiling, which permits a molecule of RNA polymerase (RP) to attach to the promoter. If no repressors (r) are bound to regulator sites transcription is initiated and the structural gene (S) is transcribed.
b) Structure of a region during transcription. RNA polymerase and nascent RNA induce further unwinding of nucleohistone and permit transcription of the entire structural gene until a terminator (T) is reached (32).

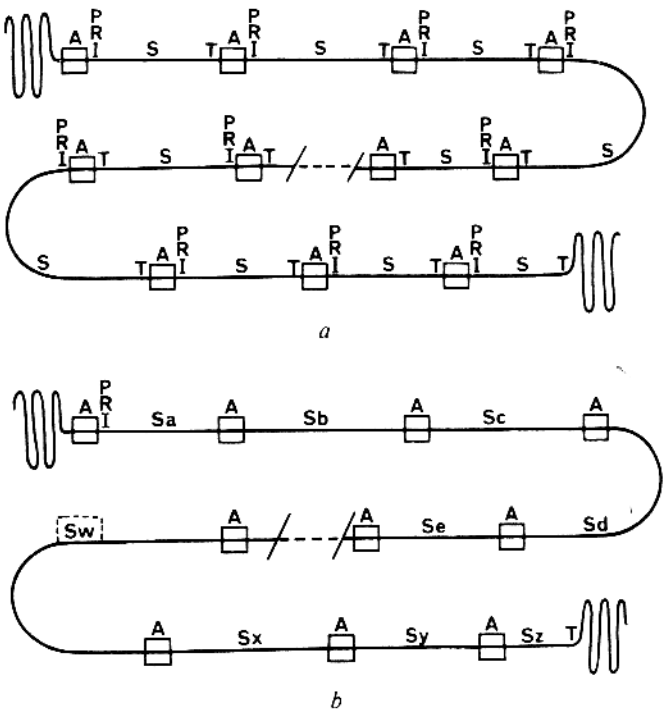

Fig. 10. a) The configuration of a region arising by simple gene reduplication. Address loci (A) with associated destabilizers alternate with cistrons each containing promoter (P), regulator (R) and initiator (I) loci, structural information (S) and a terminator locus (T). b) The configuration of a fully evolved region. Promoter, regulator and initiator loci occur only at the beginning and a terminator at the end of the region. Address loci may be distributed throughout the region as shown, or clustered at the beginning. The locus contains one "sensible" gene (Sw) and many other sequences (Sa, Sb, etc.) derived from the ancestral gene (32).

c) Experimental approaches to the study of the eukaryotic genome

Four major pathways have been followed so far in order to study the eukaryotic genome, involving, respectively, investigations on 1) the structure of chromosomes 2) the structure of chromatin 3) the transcription products 4) the nucleotide sequences in DNA. We will examine here only the methods developed for studying the nucleotide sequences in DNA. We would like to mention, however, a few important conclusions arrived at from investigations on giant chromosomes of *Diptera*.

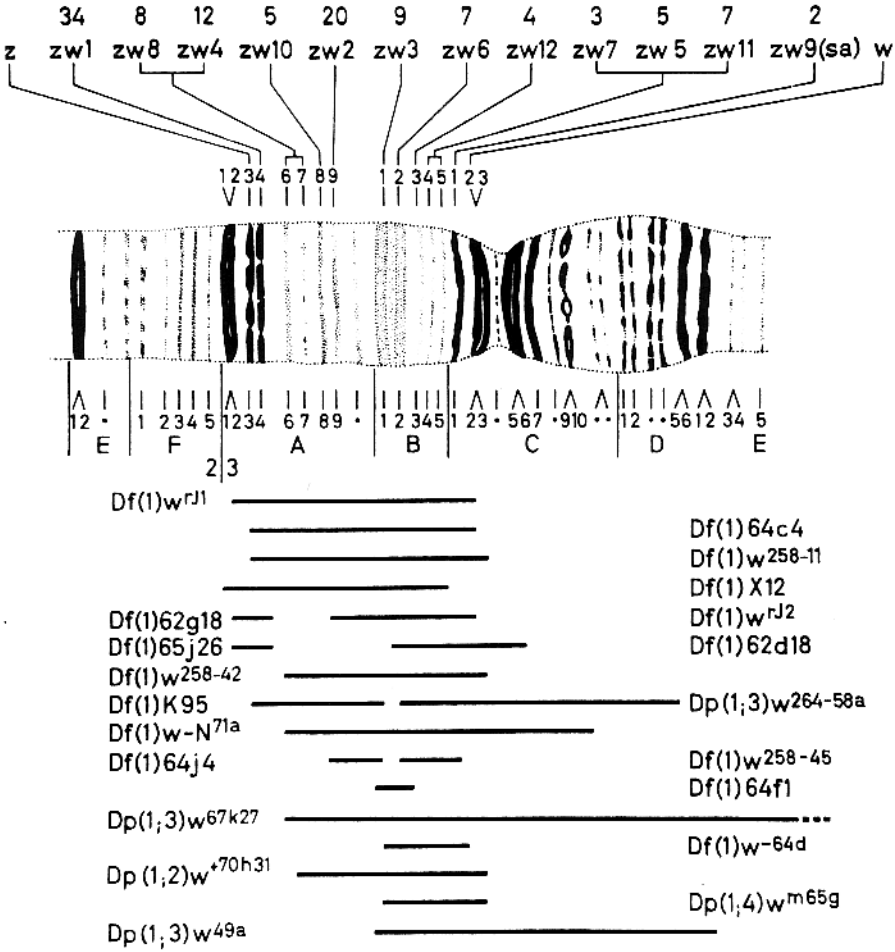

Fig. 11. Co-linearity of bands and complementation units in the *zestewhite* region of the X chromosome of *Drosophila melanogaster*, as demonstrated by JUDD *et al.* (33). Cytological map of BRIDGES slightly modified according to electron microscopic results. Z, *zeste*; W, *white*; sa, sparse arista; ZW 1–12: lethal or semilethal loci between Z and W, as defined by complementation. Number of allelic mutants found at each locus are given on top. Df, Dp: deficiencies and duplications used in the ordering and localization of complementation units (43; adapted from ref. 33.)

Each band (and/or interband), is really associated with only one essential genetic function (33, fig. 11), in spite of the fact that the DNA content per chromomere in *Drosophila melanogaster* (an animal with a very small genome size, Table I) is, on average, ten times that of a cistron coding for a typical protein.

Bands consist, largely at least, of genetic material with functions difficult to define and whose loss is often tolerated without significant consequences to the development of the individual. On the other hand, the essential portions of the genetic material – probably those coding for specific polypeptides – seem to occupy relatively short segments within each cytogenetic unit (band plus interband). These could be identical either with the interband only, or, with small "initial" of "terminal" portions of the band. These portions in themselves may, or may not, constitute independent structural subunits (34).

A very high molecular weight RNA molecule has been shown to be transcribed from a defined chromosome region, the BALBIANI ring 2, in *Chironomus tentans* salivary glands. The giant size of this transcript ($15–35 \cdot 10^6$) indicates that at least a considerable proportion of the chromomere DNA (average molecular weight $60 \cdot 10^6$) must serve as template. In fact, it is not even excluded that the whole chromomere, possibly also adjacent interchromomere regions, might be transcribed as one single transcript (35).

2. Methods for studying nucleotide sequences in DNA

Table II summarizes the methods which have been most widely used for studying nucleotide sequences in DNAs. They can be divided into two groups, according to whether the methods are based on investigations of properties related to the nucleotide sequences or on the direct study of the sequences.

a) Indirect methods

The indirect methods can be further subdivided into two classes according to whether they rely on the sequence-dependent properties of single-stranded or double-stranded DNA.

i. Sequence-dependent properties of single-stranded DNA

Kinetics of DNA renaturation: The reannealing of separated complementary single strands of DNA ideally follows a second order kinetics (fig. 12), which has been studied by several authors (36, 37). For a given total DNA concentration, one would expect that the half-period for reassociation would be proportional to the number of different types of

Table II

Methods for studying nucleotide sequences in DNAs

A. Indirect methods
   1. Sequence-dependent properties of single-stranded DNA
      a. Renaturation kinetics
      b. Cyclization after exonuclease degradation
   2. Sequence-dependent properties of double-standed DNA
      a. Buoyant density in CsCl
      b. Melting temperature
      c. Circular dichroism
      d. Silver binding
      e. Hydroxyapatite binding

B. Direct methods
   1. Sequence methods
      a. DNA polymerase repair
      b. DNA polymerase ribonubstitution
      c. RNA polymerase copy
      d. Direct sequencing
   2. Frequency methods
      a. Depurination
      b. Nearest-neighbour analysis
      c. Analysis of termini released by DNases

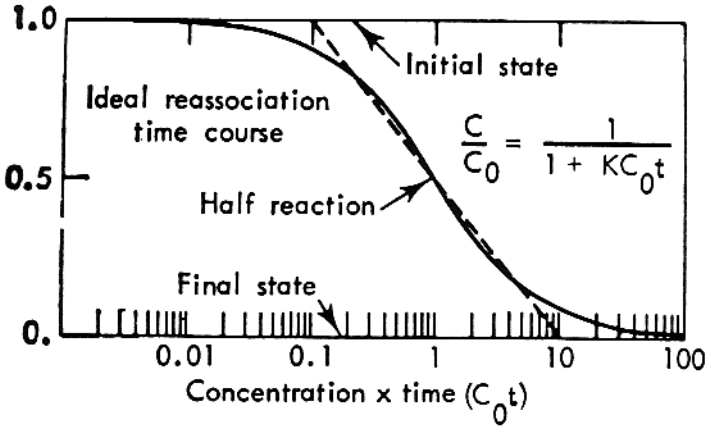

Fig. 12. Time course of an ideal, second-order reaction to illustrate the features of the log C_0t plot. The equation represents the fraction of DNA which remains single-stranded at any time after the initiation of the reaction. For this example, K is taken to be 1.0, and the fraction remaining single-stranded is plotted against the product of total concentration and time on a logarithmic scale (36).

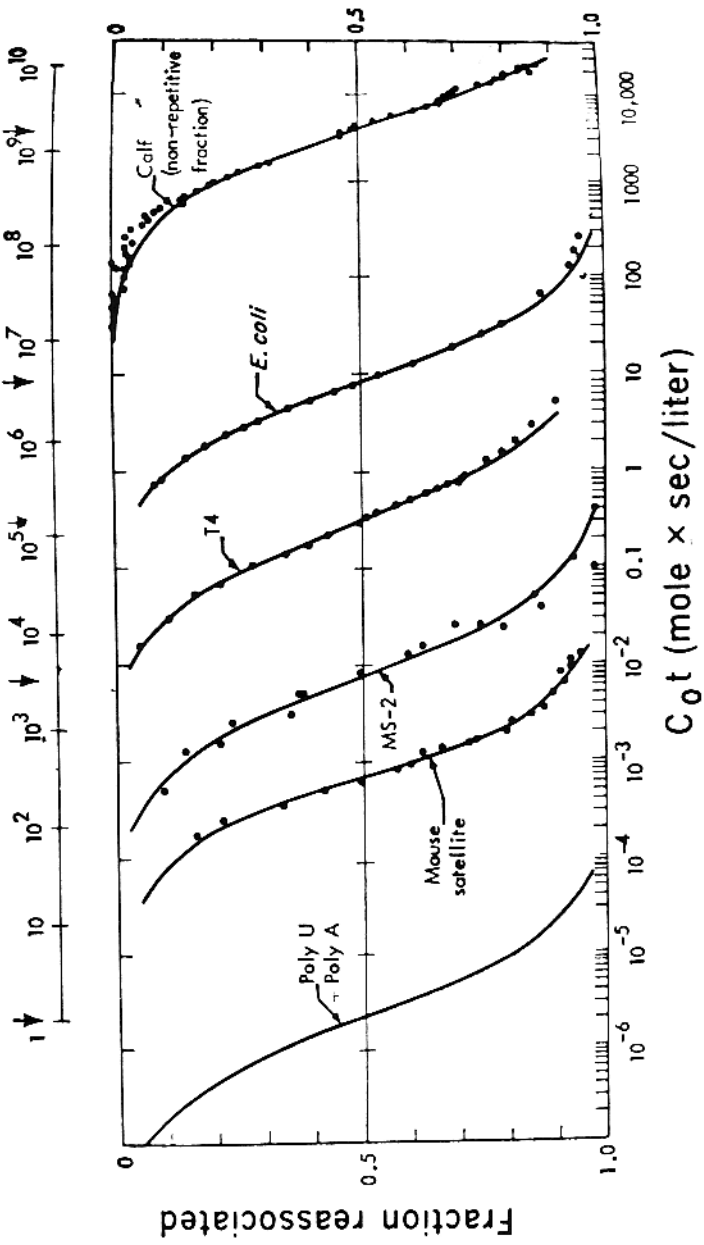

Fig. 13. Reassociation of double-stranded nucleic acids from various sources. The genome size is indicated by the arrows near the upper nomographic scale. Over a factor of 10^9, this value is proportional to the C_0t required for half reaction. For experimental details see the original paper (36).

fragments present and thus to the genome size. This expectation is exactly borne out in the case of viral and bacterial DNA (fig. 13). Eukaryotic DNAs show complex renaturation kinetics (fig. 14) and can usually be resolved into fast, intermediate and slow renaturating fractions. As

shown in fig. 13, the satellite DNA from mouse is such a fast-renaturing fraction, whereas the calf non-repetitive fraction is a slow renaturing fraction.

The major contribution of renaturation kinetics studies to our understanding of the organization of the eukaryotic genome has been the demonstration that a large fraction of the genome (about 50% in the case of calf) is made up of "repetitive" sequences, the rest being accounted for by non-repetitive or "unique" sequences. Recently, the arrangement of repetitive and non-repetitive sequences has been investigated in *Xenopus* (39) by reassociating to a Cot of 50 labeled DNA sheared to various fragment lengths with excess 450-nucleotide fragments of unlabeled DNA and by following the binding of labeled DNA to hydroxyapatide. Repetitive sequences monitored in this way are present on about 45% of the 450-nucleotide fragments. As DNA fragment length is increased, larger fractions of the DNA are found to contain repetitive elements. Up to 80% of the DNA binds at an average fragment length of 3700 nucleotides. Analysis of the data shows that a little more than 50% of the genome consists of closely interspersed repetitive and non-repetitive sequences. The average length of the repetitive sequence elements is 300 ± 100 nucleotides, while the non-repetitive sequences separating adjacent repetitive sequence elements average 800 ± 200 nucleotides.

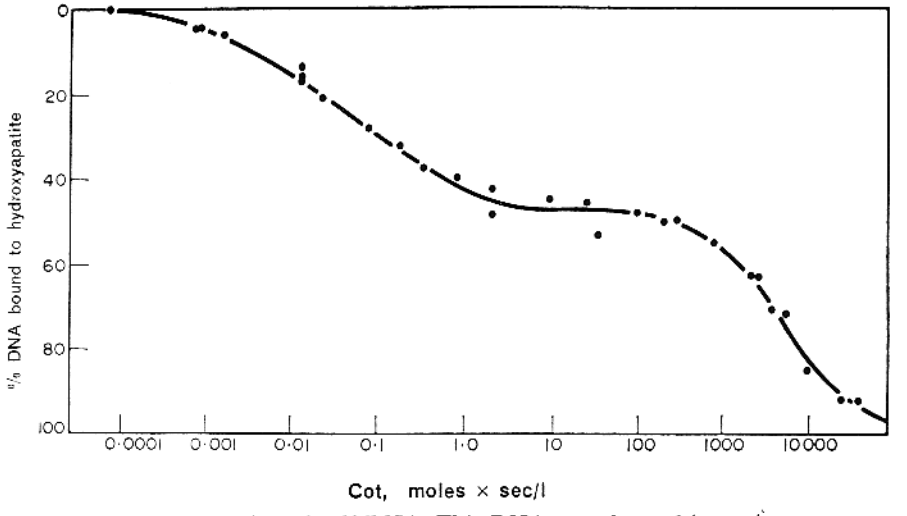

Fig. 14. The reassociation of calf DNA. This DNA was sheared into pieces approximately 400 nucleotides long, denatured and annealed at 60° C in 0.12 M phosphate buffer, and the proportion of reassociated DNA was assayed by hydroxyapatite fractionation (38).

The remainder of DNA is mainly non-repetitive, though most of it contains rare interspersed repetitive elements spaced at a minimum of 4000 nucleotides apart. A high degree of order in the arrangement of DNA sequences in the *Xenopus* genome is suggested by the alternating interspersion of relatively brief repetitive sequence and longer non-

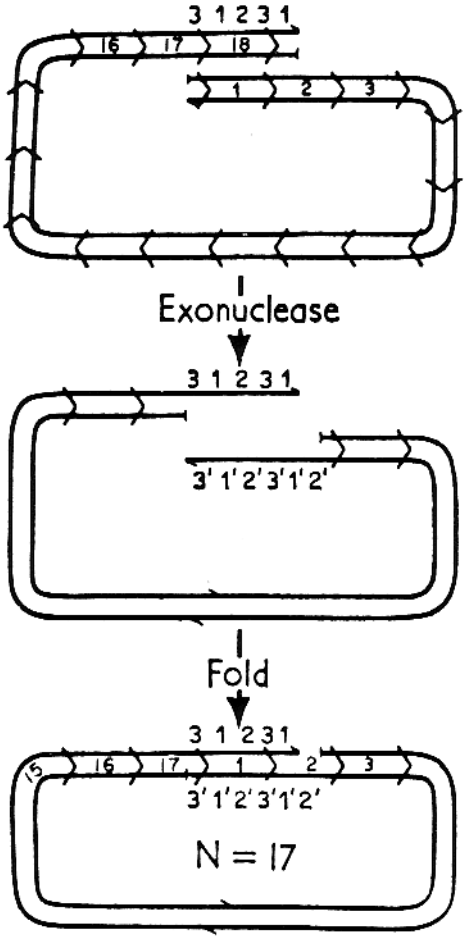

Fig. 15. Formation of "folded" circular structures from tandemly-repeating sequences (40). The diagrams illustrate how randomly-broken, tandemly-repetitious DNA fragments would be expected to form circular structures if their terminals had been partly degraded by exonuclease III. The double helix is depicted by two parallel lines, and repetitious sequences by the broad arrows >. The numbers – 1, 2, 3 – represent the nucleotide sequences, and the primed numbers – 1′, 2′, 3′ – the complementary sequence. The numbers written between the chains – 15, 16 17, 1, 2, 3 – represent the different copies of the identical sequences. The arrow → represents a 5′ end and the bar ⊣ a 3′ end of a polynucleotide chain.

repetitive sequence elements; while the latter have the size of single structural genes, the former could serve as sites for specific binding of regulatory macromolecules.

A different approach, which also is based on the reannealing of nucleotide sequences, is the cyclization of DNA after shearing and exonuclease treatment (40; fig. 15). It can be shown that such a procedure does not produce rings when used on prokaryotic DNA; in contrast, 15–30% of eukaryotic DNA can be observed as rings and complex circular forms after cyclization. The results obtained (41–45) have been interpreted in terms of Callan's model, as indicating that a significant proportion of chromosomal DNA (crudely estimated to be about 50 %) is organized into many thousands of regions (g-regions) each containing tandemly-repeating sequences.

ii. Sequence-dependent properties of double-stranded DNA

A number of physical properties of double-stranded DNA are sequence-dependent. This point was missed, for a long time because the use of essentially statistical DNAs, like bacterial DNAs, had led to the idea that the properties were simply composition-dependent. It had been known since 1962 that poly (dAT)·(dAT), poly (dA)·(dT) and poly (dG·dC) had "anomalous" buoyant densities (46), melting temperatures (47) and optical rotatory dispersions (48), compared to bacterial DNAs, but this had been attributed to the extreme base compositions of these polymers; similar "anomalies" were also found for some other repeating dinucleotides (49; fig. 16). It was only later, however, that evidence was given that 1) such "anomalies" also existed for satellite DNAs and for another natural DNA containing repetitive sequences, the mitochondrial DNA of yeast, and 2) that "anomalies" also involved binding of silver ions and binding on hydroxyapatite columns (50–53; Table III). Some of them, like the spectral properties, can be used to recognize simple repeating sequences (54–55); other ones, like buoyant density

Table III
% G·C contents

	Analysis	Buoyant density	T_m	$[\alpha]_{290}$
Mouse satellite DNA	35.2	30.6	40.8	
Guinea pig satellite DNA	38.5	45.9	40.7	
Yeast mitochondrial DNA	17.4	23.5	13.2	30.7
Mouse main DNA	40.0	40.0	40.0	
Guinea pig main DNA	39.4	39.7	39.7	

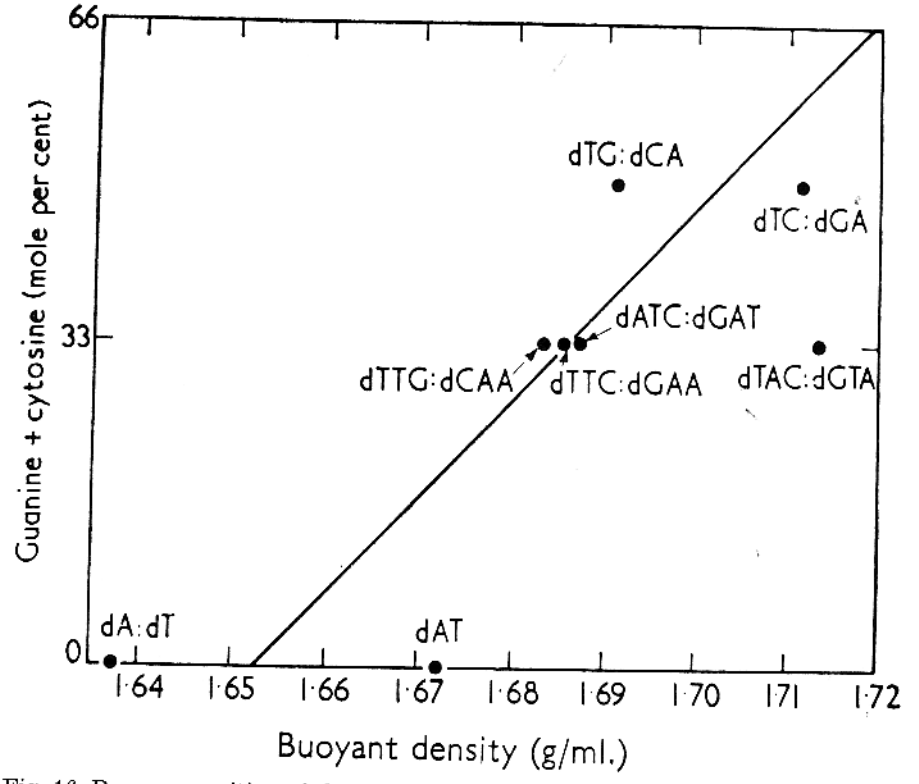

Fig. 16. Base composition of the DNA-like polymers as a function of their buoyant density values in a cesium chlorid gradient (54).

in the absence or in the presence of sequence-dependent ligands, like silver ions, are extremely important from a practical point of view, since they are the basis of very effective fractionation procedures.

We will briefly mention here a few examples of such fractionation methods, which in all cases depend upon the existence of short repetitive sequences in the DNA molecules which are separated.

Ribosomal and 5S-DNA (56)

Ribosomal DNA from *Xenopus laevis* was first isolated in 1966 by WALLACE and BIRNSTIEL (57) who had found that ribosomal RNA hybridized with heavy fractions from a preparative CsCl density gradient. The ribosomal DNA was found to consist of a series of repeating units, each of which included three major sequences: a gene for the 18 S RNA, a gene for the 28 S RNA and a "spacer" sequence that is not transcribed into RNA; the genes for the two RNA molecules are separated

by a second very small spacer sequence which is transcribed together with the genes into a 40 S precursor RNA which is subsequently processed into mature 18 and 28 S ribosomal RNAs (fig. 17). Very interestingly, while the *X. leavis* ribosomal genes show at least some sequence homology with ribosomal genes from more than 50 organisms including mammals, invertebrates, higher plants, protozoa, fungi and yeasts, the spacers sequences evolve very rapidly. If one compares the ribosomal DNAs, of *X. laevis* and *X. mülleri*, two species so closely related as to produce viable hybrid progeny, no difference can be detected among the ribosomal genes, yet the spacer sequences are largely different in the two species, the spacers being, however, very similar and perhaps identical to each other in each species.

5 S DNA was isolated in BROWN's laboratory from the light fractions of a CsCl peak of *X. laevis* DNA. It was shown that in this case, too, 5 S RNA genes were regularly interspersed with spacer sequences. As for ribosomal genes, 5 S genes were very similar and perhaps identical in *X. laevis* and *X. mülleri*, whereas the (untranscribed) spacers were different in sequence and also in length; the spacer sequences of a single species were, however, very similar but not quite identical (fig. 18).

The chromosomal localization of the ribosomal and 5 S genes was established by *in situ* hybridization. Ribosomal genes were found on only one chromosome per set, at the secondary constriction known as the nucleolar organizer, the site where the nucleolus is known to form in the actively synthesizing cell. On the other hand, most if not all of *Xenopus* chromosomes contain 5 S DNA, and it is always at the end of the chromosomes.

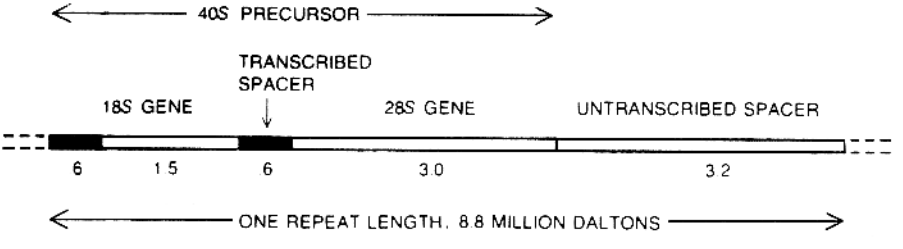

Fig. 17. Ribosomal DNA of *Xenopus* is mapped in this diagram of one repeat length based on chemical and electron-micrograph data. The gene region codes for a 40 S RNA molecule that is the precursor of the 18 S and 28 S RNAs, which are cleaved from it in the living cell by enzymes. Two small spacers in the gene region are transcribed into precursor but eliminated in the cleaving process. Then there is the main spacer region, which is not transcribed (33).

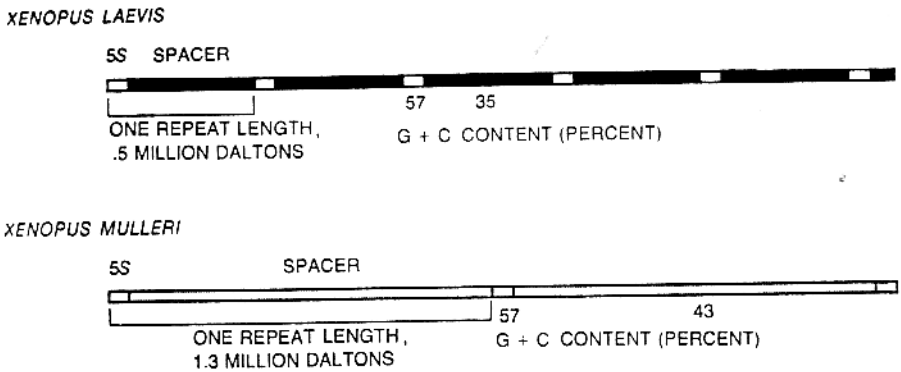

Fig. 18 Spacer regions are very different in *X. laevis* and *X. mülleri* 5 S DNA, in length as well as in nucleotide composition. The gene regions are the same length in both and appear to synthesize the same 5 S RNA. The gene accounts for about a seventh of the total repeat length in *X. laevis* but only for about an eighteenth of the much longer repeat length in *X. mülleri* (33).

Finally, it should be pointed out that both ribosomal and 5 S DNAs make up only about 0.7% of *Xenopus* DNA; this corresponds to 450 ribosomal genes and 24,000 5 S genes per genome, respectively.

Satellite DNAs (58)

Another case where rather detailed information is available is that of "satellite" DNAs. Many eukaryotic DNAs, when centrifuged to equilibrium in a CsCl density gradient show more or less separated, minor, "satellite" bands. More such components can be prepared by techniques involving differential silver, mercury or actinomycin binding. The general properties of "satellite" DNAs may be summarized as follows: 1) They usually represent less than 10–20 % of the eukaryotic genome; 2) They contain short, repetitive sequences, which have been sequenced in some cases; as a consequence, they renature rapidly, show CsCl bands and melting curves having standard deviations lower than those of bacterial DNAs, and often undergo strand separation in alkaline CsCl; 3) They do not seem to be transcribed; 4) They are localized at the centromere of one or more chromosomes; 5) They show striking changes in sequence in closely related species, indicating a rapid evolution. It is evident that "satellite" DNAs share many of the properties just described for the "spacers" above. The "satellite" DNAs isolated from the calf genome are described in the following section.

Main band DNA (59)

The resolving power of Ag^+-Cs_2SO_4 density gradient is well illustrated by recent work on the calf genome (59). This could be resolved into a number of components, which may be divided into three groups: 1) the main DNA component (1.697 g/cm³; all densities quoted are those determined in CsCl density gradients), the 1.704 and 1.709 g/cm³ components form about 50, 25 and 10% of the genome, respectively; they are characterized by having symmetrical CsCl bands and melting curves, both of which have standard deviations close to those of bacterial DNAs of comparable molecular weight, and by their G + C contents being equal to 39, 48 and 54%, respectively; after heat-denaturation and reannealing, their buoyant densities in CsCl are greater than native DNA by 12, 10 and 3 mg/cm³, respectively; 2) the 1.705, 1.710, 1.714 and 1.723 g/cm³ components represent 4, 1.5, 7 and 1.5% of the DNA, respectively, and exhibit the properties of "satellite" DNAs; their CsCl bands and melting curves have standard deviations lower than those of bacterial DNAs; after heat-denaturation and reannealing, their buoyant densities are identical to native DNA, except for the 1.705 g/cm³ component, which remains heavier by 5 mg/cm³; in alkaline CsCl, only the 1.714 g/cm³ component shows a strand separation; 3) a number of minor

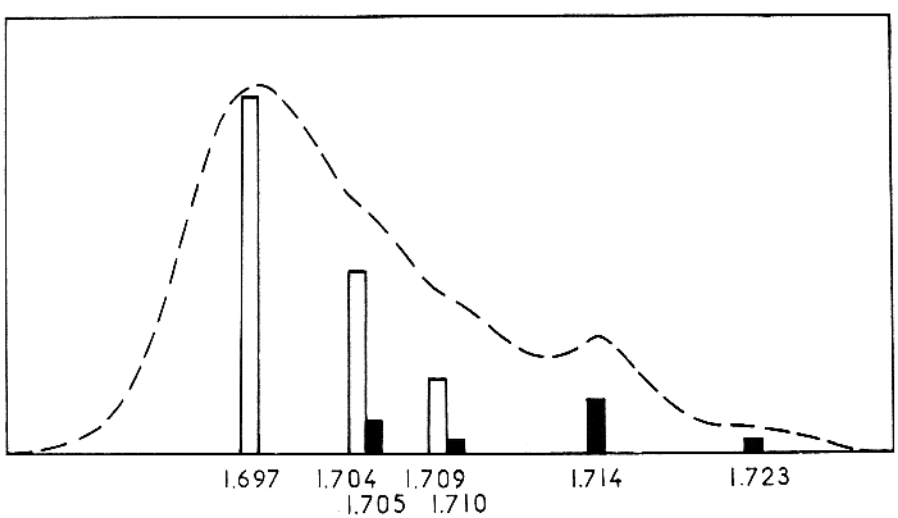

Fig. 19. A histogram of the DNA components of the bovine genome. The height of the bar is proportional to the percentage of each component; filled bars correspond to the sharp melting, empty bars to the broad melting components. The broken line is an enlarged band pattern in CsCl density gradient of total calf DNA (59).

components, forming 1% of the DNA, have been recognized, but they have not been investigated in detail; two of them (1.719 and 1.699 g/cm^3) might correspond to ribosomal and mitochondrial DNA, respectively. A histogram of the DNA components of groups 1 and 2 is shown in fig. 19. Four of these components, forming together 14% of the genome, are characterized by the high intramolecular homogeneity typical of satellite DNAs and do not deserve any additional comments. The demonstration of three well-defined components (1.697, 1.704 and 1.709 g/cm^3), forming 85% of the genome, sharing an evident intramolecular heterogeneity but endowed with distinct properties, is a new, important finding, since it is at variance with the widespread view that the bulk of the genome of higher organisms is formed by DNA molecules showing a continuous variation in their G+C content. In contrast with the satellite components, which are known to be quite different even in closely related species, the three components accounting for the "main-band" DNA may have a wider relevance. It is well known that mammalian DNAs are very close in over-all base composition, in doublet frequency (see below) and in the frequency of short oligonucleotide sequences as revealed by DNases (see next section). On the other hand, their CsCl bands show almost identical buoyant densities and a distortion on the heavy side that is not at all, or not exclusively, due to the presence of satellite components. It is worth considering, and our very recent results speak in favor of this idea, that the main band components found in the calf genome exist in all mammalian genomes.

b) Direct methods

The direct approaches to the study of nucleotide sequences in DNAs can be divided into two groups, according to whether they are aimed at the sequences themselves or at the frequency of the sequences.

i. Sequence methods

A number of approaches to the study of sequences in DNAs have been developed. They use a DNA repair copy (60), RNA copy (61) or a direct approach (62), to mention only the main possibilities. In the area of eukaryotic DNAs, sequence work has been done on the satellite DNAs of the kangaroo rat, *Dipodomys ordii* and of several species of *Drosophila* (63, 64).

ii. Frequency methods

Among the frequency methods, depurination (65) is not very informative. Suffice it to say that no large differences are seen in the pyrimid-

ine isostichs of T7 phage and calf thymus DNA. Depurination has been very useful, however, in giving the first direct evidence for the presence of short repetitive sequences in the α-satellite of guinea pig (66). Owing to a particularly fortunate situation, it was possible to isolate from the light strand the tetranucleotide T_1C_3 in a 21.5% yield, and from the heavy strand the dinucleotide T_2 in a 28.3% yield thus giving a strong hint that the hexanucleotide CCCTAA was the basic repeating unit on the light strand.

The nearest neighbor analysis method (67) gives the frequency of all doublets in a given DNA; the method involves DNA-polymerase copy

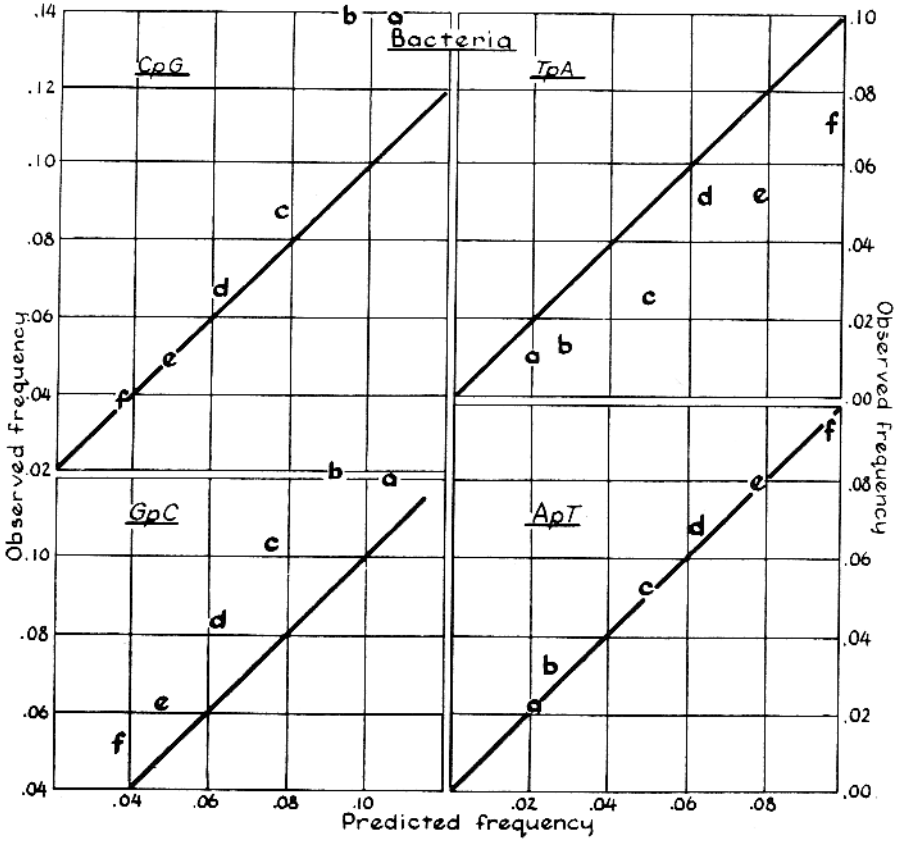

Fig. 20. Sequence frequencies in bacterial DNAs compared with values predicted from random association. a, Micrococcus lysodeikticus; b, Mycobacterium phlei; c, Aerobacter aerogenes; d, Escherichia coli; e, Bacillus subtilis; f, Haemophilus influenzae (69).

of the DNA under investigation using one αP^{32}-labeled deoxyribonucleoside triphosphate at a time; DNA is then degraded by enzymes leaving the label on the 3′ position of the nearest neighbor nucleotide. Data can be compared with statistical expectations (68; figs. 20, 21). In the case of bacterial DNAs essentially linear relationships hold between the experimental values and those based on random association (fig. 20).

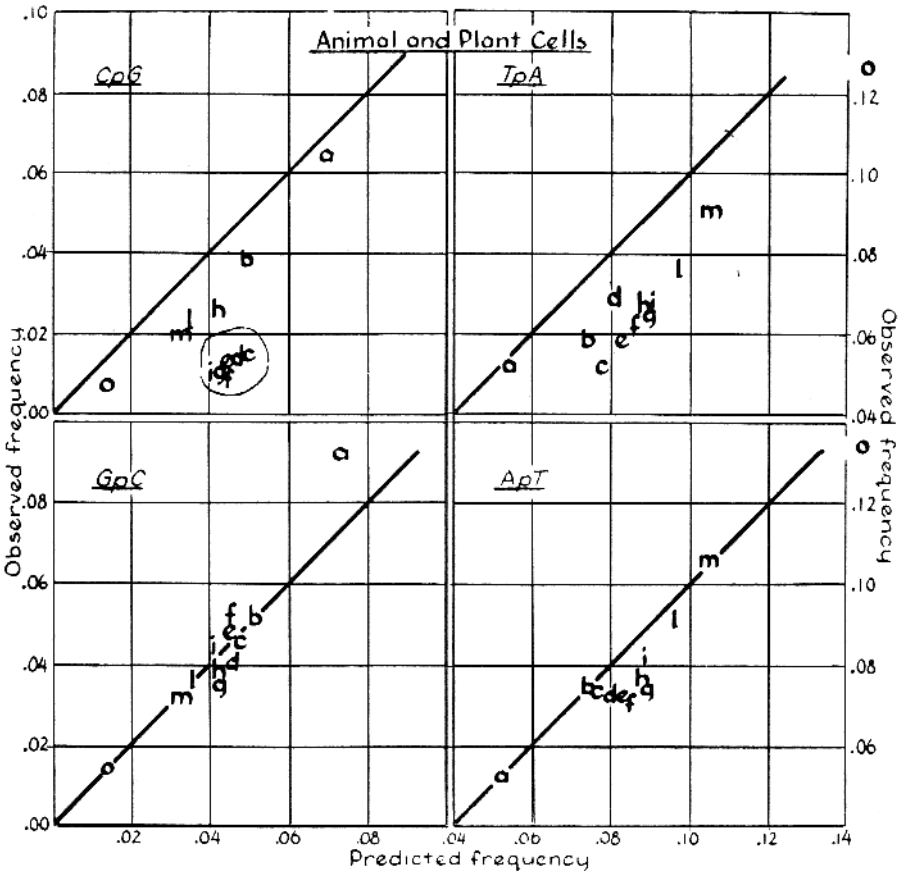

Fig. 21. Sequence frequencies in animal and plant cell DNAs compared with values predicted from random association. The product of frequencies of constituent nucleotides of a nearest neighboru pair is described by the line passing through the origin. The values represented by the lower case letters are the observed nearest neighbour frequencies. a, Chlamydomonas; b, wheat germ; c, calf thymus; d, salmon liver; e, rabbit liver; f, chicken red cells; g, mouse lymphoma; h, starfish testis; i, human spleen; l, Echinus esculentus; m, Paracentrotus lividus; o, Tetrahymena pyriformis (69).

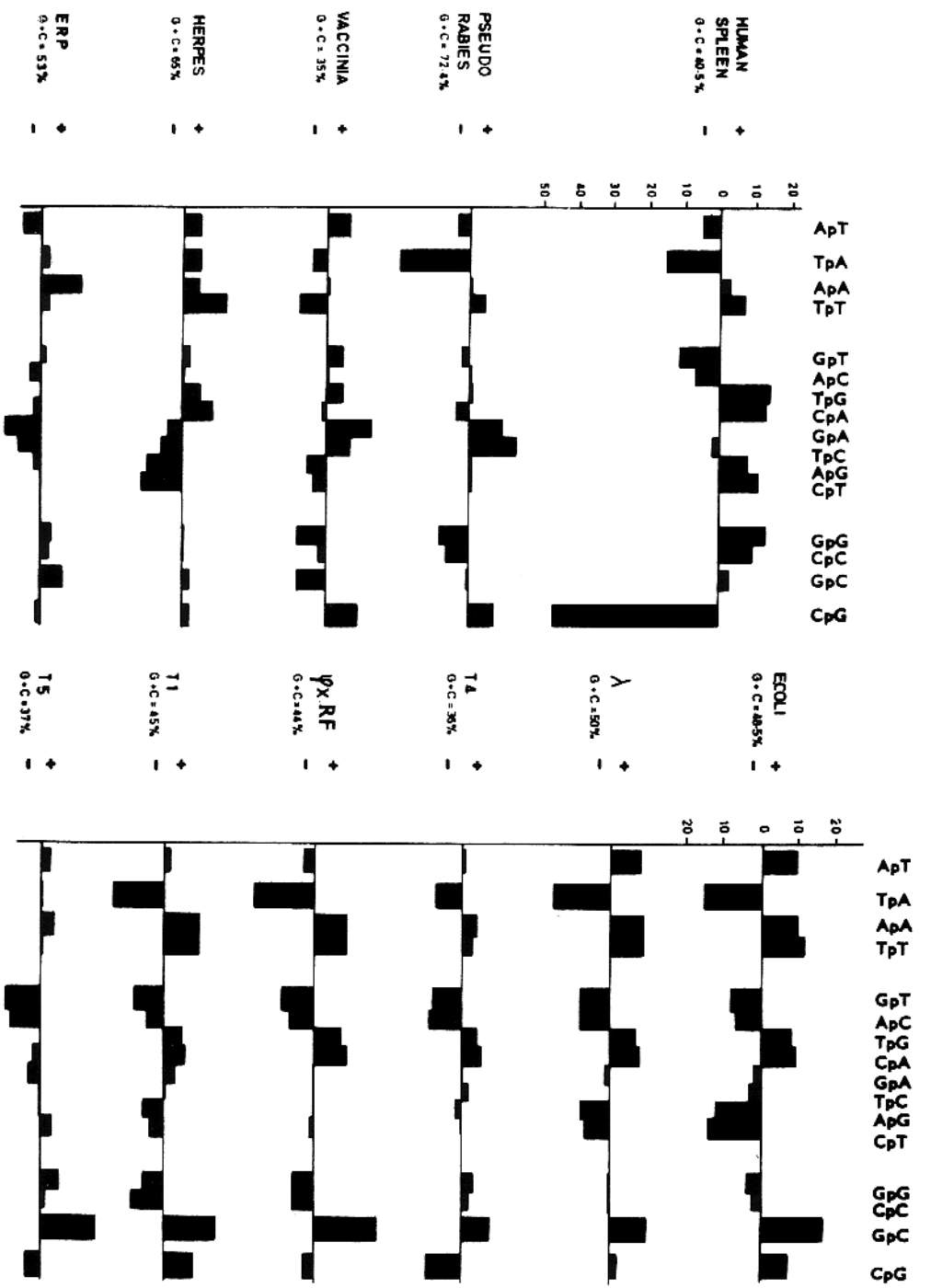

Fig. 22. Doublet frequency patterns of DNAs normalized to 50% (G + C) and expressed in terms of deviation in parts 1000 from random expectation (62.5 parts for each doublet). (G + C) contents given for each DNA are those calculated from nearest neighbour analysis ERP: equine rhino-pneumonitis (70).

In the case of eukaryotic DNAs, a very marked and characteristical CpG shortage holds for vertebrates (fig. 21). A very useful presentation of data is given by the doublet frequency patterns normalized to 50% (G + C) and expressed in terms of deviation from random expectation (70; fig. 22).

3. A new approach to the study of nucleotide sequences in DNAs

a) Specificity of DNases

i. Methodology

In order to study the specificity of DNases, it is necessary to determine the composition of the nucleotides adjacent to the breaks introduced by the enzymes. We will call *termini* the two nucleotide pairs lying on each side of the breaks; each pair is formed by a terminal and a penultimate nucleotide and is characterized by its position at the 3′ or at the 5′ end of the nucleotide chain. In fig. 23, WXYZ correspond to the 3′ penultimate, 3′ terminal 5′ terminal and 5′ penultimate nucleotides, respectively.

For the isolation and analysis of termini we have used two different approaches, the first one involving the use of column chromatographic procedures on "cold" DNA, the second one taking advantage of the possibility of radioactive labeling of the newly formed ends.

The first approach involves two basic operations, the isolation of termini and their analysis. For the isolation of termini we have developed procedures which have been described elsewhere (71) and which can be outlined as follows: a) 3′ terminal nucleotides (fig. 24; ref. 72): oligonucleotides are first dephosphorylated with spleen phosphatase B (73), then hydrolyzed with spleen exonuclease (74), an enzyme which releases 3′ nucleotides starting from the 5′ end; the 3′ terminal nucleotide is released as a nucleoside, and can be separated from the 3′ nucleotides formed; b) 5′ terminal and penultimate nucleotides (fig. 25; ref. 75): oligonucleotides are dephosphorylated, degraded with pancreatic DNase

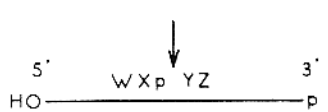

Fig. 23. A scheme of degradation of DNA by a DNase; the sequence is written in the usual 5′→3′ direction; the vertical arrow indicates the position of the break; the phosphate is arbitrarily put at the newly formed 3′ end.

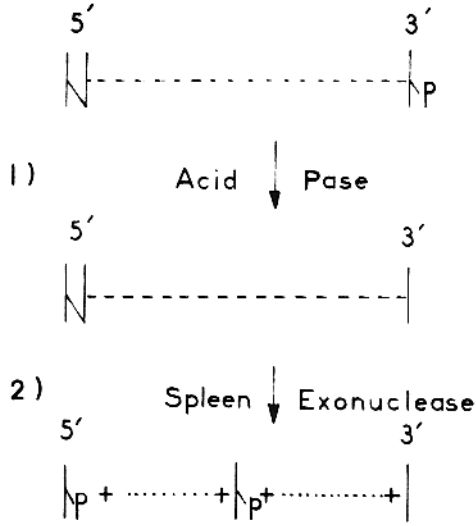

Fig. 24. Isolation of 3′ terminal nucleotide. The terminal phosphate is arbitrarily put at the 3′ end. The same procedure can be applied to 5′-Poligonucleotides; in this case dephosphorylation has the function of allowing spleen exonuclease to act.

in order to decrease their average chain length, and digested with snake venom exonuclease (75); this enzyme splits off 5′ nucleotides starting from the 3′ end; 5′ terminal dinucleoside monophosphates, being very resistant to digestion by venom exonuclease because they lack the 5′ phosphate, accumulate in the digestion mixture up to levels higher than 90% of the theoretical yield and can be separated from the 5′ nucleotides formed; they are then split with spleen exonuclease into 3′ nucleotides (corresponding to the 5′ terminals) and nucleosides (corresponding to the 5′ penultimates), which can be separated from each other; c) 3′ penultimate nucleotides (fig. 26; ref. 76): the isolation of this terminus was performed only on spleen acid DNase digests of calf thymus DNA using two methods. The first one (which can be used for any 3′ phosphate-ended digest) involved (fig. 26) splitting the digest with pancreatic DNase in the presence of Mn^{++}; the dinucleoside triphosphates, which represent about half of the products originated from the 3′ phosphate ends of the fragments, were isolated, dephosphorylated and analyzed as described above for dinucleoside monophosphates. The second method consisted in quantitatively splitting, with pancreatic DNase in the presence of Mn^{++}, the tetranucleotides (which can represent as much as 25% of the

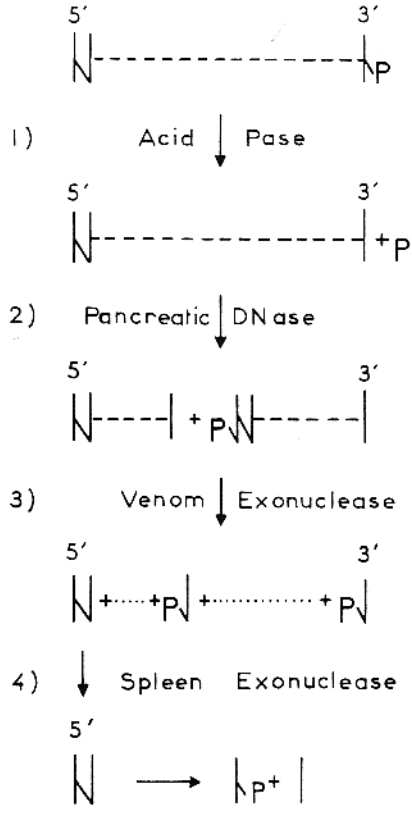

Fig. 25. Isolation of the 5′ terminal and 5′ penultimate nucleotides. The terminal phosphate is arbitrarily put at the 3′ end. The same procedure can be applied to 5′-Poligonucleotides; in this case dephosphorylation has the function of allowing venom exonuclease to stop at the terminal doublet.

isostichs in spleen acid DNase digests) into doublets, these were subsequently isolated and analyzed; it should be mentioned that the analysis of the 3′ terminal, 5′ terminal and 5′ penultimate nucleotides of tetranucleotides showed values in agreement with those obtained in total digests, suggesting that the 3′ penultimate nucleotides of tetranucleotides also were representative of those of total digests. The results obtained by this method were in agreement with those obtained by the first one.

The termini as isolated by the above procedures have been analyzed as nucleosides by one of the micromethods developed in our laboratory (77–80), the precision being of the order of 1% for the 3′ terminal and of 2% for the 5′ terminal and penultimate nucleotides.

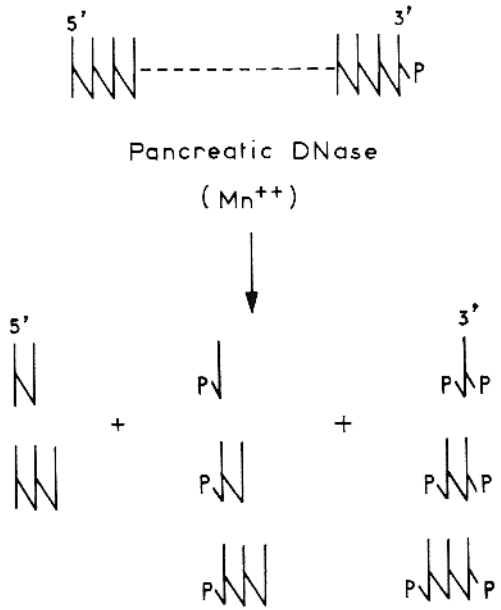

Fig. 26. Products formed by digestion of spleen acid DNase oligonucleotides with pancreatic DNase in the presence of Mn^{++}.

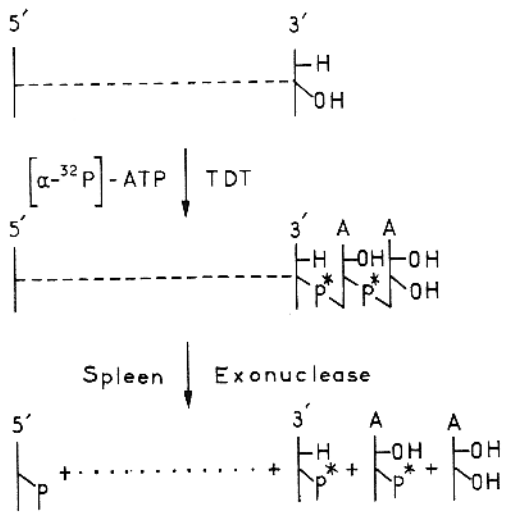

Fig. 27. Isolation of labeled 3′ terminal nucleotides using terminal deoxyribonucleotidyl transferase.

The second approach involves the enzymatic labeling with ^{32}P of the ends released by DNases: a) 3′ terminal nucleotides (fig. 27; ref. 81): oligonucleotides (previously dephoshoprylated if 3′ phosphate-ended) are used as primers for the addition of two ribonucleotides, using terminal deoxyribonucleotidyl transferase (82) to catalize the reaction (83) and α-^{32}PATP as donor; the labeled oligonucleotides are degraded with spleen exonuclease and the 3′ phosphate labeled nucleotides are separated from each other and from labeled AMP on DEAE-cellulose; b) 5′ terminal nucleotides (fig. 28; ref. 84): oligonucleotides (previously dephosphorylated if 5′ phosphate-ended) are used as acceptors of ^{32}P using polynucleotide kinase (85) to catalize the reaction and γ-^{32}PATP as donor; the labeled oligonucleotides are degraded with pancreatic DNase and venom exonuclease and the 5′ phosphate labeled nucleotides are separated from each other and from labeled inorganic phosphate and ATP by chromatography on polyethylene imine thin-layer plates. It should be mentioned that the kinase labeling procedure lends it self to the preparation of the 5′ terminal dinucleosides using *E. coli* exonuclease I (86) instead of venom exonuclease for the degradation of oligonucleotides; and that the extremely fast separation method on polyethylene imine plates (84) can also be used for the analysis of the 3′ terminal nucleotides.

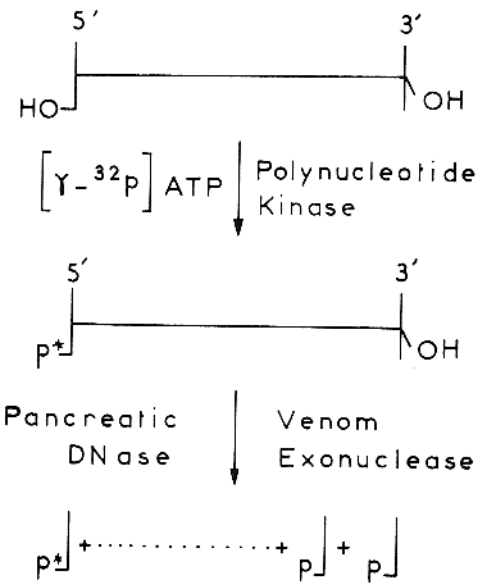

Fig. 28. Isolation of labeled 5′ terminal nucleotides using polynucleotide kinase.

Concerning the methodology for the analysis of termini, two general remarks should be made. The first one is that it can be applied to both 3′ and 5′ phosphate-ended oligonucleotides (except for the first method for the 5′ penultimate nucleotides). The second one is that it heavily relies on the quality of the techniques used for isolation and analysis, as well as on the purity of enzymes used, particularly of the phosphatase and the exonucleases.

ii. Results

Four DNases have been investigated so far: acid DNase B from hog spleen (87), acid DNase from the hepatopancreas of *Helix aspersa*, Müll. (88), bovine pancreatic DNase (89), and *E. coli* endonuclease I (90). The data obtained for the composition of the termini released from calf thymus DNA are shown in Table IV. They deserve several comments.

a) Each terminus, as released by any enzyme, has a different base

Table IV
Termini liberated from calf thymus DNA by four DNases

		3′terminals	5′terminals	5′penultimates
Spleen DNase*	T	20	11	14
	G	43	43	26
	A	29	18	52
	C	8	28	8
Snail DNase	T	16	14	38
	G	6	45	24
	A	78	10	21
	C	1	31	17
Pancreatic DNase**	T	36	38	13
	G	15	22	36
	A	31	15	30
	C	18	25	21
E. coli endonuclease I	T	41	24	28
	G	8	35	29
	A	35	17	29
	C	16	23	14

* In this case, the average chain length of oligonucleotides was equal to 15. 3′-penultimate nucleotides were also analyzed; they were T 22%, G 16%, A 46%, C 16%.

** Digestion was carried out in the presence of Mg^{++}. Results obtained in the presence of Mn^{++} are very slightly different.

composition; this never is equal to the base composition of the DNA, a finding ruling out the possibility that enzymatic degradation is random, since in this case all termini should have the base composition of the DNA. The possibility that some termini have a base composition different from that of DNA simply because they are the nearest neighbors of termini specifically recognized by the enzymes can be checked by comparing the experimental results with those expected from the nearest neighbor data. This type of control (Table V has shown that all termini released have a composition which differs from that expected from the nearest neighbor data and are indeed recognized. The only exception is given by the 5′ penultimate nucleotides released by the snail enzyme; these have the composition expected for the nearest neighbors of the 5′ terminal nucleotides and are therefore not recognized (91).

b) On the basis of the results just mentioned, we can conclude that the minimum length of the nucleotide sequences recognized by the enzymes investigated is four for spleen acid DNase, (in which case the 3′ penultimate position was also analyzed), two for snail DNase, three for pancreatic DNase and *E. coli* endonuclease I. It is likely that the actual length of the recognized sequences is not much greater than the minimum length for two reasons: 1) the presence in the final digests of oligonucleotides as short as di- and/or trinucleotides suggests that the shortest sequences which can be split are 4–6 nucleotides long; 2) the enzymes under consideration have small molecular weights ranging from 38,000 for spleen acid DNase (92), to 30,000 for snail acid DNase (88), 31,000 for pancreatic DNase (92), and about 24,000 for *E. coli* endonuclease I (94), and it is unlikely that they get in contact with long nucleotide

Table V

Average composition of sequences split by acid DNase in calf thymus DNA at $\overline{P}_n = 15$*

	3′P penultimate		3′P terminal			5′OH terminal			5′OH penultimate	
	exp.	calc.	calc.	exp.	calc.	calc.	exp.	calc.	calc.	exp.
T	22	(31)	(29)	20	(32)	(29)	11	(29)	(29)	14
G	16	(21)	(22)	43	(21)	(23)	43	(23)	(19)	26
A	46	(30)	(29)	29	(30)	(29)	18	(30)	(31)	52
C	16	(19)	(20)	8	(17)	(18)	28	(18)	(20)	8

* Values in parentheses indicate the composition of each terminus as calculated from its nearest neighbour(s). For the terminal positions calculated values could be obtained from both neighbouring positions.

sequences, as do restriction enzymes which have much higher molecular weights.

c) The conclusion that the DNases investigated here have a sequence specificity, as opposed to the single-base specificity of RNases, very probably is of general singificance. It is well known that restriction enzymes are also sequence-specific; the main differences with the DNases considered here are the greater length of the recognized sequence and the much higher specificity of the restriction enzymes.

d) The results of Table II show that the DNases tested here split specific sets of nucleotide sequences and that these sets are different for each enzyme. A rough estimate of the number of sequences forming the specific sets recognized by each enzyme can be obtained by considering the average chain length of the final DNase digests. In the case of calf thymus DNA this is comprised between 2 for pancreatic DNase-Mn^{++} (76) and 4.5 for spleen acid DNase (95); therefore, the number of sequences which can be split by the enzymes under consideration is in the range of 50% to 20% of all sequences. This estimate is correct only if all sequences are equally frequent; since this is not the case, the percentage of sequences which can be split is, in fact, higher or lower than that indicated above, but very probably not too far from it. On the other hand, the percentages of sequences which can be split may be higher than those just mentioned since the average size the final digest might be lower than those indicated above, at least for some enzymes, if the inhibitory effect of the reaction products could be completely eliminated; in addition, susceptible sequences are likely to overlap to some extent and breaks introduced by the enzymes may hinder hydrolysis of neighboring phosphodiester bonds. It seems safe to conclude, therefore, that the number of sequences split by the enzymes is so large that the sequence sets recognized by them certainly overlap with each other to some extent. Yet, the fact that the composition of termini released by the four DNases differs according to the enzyme used indicates that the sets of sequences recognized still largely differ for different enzymes.

e) The data of Table IV (80, 96–98) represent the apparent average base composition of all the sequences split by the enzymes in calf thymus DNA; this is related to the actual average base composition of all sequences split through the set of k_M and/or V_{max} values associated with the split sequences. As an example of the effect of such differences in the k_M and/or V_{max}, we can quote results obtained with poly (dAT:dAT). This polymer, which contains equal amounts of the tetranucleotides ATAT and TATA releases, upon spleen acid DNase degradation, 3′ ter-

minal A and T in a ratio 80:20, thus indicating that the two sequences are split with different k_M and/or V_{max}. The relative importance of these two factors is unknown so far, but the findings that the affinity of all DNases used (except for the pancreatic enzyme) for polyribonucleotides is comparable with, or higher than, that for DNA (88, 90, 99, 100), in spite of the large structural differences, suggests that k_M values for different sequences may not differ very much from each other and that the main factor is V_{max}. It is important to note that differences in k_M and/or V_{max} for the two sequences present in poly (dAT:dAT) do not lead to a variation in the percentage of terminal nucleotides released in the average chain length range 40 to 15, indicating that both sequences are still present at saturating levels (at the substrate concentrations used) when as much as 7% of all sequences have been split.

f) The lack of variation in the relative amounts of different termini as released by different enzymes seems to be the rule, and reflects the fact that termini derive from a very large number of sequences and are therefore associated with average k_M and V_{max} values. An additional reason may be that the majority of sequences have k_M and V_{max} values which are comprised in a narrow range. In the range of average oligonucleotide chain length 50 to 8, no change has been seen in the composition of any of the termini released from calf thymus DNA by any of the DNases investigated. (fig. 29 shows, as an example, the results obtained

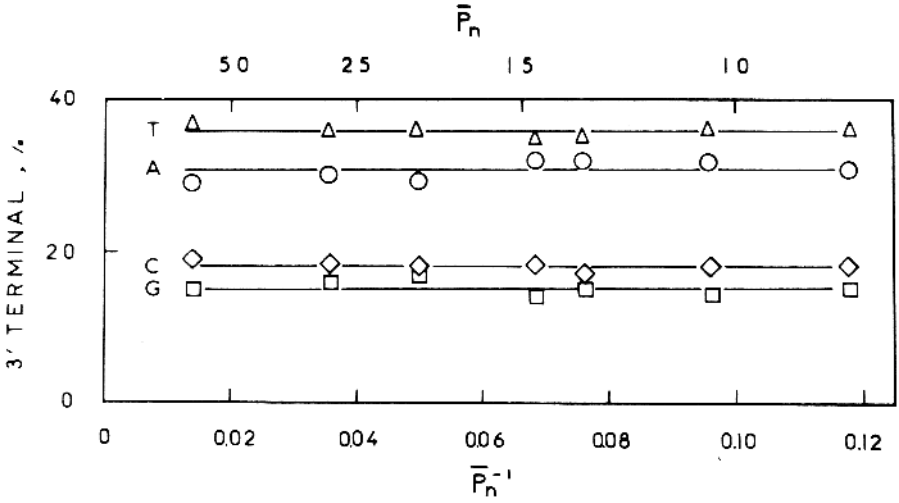

Fig. 29. 3'-terminal nucleotides as obtained at different degradation levels of calf thymus DNA by pancreatic DNase in the presence of Mg^{++}. $\bar{P}_n$ is the number average degree of polymerization of the digests expressed in nucleotide units.

with the 3′ terminals of pancreatic DNase), with the single exception of the 3′ terminals released by spleen acid DNase (see below).

g) Another implication of the invariance of the composition of termini in the 50 to 8 range of average chain length of the digests (fig. 29) is that the specificity of the DNases used is not affected by the secondary structure of the substrate. In fact, melting of double-stranded DNA fragments takes place in the average chain length range explored, under the experimental conditions used; this causes a striking slow-down of the reaction rate, due to the lower affinity of the enzymes for single-stranded *versus* double-stranded DNA, but no change in the composition of termini. The special case of spleen acid DNase is discussed in ref. 80.

b) The analysis of termini formed by DNases (101, 102)

Since DNases split specific sets of sequences, the analysis of termini provides information of the frequency of these sequences in a given DNA. In fact, the composition of termini is related: a) to the average composition of the sequences which can be split by the enzymes; b) to the k_M and V_{max} values associated with the sequences affected, and c) to their relative amount in the DNA under consideration (fig. 30).

The easiest way to check the latter effect is to determine the composition of termini as obtained from DNA's having different G + C contents and, therefore, different relative amounts of different sequences. Indeed,

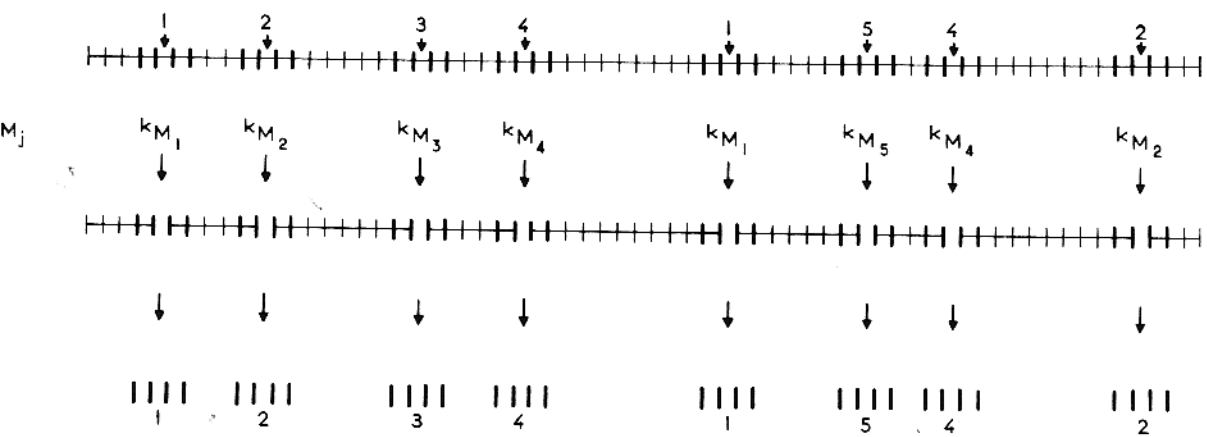

Fig. 30. Scheme of the analysis of termini. A number of sequences, indicated as tetranucleotides and numbered 1 to 5, are recognized and split with different k_M and/or V_{max}, indicated by k_1, k_2 etc. Termini are isolated from the resulting oligonucleotides, and the base compositions of each terminus, WXYZ, are determined.

the compositions of termini released by bacterial DNA's having different G + C contents are different. Furthermore, if the percentages of A, G, C and T in the termini of bacterial DNAs are plotted against their G + C contents, linear relationships are obtained (fig. 31). The choice of bacterial DNA's for looking at the effect of changes in the relative amount of sequences on the composition of termini is due to the following reasons: a) bacterial DNAs do not contain repetitive sequences and b) the doublet frequencies of bacterial DNA's, as determined by the nearest neighbor analysis, show essentially linear relationships with the frequencies predicted for random association, (68), indicating a common type of doublet distribution in these DNA's.

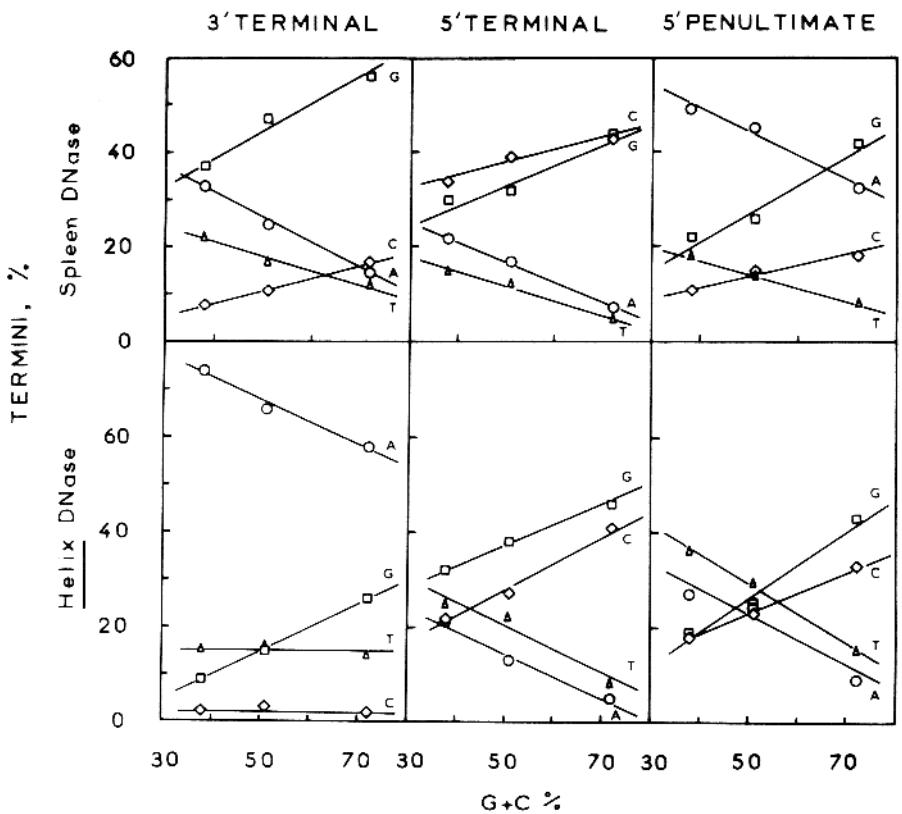

Fig. 31. The percentage of A (○), G (□), C (◇) and T (△) in the 3'-terminal, 5'-terminal and 5'-penultimate nucleotides formed by the spleen and the snail DNase from bacterial DNA's (*H. influenzae*, 38% G + C; *E. coli*, 51% G + C; *M. luteus*, 72% G + C), is plotted against the G + C contents of DNA's. Values obtained at an average degree of polymerization of the digests equal to 15 were used.

It can be expected that the composition of termini released from DNAs containing short repetitive nucleotide sequences deviates, in either direction, from the linear relationships obtained with "non-repetitive" (bacterial) DNAs. This is in fact found when "repetitive" DNAs are analyzed. Fig. 32 shows, as an example, a plot of the 5′ terminal nucleotides, as released by spleen acid DNase from poly (dAT:dAT), crab, guinea pig and mouse satellite DNAs and yeast mitochondrial DNA *versus* their G + C contents. The experimental values show strong deviations from the linear relationships of bacterial DNAs. A very convenient way of plotting such data is to plot differences in percentages of termini, as found in "repetitive" DNAs, from those expected for

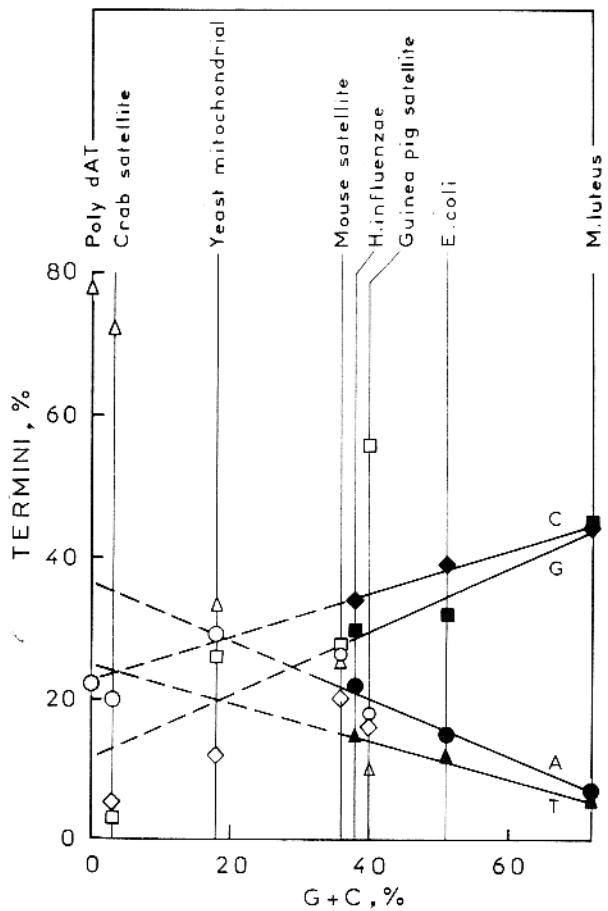

Fig. 32. Composition of 5′ terminal nucleotides as released by spleen acid DNase from "repetitive" DNAs as plotted against their G + C contents (open symbols). Data for bacterial DNAs are also shown (filled symbols).

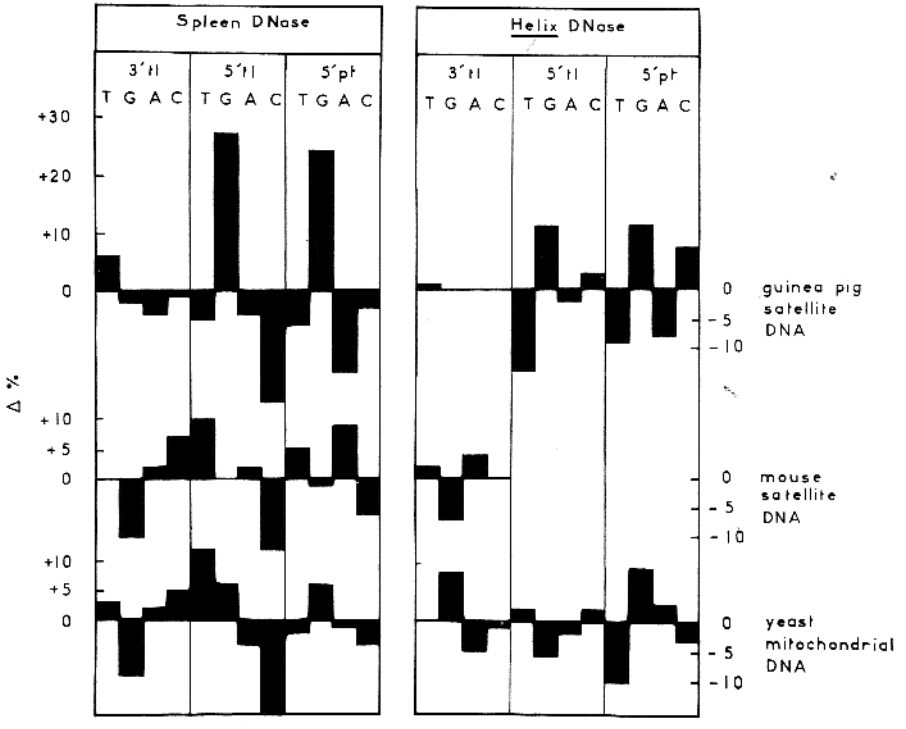

Fig. 33. Deviation patterns of three "repetitive" DNAs. The histograms show the differences between the composition of termini formed from guinea pig satellite, mouse satellite and yeast mitochondrial DNAs by spleen and snail DNases and the compositions expected for bacterial DNAs having the same G + C contents; Δ values represent differences in the percentages of each terminus.

bacterial DNAs having the same, base compositions. The derivation patterns which can thus be obtained represent a novel way of characterizing "repetitive" DNAs or, more generally, DNAs having sequence distributions different from those of the bacterial DNAs examined here. Fig. 33 shows the deviation patterns of three DNAs containing short repeated sequences: the satellite DNAs from mouse and guinea pig and the mitochondrial DNA from yeast. These results indicate that a number of short nucleotide sequences, recognized and split by the enzymes, are present in these DNAs, in amounts which are larger or smaller than those present in bacterial DNAs of comparable G + C contents. Such different amounts of short sequences are the source of the different compositions of termini shown in fig. 32 and originate the deviation patterns of fig. 33. It may be interesting to remark that deviation patterns obtained with the same DNA as degraded by different enzymes are different, since

different sets of short sequences are split, and that deviation patterns obtained with different DNAs as degraded by the same enzyme are also different since the frequency of the short sequences split by the enzyme is different in different DNAs.

As another example, fig. 34 shows deviation patterns obtained with calf, mouse and guinea pig DNAs and with yeast nuclear DNA. The result obtained with eukaryotic DNAs lead to a number of conclusions: a) a number of short sequences are present in larger or smaller amounts in the eukaryotic DNAs examined compared to bacterial DNAs of comparable G + C contents; b) mammalian DNAs have very similar devia-

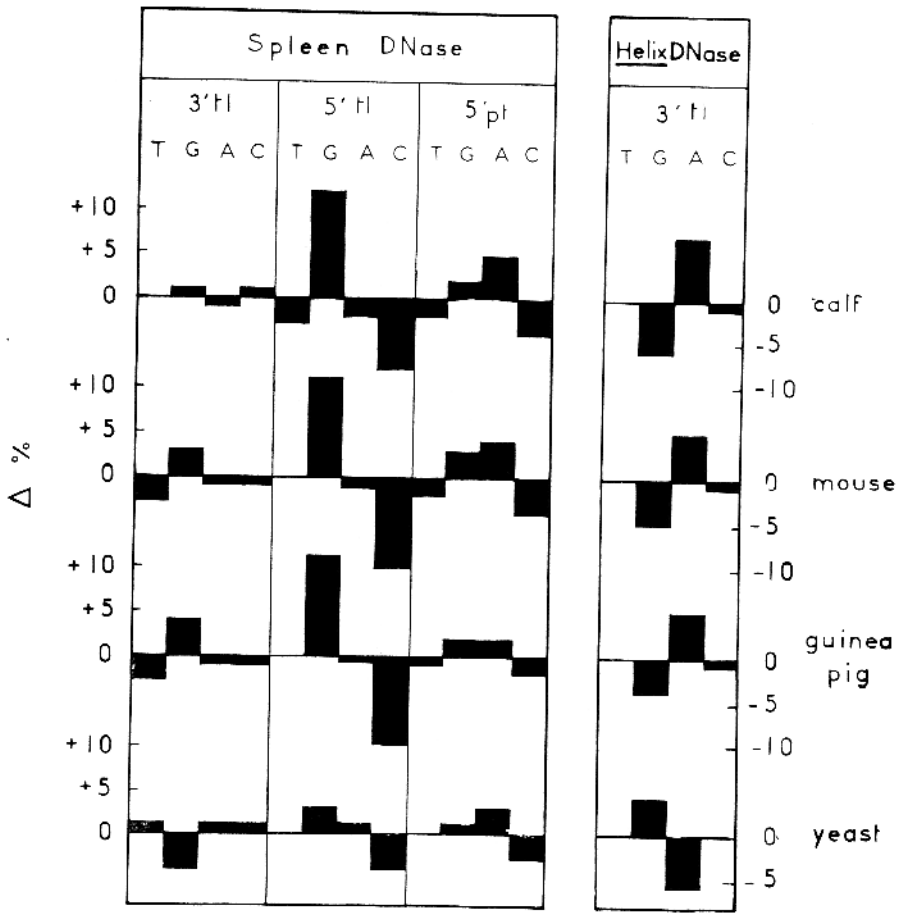

Fig. 34. Deviation patterns of four eukaryotic DNAs. In the case of yeast DNA a nuclear DNA preparation free of mitochondrial DNA was used. See legend of fig. 7 for the presentation of data.

tion patterns, whereas yeast nuclear DNA has a different deviation pattern characterized by lesser deviations from bacterial DNAs; c) as for the origin of the deviations observed in mammalian DNAs, they evidently do not arise from the repetitive sequences of the satellite DNAs they contain, since the deviation patterns of mouse and guinea pig satellite are very different from each other and from the corresponding total mammalian DNAs (compare fig. 2 and Fig. 4); the deviations must therefore originate in nucleotide sequences contained in the main (1.697 g/cm^3) DNA component and/or in components similar to the 1.704 and 1.709 g/cm^3 components observed in the bovine genome (59); such sequences must share common features in mammalian DNAs and also must be abundant, as suggested by the fact that the repetitive sequences of satellite DNAs, which only represent 5–10% of genome in the case of mouse and guinea pig, cannot be recognized in the deviation patterns of total mammalian DNAs.

In conclusion, the analysis of termini represents a new, direct approach to the study of nucleotide sequences in DNA's and, as such, it adds to the already existing methods (Table II). The main features of this new frequency method, which is particularly useful when a comparison of sequence frequencies of different DNA's is needed, are the following: a) The method is a rigorous analytical method; our present ignorance of the k_M and V_{max} values associated with each sequence is irrelevant as far as the practical use of the method is concerned, because these values are simple proportionality constants, and because the analysis actually bears on termini as derived from a number of different sequences. b) The method is very flexible. Different enzymes with different specificities can be used, leading to the analysis of different sets of sequences, thereby permitting the use of the most sensitive deviation patterns obtained; using several enzymes permits to cover the large majority, if not the totality, of nucleotide sequences in DNA's. In addition, the analysis can be limited to only one, or extended to several termini according to the degree of information required; 5′ terminal doublets can be analyzed as such, yielding more information than obtained after splitting; c) The main differences existing between the present method and another frequency method, the nearest-neighbor analysis, are 1) that the latter concerns the frequency of all doublets, whereas the analysis of termini bears on the sequences selected by the enzyme used; and 2) that the sequences recognized by the enzymes are longer than the dinucleotides investigated by the nearest neighbor analysis; a greater amount of information can therefore be obtained from the analysis of termini compared to the nearest neighbor analysis.

Discussion

Question: How correct are your figures about the genetic information in *E. coli* and man? You told us that it is a factor of about five from the protein variety estimate. Is it really correct? Wouldn't it be a factor of 50 or 100? Looking at the repetitive sequences more closely, one may find a rather large class of repetition DNA, that is not exactly repetitious; There are, possibly dependent on the mutation rate, larger or smaller deviations; so this increases the informational repertoire. However this would bring up your figure considerably and so I really wonder what kind of reproducible information is there available for its estimation?

Bernardi: If you look at the tables giving the DNA content in haploid cells you can see that mammals have about one thousand times more DNA than bacteria, and in some amphibians and dipnoans you may find even ten thousands times more.

The amount of polypeptide-coding DNA in mammals is unlikely to be 1000 times higher than in bacteria. As I mentioned, 95% of mammalian enzymes are found in *E. coli*.

Question: But there still could be multiple coding for one enzyme!

Bernardi: This may be true, but don't forget that in the case of at least one structural protein, which is the silk fibroin, it has been shown now, that there are only one or two genes; which means that in spite of the enormous amount of fibroin, which is produced in the silkgland, everything goes about by production of more messenger.

Question: Certainly the basic metabolic enzymes obviously are comparable in *E. coli* and man. But is there any reliable estimate to what extent this is valid for the other proteins?

Bernardi: The only answer I can give, is that if you trust the estimates from renaturation kinetics, usually you have something like 50% of the DNA in higher organism, which is considered "unique sequence" DNA, and which should correspond to either structural or regulatory genes. Whether this is correct or not I don't know. However, if you trust the renaturation results you should conclude that out of 50% a small percentage, maybe 5%–10%, is for protein coding and the rest is regulatory DNA.

Question: What is the procedure you are using for getting your DNA free of the enzymes after the enzymatic treatment? Wouldn't the treatment interfere with the renaturation kinetics?

BERNARDI: The procedure involves enzyme denaturation by chloroformisoamyl alcohol. This treatment does not interfere with renaturation kinetics in itself, but don't forget that the oligonucleotides we obtain are too small to do such a work.

As far as renaturation kinetics goes, we have just begun to study it, on the fractions from the main band DNA. Those fractions, as I said, have a standard deviation of the *CsCl* band and of *Tm* which are comparable to those of bacterial DNAs of the same molecular weight. We have not done any renaturation work on total DNA.

Question: You have been showing a digestion study of calf thymus DNA with different DNases. If I am correct, there are differences between satellite and main band DNA?

BERNARDI: Sure

Question: Do you see any effect of 5-methyl-cytosines on the digestions?

BERNARDI: We have no way to decide this, because we don't analyse methylcytosine.

Question: I am asking since mouse satellite DNA is known to be methylated three times more than the main band DNA. Can these differences not just be caused by the different distribution of 5-methyl-cytosine in your splits?

BERNARDI: I would not say so, since this would change the amount of cytosine in the terminal only.

Question: You can't test them?

BERNARDI: No. But assuming you have a factor of three in cytosine content, that means that your cytosine level might be different according to, whether cytosine is methylated or not. The differences are much larger than this. Don't forget that I showed deviation plots, which give the changes compared to what you find with bacterial DNAs. Those changes are really very large. Though I cannot exclude that it is possible, that methylcytosine will cause a small difference in terminals; this effect, I would say, is not important enough to cause a significant deviation.

The pattern of mouse satellite is completely different from total mouse DNA. The conclusions from the mammalian DNAs investigations show: 1) that you have repetitive sequences; 2) that the repetitive sequences, which are seen, are not those of satellite DNAs, because the deviation patterns have nothing in common with the deviation patterns of satellite

DNAs. In fact, in order to contribute to a deviation pattern you need two things: 1) you need an oligonucleotide pattern, which is different; 2) you need a certain amount of these sequences, otherwise you don't see them. If you have an exceptional DNA, let's say a pure AT which is present at a level of 1 to 4%, you just miss it. This may seem at first sight a drawback of the method in sensitivity. In fact it is extremely useful since actually in the field of eukaryotic genome, we still are basically interested in the very fundamental problems. We don't want to be disturbed by minor things. Satellites are such things for us. They are a very interesting problem in themselves. However it is not because we are going to understand the five percent of satellite DNAs, that we will to understand the other 95%, which is the real interesting point.

Question: Did you try to find a difference between the DNA of a cell and of the same cell after incubation with a tumor virus or after the integration of a tumor virus genome?

Bernardi: Not only the answer is no, but we will never try it, because we cannot pick up something which is 5 or up to 10% of a very perculiar sequence, like a satellite DNA. How are we going to pick up something which is actually much closer in sequence to the mammalian genome, like a SV 40 genome, which is present at an incredibly low level after integration? It is completely hopeless. So we are going not to do this.

Question: With SV 40?

Bernardi: Anything which is so low in amount in the genome. It is out of reach. You first need to purify a fraction to raise the amount of that particular sequence, in which you are interested, up to a sizable level, I would say at least 20% to 30%, then you have some hope to see it. However, if the sequence is very much similar to that of the host, you are not going to see it.

Question: You will find only one or two genomes of SV 40 virus in one cellular genome. You could try it with the Epstein-Barr-Virus, where you may find as many as 50 or 60 genomes. However the case of the Epstein-Barr-Virus is still in discussion. Apparently it depends on certain circumstances, how many virus genomes per cell you may get. You could amplify the genome by picking it out by density gradient centrifugation and then trying to do your investigation.

One could take messenger RNA and a RNA dependent DNA polymerase to make the genome for this m-RNA. Thus you can get the substrate for your assay,

Bernardi: One question which we would like to answer is: 1) what is the pattern of oligonucleotide sequence in polypeptide coding genes in

eukaryotes; 2) do those sequences fit with what we find in bacteria, or do we have a change in pattern in the case of mammals or vertebrates? By referring exclusively to the polypeptide coding genes, you have several experimental ways: One is to make a faithful copy of some mRNA by a reverse transcriptase, than you can take that DNA and apply our method. I don't know whether any faithful and complete copy of any mRNAs is available, may be in the case of globin. Another possibility is to take a viral genome, this is what we are doing now with SV 40, polyoma etc. – You can go even further by breaking down the SV 40 genome by a restriction enzyme into bands. You can analyse band by band. Now you can check the possibility whether all the nucleotide sequence deviations come from a particular band, and not from the polypeptide coding genes. This is a very important question which we are trying to answer.

Question: I am wondering about the control values when you are degrading bacterial DNAs. For, as you are starting with a somewhat statistical pattern, the DNase should read the DNA and look for that certain sequence to be cleaved with a certain preference. So the control values for a statistical DNA should not have a random composition in the endgroups either 5′ or 3′.

Bernardi: Now we get into the technology. What does happen when you break down a DNA with a DNase? How do the termini vary, according to fragment size? You may suspect that the DNase will start with some sequences which are highly preferred sequences, for which the K_m values and the v_{max} values are particularly favorable. However, in the range we used, which is a range below 100 nucleotides, there is no change in termini, whatever the size of the fragment may be. You may go much higher by using labeling techniques for much larger fragments, you still get the same pattern. You can take an enzyme like pancreatic DNase to break down DNA to an average size of 100 or 500 or down to 8. You always get exactly the same termini. This raises a technical point. It is clear that one finds different K_m and v_{max} values for different sequences, which we have actually shown. One has to keep in mind, that one averages out everything by taking all the 3′ ends coming from all different sequences. This average takes care of all differences. That makes it a completely constant value. If you add only a small number of sequences, it would not be so. However, as long as the number is large you average out. By taking alternating (A–T), one knows that the two different tetranucleotides ATAT and TATA are present in equal amounts. It is one of the few cases, where one knows the sequence. If the alternating (A–T)

is split with spleen DNase you find in 3′ terminal position 80% A, 20% T, which is caused by differences in K_m and/or v_{max}. We have reasons to believe, that it is essentially v_{max}. As this does not change at all, the selection is through the specificity of the enzyme. This is a special case. However, when you take a DNA with hundreds of possible sequences, you average out everything. The average value does not change. This is something of essential importance for the analysis, because if one had to worry about all possible changes, the method would not be good.

Question: One difficulty in your system is that you have used pure DNA. In this pure DNA you have a large amount of genomes. Is it possible to coat this pure DNA by protein? Is it possible to extract from perhaps a stimulated hen's oviduct, then partially and specifically break it down by adding DNases. In analogues experiments by Cox it is possible to extract chromatin from stimulated hen's oviduct which could serve for RNA polymerase from E. coli as an active template producing a special RNA; so one could imagine, that it would be also possible to adapt this procedure to your method such as to pick out stretches from genomes.

Bernardi: Yes, you are absolutely right. Of course it is possible to use the enzymes not on DNA, but on chromatin or in any system, where you block some DNA sequences. Thus, in the case of chromatin, with DNases you can preferentially break down the unprotected parts, at which you may look first. On the other hand you can protect DNA stretches in other ways, e.g. with dyes. The method is quite general. I just gave the basic idea. The applications are very numerous and any piece of DNA may be analysed.

Question: I am very much interested in hearing that you can digest DNA in chromatin. Do you mean in really may be applied in the way just indicated or is there some other objection to these experiments?

Bernardi: As far as DNase degradation is concerned, we have worked two years on chromatin degradation some years ago ourselves. I think it can be done, though probably there may be some difficulties.

Question: The difficulty, as I understand, is not really the DNA which is functional?

Bernardi: No. The difficulty lies simply in the end point definition. If you take chromatin and add enough enzymes, you can degrade chromatin down exactly to the same level as pure DNA; so you have to look very carefully to the point where to stop. I am not sure, that there is a well defined titration end point.

Question: Coming back to the class of repetitive DNA and its behaviour during evolution, this repetitive DNA is not subjected to selective pressure. It can mutate freely and thus sooner or later it would not be distinguishable any more from singular DNA. I really wonder whether most of the DNA, which cannot be accounted for of usuable genes, is simply garbage and there is no biological function whatsoever for this part of DNA. It is just playing material for further evolution. Could you comment on this point?

BERNARDI: I think that there are few facts which are well established. Satellites of even relatively closely related species may be very different. It is also true, that spacer sequences of ribosomal genes of very closed species like Xenopus laevis and Xenopus Mülleri, are different. So it is likely that there are sequences which can evolve at the much faster rate than those for polypeptide coding genes.

Now the speculations that those sequences are raw material. I don't see any fact showing that a satellite has evolved to an intermediate fraction and that such an intermediate fraction is going to evolve into a polypeptide coding gene. I am sorry, but I don't see a strong enough basis for believing that.

Question: Forgetting for a while the satellite DNA, there are apparently classes of intermediately repetitive DNA, at least this was postulated a few years ago by BRITTEN and KOHNE. This certainly could evolve by mutation to a DNA which would be indistinguishable from singular DNA. On the other hand, there are theories, that a huge amount of regulatory DNA is necessary for the function. While other people do not believe that the amount of regulatory DNA must be so large. There are thus two opposite viewpoints: 50–90% of eukaryotic DNA is either regulatory and thus functional, or it is non-sense stuff.

BERNARDI: I agree with you that this number of really necessary genes is open to question. However there is no compelling evidence favouring either one of your alternatives. That is why we want to get informations by doing experiments. These theories are useful insofar as they lead to experiments. But I don't know whether anybody could really say, that saltatory replication is a fact, which happens in evolution, or what is the amount of regulatory DNA. I think it is important, to set up new ways to study the sequences and try to understand them directly.

In my opinion renaturation kinetics has gone almost all the way, what it could give, has been given. Perhaps there will be a few minor improvements, but, always there is an end to the possibilities and potentialities of a given technique. So you have really to look for new

ways. The one we have found, is a new one; but probably more can be found. Aside of this you may go on speculating. I don't know whether it is so useful, since the basic possibilities are so many, so that you need facts.

Question: It is well known that in some cell types for example from Xenopus laevis and Triturus there are 10 times different amounts of DNA for the same chromatid, although they have the same karyotype. This means that there is 10 times more DNA for the same morphological structure. Recently it has been shown that this is not against theories on chromamer eorganization of DNA in chromatids and I just wonder whether in these two extreme cases there would be a difference in arrangement.

Bernardi: There is even a more striking case in Chironomus, where one finds 27% more DNA in one species which can breed with the other one. It has been shown that the excess of DNA is localized in a small number of bands. The excess of DNA in those bands is in a ratio of 2, 4, 8, 16 to the one in the other species. These existing problems are very well defined, but we have no explanation for it.

Question: Nevertheless eventually it can explain, what actually repeated sequences are there for. In case they are for folding and supercoiling of chromatin, than this could be accepted as an explanation.

Bernardi: This is really unlikely, since the chromosomes in those two Chironomus species are absolutely identical, and one can localize exactly the chromomeres which have increased amounts on DNA.

However this is not much help in solving the basic problem: why even in the eukaryotic species which have the least DNA increase, we still are so much off, from what one would expect from the calculation of the polypeptide coding genes. This is the issue to be solved first.

Question: I would like to apply your technique to questions of human pathology. Would you expect the sensitivity of the method to be high enough as to use it in cases with chromosomal abnormalities, e.g. in diseases like mongolism etc.

Bernardi: I don't think so. I am rather skeptical. The only faint hope is, that the method is still improvable in terms of looking at longer sequences. A completely different approach to the use of DNase would be to look at the terminal digest, which I did not dicuss at all. As it stands now the method is only good for very gross problems.

Comment: You have been understating be confining your method to gross problems. You presented amazing results, e.g. in your last table showing, that all the mammals have the same sequence deviations.

Comment: This could be seen differently, as this points to a low discrimination of the method.

References

(1) Boivin, A., Vendrely, R. and Vendrely, C.: C. R. Acad. Sci. **226**, 1061 (1948).
(2) Mirsky, A. E. and Ris, H.: J. Gen. Physiol. **34**, 451 (1951).
(3) Sober, H. A. (ed.): Handbook of Biochemistry: Selected Data for Molecular Biology. Second edition, 1972. The Chemical Rubber Company, Cleveland, Ohio.
(4) Holm-Hansen, O.: Science **163**, 87 (1969).
(5) Britten, R. J. and Davidson, E. H.: Quart. Rev. Biol. **46**, 111 (1971).
(6) Hinegardner, R.: Quoted in ref. 5.
(7) Fairbairn: Quoted in ref. 5.
(8) Nigon, V. and Bouvet, P.: Compt. Rend. Soc. Biol. **149**, 129 (1955).
(9) Sexsmith, E.: Thesis – University of Toronto 1968.
(10) Biers, K., Kunz, W. and Ribbert, D.: In: Chromosomes Today (Darlington C. D. and Lewis, K. R., eds.) vol. 2, p. 107, Plenum Press, N. Y., 1969.
(11) Ribbert, D. and Bier, K.: Chromosoma **27**, 178 (1969).
(12) Dawid, I. B. and Brown, D. D.: Develop. Biol. **22**, 1 (1970).
(13) Rothfels, K., Sexmith, E., Heimburger, M. and Krause, M. O.: Chromosoma **20**, 54 (1966).
(14) Keyl, H.-G.: Chromosoma **17**, 139 (1965).
(15) Britten, R. J. and Davidson, E. H.: Science **165**, 349 (1969).
(16) Leuchtenberger, C., Leuchtenberger, R., Vendrely, C. and Vendrely, R.: Exp. Cell. Res. **3**, 240 (1952).
(17) Ris, H., Mirsky, A. E.: J. Gen. Physiol. **33**, 125 (1949).
(18) Hinegardner, R.: Amer. Natur. **102**, 517 (1968).
(19) Atkin, N. B. and Ohno, S.: Chromosoma **23**, 10 (1967).
(20) Vendrely, R.: Compt. Rend. Soc. Biol. **143**, 1386 (1949).
(21) Gelderman, A.: unpublished data quoted in ref. 15.
(22) Ogur, M., Minckler, S., Lindegren, G. and Lindegren, C. G.: Arch. Biochem. Biophys. **40**, 175 (1952).
(23) Cairns, J.: J. Mol. Biol. **4**, 407 (1962).
(24) Bode, H. and Morowitz, H. J.: J. Mol. Biol. **23**, 191 (1967).
(25) Ben-Porat, T. and Kaplan, A. S.: Virology **16**, 261 (1962).
(26) Callan, H. G. and Lloyd, L.: Phil. Trans. Roy. Soc. (London), ser. B, biol. Sci. **243**, 135 (1960).
(27) Callan, H. G.: J. Cell. Sci. **2**, 1 (1967).
(28) Thomas, C. A., Jr.: The Neurosciences: Second Study Program (F. O. Schmitt, ed.), The Rockefeller University Press, New York, N. Y., 1970.
(29) Georgiev, G. P.: J. Theor. Biol. **25**, 473 (1969).

(30) Beermann, W.: in Developmental Studies on Giant Chromosomes (W. Beermann, ed.) p. 1, Springer Verlag, Berlin, 1972.
(31) Crick, F. H. C.: Nature (Lond.) **234**, 25 (1971).
(32) Paul, J.: Nature (Lond.) **238**, 444 (1972).
(33) Judd, B. H., Shen, M. W. and Kaufmann, T. C.: Genetics **71**, 139 (1972).
(34) Sorsa, V., Green, M. M. and Beermann, W.: Nature New Biol. **245**, 34 (1973).
(35) Daneholt, B.: Nature New Biol. **240**, 229 (1972).
(36) Britten, R. J. and Kohne, D. E.: Science **161**, 529 (1968).
(37) Wetmur, J. G. and Davidson, N.: J. Mol. Biol. **31**, 349 (1968).
(38) Britten, R. J. and Smith, J.: Carnegie Inst., Year Book **68**, 378 (1970).
(39) Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J.: J. Mol. Biol. **77**, 1 (1973).
(40) Thomas, C. A., Jr., Hamkalo, B. A., Misra, D. N. and Lee, C. S.: J. Mol. Biol. **51**, 621 (1970).
(41) Lee, C. S. and Thomas, C. A., Jr.: J. Mol. Biol. **77**, 25 (1973).
(42) Thomas, C. A., Jr. and Dancis, B. M.: J. Mol. Biol. **77**, 43 (1973).
(43) Pyeritz, R. E. and Thomas, C. A., Jr.: J. Mol. Biol. **77**, 57 (1973).
(44) Bick, M. D., Huang, H. L. and Thomas, C. A., Jr.: J. Mol. Biol. **77**, 75 (1973).
(45) Thomas, C. A., Jr., Zimm, B. H. and Dancis, B. M.: J. Mol. Biol. **77**, 85 (1973).
(46) Schildkraut, C. L., Marmur, J. and Doty, P.: J. Mol. Biol. **4**, 430 (1962).
(47) Marmur, J. and Doty, P.: J. Mol. Biol. **5**, 109 (1962).
(48) Samejima, T. and Yang, J. T.: J. Biol. Chem. **240**, 2094 (1965).
(49) Wells, R. D. and Blair, J. E.: J. Mol. Biol. **27**, 273 (1967).
(50) Corneo, G., Ginelli, E., Soave, C. and Bernardi, G.: Biochemistry **7**, 4373 (1968).
(51) Bernardi, G., Carnevali, F., Nicolaieff, A., Piperno, G. and Tecce, G.: J. Mol. Biol. **37**, 493 (1968).
(52) –, Faures, M., Piperno, G. and Slonimski, P.: J. Mol. Biol. **48**, 23 (1970).
(53) – and Timasheff, S. N.: J. Mol. Biol. **48**, 43 (1970).
(54) Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E. and Cantor, C. R.: J. Mol. Biol. **54**, 465 (1970).
(55) Allen, F. S., Gray, D. M., Roberts, G. D. and Tinoco, I., Jr.: Biopolymers **11**, 853 (1972).
(56) Brown, D. D.: Sci. Am. **229**, 21 (1973).
(57) Wallace, H. and Birnstiel, M.: Biochim. Biophys. Acta **114**, 296 (1966).
(58) Walker, P. M. B.: in Progress in Biophysics and Molecular Biology (Butler, J. A. V. and Noble, D., eds.) vol. 23, p. 145, Pergamon Press, London, 1971.
(59) Filipski, J., Thiery, J. P. and Bernardi, G.: J. Mol. Biol. **80**, 177 (1973).
(60) Wu, R. and Taylor, E.: J. Mol. Biol. **57**, 491 (1971).
(61) Takanami, M., Okamoto, T. and Sigiura, M.: Cold Spring Harbor Symp. quant, Biol. **35**, 179 (1970).
(62) Ziff, E. B., Sedat, J. W. and Galibert, F.: Nature New Biology **241**, 34 (1973).
(63) Fry, K., Poon, R., Whitcome, P., Idriss, J., Salser, W., Mazrimas, J. and Hatch, F.: Proc. Nat. Acad. Sci. USA **70**, 2642 (1973).
(64) Gall, J.: Personal communication.

(65) Chargaff E.: in the Nucleic Acids (Chargaff, E. and Davidson, N., eds.) vol. 1, p. 307, Academic Press (1955).
(66) Southern, E. M.: Nature **227**, 794 (1970).
(67) Josse, J., Kaiser, A. D. and Kornberg, A.: J. Biol. Chem. **236**, 864 (1961).
(68) Baldwin, R. L. and Kaiser, A. D.: J. Mol. Biol. **4**, 418 (1962).
(69) Swartz, M. N., Trautner, T. A. and Kornberg, A.: J. Biol. Chem. **237**, 1961 (1962).
(70) Subak-Sharpe, H., Bürk, R. R., Crawford, L. V., Morrison, J. M., Hay, J. and Keir, H. M.: Cold Spring Harbor Symp. Quant. Biol. **31**, 737 (1966).
(71) Bernardi, G., Ehrlich, S. D. and Thiery, J. P.: Methods in Enzymol. (L. Grossman and K. Moldave, eds.), in press (1973).
(72) Carrara, M. and Bernardi, G.: Biochemistry **7**, 1121 (1968).
(73) Chersi, A., Bernardi, A. and Bernardi, G.: Biochim. Biophys. Acta **246**, 51 (1971).
(74) Bernardi, A. and Bernardi, G.: Biochim. Biophys. Acta **155**, 360 (1968).
(75) Ehrlich, S. D., Torti, G. and Bernardi, G.: Biochemistry **10**, 2000 (1971).
(76) Devillers-Thiery, A., Ehrlich, S. D. and Bernardi, G.: Eur. J. Biochem. **38**, 416 (1973).
(77) Carrara, M. and Bernardi, G.: Biochim. Biophys. Acta **155**, 1 (1968).
(78) Piperno, G. and Bernardi, G.: Biochim. Biophys. Acta **238**, 388 (1971).
(79) Ehrlich, S. D., Thiery, J. P. and Bernardi, G.: Biochim. Biophys. Acta **246**, 161 (1971).
(80) Thiery, J. P., Ehrlich, S. D., Devillers-Thiery, A. and Bernardi, G.: Eur. J. Biochem. **38**, 434 (1973).
(81) Bertazzoni, U., Ehrlich, S. D. and Bernardi, G.: Biochim. Biophys. Acta **312**, 192 (1973).
(82) Yoneda, M. and Bollum, G. J.: J. Biol. Chem. **240**, 3385 (1961).
(83) Kössel, H. and Roychoudhury, R.: Eur. J. Biochem. **22**, 271 (1971).
(84) Bernardi, A.: Anal. Biochem., submitted for publication, 1974.
(85) Richardson, C. C.: Proc. Nat. Acad. Sci. U.S. **54**, 158 (1965).
(86) Lehman, I. R.: J. Biol. Chem. **235**, 1479 (1960).
(87) Bernardi, G. and Griffé, M.: Biochemistry **3**, 1419 (1964).
(88) Laval, J. and Paoletti, C.: Biochemistry **11**, 3596 (1972).
(89) Kunitz, M.: J. Gen. Physiol. **33**, 349 (1950).
(90) Lehman, I. R., Roussos, G. C. and Pratt, E. A.: J. Biol. Chem. **237**, 819 (1962).
(91) Ehrlich, S. D., Devillers-Thiery, A. and Bernardi, G.: Eur. J. Biochem. in press (1973).
(92) Bernardi, G., Apella, E. and Zito, R.: Biochemistry **4**, 1725 (1965).
(93) Lindberg, U.: Biochemistry **6**, 335 (1967).
(94) Cordonnier, C. and Bernardi, G.: Biochem. Biophys. Res. Comm. **20**, 555 (1965).
(95) Soave, C., Thiery, J. P., Ehrlich, S. D. and Bernardi, G.: Eur. J. Biochem. **38**, 423 (1973).
(96) Laval, J., Thiery, J. P., Ehrlich, S. D., Paoletti, C. and Bernardi, G.: Eur. J. Biochem., in press (1973).

(97) Ehrlich, S. D., Bertazzoni, U. and Bernardi, G.: Eur. J. Biochem., in press (1973).
(98) Ehrlich, S. D., Bertazzoni, U. and Bernardi, G.: Eur. J. Biochem., in press (1973).
(99) Bernardi, G.: Biochem. Biophys. Res. Comm. **17**, 573 (1964).
(100) Jacquemin-Sablon, A., Laval, J., Le Talaer, J.-Y., Le Pecq, J.-B., and Paoletti, C.: Compt. Rend. Acad. Sci. **259**, 2551 (1964).
(101) Bernardi, G., Ehrlich, S. D. and Thiery, J. P.: Nature New Biology **246**, 36 (1973).
(102) Ehrlich, S. D., Thiery, J. P., Devillers-Thiery, A. and Bernardi, G.: Nucleic Acids Research, in press, 1974.