ABSTRACT Accessible sites in the 5' noncoding region of the rabbit α- and β-globin mRNAs were identified and compared in deproteinized RNA and in the mRNAs engaged in translation in the reticulocyte lysate. Preparations of RNA and lysate were subjected to limited nuclease digestion by RNase T1 and Neurospora endonuclease, and the cleavage sites were analyzed by a nuclease S1 mapping procedure. The free α-globin mRNA contained few nuclease-sensitive sites and its initiation codon AUG was masked. The free β-globin mRNA contained a larger number of accessible sites and its AUG was highly exposed. The distribution of sensitive sites differed considerably in the lysate. In both mRNA species, a site near the 5' terminus became the most accessible to Neurospora endonuclease. Also the accessibility of the AUG in β-globin mRNA decreased considerably. The distribution of accessible sites on the lysate was the same when the mRNAs were undergoing rapid initiation and when initiation became limited after prolonged incubation. Inhibition of Initiation by the cap analogue 7-methylguanosine 5'-triphosphate was accompanied by increased sensitivity of some of the sites in both mRNA species. One of the accessible sites in each mRNA species had a sequence complementary to the 3'-terminal portion of the 18S ribosomal RNA.

The nature of the interactions between mRNA and ribosomal particles during polypeptide-chain initiation in eukaryotic cells is poorly understood. The cap structure at the 5' end of the mRNA chain is required for this process (1, 2). It is thought that proteins essential for the initiation process bind to this structure (3). It is also believed that the 40S ribosomal particle binds at or near the 5' terminus and then slides along the 5' noncoding region until the initiation codon AUG is reached (4). This postulated mechanism differs markedly from the one known to occur in prokaryotes, in which the 30S ribosomal subunit is positioned next to the initiation AUG by base pairing between complementary sequences in the mRNA and the 3' end of the 16S ribosomal RNA (5, 6). Sequences capable of base pairing with the 3' end of 18S rRNA also occur in the 5' noncoding region of many eukaryotic mRNAs (7), but no evidence for such mRNA–rRNA interaction in eukaryotes has been obtained so far. Binding studies with mRNA fragments have suggested that direct interaction between the ribosomal subunit and the AUG triplet may be important (8).

The secondary structure in and around the 5' noncoding region appears to play an important role in the determination of initiation efficiency. This is well-documented in the case of bacteriophage RNA translation (9, 10). In eukaryotic systems, the translation of mRNAs with a relaxed secondary structure in the 5' noncoding region is less dependent on cap-binding proteins, which are believed to promote the unwinding of the ribosomal binding site (11, 12). The two globin mRNAs have served as a useful model for the study of structural features that may affect initiation efficiency. The β-globin species, which is the more effective of the two with respect to initiation (13), has been shown to have a highly accessible initiation AUG (14, 15). This codon is not exposed in α-globin mRNA (14). This feature tends to suggest that direct interaction between the 40S ribosomal subunit and the initial AUG is an important aspect of the initiation process.

In searching for structural features of mRNA molecules that may be relevant to their function, we have taken the view that it is important to deal with the mRNA in its native configuration, preferably in a defined functional state. This has not been possible until recently, because of the need to work with pure species of deproteinized mRNA. We have used a recently developed nuclease S1 mapping technique that permits the identification of cleavage sites in mRNA molecules treated with nucleases in crude cell extracts (15). In the present study, we have mapped sites readily accessible to nucleases in the 5' noncoding region of the rabbit globin mRNAs. We have compared this aspect of the configuration in reticulocyte extracts highly active in initiation and in extracts no longer capable of initiation. The effect of the cap analogue 7-methylguanosine 5'-triphosphate (7MeGTP) was also examined to determine whether cap-binding proteins may affect the configuration. Our findings indicate that components associated with the mRNAs in the cytoplasm alter the configuration of the 5' noncoding region, that the cap analogue causes extensive unmasking in this region, and that a major accessible site has a sequence complementary to a portion of the 3' terminus of the 18S ribosomal RNA.

MATERIALS AND METHODS

Probes and Enzymes. Plasmids containing full-length α- and β-globin cDNA inserts (16), kindly provided by Paul Berg, were used for this study. The probe for α-globin mRNA was prepared by cleaving the plasmid with Taq I and isolating a 390-base-pair (bp) fragment bearing the 5'-terminal portion of the mRNA sequence (Fig. 1). The β-globin probe was prepared by first cleaving the appropriate plasmid with EcoRI and isolating a 700-bp fragment containing a portion of the β-globin mRNA sequence. The latter was cut with Hinf I to yield a 370-bp fragment that served as the probe. The fragments were end-labeled with 32P by the polynucleotide kinase reaction, as described (15). The purified probes obtained in this fashion yielded stronger signals than those obtained with the mixture of labeled restriction fragments used in the previous study.

The enzymes used for mRNA fragmentation were RNase T1 (Calbiochem–Behring) and a Neurospora endonuclease (Sigma). Nuclease S1 (Bethesda Research Laboratories) was

Abbreviations: 7MeGTP, 7-methylguanosine 5'-triphosphate; bp, base pair(s).
used for the mapping procedure. The cap analogue 7MeGTP (P-L Biochemicals) was a generous gift from David Stollar. 

**Globin mRNA Fragmentation and Nuclease S1 Mapping.** Reticulocytes were prepared from anemic rabbits and lysed as described (15). Reticulocyte RNA was isolated by subjecting a sample of the lysate to the alkaline phenol deproteinization procedure (17). The fragmentation procedure involved incubating samples of the lysate at 37°C under conditions optimal for translation (18), adding a nuclease after the appropriate time interval, and continuing the incubation for an additional 4 min. For the fragmentation of deproteinized RNA, samples were incubated in the presence of the enzyme at 37°C for 10 min, using ionic conditions similar to those used for the reticulocyte lysate: 80 mM K acetate/20 mM Hepes/0.5 mM MgCl₂. The nuclease-treated samples were subjected to deproteinization. Samples of the fragmented RNA preparations containing 1 μg of RNA were hybridized with the labeled probes as described (15). The hybrids were cleaved with 3–10 units of nuclease S1 to destroy the portions of DNA probe not covered with mRNA, and the remaining labeled probe fragments were separated by electrophoresis in a denaturing polyacrylamide gel and visualized by autoradiography (15).

**RESULTS**

**Functional States of the mRNA in Lysates Incubated Under Various Conditions.** The analysis of sites accessible to nucleases was carried out on mRNA in various functional states. Samples of reticulocyte lysate incubated under conditions standard for exogenous mRNA translation (18) maintained a high rate of initiation for at least 10 min, as indicated by the occurrence of relatively large polyribosomes after the incubation period. Incubation of the lysate in the presence of 0.5 mM 7MeGTP led to extensive loss of polyribosomes. The cap analogue appeared to cause inhibition of polypeptide-chain initiation, as indicated by the fact that most of the mRNA was then associated with smaller polyribosomes and with subribosomal particles (Fig. 2b). The use of an equivalent amount of ATP instead of the 7MeGTP had no obvious effect on the translation characteristics of the lysate. The polyribosome profile shown in Fig. 2a for the lysate incubated with ATP was very similar to the profile obtained without added ATP (not shown). The mRNA was associated with relatively large polyribosomes under these conditions, as shown by the sedimentation profile of poly(A)-containing material. The ATP control was important because excessive amounts of this compound (>1 mM) caused inhibition of translation. This may have been due to complex formation with the limited amounts of Mg ions in the incubation mixture. This control shows that the 7MeGTP effect was specific, and not due simply to excessive nucleoside triphosphate.

Inhibition of polypeptide-chain initiation was also achieved by incubating the lysate for at least 20 min, either in the presence or absence of added ATP. The extent of polyribosome loss under these conditions, however, was highly variable. A more consistent inhibitory effect was obtained when the lysate supplemented with ATP was preincubated for 20 min in ice before the 20-min incubation at 37°C. Fig. 2c shows the effect of such a treatment on the sedimentation profiles of polyribosomes and poly(A)-containing RNA. In some experiments, nearly complete conversion of polyribosomes to monomers was observed. For each nuclease S1 mapping experiment, samples of the incubation mixtures were used for sedimentation profile analysis to verify the functional state of the mRNA during the enzyme treatments. **Mapping Procedure.** The procedure involves brief exposure of the reticulocyte extracts or of free RNA to a nuclease, followed by deproteinization, hybridization of the fragmented preparations to an end-labeled probe (see Fig. 1), digestion of unhybridized segments with nuclease S1, and separation of the remaining probe fragments by gel electrophoresis (15). The sizes of these fragments correspond to the distances between the restriction site used to generate the labeled end and the sites of cleavage on the mRNA. The relationship between probe fragment size and position along the mRNA chain is shown in Fig. 5.

The mapping of RNase T1 cleavage sites yielded clusters of bands, even when the site contained a single G residue. This is due to the inability of nuclease S1 to recognize precisely the boundary between single- and double-stranded regions, as discussed previously (15). Most of the experimentally determined RNase T1 sites were within one or two nucleotides of known G positions. One such site, however (site D in Fig. 4), was mapped as being 4–5 nucleotides downstream from the likely site of cleavage. The Neurospora endonuclease cleaves preferentially at G and A residues (20), and the identification of its cleavage sites at the nucleotide level is less certain.
The mapping of intact mRNA preparations yielded two major bands (Figs. 3 and 4). In each case, the lower one (A) corresponds to the full-length mRNA. The upper one is ~20 nucleotides longer. The extra length may correspond to the poly(dC) tail in the probe. The proportions of the two bands varied in different experiments. In the case of the β-globin mRNA, preparations not exposed to nucleases also yielded minor bands (Fig. 4) that correspond to truncated molecules apparently generated by cleavages inside the cells (15). Incubation of the reticulocyte extracts for up to 20 min at 37°C did not cause any additional cleavages in the 5′ noncoding region (Figs. 3 and 4, lanes j). The truncated molecules tended to be less prominent in the incubated extracts (see lanes d, g, j, and m in Fig. 4; see also ref. 15).

**Sites Accessible to Nucleases in the Globin mRNAs.** The 5′ noncoding region of the α-globin mRNA contained few sites sensitive to the nucleases. The region around the initiation AUG codon and the beginning of the coding region were not subject to cleavage. The *Neurospora* endonuclease (nuclease N) cleaved at two sites (labeled B and D in Fig. 3). The approximate locations of these sites are shown in Fig. 5. RNase T1 caused extensive cleavage at the second G residue in the sequence (site C) and had little or no effect on the other G-containing sites. The faint band just above B, which corresponds to the first G residue in the sequence, was observed consistently in preparations from nuclease-treated lysate but not when free RNA was subjected to cleavage. Rough estimates of the extent of cleavage at each major site were obtained by subjecting autoradiograms such as the ones shown in Figs. 3 and 4 to densitometry scanning. The results show that the distribution of nuclease-sensitive sites was markedly different in the crude lysate and in deproteinized RNA (Fig. 5). The two nuclease N sites (B and D) were about equally sensitive in free mRNA, but site B was considerably more exposed in the lysate (Figs. 3 and 5). The site cleaved by RNase T1 appeared to be less exposed in the lysate. The patterns of cleavages were similar in lysate samples incubated in the presence or absence of added ATP. They were the same in lysates active in initiation and when this process was no longer effective in the lysate incubated for 20 min. The additional band in lane i was not seen in most experiments.

In the β-globin mRNA, nuclease N caused cleavages at sites scattered throughout the 5′ noncoding region (Fig. 4). The major cleavage sites are labeled C, D–F, G, and H, and their approximate locations are also shown in Fig. 5. RNase T1 caused cleavages primarily at the initiation AUG site (site I) and at the second G residue in the mRNA sequence (site B). Some cleavage also occurred at G residues in sites D and E. As in the case of the α-globin species, there were marked differences in the cleavage patterns in free RNA and in the lysate. The AUG region (sites H and I), which was the one most sensitive to the two nucleases in deproteinized RNA, was far less exposed in the lysate samples. The nuclease N site close to the 5′ terminus (site J) was also strongly affected. This site was among the ones least sensitive to this enzyme in deproteinized RNA, but it became the most exposed one in the lysate. It is unlikely that the apparent masking of the AUG region in the lysate was due to the presence of ribosomes in this region, since the same extent of masking occurred when the mRNA was being depleted of ribosomes because of limiting initiation (Fig. 4, lane k).

**Accessible Sites Complementary to the 3′ Terminal of 18S Ribosomal RNA.** Site B in α-globin mRNA, which is highly accessible to nuclease N in the lysate, appears to be located at the sequence GUCC (Fig. 5). The identification of this site seems quite reliable, since it is adjacent to the site containing G-15, which is highly accessible to RNase T1. This sequence is complementary to a portion of the 3′ terminus of the 18S ribosomal RNA (Fig. 6). The β-globin mRNA also has such a complementary sequence in an accessible state. The latter is located in sites D–F (see Figs. 5 and 6). According to the size of band cluster D in Fig. 4, the site should begin at U-26 (Fig. 5). However, cleavage by RNase T1 at the nearest G residue (G-24) leads to the same-sized band cluster (Fig. 4). It is possible, therefore, that the region accessible to nuclease N begins at the latter position, as indicated in Fig. 5 by the dotted line.

**Effect of 7MeGTP on mRNA Configuration.** Incubation in the presence of 7MeGTP caused large increases in the accessibility of many of the nuclease-sensitive sites. In the α-globin mRNA, the degree of exposure of the two sites sensitive to nuclease N was increased to the same extent. The sensitivity of site C to RNase T1 was also increased (Figs. 3 and 4).
and 5). In the β-globin species, the incubation in the presence of the cap analogue also led to very considerable increases in the accessibility of sites C and D–F to nuclease N. The sites closer to the initiation AUG were enhanced to a lesser extent (Figs. 4 and 5). The AUG region became also more sensitive to RNase T1. The other sites sensitive to the latter enzyme, however, were not enhanced. The discrepancy in the response to the two enzymes was particularly obvious in the cases of sites D and E, which became far more sensitive to nuclease N but not to RNase T1. The extensive cleavages in the lysate incubated with 7MeGTP were not due to increased activity of the nucleases in the presence of this compound. This is indicated by the fact that other sites, in the coding region and in the 3′ noncoding region, were cleaved to the same extent in the presence or absence of the inhibitor (data not shown).

**DISCUSSION**

The purpose of this study was to identify sites in the rabbit globin mRNAs that are likely to participate in the process of polypeptide-chain initiation. The approach has been to determine which portions of the 5′ noncoding regions are readily accessible to macromolecules, using sensitivity to

<table>
<thead>
<tr>
<th>RNA</th>
<th>6 min</th>
<th>6 min</th>
<th>20 min</th>
<th>10 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>T1</td>
<td>N</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
</tr>
<tr>
<td>f</td>
<td>g</td>
<td>h</td>
<td>i</td>
<td>j</td>
</tr>
<tr>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
<td>o</td>
</tr>
</tbody>
</table>

**FIG. 4.** Distribution of cleavages in β-globin mRNA. See Fig. 3 for details and lane designations.

**FIG. 5.** Identification and quantitation of cleavage sites. Locations of RNase T1 cleavage sites are indicated by open arrowheads, and approximate locations of nuclease N sites are underlined. The dotted line that extends site D–F is discussed in text. Numbers above mRNA sequences represent position along sequence, and numbers in parentheses show sizes of probe fragments that would be derived from cleavages at these sites. Capital letters identifying the sites correspond to letters in Figs. 3 and 4. Numbers under sites represent intensity of each band, as determined by densitometry scanning of autoradiograms in Figs. 3 and 4. Underlined values are for RNase T1 sites. Values for RNA are not comparable to corresponding values for lysates, since enzyme activities were different. Values for lysates incubated under different conditions are directly comparable. Messenger RNA sequences are from refs. 21 and 22.

**FIG. 6.** Possible complementary interactions between accessible sites B and D–F and 3′-terminal portion of 18S ribosomal RNA. Sequence of rRNA terminus is from ref. 7.
two nucleases as the criterion of accessibility. The *Neurospora* endonuclease, like RNase T1, recognizes polynucleotides in single-stranded configuration (23). The two enzymes were used under identical conditions, but the resulting patterns of cleavages were quite different. The difference in the behavior of the two enzymes is also illustrated by the fact that two G-containing sites in ß-globin mRNA (sites D and E) became far more sensitive to nuclease N in the presence of 7MeGTP, but their sensitivity to RNase T1 was unchanged. This indicates that the two enzymes may be sensitive to different features of RNA structure. It is also possible that sites readily accessible to RNase T1 might not be reached easily by the bulkier *Neurospora* enzyme (23). The heterogeneity in the distribution of cleavage sites indicates that a complex configuration must exist in the 5′ noncoding region of the globin mRNAs. Additional studies, using other reagents such as small chemical probes, will be required to define this configuration. The comparison of sensitive sites in deproteinized mRNA and in the mRNA as it occurs in crude reticulocyte extracts indicates that components associated with the mRNAs in the cytoplasm must cause considerable changes in the configuration. In the ß-globin mRNA, a large increase in the accessibility of a site close to the 5′ terminus is observed. The changes in ß-globin mRNA also result in a 5′-proximal site becoming the one most exposed to nuclease N. The results also show that the region around the initiation AUG codon is highly accessible in deproteinized ß-globin mRNA, but it is far less exposed in the reticulocyte extracts.

The experiments with 7MeGTP provide information on the possible role of the cap structure and of the cap-binding proteins in the initiation process. It has been suggested that proteins bound to this structure serve to unravel the 5′ noncoding region, thus making it more accessible to the 40S ribosomal subunit (11, 12). Cap analogues such as 7MeGTP interfere with polypeptide-chain initiation, apparently by binding to the proteins that normally interact with the cap structure (24). The incubations with 7MeGTP caused inhibition of the initiation process, as indicated by the partial release of ribosomes from the mRNA. The analogue also caused large increases in accessibility in the 5′ noncoding region of both mRNA species. This unmasking was not due simply to loss of ribosomes from the mRNA, as indicated by the fact that partial ribosome runoff not caused by this inhibitor did not lead to such unmasking. The cap analogue presumably causes the removal of the cap-binding proteins from the 5′ end of the mRNA. Such a removal of proteins covering the 5′-terminal region could account for the increased accessibility of this region in the treated lysate. The proteins sensitive to 7MeGTP, although essential for initiation, do not appear to have any significant impact on the configuration in the 5′ noncoding region, as indicated by the fact that their interaction with the cap analogue left the distribution of exposed sites largely unchanged. Other components, not affected by 7MeGTP, appear to be the ones involved in the determination of the native configuration in this region.

Our results show that the most readily accessible sites in both mRNA species are located near the 5′ terminus. This feature of the configuration of the 5′ noncoding region is imposed on the RNAs by cytoplasmic factors. Moreover, these are the sites most affected when initiation is inhibited by the cap analogue 7MeGTP. It seems likely, therefore, that the region near the beginning of the mRNA sequence plays a major role in the initial interaction with the ribosomal particles. The results also suggest that direct interaction with the initiation codon is less important, since this codon is in the least accessible of the exposed sites in the ß-globin mRNA and is not accessible at all in the ß-globin species. The degree of exposure of the region next to the AUG may determine some of the characteristics of the initiation process. This is suggested by the fact that this region is more accessible in the ß-globin mRNA, which is the more efficient species with respect to polypeptide-chain initiation (13). Another possible interaction is suggested by the fact that one accessible site in each mRNA species has a sequence complementary to the 3′-terminal portion of the 18S ribosomal RNA. In the ß-globin mRNA, this sequence is in the site that is highly enhanced in the lysate (site B). It is relatively less exposed in the ß-globin species (sites D-F). Such sequences capable of interacting with the rRNA terminus of the 40S ribosomal subunit have been shown to occur in a variety of eukaryotic mRNA species (7). It remains to be determined whether the high degree of accessibility of this sequence in the globin mRNAs is fortuitous or whether it has functional significance.

We thank Dr. Paul Berg for the gift of the globin cDNA clones. This work was supported by research grants from the National Institutes of Health (GM 17973) and the National Science Foundation (PCM8020712).