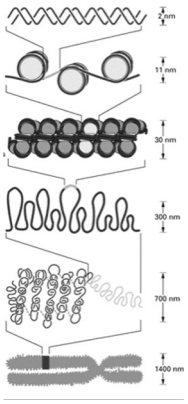
Eukaryotic Genomes & Chromatin Structure

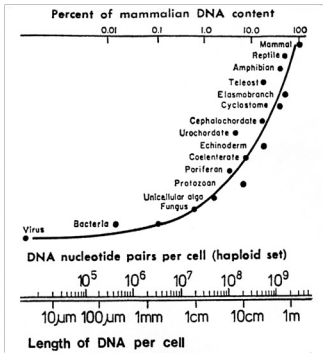
GEP 2022

Slides assembled with notes by SCR Elgin, WUSTL



1

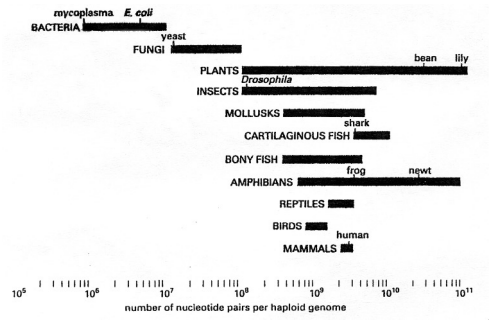
Minimum haploid DNA content - the C value paradox



R Britten and E Davidson, Science 165: 349-57, 1969

2

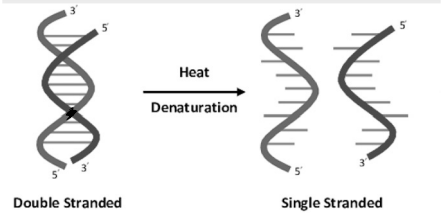
Lack of relationship between amount of DNA and organism complexity



Alberts et. al., Molec Biol Cell, 3rd ed

3

DNA denaturation/ dissociation/ melting



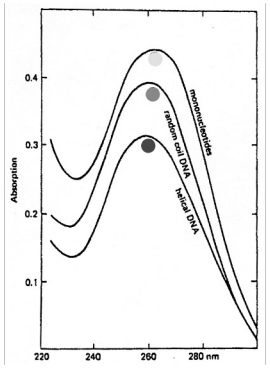
To follow the reaction and determine the melting temperature, Tm:

- Shear DNA to ~400 bp fragments
- Observe solution of DNA in spectrophotometer at 260 nm
- Use water-jacketed cuvette, raise temperature slowly

Figure taken from the Web – may be copyright.

4

Follow melting of DNA by ultraviolet absorption spectra



- Mononucleotides
- Denatured (ss) random coil DNA
- Native double helical DNA

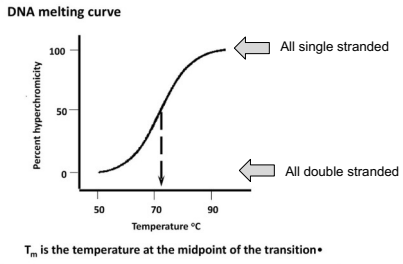
Can follow "melting" – strand dissociation with increasing temperature; ss DNA is hyperchromatic.

Can also separate ss from ds DNA (sticks) on HAP column.

Saenger (1984)

5

DNA melting curve



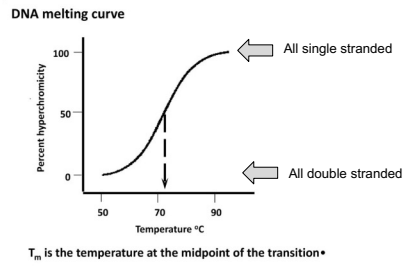
T_m is the temperature at the midpoint of the transition•

What will impact Tm? And why?

Figure taken from the Web – may be copyright.

6

DNA melting curve



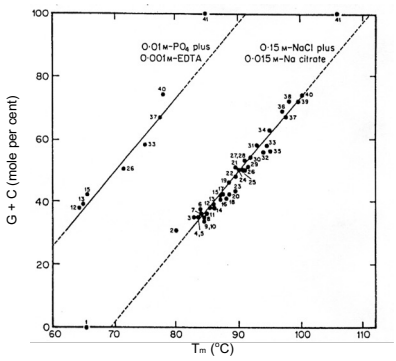
What will impact Tm? And why?
- GC content
- Buffer
- Fragment size

Figure taken from the Web – may be copyright.

7

Dependence of melting temperature (T_m) on G+C content of DNA

(also dependent on salt, pH, and fragment size)



Saenger (1984)

8

DNA denaturation can be followed by DNA renaturation/ reassociation/ hybridization

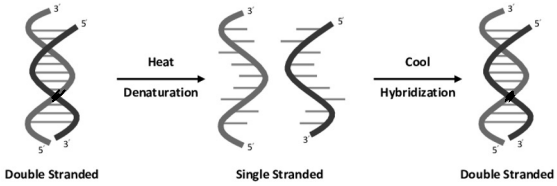


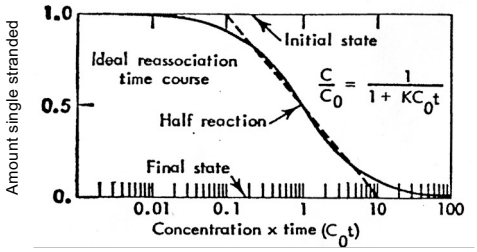
Figure taken from the Web – may be Copyright.

Protocol:
- Use DNA sheared to ~400 bp, standard buffer
- Denature DNA by heating to ~95 °C
- Drop temperature to Tm – 25°
- Follow change in OD_{260} over time

9

Ideal C_0t curve

Standard second order rate equation: limiting factor is the time it takes fragment to find a matching partner



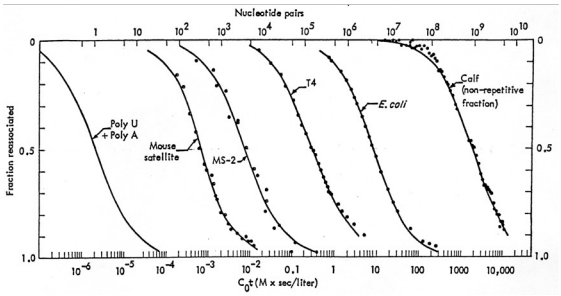
C_o = concentration (nucleotides/liter) at starting time t = 0 (seconds)

Britten R and Kohne D, Science **161:** 529-540 (1968)

10

C_0t curves of different DNA samples

Cot at 0.5 (half) of DNA renatured gives you complexity of sample (amount of unique sequence) in bp. Renaturation occurs over two orders of magnitude of Cot.

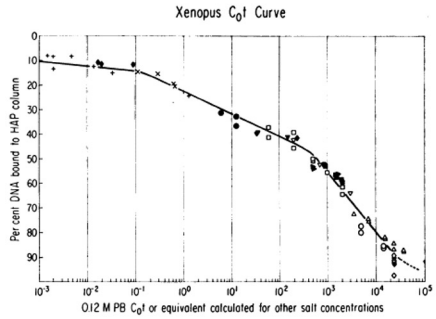


Britten R and Kohne D, Science **161:** 529-540 (1968)

11

C_0t curve of a eukaryotic genome

Curve does not indicate a simple second order reaction, but something more complex!



E.Davidson and B.Hough, PNAS **63:** 342-9. (1969)

12

What is this complex Cot curve telling us?

Some of the DNA fragments find a partner more quickly – they must be present in multiple copies!

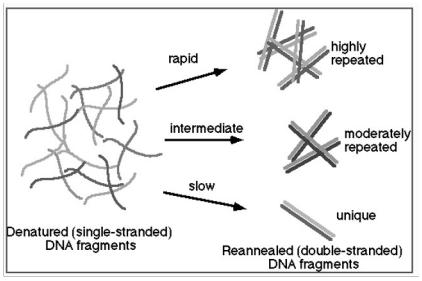
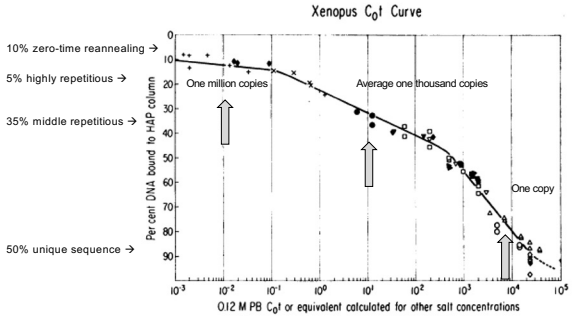


Figure taken from the Web – may be Copyright.

13

C_0t curve of a eukaryotic genome

Curve exhibits multiple frequency classes. Assuming slowest fraction to anneal is unique, you can derive relative frequency, per cent of genome, for each class



Based on Davidson and Hough, PNAS **63,** 342 (1969)

14

Recognized multi-copy genes

Table 1 Classification and properties of multigene families[a]

Campbell et al, 1975

Category	Gene products	Multiplicity	Gene or protein homology	Information content	Examples of coincidental evolution	Examples of change in family size
Simple-sequence satellites	None known	10^5–10^8	80–100%		Different satellites of *Drosophila*	Mouse satellite
Multiplicational 18S–28S RNA	RNA, protein	100–600	97–100%	One unit	Spacer regions of *X. laevis* and *X. mulleri*	
5S RNA		100–1200	97–100%	One unit	Spacer regions of *X. laevis* and *X. mulleri*	Gene number in *X. mulleri* and *X laevis*
tRNA		6–400		One unit		
histones		10–1200	87–99%	Few units	Histone mRNAs of two species of sea urchin	Gene number in different sea urchin species
Informational antibodies	Proteins	100s	30–100%	Many units	Rabbit and mouse κ chains	V_λ and V_κ in mammals
hemoglobins		~10	<75–100%	Few units	Human and cow δ chains	Human and rabbit β-like genes

[a] In all instances characteristics refer to a single closely linked set of genes.

Hood, Campbell, & Elgin, Ann Rev Genet (1975)

15

Satellite DNAs of *Drosophila melanogaster*

Satellite	Buoyant density	G-C base pairs, %	Predominant sequence	Number of copies	Proportion of genome, %	Chromosome distribution
I	1.673 (13%)	7	generally $(AAT)_n(AT)_m$ may be either 5′ A A T A A T A T A T A T 3′ T T A T T A T A T A T A or both 5′ A A T A T 3′ T T A T A (60%) and 5′ A A T A T A T 3′ T T A T A T A (40%)	6×10^5	5.2	68% on Y 22% on 4
II	1.686 (27%)	22	5′ A A C A T A G A A T 3′ T T G T A T C T T A	4×10^5	2.6	50% on Y 24% on 2 18% on 3
III	1.688		has 378 bp repeat with considerable variation		4	48% on X 44% on Y
IV′	1.705 (46%)	38	contains AAG, AG, G in ratio 5:4:1 90% is: 10% is: 5′ A A G A G A A G A G A G 3′ T T C T C T T C T C T C	1×10^6	3.8	52% on Y 33% on 2

JG Gall et. al, ca. 1975

16

Three major families of transposable elements

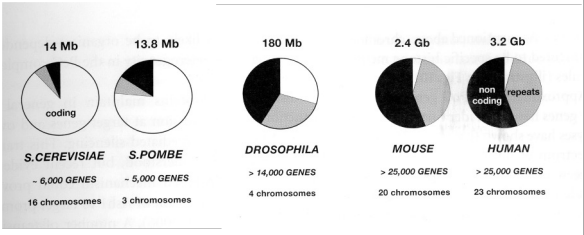
derived from DNA transposons and retroviral elements

Structure	Genes in Element	Mode of Movement	Examples
short inverted repeats at each end	encodes transposase	moves as DNA, either excising or following replicative pathway	P-element (Drosophila) Ac-Ds (maize) Tn3 and IS1 (E. coli) Tam3 (snapdragon)
directly repeated long terminal repeats (LTRs) at ends	encodes reverse transcriptase and resembles retrovirus	moves via an RNA intermediate produced by promoter in LTR	Copia (Drosophila) Ty (yeast) THE-1 (human) Bs1 (maize)
Poly A at 3′ end of RNA transcript; 5′ end is often truncated	encodes reverse transcriptase	moves via an RNA intermediate that is presumably produced from a neighboring promoter	F-element (Drosophila) L1 (human) Cin4 (maize)

after Alberts et al. 3rd ed.

17

Larger genomes have a larger proportion of repeats and non-coding DNA



Adapted from Allis et al: Epigenetics 2007

18

Repeats in the human genome

Majority of LINEs are L1; majority of SINEs are Alu repeats

Repeat class	Repeat type	Number (hg19)	Cvg	Length (bp)
Minisatellite, microsatellite or satellite	Tandem	426,918	3%	2–100
SINE	Interspersed	1,797,575	15%	100–300
DNA transposon	Interspersed	463,776	3%	200–2,000
LTR retrotransposon	Interspersed	718,125	9%	200–5,000
LINE	Interspersed	1,506,845	21%	500–8,000
rDNA (16S, 18S, 5.8S and 28S)	Tandem	698	0.01%	2,000–43,000
Segmental duplications and other classes	Tandem or interspersed	2,270	0.20%	1,000–100,000

Treangen T & Salzberg S 2012 Nature Rev Genetics 13: 36-46.

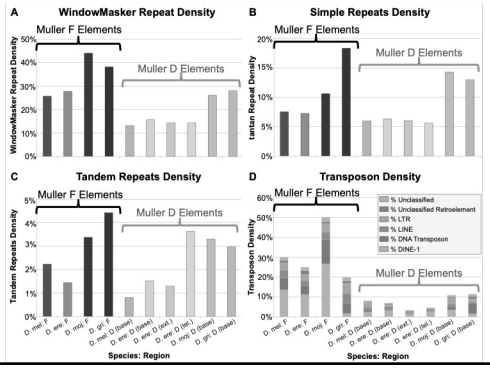
(Cvg = percentage of the genome covered by the repeat class)

Most abundant: L1 (a LINE) = 17% of genome; Alu (a SINE) = 11% to 15%.
Danger ! Mobile TEs are mutagens. Example: new muscular dystrophy mutation.
HERV-K: viral proteins from HML-2 group implicated in in ALS;
HERV-W linked with MS.
Triggering inflammatory response? Impact in cancer?

19

*Repeat distribution is not uniform:

Drosophila Muller F elements have higher repeat density than the Muller D elements and differ in repeat type



Leung et al 2015 G3 5: 719-40

20

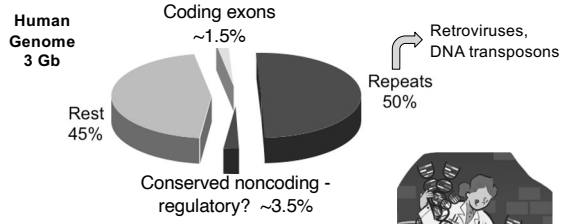
Considerations for genome sequencing

1. Satellite DNA is difficult to sequence, as there are few markers to help order reads or subclones; thus, often centromeric regions of the chromosomes remain unsequenced. New long-read technology (PacBio) is changing that.
2. Middle repetitious DNA also causes difficulties; because one finds nearly identical sequences located in different regions of the genome, mistakes can be made in assembling sequence data. Fortunately, high quality discrepancies can often identify different copies of a repeat.
3. Much of the repetitious DNA is packaged in heterochromatin, which maintains these regions in a compact and transcriptionally silent form.
4. However, in higher organisms many protein-coding genes are found embedded in repetitious DNA. Check out your favorite human gene on the UCSC Browser by taking off RepeatMasker!

SCR Elgin

21

Large eukaryotic genomes must be packaged to regulate gene expression while minimizing TE movement.



Key Questions:

- Is repetitious DNA junk or garbage ?
- How is it all packaged into a nucleus ?
- How is silencing maintained - correctly ?

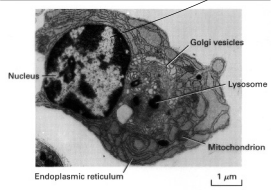
Xiaohui Xie, MIT 2007

22

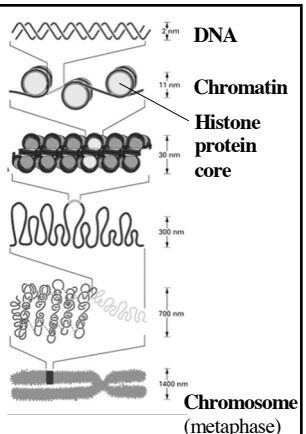
Packaging large genomes:

First step - packaging in a nucleosome array

Second - differential packaging into heterochromatin & euchromatin



Lodish et.al., Molecular Cell Biology, 4th Edition



Felsenfeld & Groudine, 2003, Nature 421: 448 – 453.

23

DNA packaging in chromatin & chromosomes

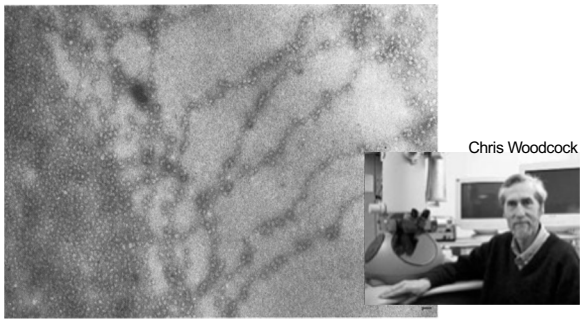
Humans: 30 miles of thread packed into a basketball!
(packing ratio = length of DNA divided by length of chromatin fiber)

			Packing Ratio	Model
Naked DNA	20A dia.	10 bp/turn	1	good
Chromatin (100 A fiber)	100A dia.	~80 bp/turn	6-7	good
Chromatin (300 A fiber)	300A dia.	6 ν, 1200 bp/turn	~40	vague (sequence dependent?)
Domains (loops, TADs)		20-100 kb/loop	~700	better (cell-type dependent?)
Chromosome		10^6 - 10^8 bp	~10,000	defined

SCR Elgin

24

Prior to 1972, chromatin was thought to be a smooth fiber – DNA coated with histones; but subunits?



Chris Woodcock

Olins et. al., J. Cell Biol, (1975) **64,** 528-537 (Electron micrograph, rat thymus nucleus).
First presentation: posters by Woodcock and by Olins & Olins at ASCB, 12/1973.

25

Peer review…..

"A eukaryotic chromosome made out of self-assembling 70A units, which could perhaps be made to crystallize, would necessitate rewriting our basic textbooks on cytology and genetics! I have never read such a naïve paper purporting to be of such fundamental significance. Definitely it should not be published anywhere!"

- Anonymous review of paper submitted by C.F.L. Woodcock, 1974, showing EM pictures of nucleosome arrays

26

CHAP. XV: Cautions in viewing Objects.
"Beware of determining and declaring your Opinion suddenly on any Object, for Imagination often gets the Start of Judgment…. Pass no Judgment upon Things over-extended by Force, or contracted by Dryness, or in any Manner out of their natural State, without making suitable Allowances."

From "The Microscope Made Easy" by Henry Baker, 1742

27

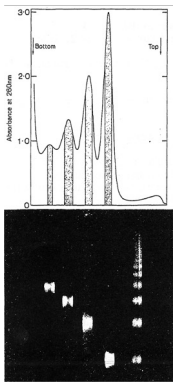
Establishing the nucleosome model..
- a paradigm shift, 1973-1975

1. Electron microscopy - images
2. Micrococcal nuclease digestion patterns
3. Knowledge of histone:histone interactions

SCR Elgin

28

Observation of defined lengths of chromatin generated by MNase



Sucrose gradient fractionation of micrococcal nuclease digestion products

- Top of gradient is on the right
- Bottom of gradient is on the left
- Fractions collected from shaded areas

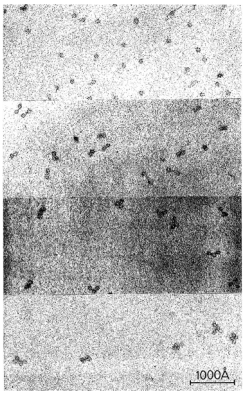
Polyacrylamide gel electrophoresis of purified DNA

- Right lane: unfractionated digest
- Left lanes: DNA purified from sucrose gradient peaks
- Cleavage indicates repeating subunit

Finch et.al., PNAS (1975) 72: 3321

29

Electron micrographs of fractions from sucrose gradient



monomer fraction

dimer fraction

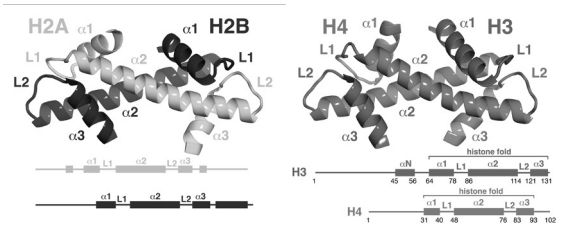
trimer fraction

tetramer fraction

Finch et.al., PNAS (1975) **72,** p3321

30

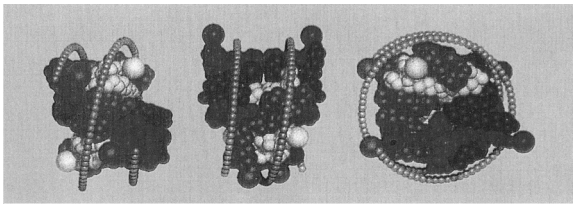
How are the beads assembled? Histone fold



Song Tan, Penn State

31

DNA on the outside of the bead? Evidence from neutron scattering, DNase I digestion



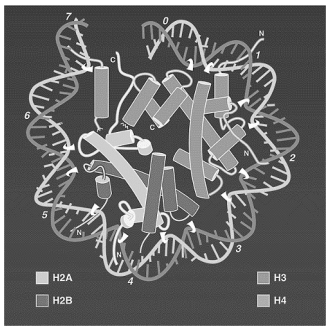
H2A
H2B
H3
H4

Axis of the DNA path in the nucleosome (not full width)

Arents et. al., PNAS (1993) **90,** 10489-93

32

The crystal structure of the nucleosome core!



Resolution: 2.8 Å

Half of the nucleosome structure is shown

One turn of the DNA helix is visible (73 bp)

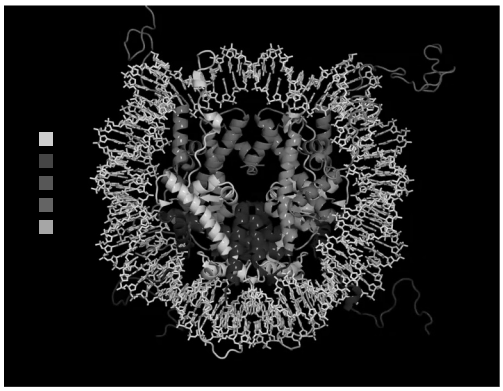
View is down the superhelix axis

Protein - DNA contacts: white hooks

Seven years effort: required symmetric DNA (inverted repeat) and histones made in *E. coli* (uniformly free of post-translational modifications)

Rhodes, Nature (1997) **389,** 231-233, after Luger et. al., Nature (1997) **389,** 251-260

33



34

Summary: Nucleosome Stoichiometry

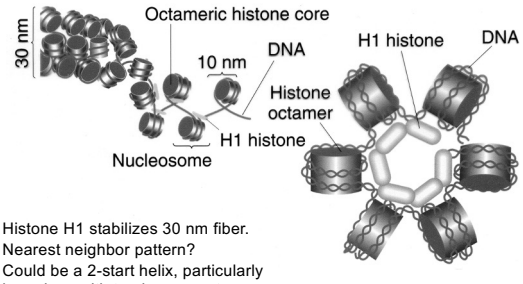
1. **Core particle:**
   - 147 bp DNA
   - histone octamer
     - tetramer $[H3 + H4]_2$
     - 2 dimers [H2A + H2B]

2. **Nucleosome** (repeating subunit)
   - 167 bp DNA (2 turns) plus ~40 bp linker → ~200 bp
   - histone octamer
     - tetramer $[H3 + H4]_2$
     - 2 dimers [H2A + H2B]
   - 1 H1 ("linker histone" - histone 1 or histone 5)

SCR Elgin

35

A model of chromatin structure: 10 nm and 30 nm fibers

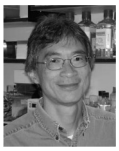


Histone H1 stabilizes 30 nm fiber.
Nearest neighbor pattern?
Could be a 2-start helix, particularly in regions with tandem repeats.

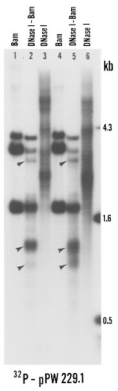
Griffiths et.al. Introduction to Genetic Analysis, 2000

36

*Given recombinant DNA probes, we can analyze the chromatin structure of specific genes:
Discovery of DH sites



Carl Wu

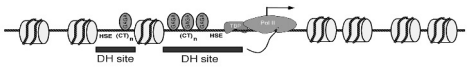


^{32}P - pPW 229.1

Wu et al, 1978 Cell16: 797.

37

*Assessing chromatin structure - genes are packaged in nucleosome arrays with the active TSS in an open region – a DH site
(or nucleosome-free region)



Analysis: MNase cuts between nucleosomes; DNase I cuts nucleosome-free sites (DH – DNase hypersensitive sites). Results shown for *hsp26.*

Mapping DH sites genome-wide identifies accessible promoters, enhancers, other regulatory sites for a given tissue.

Drawing by G. Farkas, based on Thomas & Elgin 1988 EMBO J 7: 2191

38

Histones are general repressors of template activity.

- Default state of the eukaryotic genome is "off."
- Nucleosomes block access to promoter *in vitro*
- Active and inducible genes show 5' nuclease accessible hypersensitive sites (HS sites)
- Generation of such sites is part of gene activation
- Remodeling activities are critical for activation
- Histone modification patterns reflect activity state
- Histone variants reflect activity state

SCR Elgin

39

Summary

- Eukaryotic genomes are packaged in a nucleosome array; an excellent model of the chromatin fiber based on EM results, nuclease digestion, physical measurements, and protein-protein interaction studies is available.
- The chromatin subunit (core nucleosome) has been crystallized, and the textbooks have been rewritten!
- Nucleosome packaging negatively impacts gene expression, keeping the bulk of the genome silent.
- Active Transcription Start Sites are found in Hypersensitive Sites, nucleosome-free regions; this selective access compensates for the large size of the genome.

SCR Elgin

40