PHYSICAL MAPPING OF AN EARLY SEA URCHIN GENE BATTERY FROM *PARENCHINUS ANGULOSUS*

BY : T.N. Lawson

Department of Biochemistry

University of Cape Town

M Sc June 1988

SUPERVISORS : Dr. B.T. Sewell and

Prof. C. von Holt

The University of Cape Town has been given the right to reproduce this thesis in whole or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

ACKNOWLEDGEMENTS

I wish to thank

- Prof. M.L. Birnstiel for supplying the pBH22 clone
- Prof. W. de Wet for helping me with the heteroduplex analysis
- Dr. D. van Skalkwyk and the MRC for help with statistics and for use of computer facilities
- The Electron Microscopy Unit for use of their facilities
- Prof. K. Dumbell for supplying pBR329 and pAT153
- Miss J. Rees for isolating SU1 and for all her help
- Dr. J. Retief for the plasmids used for probing
- Mrs. B. Wingfield for her helpful discussions
- Prof. K. Hardie for his help
- My family for everything they did for me
- R.S.C. Anderson for his patience and his help in preparing the diagrams
- All my friends for bearing with me throughout this thesis

and last but not least

- Dr. B.T. Sewell and Prof. C. von Holt for their constant help and encouragement.

CONTENTS

	Page
SUMMARY	1
PHYSICAL MAPPING OF AN EARLY SEA URCHIN GENE BATTERY FROM PARENCHINUS ANGULOSUS	3
1.1 INTRODUCTION	3
1.2 ELECTRON MICROSCOPY	4
1.2.1 Spreading techniques used in preparing DNA for electron microscopy	4
1.2.2 Partial denaturation mapping	6
1.2.3 Heteroduplex Mapping	10
1.3 SEA URCHIN HISTONE GENES	13
1.3.1 The Early Sea Urchin Genes	13
1.3.2 The Late Sea Urchin Histone Genes	15
1.3.3 Homology observed in Sea Urchin Histone DNA	16
THE CHARACTERIZATION OF H27	19
2.1 INTRODUCTION	19
2.2 PARTIAL DENATURATION MAPPING	19
2.3 HETERODUPLEX ANALYSIS	25
2.4 CONCLUSION	30
THE CHARACTERIZATION OF SU1	31
3.1 THE LENGTH OF SU1	31
3.2 THE PARTIAL RESTRICTION MAP OF SU1	31
3.3 THE DETERMINATION OF THE ORDER OF THE GENES IN SU1	34
3.3.1 Introduction	34
3.3.2 The order of the genes in SU1	34

3.4 THE MELTING PROFILE OF SU1 43

3.5 PARTIAL DENATURATION MAPPING 45

3.6 HETERODUPLEX ANALYSIS 49

3.7 CONCLUSION 58

MATERIALS AND METHODS 59

4.1 PREPARATION OF PLASMID DNA 59

4.2 EXTRACTION OF DNA FROM GELS 60

4.3 RADIOACTIVE LABELING 60

4.4 ELECTROPHORESIS 61

4.5 SOUTHERN TRANSFERS 61

4.6 RESTRICTION ENZYME DIGESTIONS 62

4.7 LIGATION REACTIONS 62

4.8 PROPAGATION OF PLASMIDS 62

4.9 MELTING PROFILE 63

4.10 ELECTRON MICROSCOPY 63

APPENDIX 66

5.1 DATA ANALYSIS. 66

5.1.1 Data analysis 67
5.1.1.1 Tracing and recording of length information 67
5.1.1.2 Alignment and orientation of molecules 68
5.1.2 Extraction of secondary data 69

5.2 THE USE OF THE SAS STATISTICAL PACKAGE 71

5.3 PROGRAM FOR DATA ANALYSIS 74

REFERENCES 95

INDEX TO FIGURES

	Page
Figure 1	20
Figure 2	21
Figure 3	22
Figure 4	23
Figure 5	26
Figure 6	28
Figure 7	31
Figure 8	32
Figure 9	33
Figure 10	35
Figure 11	36
Figure 12	38
Figure 13	39
Figure 14	40
Figure 15	41
Figure 16	42
Figure 17	44
Figure 18	45
Figure 19	47
Figure 20	50
Figure 21	52
Figure 22	53
Figure 23	56
Figure 24	57
Figure 25	66

INDEX TO FIGURES

	Page
Figure 1	20
Figure 2	21
Figure 3	22
Figure 4	23
Figure 5	26
Figure 6	28
Figure 7	31
Figure 8	32
Figure 9	33
Figure 10	35
Figure 11	36
Figure 12	38
Figure 13	39
Figure 14	40
Figure 15	41
Figure 16	42
Figure 17	44
Figure 18	45
Figure 19	47
Figure 20	50
Figure 21	52
Figure 22	53
Figure 23	56
Figure 24	57
Figure 25	66

INDEX TO TABLES

	Page
Table 1	24
Table 2	29
Table 3	48
Table 4	54

SUMMARY

The aim of this project was to characterise an early histone gene battery isolated from *Parenchinus angulosus*.

An early histone gene battery (named H27) which was believed to have been isolated from *Parenchinus angulosus*, appeared by restriction enzyme mapping and partial sequencing to be identical to H22, an early histone gene battery isolated from *Psammechinus miliaris*.(This latter gene was obtained from M. Birnstiel.) This was further confirmed by electron microscopy, and proved to be a convenient testing ground for the electron microscopic techniques of denaturation mapping and heteroduplex analysis.

Another gene battery (named SU1) isolated from *Parenchinus angulosus*, was then characterised using the techniques developed whilst studying H27. The restriction enzyme map of this clone is different to that of H22, indicating that differences do indeed exist between these two early histone gene batteries. SU1 also showed the expected order of the five histone genes, as determined by hybridization against the coding regions of H22.

The denaturation map of SU1 showed AT rich spacer regions and GC rich coding regions. Heteroduplex analysis indicated that the spacer regions between the H1 and H2A, the H2A and H3, and the H3 and the H2B gene coding areas are essentially nonhomologous. The H4 structural gene and corresponding spacer regions were not included in this analysis. Because it is known that all the five histones are coded for on the same strand of DNA in H22, and that each of the genes is

transcribed in the same direction, it follows that, the same holds for , at least, the H1, H2A, H3 and H2B genes of SU1.

PHYSICAL MAPPING OF AN EARLY SEA URCHIN GENE BATTERY FROM PARENCHINUS ANGULOSUS

1.1 INTRODUCTION

The various techniques available for the physical mapping of genes can be broadly grouped under the following headings:

Restriction enzyme mapping (Cohen, 1975), use of radioactive probes that can be hybridised to the DNA of interest, sequencing (Maxam and Gilbert, 1977) and electron microscopic techniques.

Electron microscopy is a powerful tool in the analysis of nucleic acid structure in that it is about the only technique in which macromolecules can be viewed directly. Accurate, quantitative measurement can be routinely achieved with very small quantities of nucleic acid and the concentration, size and physical state of a preparation can be assessed. The two DNA electron microscope techniques that will be discussed are partial denaturation mapping and heteroduplex analysis as they were important in determining the physical characterization of an early histone gene battery isolated from *Parenchinus angulosus*.

1.2 ELECTRON MICROSCOPY

1.2.1 Spreading techniques used in preparing DNA for electron microscopy

The basic protein film technique, developed by Kleinschmidt and Zahn in 1959 has, because of its simplicity and reproducibility, provided a cornerstone for the development of quantitative electron microscopy of nucleic acids. The introduction of formamide to the spreading method mixture (Westmoreland, et al., 1969) served as an important contribution. In the aqueous method, single strand DNA appears as condensed bushes (Brack, 1981, Ferguson and Davis, 1978), while in the formamide method, both single and double stranded DNA are well extended.

The principle of the basic protein film technique is as follows (Brack, 1981):

The DNA is mixed with a slightly basic protein. Several proteins such as lysozome, chymotrypsin, trypsin and histones have been tried (Davis,et al., 1971). However, of all the proteins tried, cytochrome c has given the best results (Brack, 1981; Davis, et al., 1971). The DNA-protein solution is allowed to flow down a glass ramp onto a hypophase, where the spreading takes place. The protein denatures at the air-water interphase during spreading and forms a monolayer of denatured protein in which the DNA is embedded. The non-specific and co-operative binding of the denatured protein to the DNA helps disentangle the filaments and spread out the molecules of interest. The nucleic-acid protein complexes are

adsorbed onto a support film (either carbon or collodion) on an EM grid and then stained with either uranyl acetate or phosphotungstic acid. Additional contrast is normally achieved by rotary shadowing the grids with heavy metal before visualization with the electron microscope. (Coggins (a), 1987).

The appearance of the DNA and background depends on several factors (Coggins (a), 1987), including:

the quantity and concentration of cytochrome c, ionic strength, type of salt or buffer used, pH, formamide concentration and the type and quality of the support film on the grid.

Some of these factors can be controlled, but two samples prepared in parallel can differ in appearance , as can different areas of the same grid.

This also means that consideration must also be given to the extent to which the preparation of the DNA molecule for electron microscopy has affected its contour length. The physical stretching of the nucleic acids during surface denaturation of the cytochrome c film cannot be reproducibly controlled, Also, factors such as ionic strength and formamide concentration in the hypophase and spreading solution affect the length of nucleic acid molecules (Ferguson and Davis, 1971; Brack, 1981). It is therefore necessary to add standard DNA (or RNA molecules, depending on the type of nucleic acid in the experiment) of known molecular weight to the preparation just before spreading. Circular DNA molecules are preferentially used as size standards because they are readily distinguishable

from a population of experimental molecules, which are most often linear (Coggins (a), 1987; Brack, 1981).

Instead of using cytochrome c, a detergent, the most popular being benzyl dimethyl-alkyl-ammonium chloride (BAC), can be added to the spreading solution(Brack, 1981; Coetzee and Pretorius, 1979; Vollenweider, et al., 1975). Otherwise the procedure is very similar to that of the protein monolayer spreading techniques. BAC has the advantage of not leading to any detectable thickening of the DNA strands and therefore increases the resolution of small details along a molecule, such as denatured regions. It can also be used, however only to a limited extent, because BAC is a detergent, to visualize protein molecules bound to nucleic acids (Brack, 1981; Vollenweider, et al., 1975; Coggins (a), 1987).

Despite the advantages of using a detergent such as BAC, the basic protein film technique still remains more popular. This is because the latter can be carried out in a more easily reproducible manner in most laboratories.

1.2.2 Partial denaturation mapping

The denaturation of DNA can be brought about by any factor or combination of factors that disturb the forces that hold the two polynucleotide strands of DNA together. Commonly used procedures are heat, changing pH and/or ionic conditions and the introduction into the system of reagents that lower the Tm of DNA such as formamide and sodium perchlorate.

A general idea of the base composition of a DNA molecule can be obtained by denaturation mapping, due to the differences in the stability, or reactivity, between AT- and GC- base pairs. One of the first molecules to be studied in this way was the Lambda phage DNA (Inman, 1966; Inman, 1967; Inman and Schnös, 1970). Inman incubated the double stranded DNA at "high" temperatures (between 48 and 53°C for 10 minutes) or high pH (pH 10 to 11, up to 308 min) in the presence of formaldehyde. The formaldehyde reacts with amino groups on the non hydrogen bonded bases and prevents their reassociation, thus trapping the partially melted structure and blocking renaturation.

Before preparing the DNA samples for electron microscopy, it is essential to remove most of the unreacted formaldehyde. The presence of the formaldehyde during cytochrome c spreading leads to rough background and to poor film formation (Bujard, 1970). The formaldehyde can be removed by dialysis against 0.1 M Tris (Tris (hydroxymethyl) aminomethane), 0.01 M EDTA (Ethylene diamine tetra acetic acid) (Coggins (b), 1987) or by gel filtration on a small Sephadex G-100 (Delius and Clements, 1976) or G-50 column (Coggins (b), 1987).

Instead of formaldehyde, glyoxal can be used for the fixation of denatured regions (Bick and Thomas, 1975; Johnson, 1975). Glyoxal reacts with guanine to form a reasonably stable adduct in slightly alkaline conditions. Because AT rich regions "melt out" before GC rich regions, the site of glyoxal addition is deoxyguanosines in AT rich regions.

Denaturation of DNA has also been carried out using the gene-32 protein. This protein has a selective affinity for single stranded DNA over double stranded DNA (Delius, et al., 1972). The tight, co-operative binding of gene-32 protein drastically lowers the melting temperature of double-helical DNA.

Other methods used for denaturing mapping have been the introduction of sodium perchlorate, designed so as to "avoid possible hydrolysis of alkali-labile bonds" (Delius and Clements, 1976) and the spreading of DNA at high formaldehyde concentrations (Wolfson, 1972), which is a relatively easy and quick procedure.

Denaturation mapping however, does have some major problems. Within a given stretch of DNA, not all regions are observed to denature to the same extent, and within the same experiment, in some molecules, the region of denaturation is not observed at all. The reason for this is unknown. It has been suggested that the HCHO, added to prevent renaturation of single stranded areas, causes denaturation of adjacent regions, resulting in larger than expected "bubbles" (Inman, 1966). Secondly, there is the problem of lateral aggregation, especially of the smaller loops (Coggins (b), 1987; Ferguson and Davis, 1978). If the DNA becomes thickly covered with protein, the DNA strands can stick together along their length. Lateral aggregation of two small single-stranded loops can cause them to appear as duplex DNA under the electron microscope. (Ferguson and Davis, 1978; Davis and Hyman, 1971).

Further complication can occur if the denaturation pattern along a stretch of DNA is not seen to be asymmetrical, resulting in ambiguous alignment of molecules. The orientation of molecules can be facilitated by restriction endonucleases (Coggins (b), 1987). A restriction enzyme can be used to cut outside the region of interest (in vector sequences, for example), but in such a way that the experimental molecule has an unambiguous denaturation pattern. However, this is not always possible, for example, in the case where a convenient restriction enzyme site that cuts only within the vector, but not in the region of interest, cannot be found.

Fragmentation of DNA is also observed after partial denaturation (Ruvechan, et al., 1985; Gomez and Lang, 1972; Kudler, et al., 1983). This can cause problems should the denaturation pattern of a segment of DNA be unknown. Also, the problem becomes more serious, the longer the DNA molecule to be examined, and the more nicking the DNA experiences on isolation and preparation for electron microscopy. Should this problem prove serious, conditions during the denaturation, such as time of incubation and the pH, must then be carefully selected so as to minimize fragmentation, while at the same time, to obtain the degree of denaturation required for a particular DNA.

Of all the methods available, the alkaline-formaldehyde method appears to give the best differentiation between AT- and GC- rich regions (Brack, 1981; Coggins (b), 1987; Inman and Schnös, 1970). However, because individual molecules show variations in the denaturation pattern, it is only by constructing a denaturation histogram for a whole

population of molecules, that a possible characteristic pattern can be obtained.

1.2.3 Heteroduplex Mapping

As its name implies, a heteroduplex molecule is constructed by annealing two related, yet different single strands of DNA to form a duplex structure. Experimentally, renaturation rates are found to be maximal at about 25°C below Tm. If the DNA is kept within this temperature range, the kinetic effects of intramolecular hairpin formation are minimized because most short hairpins are unstable at such temperatures (Cantor and Schimmel, 1980).

The introduction of formamide into both the preparation of the heteroduplexes and the spreading of the samples has improved both the condition for heteroduplex formation and for single strand visualisation (Brack, 1981; Ferguson and Davis, 1978). However in order to obtain unambiguous results in a heteroduplex preparation, it is necessary to have intact DNA molecules. The presence of fragmented DNA leads to branched structures and the reassociation of several molecules (Davis, et al., 1971). Also, intact molecules having long regions of nonhomology can lead to complicated aggregates due to part of one heteroduplex hybridizing to unreacted single strands or to other heteroduplexes. In order to minimize such problems, it is advisable to underrenature the DNA (~50% is optimal) and thus renaturation can sometimes be carried out with 70% formaldehyde (Wellauer, et al., 1976).

The differences between DNA molecules that can be mapped in a heteroduplex experiment fall into three classes (Coggins (b), 1987). One or more of these features may be present in a heteroduplex molecule.

(i)The insertion or deletion loop:

This occurs when two DNA sequences are identical except for an insertion or deletion of a DNA segment. The unpaired DNA forms a single stranded DNA loop.

(ii)Substitution loop:

This occurs when two DNA sequences are identical apart from a segment which is replaced by another sequence at the same position. This results in two unpaired strands interrupting an otherwise double-stranded heteroduplex molecule.

(iii)Single base changes:

Two DNA molecules which differ by a number of single base changes contain mismatched sequences, and this may result in double stranded or single stranded DNA, depending on the stringency of hybridization.

In the second and third cases especially, the problems that are experienced in partial denaturation mapping are also encountered here, although possibly not to the same extent. The single stranded bubbles are not of homogenous size and, also in a few cases, a few of the uncomplementary regions may not be observed at all (Kudler, et al., 1983). Also, on occasion it may prove difficult to orientate and align heteroduplex molecules due to a somewhat symmetric pattern of nonhomologous regions. Again, the orientation of

molecules can be greatly eased by the use of restriction endonucleases as in the case of partially denatured DNA. Vector sequences, themselves also can greatly aid in the orientation of an alignment of heteroduplexed molecules. Vectors can be chosen in such a way so that they are identical except for , say, a deletion or insertion. On renaturation, this insertion or deletion loop can be used to both align the molecules, as well as to aid in the detection of hybrids (Kinchington, et al., 1984). However, cases still arise where the orientation of molecules is difficult. If two DNA sequences in which the degree of nonhomology is to be determined, cannot be cloned into the same restriction site in a vector, and they prove to be very nonhomologous, then competition between hybridization of the vector and insert sequences may prove a serious problem (Coggins (b), 1987). In such a case, the use of a vector to unambiguously orientate the molecules would create more problems than it is worth.

Heteroduplex analysis has been successfully applied to a number of problems: It has been used to work out the arrangement of early histone genes and spacers in Sea Urchin DNA (Cohn and Kedes, 1979), the role of insertion and deletion of transposable elements in P-plasmid evolution (Villaroel, et al., 1983) and the position of specific segments of the T7 chromosome along its length following digestion with a restriction enzyme (Bick and Thomas, 1975). The relative simplicity of this electron microscope technique should still prove popular in years to come.

1.3 SEA URCHIN HISTONE GENES

1.3.1 The Early Sea Urchin Genes

The work described here involves the characterisation of an early histone gene battery isolated from *Parenchinus angulosus*. It is a useful exercise, therefore, to look at the other histone gene batteries which have been isolated from other sea urchin species. The late sea urchin histone genes will be discussed only briefly.

Until about 1979, almost all of the available information on the organization and mode of expression of histone genes had been derived from the examination of a histone gene quintet reiterated several hundred fold in the genome of the sea urchin. The ability to physically enrich this type of gene led to the early cloning and characterisation of these "early" histone genes (Hentschel and Birnstiel, 1981).

In order to understand the most fundamental level of regulation - the transcriptional level- knowledge of the organisation of the histone genes is needed. The first "early" sea urchin gene quintet to be studied in detail was H22 isolated from *Psammechinus miliaris*. The arrangement of the five coding sequences was elucidated by cleaving the 6kb fragment using restriction enzymes and hybridizing

against the individual highly purified mRNA's (Schaffner, et al., 1976). The order of the coding sequences was found to be:

H4 H2B H3 H2A H1

The relative positions of the "early" sea urchin histone genes has been found to be identical in other echinoderms studied to date (Cohn, et al., 1976; Hentschel, et al., 1981; Kedes, et al., 1975). Wu, Holmes and Davidson (1976) mapped the arrangement of genes and spacers on the chimeric plasmids pSp2 and pSp17 isolated from *Strongylocentrotus purpuratus* by hybridizing histone mRNAs with a single strand of the coding DNA and observing the positions of the RNA:DNA duplex regions along the single strand of DNA under the electron microscope. R-loop analysis has been carried out on two nonallelic histone gene clusters of *Lytechinus pictus* (Cohn and Kedes, 1979) and the same general topology of "early" sea urchin histone genes was again observed.

Each of the structural genes within the quintet is separated by an AT rich spacer region. That there is an alternation of spacers and structural genes is obvious from the electron microscopic studies. The first indication that the spacers are AT rich is gained from by the melting profile of H22 isolated from *Psammechinus miliaris*. The melting profile falls into early and late melting phases corresponding to a GC content of approximately 37% for the early melting phase and 52% for the late melting phase (Schaffer, et al., 1976; Portmann, et al., 1976). This and the fact that histone mRNAs are high in GC content (Grunstein, et al., 1973), because of the high abundance of basic amino acids (von Holt, et al., 1979; Wells, 1986), showed that the bulk of AT rich DNA does not code for histone proteins. The partial denaturation studies of

Portmann, et al., (1976) and sequencing (Schaffner, et al., 1978; Sures, et al., 1978) have verified that the spacers between each of the genes are indeed AT rich.

The apparent lack of introns observed in the electron microscopic studies is confirmed by the DNA sequencing of protein coding regions (Schaffer, et al., 1978; Sures, et al., 1978; Wells, 1986). Also, all five genes are transcribed off the same strand. The electron microscopic studies of Wu, et al., (1976) and Holmes, et al., (1977) showed that the mRNA coding for the histones H2B, H4 and H1 hybridize to one strand of pSp2 and that H3 and H2A hybridize to one strand of pSp17. That all five are coded for on the same strand was verified by the exonuclease III studies of Cohn, et al., (1976). Heteroduplex analysis of *Lytechinus pictus* against *Strongylocentrotus purpuratus* showed that the coding sequences of both sea urchin clones share considerable sequence homology (Cohn and Kedes, 1979). This, together with exonuclease III studies (Cohn and Kedes, 1979), confirmed that all the coding regions in *Lytechninus pictus* also lie on the same DNA strand.

1.3.2. The Late Sea Urchin Histone Genes

The "early" histone genes are responsible for the major portion of the histone mRNA synthesised prior to early blastula. In mid blastula however, another set of genes come into play and by gastrula, almost all of the histone mRNA is derived from the "late" histone genes. The two sets of genes are regulated independently resulting in a switch in the types of histone mRNA synthesised in the early and late embryo (Kaumeyer and Weinberg, 1986).

The late histone genes appear to be highly diverged from the early genes and their gross topology is also different. They are randomly arranged and occur as dispersed clusters (Maxon, et al., 1983; Roberts., et al., 1984; Kaumeyer and Weinberg, 1986) or as single genes (Knowles and Childs, 1986).

1.3.3 Homology observed in Sea Urchin Histone DNA

In general, the H3 and H4 proteins are more highly conserved than the other three histone proteins H2A, H2B and H1, H1 being the most variable protein, especially in its "nose" and "tail" regions. However, differences are observed in the gene sequences of even those portions of the regions coding for essentially the same protein sequence - in such cases by silent mutation predominantly in the third base of the codon. (Wells, 1986; Busslinger, et al., 1982; Levy, et al., 1982).

The spacer regions show little homology. Heteroduplex analysis of the differences between the "early" sea urchin histone genes from *Lytechinus pictus* and *Stronglyocentrotus purpuratus* indicated "pictorially" that, while coding regions share homology, the spacer regions are essentially nonhomologous (Cohn and Kedes, 1979). In the early sea urchin histone genes however, the lack of great variability in the length of the spacer regions and the observation of certain "motifs" (small base sequence homologies) indicate that there are selection pressures on the spacer sequences, as particular stretches may serve regulatory or structural

roles (Wells, 1986; Hentschel, et al., 1981). In the late histone genes, it is a little premature to talk about the length of the spacer regions. However, certain homologous motifs are again observed.

Early histone gene variant clusters also occur in sea urchins. Minor length and sequence heterogeneity exist within and between individuals in the principle gene cluster. The areas of nonhomology occur mainly in the spacer regions as has been shown by restriction enzyme analysis (Overton and Weinberg, 1976), electron microscopy (Cohn and Kedes, 1979) and sequencing (Yager, et al., 1984). Base substitution, all of which are silent codon changes, have also been detected in the H4 gene of the early histone gene clusters isolated from different individuals belonging to the species *Stronglyocentrotus purpuratus* (Yager, et al., 1984). In the late histone genes, it appears as if there are extensive differences in DNA sequence of both spacer and regions between individuals of the same species, and possibly within a particular individual (Maxson, et al., 1983).

Although sequence polymorphisms are found in histone genes of different sea urchin individuals of a particular species, the linked repeats of a cluster are, for all intents and purposes, homogeneous (Yager, et al., 1984; Overton and Weinberg, 1979). Despite the uniformity within a cluster however, variation between the repeats of different individuals can be as high as 3% of the nucleotide positions. The percentage divergence (substitutions plus deletions) of spacer DNA (H2B / H3 spacers compared) between different species of sea urchins is about 11% (except in the case of H14 isolated from *Psammechinus miliaris* and Nor 5 isolated from *Stronglyocentrotus drobochiensis*, which has a

percentage divergence of 1,3%, which is unusual (Busslinger, et al., 1982)). The variation, based on spacer divergence, found within a particular species, can therefore be about one-fourth of that found between two species.

The thermal stabilities of hybrids formed between late histone mRNAs and early histone genes are substantially lower than those of the homologous early mRNA early gene hybrids (Childs, et al., 1979). However, the coding sequences of each histone gene family (early and late) are more homologous to each other than they are to members of other histone gene families (Yager, et al., 1984; Knowles and Childs, 1986).

Electron microscopy has proved to be an invaluable tool in looking at the organisation of the early sea urchin histone genes. It is not surprising therefore, that this technique came to play a major role in the physical mapping of an early sea urchin histone gene isolated from *P. angulosus*.

THE CHARACTERIZATION OF H27

2.1 INTRODUCTION

H27, an early sea urchin histone gene battery supposedly isolated from *P. angulosus*, appeared to be a very interesting early sea urchin histone gene clone. Its restriction enzyme map is identical to H22, a histone gene battery isolated from *P. miliaris*. The regions sequenced also failed to show any differences between the two clones. Because electron microscopy allows visualisation of the gross topology of DNA, it was decided to study H27 by partial denaturation mapping and heteroduplex analysis.

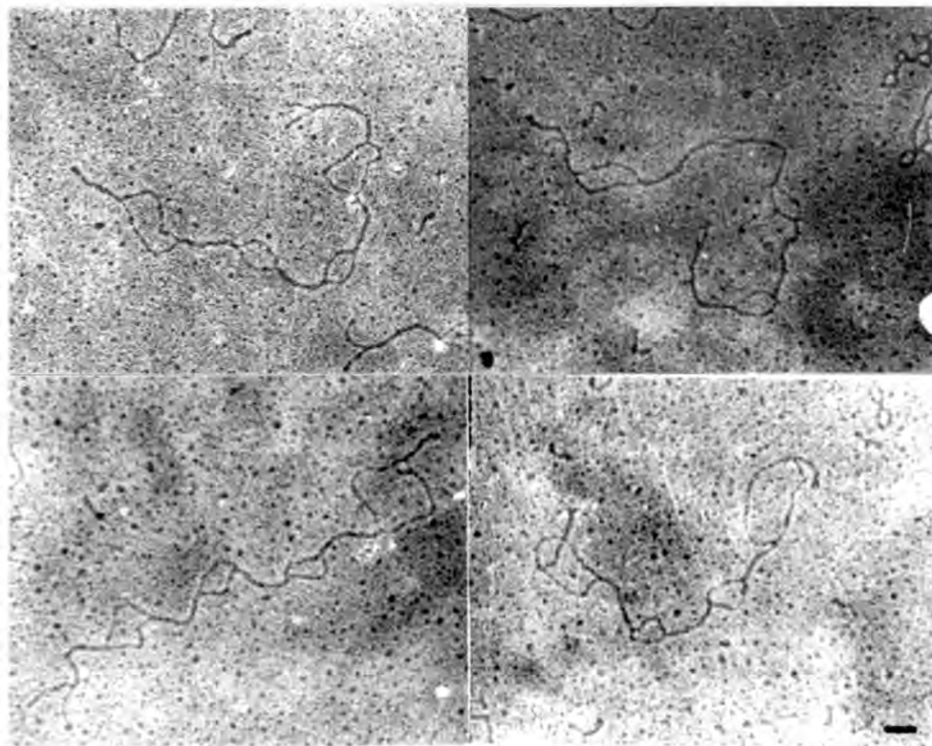
2.2 PARTIAL DENATURATION MAPPING

The chosen denaturation procedure was that introduced by Inman and Schnös in 1970, that is, alkali denaturation. This method is believed to give the best differentiation of AT and GC rich regions (Gomez and Lang, 1972; Coggins (b), 1987; Brack, 1981).

Partial denaturation of H27 and H22 was carried out at pH 10.45 for 45 minutes. Molecules, which showed a clear partial denaturation pattern were photographed at a magnification of 12 000 on a Zeiss EM 109 or 20 000 on a Jeol 200CX. PUC8 had been added as a double stranded marker. Assuming fragmentation of the chosen molecule had not occured, the lengths of experimental molecules were converted to bp by knowledge of the total length (6428 bp for both H22 and H27)

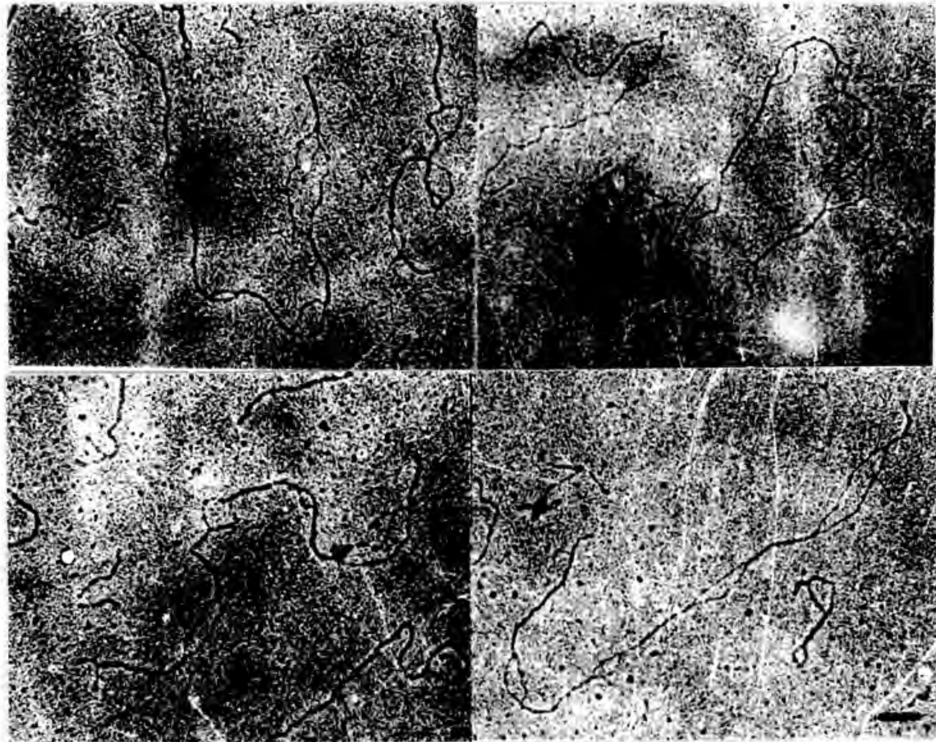
of the experimental molecules and size (2.5kb) of the standard DNA.

Both H27 and H22 showed a similar, distinctive partial denaturation pattern which allowed easy orientation of the molecules (Fig. 1 and 2). H22 had been subjected to partial denaturation analysis before (Portmann, et al., 1970) and thus the partial denaturation pattern of this particular molecule is known (Fig. 3). H22 has its longest spacer region between the H4 and H2B genes, resulting in a "large" bubble. This is followed by a series of three segments of similar size. A relatively large unmelted region, terminating, often, in an open fork structure is then observed.



bar=100nm

Fig 1.
H22 partially denatured at pH 10.45 for 45 minutes



bar=100nm

Fig 2:
H27 partially denatured at pH 10.45 for 45 minutes

H27 has a very similar, if not identical denaturation pattern. The histogram average denaturation maps of H22 and H27 are shown in 4. In figure 4, it is noticed that H27, with inverted ordinate, fits the upper map, that is of H22, almost identically.

Fig 3:
A summary of the physical characteristics of H22. The boxes above the line indicate the position of denatured areas, whereas those below the line indicate the position of the structural genes. The approximate position of the denatured regions as determined in this thesis (shown above) correlates well with that determined by Portmann et al., 1976.

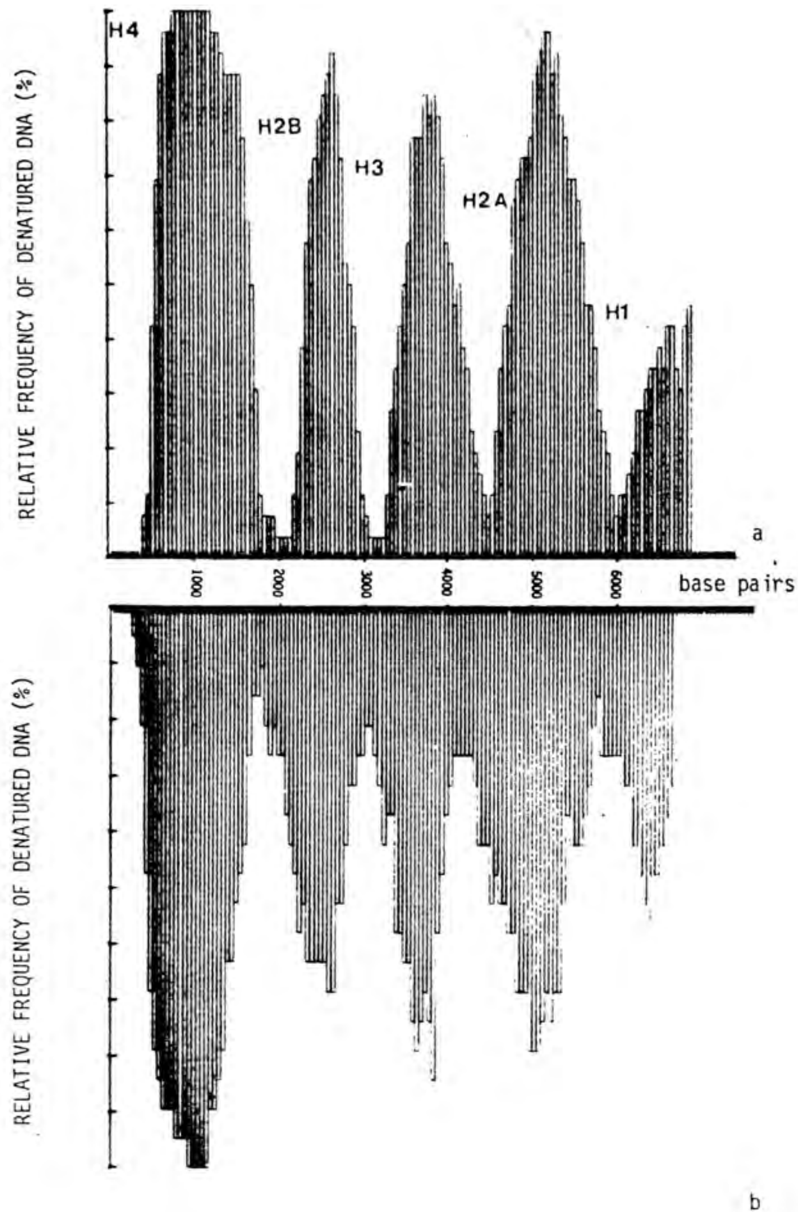


Fig 4:
The weight average histograms of a) H22 and b) H27, both denatured at pH 10.45 for 45 minutes.
20 molecules of H22 were measured.
20 molecules of H27 were measured.
The histogram of H27 has been inverted, so as to indicate the similarity in the denaturation pattern of H22 and H27.

The average length of the DNA in the "bubbles" and double-stranded regions, given in table 1 , was calculated from the data used for the histograms shown in figure 4, and which was subjected to statistical analysis using SAS (section 5.2). A comparison was carried out between the sequence of the H22 histone gene battery and H22. The agreement is such that confidence can be placed in the values calculated for the AT and GC rich region of H22 using the SAS system.

If one takes into account the fluctuations that can occur between two denaturation experiments, despite using the same solutions, and trying to keep the conditions identical, then the values calculated for the AT and GC rich regions of H27 correlate well with that of H22.

	H4 (bp)	spacer (bp)	H2B (bp)	spacer (bp)	H3 (bp)	spacer (bp)	H2A (bp)	spacer (bp)	H1 (bp)	spacer (bp)
H22	467 ±60	1016 ±100	661 ±60	439 ±84	600 ±72	529 ±56	670 ±89	740 ±100	808 ±84	327 ±44
H27	460 ±114	1002 ±192	676 ±127	468 ±145	706 +146	430 ±96	798 ±135	727 ±143	827 ±150	303 ±68

Table 1:
The lengths in base pairs of the AT and GC rich regions of H22 and H27 as determined using SAS (section 5.2). The denaturation of the DNA was carried out at pH 10.45 for 45 minutes. The GC rich regions are identified by the structural histone genes that occur there. The AT rich areas are considered as the "spacers" between these gene regions.

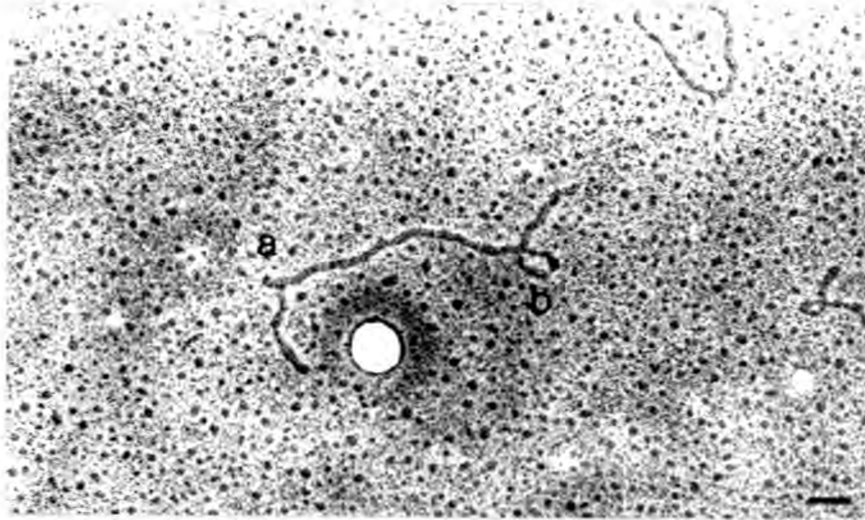
Denaturation mapping, however, is not sensitive enough to detect small differences between the two DNA sequences. It gives a gross indication of the AT rich regions in a specific segment of DNA. From the partial denaturation mapping, therefore, it can be said that H22 and H27 show a very similar pattern of AT and GC rich regions.

2.3 HETERODUPLEX ANALYSIS

If the DNA molecules to be compared by heteroduplex analysis are very closely related, it can be extremely difficult to determine which of the molecules seen on the grid are heteroduplexes, and which are homoduplexes. In order to overcome such a problem, H22 and H27 were subcloned into the two different, yet clearly related plasmids, pAT153 and pBR329. Deletion loops, corresponding to differences between these plasmids can be used to identify and orientate the heteroduplexed DNA molecules (Kinchington, et al., 1984).

The plasmid cloning vectors pAT153 (Twigg, et al., 1980) and pBR329 (Covarrubias and Bolivar, 1982) were each derived from pBR322. They differ from each other in two respects. Firstly, pBR329 has an additional sequence of 1151 nucleotide pairs which code for chloramphenical transacetylase. Secondly, the two plasmids have a non-identical deletion sequence which is present in pBR322, and is concerned with plasmid mobilization. The shorter deletion is in pAT153 and results in that plasmid having an additional 384 nucleotide pairs which are absent in pBR329, A heteroduplex of pAT153 and pBR329 is shown in figure 5. The plasmids had been linearized with Sal 1. The positions of the two deletion

loops can be clearly seen, and can be used as internal markers.



bar=100nm

Fig 5:
Hybrid of the plasmids PBR329 and PAT153,
showing the
a) deletion and
b) insertion loops.
Both plasmids were linearized with Sal 1.

The hybrids of H22 and H27 were easily detected due to the observable differences between pAT153 and pBR329. When spread at a formamide concentration of 50%, they appeared totally homologous, as is indicated in figure 6 a (i).

Areas that have mismatched base pairs, but that are closely homologous, may be detected if the heteroduplex molecule is parially denatured. In homoduplex DNA, bubbles will occur in regions rich in adenine and thymine, whereas in heteroduplexed DNA molecules, short sequences of mis-matched bases will contribute to a lowering of the melting temperature in localised regions and will thereby generate "bubbles" on partial denaturation (Kinchington, et al., 1984). A

comparison of denaturation thresholds between corresponding homo- and hetero- duplexes and the position of the "bubbles" produced should localise small regions of heterogenicity.

Formamide lowers the melting temperature (Tm) of DNA, 1% formamide reducing the Tm by 0.72°C (McConoughy, et al., 1969). It therefore can be used as a denaturant (Kinchington, et al., 1984; Davis and Hyman,1971). H22 and the heteroduplexes between H22 and H27 were spread at increasing formamide concentrations. The formamide concentration at which strand separation was first observed was the same in both cases (table 2). This indicated that H22 and H27 are, if not totally homologous, very, very closely homologous.

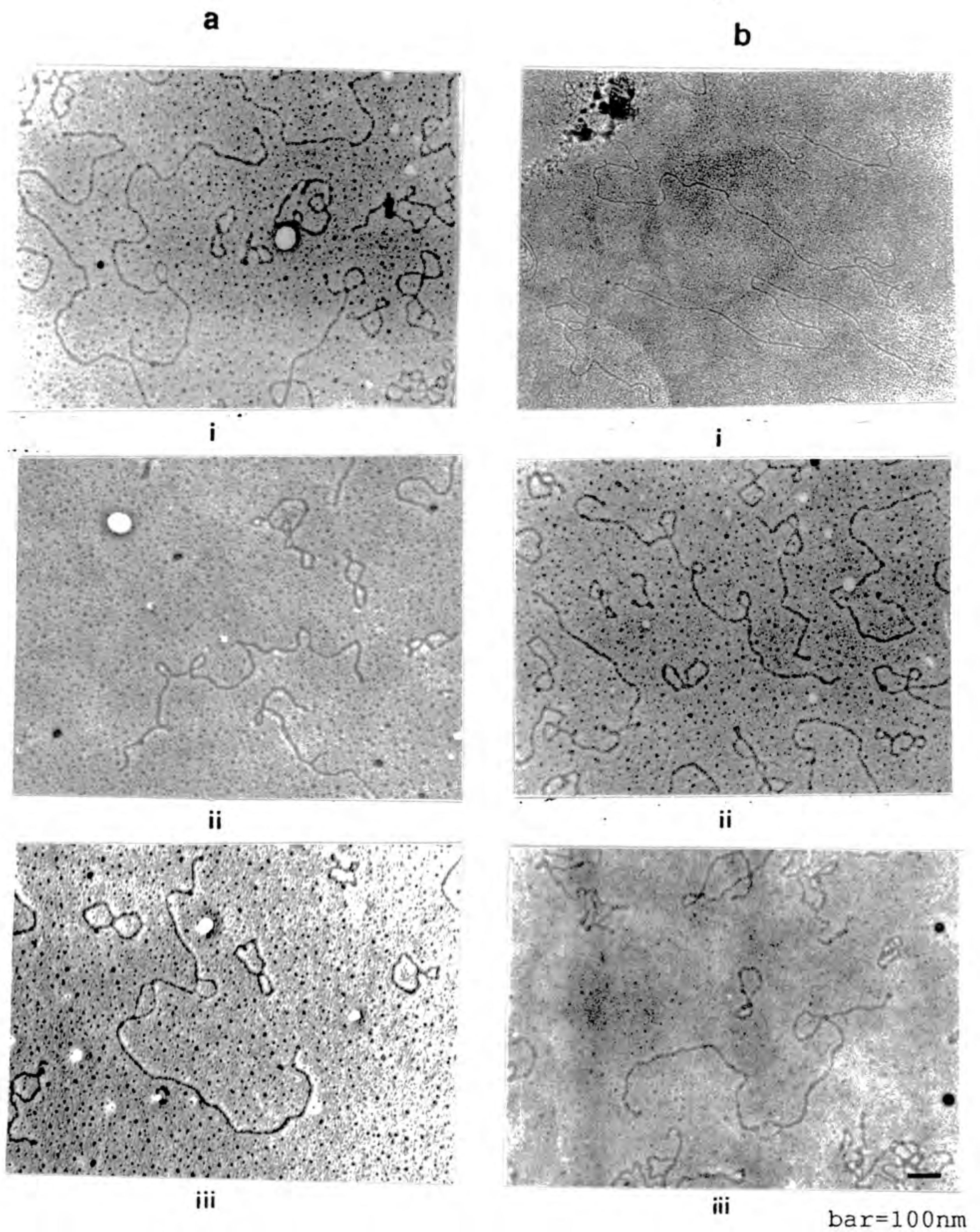


Fig 6:
a) Heteroduplex molecules formed by hybridizing H27 (cloned into PAT153) and H22 (cloned into PBR329). Both were cut with Sal 1.
i] 50% formamide
ii] 70% formamide
iii] 75% formamide

b) H22, cloned into PBR329, cut with Sal 1, and spread at increasing formamide concentrations.
i] 50% formamide
ii] 70% formamide
iii] 75% formamide

Insertion loop to bubble (bp)	Bubble (bp)	Bubble to end of molecule (bp)
772+96	759+166	5713+118

a

Beginning of molecule to bubble (bp)	Bubble (bp)	bubble to end of molecule (bp)
4443+99	745+130	5614+101

b

Table 2:
Comparison of heteroduplexed H22/H27 (a) and homoduplexed H22 (b) spread at a formamide concentration of 75%.

2.4 CONCLUSION

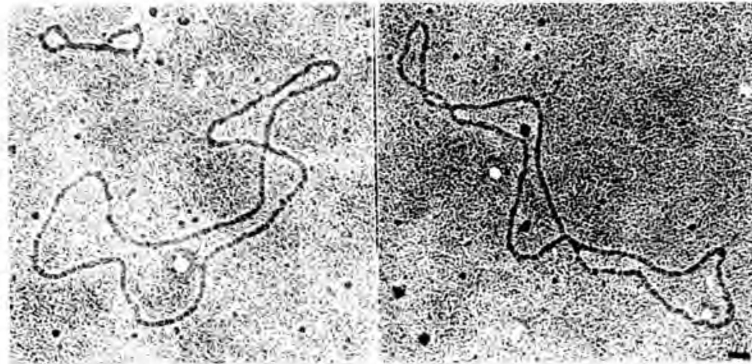
The close homology between H22 and H27, especially in the spacer regions is highly surprising. These two early sea urchin histone gene clones were supposed to have been derived from two different species of sea urchin, *Psammechinus miliaris* (H22) and *Parenchinus angulosus* (H27).

The most likely explanation for the lack of detectable differences between the two clones is that they are in fact the same. This may have occured by contamination during the isolation of H27 during which H22 was used as a probe.

The exercise of comparing H22 and H27, although it led to a rather disappointing conclusion did provide a useful testing ground for the techniques.

THE CHARACTERIZATION OF SU1

3.1 THE LENGTH OF SU1



bar=100nm

Fig 7:
The closed circular plasmid of SU1, consisting of a histone gene battery isolated from Parenchinus angulosus and cloned into pBR322. 22 such molecules were measured. The histone gene battery was determined to be 7.5±.2 Kb in length.

The chimeric plasmid SU1, consisting of an early histone gene battery isolated from *P. angulosus*, cloned into the Bam H1 site of pBR322, was prepared for EM using cytochrome C spreading. (fig.7). (Brack, 1981). SU1 was determined to be 7.5 ± .2kb long.

3.2 THE PARTIAL RESTRICTION MAP OF SU1

Because a restriction endonuclease recognises a specific base sequence in a DNA molecule, a particular restriction enzyme will generate a unique family of fragments for a particular DNA molecule. Knowledge of the position of the cuts in the DNA of interest - that is, the restriction map - allows easier manipulation of the DNA.

The molecular size of the restriction enzyme fragments of SU1 were determined by comparison with the Hind III - EcoR1 restriction fragments of Lambda DNA. Figure 8 shows some of the restriction enzyme digests that were used in the derivation of the restriction map for SU1. A reverse transparency was scanned using a densitometer, and the sizes of the restriction fragments determined. The partial restriction map of SU1, so derived, is shown in figure 9.

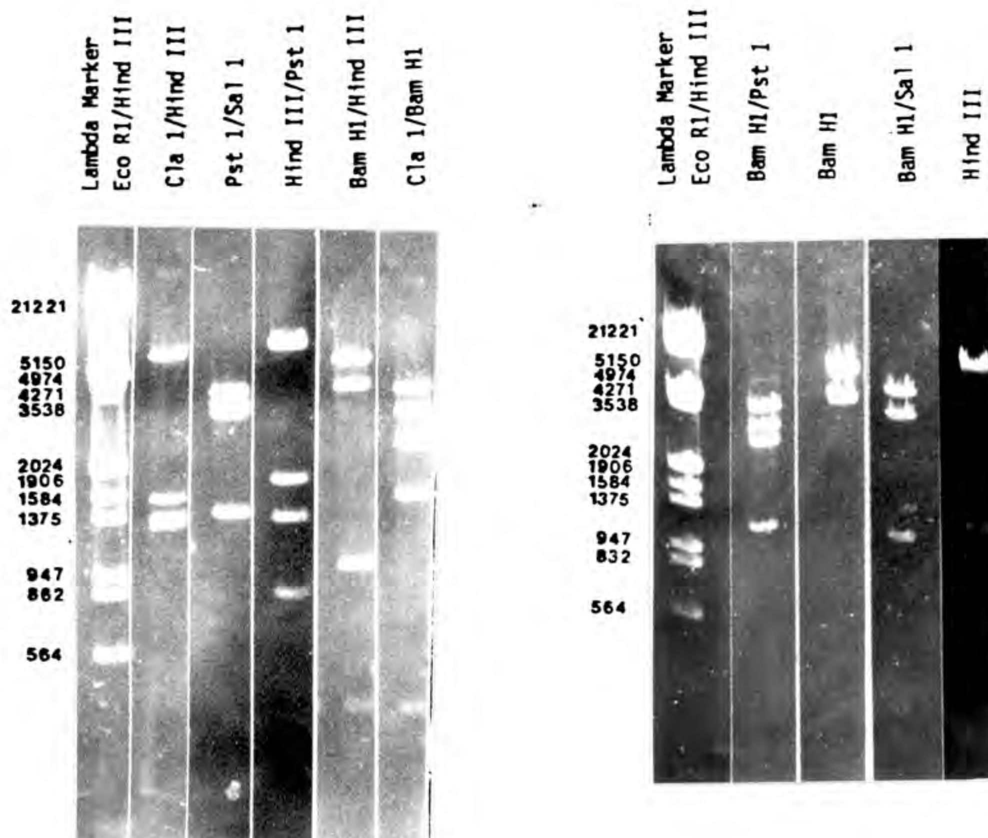


Fig 8:
Double restriction enzyme digests that were used in determining the partial restriction enzyme map of SU1.

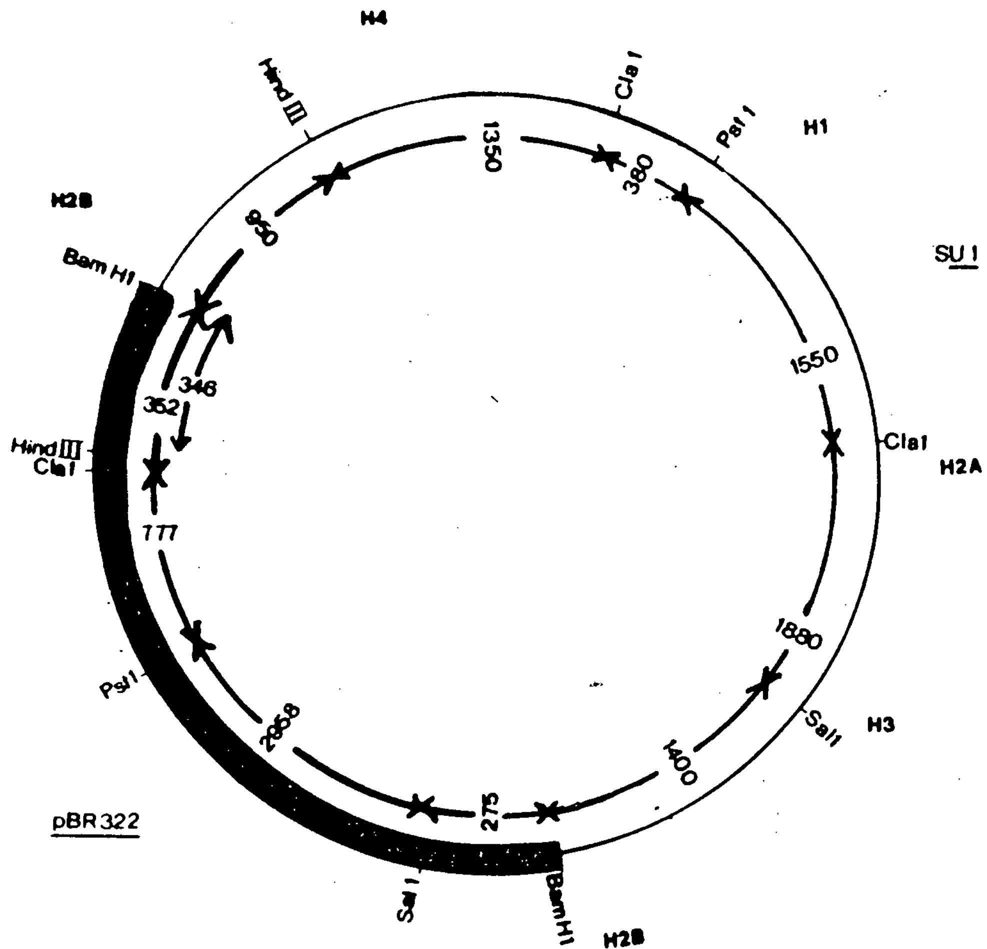


Fig 9:
The partial restriction enzyme map of SU1.
The restriction enzyme fragments on which the structural histone gene coding sequences lie, as determined in section 3.3, are also shown in the above diagram. The order of the genes is as in all other sea urchin species studied to-date.

3.3 THE DETERMINATION OF THE ORDER OF THE GENES IN SU1

3.3.1 Introduction

The method of choice in determining the order of the genes in SU1 was similar to that used by Cohn, et al., in 1979, where the order of two, non-allelic histone gene clusters of an individual sea urchin from *Lytechinus pictus* was determined by hybridization of ^{32}P rich labelled restriction fragments to known coding sequences from *Stronglylocentrotus purpuratus*. The ready availability of subcloned gene coding sequences of all except H2B of the H22 early histone gene battery from *Psammechinus miliaris* greatly eased this particular task.

3.3.2 The order of the genes in SU1

SU1 was subjected to double restriction enzyme cleavages, run on 1% agarose gels and blotted onto hybond N. These filters were then hybridized to one of the five ^{32}P nick-translated histone gene coding sequences obtained from H22. All, except the H2B probe, had been subcloned into pBR322 (figure 10). The histone-coding sequences were separated from pBR322, which would have given confusing results if still present, by low melting gel electrophoresis. The H2B coding sequence, which had not been subcloned, was prepared by cutting H22 with Pst 1 and Bam H1, and purifying the required fragment by low melting gel electrophoresis (figure 11).

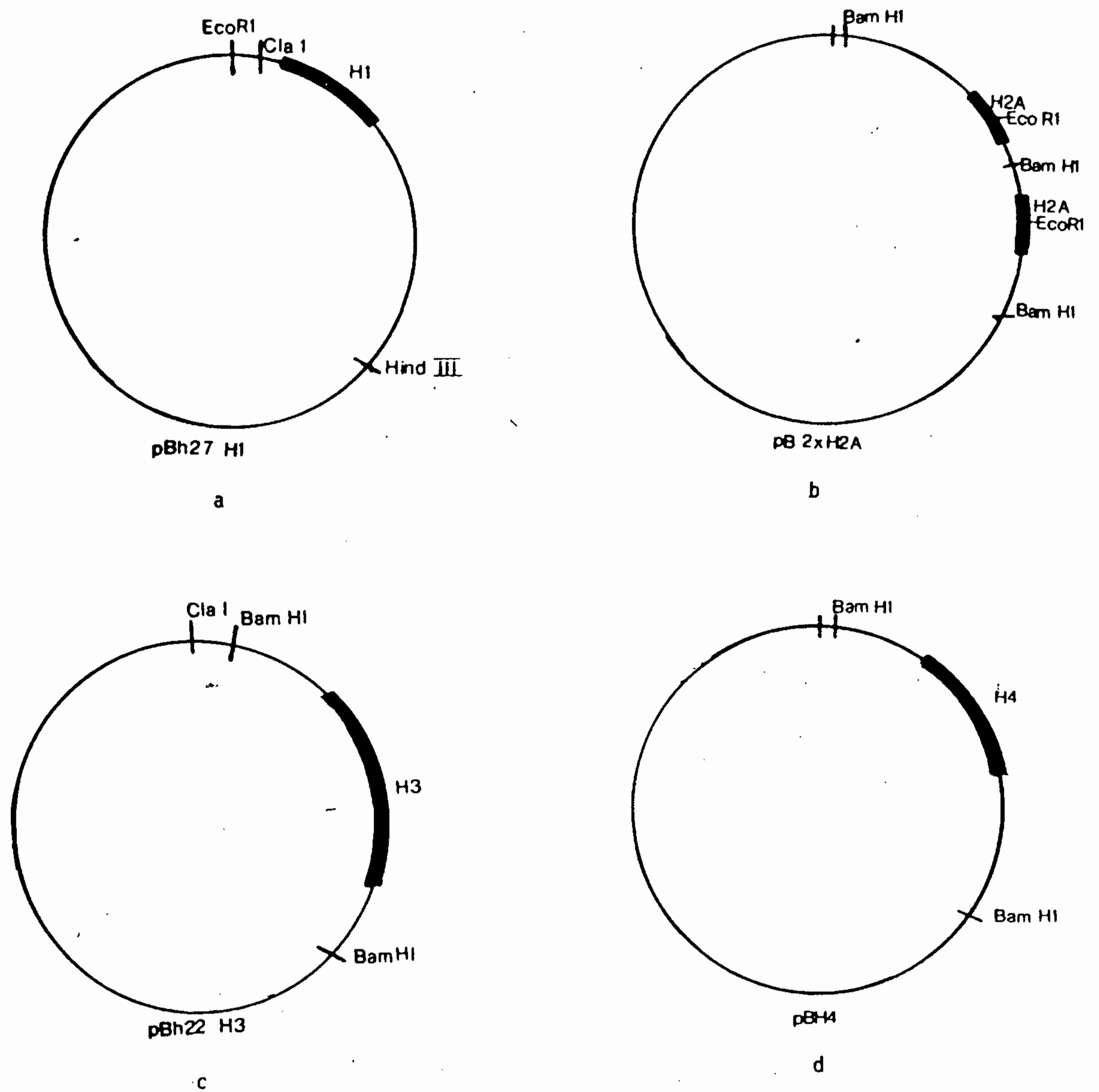
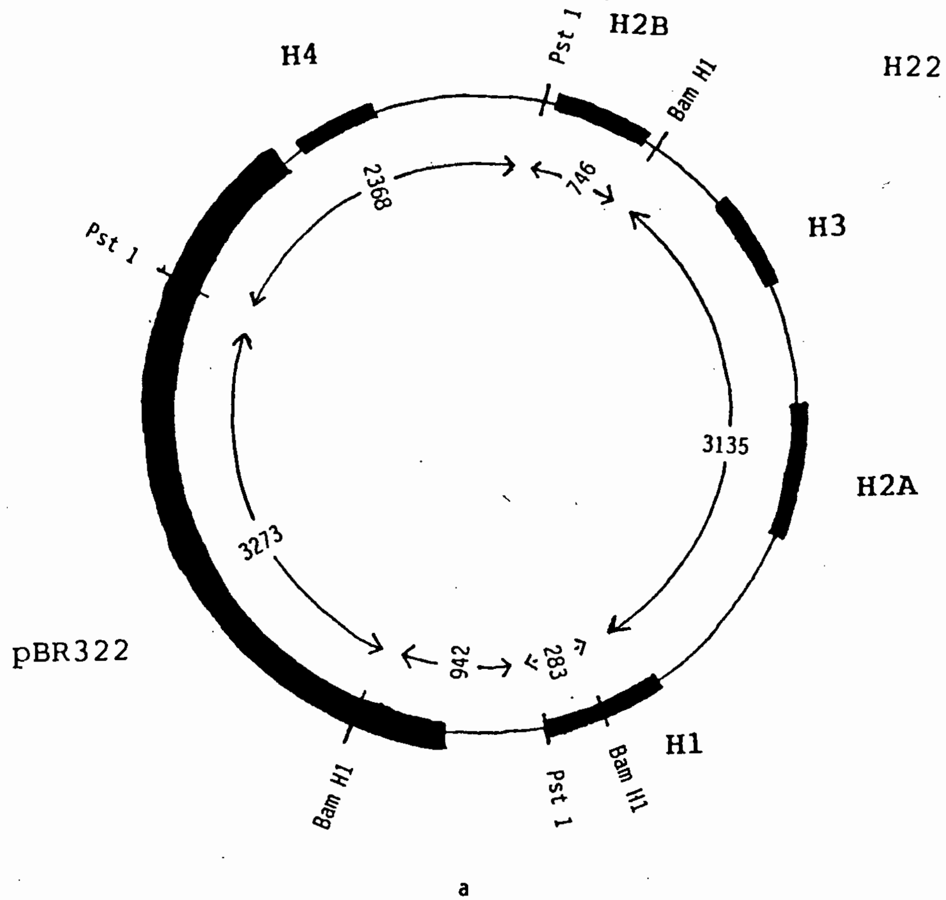


fig 10:
Schematic presentation of the plasmids used in the probing for the a) H1, b) H2A, c) H3 and d) H4 structural genes of SU1.



a

pBR322/H22

cut with Pst 1 and Bam H1

low melting gel electrophoresis

label with 32-Phosphorus

H2B PROBE

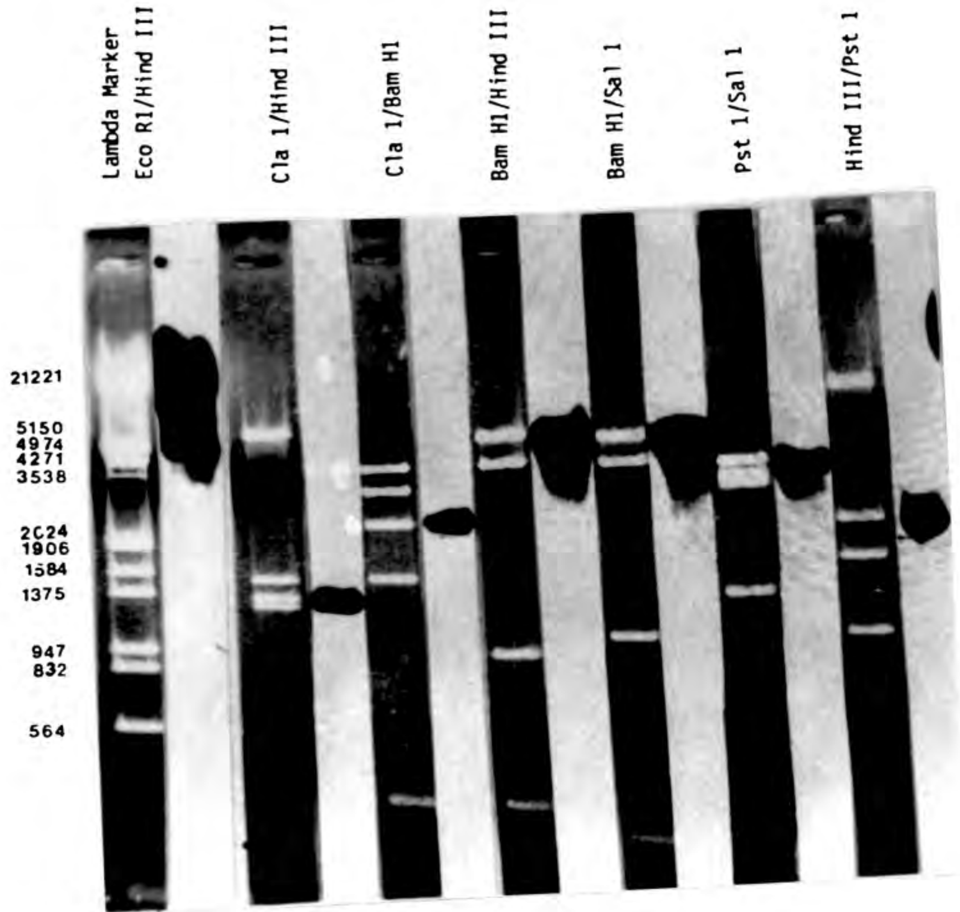
b

Fig 11:
a) Schematic presentation of the fragments that would result from cutting the chimeric plasmid pBR322/H22 with Pst 1 and Bam H1. The approximate positions of the histone gene coding sequences within H22 are indicated by thickened lines. The 746 base pair fragment containing the H2B gene is well resolved on a 1% agarose gel.

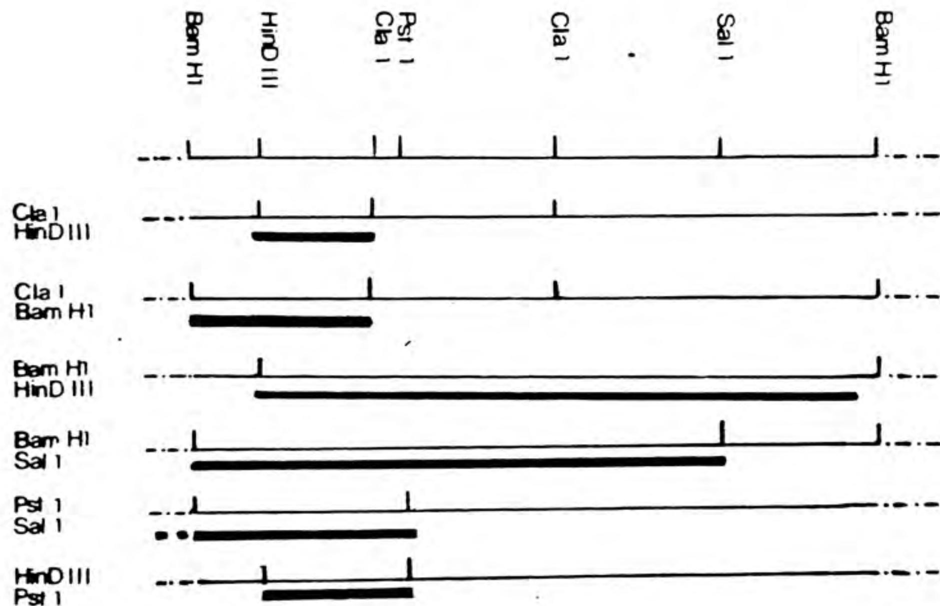
b) Flow diagram indicating as to how the H2B probe was prepared.

By considering the restriction fragments to which the radiolabelled probes hybridize, it is possible to determine the order of the histone genes in SU1. The H4 gene probe hybridizes to only one restriction fragment in each of the double digests. (figure 12). Futher, the digestions indicate that the H4 coding sequence is situated somewhere between the Hind III and the first Cla 1 cutting site of SU1. Similarily H1 lies between the Pst 1 and the second Cla 1 cutting site of SU1. (figure 13); H2A between the second Cla 1 cutting site and the Sal 1 cutting site. (fig 14). H3 appears to lie on the same DNA restriction fragment as H2A. However, on studying the autoradiograph, it appears that the Bam H1 /Sal 1 and Pst 1 /Sal 1 restriction digests have a second faint band as indicated in figure 15. These faint bands would result if Sal 1 cuts within the H3 coding region. However, the majority of the H3 coding sequence does indeed lie on the Cla 1/Sal 1 restriction fragment.

The hybridisation pattern observed for H2B in fig 16 can be explained by considering that Bam H1 actually cuts in the H2B coding sequence. This has been verified by partial sequencing of SU1 (J. Rees, Unpublished results). About 12 nucleotides occur before the H2B/H4 spacer. The majority of the H2B coding sequences follows the H3/H2B spacer region.



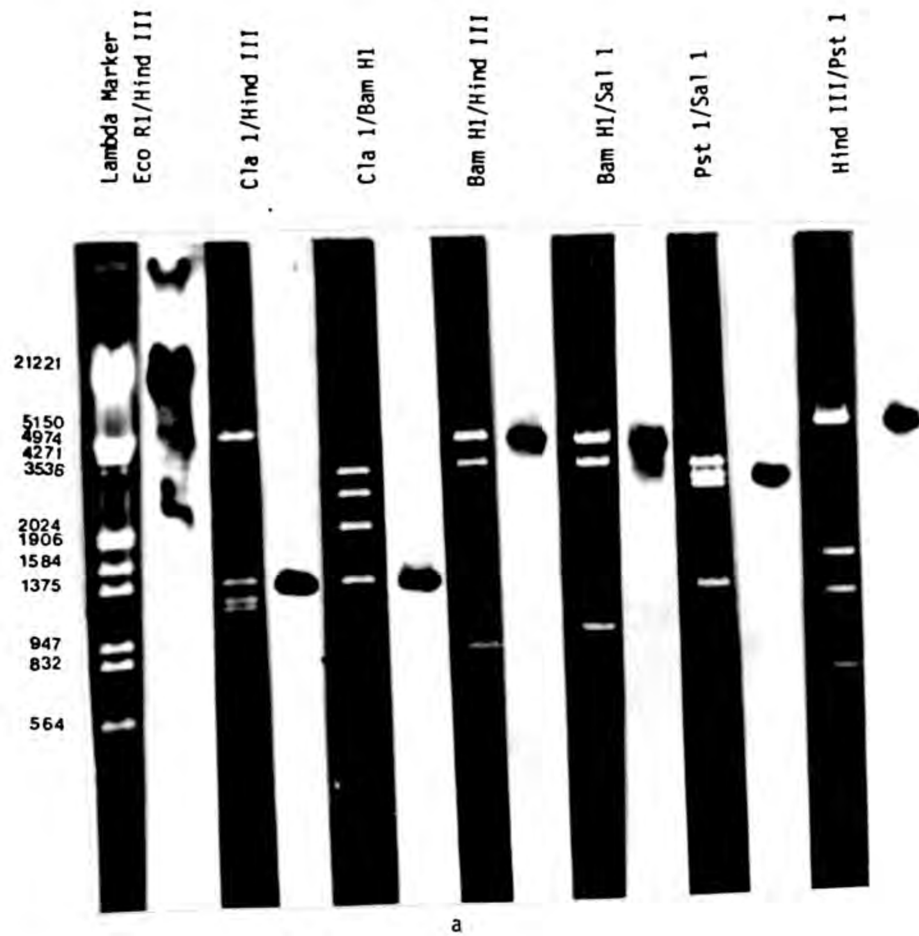
a



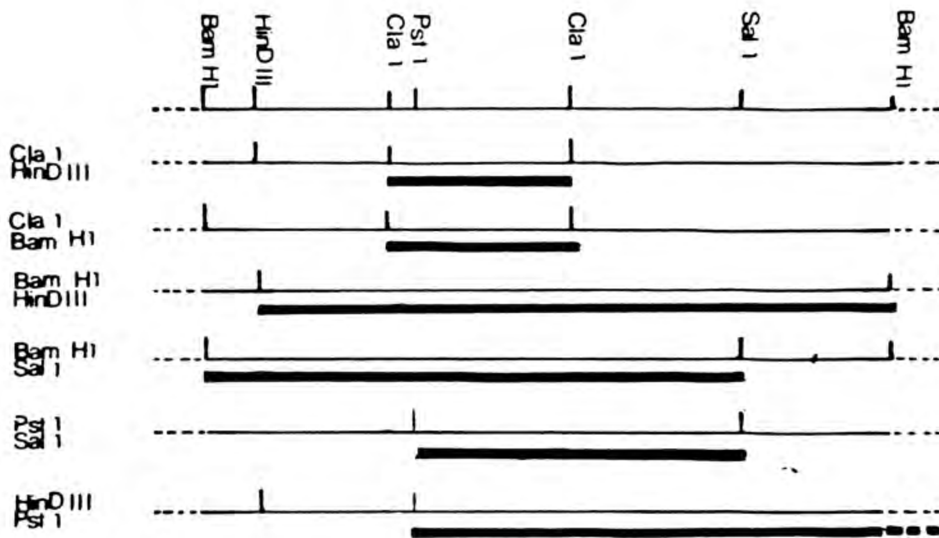
b

Fig 12:
a) The probing of restriction enzyme digests of SU1 to determine the position of the H4 structural gene coding sequence.

b) A schematic presentation of the restriction fragments to which the probe hybridizes. pBR322 is not shown as it is not important in the positioning of the histone gene coding sequence.



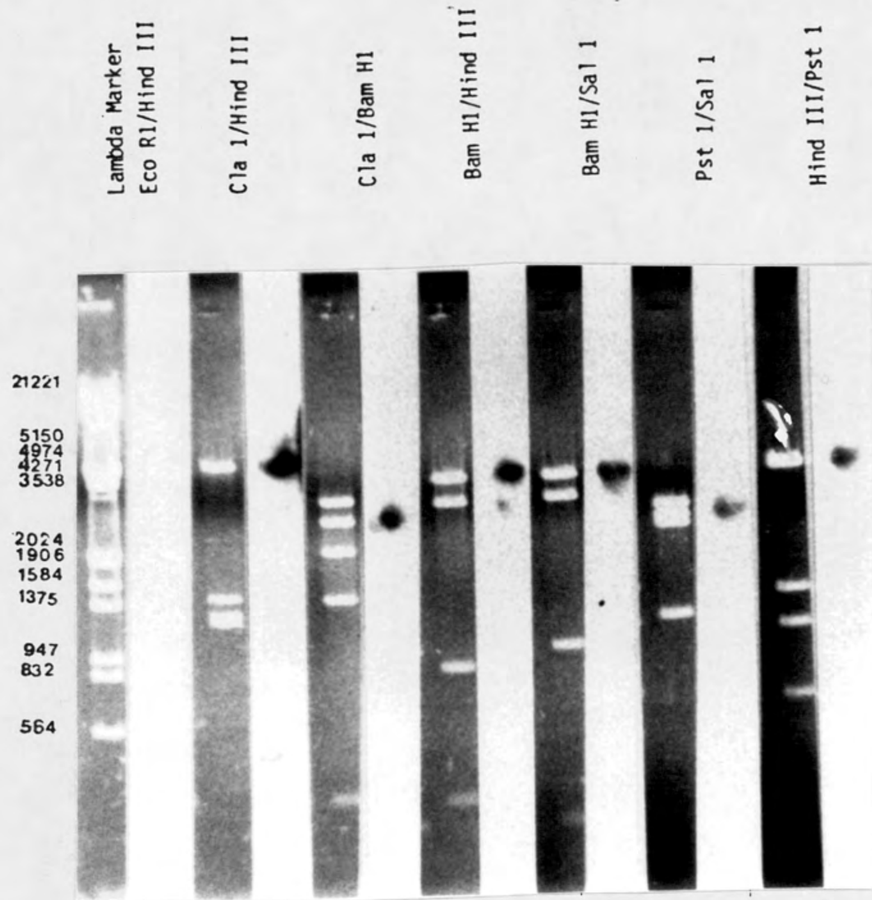
a



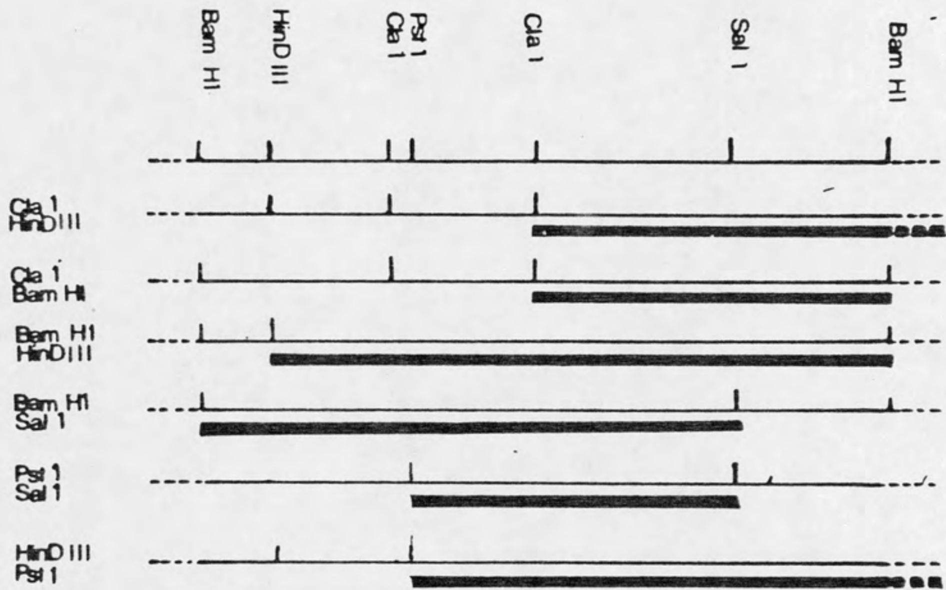
b

Fig 13:
a) The probing of restriction enzyme digests of SU1 to determine the position of the H1 structural gene coding sequence.

b) A schematic presentation of the restriction fragments to which the probe hybridizes. As in figure 12, pBR322 is not included in the diagram.



a

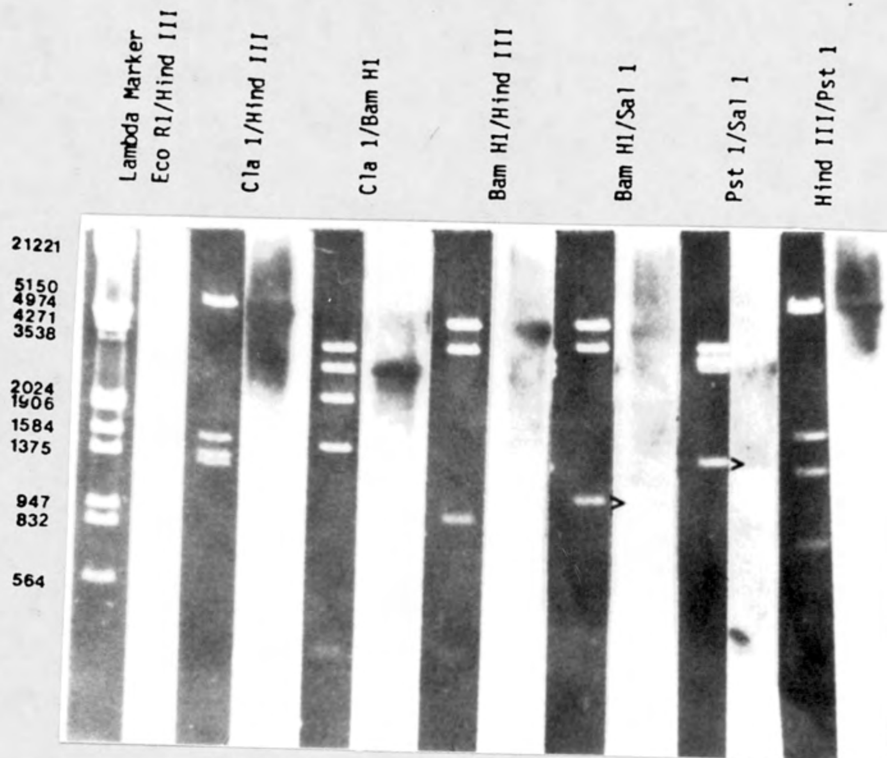


b

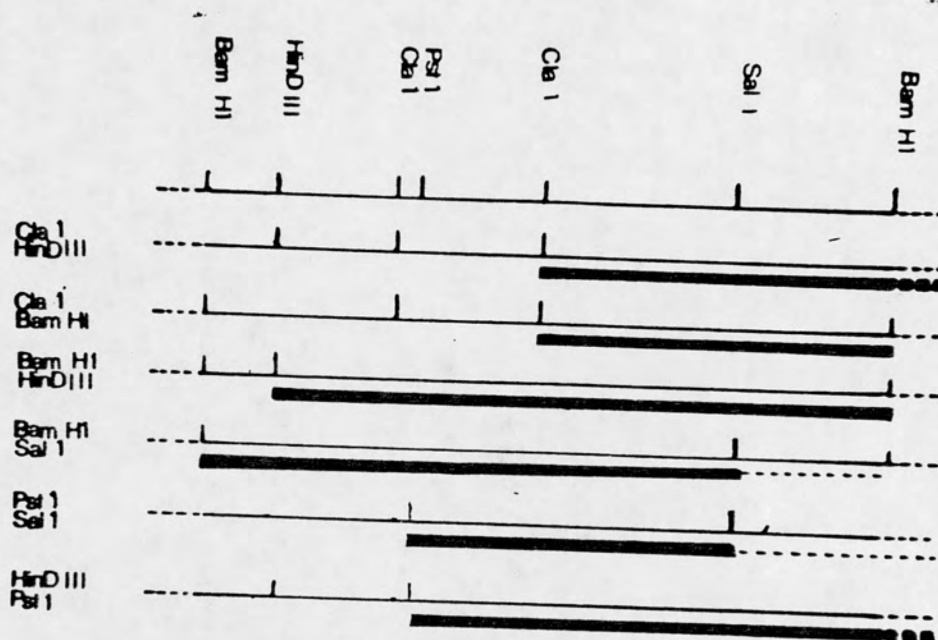
Fig 14:

a) The probing of restriction enzyme digests of SU1 to determine the position of the H2A structural gene coding sequence.

b) A schematic presentation of the restriction fragments to which the probe hybridizes. pBR322 is not considered in the above diagram.



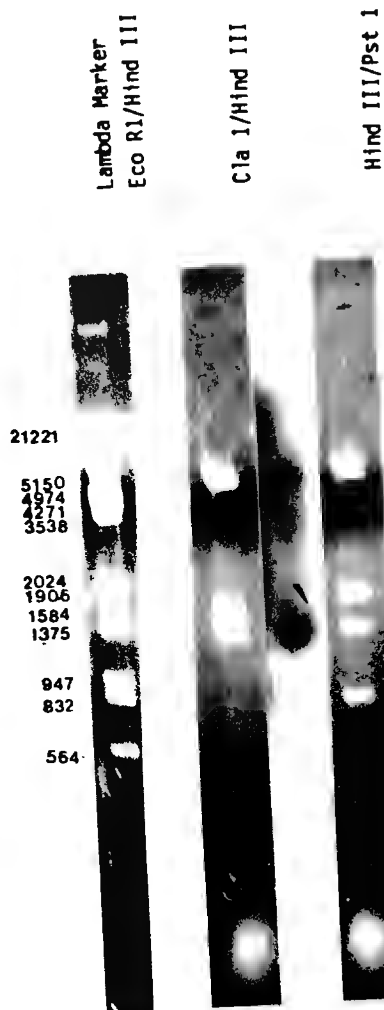
a



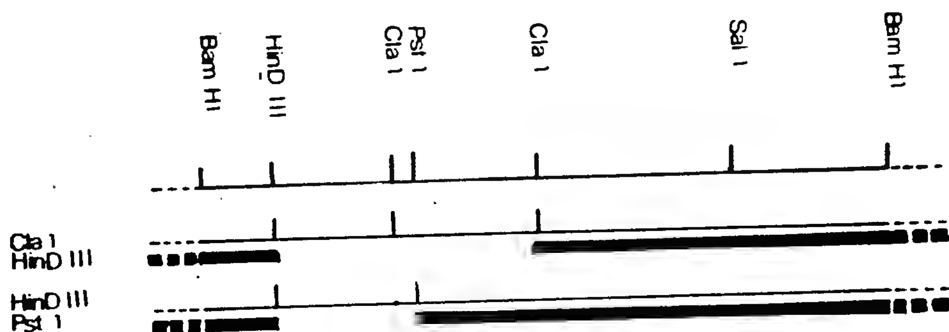
b

Fig 15:
a) The probing of restriction enzyme digests of SU1 to determine the position of H3 on SU1. In the Bam H1/Sal 1 and the Pst 1/Sal 1 digests, there appears to be a second faint band, as indicated by >.

b) A schematic presentation of the restriction fragments to which the probe hybridizes. The dashed lines indicate the fragments to which the probe hybridizes weakly. pBR322 is not considered in the above diagram.



a



b

Fig 16:
a) The probing of restriction enzyme digests of SU1 to determine the position of H2B on SU1.

b) A schematic presentation of the restriction fragments to which the probe hybridizes. Again pBR322 is not necessary in the above diagram.

The order of the genes in SU1 is thus:

H2B H4 H1 H2A H3 H2B

This is identical to that observed in all the other sea urchin species studied to date, assuming that the polarity and direction of transcription is the same as observed in other sea urchin species. This appears to be so from the partial sequencing of SU1 (J Rees, unpublished results).

In the early sea urchin histone gene batteries studied to date, it appears that the histone gene quintet is arranged in tandem. It can thus be concluded that the H2B portions of SU1 probably constitute the whole of the H2B gene coding sequence unless there are two or more Bam H1 cutting sites in the H2B gene of the particular clone isolated.

3.4 THE MELTING PROFILE OF SU1

The thermal denaturation temperature is related to the base composition of DNA. When DNA is heated in solution, a sharp increase in its extinction coefficient occurs at the temperature where DNA changes from being double-stranded to being denatured. The temperature, corresponding to the midpoint of the absorbance rise, the Tm, is linearly related to the average DNA base composition (Marmur and Doty, 1962). When thermal denaturation of DNA samples is carried out in 0.1 x SSC, (0.3 M Na citrate,3 M NaCl) the GC content for samples can be calculated from the equation:

$$GC = (Tm - 53.9)2.44$$ (Mandel and Marmur, 1968)

The early histone gene batteries from sea urchins consist of two major types of DNA - GC-rich coding, and AT-rich spacer DNA sequences (Portmann, et al., 1976). The melting profile of the histone DNA falls into early and late melting halves.

In H22 from *Psammechinus miliaris*, when heated in 0.1xSSC pH8, the DNA melts at 65.6, 68 and 74.8°C respectively, corresponding to a GC content of 34%, 40% and 52% (Portmann, et al., 1976).

The melting profile of SU1 (fig 17) shows the expected early and late melting halves. The melting temperatures in 0.1 x SSC pH8 are 69,6 and 71° for the early half and 78° for the late melting half. Using eqn (1), the GC content of SU1 was found to be 38,3%, 41.7% and 58.8% respectively. SU1 thus shows a higher GC content than H22.

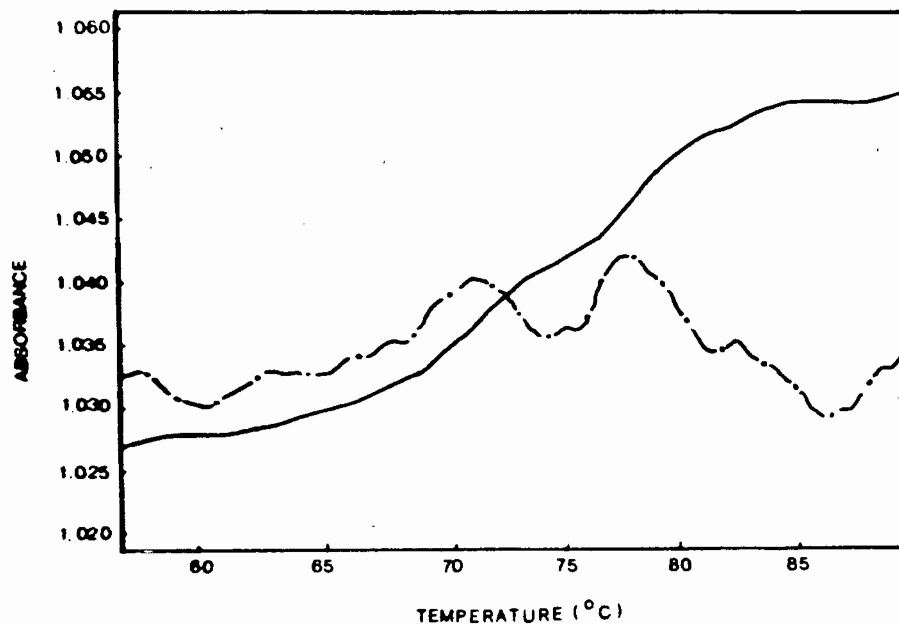
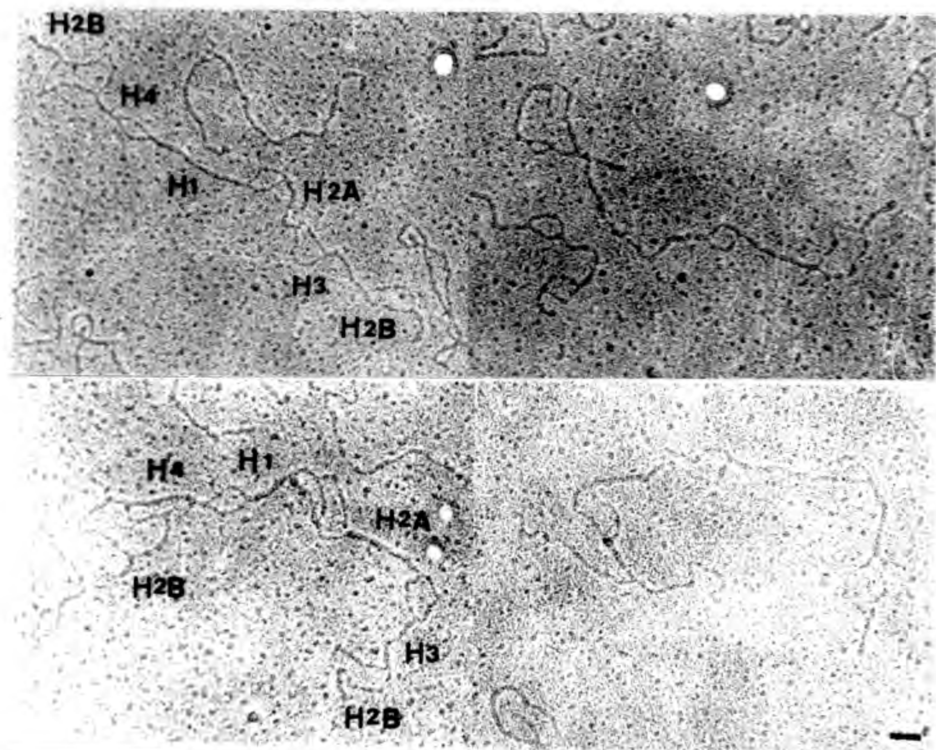


Fig 17:
The melting profile of SU1 (——). The Tm for each transition was determined by looking at the first derivative of absorbance versus temperature (-·-·).

3.5 PARTIAL DENATURATION MAPPING

Partial denaturation of SU1 and H22 was carried out using the method of Inman and Schnös (1970). Electron micrographs of the partially denatured SU1 are shown in figure 18. The expected denaturation map is observed.



bar=100nm

Fig 18:
SU1 partially denatured at pH 10.45 for 45 minutes

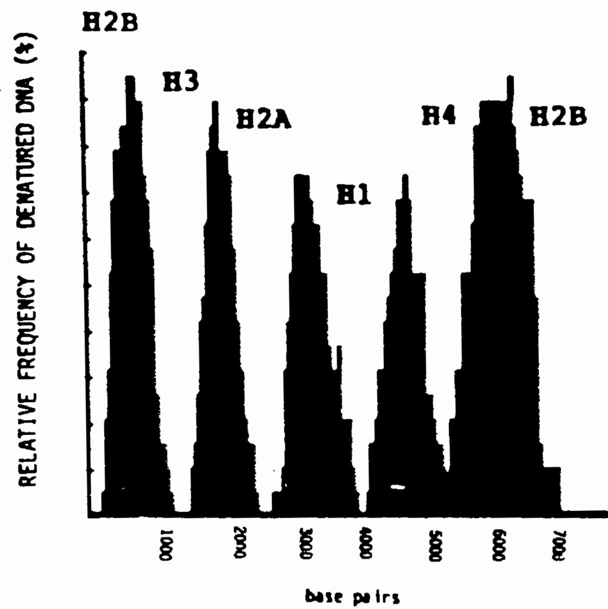
The orientation of the denatured molecules of H22 has already been discussed in section 2.2.

If SU1 has very similar characteristics as H22, then the spacers between the genes, the order of which had been determined in section 3.3, should be AT rich. SU1 showed the expected denaturation pattern. There is a large "bubble" towards one side of the molecule, preceded by a short double stranded region. In some cases, a large open fork structure is noticed where this short double stranded region has been totally denatured (fig 18). The other end of the molecule terminates in a double stranded region, similar in size to the other double stranded regions observed. This distinctive denaturation pattern allowed easy orientation of SU1.

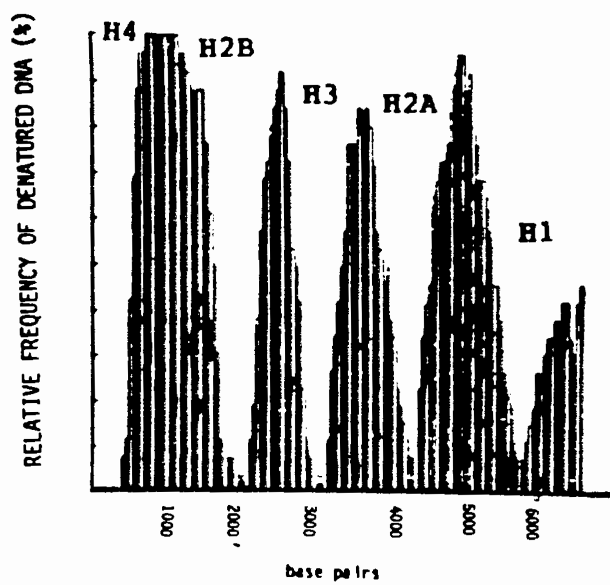
Both single (M13) and double stranded (PUC8) standard DNA were included in the spreading mixture of denatured SU1. The lengths of the single- and double-stranded regions of the experimental molecule were measured, and these lengths were converted into base pairs by comparison with the appropriate marker DNA. The molecules were then aligned and the weight average histogram plotted (fig. 19).

The ordering of the genes and the spacers along SU1 is expected to be as indicated in figures 18 and 19. The five genes coding for the five histone proteins alternate with AT-rich spacer. From the partial sequencing of SU1, it is known that the majority of the H2B gene is situated following the H2B/H3 spacer. By taking this into account, the ordering of the genes and spacers along the partially denatured SU1 was worked out. The largest "bubbles" coincide with the H4/H2B spacer, and the longest undenatured region corresponds to that portion of the DNA sequence where the H1 coding region is expected to occur. Thus, the partial denaturation map of SU1

strongly resembles that of H22, which has been redrawn in fig. 19 (b).



a



b

Fig 19:
The weight average histograms of
a) SU1 denatured at pH 10.45 for 45 minutes
(19 molecules measures)
b) H22 redrawn from figure 3.

Table 3 shows a comparison of the AT and GC rich regions of SU1 and H22, corresponding to the genes and spacer regions. In general, the AT rich regions of SU1 are longer than the corresponding AT rich regions of H22, while the GC- rich regions are of similar length.

H2B (bp)	spacer (bp)	H3 (bp)	spacer (bp)	H2A (bp)	spacer (bp)	H1 (bp)	spacer (bp)	H4 (bp)	spacer (bp)	H2B (bp)
301 ± 47	694 ± 94	664 ± 105	651 ± 85	713 ± 180	749 ± 179	823 ± 165	539 ± 94	570 ± 108	1067 ± 254	319 ± 156

a

H4 (bp)	spacer (bp)	H2B (bp)	spacer (bp)	H3 (bp)	spacer (bp)	H2A (bp)	spacer (bp)	H1 (bp)	spacer (bp)
467 ± 60	1016 ± 100	661 ± 60	439 ± 84	600 ± 72	529 ± 56	670 ± 89	740 ± 100	808 ± 84	327 ± 44

b

Table 3:
a) The lengths in base pairs of the AT and GC rich regions of the histone gene battery insert of SU1. The values were determined for molecules denatured at pH 10.45 for 45 minutes.

b) The lengths quoted for H22 are from Table 1. The GC rich regions are labelled by the structural histone genes that occur there. The AT rich regions are identified as the "spacers" between these histone coding regions.

3.6 HETERODUPLEX ANALYSIS

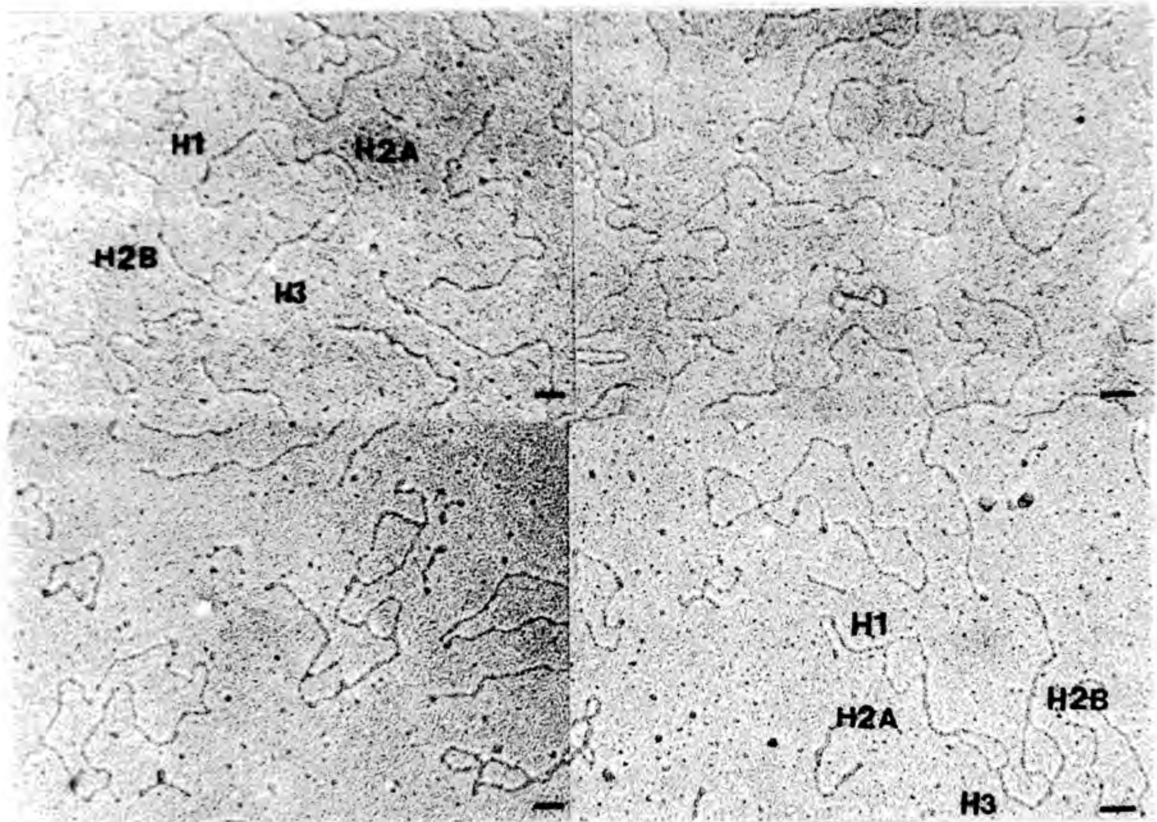
The question as to whether two nucleic acid strands have complementary sequence can be answered by hybridization experiments. Heteroduplex analysis is such a type of experiment. The regions of homology and nonhomology are recognisably different in suitable electron microscope preparations.

The DNA sequence of an early histone gene battery isolated from the sea urchin *Lytechinus pictus* has been compared to that of *S. purpuratus* by heteroduplex analysis (Cohn and Kedes, 1979). The spacer DNA segments shared little sequence homology with *L. pictus*; however, the analogous coding (and possibly flanking) regions have conserved their sequences. Similar results are obtained in the heteroduplexes that are observed when SU1 is hybridised against H22.

Hybridisation of the H22 and SU1 clones could not be carried out directly. H22 had been cloned into the Hind III site of pBR322 and thus the order of the genes relative to the origin of pBR322 is H4, H2B, H3 and H2A and H1 (Schaffner, et al., 1976). SU1 had been cloned into the Bam H1 site of pBR322 and the order of the genes relative to the origin of pBR322 is part of the H2B, H4, H1, H2A, H3 and the remainder of the H2B gene. Also, while the H22 containing chimeric plasmid could be linearized by cutting pBR322 with a restriction enzyme such as Sal 1, SU1 could not be as easily linearised. At this stage a restriction enzyme that cuts only in the pBR322, but not in the SU1 histone gene battery is unknown.

An Eco R1/Hind III double digest of H22 results in a fragment containing the H2B, H3, H2A and H1 coding sequences. A Pst 1/Bam H1 double digest of SU1 results in a similar fragment, but with a small section of the H2B gene missing due

to Bam H1 cutting within the H2B coding sequence. These two fragments were hybridised against each other, and the resulting heteroduplexes are shown in figure 20.



bar=100nm

Fig 20:
Heteroduplex DNA formed by annealing part of H22 and part of SU1.(H4 and the relevant spacer regions were excluded). The genes show considerable homology, whereas the spacer regions are essentially nonhomologous. M13 was included as a single strand marker and pUC8 as a double strand marker.

The hybrids showed four double stranded regions separated by "bubbles". Orientation of the hybrid molecules was difficult as the nonhomologous regions were not obviously asymmetrical. An attempt was made to align and orientate the molecules according to maximum overlap with the same orientation of the nonhomologous "bubbles". This however, did not achieve satisfactory orientation of all the hybrid molecules, as it appeared as if some molecules could be orientated in either direction.

The above problem was thus approached as follows. Molecules, which showed a distinct and similar heteroduplex pattern, were aligned and orientated, and the values for the single and double stranded regions worked out. These values were then stored and the remainder of the heteroduplex molecules orientated by looking at the Euclidean sum of each hybrid molecule orientated in either direction. The Euclidean sum is defined as:

$$\Sigma(x_i - x)^2$$

where x_i is the value of each single and double stranded region of the hybrid molecule in question, and x the values of the stored single and double stranded regions. The orientation that showed the lowest Euclidean sum from the stored values was chosen.

The measurements of the heteroduplexes were than subjected to statistical analysis using the SAS package (section 5.2). The double stranded regions corresponding to the H2B and H1 coding sequences were not involved in the analysis because of the difficulty in determining the start of the double stranded regions from the single stranded at the beginning and end of the heteroduplexed molecules. There is expected to be a positive correlation between each of the single strands constituting a "bubble" i.e. as the one single strand increased in length so does the other. A negative correlation

is expected between the strands constituting the "bubbles" and the homologous regions i.e. as the "bubble" increases in size, the homologous regions decrease, and vice versa. Molecules, that obviously disturbed the expected correlation matrix were either removed if the values differed to a noticeable extent from the majority of measurements, or reorientated as required (fig 21).

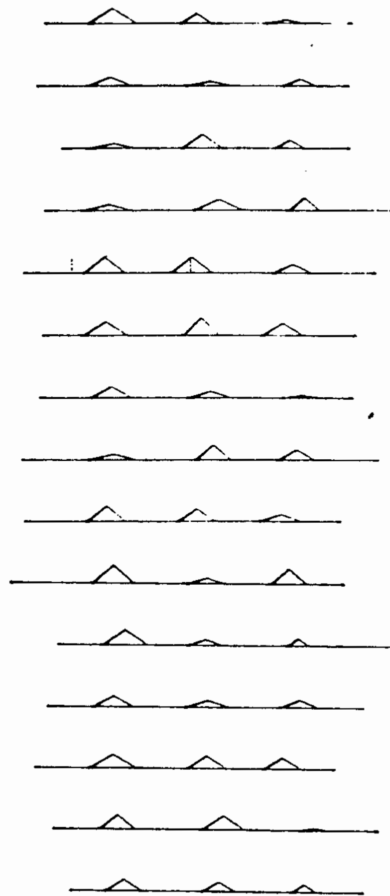


Fig 21:
The alignment of the heteroduplexed molecules. The orientation of the molecules was determined by using the SAS package and by considering the Euclidean sum of each orientation. The triangles indicate the position and the relative sizes of the nonhomologous bubbles. (section 5.3)

It was not immediately obvious as to which genes and which spacers corresponded to which double stranded and which single stranded regions of the hybrid molecule. One factor that aided in the ordering of genes and spacers along the hybrid molecule was that the cutting of H22 with Hind III and Eco R1 results in about 800bp of the H4/H2B spacer and the whole of the H2B gene being present to one end of the molecule and about 500bp of the H1/H4 spacer being present to the other end of the molecule. If one considers that the whole of the H1 gene of SU1, plus possibly some of the spacer, is present in the SU1 fragment used in the heteroduplex experiments - then in at least some of the hybrid molecules, a single strand of about 800 to 900 bp should be observed to one end, and a single strand of about 500 bp or less to the other end. Such molecules were observed, and thus the heteroduplexes ordered as to genes and spacer regions, as indicated in fig 22.

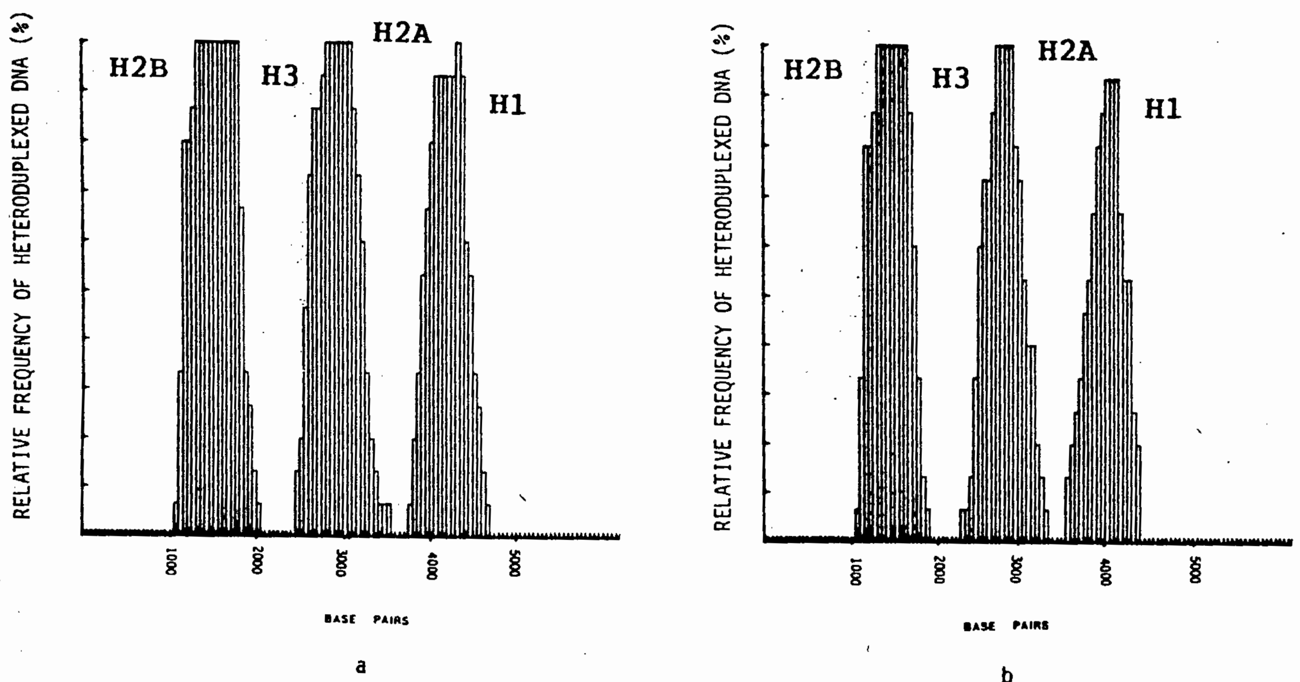


Fig 22:
Weight average histogram of the heteroduplexed DNA. The molecules used in drawing the histograms were first aligned and orientated as shown in fig. 21.
a) The larger of the two single strands considered.
b) The smaller of the two single strands considered.

It must be remembered, however, that it is difficult to prevent the nicking of DNA, and the more processes the DNA undergoes in preparation for heteroduplex analysis, the greater the chance of nicking (Kudler, et al., 1983). The majority of heteroduplexes, therefore, did not show an unambiguous orientation, as has already been discussed. It is therefore possible that some of the heteroduplexes have been orientated in the incorrect direction. Also, due to the difficulty of determining the end of a single strand and the beginning of a double stranded region, or vice versa, at the beginning and/or end of a heteroduplexed molecule in many cases, the values quoted for H2B and H1 in Table 4 are subject to error.

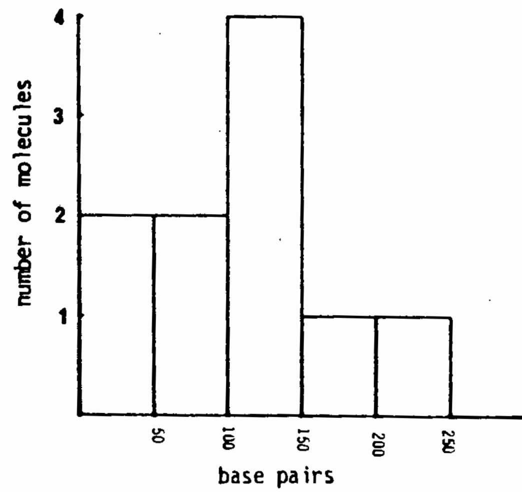
H2B (bp)	H2B/H3 spacer (bp)	H3 (bp)	H3/H2A spacer (bp)	H2A (bp)	H2A/H1 spacer (bp)	H1 (bp)
	718 $\pm$57		655 $\pm$78		550 $\pm$51	
685 $\pm$188		724 $\pm$102		687 $\pm$78		664 $\pm$214
	589 $\pm$57		537 $\pm$60		466 $\pm$54	

Table 4:
The lengths calculated for the double stranded regions and for the single strands constituting each bubble of the hybrids that were produced by annealing together part of SU1 containing the H1, H2A, H3 and part of the H2B structural gene coding sequences, and part of H22 produced by cutting pBR322/H22 with Eco R1 and Hind III. The data was derived from the use of both SAS and the weight average histograms. The homologous regions are identified by the structural histone genes that occur there. The nonhomologous regions are identified as the relevant spacer regions.

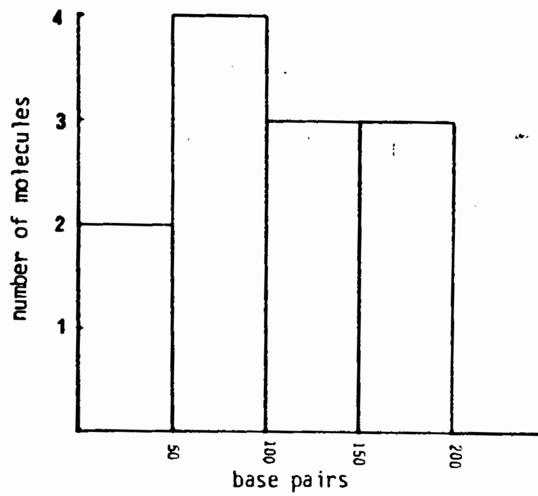
The initial measurements of the heteroduplexed molecules (Table 4) indicate that there is a difference in the lengths of the H3/H2B, H2A/H3 and possibly the H2A/H1 spacers of H22 and SU1. To verify this observation, histograms of the

absolute value of the differences between the two single strands constituting one "bubble", were drawn (fig. 23). It appears that the single strands constituting the first and second "bubbles", i.e. those "bubbles" corresponding to the H3/H2B spacer and the H2A/H3 spacer are of different lengths. The single strands constituting the third "bubble" (the H2A/H1 spacer) appears to be of similar length.

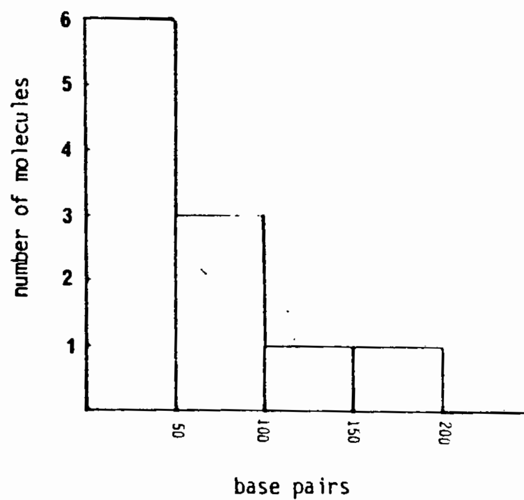
Comparison of the lengths calculated for the AT rich and GC rich regions of H22 and SU1, as determined by denaturation mapping (Table 3), and that of the heteroduplexes, leads to the conclusion that in both the H2B/H3 and the H2A/H3 nonhomolgous "bubbles", the longer of the two single strands is from SU1, and the shorter from H22. In the case of the nonhomologous "bubble" corresponding to the H2A/H1 spacer, however, the length of the single strands on each of the "bubble" is of similar length. However,the single strands constituting the nonhomologous "bubble" corresponding to this particular spacer region is shorter than that observed in either the denatured H22 or SU1. Therefore there is a region of about 200 bp in this particular spacer that is homologous under the conditions of the experiment, as compared to the two other spacer regions looked at.



a



b



c

Fig 23:
Histograms of the absolute value of the differences between the shorter single strand and the larger single strand of the heteroduplexes.
a) The H2B/H3 spacer.
b) The H3/H2A spacer.
c) The H2A/H1 spacer.

Despite the many problems experienced in analysing the heteroduplexed molecules, a certain amount of useful information could be derived. The heteroduplexed molecules served to prove that, at least for the H1, H2A, H3 and H2B coding sequences of SU1, all four genes lie on the same strand of DNA and that these four genes are transcribed in the same direction. Knowledge of the position of the structural gene coding sequences in H22, allows tentative positioning of the H1, H2A, H3 and part of the H2B structural gene coding sequences in SU1, as indicated in Fig. 24. From table 4, it can be seen that the GC-rich regions are homologous, except in the case of the GC-rich region in which the H1 structural gene occurs. Here about 200 base pairs of the AT- rich region lying between the H2A and H1 structural gene coding areas also appears to be homologous. Taking this into account, the H2B gene coding sequence of SU1 can be located before base pair 117, the H3 gene coding sequence between base pairs 1170 and 1580, that of H2A between base pairs 2520 and 2900 and that of H1 between base pairs 3800 and 4420.

Fig 24:
A summary of the physical characteristics of SU1 is shown in the above diagram. The GC rich regions of the histone gene battery quintet is expected to contain the histone gene coding sequences. The GC- rich regions and the restriction fragments on which these structural gene coding sequences are expected to occur, correlate. The heteroduplex analysis of SU1 against H22 gave the approximate positions of the H2B, H3, H2A, and H1 structural genes. Also, it was shown that these four structural genes all occur on the same strand of DNA and that these four genes are transcribed in the direction as shown.

3.7 CONCLUSION

SU1 proved to be an excellent example in introducing one to the problems associated with denaturation, and especially heteroduplex mapping. The statistical analysis enabled the rejection of spurious measurements and the determination of the lengths of the single and double stranded regions of both the denatured and heteroduplexed DNA.

SU1 has the properties of an early histone gene battery. The order of the structural genes is as has been observed for all species studied to date. There is an alternation of AT rich spacer regions and GC rich gene areas. The DNA coding for H2A, H2B, H3 and H1 is very homologous to the corresponding DNA sequences in H22 and thus to the structural histone genes of most other species (the H4 histone gene was not studied). Nonhomology of spacer regions has been noticed in other sea urchin histone gene batteries, both by heteroduplex mapping (Wu, et al., 1976) and sequencing (Sures, et al., 1978, Wells, 1976). Further, the coding sequences of the structural genes studied all lie on the same strand and are transcribed in the same direction. A summary of the physical characteristics of SU1, as determined in this thesis is given in fig. 24.

MATERIALS AND METHODS

4.1 PREPARATION OF PLASMID DNA

pBR322 and derivatives in HB101 were prepared as follows: A culture was prepared by innoculating 0,3 ml glycerol culture into 100 ml of L-broth with the required antibiotic. The culture was incubated at 37°C for 16 hours with vigorous shaking. Of this culture, 8 ml was innoculated into 800 ml L-broth and incubated at 37°C, while shaking until the cells had reached optical density at wavelength 650 nm of 0.2 (about 4 hrs). To amplify the plasmids, 2 ml chloramphenicol (87.5 mg/ml in absolute ethanol) was added to 200 ml of L-broth and this added to the culture. The culture was further incubated for 16 hrs at 37°C with shaking. The cells were harvested by centrifugation for 30 minutes at 10000 rpm in a GSA Sorval rotor at 4°C.

The pellets were resuspended in 8 ml 50 mM Tris/HCl pH 7.6. 1 ml lysosome was added from a 10 mg/ml freshly made up stock solution. The mixture was incubated for 30 minutes at 4°C with vigorous shaking. To this, 1 ml of 500 mM EDTA pH 7.5 was added. The mixture was further incubated at 4°C with vigorous shaking. 0,1 ml RNase A was added from a 20 mg/ml solution in 0.3 M sodium acetate pH 5.5, followed immediately by 0,2 ml 10%(w/v) Triton. This was incubated at 4°C for 20 minutes with gentle shaking. The lysed cells were then centrifuged for 45 min at 17 000 rpm in a Sorvall SS34 rotor at 4°C. The supernatant was extracted twice with phenol. The extraction was repeated with chloroform, followed by ethanol or isopropanol precipitation.

The pellet was dissolved in a caesium chloride solution containing 0,84 g/ml CsCl, 0,2 mg/ml ethidium bromide, 1 mM EDTA and 10 mM Tris/HCl pH 7.6. The refractive index was adjusted to 1.3888. The gradient was centrifuged for 16 hrs

at 55 000 rpm in 65VTi rotor at 20°C. The supercoiled fraction was collected by piercing the tube with a hypodermic needle attached to a 1 ml syringe (Maniatis, et al., 1982). The ethidium Bromide was removed by iso-amyl alcohol extraction and the caesium chloride by dialysis against 1 mM EDTA, 10 mM Tris/HCl pH 7.6.

4.2 EXTRACTION OF DNA FROM GELS

DNA restriction fragments were isolated from low melting point agarose gels according to Maniatis, et al., 1982.

4.3 RADIOACTIVE LABELING

Nick-labelling reactions were carried out by a modified method of Rigby, et al., (1977). Each reaction was carried out in a total volume of 45 µl containing 1 µg DNA, 0.02 mM dATP, 0.02 mM dCTP, 0.02 mM dTTP, 6 mM $MgCl_2$, 10 mM mercaptoethanol, 5 mg BSA, 50 mM Tris/HCl pH7.8, 1 µl ^{32}P-dGTP (approximately 1000 Ci/mmole), 4 x 10^{-4} units DNase I (freshly diluted stock), 2 units DNA polymerase. The reaction was monitored by TCA precipitation. During the labelling process, a 1 µl aliquot was place on a Whatman DE 81 filter. The Cerenkoff radiation of the aliquot was determined in the tritium channel of a scintillation counter. The filter was then placed on a scintered glass filter and washed with 2 ml 10% (W/V) TCA, followed by 2 washes of 15 ml 5% (W/V) TCA, and recounted to determine the percentage incorporation. The reaction was allowed to proceed at 16°C until the incorporation of the radioactive nucleotides reached a plateau as determined by the TCA precipitation. Normally greater than 50% of the radioactive nucleotide is incorporated. All un-incorporated nucleotides were removed by spun columns. (Maniatis, et al., 1982).

4.4 ELECTROPHORESIS

Double stranded DNA fragments were analysed on horizontal agarose gels in TBE buffer (0.089 M Tris-borate, 0.089 M boric acid, 0.002 M EDTA, pH8). The agarose was dissolved in boiling TBE. The gel mixture was then cooled to 50°C and ethidium bromide added from a stock solution (10 mg/ml in H_2O) to a final concentration of 1 µg/ml. The gel chambers were constructed essentially according to Maniatis, et al., (1982). All agarose type IV was obtained from Sigma. To monitor restriction enzyme digests miniature gels (85 x 125 mm) containing a 50 ml agarose solution were run at 120 V constant voltage in the presence of 1 µg/ml ethidium bromide. The samples were applied in sample application buffer (50% [w/v] glycerol, 0.1 mM EDTA and 0.01% [w/v] bromophenol blue). The progress of the electrophoresis was followed by using a hand-held UV light source.

When the gels were to be photographed, they were visualised on an ultra-violet light box and photographed through a red film on Kodak Panatomic X film. The film was processed using D76 or D19 developer according to the manufacturers instructions.

4.5 SOUTHERN TRANSFERS

Southern transfers were carried out according to Maniatis, et al., 1982. Gels were denatured for 1 hour in denaturing solution (1.5 M NaCl, 0.5M NaOH) followed by 1 hour in renaturing solution (1 M Tris/HCl pH 8.0, 1.5 M NaCl). The DNA was transferred onto Hybond N according to the manufacturers instructions. The DNA was cross-linked to the hybond N by 5 minutes exposure to UV light.

Hybridisations were carried out in sealed plastic bags according to Maniatis, et al., (1982). Prehybridisation and hybridisations were carried out at 65°C.

4.6 RESTRICTION ENZYME DIGESTIONS

Unless otherwise indicated, restriction enzymes were obtained from Amersham and digestions carried out according to the manufacturer's instructions.

4.7 LIGATION REACTIONS

The insert to vector ratio of the ligation reactions was calculated according to Maniatis, et al., (1982). Ligation reactions were carried out overnight at 16°C. The vectors were first treated with bacterial alkaline phosphatase, followed by phenol extraction and a spun column to remove ammonium sulphate, present in the bacterial alkaline phosphatase buffer and which could possibly inhibit the ligation reaction.

4.8 PROPAGATION OF PLASMIDS

HB101 (F-, hsd 20 (r^{-B},m^{-B}), rec A 13, ara -14, pro A2, Lac VI, gal K2, rps L20 (SM^{v}), xyl -5, m +1-1, sup E44, -) competent cells were prepared by innoculating 5 ml of L-broth with a loop from a stock glycerol culture and incubating overnight at 37°C with shaking. 1 ml of this culture was innoculated into 100 ml L-Broth (Maniatis, et al., 1982) and incubated for 2 hours at 37°C. This was followed by chilling the cells on ice for 10 minutes. The cells were pelleted in a Sorvall SS34 rotor at 4000 rpm at 4°C for 5 minutes. The pellet was resuspended in ice cold CT buffer (50 mM CaCl2, 1 mM Tris/HCl pH 8.0) and left on ice for 15 minutes. These cells were centrifuged down as before and the pellet was resuspended in

6 ml ice cold CT buffer containing 20% (w/v) glycerol. The resuspended cells were aliquotted and stored at -80°C.

Each ligation reaction was carried out with suitable controls (Maniatis, et al., 1982). The transformation mixture was made up by addition of a maximum of 100 mg of plasmid DNA suspended in 10 µl 1 mM EDTA, 10 mM Tris/HCl pH 7.6 to 200 µl competent cells. The mixture was mixed gently and incubated on ice for 30 minutes. 1 ml L-broth was added with gentle agitation and the cells incubated at 37°C for 1 hour without shaking.
100 µl of the transformed cells were plated out on Petri dishes containing L-broth with 1,5% (w/v) Bacto-agar and a suitable antibiotic (Maniatis, et al., 1982).

The insertion of DNA fragments into pBR322 was determined by the inactivation of the relevant antibiotic site. Antibiotic sensitive colonies were grown up and screened for the correct sized insert. The identity of the insert was confirmed by restriction enzyme digestion of the insert.

Bacterial strains were maintained as glycerol cultures. A colony was picked from a streaked plate and innoculated into 2 ml L broth and a suitable antibiotic. This culture was incubated overnight at 37°C with shaking and 2.4 ml 80% (w/v) glycerol was added. The cels were well mixed and maintained at -20°C for storage up to 6 months or -80°C for long term storage.

4.9 MELTING PROFILE

The method of Mandel and Marmur (1968) was used. The melting was carried out in 0.1 x SSC (0.3 M Na Citrate, 3 M NaCl), pH 8 in a Hewlett Packard 8450A Diode array spectrophotometer.

4.10 ELECTRON MICROSCOPY

The denaturation of DNA was carried out essentially by the method of Inman and Schnos 1970. Each denaturation was carried out in a total volume of 100 µl containing 5 µg DNA, 0.02 M Na_2CO_3, 0.005 M Na_2 EDTA and 10% formaldehyde, which had been prepared by the method of Figerio, et al., 1969. To this solution 5 µl 1 N NaOH was added. This solution was left at 37°C for 45 minutes and then cooled for 10 minutes on ice. When cold, the denatured DNA was passed through 2 spun columns that had been equilibriated with 10 mM Tris HCl pH 8.5, 10 mM Na_2EDTA. The denatured DNA was then stored at -20°C until prepared for electron microscopy.

Heteroduplex DNA was prepared essentially by the method of Davis, Simon and Davidson (1977). Equimolar amounts of linearised plasmid (100 mg of each) was dissolved in 20 µl Roberts buffer (6 mM Tris/HCl pH 7.5, 6 mM NaCl and 1 mM EDTA)[W. de Wet - personal communication]. Denaturation was achieved by the addition of 2.5 µl 2 M Tris/HCl pH 7.2 and

25 µl formamide (3x recrystallised) was added and the hybrids allowed to form for a minimum of 1/2 an hour and a maximum of 4 hours. Thereafter, the DNA was prepared for electron microscopy.

The DNA was spread according to the method of Davis, Simon and Davidson (1977). Typically the hyperphase consisted of 300 ng of DNA and 100ng of size markers (PUC8 was used as a double stranded marker and M13 was used as a single stranded marker) 100 mM Tris pH 8.5, 10 mM Na_2EDTA and formamide at a specified concentration in the range of 50 - 80% in a total volume of 50 µl. To this 2 µl 1 mg/ml cytochrome C (Type V, Sigma, prepared freshly) was added. The DNA was spread at room temperature in a perspex hood onto the hypophase containing 10 mM Tris/HCl pH 8.5, 1mM Na_2EDTA and formamide at a concentration 30% less than the hyperphase. The cytochrome c

film was picked up on a collodion-coated grid (Kleinschmidt 1968), stained with uranyl acetate in 90% alcohol, rinsed in 90% alcohol, dried in 2-methylbutane and shadowed with gold(40%)/palladium(60%) on a rotating stage and coated with carbon. The grids were viewed in Zeiss 109 electron microscope or a Jeol 200 CX and at least 20 molecules were photographed for each experiment

APPENDIX

5.1 DATA ANALYSIS

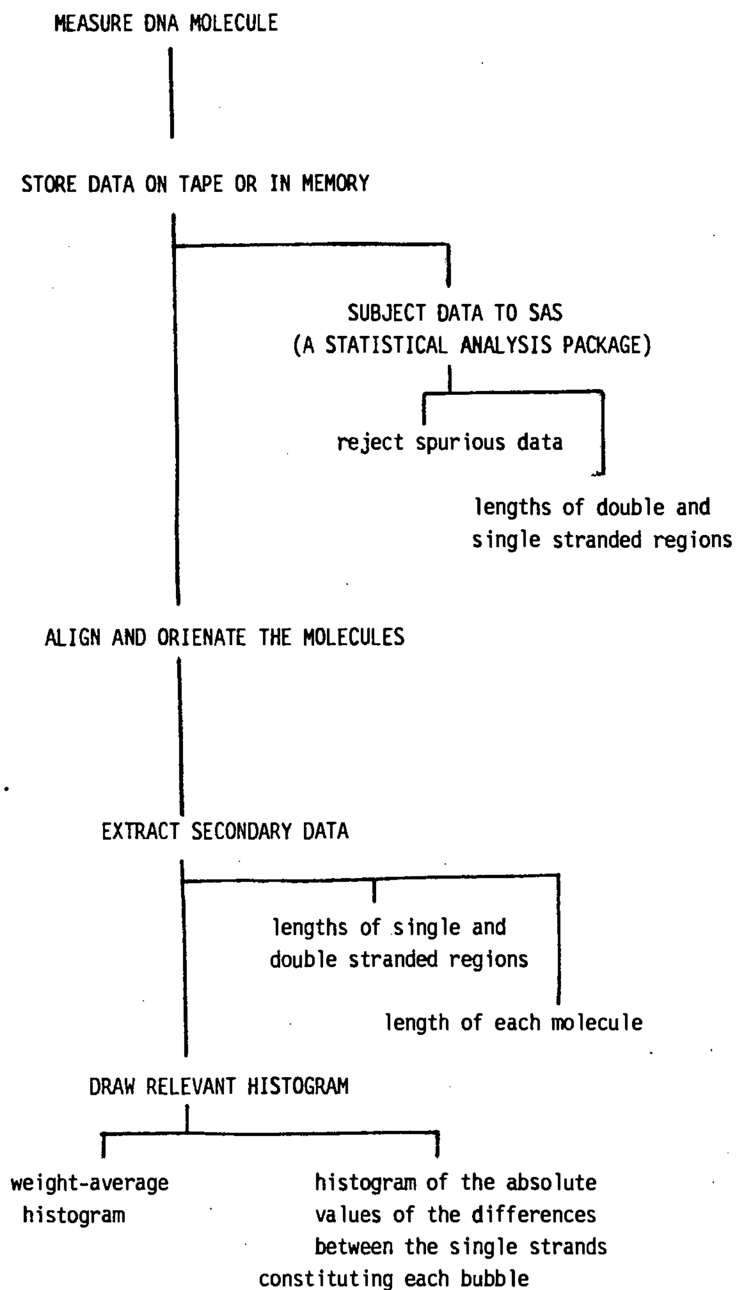


Fig 25:
A flow diagram summarizing data analysis.

"Obtaining DNA length measurements from electron micrographs is an error prone and tedious task" (Littlewood and Inman, 1982). The truth of this statement is painfully brought home to anyone who attempts a denaturation or heteroduplex map. It is thus advisable to delegate as much of the procedure as possible to a computer, without the loss of subjective decisions on the part of the operator in the analysis of electron micrographs.

5.1.1 Data analysis

A photograph of the DNA molecule is traced by hand using a digitiser which sends a stream of X-Y co-ordinates to the computer (Tektronix 4051) and a computer programme constantly calculates the distances moved by the cursor. The distances moved between any two points on a molecule and the information as to whether the DNA is single or double stranded is communicated to the computer by the coloured keys on the cursor. This length information can then be stored, manipulated, plotted and printed in various ways; secondary data extracted and final versions of the molecules can be plotted. After any given set of molecules have been processed, the composite data is expressed in a histogram or histograms.

5.1.1.1 Tracing and recording of length information

Length measurements using digitizer systems can result in considerable overestimation of the actual value (Cornelisse, et al., 1984). The use of standard DNA, however should minimize this effect, because both the standard DNA, and the experimental molecule will be overestimated to the same degree.

The molecular size of the experimental molecule is calculated from the ratio of its contour length to the average contour length of one or more standard molecules on the same photograph, multiplied by the known molecular weight or size (in bp) of the standard (Coggins (b), 1987). The monitoring as to whether single stranded or double-stranded marker DNA, or the experimental molecule is being measured is carried out via the coloured keys of the cursor. These keys are also used to indicate the start and end of molecules, denatured regions in denatured DNA and nonhomologous regions in heteroduplexed DNA. When the user indicates the end of a molecule, the length data for the various segments, along with the information as to what they represent, are stored in the memory of the computer. When all the molecules have been measured, the data can be stored on magnetic tape should the operator so require. The data can now be analysed using the SAS statistical package. (section 5.2). At a later stage, data can be retrieved from the magnetic tape for futher manipulations.

The computer programme uses the hypotenuse rule to calculate lengths, based on the incoming stream of X-Y co-ordinates.

5.1.1.2 Alignment and orientation of molecules

As has been previously discussed in the introduction, it is sometimes necessary in some denaturation and heteroduplex experiments to be able to both orientate and align the molecules. The computer has been programmed to draw each molecule on the screen, with the denatured or nonhomologous regions indicated by a double linear triangle. Molecules can be reversed, removed and/or moved so that the best alignment of the denatured or nonhomologous regions, as judged by the operator, is achieved. This information can then be plotted out, if so wanted. The newly aligned and orientated molecules

are then stored on tape with the relevant information for further manipulation.

5.1.2 Extraction of secondary data

After orientation and alignment of the molecules, various other types of information can be extracted, for example, the position or size of denatured or nonhomologous sites and the lengths of each of the experimental molecules. Graphs or histograms of such secondary data can be produced.

If one is considering partially denatured molecules, a convenient way of presenting the extent and frequency of denatured regions is by plotting a weight average histogram (Coggins (b), 1987; Littlewood and Inman, 1982). Each molecule is divided up into a number of segments, as specified by the operator, and a histogram is constructed of rectangles for each of these segments. The height of each rectangle is proportional to the fraction of molecules that show denaturation in a site covering the segment.

A similar type of histogram can be plotted for heteroduplex molecules, should substitution loops or "bubbles" due to single base changes, be present. A problem does occur in that one is not sure as to whether both single strands constituting the "bubble" are of equivalent size. This can be overcome by plotting two histograms, one of the larger of the two single strands constituting the "bubbles" along a molecule, and the other the smaller of the two single strands.

The question as to whether the two single strands constituting a "bubble" in a heteroduplex are really different in size can be answered by looking at the absolute values of the differences in length between the two single strands. Supposing the two sides of the "bubble" are equal, it is then

easy to show (Wu, et al., 1976) that the distribution curve of the difference between the two sides of the "bubble" would fit the equation:

$$f[(x_1-x_2)] = A \exp[-1/2(x_1-x_2)^2/(\sigma_1^2 + \sigma_2^2)]$$

where σ_1 and σ_2 are the standard deviations in the length of x_1 and x_2. The maximum value of $f(x_1-x_2)$ would be for $x_1-x_2 = 0$, and it should fall off as a Gaussian curve. Should this value not be zero, then it can be assumed that the two single strand lengths are different. A histogram can thus be plotted of (x_1-x_2) for each of the nonhomologous "bubbles", the segment size being specified by the operator, and conclusions be drawn as to the lengths of the single strands constituting the "bubble".

It must be noted that the programme has only been written to handle the data of the experiments in this project. Further modifications may have to be made for a different set of experiments. Deletion loops, for example, were not experienced in the heteroduplex, and thus the programme is not yet equipped to handle such data.

5.2 THE USE OF THE SAS STATISTICAL PACKAGE

SAS, a software system for data analysis was introduced, so as to futher refine the results. This system provides a useful statistical tool for looking at the data prior to final manipulations. Molecules that do not follow the general trend can be detected, erroneous measurements picked up, and general statistics on the data carried out.

The double and single stranded regions of each molecule are negatively correlated, that is, as the single stranded lengths increase,the lengths of the double stranded regions decrease, and vice versa. This observation is borne out by both the plotting of the total single strand lengths versus the total double strand length and by the correlation matrix.

The package also provides a stem leaf plot, so that one can access by eye, the skewness of distribution of the values for each double and single stranded regions. In many cases, a skew distribution, both of the positive and negative kind, is observed. This is a reflection of the data set. As has already been discussed, each individual molecule does not show the same degree of denaturation as another molecule on the same grid (Section 1.2.2). Within an AT rich region, there may exist a section that is more easily denatured. The choice of molecules in the data set may be such that such a phenomenon is observed. This phenomenon may also result in two or more length populations for a particular stretch of denatured DNA being noted.

The first clue as to an erroneous measurement, or a molecule that does not follow the general trend, is in the plot of the total double stranded vs total single stranded regions (and/or in the total larger single strand vs total

smaller single strand in the case of heteroduplexed DNA), where there appears a data point or points such that the correlation observed is not as expected. This is usually reflected in the correlation matrix. The stem leaf plot also helps in the detection of erroneous measurements by showing outlying values such that one is made aware that there is a value that is significantly larger or smaller that the majority of other measurements for a particular segment of single stranded, denatured or double stranded DNA. One must go back to the original data to check that there is not a valid reason for this. A molecule with a somewhat more denatured region, may have the lengths of the corresponding double stranded regions shorter to a corresponding extent. The decision now must be made as to whether such a molecule is introducing a significant error or not. The usual method is to remove the molecule, and to see if the correlation matrix is now as expected. If not, there are other significant errors in the data. The measurements of the molecule can be returned to the data set, and the data again analysed.

Because there is a dependence of the length of the double stranded regions on that of the single stranded regions, and vice versa, it should be possible to carry out linear regression analysis on the data values, so as to narrow down the confidence interval. Each double stranded region was assumed to be affected by the length of the single stranded regions ajacent on each side, and vice versa. The corrected standard deviation, that is, the root mean square error, is derived from the "corrected" data points, i.e. the data points are corrected so that each point lies on a line derived by fitting the available data points to a line in such a way that the squared deviations between the the points and the line are a minimum value.

In the case of the heteroduplexed DNA, the linear regression was carried out is as for the denatured DNA, but assuming that both the single strands on adjacent sides affect the length of the double stranded region (negative correlation). Each single strand's length was assumed to be affected by the length of the double stranded regions (negative correlation) and by the length of the single strand immediately opposite (positive correlation).

The major limitation in the use of SAS was that specific molecules have to be chosen from the total data set. The number of single and double stranded regions were specified, resulting in molecules having more, or less denatured regions not being included in the data analysis.

5.3 PROGRAM FOR DNA MEASUREMENT AND DATA DISPLAY

```
100 REM D saves double stranded lengths
110 REM S saves single stranded lengths-one side loop
120 REM G saves single stranded lengths-other side loop
130 REM O -counter end of molecule
140 REM B1-distance molecule moved during alignment
150 REM B-the size of each loop
160 REM B3-the std of B
170 REM E-the size of each loop-other side
180 REM E3-the std of E
190 REM A2-saves size double stranded lengths
200 REM A3-the std dev of A2
210 REM C-counter B,E
220 REM D1-counter A2
230 REM
240 REM----------------------------------------------------
250 INIT
260 E1=0
270 P0=4
280 T0=8
290 S0=32
300 P1=1
310 PRINT @S0,18:5
320 PRINT @S0,26:1
330 S6=100000
340 PRINT "JJ"
350 REM
360 REM
370 C1=0
380 C2=0
390 C3=0
400 A=0
410 DELETE D,S,G,O,U,B1,B,E,C,A2,D1,B3,A3
420 DIM D(150),S(150),G(150),O(50),B1(150)
425 DIM B(20),E(20),C(20),A2(20)
430 DIM D1(20),B3(20),A3(20),E3(20),Y5(20)
440 B=0
450 E=0
460 C=0
470 A2=0
480 D=0
490 D1=0
500 B3=0
510 A3=0
520 S=0
530 G=0
540 B1=0
550 E3=0
560 T=1
570 A1=0
580 O=0
590 Y5=0
```

```
600 REM --------------------------------------------------
610 PAGE
620 PRINT @SO:"PROGRAM TO MEASURE DENATURED OR HETERODUPLEX DNA"
630 PRINT "JJ"
640 PRINT "Do you want to read off tape?"
650 INPUT S$
660 IF S$="y" THEN 2900
670 REM---------------------------------------------------
680 PRINT "Do you want to enter data via keyboard?"
690 INPUT S$
700 IF S$="y" THEN 3580
710 REM --------------------------------------------------
720 PRINT "Is the DNA denatured? (y or n)"
730 INPUT Q$
740 REM --------------------------------------------------
750 REM marker DNA
760 PRINT "Place photograph on tablet.  First std DNA shall"
770 PRINT "be measured.  Then the experimental molecule/s"
780 PRINT "shall be measured."
790 PRINT @SO:"Enter the magnification of the print"
800 INPUT M5
810 PRINT "JJ"
820 Q=1
830 REM double strand correction factor
840 Q1=1
850 REM single strand correction factor
860 V=0
870 X8=0
880 L=0
890 PRINT "Are you using only d.s. std?"
900 INPUT L$
910 PAGE
920 IF L$="y" THEN 940
930 PRINT "Press green button to measure s.s. DNA"
940 PRINT "Press yellow button to measure d.s.std"
950 PRINT "When have measured all, press white button"
960 PRINT @TO:"O"
970 WINDOW 189,3580,176,3769
980 INPUT @TO:X,Y,Z
990 IF Z<>0 THEN 1020
1000 GO TO 980
1010 PAGE
1020 IF Z=1 OR Z=8 THEN 1060
1030 IF Z=2 THEN 1340
1040 IF Z=0 THEN 1020
1050 GO TO 1020
1060 R=Z
1070 MOVE X,Y
1080 X1=X
1090 Y1=Y
1100 INPUT @TO:X,Y,Z
1110 IF Z=2 THEN 1340
1120 IF Z=0 THEN 1210
1130 X2=X1-X
```

```
1140 Y2=Y1-Y
1150 L1=X2*X2+Y2*Y2
1160 L=L+SQR(L1)
1170 DRAW X,Y
1180 X1=X
1190 Y1=Y
1200 GO TO 1100
1210 IF R=1 THEN 1230
1220 IF R=8 THEN 1270
1230 PRINT @4:"ds std",L
1240 Q=Q+L/M5*S6
1250 L=0
1260 GO TO 1320
1270 PRINT @4:"ss std",L
1280 Q1=Q1+L/M5*S6
1290 L=0
1300 X8=X8+1
1310 GO TO 910
1320 V=V+1
1330 GO TO 910
1340 IF L$="y" THEN 1360
1350 Q1=Q1/X8
1360 Q=Q/V
1370 PAGE
1380 PRINT "Have you measure all stds?"
1390 INPUT S$
1400 IF S$="y" THEN 1450
1410 X8=1
1420 V=1
1430 GO TO 910
1440 REM************************************
1450 PRINT "What is the no of bp for d.s. std?"
1460 INPUT M3
1470 Q=M3/Q
1480 PRINT @4:"The d.s. corr factor is";Q
1490 IF L$="y" THEN 1540
1500 PRINT "What is the no of bp for ss std?"
1510 INPUT M4
1520 Q1=M4/Q1
1530 PRINT @4:"the ss corr factor is";Q1
1540 REM ---------------------------------------------------
1550 REM EXPERIMENTAL MOLECULES
1560 PAGE
1570 PRINT @S0:"MEASUREMENT OF DNA"
1580 PRINT "JJ"
1590 GO TO 1610
1600 PAGE
1610 R=0
1620 L=0
1630 W=0
1640 REM************************************
1650 PRINT "Press the yellow button to measure d.s. DNA"
1660 PRINT "Press green button to measure s.s.DNA"
1670 PRINT "Note: measure both sides of the denatured or hybrid bubbles"
```

```
1680 PRINT "However, release green button between measurements"
1690 PRINT "If only one ss to be measured, then press"
1700 PRINT "blue button after using green button"
1710 PRINT "Use white button to indicate the end of a molecule"
1720 PRINT "when all measured,then press yellow/white keys"
1730 PRINT @T0:"0"
1740 WINDOW 189,3580,176,3769
1750 INPUT @T0:X,Y,Z
1760 IF Z<>0 THEN 1780
1770 GO TO 1750
1780 PAGE
1790 IF Z<>8 THEN 1830
1800 C1=C1+1
1810 D(C1)=0
1820 REM
1830 IF Z=1 THEN 1930
1840 IF Z=2 THEN 2400
1850 IF Z=3 THEN 2390
1860 IF Z=4 THEN 1890
1870 IF Z=8 THEN 1960
1880 IF Z=0 THEN 1830
1890 C3=C2
1900 REM
1910 L=0
1920 GO TO 2270
1930 C1=C1+1
1940 REM
1950 GO TO 2020
1960 IF A=0 THEN 2000
1970 C3=C3+1
1980 A=A+1
1990 GO TO 2020
2000 C2=C2+1
2010 A=A+1
2020 MOVE X,Y
2030 X1=X
2040 Y1=Y
2050 R=Z
2060 W=Z
2070 INPUT @T0:X,Y,Z
2080 IF Z=2 THEN 2400
2090 IF Z=0 THEN 2190
2100 X2=X1-X
2110 Y2=Y1-Y
2120 L1=X2*X2+Y2*Y2
2130 L=L+SQR(L1)
2140 DRAW X,Y
2150 X1=X
2160 Y1=Y
2170 GO TO 2070
2180 REM************************************
2190 IF R=1 THEN 2210
2200 IF R=8 THEN 2240
2210 L=L*Q/M5*S6
```

```
2220 D(C1)=L
2230 GO TO 2320
2240 L=L/M5*S6
2250 REM
2260 IF A=1 THEN 2300
2270 G(C3)=L
2280 A=0
2290 GO TO 2320
2300 S(C2)=L
2310 IF Z=4 THEN 2330
2320 PRINT "."
2330 L=0
2340 L1=0
2350 IF Z=2 THEN 2400
2360 INPUT @T0:X,Y,Z
2370 IF Z<>0 THEN 1830
2380 GO TO 2360
2390 W=Z
2400 T=T+1
2410 A1=A1+1
2420 IF C1>C2 THEN 2450
2430 IF C2>C1 THEN 2500
2440 GO TO 2530
2450 C2=C2+1
2460 C3=C3+1
2470 S(C2)=0
2480 G(C3)=0
2490 GO TO 2420
2500 C1=C1+1
2510 D(C1)=0
2520 GO TO 2420
2530 O(T)=C1
2540 PAGE
2550 IF W=3 THEN 2590
2560 PRINT "Is there another exp. molecule?"
2570 INPUT S$
2580 IF S$="y" THEN 1600
2590 IF Q$="n" THEN 2780
2600 REM***********************************
2610 REM
2620 U=0
2630 X3=0
2640 REM
2650 X4=0
2660 PRINT @S0:"what is the length of denatured DNA in bp?"
2670 INPUT M1
2680 FOR I=O(T-1)+1 TO O(T)
2690   X3=D(I)+X3
2700   X4=(S(I)+G(I))/2+X4
2710 NEXT I
2720 X3=M1-X3
2730 Q1=X3/X4
2740 X4=0
2750 X3=0
```

```
2760 PRINT @4:"The ss corr factor is:",Q1
2770 REM************************************
2780 FOR J=T-A1 TO T-1
2790    FOR I=O(J)+1 TO O(J+1)
2800             S(I)=S(I)*Q1
2810             G(I)=G(I)*Q1
2820    NEXT I
2830 NEXT J
2840 A1=0
2850 IF W=3 THEN 3230
2860 PRINT "Next photograph"
2870 GO TO 820
2880 GO TO 3140
2890 REM ----------------------------------------------------
2900 REM read data from tape
2910 PRINT "Enter file number"
2920 INPUT X
2930 FIND X
2940 FOR I=1 TO 150
2950   INPUT @33:B1(I)
2960   INPUT @33:D(I)
2970   INPUT @33:S(I)
2980   INPUT @33:G(I)
2990 NEXT I
3000 FOR I=1 TO 50
3010   INPUT @33:O(I)
3020 NEXT I
3030 FOR I=1 TO 20
3040   INPUT @33:B(I)
3050   INPUT @33:B3(I)
3060   INPUT @33:E(I)
3070   INPUT @33:E3(I)
3080   INPUT @33:A2(I)
3090   INPUT @33:A3(I)
3100 NEXT I
3110 INPUT @33:T
3120 INPUT @33:Q$
3130 REM----------------------------------------------------
3140 REM to put the longer ss length in s
3150 REM and the shorter in g
3160 FOR I=1 TO 150
3170    IF S(I)=>G(I) THEN 3210
3180    K=G(I)
3190    G(I)=S(I)
3200    S(I)=K
3210 NEXT I
3220 REM----------------------------------------------------
3230 REM data to line printer
3240 PRINT @4:T
3250 PRINT @P0:"The first number printed is the length of ds measured"
3260 PRINT @P0:"The second is ss DNA"
3270 PRINT @P0:"The third is the other side of ss bubble"
3280 PRINT @P0:"The fourth is the average of the ss measurements"
3290 P=0
```

```
3300 Z=0
3310 FOR J=2 TO T
3320   P=Z+1
3330   Z=O(J)
3340   FOR I=P TO Z
3350       PRINT @PO: USING 3420:I,D(I)
3360       PRINT @PO: USING 3420:I,S(I)
3370       PRINT @PO: USING 3420:I,G(I)
3380       IF G(I)<>0 THEN 3410
3390       PRINT @PO: USING 3430:I,S(I)
3400       GO TO 3440
3410       PRINT @PO: USING 3430:I,(S(I)+G(I))/2
3420       IMAGE"I",5d,"."10d,s
3430       IMAGE"I",5d,"."10d
3440   NEXT I
3450   PRINT @PO:"J"
3460 NEXT J
3470 L$=CHR(13)
3480 PRINT @PO:L$,"L"
3490 IF E1=1 THEN 4070
3500 REM ------------------------------------------------
3510 PAGE
3520 PRINT "Do you wish to save on tape?"
3530 INPUT L$
3540 IF L$="y" THEN 7220
3550 REM--------------------------------------------------
3560 PRINT "Do you wish to enter data via keyboard"
3570 INPUT S$
3580 IF S$="n" THEN 3770
3590 C1=O(T)+1
3600 GO TO 3640
3610 PRINT "any more data?"
3620 INPUT S$
3630 IF S$="n" THEN 3770
3640 T=T+1
3650 PRINT "inp,d,s,g"
3660 INPUT X,Y,Z
3670 D(C1)=X
3680 S(C1)=Y
3690 G(C1)=Z
3700 C1=C1+1
3710 PRINT "End of molecule?"
3720 INPUT S$
3730 IF S$="n" THEN 3650
3740 O(T)=C1-1
3750 GO TO 3610
3760 REM ------------------------------------------------
3770 REM to work out the length of DNA molecules
3780 REM and to find max
3790 X4=0
3800 FOR I=2 TO T
3810       X3=0
3820       FOR J=O(I-1)+1 TO O(I)
3830           X3=D(J)+(S(J)+G(J))/2+X3
```

```
3840        NEXT J
3850        PRINT @4:"Length of molecule",I-1,X3
3860        X4=X4 MAX X3
3870 NEXT I
3880 PRINT @4:"The longest molecule's length",X4
3890 REM-----------------------------------------------
3900 PAGE
3910 PRINT "Do you want to draw molecules as already aligned?"
3920 INPUT K$
3930 IF K$="y" THEN 4070
3940 REM ----------------------------------------------
3950 REM to orientate and align molecules
3960 PRINT "the program at this stage can only reverse"
3970 PRINT "heteroduplex's observed for sul(t)/h22(t)"
3980 PRINT "align all"
3990 PRINT "To help in caculating length of double and"
4000 PRINT "single stranded regions, enter"
4010 PRINT "the total no of loops and double stranded"
4020 PRINT "areas to be considered"
4030 PRINT "loops?"
4040 INPUT X6
4050 PRINT "double stranded areas?"
4060 INPUT X7
4070 WINDOW 0,X4+500,0,100
4080 VIEWPORT 65,130,0,100
4090 E1=0
4100 J=1
4110 PAGE
4120 IF K$="y" THEN 6610
4130 IF Q$="y" THEN 4330
4140 PRINT "Do you want to orientate with std dev and euclidean sum?"
4150 INPUT M$
4160 IF M$="n" THEN 4330
4170 PRINT "In order to calculate the Euclidean sum, it"
4180 PRINT "is necessary to have mean or known values entered"
4190 PRINT "for each of the single and double stranded regions."
4200 PRINT "For the heteroduplexes, the double and single stranded"
4210 PRINT "areas at the beginning and end of the molecules are"
4220 PRINT "to be ignored."
4230 PRINT "Enter:longer single strand followed by shorter single"
4240 PRINT "strand for the first bubble, followed by the double"
4250 PRINT "stranded region.  Repeat for the second and third"
4260 PRINT "bubbles."
4270 DELETE F1
4280 DIM F1(8)
4290 FOR I=1 TO 8
4300   INPUT L
4310   F1(I)=L
4320 NEXT I
4330 REM************************************
4340 PAGE
4350 FOR I=2 TO T
4360   IF Q$="n" THEN 4480
4370   L=B1(I)
```

```
4380   MOVE B1(I),100-J*10
4390   FOR K=O(I-1)+1 TO O(I)
4400        DRAW D(K)+L,100-J*10
4410        DRAW D(K)+(S(K)+G(K))/2+L,100-J*10
4420        MOVE D(K)+L,100-J*10+2
4430        DRAW D(K)+(S(K)+G(K))/2+L,100-J*10+2
4440        MOVE D(K)+(S(K)+G(K))/2+L,100-J*10
4450        L=D(K)+(S(K)+G(K))/2+L
4460   NEXT K
4470   GO TO 4700
4480   S1=0
4490   B1(I)=0
4500   R=O(I-1)+1
4510   R1=O(I)
4520   FOR K=R TO R1
4530        IF K=R OR R1 THEN 4560
4540        S1=S1+(S(K)+G(K))/2+D(K)
4550        GO TO 4570
4560        S1=S1+(S(K)+G(K))/2
4570   NEXT K
4580   S1=S1/2
4590   B1(I)=X4/2-S1-D(R)-500
4600   L=B1(I)
4610   MOVE B1(I),100-J*10
4620   FOR K=O(I-1)+1 TO O(I)
4630        DRAW D(K)+L,100-J*10
4640        DRAW D(K)+G(K)+L,100-J*10
4650        MOVE D(K)+L,100-J*10
4660        DRAW D(K)+L+G(K)/2,100-J*10+SQR((S(K)/2)^2-(G(K)/2)^2)/100
4670        DRAW D(K)+L+G(K),100-J*10
4680        L=D(K)+L+G(K)
4690   NEXT K
4700   HOME
4710   IF Q$="y" THEN 5400
4720 REM*****************************
4730   L9=0
4740   L=0
4750   FOR K=1 TO 8
4760   IF K=1 THEN 4780
4770        Y5(K-1)=F1(K-1)+D(O(I-1)+1+L)
4780        Y5(K)=F1(K)+S(O(I-1)+1+L)
4790        Y5(K+1)=F1(K+1)+G(O(I-1)+1+L)
4800        K=K+2
4810        L=L+1
4820   NEXT K
4830   L=0
4840   FOR K=9 TO 16
4850        IF K=9 THEN 4870
4860        Y5(K-1)=F1(K-9)^2+D(O(I-1)+1+L)^2
4870        Y5(K)=F1(K-8)^2+S(O(I-1)+1+L)^2
4880        Y5(K+1)=F1(K-7)^2+S(O(I-1)+1+L)^2
4890        K=K+2
4900        L=L+1
4910   NEXT K
```

```
4920   PRINT @4:"To align, std dev each orientation"
4930   FOR K=1 TO 8
4940        FOR L=1 TO 3
4950             IF J=3*L THEN 4990
4960        NEXT L
4970        PRINT @4:"ss",SQR(Y5(K+8)/2-(Y5(K)/2)^2)
4980        GO TO 5000
4990        PRINT @4:"ds",SQR(Y5(K+8)/2-(Y5(K)/2)^2)
5000   NEXT K
5010   IF L9=10 THEN 5230
5020   L=0
5030   FOR K=1 TO 8
5040        IF K=1 THEN 5060
5050        Y5(K-1)=F1(K-1)+D(O(I)-(1+L))
5060        Y5(K)=F1(K)+S(O(I)-(1+L))
5070        Y5(K+1)=F1(K+1)+G(O(I)-(1+L))
5080        K=K+2
5090        L=L+1
5100   NEXT K
5110   L=0
5120   FOR K=9 TO 16
5130        IF K=9 THEN 4870
5140        Y5(K-1)=F1(K-9)^2+D(O(I)-(1+L))^2
5150        Y5(K)=F1(K-8)^2+S(O(I)-(1+L))^2
5160        Y5(K+1)=F1(K-7)^2+S(O(I)-(1+L))^2
5170        K=K+2
5180        L=1
5190   NEXT K
5200   PRINT @4:"other orientation"
5210   L9=10
5220   GO TO 4930
5230   L=0
5240   Q8=0
5250   FOR K=1 TO 8
5260        IF K=1 THEN 5280
5270        Q8=Q8+(D(O(I-1)+(1+L))-F1(K-1))^2+(S(O(I-1)+(I+L))-F1(K))^2
5280        Q8=Q8+(G(O(I-1)+(1+L))-F1(K+1))^2
5290        K=K+2
5300   NEXT K
5310   PRINT @4:"Euc",Q8
5320   FOR K=1 TO 8
5330        IF K=1 THEN 5280
5340        Q8=Q8+(D(O(I)-(1+L))-F1(K-1))^2+(S(O(I)-(I+L))-F1(K))^2
5350        Q8=Q8+(G(O(I)-(1+L))-F1(K+1))^2
5360        K=K+2
5370   NEXT K
5380   PRINT @4:"Euc,other orientation:",Q8
5390   PRINT "__"
5400 REM to remove a molecule
5410   PRINT "Do you want to remove molecule?"
5420   INPUT S$
5430   IF S$="n" THEN 5570
5440   L=O(I)-O(I-1)
5450   FOR K=O(I-1)+1 TO 150-L
```

```
5460        D(K)=D(K+L)
5470        S(K)=S(K+L)
5480        G(K)=G(K+L)
5490   NEXT K
5500   FOR M=I TO T-1
5510        O(M)=O(M+1)-L
5520   NEXT M
5530   T=T-1
5540   E1=1
5550   I=I-1
5560   GO TO 5590
5570   PRINT "_"
5580 REM***********************************
5590   IF Q$="y" THEN 5830
5600   PRINT "Do you want to reorientate the molecule?"
5610   INPUT S$
5620   IF S$="n" THEN 5830
5630   FOR K=O(I-1)+1 TO O(I)
5640        V=D(K)
5650        W=S(K)
5660        A=G(K)
5670        D(K)=D(O(I)-(K-(O(I-1)+1)))
5680        IF K=O(I)-1-(K-(O(I-1)+1)) THEN 5730
5690        S(K)=S(O(I)-1-(K-(O(I-1)+1)))
5700        G(K)=G(O(I)-1-(K-(O(I-1)+1)))
5710        S(O(I)-1-(K-(O(I-1)+1)))=W
5720        G(O(I)-1-(K-(O(I-1)+1)))=A
5730        D(O(I)-(K-(O(I-1)+1)))=V
5740        IF K=O(I)-1-(K-(O(I-1)+1)) THEN 5760
5750   NEXT K
5760   J=J+1
5770   IF J=9 THEN 5790
5780   GO TO 4360
5790   PAGE
5800   J=1
5810   GO TO 5780
5820 REM***********************************
5830   PRINT "_"
5840   PRINT "Do you want to move the molecule?"
5850   INPUT S$
5860   IF S$="n" THEN 5980
5870   E1=1
5880   PRINT "How much wish to move molecule?"
5890   INPUT X
5900   B1(I)=X+B1(I)
5910   J=J+1
5920   IF J=9 THEN 5940
5930   GO TO 5960
5940   PAGE
5950   J=1
5960   IF Q$="n" THEN 4600
5970   GO TO 4360
5980   PRINT "__"
5990   E1=0
```

```
6000 REM************************************
6010   PRINT "Assuming bubbles from left to right"
6020   PRINT "are numbered from 1, enter no. of"
6030   PRINT "to be used to caculate size of each"
6040   PRINT "bubble"
6050   PRINT "inp ""0"" when finished"
6060   FOR K=1 TO X6
6070        INPUT Y
6080        IF Y=0 THEN 6200
6090        IF Q$="y" THEN 6160
6100        B(Y)=G(Y+O(I-1))+B(Y)
6110        B3(Y)=G(Y+O(I-1))^2+B3(Y)
6120        E(Y)=S(Y+O(I-1))+E(Y)
6130        E3(Y)=S(Y+O(I-1))^2+E3(Y)
6140        C(Y)=C(Y)+1
6150        GO TO 6190
6160        B(Y)=(S(Y+O(I-1))+G(Y+O(I-1)))/2+B(Y)
6170        B3(Y)=((S(Y+O(I-1))+G(Y+O(I-1)))/2)^2+B3(Y)
6180        C(Y)=C(Y)+1
6190   NEXT K
6200   IF Y<>0 THEN 6240
6210   IF K=1 THEN 6230
6220   K=K-1
6230   K=K+2
6240   K=K-1
6250   FOR M=1 TO X6-K
6260        PRINT "_"
6270   NEXT M
6280   PRINT "Similarly for the double stranded regions"
6290   FOR K=1 TO X7
6300        INPUT Y
6310        IF Y=0 THEN 6360
6320        A2(Y)=A2(Y)+D(Y+O(I-1))
6330        A3(Y)=D(Y+O(I-1))^2+A3(Y)
6340        D1(Y)=D1(Y)+1
6350   NEXT K
6360   J=J+1
6370   IF J<9 THEN 6400
6380   J=1
6390   PAGE
6400 NEXT I
6410 REM -------------------------------------------------
6420 PAGE
6430 FOR I=1 TO X6
6440   IF Q$="y" THEN 6510
6450   B(I)=B(I)/C(I)
6460   E(I)=E(I)/C(I)
6470   B3(I)=SQR(ABS(B3(I)/C(I)-B(I)^2))
6480   E3(I)=SQR(ABS(E3(I)/C(I)-E(I)^2))
6490   PRINT @4:"single strand,1",B(I),"2",E(I),"std dev,1",B3(I),E3(I)
6500   GO TO 6540
6510   B(I)=B(I)/C(I)
6520   B3(I)=SQR(ABS(B3(I)/C(I)-B(I)^2))
6530   PRINT @4:"<single strand>",B(I),"std dev",B3(I)
```

```
6540 NEXT I
6550 FOR I=1 TO X7
6560   A2(I)=A2(I)/D1(I)
6570   A3(I)=SQR(ABS(A3(I)/D1(I)-A2(I)^2))
6580   PRINT @4:"double strand length",A2(I),"std dev",A3(I)
6590 NEXT I
6600 REM***********************************
6610 PAGE
6620 PRINT "Do you want to draw on plotter or screen,(p,s,n(neither))"
6630 INPUT S$
6640 IF S$="n" THEN 7060
6650 IF S$="s" THEN 7000
6660 J=1
6670 FOR I=2 TO T
6680   IF Q$="n" THEN 6800
6690   L=B1(I)
6700   MOVE @P1:B1(I),100-J*10
6710   FOR K=O(I-1)+1 TO O(I)
6720        DRAW @P1:D(K)+L,100-J*10
6730        DRAW @P1:D(K)+(S(K)+G(K))/2+L,100-J*10
6740        MOVE @P1:D(K)+L,100-J*10+2
6750        DRAW @P1:D(K)+(S(K)+G(K))/2+L,100-J*10+2
6760        MOVE @P1:D(K)+(S(K)+G(K))/2+L,100-J*10
6770        L=D(K)+(S(K)+G(K))/2+L
6780   NEXT K
6790   GO TO 6900
6800   L=B1(I)
6810   MOVE @P1:B1(I),100-J*10
6820   FOR K=O(I-1)+1 TO O(I)
6830        DRAW @P1:D(K)+L,100-J*10
6840        DRAW @P1:D(K)+G(K)+L,100-J*10
6850        MOVE @P1:D(K)+L,100-J*10
6860        DRAW @P1:D(K)+L+G(K)/2,100-J*10+SQR((S(K)/2)^2-(G(K)/2)^2)/100
6870        DRAW @P1:D(K)+L+G(K),100-J*10
6880        L=L+D(K)+G(K)
6890   NEXT K
6900   IF J=9 THEN 6920
6910   GO TO 6970
6920   PRINT "press c to continue"
6930   INPUT X$
6940   PAGE
6950   J=1
6960   GO TO 6980
6970   J=J+1
6980 NEXT I
6990 GO TO 7020
7000 P1=S0
7010 GO TO 6660
7020 P1=1
7030 PRINT "_"
7040 GO TO 6620
7050 REM ---------------------------------------------------
7060 PRINT "do you want to align/orientate?"
7070 INPUT X$
```

```
7080 K$="n"
7090 IF X$="n" THEN 7190
7100 B=0
7110 B3=0
7120 C=0
7130 A2=0
7140 A3=0
7150 D1=0
7160 B1=0
7170 GO TO 3950
7180 REM -------------------------------------------------
7190 PRINT "do you wish to save on tape?"
7200 INPUT S$
7210 IF S$="n" THEN 7470
7220 PRINT "enter file number"
7230 INPUT X
7240 FIND X
7250 MARK 1,25000
7260 FIND X
7270 FOR I=1 TO 150
7280   PRINT @33:B1(I)
7290   PRINT @33:D(I)
7300   PRINT @33:S(I)
7310   PRINT @33:G(I)
7320 NEXT I
7330 FOR I=1 TO 50
7340   PRINT @33:O(I)
7350   NEXT I
7360   FOR I=1 TO 20
7370   PRINT @33:B(I)
7380   PRINT @33:B3(I)
7390   PRINT @33:E(I)
7400   PRINT @33:E3(I)
7410   PRINT @33:A2(I)
7420   PRINT @33:A3(I)
7430 NEXT I
7440 PRINT @33:T
7450 PRINT @33:Q$
7460 PRINT "data stored on tape"
7470 END
```

```
4 GO TO 130
8 GO TO 400
16 GO TO 2280
20 GO TO 2640
100 S9=1
110 PRINT "Need to read in data file derived using program 4 or 6"
120 PRINT "to be able to use this program"
130 DELETE B1,D,S,G,O,B,B3,E,E3,A2,A3
140 DIM B1(150),D(150),S(150),G(150),O(50)
150 DIM B(20),B3(20),E(20),E3(20),E3(20),A2(20),A3(20)
160 PRINT "Enter file number"
170 INPUT X
180 FIND X
190 FOR I=1 TO 150
200   INPUT @33:B1(I)
210   INPUT @33:D(I)
220   INPUT @33:S(I)
230   INPUT @33:G(I)
240 NEXT I
250 FOR I=1 TO 50
260   INPUT @33:O(I)
270 NEXT I
280 FOR I=1 TO 20
290   INPUT @33:B(I)
300   INPUT @33:B3(I)
310   INPUT @33:E(I)
320   INPUT @33:E3(I)
330   INPUT @33:A2(I)
340   INPUT @33:A3(I)
350 NEXT I
360 INPUT @33:T
370 INPUT @33:Q$
380 GO TO 2200
390 REM-----------------------------------------------------
400 REM variables are:x5,x6,x4,x3,c4,w,w1,h,s9
410 REM array variables are:u,m,b,e,k
420 REM to draw histogram
430 DELETE U,M,B2,E4
440 DIM B2(50)
450 DIM U(50)
460 U=0
470 B2=0
480 X4=0
490 X5=0
500 X6=0
510 X8=0
520 FOR I=2 TO T
530   FOR J=O(I-1)+1 TO O(I)
540        IF S(J)=0 THEN 570
550        X5=X5+1
560        X6=X6 MAX X5
570   NEXT J
580   IF U(X5)<>0 THEN 600
590   U(X5)=X5
```

```
600   B2(X5)=B2(X5)+1
610   X5=0
620 NEXT I
630 FOR I=1 TO X6
640   X5=X5 MAX B2(I)
650 NEXT I
660 FOR I=1 TO X6
670   IF X5<>B2(I) THEN 690
680   X4=I
690 NEXT I
700 PRINT @4:"The max no of bubbles",X6
710 PRINT @4:"The no of molecules",B2(X6)
720 PRINT @4:"the maj of molecules have bubbles",U(X4)
730 PRINT @4:"The no of molecules",X5
740 PRINT "How many bubbles do you want to use?"
750 INPUT X6
760 C4=B2(X6)
770 X4=0
780 X3=0
790 DIM U(50),M(50),B2(50),E4(50),K(900)
800 U=0
810 B2=0
820 E4=0
830 REM***********************************
840 REM to determine the total length of each molecule
850 REM and the std dev to each bubble
860 FOR J=2 TO T
870   X5=0
880   FOR I=O(J-1)+1 TO O(J)
890        IF S(I)=0 THEN 910
900        X5=X5+1
910   NEXT I
920   IF X5=X6 THEN 930
930   FOR I=O(J-1)+1 TO O(J)
940        IF G(I)<>0 THEN 980
950        X3=D(I)+S(I)+X3
960        L=S(I)
970        GO TO 1000
980        X3=D(I)+(S(I)+G(I))/2+X3
990        L=(S(I)+G(I))/2
1000       IF X5<>X6 THEN 1080
1010       IF S(I)=0 THEN 1080
1020       IF O(J)-(O(J-1)+1)>X6 THEN 1080
1030       Z=O(J)
1040       B2(X6-(Z-I)+1)=B2(X6-(Z-I)+1)+X3-L
1050       E4(X6-(Z-I)+1)=E4(X6-(Z-I)+1)+(X3-L)^2
1060 REM B2 remembers the approx distance to each bubble
1070 REM E4 is the std dev of the above
1080   NEXT I
1090   PRINT @4:"length of molecule",J-1,X3
1100   U(J)=X31110 X4=X4 MAX X3
1120   X3=0
1130 NEXT J
1140 FOR I=1 TO X6
```

```
1150   B2(I)=B2(I)/C4
1160   E4(I)=SQR(E4(I)/C4-B2(I)^2)
1170   PRINT @4:"approx length to bubble",I,":",B2(I)
1180   PRINT @4:"std dev",E4(I)
1190   PRINT @4:"J"
1200 NEXT I
1210 REM************************************
1220 REM to work out average total length
1230 FOR J=2 TO T
1240   X3=X3+U(J)
1250 NEXT J
1260 X3=X3/(T-1)
1270 PRINT @4:"The average length is",X3
1280 PRINT @4:"The longest molecule's length is",X4
1290 REM************************************
1300 REM histogram looks at ss areas
1310 PRINT "How wide do you want the bars(bp)?"
1320 INPUT W
1330 K=0
1340 H=X3/W
1350 H=INT(H)
1360 DELETE N
1370 N=1
1380 L=0
1390 P=0
1400 Q=0
1410 R=0
1420 IF Q$="y" THEN 1470
1430 PRINT "The first histogram drawn will be the larger of"
1440 PRINT "the two ss lengths measured.  After plotting on plotter,"
1450 PRINT "the second histogram will be using the smaller"
1460 PRINT "of the two ss measured"
1470 FOR I=2 TO T
1480   L=R+1
1490   W1=W
1500   FOR J=L TO O(I)
1510        IF J=R+1 THEN 1540
1520        Q=P+D(J)
1530        GO TO 1550
1540        Q=P+B1(I)+D(J)
1550        IF S(J)=0 THEN 1670
1560        IF Q$="y" THEN 1590
1570        IF S9=1 THEN 1640
1580        IF S9=2 THEN 1690
1590        IF G(J)<>0 THEN 1620
1600        P=Q+S(J)
1610        GO TO 1670
1620        P=Q+(S(J)+G(J))/2
1630        GO TO 1670
1640        P=Q+S(J)
1650        GO TO 1670
1660        P=Q+G(J)
1670        IF W1<=Q THEN 1700
1680        IF W1<P THEN 1730
```

```
1690        GO TO 1780
1700        N=N+1
1710        W1=N*W
1720        GO TO 1670
1730        N=N+1
1740        K(N)=K(N)+1
1750        W1=N*W
1760        IF W1>P THEN 1780
1770        GO TO 1730
1780   NEXT J
1790   N=1
1800   Q=0
1810   P=0
1820   R=O(I)
1830 NEXT I
1840 PAGE
1850 REM***********************************
1860 REM histogram on screen
1870 VIEWPORT 30,100,20,82
1880 WINDOW 0,X4+1500,0,100
1890 AXIS W,10,0,0
1900 X=-W
1910 FOR I=1 TO H+100
1920   X=X+W
1930   MOVE X,0
1940   DRAW X,K(I)/(T-1)*100
1950   DRAW X+W,K(I)/(T-1)*100
1960   DRAW X+W,0
1970 NEXT I
1980 REM***********************************
1990 REM histogram on plotter
2000 MOVE -10,-10
2010 PRINT "Do you want to print on plotter?"
2020 DELETE F$
2030 INPUT F$
2040 IF F$="n" THEN 2190
2050 AXIS @1:W,10,0,0
2060 X=-W
2070 FOR I=1 TO H+100
2080   X=X+W
2090   MOVE @1:X,0
2100   DRAW @1:X,K(I)/(T-1)*100
2110   DRAW @1:X+W,K(I)/(T-1)*100
2120   DRAW @1:X+W,0
2130 NEXT I
2140 IF Q$="y" THEN 2180
2150 S9=S9+1
2160 IF S9=2 THEN 1310
2170 S9=1
2180 GO TO 2200
2190 REM---------------------------------------------------
2200 PAGE
2210 I=MEMORY
2220 PRINT "I_UDK1...read in data file"
```

```
2230 PRINT "I_UDK2...weight average histogram"
2240 PRINT "I_UDK4...mean and std dev each side bubbles"
2250 PRINT "I_UDK5...abs histogram"
2260 END
2270 REM---------------------------------------------------
2280 REM to work out the mean and std dev of each side ofbubble
2290 F=0
2300 M=0
2310 E4=0
2320 X7=0
2330 U=0
2340 A1=0
2350 A4=X4
2360 PRINT @4:"J"
2370 FOR I=2 TO T
2380   FOR J=1 TO X6
2390       FOR J1=O(I-1)+1 TO O(I)
2400           X7=D(J1)+(S(J1)+G(J1))/2+X7
2410           IF X7>B2(J) THEN 2430
2420           GO TO 2440
2430           IF B2(J)+(S(J)+G(J))/2>X7-(S(J)+G(J))/2 THEN 2450
2440       NEXT J1
2450       IF F=0 THEN 2500
2460       M(J)=M(J)+G(O(I-1)+J)
2470       E4(J)=E4(J)+G(O(I-1)+J)^2
2480       U(J)=U(J)+1
2490       GO TO 2530
2500       M(J)=M(J)+S(O(I-1)+J)
2510       E4(J)=E4(J)+S(O(I-1)+J)^2
2520       U(J)=U(J)+1
2530       X7=0
2540   NEXT J
2550 NEXT I
2560 FOR I=1 TO X6
2570   PRINT @4:"average one side of bubble",M(I)/U(I)
2580   PRINT @4:"std dev:",SQR(E4(I)/U(I)-(M(I)/U(I))^2)
2590 NEXT I
2600 F=F+1
2610 IF F=1 THEN 2300
2620 GO TO 2200
2630 REM---------------------------------------------------
2640 REM Abs value histograms
2650 M=0
2660 FOR V=1 TO X6
2670   K=0
2680   PAGE
2690   FOR I=2 TO T
2700        X7=0
2710        FOR J1=O(I-1)+1 TO O(I)
2720             X7=D(J1)+(S(J1)+G(J1))/2+X7
2730             IF X7>B2(V) THEN 2750
2740             GO TO 2760
2750             IF B2(V)+(S(J1)+G(J1))/2>X7-(S(J1)+G(J1))/2 THEN 2880
2760        NEXT J1
```

```
2770        M(I)=ABS(S(J1)-G(J1))
2780   NEXT I
2790 REM************************************
2800 REM mean and std dev
2810   U1=M(2)
2820   U2=M(2)
2830   S1=M(2)
2840   S2=M(2)*M(2)
2850   FOR I=2 TO T
2860        PRINT I
2870        U1=U1 MAX M(I)
2880        U2=U2 MIN M(I)
2890        S1=S1+M(I)
2900        S2=S2+M(I)*M(I)
2910   NEXT I
2920   M1=S1/(T-1)
2930   S3=SQR(ABS(S2/(T-1)-M1^2))
2940   PRINT @4:"mean",M1," std dev",S3
2950   PRINT @4:"min",U2," max",U1
2960 REM************************************
2970   PRINT "How wide do you want the bars?"
2980   INPUT S4
2990   PRINT "Where do you want x to start?"
3000   INPUT X0
3010   PAGE
3020   N2=INT((U1-X0)/S4)+1
3030 REM n2=no of bars
3040   FOR I=2 TO T
3050        FOR J1=1 TO N2
3060             IF X0+(J1+1)*S4=>M(I) AND M(I)=>X0+J1*S4 THEN 3080
3070             GO TO 3100
3080             K(J1)=K(J1)+1
3090             GO TO 3110
3100        NEXT J1
3110   NEXT I
3120   Y1=0
3130   FOR I=1 TO N2
3140        Y1=Y1 MAX K(I)
3150   NEXT I
3160 REM************************************
3170 REM histogram on screen
3180   VIEWPORT 30,100,20,82
3190   WINDOW X0,X0+N2*S4,0,Y1
3200   AXIS S4,1,0,0
3210   X=X0-S4
3220   FOR I=1 TO N2
3230        X=X+S4
3240        MOVE X,0
3250        DRAW X,K(I)
3260        DRAW X+S4,K(I)
3270        DRAW X+S4,0
3280   NEXT I
3290 REM************************************
3300 REM to draw on plotter
```

```
3310    HOME
3320    PRINT "Do you want to print on plotter?"
3330    DELETE F$
3340    INPUT F$
3350    IF F$="n" THEN 3460
3360    AXIS @1:S4,1,X0,0
3370    X=X0-S4
3380    FOR I=1 TO N2
3390         X=X+S4
3400         MOVE @1:X,0
3410         DRAW @1:X,K(I)
3420         DRAW @1:X+S4,K(I)
3430         DRAW @1:X+S4,0
3440    NEXT I
3450    PAGE
3460 NEXT V
3470 GO TO 2200
```

REFERENCES

Bick M.D. and Thomas C.A. "A New method for mapping nucleic acid sequence homology by Electron Microscopy." Nucleic Acid Res. 2 (4), 509, 1975.

Brack C. "DNA Electron Microscopy." Critical Rev. in Biochem. 10, 113, 1981.

Bujard H. "Electron Microscopy of Single-stranded DNA". J. Mol. Biol. 49, 125, 1970.

Busslinger M., Rusconi S. and Birnstiel M.L. "An unusual evolutionary behaviour of a sea urchin histone gene cluster." EMBO 1 (1), 27, 1982.

Cantor C.R. and Schimmel P.R. Biophysical Chemistry Part III: The Behavior of Biological Macromolecules, W.H. Freeman and Company, San Francisco, 1980, 1226.

Childs G., Maxson R. and Kedes L.H. "Histone Gene Expression during Sea Urchin Embryogenesis: Isolation and Characterisation of Early and Late Messenger RNAs of *Strongylocentrotus purpuratus* by Gene-Specific Hybridization and Template Activity." Dev. Biol. 73, 153, 1979.

Coetzee W.F. and Pretorius G.H.J. "Factors which influence the Electron Microscopic Appearance of DNA when Benzyldimethylalkyl ammonium chloride is used." J. of Ultrastructural Res. 67, 33, 1979.

Coggins L.W. (a) "Preparation of nucleic acids for electron microscopy", in Electron Microscopy in molecular Biology- A practial approach, Sommerville J. and Scheer U., Ed., IRL Press Oxford Washington, 1987, 1.

Coggins L.W. (b) "Denaturation and hybridization of nucleic acids", in Electron Microscopy in molecular Biology- A practial approach, Sommerville J. and Scheer U., Rd., IRL Press Oxford Washington, 1987, 31.

Cohen S.N. "The manipulation of Genes." Scientific American (July), 25, 1975.

Cohn R.H. and Kedes L.H. "Nonallelic histone Gene Clusters of Individual Sea Urchins [*Lytechinus pictus*]: Mapping of Homologies in Coding and Spacer DNA." Cell 18, 855, 1979.

Cohn R.H., Lowry J.C. and Kedes L.H. "Histone Genes of the Sea Urchin [*S. Purpuratus*] cloned in E.Coli: Order, Polarity, and Strandedness of the five histone-coding and Spacer Regions" Cell 9, 147, 1976.

Cornelisse J.T.W.A. and van den Berg T.J.T.P. "Profile boundary length can be overestimated by as much as 41% when using a digitizer tablet." J. of Microscopy 136, 341, 1984.

Covarrubias L. and Bolivar F. "Construction and Characterization of New Cloning Vechicles: vi. plasmid pBR329, a new derivative of pBR328 lacking the 482-base-pair inverted duplication." Gene 17, 79, 1982.

Davis R.W. and Hyman R.W. "A Study in Evolution: the DNA Base Sequence Homology between Coliphages T7 and T3." J. Mol. Biol. 287, 1971.

Davis R.W., Simon M. and Davidson N. "Electron Microscope Heteroduplex Methods for Mapping Regions of Base Sequence Homology in Nucleic Acids." Methods Enzymol,. XXI D, 413, 1971.

Delius H. and Mantell N.J. "Characterization by Electron Microscopy of the complex formed between T4 Bacteriophage Gene 32- protein and DNA." J. Mol. Biol. 67, 341, 1972.

Ferguson J. and Davis R.W. "Quantitative electron microscopy of nucleic acids", in Advanced Techniques in Biological Electron Microscopy. II. Koehler J. K.,Ed., Springer-Verlag, Berlin-Heidelberg, 1978, 123.

Frigero N.A and Shaw M.A. "A simple method for the detection of gluteraldehyde" Histochem. Cytochem. 17, 176, 1969.

Gomez B. and Lang D. "Denaturation Map of Bacteriophage T7 DNA and Direction of DNA Transcription." J. Mol. Biol. 70, 239, 1972.

Gross K., Schaffner W., Telford J. and Birnstiel M. "Molecular Analysis of the Histone Gene Cluster of *Psammechinus miliaris* : III. Polarity and asymmetry of the histone-coding Sequences." Cell 8, 479, 1976.

Hentschel C.C. and Birnstiel M.L. "The Organisation and Expression of Histone Gene Families." Cell 25, 301, 1981.

Holmes D.H., Cohn R.H., Kedes L.H. and Davidson N. "Positions of Sea Urchin [*Strongylocentrotus purpurates*] Histone Genes Relative to Restriction Endonuclease Sites on the Chimeric Plasmids pSp2 and pSp17." Biochem. 16 (7), 1504, 1977.

Inman R.B. "A Denaturation Map of the Phage DNA Molecule determined by Electron Microscopy." J. Mol. Biol. 18, 464, 1966.

Inman R.B. "Some factors affecting Electron Microscopy length of Deoxyribonucleic Acid." J. Mol. Biol. 25, 209, 1967.

Inman R.B. and Schnos M. "Partial Denaturation of Thymine- and 5-Bromouracil- containing DNA in Alkali." J. Mol. Biol. 49, 93, 1970.

Johnson D. "A new method of DNA denaturation mapping" Nucl. Acid. Res. 2 (11), 2049, 1975.

Kaumeyer J.F. and Weinberg E.S. "Sequence, organisation and expression of late Embryonic H3 and H4 histone genes from the sea urchin *Strongyloncentrotus purpuratus*." Nucl. Acid. Res. 14 (11), 4557, 1986.

Kedes L.H., Cohn R.H., Lowry J.C., Chang A.C.Y. and Cohen S.N. "The Organisation of Sea Urchin Histone Genes." Cell 6, 359, 1975.

Kinchington D., Dollery A., Greenaway P. and Dumbell K. "The detection of subtle differences between different orthopox virus genomes by Heteroduplex analysis." Virus Res. 1, 351, 1984.

Kleinschmidt A.K. "Monolayer technique in electron microscopy of nucleic acid molecules." Methods Enzymol., 12b, 316, 1968.

Kleinschmidt A.K. and Zahn R.K. "Ueber Deoxyribonucleinsaure - Molekeln in Protein-Mischfilmen" Z. Naturforsch. 14b, 770, 1959.

Knowles J.A. and Childs G.J. "Comparison of the late H1 histonic genes of the sea urchins *Lytechinus pictus* and *Strongylocentrotus purpuratus*." Nucl. Acid. Res. 14 (2), 8121, 1986.

Kudler L., Jones T.R., Russel R.J. and Hyman R.W. "Heteroduplex Analysis of Cloned Fragments of Herpes Simplex Virus DNAs." Virology 124, 86, 1983.

Littlewood R.K. and Inman R.B. "Computer-assisted DNA length measurements from electron micrographs with special reference to partial denaturation mapping." Nucl. Acid. Res. 10 (5), 1691, 1982.

Mandel M. and Marmur J. "Use of Ultraviolet Absorbance-Temperature Profile for determining the Guanine plus Cytosine content fo DNA." Methods Enzymol. 12b, 195, 1968.

Maniatis T. and Fritsch E.F. ; Sambrook J. Molecular Cloning, A Laboratory Manual, Cold Spring Harbour Laboratory 1982.

Marmur J. and Doty P. "Determination of The Base Composition of Deoxyribonucleic Acid from its Thermal Denaturation Temperature." J. Mol. Biol 5, 109, 1962.

Maxam A.M. and Gilbert W. "A new method for sequencing DNA." Proc. Natl. Acad. Sci. USA 74, 560, 1977.

Maxson R., Cohn R. and Kedes L. "Expression and organization of Histone genes." Ann. Rev. Genet. 17, 239, 1983.

Maxson R., Mohun T., Gormezano G., Childs. and Kedes L. "Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin." Nature (London) 301, 120, 1983.

McConaughy B.L., Laird C.D. and McCarthy B.J. "Nucleic Acid Reassociation in Formamide." Biochemistry 8 (8), 3287, 1969.

Overton G.C. and Weinberg E.S. "Length and Sequence Heterogenicity of the Histone Gene Repeat Unit of the Sea Urchin *S. purpuratus*." Cell 14, 247, 1978.

Portmann R., Schaffner W. and Birnstiel M. "Partial denaturation mapping of cloned histone DNA from the sea urchin *Psammechinus miliaris*." Nature (London) 264, 31, 1976.

Pretorius G.H.J. and Coetzee W.F. "Proteus mirabiles Phage 5006M: a Physical Characterisation." J. Gen. Virol. 45, 389, 1979.

Rigby P.W.J., Dieckmann M., Rhodes C. and Bery P. "Labelling deoxyribonucleic acids to high specific activity "in vitro" by nick translation with DNA Polymerase 1." J. Mol. Biol. 113, 237, 1977.

Roberts S.B., Weisser K.E. and Childs G. "Sequence Comparisons of Non-allelic Late Histone Genes and their Early Stage Counterparts." J. Mol. Biol. 174, 647, 1984.

Ruyechan W. T., Casey T.A., Reinhold W., Weir A.C., Wellman H., Strauss S.E. and Hay J. "Distribution of G + C rich regions in Varicella-Zoster Virus DNA." J. Gen. Virol. 66, 43, 1985.

Schaffner W., Gross K., Telford J. and Birnstiel M. "Molecular Analysis of the Histone Gene Cluster of *Psammechinus miliaris* : 11. The arrangement of the five histone-coding and Spacer sequences." Cell 8, 471, 1976.

Schaffner W., Kunz G., Daetwyler H., Telford J., Smith H.O. and Birnstiel M.L. "Genes and spacers of Cloned Sea Urchins Histone DNA analysed by sequencing." Cell 14, 655, 1978.

Sures I., Lowry J. and Kedes L.H. "The DNA Sequence of Sea Urchin [*S. purpuratus*] H2A, H2B and H3 Histone Coding and Spacer Regions." Cell 15, 1033, 1978.

Twigg A.J. and Sherratt D. "Trans-complementable copy-number mutants of plasmid Col E1." Nature (London) 283, 216, 1980.

Villaroel R., Hedges R.W., Maenhaut R., Leemans J., Engler G., van Montagu M. and Schell J. "Heteroduplex Analysis of P.Plasmid Evolution: The Role of Insertion and Deletion of Transposable Elements." Mol. Gen. Genet. 189, 390, 1983.

Vollenweider H.J., Sogo J.M. and Koller T. "A Routine Method for Protein-free Spreading of Double- and Single-stranded Nucleic Acid Molecules." Proc. Natl. Acad. Sci. USA 72 (1), 83, 1975.

Von Holt C., Strickland W.N., Brandt W.F. and Strickland M.S. "More Histone Structures." FEBS letters 100 (2), 201, 1979.

Weinberg E.S., Overton G.C., Hendricks M.B., Newrock K.M. and Cohn L.H. "Histone Gene Heterogeneity in the Sea Urchin *Strongylocentrotus purpuratus* ." Symposia on Quantitative Biology XLII (2), 1093, 1977.

Wellauer P.K., Reeder R.H., Dawid I.B. and Brown D.D. "The Arrangement of Length Heterogeneity in Repeating Units of Amplified and Chromosomal Ribosomal DNA from Xenopus laevis." J. Mol. Biol 105, 487, 1976.

Wu M., Holmes D.S., Davidson N., Cohn L.H. and Kedes L.H. "The Relative Positions of Sea Urchin Histone Genes on the Chimeric Plasmids pSp2 and pSp17 as studied by E.M." Cell 9, 163, 1976.

Westmore B.C., Szylbalski W. and Ris H. "Mapping of deletions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy." Science 163, 1343, 1969.

Wells D.E. "Compilation analysis of histones and histone genes" Nucl. Acid. Res. 14, r119, 1986.

Wolfson J., Dressler D. and Magazin M. "Bacteriophage T7 DNA Replication: A Linear Replicating Intermediate." Proc. Natl. Acad. Sci. USA 69 (2), 499, 1972.

Yager L.N., Kaumeyer J.F. and Weinberg E.S. "Evolving Sea Urchin Histone Genes - Nucleotide Polymorphisms in the H4 Gene and Spacers of *Strongylocentrotus purpuratus* ." J. Mol. Evol. 20, 215, 1984.

18 APR 1988