COMPARATIVE STUDIES ON THE SATELLITE DNA OF RELATED RODENT SPECIES

JAMES ALLAN, B.Sc.

Thesis submitted for the degree of

Doctor of Philosophy.

University of Edinburgh.

APRIL 1974.

CONTENTS

SUMMARY.

The A+T rich satellite DNAs of Apodemus sylvaticus, Apodemus flavicollis and Apodemus agrarius were purified on caesium chloride density gradients. Each satellite DNA was characterised with respect to its properties when subjected to density gradient centrifugation, thermal denaturation and reassociation. From the rate at which A. sylvaticus and A. flavicollis satellite DNAs renature, a kinetic complexity of 3-400 base pairs has been estimated.

The melting properties of sonicated A. agrarius satellite DNA suggest that it consists of two interspersed sequence components which alternate at least once every 500 base pairs. Other evidence is described which supports the hypothesis that this satellite consists of a repeating unit 370 base pairs long. In contrast, the satellite DNA renatures at a rate equivalent to a kinetic complexity of about 100,000 base pairs. It is suggested that the high extent of mismatching found in the renatured duplex was responsible for slowing down the reassociation reaction.

Partial sequence analyses of RNA transcripts of the DNA suggests that all three satellite DNAs are related in sequence. The results are consistent with the interpretation that A. sylvaticus, A flavicollis and A agrarius satellite DNAs are derived from a common, highly diverged simple sequence with a repeat length much shorter than 300 base pairs.

The possible evolutionary relationships of the three satellite DNAs are discussed.

1. INTRODUCTION

The DNA of all eukaryotic organisms studied to date can be divided into two types of molecules; those sequences present in only one copy within the genome (unique sequences) and those present in more than one copy (repetitious sequences). The proportion of the DNA of higher organism genomes that is in the form of repetitive sequences varies considerably even between closely related species (Britten and Kohne, 1968; Laird and McCarthy,1969; Hennig and Walker, 1970; see Bostock, 1972). In many cases the repetitious sequences within one organism may be subdivided with respect to the length of the sequences repeated, and the frequency with which they occur in the genome (Britten et al., 1968; Britten, 1968).

The experimental evidence for the existence of repeated sequences comes from the measurement of the rate of reassociation of the single strands of homologous DNA. Marmur and Doty (1961) were the first to discover that denatured DNA is capable of reforming, in a sequence specific manner, a duplex with the properties of the native molecule. A more detailed investigation of the kinetics of renaturation by Wetmur and Davidson (1968) showed that the reaction is second order and follows a two step mechanism involving nucleation followed by zippering. Nucleation, which is rate limiting, explains the dependance of the rate of the reaction upon temperature, salt concentration and nucleic acid concentration (Britten and Kohne, 1966; Wetmur et al., 1968). After each successful nucleation the amount of zippering that can take/

take place depends upon the molecular weight of the DNA, and the rate of renaturation is found to be proportional to the square root of molecular weight. However the most important aspect of this work was the demonstration by Wetmur and Davidson (1968) that the rate of renaturation of a DNA is inversely proportional to its genomic complexity, which they defined as the total number of DNA base pairs in non repeating sequences.

Using a variety of DNAs of differing complexity, Waring and Britten (1966) and Britten et al., (1966) made the original observation of the above relationship but were surprised to find that the DNAs of some eukaryotic organisms, such as mouse, did not renature as predicted. Although they had sheared their DNA in order to avoid the rapid formation of high molecular weight networks, (Bolton, Britten, Cowie, Roberts, Szafranski and Waring, 1966) significant amounts of renaturation took place at rates much faster than expected for DNAs of such large genomic complexity. To account for this phenomenon Britten et al. (1966) proposed that some eukaryotic DNAs may contain sequences present in more than one copy and thus, because they are at a relatively higher concentration, they will renature more rapidly than those sequences present only once. Indeed this had been the explanation for the formation of high molecular weight networks upon renaturation of eukaryotic DNA (Bolton et al., 1966).

There are two advantages to be gained by expressing renaturation kinetics in terms of Cot values. Cot is the product of the initial DNA concentration in moles of nucleotide per litre /

litre and time in seconds. By plotting the Cot of a reaction on a logarithmic scale DNA fractions differing in their rates of renaturation by a factor of 10^6 can be conveniently represented. Also the Cot value at 50% renaturation (Cot $\frac{1}{2}$) is equivalent to the reciprocal of the rate constant, k_2, and is therefore directly proportional to sequence complexity.

Most higher organisms show a broad spread with respect to the frequency of sequence repetition within their genomes. In general however, three classes of molecules are often distinguishable. Those sequences which renature in the range of Cot values 10^2 - 10^4 are thought to be present only once within the genome and constitute the slow fraction. Present in all higher organisms DNAs is an intermediate fraction which is characterised by its renaturation between Cot values of 10^{-2} and 10^2. These sequences are thought to be present in 10^2 - 10^5 copies. The remainder of the DNA which reassociates with a Cot of less than 10^{-2} is the fast fraction and is not always present or apparent in every higher organism DNA. The fast fraction represents sequences present in more than 10^5 copies. These fractions are not always individually distinguishable and often overlap each other as is the case, for example, in Xenopus laevis and salmon DNA (Britten et al., 1968).

In some cases we are able to estimate the length or complexity of the sequences that are repeated from the percentage of DNA represented by a specific frequency of renaturation. For example in mouse, 10% of the DNA renatures with an average Cot value of 10^{-3} and must therefore be made up of sequences about 300 base pairs long repeated 10^6 times. However, we are only able/

able to make such a conclusion because we know that mouse fast fraction is a homogeneous class of sequences, unlike the 20% intermediate fraction of mouse DNA which is probably made up of a variety of sequence lengths present in different frequencies.

We are able to define mouse fast fraction as a homogeneous class of sequences through significant discoveries made in another field of DNA studies.

When analysed by density gradient centriguation the DNA of higher organisms rarely gives the homogenous banding pattern characteristic of bacteria or phage DNAs (Meselson, Stahl and Vinograd, 1957). More surprising, however, was the discovery that some higher organism DNAs, when subjected to density gradient centrifugation, exhibited the presence of minor components banding in positions well removed from the majority of the DNA. For example, mouse DNA was shown to contain a light banding component (Kit, 1961 a,b), and heavy components were found in guinea pig (Kit, 1961 b) and calf DNA (Sueoka, 1961) on analysis in neutral caesium chloride gradients.

Although the term 'satellite' DNA had been coined to describe minor components found in caesium chloride density gradients, the term had to be redefined (Walker, 1971) when it was discovered that certain DNA fractions bind heavy metals, such as Hg^{++} and Ag^{++} to an extent different from that of the bulk DNA (Nandi, Wang and Davidson, 1965) and that these fractions could afterwards be separated in caesium sulphate gradients. Using this technique, Corneo et al. (1968 a,b; 1970 a,b) were able to demonstrate the existence of satellite components in the DNA of man, guinea pig and calf, which are not visible in neutral caesium chloride/

chloride gradients.

Naturally, because of their widespread occurence in the DNA of higher organisms (Arrighi, Mandel, Bergendahl and Hsu, 1970; Hennig et al., 1970) and the relative ease with which they could be isolated after centrifugation in fixed angle rotors (Flamm, Bond and Burr, 1966 a), satellite DNAs became the subject of a great deal of intensive research. As a result, satellite DNAs have been shown to possess, in many cases, a number of physical properties which distinguish them from the bulk DNA.

Although the density at which DNA bands in a caesium chloride gradient is a function of its base composition for bacterial DNA (Schildkraut, Marmur and Doty, 1962) many satellite DNAs band in positions characteristic of DNA of higher G + C content than has been chemically determined for those satellites (Walker, 1971). The ability of bacterial DNA to bind heavy metals is also a function of G + C content. However, the three satellite DNAs of guinea pig, which have similar base compositions and identical densities in caesium chloride (Flamm, Walker and McCallum, 1969 b), band as quite different densities when complexed with heavy metals and analysed in caesium sulphate gradients (Corneo et al., 1970a). Due to their abnormal bias in G+T content, it is often possible to isolate the single strands of satellite DNAs on alkaline caesium chloride gradients (Flamm, McCallum and Walker, 1967; Corneo et al., 1968b) although this is certainly not always the case (Hennig et al., 1970; Walker, Flamm and McLaren, 1968). Furthermore although the temperature at which DNA in solution denatures is normally a function of its base/

composition (Marmur and Doty, 1961), some satellite DNAs, including mouse (Flamm et al., 1967), guinea pig (Corneo et at., 1968a), and a number of crustacea satellite DNAs (Skinner, Beattie, Kerr and Graham, 1970) exhibit thermal transitions at temperatures higher than expected from their known G+C content.

Recently, bacterial restriction enzymes have been used to study the structure of satellite DNAs (Southern 1974a; Cooke, 1974). These enzymes cleave double stranded DNA in both strands, the breaks occuring within a short defined sequence (Kelly and Smith, 1970) which is characteristic of the enzyme used. The fragments produced can be separated by gel electrophoresis, which enables their size and yield to be calculated. As satellite DNA is usually a tandemly repeating sequence, restriction enzymes may have either of two actions upon them: either the DNA will remain high molecular weight after enzyme treatment if the repeating unit does not contain the restriction site; or, if the restriction site is contained within the repeating sequence, then the satellite should be digested into fragments with the length of the repeating unit. By this method, Southern(1974a) has estimated that the basic repeat length of mouse satellite DNA is 240 base pairs. However, from the pattern of fragments produced by the original digest, and the renaturation behaviour of the 240 base pair fragment, the presence of another repeating unit is indicated. This sequence appears to be only 120 base pairs long and is contained within the longer repeat. However, these discrepancies in repeat length with the original estimate of about 300 base pairs (Waring et al.,1966) for mouse satellite DNA are small/

compared to the estimate obtained from partial sequence analysis of that DNA.

Analysis of the pyrimidine oligonucleotides produced by dephenylamine-formic acid degradation of the DNA (Burton and Peterson, 1960) remains the only direct method of sequencing DNA. Alternatives include sequence analysis of either copy RNA, made by transcription of the DNA with RNA polymerase, or of ribo-substituted DNA copies. In all cases the analysis is simplified if the single strands of the satellite can be isolated.

To date, ten satellite DNAs have been partially or completely sequenced and in every case the basic repeating sequence is very simple. Pyrimidine tract analysis of guinea pig α satellite suggests that the DNA consists of a sequence only six basepairs long, $\frac{CCCTAA}{GGGATT}$, which has been repeated many times (Southern, 1970). However, it is also evident from the analysis that a great deal of sequence divergence has taken place in the DNA due to the introduction of mutations. A similar anlaysis of mouse satellite DNA (Biro, Carr-Brown, Southern and Walker, 1974), although indicating a more complex sequence, suggests a repeat length of about 12 base pairs. However, analysis of the RNA transcribed from this DNA, shows that some of the short tracts from the pyrimidine tract analysis derive from sequences related to the longer tracts so that the length of the basic repeating sequence is almost certainly shorter than 12 base pairs.

The pyrimidine tract analysis of two other satellites of guinea pig (Southern, 1973) and three satellite DNAs of Dipodomys ordii (Szala and Southern, 1974) suggest that these DNAs have also evolved/

evolved from very short repeating sequences. More recently Gall and Atherton (1973) have managed a complete sequence analysis of the three satellite DNAs which constitute approximately half the genome in Drosophila virilis. By sequencing cRNA transcripts of the isolated strands of the satellite DNAs they were able to show that the DNAs are in all cases repeating heptanucleotides related to each other by simple base pair changes.

Perhaps the most significant aspect of the sequence analyses of satellite DNAs is the contribution that they have made to our understanding of the origin and evolution of repeated sequences. It had been generally accepted that satellite DNAs were produced rapidly in evolution and thus they were apparently species specific (Hennig et al.,1970). However, in the case of guinea pig α satellite, the former statement is far from true. Many of the diverged pyrimidine tracts found in the sequence analysis of this DNA are present in quantities significantly greater than would be expected from the random introduction of mututions into a pure repeating hexanucleotide. To explain this phenomenon Southern (1970) proposed that amplifications had occured in stages and that mutations were introduced between the steps of multiplication. By this mechanism, distinct satellite DNAs can arise from the same ancestral simple sequence if different sections of the sequence are taken from subsequent multiplications. This appears to be the case for guinea pig satellite DNAs II and III which share a common ancestral sequence but have different physical properties (Southern 1973).

Satellite DNAs from distinct yet related species can also share/

share a common ancestral sequence. Sutton and McCallum (1972) were able to detect substantial amounts of cross reassociation, at low stringency conditions, between the separated strands of Mus musculus satellite DNA and five satellite DNAs of three other related Mus species. The duplexes formed had low stability but, surprisingly, melted over a very narrow temperature range some 20-25^{o} below the melting temperature of the native duplex. This result indicates that although the reassociated duplexes are highly mismatched, the mismatching is evenly spread over the whole length of the duplex. Sutton et al.(1972) interpreted these results as indicating that the satellite DNAs of the Mus species had all arisen from a common ancestor, which consisted of a highly repetitious but highly diverged sequence.

Sequence analysis of satellite DNAs also highlights the fact that although satellites can show distinct physical properties they may be very closely related in sequence. This is most strikingly demonstrated by the three Drosophila virilis satellite DNAs. Although they differ only in one base pair from the common heptamer sequence (Gall et al., 1973) they are easily distinguished with respect to a number of physical properties (Blumenfeld, 1973). Also, the hybrids formed upon renaturation of the separated strands of the three satellite DNAs have greatly reduced stability. In fact there appears to be no duplex formation between the strands of satellites II and III, the only case in which there is a two base pair difference within the heptamer (Blumenfeld, 1973).

Sequencing techniques are of course the ultimate method for the determination of the structure and sequence relationships of repetitive DNA. However, this method can only be applied successfully/

successfully to the study of a relatively simple sequence. For this reason repetitive DNAs are more frequently characterised with respect to their renaturation behaviour. Nevertheless this method also has disadvantages inherent in the techniques used and the interpretation of the results obtained.

Renaturation is often monitored by absorbance changes in a conventional spectrophotometer and therefore relatively high concentrations of DNA have to be used. As most satellite DNAs renature rapidly, it is necessary to slow the reaction down by reducing the salt concentration. Although the dependence of renaturation rate on salt concentration has been determined (Wetmur et al., 1968; Britten and Smith, 1970), this relationship may not apply to the special sequences found in satellite DNAs. By working with very low concentrations of radioactively labelled DNA and using hydroxylapatite (Tiselius, Hjerten and Levin,1956) to distinguish between single and double stranded DNA (Miyazawas and Thomas, 1965; McCallum and Walker, 1967) the effects of salt concentration can be overcome. However this method is not capable of distinguishing between partially single stranded duplex molecules and native molecules (Sutton, 1971) and will tend therefore to overestimate the extent of renaturation.

However, the above problems, although of some importance, have been dwarfed by the controversy surrounding the effect that mismatching may have upon the rate of renaturation and more recently by the suggestion that highly repetitive DNAs do not obey one of the principles of renaturation, stated by Wetmur and Davidson (1968). These two problems, discussed below throw doubt on the use of measurements/

measurements of reassociation rates as a means of measuring the sequence complexity of repeated sequences.

When a reassociated, repetitive fraction of DNA is subjected to thermal denaturation, it very rarely melts at the same temperature as the native duplex. For example, the Δ Tm (Difference in melting temperature between a native and reassociated DNA) of mouse satellite DNA is about 5°C (Sutton, 1971) and for guinea pig $\propto$ satellite is closer to 23°C (Corneo et.al., 1970a). The reduction in thermal stability is principally due to imperfect sequence complementarity between the reassociated strands and thus supplies a convenient measure of the extent of mismatching in reassociated duplexes. The most commonly used conversion factors vary from 1% (Ullman, 1970; Wang and Kallenbach, 1971; McCarthy and Farquhar, 1972) up to 1.5% (Laird, McConnaughty and McCarthy, 1970) mismatching for each 1°C drop in melting temperature.

Southern (1971) was the first to suggest that this imperfect pairing during reassociation could have a very significant effect upon the rate of renaturation. His predictions were apparently confirmed by Sutton and McCallum (1971) who, after fractionating reassociated mouse satellite DNA into a number of classes of varying degrees of mismatch, found that their rates of renaturation were highly dependent upon the Δ Tm they exhibited with respect to native DNA. They estimated a 6 fold reduction in renaturation rate for a 10°C.Δ Tm. However, McCarthy and Farquhar (1972) and Bonner, Brenner, Neufeld and Britten (1973) were unable to measure such a large effect of mismatch upon rate using both deaminated bacterial and phage DNAs, and also cross-reassociation of/

of closely related bacterial DNAs. Their estimates were a more conservative 2-3 fold reduction in rate for $10^{\circ}C \Delta Tm$.

Recently Hutton and Wetmur (1973) have shown that on renaturing a crude preparation of mouse satellite DNA and following this by exonuclease S_1 digestion, (Sutton, 1971) the molecular weight of the reacting molecules is reduced. If this procedure is repeated over a number of cycles of renaturation the molecular weight continues to decrease. However, the surprising aspect of this work was that as the molecular weight of the renaturing molecules decreases, the rate constant for the reaction increases. A possible explanation for this peculiar result has been proposed by Chilton (1973). She argues that for a short repeated sequence within a longer DNA strand, k_2 will not exhibit the usual molecular weight dependence (Wetmur et.al., 1968). Increasing the molecular weight of the reacting strands will not linearly increase the length of duplex formed per successful nucleation. The length of the duplex formed should instead be proportional to the repeat length of the DNA. This being the case, then the rate of renaturation, rather than being directly proportional to, should be inversely proportional to the square root of the molecular weight. Indeed Hutton and Wetmur's (1973) results do obey such a relationship.

We shall now turn our attention to the aspects of satellite DNAs which are more related to their possible function. Although it had been known for some time that satellite DNAs are not confined to one type of tissue (Kit, 1961 b, 1962; Szybalski,1961; Chun and Littlefield, 1963), it was not until 1966 that a number of/

of workers showed that they are of nuclear origin (Corneo et al., 1966; Borst et al., 1966; Flamm et al., 1966 b). Maio and Schildkraut (1969) were able to show that satellite DNA is equally distributed amongst the different size classes of chromosomes, although earlier they had shown that the sequences are more concentrated in nucleolus associated chromatin (Schildkraut and Maio, 1968). By using the technique of cytological in situ hybridization, Jones (1970) and Pardue and Gall (1970) confirmed the presence of mouse satellite sequences on all types of mouse chromosomes with the possible exception of the Y chromosome. The more significant aspect of this work was the discovery that hybridization is concentrated primarily in the centromeric heterochromatin. However, in situ hybridization is not a sensitive technique. This is shown by the low level of hybridization achieved, equivalent to saturation of only 8% of the satellite sequences (Jones, 1970). The results, therefore, do not preclude the possibility that lower concentrations of satellite sequences may be spread throughout other regions of the chromosome. In fact the results of Hennig, Hennig and Stern (1970) suggest that for three closely related Drosophila species, "young" satellites are found first in the centric heterochromatin but later in their evolution they appear at many other sites. The association of satellite sequences with other types of sequence has also been inferred by Flamm et al., (1969a) who found significant amounts of satellite sequences retained within the main band after centrifugation on density gradients.

The close relationship between satellite DNA and heterochromatin is also evident from the fact that the 70% of nuclear DNA

extracted by washing chromosomes in high salt does not contain the satellite sequences, (Maio et al., 1969) which are retained in the resistant fraction. These observations have been extended by Yasmineh and Yunis (1969, 1970a), who have isolated chromatin fractions from nuclei by mild ultrasonic treatment followed by differential centrifugation. They resolved three fractions which they called heterochromatin, intermediate and euchromatic DNA. Satellite DNA is greatly enriched in the heterochromatic fraction for mouse (Yasmineh and Yunis 1969, 1970a), guinea pig (Yunis and Yasmineh, 1970 b) and calf (Yasmineh and Yunis 1971). However neither Mattochia and Cummings (1970) nor Bostock and Prescott (quoted in Bostock, 1972) have been able to find such high satellite enrichment in heterochromatin.

Lima-de-Faria (1969) and Ohno, Kaplan and Kinosita (1957) have shown that mouse heterchromatin is replicated late in the S phase of the cell cycle. The satellite DNAs of mouse (Bostock and Prescott, 1971; Tobia, Schildkraut and Maio, 1970) and Kangaroo rat (Bostock, Prescott and Hatch, 1972) have also been shown to replicate late in S phase. Late replication is one definition of heterochromatin, and this shared property strengthens the association between the satellite DNAs and heterochromatin.

Before going on to describe the possible functions of repetitive DNA we should mention a number of mechanisms by which these sequences may be produced. With respect to the relatively large differences in the size of sequences that are amplified, it seems probable that more than one method may be used for their production. Given a short oligodeoxyribonucleotide as a template, bacterial DNA polymerase is capable of producing a long chain of tandemly linked/

linked copies of the template (Wells et al., 1967). This structure is similar to the basic sequence of satellite DNAs. The amplification of longer sequences may be more successfuly accomplished by continuous replication from a circular DNA as is the case for phage DNA replication (Gilbert and Dressler, 1968). In the latter case the circular DNA may take the form of a loop projecting from the chromosome (Keyl, 1965). Finally, Smith (1973) has shown that repeated unequal crossing over in a family of serially duplicated sequences can lead to the gradual 'turn over' of these sequences. The result of this phenomenon is that the surviving sequences are all descendants of only one of the original sequences and are related to each other by a highly structured family tree, in which the sequences fall into a few relatively distantly related subgroups of closely related sequence.

Broadly speaking, the theories proposed to account for the presence of repetitious sequences within the genomes of eukaryotes can be divided into four classes. They have been reviewed by both Walker (1971) and Bostock (1972).

A haploid genome containing 3×10^9 base pairs is probably unable to withstand a mutational load equivalent to 8.3×10^{-9} changes per amino acid site per year (Corbin and Uzzell, 1970) if all of the base pairs are in coding sequences (Muller, 1967; Ohno, 1971). For this reason Ohno (1970; 1973) has concluded that the bulk of DNA in the mammalian genome must have no function other than being present in large amounts. Functionless DNA base sequences may be useful in a negative way. Chromosome breaks often occur in mammalian cells and the rejoining of the broken ends may be/

be accompanied by deletion, addition and substitution of base pairs. These events will be harmless so long as the break occurs within a function-less base sequence. Similarly large chromosomal changes, which often accompany speciation, could be accommodated within the satellite rich centromeric heterochromatin.

Callan (1967) in his 'Master-Slave' hypothesis also attempts to explain the large and variable DNA content of eukaryotic genomes. Basically, this model predicts that genes are present in many exact copies which are tandemly linked. Both Whitehouse (1967) and Thomas (1970) have expressed support for this theory and recently Thomas, Hamkalo, Misra and Less (1970) have provided corroborative experimental evidence. Although a number of coding sequences are known to be present in more than one copy (Birnstiel, Chipchase and Spiers, 1971; Kedes and Birnstiel, 1971) there are three lines of evidence which argue against this model as a general structure for all genes and an explanation for the high and variable C-value of higher organisms. In many instances the majority of the DNA of higher organisms is present only as unique sequences. Also, although repetitious DNA is known to be transcribed to some extent (Church and McCarthy, 1967; Paul and Gilmour, 1968) the product probably does not leave the nucleus to be translated (Shearer and McCarthy, 1967; McCarthy, Shearer and Church, 1969). Finally, if the genome is composed of exact tandem repeats, then the fragments produced by digestion with a restriction endonuclease should not be able to form circles such as those described by Thomas et al., (1970). However guinea pig, mouse and HeLa cell DNA all form a high proportion of circles after such treatment (Southern and R oizes 1973).

There are a number of theories which suggest that repetitive DNA may be involved in the control of protein synthesis. Georgiev (1969) has proposed a model which is an extension of the bacterial operon structure (Jacob and Monod, 1961) in which protein molecules control transcription by binding to acceptor sites adjacent to the structural gene. In this model many acceptor sites are responsible for the control of each gene and may also be shared in common with other genes. Britten and Davidson (1969) have proposed a more complex scheme involving RNA as the regulating molecule. This RNA may be transcribed from repeated sequences and the acceptor sites, as in the previous model, may be held in common between many structural genes. Paul (1972) prefers to retain the idea of multiple gene copies with the distinction that these sequences may diverge during relaxed selection. However the controlling elements common to each gene copy may not diverge, and will be maintained as a homogeneous family of repeated sequences. Each of these models accounts for the presence of repeated sequences, their interspersion between non-repeated DNA (Britten and Smith, 1970) and also the properties of nuclear heterogeneous RNA.

However there are difficulties with the above models. The majority of repetitive sequences do not form highly stable duplexes upon renaturation (Sutton, 1971) and therefore show no indication that they have been selectively maintained within the genome. Much of the intermediate DNA has possibly evolved from satellite sequences (Walker, 1971) and will therefore be in the form of highly diverged highly repeated very short sequences. It is difficult to envisage such sequences showing specific binding or coding properties with respect to proteins.

In order to overcome these problems, transcriptional control models which are not dependent upon sequence recognition have been developed. Crick (1971) proposes that repetitive sequences adjacent to recognition sites for regulatory activators may be involved in local unwinding of these sites to aid in transcription. In this context the important feature of the repeated DNA is not its sequence but rather its secondary structure. Sutton (1972) derives his model from the observation that newly formed repeating sequences, such as satellites, are often found in highly condensed inactive heterochromatin. He proposes that as these sequences diverge, a gradual weakening in the extent of condensation in chromatin will occur and that a gradation of packing will be the result. He suggests that such structures may act as a coarse control for transcription.

Finally, Walker, Flamm and McLaren (1969) have proposed a number of functions specifically for satellite DNAs. They argue that as these sequences are not transcribed (Flamm et al., 1969 a; Walker, 1971) they may only operate in a structural sense. The functions proposed include (1) sites for initiation of DNA replication, (2) special regions for recombination, (3) pairing sites in meiosis, (4) centromere recognition sites, and (5) sequences involved in determining the complex and changing patterns of chromosome structure by providing specific folding sites.

It is clear from the diversity of the theories described above that our knowledge of the organisation of higher organism genomes is at an early stage in development. We have much to learn about the structure of the different classes of repetitive sequences especially/

especially in terms of their possible evolutionary relationship. If satellite DNAs are evolutionary precursors to the intermediate and unique fractions of DNA, a more detailed understanding of their structure and evolution will surely reflect on the significance of the other fractions.

Apodemus sylvaticus, Apodemus flavicollis and Apodemus agrarius are three closely related species of wood mice. Each has a light satellite DNA which can be isolated on caesium chloride gradients (Hennig et al., 1970). The purpose of this study was to isolate and characterise each satellite DNA with respect to their physical properties. To determine whether the satellites are related in sequence, a preliminary sequence analysis was also conducted.

Abbreviations .

EDTA	Ethylene diamine tetracetic acid disodium salt.
Tris.	Tris (hydroxymethyl) amino methane.
S.S.C.	Standard saline citrate (0.15m-NaCl, 0.015m-tri-sodium citrate).
1m-phosphate	0.5m-Di-sodium hydrogen orthophosphate
buffer	0.5m -Sodium di-hydrogen orthophosphate.
Cs_2SO_4	Caesium sulphate.
CsCl	Caesium chloride.
A,G,C,U.	Adenosine, guanosine, cytidine, uridine.
ATP,UTP etc.	5' - (pyro) triphosphate of adenosine, uridine, etc.

2. MATERIALS AND METHODS

2.1 Materials

All chemicals used, where possible, were of 'Analar' grade from the British Drug House Ltd., London. The sources of all other chemicals and materials, with the exception of those listed below, are indicated in the text.

I.a. Hydroxyapatite preparation

Large crystals of hydroxyapatite were prepared by the procedure of Tiselius et. al. (1956) as modified by Miyazawa and Thomas (1965).

I.b. Preparation of E. Coli RNA polymerase

E. Coli RNA polymerase was prepared by the method of Burgess (1969) involving the glycerol gradient centrifugation procedure.

I.c. Preparation of thin layers.

Polyethylenemine impregnated cellulose layers were prepared by the procedure of Randerath and Randerath (1967). The support for the layers was subbed backing of X-ray films (kindly supplied by Kodak). After spreading and drying, the layers were stored at room temperature for at least one week before further use. For RNase T_1 fingerprints the size of the layers was 15 x 12.5 inches. Smaller layers (7.5 x 7.5 inches or 11 x 11 inches) were used for the analysis of pancreatic RNase digests. Before chromatography the layers were washed in distilled water (1 min.), 10% (w/v) NaCl(5 min.), distilled water (1 min.), 2 m-formic acid (pH 2.3 with pyridine) 10 min.) and finally in distilled water.

I.d Scintillation Fluid.

Tolulene based scintillation fluid was prepared by dissolving 5g. 2-5- Diphenyloxazole (PPO) and 300 mg. 1,4 bis-2(4- methyl-5-

Phenyloxazole)- Benzene (POPOP) in 1 litre of tolulene.

I.e. Mice

Apodemus sylvaticus, originating from Midlothian (Scotland), Apodemus flavicollis, from Yugoslavia and Apodemus agrarius from Bohemia were bred in the laboratory.

2.II METHODS

II.a. DNA preparation

DNA was prepared from liver, kidney, testes and spleen by extraction with cold phenol and sodium dodecyl sulphate as described by Walker and McLaren (1965). For caesium sulphate density centrifugation the DNA was further purified by iso-propanol precipitation. This was achieved by slowly adding 0.58 volume (of DNA solution) of iso-propanol to the DNA solution made 0.3 M in sodium acetate. The DNA was spooled out on a glass rod and redissolved in distilled water.

II.b. Preparative density gradient centrifugation.

The satellite fractions of the DNAs were isolated from preformed CsCl density gradients (Brunk and Leick, 1969). The gradients were formed by layering a solution of CsCl at a density of 1.62 gms./ml., and containing 400-500 µg DNA, onto the same volume of CsCl solution at a density of 1.8 gm./ml. The two layers were then covered with paraffin oil and centrifuged at 26,000 r.p.m. for 70 hours in a fixed angle rotor (Flamm et. al., 1966a) on an M.S.E. 65 centrifuge.

0.2 ml. fractions were collected from the gradients by means of a finger pump connected to a perspex float, which had been carefully lowered onto the surface of the solution. the U.V. absorbance/

absorbance of each fraction was measured on a Beckmann DB-G spectrophotometer. The satellite fractions were collected and pooled.

By diluting the pooled fractions and making the CsCl solution up to a density of 1.62 gm/ml., the gradients could again be preformed and the fractionation repeated. Altogether 3 cycles were performed with A. sylvaticus and A. flavicollis but only two cycles were necessary with A. agrarius.

LL.c.Analytical density gradient centrifugation.

Analyses of the DNA at various stages throughout the fractionation procedure were carried out on a M.S.E. analytical ultracentrifuge equipped with standard U.V. optics. The solution to be centrifuged was in each case brought to a DNA concentration of 0.06 O.D. units/ml. at an initial CsCl (Harshaw grade) density between 1.7050 and 1.7213 gm./ml.

Micrococcus lysodeikticus DNA at a concentration of 1 μg/ml. was added as a marker.

Centrifugation was carried out for at least 17 hours at 45,000 r.p.m. At equilibrium photographs were taken on Kodak commercial fine grain film and were anlysed on a Joyce-Lobel double beam recording microdensomiter, Mark IIIC.

The buoyant densities of the specimen DNAs were determined by the method of Szybalski (1968).

$Ag+Cs_2SO_4$ density gradients were prepared by the procedure of Jensen and Davidson (1966) as modified by Corneo et al. (1970b). The DNA, dissolved in 0.1m-Na_2SO_4 was added to 0.1m borate buffer pH 9.2. 10^{-3} m-$AgClO_4$ was then added to give a final Ag^+ ion to DNA-phosphate molar ratio of 0.27, and the density was adjusted to

1.508 gm./ml. with Cs_2SO_4 at 1.25 gm./ml. (w/v).

Alkaline CsCl gradients were prepared by adding 0.02 ml. of 1% SLS and 0.04 ml. of 1m-NaOH to 0.4 ml. of DNA solution followed by enough CsCl (Harshaw grade) to bring the density to 1.75 gm./ml.

II.d.Treatment of DNA prior to optical studies

Pooled satellite fractions from the final preparative centrifugation were dialysed overnight against 0.12m-phosphate buffer, pH 6.8. The molecular weight of the DNAs was reduced by sonication. The samples, cooled on ice, were given two ten second treatments with a Dawe Soniprobe at a setting of 4. The sample was allowed to cool for 1 minute between treatments. To remove U.V. absorbing material other than DNA, the samples were passed over a 0.5 ml. hydroxyapatite column. The satellite DNAs, eluted with 0.3m-phosphate buffer were dialysed overnight against 0.01m-NaCl, 0.001m-Tris/HCl (pH8), 0.0001m-EDTA. Before melting, all samples were dearated and finally given a short centrifugation to remove dust particles.

II.e.Melting and Reassociation Experiments.

All melting and reassociation experiments were performed in a Zeiss PMQ11 spectrophotometer equipped with a thermostated, circulating water bath. Melting was carried out in 0.01m-NaCl, 0.001m-Tris/HCl(pH8), 0.0001m-EDTA. The temperature was increased at a rate of 0.5° per minute. Absorbance readings, taken at 260 μm, were corrected for thermal expansion.

For reassociation experiments, DNA samples in the melting buffer, were heat denatured at90° for five minutes and then equilibrated for 5-10 minutes at the reassociation temperature. Reassociation was initiated by the addition of an appropriate

volume (not more than 5% of the sample vol.) of 1m or 5m-NaCl, followed by rapid mixing (by inversion). The first reliable reading could be taken 15 secs. after the salt was added. All Cot values (product of initial DNA concentration and time) have been corrected to the equivalent value in 0.18m-Na^+, using the tables of Britten and Smith (1970).

II.f.Transcription of Satellite DNAs.

The satellite DNAs were transcribed with E. Coli. RNA polymerase prepared by the method of Burgess (1969) involving the glycerol gradient centrifugation procedure. 2µg of DNA were incubated with 5 units of RNA polymerase in a reaction mixture of 50µl , containing 0.04m-Tris/HCl (pH 7.9 at 25°), 0.01m-$MgCl_2$, 0.1 mm-EDTA, 0.1mm-dithiothreitol, 0.15m-KCl, 0.04 mm$K_2H_2PO_4$, 0.1 mm-GTP, CTP and ATP and 0.1 mm ^{32}P,∝ - labelled UTP (specific activity of 10 Ci./m.mole). When labelling with ^{32}P,∝ -labelled ATP the specific activity was 1 Ci./m.mole. After incubation at 37° for 25 mins., 50µg /ml. DNase (Worthington RNase free) was added and incubation was continued at 37° for a further 20 mins. The sample was then chilled on ice, diluted to 2mls. with 0.2% SDS, 0.1m-NaCl, and extracted three times with buffer saturated, distilled phenol. The RNA was then precipitated overnight with two vols. of ethanol at -20°. After pelleting, the cRNA was redissolved in distilled water and passed through a column of Sephadex G-75 to separate the transcribed RNA from free triphosphates and oligonucleotides. The excluded peak was pooled, dried and stored at - 20°.

11.g.Fingerprinting of cRNAs.

Approximately 500 x 10^3 cpm of cRNA, with enough t-RNA from

E. coli. to make 10 µg total RNA, was dissolved in 10 µl, of 10 mm-Tris/HCl (pH 7.4), 0.1 mm-EDTA and treated with RNase T_1 (Sankyo) or pancreatic RNase (Worthington) at an enzyme to substrate ratio of 1:20 for 30 mins. at 37°. The reaction was performed on a polythene sheet under a cap sealed with silicone grease, and was terminted by drying rapidly in a vacuum. Electrophoresis was performed on cellulose acetate strips in either 7m-urea, 5% acetic acid (pH 4.3 with pyridine) for RNase T_1 digests or 7m-urea, 5% acetic acid (pH3.5) for pancreatic RNase digests, at 5 kV for 2.5 hours and 3 kV for 1.0 hour respectively (Sanger and Brownlee, 1967). The oligonucleotides were transferred by a modified procedure giving quantitative transfer (Southern, 1974b) to pre-washed layers of cellulose impregnated with polyethylenemine (Randerath *et al*., 1967). These were then developed with 1.6m-formic acid (pH 3.8 with pyridine) for RNase T_1 digests or 2.0m-formic acid (pH 3.5 with pyridine) for pancreatic RNase digests. The plates were dried by evaporation at room temperature and exposed to X-ray film (Kodak Blue Brand) for 1-3 days. The nucleotides were located, cut out and their radioactivity was measured. Nucleotides eluted from the layer with 2.0m-triethylammonium carbonate, were dried down on a polythene sheet, taken up in distilled water and dried down again. This procedure was repeated three times to remove traces of the eluting agent.

Further analysis of the RNase T_1 nucleotides were carried out by digestion with pancreatic RNase. Sequences from pancreatic RNase digests were further analysed by digestion with RNase T_1. The conditions used for enzyme digestions were as above. The products from either of these secondary digestions were separated

by electrophoresis on Whatman DE81 paper at pH 3.5 (Sanger et al., 1967).

II.h. Nearest neighbour analysis.

Nearest neighbour analysis of the total cRNA and of the primary or secondary digestion products was accomplished by alkaline digestion followed by ionophoresis. The RNAs or nucleotides were digested in 10µl of 0.2m-NaOH at 37° for 16 hours in a sealed capillary. After incubation the products were separated on Whatman No.52 paper by electrophoresis at pH 3.5 and 60 volts/cm. for 1-1.5 hours (Sanger et al., 1967). The paper was then dried and exposed to X-ray film. The products were located, cut out and their radioactivity was measured.

3. RESULTS AND DISCUSSION

The presentation and discussion of results has been divided into three sections. The satellite DNA of A. agrarius displays many physical properties quite distinct from those of the other two satellite DNAs and indeed from any other satellite DNA that has been described. The physical properties of this DNA are therefore dealt with separately. With respect to their partial sequence analyses the three satellite DNAs have been dealt with together.

3.1 The physical properties of A. sylvaticus and A. flavicollis satellite DNAs

I.a. Density gradients

The analyses of the DNAs in density gradients of caesium chloride of the species used here, have been described elsewhere (Hennig et al., 1970). A. sylvaticus and A. flavicollis DNAs both contain light satellites amounting to 8% and 6% of the total DNA respectively (Fig. 1a, Table 1). Analysed as silver complexes on caesium sulphate gradients A. sylvaticus revealed two satellite components whereas A. flavicollis had only one (Fig. 1b). It has not, however, been confirmed that the silver-caesium sulphate satellites correspond to those seen in caesium chloride gradients.

The satellites were purified in a preparative centrifuge on gradients of caesium chloride (Figs. 2 and 3). After three cycles of purification both satellite DNAs gave a single, symmetrical band in analytical gradients of caesium chloride, though it is possible that they were contaminated with other DNAs having the same buoyant density. This problem is more serious with A. sylvaticus which is closer to the main band.

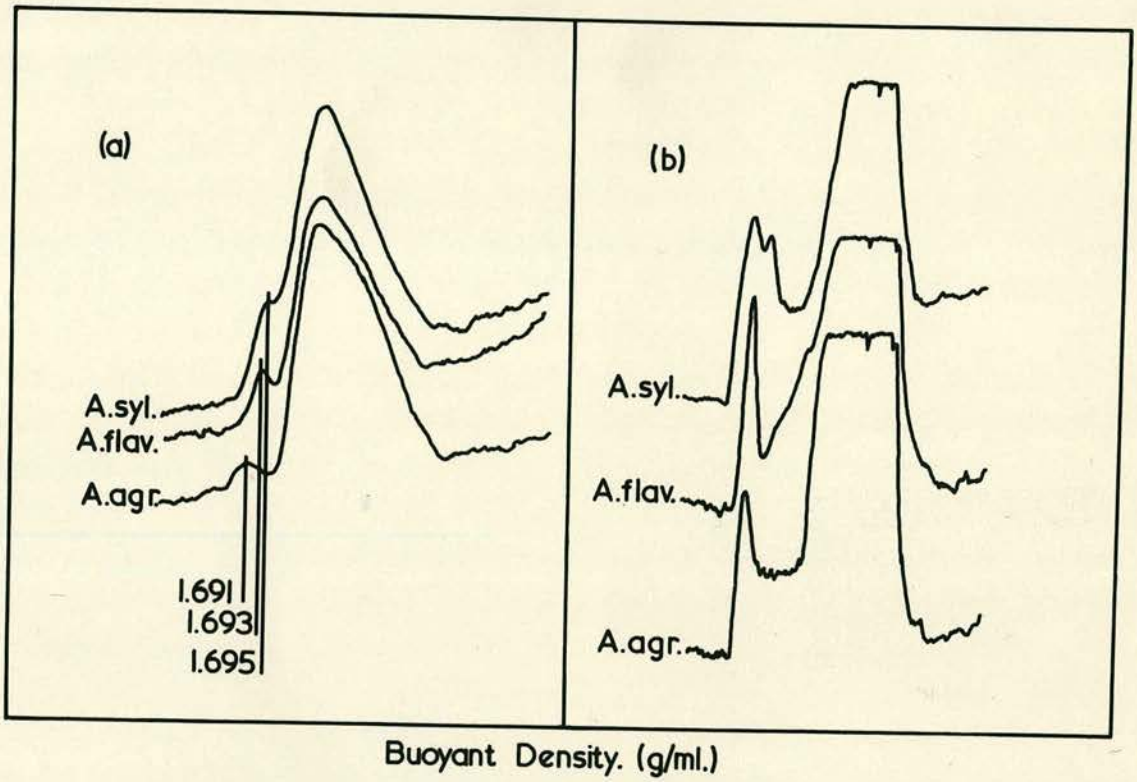

Fig.1. (a) Neutral CsCl gradient centrifugation of total Apodemus DNAs. The marker DNA is not shown.

(b) Ag^{+}/Cs_2SO_4 gradient centrifugation of total Apodemus DNA.

Table 1

Caesium chloride densities of Apodemus satellite DNAs (g./ml.).

	A. sylvaticus	A. flavicollis	A. agrarius
Relative Amount (%) -	8	6	4
Native	1.695	1.693	1.691
Renatured	1.699	1.697	1.699

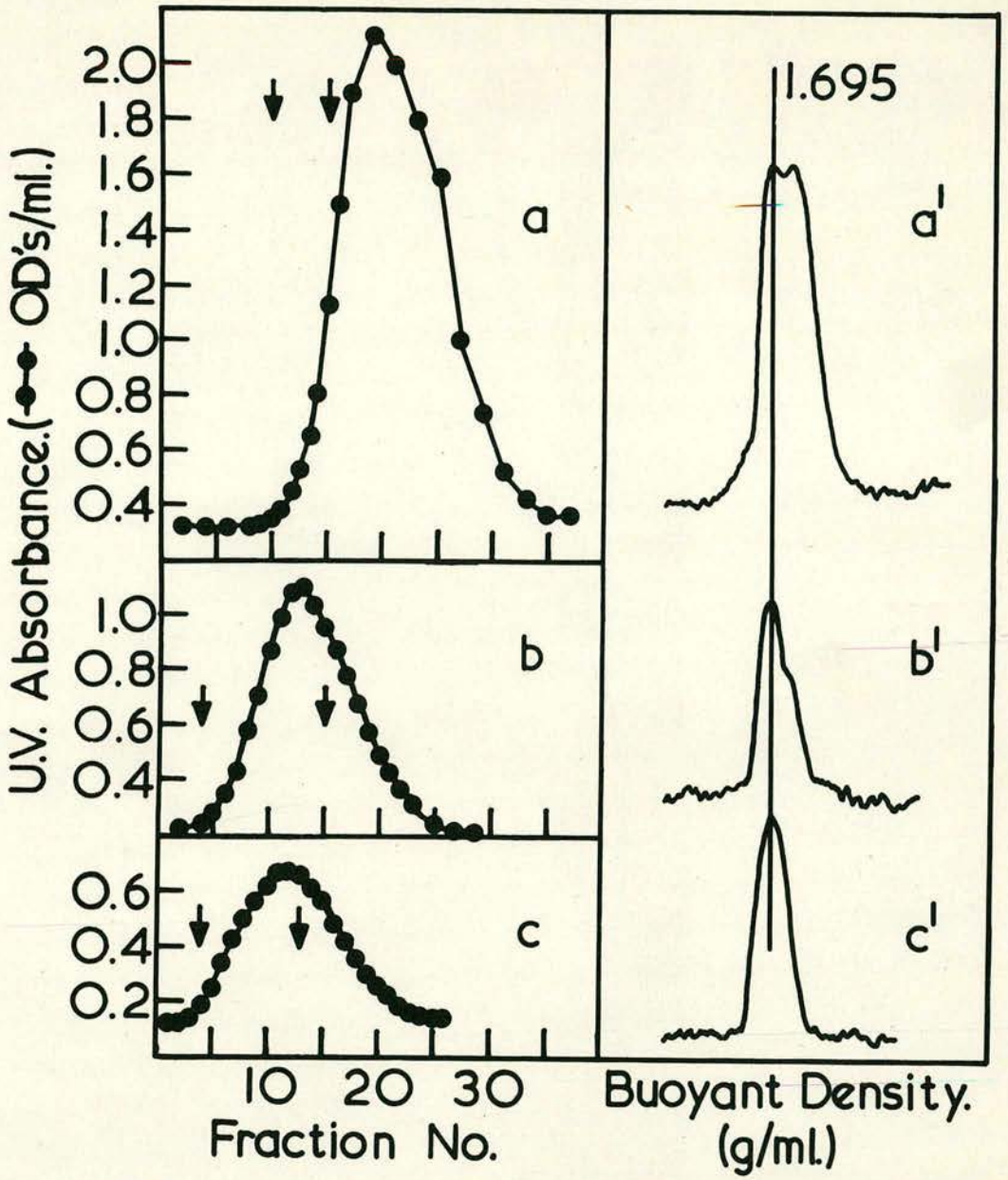

Fig. 2. Preparative purification of A.sylvaticus satellite DNA on neutral CsCl gradients. A sample of the pooled satellite fractions (between the arrows) was analysed by analytical ultracentrifugation (a^1, b^1, c^1) after each cycle of purification (a, b, c).

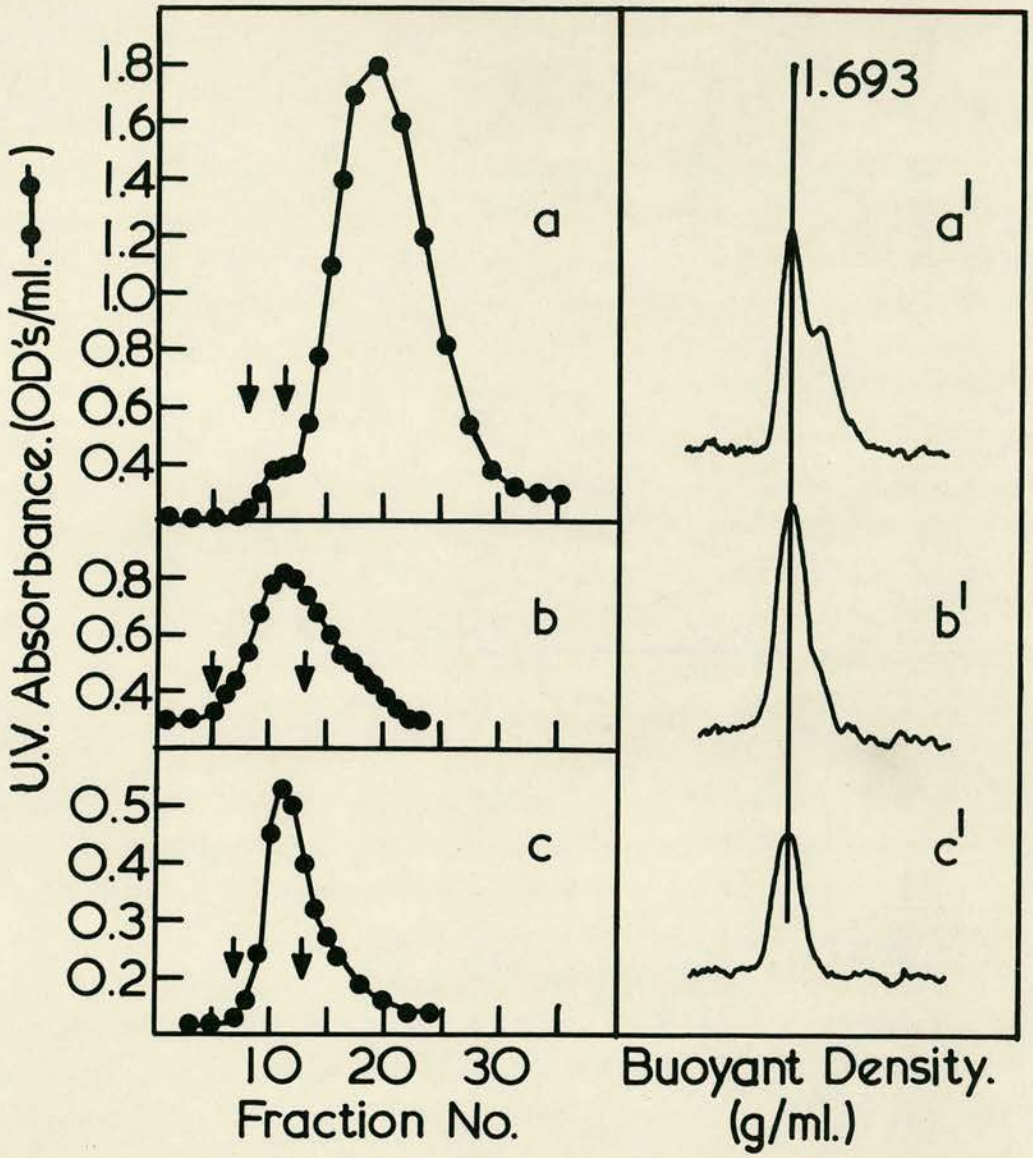

Fig. 3 Preparative purification of A. flavicollis satellite DNA on neutral CsCl gradients. Details as in Fig. 2

A better idea of the homogeneity of the satellite DNAs was obtained by analysing them in density gradients after sonication, denaturation and reassociation. (Fig 4b). It is unlikely that contaminating DNAs would reassociate to give duplexes with the same buoyant density as the satellite DNA.

In alkaline caesium chloride gradients in the analytical ultracentrifuge, A. flavicollis satellite DNA showed some strand separation whereas A. sylvaticus satellite DNA gave a single band (Fig. 4a). It was not possible to separate the strands of A. Flavicollis satellite DNA on a preparative gradient.

I.b. Melting of Native Satellite DNAs.

A. sylvaticus and A. flavicollis satellite DNAs gave sharp, unimodal transitions with a mid-point of denaturation at 67.5° when melted in 0.01 m-Na^+ (Fig. 5, Table 2). Both Tm (Marmur and Doty, 1962) and buoyant density (Schildkraut et. al., 1962) are related to base composition for bacterial DNAs. The same relationships do not hold good for all satellite DNA's (Walker, 1971) or for synthetic polymers with simple repeating sequences (Wells et.al., 1965; Riley et al., 1966). For the satellite DNAs used here we have no chemical measurement of the base composition. However, the G+C contents calculated from the Tm and buoyant density are not in agreement (Table 3). The melting temperature would suggest a G+C content of 42% for both DNAs whereas the buoyant densities suggest base compositions of 34.7 and 33.7% G+C for A. sylvaticus and A. flavicollis respectively.

I.c. Renaturation of the Satellite DNAs

Preliminary experiment showed that A. sylvaticus and A. flavicollis light satellite DNAs reassociate very rapidly. Rapid

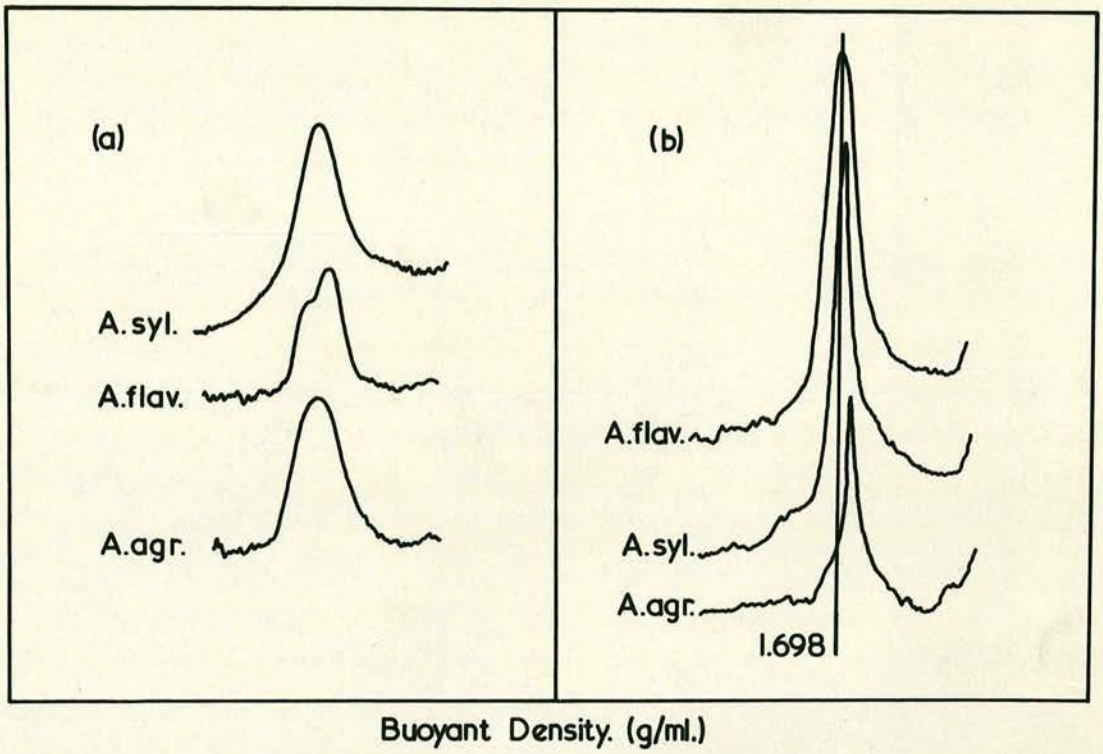

Fig. 4(a) Alkaline CsCl gradient centriguation of purified Apodemus satellite DNAs.

(b) Neutral CsCl gradient centriguation of reassociated, sonicated Apodemus satellite DNAs. The marker DNA is not shown.

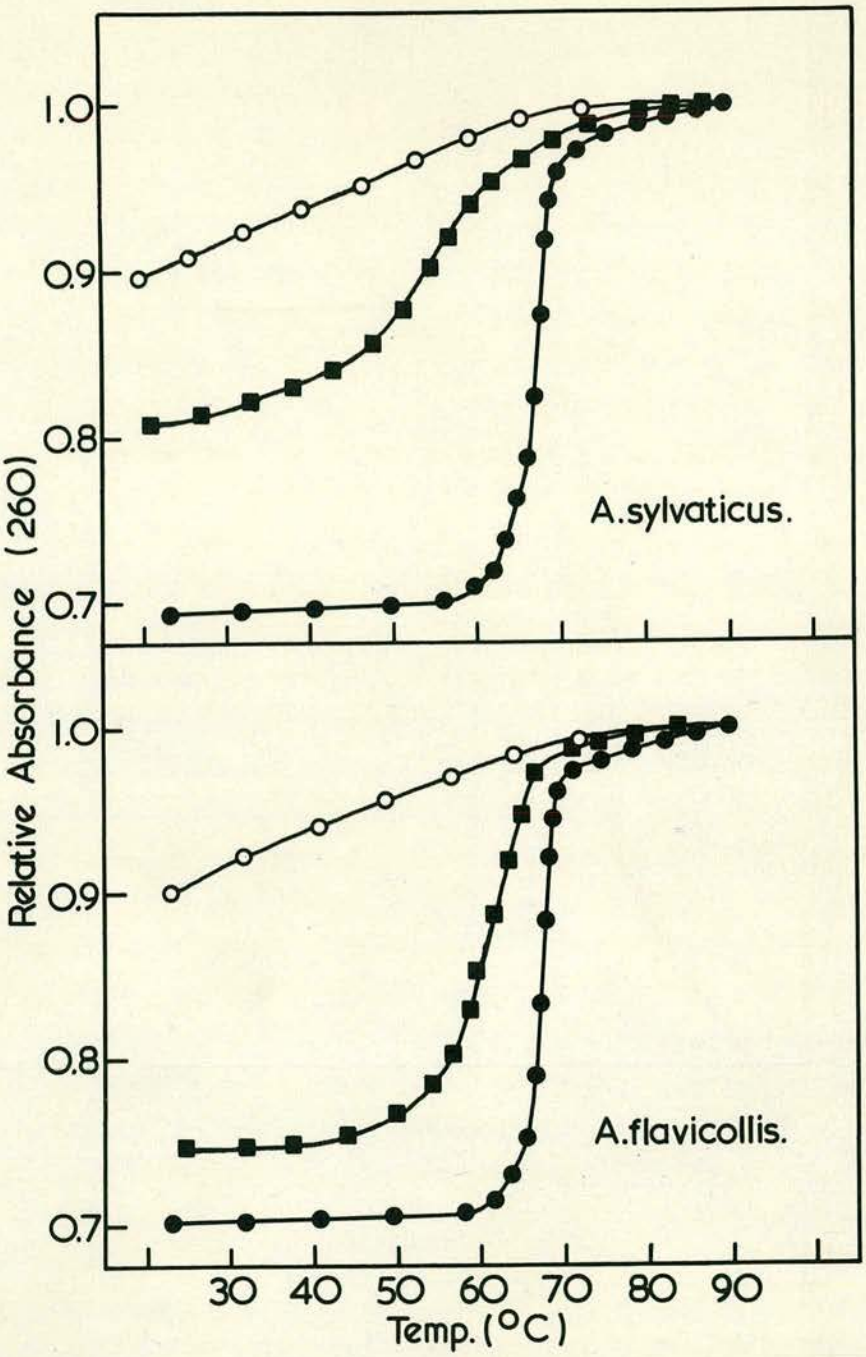

Fig. 5 Optical melting curves of isolated Apodemus satellite DNAs in 0.01m $-Na^{+}$.

—●—● Native DNA; —■—■ reassociated DNA;
—○—○ cooling curve.

Table 2

Melting temperatures of Apodemus Satellite DNAs in 0.01m-Na^+.(^oC)

	A sylvaticus	A flavicollis	A agrarius I	A agrarius II
Native Tm.	67.5	67.5	63.5	70.0
Renatured Tm (60^o reaction)	56.5	61.5	48.0	
Renatured Tm (57^o reaction)	-	-	42.0	
Δ T.m. (60^o reaction)	11.0	6.0	18.0*	
Δ Tm (57^o reaction)	-	-	24.0*	

* Relative to Tm of 66^o.

I - low melting component

II - high melting component

Table 3

Estimated G+C contents of Apodemus Satellite DNAs (%)

	A sylvaticus	A flavicollis	A agrarius
Calculated from density in CsCl.+	34.7	33.7	31.6
Calculated from melting temperature ++	42.0	42.0	37.3*

+ Schildkraut et al. (1962).

++ Marmur and Doty (1961).

* Average of two melting components.

reassociation rates can be measured using hydroxylapatite (Bernardi, 1965; Miyazawa and Thomas, 1965), if the rate is slowed down by using dilute solutions of DNA. This method requires radioactive DNA. Alternatively, special apparatus capable of following rapid reactions optically may be used, or the reaction may be followed in a standard spectrophotometer if the rate is slowed by lowering the ionic strength of the solution. As neither radioactive DNA nor apparatus for measuring rapid reactions were available, the latter method was used to study these satellite DNAs. Sutton (personal communication)has shown that mouse satellite DNA obeys the same dependence of reassociation rate on ionic strength as do bacterial DNAs and therefore the tables of Britten and Smith (1970) have been used to adjust the estimates of the reassociation rate constants to the standard salt concentrations used by other workers.

Estimation of the rate constant of reassociation requires a measurement of the final extent of the reaction. However, neither A. sylvaticus nor A. flavicollis satellite DNAs returned to the optical density of the native duplex (Fig. 6). There are two possible causes for the reduced hypochromicity: for both satellite DNAs the reassociated duplexes are not perfectly base paired as is shown in the next section, and this will give some reduction in the hypochromicity; secondly, as already mentioned, the purified satellites may not be homogeneous and the 'impurities' may not reassociate at the rapid rate of the satellite sequences. This latter problem is likely to be serious only with A. sylvaticus satellite DNA which may be contaminated with main band DNA. Two rate constants for the reactions, have been calculated first making the assumption that the true end point of the reaction would have been a return to the native optical density, as suggested by Wetmur and Davidson (1968), and secondly they have been calculated

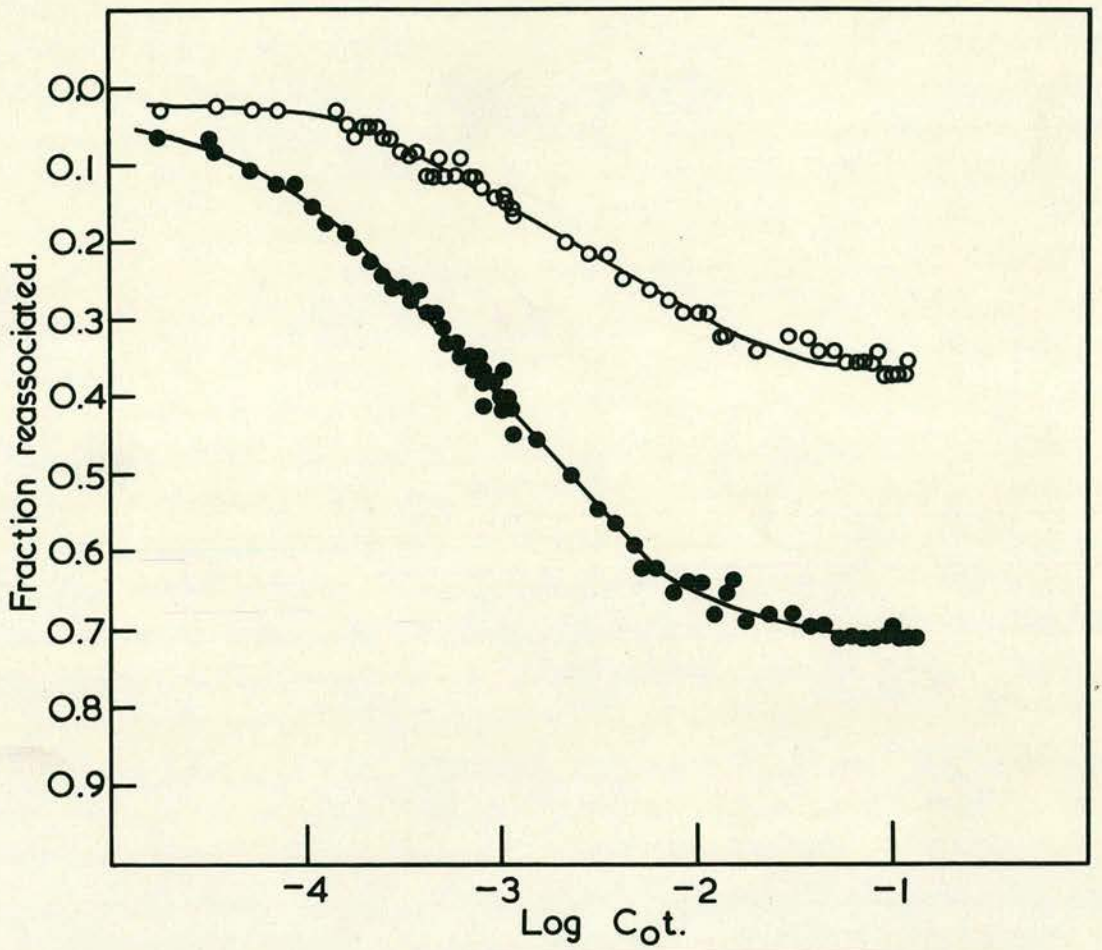

Fig. 6 Optical reassociation curves of isolated A. sylvaticus (—○—○—) and A. flavicollis (—●—●) satellite DNAs. All initial absorbances were measured in 0.01m-Na^+ and corrected for dilution. Both satellites were reassociated in 0.04m-Na^+ at 51°C and also in 0.18m-Na^+ at 60°C.

using the end point of the reaction actually achieved (Table 4). The first value is likely to be an underestimate and the second an overestimate of the rate constant.

The renaturation of A. sylvaticus and A. flavicollis are represented in Figs. 7 and 8 in the form described by Wetmur and Davidson (1968). On such a plot, the renaturation of a bacterial DNA would follow a straight line relationship. From the slope of the line the rate constant for reassociation can be determined. For A. sylvaticus and A. flavicollis no single straight line can be drawn through the points and therefore the rate constants for these reactions have been determined by taking the reciprocal of the Cot½ value (Fig. 6).

L.d. Properties of the reassociated satellite DNAs.

Both satellite DNAs reassociate to give imperfectly matched duplexes (Fig. 5). An estimate of the extent of mismatching is given by the decrease of Tm of the reassociated duplex from that of the native DNA. Calculated from the Δ Tm values using the relationship of 1% mismatching per 1° drop in Tm (Bonner et al., 1973) the proportion of mismatched base pairs are: 11% for A. sylvaticus; 6% for A. flavicollis (Table 2). From these results it would be expected that the density difference between the native and reassociated DNAs in neutral caesium chloride would be less for A. flavicollis than for A. sylvaticus. In fact both DNAs gave density shifts of 4 mg./ml. (Table 1).

I.e. Discussion of Results

With respect to their physical properties, the only significant difference between the satellite DNAs of A. sylvaticus and A. flavicollis is in the extent and quality of duplex formation during

Table 4

Rate constants (k2) for the renaturation of the Apodemus Satellite DNAs expressed as $\frac{1}{Cot_{\frac{1}{2}}}$ (litres/moles x seconds).

	A sylvaticus	A flavicollis	A agrarius	
	I	I	I	II
k2*	5.3×10^2	1.3×10^3	1.4	0.5
k2**	1.4×10^3	1.8×10^3	-	-

I Renaturation at 60° in 0.18m-Na^+.

II Renaturation at 57° in 0.18m-Na^+.

* Calculated on the basis that 100% reaction will result in a return to the native optical density.

** Calculated on the basis that 100% reaction has been achieved.

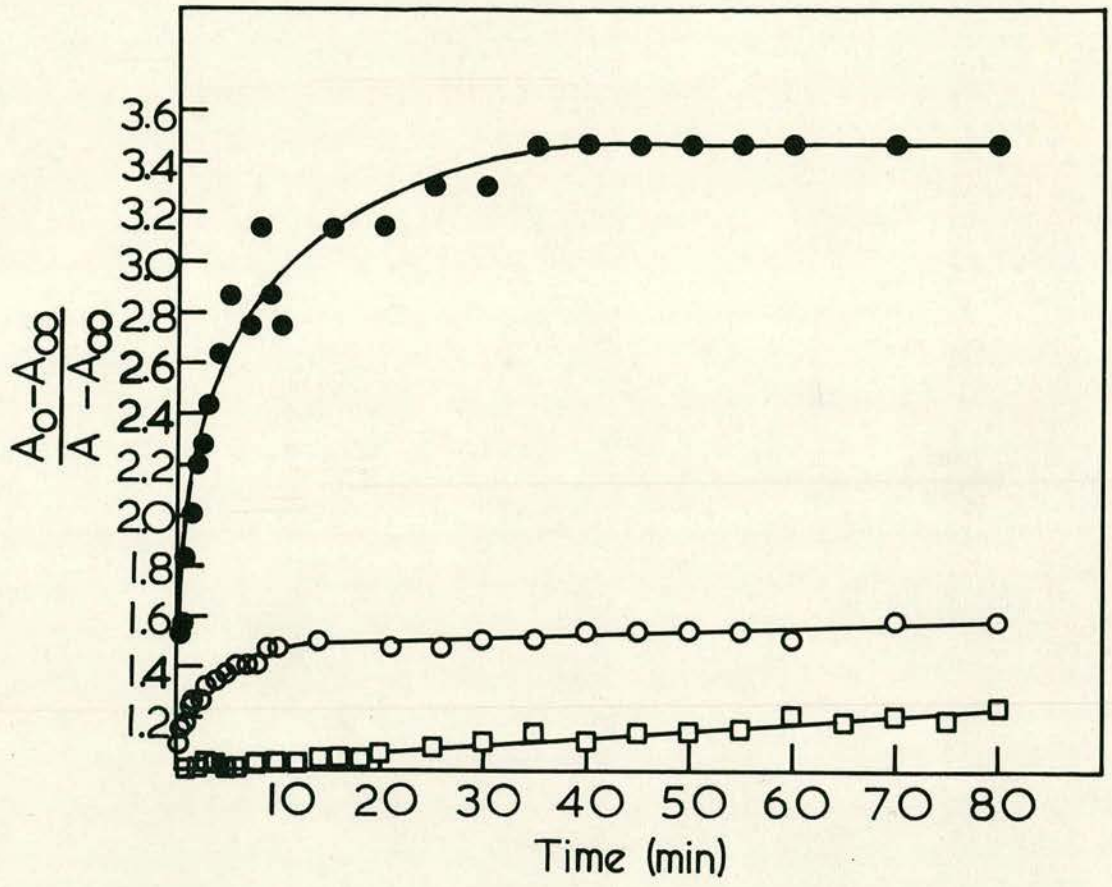

Fig. 7 Optical reassociation curves of isolated Apodemus satellite DNAs plotted as described by Wetmur and Davidson (1968). Renaturation was in 0.18m-Na^+ at 60°C.

—○—○— A. sylvaticus; —●—●— A. flavicollis;
—□—□— A. agrarius

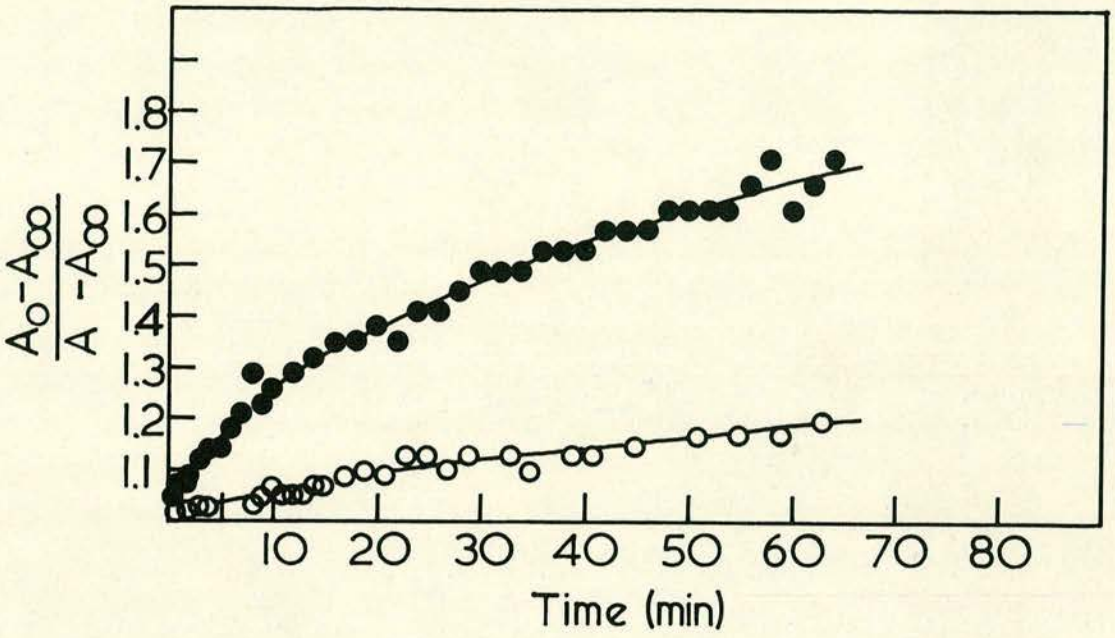

Fig. 8 Optical reassociation of isolated A. sylvaticus —○—○— and A. flavicollis —●—●— satellite DNAs in 0.04m-Na^+ at 51°C., plotted as described by Wetmur and Davidson (1968).

Approximately 70 mins. on the above scale is equivalent to 1 min. on that of Fig. 7

reassociation. Upon correction for the extent of renaturation both DNAs have kinetic complexities of about 3-400 base pairs. However before discussing the results obtained in this study it is important that a brief mention be made about the results of the restriction studies being carried out on the satellite DNAs of A. sylvaticus and A. flavicollis.

The pattern of fragments produced by digestion with the bacterial restriction enzyme, HindIII, are very similar for both DNAs (Cooke, 1974). Each DNA gives a pattern of bands with lengths of 340, 2 x 340 and 3 x 340 base pairs. This is strong evidence in favour of a sequence relationship between A. sylvaticus and A. flavicollis satellite DNAs. Two other points of this work are important. When A. sylvaticus satellite is digested with the enzyme only about 80% of the DNA goes into the bands. The remaining 20% may well be contaminating main band DNA and therefore partly responsible for the reduced hypochromicity upon renaturation of the satellite DNA. Furthermore when the DNA from the smallest A. sylvaticus restriction band is isolated and renatured it goes to within 80% of completion. Taking these results into consideration it would be expected that the A. sylvaticus satellite DNA used in the present work would achieve at least 60% renaturation even taking contaminatio n into account. Clearly another factor must be involved in reducing the extent of renaturation. The restriction fragments of A. sylvaticus satellite DNA are homogeneous in size and reassociate to form duplexes matched from one end to the other. They do not form concatemers or high molecular weight networks. However such structures can be formed from sonicated DNA because on average, the length of the reacting molecules is longer than the

sequence complexity. Indeed there is evidence that they do form since the reassociated DNAs form sharp bends in CsCl gradients. Combined with relatively high extent of mismatching, these structures may prevent long stretches of sequence from forming duplexes and hence reduce the extent of renaturation measured from the hyprochromicity.

The duplex formed upon renaturation of the smallest restriction band of A. sylvaticus satellite DNA melts in two steps (Cooke, 1974). The majority (80%) of the duplex melts with a Δ Tm similar to that reported here but the remaining 20% melts with a Δ Tm of only 2°. It seems likely that the inhomogeneity of the sonicated sample used in this study, combined with the low extent of renaturation achieved, is responsible for the absence of a step in the melting curve of the reassociated DNA.

Cooke was also able to show that on remelting and renaturing the low stability duplexes only, the product contained, once again, both high melting and low melting duplexes. He interpreted this result as indicating that A. sylvaticus satellite DNA contains more than one family of repeated sequences and that th e sequences of different families are able to cross reassociate to form poor duplexes. The proportions of high stability duplexes, formed by renaturation of members of the same sequance family, and of low stability duplexes, formed by renaturation between members of different sequence families, suggest that about five such sequence families are present in A. sylvaticus satellite DNA.

3.II THE PHYSICAL PROPERTIES OF A. AGRARIUS SATELLITE DNA.

3.IIa. Density gradients

Apodemus agrarius has a light satellite DNA amounting to about 4% of the total DNA, with a buoyant density of 1.691 gm./ml. (Table 1, Fig. 1a). It is well separated from the bulk of the DNA and needed only two cycles of preparative purification before giving a single, symmetrical band in an analytical gradient of caesium chloride (Fig. 9). When analysed as a silver complex on a caesium sulphate gradient, the total DNA again showed a light satellite (Fig. 1b). It has not, however, been confirmed that the silver caesium sulphate satellite corresponds to the one seen in caesium chloride gradients. On centrifugation in an alkaline caesium chloride gradient there was no evidence to suggest strand separation (Fig. 4a).

II.b Melting of Native Satellite DNA.

When melted in 0.01m - Na^+ A. agrarius satellite DNA gave a two step transition (Fig. 10). Approximately 70% of the hyperchromicity resulted from a sharp melt, the mid-temperature of which was 63.5°. The remainder of the DNA melted over a wider range with a Tm of 70.0°. From Marmur and Doty's (1962) relationship between Tm and base composition, the base composition of the low and high melting fractions of A. agrarius satellite DNA are calculated as 32.2% and 49.3% G+C respectively. Given these values, and taking into consideration the different amounts of the two fractions, the average G+C content for the satellite DNA would be 37.3%. This is to be compared with a value of 31.6% calculated from the buoyant density of the satellite (Table 3). It has already been mentioned that synthetic polymers with simple repeating sequences do not always

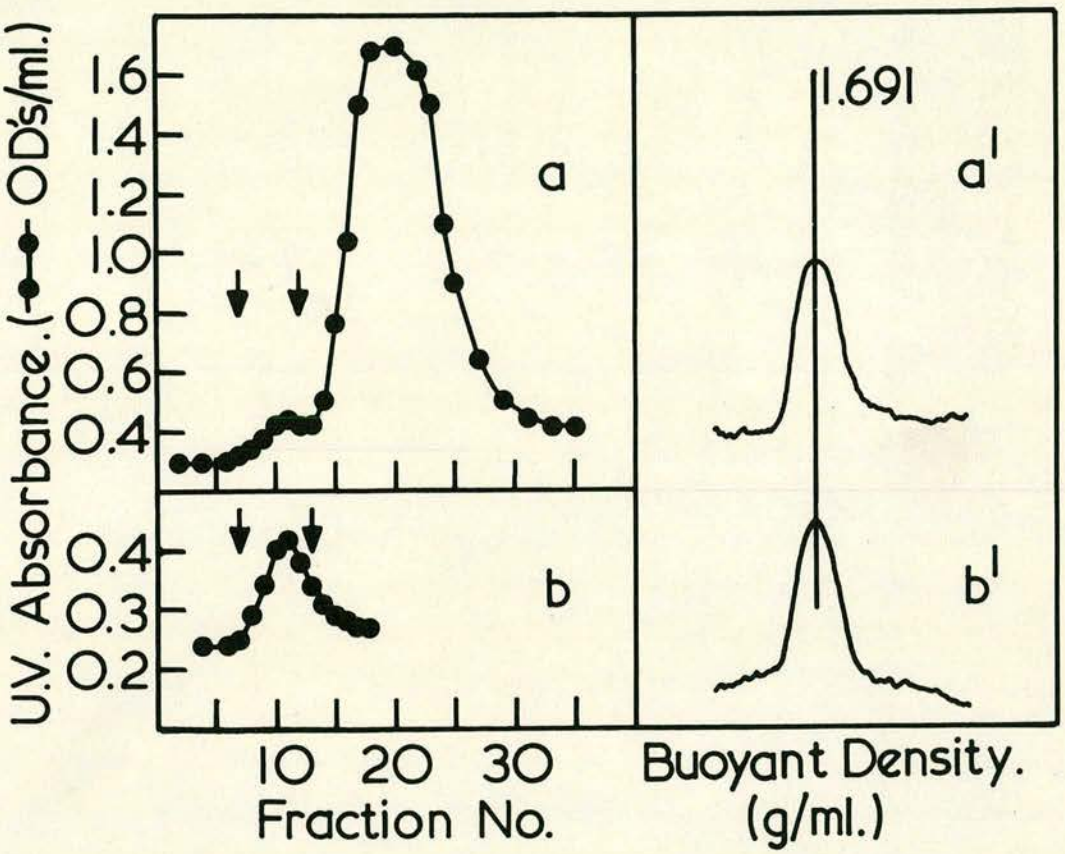

Fig. 9 Preparative purification of A. agrarius satellite DNA on neutral CsCl gradients. Details as in Fig. 2

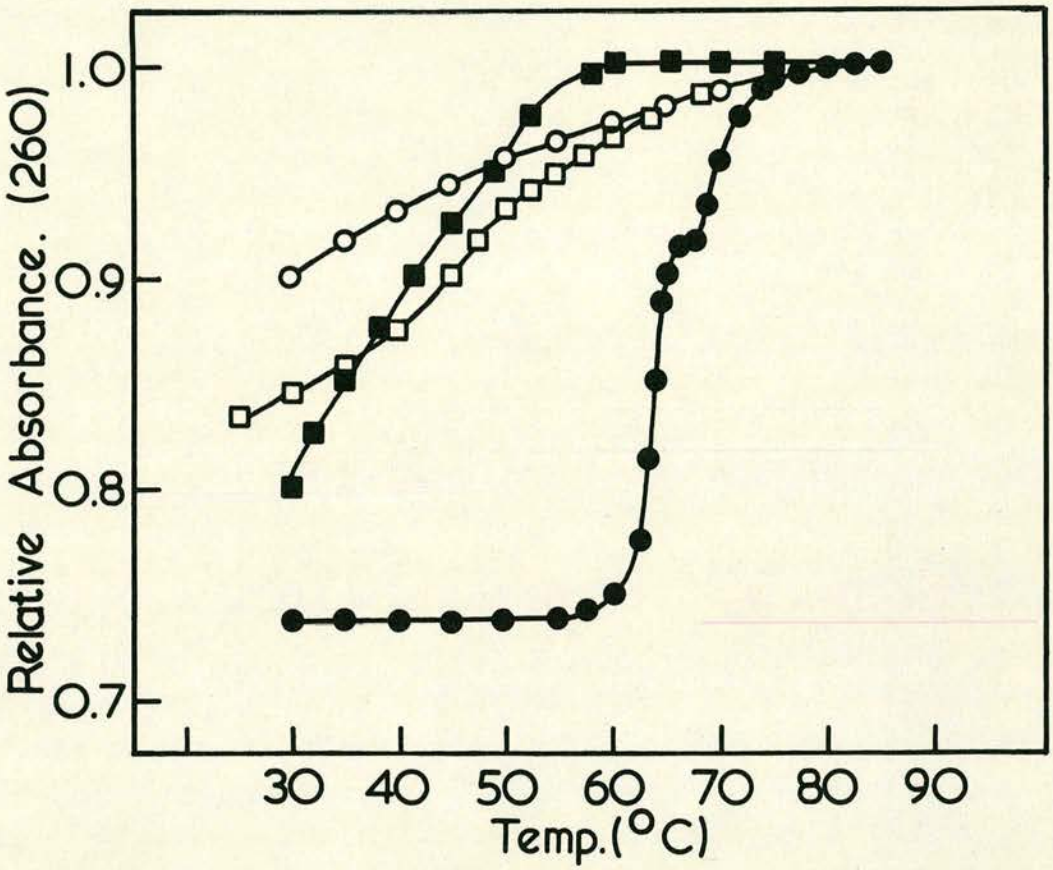

Fig. 10. Optical melting curves of isolated A. agrarius satellite DNA in 0.01 m - Na^+.

Native DNA; DNA reassociated at 60°C;
DNA reassociated at 57°C; cooling curve.

obey the relationship between Tm and G+C content. For example, alternating poly d (A-T) has a Tm 7.5° lower than poly dA. poly dT. (Riley et.al., 1966). Although it cannot be ruled out that A. agrarius satellite DNA may contain a high G+C fraction, the discrepancy in base composition from the two estimates above coupled with the partial sequence analysis described later, suggests that the two step melt may result from sequence differences in the two fractions rather than a base compositional difference.

II.c. Renaturation of A. agrarius satellite DNA

As A. agrarius satellite DNA renatures much more slowly than A. sylvaticus or A. flavicollis, it was found necessary to increase the rate of reassociation by increasing the ionic strength of the DNA solution. The tables of Britten and Smith (1970) have been used to adjust the estimates of reassociation rate constants to the standard salt concentration of 0.18m-Na^+.

The rate of renaturation of A. agrarius satellite DNA was followed at two different temperatures, 57° and 60°. The reaction at 60° in 0.18m-Na^+ is equivalent to working 22.5° and 29.0° below the melting temperatures of the low and high melting fractions respectively.

It is clear from Fig. 11, that at both temperatures reassociation was still in progress when the experiments were stopped. As no real end point of the reactions is available, the rate constants have been calculated making the assumption that the true end point of the reaction would have been a return to the native optical density. Based upon this assumption, the $Cot_{\frac{1}{2}}$ value of the reaction at 57° would be $\frac{7}{2} \times 10^{-1}$ and at 60° would be 2. These values represent sequence complexities of about 0.5 and 1.5×10^6

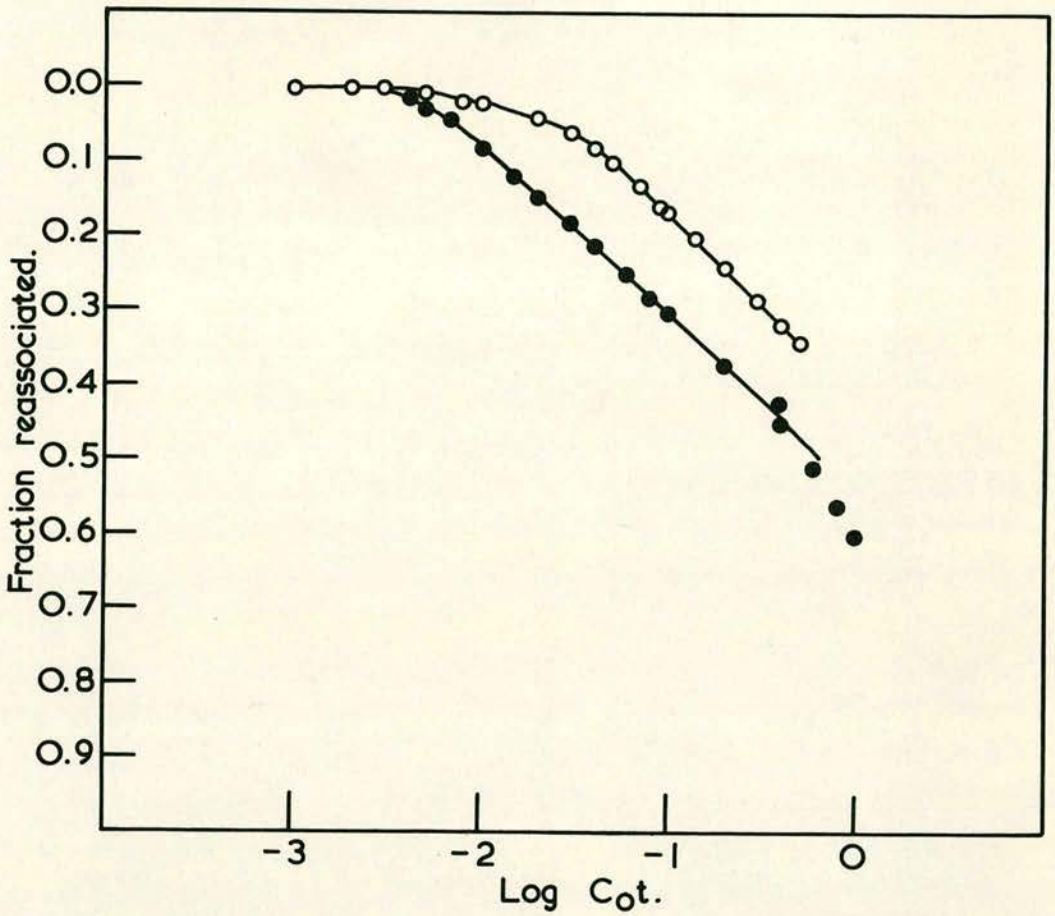

Fig. 11 Optical reassociation curves of isolated A. agrarius satellite DNA. Renaturation was in 0.18 m-Na^+ at 60° (—○—○—) and 57°C (—●—●—).

All initial absorbances were measured in 0.01 m - Na^+ and corrected for dilution.

base pairs respectively. However, a complexity of 5×10^4 base pairs would be obtained if the 57° reaction was complete at 60% reassociation.

II.d. Properties of the Reassociated Satellite DNA

Fig. 10 shows the melting curves of the reassociated duplexes formed at 57° and 60°. They have Δ Tm values of 24° and 18° respectively, taking the Tm of the native satellite as 66°. Clearly the satellites have formed very poorly matched duplexes. On analysis in a ceasium chloride gradient the reassociated duplex, formed at 57°, gave a hypersharp band with a density of 1.700 gm./ml. This is 8 mg./ml. heavier than that of the native satellite DNA which is what one would expect for such a poorly matched duplex. It was rather surprising however that only one band was formed. It may be inferred from this result that most molecules had entered into duplex formation even though the hypochromicity was only 60%. It is unlikely that the formation of such a highly mismatched duplex would result in the return of the optical density to that of the native DNA. If this conclusion is correct, then this result lends more weight to a kinetic complexity estimate of around 50,000 base pairs.

II.e. 'Snapback' experiment

The two step melt of A. agrarius satellite DNA could result from a variety of configurations within the DNA molecules, the two extremes of which are represented diagrammatically in Fig. 12. In Model A, the sequences responsible for the high and low melting fractions in the satellite DNA alternate regularly with each other. On the other hand, the sequences may be grouped together in large blocks and this extreme is shown in Model B.

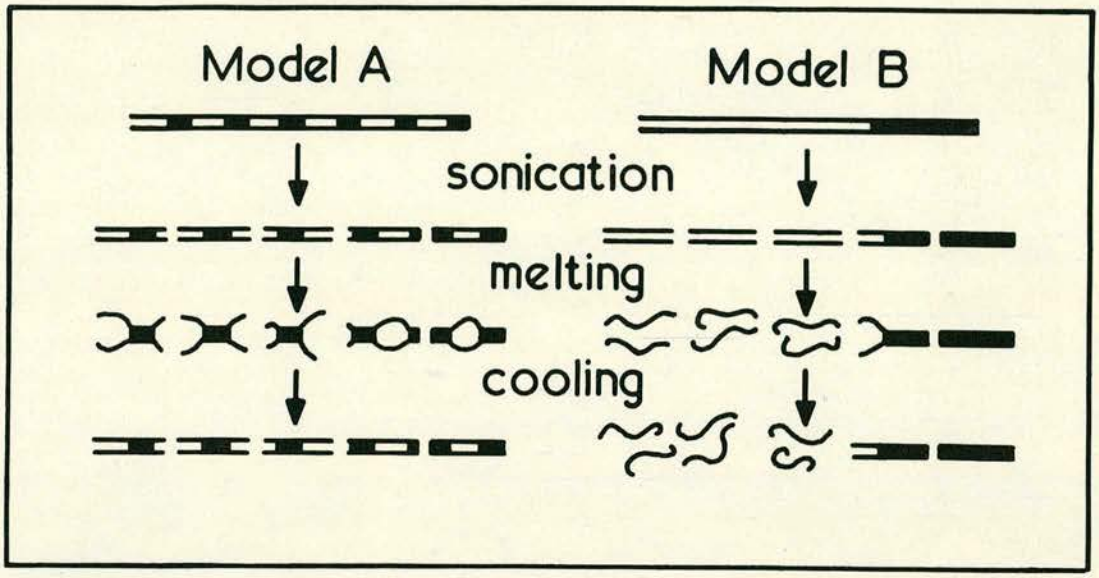

Fig. 12 Two interpretations of the organisation of A. agrarius satellite DNA. The two models illustrate the different structures adopted by the molecules after sonication, melting and cooling.

= low melting double stranded DNA;

▬ high melting double stranded DNA;

≈ single stranded DNA.

A simple experiment was conducted to determine which of the above models applied in the base of A. agrarius satellite DNA. The DNA was sonicated under conditions which normally produce molecules with an average size of 500 base pairs. The DNA was then melted in 0.01m-NaCl. At various stages during the melting transition (a, b, c, etc. Fig. 13) the sample was removed from the spectrophotometer and placed on ice for a few minutes. The sample was then returned to the spectrophotometer at 40° and the absorbance or the extent of denaturation, was again measured (a', b', c', etc. Fig. 13). This procedure was repeated a number of times gradually increasing the extent of denaturation before cooling, until eventually the sample was completely denatured. Fig. 13 shows the results of this experiment.

Even when the sample was denatured by more than 90% (d) and then cooled, the absorbance returned to within 25% of the native optical density (d'), and when remelted the denaturation transition was very similar to that of the native DNA. It should be noted that in such low ionic conditions and over such a short period of time the DNA would not reassociate if complete strand separation occurred.

Only Model A predicts the results obtained above. The denatured low melting fraction must be attached to a high melting block which holds the complementary DNA strands in register until, when the temperature is lowered, the molecule returns to its native state by 'zippering' up from the duplex region. Return to the native optical density was almost complete when only the low melting fraction was denatured (b, Fig. 13). It is concluded from this that the high melting blocks must occur at least once every 500 base pairs.

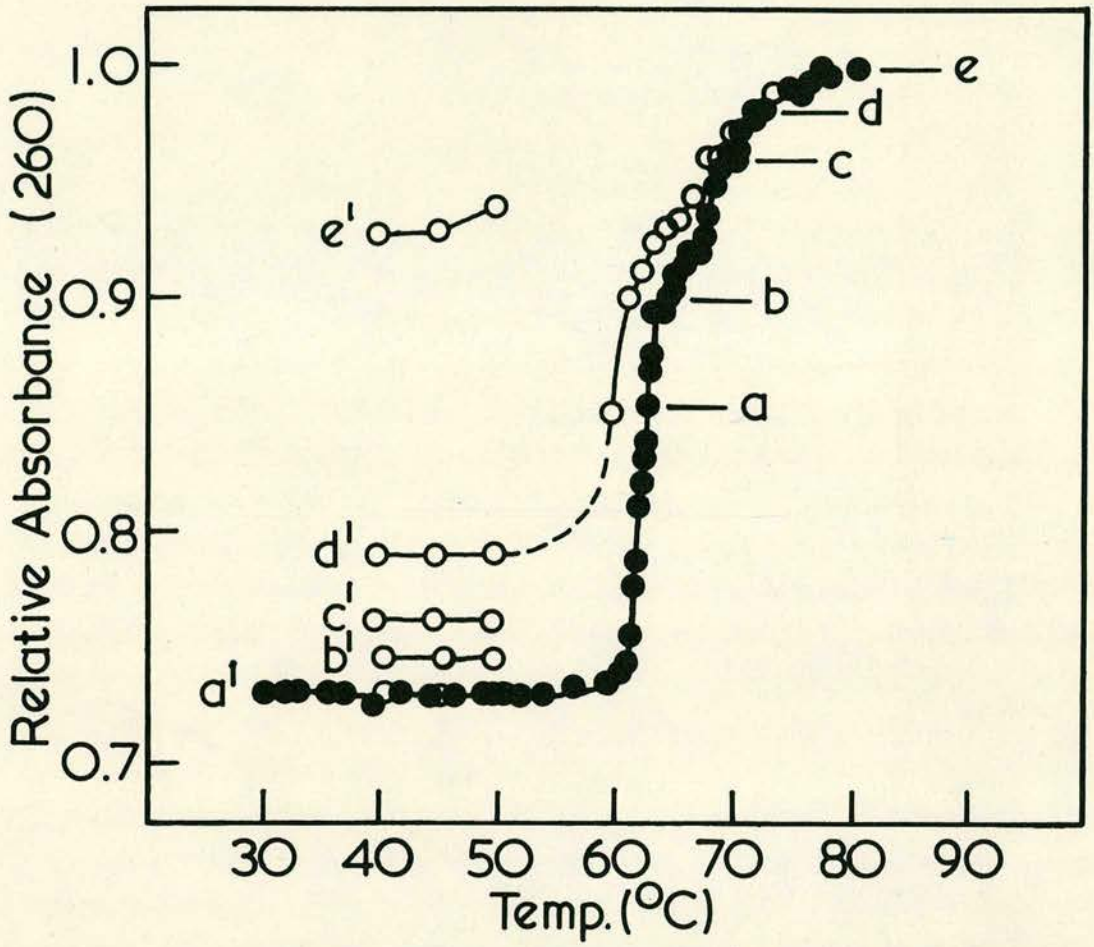

Fig. 13 'Snapback' experiment with A. agrarius satellite DNA.

The DNA was melted in 0.01 m-Na^+ (—●—● —) up to the temperature indicated (a). The sample was then cooled on ice for a few minutes and the absorbance was remeasured (a^1 —○—○—). The procedure was repeated, gradually increasing the extent of denaturation ($b, b^1; c, c^1$; etc.)

11.f. Discussion of Results

The 'snapback' experiment described above strongly indicates that A. agrarius satellite DNA consists of a repeating sequence less than 500 base pairs long. However, more decisive evidence for this conclusion has been provided by the experiments of L. Polito (reported in Allan et al., 1974).

The efficiency with which exonuclease treated sheared fragments of DNA form circles is dependent upon the ability of the single stranded ends of one fragment to renature with each other (Thomas, Lee, Pyeritz and Bick, 1972). Circle formation therefore reflects sequence repetition within the DNA fragments. The reaction is also a function of the length of the repeating sequence, the extent of exonucleolytic hydrolysis, and the degree of sequence divergence within the repeating sequence.

Circles have been formed from both needle-sheared and sonicated A. agrarius satellite DNA (Allan et al., 1974). The proportion of circles formed with needle-sheared A. agrarius satellite DNA was only half of that formed from mouse satellite DNA under the same conditions. This result may reflect either a longer sequence complexity or a higher extent of sequence divergence within the repeating sequence of A. agrarius satellite DNA. In view of the low thermal stability of the duplexes of reassociated A. agrarius satellite DNA, the latter is probably the case. The fact that circles were also formed with sonicated A. agrarius satellite DNA seems to indicate at first sight that the repeat length is shorter than the circumference of the circles, because only a small proportion of the DNA was exposed by exonuclease treatment and as Thomas et al., (1972) have argued, the

probability of forming a circle becomes significant only when the length exposed is a substantial fraction of the repeat length. The circumference of the circles formed is therefore a maximum measure of the repeat length.

When the size of the sonicated fragments and the circles formed from them were measured in the electron microscope, a very surprising result was obtained. The molecules fell sharply into size classes with groups at 0.13, 0.26, 0.39 and 0.5 microns. The spacing of the size classes was the same as that of the circles (Fig. 14).

It is generally thought that sonication of DNA involves random breakage of the molecules, which, under the conditions used by Allan et al., (1974),would result in a broad distribution of DNA fragments ranging in size from about 200 up to about 1000 base pairs with a mean length of 5-600 base pairs (Mitchell, 1973). In this respect the result obtained on sonicating A. agrarius satellite DNA is surprising in three ways. Firstly, the fragments do not give a broad distribution of lengths even within the size classes formed. Secondly, the fact that there are distinct size classes and that these are multiples of the smallest class produced is also unexpected. Finally, the smallest group of molecules formed has a mean size of about 370 base pairs which is quite different from the 5-600 base pairs expected.

Although Thomas et al., (1972) argue against the possibility of shear breaks occurring at or near special sequence related sites, the above evidence is not only strongly in favour of such sites being present, but suggests that they are evenly distributed throughout A. agrarius satellite DNA and occur about once every 370

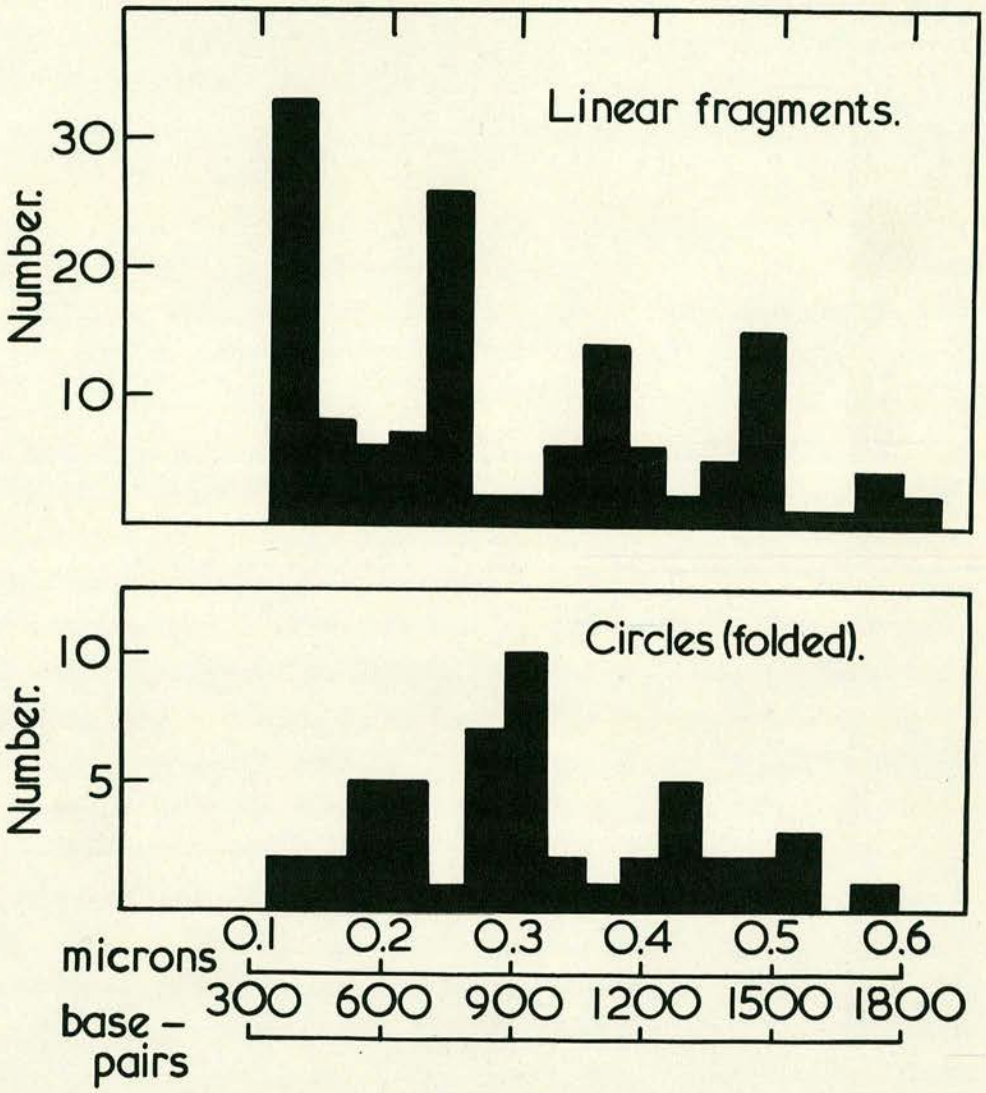

Fig. 14 Top; distribution of sizes of fragments produced by sonication of A. agrarius satellite DNA.

Bottom: distribution of sizes of folded circled produced by renaturation of exonuclease treated sonicated A. agrarius satellite DNA.

base pairs. After treatment of the sonicated fragments with exonuclease, the fragments form circles. The circles also fall into distinct size classes, separated by about 370 base pairs (Fig. 14).

These results and the 'snapback' experiment strongly suggest that A. agrarius satellite DNA consists of a repeating sequence about 370 base pairs long. However, the kinetic complexity of the DNA, derived from its renaturation rate constant, is 50,000 base pairs at a minimum. There are two possible reasons for this discrepancy. They differ according to which value is taken for the effect of mismatching on the rate of reassociation of the DNA.

Firstly, if the effect of mismatching on rate is small as suggested by Bonner et al., (1973), McCarthy et al., (1972) and Wang et al., (1971), then A. agrarius satellite DNA must consist of more than one family of repeated sequences. The effect of mismatching upon the rate of renaturation found by the above authors is equivalent to a five fold reduction in rate for a $20^{\circ}C$ Δ Tm. After correction by this factor, the kinetic complexity of A. agrarius satellite DNA is reduced to 10,000 base pairs. A DNA consisting of 25 - 30 families of sequences, all with a complexity of 3-400 base pairs and each family reassociating to give duplexes with a Δ Tm of 20°, would reassociate with an average complexity of 10,000 base pairs. Such a situation might exist for A. agrarius satellite DNA.

The duplexes formed by reassociation of A. agrarius satellite DNA at 60° melt over a broad temperature range (Fig. 10) indicating the presence of both well matched and poorly matched duplexes. However, the duplexes formed by renaturation at 57° give a more

homogeneous melting curve with a higher average Δ Tm (Fig. 10, Table 2). These results support the idea that A. agrarius satellite DNA consists of many sequence related families. At 57° many families may cross-reassociate to form highly mismatched duplexes. However, at the higher stringency conditions of 60° renaturation may be limited to members of the same or very closely related sequence families, giving well matched duplexes, and only a few families will cross-reassociate to form poorly matched duplexes. Also, if this is the case, the concentration of molecules able to renature will be higher at 57° than at 60° and the rate of the former reaction should be correspondingly faster. The rate of renaturation at 57° was three times faster than at 60° (Fig. 11, Table 4). It should be noted of course, that the estimate of 25-30 sequence families in A. agrarius satellite DNA was based on the assumption that they could not cross-reassociate.

If A. agrarius satellite DNA contains many sequence families they must have closely related physical properties. The 'snapback' experiment (Fig. 13), the sonication results (Fig. 14) and the relative homogeneity of both native (Fig. 1a) and reassociated (Fig. 3a) satellite DNA when banded in caesium chloride gradients all argue for physical homogeneity. This possibility is made more plausible by the work of Smith (1973) who showed by computer simulation that repeating sequences can be separated into sequence related families by unequal crossing over.

Secondly, it remains possible that mismatching has a much greater effect upon the rate of renaturation than that proposed by Bonner et al., (1973). Sutton and McCallum (1971), using a correction factor derived from studies on mouse satellite DNA,

estimate a corrected sequence complexity for A. agrarius satellite DNA of only 330 base pairs. This figure is in close agreement with the estimate of 370 base pairs obtained in the present study. Kallenback and Drost (1972) used synthetic polynucleotides to study the effect of mismatching on rate and they showed that the rate of reassociation for highly mismatched duplexes is very sensitive to the conditions used for reassociation; and this dependence was emphasised in the work of Bonner et al., (1973) too. In addition to renaturation at 57° and 60°, A. agrarius satellite DNA was also reassociated at 50° in some preliminary experiments. In this case, although the initial rate of renaturation was faster than the reaction at 57° (but thereafter much slower) the structures formed on reassociation gave a melting profile which indicated that very little, if any, duplex had been formed. These experiments with A. agrarius satellite DNA show that even when a variety of reassociation conditions is used for rate measurements, it is still difficult to get a meaningful estimate of sequence complexity for sequences which reassociate to give poorly matched duplexes. The implications of this for the families of repeated sequences in higher organism DNA are discussed more fully in a later section.

3.III PARTIAL SEQUENCE ANALYSIS

IIIa Transcription with E Coli RNA polymerase

Direct sequence analysis of DNA is still much more difficult than analysis of RNA largely because of the lack of methods for specific degradation. For the Apodemus satellite DNAs there was the added problem of making DNA uniformly labelled with ^{32}P, because there are no established cell lines for these organisms. The alternatives were to make DNA copies and use ribo-substitution to allow for base specific degradation, as was used by Fry et al., (1973) for the analysis of the satellite DNAs from the kangaroo rat, or to make RNA copies with RNA polymerase and use the standard procedures for sequence analysis of RNA (Gall et al., 1974). The latter method was chosen because the techniques involved were already well established in this laboratory.

RNA polymerase from E coli which had been prepared by a method which probably does not remove the sigma factor (Burgess, 1969) was used for transcription. The native, high molecular weight satellite DNAs were transcribed in the presence of either UTP or ATP labelled with ^{32}P in the α phosphate, together with the other three non-radioactive ribonucleotide triphosphates.

Since it was not possible to obtain the isolated single strands of the satellites it was hoped that transcription might, under the conditions used, be highly asymmetric, as is the case for mouse satellite DNA (Biro et al., 1974). This would simplify the analysis of the RNA. It turned out, however, that for all three satellite DNAs the RNA produced was resistant to RNase T_1, thought not to pancreatic RNase (Fig. 15). The transcripts from

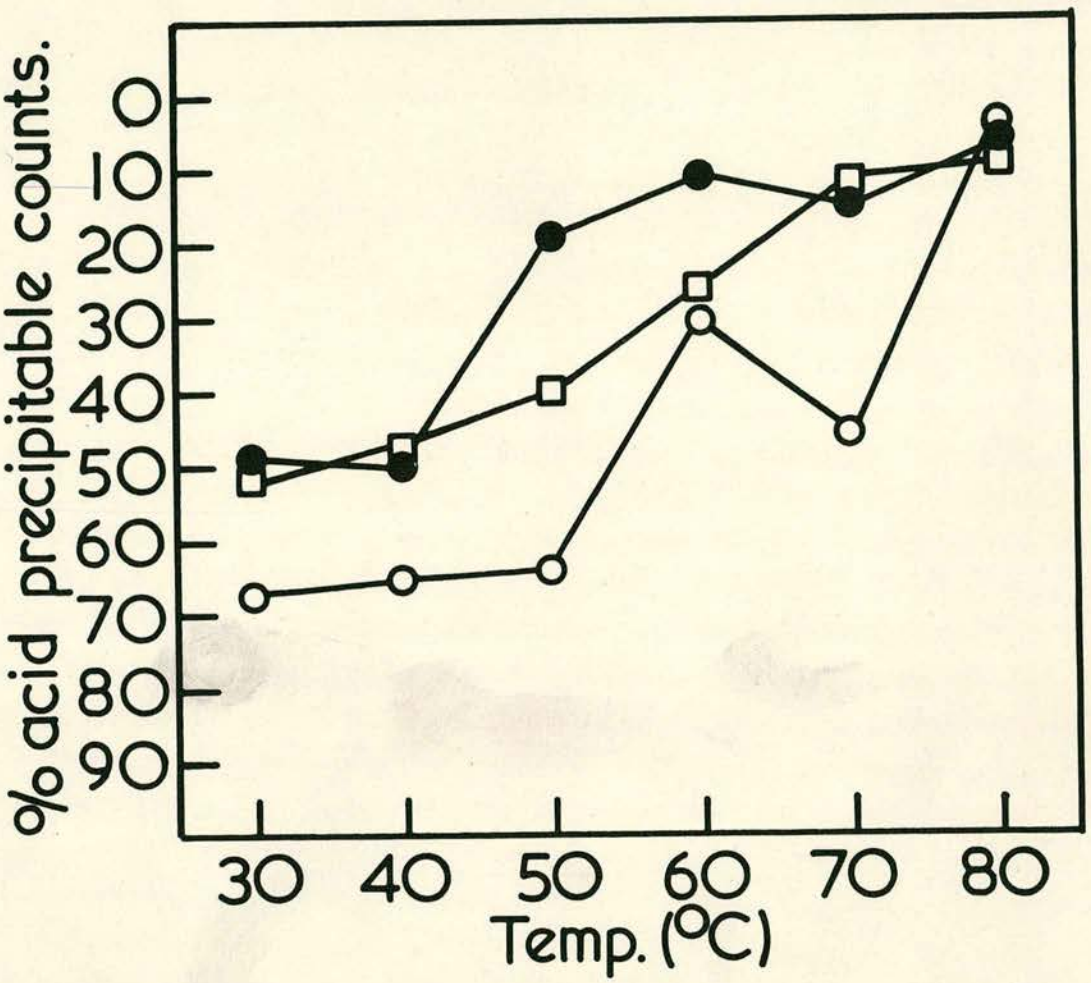

Fig. 15 Melting curves of UTP labelled RNA transcripts of Apodemus satellite DNAs, as assayed by resistance to RNase T_1. Samples in 0.01 SSC were heated to the temperature indicated and then digested with RNase T_1 under the conditions described in section 2,11.g.

Duplicate samples of the digested RNA were spotted on Whatman GFA filters. One sample was then precipitated with 5% trichlorocetic acid the other served as a control. The filters were washed, dried and counted in scintillator.

—□—□— A. Sylvaticus cRNA; —○—○— A. flavicollis cRNA;
—●—● A. agrarius cRNA

A. sylvaticus and A. agrarius satellite DNAs were found to be 50% resistant to RNase T_1 whereas that from A. flavicollis was 70% resistant. Heating to 100° followed by rapid cooling rendered the RNAs labile to RNase T_1.

RNA-RNA duplexes melt at higher temperatures than the corresponding DNA-DNA duplexes (Bishop, 1969). In 0.01 m-SSC a perfectly base paired RNA-RNA duplex, transcribed from the Apodemus satellite DNAs would have a melting temperature of 60°-70°. When assayed by resistance to RNase T_1, the cRNAs of the Apodemus DNAs melted between 40° to 70°, but were completely digested by pancreatic RNase at 40°. As the duplexes are clearly not perfectly matched it is unlikely that they were produced by displacement of RNA from RNA-DNA hybrids during transcription (Bishop, 1969). A more probable explanation is that both strands of the DNA are copied during transciption and that subsequently the RNA product renatures to form a poorly matched duplex.

III.b. Nearest neighbour and Partial Sequence Analyses

Analysis of the nucleotide compositions of the Apodemus cRNAs after alkaline hydrolysis identified the nucleotide to the 5' side of the $\alpha\ ^{32}P$ labelled nucletide. The results (Table 5) indicate significant clustering of A and U residues which was later confirmed by further analysis of the products of RNase T_1 and pancreatic RNase digests.

The products of digestion by RNase T_1 and pancreatic RNase were analysed on a two dimension system involving ionophoresis and ion-exchange chromatography (Southern and Mitchell, 1971; Sanger and Brownlee, 1967). A total of four analyses (two for

Table 5

Nearest Neighbour Analysis of RNA transcripts of the Apodemus Satellite DNAs.

I. UTP labelled cRNAs.

	A. Sylvaticus	A flavicollis	A agrarius
.% transfer to A	21.2	20.6	34.9
.% transfer to U	44.4	39.0	37.5
. " C	17.5	18.6	15.2
. " G	15.5	20.8	12.4

II ATP labelled cRNAs

	A. sylvaticus	A flavicollis	A agrarius
.% transfer to A	47.2	39.9	42.0
" U	16.9	22.3	22.2
" C	17.2	18.6	14.9
" G	18.7	19.2	20.9

each type of label) were required for each satellite DNA.

The 'fingerprints' (Plates 1-4) are rather complex but some features are apparent. All three DNAs show a pattern formed by a number of strong spots with an overlying pattern of weaker spots. This feature is most clearly demonstrated in the fingerprints of the RNase T_1 digests of cRNAs from A. sylvaticus and A. flavicollis (Plates 1 and 3). The predominant pattern for A. sylvaticus and A. flavicollis DNAs is very similar and most of the strong spots can be matched between these fingerprints. However, the pattern for A. agrarius satellite DNA is clearly much more complex and does not bear a strong initial resemblance to either of the other satellite DNAs. Further analysis of the major spots, described later, suggests that they are derived from the basic repeating sequence of the satellite DNA and therefore the overlying weaker pattern may have been formed as a result of divergence within this sequence. The stronger overlying pattern in A. sylvaticus DNA, relative to A. flavicollis DNA, may be partly due to main band contamination of this sample.

For each DNA, the major products after digestion of the UTP labelled RNA with RNase T_1 were long nucleotides which remained close to the origin of the chromatograms (Plate 1). From their position on the fingerprints these spots must be rich in A and U and may also contain C residues. This feature of the DNAs was confirmed upon analysis of the same RNA with pancreatic RNase. Products of the type A_nU constituted 53.2%, 47.3% and 59.4% of the total radioactivity of the pancreatic RNase fingerprints of A. sylvaticus, A. flavicollis and A. agrarius respectively (Plate 2, Table 7).

RNase T_1 digestion of the RNAs labelled with ATP gave similar fingerprints (Plate 3), in which nucleotides rich in A and U and containing C residues were again most prominent. Both RNase T_1 and pancreatic RNase analyses of the A-labelled RNA again showed that long runs of A residues are common to each satellite DNA (Plates 3 and 4).

It should be stressed that the major patterns of spots on the fingerprints result not only from the use of ATP and UTP as the radioactive triphosphates but also genuinely reflect clustering of the A and U residues, since U rich spots are apparent in RNA labelled with ATP and vice versa. (Compare Plates 1 and 3).

All of the strong and many of the weaker nucleotides were eluted from the thin layers and analysed in more detail by digesting RNase T_1 products with pancreatic RNase and vice versa. The results of these secondary analyses are shown in Tables 6 - 9. The majority of first digestion nucleotides were also analysed by nearest neighbour analysis. This latter technique was used primarily to determine the base composition of the nucleotides rather than their sequence.

III.C.Discussion

In discussing the results of the partial sequence analysis of the Apodemus satellite DNAs we are primarily concerned with attempting to answer two questions:

(1) Is each of the Apodemus satellite DNAs derived from very short repeating sequences, analogous to those of guinea pig α (Southern 1970) and mouse (Biro et al., 1974) satellite DNAs?

(2) If the Apodemus satellite DNAs consist of short repeating units, do these repeats exhibit a similarity in their base sequence

Plates 1 - 4 Separation of RNase T_1 and pancreatic RNase digests of $\alpha\ ^{32}$P-UTP and $\alpha\ ^{32}$P-ATP labelled cRNAs of the Apodemus satellite DNAs. The digests were separated by electrophoresis on cellulose acetate and chromatography on polyethyleneimine cellulose as described in section. 2.11.g

Plate 1. RNase T_1 digests of UTP labelled cRNAs.*

Plate 2. Pancreatic RNase digests of UTP labelled cRNAs.

Plate 3. RNase T_1 digests of ATP labelled cRNAs.

Plate 4. Pancreatic RNase digests of ATP labelled cRNAs.

* In this analysis some of the nucleotides on the A.flavicollis cRNA fingerprint have been eluted from the end of the thin layer. The data in Table 6 relating to the missing spots was obtained from a duplicate experiment.

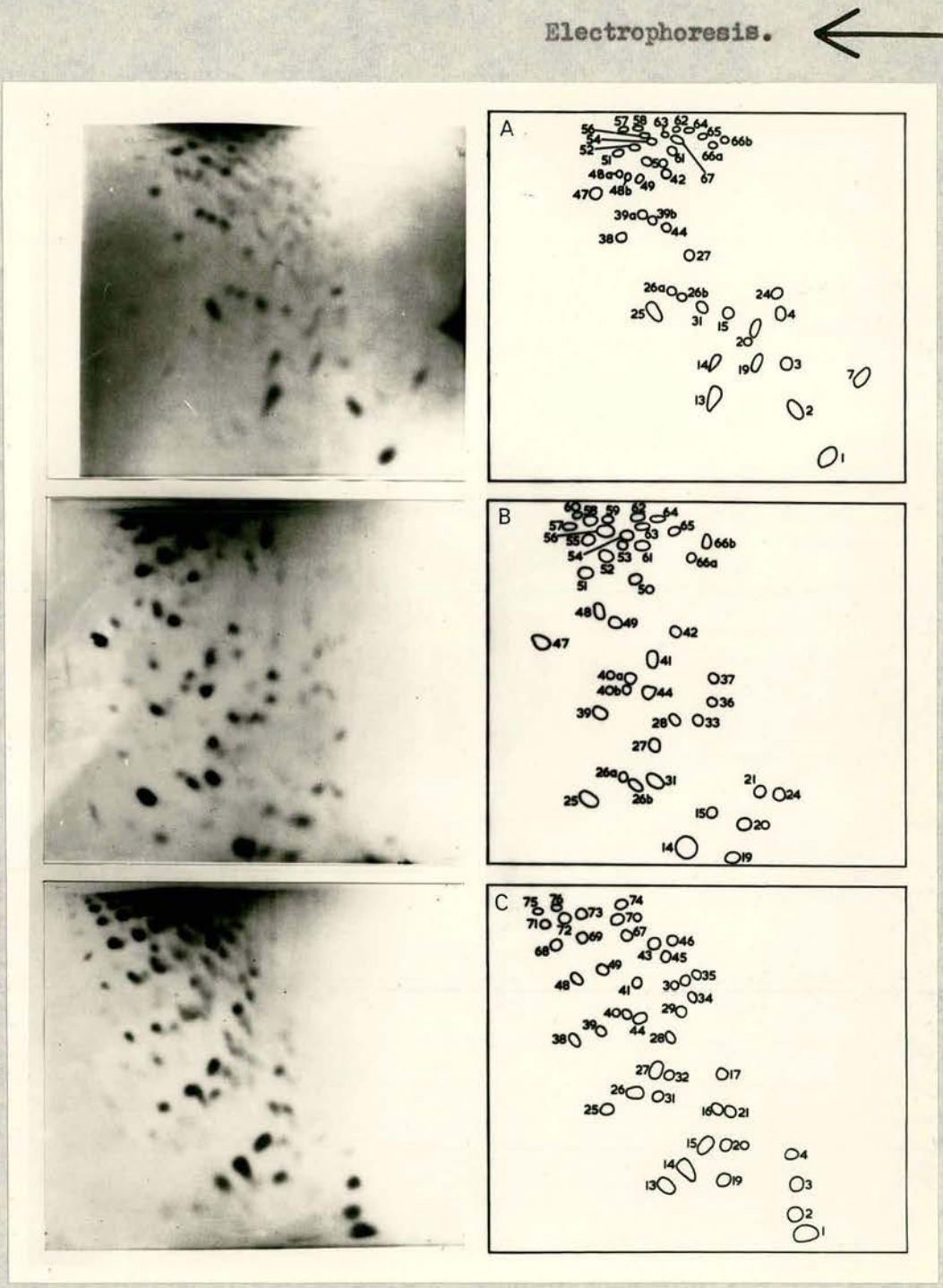

Plate I. A. A. sylvaticus.
B. A. flavicollis.
C. A. agrarius.

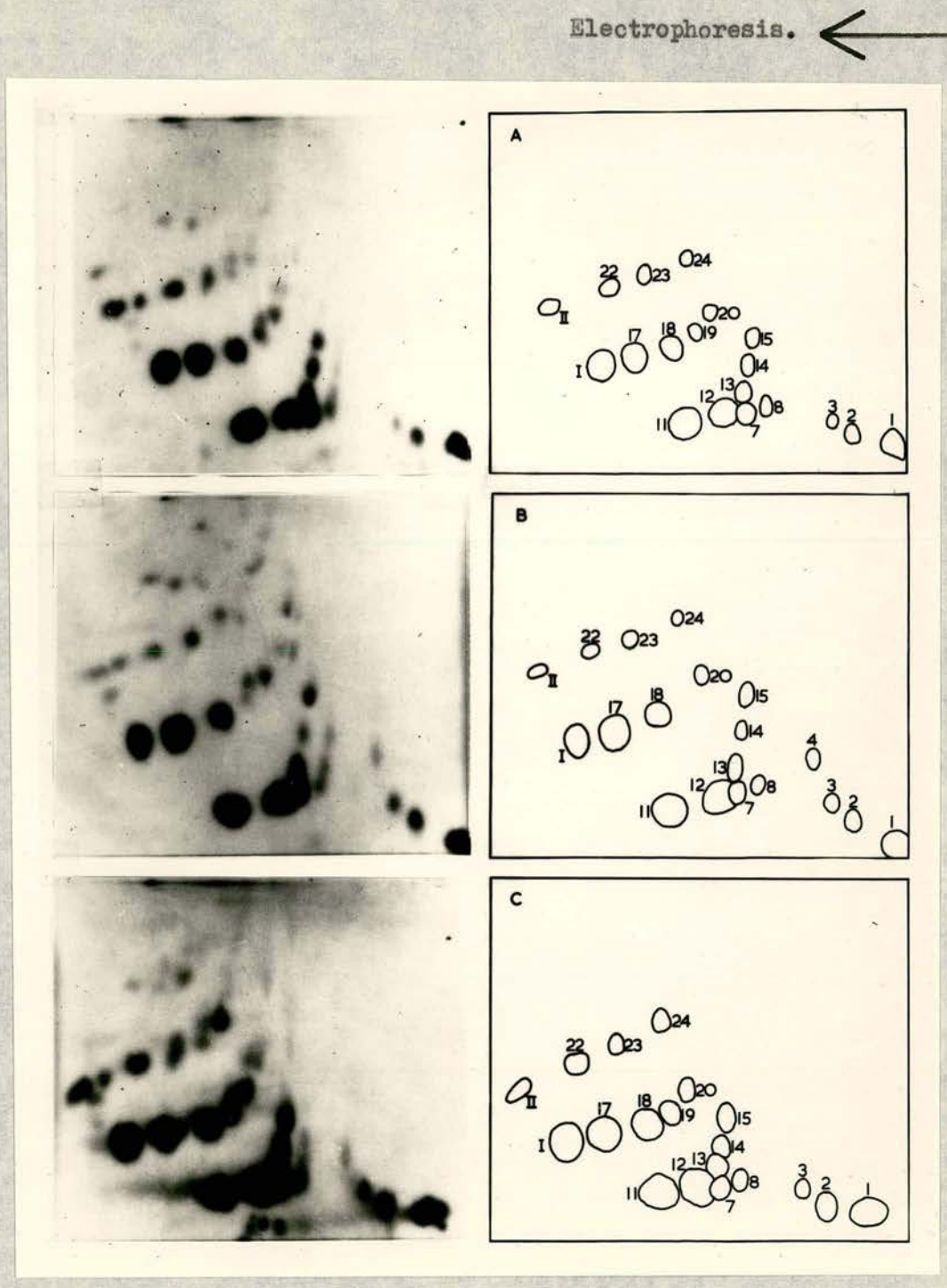

Plate 2. A. A. sylvaticus.
B. A. flavicollis.
C. A. agrarius.

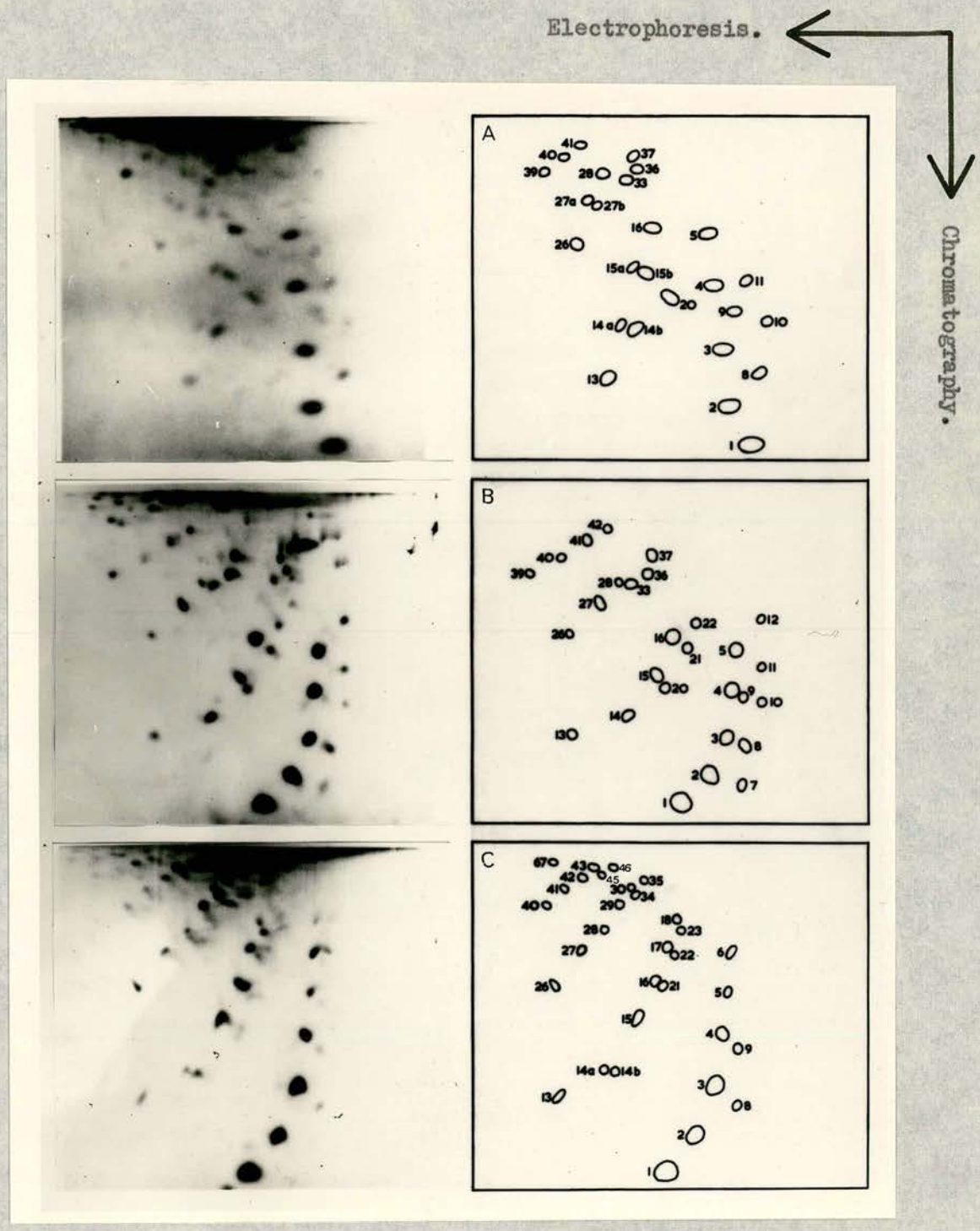

Plate 3. A. A.sylvaticus.
B. A.flavicollis.
C. A.agrarius.

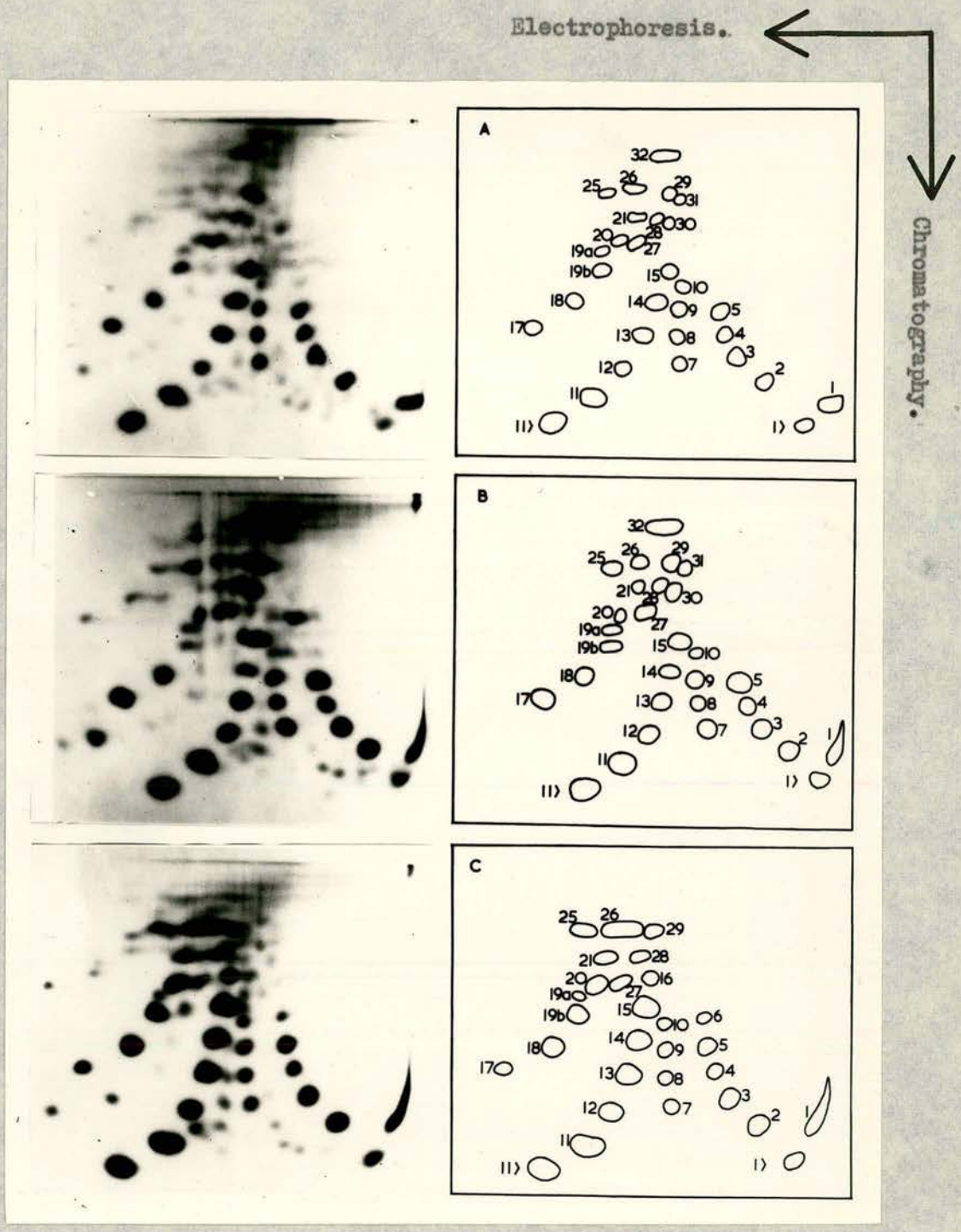

Plate 4. A. A.sylvaticus.
B. A.flavicollis.
C. A.agrarius.

Tables 6 - 9 — **Characterisation of nucleotides from fingerprints of ribonuclease digests**

Composition:- The composition of each spot was determined partly from its fingerprint position, nearest neighbour analysis and secondary digestion products.

%: The amounts of radioactivity (^{32}P) in each spot is presented as the percentage of the total radioactivity on the respective fingerprint.

Secondary Products:- Nucleotides from RNase T_1, digests were analysed by digestion with pancreatic RNase. Nucleotides from pancreatic RNase digests were analysed by digestion with RNase T_1. The secondary digestion products were separated by electrophoresis (section 2.11.g.) The secondary products are ordered by their relative amounts (visually estimated).

Table 6. Nucleotides from RNase T_1 fingerprints of α ^{32}P-UTP labelled cRNAs.

Table 7 Nucleotides from pancreatic RNase fingerprints of α ^{32}P-UTP labelled cRNAs.

Table 8 Nucleotides from RNase T_1 fingerprints of α ^{32}P-ATP labelled cRNAs.

Table 9 Nucleotides from pancreatic RNase fingerprints of α ^{32}P-ATP labelled cRNAs.

Table 6

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
1	G	5.1	G	3.3	G	2.9	G
2	AG	2.4	AG	1.8	AG	1.4	AG
3	A_2G	-		-		1.6	A_2G
7	CG	0.2	G	1.2	G	---	
13	UG	3.5	G	3.2	G	1.7	G
14a	AUG	(		(		(	
14b	AUG	(1.9	AU,G	(3.6	AU,G	(3.1	AU,G
		(		(		(	
15a	A_2UG	(		(		(	
15b	A_2UG	(0.7	A_2U,AU,G	(0.8	A_2U,AU,G	(1.8	A_2U,AU,G
		(		(		(	
16	A_3UG	-	-	-		1.0	A_3U,A_2U,AU,G
17	A_4UG	-		-		1.2	G,AU,A_2U,A_4U
19	CUG	1.3	C,G	0.7	C,G	1.6	C,G
20	ACUG	1.1	AU,G,C,AC	1.2	G,AU	---	
21	A_2CUG			0.8	AU,AG	0.7	A_2C,A_2U,AU
24	----	0.7	AU,G	-		---	
25	U_2G	3.4	G,U	5.7	U,G	1.2	U,G
26a	AU_2G	0.7	AU	0.7	AU	(	
26b	AU_2G	0.8	AU,U,C	1.5	AU,G	(2.1	AU,G

Table 6 (Contd.)

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
27	A_2U_2G	0.6	AU	1.7	A_2U,AU	1.4	A_2U,AU,G,U
28	A_3U_2G	-		1.2	AU,A_2U,G	1.0	A_3U
29	A_4U_2G	-		-		1.2	A_4U
30	A_5U_2G	-		-		0.8	AU,A_2U
31	CU,G	1.5	C.U.G	2.5	U.C.G	0.7	C,U
32	ACU_2G	-		-		0.8	AU
33	---	-		1.1	A_2G,AU,AC,G	---	
34	A_3CU_2G	-		-		1.0	AU,A_2U
35	A_4CU_2G	-		-		1.0	AU,A_2U,A_3X
36	----	-		0.8	A_2C	---	
37	----	-		1.0	U,G,AU	---	
38	U_3G	0.7	U	-		0.9	U,G
39a	AU_3G	1.1	AU,U,G	(		(	
39b	AU_3G	1.1	U,AU	(1.9 (	AU,U	(1.0 (	AU
40a	A_2U_3G	-		0.9	AU,U,G	---	
40b	A_2U_3G	-		0.9	AU,U,	---	
41	A_3U_3G	-		1.3	U,AU,A_2U	0.8	A_2U,AU
42	A_4U_3G	1.6	U,C,AU	1.0	U,G,C,AU	---	
43	A_5U_3G	-				2.3	AU,A_4U
44	U_3CG	1.2	U,C,G	2.2	U,C,G	0.9	AU,A_2U

Table 6 (Contd.)

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
45	A_5CU_3G	-		-		1.1	AU
46	A_6CU_3G	-		-		1.8	AU,A_2U
47	U_4G	1.0	U	3.9	U	—	
48	AU_4G	0.7	U,AU	1.4	A_2U,U,AU	0.9	AU,U,G
49	$A_2U_4G(C)$	0.9	AU,U,C	1.7	U,C	1.0	AU,U,A_2U
50	—	1.7	U,AU,A_2U	0.9	A_2U,AU,U	-	
51	—	1.8	U,C,AU	2.2	AU,U,C	-	
52	—	1.6	AU,U,A_2U	1.4	A_2U,AU,U,C	-	
53	—	-		0.6	U,AU,A_4U	-	
54	—	1.5	AU,U,C,A_2U	1.7	AU,A_3U,U,C,G	-	
55	—	-		1.2	U,C,AU	-	
56	—	1.9	AU,U,C,A_2U	2.4	U,C,AU,A_2U	-	
57	—	2.5	Au,U,A_2U	1.4	U,C,AU	-	
58	—	2.8	AU,A_2U,U	2.5	U,A_2U,AU	-	
59	—	-		1.4	AU,C,U,A_2U	-	
60	—	-		0.6	AU,U,C	-	
61	—	1.5	AU,A_2U,U	1.7	U,AU,A_2U	-	
62	—	2.5	AU,A_4X,A_2U,U	3.0	U,AU,A_2U,A_4X	-	
63	—	1.8	AU,U,A_3U,A_2U	1.0	AU,U,A_3U,A_2U,C	-	
64	—	1.7	AU,U	1.6	AU,U,G,C	-	

Table 6 (Contd.)

	Apodemus sylvaticus			Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
65	---	1.3	AU,U,A_2U,C	1.0	U,G	-	
66a	---	1.0	AU,U	-		-	
66b	---	0.7	AU,A_2U,C	0.9	AU	-	
67	---	-				1.1	AU,U,A_2U,A_3U
68	---	-				1.3	AU,U
69	---	-				1.1	AU,U
70	---	-				0.9	AU,U
71	---	-				0.6	AU,U
72	---	-				1.0	AU,U
73	---	-				1.5	AU,U,C
74	---	-				1.4	U,C,G,AU
75	---	-				0.7	U,AU
76	---	-				0.7	AU,U

Table 7

	Apodemus sylvaticus			Apodemus flavicollis			Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products	
1	C	8.8	C	9.1	C	7.4	C	
2	AC	1.5	AC	1.7	AC	2.2	AC	
3	A_2C	0.5	A_2C	0.8	A_2C	0.8	A_2C	
4	A_3C	-		0.4	A_3C	-		
7	GC	4.6	C	3.7	C	4.1	C	
8	AGC	1.2	C	0.7	C	0.9	C	
11	U	20.7	U	18.5	U	15.1	U	
12	AU	11.5	AU	13.5	AU	23.8	AU	
13	A_2U	2.5	A_2U	3.0	A_2U	4.6	A_2U	
14	A_3U	1.7	A_3U	0.6	A_3U	1.9	A_3U	
15	A_4U	1.0	A_4U	1.2	A_4U	1.9	A_4U	
17	GU	14.0	G,U	21.1	G,U	8.5	G	
18	AGU	3.7	AU,AG	4.3	AU,AG	5.8	AU,AG	
19a	A_2GU	(				(		
19b	A_2GU	(1.0	A_2U,AU	-		(1.2	AU,A_2U	
20	A_3GU	(0.6	A_3U,A_2U,AU	0.8	A_3U,A_2U,AU	(1.3	A_2U,A_3U	
22	G_2U	1.8	G	1.0	G	1.2	G	
23	AG_2U	1.0	AG	1.2	AG	0.6	AG,AU,G	
24	---	-		-		0.9	AU	
I	U_2	13.8	U	9.7	U	10.9	U	
II	U_3	2.0	U	0.8	U	1.3	U	

Table 8

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
1	G	-	G	6.7	G	8.8	G
2	AG	-	AG	6.4	AG	5.9	AG
3	A_2G	-	A_2G	4.6	A_2G	9.0	A_2G
4	A_3G	-	A_3G	5.0	A_3G	3.0	A_3G
5	A_4G	-	A_4G	5.1	A_4G	1.2	A_4G
6	A_5G	-		-		1.6	A_5G
7	CG	-		0.7	G	-	
8	ACG	-	C,G,AG	1.0	G,AG,C	1.0	C,AG,G
9	A_2CG	-	A_2C	0.6	A_2C,AC	0.8	AC,A_2G,AG,C
10	----	-		0.6	C,G	-	
11	----	-		0.9	AC,C	-	
12	----	-		1.3	A_2C,C,G	-	
13	UG	-	G	0.9	G	1.1	G
14a	AUG	-	G	(		0.8	G
14b	AUG	-	G	(1.9	U,AG,G	1.1	U,AG
15a	A_2UG	-	A_2G,U	(		2.3	AU,A_2U,G
15b	A_2UG	-	A_2U	(2.0	A_2G,U,A_2U,AU	1.0	A_2G,A_2U
16	A_3UG	-	A_3U,U	4.0	A_3G,U	1.1	A_3U,AU
17	A_4UG	-		-		1.4	A_4X
18	A_5UG	-		-		0.7	A_5X

Table 8 (Contd.)

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
20	XCUG	-	U,C	1.1	U	-	A_5X
21	A_2CUG	-		1.1	C,AU,A_2G	1.6	A_2C,A_2U,G
22	A_3CUG	-		1.0	AC,U	0.7	A_2G,A_2U,AC
23	A_4CUG	-		-		0.6	A_4X
26	AU_2G	-		0.8	AU,U	1.3	U,G,AC
27	A_2U_2G	-		1.3	U,A_2G,A_2U,AU	0.9	A_2U,G
28	A_3U_2G	-		0.7	A_2U,U	0.5	A_2U
29	A_4U_2G	-		-		1.1	A_4U,G
30	A_5U_2G	-		-		0.6	A_4X,A_2U,AU
33	---	-	U,A_2U,A_2G	1.1	U,A_2G,A_2U,AG	-	
34	A_3CU_2G	-		-		0.8	AU,A_2U,A_3U
35	A_4CU_2G	-		-		0.8	A_2U,AU
36	---	-	U,A_2C,G	1.8	A_2C,G,U	-	
37	---	-	A_2C,U,AC	1.1	A_2C,AC,U	-	
39	AU_3G	-	U	0.8	U	-	
41	A_3U_3G	-	U,AU	0.7	U	-	
42	A_4U_3G	-		0.6	A_2C,U	0.6	A_4U
43	A_5U_3G	-		-		1.2	A_4U,AU
45	A_5CU_3G	-		-		0.7	AU,A_2U,A_3U,A_4U
46	A_6CU_3G	-		-		0.8	AU,A_2U,A_4U
67	---	-		-		0.5	U,G,AU,A_2U,A_3U

Table 9

		Apodemus Sylvaticus		Apodemus Flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
1	C	2.0	C	2.0	C	2.1	C
2	AC	4.1	AC	6.3	AC	4.2	AC
3	A_2C	3.8	A_2C	4.6	A_2C	3.3	A_2C
4	A_3C	2.6	A_3C	1.7	A_3C	1.2	A_3C
5	A_4C	2.5	A_4C	3.8	A_4C	2.3	A_4C
6	A_5C	-		-		0.6	A_5C
7	GC	2.1	C	2.3	C	1.0	c
8	AGC	2.4	G,C,AC	1.3	G.C	1.3	G.AC,C
9	A_2GC	1.9	AG,A_2G,A_2C,AC	1.4	A_2G,AG	1.5	AG,A_2C,A_2G
10	A_3GC	1.0	A_3X	0.5	A_2G	0.8	AG,A_2C
11	U	5.3	U	5.6	U	6.5	U
12	AU	1.6	AU	2.6	AU	3.7	AU
13	A_2U	3.6	A_2U	2.9	A_2U	5.0	A_2U
14	A_3U	4.0	A_3U	1.1	A_3U	4.9	A_3U
15	A_4U	3.2	A_4U	2.2	A_4U	6.3	A_4U
16	A_5U	-		-		1.3	A_5U
17	GU	2.4	U	4.7	U	1.4	U
18	AGU	2.1	G	2.0	G,U	4.7	AU,G,U
19a	A_2GU	0.7	A_2G	0.6	A_2G	0.5	A_2G
19b	A_2GU	1.4	A_2U,AG,C	0.5	A_2U,AG,C	2.8	AG,AU,A_2U,G
20	A_3GU	1.0	A_3U	0.5	A_3U	3.2	A_2U

Table 9 (Contd.)

		Apodemus sylvaticus		Apodemus flavicollis		Apodemus agrarius	
Spot No.	Composition	%	Secondary Products	%	Secondary Products	%	Secondary Products
21	A_4GU	1.2	A_3G	0.5	A_2G,A_2U	1.5	A_4X
25	—	0.5	Origin,G	0.9	Origin	1.6	A_2G,AG
26	—	0.8	A_3C,A_2G,AG,G	0.5	A_3C,G	4.9	A_2G,AG
27	—	1.5	A_2G,AG	1.6	A_2G,AG	1.0	A_2U,A_2G,AG
28	—	1.2	AG,A_2U,A_2U,G	1.4	A_4X,A_3U	1.1	A_4X,A_3U
29	—	1.8	A_3U,A_2G,G	2.2	A_3U,A_3G,A_2G	1.3	A_4X,A_2G,AG
30	—	0.6	A_3C,A_2G,AG,G	1.5	A_3C,A_3G,G,A_2G	-	
31	—	0.7	Origin,G	0.8	Origin,G	-	
32	—	2.1	A_3U,A_2G.G	2.4	A_3U,A_3G	-	
1	Cyclic C	-		-		-	
11	Cyclic U	6.6		6.2		5.6	

between the three species? In other words, is the basic repeating sequence for each satellite DNA derived from a common ancestral short sequence?

Some satellite DNAs appear to have evolved by a step-wise multiplication of stretches of DNA consisting of a tandemly repeating short sequence (Southern, 1970; Sutton et al., 1972). Base changes introduced into the simple sequence are amplified during subsequent multiplication steps. In the sequence analysis of a DNA which has evolved in this manner, divergence within the short repeating sequences manifests itself as related nucleotides on a fingerprint. For example, more than half of the pyrimidine tracts obtained by diphenylamine - formic acid digestion of either the light or heavy strand of guinea pig α satellite DNA are related to the main hexamer sequence by only one base change. (Southern, 1970).

From the limited analysis of the Apodemus satellite DNAs it was not possible to determine the sequence of more than a few relatively short nucleotides. However, comparison within each species, of the secondary products clearly reveals that many of the nucleotides on each fingerprint are related in sequence.

Consider for example, the secondary products obtained from the RNase T_1 fingerprints of the UTP labelled cRNAs (Table 6). From the A. sylvaticus analysis spots numbered 50, 52, 57, 58 and 61, accounting for 10.1% of the total radioactivity gave the same secondary digestion products, these being U, AU and A_2U. Furthermore, another 12 spots from the same analysis, amounting to 16.1% of the radioactivity, gave products differing from those above by one one base. Similarly for A. flavicollis spots 41,

48, 50, 58, 61 and 62 totalling 10.8% of the radioactivity also gave U, AU and A_2U as their only secondary digestion products. Also nucleotides differing by only one base in their secondary products were present in 13 spots amounting to 18.2% of the total radioactivity of this RNA. However, sequence relationships between the nucleotides on the fingerprints was best illustrated in the case of A. agrarius. More than half of the nucleotides obtained from this RNA gave AU as their main product after further analysis. Furthermore, spots amounting to 36% of the total radioactivity gave only products of the type AnU and occasionally a G residue. Taken together, the results from the above analyses, with those from the other three analyses of each satellite DNA strongly suggest that each of the Apodemus satellites are derived from short repeating sequences.

When analysing fingerprints of uniformly ^{32}P labelled nucleic acids the frequency with which specific nucleotides occur within a fixed length of nucleic acid can be simply determined (Southern, 1970). By this method it is sometimes possible to estimate the length of the ancestral sequence of the nucleic acid. However, in order to apply the same method to fingerprints of nucleic acids labelled only in one nucleotide it is necessary to know the number of phosphate groups contributed by the labelled nucleotide. These phosphate groups may be internal or at the 3' end by transfer. In this preliminary investigation the nucleotides were not analysed so rigorously, and therefore an estimation of the sequence complexity could not be made. Nevertheless, some idea of the complexity of a sequence is given by the number of spots seen on fingerprints. For example RNase T1

digests of 5s (about 120 nucleotides long) RNAs give around twenty spots whereas 16s RNAs (about a thousand nucleotides long) give more than one hundred. Sequence complexity is less simply defined for heterogeneous sequences. For example the ovarian 5s RNA of Xenopus laevis gives a pattern of strong spots arising from the principle sequence and is overlaid with a pattern of weaker spots arising from minor sequences (Ford and Southern, 1973). The situation is even more complex for satellite DNAs having more than one level of complexity (Biro et al, 1974). For the Apodemus satellite DNAs at least two such levels of complexity are suggested. The many strong spots are derived from divergence within the ancestral sequence and the overlying pattern of weaker spots comes from divergence within a more recently formed longer repeat. Another significant complication in the case of the Apodemus analyses is that the cRNAs from the Apodemus DNAs are double stranded. Therefore digestion products from both strands are present on the fingerprints.

Although in the preceding discussion we have considered the satellite DNAs separately, the partial sequence analysis suggests that they share a common ancestral sequence. To illustrate this point it is necessary to compare the secondary digestion products of corresponding nucleotides from equivalent fingerprints of the different species. For A. sylvaticus and A. flavicollis this task is relatively easy as their fingerprints are not only comparatively simple but are also very similar. However, A. agrarius cRNA gave more complex maps and thus matching of corresponding nucleotides between this species and the other two

was more difficult. In fact many of the spots on the A. agrarius fingerprints have been given different numbers (67 upwards on Plate 1c) because they could not be matched with any certainty to the nucleotides in corresponding positions on the fingerprints of A. sylvaticus and A. flavicollis cRNAs (spots numbered 50-66, Plates 1a and b). As a result only a broad comparison can be made between the products of the large nucleotides from the A. agrarius fingerprints and those of A. sylvaticus and A. flavicollis. Nevertheless, the analyses do suggest a strong similarity in the basic sequence of the satellite DNAs of the three species. Consider for example the general secondary digestion products obtained from the nucleotides of the RNase T_1, UTP labelled cRNA fingerprints (spots numbered > 40, Plate 1, Table 6). The outstanding products for all three satellite cRNAs were AU and U, although C residues and products of the type A_nU figured prominently in the A. flavicollis analysis and to a lesser extent in the A. sylvaticus analysis. More specific examples can be given to illustrate the similarity between A. sylvaticus and A. flavilcollis satellite DNAs. In these latter analyses spots number 42, 44, 50, 51, 56, 58, 61 and 62 gave the same secondary digestion products and a further 8 nucleotides differed with respect to only one base. These results strongly suggest that the Apodemus satellite DNAs share a common ancestral sequence.

4. CONCLUSION

The repeated sequences found in higher organism DNA must be the result of a process involving multiplication or accumulation. Very little is known about this process. The molecular mechanism involved may be enzymatic copying or unequal crossing over. With respect to the time scale of the process, copies may be made all in one burst of saltatory replication (Britten and Kohne, 1968) or accumulate gradually over many generations (Southern, 1970). Some of these questions may be approached by studying the relationships between the different members of the set of sequences which form a single satellite DNA and the relationships between the sequences of different satellite DNAs.

The satellite DNAs of A. sylvaticus, A. flavicollis and A. agrarius appear to derive from the same ancestral sequence and yet have very different physical properties. It is easy to see how repeated sequences with different physical properties can arise from the same ancestral sequence. If the ancestral sequence is itself a diverged repeated sequence and different portions are taken to produce three new repeated sequences, these will have properties which reflect the qualitative differences in the portions taken for multiplication. It is a corollary to this argument that the portions taken for multiplication are short enough that a non-random selection of the divergence is made.

The set of sequences which go to make the satellite DNA of A. sylvaticus is less diverged than that in A. agrarius, implying that most of these copies have accumulated since the two species diverged. Assuming that the Δ Tm reflects the average divergence

of the set and that base changes accumulate at a rate of 1% per million years (Kohne, 1970; Laird et al., 1969), the age of the A. sylvaticus set of sequences is around 6 million years. Applying the same assumptions to the satellite DNA of A. agrarius it would appear that these sequences are about 5 million years older than those in A. sylvaticus. It is possible therefore that the set of sequences which make up the A. agrarius satellite DNA were present in the common ancestor to A. agrarius and A. sylvaticus and that the A. sylvaticus sequences were made by a multiplication from a small portion of this set. In other words, the A. sylvaticus satellite is less diverged because it was produced more recently than the A. agrarius satellite DNA. In this case, the other A. agrarius sequences must have been deleted in the line to A. sylvaticus. In a similar manner one can propose a scheme whereby the satellite DNA of A. flavicollis, which appears to be 3 million years younger than that of A. sylvaticus, may also have evolved from the diverged sequences of A. agrarius satellite DNA.

Although A. sylvaticus satellite DNA renatures to form a duplex with a Δ Tm of 11°C, suggesting an age of 6 million years, it is unlikely that the poorly formed duplex actually reflects divergence within only one set of sequences. Cooke (1974) has suggested that this satellite DNA contains a number of sequence related families which are able to renature with each other to form poorly matched duplexes. This is supported by the fact that the pattern of restriction fragments found on analysis of A.sylvaticus satellite DNA suggest an age of only 2 - 3 million years for the most recently formed repeats. Therefore the satellite DNAs of A. sylvaticus and A. flavicollis must have been formed at or about the same time.

These results imply that a number (probably > 4) of related sequences may have been taken for separate multiplications in the line to A. sylvaticus while only one sequence was involved in the formation of A. flavicollis satellite DNA. This mechanism has been proposed previously to account for the relationships seen between the satellite DNAs of Mus species (Sutton and McCallum, 1972), the satellite DNAs II and III of guinea pig (Southern 1973) and bear some analogies to the mechanisms proposed to account for the evolution of multiple copies of ribosomal DNA (Brown, Wensink and Jordan, 1972; Ritossa, 1972). If A. sylvaticus satellite DNA evolved in the above manner then the Δ Tm of the cross-reassociated molecules reflects the divergence between the selected sequences at the time they were taken for multiplication and suggests that they were then 3 - 4 million years old. However, this is a very unreliable estimate as in all probability the sequences chosen for multiplication would not have been a representative sample of the diverged precursor.

There are indications from the studies on A. agrarius satellite DNA that it also contains many sets of sequences, and therefore may have evolved in the same manner as A. sylvaticus satellite DNA. However, in this instance very little can be concluded about either the number of families of sequences involved or the extent to which they are able to cross-reassociate with each other. This problem stems from the fact that when studying such highly mismatched duplexes it is difficult to estimate just how much the rate of renaturation is influenced by the quality of the duplex being formed. It should be mentioned in this context, that the Δ Tm values

measured for the reassociated A. agrarius satellite DNA may underestimate the real extent of divergence between the reacting sequences. This is based on the assumption that those sequences which have not renatured during the course of the reaction were to some extent prevented from doing so because they were even more mismatched than those sequences which did renature. Kohne(1970) has stressed the importance of this point when using reassociation techniques to determine sequence homologies between the DNAs of related species.

Nevertheless, it may be that within each of the sequence families of A. agrarius satellite DNA the extent of divergence is no more than that found in A. flavicollis satellite DNA and that the poorly matched duplexes formed upon reassociation are a result of cross-reassociation between members of different families. In this case the satellite of A. agrarius may have been formed at the same time as the other Apodemus satellite DNAs. Alternatively, as already suggested, A. agrarius satellite DNA may be the common highly diverged ancestral precursor to the satellite DNAs of A. sylvaticus and A. flavicollis.

The strong implication from the experiments with A. agrarius satellite DNA is that it contains a repeat every 3-400 base pairs and that reassociation kinetics overestimates the sequence complexity by a factor of about 300. A number of possible structures for the satellite can be proposed to explain this anomaly. Three such proposals, listed below, are extreme cases and the true structure of A. agrarius satellite DNA may lie between any two of the three.

A. agrarius satellite DNA may consist of:-

(1) A single set of highly diverged repeating sequences with a complexity of 300 base pairs.

(2) A large number of sets of sequences which are able to cross-reassociate with each other.

(3) A large number of sets of sequences which are not able to cross-reassociate with each other.

A number of properties of A. agrarius satellite DNA are not in agreement with the last proposal. The dramatic increase in the rate of reassociation for a small reduction in temperature, the differences in the homogeneity and stability of the duplexes formed at different temperatures of reassociation and finally, the banding pattern of the reassociated satellite DNA indicate that many of the sets of sequences are able to cross reassociate with each other. The homogeneity of the satellite is further emphasised by the sharp melting curve, of both components, of the satellite DNA. It should be mentioned however, that although A. agrarius satellite DNA appears to be relatively homogeneous with respect to its physical properties, this cannot be taken as suggesting a limited extent of divergence within the families of sequences. The presence of high melting regions and sonication - susceptible sites, evenly distributed throughout the DNA, are properties of the DNA which are probably related to structural features of the DNA resulting from localised sequence differences. These regions are therefore likely to be able to sustain a high degree of divergence without loosing their inherent properties.

In terms of renaturation behaviour there is no real distinction between proposals (1) and (2) above, as a single set of highly diverged sequences may be regarded as comprising of a number of sets

with differing degrees of divergence and therefore having varying capacities for cross-reassociation. If mismatching has only a small effect upon the rate of reassociation as suggested by Bonner et. al., (1973) and others (Wang and Kallenbach, 1971; McCarthy and Farquhar, 1972), and the structure of A. agrarius is as contained in proposals (1) or (2) above, then the DNA should renature 50 times faster than was actually found. Therefore, it is strongly indicated by these studies that the rate of renaturation of A. agrarius satellite DNA is highly dependant on the extent of mismatching in the duplexes formed.

Part of the problem with respect to the mismatching/rate controversy may lie in the design of the experiments which have been used in this field of inquiry. Bonner et al., (1973) determined the maximum rate of reassociation of their mismatched duplexes at optimum temperatures, and estimated the effect of mismatching upon the rate of reassociation from these rates. Sutton and McCallum (1971), however, determined the rates of reassociation of the mismatched duplexes at a fixed temperature, 25° below the Tm of the native DNA. A brief review of the literature reveals that when determining the rate of reassociation of a DNA most workers adopt the latter technique and operate at a temperature 25° below the Tm of the native duplex. In fact when studying samples which have been isolated in a non-native state (kinetic fractions) the standard procedure is to renature under conditions equivalent to 0.18m-Na^+ at 60°.

The rate of renaturation of very highly mismatched duplexes is critically dependant on the temperature of reassociation

(Kallenbach and Drost, 1972). A. agrarius satellite DNA does show a dramatic increase in its rate of reassociation as the temperature is decreased. However on renaturation at 50° (34° below the Tm of the satellite DNA) the results suggest that only extremely unstable duplexes are formed at this temperature. This observations raises another problem which may be relevant to the studies of mismatched duplexes. It is possible that the optimum rate of reassociation of the diverged, 3-400 base pair repeat in A. agrarius satellite DNA is at a temperature well below what was actually used. The partial sequence analysis of this DNA suggests that within the 3-400 base pair repeat, there is a much shorter repeating sequence. Although at this lower level of complexity the divergence between the sequences may be quite high, at very low stringency conditions of reassociation these sequences may begin to determine which register of reassociation is involved in duplex formation. In other words, upon reassociation at low temperatures, duplexes may form in the shorter register. The latter register is unlikely to be the same as for the longer repeat. The resulting structures will be very short, highly mismatched duplexes with long single stranded tails which are unable to zipper up. In effect, reassociation in the conventional sense will be prevented from taking place. One might also predict that a similar situation may arise in other DNAs which contain more than one level of complexity. A number of satellite DNAs have this type of sequence complexity hierarchy and may not, therefore, obey the relationship found for mismatching upon the rate of reassociation determined from non-repetitive DNAs.

The properties of A. agrarius satellite DNA also contribute to our present interpretation of the evolution of repeated sequences. The evolutionary origins of satellite DNAs either by the formation of a tandemly repeating short sequence (Blumenfeld, 1973; Gall and Atherton; 1973) or by the amplification of a portion(s) of a diverged repeating short sequence (Southern, 1973; Sutton and McCallum, 1972) have already been discussed in some detail. These fastest reassociating sequences appear to have been introduced into the genome most recently, since the broad pattern found for their reassociation is that they give less mismatched duplexes than those repeated sequences which reassociate more slowly. It is generally accepted that the more slowly reassociating sequences, such as the intermediate DNA, evolve from the faster reassociating sequences. However, there is disagreement over the way in which these changes are thought to take place. In the interpretation of Britten and Kohne (1967) the sequences in the fast reassociating DNA are present in more copies than in more slowly reassociating DNA. Therefore a loss of many copies of a fast fraction will reduce the concentration of these sequences and they will subsequently renature more slowly. Alternatively a single family of rapidly renaturing sequences may evolve into a set of different families with sequences which will reassociate only within a family and not between families. This pathway seems to have been followed in the evolution of ribosomal DNA (Brown et al., 1972). A third method is that random divergence of a fast reassociating family can slow down the rate of reassociation to such an extent that it becomes an intermediate family.

The properties of A. agrarius satellite DNA are best explained by this last method. Furthermore, the general application of the first two proposals to the evolution of repeated sequences is not borne out by experimental observation. A wide study of the cross-reassociation of repeated sequences from related species of rodents (Rice, 1972) and primates (Gummerson, 1972) show that there is no major deletion of repeated sequences as species diverge. Also, in the absence of new amplification events, the divergence of repeated sequences is a random process. Therefore, the broad pattern of changes occuring in the repeated sequences can be explained simply as a progression from the introduction of the fast fraction into the genome, through the intermediate fraction and ultimately into the unique fraction (McCarthy and Farquhar, 1972), the changes being brought about by increasing divergence. The progression is cyclical in that new repeated sequences are introduced by multiplying sections taken out of old repeated sequences.

References.

Allan,J., Polito,L., Southern,E.M. and Duncan,G. (1974) In preparation.

Arrighi,F.E., Mandel,M., Bergendahl,J. and Hsu,T.C. (1970). Biochemical Genetics, 4, 367.

Bernardi,G. (1965). Nature, 206, 779.

Birnstiel,M.L., Chipchase,M. and Spiers,J. (1971). Prog. Nucleic Acid Res. Mol. Biol., 11, 351.

Biro,P.A.B., Carr-Brown,A., Southern,E.M. and Walker,P.M.B.(1974) In preparation.

Bishop,J.O. (1969). Biochim. Biophys. Acta.,174, 636

Blemenfeld,M. (1973). Cold Spring Harbor Symposia on Quantitative Biology, 38.

Bolton,E.T., Britten,R.J., Cowie,D.B., Roberts,R.B., Szafranski,P. and Waring,M.J. (1966) Carnegie Inst. Year Book, 64, 316.

Bonner,T.I., Brenner,D.J., Neufeld,B.R. and Britten,R.J.(1973). Cold Spring Harbor Symposia on Quantitative Biology, 38.

Borst,P. and Ruttenberg,G. (1966). Biochim. Biophys,Acta.,114, 643.

Bostock,C.J. and Prescott, D.M. (1971) Exp. Cell Res.,64, 267.

Bostock, C.J., Prescott,D.M. and Hatch, F.T. (1972). Exp. Cell Res., 74, 487.

Bostock,C.J.(1972). Advances in Cell Biology, 2, 1853. Ed. Prescott, D.M. Goldstien,L. and McConkey,E.New York: Appleton Century Crofts.

Britten,R.J. (1968). Carnegie Inst. Year Book, 67, 327.

Britten, R.J. and Davidson,E.H. (1969). Science, 165, 349.

Britten,R.J. and Kohne,D.E. (1966). Carnegie Inst.Year Book,65, 78.

Britten,R.J. and Kohne,D.E.(1967). Carnegie Inst.Year Book,66, 73.

Britten,R.J. and Kohne,D.E.(1968). Science, 161, 529.

Britten,R.J. and Smith,J. (1970) Carnegie Inst. Year Book, 68, 378.

Brown,D.D., Wensink,P.C. and Jordan,E. (1972). J. Mol. Biol.,63, 57.

Brunk,C.F. and Leick,V. (1969). Biochim.Biophys.Acta.,179, 136.

Burgess,R.R. (1969). J. Biol. Chem., 244, 6160.

Burton,K. and Petersen,G.B.,(1960). Biochem.J. 75, 17

Callan,H.G. (1967). J. Cell Science, 2, 1.

Chun,E.H.L. and Littlefield,J.W. (1963). J. Mol. Biol.,7, 245.

Church,R.B. and McCarthy,B.J. (1967). J. Mol. Biol., 23, 459.

Cooke, H.J. (1974). In preparation.

Corbin,K.W. and Uzzell,T. (1970). Amer. Naturalist,104, 37.

Corneo,G., Ginelli,E., Soave,C. and Bernardi,G. (1968a). Biochemistry, 7, 4373.

Corneo, G., Ginelli,E. and Polli,E. (1968b). J. Mol. Biol.,33, 331.

Corneo, G., Ginelli,E. and Polli,E. (1970a). Biochemistry,9, 1565.

Corneo,G., Ginelli,E. and Polli, E. (1970b). J. Mol. Biol.,48, 319.

Corneo,G., Moore,C., Sanadi,D.R., Grossman,L.T. and Marmur,J. (1966). Science, 151, 687.

Crick, F.H.C., (1971). Nature, 234, 25.

Flamm, W.G., Bond,H.E. and Burr,H.E. (1966a). Biochim. Biophys. Acta., 129, 310.

Flamm,W.G., Bond,H.E., Burr,H.E. and Bond, S.B. (1966b). Biochim. Biophys. Acta., 123, 652.

Flamm, W.G., McCallum,M. and Walker,P.M.B. (1967). Proc. Natl. Acad. Sci., U.S., 57, 1729.

Flamm,W.G., Walker,P.M.B. and McCallum,M. (1969a). J. Mol. Biol., 40, 423.

Flamm,W.G., Walker,P.M.B. and McCallum,M. (1969b). J. Mol. Biol., 42, 441.

Ford, P.J. and Southern, E.M. (1973) Nature New Biology,241, 7.

Fry,K., Poon,R.,Whitcome,P.,Idriss,J., Salser,W.,Mazrimas,J. and Hatch,F. (1973) Proc.Natl.Acad.Sci.,U.S. 70, 2642.

Gall, J.G. and Atherton,D.D.(1973) Cold Spring Harbor Symposia on Quantitive Biology. 38.

Georgiev, G.P. (1969) J. Theoret. Biol.,25, 473.

Gilbert, W. and Dressler,D. (1968) Cold Spring Harbor Symposia on Quantitive Biology, 33, 473.

Gummerson, K.S. (1972) Ph.D.Thesis, Johns Hopkins University.

Hennig, W., Hennig,I. and Stein,H. (1970) Chromosoma, 32, 31.

Hennig, W. and Walker, P.M.B. (1970) Nature, 225, 915.

Hutton, J.R. and Wetmur, J.G.(1973) Biochem. Biophys.Res.Comm. 52, 1148.

Jacob, F. and Monod,J. (1961) J. Mol. Biol., 3, 318.

Jensen, R.H. and Davidson, N. (1966) Biopolymers. 4, 17.

Jones, K.W. (1970) Nature 225, 912.

Kallenbach, N. and Drost, S.D. (1972) Biopolymers. 11, 1613.

Kedes, L. and Birnstiel, M.L. (1971) Nature 230, 165.

Kelly, T.J. Jr. and Smith, H.O. (1970) J. Mol. Biol. 51, 383.

Keyl, H-G. (1965). Chromosoma, 17, 139.

Kohne, D.E. (1970) Quart. Rev. of Biophys. 33, 327.

Kit, S. (1961a) In the Molecular Basis of Neoplasia, p.133,M.D.Anderson. Hospital and Tumor Institute, University of Texas Press, Austin, Texas.*

Kit, S. (1961)J. Mol. Biol. 3, 711.

Kit, S. (1962) Nature, 193, 274.

Laird, C.D. and McCarthy, B.J. (1969) Genetics, 63, 865.

Laird, C.D., McConnaughty, B.L. and McCarthy, B.J. (1970) Nature, 224, 149.

Lima-de-Faria, A. (1969) Handbook of Molecular Cytology, p 277 Ed. Lima-de-Faria. North-Holland Publ. Co.

Maio, J.J. and Schildkraut, C.L. (1969) J. Mol. Biol., 40, 203.

Marmur, J. and Doty, P. (1961) J. Mol. Biol., 3, 585.

Marmur, J. and Doty, P. (1962) J. Mol. Biol., 5, 109.

Mattoccia, E. and Comings, D.E. (1970) Chromosoma, 30, 202.

McCallum, M. and Walker,P.M.B. (1967) Biochem. J., 105, 163.

McCarthy, B.J. and Farquhar, M.N. (1972) In Evolution of Genetic systems, Brookhaven Symposium No. 23 p.1. New York: Gordon and Breach.

McCarthy, B.J., Shearer, R.W. and Church, R.B. (1969) In Problem in Biology: RNA in Development. Ed. Hanly, E.W., Salt Lake City, University of Utah. Press, p.285.

Meselson, M, Stahl, F.W. and Vinograd, J. (1957) Proc. Natl. Acad. Sci. U.S. 43, 581.

Mitchel, R.M. (1973) M.Sc thesis, University of Edinburgh.

Miyazawa, Y. and Thomas, C.A. Jr. (1965) J. Mol. Biol. 11, 223.

Muller, H.J. (1967) Heritage from Mendel, 419 (R.A. Brink, editor) Madison: University of Wisconsin Press.

Nandi, U.S. Wang, J.C. and Davidson, N. (1965) Biochem., 4, 1687.

Ohno, S. (1970) Evolution by Gene Duplication. Springer, Heidelberg.

Ohno, S. (1971) Nature, 234, 134.

Ohno, S. (1973) Modern Aspects of Cytogenetics: Constitutive Heterochromatin in Man. Symposia Medica Hoeschst 6. F.K. Schattauer Verlag. Stuttgart - New York.

Ohno, S, Kaplan, W.D. and Kinosita, R. (1957) Exptl. Cell Res. 13,358.

Pardue, M.L. and Gall, J.G. (1970) Science 168, 1356.

Paul, J. (1972) Nature 238, 444.

Paul, J. and Gilmour, R.S. (1968) J. Mol. Biol. 34, 305.

Randerath, J. and Randerath, E. (1967) In Methods of Enzymology, vol 12a, p 323. Ed. Grossman, L. and Moldave, K. New York and London: Acadmic Press.

Rice, N. (1972) Evolution in Genetic Systems , BrookhavenSymposium, No. 23, P. 44 (New York: Gordon and Breach).

Riley, M. Maling, B. and Chamberlin, M.J. (1966) J. Mol. Biol. 20, 359.

Ritossa, F.M. (1972) Nature New Biology, 240, 109.

Sanger, F. and Brownlee, G.G. (1967) In Methods of Enzymology, vol 12A, p. 361 Ed. Grossman, L. and Moldave, K. New York and London: Academic Press.

Schildkraut, C.L. and Maio, J.J. (1968) Biochem. Biophys. Act.161, 76.

Schildkraut, C.L., Marmur, J. and Doty,P. (1962) J. Mol. Biol. 4, 430.

Shearer, R.W. and McCarthy, B.J. (1967) Biochem. 6, 283.

Skinner, D.M., Beattie, W.G., Kerr, H.S. and Graham, D.E. (1970) Nature, 227, 837.

Smith, G.P. (1973) Cold Spring Harbor Symposium on Quantitative Biology, 38

Southern, E.M. (1970) Nature, 227, 794.

Southern, E.M. (1971) Nature New Biology, 232, 82.

Southern, E.M. (1973). In Modern Aspects of Cytogenetics: Consitutive Heterochromatin in Man. Symposia Medica Hoechst 6. F.K. Schattauer Verlag. Stuttgart - New York.

Southern, E.M. (1974a) In preparation.

Southern, E.M. (1974b) In preparation.

Southern, E.M. and Mitchell, A.R. (1971) Biochem. J. 123, 613.

Southern, E.M. and Roizes, G.P. (1973) Cold Spring Harbor Symposia on Quantitative Biology , 38.

Sueoka, N. (1961) J. Mol. Biol., 3, 31.

Sutton, W.D. (1971) Bioch im. Biophys. Act. 240, 522.

Sutton, W.D. (1972) Nature New Biology, 237, 70.

Sutton, W.D. and McCallum, M. (1971) Nature New Biology, 232, 82.

Sutton, W.D. and McCallum, M. (1972) J. Mol. Biol. 71, 633.

Szala, S. and Southern, E.M. (1974) In preparation.

Szybalski, W. (1961) In the Molecular Basis of Neoplasia, p.164. M D Anderson Hospital and Tumor Institute, University of Texas Press, Austin, Texas.*

Szybalski, W. (1968), Fractions, I 1. Spinco Division of Beckman Instruments Inc.

Thomas C.A. Jr. (1970) In The Neurosciences: A Study Program, Schmitt, F.O. Ed. Rockefeller University Press, New York.

Thomas, C.A. Jr., Hamakalo, B.A., Misra, D.N. and Lee , C.S. (1970) J. Mol. Biol. 51, 621.

Tiseluis, A., Hjerten, S. and Levin, O. (1956). Arch, Biochem. Biophys., 65, 132.

Tobia, A, Schildkraut, C.L. and Maio, J.J. (1970) J. Mol. Biol., 54, 499.

Ullman, J.S. (1970), Ph.D Thesis, University of Washington, Seattle.**

Walker, P.M.B. (1971). Prog. Biophys. and Mol. Biol., 23, 145.

Walker, P.M.B., Flamm, W.G. and McLaren, A. (1969) In Handbook of Molecular Cytology. Lima-de-Faria, A, Ed., Amsterdam and London, North-Holland Publishing Co., p.52.

Walker, P.M.B. and McLaren, A. (1965) Nature, 208, 1175.

Wang, A.C. and Kallenback, N.R. (1971) J. Mol. Biol., 62, 591.

Warring, M. and Britten, R.J. (1966) Science, 154, 791.

Wells, R.D. Ohtsuka, E. and Khorama, H.G. (1965) J. Mol. Biol., 14, 221.

Wetmur, J.H. and Davidson, N. (1968) J. Mol. Biol., 31, 349.

Whitehouse, H.L.K. (1967) J. Cell Science, 2, 9.

Yasmineh, W.G. and Yunis, J.J. (1969) Biochem. Biophys. Res. Comm., 35, 779.

Yasmineh, W.G. and Yunis, J.J. (1970a) Exp. Cell. Res., 59, 69.

Yunis, J.J. and Yasmineh, W.G. (1970b) Science, 168, 263.

Yasmineh, W.G. and Yunis, J.J. (1971) Exp. Cell. Res., 64, 41.

* Quoted in Walker (1971)

** Quoted in Bonner (1973).

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor P.M.B. Walker and Dr. E.M. Southern for their guidance and advice during the course of this study. I also wish to thank all the members of the M.R.C. Mammalian Genome Unit.