

(1)

10.27.86.

[SLIDE!] Short segment of a DNA molecule. ~20 bp

This is, of course, the fundamental structure from which and around which contemporary biologists think about living things.

The totality of the DNA in a cell, be it a ~~single cell free-living, organism composed~~ single independent cell that is itself an organism (bacteria, yeast) or one of billions of cells in a complex multi-cellular organism (man, corn plant) is called the genome.

Mr. Mrs. Fellow members: - gratitude.

The students of the late 1980's take the DNA molecule and the major biological principles that flow from it for granted.

But, ^{it was only} in the late 1940's and early 1950's, when I was an undergraduate and still thought that I wanted to be a chemist, that the first of these principles was established.

Avery, then Hershey and Chase, established that genes, genomes, the informational systems that encode the competencies of living things resides in DNA.

I was only vaguely aware of ~~these~~¹⁵ notions as I took the first step toward biology - entering graduate studies in biochemistry, ~~not~~ all

~~biology.~~

While I was a graduate student, busy with enzymology and phosphoproteins, the second gigantic step was taken. Watson and Crick proposed the double-helical, base paired structure for DNA. ~~had~~ ~~had~~ ~~been~~ ~~now~~, There was ~~was~~ no way to avoid being challenged by the intellectual and experimental attractions of this macromolecule and its close relative, RNA.

Upon completion of the Ph.D. degree I entered the world of nucleic acids by becoming a post-doctoral fellow with Leon Heppel at the NIH. The year was 1956.

It was a small world, a few laboratories scattered about the ~~globe~~^{globe}. For me, it was the start of 30 years of unflagging excitement and ~~the start of association with~~^{the start of association with} ~~it introduced~~^{have} me to a group of people that ~~were~~^{have} to become my most cherished colleagues, and friends. ~~for these same 3 decades~~: These days when, in a year's time, 1000's of people attend 10's of meetings on nucleic acids each year, it is amusing to remember that in the late 50's, ~~the nucleic acids group was given one~~, we had one, morning, greedily donated by those who ran the week long Gordon Conference on proteins.

The ~~key~~^{central} contribution of the mid-fifties - ^{to last}
~~the recognition~~^{the recognition} was ~~the demonstration~~^{the demonstration} that enzymes were the key to dealing with nucleic acid structure ~~to~~
~~this day, purely chemical methods do not create~~^a although we were all unsuspecting of just how versatile, and in reaction and specificity such enzymes would prove to be. ~~as~~

My own work in those post-doctoral years and in the first decade of independent scientific work was to study the mechanism of the synthesis of polyribonucleotides (RNA if you will) by PNase, the enzyme discovered in the mid-fifties by S. Ochoa + M. Grunberg-Manago.^{first polymerase} By 1960, I had learned how to make a variety of polyribonucleotides, for that reason, ~~the~~^{3rd} when the 3rd major principle of modern biology was established I could be a participant, not a distant admirer.

(4)

In est efforts to establish the nature of the genetic code.
3 base pairs specifies 1 amino acid.
including 1 codon "start". 3 codons "end".

genetic code is
order of bases, char.
order of bases
nature of regular
code
is established.

I joined with [redacted] to assist our colleague Marshall Nirenberg along with several other NIH scientists to put aside my own experiments & in order to prepare a whole array polyribon-RNAs of different structures. My own contribution was to prepare RNAs of different structures. We thought it was hi-tech work. From our current vantage point it seems almost trivial, embarrassing even to mention.

My interest in the mechanism of polymerization and depolymerization continued through the 60's. Along with the first post-doctoral fellow to join my group, I demonstrated that enzymatic synthesis of long polymer can occur by what we called a processive mechanism. Processivity is now known to be quite common. Consider the following. First, an enzyme catalyzing the joining of the first 2 monomer units. There is then a choice. Drop this dimer and then either synthesize another, or add a third unit to the original. After each addition, the enzyme continues and continues to release the product after each addition and then pick randomly between extension of existing chains or starting a new one. In this model, all the new chains grow together. The second, the processive model, the enzyme hangs on to the new chain once it is started and adds to it new monomer units. When the chain grows long enough, the enzyme drops it and initiates synthesis of another chain. Here, there are a few very long molecules even very early in the reaction.

* Completely symmetrical arguments can be made for the degradation of ~~the~~ chains by processive or random mechanisms.

The mechanism work on polyribonucleotides polynucleotide phosphorylase and κ bacterial ribonuclease was ~~rewarding~~, but ~~as~~ my frustration grew as approach after approach failed to define the ~~physical~~ functional role of these enzymes in bacteria. It was time to shift gears. This change was initiated with a sabbatical year at the Weizmann Institute - where I spent the time learning about mammalian cells and viruses whose genomes are DNA. Returning to the NIH, ~~as~~ I spent the next few years trying out various ideas and systems. ~~Finally all these~~ ~~These~~ efforts could not escape. Soon it became obvious During this time the powerful methods we now refer to as 'recombinant DNA techniques' ~~and/or~~ Mention: complete dependence on certain reagents 'genetic engineering' emerged. And as they increasingly became ~~a~~ their extraordinary ~~power~~ capability ~~to~~ permit ^{to} bicker fruit with previously intractable problems became increasingly clear, it proved ^{their} impossible if not ~~downright~~ foolish to resist being ~~seduced by them~~. Especially because they seemed applicable to a truly peculiar puzzle, a ^{important} puzzle ~~problem~~ enough to be given a name. The C-value paradox.

(6)

The paradox is simple to describe.

Average size of a gene \sim 1200 bp (400 codons).

SLIDE: E. coli (bacteria), \sim 4000 genes $\rightarrow \sim 6 \times 10^6$ bp.

Complex multi-cellular (fly, mammal), $\leq 10^5$ genes $\rightarrow 1.2 \times 10^8$ bp

What is all the extra DNA?

1. Introns.

2. $\sim 10^{20}$ yrs ago, first hint that ~~on~~ a large portion of 'extra' DNA is not informational in the classical genetic sense.

Repeated DNA sequences. Segments 2 bp $\rightarrow 10^3$ where more or less the same order of bases is repeated. Rec. DNA techniques have permitted us to begin to analyze.

First step: classification. High, middle, occasional
interested in highly-repeated $> 10^4$. Cenozoic era (implies)

SLIDE: Tandem & Interspersed

SLIDE: Types of repeats.

SLIDE: Amounts of repeats

Our most recent efforts - with LINEs. LINE-1.

SLIDE: Summary of LINE-1 in all mammals. Define

SLIDE: STRUCTURE OF THE LONGEST LINEs. Pseudogenes. Impure genes.

SLIDE: MAJOR QUESTIONS.

SLIDE: MECHANISM OF PSEUDOGENE FORMATION. RNA \rightarrow DNA

Various problems: ① general - reverse transcriptase (viral).

② RNA, where from? functional requirement

- progress: putative RNA. cytA^t

progress on reverse transcriptase - unexpected quirk

SLIDE: STRUCTURE OF LONGEST LINE. / encodes reverse transcriptase?