Organization of the Unique and Repetitive Sequences in Feather Keratin Messenger Ribonucleic Acid†

Trevor J. Lockett,‡ David J. Kemp,§ and George E. Rogers*

ABSTRACT: Feather keratin messenger ribonucleic acid (mRNA) was transcribed into complementary deoxyribonucleic acid (cDNA) by using the reverse transcriptase of avian myeloblastosis virus and also by using *Escherichia coli* DNA polymerase I, and cDNA averaging 550 and 150 bases in length, respectively, was obtained. Studies on the kinetics of hybridization and reannealing of the two cDNAs and of the thermal stabilities and densities of the duplexes formed demonstrated quite distinct properties in the two cDNAs. In all respects, the long cDNA behaved as if it contained sequences which are repeated many times in inexact copies in the chick genome, while the short cDNA apparently behaved as unique sequences. Competition hybridization experiments demonstrated that the observed results could not be due to the different enzymes copying different subsets of the mRNA population. Furthermore, when short cDNA was used to prime cDNA synthesis using reverse transcriptase, the resulting long cDNA exhibited reassociation kinetics comparable to those for oligo(dT)-primed long cDNA. Of two possible models which are proposed in order to explain these observations, it is considered most likely that keratin mRNA molecules have distinct regions of unique and repetitive sequences and hence have diverged from each other in a regionally nonrandom manner. Long cDNA molecules contained a G+C-rich sequence which was not present in short cDNA molecules. Fragments of the chicken genome up to 140 kb in length which were able to reassociate with long keratin cDNA were also relatively rich in G+C, suggesting that keratin genes are clustered in the genome.

All eucaryotic mRNAs which have been characterized in detail are known to contain more bases than are necessary for coding, in addition to poly(A) (Gould & Hamlyn, 1973; Berns et al., 1972; Brownlee et al., 1973; Milstein et al., 1974; McReynolds et al., 1978). Since these untranslated regions are ubiquitous, they presumably have important biological functions. Early sequence studies (Milstein et al., 1974; Cheng et al., 1976; Proudfoot & Brownlee, 1974) demonstrated that rabbit β-globin, chick ovalbumin, and mouse immunoglobulin mRNAs all contain long untranslated sequences adjacent to the poly(A) tracts at the 3′ termini. Similarities in both sequence and possible secondary structure were evident indicating that there may be general structural features common to all eucaryotic mRNAs (Proudfoot & Brownlee, 1976). More recent sequence analysis has shown a high degree of homology between the 3′ noncoding sequences of human and rabbit α-globin mRNAs (Proudfoot et al., 1977) as well as between the 3′ noncoding sequences of the β-globin mRNAs of these two mammals (Proudfoot, 1977; Proudfoot & Longley, 1976). In addition, the 5′ noncoding sequences of rabbit and human α- and β-globin mRNAs respectively show a similar high degree of homology (Baralle, 1977a,b). The significance of this homology was discussed by Proudfoot & Brownlee (1976). These sequencing studies have led to the speculation that the untranslated sequences may be under less selective pressure than the translated sequences while having diverged far less than DNA species such as satellite DNA, whose function may be regarded as not sequence specific (Southern, 1975).

Feather keratin mRNA has distinctive properties relevant to studies on the organization of different classes of sequence. Feather keratin consists of a large number of homologous polypeptide chains, all containing ~100 amino acids but differing in primary structure by multiple amino acid substitutions (O'Donnell, 1973; O'Donnell & Inglis, 1974; Kemp et al., 1975). Highly purified keratin mRNA (Partington et al., 1973; Kemp et al., 1974a,b) was ~0.8 kb long, indicating the presence of ~0.5 kb in addition to the keratin coding sequence of 0.3 kb. Keratin mRNA acted as a template for synthesis of complementary DNA by the DNA polymerase of avian myeloblastosis virus (AMV) (Kemp et al., 1975; Kemp, 1975). Hybridization of keratin cDNA to keratin mRNA indicated that the sequence complexity of keratin mRNA was 25–35 times that expected for a single molecular species (Kemp, 1975). The apparent rate of hybridization was ~10 times greater when measured on hydroxylapatite (HAP) than when measured with nuclease S1, and the hybrids increased in thermal stability with increasing R_0t. These results were explained by the hypothesis that each of the multiple homologous keratin mRNA species contained a common (but nonidentical) sequence (reiterated sequence) covalently attached to a sequence so different in base sequence from that of any other keratin mRNA species (unique sequence) that it appeared unique in the chick genome under the experimental criteria used. It was estimated that there are ~100–240 copies of the reiterated sequence in the chick genome and suggested that this sequence is the keratin coding sequence (Kemp, 1975).

We show here that shorter keratin cDNA molecules obtained by transcribing keratin mRNA with *Escherichia coli* DNA polymerase I (Pol I) are apparently greatly enriched in the unique sequence relative to reiterated sequences and hence that unique sequences are located near the 3′ terminus of keratin mRNA.

A G+C-rich sequence was present on long but not on short keratin cDNA molecules. This G+C-rich sequence could also be detected in randomly sheared native DNA fragments of

† From the Department of Biochemistry, The University of Adelaide, Adelaide, South Australia 5001. *Received April 19, 1979; revised manuscript received July 24, 1979.* This work was supported by a postdoctoral fellowship from The University of Adelaide (to D.J.K.) and a grant from the Australian Research Grants Committee.

‡ Present address: Department of Genetics, Rockefeller University, New York, NY 10021.

§ Present address: The Walter and Eliza Hall Institute for Medical Research, Parkville, Melbourne, Australia.

0006-2960/79/0418-5654$01.00/0

the chick genome, and very long fragments containing sequences which could anneal to keratin cDNA were richer in G+C than the bulk of the chicken genome.

Materials and Methods

Preparation of mRNAs, cDNAs, and DNA. Keratin and rabbit globin mRNAs were prepared from EDTA-treated polysomes as described by Kemp et al. (1974a). AMV cDNAs were prepared and purified as described by Kemp (1975). Pol I cDNAs were synthesized in a system (50 μL) containing glycine buffer (67 mM, pH 9.2), KCl (50 mM), dATP, dTTP, and dGTP (50 μM each), [^{3}H]dCTP (20 μM) (sp act. 26.2 Ci/mmol), β-mercaptoethanol (10 mM), $MnCl_2$ (0.5 mM), dT_{10} (2.5 μg/mL), mRNA (0.6 μg), actinomycin D (100 μg/mL), and *E. coli* DNA polymerase I (1 μg) incubated for 90 min at 37 °C and purified as for AMV cDNA (O. Bernard, personal communication). Chick erythrocyte chromosomal DNA was prepared essentially as described by Marushige et al. (1968) and sonicated to an average length of ~500 bases as determined by electron microscopy. Chick DNA was labeled by nick translation with *E. coli* DNA polymerase I in the presence of DNase I as described by Schachat & Hogness (1973), using [^{3}H]dCTP (sp act. 26 Ci/mmol) as the only labeled nucleoside triphosphate. After being labeled, the DNA was purified as for cDNA. Pol I cDNA was elongated as described by Crawford & Wells (1976).

mRNA–cDNA Hybridizations. Hybridizations were carried out in 10 mM Tris, pH 7.4, 1 mM EDTA, 0.18 M NaCl, and 0.05% sodium dodecyl sulfate ($NaDodSO_4$) at 60 °C, nuclease S1 assays were carried out in 0.03 M NaOAc, 0.05 M NaCl, 0.001 M $ZnSO_4$, and 5% glycerol (pH 4.6 at 45 °C), HAP assays were carried out in 0.18 M sodium phosphate at 60 °C, and thermal stability determinations on HAP were as described by Kemp (1975).

Competition hybridizations were carried out in the same buffer as described above. Twenty-five microliter aliquots of a hybridization solution containing a 100:40:1 molar ratio of cold AMV cDNA/keratin mRNA/^{3}H-labeled Pol I cDNA or a 40:1 molar ratio of mRNA/Pol I cDNA were sealed in sterile 50-μL microcapillaries, placed in boiling water for 1 min, and then transferred to a 60 °C water bath. Aliquots were removed at different times, snap chilled in ice water, and stored frozen at −20 °C. When the last sample had been removed, earlier samples were thawed and assayed by nuclease S1 as previously described (Kemp, 1975).

cDNA–DNA Reassociations. Reassociation of keratin AMV, Pol I, and elongated Pol I cDNAs with sonicated chick erythrocyte DNA was performed at 60 °C in the same hybridization buffer as described above, using a ratio of 2000–3000 cpm of cDNA to 1 mg of DNA, which represents a three- to fivefold excess of genomic sequences, as described by Kemp (1975). Nuclease S1 assays at 37 °C in 0.03 M NaOAc, 0.3 M NaCl, 0.001 M $ZnSO_4$, and 5% glycerol (pH 4.6), and HAP assays and thermal stability determinations on HAP were carried out as described by Kemp (1975). Reassociation of nick-translated DNA was performed at a ratio of 3000 cpm to 1 mg of unlabeled DNA.

Preparation of Keratin AMV and Pol I cDNA Duplexes for Density Gradient Centrifugation and Centrifugation Conditions. AMV or Pol I cDNA duplexes were prepared by incubation to a C_0t of 2×10^4 mol s L^{-1} as above. Samples (100 μL) containing 1 mg of DNA and 3000 cpm of cDNA were added to 4.0 mL high-salt nuclease S1 assay buffer and incubated at 37 °C for 15 min with 100 units of nuclease S1. The reaction was stopped by addition of 0.25 mL of 0.4 M sodium phosphate, pH 7.0, and the mixture was dialyzed against three changes of buffer (10 mM Tris-HCl, pH 8.4, 1 mM EDTA, and 10 mM NaCl). Chick DNA (11 kb) was added to a total of 1 mg of DNA, and buffer and solid CsCl were added to a volume of 12 mL and a density of 1.6 g/cm^3. The mixtures were chilled on ice, and 500 μg of actinomycin D (a gift from Merck) was added. The refractive index was adjusted to 1.3903 at 20 °C. The mixtures were centrifuged in a Beckman Ti50 rotor at 32000 rpm for 60 h at 20 °C. The gradients were fractionated by upward displacement, and 0.5-mL fractions were collected. The A_{260} of each fraction was determined, and the amount of DNA in each fraction was made constant by addition of sonicated calf thymus DNA. Actinomycin D was extracted with CsCl-saturated 2-propanol. H_2O (1.5 mL) and 20% trichloroacetic acid (Cl_3AcOH) containing 1% sodium pyrophosphate (2.0 mL) were added to each fraction. The precipitates were collected on Whatman GF/A filters, rinsed twice with 5% Cl_3AcOH–1% sodium pyrophosphate and twice with ether, dried, and counted in a toluene-based scintillation fluid.

In the density analysis of different size classes of chick DNA, actinomycin D–CsCl gradients using 1 mg each of 0.5-, 11-, and 140-kb DNA were prepared as described above. Netropsin sulfate–CsCl gradients were prepared in the same way except that the gradients were made up to a density of 1.64 g/cm^3 and the ratio of DNA/netropsin sulfate was 1:1.6 (w/w). Centrifugation conditions were as above. Netropsin sulfate was a gift from Dr. H. Thrum.

Reassociation of Keratin cDNA with Chick DNA after Centrifugation in CsCl. Gradients were fractionated into 0.5-mL fractions by upward displacement. After determinations of A_{260} and density, all fractions were made up to the same DNA concentration by the addition of calf thymus DNA and divided into two equal amounts to provide duplicates. Water (2.5 mL) was added and the DNA precipitated with 2.5 volumes of ethanol overnight at −20 °C. DNA was collected by centrifugation (3000 rpm, 4 °C, 30 min), dried in vacuo, dissolved in 50 μL of 0.18 M NaCl, 0.01 M Tris-HCl, pH 7.0, 0.001 M EDTA, and 0.5% $NaDodSO_4$, and sonicated for 90 s. About 2000 cpm of keratin AMV cDNA was added to each fraction, and the mixtures were sealed in 100-μL capillary tubes, placed in boiling water for 2 min, and then incubated at 60 °C until the chick DNA in the peak fraction had reached a C_0t of 10^3 mol s L^{-1}. Each mixture was then assayed for resistance of the cDNA to S1 nuclease by using low-salt S1 assay buffer (Kemp, 1975). Duplicates were averaged.

Results

Reverse Transcriptase and DNA Polymerase I Give Different Size cDNA Copies of the Same mRNA Population. Keratin and rabbit globin cDNAs were prepared by using reverse transcriptase (AMV cDNA) and also by using *E. coli* DNA polymerase I (Pol I cDNA). The lengths of the keratin AMV cDNA and Pol I cDNA were estimated by polyacrylamide gel electrophoresis in the presence of formamide (Pinder et al., 1974) with RNA markers (Figure 1). Both cDNA preparations were heterogeneous with regard to length. However, the main peak of keratin AMV cDNA was ~0.55 kb long while that of keratin Pol I cDNA was ~0.15 kb long. Globin AMV and Pol I cDNAs were ~0.5 and ~0.16 kb long, respectively. It has previously been reported that cDNAs transcribed with Pol I are shorter than those transcribed with reverse transcriptase (Milstein et al., 1974; Proudfoot & Brownlee, 1974; Modak et al., 1973; Gulati et al., 1974; Rabbitts, 1974; Crawford et al., 1977). Synthesis of cDNA

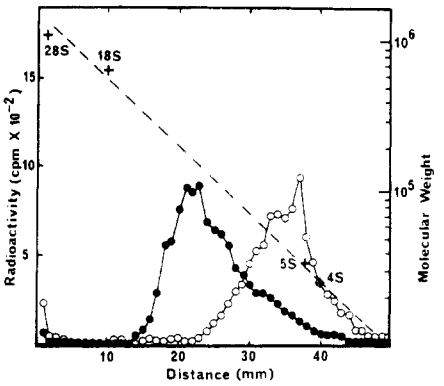

FIGURE 1: Polyacrylamide gel electrophoresis of keratin AMV cDNA and Pol I cDNA in the presence of formamide. Aliquots of AMV cDNA and Pol I cDNA were coelectrophoresed with the marker RNAs on 4% polyacrylamide gels in formamide containing 0.02 M barbital, pH 9.0, by the procedure of Pinder et al. (1974). After electrophoresis for 1 h at 5 mA/gel, the gels were stained with 0.05% Toluidine blue, photographed, and cut into 1-mm slices. The slices were counted in toluene–NCS scintillation fluid. The counts per minute scale for Pol I cDNA was half that shown. (●) AMV cDNA; (○) Pol I cDNA. The positions of markers (rabbit reticulocyte 28S, 18S, 5S, and 4S RNAs) are shown. Similar results were obtained by using *E. coli* 5S, 16S, and 23S RNA as markers (not shown).

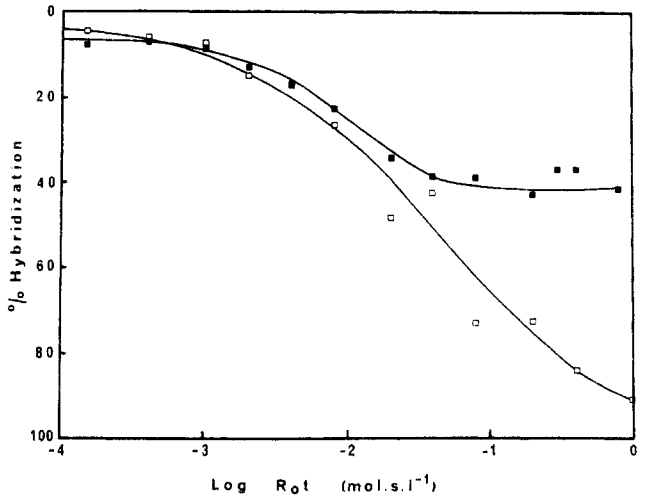

FIGURE 2: Hybridization kinetics of ^{3}H-labeled Pol I cDNA to keratin mRNA in the presence and absence of cold competing AMV cDNA. Hybridization solutions containing a 100:40:1 molar ratio of cold AMV cDNA/mRNA/^{3}H-labeled Pol I cDNA and a 40:1 molar ratio of mRNA/^{3}H-labeled Pol I cDNA were divided into 25-μL aliquots, and hybridizations were performed as described under Materials and Methods. Hybridizations were assayed by nuclease S1, as described by Kemp (1975). Curves were normalized to remove background, and the normalization factor is shown in parentheses below for each curve. The background was determined as the percentage of cDNA resistant to nuclease S1 digestion or which bound to HAP at zero time. The normalization factor was (% hybrid − % background)/(100 − % background). Hybridization in the presence of cold AMV competitor (■) (5%); hybridization in the absence of cold AMV competitor (□) (5%).

by both enzymes was completely dependent on the presence of oligo(dT) in the reverse transcription system and therefore presumably initiates at the poly(A) region of mRNA (Milstein et al., 1974; Proudfoot & Brownlee, 1974; Crawford et al., 1977). Since the poly(dT) segments of chick globin AMV and Pol I cDNAs, prepared under conditions identical with those used here, average 18–20 bases in length (Crawford et al., 1977), the keratin AMV cDNA and Pol I cDNA are presumably complementary to regions of keratin mRNA, extending on average ~0.53 and ~0.13 kb from the poly(A) tract, respectively.

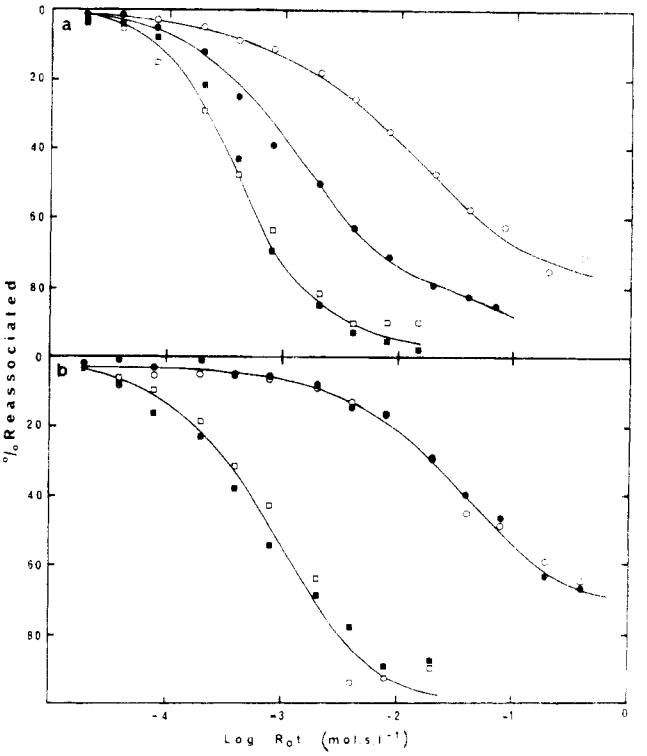

FIGURE 3: Hybridization of keratin and rabbit globin mRNA to keratin and globin AMV cDNAs and Pol I cDNAs. Results for hybridizations with keratin mRNA are the average of two or three independent experiments with different batches of mRNA and cDNA. Normalization factors determined as in Figure 2 are shown below in parentheses. (a) AMV cDNA hybridizations: (○) keratin AMV cDNA–keratin mRNA, assayed with nuclease S1 (0%); (●) keratin AMV cDNA–keratin mRNA, assayed with HAP (10%); (□) globin AMV cDNA–globin mRNA, nuclease S1 (5%); (■) globin AMV cDNA–globin mRNA, HAP (20%). (b) Pol I cDNA hybridizations: (○) keratin Pol I cDNA–keratin mRNA, nuclease S1 (3%); (●) keratin Pol I cDNA–keratin mRNA, HAP (4%); (□) globin Pol I cDNA–globin mRNA, nuclease S1 (0%); (■) globin Pol I cDNA–globin mRNA, HAP (10%).

For certainty that the AMV reverse transcriptase and DNA polymerase I were copying the same set of mRNA molecules, the AMV and Pol I cDNAs were tested for the presence of common sequences by studying the competition between the two cDNAs for sequences in the mRNA during hybridization. Figure 2 shows the hybridization of ^{3}H-labeled Pol I cDNA to mRNA in the presence and absence of cold competing AMV cDNA. The molar ratio of cold AMV cDNA/mRNA/labeled Pol I cDNA was 100:40:1. Under these conditions, if Pol I cDNA and AMV cDNA represented copies of the same mRNA population, then in any given hybridization event the chance of a Pol I cDNA molecule forming the hybrid relative to that of an AMV cDNA molecule is 40/100 or 0.4. Thus, the percentage of Pol I cDNA hybridized to mRNA at completion in the presence of cold competitor should be 40% of that obtained in the absence of competitor (Gambino et al., 1974). This theoretical value is in complete agreement with the value derived experimentally. The Pol I and AMV cDNAs therefore represent different length copies of the same mRNA population.

Keratin cDNAs with Different Lengths Hybridize to Keratin mRNA with Very Different Kinetics. Figure 3 shows the hybridization of keratin AMV cDNA and Pol I cDNA to keratin mRNA, each assayed both by resistance of the hybrids to nuclease S1 and by HAP chromatography. The hybridization of rabbit globin AMV cDNA and Pol I cDNA to rabbit globin mRNA is also shown to provide kinetic standards.

As described in detail previously (Kemp, 1975), keratin AMV cDNA hybridized to keratin mRNA with a $R_0t_{1/2}$ value of $\sim 9 \times 10^{-3}$ mol s L^{-1} when assayed with nuclease S1 (Figure 3a) compared with the $R_0t_{1/2}$ value of $\sim 4 \times 10^{-4}$ mol s L^{-1} for hybridization of globin mRNA to globin AMV cDNA. Keratin and globin Pol I cDNAs (Figure 3b) hybridized with $R_0t_{1/2}$ values of 3×10^{-2} and 8×10^{-4} mol s L^{-1}, respectively. These higher $R_0t_{1/2}$ values are consistent with the length of the Pol I transcripts being shorter than the AMV transcripts. Since the α- and β-globin mRNAs and cDNAs do not cross-hybridize (Kacian et al., 1973; Housman et al., 1973), the total sequence length of globin mRNA is about $2 \times 600 = 1200$ nucleotides, of which about $2 \times 150 = 300$ are presumably represented in globin Pol I cDNA. By a correction for this amount of the mRNA which can form hybrids, the $R_0t_{1/2}$ of globin mRNA with Pol I cDNA is equal to $(8 \times 10^{-4})(300/1200) = 2.0 \times 10^{-4}$ mol s L^{-1}. Similarly, the corrected $R_0t_{1/2}$ for the reacting sequences in keratin mRNA with keratin Pol I cDNA is equal to $(3 \times 10^{-2})(150/800) = 5.6 \times 10^{-3}$ mol s L^{-1}. The number of non-cross-reacting sequences 150 bases long in keratin mRNA is therefore $(5.6 \times 10^{-3}/2.0 \times 10^{-4})(300/150) = 56$. This estimate of 56 different keratin mRNA species is somewhat higher than estimated previously (25–35 different species) with keratin AMV cDNA (Kemp, 1975). This result is consistent with the previous conclusion that the AMV cDNA transitions involved both unique and repeated sequences.

When assayed by HAP chromatography (parts a and b of Figure 3), the rates of hybridization of globin mRNA to either globin AMV cDNA or globin Pol I cDNA were very similar to the rates of the same reactions measured with nuclease S1. When assayed on HAP, keratin AMV cDNA gave two transitions, as observed previously (Kemp, 1975). The $R_0t_{1/2}$ of the major transition was $\sim 9 \times 10^{-4}$ mol s L^{-1}, 10 times lower than that determined with the nuclease S1 assays. In contrast, however, the rate of hybridization of keratin mRNA to keratin Pol I cDNA was indistinguishable whether measured with nuclease S1 or by HAP chromatography (Figure 3b).

Thermal Stabilities of Long Keratin cDNA–mRNA Hybrids Can Be Lower Than Those of Short cDNA–mRNA Hybrids. As shown previously, keratin AMV cDNA hybrids contained two components with T_m values of ~ 72 and ~ 90 °C, and the average thermal stability of the hybrids increased with increasing R_0t (Figure 4a). The same data are shown in Figure 4b, replotted after multiplying the points in each curve by the fraction of the total cDNA which had bound to HAP at that R_0t value. When plotted in this form, it is clear that the total amount of material of $T_m = 80$ °C or less increased during the early part of the reaction ($R_0t = 8 \times 10^{-4}$ and 8×10^{-3} mol s L^{-1}) and then decreased late in the reaction (by $R_0t = 8 \times 10^{-2}$ mol s L^{-1}). Therefore, some of the initial hybrids of low thermal stability must have been transferred into hybrids of higher thermal stability. Points taken at R_0t values between those shown confirmed this observation.

Keratin Pol I cDNA hybrids (Figure 4c) changed in thermal stability with increasing R_0t also, but much less than the AMV cDNA hybrids. Replots of these curves (Figure 4d) demonstrate that there was no significant transfer of hybrids from low thermal stability to high thermal stability, in contrast to the result with AMV cDNA hybrids. The maximum T_m of keratin Pol I cDNA hybrids was ~ 83 °C at $R_0t = 4 \times 10^{-1}$ mol s L^{-1}, compared with 90 °C for keratin AMV cDNA hybrids at $R_0t = 8 \times 10^{-2}$ mol s L^{-1}. The thermal stability of keratin Pol I cDNA hybrids was ~ 3–4 °C lower than that of globin Pol I cDNA hybrids (87 °C), which in turn was ~ 4 °C lower than that of globin AMV cDNA hybrids (91 °C; Kemp, 1975). Since the lowered thermal stability of globin Pol I cDNA hybrids is undoubtedly due to their low molecular weights (Modak et al., 1973), the T_m of the keratin Pol I cDNA hybrids is equivalent to at least 87 °C for longer sequences. The remaining difference of 3 °C may result from differences in the G+C content of the different molecules (see below).

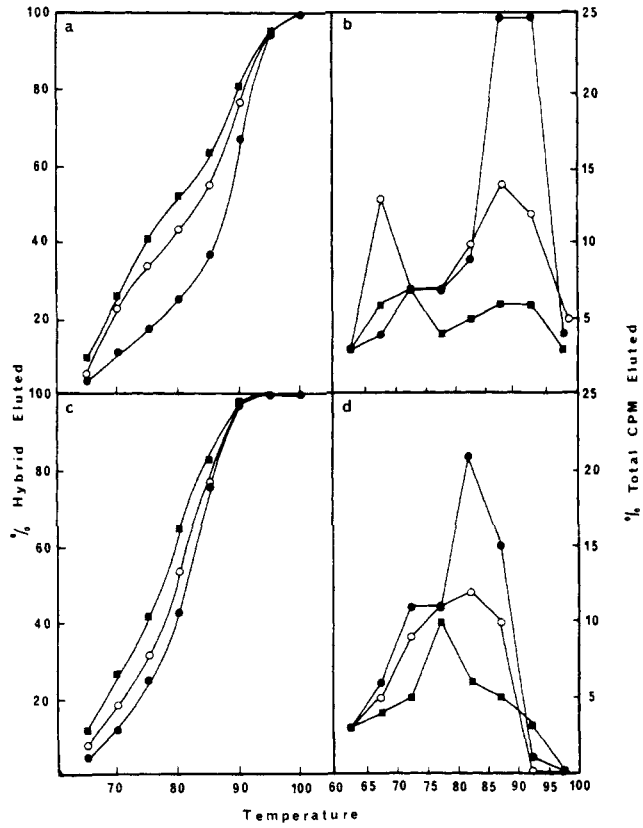

FIGURE 4: Thermal stabilities of keratin AMV cDNA and Pol I cDNA hybrids. Hybridization mixtures of mRNA and cDNA were incubated to the various R_0t values shown, and the thermal stabilities of the hybrids formed at each R_0t value were determined on hydroxylapatite exactly as described (Kemp, 1975). (a) Keratin AMV cDNA hybrids after incubation to R_0t values (mol s L^{-1}) of (■) 8×10^{-4}, (○) 8×10^{-3}, and (●) 8×10^{-2}. (b) The same data as in (a) replotted to show the percentage of input cDNA eluted at the temperature intervals indicated. It should be noted that the sum of the individual values in each plot equals the proportion of the input cDNA hybridized at the R_0t values indicated. (c) Keratin Pol I cDNA hybrids after incubation to R_0t values of (■) 2×10^{-2}, (○) 8×10^{-2}, and (●) 4×10^{-1}. (d) The same data as in (c) replotted as for (b).

Only the Long Keratin cDNA Hybridizes to Repetitive Sequences in the Chick Genome. Figure 5 shows the reassociation kinetics of keratin AMV cDNA and Pol I cDNA in the presence of a vast excess of chick erythrocyte DNA. Keratin AMV cDNA reassociated in a biphasic curve with major transitions at $C_0t_{1/2}$ values of $\sim 4 \times 10^1$ and $\sim 6 \times 10^3$ mol s L^{-1} when reassociation was measured by resistance of the cDNAs to digestion with nuclease S1 (Figure 5a). In contrast, less than 10% of keratin Pol I cDNA had reassociated by a $C_0t_{1/2}$ of 4×10^1 mol s L^{-1} and there was only one major transition with a $C_0t_{1/2}$ of 9.5×10^2 mol s L^{-1}, a value indistinguishable from that of the unique sequence fraction of chick erythrocyte DNA (Figure 5a).

When assayed on HAP (Figure 5b), keratin AMV cDNA again reassociated in a biphasic curve, the major transitions occurring at $C_0t_{1/2}$ values of $\sim 7 \times 10^0$ and $\sim 9 \times 10^2$ mol s L^{-1}, respectively. Keratin Pol I cDNA reassociated with a single transition at a $C_0t_{1/2}$ of $\sim 4.7 \times 10^2$ mol s L^{-1}, a value

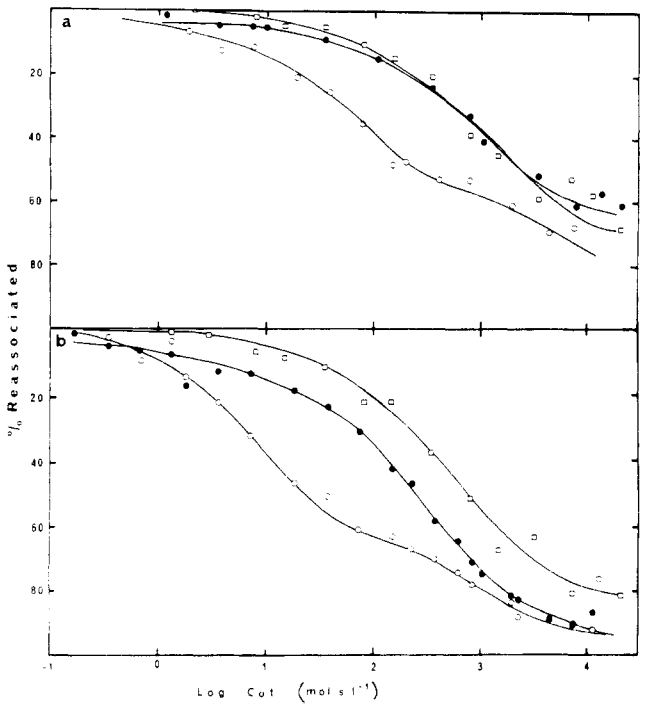

FIGURE 5: Reassociation of keratin AMV and Pol I cDNAs with chick erythrocyte chromosomal DNA. Normalization factors determined as in Figure 2 are shown below in parentheses. (a) Nuclease S1 assays: (○) keratin AMV cDNA (0%); (□) keratin Pol I cDNA (10%); (●) chick erythrocyte DNA, labeled by nick translation (14%). (b) HAP assays: (○) keratin AMV cDNA (10%); (□) keratin Pol I cDNA (11%); (●) chick erythrocyte DNA (A_{260}) (18%).

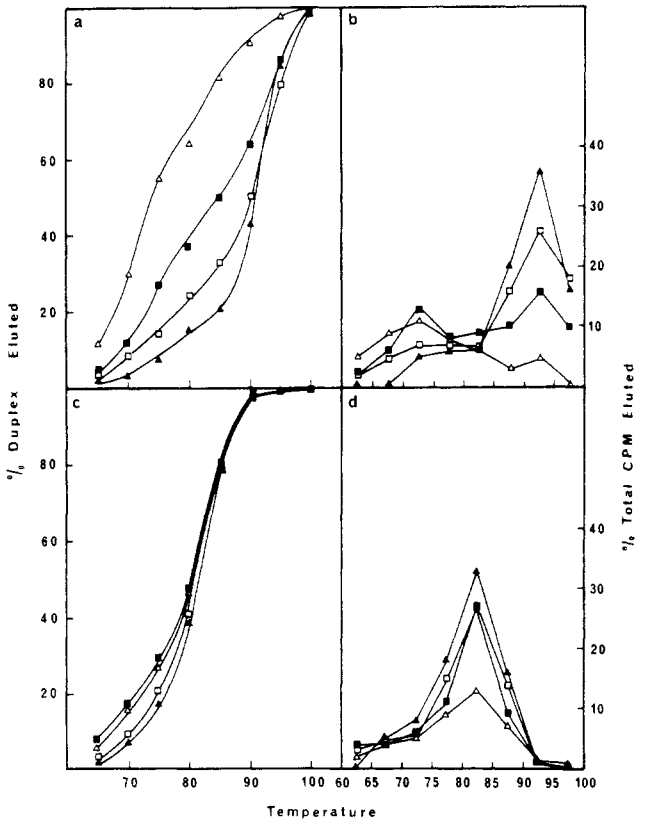

FIGURE 6: Thermal stabilities of keratin AMV and Pol I cDNA reassociated duplexes. Reaction mixtures identical with those of Figure 5 were incubated to the various C_0t values shown and the thermal stabilities determined. (a) Keratin AMV cDNA duplexes after incubation to C_0t values (mol s L^{-1}) of (△) 1.9×10^{-1}, (■) 5.7×10^2, (□) 3.75×10^3, and (▲) 1.14×10^4. (b) The same data as in (a) replotted to show the percentage of input cDNA eluted at the temperature intervals indicated. It should be noted that the sum of the individual values in each plot equals the proportion of the input cDNA reassociated at the C_0t values indicated. (c) Keratin Pol I cDNA duplexes after incubation to C_0t values (mol s L^{-1}) of (△) 4.9×10^2, (■) 9.3×10^2, (□) 3.1×10^3, and (▲) 1.6×10^4. (d) The same data as in (c) replotted as in (b).

again indistinguishable from that of the unique sequence fraction of chick erythrocyte DNA (Figure 4b). The differences in reassociation of AMV and Pol I cDNAs were reproducible with different batches of each cDNA and also when the cDNAs were each annealed with the same batch of sheared chick erythrocyte DNA, ruling out the possibility that the differences were caused by variations in the fragment size of the chick erythrocyte DNA.

The thermal stabilities of the keratin AMV cDNA duplexes changed markedly with C_0t (Figure 6a), two major components with T_m values of ~73 and ~93 °C being evident. Replots of these curves (Figure 6b) demonstrate clearly that more than 60% of the duplexes with T_m values less than 85 °C which were present at $C_0t = 1.9 \times 10^1$ mol s L^{-1} had transferred into duplexes of high thermal stability by $C_0t = 1.14 \times 10^4$ mol s L^{-1}. In contrast, there was a much smaller change in thermal stability of keratin Pol I cDNA duplexes with C_0t (Figure 6c) and no significant transfer of duplexes from low thermal stability to high thermal stability (Figure 6d). One major component with a T_m of ~83 °C was evident, although in addition there were some duplexes with T_m values lower than 70 °C.

When ^{3}H-labeled Pol I cDNA was used as primer in an AMV cDNA synthesizing system containing all four unlabeled deoxynucleoside triphosphates (Crawford & Wells, 1976), labeled products ranging from 200 to 650 nucleotides in length were obtained, as estimated from their sedimentation velocity through linear 5–20% sucrose gradients relative to AMV cDNA and Pol I cDNA markers (results not shown). The longer molecules, comprising 20% of the total radioactivity, were selected off of sucrose gradients, and the HAP-assayed reassociation kinetics of these elongated molecules were compared with those of the original Pol I cDNA (Figure 7a). It is apparent from these data that extension of the short probe has caused an increased rate of reassociation. The reassociation kinetics observed for the elongated probe are comparable with those obtained with AMV cDNA with a broad transition replacing the biphasic one shown in Figure 5b.

The radioactive label is confined to the 150 nucleotides adjacent to the 5′ terminus of these elongated molecules, and since the Pol I cDNA shows unique reassociation kinetics (Figure 5b), the additional cold nucleotides added by the AMV polymerase, as directed by the mRNA template, must represent a reiterated sequence in the chick genome. Elongation products with lengths intermediate between 200 and 600 bases showed increasing rates of reassociation with increasing length (Figure 7b). These results suggest that the 3′-terminal 150 nucleotides of keratin mRNA represent unique sequences in the chick genome, covalently linked to reiterated sequences located further toward the 5′ end of the messenger.

Long Keratin cDNA Molecules Contain a G+C-Rich Sequence Which Is Not Present in Short Keratin cDNA Molecules. Since the T_m values of AMV cDNA and Pol I cDNA hybrids showed a greater difference than expected on the basis of cDNA length, it was possible that these cDNAs differed in G+C content. For determination of whether this was so, the density of duplexes of keratin AMV cDNA and Pol I cDNA formed by reassociation of the cDNAs with chick erythrocyte DNA to high C_0t was investigated. The duplexes were treated with nuclease S1 before actinomycin D–CsCl gradient centrifugation in order to remove single-stranded

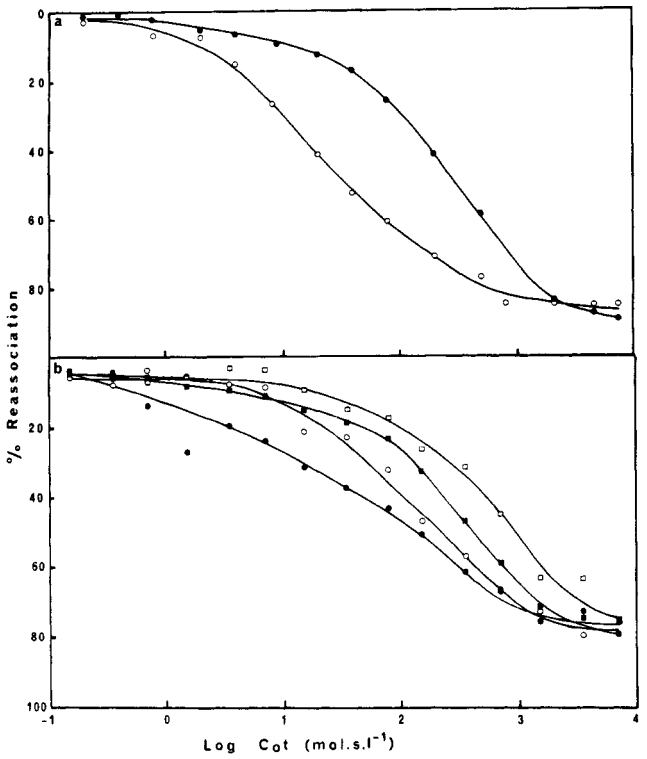

FIGURE 7: Reassociation of AMV reverse transcriptase elongated Pol I cDNA with chick erythrocyte chromosomal DNA. All assays were carried out on HAP. The average size of the elongated Pol I cDNA size classes was estimated from their velocity of sedimentation through linear 5–20% sucrose gradients relative to AMV cDNA and Pol I cDNA markers run on parallel gradients. Normalization factors as determined in Figure 2 are shown below in parentheses. (a) (●) Pol I cDNA (10%); (○) 550-base elongated Pol I cDNA (13%). (b) (□) A_{260} (18%); (■) 220 bases (0%); (○) 350 bases (8%); (●) 550 bases (21%).

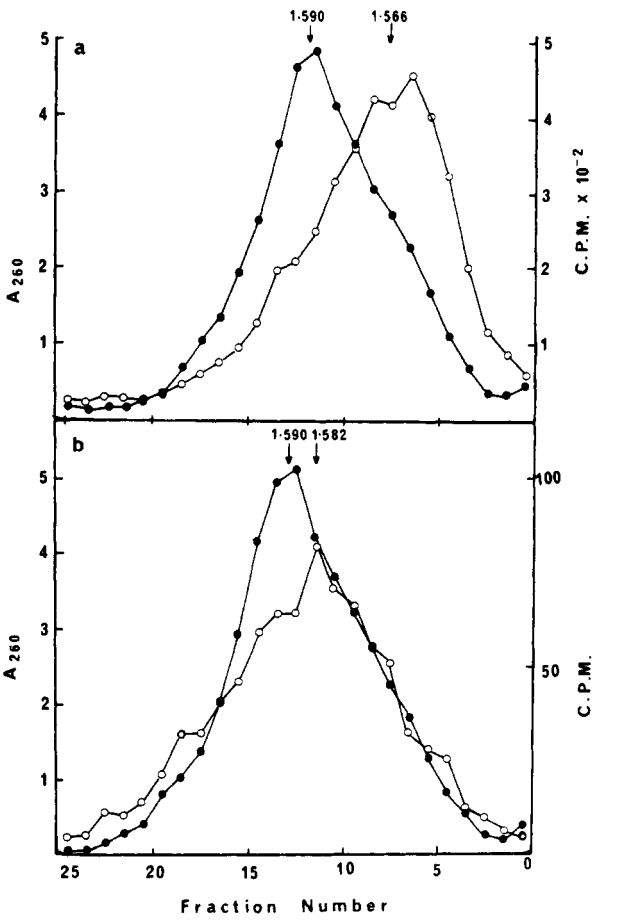

FIGURE 8: Centrifugation of keratin AMV and Pol I cDNA reassociated duplexes on CsCl–actinomycin D gradients. Duplexes formed at $C_0t = 2.0 \times 10^4$ mol s L^{-1} were treated with nuclease S1 and then centrifuged on CsCl–actinomycin D gradients as described under Materials and Methods. (a) (○) Keratin AMV cDNA duplexes (cpm); (●) chick erythrocyte DNA (A_{260}). (b) (○) Keratin Pol I cDNA duplexes (cpm); (●) chick erythrocyte DNA (A_{260}).

regions, mismatched regions, and oligo(dT) tails on the cDNA molecules.

The main band of reassociated chick erythrocyte DNA had a density of 1.590 g/cm^3 on actinomycin D–CsCl gradients (Figure 8). Keratin AMV cDNA duplexes banded at an average density of ~1.566 g/cm^3 (Figure 8a). In contrast, keratin Pol I cDNA duplexes banded at an average density of 1.582 g/cm^3 (Figure 8b). Since actinomycin D binds preferentially to G+C-rich sequences (Kersten et al., 1966), resulting in a displacement of such sequences to the light side of CsCl gradients, keratin AMV cDNA must contain a G+C-rich sequence which is not present in keratin Pol I cDNA. Although there was slight variation in the apparent density of the main-band DNA in independent experiments, the variation in density difference between main-band DNA and cDNA duplexes was 0.002 g/cm^3 or less in three independent experiments. When the chick erythrocyte DNA was replaced with calf thymus DNA, no radioactivity above background was detectable on the gradients (results not shown). Since the duplexes had been treated with nuclease S1 prior to centrifugation, we assume that the apparent densities of the cDNA duplexes were not significantly affected by the presence of single-stranded or mismatched regions or by the presence of oligo(dT) tails.

Fragments of the Chick Genome Which Contain Keratin Genes Are Also Rich in G+C. Since the above results demonstrated that long keratin cDNA molecules contained a G+C-rich sequence, this sequence should be demonstrable in sheared fragments of native DNA from the chick genome. The buoyant density in actinomycin D–CsCl gradients of sheared fragments of chick DNA-containing sequences complementary to keratin AMV cDNA was therefore investigated. Chick erythrocyte DNA samples were sheared to average fragment sizes of 0.55, 11, and 140 kb, respectively, and subjected to equilibrium centrifugation on CsCl in the presence of actinomycin D. The gradients were fractionated, and each fraction was assayed for the presence of keratin sequences by hybridization with keratin AMV cDNA, as described under Materials and Methods.

Chick DNA fragments of length 0.5 kb banded on actinomycin D–CsCl gradients in a broad distribution with the peak at a density of 1.598 g/cm^3 (Figure 9a). These fragments which could reanneal with keratin AMV cDNA banded with a distinctively lower density, the peak being at 1.580 g/cm^3, a result compatible with the observations reported above for reassociated duplexes. Since a significant fraction (up to 20%) of the keratin AMV cDNA had reannealed to the peak fraction and the maximum resistance of keratin cDNA duplexes to nuclease S1 under the digestion conditions used here was 50–55% (Kemp, 1975), then the density distribution observed must reflect the properties of at least 40% of the cDNA preparation rather than those of a minor contaminant. Since the cDNA preparations never showed any detectable complementarity to rRNA and the keratin mRNA used as template was judged to be >95% pure by cell-free translation (Kemp et al., 1974b), the G+C-rich genomic sequences must be segments of keratin genes.

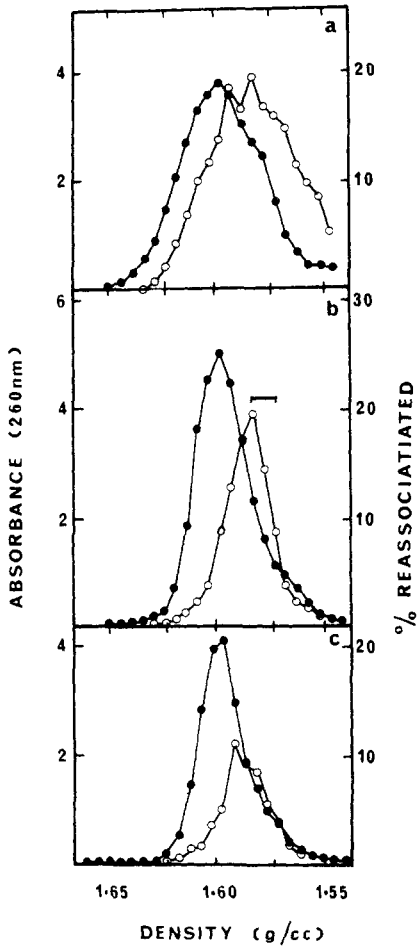

FIGURE 9: Actinomycin D–CsCl gradients showing the change in density of DNA-carrying keratin genes with length of the DNA. 1 mg of chick DNA of length 0.5 (a), 11 (b), and 140 kb (c) was centrifuged to equilibrium in CsCl in the presence of actinomycin D (DNA/actinomycin D = 2:1 w/w). The position at which the keratin genes band was determined by reassociation studies using keratin AMV cDNA and nuclease S1. (●) A_{260}; (○) percent of counts reassociated.

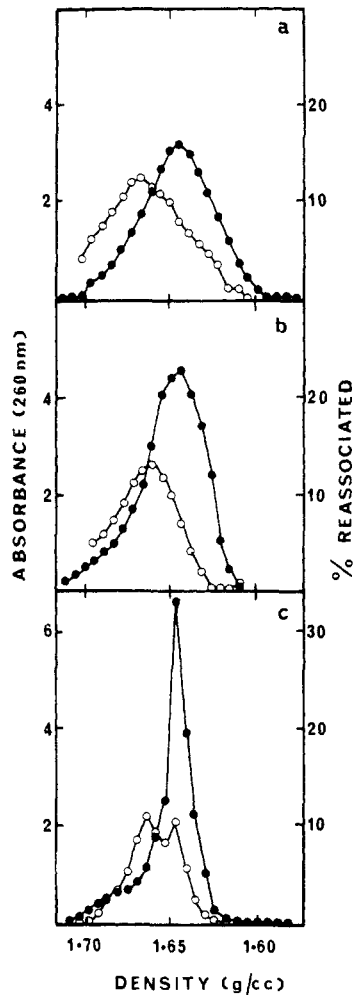

FIGURE 10: Netropsin–CsCl gradients showing the change in density of DNA-carrying keratin genes with molecular weight of the DNA. 1 mg of chick DNA of length 0.5 (a), 11 (b), and 140 kb (c) was centrifuged to equilibrium in CsCl in the presence of netropsin sulfate (DNA/netropsin sulfate = 1:1.6 w/w). The position at which the keratin genes band was determined by reassociation studies using keratin AMV cDNA and S1 nuclease. (●) A_{260}; (○) percent of counts reassociated.

Chick DNA fragments of much greater length (11 and 140 kb) which could reanneal with keratin cDNA also banded with distinctively lower buoyant densities in actinomycin D–CsCl gradients than the main band of chick DNA (parts b and c of Figure 9). Since high concentrations of the large DNA fragments on the gradients were necessary, thereby inviting artefacts in observed densities (Brown & Weber, 1968), the results were confirmed by testing three predictions. Firstly, if long DNA molecules containing keratin genes are richer in G+C content than the bulk of chick DNA, then these molecules should exhibit a higher buoyant density than the bulk of chick DNA when centrifuged to equilibrium on CsCl gradients in the presence of the antibiotic netropsin sulfate (Zimmer et al., 1971). The test of this prediction for each of the three sizes of DNA fragments is shown in Figure 10. Clearly, the fragments which could anneal with keratin AMV cDNA banded at higher densities than the bulk of chick DNA in each case.

Secondly, CsCl centrifugation in the presence of actinomycin D should lead to an enrichment of keratin genes in the DNA fraction of higher G+C content. For a test of this, DNA fractions from 11-kb chick DNA were pooled from CsCl–actinomycin D gradients, as indicated by the bar in Figure 9. The actinomycin D was removed, the DNA was recentrifuged in netropsin sulfate–CsCl gradients, and keratin sequences were reselected as in Figure 10b. By this procedure, a ninefold enrichment of keratin sequences was obtained, as determined by a consideration of the amount of DNA in the final fraction and the proportion of the total starting amount of keratin sequences still present in that fraction (data not shown).

Thirdly, since actinomycin D and netropsin sulfate amplify the fractionation of DNA in CsCl on the basis of G+C content, the resolution between main-band and keratin sequences would be expected to be significantly reduced by CsCl gradient centrifugation in the absence of either antibiotic. When this was carried out, keratin sequences banded to the heavy side of the bulk of chick DNA (a shift of 0.012 g/cm^3) for DNA of length 0.5 kb while for both 11- and 140-kb size classes the peak of keratin AMV cDNA hybridization was coincident with the A_{260} peak (data not shown).

Discussion

Arrangement of the Unique and Repetitive Sequences in Keratin mRNA. The data described demonstrate that the hybridization behavior of keratin AMV cDNA averaging 0.55 kb in length is quite different from that of keratin Pol I cDNA averaging 0.15 kb in length by all of the criteria examined. The observation that the apparent rate of hybridization of

keratin AMV cDNA to keratin mRNA was ~10 times faster when measured on HAP than when measured by nuclease S1 resistance was previously explained (Kemp, 1975) by proposing that each keratin AMV cDNA molecule contained a region of very similar sequence which could form poorly matched hybrids with the corresponding regions of nonidentical keratin mRNA molecules. An alternative interpretation would propose that each AMV cDNA molecule contained a short faithfully conserved region which could form short perfectly matched hybrids with the corresponding regions of nonidentical keratin mRNA molecules. We will denote this sequence the repetitive sequence in the sense that it must occur many times in the chick genome, while avoiding detailed description of the nature of this repetitive sequence (i.e., whether it is a short common repeat or a longer mismatching repeat). The observation that the same cDNA molecules present in these low melting point hybrids early in the reactions were later transferred into hybrids of much greater thermal stability can be explained by proposing that each keratin AMV cDNA molecule also contains a sequence which has diverged to such an extent that it cannot cross-hybridize with nonidentical keratin mRNA molecules and therefore behaves as if it is unique in the genome. This model therefore assumes that the degree of sequence divergence between different keratin mRNA molecules is nonrandom and varies greatly in different regions of the keratin mRNA.

An alternative model may also be considered in relation to the data. In this model it is assumed that the degree of sequence divergence between different keratin mRNA species is distributed randomly along the molecules. This "random divergence" model can account for the observed behavior of keratin AMV cDNA since RNA sequences which are annealed to DNA can be displaced from their hybrids by other molecules, provided that more stable complexes can be formed (Beckmann & Daniel, 1974). In fact, the latter process can be invoked to explain the high thermal stability of the hybrids formed late in the reaction by using either of the two models.

The two models lead to different predictions about the behavior of Pol I cDNA, however. Whereas the unique and repetitive sequence model predicts all the observed properties of the Pol I cDNA molecules, the random divergence model predicts, in principle, that the Pol I cDNA should be equally as repetitive in nature as the AMV cDNA. While the latter was clearly not the case, the interpretation is complicated by the fact that short duplexes are intrinsically less stable than longer duplexes. It is therefore possible that mismatched duplexes of Pol I cDNA did not form because their thermal stability is too low relative to the incubation temperature. A quantitative evaluation of this possibility is necessary in order to distinguish between the two models.

Southern (1975) demonstrated that restriction endonuclease fragments of mouse satellite DNA ranging in length from 120 to 600 bases could all reassociate to form mismatched duplexes. The depression of the T_m values of these duplexes from those of the native molecules due to mismatching was constant, irrespective of the length of the fragments. The depression of T_m values of the shortest reassociated duplexes relative to long native duplexes could be accounted for as the sum of the depression due to chain length and the constant mismatching factor of 6 °C. These data can be used to predict the thermal stability of the keratin Pol I cDNA duplexes expected from the random divergence model. The maximum T_m of keratin Pol I cDNA duplexes was lower by 7 °C than that of the corresponding AMV duplexes (Figures 4 and 6). The difference can be partially accounted for by a factor of 3–4 °C due to reduced length. The remaining difference of 3–4 °C can presumably be attributed to the difference in G+C content of the AMV and Pol I cDNA duplexes (Figure 8). The T_m values of mismatched Pol I cDNA duplexes should therefore be 7 °C below those of mismatched AMV cDNA duplexes (72–73 °C; Figures 4 and 6) or ~65–66 °C. Clearly, the rate of formation of many of these duplexes would be negligible (Bonner et al., 1973). However, because of the broad range of T_m values exhibited by mismatched AMV cDNA duplexes (Figures 4 and 6), about half of the mismatched Pol I cDNA duplexes would be expected to have T_m values distributed between 65 and 75 °C. Such duplexes should form at a rate comparable to those of the mismatched AMV cDNA duplexes (Bonner et al., 1973) after correcting for the effect of length on rate.

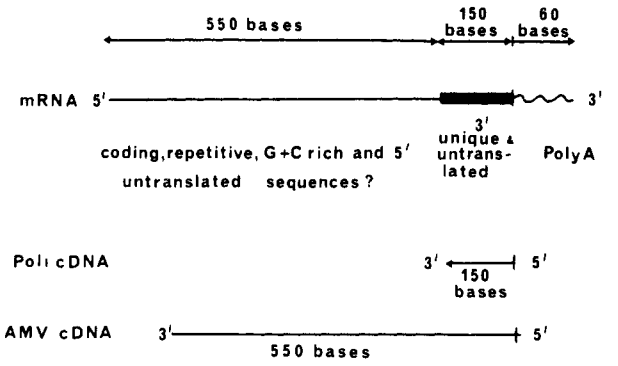

FIGURE 11: Model for the arrangement of sequences in keratin mRNA. The 150 nucleotides adjacent to the 3′ poly(A) tract constitute a unique untranslated sequence. Further to the 5′ end of the mRNA and covalently linked to this 3′ unique sequence is a repetitive sequence (short and faithfully conserved or longer and mismatching), a G+C-rich sequence, the coding sequence, and presumably a 5′ untranslated sequence. The currently available data permit no further conclusions about the sequence arrangement within this 5′ region of 550 bases. The orientations and approximate sizes of AMV and Pol I cDNAs are also shown.

The random divergence model therefore predicts that assay of keratin Pol I cDNA–mRNA hybridization on HAP, but not with nuclease S1, should result in a biphasic curve with the first transition (~35% of the total cDNA) at a $R_0t_{1/2}$ of $\sim 3 \times 10^{-3}$ mol s L^{-1}. No such transition was evident (Figure 3b). Since the percent hybridization observed using HAP was identical with that observed using nuclease S1 within an accuracy of ±2.5% at any point, an early transition much smaller than that predicted should have been detectable. We conclude that the random divergence model is incompatible with the data and hence that it is most probable that keratin mRNA contains distinct regions of unique and repetitive sequence.

Model for the Sequence Organization of Keratin mRNA. The data available suggest a model for the sequence organization in keratin mRNA (Figure 11). The experimental data, assumptions, and deductions upon which this model is based are summarized below.

(a) Keratin mRNA is ~0.8 kb long (Kemp et al., 1974a), contains a 3′ poly(A) tract of average length 60 nucleotides, and a 5′ m^7G cap structure (Morris & Rogers, 1979).

(b) At least 150 bases adjacent to the poly(A) tract are unique sequences (present results). We assume that the actual length is 150 bases and designate this region as 3′ unique sequences.

(c) About 30–50% of the length of each keratin AMV cDNA molecule 550 bases long is composed of repetitive sequences, since in separate experiments 30–50% of the radioactivity in these molecules appears as a fast transition resistant to nuclease S1 at high-salt concentration when they are

reannealed with chick erythrocyte DNA (Figure 5a).

(d) The reassociation kinetic analysis of elongated cDNAs of intermediate length (Figure 7b) also suggests that the repetitive sequence is fairly long and extends to within close proximity (100 nucleotides) of the 3′ unique region of the mRNA. The presently available data do not allow any further conclusions to be made about the nature, size, or position of this repetitive sequence. Most or all of the repetitive sequences are covalently attached to unique sequences since elongated Pol I cDNA shows reiterated reassociation kinetics when reannealed with chick erythrocyte DNA (Figure 7).

(e) Low melting point duplexes of the repetitive region were depressed in T_m by 18–20 °C from well matched duplexes (Figures 4 and 6). Therefore, the sequences are mismatched at ~20% of the bases (Bonner et al., 1973) and hence have diverged by ~10% in sequence (Southern, 1971). Alternatively, the repetitive region is very short and highly conserved.

(f) Keratin chains are ~100 amino acids long (O'Donnell, 1973; Walker & Rogers, 1976). They do not appear to be synthesized as larger polypeptide precursors, since the molecular weight of keratin translated from keratin mRNA in both rabbit reticulocyte and wheat embryo cell-free systems was identical with that of native keratin and the N-terminal sequence was identical (Partington et al., 1973; Kemp et al., 1974b). Therefore, the keratin coding sequence is ~300 bases long.

(g) Complete amino acid sequences have been determined for feather keratin chains from other species (O'Donnell, 1973; O'Donnell & Inglis, 1974), and partial sequences of tryptic peptides from embryonic chick feather keratins have also been determined (Walker & Rogers, 1976; Kemp et al., 1975). While regions of ~10 amino acids at the N and C termini vary considerably in sequence, the internal section of ~80 amino acids is highly conserved in those keratin chains which have been sequenced. The limited sequence studies on embryonic chick feather keratins in this internal region demonstrate that the sequences of four different N-terminal tryptic peptides were identical in the region from amino acids 12–28 except for a single proline → serine substitution in one of the four sequences. In the absence of further sequence data, we conclude that the amino acid sequences of the internal region of all embryonic chick feather keratins are very similar. If these regions were identical in amino acid sequence, then they could differ by a maximum of ~17% in their mRNA sequence (Kohne, 1970). From (e), if the low melting point of early duplexes was due to mismatching, then the repetitive sequences in the keratin mRNA would have diverged by ~10%, a value compatible with the suggestion that the repetitive sequence in keratin mRNA is in the coding sequence.

(h) From the amino acid composition of feather keratin (Walker & Rogers, 1976), it can be predicted that the DNA coding for keratins should have a G+C content of ~54.7%, compared with 41% for total chick DNA. Such a high G+C region is present further than 150 bases from the poly(A) region, again suggesting that the coding sequence is in this region.

The model for keratin mRNA sequence organization is similar to those determined for globin, immunoglobulin, and ovalbumin mRNAs by direct sequencing methods (Cheng et al., 1976; Proudfoot & Brownlee, 1976; Proudfoot et al., 1977), in that all four mRNAs appear to have a long untranslated sequence adjacent to the poly(A) tract. While the function of untranslated sequences is obscure, the present results suggest that a considerable degree of sequence divergence can be tolerated in this region, even though some elements of sequence and structure may be conserved as suggested by Proudfoot & Brownlee (1976). In addition, the total length of each of the different keratin mRNA species has been highly conserved during their evolution, suggesting that the length of the untranslated sequences has been conserved. These apparently conflicting data can be reconciled by speculating that within highly variable 3′ untranslated sequences of mRNAs, short regions of conserved sequence exist. The latter may result in a common general secondary and/or tertiary structure which is necessary for function, analogous to the situation with different tRNA species.

Keratin mRNA constitutes an example of a mRNA containing covalently linked unique and reiterated sequences, as does κ-chain immunoglobulin mRNA (Seidman et al., 1978). In the case of κ-chain mRNA, many different V-region sequences, some of which cross-hybridize, can be translocated to one or a few C-region (i.e., unique) sequences. In contrast, there are many different keratin genes, the unique region of each being so different that it cannot cross-hybridize to the unique region of any other keratin gene. The unique and repetitive elements of these two multigene families therefore reflect a very different underlying genetic organization. Keratin genes are closely analogous to the genes for chorion proteins (Regier et al., 1978) since both systems are multigene families, the members of which have diverged significantly in sequence. If there are many such multigene families in eucaryotes, then they may account for a significant proportion of the moderately repetitive nonidentical sequence fraction of the eucaryote genome.

Arrangement of Keratin Genes in the Chick Genome. As expected from the observed high G+C content of the sequence within keratin AMV cDNA, short duplex fragments of the chick genome which were complementary to keratin AMV cDNA were rich in G+C. In addition, we found that those genomic fragments averaging 150× the length of keratin mRNA and complementary to keratin AMV cDNA also had a higher G+C content than the bulk of chick DNA. We have no direct evidence that the G+C-rich sequence within keratin AMV cDNA confers the high G+C content on these long segments of the genome. Nevertheless, the data place restrictions upon possible models for the organization of keratin genes. If keratin genes are clustered in the genome such that many of them are present on the 140-kb fragments, then the high G+C content of such fragments might derive from that G+C-rich sequence which we have identified within keratin mRNA. Alternatively, if keratin genes are so distant that there is no more than one per fragment, it is necessary to postulate that in most of these dispersed locations keratin genes are surrounded by long stretches of DNA of high G+C content. If so, such regions would account for ~0.5–2.5% of the total chick genome. In the absence of further data, it therefore seems most reasonable to propose that keratin genes are clustered and closely linked in the chicken genome. This proposal should be testable by using the current recombinant DNA technology.

Acknowledgments

We thank J. R. E. Wells for the generous gift of purified AMV polymerase. AMV was donated to J. R. E. Wells by the National Institutes of Health Cancer Program. We thank W. J. Peacock of the C.S.I.R.O. Division of Plant Industry, Canberra, for donating purified *E. coli* DNA polymerase I and L. Crocker for technical assistance.

References

Baralle, F. E. (1977a) *Cell 10*, 549–558.

Baralle, F. E. (1977b) *Cell 12*, 1085–1095.
Beckmann, J. S., & Daniel, V. (1974) *J. Mol. Biol. 89*, 355–362.
Berns, A. J. M., Strous, G. J. A. M., & Bloemendal, H. (1972) *Nature (London), New Biol. 236*, 7–9.
Bonner, T. I., Brenner, D. J., Neufield, B. R., & Britten, R. J. (1973) *J. Mol. Biol. 81*, 123–135.
Brown, D. D., & Weber, C. S. (1968) *J. Mol. Biol. 34*, 661–680.
Brownlee, G. G., Cartwright, E. M., Cowan, N. J., Jarvis, J. M., & Milstein, C. (1973) *Nature (London), New Biol. 244*, 236–240.
Cheng, C. C., Brownlee, G. G., Corey, N. H., Doel, M. T., Gillam, S., & Smith, M. (1976) *J. Mol. Biol. 107*, 525–547.
Crawford, R. J., & Wells, J. R. E. (1976) *Mol. Biol. Rep. 3*, 167–173.
Crawford, R. J., Scott, A. C., & Wells, J. R. E. (1977) *Eur. J. Biochem. 72*, 291–299.
Gambino, R., Kacian, D., O'Donnell, J., Ramirez, F., Marks, P. A., & Bank, A. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 3966–3970.
Gould, M. J., & Hamlyn, P. H. (1973) *FEBS Lett. 30*, 301–304.
Gulati, S. C., Kacian, D. L., & Spiegelman, S. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1035–1039.
Housman, D., Forget, B. G., Skoultchi, A., & Benz, E., Jr. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 1809–1813.
Kacian, D. L., Gambino, R., Dow, L. W., Grossbard, E., Natta, G., Ramirez, F., Spiegelman, S., Marks, P. A., & Bank, A. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 1809–1813.
Kemp, D. J. (1975) *Nature (London) 254*, 573–577.
Kemp, D. J., Partington, G. A., & Rogers, G. E. (1974a) *Biochem. Biophys. Res. Commun. 60*, 1006–1014.
Kemp, D. J., Schwinghamer, M. W., & Rogers, G. E. (1974b) *Mol. Biol. Rep. 1*, 441–446.
Kemp, D. J., Walker, I. D., Partington, G. A., & Rogers, G. E. (1975) in *The Eukaryote Chromosome* (Peacock, W. J., & Brock, R. D., Eds.) pp 205–216, ANU Press, Canberra, Australia.
Kersten, W., Kersten, H., & Szybalski, W. (1966) *Biochemistry 5*, 236–244.
Kohne, D. E. (1970) *Q. Rev. Biophys. 3*, 327–375.
Marushige, K., Brutlag, D., & Bonner, J. (1968) *Biochemistry 7*, 3149–3155.
McReynolds, L., O'Malley, B. W., Nisbet, A. D., Forthergill, J. E., Givol, D., Fields, S., Robertson, M., & Brownlee, G. G. (1978) *Nature (London) 273*, 723–728.
Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M., & Proudfoot, N. J. (1974) *Nature (London) 252*, 354–359.
Modak, M. J., Marcus, S. L., & Cavalieri, L. F. (1973) *Biochem. Biophys. Res. Commun. 55*, 1–7.
Morris, C. P., & Rogers, G. E. (1979) *Mol. Biol. Rep. 5*, 145–149.
O'Donnell, I. J. (1973) *Aust. J. Biol. Sci. 26*, 415–437.
O'Donnell, I. J., & Inglis, A. S. (1974) *Aust. J. Biol. Sci. 27*, 369–382.
Partington, G. A., Kemp, D. J., & Rogers, G. E. (1973) *Nature (London), New Biol. 246*, 33–36.
Pinder, J. C., Staynov, D. J., & Gratzer, W. B. (1974) *Biochemistry 13*, 5373–5378.
Proudfoot, N. J. (1977) *Cell 10*, 559–570.
Proudfoot, N. J., & Brownlee, G. G. (1974) *Nature (London) 252*, 359–362.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Proudfoot, N. J., & Longley, J. I. (1976) *Cell 9*, 733–746.
Proudfoot, N. J., Gillam, S., Smith, M., & Longley, J. I. (1977) *Cell 11*, 807–818.
Rabbitts, T. H. (1974) *FEBS Lett. 42*, 323–326.
Regier, J. C., Kafatos, F. C., Goodfliesh, R., & Hood, L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 390–394.
Schachat, F. H., & Hogness, D. S. (1973) *Cold Spring Harbor Symp. Quant. Biol. 38*, 371–381.
Seidman, J. G., Leder, A., Nau, M., Norman, B., & Leder, P. (1978) *Science 202*, 11–17.
Southern, E. M. (1971) *Nature (London), New Biol. 232*, 82–83.
Southern, E. M. (1975) *J. Mol. Biol. 94*, 51–69.
Walker, I. D., & Rogers, G. E. (1976) *Eur. J. Biochem. 69*, 329–339.
Zimmer, C. H., Reinert, K. E., Luck, G., Wahnert, U., Laber, G., & Thrum, H. (1971) *J. Mol. Biol. 58*, 329–348.