Intramolecular base composition heterogeneity of human DNA

Douglas L. Vizard †, Frank P. Rinehart *, Carol M. Rubin ** and Carl W. Schmid**


Received 22 July 1977

ABSTRACT

The intramolecular base composition heterogeneity of human DNA has been investigated by electron microscopic observations of partially denatured structures and by equilibrium solution thermal denaturation techniques. DNA sequences having an average length of less than 2000 base pairs are found to be heterogeneous in base composition. These heterogeneous sequences occupy a minimum of 67 to 81% of the human genome.

INTRODUCTION

It is well known that eukaryotic DNA is heterogeneous in base composition. It is possible that this heterogeneity arises from either a limited number of DNA components of different base composition or from a continuous distribution of sequences of different base composition. Thiery, Macaya and Bernardi have found that the equilibrium concentration distributions of several eukaryotic DNAs in buoyant CsCl density gradients can be represented as the sum of a few DNA components of different buoyant densities. These same components have been shown to adequately represent the density distributions of DNA fractions selected from preparative Ag⁺/Cs₂SO₄ buoyant density gradients. It was also noticed that the same major components are representative of both the mammalian and avian genomes. If these conclusions are correct, they have several important implications. As one example, it might be imagined that a component is composed of sequences which share common biological functions and that these functions are identical even in highly divergent organisms. It has also been proposed that these components are essentially homogeneous in base composition since their equilibrium band widths exhibit little density heterogeneity. At a certain molecular weight level, sequences of different base compositions must be covalently joined, and at this molecular weight level, we would expect these components to exhibit density heterogeneity in either their band widths or band shape. A molecular weight study of the putative
components suggested that they are homogeneous in base composition down to a DNA fragment size of 1 to 2 x 10^6 daltons.³

The proposal that eukaryotic DNA is paucidisperse in base composition seems inconsistent with the sequence interspersion patterns which have been observed for the eukaryotic genome. In all higher organisms that have been tested, repeated DNA sequences and single copy DNA sequences have been found to be mutually interspersed.⁴ The level of this interspersion falls into two broad classes as typified by the DNA sequence organizations observed in Xenopus and Drosophila.⁴ In the case of Drosophila and certain other insects, long repeated DNA sequences are found to be interspersed with very long single copy DNA sequences.⁵⁶ Among those organisms tested, the Drosophila sequence organization is the exception and the more prevalent DNA sequence organization is that typified by Xenopus.⁴ It is certain that most of the organisms studied by the buoyant density analysis described above would have a DNA sequence organization similar to that in Xenopus. For our present purposes, the most relevant feature of the DNA sequence organization in Xenopus is the short period interspersion pattern.⁴ Since human DNA will be used in this study, we will describe the short period interspersion pattern as observed in human. In at least 50% of the human genome, 300 nucleotide long repeated DNA sequences are interspersed with short single copy sequences.⁷ The average length of the short single copy sequences is less certain and our best estimate of this value is in the range of 1500 to 2000 nucleotides.⁷⁸ Thus human DNA is organized at the level of sequences having a combined molecular weight of approximately 1.4 x 10^6 daltons (1800 to 2300 bp).

It is difficult to reconcile the short period interspersion pattern with the proposal that eukaryotic DNAs are largely composed of a few components which are homogeneous in base composition. Unless the base compositions of the repeated sequences and single copy sequences involved in short period interspersion happen to be very similar, it would be expected that human DNA fragments having a molecular weight of 1.4 x 10^6 daltons would be heterogeneous in base composition. Available evidence indicates that the interspersed repeated sequences in human tend to be relatively more (G+C)-rich compared to the interspersed single copy sequences.⁹¹⁰ It seems likely that short fragments of human DNA should be polydisperse in base composition, a conclusion at odds with that reached by Bernardi and co-workers.³
In the hope of resolving these apparent discrepancies, we have undertaken a study of the intramolecular base composition heterogeneity of human DNA. This investigation has been conducted by complementary techniques in our separate laboratories. We have examined DNA molecules spread for electron microscopy under partially denaturing conditions. This approach provides an estimate of the sequence lengths characterizing intramolecular compositional heterogeneity and an estimate of the fraction of the human genome occupied by intramolecularly heterogeneous sequences. We have also employed high resolution thermal denaturation, which can detect the thermal transitions of discrete subcomponents present in the DNA, and variations of this procedure, which describe the intramolecular compositional heterogeneity of that fraction of the genome occupied by heterogeneous sequences.

METHODS AND MATERIALS

Human DNA for the electron microscope studies was extracted from human placental tissue as described by Schmid and Deininger and is the same sample which was used to characterize the lengths and spacing of repetitious sequences by electron microscopy.

Samples were spread for electron microscopy from denaturing concentrations of formamide. The hyperphase contained 0.1 M Tris, 0.01 M EDTA (pH 8.4) and indicated amounts of formamide (75%, 85% and 90%). The hypophase contained 0.01 M Tris, 0.001 M EDTA (pH 8.4) and 30% less formamide than the hyperphase. Under these conditions the hyperphase and hypophase are approximately isodenaturing. The partially-denatured DNA was picked up on a parlodian-coated copper grid, stained with uranyl acetate, and shadowed with platinum and palladium, 80%:20%.

All photomicrographs were taken at a magnification of 15,000 on an AEI 6B electron microscope. Length measurements were performed on a six-fold enlargement of the negative. All statistics were based on randomly visualized molecules. Single strand lengths were calibrated relative to denatured DNA spread from 85% formamide.

Human DNA for the solution denaturation studies was prepared from peripheral leukocytes (leukemic), human placenta, or HeLa S3 cells; these DNAs demonstrated no significant differences in melting profiles. The purification method involved nuclear isolation in the presence of EDTA and spermidine, RNase and pronase-SDS treatment, phenol extraction, alcohol precipitation, and subsequent dialysis into the buffered salt solution. DNA samples to be denatured in 50% DMSO (dimethylsulfoxide) solution were
dialized into a twice-concentrated salt solution and subsequently diluted
with DMSO. This method of purification gives quantitative yields of DNA
from cells of many origins, without creating significant DNA degradation;
further characterization of this procedure will be published separately
(D. Vizard, in preparation).

Details of the thermal denaturation procedure and data treatment can
be found in previous studies.\textsuperscript{13,14} The results of the solution denatura-
tion experiments are demonstrated here using the thermal derivative of
hyperchromicity at 270 nm, which more clearly describes the DNA sequence
heterogeneity. The profiles are corrected for solution heat dilation. The
method of heating the DNA was continuous at a rate (0.2-0.4°C/min) on which
the results do not depend; the experimental conditions are consistent with
those of equilibrium thermal denaturation.

\textbf{RESULTS}

Human DNA has been prepared for electron microscopy under conditions
employing sufficiently high concentrations of formamide to partially denature
the DNA.\textsuperscript{11,12} Under the conditions used, three classes of structures are
observed: completely denatured DNA, completely duplex DNA, and partially
denatured DNA, Fig. 1. Partially denatured structures are easily recog-
nized by the presence of either single-strand bubbles along the molecule
or single stranded forks terminating a molecule. The distinction between
completely undenatured and completely denatured molecules is usually clear,
but is occasionally a matter of judgement, Fig. 1. Our principal con-
clusions rely primarily on observations and measurements of the easily
identified partially denatured structures rather than on the less certain
distinction between double and single stranded DNA.

Random fields of DNA molecules, mounted under partially denaturing
conditions, were photographed and the lengths of all structures on these
fields were measured. The results of such measurements for each of three
different formamide concentrations are summarized in Table I. We note
that those molecules classified as being partially denatured are able
to completely reform base pairing under less denaturing conditions.
Qualitatively, the mass fraction of completely denatured molecules
increases as the formamide concentration is increased, and the fraction
of pure duplex decreases, Table I. The mass fraction of partially denatured
molecules goes through a maximum. There are at least two reasons why the
optimum value we observe, 67%, should be taken as the minimum mass fraction
which might be found in partially denatured structures. First, it is
Figure 1. Photomicrographs of typical molecules visualized in the 85% formamide spreading conditions. The bar shows the length equivalent to 1000 bp of duplex. Structures judged as being completely single-stranded are marked "S", as pure duplex, "D." Examples of very small bubbles are marked "B".
likely that the mass fraction of partially denatured structures goes through a maximum at some formamide concentration other than 85%, Table I. Second, if longer DNA samples were studied, it is probable that partially denatured molecules would be attached to molecules classified here as being completely duplex or completely denatured. The thermal stability studies discussed below suggest that the maximum observable mass fraction of partially denatured molecules is indeed higher than 67%.

There are two explanations for the existence of the partially denatured structures. The first is that the duplex regions have a higher (G+C) content than the single strand regions and consequently have a higher thermal stability. The second interpretation is that denatured and native regions are thermodynamically equivalent and that they are in dynamic equilibrium prior to mounting for electron microscopy. The probability of this second interpretation depends on the transition breadth of a homogeneous sequence under the conditions used for electron microscopy. This breadth has not been quantitatively determined for the electron microscope spreading conditions employed here. It is, however, instructive to compare the thermal stability of T7 DNA in solution to its stability under formamide mounting conditions. Using the rule that a 1% increase in formamide concentration has the same effect on the thermal stability of DNA as a 0.7°C increase in temperature, the thermal stability of duplex T7 DNA as judged by electron microscopy is found to be similar to the thermal stability of T7 DNA in solution. In solution, near the melting temperature, a 50% increase in the extent of denaturation of T7 DNA occurs over a 2.4°C increase in temperature. Under the electron microscope spreading conditions, a 50% increase in the extent of denaturation of T7 DNA occurs over a 4% increase in formamide concentration. Again using

Table I: SUMMARY OF RANDOM FIELD MEASUREMENTS OF HUMAN DNA SPREAD AT THREE FORMAMIDE CONCENTRATIONS.

<table>
<thead>
<tr>
<th>Per Cent Formamide</th>
<th>75%</th>
<th>85%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% (By Mass) of Bases Denatured</td>
<td>12.4</td>
<td>65.9</td>
<td>79.1</td>
</tr>
<tr>
<td>% (By Mass) of Molecules Completely Single Stranded</td>
<td>4.1</td>
<td>22.0</td>
<td>60.8</td>
</tr>
<tr>
<td>% (Mass) Pure Duplex</td>
<td>50.1</td>
<td>11.0</td>
<td>6.9</td>
</tr>
<tr>
<td>% (Mass) Partially Denatured</td>
<td>45.8</td>
<td>66.9</td>
<td>32.2</td>
</tr>
<tr>
<td>Percent of Bases Denatured in Partially Denatured Molecules</td>
<td>18.3</td>
<td>65.5</td>
<td>56.8</td>
</tr>
<tr>
<td>Number of Molecules Counted</td>
<td>119</td>
<td>136</td>
<td>275</td>
</tr>
</tbody>
</table>
the rule that 1% formamide corresponds to 0.7°C, the overall transition breadths of T7 DNA under these two sets of experimental conditions are in satisfactory agreement. Human DNA near its formamide melting concentration is half-denatured by a 10% increase in formamide concentration, Table I. This transition breadth is substantially larger than the 4% increase in formamide concentration needed to half-denature T7 DNA, which is in itself very heterogeneous in base composition.\textsuperscript{11} That the temperature change necessary to half-denature the human DNA depicted in the melting profile of Fig. 5 is 6.6°C, also correlates with the extent of denaturation found in Table I. We take these observations to mean that the breadth of the melting range in Table I is evidence for base composition heterogeneity of human DNA and that the partially denatured structures are largely the result of base composition heterogeneity. This interpretation is strongly supported by the thermal stability studies described below. According to this interpretation, the duplex regions in partially denatured structures are (G+C)-rich relative to the denatured regions.

We wish to analyze the sequence lengths which characterize the partially denatured structures. Since we find an optimum number of partially denatured structures in the 85% formamide spreading conditions, Table I, we will restrict our attention to this formamide concentration. The center-to-center distance between duplex regions on either side of a denaturation bubble should give a measure of the relative spacing of the most (G+C)-rich sequences within a DNA molecule. All structures encountered which contained such a center-to-center distance were measured and the distribution of lengths of such structures is shown in Fig. 2. The distribution is narrow and has a number average molecular length of 2450 bp (base pairs). Possible interpretations of this length distribution are considered in detail in the Discussion Section.

The duplex regions involved in partially denatured structures in 85% formamide have a number average length of 660 bp, Fig. 3. There is no reason to expect that the lengths observed in this distribution are especially unique. In lower formamide concentrations, the denaturation loop sizes are smaller and consequently the duplex regions are longer. Conversely, at higher formamide concentrations, the average length of duplex regions becomes shorter. We, therefore, expect and observe that the distribution of duplex lengths depends on the formamide concentration. We also wish to note that a large fraction of the duplex regions have a length of about 300 nucleotides, Fig. 3. This point will also be considered in
Figure 2. The center-to-center contour length between the two duplex regions on either side of a denaturation loop was measured in partially denatured structures observed in random fields of photomicrographs. The distribution includes 182 such structures observed in 51 separate molecules. The number average length is 2450 bp. Two structures with lengths between 10,000 bp and 11,000 bp are not illustrated. The DNA was prepared for electron microscopy using the 85% formamide spreading conditions.

Figure 3. The length of the duplex regions in random fields containing partially denatured structures, as defined by the presence of a junction of two single strand forks on each end of the duplex, is illustrated. (The 85% formamide spreading condition is used.) The number average length of 259 duplex regions in 80 different molecules is 660 bp. Six duplex lengths (2700 bp < L < 9600 bp) are not depicted on the histogram.
the Discussion Section. The distribution of the single strand lengths of the denaturation bubbles in partially denatured structures is presented in Fig. 4. This length is variable and must also depend on the formamide concentration as described above for the duplex length. At lower formamide concentrations, denaturation is just initiated and the single strand bubbles are small. At higher formamide concentrations two or more bubbles might merge giving rise to one larger bubble. The presence of multiple bubbles in partially denatured structures is very common, Fig. 1. The 227 bubbles reported in Fig. 4 were observed in 63 partially denatured molecules.

Occasionally large bubbles are found next to short duplex regions or small bubbles are found next to long duplex regions. This suggests the possibility that there might be a rather definite spacing of the (G+C)-rich sequences. This possibility would also be consistent with the relatively narrow distribution of lengths between the centers of adjacent duplex regions, Fig. 2. To further test this possibility, we compared the lengths of single strand bubble regions to the average lengths of their two immediately adjacent duplex regions using a scatter diagram.

![Figure 4](image-url)

**Figure 4.** The length of the denaturation bubbles as defined by the junctions with duplex regions on each end of two single strands is illustrated. This length was measured for partially denatured structures found in random fields of human DNA mounted for electron microscopy under the 85% formamide spreading conditions. The number average length of 227 such structures on 63 different molecules is 1749 nucleotides. Two bubbles (10,000 bp < L < 11,000 bp) are not depicted on the distribution.
correlation was found between these two values and we will not show this comparison here. It is not certain if this lack of correlation signifies that the length of single strand DNA within a bubble and the adjacent duplex length are uncorrelated, or if the data and analysis are too crude to reveal a relationship between bubble length and adjacent duplex length.

In summary, the electron microscope results indicate substantial intramolecular base composition heterogeneity occurring throughout 67% of the human genome at a molecular weight level of $1.5 \times 10^8$ daltons (2450 bp). These results are verified and extended by the following thermal stability studies. A high resolution melting profile of human DNA is shown in Fig. 5. A more complete analysis of the minimum number of components necessary to describe the heterogeneity in thermal stability of human DNA will be presented elsewhere (D. Vizard, manuscript in preparation). For the present purposes, it suffices to note that there are two broad classes of sequences in human DNA, the major (A+T)-rich class, which has

![Human DNA Re-denaturation](image)

Figure 5. The thermal derivative profile of preheated human DNA in 30 mM Na$^+$ (Cl$^-$, 1 mM cacodylate, pH 7, $10^{-4}$ EDTA). Every fifth point is designated. The profiles corresponding to the preheated samples have been calculated on the basis of the absorption of the DNA prior to preheating, so that the area described by a profile is directly proportional to the amount of duplex DNA that undergoes denaturation.
a lower thermal stability, and the minor (G+C)-rich class, which has higher thermal stability, Fig. 5.

We wish to test for the interspersion of sequences of different base compositions. The major class of (A+T)-rich sequences is denatured by preheating the DNA to 78°C, Fig. 5. Preheating times of between 5 and 10 minutes sufficed to achieve equilibrium partial denaturation. Upon cooling the preheated DNA, most of the denatured bases reform base pairs as demonstrated by the melting profile of the preheated sample. This is an example of intramolecular renaturation\textsuperscript{15}, and demonstrates the interspersion of (G+C)-rich sequences with the major class of (A+T)-rich sequences. The detailed question of possible ordering of special classes of (A+T)-rich sequences will be the subject of a subsequent study (D. Vizard, manuscript in preparation). That portion of the DNA (approximately 10%) which did not reform duplex molecules after preheating the sample to 78°C can be ascribed to two causes: It could be a fraction of the genome consisting of relatively homogeneous (A+T)-rich DNA sequences, or it may reflect artifactual strand separation resulting from the hydrolysis of partially denatured structures. More precise data demonstrate that the fraction of irreversibly denatured DNA is actually 5% of the genome and probably corresponds to human DNA satellites. Even when the DNA is preheated to 82°C, so that only 2.6% of the bases are paired, a substantial fraction of the DNA, 53%, is not strand separated, Fig. 5. As a control, it is demonstrated that the complete denaturation of the DNA (preheating to 90°C) leads to total strand separation, and it is also apparent that no significant bimolecular reassociation has occurred during the time course of these experiments, Fig. 5.

The melting profiles of the preheated samples demonstrate an interspersion of the minor class of (G+C)-rich sequences with the major class of (A+T)-rich sequences. A difficulty of interpreting these experiments purely in terms of intermolecular sequence heterogeneity is that elevated temperatures may cause hydrolysis and subsequent strand separation of partially denatured DNA. The effect of DNA hydrolysis can be minimized by carrying out the experiment in the presence of a denaturant; dimethylsulfoxide has been successfully used for this purpose. The partial denaturation and re-denaturation profiles for human DNA in 50% DMSO are shown in Fig. 6. The relationship between the melting temperature and the (G+C) content of selected bacterial DNAs of different base composition is superimposed in the same figure in order to provide an estimate of the compositional heterogeneity of the DNA in this solvent. This figure clearly
demonstrates that human DNA whose molecular weight continuity averages 24,000 nucleotide bases (8 x 10^8 daltons alkaline single strand), when preheated to 53.5°C, corresponding to 92.9% denaturation, is only about 11.7% strand separated. As a minimum value, we estimate that 81.2% (100% - 11.7% - 7.1%) of the genome must be associated with partially denatured structures. In arriving at this value, we exclude the 11.7% which strand separated and we also ignore any contribution to the partially denatured structures of the undenatured duplex fraction, which is 7.1%. Electron microscopy and the solution denaturation studies both demonstrate intramolecular base composition heterogeneity in a minimum of 67% to 81% of the genome.

A rough estimate of the difference in base composition between the (A+T)-rich regions and their interspersed (G+C)-rich regions can be obtained from the data in Fig. 6. At 53.5°C, sequences which are at least 55% in (G+C) composition are undenatured. The denatured regions have a

![Figure 6](image-url)

Figure 6. The high resolution thermal denaturation profile of human DNA in 50% (v/v) DMSO, 30 mM Na^+(Cl^-, 1 mM cacodylate, pH 7, 10^{-4} M EDTA). The closed symbols designate the original denaturation profile of the DNA, which terminates at 53.5°C, corresponding to 92.9% denaturation. The re-denaturation curve is designated by the open symbols and is based upon the absorption of the DNA prior to the original partial denaturation. The resolution in this profile (all data points are signified) is sufficient to describe the inherent features of the profile. The base composition dependence of the melting temperature of DNA in this solvent is also shown (right hand ordinate).
broad range of base compositions from about 24% to 55% (G+C) with an average value of about 39% (G+C).

**DISCUSSION**

In the case of simpler bacteriophage DNA, the high resolution thermal denaturation technique and the electron microscope denaturation technique are capable of providing very detailed sequence information. When applied to a more complex eukaryotic DNA, as in this study, the results of these methods are of necessity more limited. Both methods, however, provide substantial evidence for intramolecular base composition heterogeneity throughout at least 70% of the human genome. By electron microscopy, we find that the number average distance between the centers of two (G+C)-rich sequences is 2450 nucleotides and that the average single-strand loop size in 85% formamide is 1750 nucleotides.

The physical significance of the measured lengths is somewhat obscured by their dependence upon the extent of DNA denaturation. However, central to any argument of physical significance will be the statistical and thermodynamic aspects of the denaturation process. While theoretical estimates for random DNA sequences suggest that the average cooperatively denaturing sequence will encompass 400 bp, experience with bacteriophage DNA suggests that this length is about 900 bp. Note the 900 bp figure is closer to the mode of bubble lengths (Fig. 4). The distribution of observed lengths of denatured regions would in principle include neighboring bubbles that have merged to form enlarged regions of denaturation. For randomly arranged sequences averaging 900 bp in length, an average denaturation bubble length of 1800 nucleotides is expected if two-thirds of these sequences are denatured (corresponding to 85% formamide). Although the expected and measured bubble lengths agree, the measured duplex length (660 bp) is well below expectation.

We can attribute little significance to the above comparisons since the arguments are based upon randomly arranged sequences. Clearly, an alternative interpretation of the electron microscopy data is that the (A+T)-rich sequences are slightly clustered and the bubbles are correspondingly enlarged. This interpretation also implies that (G+C)-rich sequences are more dispersed than in randomly arranged DNA sequences, leading to smaller apparent duplex lengths. This explanation is consistent with relatively (G+C)-rich intermediate repeat sequences interspersed with (A+T)-rich single copy sequences. If this explanation is correct, then the partially denatured DNA molecules characterized here may be another mani-
Festation of the well-known short period sequence interspersion pattern observed in most eukaryotic genomes. We believe these conclusions can be tested only by characterizing the classes of sequences present in the short duplex regions of the partially denatured structures.

A major concern is the relationship of the intramolecular base composition heterogeneity observed here to the proposal that eukaryotic DNAs are composed of only a small number of homogeneous components. It is possible that DNA fragments having a length significantly greater than 2000 nucleotides could be intermolecularly homogeneous with respect to base composition, but at the same time could be intramolecularly heterogeneous with respect to base composition. In this case we need only imagine that all of the (A+T)-rich regions have approximately the same base composition and, similarly, all of the (G+C)-regions share the same base composition. These experiments do not test this possibility for long DNA fragments. For DNA fragment lengths of less than 2400 nucleotides (1.6 x 10^6 daltons) the results of this study predict a significant amount of intermolecular base composition heterogeneity. The proposal that eukaryotic DNAs consist of a small number of components which are relatively homogeneous in base composition is in part based on studies of DNA samples having a molecular weight in this range, 1 to 2 x 10^6 daltons. We wish to estimate the intermolecular base composition heterogeneity of DNA samples having a molecular weight of 1 x 10^6 daltons (1500 bp).

For simplicity, we assume that a DNA component consists of a regular interspersion of two types of sequences, one having a length of 500 nucleotides and a base composition of 57% (G+C) and the other having a length of 1500 nucleotides and a base composition of 39% (G+C). This numerical description of the intramolecular base composition heterogeneity corresponds to the values observed in this work. A more accurate description of the intramolecular base composition heterogeneity would include a distribution of DNA sequence lengths and distributions of base compositions in these sequences.

The intermolecular base composition heterogeneity of a sample also depends on the distribution of DNA fragment lengths in that sample. At present we do not have an accurate measure of these distributions so that it is necessary to represent these parameters by discrete values. We believe that this highly simplified model provides a realistic estimate of the intermolecular base composition heterogeneity that might result from random fragments of human DNA. If this component is randomly cleaved into 1500
nucleotide-long fragments (1 x 10^6 daltons), then half of the fragments would have a homogeneous base composition of 45% (G+C) and the other half would have a heterogeneous base composition of an average of 42% (G+C). For the purpose of analyzing the band width of DNA at sedimentation equilibrium, we require the mean square breadth of the base composition distribution. For the numerical model described above, the root mean square breadth is 1.9% (G+C). The width of a DNA band depends on its molecular weight and the mean square difference in base composition. For a sample with a molecular weight of 1 x 10^6 daltons (1.33 x 10^6 daltons for CsDNA) the base composition heterogeneity predicted here (± 1.9% (G+C)) would increase the band width of one component by only 6.3%. A 6% increase in the band width of an unresolved component would be hard to detect. We believe the analysis of the density gradient sedimentation equilibrium distributions would be insensitive to the level of intramolecular base composition heterogeneity observed here.

The density distributions of eukaryotic DNAs appear to be continuous. While it is possible to mathematically represent these distributions with a limited number of gaussian components, we conclude that these components must consist of molecules which are heterogeneous with respect to base composition. We have shown that the densities of the putative components would reflect compositional averages of internal sequences which have very different base compositions. There is no reason to suspect that molecules which happen to have similar base compositions would also have similar biological functions.

ACKNOWLEDGEMENTS

This research was supported in part by grants GM23067 and GM21346 awarded by the National Institute of General Medical Sciences, Department of Health, Education and Welfare.

To whom reprint requests should be addressed.

REFERENCES