Volume 14 Number 14 1986 Nucleic Acids Research

Secondary structure model for mouse β^{Maj} globin mRNA derived from enzymatic digestion data, comparative sequence and computer analysis

Raymond E.Lockard, Kathleen Currey[1], Michelle Browner[2], Charles Lawrence[2] and Jacob Maizel[3]

Department of Biochemistry, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, [1]Department of Neurology, University of Utah, Salt Lake City, UT 84132, [2]Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 and [3]Laboratory of Mathematical Biology, National Cancer Institute-FCRF, Frederick, MD 21701, USA

Received 27 February 1986; Revised and Accepted 21 May 1986

ABSTRACT

A model for the secondary structure of mouse β^{Maj} globin messenger RNA is presented based on enzymatic digestion data, comparative sequence and computer analysis. Using 5'-^{32}P-end-labeled β globin mRNA as a substrate, single-stranded regions were determined with S_1 and T_1 nucleases and double-stranded regions with V_1 ribonuclease from cobra venom. The structure data obtained for ca. 75% of the molecule was introduced into a computer algorithm which predicts secondary structures of minimum free energy consistent with the enzymatic data. Two prominent base paired regions independently derived by phylogenetic analysis were also present in the computer generated structure lending support for the model. An interesting feature of the model is the presence of long-range base pairing interactions which permit the β globin mRNA to fold back on itself, thereby bringing the 5'- and 3'-noncoding regions within close proximity. This feature is consistent with data from other laboratories suggesting an interaction of the 5'- and 3'-domains in the mammalian globin mRNAs.

INTRODUCTION

Our knowledge of eukaryotic messenger RNA higher order structure is scant considering its importance during gene expression (1-4). The identification of sites within specific mRNA molecules important during protein synthesis, as well as, those recognized by mRNA-bound proteins, translation factors, and ribosomes is essential in resolving the complexities of the eukaryotic translational process. We have analyzed the structure of mouse β^{Maj} globin mRNA using both single- and double-strand specific enzymatic structure probes. Cleavage profiles using 5'-^{32}P-end-labeled mRNA renatured under a variety of buffer and temperature conditions indicated one predominant, stable secondary structure. The structure data obtained was introduced into a computer algorithm which predicts secondary structures of minimum free energy consistent with the enzymatic data (5). A search for possible phylogenetically conserved structural features within mammalian β-like globin mRNA sequences was carried out. Two major base paired regions were derived which also were present in the

computer-generated model (6). A combination of these two approaches for deriving secondary structure models will likely be useful for other informational RNA molecules.

MATERIALS AND METHODS

MATERIALS

Tobacco acid pyrophosphatase (TAP) was purchased from Bethesda Research Labs. Both calf intestinal alkaline phosphatase (CIAP) and T_4 polynucleotide kinase were supplied by Boehringer-Mannheim and used without further purification. Neurospora crassa endonuclease was obtained from P.L. Biochemicals, Inc. S_1 nuclease was a kind gift from Dr. George Pavlakis, NIH. [γ-^{32}P]-ATP was prepared from carrier-free $^{32}P_i$ (ICN) by the procedure of Johnson and Walseth (7). V_1 nuclease was prepared from Naja naja oxiana cobra venom as previously described (8).

METHODS

Purification and 5'-^{32}P-End-Labeling of Mouse β Globin mRNA

Isolation of reticulocytes and preparation of reticulocyte polysomes have been previously described (9). Large scale purification of globin mRNA from pelleted polysomes was carried out as previously detailed (10). The $m^7G5'ppp5'$ cap structures of mouse globin mRNA were enzymatically removed using tobacco acid pyrophosphatase and calf alkaline phosphatase as previously described (11). 0.50 A_{260} units of decapped and dephosphorylated mouse globin mRNA were incubated with 8 units of T_4 polynucleotide kinase and 500 pmol of [γ-^{32}P]-ATP (7000 Ci/mmol) in a 15 µl reaction mixture containing 20 mM Tris-HCl (pH 8.5), 5 mM $MgCl_2$, 10 mM DTT, and 10% $(CH_3)_2SO$. The mixture was incubated for 50 minutes at 37°C. Reactions were terminated by addition of 40 µl of loading solution (99% formamide with 0.1% w/v xylene cyanol and bromophenol blue dyes and 25 mM EDTA) and loaded onto a preparative 3.5% polyacrylamide gel in 7 M urea (40 cm long x 20 cm wide x 0.15 cm thick) (3,10). After gel electrophoresis the 5'-^{32}P-end-labeled β globin mRNA molecules were localized by autoradiography, excised, recovered by electrophoretic elution, and ethanol precipitated as previously detailed (10,12).

Digestions for Secondary Structure Analysis

Purified 5'-^{32}P-labeled β globin mRNA was carefully renatured prior to nuclease digestion according to either of two protocols: 1) preincubation in corresponding digestion buffer for 10 minutes at 65°C followed by slow cooling to 25°C, or 2) preincubation in 20 mM Tris-HCl, pH 7.5 for 4 minutes at 85°C followed by addition of corresponding digestion buffer and slow

cooling to 25°C. Both conditions generated the one major mRNA conformer. Digestions with S_1 nuclease were performed at 37°C in 20 µl reactions containing 40 mM NaOAc, pH 4.5, 5 mM $MgCl_2$, 50 mM KCl, 1 mM $ZnSO_4$ at an enzyme-to-substrate ratio of 6 x 10^{-3} units/µg carrier tRNA. V_1 nuclease digestions were carried out at 37°C in either 15 or 20 µl reactions containing 20 mM Tris-HCl, pH 7.2, 10 mM $MgCl_2$, 200 mM NaCl with 1 x 10^{-6} units of enzyme/µg carrier tRNA. Structural digestions with T_1 RNase were performed at 37°C for 5 minutes in two separate 5 µl reaction mixtures containing 15 mM NaOAc, pH 5.0, 15 mM $MgCl_2$, 300 mM NaCl with 3 x 10^{-4} and 3 x 10^{-5} enzyme units/µg carrier tRNA, respectively. All digestion mixtures contained (5-10) x 10^4 cpm of 5'-^{32}P-end-labeled mRNA. Aliquots from reaction mixtures were terminated by addition of an equal volume of "stop solution" (9 M urea, 10 mM EDTA, 0.5 µg/µl tRNA carrier, 5 mM ATP, 0.05% xylene cyanol and 0.05% bromophenol blue). Enzymatic digests were electrophoretically fractionated in adjacent lanes of either 15% or 10% polyacrylamide sequencing gels run in 95% formamide (8,13), or 6.5% acrylamide gels run in 7 M urea with 100 mM Tris-Borate-4 mM EDTA, pH 7.2, buffer. Densitometric scans of structural digests from autoradiograms of polyacrylamide sequencing gels were performed using a Beckman DU-8B scanning spectrophotometer.

Computer Analysis of Structure Data

Structure data was analyzed on a VAX 11/750 computer using the algorithm RNA-FOLD developed and provided by Dr. Michael Zuker, National Research Council of Canada, Ottawa, Canada (5). Stacking energies were identical to those described by Salser (14), and loop destabilizing energies were the same as described by Cech et al. (15).

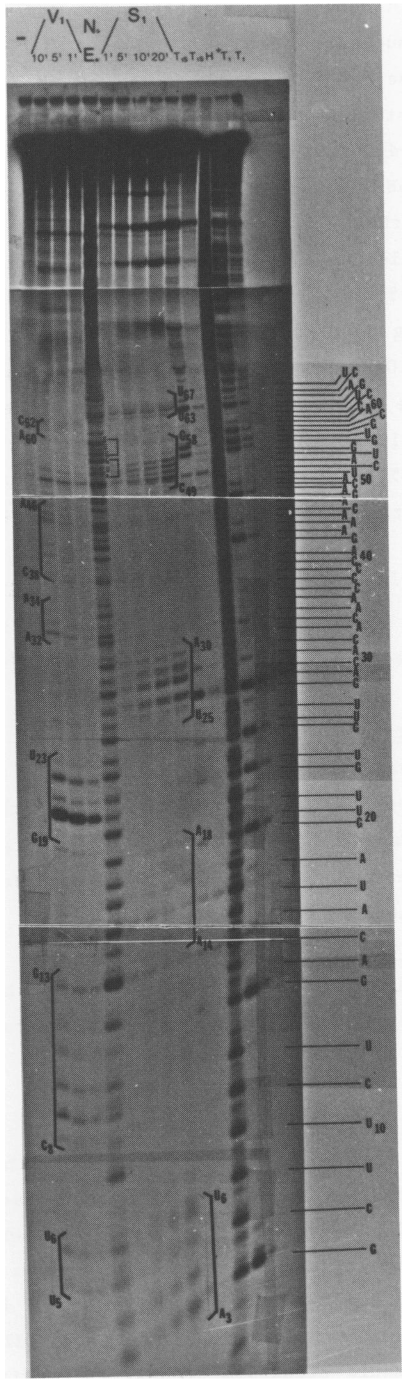
Detection of Phylogenetically Conserved Base Pairing

The computer analysis of conserved secondary structures within the β globin mRNA sequences was performed as previously detailed (6).

RESULTS

Structure Analysis of 5'-^{32}P-Labeled β Globin mRNA

Mouse globin mRNA was enzymatically decapped with tobacco acid pyrophosphatase, dephosphorylated with alkaline phosphatase, and 5'-terminally labeled with ^{32}P in 10% $(CH_3)_2SO$ using [γ-^{32}P]-ATP and T_4 polynucleotide kinase. The 5'-^{32}P-labeled β mRNA was resolved from α mRNA by electrophoresis on 40 cm long 3.5% polyacrylamide gels run in 7 M urea and the individual β band excised and extracted after autoradiography (3). Prior to digestion with enzymatic structure probes, the purified ^{32}P-end-labeled



β mRNA was carefully renatured in the corresponding digestion buffers (see MATERIALS AND METHODS). Preincubation of the RNA for ten minutes at 65°C in digestion buffer followed by slow cooling to 25°C, generated one major conformer.

Fig. 1 shows a structure analysis of the 5'-terminus for mouse β mRNA. The 5'-^{32}P-labeled β globin mRNA was digested with single-strand specific S_1 and T_1 nucleases and double-strand specific V_1 RNase. Using a fixed enzyme concentration for both S_1 and V_1 nucleases, aliquots were removed from the reaction mixture at five minute intervals so as to distinguish primary from secondary cutting. These partial digestion products were electrophoresed in parallel on polyacrylamide sequencing gels run in 95% formamide. Guanosine-specific T_1 RNase was used in the presence of a high salt buffer to complement S_1 nuclease which has a reduced specificity for G's (8,16). Since both S_1 and V_1 generate 3'-hydroxyls at their site of cleavage, a complete ladder of ^{32}P-end-labeled fragments with 3'-hydroxyls can easily be generated using Neurospora crassa endonuclease. When loaded onto a lane between both S_1 and V_1 digests, this ladder facilitates simultaneous interpretation of both single- and double-strand specific cleavages (8).

In Fig. 1 interpretation of which regions of the β globin mRNA are single-stranded and which are double-stranded is straightforward due to the alternating cutting pattern of both S_1 and V_1 nucleases. The V_1 nuclease fraction from Naja naja oxiana venom, unlike the previously characterized "cobra venom ribonuclease" fraction (17), appears to cleave more extensively within double-helical regions (8), thereby facilitating interpretation of the extent of base pairing. Primary cleavages with both S_1 and V_1 nucleases were ascertained when only 10-15% of the RNA was digested, corresponding to incubation times of 1-5 minutes. Generally, the most difficult region to interpret in a structural digest is that

Figure 1. Autoradiogram of partial digests on 5'-[^{32}P]mouse β globin mRNA electrophoresed on a 15% polyacrylamide slab gel run in 95% formamide. Gel dimensions were 0.40 mm thick x 33 cm wide x 100 cm long. Partial digests contained the following amounts of enzyme per μg RNA. From left to right: (-) minus enzyme; (V_1)V_1 nuclease 1 x 10^{-6} units; 10 min, 5 min, 1 min; (N.E.) Neurospora crassa endonuclease, 5 x 10^{-2} units; (S_1)S_1 nuclease 6 x 10^{-3} units, 1 min, 5 min, 10 min, 20 min; (T_{1s}) T_1 RNase under nondenaturing conditions 3 x 10^{-4} units, 3 x 10^{-5} units; (H+) controlled acid hydrolysis; (T_1) T_1 RNase under denaturing conditions 5 x 10^{-3} units, 5 x 10^{-4} units. Bracketed nucleotide sequences at left indicate regions digested by nuclease V_1 and at right regions digested by S_1 and T_1 nuclease.

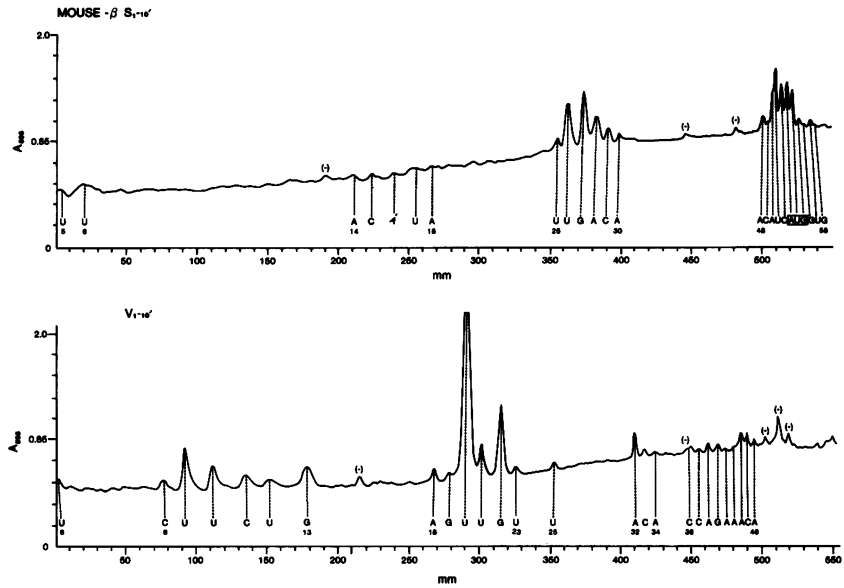


Figure 2. Densitometric scans of S_1 and V_1 partial digests from Figure 1 for mouse β globin mRNA. (-) Indicates band also present in minus enzyme control.

area contiguous to a ^{32}P-end-labeled terminus, due to possible comigration of secondary scissions from previously fragmented RNA. In mouse β mRNA (Fig. 1) there appears to be coincident digestion with both S_1 and V_1 enzymes of nucleotides U_5-U_6. The relatively similar intensities of the coincident S_1 and V_1 cleavages at all digestion times suggests that these nucleotides are either involved in weak base pairing or are a result of secondary cutting of previously fragmented RNA. A densitometric scan of the S_1 and V_1 cleavages for β globin mRNA is shown in Fig. 2.

Fig. 3 shows a structure analysis of an interior portion of mouse β mRNA. Again interpretation of the S_1 and V_1 cleavages is direct. There appears to be some weak cleavage of nucleotides A_{78}-C_{81} by S_1 nuclease and of nucleotides C_{84}-U_{91} by V_1 RNase. Those scissions are only noticeable at later digestion times (10 and 20 minutes) suggesting that they are a result of secondary cutting.

A representative gel displaying a structure analysis of the 3'-coding portion of mouse β mRNA is shown in Fig. 4. The mouse β mRNA can be mapped up to nucleotide C_{442}. Although the A_{336}-C_{341} region is strongly cleaved by S_1 nuclease, two relatively weak V_1 cleavages are present at nucleotides

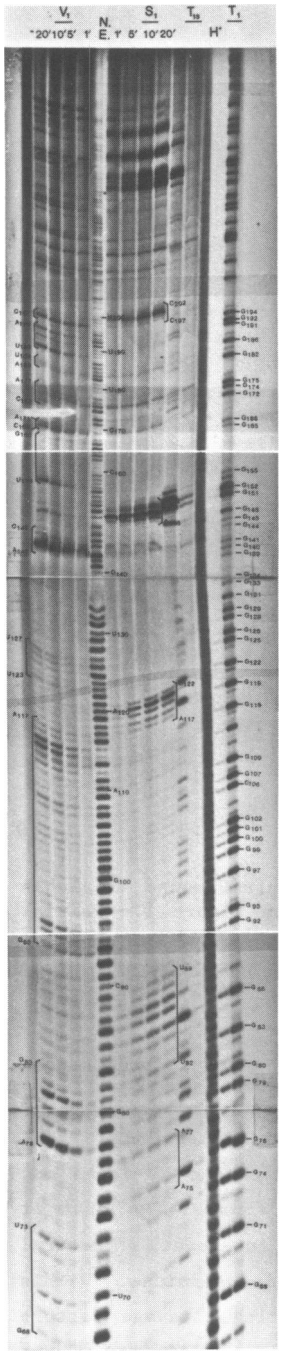


Figure 3. Autoradiogram of partial digests on 5'-[^{32}P]mouse β globin mRNA of nucleotides #68 to #202. Digestion conditions were identical to those in Figure 1. Partial digests were electrophoresed on a 10% polyacrylamide slab gel run in 95% formamide. Gel dimensions were 0.4 mm thick x 33 cm wide x 140 cm long.

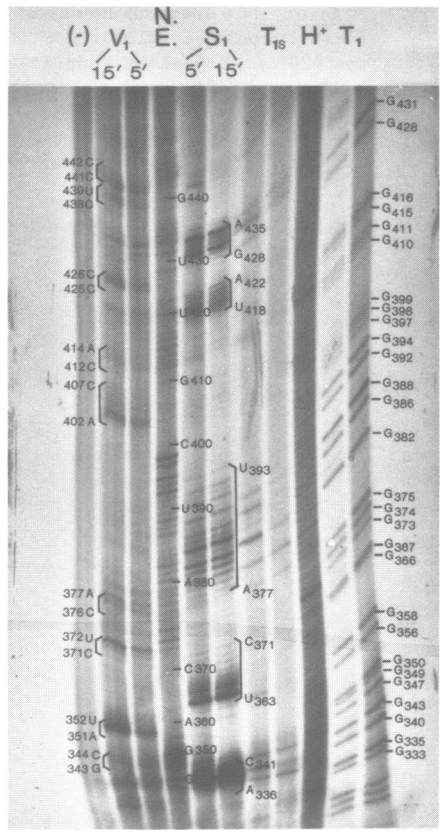


Figure 4. Autoradiogram of partial digests on 5'-[^{32}P]mouse β globin mRNA. Digestion conditions were identical to those in Figure 1. Partial digests were electrophoresed on 6.5% polyacrylamide slab gels run in 7 M urea and 100 mM Tris-Borate-4 mM EDTA, pH 7.2, buffer. Gel dimensions were 0.4 mm thick x 20 cm wide x 110 cm long.

C_{337} and A_{338}. It is not clear whether these two bands result from secondary cutting by V_1 or are actually indicative of possible tertiary interactions which also can be recognized and cleaved by V_1 nuclease (8). The two bands corresponding to nucleotides C_{434} and A_{435} in the V_1 lanes may have been generated radiolytically. It is apparent from Fig. 4 that it becomes increasingly more difficult to confidently interpret structural digests beyond 450-460 nucleotides from a ^{32}P-end-labeled terminus. Within all of the polyacrylamide sequencing gels an alternating cleavage pattern was obtained with S_1 and V_1 nucleases suggesting the existence of one

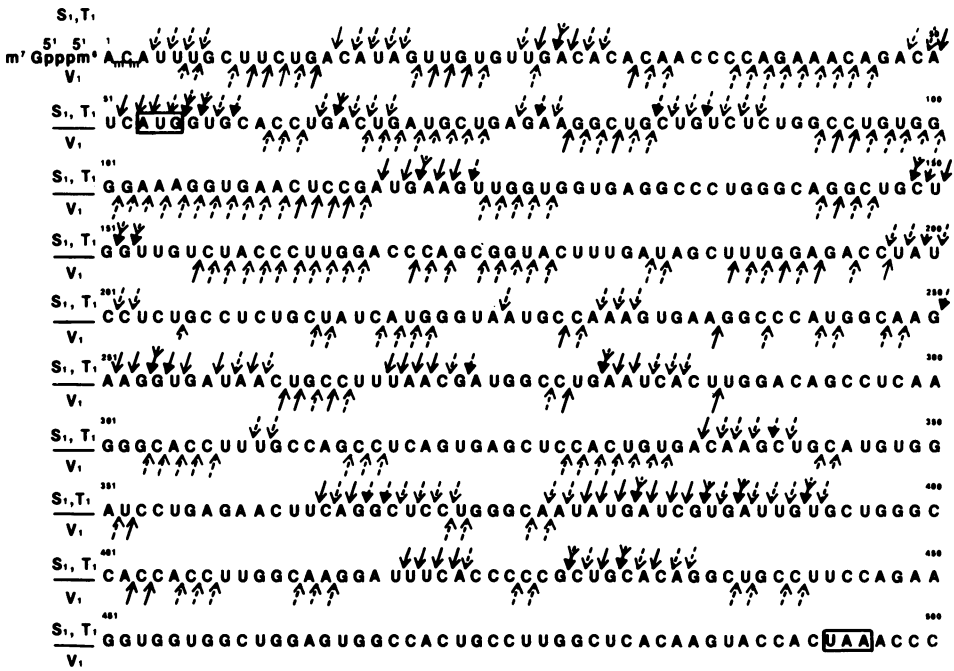


Figure 5. Structure map for mouse β globin mRNA. Single-strand specific S_1 and T_1 cleavages are indicated at the top of the nucleotide sequence and double-strand specific V_1 cleavages at the bottom.

predominant stable secondary structure. Similar results have been obtained with apolipoprotein II mRNA (4). We have found that RNA molecules having more than one conformation generate uninterpretable data on sequencing gels giving simultaneous cleavage in many regions of the molecule with both S_1 and V_1 enzymes. We have determined this to be the case for both E. coli and mouse 5S RNA each of which will renature into two different conformations (18,19).

Computer Analysis of Structure Data

Fig. 5 summarizes the structure data obtained for mouse β mRNA. The secondary structure was derived using a computer algorithm which incorporates the S_1, T_1 and V_1 data, and then utilizes experimentally derived thermodynamic parameters to generate a structure of minimum free energy compatible with the enzymatic data (5,15). Nucleotides cleaved by both S_1 and V_1 nucleases were not included in the database. Fig. 6 shows the resulting structure model for mouse β globin mRNA. The enzymatic structure

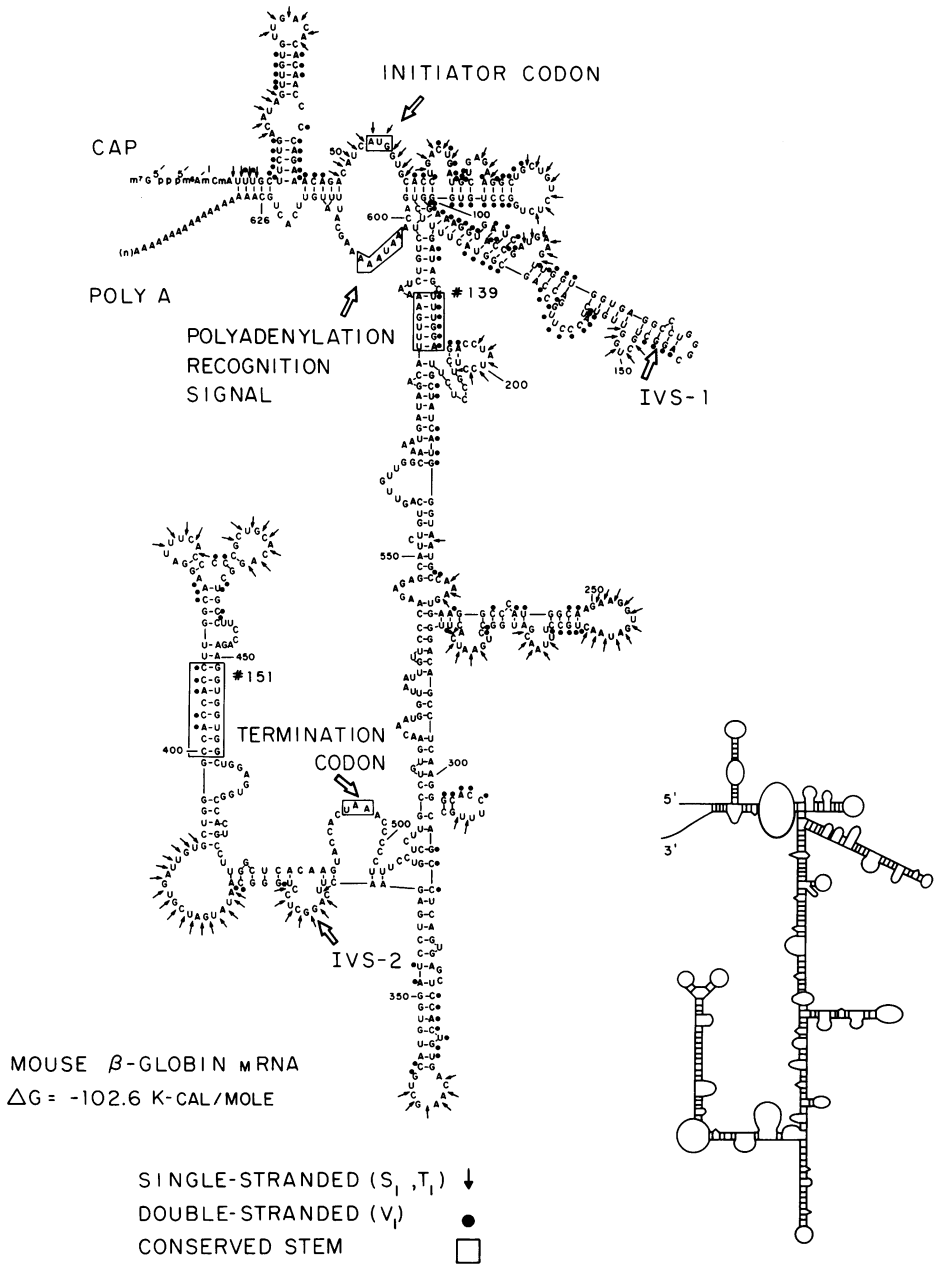


Figure 6. Secondary structure model for mouse β globin mRNA. Phylogenetically conserved stems are boxed.

data is superimposed on the model. Regions designated as IVS-1 and IVS-2 specify points where the two mouse globin introns were likely excised (20).

In the model the percentage of nucleotides actually base paired at 37°C is 55%, in general agreement with previous estimates based on optical melting profiles (21). The free energy estimate for β mRNA is -102.6 Kcal/mole which is higher than that obtained using thermodynamic parameters alone. Without data the program attempts to maximize for the most G + C rich base pairing and generates a structure with a lower free energy value. This observation has already been noted for 5S and ribosomal RNAs (1) and underscores the importance of utilizing ancillary data as a guide in conjunction with thermodynamic parameters in generating structure predictions.

Comparative Sequence Analysis of β Globin mRNA Sequences

Comparative sequence analysis has proven to be a powerful tool in developing secondary structure models for structural RNAs because it can be used to identify and verify regions of potential base pairing which are conserved in evolution (22-24). Since mRNAs spend most of their time being traversed by ribosomes, it could be argued that there is reduced evolutionary pressure to conserve specific structural features within informational RNAs. Furthermore, mRNAs have an additional constraint on sequence variability because they also encode a protein sequence. However, it is likely that there are some conserved structural features in mRNAs that are important in the modification, metabolism, and regulation of the molecule. We have analyzed the β-like globin mRNA sequences in mammals for phylogenetically conserved base paired interactions. As previously reported (6), potential helices with a stem of at least five base pairs (allowing G:U pairing) were found using comparative sequence analysis. Those helices showing compensating base changes (two base transitions or transversions) were additionally screened to determine if they were conserved primarily as a result of conservation of amino acid sequence (6). Two base paired regions were found which are supported by compensating base changes as shown in Fig. 7. Helix #151 was one of the longest conserved stems found with a length of eight base pairs. Five nucleotides within the 5' side of the helix are also digestible with double-strand specific V_1 nuclease. Helix #139 is a six base paired stem between two remote regions of the RNA. Six nucleotides at the 5'-side are cleavable with V_1 nuclease. Both of these phylogenetically conserved stems were

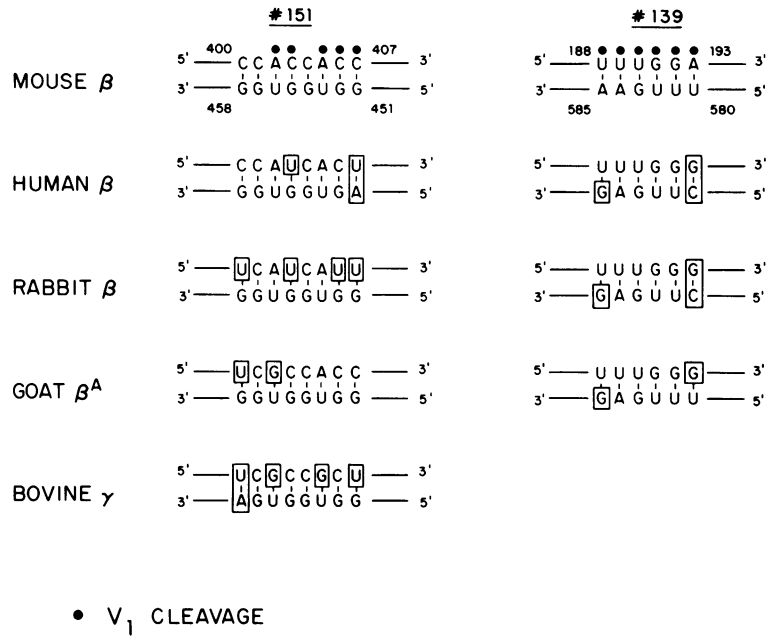


Figure 7. Conserved secondary structures in mammalian β globin mRNAs. The corresponding base paired regions observed in each β globin sequence are shown. Boxed nucleotides indicate transitions and compensating base changes.

also present in the computer generated model shown in Fig. 6, lending support for the model and the long-range interactions between 3'-noncoding and coding portions of the molecule.

DISCUSSION

The work reported here demonstrates one way of generating working secondary structure models for informational RNAs by coupling biochemical data with comparative sequence and computer analysis. The generation of a complete secondary structure model for mouse β globin mRNA derived from phylogenetic data alone has proven difficult (6). While conserved base pairing interactions are readily found in both α and β globin mRNA families, their possible existence in native mRNA structure is not well supported by the presence of compensating base changes. This appears to be due in part to the fact that sequences of the globin mRNA family have not diverged significantly enough to reveal the conserved base pairing interactions among a background of potential base paired regions which

are not involved in forming native structure. In the present study, the two most likely phylogenetically conserved helices shown in Fig. 7 were also generated by computer analysis utilizing the biochemical data, lending support for these structural features in the model.

When the 5'-^{32}P-end-labeled mouse β globin mRNA was renatured under a variety of buffer and temperature conditions, only one predominant, stable secondary structure was evident. Similar results have been obtained for apolipoprotein II mRNA (4), suggesting that mRNA molecules can renature into one thermodynamically favored conformation. This result is not too surprising because mRNAs are indeed dynamic structures becoming partially unfolded when translationally active and refolded when ribosome-free.

The minimum free energy structure for mouse β globin mRNA (Fig. 6) was derived from enzymatic digestion data using 5'-^{32}P-end-labeled mRNA. In an effort to obtain structure data for the unmapped 3'-terminal portion of the mRNA, site-specific RNase H deadenylation was carried out so to specifically label a unique 3'-terminus with ^{32}P for enzymatic structure probing. Even though this approach was used successfully with rabbit globin mRNA (25), mouse β globin mRNA was refractory under identical conditions.

The most interesting feature of the model presented in Fig. 6 is the presence of long-range base pairing interactions which permit the β globin mRNA to fold back on itself, thereby bringing the 5'- and 3'-noncoding regions within close proximity. Electron microscopic evidence for "folded-back" HeLa cell mRNAs has been presented by Hsu and Coca-Prados (26). More recently, Rosenthal and coworkers revealed by electron microscopy that 40% of rabbit globin mRNA molecules prepared by proteinase K digestion of polyribosomes had "folded-back" or "circular" configurations (27). Similar results were previously noted for rabbit and duck globin mRNA-protein particles (28). Such "folded-back" conformations for globin mRNA molecules are reminiscent of earlier studies showing "circular" configurations for rabbit reticulocyte polyribosomes viewed by platinum shadowing (29). Recently, Liebhaber and Kan have presented evidence supporting a role for the 3'-noncoding region of human α globin mRNA in translational initiation (30). These results are consistent with possible base pairing interactions between the 5'- and 3'-noncoding regions in human α globin mRNA. Such interactions have already been implicated in controlling the translational efficiency of zein mRNA in Zea mays (31). The working model proposed in this report should assist in the design

and interpretation of future experiments on the structure analysis of mouse β globin mRNA within messenger ribonucleoprotein particles.

ACKNOWLEDGEMENTS

We thank George Pavlakis and Mel Billingsly for generously providing materials and procedures, and Madhavan Kutty for technical assistance. We are indebted to Hugh Robertson, Ming-Ta Hsu, and Michael Zuker for helpful comments. We thank Terri McKernan for preparing this manuscript. This work was supported by Grant GM35101 from the United States Public Health Service to R.E.L. and Grant AI19578 to C.L.

REFERENCES

1. Auron, P.E., Rindone, W.P., Vary, C.P.H., Celentano, J.J., and Vournakis, J.N. (1982). Nucleic Acids Research 10, 403-418.
2. Gehrke, L., Auron, P.E., Quigley, G.J., Rich, A., and Sonenberg, N. (1983). Biochemistry 22, 5157-5164.
3. Pavlakis, G.N., Lockard, R.E., Vamvakopoulos, N., Rieser, L., RajBhandary, U.L., and Vournakis, J.N. (1980). Cell 19, 91-102.
4. Shelness, G.S. and Williams, D.L. (1985). J. Biol. Chem. 260, 8637-8646.
5. Zuker, M. and Stiegler, P. (1981). Nucleic Acids Research 9, 133-148.
6. Browner, M.F. and Lawrence, C.B. (1985). Nucleic Acids Research 13, 8645-8660.
7. Johnson, R.A. and Walseth, T.F. (1979). Advances in Cyclic Nucleotide Res. 10, 135-165.
8. Lockard, R.E. and Kumar, A. (1981). Nucleic Acids Research 9, 5125-5140.
9. Lockard, R.E. and Lingrel, J.B. (1972). J. Biol. Chem. 247, 4174-4179.
10. Lockard, R.E. and RajBhandary, U.L. (1976). Cell 9, 747-760.
11. Lockard, R.E., Rieser, L., and Vournakis, J.N. (1981). In Gene Amplification and Analysis, Vol. 2, Structural Analysis of Nucleic Acids (J.G. Chirikjian and T.S. Papas, eds.). Elsevier/North Holland, 229-251.
12. Lockard, R.E., Alzner-Deweerd, B., Heckman, J.E., MacGee, H., Tabor, M.W., and RajBhandary, U.L. (1978). Nucleic Acids Research 5, 37-52.
13. Connaughton, J.F., Rairkar, A., Lockard, R.E., and Kumar, A. (1984). Nucleic Acids Research 12, 4731-4745.
14. Salser, W. (1977). Cold Spring Harbor Symp. Quant. Biol. 42, 993-1002.
15. Cech, T.R., Tanner, N.K., Tinoco, I., Weir, B.R., Zuker, M., and Perlman, P.S. (1983). Proc. Natl. Acad. Sci. U.S.A. 80, 3903-3907.
16. Wurst, R.M., Vournakis, J.N., and Maxam, A. (1978). Biochemistry 17, 4493-4499.
17. Favorova, O.O., Fasiolo, F., Keith, G., Vasilenko, S.K., and Ebel, J.-P. (1981). Biochemistry 20, 1006-1011.
18. Kao, T.H. and Crothers, D.M. (1980). Proc. Natl. Acad. Sci. U.S.A. 77, 3360-3364.
19. Toots, I., Misselwitz, R., Böhm, S., Welfle, H., Villems, R., and Saarma, M. (1982). Nucleic Acids Research 10, 3381-3389.
20. Konkel, D.A., Tilghman, S.M., and Leder, P. (1978). Cell 15, 1125-1132.
21. Holder, J.W. and Lingrel, J.B. (1975). Biochemistry 14, 4209-4215.

22. Woese, C.R., Gutell, R., Gupta, R., and Noller, H.F. (1983). Microbiology Reviews 47, 621-669.
23. Stiegler, P., Carbon, P., Ebel, J.-P., and Ehresmann, C. (1981). Eur. J. Biochem. 120, 487-495.
24. Fox, G.E. and Woese, C.R. (1975). Nature 256, 505-507.
25. Vary, C.P.H. and Vournakis, J.N. (1984). J. Biol. Chem. 259, 3299-3307.
26. Hsu, M.-T. and Coca-Prados, M. (1979). Nature 280, 339-340.
27. Ladhoff, A.M., Uerlings, I., and Rosenthal, S. (1981). Mol. Biol. Rep. 7, 101-106.
28. Dubochet, J., Morel, C., Lebleu, B., and Herzberg, M. (1973). Eur. J. Biochem. 36, 465-472.
29. Warner, J.R., Rich, A., and Hall, C.E. (1962). Science 138, 1399-1403.
30. Liebhaber, S.A. and Kan, Y.W. (1982). J. Biol. Chem. 257, 11852-11855.
31. Spena, A., Krause, E., and Dobberstein, B. (1985). EMBO Journal 4, 2153-2158.