MIT Apnl 1972

Chromosomes

Pair of lectures

lectures technical – assume background in mol. biology – subject v. confused + major problems unsolved.

Higher organisms ~ mainly higher eucaryotes.
(but forgin dinoflagellates and esp fungi v. interesting)

Major differences from procaryotes

① large (sometimes v. large) amounts of DNA
to be made more precise later.

② Histones : basic proteins. usually 1:1 mass ratio
5 main types (but modifications)

③ HnRNA – large, often v. large MW
– rapid turnover, or least in part
– (mainly) confined to the nucleus

[also quite a bit of small MW RNA, of unknown function]

HnRNA claimed by Georgiev to be complexed with a virus-like "structural" protein – informofers.

Nature of higher organisms

– many different major cell types eg. muscle,
nerve cells – many types
fibroblasts
(not micro-effects such as positional information)

– complex cell-cell interactions to build organisms.

All this would suggest not only more types of proteins (enzymes, muscle + collagen, cell surface) but also more complex control mechanisms (Britten + Davidson)

Thus there appear to be major switches

eg myoblast → → fused muscle cells

"determination" – can be hormone triggered in certain cases.

but is this "irreversible"? ? may depend on how hard one tries.

but nucleus appears to be reversible – Gurdon's experiment

not so easy to switch a whole cell [though can be switched on

ie from an undifferentiated to a differentiated state]

[but switch of a cell → nerve or muscle is rare? unknown?]

Amount and Nature of DNA

Drosophila 2×10^8 b.p.

Drosophila 30-40 times E. coli.

Man perhaps 1000 times E. coli.

but amphibia (eg newts) and certain plants (eg lilies) very much higher.

also closely related species can differ v. much in amount of DNA.

my main concern is with Drosophila because

① not excessive DNA

② genetics v. good

③ cytogenetics eg giant polytene chromosomes.
eg. in salivary glands.

Nature of DNA

annealing tests (ie melting double helix + seeing how fast it comes together again) - Cot curve Roy Britten.

Cot for procaryote follows simple kinetics and allows estimate of (kinetic) complexity. &

perhaps of annealing — remelting temperature.

① Simple sequence DNA (≙ satellite DNA)

eg poly AT in crabs

Can be appreciable amounts eg 10 % (up to 40 %)

Sequence not exactly simple repetition but probably based on, or derived from, a simple repetition. — Grunz Pos.

Can vary greatly from species to species, even if fairly closely related.

mostly

prob. located not near the chromomere (heterochromatin)

usually not transcribed (but collagen, silk, simple coiled coils)

② Intermediate DNA

not very simple sequences; but many copies; probably not all identical. Amount varies in diff. species — prob not a lot in Drosophila.

"Unique" DNA 60–70% ? in Drosophila.

~~Of~~ sequences appear to be all different.

how unique is unique?

start off all the same : put in mutations at random.

after a time becomes "unique" how many mutations needed?
(may depend on type of mutation)

Thus "unique" doesn't imply completely unrelated.

Laird's results : much of Drosophila DNA unique

and kinetic complexity gives correct answer

(Charlie T's reservations?)

~~Bout~~ But are there really reiterations?

Charlie Thomas' work

impossible to summarise all this work

example : make double-stranded : nibble away with
an exonuclease (draw) . anneal.

forms high prop. of circles.

even for Drosophila , even for polytene Drosophila

His general conclusion.

must be some degree of tandem repetition.

typical repetition length ~ ~~3,000~~ 1500 bases.

typical region length ~ perhaps 20 times this

5

~~but ① exact~~ but depends on certain assumptions

eg. breakage "random" – now some evidence for this.

does say exactly how exact the repeat is.

doesn't ~~the~~ think it can be "small" exact repeats
interspersed with unrelated sequences.

but difficult to deal with ~~a~~ more complex models

eg. mixed models.

but basically with have the Laird – Thomas dilemma.

"no" repetition v. much ~~a~~ repetition.

Nearest Neighbour analysis Subek – Sharpe.

(mainly vertibrates) ~~difference~~ essentially total DNA, but has now be

done on various fractions.

big ~~to~~ difference of CpG : also of ApT

he explains by saying ① ~~the~~ of the arginine codes tit used very little

② chain termination ~~by~~ triplets rare.

from this + amino acid composition can deduce rather well

the double frequencies. Therefore concludes that most of the

DNA is potentially coding for proteins.

Counter argument : CpG forbidden : case of 5 me C.

may have been derived from coding sequences, but not now ~~used~~

necessarily now used for them.

Giant Polytene Chromosomes

Shall assume

① ~~...~~ mononeme hypothesis (for single chromosome)

② "resemble" interphase chromosomes – the ~~important~~

ones we really want to know about
(mitotic chromosomes mainly inactive)

main features : band and interbands.

most (95% of DNA ?) in bands and histone ~ 5,000 bands counted in Drosophila.

interbands – exact ~~amount~~ density not known

– ambiguity about length of bands.

But could not this be some convenient way of just packaging ?

~~Ben~~

Genetic + cytological results.

Results of Judd (Shen + Kaufman) – in press.

Segment of the X chromosome of Drosophila

3A2 to 3C2 between zeste and white locus.

~~pic~~ mutants picked up against a deletion [trick to allow males with the deletion to survive]
ie could pick up (recessive) lethals.
various mutagens. X-rays, EMS + nitrosoguanidine.

pickup about. ~ 120 (single-site mutants)

complementation groups (= unit of function)
got 12 between the two markers.
~~...~~ 12 no. of bands. Some ~~...~~ doubts – used mainly Berendes ~~...~~ map.

PP/CRI/E/1/2015

major result

for 1 : 1 correlation between "bands"

and complementation groups.

importance of result

although ① may not be completely exact

② may not apply to all genes

nevertheless a major result – cannot be laughed off.

other workers. Hochman on 4th chromosome.

36, 37 vital loci + 5 where visible mutants occur

≃ 50 bands.

also other work on proximal portion of X-chromosome.

again gives a reasonable correlation.

Lifschytz + Falk + Schalet, Lefevre + Singer

We can now state the dilemma of the amount of DNA more sharply

average band Drosophila has ≃ 30,000 n.p.

average size to code for a protein ~ 1000 pp.

Thus major question is, what is all this DNA for.

HnRNA

in higher eukaryotes, large size.

implausible molecule

at least partly turning (over rapidly) and doesn't get into cytoplasm..

discourse efficiency argument

Some part of it is mRNA precursor.

plausible model ppp — mRNA — poly A

Georgiev

claim for repeated sequences in this region.

puffing : true puffing + limited puffing.

Grossbach's analysis of salivary gland secretory proteins.

Shows strong correlation between a puff and a polypeptide.

~~Q~~ ~~[illegible]~~

must now consider function.

① puffing: true puffing and limited puffing.

Grossbach's analysis of salivary gland secretory proteins

~~[illegible]~~ shows strong correlation between a puff and a polypeptide.

② HnRNA — not only nuclear RNA - small molecule, some homogeneous, 10^6 copies.

a most unlikely class of molecules. because

"wateful" - efficiency.

① large size

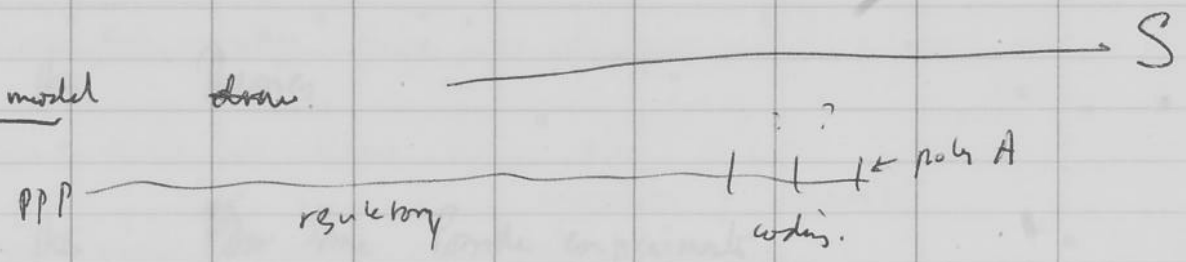
② rapid turnover.

accepted that mRNA occurs among these sequences.

Georgiev's model but ~~most~~ most HnRNA doesn't get into cytoplasm

(10% is mRNA?) turn-over - but may be complex

Georgiev's model ~~[illegible]~~



evidence suggestive, but falls short of proof.

even so, why is regulatory so big? — Britten & Davidson - but still v. large.

why is it repetitious?

Population Geeticist results

Crow, Nimura et

mutation rate for p fibrinopeptides

~~Questions on statements of facts~~

conclusion either ① fewer genes than might be expected

or ② only a few % of bases matter.

(2) Types of theory

~~Master-slave model. says replicas are exact~~

PP/CRI/E/1/20/5

Second Lecture

MIT April 72
~~Second lecture.~~

General remarks

Omit DNA replication and recombination.
and processes in cytoplasm (and transport there)

Sketch of previous lecture

1 "band" = 1 gene.

band too big.

compatible with HnRNA.

latter has mRNA at r.h. end.

Plan of this lecture.

more evidence (a) chromosomal protein
(b) repititive "genes"

then Theories

then Some Possible experiments.

~~[illegible]~~ (2)

~~[illegible]~~ April 72

What is associated with this DNA?

Chromosomal proteins

histones F1, F2A1, F2A2, F2B, F3.

21,000				11,200
(200)	102	140	125	101

indep. of base sequence

depn on base sequence.

Thus one histone, on average, every 25 base-pairs.

roughly equal amounts by mass, not by number.

basic-groups ≃ phosphate groups. ~~[illegible]~~ but evidence suggests that less not all ($\frac{1}{2}$?) DNA phosphates neutralised. — double nature of histones molecules.

Felsenfeld, Iszaki, Mirsky

non-histone protein

problem of artefacts

perhaps, by mass, half of much as histones

but as MW higher ① considerably fewer by number.

② some there in large amounts.

Thus the ones easily seen may only be about 50 (to 100) types say, or less.

but may be many more, rather important ones, there in smaller amounts.

Structure of chromatin

e/m data very confusing: almost impossible to summarize difficulty of artefacts.

X-ray data: "chromatin" (fairly concentrated) does give a "set" of X-ray spacings: 110Å, or submultiples. — rings

110
55
37
27
22

Wilkins + Pardon helix — explain helical diffraction.

but difficult to get orientation.

Thus helix improved: should consider other structures, ~~eg sheets~~

~~[illegible]~~

X-ray data can be got with histones alone (not F1? others?). ~~[illegible]~~

~~[illegible]~~

PP/CRI/E/1/2015

(3)

Repititive genes

eg. rRNA in Xenopus

Spacer sequences.

The two species X. leavis + mulleri.

suggest tandem replication in evolution.

also evidence from Chironomus ~~for~~ (Kyle)

whole whole band replication. in evolution.

need for many copies ~~in~~ when no amplification

case of genes for histones (very rapid syn. needed)

but haemoglobin case (esp. β chain in man) – much Hb needed.

~~[illegible]~~ could not easily detect many minor species.

but mutants suggest one coding sequence

[but Master Slave]

Are there other amplification ~~[illegible]~~ process (apart from nucleolar amplification in amphibia) ?

RNA – RNA RNA – DNA – DNA DNA → RNA

BudR Gene Bell

no evidence for operons

ie what little evidence there is suggests that

genes which might be in an operon are separate

(exception for one Hb case).

(4)

Theories

bands { (1) "Junk" – unlikely in strict sense.

(2) Evolutionary reserve in bands.

for unreasonable

two types of novelty – novelty – from existing coding sequences

– true novelty – from mutated simple sequences or other-phase – reading.

but not obvious that has a short-term evolutionary advantage

Some compatible with other theories.

(3) from or interbands

(3) could be for easy translocation to interbands = "junk"

[recombination Lefevre evidence, suggests prop. to DNA]

(4) John Paules Theory (in press?)

interbands = control elements

bands = multiple, drifting, copies

one or more used as mRNA.

PP/CRVE/1/2015

(5)

(5) My theory ~~for~~ Nature 234, 25 (1971)

(a) Interbands are coding sequence ——— S

av. length of ~~int~~ DNA in interbands 10^3 – two estimates.

[unpublished evidence Green + Beermann
white locus – ~~two~~
interband + small amount of band]

– ~~de~~ argument difficult – only one case.

(b) Nature of recognition sites

basically one of two theories of
Britten + Davidson (Fig 1A) ~~Science~~

ie multiple control sites.

Unpairing postulate ——— S

meaning of figure.

argument about cavities.

recognition site in procaryotes – Gierer ——— S

~~for~~ ~~to~~ Abundant versatility

Thus implies unwinding

(c) histones build a structure or structure
which unwinds
example ——— S

possible sheet structures (polyethylene – Leonard Lehman)

but little like it in e/m pictures.

(6)

Thus most of DNA used to build a framework
(framework for unwinding)

Explains why exact base sequence doesn't matter.

but ~~[illegible]~~ origin by random replication + evol. drift.

Explains why some degree of
random repetition.

Control regions

how to localize → [heterochromatin = unfixed folding]

Thus prob. need an unwinding sequence to
nucleate
+ an unwinding acidic protein.

~~An~~ multiple operon sites base sequences not
easy to change

~~Cont~~
∴ could have a degree of exact repeat

multiple promoter sequence

prob. an ~~ex~~ exact repeat
(diff. from Georgiev model)

Logic of control mechanism — topic omitted — +ve, negative
— and, or.

complications due to rapid synthesis
altering band structure.

Nature of recognition molecules — RNA convenient
— protein (Orgel's rule)

PP/CRI/E/1/20/5

(7)

Present Evidence

All very weak : eg. 8 interband = coding — ~~Green & Beermann~~ White Wem. (interband + big & small)

unpaired sequences an easy explanation of

the pairing up of giant chromosomes.

Frenster some evidence for S.S. regions, but should be repeated.

Future experiments.

The Genetic co-linearity expt

clustering of mutants on map

but not, for example, on rosy (xanthine dehydrogenase)

Where are "amino acid" mutants? — lack of data in Drosophila.

location in gene

direction on map. → for one gene

— for adjacent genes.

Not easy to get suitable gene

and suitable protein

but now necessary to combine fine & structure

genetics with exact protein chemistry.

Have control mutants any special characteristic?

(8)

The Peter Walker expt i.e. ribosome initiation complex (Stents) (Bretscher). (Floyd Robertson).

(a) on DNA. expt.
diff. in interpretation.

(b) on HnRNA.

The anti-messenger RNA expt

(1) to get location of mRNA sequence

(2) to purify a HnRNA
and look at the ~~m~~ l.h. part.

repeat promoter sequences. ?
"hairpins" ?

looks like one of the best ways into the chromosome.

Chromatin Structure

X-ray studies (e/m studies)

what do histones do - covering ?
distorting ?
linking ?

model expts on
simple DNA.
phage DNA.

looking for single-strand regions

do are they rich in 5mc C. ? etc.

PP/CRI/E/1/2015

(9)

General conclusion

- Subject in a confused state
- many lines of attack, only a few with results we can even partly accept.

  but techniques improving rapidly

- necessity to reconcile different approaches

  ∴ theories useful

  but no compelling theoretical arguments yet

Also may be missing elements (as mRNA was missing)

eg. small MW nuclear RNA.

nature of chromosomal proteins

- informational DNA of Dr. Gene Bell

Thus need for caution.

But problem essential to solve in, at least in outline, before we get to the major problem — how are the major switches in development made, how do hormones act, etc.

Fascinating subject but very difficult to lecture on.