To the Graduate Council:

I am submitting herewith a dissertation written by Nelwyn Tullos Christie entitled "Analyses of Highly Repetitive DNAs in the Genome of the Red Crab Geryon quinquedens." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Dorothy M. Skinner

Dorothy M. Skinner, Major Professor

We have read this dissertation and recommend its acceptance:

W. S. Riggsby

Robert K Tajemura

E. H. Grell

Accepted for the Council:

L. Evans Roth

Vice Chancellor
Graduate Studies and Research

Thesis
79b
.C375
cop.2

ANALYSES OF HIGHLY REPETITIVE DNAs IN THE GENOME OF THE RED CRAB GERYON QUINQUEDENS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Nelwyn Tullos Christie

August 1979

1398041

ACKNOWLEDGMENTS

My gratitude for training in the Graduate School of Biomedical Sciences is extended at large to the Biology Division of the Oak Ridge National Laboratory for the excellent research facilities provided. Appreciation is also extended to Oak Ridge Associated Universities for my first research experience as summer research participant at the Biology Division. Individuals who have been instrumental in my scientific development have been numerous. Close associates with whom I have worked include William Gude, Grant Brewen and Julian Preston. It was the interest generated in cytogenetics while working with the latter two which spurred me to pursue further training. Early research experiences in the Graduate School of Biomedical Sciences were in the laboratories of E. H. Grell and Shankar Mitra. From each I gained a different perspective of genetics, my chosen field. My appreciation to my husband and two daughters is especially keen for their constant support and understanding of the demands of my endeavor. Bob Fujimura and Stuart Riggsby are to be commended for their searching questions and critical reviews of my manuscripts. Appreciation is extended to Nette Crowe for her cooperation with the art work and for her delightful wit. Special thanks is due Carrie Chambers and Maria Schell from our laboratory for their scientific expertise and endearing friendship. Many of the initial struggles and joys of my research were shared with my friend and fellow graduate student, Christie Holland. I thank my advisor, Dorothy Skinner, for her wise council, her vision in achieving research goals, and most of all, for her constant expectation of continual improvement.

ABSTRACT

Previous studies on the genome organization of brachyurans (true crabs) have demonstrated the presence of large amounts of highly repetitive DNA either by density separations or reassociation techniques. Demonstration of highly repetitive DNA interspersed with single copy DNA distinguishes the genome of the red crab, Geryon quinquedens, from other eukaryotes which have been examined in this way. This interspersion feature has not been investigated for other brachyurans. The general organization of the genome of the red crab is the short period interspersion pattern typical of organisms in a wide range of taxonomic groups. The experiments in Part I define this general pattern and also show that fractions of both repetitive components, present at 4×10^6 and 8×10^5 copies per genome, are interspersed with single copy DNA. After isolation of these kinetic fractions and digestion with HindIII restriction endonuclease, several fragments of discrete size are produced. In Part II these fragments are shown to be elements of a multimeric series based on an 81 nucleotide (nt) fragment. The relative amounts of each of the multimers suggest that random mutation alone cannot explain a distribution of restriction sites featuring enhanced amounts of dimer, tetramer and octamer. Restriction endonuclease analyses of the octamer indicate the presence of large amounts of a divergent octamer. The pathway of evolutionary change which emerges is amplification of DNA with an 81 nt repeat unit followed by divergence followed by a second amplification of only a fraction of the original repetitive sequences.

In Part III the sequences of the basic repeating unit in the HindIII multimeric series are presented. Their characteristics are discussed relative to those of highly repetitive DNAs in other organisms. Indications of nonrandom sequence alterations in the multimers support the idea that divergent sequences have been amplified. Also present are examples of random evolutionary changes. In addition, divergence estimates based on HhaI digestion have been performed for individual multimers as a comparison to the one obtained from HindIII digestion of total DNA (Part II). The low level of divergence (4%) due to random mutation indicated from quantitation of the products of HhaI digestion is incompatible with the large amounts of dimer, tetramer and octamer produced by HindIII digestion being formed by random mutation alone. Either selective mutation or selective amplification must be invoked to explain these results. The large number of constraints required for selective mutation to have produced the observed distribution of HindIII multimers makes a model of selective amplification more plausible.

TABLE OF CONTENTS

CHAPTER	PAGE
INTRODUCTION	1
PART I. THE ORGANIZATION OF THE GENOME OF THE RED CRAB, GERYON QUINQUEDENS	
I. INTRODUCTION	7
II. MATERIALS AND METHODS	9
Purification of DNA and Separation of Satellites	9
Preparation of Labeled DNA	9
DNA Fragmentation and Sizing	10
Reassociation of DNA Fragments	10
Hyperchromicity Measurements	11
Measurements of Duplex Length of Reassociated Molecules	11
Restriction Endonuclease Analyses	12
III. RESULTS	13
Reassociation Kinetics of 420 Nucleotide Fragments	13
Interspersion Studies by Hyperchromicity Measurements and S1 Digestion	14
Estimate of Amounts and Lengths of Repetitive Components	15
Organization of Repetitive DNA Classes and Single Copy DNA	16
Further Characterization of the Repetitive DNA	19
IV. DISCUSSION	22
LIST OF REFERENCES	26
APPENDICES	30
A. TABLES	31
B. FIGURE LEGENDS AND FIGURES	36
C. EXTENDED PROCEDURES	50
Labeling DNA by Iodination	51
Actinomycin D/CsCl Gradients	53
Isolation of HindII and HindIII Restriction Enzymes	54
Restriction Endonuclease Digestion of Kinetic Fractions	59

CHAPTER PAGE

PART II. SEQUENCE ORGANIZATION OF A MULTIMERIC SERIES OF RESTRICTION FRAGMENTS FROM TOTAL DNA

I. INTRODUCTION 62

II. MATERIALS AND METHODS 64

Restriction Enzyme Analyses and Isolation of Restriction Fragments 64
Quantification of the DNA in the Restriction Fragments 65
Electrophoretic Size Determination of the Restriction Fragments 65

III. RESULTS 67

Organization of HindIII Sites in Total DNA 67
Quantification of HindIII Fragments in Total DNA 68
Analyses of the Multimers 70

IV. DISCUSSION 80

LIST OF REFERENCES 84

APPENDICES 87

A. TABLES 88

B. FIGURE LEGENDS AND FIGURES 92

C. EXTENDED PROCEDURES 112

Preparative Isolation of Geryon HindIII Fragments 113
Isolation of ØX174-am^3 RF 116
Determination of Error in Sizing Method on Acrylamide Gels 121

PART III. SEQUENCE CHARACTERIZATION OF THE BASIC REPEATING UNIT OF A MULTIMERIC SERIES OF RESTRICTION FRAGMENTS

I. INTRODUCTION 123

II. MATERIALS AND METHODS 126

Preparation of DNAs 126
Electrophoretic Sizing of Restriction Fragments 126

CHAPTER PAGE

DNA Sequencing 127
Quantitation of Relative Amounts of Restriction Products 127

III. RESULTS 128

Sequences of the Basic Repeat Unit 128
Organization of the Multimers 129
Sequence Variants Indicating Amplification of Divergent Sequences 131
Sequence Variants Consistent with Random Mutation 133
Estimate of Divergence in the Multimers 136

IV. DISCUSSION 140

LIST OF REFERENCES 145

APPENDIX 148

VITA 161

LIST OF TABLES

TABLE PAGE

PART I

1. Kinetic Analysis of Geryon DNA Reassociation 32

2. Duplex Content of DNA Fragments Reassociated to Cot 2.0 33

3. Fraction of Fragments Containing Repetitive Sequences 34

4. Estimates of the Amount of Each Component in the Genome 35

PART II

1. Estimate of the Amount of HindIII Fragments in the Genome 89

2. Variations in Estimates of Fragment Size in 7% Polyacrylamide Gels 90

3. Variations in Estimates of Fragment Size in 5% Polyacrylamide Gels 91

LIST OF FIGURES

FIGURE PAGE

PART I

1. Reassociation Kinetics of Geryon DNA 37
2. Thermal Dissociation of Reassociated DNA 39
3. Size Distribution of S1 Resistant Sequences of DNA 40
4. Fractions of Geryon DNA Containing Repetitive Elements as a Function of Fragment Length 42
5. Analysis of Single Copy DNA Adjacent to the Very Highly Repetitive Component 43
6. Model E Analytical Ultracentrifuge Tracings of Total DNA 45
7. Restriction Endonuclease Digestion of Three Kinetic Fractions and Native DNA 46
8. Possible Arrangements for Sequences of the Very Highly Repetitive Component 48

PART II

1. Gel Electrophoresis of Restriction Endonuclease Digests of Total Geryon DNA 93
2. HindIII Digest of Total Geryon DNA 94
3. Test for Random Inactivation of HindIII Sites in Total DNA 97
4. Autoradiogram of HhaI Digestion Products of Monomer, Dimer, Trimer, and Tetramer DNA 99
5. Proposed Organization for the Original Spacings of Restriction Endonuclease Sites in the Geryon Sequences Isolated by Digestion with HindIII 100
6. Autoradiogram of HaeIII Digestion Products of Monomer, Dimer, Trimer, and Tetramer DNA 102
7. Autoradiogram of HhaI Digestion Products of Pentamer, Hexamer, and Heptamer DNA 103

FIGURE PAGE

8. Restriction Endonuclease Analysis of Octamer and Decamer DNA 105

9. Proposed Organization of Subset 2 Fragments Obtained from HhaI Cleavage of Octamers 108

10. Autoradiogram of Restriction Endonuclease Analyses of Decamer DNA 110

PART III

1. The Predominate Nucleotide Sequence of the Monomer of the HindIII Multimeric Series 149

2. Organization of Sequences in the HindIII Dimer 151

3. Representative Autoradiograms of DNA Sequencing Gels 153

4. Autoradiogram of Digestion Products of the HindIII Monomer and Dimer 156

5. Distribution of HhaI Sites in HindIII Multimers 158

6. Distribution of HhaI Sites in the HindIII Octamer 160

INTRODUCTION

An initial impetus in studies of genome organization was an understanding of the regulation of gene expression in eukaryotes. Early observations suggested that a general pattern of organization was common to most eukaryotes (Davidson et al., 1973, and Graham et al., 1974, Part I; Schmid and Deininger, 1975;[1] Flavell et al., 1974[2]). This pattern of "short period" interspersion consisted of repetitive DNA of approximately 300 nt (nucleotides) adjacent to single copy DNA of approximately 1000 nt. Since the single copy length was close to the average gene length, the short repetitive regions were proposed to be regulatory elements directing the time of gene expression during development of the organism (Davidson and Britten, 1973[3]). An exception to this pattern was Drosophila melanogaster in which an average of 5000 nt of repetitive DNA was juxtaposed to approximately 10,000 nt of single copy DNA (Manning et al., 1975, Part I). Although the next report of "long period" interspersion was also found in the Class Insecta (Crain et al., 1976, Part I) later studies described this pattern in the water mold (Hudspeth, 1977, Part I), the Syrian hamster (Moyzis, 1977, Part I) and the chicken (Arthur and Straus, 1978, and Eden and Hendrick, 1978, Part I).

---

[1] C. W. Schmid and P. L. Deininger, Cell, 6 (1975), 345-358.

[2] R. B. Flavell, M. D. Bennett, J. B. Smith, and D. B. Smith, Biochemical Genetics, 12 (1974), 257-268.

[3] E. H. Davidson and R. J. Britten, Quarterly Review of Biology, 48 (1973), 565-613.

Identification of long period interspersion patterns in diverse taxonomic groups made it difficult to correlate an overall pattern of short period interspersion as a necessary requirement for gene regulation.

There had always been a problem with correlating mechanisms of gene regulation with the overall organization of the DNA. This was the large size of eukaryotic genomes, ranging from 10^7 to 10^{10} base pairs (Lewin, 1974[4]). Estimates of the number of genes required for growth and development of the organism were between 5×10^3 and 10^5, representing no more than 5 to 15% of the genome. The number of genes was approximated in two ways. First, from the tolerable mutation loads which would allow production of a reasonable number of viable progeny, Ohta and Kimura (1971)[5] obtained an estimate of 5 to 15% of the genome as essential for development. Secondly, from measurements of the number of mRNA molecules at different times in development and the cross-reactivity of mRNA from different stages, similar percentages were obtained (Turner and Laird, 1973,[6] Galau et al., 1976[7]). Even when the three classes of DNA in eukaryotes are considered (highly repetitive, moderately repetitive, and

---

[4]B. Lewin, Gene Expression, 2 (London, New York, Sydney, Toronto: John Wiley and Sons, 1974), 151.

[5]T. Ohta and M. Kimura, Nature, 233 (1971), 118-119.

[6]S. H. Turner and C. D. Laird, Biochemical Genetics, 10 (1973), 263-274.

[7]G. A. Galau, W. H. Klein, M. M. Davis, R. J. Britten, and E. H. Davidson, Cell, 7 (1976), 487-505.

single copy; Britten and Kohne, 1968[8]), the amount of single copy DNA is typically in excess of 15%. The picture which emerges of the eukaryotic genome is one in which 85 to 95% of the DNA has no obvious reason for being there.

Within the large, presumably non-coding fraction, attention has been focused on repetitive DNAs because they are easily isolated and can be localized on the chromosomes. One of the best clues to their function is their localization to heterochromatic regions characterized by a high degree of condensation compared to euchromatin where most of the genes reside. Heterochromatin appears to be a highly variable part of the chromosome and has been implicated in speciation. Many if not most of the karyotypic differences between closely related species can be related to heterochromatic changes (White, 1969, Part II; Lejeune et al., 1973;[9] Arnason et al., 1978[10]). Similar changes associated with speciation occur in satellite DNAs which have been localized to heterochromatin (Holmquist, 1975, and Gall and Atherton, 1974, Part III; Arnason et al., 1978[10]). The relationship between satellites and phylogeny is not always clear. Gosden and coworkers[11] have shown that species considered

---

[8]R. J. Britten and D. E. Kohne, Science, 161 (1968), 529-540.

[9]J. Lejeune, B. Dutrillaux, M. Rothore, and M. Prieur, Chromosoma, 43 (1973), 423-444.

[10]U. Arnason, I. F. Purdom, and K. W. Jones, Chromosoma, 66 (1978), 141-159.

[11]J. R. Gosden, A. R. Mitchell, H. N. Seuanez, and C. A. Gosden, Chromosoma, 63 (1977), 253-271.

distantly related by morphological criteria share satellites not present in intervening species. It must therefore be concluded that satellite alterations alone are not preemptive for speciation but may predispose an organism toward genetic isolation. An appropriate genetic change must coincide with suitable environmental forces to produce an effective speciation event.

The persistence of highly repetitive DNAs throughout long periods of evolution argue for selective forces maintaining them or perpetuating them. The selective force for satellite DNAs may not be applied to a particular sequence but rather to a general class of DNA. Certain sequences may appear to be maintained by periodic amplifications within this general class which continually perpetuate them in newly emerging species. Those taxonomic groups in which amplification is a frequently recurring process would best simulate conservation of a particular sequence. A plausible rationale for the survival of this class of DNA is that it promotes genetic variations by facilitating chromosomal rearrangements and producing heterochromatic polymorphisms. Production of a large number of diverse species will increase the chances of survival of any taxonomic group and simultaneously that of its highly repetitive DNA. A feature of the eukaryotic genome which increases the potential for genetic variation might require no other reason to be maintained throughout evolution.

As particular genomes are examined in great detail, these observations may be analagous to looking at a cyclic process of DNA sequence evolution which has fortuitously stopped at a particular place.

Amplification of highly repetitive DNAs may occur from a sequence in either single copy DNA or middle repetitive DNA followed by divergence to middle repetitive DNA and further alteration to single copy DNA. Chromosomal deletions or rearrangements yielding acentric fragments can produce losses from the genome. Superimposed on cycles of amplification and divergence, some chromosomal rearrangements may produce dispersal of new sequences to various chromosomes within the karyotype. The purpose of the research reported here has been a detailed analysis of the organization of the DNA in a single species, the red crab, Geryon quinquedens. This organism was chosen for extensive study because of its large amounts of highly repetitive DNA although lacking satellite DNAs isolable in neutral CsCl. In Geryon DNA an exception to the typical arrangement of highly repetitive DNAs in long continuous arrays is demonstrated by the interspersion of a fraction of this class with the single copy DNA in lengths of approximately 300 nt. Restriction enzyme analyses and DNA sequencing have been used to describe a portion of the highly repetitive DNA in this organism. These studies indicate plausible pathways of evolution for this class of DNA.

PART I

THE ORGANIZATION OF THE GENOME OF THE RED CRAB, *GERYON QUINQUEDENS*

CHAPTER I

INTRODUCTION

Although the interspersion of single copy with repeated DNA is a general pattern of organization for eukaryotic genomes, there is considerable variability in the length of the two classes of sequences. As originally described for Xenopus laevis (Davidson et al., 1973), most organisms have 70-80% of their single copy DNA in lengths of 800-2000 nucleotides (nt) followed by repetitive sequences 300 nt long (short period interspersion). A distinctly different pattern (long period interspersion) is characterized by more than ten-fold greater lengths of both repetitive and single copy sequences. Until recently the long period interspersion pattern was thought to be an anomaly found in the fruit fly (Manning et al., 1976) and the honeybee (Crain et al., 1976) but not in all insects (Crain et al., 1976; Efstratiadis et al., 1976). It has now been demonstrated in the genomes of the water mold (Hudspeth et al., 1977), the Syrian hamster (Moyzis et al., 1978) and the chicken (Arthur and Straus, 1978; Eden, 1978).

It seemed appropriate to examine individual differences and similarities of genome organization within a particular taxon. For example, the genomes of the true crabs (brachyurans) Libinia emarginata, Cancer borealis, and Gecarcinus lateralis, are all similar in their scarcity of sequences of slow and intermediate repetitive frequencies, (i.e., those that reassociate 10 to 1000 times the rate of single copy DNA; Vaughn, 1975; Holland and Skinner, 1977). Highly repetitive

satellite DNAs account for 11 to 30% of these genomes. For comparison, we have studied the organization of the genome of Geryon, a primitive brachyuran which lacks patent satellites but contains large amounts of highly repetitive DNA.

CHAPTER II

MATERIALS AND METHODS

Purification of DNA and Separation of Satellites. Frozen testes were purchased from the Supply Department of the Marine Biological Laboratory. DNA was isolated by the MUP procedure (Britten et al., 1974) and stored in 10^{-3} M EDTA and 10^{-4} M NaCl. Heavy metal gradients, analytical and preparative, were prepared according to Skinner and Beattie (1973). CsCl gradients containing actinomycin D were prepared as described by Kersten et al. (1966) by the procedure given in Appendix C. Actinomycin D was dissolved in water immediately before use. The density of the CsCl gradients were adjusted to approximately 1.7 g/ml; the density of Cs_2SO_4 gradients, to 1.5 g/ml. Centrifugation was carried out in a Spinco Model E analytical ultracentrifuge for 20 hours at 44,770 rpm, 25°C.

Preparation of Labeled DNA. Tritium-labeled DNA was prepared in vitro by repair synthesis using DNA polymerase I of *E. coli* (Englund, 1971; Holland and Skinner, 1977). The modal single-stranded size of the labeled product ranged from 420 to 560 nt. Discrete size classes were selected from isokinetic alkaline sucrose gradients (Noll, 1969). Since DNAs labeled with DNA polymerase I exhibit abnormally high zero-time binding, all tritium labeled preparations were denatured and rapidly passed through hydroxyapatite at Cot < 10^{-4} before use in reassociation experiments.

Cytosine residues in sheared and unsheared DNA were iodinated according to Commerford (1971) with several modifications (Getz, 1972;

Orosz and Wetmur, 1974; Appendix C). Immediately after denaturation in an ethylene glycol bath (5 min, 104°) 25 µg of DNA was added to a reaction mixture of $Na^{125}I$ and thallium chloride (2×10^{-4} M in 0.5 M sodium acetate buffer). Iodination was carried out at 60° for 5 min, conditions which yield an equivalent Cot of 0.05. For some experiments native DNA was iodinated. The iodinated DNA was dialyzed against 0.4 M NaCl, 15 mM phosphate buffer (pH 6.0), and 0.2 mM EDTA and then placed at 90°, 15 min, to remove the unstably bound iodine. Specific activities ranged from 3×10^4 - 1.5×10^5 cpm/µg for denatured DNA and 1-2 $\times 10^4$ cpm/µg for native DNA. Iodinated DNAs were used within two weeks after iodination. All iodinated molecules used in reassociation experiments were selected from preparative alkaline sucrose gradients.

DNA Fragmentation and Sizing. DNA fragments of various sizes were obtained by shearing in a Virtis 60 homogenizer (Davidson et al., 1973; Holland and Skinner, 1977). Unlabeled molecules were sized by velocity sedimentation in 0.9 M NaCl + 0.1 M NaCl or in 1 M NaCl (Markham, 1967). Labeled molecules were sized in alkaline sucrose gradients (Noll, 1969). Fragment sizes given were calculated according to the equations of Studier (1965) and represent an average of 2 independent determinations.

Reassociation of DNA Fragments. DNA samples in 0.12 M phosphate buffer (PB), pH 6.8, were denatured at 104° for 5 minutes in an ethylene glycol bath, placed at 60°C for reassociation and analyzed by hydroxyapatite chromatography (Britten et al., 1974). The reassociation of DNA samples in 0.4 M PB at 64°C allowed a reduction in time of incubation while

maintaining criteria equivalent to reassociation in 0.12 M PB. Unlabeled driver DNA was usually added to tracer DNA in a hundred fold excess except for very low Cots when tracer DNA alone was used.

Hyperchromicity Measurements. Thermal denaturation profiles were obtained by melting DNA samples in 0.12 M PB in a Beckman Acta III C with an attached Haake bath programmed for linear temperature increase of 45°C per hour. The absorbance at 260 nm was monitored for 1-3 samples. A standard DNA, the main component DNA of Gecarcinus lateralis, was included with every cycle. The T_m was taken as the temperature at which 50% of the hyperchromicity was achieved. Hyperchromicity was calculated as the fraction of total absorbance at 100° which was achieved between 55°C and 100°C $\{H = A_{260}\ (100^{\circ}C) - A_{260}\ (55^{\circ})/A_{260}\ (100^{\circ})\}$.

Measurements of Duplex Length of Reassociated Molecules. The fraction of base pairs in duplex was measured for reassociated samples by digestion with S1 nuclease, specific for single strands (Vogt, 1973). Samples were reassociated in 0.3 M NaCl and 0.01 M Pipes buffer at 64°, diluted by 1/2 with 0.2 mM zinc acetate, pH 4.5, before digestion with S1 nuclease. Mercaptoethanol was also added to final concentration of 2 mM. The digestion was carried out at 37° for 45 min. These conditions remove ~ 98% of the single-stranded ends while leaving duplex regions intact (Britten et al., 1977; Holland and Skinner, 1977). The extent of digestion was held constant by maintaining a constant ratio of enzyme to DNA. This required the addition of unlabeled E. coli DNA to incubation mixtures reassociated to Cot 10^{-5} and 2×10^{-3}.

After S1 digestion, reassociated duplexes collected from hydroxyapatite were analyzed by chromatography on Biogel agarose $A_{50}m$ (Bio-Rad, 100-200 mesh). The gel was poured over a support of 5 and 6 mm glass beads to 2.3 × 64 cm dimensions.

Restriction Endonuclease Analyses. AluI was isolated according to H. Heynecker by a procedure similar to that of Roberts et al. (1976). HindIII was isolated according to Smith (1974) by the procedure given in Appendix C. Restriction endonuclease digestions were carried out as described (Roberts et al., 1976). Before digestion, samples of reassociated DNA were adjusted to 0.6% formaldehyde to prevent further reassociation (Feldman, 1973; Appendix C). Dilute samples were concentrated by adsorption to hydroxyapatite followed by ethanol precipitation. For digestion of kinetic fractions the amount of DNA in duplex was calculated from the S1 resistance of similarly isolated fractions. Restriction fragments were sized by electrophoresis on polyacrylamide gels (3.5-5%) as described (Maniatis et al., 1975; Chambers et al., 1978). DNA fragments less than 200 nt long were sized by co-electrophoresis with ØX174 DNA fragments of known sizes obtained by AluI digestion of specific HindII + III fragments (Maniatis et al., 1975).

CHAPTER III

RESULTS

Reassociation Kinetics of 420 Nucleotide Fragments. Reassociation kinetics of 420 nt fragments (Figure 1, Appendix B) showed that Geryon DNA contains two general classes of sequences, highly repetitive (60%) and presumptive single copy (35%); an additional 5% of the DNA was unreactive at Cot 4×10^4. The solid curve in Figure 1 (Appendix B) describes a least-squares analysis of the data fit by computer for three second-order components which subdivides the highly repetitive DNA into two fractions (Britten et al., 1974). Table 1 (Appendix A) lists the characteristics of these components. By these analyses very highly and highly repeated sequences comprise 22% and 34% of the DNA respectively; approximately 4% of the DNA bound to hydroxyapatite at Cot 10^{-5} ("zero-time binding DNA"). In other eukaryotic genomes, foldback sequences (palindromes) have been a major component of this fraction (Wilson and Thomas, 1974).

The stated reiteration frequencies in Table 1 (Appendix A) are average values that may represent a range of repetition frequencies. They were calculated as the ratios of the rates of reassociation for the repetitive components and the assumed single copy component as determined in unfractionated DNA (Britten et al., 1974; Angerer et al., 1974). Despite some uncertainty in the calculated frequencies, these values permit a better comparison with components in other genomes than do the rate constants alone. The Cot curve shares with those of other true crabs a most striking feature: the seemingly complete absence of middle repetitive

sequences that reassociate between Cots 2 and 100 (Vaughn, 1975; Holland and Skinner, 1977).

Interspersion Studies by Hyperchromicity Measurements and S1 Digestion. If repetitive sequences are interspersed with single copy DNA, reassociation of fragments longer than the length of the repetitive sequences to Cot values at which only repetitive sequences reassociate yields duplexes with single-stranded tails (Graham et al., 1974). The fraction of paired bases in reassociated DNA can be determined from hyperchromicity measurements or by digestion of single-stranded ends by S1 nuclease (Graham et al., 1974; Angerer et al., 1974). To this end, different sized fragments were incubated to Cot 2 to permit reassociation of the repetitive sequences only. The hyperchromicity of reassociated 250, 420 and 1700 nt fragments respectively was 88, 63 and 53% that of native DNA (Figure 2, Appendix B). The greatly reduced hyperchromicity of the longer fragments is consistent with the interspersion of short repetitive sequences and single copy DNA.

The percent of the genome estimated as repetitive by collecting DNA reassociated to Cot 2 on hydroxyapatite includes single copy sequences attached to reassociated duplexes. The duplex content of reassociated 420 nt fragments obtained from melting profiles allowed a correction of the estimate for the amount of repetitive DNA derived from kinetic measurements (Table 2, Appendix A). Although after reassociation to Cot 2, 51% of the 420 nt fragments adsorbed to hydroxyapatite only 59% of the bases in the bound DNA were paired; therefore, only 30% of the

genome (0.51×0.59) is repetitive. Similar calculations for 250 and 1700 nt fragments gave values of approximately 39% repetitive DNA.

Estimate of Amounts and Lengths of Repetitive Components. A more accurate measure of the amount of DNA in each repetitive component was obtained by digesting single-stranded ends of reassociated fragments with S1 nuclease. Following isolation on hydroxyapatite, the S1 resistant duplexes were sized by chromatography on agarose $A_{50}m$. The size distribution for zero-time binding DNA in 2000 nt fragments was determined after reassociation to Cot 10^{-5} (Figure 3A, Appendix B). Although these sequences ranged in size from less than 300 to 1200 nt, the majority were approximately 300 nt. The DNA reassociating between Cot 10^{-5} and 2×10^{-3} (9% of the genome) is composed of zero-time and very highly repetitive sequences. The size distribution for the latter was corrected for the contribution of zero-time binding DNA (5% of the genome) to obtain the distribution of only the very highly repeated sequences (lower curve, Figure 3B, Appendix B). Approximately 80% of the very highly repeated sequences ranged in size from 300 to 500 nt. By correcting for the amount of DNA in duplex at Cot 2×10^{-3} recovered in the Cot 2 fraction, it was found that 29% of the genome was comprised of highly repetitive DNA (lower curve, Figure 3C, Appendix B). Approximately 60% of that component (17% of the genome) was arranged in repetitive clusters longer than 1200 nt. The remainder of the fraction was estimated as stretches of 300 nt interspersed with single copy DNA.

Although the data on ^{125}I-iodinated DNA included in the Cot curve (Figure 1, Appendix B) were superimposable on those from either Pol I

labeled or unlabeled DNA, a report of destabilization of DNA duplexes by the presence of 5-iodocytosine (Anderson and Folk, 1976) prompted us to determine the susceptibility to S1 nuclease of reassociated repeated regions in iodinated Geryon DNA. As a control unlabeled 2000 nt fragments were reassociated to Cot 2 and treated with S1 nuclease. The size distribution of the unlabeled duplexes (data not shown) was nearly identical to that of iodinated duplexes isolated in the same manner (Figure 3C, Appendix B). We conclude that the reassociation kinetics of iodinated DNA did not yield an underestimate of the lengths of the repetitive sequences or an overestimate of the fraction of interspersed repetitive sequences.

Organization of Repetitive DNA Classes and Single Copy DNA. A measure of the spacing of repetitive sequences is obtained by reassociating different sized fragments to the same Cot in the presence of an excess of short fragments (250-400 nt) as driver DNA (Davidson et al., 1973; Graham et al., 1974). For the Geryon genome the spacing of the sequences of both repetitive components was measured simultaneously by reassociation to Cot 2 (Figure 4, Appendix B; Table 3, Appendix A). Corrections were made for zero-time binding DNA and adjacent sequences as for other genomes in which such sequences are interspersed with both single copy and repetitive sequences (Davidson et al., 1973; Angerer et al., 1974). When 2000 nt long fragments were reassociated to Cot 2 in the presence of a 100-fold excess of short unlabeled fragments, 89% bound to hydroxyapatite. This result indicates that most of the genome is organized in a short period interspersion pattern with repetitive sequences

interspersed every 800-1000 nt as in many other eukaryotes (Walbot and Dure, 1976; Efstratiadis et al., 1976; Holland and Skinner, 1977). No more than 11% of the genome contains stretches of single copy DNA longer than 800-1000 nt. Interspersion measurements also provided an independent measure of the amount of repetitive DNA. The Y-intercept of the line for the amount of DNA bound when the fragments are less than 1000 nt is an estimate of the amount of duplex DNA which reassociates by the specified Cot (Graham et al., 1974). Since essentially all the repetitive DNA reassociates by Cot 2, the amount of repetitive DNA is 35% excluding zero-time binding DNA.

The interspersion of the very highly repetitive component alone was measured with 300 nt fragments of driver DNA at Cot 2×10^{-3} by which less than 5% of the highly repeated component would have reassociated. A significant amount of the very highly repeated fraction was interspersed with other sequences in the genome. Approximately 30% of the genome is arranged as very highly repeated sequences interspersed with either highly repeated or single copy sequences averaging 1000 nt (Figure 4, Appendix B; Table 3, Appendix A). The negligible increase in hydroxyapatite binding of fragments longer than 1000 nt indicates that longer spacings between these sequences are uncommon. This interspersion pattern for the very highly repeated sequences was confirmed by determining the fraction of DNA bound to hydroxyapatite at Cot 2×10^{-3} for 100, 800 and 1700 nt fragments that had been iodinated without denaturation. The data points fell on the same curve derived from similarly sized fragments denatured before iodination, indicating that reassociation during the iodination reaction did not cause significant underiodination of very

highly repeated sequences. The sequences adjacent to the very highly repeated component could include repeated sequences that reassociated after Cot 2 × 10^{-3}, single copy sequences, or both. The Y-intercept of the line for the amount of DNA fragments bound at Cot 2 × 10^{-3} for lengths less than 1000 nt corroborates the S1 digestion data in Section 3. Taken together, they indicate that only 4% of the genome is comprised of sequences of the very highly repeated component rather than 22% as calculated from kinetic analyses.

By contrast the interspersion of the very highly repeated fraction with single copy DNA was demonstrated by recovering from hydroxyapatite DNA that had been reassociated to Cot 2 × 10^{-3}. This fraction which contains zero-time binding and very highly repetitive DNAs and their adjacent sequences, was further reassociated to Cot 2 and digested with S1 nuclease. If the very highly repeated component is interspersed with single copy DNA, the fraction recovered at Cot 2 will contain S1 sensitive DNA. Subtracting from the total S1 sensitive DNA (33%; Figure 5B, Appendix B) in the isolated fraction, the 15% S1 sensitive DNA adjacent to zero-time sequences (Figure 5A, Appendix B), we obtained an estimate for the amount of single copy DNA adjacent to the very highly repeated component. Eighteen percent is a minimum estimate since some of the DNA which abuts zero-time binding DNA may be highly repetitive. Verification of this value was obtained by comparing estimates of the amount of repetitive DNA from the two types of experiments (Direct and Fractionation, Figure 5, Appendix B). After incubation of 2000 nt fragments directly to Cot 2, 38% of the DNA was in duplex (Figure 5A, Appendix B), after the two-step reassociation experiment (Figure 5B, Appendix B), 34% (20%+14%)

had formed duplexes. Despite the 4% difference in the two estimates of the amount of repetitive DNA, the data clearly show interspersion of the very highly repetitive component with single copy DNA.

Further Characterization of the Repetitive DNA:

a. Density gradient centrifugation in the presence of ligands. Although 17% of the Geryon genome is arranged as long repetitive sequences (Figure 3C, Appendix B), the total DNA formed a single band (1.700 g/cm^3) in neutral CsCl gradients (Figure 6A, Appendix B). Three minor satellites were observed on centrifugation of the DNA in the presence of either Hg^{2+} or Ag^{+}/Cs_2SO_4 gradients (Figure 6B, C; Appendix B). Two satellites comprising 3-5% of the genome were detected in the presence of Hg^{2+} while a single satellite comprising 1-2% of the genome was seen in the presence of Ag^{+}. No additional satellites were detected with actinomycin D at a ratio as high as 1:1. In sum, the cryptic satellites could account for less than one third of the long repetitive sequences detected by the kinetic analyses.

While these results do not exclude the presence of long tandem arrays of short repeating sequences, they do indicate a greater probability that the majority of the sequences in the long repetitive class have a longer basic repeat. The probability of a highly repetitive DNA being separated from the bulk of the DNA in density gradients is greater for a short repeating sequence in which an overall bias in base composition is more likely. The longer and more complex the repeat, the greater the probability that the density will be the same as the major component, even in the presence of base-specific ligands.

b. Digestion with restriction endonucleases. Another means of excising and characterising the repetitive sequences in Geryon DNA, all of which are present at frequencies characteristic of satellites, is by digestion with restriction enzymes. Such enzymes have been used to excise highly repetitive DNAs from total DNA of numerous eukaryotic genomes (Botchan, 1974; Cooke, 1975; Gruss and Sauer, 1975; Mowbray et al., 1975; Rosenberg et al., 1978; Manuelidis and Wu, 1978).

Following a series of digestions with numerous restriction enzymes, AluI and HindIII were selected since they yielded fragments of convenient length for subsequent sequencing and their recognition sites were similar. The AluI site is comprised of the four central nucleotides of the HindIII site. Accordingly, the HindIII fragments must include many but not necessarily all of the AluI fragments. Treatment of total Geryon DNA with HindIII (Figure 7A, Lane 4; Appendix B) and AluI (Figure 7B, Lane 4; Appendix B) produced a background of heterogeneously sized fragments with large amounts of fragments of discrete sizes superimposed on the background. Treatment of each of the three kinetic fractions with AluI yielded different amounts of the discrete fragments (Figure 7B, Appendix B). Zero-time binding DNA contained none (Lane 1); the very highly repeated (Lane 2) and highly repeated (Lane 3) fractions contained large amounts of 75 and 150 nt fragments. Because of the lower specificity of the AluI site (i.e., the shorter recognition sequence may be expected to occur at higher frequency) there was a higher background of heterogeneously sized fragments in that digest than in the HindIII digest.

We have begun sequence analyses (Maxam and Gilbert, 1977) of the HindIII fragments because of the lower background. As in highly reiterated human (Manuelidis and Wu, 1978) and simian DNAs (Rosenberg et al., 1978), within each size class there are several populations of molecules since there is more than one residue at certain positions. Nevertheless, there is extensive similarity among the molecules in any one size class. Furthermore, evidence to date indicates considerable homology among the fragments of different sizes. Even from these initial results it is clear that the sequences are not composed of short, simple repeats as found in several crustacean satellite DNAs (Skinner, 1967; Skinner et al., 1974; Skinner, 1977; Chambers et al., 1978). The partial sequence homology and the variation in copy number of closely related sequences may indicate extensive alterations in a previously homogeneous reiteration. These results will be discussed in detail elsewhere.[1]

---

[1]The sizes of the HindIII fragments are more precisely determined in Parts II and III. The 75 and 150 nt fragments correspond to 81 and 162 nt fragments which are the monomer and dimer of a multimeric series.

CHAPTER IV

DISCUSSION

One distinctive characteristic of Geryon genome is the interspersion of single copy DNA with both a highly repeated component and an even more highly repeated component; interspersion of such highly repetitive DNA with single copy DNA has not been observed previously. The sequences of the highly repeated component are subdivided into short stretches of 300 nt interspersed with single copy DNA and longer stretches greater than 1200 nt whose interspersion pattern has not been defined. Although the highly and the very highly repetitive components account for 29 and 4% of the DNA respectively (Table 4, Appendix A), as in other brachyuran genomes, middle repetitive sequences are few in number or absent. Another distinctive characteristic is the 10 to 1000 times greater repetition frequency of the major repetitive component than has been found in other genomes. In most genomes studied to date, the frequencies of the major repetitive classes range from 10 to 5000 copies (Davidson et al., 1973; Graham et al., 1974; Angerer et al., 1975; Efstratiadis et al., 1976) while in Geryon there are as many as 8.3×10^4 copies of the major repetitive component. The repetition frequency of the minor component is similar to that observed for Aplysia DNA (Angerer et al., 1975). In the latter genome, however, the minor component is not interspersed with single copy DNA. The rate constants for reassociation of the two interspersed repetitive components of Geryon are more similar to those of DNAs isolable as satellites (Waring and Britten, 1966; Cordeiro-Stone and

Lee, 1976) than to those for the major interspersed repetitive components in other genomes. In addition, the reassociated repetitive sequences of Geryon exhibit only 4% mismatch (assuming 1% mismatch for every degree lowering of the T_m), a value also characteristic of reassociated patent satellite DNAs (Cordeiro-Stone and Lee, 1976) but less than that of other reassociated repetitive DNAs (Graham et al., 1974; Angerer et al., 1975; Walbot and Dure, 1976). Furthermore, the pairing of the reassociated repetitive sequences in Geryon DNA is sufficiently faithful to allow recognition and cutting by restriction enzymes under standard conditions. Although the restriction fragments could be obtained from long repetitive sequences, there is no a priori reason to assume that they are excluded from the short repetitive sequences.

Because of the extensive interspersion of single copy DNA with repetitive DNA in the Geryon genome the amount of single copy sequences was underestimated by the usual kinetic analyses. From either interspersion measurements or S1 digestion (Table 4, Appendix A) the more accurate value of 57% was obtained. Although the very highly repeated component accounts for only 4% of the genome, it is interspersed with approximately 25% of the genome, including single copy DNA and probably some of the highly repeated sequences. Several possible arrangements of both repetitive components with single copy DNA are shown in Figure 8A and B, Appendix B. A fraction of the highly repeated component is also interspersed only with single copy DNA. Two other proposed arrangements show clustering of repetitive sequences as might occur in cryptic satellites (Figure 8C,D; Appendix B). Short runs (300-500 nt) of the very highly repeated component could be adjacent to sequences of the highly

repeated component which, in turn, could contain internal repeats related but not identical to those of the very highly repeated sequences. From the partial sequence homologies observed for three different HindIII fragments (Part III) this postulated organization might involve the two classes of sequences containing AluI sites, the abundant sequences juxtaposed with those of less abundance. Alternatively, a cryptic satellite might consist primarily of very highly repeated sequences, i.e. only the abundant sequences containing the AluI sites.

Many but not all brachyuran genomes have relatively high percentages of patent satellite DNAs (Sueoka, 1961; Skinner, 1967; Skinner and Kerr, 1971; Beattie and Skinner, 1972). Of the crustacean genomes whose organization has been examined, all have at least 25% highly repeated DNA. In Libinia emarginata satellites constitute 11% of the genome which has a total of 25% highly repetitive DNA; in Cancer borealis satellites account for 30% of the DNA with an additional 10% kinetically defined as highly repetitive sequences (Vaughn, 1975). Approximately 55% of the DNA of Gecarcinus lateralis is in long repetitive sequences, half of which are isolable as satellites (Skinner, 1967; Holland and Skinner, 1977). In the Geryon genome the highly repeated DNA is nearly equally divided between long and short repetitive sequences. If, as suggested (Britten and Davidson, 1971), short repetitive sequences are derived from long repetitive sequences, the interspersion of the very highly repeated sequences with single copy DNA in the Geryon genome may indicate a more rapid transfer of long repetitive sequences to single copy regions as compared to other eukaryotes. While it is likely that the same general

mechanisms of DNA evolution occur in all eukaryotes, it is this abundance of both satellite DNAs and other highly repetitive fractions in crustacean genomes which may aid the study of possible relationships among classes of repeated DNAs.

Studies of the DNAs of distantly related crustaceans, such as shrimps and lobsters, will establish whether the organization of crab genomes is typical of the entire Class Crustacea.

LIST OF REFERENCES

LIST OF REFERENCES

Anderson, D. M., and W. R. Folk. 1976. Biochemistry 15: 1022-1030.

Angerer, R. C., E. H. Davidson, and R. J. Britten. 1975. Cell 6: 29-39.

Arthur, R. R., and N. A. Straus. 1978. Can. J. Biochem. 56: 247-263.

Beattie, W. G., and D. M. Skinner. 1972. Biochim. Biophys. Acta 281: 169-178.

Botchan, M. R. 1974. Nature 251: 288-292. London.

Britten, R. J., and E. H. Davidson. 1971. Quart. Rev. Biol. 46: 111-138.

Britten, R. J., D. E. Graham, F. C. Eden, D. M. Painchard, and E. H. Davidson. 1976. J. Mol. Evol. 9: 1-23.

Britten, R. J., D. E. Graham, and B. R. Neufeld. 1974. L. Grossman and K. Moldave, Eds. Methods in Enzymology 29E: 363-418. Academic Press, New York.

Chambers, C. A., M. P. Schell, and D. M. Skinner. 1978. Cell 13: 97-110.

Commerford, S. L. 1971. Biochemistry 10: 1993-1999.

Cooke, H. J. 1975. J. Mol. Biol. 95: 87-99.

Cordeiro-Stone, M., and C. S. Lee. 1976. J. Mol. Biol. 104: 1-24.

Crain, W. R., E. H. Davidson, and R. J. Britten. 1976. Chromosoma 59: 1-12. Berlin.

Davidson, E. H., B. R. Hough, C. S. Amenson, and R. J. Britten. 1974. J. Mol. Biol. 77: 1-23.

Eden, F. C., and J. P. Hendrick. 1978. Biochemistry 17: 5838-5844.

Efstratiadis, A., W. R. Crain, R. J. Britten, E. H. Davidson, and R. C. Kafatos. 1976. Proc. Natl. Acad. Sci. 73: 2289-2293. USA.

Englund, P. T. 1971. Procedures in Nucleic Acid Research. G. Cantoni and D. Davies, Eds. Harper & Row, New York, pp. 864-874.

Feldman, M. Ya. 1973. Progress in Nucleic Acid Research and Molecular Biology. J. N. Davidson and W. E. Cohn, Eds. Academic Press, New York and London, Vol. 13, pp. 1-49.

Getz, M. J., L. C. Altenburg, and G. F. Saunders. 1972. Biochim. Biophys. Acta 287: 485-494.

Graham, D. E., B. R. Neufeld, E. H. Davidson, and R. J. Britten. 1974. Cell 1: 127-137.

Gruss, P., and G. Sauer. 1975. FEBS Letters 60: 85-88.

Holland, C. A., and D. M. Skinner. 1977. Chromosoma 63: 223-240. Berlin.

Hudspeth, M. E. S., W. E. Timberlake, and R. B. Goldberg. 1977. Proc. Natl. Acad. Sci. 74: 4332-4336. USA.

Kersten, W., H. Kersten, and W. Szybalski. 1966. Biochemistry 5: 236-244.

Maniatis, T., A. Jeffrey, and H. van deSande. 1975. Biochemistry 14: 3787-3794.

Manning, J. E., C. W. Schmid, and N. Davidson. 1975. Cell 4: 141-155.

Manuelidis, L., and J. C. Wu. 1978. Nature 276: 92-94. London.

Markham, R. 1967. Methods in Virology. K. Maramorosch and H. Koprowski, Eds. Academic Press, New York, pp. 1-39.

Maxam, A., and W. Gilbert. 1977. Proc. Natl. Acad. Sci. 74: 560-564. USA.

Mowbray, S. L., S. A. Gerbi, and A. Landy. 1975. Nature 253: 367-370. London.

Moyzis, R., J. Bonnet, and P. O. P. Ts'o. 1977. J. Cell Biol. 75: 130a.

Noll, H. 1969. Techniques in Protein Biosynthesis. P. N. Campbell and J. R. Sargent, Eds. Vol. 2. Academic Press, New York, pp. 101-179.

Orosz, J. M., and J. G. Wetmur. 1974. Biochemistry 13: 5467-5473.

Roberts, R. J., P. A. Myers, A. Morrison, and K. Murray. 1976. J. Mol. Biol. 102: 147-165.

Rosenberg, H., M. Singer, and M. Rosenberg. 1978. Science 200: 397-402.

Skinner, D. M. 1967. Proc. Natl. Acad. Sci. 58: 103-110. USA.

Skinner, D. M. 1977. BioScience 27: 790-796.

Skinner, D. M., and W. G. Beattie. 1973. Proc. Natl. Acad. Sci. 70: 3108-3110. USA.

Skinner, D. M., W. G. Beattie, F. R. Blattner, B. P. Stark, and J. E. Dahlberg. 1974. Biochemistry 13: 3930-3937.

Skinner, D. M., and M. S. Kerr. 1971. Biochemistry 10: 1864-1872.

Smith, H. O. 1974. Methods in Molecular Biology. R. B. Wickner, Ed. Vol. 7. Academic Press, New York and London, pp. 71-85.

Studier, F. W. 1965. J. Mol. Biol. 11: 373-390.

Sueoka, N. 1961. J. Mol. Biol. 3: 31-40.

Vaughn, J. C. 1975. Chromosoma 50: 243-257. Berlin.

Vogt, V. M. 1973. Eur. J. Biochem. 33: 192-200.

Walbot, V., and L. S. Dure. 1976. J. Mol. Biol. 101: 503-536.

Waring, M., R. J. Britten. 1966. Science 154: 791-794.

Wilson, D. A., and C. A. Thomas, Jr. 1974. J. Mol. Biol. 84: 115-144.

APPENDICES

APPENDIX A

TABLES

Table 1

KINETIC ANALYSIS OF GERYON DNA REASSOCIATION

COMPONENT	FRACTION OF TOTAL DNA	LIMITS OF REASSOCIATION (Cot)	AVERAGE RATE CONSTANT ($1\ M^{-1}\ sec^{-1}$)	$Cot_{1/2}$	AVERAGE REPETITION FREQUENCY
ZERO-TIME BINDING DNA	0.04	10^{-5}	–	–	–
VERY HIGHLY REPETITIVE	0.22	$10^{-5} - 2\times10^{-3}$	984.0	1×10^{-3}	3.9×10^{6}
HIGHLY REPETITIVE	0.34	$2\times10^{-3} - 2$	20.7	4.8×10^{-2}	8.3×10^{4}
SINGLE COPY	0.35	$300 - 3\times10^{4}$	2.5×10^{-4}	4×10^{3}	1

Table 2

DUPLEX CONTENT OF DNA FRAGMENTS REASSOCIATED TO Cot 2.0

FRAGMENT LENGTH (NUCLEOTIDES)	250	420	1700
FRACTION BOUND TO HYDROXYAPATITE (F)	0.45	0.51	0.81
HYPERCHROMICITY (H)[a]	0.241	0.172	0.145
T_m (°C)	76	75.5	76
AVERAGE DUPLEX CONTENT FROM HYPERCHROMICITY OF BOUND FRAGMENTS (D)[b]	0.86	0.59	0.48
PERCENT OF GENOME, (D X F)	0.39	0.30	0.39

[a]Calculated as (A_{260} at 100°C - A_{260} at 55°C)/A_{260} at 100°C.

[b]Calculated as $(H - 0.025)/(H_{native} - 0.025)$. A correction of 2.5% is made for the hyperchromicity of single strands. $H_{native} = 0.275$.

Table 3

FRACTION OF FRAGMENTS CONTAINING REPETITIVE SEQUENCES

FRAGMENT LENGTH (NUCLEOTIDES)	FRACTION BOUND TO HYDROXYAPATITE (F)			CORRECTED FRACTION BOUND (F_C)[a]	
	Cot 10^{-5}	Cot .002	Cot 2	Cot .002	Cot 2
110	0.09	0.2	0.38	0.12	0.31
300	0.08	0.2	0.58	0.14	0.56
400	0.09	0.23	0.62	0.16	0.63
500	0.19[b]	0.33[b]	0.78[b]	0.18	0.72
800	0.21[b]	0.48[b]	0.87[b]	0.35	0.84
1100	0.17[b]	0.46[b]	0.85[b]	0.36	0.82
1700	0.24	0.46	0.84	0.29	0.79
2000	0.21[b]	0.5[b]	0.89[b]	0.36	0.86

[a]The fraction bound to hydroxyapatite (F) is corrected for zero-time sequences (Z) according to the formula, $F_c = \frac{F - Z}{1 - Z}$, where F_c is the corrected fraction bound.

[b]The fraction bound to hydroxyapatite is an average of two determinations that differ by 3% or less.

Table 4

ESTIMATES OF THE AMOUNT OF EACH COMPONENT IN THE GENOME

COMPONENT	KINETIC[a] (420 nt)	INTERSPERSION[b] (110-2000 nt)	S_1[c] (2000 nt)
ZERO-TIME BINDING DNA	4.0%	6.0%	5.0%
VERY HIGHLY REPETITIVE	22.0	4.0	4.0
HIGHLY REPETITIVE	34.0	28.0	29.0
SINGLE COPY	35.0	57.0	57.0
UNREACTIVE	5.0	5.0	5.0

[a]Derived from the least squares fit to the data for the reassociation of 420 nucleotide fragments assuming three second order components.

[b]Determined from reassociation of labeled fragments of increasing length with 300 nucleotide driver DNA to Cots 10^{-5}, 2×10^{-3}, and 2.

[c]S_1 resistance at Cot 10^{-5}, 2×10^{-3}, and 2 corrected for each repetitive component. Undigested DNA at Cot 2 is corrected for unreactive sequences to estimate single copy DNA.

APPENDIX B

FIGURE LEGENDS AND FIGURES

Figure 1. Reassociation kinetics of Geryon DNA.

420 nt fragments of DNA were denatured, reassociated, and assayed on hydroxyapatite, unlabeled (0); Pol I labeled (●); ^{125}I-iodinated (ℑ). The solid line represents the solution of a least squares computer analysis which also resolved the second order components shown by the dashed lines. As a standard, the Cot 1/2 of iodinated E. coli DNA (Δ), was measured.

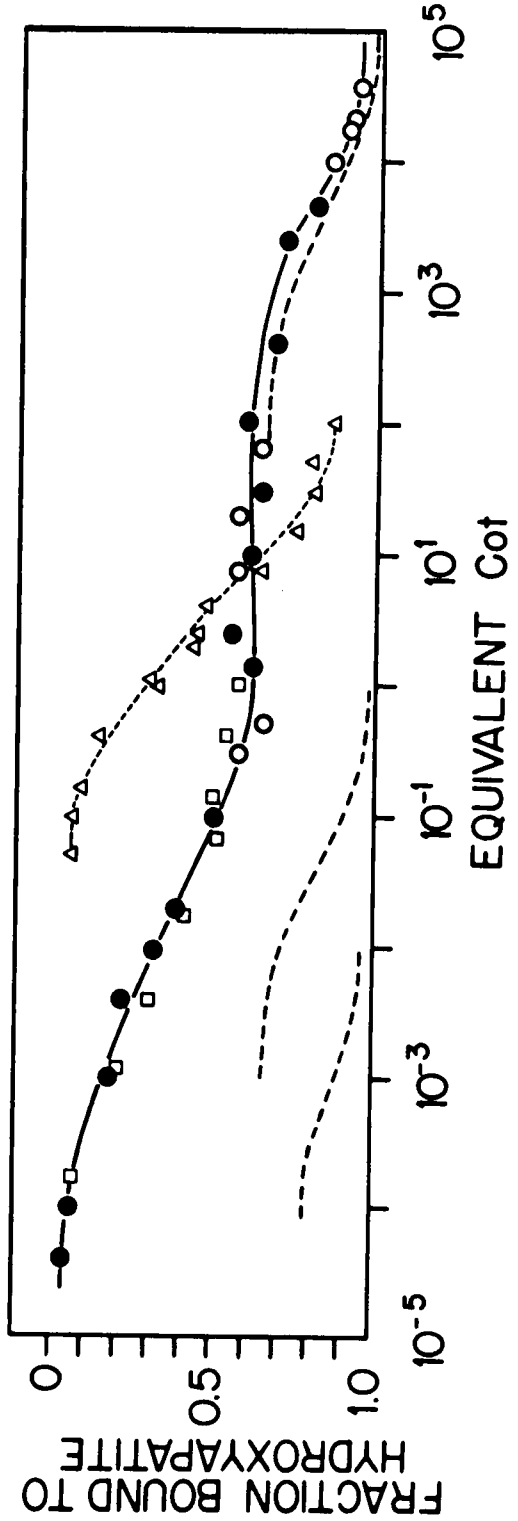

FRACTION BOUND TO HYDROXYAPATITE
0
0.5
1.0
10^{-5}
10^{-3}
10^{-1}
10^{1}
10^{3}
10^{5}
EQUIVALENT Cot

Figure 1

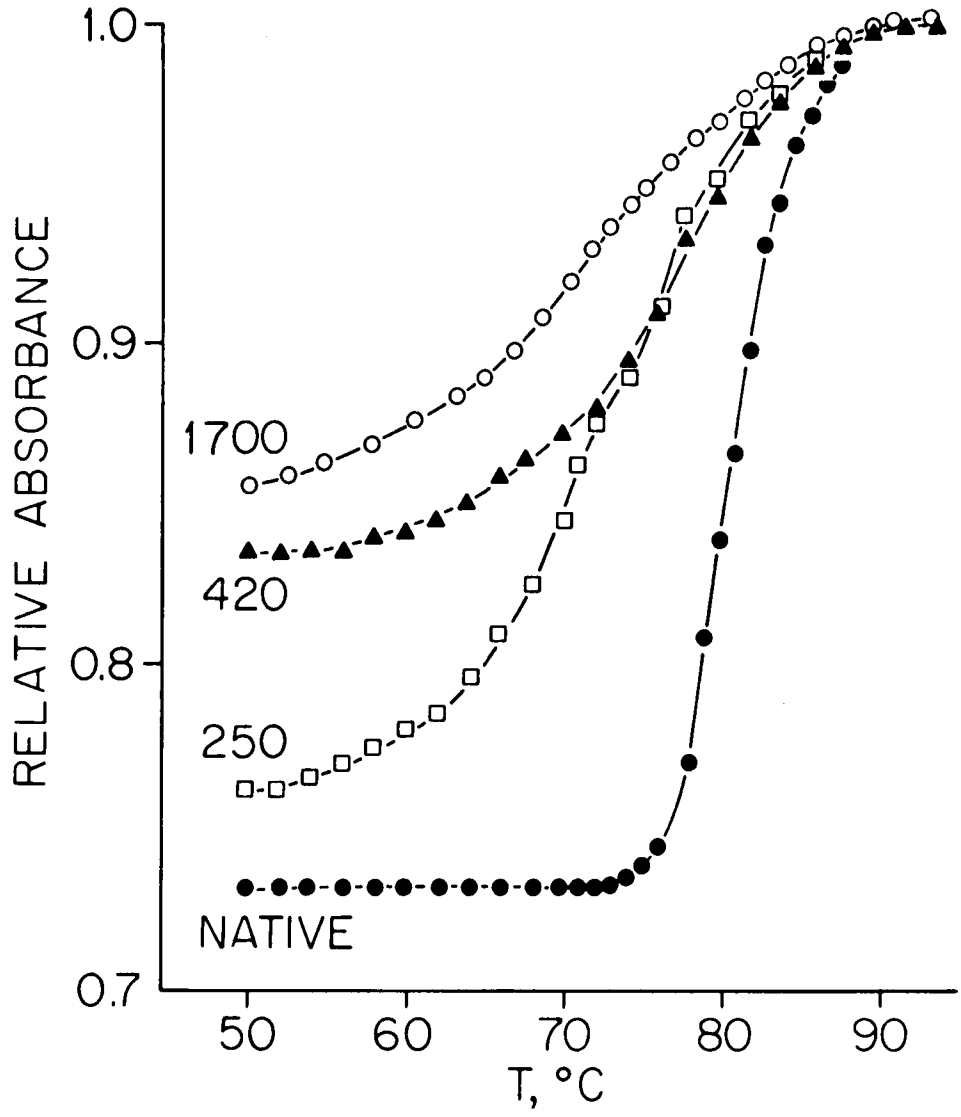


Figure 2. Thermal dissociation of reassociated DNA.

DNA fragments of 250, 420, and 1700 nt were denatured, reassociated to Cot 2 and melted. A melt of native DNA is included for comparison.

Figure 3. Size distribution of S1 resistant sequences of DNA.

^{125}I-iodinated 2000 nt fragments were reassociated to (A) Cot 10^{-5} (B) Cot 2×10^{-3} or (C) Cot 2 and digested with S1 nuclease. Positions are indicated for an included marker (Inc; ^{3}H-uridine), DNA fragments 450 and 1200 nt, and an excluded marker (Exc; 4000 nt). The lower curve of (B) shows the size distribution of repetitive sequences which reassociate between Cot 10^{-5} and Cot 2×10^{-3}, i.e., the very highly repetitive component. The lower curve of (C) represents the size distribution for sequences which reassociate between Cot 2×10^{-3} and Cot 2, i.e., the highly repetitive component.

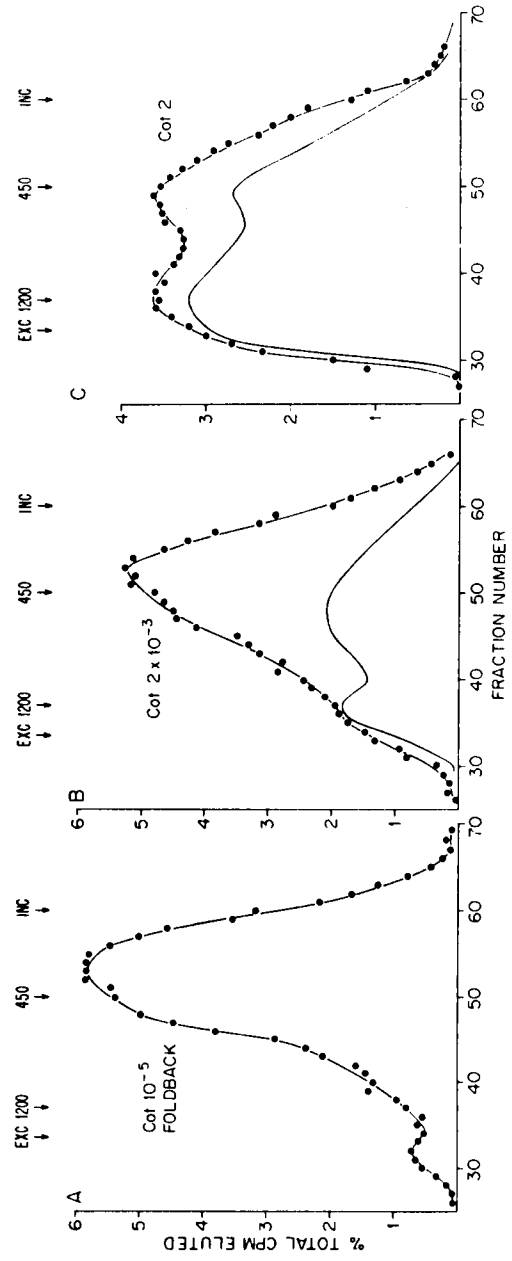

% TOTAL CPM ELUTED
FRACTION NUMBER
A
B
C
EXC 1200
450
INC
Cot 10⁻⁵
FOLDBACK
Cot 2 x 10⁻³
Cot 2

Figure 3

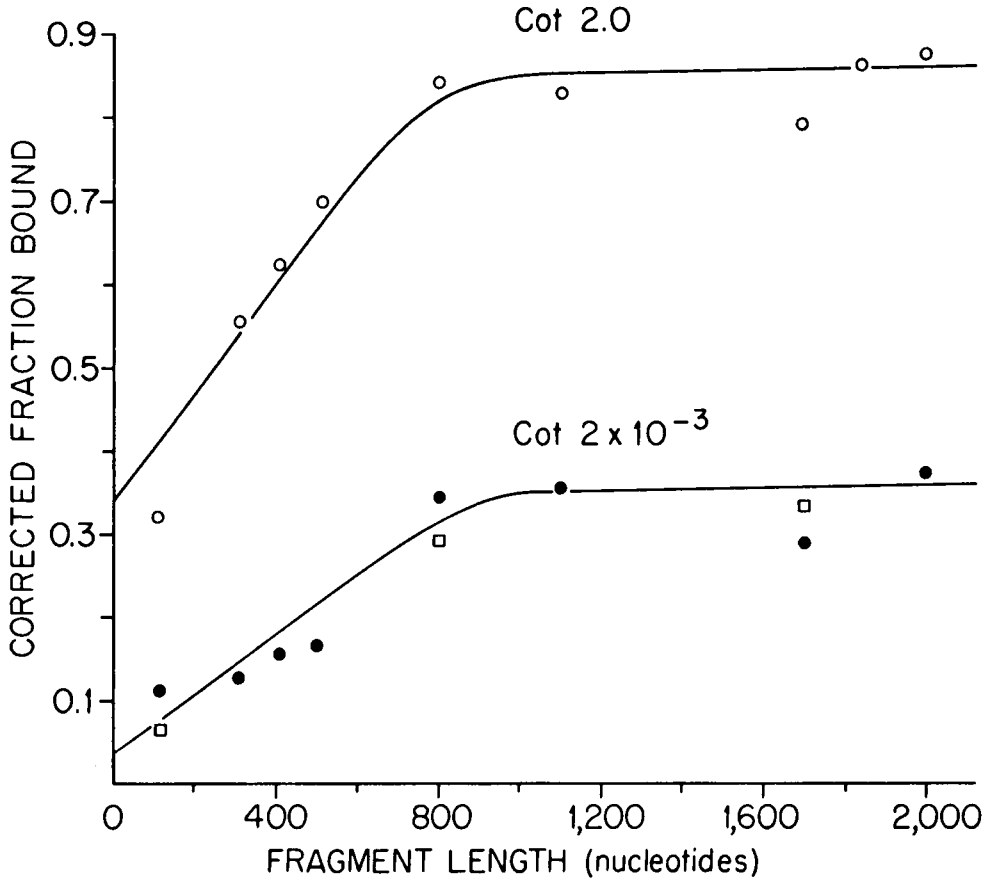


Figure 4. Fractions of Geryon DNA containing repetitive elements as a function of fragment length.

^{125}I-iodinated DNA fragments ranging in size from 110 to 2000 nt as determined by alkaline sucrose gradients were combined with a 100-fold excess of unlabeled 300 nt driver DNA. Reassociation, monitored by hydroxyapatite chromatography, was performed to driver Cot 2 with tracer DNA iodinated after denaturation (0) and to driver Cot 2 × 10^{-3} for tracer DNA iodinated with (●) and without (□) denaturation. See Table 3 for calculation of the corrected fraction bound.

Figure 5. Analysis of single copy DNA adjacent to the very highly repetitive component.

A. Direct Measurements: The percent of ^{125}I-iodinated DNA fragments containing a duplex was monitored by hydroxapatite (HAP) chromatography. The percent of the genome in duplex at the different Cots was measured as the S1 resistance of DNA fragments.

B. Fractionation Measurements: ^{125}I-iodinated DNA fragments with and without very highly repetitive sequences were isolated by reassociation to Cot 2×10^{-3}, denatured and reassociated to Cot 2 in the presence of driver DNA of the same size. The S1 resistance of the DNA fragments was determined after the second reassociation. SC refers to single copy DNA, HR to highly repetitive DNA, VHR to very highly repetitive DNA and Z to zero-time binding DNA.

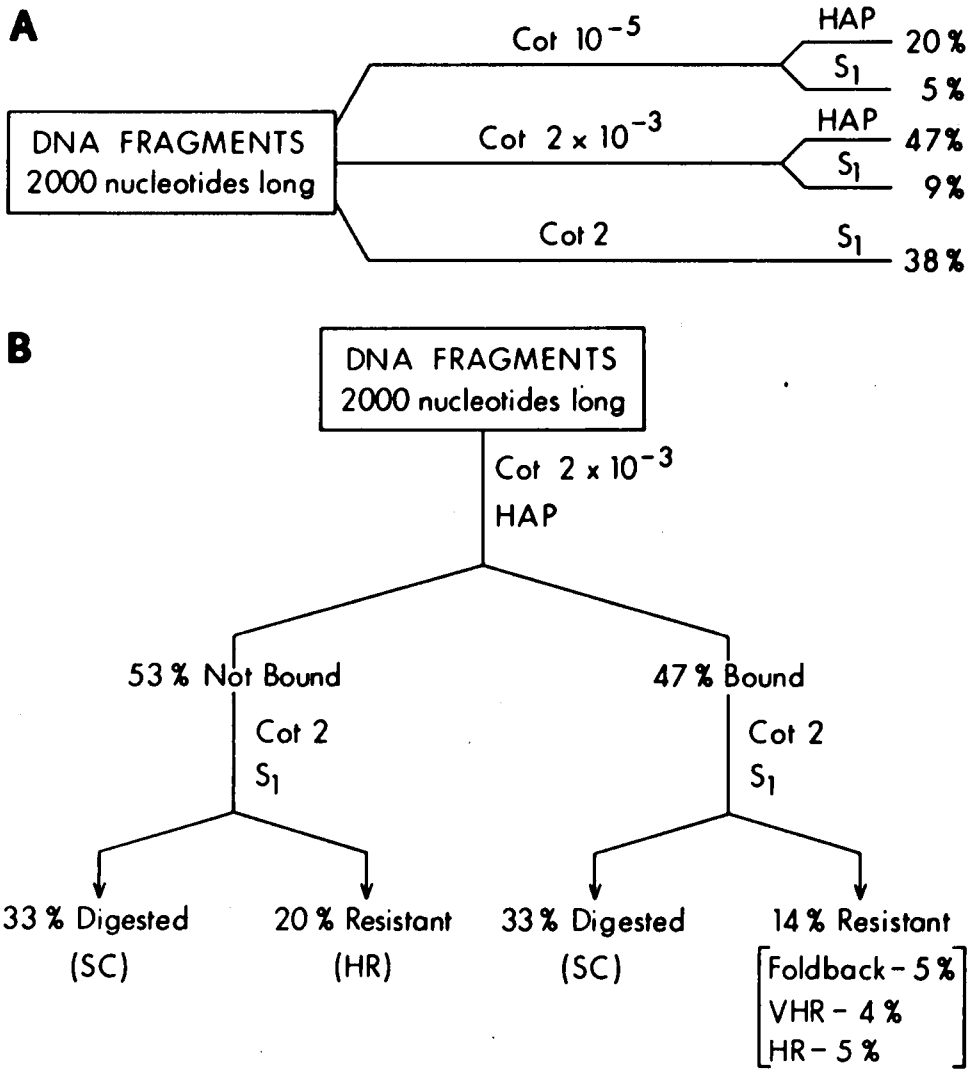

A
DNA FRAGMENTS
2000 nucleotides long
Cot 10^{-5}
HAP
20 %
S_1
5 %
Cot 2×10^{-3}
HAP
47 %
S_1
9 %
Cot 2
S_1
38 %
B
DNA FRAGMENTS
2000 nucleotides long
Cot 2×10^{-3}
HAP
53 % Not Bound
47 % Bound
Cot 2
S_1
Cot 2
S_1
33 % Digested
(SC)
20 % Resistant
(HR)
33 % Digested
(SC)
14 % Resistant
Foldback - 5 %
VHR - 4 %
HR - 5 %

Figure 5

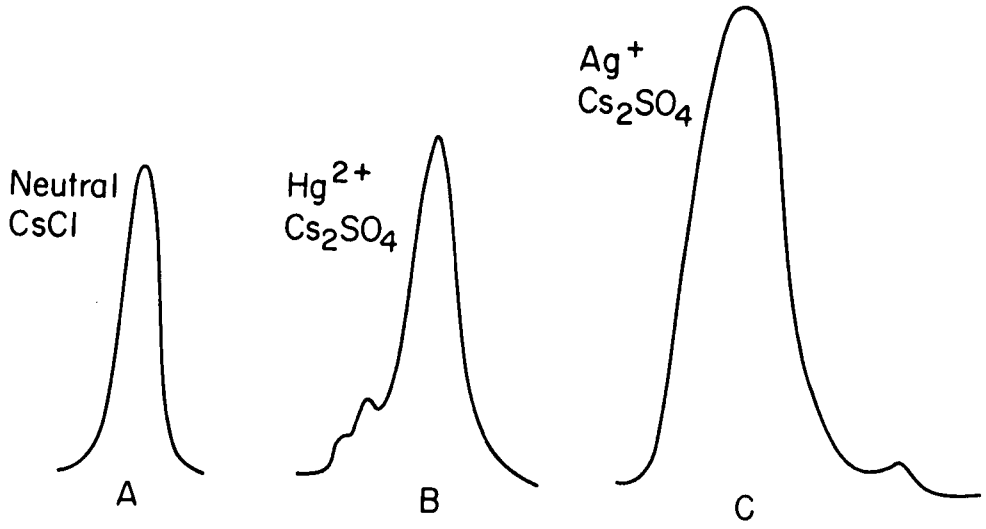


Figure 6. Model E analytical ultracentrifuge tracings of total DNA.

A. 3 μg of DNA in neutral CsCl.

B. 5 μg of DNA in neutral Cs_2SO_4 with Hg^{2+} (rf = 0.26).

C. 15 μg of DNA in neutral Cs_2SO_4 with Ag^+ (rf = 0.19).

Figure 7. Restriction endonuclease digestion of three kinetic fractions and native DNA.

A,B. (1). Zero-time binding DNA was isolated after rapid addition of formaldehyde to heat-denatured 1500 nt fragments.

(2). Sequences of the very highly repetitive component were isolated by reassociation of 1500 nt fragments to Cot 2×10^{-3}.

(3). Highly repetitive sequences were collected by further reassociation to Cot 2 of the single-stranded DNA from a Cot 2×10^{-3} sample. Each kinetic fraction was 5 µg of duplex DNA.

(4). Native DNA (6 µg) was digested with either HindIII (A) or AluI (B). Electrophoresis on polyacrylamide gels was for 1.5 hr for A, Lanes 1-4, and for B, Lanes 1-3, and was for 2 hr for B, Lane 4.

C. Densitometer tracings of the AluI digests of very highly repetitive (1;B, Lane 2) and highly repetitive (2; Lane 3) sequences.

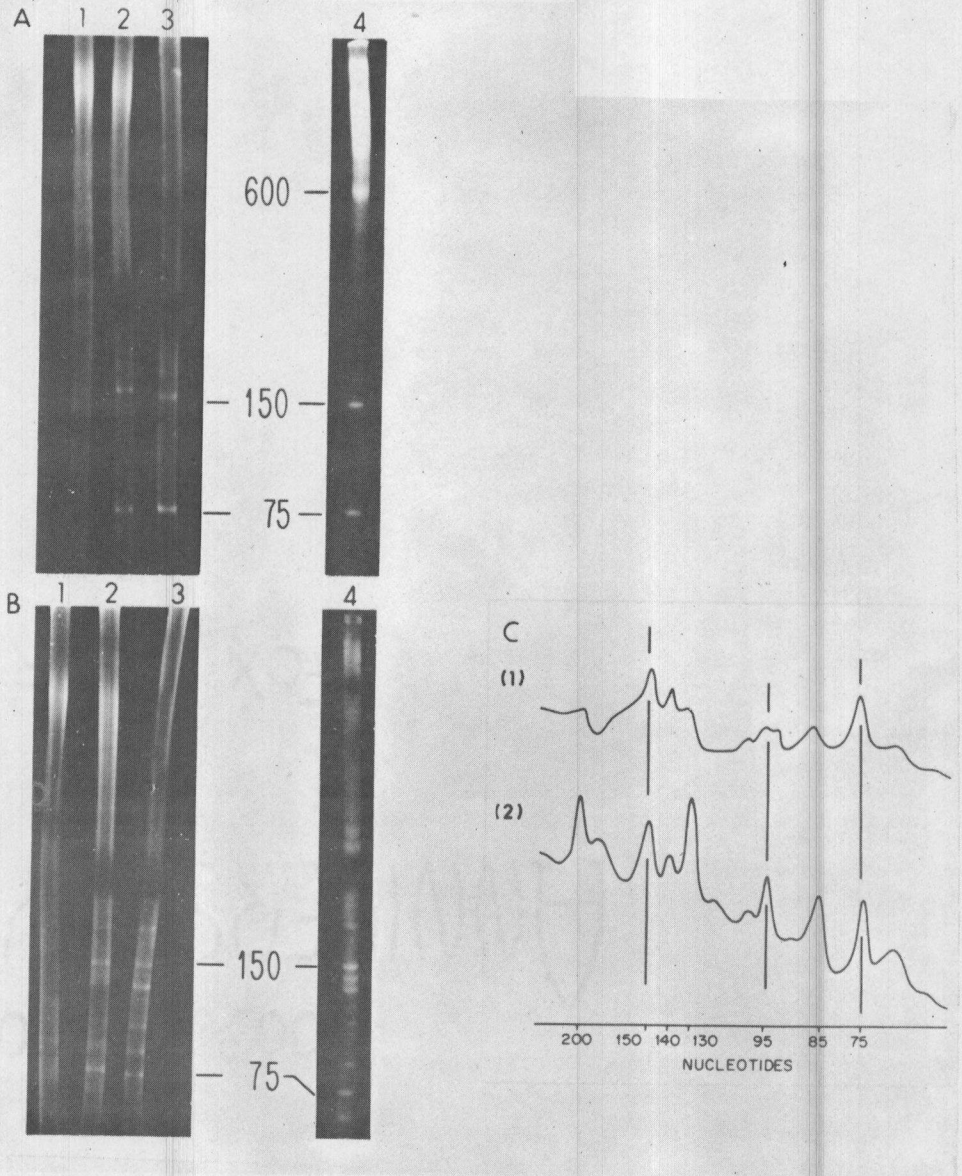

A
1
2
3
4
600
150
75
B
1
2
3
4
150
75
C
(1)
(2)
200
150
140
130
95
85
75
NUCLEOTIDES

Figure 7

Figure 8. Possible arrangements for sequences of the very highly repetitive component.

Symbols are as follows: (---) VHR, very highly repetitive sequences; (∿∿) HR, highly repetitive sequences; and (——) SC, single copy sequences.

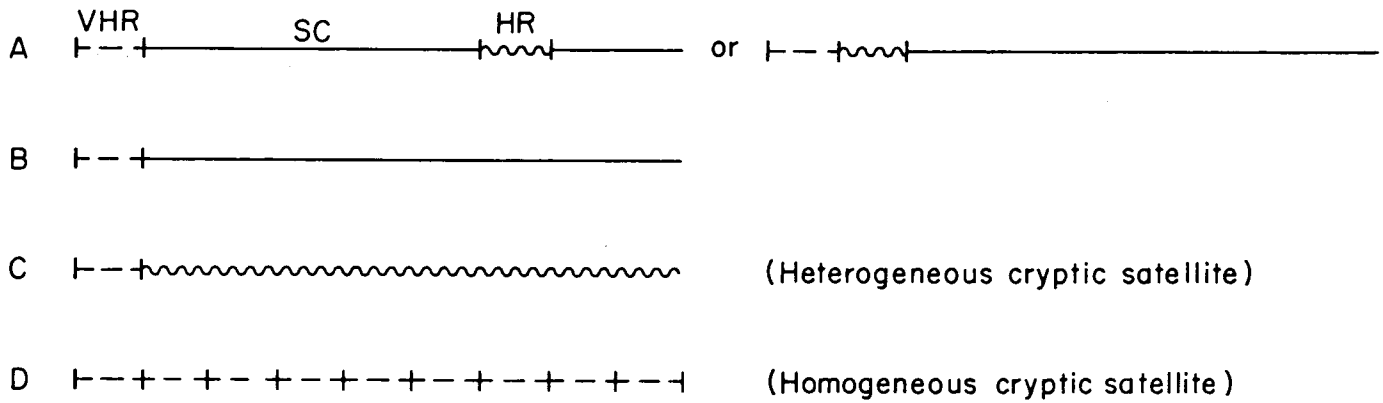

VHR
SC
HR
A
or
B
C
(Heterogeneous cryptic satellite)
D
(Homogeneous cryptic satellite)

Figure 8

APPENDIX C

EXTENDED PROCEDURES

Labeling DNA by Iodination

Reagents:

1. $Na^{125}I$ was ordered from Amersham/Searle in 1 mCi lots. It was shipped in 0.010 ml dilute NaOH, pH 8-11. Specific activity was always given as a range of 11-17 mCi/µg. Before use 0.100 cc of 6×10^{-5} M KI was added.
2. $TlCl_3$ in 0.5 M $NaC_2H_3O_2$ (NaAc). The pH of NaAc is adjusted to 4.6-4.8 with 1 M $HC_2H_3O_2$.
3. I buffer:

   0.4 M NaCl

   0.015 PB, pH 6.0 (phosphate buffer)

   0.0002 M EDTA (disodium ethylene diamine tetraacetate)

   PB:

   Make 1 M $NaH_2PO_4 \cdot 1\ H_2O$, and

   1 M $NaHPO_4 \cdot 12\ H_2O$.

   For pH 6.0 mix in a ratio of dibasic:monobasic of 1:2.

REACTION CONDITIONS

	REAGENT	REACTION CONCENTRATION	STOCK CONCENTRATION	AMOUNT ADDED
1.	DNA	15-20 µg/ml	Variable	
2.	$Na^{125}I$	4×10^{-7} M (~ 120 µCi)		0.012
3.	$TlCl_3$	2×10^{-3} M	2×10^{-2} M	0.120
4.	Glass distilled water (GDW)			
				1.200 ml Final vol.

Procedure:

1. Measure GDW and $Na^{125}I$ into Pyrex reaction tube which has been siliclad. Leave in ice.
2. In a second siliclad reaction tube denature DNA in an ethylene glycol bath (101°-105°) for 4 to 5 min. Do not put in ice after denaturation.
3. During denaturation add $TlCl_3$ to reaction tube containing $Na^{125}I$.
4. Measure DNA solution into reaction tube in ice. Rinse denaturation tube with volume of GDW equal to that lost during denaturation. Add this to reaction tube.
5. Place at 60° for 5 min.
6. Chill in ice. Transfer reaction mix to dialysis sac (smallest size). Dialyze against I buffer overnight with a total of 3 changes, using a minimum of 4-hour intervals.
7. Transfer iodinated sample from dialysis tubing to a siliclad reaction tube. Place at $90^{\circ}C$, 15 min. Sample 1 or 2µl is checked for total incorporation by liquid scintillation counting.
8. Dialyze against buffer of choice. After dialysis, sample is again counted for the amount of stable incorporation. The difference between this count and the count in No. 7 gives the percent of iodine which is unstably bound (~ 40%).

Actinomycin D/CsCl Gradients

Procedure:

1. The amount of actinomycin D is dependent on the amount of DNA used. Typical ratios for density separations are between 1:1 and 1:3 for the number of actinomycin D molecules to the number of base pairs of DNA. These ratios are determined from the following values:

   Actinomycin D (mol. wt.=1200): 1 μg = 5×10^{14} molecules, and

   DNA (mol. wt./base pair = 650): 1 μg = 1×10^{15} base pairs.

   For a ratio of 1:1 for the number of molecules of actinomycin D to the number of base pairs of DNA, use 1 μg actinomycin D for 0.5 μg DNA.

2. Actinomycin D is dissolved in acetone, using as small an amount as possible (one drop preferably). Then add water to a sufficient volume to read the optical density.

3. The concentration of actinomycin D is determined from the following relationship:

   A_{443} = 2 for a concentration of actinomycin D of 100 μg/0.6 ml.

4. An average sample (ratio 1:3) for the Model E ultracentrifuge is given below:

   0.005 ml Tris, 0.1 M, pH 8.0
   0.037 ml 5 μg DNA, stock solution = 135 μg/ml
   0.048 ml 3.4 μg actinomycin D, stock solution = 70 μg/ml
   glass distilled water to make 0.5 ml.

   0.66 g CsCl is dissolved in the above sample and the refractive index is adjusted to 1.4005 to 1.4010.

Isolation of HindII and HindIII Restriction Enzymes

Reagents:

1. Lysozyme, dissolved in lysis buffer.
2. Lysis buffer

   0.025 M Tris, pH 8.0

   0.005 M MSH (mercaptoethanol)

   0.001 M NaN_3 (sodium azide)

   25 µg/ml PMSF (phenyl methyl sulfonyl fluoride)[1]
3. Streptomycin sulfate, freshly dissolved.
4. Ammonium sulfate, crystalline
5. Potassium phosphate buffer (KPB)

   0.01 M KPB, pH 7.4 {4:1, dibasic:monobasic}

   0.001 M NaN_3

   0.005 M MSH

   25 µg/ml PMSF
6. Solution for gradient elution of P-11 column

   0.15 M KCl

   0.01 M KPB

   0.001 M NaN_3

   0.005 M MSH

   25 µg/ml PMSF

---

[1]PMSF; a selective enzyme inhibitor. Make up stock solution in absolute alcohol.

7. Solution for gradient elution of P-11 column

   0.4 M KCl

   0.01 M KPB

   0.001 M NaN_3

   0.005 M MSH

   25 µg/ml PMSF

8. Equilibrating buffer for DE 52 column

   0.010 M Tris, pH 8.0

   0.005 M MSH

   0.001 M NaN_3

   25 µg/ml PMSF

9. Elution buffer for DE 52 column

   0.010 M Tris, pH 8.0

   0.150 M KCl

   0.005 M MSH

   0.001 M $NaNa_3$

   25 µg/ml PMSF

10. Dialysate

   0.010 M Tris, pH 8.0

   0.001 M NaN_3

   50% glycerol

Columns:

1) Sephadex G-25 (2.5×16 cm)

2) Phosphocellulose P-11 (1.0×38 cm)

3) DEAE cellulose DE 52 (size dependent on amount of protein)

Equipment:

Sonicator

Fraction collector

Procedure:

1. Resuspend 30 g of cells in 100 ml of lysis buffer in a small beaker. Keep the suspension on ice or in the cold room at all times. After the cells are in solution, add 1 ml lysozyme (6 mg). Stir gently. Let stand 1 hour.
2. Sonicate the solution as follows:
   a. Tune the instrument (Biosonic) by ear to the point at which the sound is most intense.
   b. Sonicate in 1.5 min bursts with the sample beaker immersed in an ice bath. Keep the temperature below 8°.
   c. During each burst continuously swirl the beaker. Viscosity is usually reduced after 4.5 to 6 min.
3. Spin the sonicate at 9,000 rpm in Corex tubes for 1 hr. If plastic tubes are used, this and subsequent spins can be performed at 13,500 rpm for 30 min.
4. The nucleic acids are precipitated with a 10% streptomycin sulfate solution. For each 1800 A_{260} units 1 ml of solution is added. The solution is added dropwise to the sonicate with continuous stirring over a 20 min period. Stirring is continued 30 min after addition is complete. Let stand overnight. This step can be done in the cold room.

5. Spin in Corex tubes at 9,000 rpm for 1 hr.
6. For the first ammonium sulfate precipitation dilute the supernatant by 2.3 times the original volume with lysis buffer and simultaneously bring the solution to 0.3 M NaCl. Add ammonium sulfate over a 30 min period to bring the solution to 50% saturation (0.313 g/ml). Continue stirring an additional 30 min.
7. Spin 9,000 rpm for 1 hour.
8. Add ammonium sulfate to the supernatant to 70% saturation (0.16 g/ml). Add as for previous precipitation. Let stand 30 min.
9. Spin at 9,000 rpm for 1 hour. Save pellet.
10. Dissolve pellet in KPB (Solution 5).
11. Load sample onto a G-25 column and elute with KPB (Solution 5).
12. Collect 2 ml fractions and pool the amber-colored fractions (approximately 20 ml).
13. Load pooled fractions onto a P-11 column equilibrated with KPB (Solution 5). After loading the sample volume, wash with 50 ml of equilibrating buffer. Develop column with 300 ml total volume of a linear gradient of 0.15 M NaCl to 0.4 M NaCl using Solutions 6 and 7. Collect 4 ml fractions. This column may require 16 to 30 hr to run.
14. Fractions may be sampled by the clot test (see procedure 18 below) or by digestion of a susceptible DNA followed by gel electrophoresis. Using the column dimensions indicated above a fraction of the HindII will elute prior to fractions of combined HindII + III enzymes. Fractions of pure HindIII are eluted next.

15. If separation of HindII from HindIII is desired, the appropriate fractions are desalted on the G-25 column equilibrated with Solution 8. The desalted sample is applied to a DE 52 column. For each 2 grams of protein 1 ml of cellulose is used, assuming 1 mg protein for each A_{280} unit. HindII appears in the 0.01 M Tris-HCl (Solution 8) breakthrough volume while HindIII is eluted with 0.01 M KCl (Solution 9).

16. Pooled enzyme fractions are dialyzed in the cold room against Solution 10.

17. All fractions are stored at -20°.

18. The clot assay. The following solutions are required: buffer (0.01 M Tris, pH 7.5; 0.05 M NaCl; 0.01 M $MgCl_2$), 10% trichloroacetic acid (w/v), 0.3% bovine serum albumin (BSA; w/v), and phenol-extracted calf thymus DNA (1.5 mg/ml).

    Aliquots of column fractions (10-50 µl) are added to 0.1 ml buffer containing 30 µg DNA. After incubation at 37° for 20 min, 1 drop (0.05 ml) BSA solution and 0.2 ml of 10% TCA solution are added. After mixing, a milky precipitate indicates restriction activity and a clotted precipitate indicates absence of activity.

Restriction Endonuclease Digestion of Kinetic Fractions

Reagents:

1. Standard buffers used for DNA reassociation (see Chapter II).
2. Formaldehyde, 37%.

Procedure:

1. Shear DNA to a fragment size of 1500 to 2000 nt in a Virtis 60K homogenizer at 14,000 to 17,000 rpm for 25 min. The DNA sample is adjusted to 33 ml of 0.01 M phosphate buffer (PB), pH 6.8. The shear vessel is submerged in an ice bath.
2. Denaturation and reassociation:
   a. Foldback. Denaturation is performed in Corex tubes (siliclad) in 10 ml aliquots of DNA solution, 0.12 M PB. Immediately after denaturation, 10 ml cold 1.2% formaldehyde in 0.12 M PB is added. The sample is applied to a hydroxyapatite (HAP) column and the duplex fraction is eluted with 0.4 M PB and concentrated by ethanol precipitation.
   b. Very highly repetitive component. The time of reassociation to reach Cot 2×10^{-3} is calculated according to Britten et al. (1974). In general,

$$\text{Cot} = ML^{-1}t \text{ (in seconds)} = \frac{OD}{2} \times t \text{ (in minutes)}, \text{ where}$$

M is moles, L is liter, t is time and OD is optical density. Typically the time for this fraction is approximately a minute. This brief reassociation time is accomplished by denaturing as in (a) and applying the sample to the HAP column as soon as the sample reaches 60°. After the duplex fraction is eluted as in (a),

the collected fractions are adjusted to 0.6% formaldehyde to prevent further reassociation in subsequent steps. The DNA can be concentrated by ethanol precipitation.

c. Highly repetitive component. The fraction not bound at Cot 2 × 10^{-3} in (b) is further reassociated to Cot 2, applied to a HAP column and eluted as above. The sample is adjusted to 0.6% formaldehyde and concentrated by ethanol precipitation.

3. Digestion with restriction endonucleases. The amount of enzyme required is determined by calculating the expected fraction of each sample in duplex determined from S1 digestion measurements. Expected recoveries of duplex DNA are as follows:

a. Foldback. Of the starting sample 20% is bound to HAP with 5% of the starting sample in duplex.

b. Very highly repetitive component. Of the starting sample 30% is bound to HAP with 9% of the starting sample in duplex. This represents 4% as foldback and 4% as the very highly repetitive sequences.

c. Highly repetitive component. Further reassociation of the fraction not bound at Cot 2 × 10^{-3} to Cot 2 results in approximately 20% of the initial sample in duplex as the highly repetitive sequences.

Conditions for digestion were the same as in Chapter II.

5. The reaction products were displayed on 5% polyacrylamide tube gels. Dye markers were omitted to avoid obscuring the bands of DNA.

PART II

SEQUENCE ORGANIZATION OF A MULTIMERIC SERIES OF RESTRICTION FRAGMENTS FROM TOTAL DNA

CHAPTER I

INTRODUCTION

The size of the eukaryotic genome greatly exceeds estimates of the amount of DNA required for growth and development of the organism; no more than 5 to 15% is thought to be involved in a protein coding function. The remainder of the genome, while possibly involved in chromosome structure, is quite likely exerting an influence on gene expression. Within the large non-coding fraction, attention has been focused on repetitive DNAs because of their ease of isolation and chromosomal localization. In the four brachyuran ("true crab") genomes which have been examined, highly repetitive sequences comprise 25 to 41% of the genome (Vaughn, 1975; Holland and Skinner, 1977; Christie and Skinner, 1979). In each case except the red crab, Geryon quinquedens, these sequences include patent satellite DNAs. The most striking characteristic common to these four species is that sequences typically classified as middle repetitive are present only in minimal amounts. This is tantamount to stating that the majority of the repetitive sequences in these genomes, greater than 90%, can appropriately be termed highly repetitive. While this observed skew in the distribution of repetition frequencies may or may not be a characteristic feature of crab genomes, it facilitates fine structure analyses of this class of DNA. Detailed structural analyses should be useful in elucidating origins for highly repetitive DNAs. The structure of a satellite fraction from the hermit crab Pagurus pollicaris presents strong evidence that unequal crossovers occur in highly repetitive DNAs

(Chambers et al., 1978). The marked diversities in the structural features of satellite DNAs (Skinner, 1977) suggest that more than one mechanism may be involved in their formation.

Although the DNA of the red crab Geryon lacks patent density satellites, it contains several classes of sequences with characteristics similar to satellites DNAs. Both the reassociation rate and the measured mismatch (4%) of the entire class of repetitive DNA (Christie and Skinner, 1979) are similar to values observed for satellite DNAs (Cordeiro-Stone and Lee, 1976). In addition digestion of Geryon DNA by several restriction endonucleases yields extensive arrays of discretely sized fragments. Earlier we identified a series of fragments by HindIII digestion of total DNA whose sizes were multimers of a basic repeat. The present study was undertaken to explore the relationship of the sequences of the different multimers. The relative amounts of each larger multimer present in several satellite DNAs is reduced compared to the next smaller multimer (Southern, 1975; Altenberger et al., 1977; Fittler, 1977). This result is expected following the random loss of restriction sites that had been spaced uniformly. In Geryon DNA there are much larger amounts of both the tetramer and the octamer than expected. These results suggest a selective amplification of specific fractions of the DNA; amplification of the entire set of HindIII sequences would have maintained multimer ratios determined by the degree of divergence since the initial amplification of the 81 nt repeating unit had occurred.

CHAPTER II

MATERIALS AND METHODS

Testes of Geryon quinquedens were obtained frozen from the Marine Biological Laboratory, Woods Hole, Massachusetts. DNA was extracted by a modified "MUP" procedure (Britten et al., 1974). Both labeled and unlabeled DNAs were used for restriction enzyme analyses. DNA was labeled with γ-^{32}P-ATP (S.A., 2000-3000 Ci/mmol; ICN) at the 5' termini after treatment with bacterial alkaline phosphatase (Worthington) by the action of T4 polynucleotide kinase (P. L. Biochemicals, Inc.).

Restriction Enzyme Analyses and Isolation of Restriction Fragments. HindIII and AluI were isolated according to Smith (1974). Other restriction endonucleases were purchased from New England Biolabs. Analytical digests (15 and 20 µl) were brought to 7-10% glycerol and 0.025% each of xylene cyanol ff and bromphenol blue and applied directly to polyacrylamide gels (5, 7 or 10%). Preparative digests (Appendix C) were ethanol precipitated and solubilized in electrophoresis buffer before application to the gels. Typically, 50 µg of DNA was applied to wells 1.2 × 0.2 cm for preparative isolations. Electrophoresis was carried out in a vertical apparatus with 90 mM Tris-borate buffer, pH 8.3, and 2.5 mM EDTA (Maniatis et al., 1975). Individual fragments were excised from the gels and recast with 7% polyacrylamide in 7 mm (O.D.) glass tubes containing a 10% polyacrylamide plug. Fragments were electroeluted at 4°C into dialysis bags and precipitated before redigestion by other restriction endonucleases.

Quantification of the DNA in the Restriction Fragments.

Quantification of the amounts of HindIII fragments was accomplished using optical density scans of total DNA (6 μg) which had been digested with HindIII and electrophoresed on a 5% acrylamide gel. Since the film response to impinging light is represented by a sigmoidal function (Prunell et al., 1977), parallel gels of ØX174 DNA (1 μg) digested with AluI were photographed simultaneously. The amount of ØX174 DNA and the photographic conditions (aperture settings, time of exposure) were adjusted until the ratios for the areas under selected DNA peaks agreed with independent determinations of the amount of DNA in the corresponding fragments (Vereijken et al., 1975). This assures that the response of the film was in the linear range and justifies the use of peak areas to determine the amount of DNA in the small (<400 nt) HindIII fragments of Geryon. The isolation procedure for ØX174 DNA is given in Appendix C.

Measurements of the relative amounts of the fragments were made by 5' end-labeling a HindIII digest of total DNA and electrophoresing on a 7% polyacrylamide gel. The gels were sliced into 1-mm slices, incubated overnight with 0.2 ml of $HClO_4$:H_2O_2 (1:2 by volume), and counted in ACS scintillation fluid (Amersham/Searle). The total cpm in the major peaks ranged from 5.2×10^4 to 2.8×10^5. Estimates of the percent of each fragment were made from the lengths of the fragments, their relative amounts, and the genome size.

Electrophoretic Size Determination of the Restriction Fragments.

The size of the restriction fragments were obtained by comparison to HinfI fragments of ØX174 (24 to 726 nt). A least squares regression

line was calculated for plots of the log of the known HinfI sizes in nucleotides (Sanger et al., 1977) versus the distance migrated (Appendix C). Certain electrophoretic conditions required that separate lines be fitted for fragment sizes above and below 100 nt. The linear correlation coefficients were between -0.992 and -1.0. The deviations between the sizes obtained from DNA sequencing data (Sanger et al., 1977) and the sizes estimated from the linear regression line were only 1 to 5 nt for fragments less than 200 nt but increased for larger fragments. When electrophoresis was maximized for the larger fragments, the fractional deviations in size were comparable to those of the small fragments. Absolute deviations will increase in direct proportion to fragment length since sizes are determined from a logarithmic scale (Schroeder and Blattner, 1978). Greatest accuracy was obtained for ØX174 fragments of 500 or more nucleotides when the fragments were electrophoresed to the central portion of 5% polyacrylamide gels.

CHAPTER III

RESULTS

When total DNA is digested with any of several enzymes, fragments of discrete sizes are observed. The fragments are superimposed on a background of DNA fragments having a continuous range of sizes (Figure 1, Appendix B). Most of the enzymes used had recognition sites of 4 nucleotides; the sites occur rather frequently throughout the genome. Digests of total DNA with SmaI and BamI (Figure 1, Lanes 3 & 7; Appendix B) contain no detectable fragments of discrete size classes.

Organization of HindIII Sites in Total DNA. Treatment of total DNA with HindIII (Figure 2A, Lanes 2 & 4; Appendix B) produced a relatively small number of discretely sized fragments and only a slight background of heterogeneously sized DNA; there were almost no heterogeneous fragments smaller than 300 nt. The sizes of each of the five major fragments was determined as described in Materials and Methods relative to fragments of ØX174 DNA (Figure 2A, Lanes 1 & 3; Appendix B). The lengths of the five fragments conform to the proposed organization (see below) derived from restriction enzyme analyses and direct DNA sequencing. The slight variations in the sizes of the fragments obtained for different electrophoresis conditions are within reasonable experimental error. The smallest fragment (V) is approximately 81 nt; the larger fragments are multiples of this basic unit. Fragments III and IV are very close to the lengths expected for a tetramer and a dimer respectively. Two fragments,

I and II, are both close to the length of an octamer. There are also minor fragments at sizes expected for trimers, pentamers, hexamers and heptamers. Sequences of all the multimers from the dimer to the heptamer and at least one of the octamers are very closely related. There are no nonomers and the decamers may be unrelated to the basic 81 nt repeat. The decamer may be related to one of the octamer fragments and possibly to a minor fragment of 196 nt which is also unrelated to the basic repeat.

Quantification of HindIII Fragments in Total DNA. The relative amounts of the fragments were determined in one of two ways, either from scans of gels containing optical density amounts of DNA or by counting labeled fragments displayed on a gel. Because of the large amount of heterogeneously sized fragments in the first third of the gel near the origin, estimates of the relative amounts of the larger fragments were determined with labeled DNA. A direct estimate of the amount of the fragments smaller than fragment III was made from optical density scans of a total HindIII digest (Figure 2B, Appendix B) using scans of comparably sized AluI fragments of ØX174 DNA as standards.

If it is assumed that cleavage sites for HindIII were once spaced uniformly, a test for the random loss of sites can be performed (Slack, 1974; Southern, 1975). The proportion of each multimer (n=1 for a monomer) will be given by the probability of a run of (n-1) adjacent alterations in an array of regularly spaced restriction sites. Such a probability distribution predicts that each successively larger multimer

will be present in smaller amounts. Figure 3, Appendix B, shows the expected amounts of multimers for different degrees of divergence (dashed lines; a, 3.7%; b, 9.5%; and c, 15.7%) and the experimentally determined amounts of HindIII fragments in Geryon DNA (solid line). The relative amounts of the HindIII multimers remain the same even with increases of either enzyme concentration or time of incubation. The ratios of the amounts of the different multimers (Table 1, Appendix A) after complete digestion are different from those of various satellite DNAs (Southern, 1975; Altenberger et al., 1977; Fittler, 1977) and total DNAs (Manteuil et al., 1975; Cooke, 1975). The deviation of the Geryon fragments from a linear plot indicates that the distribution of HindIII sites (solid line, Figure 3; Appendix B) cannot be explained simply by random alteration of a formerly homogeneously spaced array of sites. The amounts of dimer, tetramer, and octamer best fit line (c) which represents 92% altered sites. This level of divergence cannot reflect the overall pattern since 62, 50, and 12% respectively of these fragments were digested to 47 and 40 nt fragments by HhaI. If a divergence of 15.7% was present in the fragments, more of the original HhaI sites would be altered.

Several explanations are possible for this nonrandom distribution of enzyme sites. Since the Geryon HindIII fragments were obtained from digests of total DNA the presence of a multimeric series based on a single repeating unit cannot be assumed from the sizes alone. It is necessary to demonstrate a sequence relationship for this series of fragments. Experiments were initiated which demonstrated that such a relationship

exists. After isolation from polyacrylamide gels, the HindIII fragments were individually labeled at the 5' termini by polynucleotide kinase. Direct sequencing of the 81 nt monomer showed that both an HhaI site and a HaeIII site were present (Part III). If the HindIII fragments represent a multimeric series based on an 81 nt repeat, digestion of the fragments with either enzyme would produce fragments whose sizes are consistent with the same restriction sites at the same position of each basic repeat of the multimers.

Analyses of the Multimers

A. The Monomer, Dimer, Trimer, and Tetramer. Initial digests of the monomer (V) and dimer (IV) with several restriction enzymes MboI, TaqI, HpaII and AluI left the majority of the DNA of each fragment undigested (data not shown). By contrast, with either HhaI or HaeIII from 50 to 74% of each fragment was digested indicating that these sites are present in only a fraction of each multimer. The digests with either enzyme contained two major and several minor fragments. HhaI digests of the monomer, dimer, trimer, and tetramer (III) are shown in Figure 4, Lanes 1, 2, 4, and 5 (Appendix B). The major products are fragments of 47 and 40 nt which represent between 30 and 60% of each digest. Direct sequencing by the chemical degradation method (Maxam and Gilbert, 1977) of these major fragments from HindIII fragments III, IV and V has shown that the same sized fragments from either the monomer, dimer, or tetramer have essentially the same sequences (Part III). The 47 and 40 nt fragments correspond to the two 5' termini of the monomer. An overlap of

four nucleotides for the HindIII site and two for the HhaI site results in the 81 nt repeating unit.

Although the reduced amounts of the trimer prevented its sequencing, sequence homology between it and the monomer, dimer, and tetramer is indicated by the presence of similarly sized major and minor fragments in its HhaI digest. Digests of the dimer, trimer and tetramer contain fragments 130 and 127 nt in length while fragments of 212 nt are also present in the trimer and tetramer. The sizes of these minor fragments correspond closely to those expected if one or two HhaI sites, respectively, were lost from a set of sites formerly spaced every 81 nt.

The proposed organization for the original spacing of HindIII, HhaI, and HaeIII sites is shown in Figure 5A (Appendix B) from which the organization of the monomer and its multiples can be deduced. HhaI fragments (Figure 5B, Appendix B) other than those of 47 or 40 nt have lost one or more sites. HhaI cleavage of a population of 5' end-labeled trimers (243 nt) would produce fragments of 47, 128, and 209 nt from the 5' labeled termini shown at the left and fragments of 40, 121, and 202 nt from the 5' labeled termini shown at the right. Any 81 nt fragment or its multiples would be internal; without a 5' labeled end they would escape detection in an autoradiogram. The presence in the HhaI digest of fragments whose sizes closely approximate those of fragments that have lost one site (128 nt from the 47 nt end and 121 nt from the 40 nt end) and two sites (209 nt from the 47 nt end and 202 nt from the 40 nt end) supports this organization. The sizes of the observed fragments are within reasonable limits of experimental variation expected for the

multimeric series. Only for the smallest pair of fragments, 47 and 40 nt, was a seven nucleotide difference in size sufficient for electrophoretic separation. In the digest of the tetramer (Figure 4, Lane 5; Appendix B) fragments of the size expected following the loss of 3 HhaI sites (290, 283 nt) may be obscured by the undigested fraction of tetramer DNA.

This undigested fraction of each HindIII fragment presumably represents sequences which have lost all the HhaI sites although a completely different repeating unit cannot currently be excluded. Since the HindIII fragments were obtained from total DNA, fragments of the same length may be derived from two or more distinct sequence classes. Alternatively, for the larger multimers the undigested fraction may be derived from background fragments of heterogeneous sizes. In autoradiograms this fraction will appear to comprise a larger amount of the total digest because it retains label at both 5' ends while each HhaI digestion product is labeled at only one end. That the undigested subset represented incomplete digestion was eliminated by increasing both the amount of enzyme and the time of incubation.

Digestion by HaeIII of the monomer and the next three consecutive multimers, all of them 5' end-labeled DNA, (Figure 6, Lanes 1, 2, 4 & 5; Appendix B) produced the predicted array of HaeIII fragments (Figure 5C, Appendix B). The pairs of fragments present in the HaeIII digests, one for each 5' end, correspond to those that have lost either no HaeIII sites (75, 10 nt), one site (156, 91 nt), two sites (237, 172 nt) or three sites (318, 253 nt). The 10 nt fragments migrated considerably

faster than the other fragments and are not shown in the figure. The digestion products recovered from HhaI or HaeIII digests of the smallest four oligomers is consistent with the original sequence having contained a repeating unit of 81 nt.

Fifty to 75% of each of the HindIII fragments thus far described is digested by either HhaI or HaeIII. That the subset of sequences containing the HhaI sites almost completely overlaps the subset containing the HaeIII sites was shown by HaeIII digestion of those fractions of the monomer or dimer which were undigested by HhaI. Very little digestion occurred. It is unlikely that these results are due to the inhibition of the restriction enzyme digestion since treatment of the fractions undigested by HhaI gave the same result with three times as much HaeIII. Rather, it appears that the majority of the subset of sequences containing HaeIII sites had been removed by HhaI digestion. This is a surprising result. Random mutation in a homogeneous population of repeating units would not be expected to lead to elimination of both sites in one-fourth to one-half of the sequences comprising one multimer and to both sites being present in the remaining sequences. This distribution of HhaI and HaeIII sites suggests two classes of sequences, one considerably more divergent than the other. We conclude that an unusually large amount of the sequences of both the monomer and dimer size classes have lost both the HhaI and the HaeIII recognition sites.

B. The Pentamer, Hexamer, and Heptamer. As seen in Table I (Appendix A) and Figure 3 (Appendix B) the fragments corresponding to the lengths expected for a pentamer, hexamer, or heptamer comprise very

small fractions of a total HindIII digest. Because of the small amounts of each of these fragments and the relatively large background of heterogeneous fragments of similar size, quantification was less precise. At least 50% of each of these fragments is digested by HhaI yielding as major digestion products the 47 and 40 nt fragments which are characteristic of the 81 nt repeating unit (Figure 7, Lanes 1, 2 and 3; Appendix B). The lengths of the minor digestion products can be due to the loss of one or more of the sites in the 81 nt spacing as for the other four multimers. Thus, the sequence organization in these fragments is similar to that in the smaller multimers.

C. The Octamer. The organization of the octamer fragments I (675 nt) and II (648 nt) was examined by digestion of co-isolated fragments with HhaI (Figure 8A, Lane 1; Appendix B) and HaeIII. The sizes of the HhaI cleavage products of co-isolated fragments were essentially the same as for digests of fragment I (Figure 8A, Lane 2; Appendix B) or fragment II (not shown). Similarly, there was no detectable difference in the number or size of the products between HaeIII digests of separated and co-isolated octamer fragments; only the HaeIII digest of fragment I is given in Figure 8A, Lane 4 (Appendix B). In the subsequent discussion, "octamer" will be used to designate DNA co-isolated as the 675 and 624 nt fragments. Both the HhaI and HaeIII digestion patterns of the fragments from the octamers are more complex than those of the smaller multimers. Apparently several sequence subsets are present and only a fraction of the DNA of the octamers is clearly related to the smaller multimers.

Although the sizes of most HhaI digestion products of the octamers can be accounted for by the organization shown in Figure 5 (Appendix B), there are two deviations from the results expected for an original repeat length of 81 nt. First, the two fragments of 80 and 142 nt do not fit the proposed organization. All the other digestion products, as indicated in Figure 8A, Lanes 1 and 2 (Appendix B) correspond to the proposed organization for the smaller multimers. Second, there is an inordinately large amount of one of the size classes. Quantification of the amount of each product in an HhaI digest of octamer fragments shows that 52% of the DNA is digested to a family of fragments that form a broad band with a mean of approximately 270 nt (Figure 8A, Lanes 1 and 2; Appendix B). This result differs significantly from the expected amount of the size class representing the loss of three adjacent HhaI sites from each 5' end. Because of the decreased probability of altering neighboring sites by random mutation, there should be larger amounts of the fragments of sizes produced by the loss of one or two adjacent sites. The very large amount of this fragment is most easily explained as a selective amplification of a sequence subset in which the divergence from the 81 nt basic repeat is greater than that in the other HindIII fragments. Alternatively, the sequence in the 81 nt repeat and the sequences in the 270 nt DNA might represent distinct variations stemming from an earlier basic repeat unit in which case there should be few homologies between these two DNAs.

The population of molecules comprising the octamers can be assigned to four groups[1]: Subset 1, restriction fragments whose sizes can be predicted from the 81 nt repeat; Subset 2, fragments of predicted sizes present in anomalously large amounts, i.e., the 270 nt class; Subset 3, fragments of unexpected sizes, i.e., the 80 and 142 nt fragments; and Subset 4, the fraction undigested by HhaI. The latter three categories may represent diversity in the types of alterations which have occurred in the basic repeat, including amplification, introduction of new HhaI sites, and the loss of the 81 nt spacing for HhaI sites. Although the precise sequence relationship of these categories to the 81 nt repeat is not precisely defined, they will be referred to collectively as the "divergent HindIII sequences." Subsets 1, 3 and 4 are also present in the smaller multimers; however Subset 3 accounts for a lower percentage of the smaller multimers.

The organization of the 270 nt fragments was investigated as follows: co-isolated octamers were 5' end-labeled, digested with HhaI, precipitated and treated with TaqI (Figure 8B, Lane 1; Appendix B). Digestion with both enzymes produced fragments whose sizes were the same as those observed when the octamers were digested by HhaI alone. A large fraction of the 270 nt fragments contain TaqI sites since very few fragments of this size are present in the double digest. Unique to the

---

[1]The assignment of fragments to Subsets is based on their relationship to the proposed organization, Figure 5, Appendix B. In Part III fragments conforming to this organization (Subset 1) will be assigned to Classes 0 to 7 depending on the number of adjacent restriction sites which have been altered by random mutation.

double digest is the presence of a 138 nt fragment. This is the major digestion product of this digest indicating that a significant fraction of octamers have TaqI sites 138 nt from either one or both of HindIII sites. Although the basic 81 nt repeating unit did not contain a TaqI site they might be expected to appear at many positions as the sequences diverge away from the 81 nt repeat. The presence of large amounts of the TaqI/138 nt fragment in the octamers can be explained in two ways: either TaqI sites are added to the basic 81 nt repeat by random divergence followed by amplification or they are an amplification of a distinctly different repeating unit. Digestion of monomer, dimer, and trimer DNAs by TaqI yields no major digestion products. The small amounts of minor fragments recovered from these DNAs can be attributed to divergence from the basic 81 nt repeat. When the decamer is digested by TaqI (Figure 8B, Lane 3; Appendix B), the 138 nt fragment is present but not as a major product. If one of the octamer fragments is related to the decamer, it is not a simple relationship (see below also).

Digestion of the octamer fragments by HaeIII corroborates the conclusions made from HhaI digestion. The sizes of the products from HaeIII digestion (Figure 8A, Lane 4; Appendix B) can be ascribed to the same categories as the products of HhaI digestion. With HaeIII digestion Subset 2, fragments of expected size but anomalously high in amount, is comprised of fragments of 350 and 320 nt. These correspond to proposed sizes of 334 and 318 nt both of which would be produced by cleavage at the same site. The next six smaller fragments in the HaeIII digest are clearly related to the proposed organization containing an

81 nt repeat, i.e., Subset 1 DNA. The remaining fragments of sizes 55 to 24 nt do not appear to be related to the basic repeating unit and are therefore Subset 3 DNA. Undigested DNA is also present (Subset 4 DNA). The major digestion products are the 24 nt fragment in Subset 3 and the two large fragments in Subset 2. A diagram is presented (Figure 9, Appendix B) for the proposed organization of Subset 2 DNA in which a cluster of sites from the basic 81 nt repeat remain. Assuming that the Subset 2 DNA from both HhaI and HaeIII digests overlap as in the monomer and dimer DNAs, the specific increases in the amounts of larger HaeIII and HhaI products strenghtens the possibility that an amplification of a sequence set somewhat divergent from the basic 81 nt repeat has occurred.

D. The Decamer. The cleavage products produced by digestion of labeled decamers with HhaI, TaqI and HaeIII were most similar to those sequences in the octamers which had diverged from the basic 81 nt repeat. In the HhaI digest (Figure 10, Lane 2; Appendix B) the fragment sizes characteristic of the 81 nt repeat (viz. the 47 and 40 nt fragments) were not seen. Also missing were fragments of the sizes that would occur if one or two adjacent sites were modified. The 270 nt fragments that comprise the bulk of the HhaI digest of the octamers account for only a minor fraction of this digest. The 80 and 142 nt fragments of the divergent Subset 3 DNA are also present. Fragments of 165, 138 and 88 nt are recovered both from TaqI digests of decamers and from HhaI/TaqI double digests of octamers. All TaqI fragments belong to the divergent Subset 3 DNA since no TaqI sites are in the basic repeat. HaeIII digestion of decamer DNA (Figure 10, Lane 4; Appendix B) did not yield the

pairs of fragments expected for Subset 1 DNA. Rather, a fragment of 55 nt (Subset 3) was observed as a major fragment. In summary, many similarities between the decamer and the octamer appear to be in the subsets of sequences that are not related to the basic 81 nt repeat.

CHAPTER IV

DISCUSSION

The distribution of amounts of DNA in the multimeric series of HindIII fragments digested from total Geryon DNA deviates from the typical distribution of a monomer and its multiples as observed in isolated satellite DNAs (Southern, 1975; Altenberger et al., 1977; Fittler, 1977) and in satellite DNAs digested from total DNAs (Manteuil et al., 1975; Cooke, 1975). This organization could not have arisen from random mutation alone but rather is indicative of selective amplifications of subsets of the DNA. For selective mutation to give these results would require a complex set of constraints in which particular sites in a linear array would be protected from mutation (See Chapter IV, Part III). Selective amplification is simpler to envisage and could occur in at least two ways. First, amplification of a particular subset, representing a relatively small fraction of the total HindIII sequences, might occur after its translocation to a different chromosome. Alternatively, only a small fraction of a repeated sequence at a single location might be selected for amplification. Since the sequences are digested from total DNA it cannot be assumed that they are contiguous. Rather, they may be interspersed with single copy DNA which has been shown to abut two highly repetitive components of very different frequencies (Christie and Skinner, 1979). In addition, if these sequences were contiguous a difference in density would be expected since their {G+C} content is 52%. This base composition corresponds to a density of 1.72 g/cm^3 compared to 1.70 g/cm^3 for the total DNA.

There could also be several types of amplification. It could occur in a subset which had diverged from the basic repeat by a saltatory mechanism (Britten and Davidson, 1971) or by an unequal crossover mechanism (Smith, 1973; Chambers et al., 1978). In either case to account for the relative amounts of the HindIII multimers it must be stipulated that only a small fraction of related repetitive sequences has undergone amplification. Either mechanism could increase a divergent set of sequences by amplifying a small subset, e.g., several thousand nucleotides. The apparent amplification of a divergent octamer indicates that the amplified segment is greater than 648 nt though several thousand nucleotides might be required for extrachromosomal replication by a rolling circle mechanism. A segment of this size would still be small enough such that its ratio of multimer classes would not reflect that of the entire population of sequences. Divergence of the original sequence and subsequent amplification of the entire class would preserve ratios expected for simple random inactivation of regularly spaced restriction endonuclease sites. Some form of selective amplification must be postulated to explain the observed ratios in this DNA.

The picture that emerges from a study of Geryon DNA for the formation of highly repetitive sequences is a cyclic pattern of amplification followed by divergence followed by reamplification of the divergent sequences. This assessment is based on (1) the enhanced amounts of some multimers and (2) the presence of large amounts of divergent sequences. In addition to enhanced amounts of particular fragments in digests of octamers, all multimers contain unexpectedly

large amounts of fragments that are resistant to subsequent digestion with either HhaI or HaeIII. Loss of the sites by random changes alone would have yielded a much smaller fraction of undigested fragments. Except for the limitations which can be placed on the size of the segment selected for amplification, this evolutionary pattern is as described by Southern for guinea pig α-satellite (1970). In contrast, Fry and Salser (1977) have suggested that a library of satellite sequences are conserved by their association with genetic mechanisms fostering sympatric speciation[1]. Chromosomal rearrangements, especially those facilitated by heterochromatic regions, have also been implicated in speciation (White, 1969). Conservation of specific sequences is probably not necessary for a role in sympatric speciation. Heterochromatin and thus the highly repetitive DNAs localized to this region may facilitate chromosomal rearrangements which increase the potential for variant species. Increases in genetic variability are directly related to increased survival of a particular taxonomic group and that survival will perpetuate a class of highly repeated DNAs. In addition, it is not necessary that the amplification of a particular DNA sequence initiates speciation but only that it fosters it.

Out data indicate a pattern of multiple amplifications of relatively small size compared to the complete set of related sequences.

---

[1]In the sympatric model of evolution, two biological races adapt to different ecological niches within the same geographic range and form distinct species. Alternatively, in the allopatric model, geographic isolation precedes genetic isolation.

A single amplification would have produced the observed distribution of HindIII fragments only if a contiguous piece of DNA contained two octamers for every tetramer and no trimers, pentamers, hexamers, or heptamers. This is quite unlikely since the pre-amplified number of octamers would be very small. Small amplifications might be fairly frequent while an amplification of sufficient magnitude to separate species might be relatively rare. Multiple amplifications of the highly repetitive regions would have the same effect as conserving particular sequences.

LIST OF REFERENCES

LIST OF REFERENCES

Altenberger, W., W. Horz, and H. G. Zachau. 1977. Eur. J. Biochem. 73: 393-400.

Britten, R. J. and E. H. Davidson. 1974. Quart. Rev. Biol. 46: 111-138.

Britten, R. J., D. E. Graham, and B. R. Neufeld. 1974. Methods in Enzymology. L. Grossman and K. Moldave, Eds. Vol. 29E. Academic Press, New York, pp. 363-418.

Chambers, C. A., M. P. Schell, and D. M. Skinner. 1978. Cell 13: 97-110.

Christie, N. T. and D. M. Skinner. 1979. Nucleic Acids Res. 6: 781-796.

Cooke, H. J. 1975. J. Mol. Biol. 95: 87-99.

Cordeiro-Stone, M. and C. S. Lee. 1976. J. Mol. Biol. 104: 1-24.

Fittler, F. 1977. Eur. J. Biochem. 74: 343-352.

Fry, K. and W. Salser. 1977. Cell 12: 1069-1084.

Holland, C. A. and D. M. Skinner. 1977. Chromosoma 63: 223-240. Berlin.

Maniatis, T., A. Jeffrey, and H. van deSande. 1975. Biochemistry 14: 3787-3794.

Manteuil, S., D. H. Hamer, and C. A. Thomas, Jr. 1975. Cell 5: 413-422.

Musich, P. R. and D. M. Skinner. 1972. J. Cell Biol. 55: 184a.

Prunell, A., F. Strauss, and B. Leblanc. 1977. Anal. Biochem. 78: 57-65.

Sanger, F., G. M., Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, and M. Smith. 1977. Nature 265: 687-694. London.

Schroeder, J. L. and F. R. Blattner. 1978. Gene 4: 167-174.

Skinner, D. M. 1977. BioScience 27: 790-796.

Slack, J. 1974. Biopolymers 13: 2241-2264.

Smith, G. 1973. Cold Spring Harbor Symp. Quant. Biol. 38: 507-513.

Southern, E. M. 1970. Nature 227: 794-798. London.

Southern, E. M. 1975. J. Mol. Biol. 94: 51-69.

White, M. J. D. 1969. Annual Review of Genetics. H. L. Roman, L. M. Sandler, and A. Campbell, Eds. Vol. 3. Annual Reviews, Palo Alto, pp. 75-98.

Vaughn, J. C. 1975. Chromosoma 50: 243-257. Berlin.

Vereijken, J. M., A. D. M. van Mansfield, P. D. Bass, and H. S. Jansz. 1975. Virology 68: 221-233.

APPENDICES

APPENDIX A

TABLES

TABLE 1

ESTIMATE OF THE AMOUNT OF HINDIII FRAGMENTS IN THE GENOME

HINDIII FRAGMENT (nt)	RATIO OF COPIES/GENOME[a]	PERCENT OF THE GENOME[b]	COPIES/ GENOME[c]
81	1.00	0.25	1.2×10^5
162	0.98	0.50	1.2×10^5
243	0.10	0.075	1.2×10^4
324	0.35	0.35	4.2×10^4
405	0.005	0.006	6.0×10^2
486	0.005	0.008	6.0×10^2
567	0.008	0.014	9.6×10^2
648	0.75	1.50	9.0×10^4
675	0.75	1.50	9.0×10^4
810	0.47	1.20	5.6×10^4

a. Determined from the amount of radioactivity in each fragment relative to that in monomer DNA after 5' end-labeling of a HindIII digest of total Geryon DNA.

b. Determined from optical density scans of HindIII digestion products separated in 5% polyacrylamide tube gels and stained with ethidium bromide, using AluI fragments of ØX174 DNA as a standard.

c. Calculated from the percent of each fragment in the genome, length of the fragment, and a genome size for Geryon determined by reassociation kinetic data (Britten et al., 1974). Using a ratio of the rate of reassociation for Geryon single copy DNA (Christie and Skinner, 1979) to that of Gecarcinus lateralis (Holland and Skinner, 1977) and the genome size of G. lateralis determined independently (Musich and Skinner, 1972) the size of the Geryon genome was calculated to be 2.4 pg. Both single copy rates were computer-derived values of the expected rates of isolated single copy DNA as determined from reassociation of total DNA.

TABLE 2

VARIATIONS IN ESTIMATES OF FRAGMENT SIZE
IN 7% POLYACRYLAMIDE GELS

FRAGMENT NUMBER	KNOWN FRAGMENT SIZE[1] (nt)	DEVIATIONS FROM KNOWN SIZES FOR DIFFERENT ELECTROPHORETIC RUNS[2]					
		1	2	3	4	5	6
1	726	- 4	-42	-44	-29	-35	-34
2	716	-56	-77	-86	-84	-81	-66
3	550	+45	+20	+26	+24	+34	+38
4	500	+10	+14	+14	+12	+15	+12
5	417	-24	+39	+35	+37	+34	+34
6	311	+13	+19	+25	+18	+13	+ 5
7	240	+10	+13	+18	+16	+12	+ 9
8	200	- 1	- 1	- 2	- 1	- 9	- 6
9	151	+ 2	+ 1	+ 1	-10	+ 4	- 5
10	139	- 1	- 3	- 4			
11	118	- 3	- 6	- 6			
9	151				0	- 2	- 1
10	139				0	- 2	- 1
11	118	- 4	- 4	- 8	0	0	- 1
12	101	- 1	- 2	- 4	- 1	0	0
13	82	+ 2	+ 3	+ 1	+ 1	0	+ 2
14	66	+ 2	+ 2	+10	0	- 1	+ 1
15	48	+ 1	+ 1	- 1	0	0	- 1
16	42	0	- 1	- 1	0	- 1	- 1
17	24	- 1	0	0			

[1]HinfI digestion products of ØX174 DNA.

[2]See Appendix C for details of the procedure.

TABLE 3

VARIATIONS IN ESTIMATES OF FRAGMENT SIZE IN 5% POLYACRYLAMIDE GELS

FRAGMENT NUMBER	KNOWN FRAGMENT SIZE[1] (nt)	DEVIATIONS FROM KNOWN SIZES FOR DIFFERENT ELECTROPHORETIC RUNS[2]			
		1	2	3	4
1	726	+30	- 2	+29	+41
2	716	-52	-31	-37	-60
3	550	+19	+32	+11	+17
4	500	- 7	+ 7	- 7	-27
5	417	+ 3	-11	+ 4	+24

[1] HinfI digestion products of ØX174 DNA.

[2] See Appendix C for details of the procedure.

APPENDIX B

FIGURE LEGENDS AND FIGURES

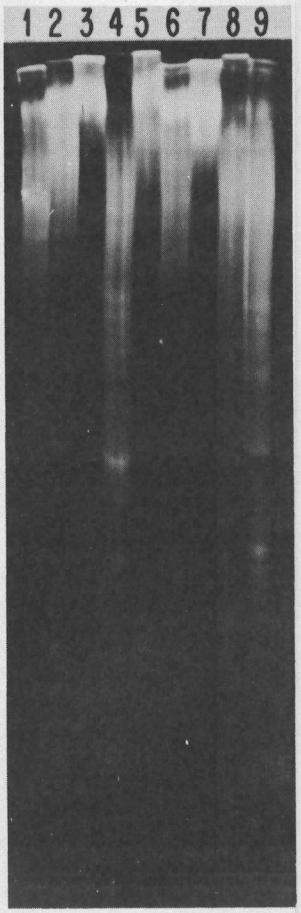


Figure 1. Gel electrophoresis of restriction endonuclease digests of total Geryon DNA.

Samples of DNA (6 µg) were digested with either (1) HhaI, (2) TaqI, (3) SmaI, (4) HinfI, (5) HgaI, (6) HpaII, (7) BamI, (8) HaeIII or (9) MboI. The digestion products were electrophoresed on 5% polyacrylamide tube gels and stained with ethidium bromide.

Figure 2. HindIII digest of total Geryon DNA.

A. HinfI digests of ØX174 DNA (Lanes 1,3) and HindIII digests of Geryon DNA (Lanes 2,4) were labeled as described and electrophoresed on a 5% (Lanes 1,2) or a 7% (Lanes 3,4) polyacrylamide slab gel.

B. A HindIII digest of Geryon DNA was electrophoresed on a 5% polyacrylamide tube gel. The gel was stained with ethidium bromide, photographed, and scanned.

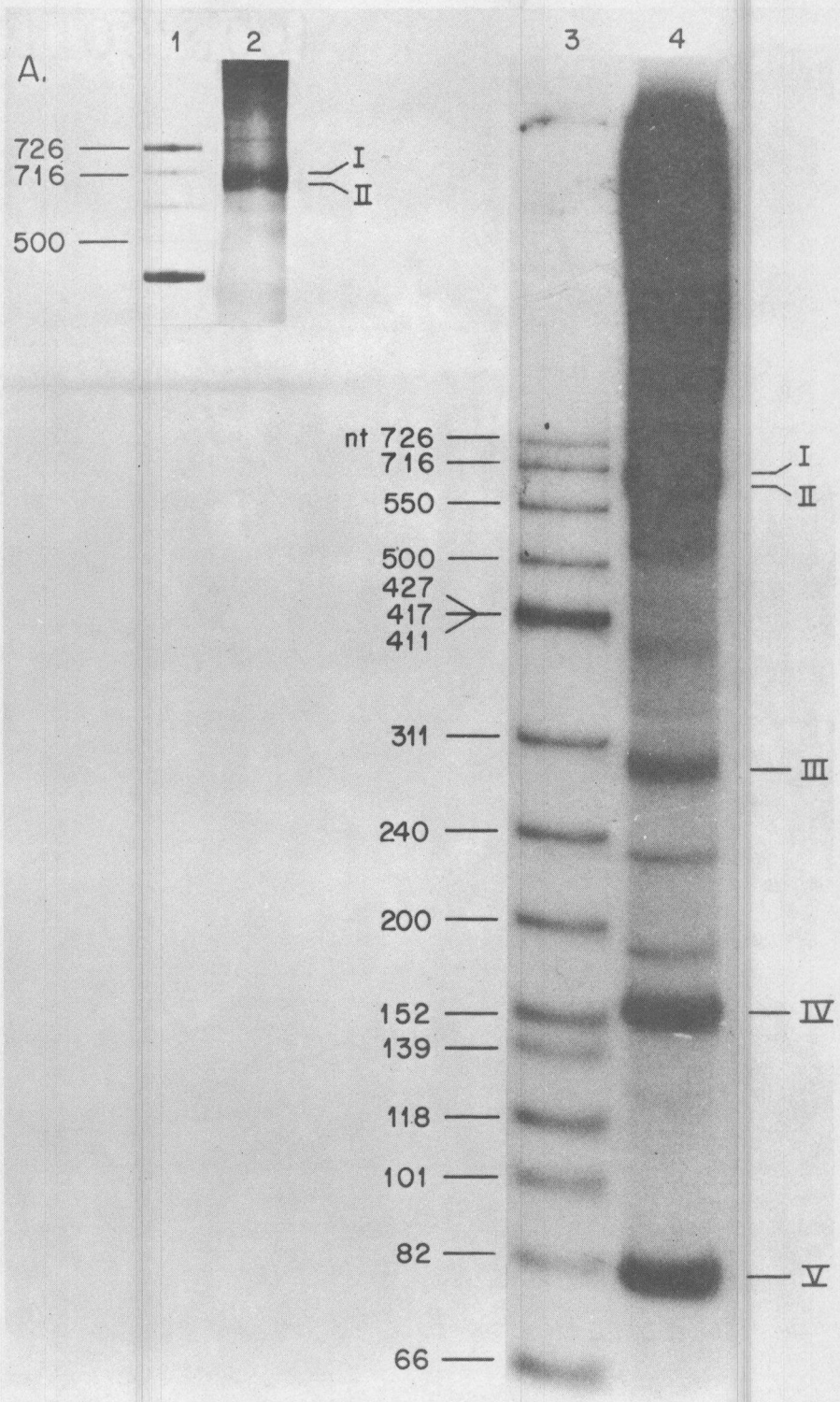

A.
1
2
3
4
726
716
500
I
II
nt 726
716
550
500
427
417
411
311
240
200
152
139
118
101
82
66
I
II
III
IV
V

Figure 2

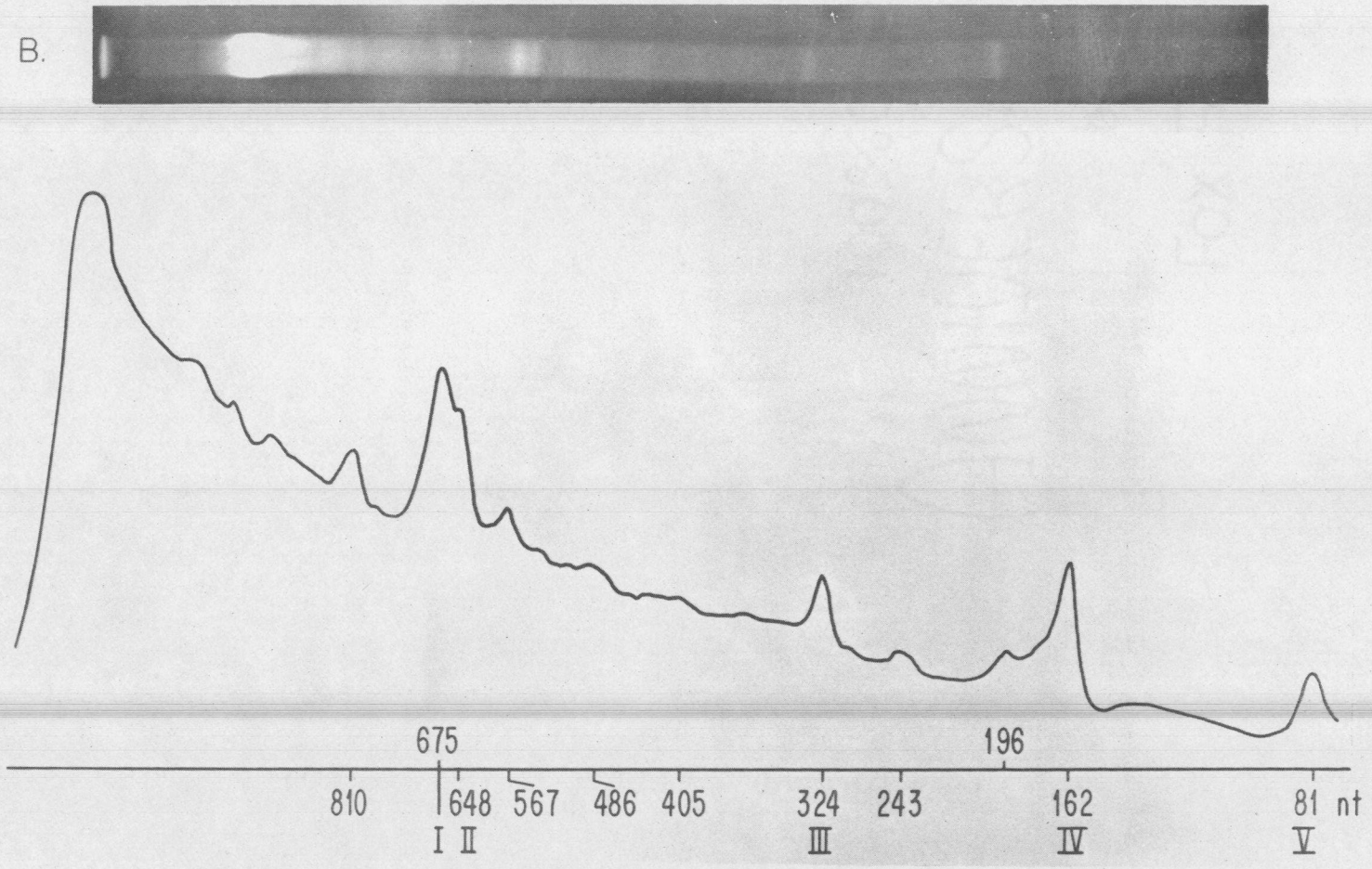


Figure 2 (continued)

Figure 3. Test for random inactivation of HindIII sites in total DNA.

If it is assumed that a regularly spaced array of restriction sites is altered by random divergence to give dimers, trimers, and larger multimers of the basic unit, the fraction f of DNA in each multimer is given by

$$f = n(1-p)^2 \, p^{(n-1)}, \quad \text{or}$$

$$\log f/n = 2 \log (1-p) + (n-1) \log p,$$

where $n = 1$ for monomers, etc., and p is the fraction of modified sites. The plot of $\log f/n$ versus $n - 1$ is a straight line when sites are inactivated at random, assuming that a change of one base pair prevents enzymatic cleavage. The slope is given by $\log p$ and the intercept by $2 \log (1-p)$. The term f/n has been replaced by c_n, the number of copies of each multimer relative to the monomer. The latter term is determined directly from ratios of radioactivity in each multimer.

The dashed lines are three theoretical plots for different degrees of divergence in a once uniform repeat; (a) 3.7%, (b) 9.5%, (c) 15.7%. Divergence is defined as the fraction of altered base pairs and was calculated according to Southern (1975) for the following fractions of modified sites, (a) 22%, (b) 56% and (c) 92.5%. The solid line gives the relative proportions of DNA determined experimentally for HindIII fragments of total Geryon DNA (see Table 1, Appendix A).

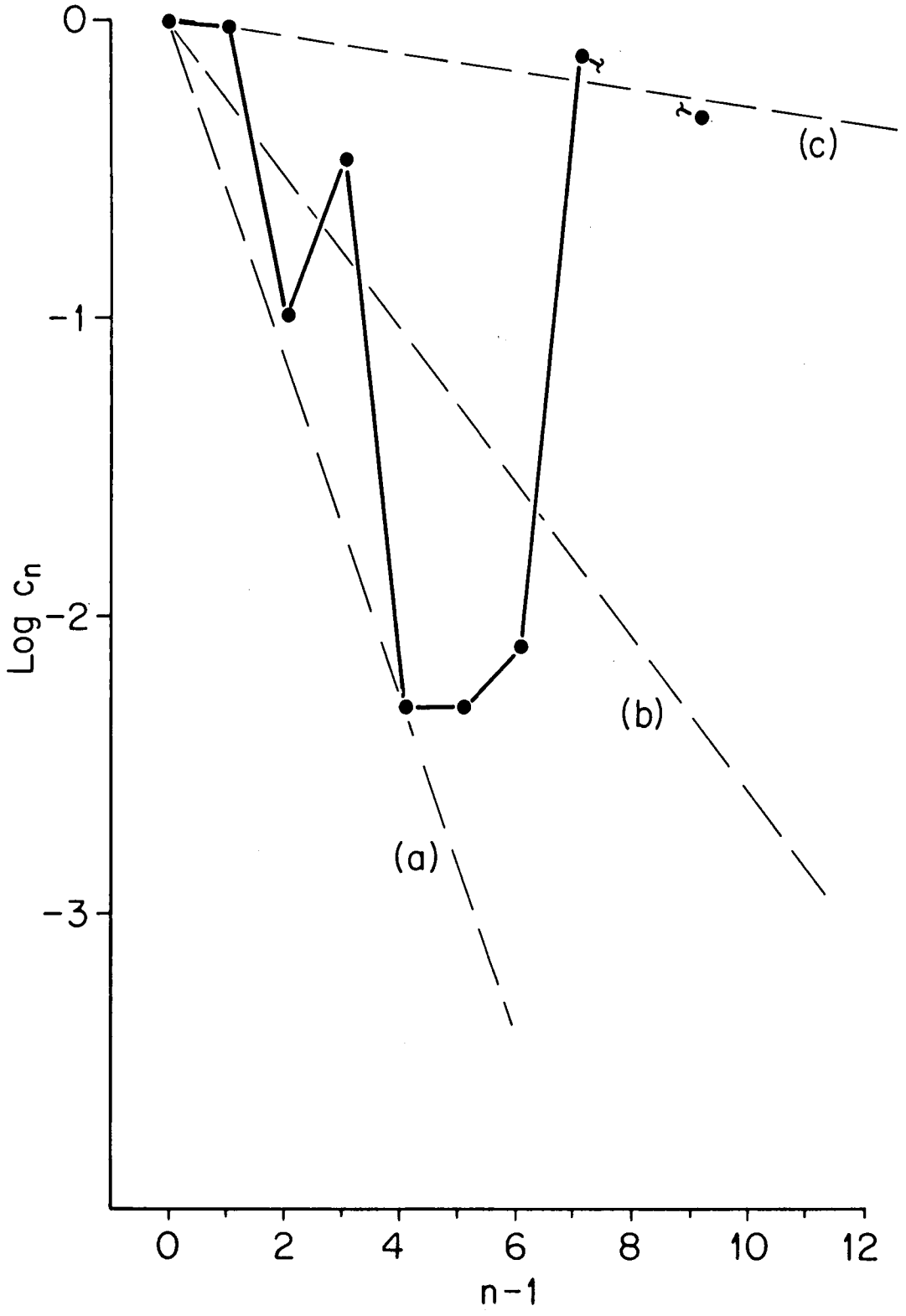


Figure 3

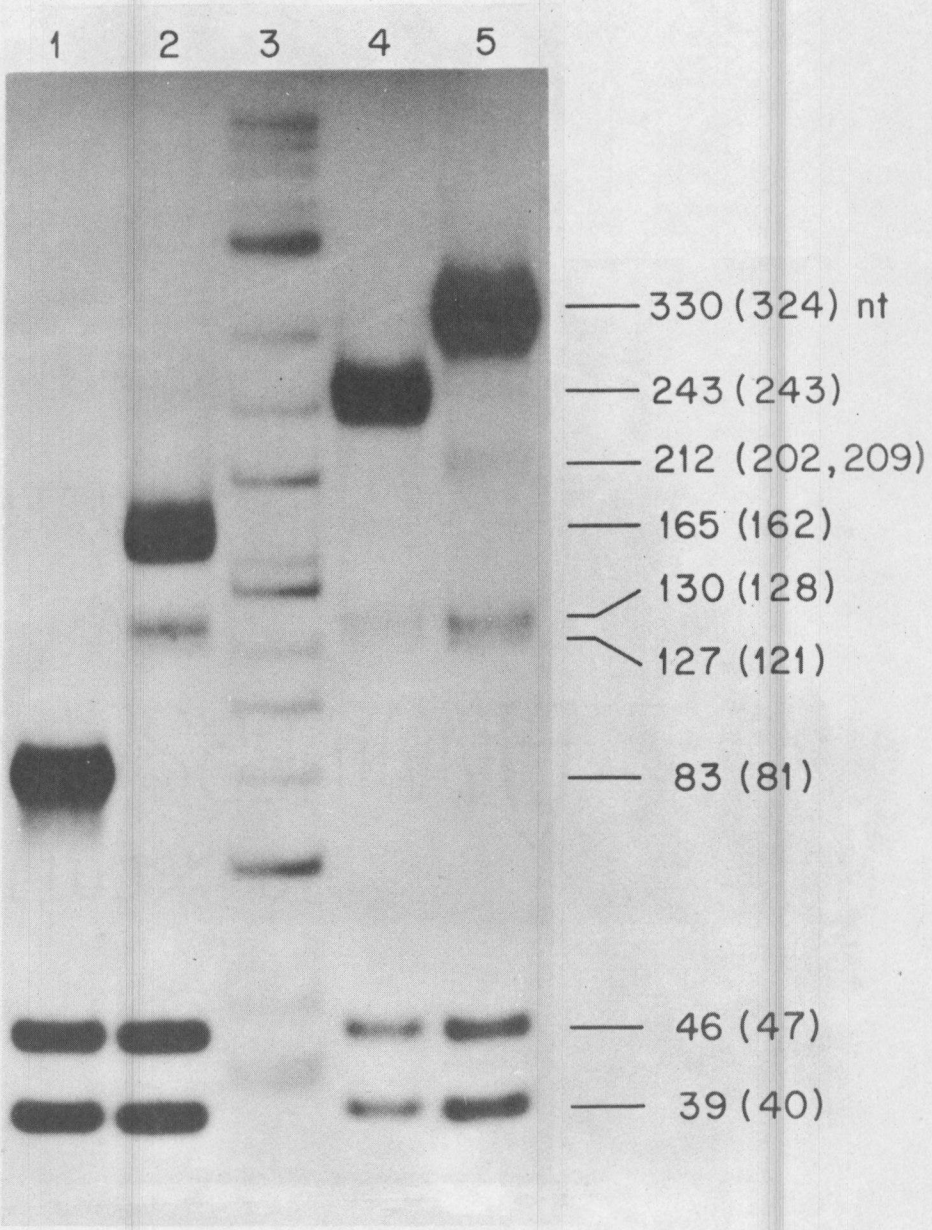


Figure 4. Autoradiogram of HhaI digestion products of monomer, dimer, trimer, and tetramer DNA.

Products of HhaI cleavage of 5' end-labeled monomer (Lane 1), dimer (Lane 2), trimer (Lane 4) and tetramer (Lane 5) DNA were electrophoresed on a 7% polyacrylamide gel. A HinfI digest of restriction fragments of ØX174 was included (Lane 3) to permit determination of fragment sizes. Measured sizes are given first with proposed sizes (Figure 5, Appendix B) in parentheses.

Figure 5. Proposed organization for the original spacings of restriction endonuclease sites in the Geryon sequences isolated by digestion with HindIII.

A. A diagrammatic representation of the sites for restriction endonucleases, HindIII (0), HhaI (X), and HaeIII (▲), in the sequences proposed as the original amplified DNA. At each HindIII site the size of a multimer is given in nucleotides (nt) as measured from the left of the diagram and assuming the internal HindIII sites have been inactivated.

B. The array of expected restriction fragments after HhaI cleavage of 5' end-labeled HindIII fragments which arise due to random divergence from the homogeneous repeating unit. Fragments missing more than 5 sites are not shown but may be present in undetectable quantities in digests of appropriate multimers.

C. The array of expected restriction fragments for HaeIII cleavage of 5' end-labeled HindIII fragments as for HhaI sites in B. In the text the arrays in (B) and (C) are referred to as Subset 1 fragments, i.e., those whose presence can be explained by random divergence.

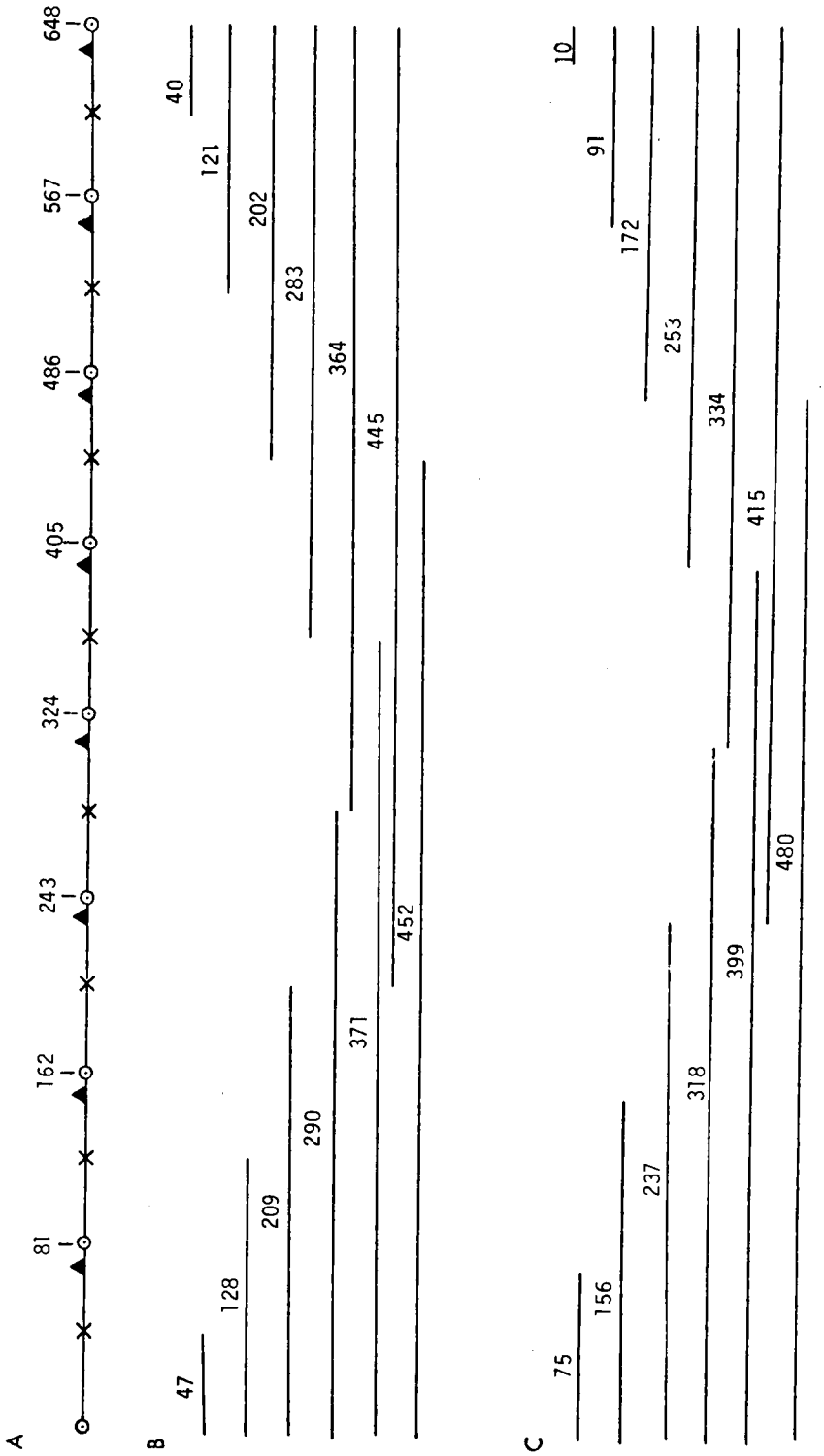

A
81
162
243
324
405
486
567
648
B
47
128
209
290
371
452
40
121
202
283
364
445
C
75
156
237
318
399
480
10
91
172
253
334
415

Figure 5

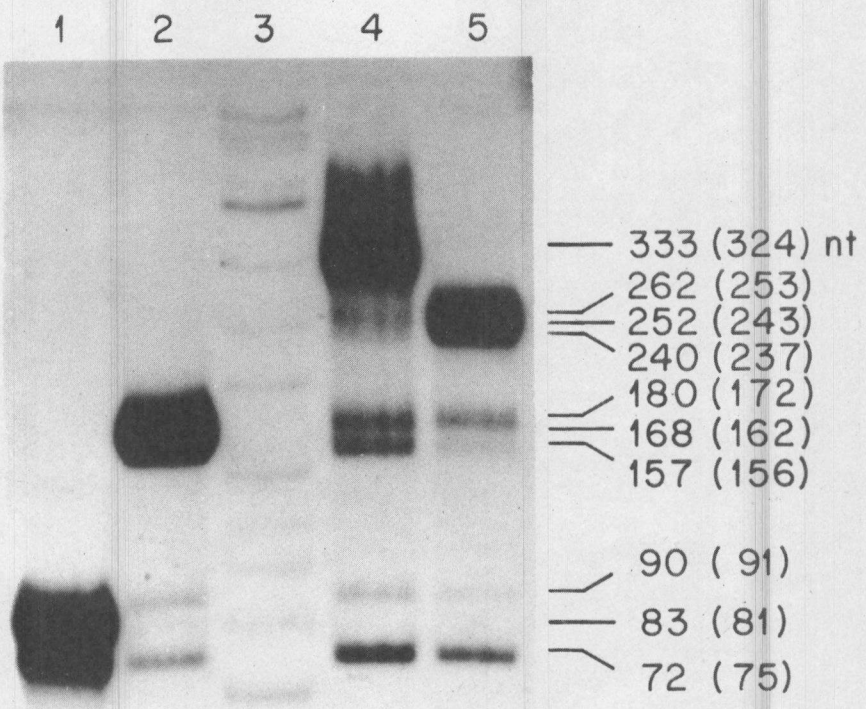


Figure 6. Autoradiogram of HaeIII digestion products of monomer, dimer, trimer, and tetramer DNA.

Products of HaeIII cleavage of 5' end-labeled monomer (Lane 1), dimer (Lane 2), tetramer (Lane 4), and trimer (Lane 5) DNA were electrophoresed on a 10% polyacrylamide gel. A HinfI digest of ØX174 DNA was included (Lane 3). Proposed sizes (Figure 5, Appendix B) are given in parentheses.

Figure 7. Autoradiogram of HhaI digestion products of pentamer, hexamer, and heptamer DNA.

Products of HhaI cleavage of 5' end-labeled pentamer (Lane 1), hexamer (Lane 2), and heptamer (Lane 3) were electrophoresed on a 7% polyacrylamide gel. For sizing a HinfI digest of ØX174 DNA was included (Lane 4). Proposed sizes (Figure 5, Appendix B) are given in parentheses.

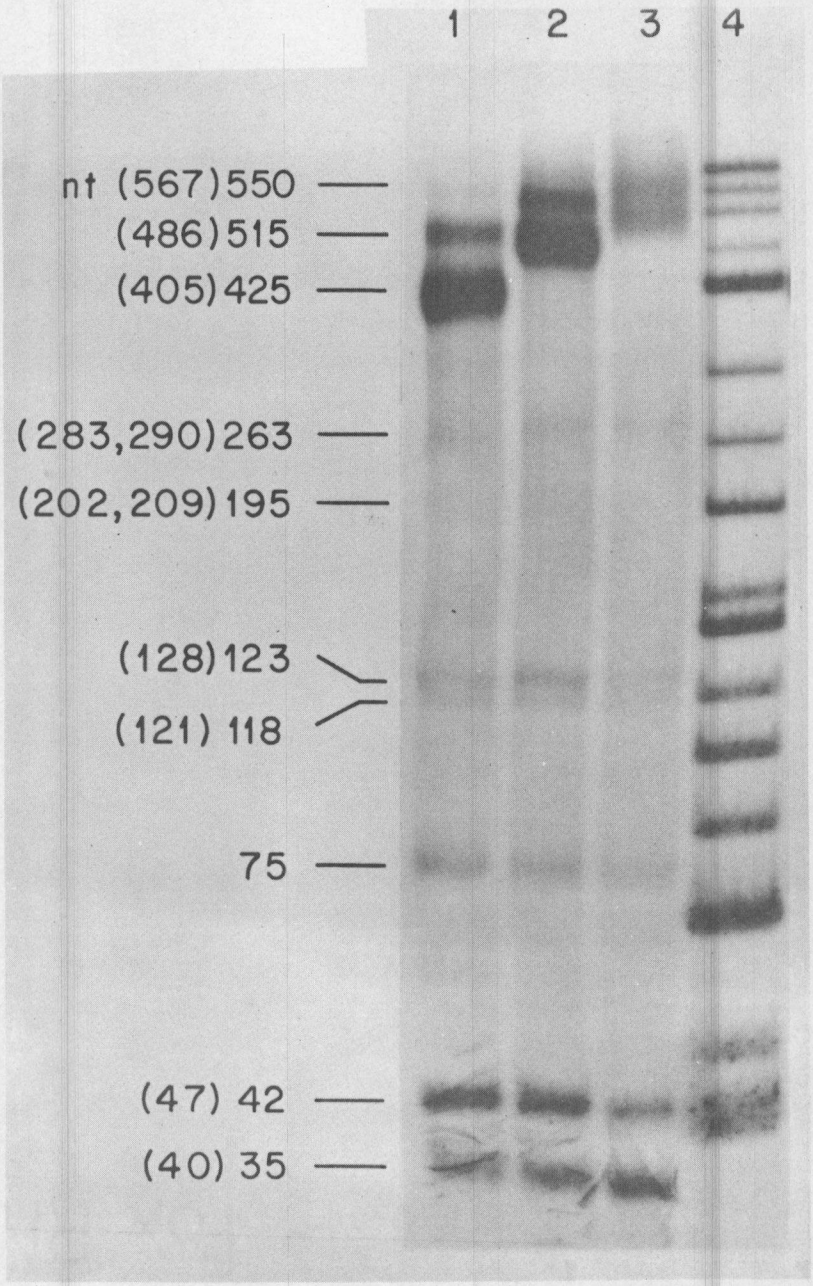

1
2
3
4
nt (567)550
(486)515
(405)425
(283,290)263
(202,209)195
(128)123
(121) 118
75
(47) 42
(40) 35

Figure 7

Figure 8. Restriction endonuclease analysis of octamer and decamer DNA.

HhaI digestion products of co-isolated octamer fragments (Lane 1) were electrophoresed on a 7% polyacrylamide gel and HhaI digestion products of fragment I (Lane 2) were electrophoresed on a 5% polyacrylamide gel. ØX174 DNA digested by HinfI (Lane 3) was co-electrophoresed on a 5% polyacrylamide gel with HaeIII digestion products of fragment 1 (Lane 4). Samples of octamer DNA digested first with HhaI, then with TaqI, (Lane 5); ØX174 DNA digested with HinfI (Lane 6); and decamer DNA digested with TaqI (Lane 7) were co-electrophoresed on a 7% polyacrylamide gel. Proposed sizes (Figure 5, Appendix B) are given in parentheses.

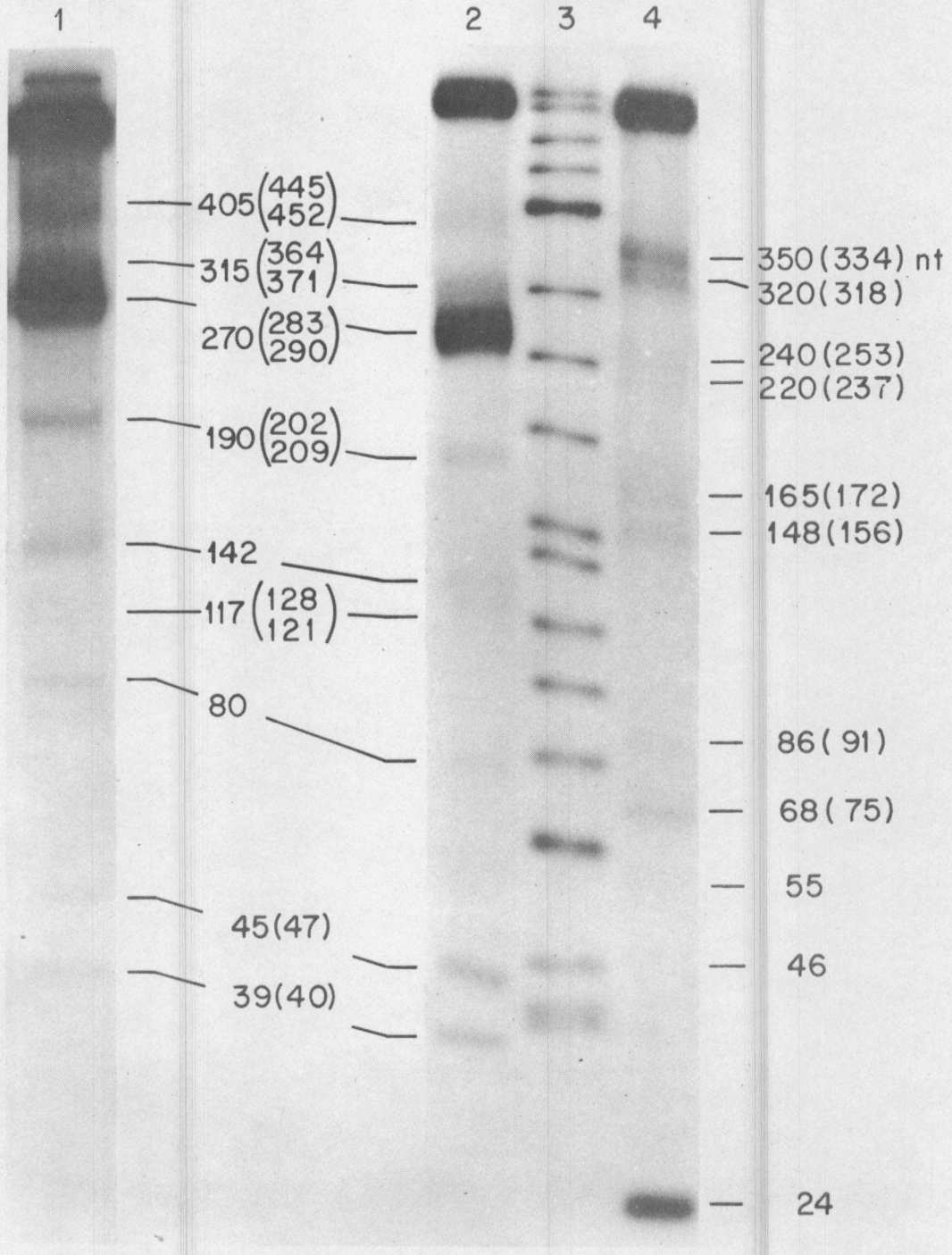

A.
1
2
3
4
405 (445 452)
315 (364 371)
270 (283 290)
190 (202 209)
142
117 (128 121)
80
45(47)
39(40)
350(334) nt
320(318)
240(253)
220(237)
165(172)
148(156)
86(91)
68(75)
55
46
24

Figure 8

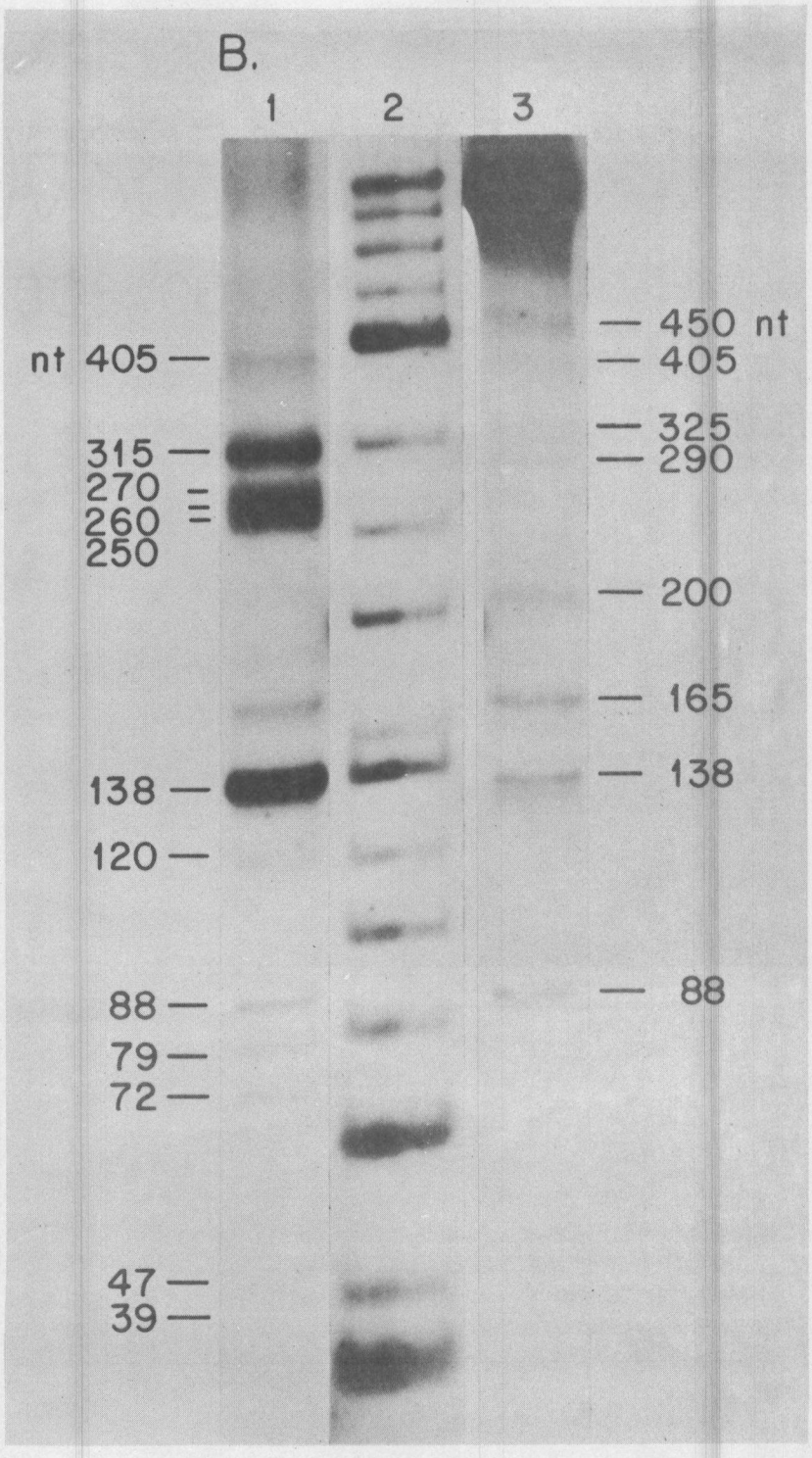


Figure 8 (continued)

Figure 9. Proposed organization of Subset 2 fragments obtained from HhaI cleavage of octamers.

A. The proposed organization of the initial spacing of restriction endonuclease sites is shown. Symbols are given in Figure 5, Appendix B.

B. The proposed organization of Subset 2 DNA is given which is based on analyses discussed in the text.

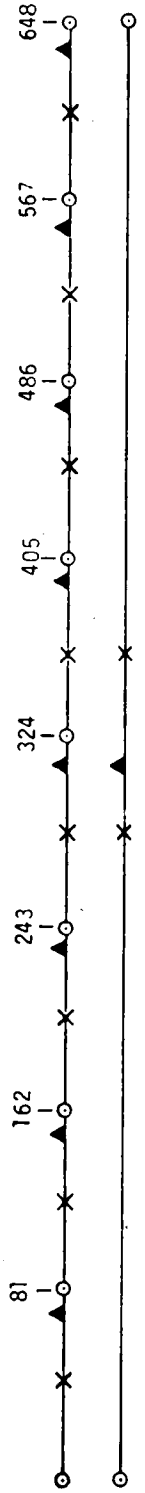


Figure 9

Figure 10. Autoradiogram of restriction endonuclease analyses of decamer DNA.

HinfI fragments of ØX174 (Lane 1) were co-electrophoresed with HhaI digestion products of 5' end-labeled decamer DNA (Lane 2) on a 7% polyacrylamide gel. A sample of the same control DNA (Lane 3) was co-electrophoresed on a 10% polyacrylamide gel with HaeIII digestion products of the labeled decamer (Lane 4).

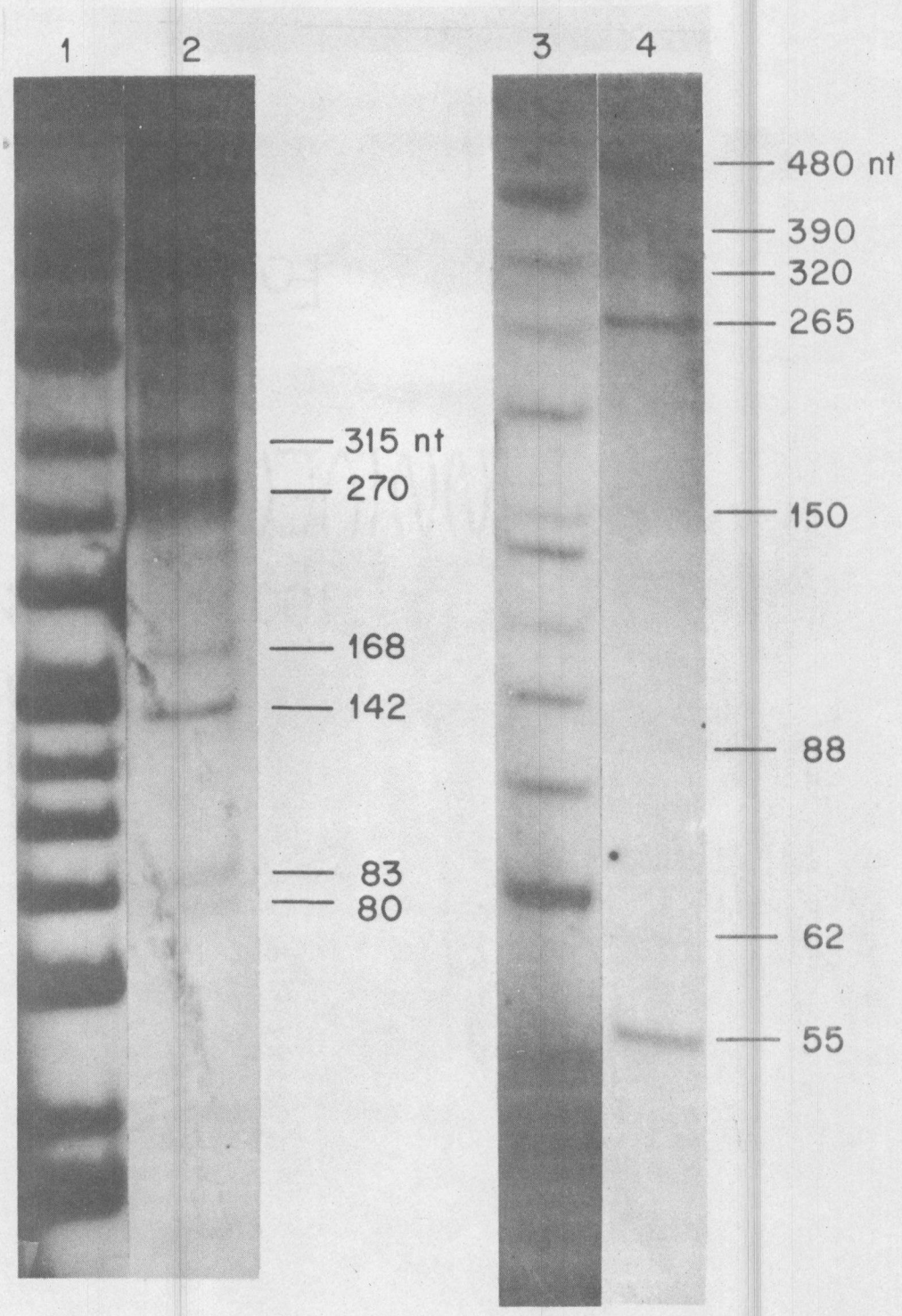


Figure 10

APPENDIX C

EXTENDED PROCEDURES

Preparative Isolation of Geryon HindIII Fragments

Procedure:

1. Concentration of DNA. DNA solutions were precipitated with ethanol to concentrate before application to a Sephadex G-100 column, 2.5 × 14.5 cm. This chromatographic step was used to remove calcium phosphate salts obtained in the isolation of the DNA by the MUP procedure. The EDTA used to inhibit nucleases during isolation convert the crystalline form of calcium phosphate known as hydroxyapatite (HAP) to other forms which coprecipitate with DNA and inhibit digestion by restriction enzymes. If necessary, fractions from the column are concentrated by ethanol precipitation. Attempts to remove these salts by extensive dialysis against 0.1 M EDTA did not produce satisfactory results. Digestion of the dialysed samples after concentration by ethanol precipitation was incomplete.
2. Analytical Digest. Test digests were performed using only 10 to 15 µg DNA to determine effective reaction conditions for DNA and enzyme concentration. The concentration of DNA in the reaction ranged from 30 to 200 µg/ml depending on the purity of the DNA sample. For those samples removed from HAP at high DNA concentration the contamination by salts of calcium phosphate was less. The enzyme concentration in the reaction was determined empirically for each enzyme preparation. Other reaction requirements for HindIII were 6 mM Tris-HCl (pH 7.4), 6 mM $MgCl_2$, 60 mM NaCl, 6 mM mercaptoethanol and 100 µg/ml bovine serum albumin. Digests were performed at 37° for 1 hour.

3. Preparative Digest. Larger amounts of DNA were digested using conditions identical to the analytical digests. Between five and six milliliters of final reaction volume were used for each Corex tube. Ethanol precipitation was then performed in the same tube. Samples were spun at 9,000 rpm for 1 hr. The total amount of DNA digested is determined by the number of gels which can be electrophoresed and processed in one day. Typically a total of 6 to 10 mg of DNA was prepared and digested over several days.

4. Preparative Electrophoresis. Pellets were taken up in electrophoresis buffer (90 mM Tris-borate, 2.5 mM EDTA). The samples were adjusted to 10% glycerol. No more than 50 µg was loaded on each well (9 × 1.2 mm) of a 7% polyacrylamide gel. Electrophoresis was preformed at 350 volts for approximately four hours. A sample containing no DNA but with 0.025% each of xylene cyanol ff and bromphenol blue was loaded in one well of each gel. The bromphenol blue marker runs slightly in front of the HindIII/81 nt fragment.

5. Excision of DNA Fragments from the Gel. The gel is stained with ethidium bromide (3 µg/ml) and viewed with a UV lamp. The bands of DNA are marked on a piece of plastic wrap which has been stretched tautly over the gel. Safety glasses should be worn for this step and also a full face mask if more than a minute is required. With the UV lamp off, bands are cut from the gel with a sharp scalpel or razor blade. Bands corresponding to a single size are pooled in a siliclad tube. These can be stored at -20° until electroelution.

6. Electroelution. Gel strips pooled for each fragment size are cut to a size small enough to allow easy packing into a 10 cm segment of glass tubing (6 mm OD, 4.5 mm ID) which has a 1.5 cm plug of 10% polyacrylamide. After the tubing was filled to a height of approximately 6 or 7 cm, 7% acrylamide was added and polymerized around the loose gel strips. Electrophoresis was performed in the cold room at 200 to 250 volts, using the buffer in No. 4 above. The time of electroelution was varied from 1.5 hours for an 81 nt fragment to 5.5 or 6 hours for fragments between 600 and 800 nt. The DNA is precipitated by addition of 2.1 to 2.3 volumes of ethanol and centrifuged at 45,000 rpm in an SW 50.1 rotor, at 0°.

Isolation of ØX174-am^3 RF

Medium:

3XD

Na_2HPO_4	2.1 g
KH_2PO_4	0.9 g
NH_4Cl	1.0 g
$MgSO_4 \cdot 7H_2O$	0.3 g
Casamino acids	15.0 g
Gelatin	10.0 mg
0.1 M $CaCl_2$	3.0 ml
Glycerol	30.0 ml

H_2O to 1 liter

Add thymidine to a final concentration of 4 μg/ml.

Solutions:

1. Sterile $CaCl_2$, 0.2 M.
2. Chloramphenicol, 10 mg/ml. Dissolve in absolute ethanol; then dilute to 50% with glass distilled water.
3. Sucrose, 10%, in 0.05 M Tris, pH 8.0.
4. Lysozyme. Make fresh each time.
5. NET

   0.100 M NaCl

   0.002 M EDTA (disodium ethylene diamine tetraacetate)

   0.050 M Tris, pH 8.0
6. T-E

   0.010 M Tris, pH 8.0

6. T-E (Continued)

   0.001 M EDTA

7. RNase (pancreatic), 5 mg/ml
8. Sarkosyl, 10%
9. Phenol, equilibrated with borate
10. Sodium acetate, 3 M
11. CsCl
12. Ethidium bromide, 5 mg/ml
13. Sodium dodecyl sulfate, 20%
14. KCN, 0.1 M

Procedure:

1. Start innoculum of E. coli cells (H502) in 10 ml of 3XD media + thymidine (4 μg/ml) from an agar slant stored at room temperature or from cells frozen in 17% glycerol stored at -20°.
2. Let stand at 37° until culture reaches an absorbancy 0.1 at 650 nm. Then begin shaking to aerate culture and grow overnight. (Since aeration is important, fill only 1/2 volume of Erlenmeyer flask with medium, e.g., 1 liter (l) medium in a 2 l flask.)
3. Use a 10 ml culture to start a larger innoculum (~150 mls for 2 l culture). The number of cells added to the medium should give a final concentration of 3×10^7 cells/ml.

   $1.0\ A_{650} = 10^9$ cells/ml

   $0.1\ A_{650} = 10^8$ cells/ml

   Note: Growth in successive stages (10 ml to 150 ml to 2 l) is presumably faster than overdiluting original 10 ml innoculum into the 2 l volume.

4. Growth of large culture. H502 doubling time is 2 to 2.5 hr until the A_{650} reaches 0.5-.6. The time required to grow from A_{650} of 0.1 to 0.7 is 6 to 7.5 hr. When a large culture is required, approximately one-half the final desired volume may be cultured overnight (~8-10 hrs). The A_{650} after overnight culture is usually 0.6-0.7. After diluting culture by one-half, an A_{650} of 0.5 can be reached in ~2 to 2.5 hr.

5. Initiate growth of phage after culture reaches A_{650} of 0.5.
   a. Add 10 ml 0.2 M $CaCl_2$ per liter. Swirl during addition to avoid precipitation of calcium phosphate salts.
   b. Add ϕXam[3] (CsCl purified) diluted with 5 ml 3XD medium. Use a multiplicity of 5-10 phage particles per cell.
   c. After 15 minutes, add 2 ml chloramphenicol per liter.
   d. After 20 min, chill 5 min and spin down cells in large buckets for Sorvall centrifuge. Wash once with NET + 0.01 M KCN. Wash again with NET.

6. Lyse cells and precipitate bacterial DNA. The separation of bacterial DNA is accomplished by centrifugation in an SW 27 rotor (use ~36 cc/tube Ti60 type).

   For 2 l Medium:

   a. Suspend cells in 80 cc 10% sucrose in 0.01 M Tris, pH 8.0 (13.3 ml each tube).
   b. Lyse cells by adding: 60 cc 0.1 M EDTA (10 ml each tube)
      20 cc lysozyme, 4 mg/ml (3.3 ml each tube)

   and putting solution at 37°, 30 minutes.

c. Add 12 ml 25% sodium dodecyl sulfate (2 ml each tube). Place at 37° for 15 min.

d. Add 45 ml 5 N NaCl (7.5 ml each tube). Place in ice in the coldroom overnight.

e. Spin 45 min, 25,000 rpm, 6° in an SW 27 rotor.

f. Pour solution through cheese cloth into an Erlenmeyer flask to trap the bacterial DNA.

7. Extract Phage DNA with equal volume of borate-equilibrated phenol. Reextract phenol layer with small volume of borate.

8. Ethanol precipitate aqueous phase in large Sorvall buckets (~230 cc). Store at -20° for 2 hours or overnight.

9. Spin 30 min to 1 hr in large Sorvall rotor, 4°.

10. Resuspend pellet in T-E (~2.5 mls).

11. Add 0.3 cc RNase (pancreatic), 2.5 mg/ml. Place at 37°, 60 min. T_1 RNase may also be added at 25 µg/ml.

12. Add 5 N NaCl to 0.1 N, final concentration. Add 10% Sarkosyl to 0.5% final concentration.

13. Extract DNA with borate-equilibrated phenol. Reextract phenol layer with T-E (Solution 6) and combine with first aqueous layer.

14. Apply total aqueous layer $\leq$ 3 mls to G-100, equilibrated with T-E (Column dimensions : 1 × 18 cm). Collect 1 ml fractions. Pool fractions 29-42; digested products begin in fraction 61.

15. Add $NaC_2H_3O_2$ to 0.2 to 0.3 M and ethanol precipitate.

16. Separation of RF from single-stranded form by CsCl-ethidium bromide gradient. Put all DNA in one gradient (as much as 600 µg). Use 4 mls T-E to dissolve precipiate; add 4.4 gms CsCl and dissolve

(η=1.3950). Add 0.4 ml ethidium bromide (5 mg/ml) very slowly with stirring (η=1.3910). Volume will be approximately 5.5 ml. Spin 40 hours, 34,000 rpm (Ti50), 23°.

17. Collection of ethidium bromide gradient is done with yellow safe light. Puncture alcohol-cleaned tube below band (RF-I) and let gradient drip down side of tube into collection tubes or plastic divided dish.

18. Extract the ethidium bromide with isoamyl alcohol (IA) which has been equilibrated with 1 M Tris, pH 8.0. Several extractions (5-6) are usually necessary. Use a conical tube for maximum interface area. The IA layer will be on top. Extract slowly with gentle shaking until IA layer is also clear since any ethidium bromide in IA which is carried to the dialysis step will rebind to the DNA. Small amounts of ethidium bromide may be dialyzed away using SSC-0.001 M EDTA. The final dialysate is 0.01 M Tris, 0.001 M EDTA and 0.005 M NaCl.

Determination of Error in Sizing Method on Acrylamide Gels

Procedure:

1. Select the linear regression program from the COSAP statistical package provided for the Biology Division PDP-11 computer.
2. Use the migration distance (cm) as the independent variable and the log of the known size of the fragment of ØX174 DNA as the dependent variable. When this assignment is used, the estimated value of the dependent variable will indicate the variation in size determinations for each fragment.
3. The estimated value of the dependent variable will be the logarithm of the size of each fragment for the best line determined by a least squares analysis. The antilog of this value will be the size estimated for that particular electrophoretic run.
4. Deviations are determined as the positive or negative differences between known sizes and those determined by the regression analysis. The observed deviations in fragment size for two different gel concentrations are given in Tables 2 and 3, Appendix A. The buffer system was the same as that described in Chapter II.

PART III

SEQUENCE CHARACTERIZATION OF THE BASIC REPEATING UNIT OF A MULTIMERIC SERIES OF RESTRICTION FRAGMENTS

CHAPTER I

INTRODUCTION

Highly repetitive DNAs are an integral, if not essential, part of most eukaryotic genomes. They have been described in a wide range of taxonomic groups, including plants, animals and prokaryotes (Skinner, 1977). Their detection has included methods of density seqarations (Kit, 1961; Sueoka, 1961; Skinner, 1967), denaturation and reassociation of total or main band DNAs (Beattie and Skinner, 1972; Christie and Skinner, 1979), and restriction endonuclease digestion (Manuelidis and Wu, 1978; Christie and Skinner, Part II). The functional significance of these DNAs is unknown though numerous suggestions have been offered. Within the brachyurans (true crabs) there are examples of satellites with short repeats of only 2 or 4 nucleotides (nt) (Swartz et al., 1962; Skinner et al., 1974). The repeating unit in satellite II of the hermit crab is a heptamer followed by a variable number of trimers, three to twelve (Chambers et al., 1979). One distinguishing feature of these DNAs compared to those described for other taxonomic groups is the homogeneity of the repeating units. In each of the above cases the predominate sequence represented ninety percent or greater of the total sequences. For guinea pig α-satellite (Southern, 1970) or mouse satellite (Biro et al., 1975) the predominate sequence was representative of only 50% of each DNA. Nevertheless, the basic unit for both DNAs was determined to be less than twenty nucleotides. Recently, within the primates repetitive sequences of much greater complexity have been described. A basic repeat

length of approximately 170 nt is obtained both for African green monkey α-satellite (Rosenberg et al., 1978) and for human repetitive DNA (Manuelidis and Wu, 1978). Of even greater interest, two-thirds of the sequence in one is the same as that in the other. The multimers digested from total Geryon DNA are similarly complex, although the repeat length is only one-half that of the primate basic unit.

The structural diversity of satellites and other highly repeated DNAs suggests that several mechanisms may be responsible for their formation. The organization of both mouse satellite (Southern, 1975) and satellite II of the hermit crab (Chambers et al., 1978) indicate that unequal crossing over likely occurs in highly repeated DNAs. The presence of satellite sequences in only one or a few species in a group of related species argues for frequent amplifications producing sizeable increases in the DNA content (Jones, 1977; Barnes et al., 1978). In contrast, loss of satellites apparently coincident with speciation has been observed for the related Drosophila virilis group (Holmquist, 1975). Evolutionary alteration of restriction endonuclease sites in satellite DNAs has been frequently observed and attributed to random mutation of an amplified set of sequences (Southern, 1975; Manteuil et al., 1975; Cooke, 1975). Random mutation can easily explain the presence of multimers of a basic repeat length. Assuming a given extent of random divergence certain proportions of different multimers can be predicted (Southern, 1975). Our analysis of the relative amounts of multimers digested from the total DNA of the red crab, Geryon quinquedens is incompatible with formation of larger multimers only by random loss of restriction sites previously

spaced uniformly (Part II). In this report we present the sequences of the basic repeating unit. Some sequence variants occur too frequently to be explained by random mutation alone. The complexity of this DNA (81 nt) is considerably greater than previously reported for crab DNAs (Swartz et al., 1962; Skinner et al., 1974; Chambers et al., 1978) other than a (G+C)-rich satellite in the Bermuda land crab (Schell et al., 1978). Also a cryptic satellite in the hermit crab (Hinkle, Chambers, and Skinner, unpublished observations) is quite complex. The similarity in complexity of these DNAs to simian and human repeated DNAs (Rosenberg et al., 1978; Manuelidis and Wu, 1978) argues for a universality in the mechanisms of origin for many highly repeated DNAs and probably also in their functions. The range of sequence complexities observed in these DNAs coupled with the localization of both simple and complex repetitive DNAs to the same centromeric regions (Manuelidis, 1978) indicate that several mechanisms may be acting to produce a continuum of complexities. Our studies indicate that the pathway for evolutionary change in the repetitive DNA of Geryon has been amplification of a complex repeat unit, divergence from that sequence, and subsequent reamplification of divergent subsets.

CHAPTER II

MATERIALS AND METHODS

Preparation of DNAs. A slight modification of the "MUP" procedure (Britten et al., 1974) was used to isolate DNA from testes of Geryon quinquedens obtained frozen from the Marine Biological Laboratory, Woods Hole, Massachusetts. Multimeric DNA fragments were obtained from total DNA by restriction endonuclease digestion with HindIII isolated according to Smith (1974). Digestion with other restriction endonucleases (Biolabs) were performed by recommended assay procedures. Analytical and preparative digests were electrophoresed on polyacrylamide gels as previously described (Christie and Skinner, in preparation). For preparative samples individual fragments were excised from the gel and electroeluted into dialysis bags at $4^{o}C$. Labeled fragments were obtained by treatment with bacterial alkaline phosphatase (Worthington) and 5'-phosphorylation using γ-^{32}P-ATP (S. A. 2000-3000 Ci/mmol; ICN) and T4 polynucleotide kinase (P. L. Biochemicals, Inc.).

Electrophoretic Sizing of Restriction Fragments. Sizes of restriction fragments were obtained by co-electrophoresis with HinfI fragments of ØX174. For each electrophoretic run a least squares regression line was calculated for plots of the logs of the known HinfI sizes in nucleotides (Sanger et al., 1977) versus the distance migrated. Slight adjustments in the sizes of Geryon fragments obtained by comparison to ØX174 fragments were made after DNA sequencing was performed. The size of the monomer

was 81 nt, as obtained from sequencing data though the size determined by electrophoresis varied ±5 nt.

DNA Sequencing. Base-specific cleavage reactions were performed on 5'-end labeled DNA (Maxam and Gilbert, 1977; 1979). Discrimination between guanine and adenine have included methylation by dimethylsulfate and differential release of each base (reactions of G>A, or A>G) as well as assessment of guanine residues by the dimethylsulfate reaction (G) and determination of adenines and guanines by acid induced depurination (G+A). Discrimination of cytosine and thymine required hydrazinolysis with (C) or without (C+T) addition of salt. In the strand scission step sodium hydroxide is replaced by piperidine which can be removed by lyophilization of the reaction products. Samples were loaded in 99% foramide onto 8, 12, or 20% polyacrylamide gels (0.05 cm thick) containing 7M urea.

Quantitation of Relative Amounts of Restriction Products. HhaI digestion products of each HindIII multimer were quantitated after 5' end-labeling of each multimer. Digests were electrophoresed on 7% polyacrylamide gels and each lane was sliced into 1-mm slices, incubated overnight at 65°C in 0.2 ml of $HClO_4$:H_2O_2 (1:2 by volume), and counted in ACS scintillation fluid (Amersham/Searle).

CHAPTER III

RESULTS

Sequences of the Basic Repeat Unit. A multimeric series of HindIII fragments which includes sizes from monomer to octamer length can be digested from total Geryon DNA (Christie and Skinner, in preparation). In HhaI digests of each multimer the two major products are 47 and 40 nt, termed fragments a and b. Sequencing of these two fragments from the monomer, dimer, and tetramer indicated that the same sized HhaI fragment from each multimer had essentially the same sequence. A repeat length of 81 nt is obtained because of the overlap of four nucleotides for the HindIII site and the two nucleotide overlap for the HhaI site. Fragment a was read from the labeled adenine (A*) of the upper strand in Figure 1, Appendix A; fragment b, from the labeled adenine of the lower strand. The sequences in the basic unit are more complex than repeated DNA in other crustaceans in that they lack short repeating units internal to the 81 nt repeat. There are tandemly repeating trimers from position 9 to 14 and from position 18 to 23, but no overall pattern related to a short repeat of three or four nucleotides can be discerned. In one strand there are runs (4 to 8 nt) of pyrimidines and conversely, on the other strand, runs of purines. The sequence arrangement indicates that intra-strand hybridization is unlikely. That this DNA reassociates in the correct sequence register is known from previous experiments. Fragments of monomer and dimer size are produced by HindIII digestion of both second-order repetitive components but not by digestion of the foldback fraction (Christie and Skinner, 1979).

Organization of the Multimers. The sequence in the monomer was confirmed as the basic repeat unit by examination of minor fragments produced by HhaI; for the dimer, fragments slightly greater than 120 nt are produced in addition to those of 47 and 40 nt (Figure 2A, Lane 1; Appendix A). These can be explained by the loss of one of the two expected HhaI sites in the dimer (Figure 2B; Appendix A). The predicted fragments of 128 and 121 nt would not be expected to migrate as two distinct bands in a 7% polyacrylamide gel. Sequence analyses of this size class indicated that it was a mixture of DNA species. Bands appeared at nearly all positions in every sequence ladder. Since the dimer was labeled at the 5' ends before HhaI restriction, the reading order for the 128 nt fragment would be a-b-a and for the 121 nt fragment, b-a-b. This sequence organization resulted in an overlap of fragment a and b sequences. Sequences of the two expected fragments could be distinguished in certain regions such as the distinctive run of seven thymine residues from position 24 to 30 in fragment a. The sequence 5'-TGTGGCA-3' (positions 62 to 56, lower strand) from the 121 nt fragment was superimposed on the run of thymine residues in the initial a sequences of the 128 nt fragment. For the 121 nt fragment reading would begin at the labeled 5' end on the indicated adenine residue (A*) at the right side of the HhaI/40 nt fragment (Figure 1; Appendix A).

The 128 and 121 nt fragments are examples of sets of minor fragments observed in digests of all multimers which are clearly related to the basic repeating unit (Part II). In HhaI digests of all multimers minor fragments can be identified for sizes expected from the loss of one or

more adjacent HhaI sites. The presence of these fragments corroborates the conclusion that an original population of 81 nt monomers has been eroded by mutation. From the proposed organization it is predicted that fragments of 81 nt should be produced by HhaI digestion of all multimers above the monomer in size. When the 5' ends of the dimer were labeled before digestion, the only HhaI digestion products of the dimer (Figure 2A, Lane 1; Appendix A) are the 47 and 40 nt fragments and the mixed class of 128 and 121 nt fragments. When the HhaI digest of the dimer is treated with alkaline phosphatase and relabeled with polynucleotide kinase, a class of 81 nt fragments is readily detectable (Figure 2A, Lane 2; Appendix A). The presence of internal digestion products of the predicted size also corroborates the proposed organization based on an 81 nt repeat.

Several lines of evidence indicate that the fragments of each multimeric size class are not homogeneous populations of identical sequences. Instead, various subsets are present due to minor differences both in sequence and fragment length. An enzyme resistant fraction is a constant feature of each multimer even with large increases in enzyme concentration and time of incubation. This is true even for the monomer and dimer which are isolated from a region of the gel essentially free of heterogeneously sized HindIII digestion products. As indicated above in the HhaI digest of the dimer, fragments are present whose lengths can be explained as an alteration in one or more restriction sites. This suggests that the heterogeneity in each multimer is due to divergence from a formerly homogeneous population of monomers. From the relative amounts of each HindIII multimer it was postulated that selective

amplification of a small subset of the sequences containing the HindIII site had occurred (Part II). This would explain the unexpectedly large amounts of dimer, tetramer, and octamer as compared to the very small amount of trimer, pentamer, hexamer, and heptamer. In addition, the amount of dimer was also higher than expected for a simple model of random inactivation of uniformly spaced restriction sites (Slack, 1974; Southern, 1975). The number of dimers in the genome nearly equaled that of the monomer. For this number of dimers to be produced by random divergence alone, would require that 92% of the HindIII sites be altered and that the overall divergence in this DNA be approximately 16%. While this level of divergence is quite high, it would not be detected on sequencing gels if mutations occurred completely at random. Alterations at every position in the basic repeat occurring randomly in the entire array of repeating units would appear only as background in the sequence ladders of the autoradiograms. Evidence from studies of multimeric series digested from isolated satellite fractions (Southern, 1975; Altenberger et al., 1977; Fittler, 1977) support the conclusion that mutation occurs at random in clustered highly repetitive DNAs.

Sequence Variants Indicating Amplification of Divergent Sequences. Heterogeneity which can be observed from sequence analyses is the presence of sequence subsets in individual multimers. The 81 nt monomer sequence as presented in Figure 1, Appendix A, was determined by sequencing the two major HhaI digestion products, the 47 and 40 nt fragments from the monomer, dimer and tetramer, and also the HaeIII/75 nt product from the monomer (Figure 3, Appendix A). Certain features

distinguish the sequences of the dimer from those of the monomer. These differences indicate the presence of one or more sequence subsets either more abundant or unique to the dimer. First, the pattern observed in the sequencing gel at the 3' end of the HhaI/47 nt fragment from the dimer indicates the presence of variant sequences. The presence of more than one strong band at the 3' end of the molecule is potentially indicative of length heterogeneity; the observed pattern is compatible with three or more differently sized fragments. The relative amount of DNA for each presumptive band of uncleaved fragments indicate that the variant sequences are not trivial fractions. With two different sets of sequencing reactions this same result was obtained.

When the sequences of the monomer, dimer, and tetramer are compared, other variations are observed. At positions 52 and 60 in the HaeIII/75 nt fragment of the monomer base pairs of adenine and thymine (A/T) are found. Although A/T base pairs also predominate at these positions in the HhaI/40 nt fragment of the dimer, guanine and cytosine (G/C) base pairs are also detected. The nucleotides at these positions for the HaeIII/75 nt fragment, read from the upper strand (Figure 1, Appendix A), were adenines; thus the presence of a subset with guanine residues would be readily detected as bands in the A+G channel. Therefore, the dimer contains sequence variants not present in comparable amounts in the monomer. Presumably the original sequence had A/T base pairs at these positions since the tetramer sequence matches that of the monomer. Another sequence variation was found at position 26 of the dimer. A guanine residue occurred as a minor sequence coincident with the thymine

residue found in the major sequence. While there was background radioactivity in the G+A channel for all the thymines between positions 24 and 30, a band in the G channel occurs only at position 26. While this sequence variant was not readily detected in the sequences of the monomer, it might be present in lesser amounts. This variant was detected in four different sets of sequencing reactions performed on the dimer. For the sequences of the tetramer this variant could not be readily distinguished, possibly because of an even higher background in these sequences compared to the monomer and dimer.

Conclusive assignments of sequences require multiple trials to eliminate error due to the artefactual potential in the numerous procedures required to produce a sequencing gel. Maxam and Gilbert (1979) list fifteen problems which can lead to aberrant banding patterns. The assessment of sequence variants in the dimer was based on repeated observations in numerous sequencing gels utilizing separately prepared reaction sets. An alternative explanation for the multiple dark bands in every channel at several positions near the 3' end of the HhaI/47 nt fragment of the dimer would be aberrant cutting by the HhaI restriction endonuclease at sites other than 5'-GCG$\overset{\downarrow}{C}$-3'. It seems unlikely that there will be aberrant cutting in repeated trials. It is more plausible to assume that there are multiple subsets in the dimer which have HhaI recognition sites at several different positions.

Sequence Variants Consistent with Random Mutation. Minor subsets can also be demonstrated by the susceptibility of the multimers to

restriction enzymes whose recognition sites are not in the predominate sequence derived from sequencing gels. While the site for the AluI enzyme is the four central nucleotides of the HindIII site, there are no AluI sites located internally within the basic repeat. Treatment of the dimer with AluI produces several fragments which collectively comprise less than 5% of the dimer DNA (Figure 4, Lane 1; Appendix A). The sequences of the AluI/45 nt fragment appear identical to those of the HhaI/47 nt fragment. A single base change of guanine to thymine at position 47 would produce the AluI recognition site of 5'-AG↓CT-3' and yield a 45 nt fragment. Sequence analyses of the AluI/83 nt fragment from the dimer indicate that this class is a mixture of at least two fragments. It is not unexpected that a fraction of dimer fragments will have lost the central HindIII site by alteration of either the first or last nucleotide in the site. Cleavage at the AluI site in the center of the dimer would produce two labeled fragments of 83 nt, one from each 5' terminus. Random mutation might easily have produced this fraction. Only a small percentage of these fragments is expected because of the greater probability of altering one of the four central nucleotides in the HindIII site and thus losing the AluI site.

The presence of restriction sites in small subsets of each multimer is quite clear for other enzymes as well. A HaeIII restriction site is part of the basic repeat unit at positions 74 to 77. The dimer presumably contained two original HaeIII sites and should be digested by HaeIII to give fragments corresponding to the presence of either one or both sites. Fragments of the expected sizes of 10, 75, and 91 nt are obtained

in the digest but fragments of 21, 23, and 28 nt are also observed (Figure 4A, Lane 2; Appendix A). The latter fragments represent only a few percent of the total radioactivity in the digest. Their presence can most easily be explained as alterations in the b set of sequences. Several HaeIII sites could have arisen by one or two nucleotide changes in the region between positions 50 and 70. A change of cytosine to guanine at position 64 would give a 21 nt fragment from the labeled 5' end of the lower strand and a change of adenine to guanine at position 62 would give a 22 nt fragment which likely corresponds to the minor fragment of 23 nt. In addition, a single nucleotide change of thymine to guanine at position 56 could account for the 28 nt fragment. A similar interpretation can be made for the HhaI minor fragments of less than 30 nt. An HhaI digest of the monomer is shown in Figure 4B (Appendix A). The major digestion products are the 47 and 40 nt fragments discussed above. Two minor fragments of 20 and 28 nt are present although the 28 nt fragment is barely detectable. These minor HhaI fragments would be labeled at the A* of the lower strand and would correspond to single nucleotide changes at position 65 for the 20 nt fragment and 56 for the 28 nt fragment. These changes would yield 22 and 30 nt fragments, respectively. In summary, the minor fragments produced by either HhaI or HaeIII cleavage at sites not found in the basic repeat can be explained by random mutation. Thus, while random mutational alterations cannot be detected by sequence analyses they are detectable by the appearance of "new" restriction sites within the basic repeat unit. As expected, these arise most often by single nucleotide changes. As long as the proportion of fragments containing

these "new" sites is only a few percent of the total sequences, their presence can be readily explained by random mutation.

Estimate of Divergence in the Multimers. The extent of the divergence in classes of repetitive DNAs can be determined by examination of the organization of restriction sites (Slack, 1974; Southern, 1975). This approach yields the percent of altered sites in an array of sites originally spaced uniformly. From the fraction of sites which have been altered the divergence in the entire sequence can be calculated. Assessment of a set of restriction sites in a repetitive DNA is analogous to selecting a small population of nucleotides from which to predict the amount of change which has occurred in the entire population. As divergence within the sequences increases, the number of sites altered at adjacent positions will increase. As a result, a multimeric series of fragments will be produced from the original population of monomer. Regardless of the overall extent of divergence, the relative amount of each successively larger multimer should be less than the preceding multimer. Other multimeric series digested from either satellites (Southern, 1975; Altenberger et al., 1977; Fittler, 1977) or total DNA (Manteuil et al., 1975; Cooke, 1975) were compatible with formation by random mutation alone. For the HindIII multimeric series of Geryon total DNA the plot of the log of the relative amount of each multimer versus (n-1), where n = 1 for a monomer, deviates significantly from the typically observed straight line (Part II). In particular, there are enhanced amounts of dimer, tetramer, and octamer. This result coupled with specific indications of amplified divergent sequences in the dimer

and in the octamer (see below) indicate that random mutation alone cannot explain the observed distribution of restriction sites. Selective amplification of a relatively small DNA segment, several thousand nucleotides, is offered as an explanation.

Due to the indication of a nonrandom spacing of HindIII sites in the Geryon genome, the distribution of HhaI sites within specific HindIII fragments was investigated. Quantitation of this distribution for the monomer, dimer, trimer and tetramer was performed in the following way. HhaI digestion products of HindIII fragments were quantified according to the number and arrangement of HhaI sites. Fragments are divided into Classes[1] depending on the number of HhaI sites which had been altered. For each multimer, Class 0 fragments are those with no altered HhaI sites (always the 47 and 40 nt fragments) while Class 1 fragments have one altered site. Classes 2, 3, and 4 include fragments having lost two, three or four adjacent HhaI sites, respectively. For HhaI digestion of the monomer, Class 1 will be the enzyme resistant fraction while for the dimer it will be fragments of 121 and 128 nt (see Figure 2, Appendix A). Since 5' end-labeled multimers were used, the sizes of the fragments in each Class are the distances from each labeled adenine in the HindIII sites to the respective HhaI sites. The relative amounts of each Class are plotted in Figure 5, Appendix A, for the first four multimers. On each graph there are two dashed lines representing two different levels

[1]The fragments in these Classes comprise the category defined as Subset I in Part II, i.e., those fragments clearly related to the basic repeat sequences specified in Figure 1, Appendix A of Part III.

of divergence, (a) 3.7% and (b) 15.7%. These two lines are included to allow a comparison to the distribution of HindIII sites in the total genome (Part II). For the HindIII multimeric series the relative amounts of trimer, pentamer, hexamer, and heptamer best fit line (a) while the dimer and octamer best fit line (b). The tetramer value is intermediate.

Examination of the distribution of HhaI sites reveals two points. First, there is always a large amount of enzyme resistant fragments (indicated as uncut in the figure). The difficulties in eliminating technical problems such as incomplete digestion do not allow firm conclusions to be made from this result. Nevertheless, unexpectedly large amounts of resistant DNA was also observed by Carlson and Brutlag (1977) for the Drosophila melanogaster 1.688 (g/cm^3) satellite. This result, together with the presence of products with a length of 15 monomer units in complete digests of the satellite, prompted these authors to conclude that random inactivation of restriction sites was insufficient to explain the organization of this satellite. Such resistant fractions may well be the result of an amplification of a DNA segment in which the original spacing of restriction sites has been lost.

The second and more easily interpretable result obtained from these analyses is the indication of the extent of divergence of HhaI sites within the HindIII sequences. A precise determination of divergence of the HhaI sites could not be made because of the small number of sites within the first four multimers for which quantitation was performed. Multimers of five, six, and seven monomer lengths were present in too few copies to be used for quantitation. The octamer was not included in this

analysis due to enhance amounts of Class 3 fragments which will be discussed later. Only the dimer, trimer, and tetramer will be considered for an estimate of divergence in HhaI sites since an estimate from the monomer would depend on the enzyme resistant fraction. It is assumed that even if a HindIII multimer was not digested to completion, that the observed amounts of each Class would not deviate from their actual relative relationships and, therefore, that the observed slopes would be indicative of the sequence divergence. The relative amounts of each Class of fragments for the dimer, trimer, and tetramer are most compatible with the lower divergence of line (a). These results support the idea that the divergence in at least one subset of the HindIII fragments is near 4% and that the large amounts of dimer, tetramer, and octamer are due, most likely, to selective amplification.

The nature of at least one amplification event is indicated from a quantitation of the HhaI products of the octamer (Figure 6, Appendix A). Over fifty percent of the digest is comprised of Class 3 fragments, those which have presumably lost three adjacent HhaI sites. Amplification of a DNA segment containing a divergent octamer would yield this result. Clearly less divergent octamers are also present since Classes 0, 1, and 2 are also present. The relative amounts of these latter Classes are most compatible with a divergence of 4% similar to the second, third, and fourth multimers. We therefore conclude that the divergence due to random mutation in this repetitive DNA is only 4% and that the presence of unusually large amounts of particular HindIII multimers results from amplification of a DNA segment containing divergent multimers. It is also possible that the large amounts of enzyme (HhaI) resistant fractions in each multimer have resulted from amplification of divergent sequences.

CHAPTER IV

DISCUSSION

The Geryon sequences in the HindIII multimers are more similar in complexity to those observed in repetitive DNAs of the African green monkey and man (Rosenberg et al., 1978; Manuelidis and Wu, 1978) than those previously observed for crab satellites (Skinner, 1967; Skinner et al., 1974; Chambers et al., 1978; Skinner, 1977). Also in the Geryon multimers, as in the primate repetitive DNAs the ratio of CG to GC neighbors greatly exceeds values in total DNA for the mouse and guinea pig (Russell et al., 1976). The total DNAs of mouse and guinea pig have a marked deficiency of CG neighbors compared to that expected from their base composition. In contrast, satellites from either organism are only slightly reduced from expected values rather than increased, as for the primate and Geryon repetitive DNA. Although a general pattern for bulk DNAs featuring a CG deficiency has been suggested (Russell et al., 1976), there does not appear to be a trend for highly repetitive DNAs. For Geryon the CG to GC ratio is nearly equal. Whatever forces act to reduce the CG neighbors in bulk DNA do not seem to be acting in this repetitive DNA. The relative proportion of these two types of nearest neighbors in a repeated DNA may depend only upon the particular sequence chosen for amplification. As yet there is no strong evidence for evolutionary selection of specific repetitive sequences. Should a selection on repeat units exist, a deficiency of CG neighbors does not appear important.

The relative proportions of HindIII multimers in total Geryon DNA cannot be explained solely by a random accumulation of mutations in a homogeneous array of monomers. Rather than progressively lesser amounts of each succeedingly larger multimer, there are enhanced amounts of dimer, tetramer and octamer (Part II). From the observed proportions of the multimers it was not possible to determine the overall divergence in these sequences. The relative amounts of the trimer, pentamer, hexamer, and possibly the heptamer could result from approximately 4% divergence, while the amounts of the tetramer were consistent with approximately 10% divergence and the amounts of dimer and octamer require approximately 16%. Examination of the distribution of HhaI sites was performed to give an independent assessment of divergence. This distribution best fits a divergence of 4% and is quite incompatible with a divergence of 16%. This extent of divergence is much too low to explain the large amounts of particular multimers observed. While these enhanced proportions may be due to mutation at specific sites, a simpler interpretation is selective amplification of divergent sequences. For DNA regions to vary in accessibility to mutation, the nucleosome structure might dictate protected regions, e.g., sites every 162 nt. Enhanced amounts of dimer (162 nt), tetramer (2 × 162 nt), hexamer (3 × 162 nt), and octamer (4 × 162 nt) might then be expected. The amount of hexamer is not enhanced but is equivalent to that expected for 4% divergence. Specific sites for mutation would require precise phasing of the repeating DNA units with the proteins in the chromatin complexes. This has been suggested for several repetitive DNAs (Maio et al., 1977). The presence of

more octamers than hexamers in Geryon DNA requires the additional constraint of a multinucleosome complex with every fourth nucleosome being less accessible to mutation than every third. While not prohibitive, these constraints seem unlikely. Alternatively a model of selective amplification requires only that the amplified segment of DNA be sufficiently small to be unrepresentative of the entire class in its relative amounts of multimers. This might easily be the case if amplification occurs by rolling circle replication with a DNA segment several thousand nucleotides long. A segment of this length is still quite small compared to the total quantity of HindIII sequences ($\sim 3 \times 10^8$ nt).

Nonrandom distribution of restriction sites has been previously observed for mouse satellite in which only a small fraction of the basic repeat units contained HaeIII sites (Southern, 1975). This fraction represented about 5% of the total satellite. For Geryon the sequences characterized by nonrandom distribution of restriction sites represent approximately 20% of the HindIII sequences, assuming that one-half of each of the current amounts of dimer, tetramer, and of one of the octamers was produced by amplification of divergent sequences. Although it is likely that the same general mechanisms for amplification occur in all or most eukaryotes, some specialization within particular taxonomic groups is likely. The genomes of several crabs have already been shown to have large amounts of highly repetitive DNA (Vaughn, 1975; Holland and Skinner, 1978; Christie and Skinner, 1979). Other arthropods, specifically, certain insects have quite large amounts of highly repetitive DNA (Gall and Atherton, 1974; Barnes et al., 1979). While such examples are not

considered indicative of a requirement for large amounts of highly repetitive DNAs they might well suggest active mechanisms for their formation. By contrast the genomes of six primates, including man, contained less than 10% highly repetitive DNA (Jones, 1977).

The homogeneity of repeating units in certain arthropod satellite DNAs can also be contrasted with the heterogeneity in rodent satellite DNAs (see Chapter I). This feature may again imply that an unusually high rate of amplifications occurs in the repetitive DNAs of crabs. Multiple rounds of amplification at frequent time intervals would have the effect of maintaining a particular set of sequences. For example, crab poly d(A-T) has been demonstrated in widely divergent species (Beattie and Skinner, 1972). Simultaneously it is expected that the inclusion of a divergent multimer in the segment for amplification will increase the magnitude of the divergent sequences. This would give an unexpectedly large proportion of enzyme resistant fragments as we observe. Additional support for this idea is derived from the very low divergence observed in a fraction of three HindIII multimers of Geryon. Amplification of both divergent and non-divergent multimers would effectively maintain a portion of the original subset and simultaneously magnify a particular divergent subset. "Divergent" and "non-divergent" refer only to the set of restriction sites being tested; other modifications would be undetected. Since each of the three multimers contain fractions showing 4% divergence, it is likely that a single amplification contained a dimer, trimer, and tetramer which had retained the original spacing of HhaI sites. This would indicate that the minimum amount of amplified DNA was the combined

length of these three multimers (729 nt). The data do not require that these three multimers were contiguous for amplification. In contrast, from the presence of large amounts of a divergent octamer it is clear that a continuous length of at least 648 nt was amplified. Whether this latter amplification occurred at the same time as that of the smaller multimers cannot be determined. These estimates of the length for an amplification unit are minimal since a length of several thousand nucleotides would still be small enough so that the distribution of restriction sites in the amplified DNA might not reflect the distribution in the entire class of sequences originally related.

LIST OF REFERENCES

LIST OF REFERENCES

Barnes, S. R., D. A. Webb, and G. Dover. 1978. Chromosoma 67: 341-363. Berlin.

Beattie, W. G. and D. M. Skinner. 1972. Biochim. Biophys. Acta 281: 169-178.

Biro, P. A., A. Carr-Brown, E. M. Southern, and P. M. B. Walker. 1975. J. Mol. Biol. 94: 71-86.

Britten, R. J., D. E. Graham, and B. R. Neufeld. 1974. Methods in Enzymology. L. Grossman and K. Moldave, Eds. Vol. 29E. Academic Press, New York, pp. 363-418.

Carlson, M. and D. Brutlag. 1977. Cell 11: 371-381.

Chambers, D. A., M. P. Schell, and D. M. Skinner. 1978. Cell 13: 97-110.

Christie, N. T. and D. M. Skinner. 1979. Nucleic Acids Res. 6: 781-796.

Cooke, H. J. 1975. J. Mol. Biol. 95: 87-99.

Fittler, F. 1977. Eur. J. Biochem. 74: 343-352.

Gall, J. G. and D. D. Atherton. 1974. J. Mol. Biol. 85: 633-664.

Holland, C. A. and D. M. Skinner. 1977. Chromosoma 63: 223-240. Berlin.

Holmquist, G. 1975. Nature 257: 503-506. London.

Jones, K. W. 1977. Molecular Structure of Human Chromosomes. J. J. Yunis, Ed. Academic Press, New York, San Francisco, and London, pp. 295-326.

Kit, S. 1961. J. Mol. Biol. 3:711-716.

Maio, J. J., F. L. Brown, and P. R. Musich. 1977. J. Mol. Biol. 117: 637-656.

Manteuil, S., D. Hamer, and C. A. Thomas, Jr. 1975. Cell 5: 413-422.

Manuelidis, L. 1978. Chromosoma 66: 23-32. Berlin.

Manuelidis, L. and J. Wu. 1978. Nature 276: 92-94. London.

Maxam, A. and W. Gilbert. 1977. Proc. Natl. Acad. Sci. 74: 560-564. USA.

Maxam, A. and W. Gilbert. 1979. In preparation.

Rosenberg, H., M. Singer, and M. Rosenberg. 1978. Science 200: 394-402.

Russell, G. J., P. M. B. Walker, R. A. Elton, and J. H. Subak-Sharpe. 1976. J. Mol. Biol. 108: 1-23.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe, and M. Smith. 1977. Nature 265: 687-694. London.

Schell, M. G., C. A. Chambers, and D. M. Skinner. 1978. J. Cell Biol. 79: 143a.

Skinner, D. M. 1967. Proc. Natl. Acad. Sci. 58: 103-110. USA.

Skinner, D. M., W. G. Beattie, F. R. Blattner, B. P. Stark, and J. E. Dahlberg. 1974. Biochemistry 13: 3930-3937.

Skinner, D. M. 1977. BioScience 27: 790-796.

Slack, J. 1974. Biopolymers 13: 2241-2264.

Smith, H. O. 1974. Methods in Molecular Biology. R. B. Wickner, Ed. Vol. 7. Academic Press, New York and London. pp. 71-85.

Southern, E. M. 1970. Nature 227: 794-798. London.

Southern, E. M. 1975. J. Mol. Biol. 94: 51-69.

Sueoka, N. 1961. J. Mol. Biol. 3: 31-40.

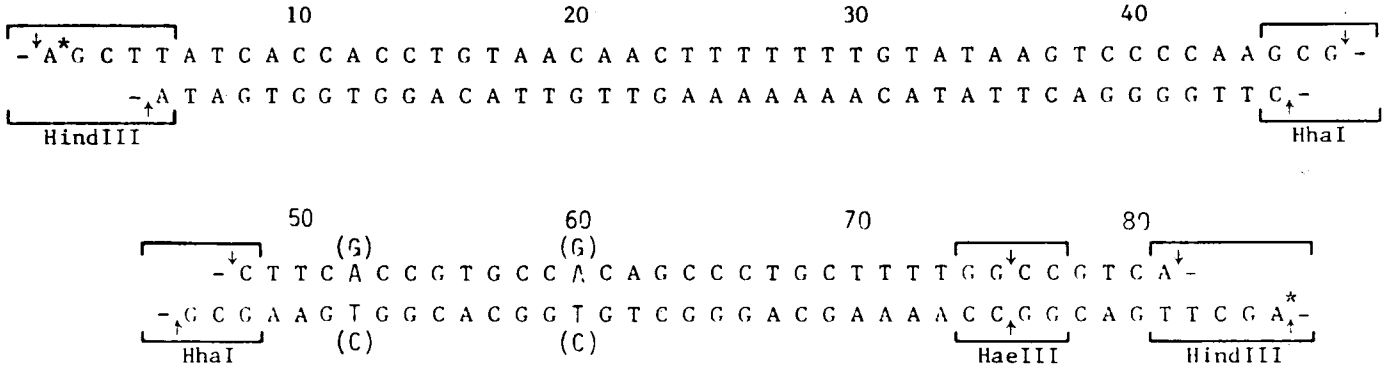
Swartz, M. N., T. A. Trautner, and A. Kornberg. 1962. J. Biol. Chem. 237: 1961-1967.

APPENDIX

Figure 1. The predominate nucleotide sequence of the monomer of the HindIII multimeric series.

Sites for restriction endonuclease cleavage are shown. For sequence analyses performed on HhaI digestion products, sequences for each fragment were read from the indicated nucleotide (A*). Two sequence variants of the dimer are shown in parentheses at positions 52 and 60.

Fragment a (HhaI/47nt)



Fragment b (HhaI/40nt)

Figure 1

Figure 2. Organization of sequences in the HindIII dimer.

A. The dimer is labeled at the 5' termini either before (1) or after (2) digestion with HhaI.

B. The original sequence pattern (1) is shown with restriction endonuclease sites given for HindIII (0), HhaI (X), and HaeIII (▲). A representative of the HindIII dimer fragments is presented (2) which contains all the original sites except the central HindIII site. HhaI digestion of 5' end-labeled dimer yields Classes of fragments (3). Classes 0, 1 and 2 have either zero, one or two altered sites.

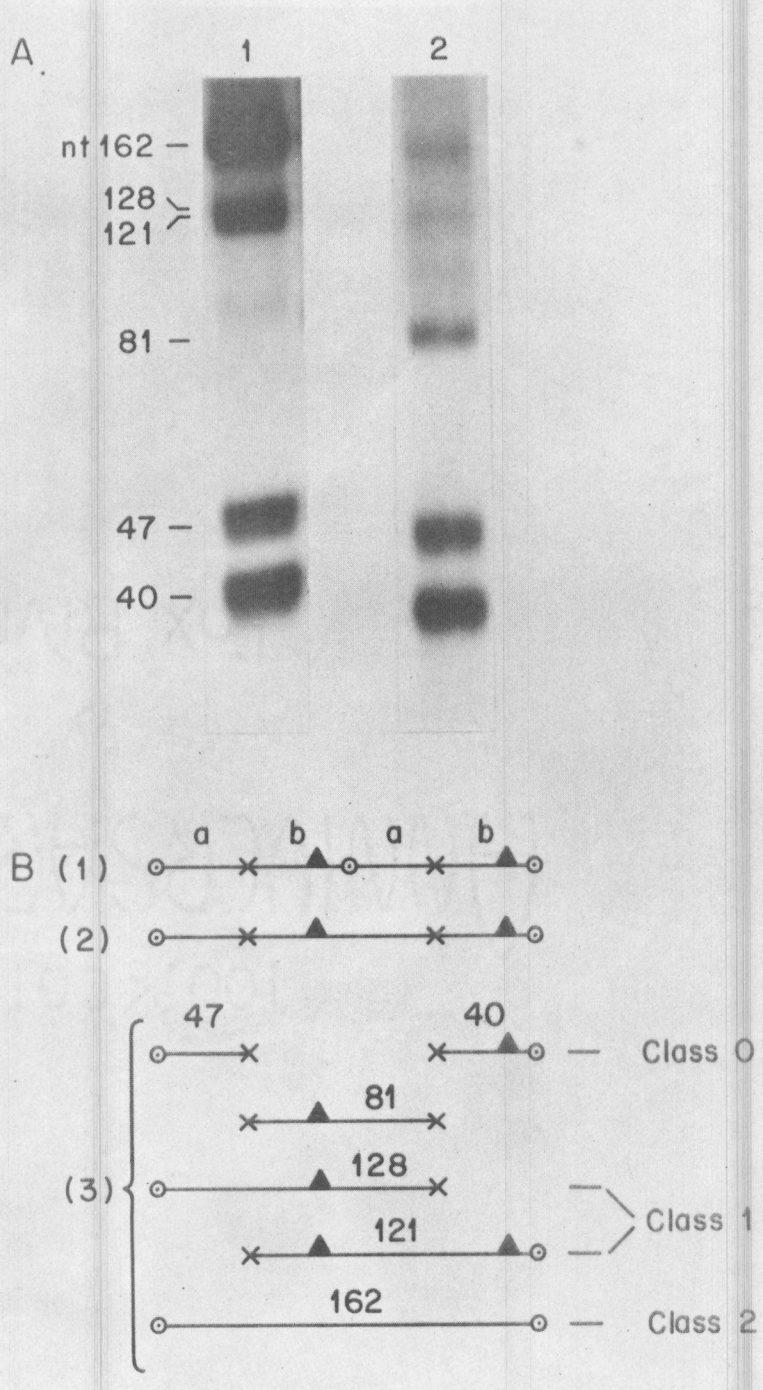

A
1
2
nt 162
128
121
81
47
40
B (1)
a
b
a
b
(2)
(3)
47
40
Class 0
81
128
121
Class 1
162
Class 2

Figure 2

Figure 3. Representative autoradiograms of DNA sequencing gels.

Base-specific reactions for fragments HhaI/40 nt (A), HhaI/47 nt (B) and HaeIII/75 nt (C) were applied to 12% polyacrylamide-urea gels. The positions are indicated for reference to the complete sequence given in Figure 1, page 149. The base-specific reactions used (G, G+A, C, C+T) are given in Materials and Methods.

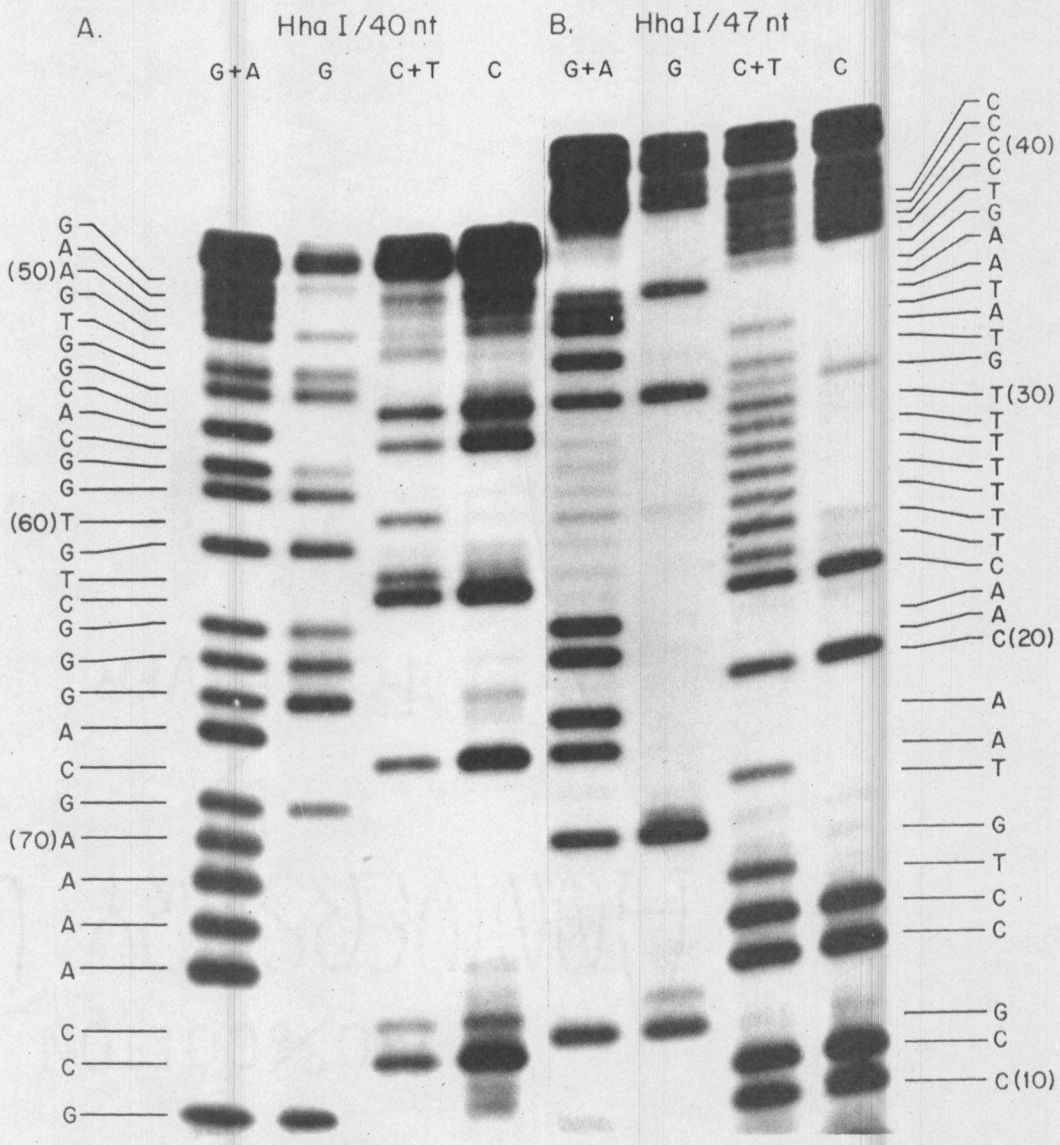

A.
Hha I/40 nt
G+A
G
C+T
C
B.
Hha I/47 nt
G+A
G
C+T
C
G
A
(50)A
G
T
G
G
C
A
C
G
G
(60)T
G
T
C
G
G
G
A
C
G
(70)A
A
A
A
C
C
G
C
C
C(40)
C
T
G
A
A
T
A
T
G
T(30)
T
T
T
T
T
T
C
A
A
C(20)
A
A
T
G
T
C
C
G
C
C(10)

Figure 3

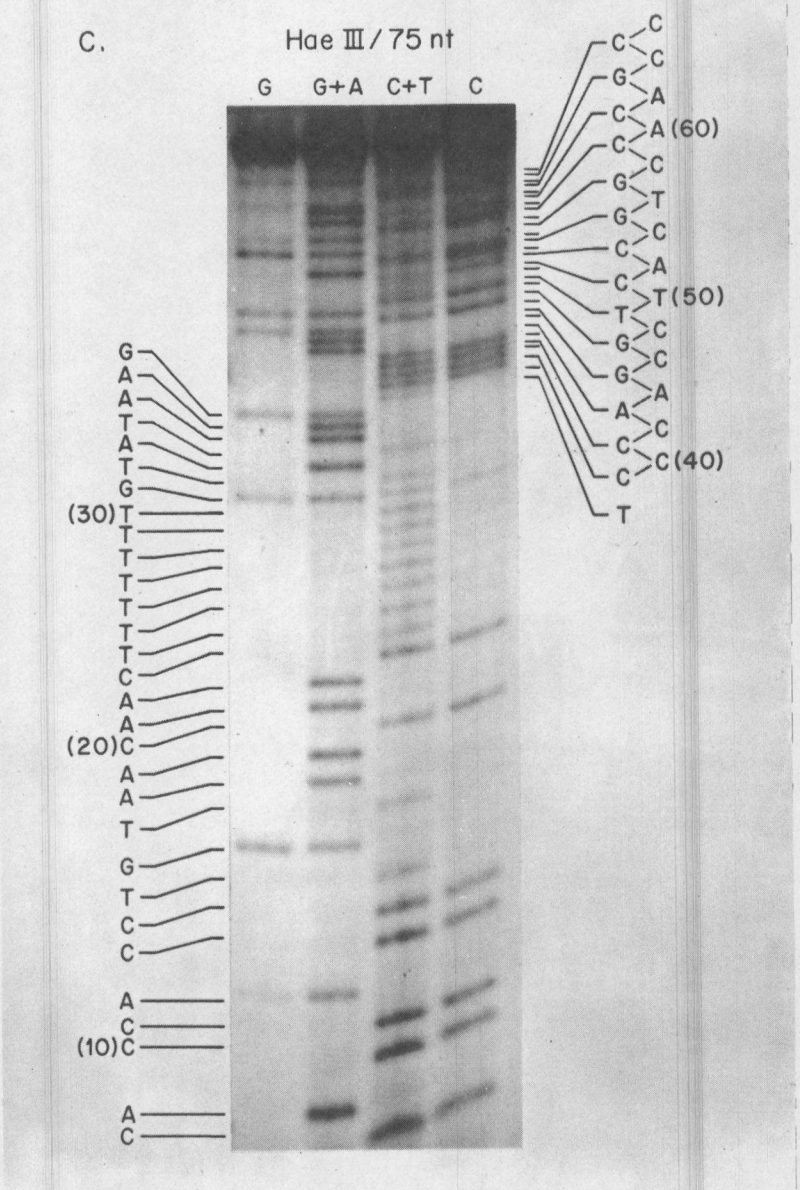


Figure 3 (continued)

Figure 4. Autoradiogram of digestion products of the HindIII monomer and dimer.

The dimer (A) was digested with AluI (Lane 1) or HaeIII (Lane 2) and electrophoresed on a 7% polyacrylamide gel. The monomer (B) was digested with HhaI and electrophoresed on a 7% polyacrylamide gel. Both multimers were 5' end-labeled before digestion.

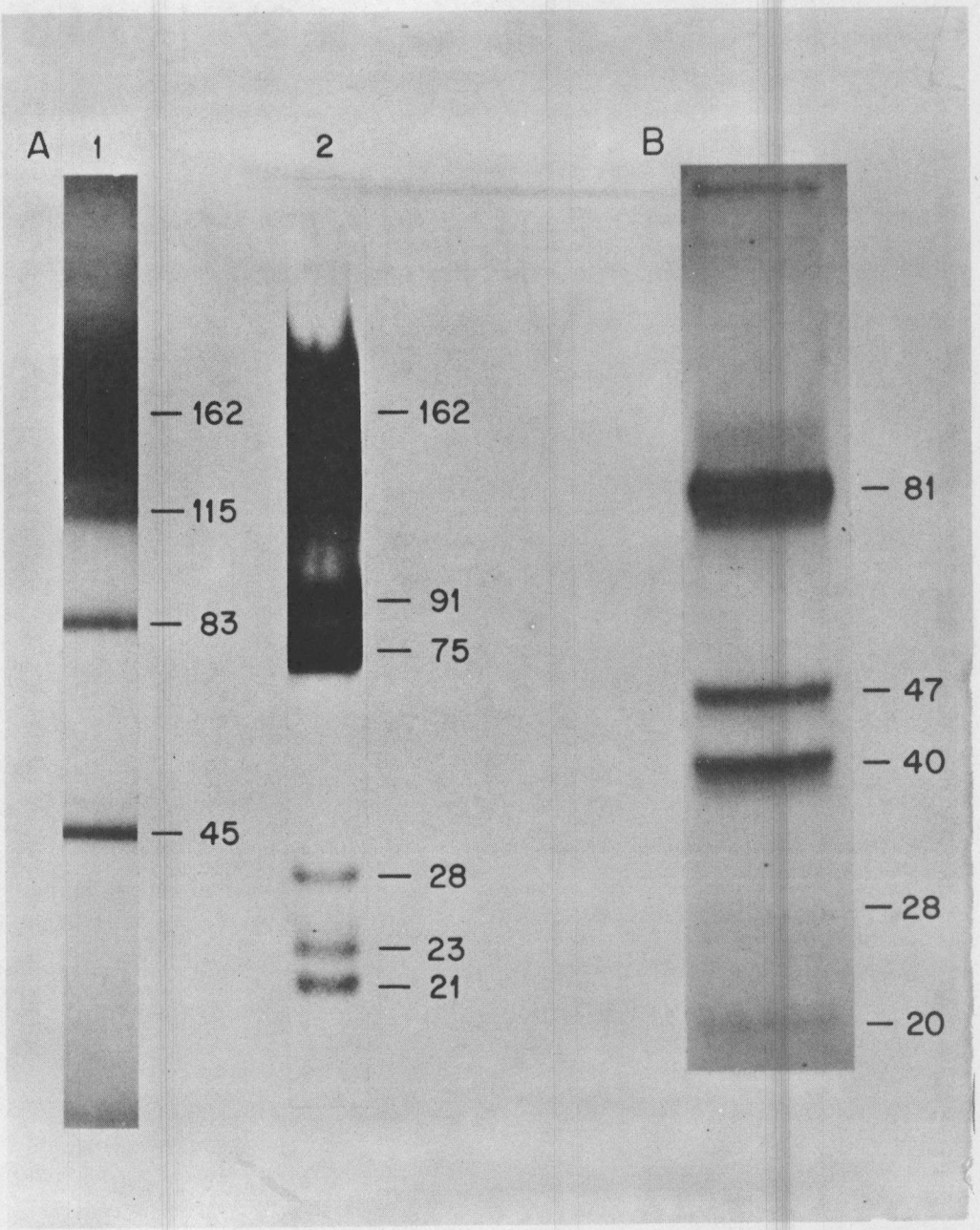


Figure 4

Figure 5. Distribution of HhaI sites in HindIII multimers.

The relative amount, a_n, of each HhaI digestion product of the monomer (A), dimer (B), trimer (C), and tetramer (D) were quantitated (see Materials and Methods) and plotted (solid line) according to the arrangement of HhaI sites. For the dimer (Figure 2, page 151) Classes of fragments are produced in HhaI digests of the other multimers in which cleavage has occurred at the first through the fourth site from each labeled 5'-terminus depending on the number of internal altered sites. The number of Classes in each multimer depends on its size. The plot of log a_n, versus s, the number of adjacent altered sites per fragment, should yield a straight line if the distribution of HhaI sites is due entirely to random mutation of a previously uniform array of sites. This plot is an adaptation from Southern (1975) for the fraction f of DNA in each multimer given by

$$f = n(1-p)^2\, p^{(n-1)}, \text{ or}$$

$$\log f/n = 2 \log (1-p) + (n-1) \log p$$

where $n = 1$ for monomers, etc. and p is the fraction of altered sites, a_n has been substituted for f/n and s, for $n-1$. The terms s and $(n-1)$ are interchangeable since the nth monomer will have $(n-1)$ altered sites. The dashed lines are plots of log c_n versus s for divergence values of (a) 3.7% and (b) 15.7% where c_n is the number of copies of each multimer relative to the monomer, i.e., $c_n = 1$ when $n = 1$. The terms f/n, a_n and c_n are all number fractions and therefore can be used to determine p from the slope. The divergence d as the fraction of altered nucleotides is determined by $p \times 1/t$ where t is the number of nucleotides in the recognition site of the restriction endonuclease. Uncut refers to the enzyme resistant fraction. For the tetramer, Class 3 fragments were not detected.

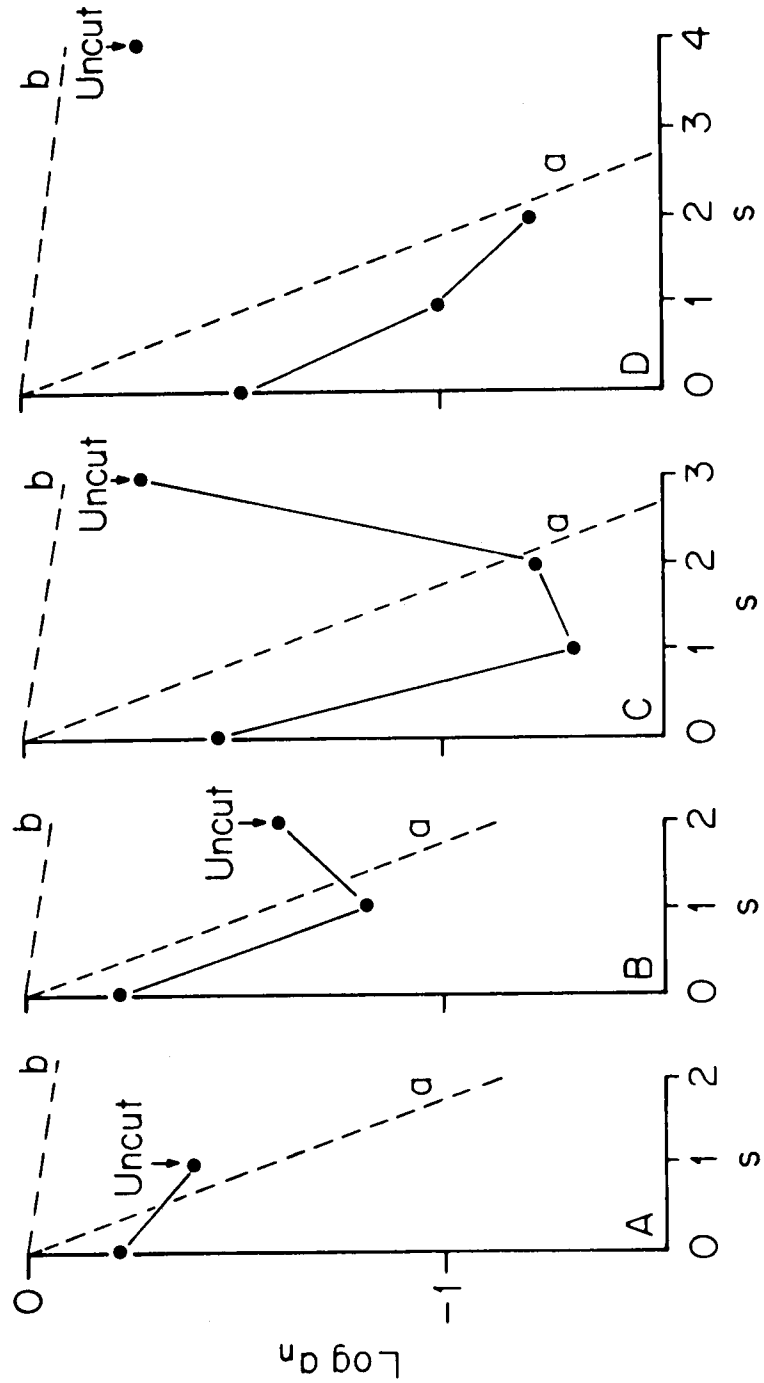


Figure 5

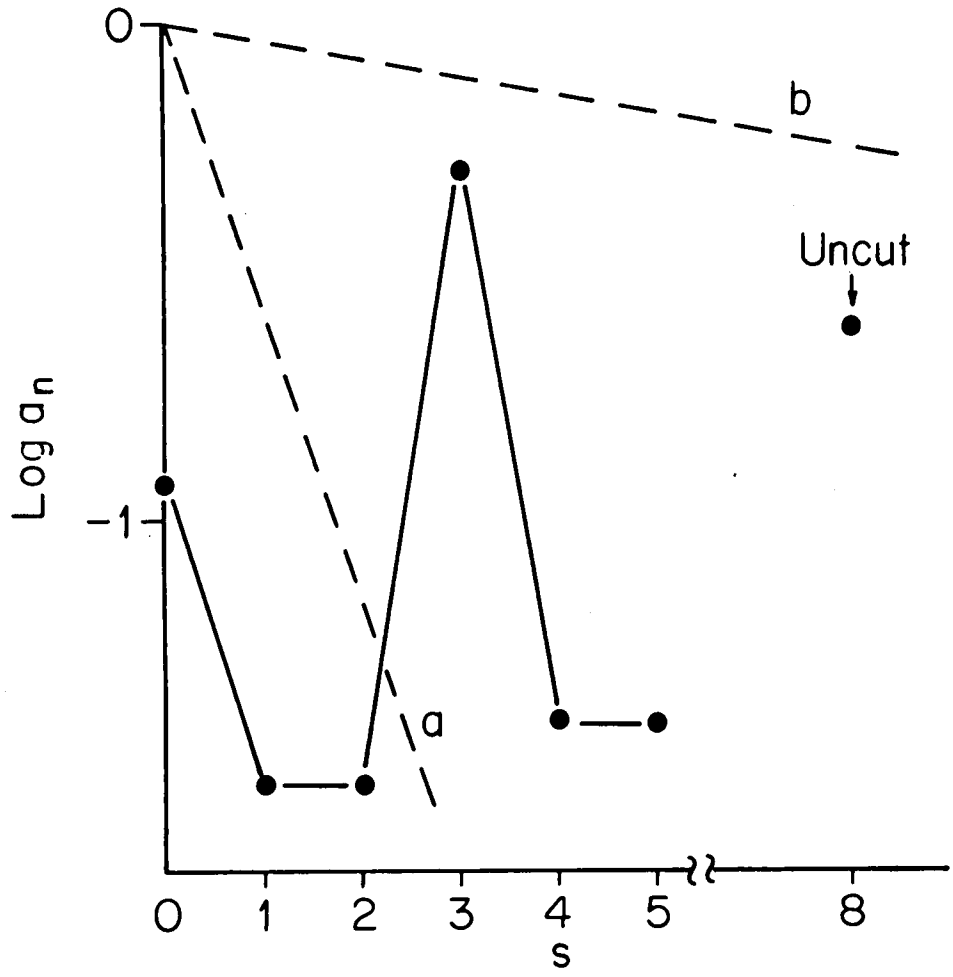


Figure 6. Distribution of HhaI sites in the HindIII octamer.

The relative amount a_n, of each Class of HhaI digestion products from the octamer was plotted versus s, the number of adjacent altered sites, as in Figure 5, page 158. Fragments of Classes 6 and 7 were not detected.

VITA

Nelwyn Tullos Christie was born in Ruston, Louisiana, on December 12, 1942. She received the Bachelor of Arts degree in August, 1964 from Northeast Louisiana University, Monroe, Louisiana. In September, 1974, she entered the predoctoral training program of The University of Tennessee-Oak Ridge Graduate School of Biomedical Sciences in the Biology Division of Oak Ridge National Laboratory, Oak Ridge, Tennessee. She received the degree of Doctor of Philosophy in August, 1979.