Structures of Human and Rabbit β-Globin Precursor Messenger RNAs in Solution[†]

John Teare and Paul L. Wollenzien*

E. A. Doisy Department of Biochemistry and Molecular Biology, St. Louis University Medical Center, St. Louis, Missouri 63104

Received January 10, 1989; Revised Manuscript Received April 10, 1989

ABSTRACT: The structures in solution of human and rabbit β-globin precursor messenger RNAs containing their first intervening sequence have been investigated. This was accomplished by chemical probing experiments to determine sites of potential base pairing, and by cross-linking experiments to determine the sites of long-range interactions. Secondary structures for both molecules were predicted by using this information. Both molecules are arranged into two separate domains. The first domain, consisting of the first exon, contains several long-range interactions between the beginning of the molecule and sites adjacent to the donor splice site and a partially conserved stem/loop structure. The second domain contains part of the intervening sequence and the beginning of the second exon. The secondary structures involved in the second domain are different in the two molecules. These studies indicate a lack of connection between the donor and acceptor splice sites in these two molecules on the level of the secondary structure. Furthermore, given the absence of strongly conserved structures, it is unlikely that there could be any strict requirements for secondary structures that influence splice site usage.

Higher eukaryotic genes are often interrupted by noncoding intervening sequences that must be removed after transcription by the process of RNA splicing (Sharp, 1987). The splicing reaction occurs in a multicomponent ribonucleoprotein complex, the spliceosome, composed of proteins isolated from heterogeneous ribonucleoprotein particles (Swanson & Dreyfus, 1988), the Sm class of small nuclear ribonucleoproteins (Grabowski et al., 1985; Chabot et al., 1985; Frendewey & Keller, 1985; Bindereif & Green, 1987; Kramer, 1987, 1988), and other protein factors (Sharp, 1987; Konarska & Sharp, 1987; Ruskin et al., 1988). Since no examples have been found in which isolated nuclear precursor messenger RNA (pre-mRNA) undergoes the splicing reaction by itself, it is presumably the assembly of the spliceosome that positions the splice sites such that accurate cleavage and exon ligation can occur.

Sequences in the pre-mRNA at the 5′ and 3′ ends of the intron are required for denoting splice sites, and there is evidence that components of the spliceosome interact directly with these sequences (Grabowski et al., 1985; Kramer, 1987). However, sometimes sequences surrounding the splice sites are important for splice site selection (Khoury et al., 1979; Somasekhar & Mertz, 1985; Reed & Maniatis, 1986; Nelson & Green, 1988). To test directly if secondary structure in a pre-mRNA could determine the pattern of splice site usage, Solnick and Lee (1987) inserted sequences into the adenovirus tripartite leader capable of forming secondary structures that loop out the middle exon. They found that the splicing of this exon was skipped both in vivo and in vitro if the secondary structure that sequesters the middle exon was long enough. Similarly, Eperon et al. (1988) found that a stem/loop structure, which sequesters a 5′ splice site within its loop region, would inhibit selection of that splice site in vivo provided that the distance between the complementary sequences that form the stem was below a certain value. Thus, secondary and/or tertiary structure in the precursor mRNA could play an important role in the process of splice site selection, and could possibly contribute in regulation of alternative splicing.

In this report, we have investigated the structures of human and rabbit β-globin pre-mRNAs in order to determine if common structural motifs exist that might be related to splicing. These two molecules have been analyzed under ionic conditions in which they would be assembled into a spliceosome; this is the conformational state of the RNA molecules before they are engaged by any of the components of the nuclear extract. Therefore, if the molecules contain structures that help direct the initial steps of spliceosome assembly, they should be present under this condition. We chose to study these two pre-mRNAs because they have already been the focus of numerous studies aimed at determining the sequences that influence splice site utilization (Reed & Maniatis, 1986; Nelson & Green, 1988; Parent et al., 1987). Our analysis reveals a common structural organization between the human and rabbit precursor mRNAs. These results are discussed in relation to the possible function of such structures in splice site selection and spliceosome assembly in vitro.

Materials and Methods

Materials. Dimethyl sulfate (DMS) and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate (CMCT) were obtained from Aldrich, and kethoxal was purchased from Organic Esearch. Aminomethyltrimethylpsoralen (AMT) was from HRI Associates. Nucleotides, deoxynucleotides, and dideoxynucleotides were from Pharmacia. Avian myeloblastosis virus (AMV) reverse transcriptase was purchased from Life Sciences, SP6 RNA polymerase was from Promega, and T7 RNA polymerase was purified according to Davanloo et al. (1987). Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer.

[†] This work was supported by grants from the National Institutes of Health (GM35410) and by an American Cancer Society Institutional Research Grant (IN-124).

* Address correspondence to this author at the Department of Biochemistry and Molecular Biology, St. Louis University Medical Center, 1402 S. Grand Blvd., St. Louis, MO 63104.

0006-2960/89/0428-6208$01.50/0

pSP4-Δ6-HBG was a gift from Dr. Carlos Goldenberg, and pGEM-RBG was a gift from Dr. Argiris Efstratiadis and Annette Parent.

Plasmids and in Vitro Transcription. The 2.1-kb *Hin*dIII/*Pst*I fragment containing the human β-globin gene from SP64- 6-HBG (Krainer et al., 1984) was cloned into the *Hin*dIII to *Pst*I site of pTZ19R (Genescribe). The resulting plasmid, p19R-HBG, and the plasmid containing the rabbit β-globin gene, pGEM-RBG (Parent et al., 1987), were linearized at the *Bam*HI site in the second exon to give run-off transcripts of 495 and 500 nucleotides, respectively. In vitro transcriptions were performed as described previously (Wollenzien et al., 1987). Labeling of RNA was accomplished by incorporation of [^{32}P]UTP or [^{32}P]GTP with T7 or SP6 RNA polymerase, respectively.

Chemical Modification. Chemical modification of the human and rabbit pre-mRNAs was performed as described (Moazed et al., 1986), with some modifications. Five micrograms of unlabeled RNA was modified in 50 μL of assembly buffer (12 mM Hepes, pH 7.9, 60 mM KCl, 3.2 mM $MgCl_2$, 1 mM DTT, and 12% glycerol) at 4 and at 37 °C for native conditions or in 50 μL of 50 mM cacodylate, pH 7.0, and 1 mM EDTA at 37 and 90 °C for denaturated conditions (Romby et al., 1987). DMS was added to the RNAs as a water-saturated solution (22 mM), kethoxal (37 mg/mL) was added in 20% ethanol, and CMCT (42 mg/mL) was added in either assembly buffer or cacodylate/EDTA buffer. The concentrations of the reagents and the time of exposure were adjusted empirically to give a similar degree of modification under the different conditions. The locations and degree of modifications were determined by a series of primer extension reactions using AMV reverse transcriptase and oligonucleotide primers (Inoue & Cech, 1985; Moazed et al., 1986).

Cross-Linking. Direct AMT cross-linking and cross-linking of RNAs with AMT monoadducts have been described previously (Wollenzien et al., 1987). For preparative-scale cross-linking at 4 °C, 100 μg of ^{32}P-labeled pre-mRNA [(5×10^5)–(5×10^6) cpm/μg] was incubated in 0.4 mL of assembly buffer (1.6 μM final concentration) with 2.6 μM AMT at 4 °C for 10 min.

For reactions at elevated temperatures, AMT monoadducts were first placed on the RNA at 4 °C. RNA (0.21 μM) in assembly buffer was incubated with 0.32 mM AMT for 10 min at 4 °C. The sample was irradiated with 390-nm light (3 mW/cm^2) for 1 h at 4 °C under an N_2 atmosphere. The RNA was separated from unincorporated AMT by ethanol precipitation. Both the direct cross-linking and monoaddition conditions resulted in the covalent attachment of about 1 mol of AMT/mol of RNA. Note that because of the higher concentration of AMT that was used for the monoaddition reaction, it is likely that a larger number of sites contain adducts. The monoadduct samples were then incubated at the indicated temperature before cross-linking with 365-nm light. After cross-linking, the RNA was precipitated, washed with 70% ethanol, and resuspended in 80% deionized, distilled formamide.

The RNA was initially fractionated on a 5% polyacrylamide [20:1 acrylamide/bis(acrylamide)] gel in 1× TBE buffer/8.3 M urea at 36 V/cm for 24 h. The cross-linked RNA was visualized by autoradiography, and the gel was cut into 16–24 fractions with each fraction containing a major cross-link. The RNA was eluted (Wollenzien et al., 1987), resuspended in 80% formamide, and subjected to a second fractionation on a 8% polyacrylamide (20:1) gel in 1× TBE/95% formamide at 36 V/cm for 30 h. For different samples, 23–45 cross-linked RNA species were eluted. Total yields of each fraction were determined by Cherenkov counting and ranged from 4 to 100 ng. A sample of each fraction was electrophoresed on a 5% polyacrylamide gel in a 1× TBE buffer/8.3 M urea gel to determine its purity and on a 1.7% agarose (nondenaturing) gel to determine if the RNA contained intermolecular cross-links. UV cross-linking was also done on the human pre-mRNA; 100 μg of ^{32}P-labeled pre-mRNA was equilibrated in 0.5 mL of assembly buffer for 10 min at 4 °C and then was exposed to 300-nm light (Fotodyne) for 30 min at 4 °C under an N_2 atmosphere. It was then fractionated as described above.

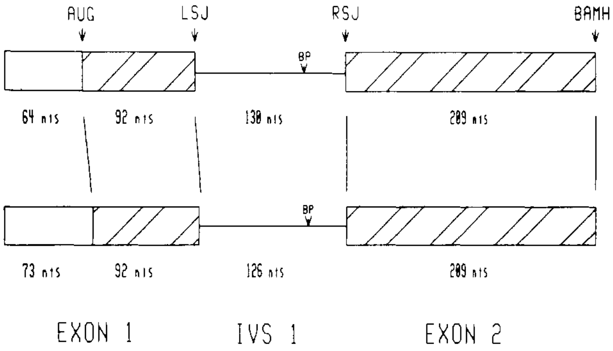

FIGURE 1: Primary structures of the human β-globin and rabbit β-globin pre-mRNAs. The RNAs were synthesized by run-off transcription of the plasmids pTZ19R-HBG and pGEM-RBG that were linearized at a *Bam*HI site located in the 3′ end of exon 2. The protein coding regions, indicated by the hatched areas, are identical in length and show a 90% homology at the nucleotide level. The 5′ nontranslated leader regions and the intervening sequences differ in length and show only a 66% conservation of nucleotides.

The locations of the cross-links in the purified fractions were determined by a series of primer extension reactions carried out as described (Wollenzien et al., 1987; Wollenzien, 1988). The oligonucleotides used as primers were complementary to the following regions: nucleotides 85–102, 176–193, 245–263, 273–290, 332–348, 348–366, 429–449, 459–477, and 474–490 for human pre-mRNA and nucleotides 94–111, 196–213, 255–273, 271–290, 338–356, 353–371, 425–443, 464–482, and 479–495 for rabbit pre-mRNA. Some cross-links in fractions containing small loop sizes were determined after the sample was subjected to photoreversal of the cross-link with 300-nm light. For this reaction, RNA in 50 μL of 2 mM EDTA, pH 8.0, in a quartz cuvette was placed directly in front of a Fotodyne transilluminator. The sample was irradiated for 2 min under N_2 gas and then was ethanol precipitated before analysis by reverse transcription.

Secondary Structure Prediction. PCFOLD, an RNA secondary structure prediction program (Zuker & Sankoff, 1986), was provided by Dr. Michael Zuker. Incorporation of the solution structure data into the computer structure prediction analysis was through the auxiliary information function in the program. For both molecules, the entire sequence was examined with and without the experimentally determined constraints. In addition, overlapping 100-nucleotide subsections of the molecules were examined to find stable local structures. To identify some potential long-range interactions, distant segments of the molecules were joined noncovalently (through the auxiliary information function) and folded.

Results

Chemical Modification of the Precursor mRNA. The primary structures of the human β-globin and rabbit β-globin pre-mRNAs used in these experiments are diagrammed in Figure 1. The RNAs were synthesized in vitro from the

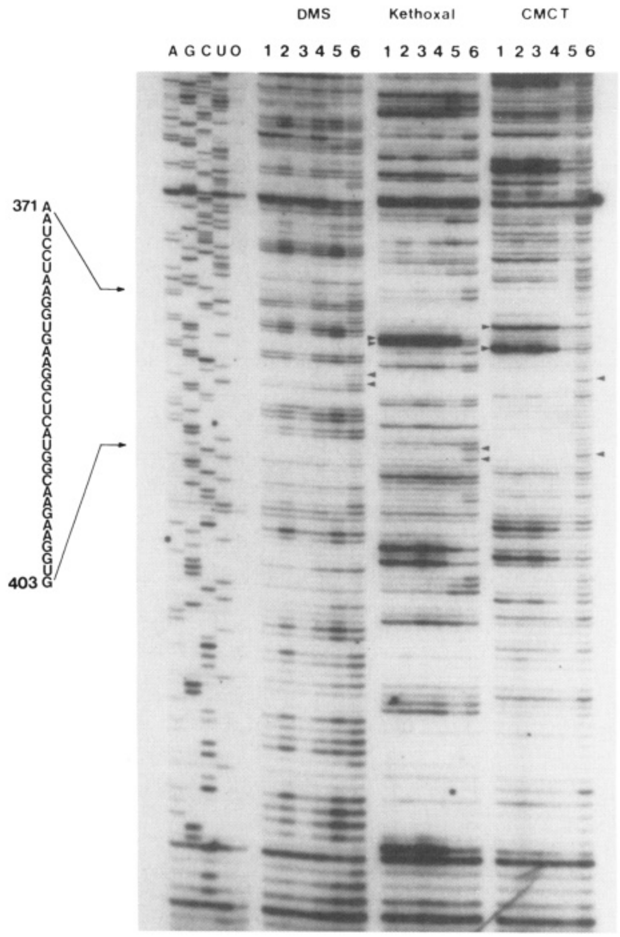
DMS
Kethoxal
CMCT
A G C U O
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
371
A A U C C U A A G G U G A A G G C U C A U G G C A A G A A G G U G
403

FIGURE 2: Reverse transcription analysis of the chemical probing of the rabbit pre-mRNA. The 5′ end-labeled oligonucleotide used for primer extension hybridizes to nucleotides 464–482 of the rabbit pre-mRNA. Sequencing reactions (lanes A, G, C, and U) and a control reaction (lane O) were done on untreated RNA. Chemically modified rabbit pre-mRNA was examined in the other lanes shown. DMS: lanes 1 and 3, 4 °C in assembly buffer; lanes 2 and 4, 37 °C in assembly buffer; lane 5, 37 °C in cacodylate/EDTA buffer (semidenatured conditions); lane 6, 90 °C in cacodylate/EDTA buffer (denatured conditions). Kethoxal: lanes 1 and 3, 4 °C in assembly buffer; lanes 2 and 4, 37 °C in assembly buffer; lane 5, 37 °C in cacodylate/EDTA buffer; lane 6, 90 °C in cacodylate/EDTA buffer. CMCT: lanes 1 and 3, 4 °C in assembly buffer; lanes 2 and 4, 37 °C in assembly buffer; lane 5, 37 °C in cacodylate/EDTA buffer; lane 6, 90 °C in cacodylate/EDTA buffer. The arrows pointing left indicate nucleotides protected from chemical modification under native conditions. Arrows pointing right indicate nucleotides that exhibit enhanced reactivity toward the chemical reagents under native conditions.

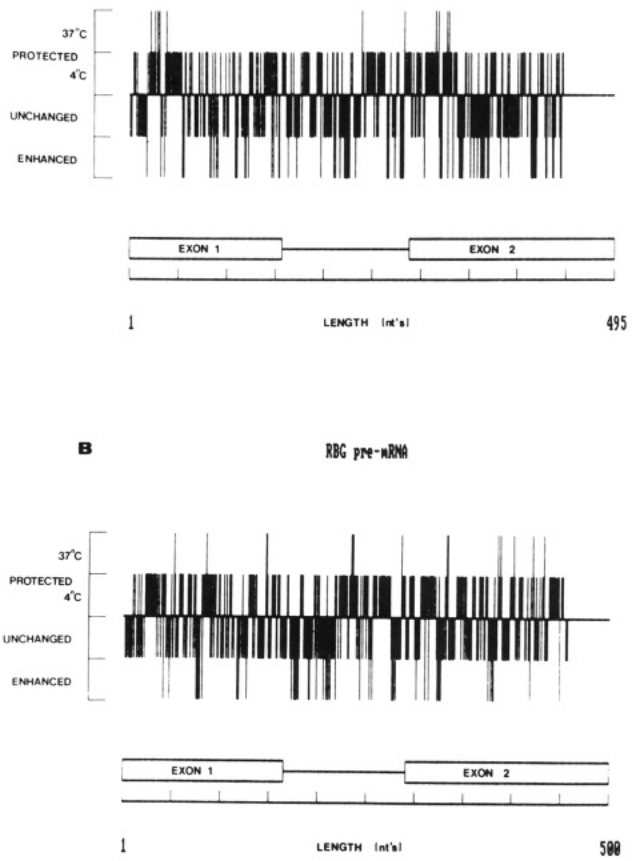

FIGURE 3: Summary of the chemical probing data for the human β-globin pre-mRNA and the rabbit β-globin pre-mRNA. Each nucleotide of human β-globin (panel A) or rabbit β-globin (panel B) that was scored for reactivity to the DMS, kethoxal, and CMCT is presented as a line of the histogram. Nucleotides protected from chemical modification under native conditions at 4 and 37 °C are represented by the short and long lines pointing upward. The short lines pointing downward represent the nucleotides that were unchanged in their chemical reactivities between the native and denatured states. The locations of the nucleotides with enhanced chemical reactivities under native conditions are indicated by long lines pointing downward.

plasmids pTZ19R-HBG and pGEM-RBG, both linearized at *Bam*HI sites located 17 nucleotides upstream from the 3′ end of the second exon. The resulting transcripts of 495 and 500 nucleotides encode the first exon, the first intron, and most of the second exon of each gene.

To determine which residues of the pre-mRNAs were involved in base pairing interactions when the RNAs were in spliceosome assembly buffer, we performed a series of chemical probing experiments. Dimethyl sulfate (DMS), kethoxal, and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate (CMCT) were used to identify base-paired nucleotides as they do not react with bases involved in hydrogen-bonding interactions (Noller et al., 1987). Pre-mRNAs were incubated in assembly buffer (see Materials and Methods) at 4 or 37 °C and were treated with one of the three reagents. Control RNA samples that establish the inherent reactivity of each base under semidenatured and denatured conditions were made by resuspending the RNA in a cacodylate/EDTA buffer and chemically modifying it at 37 and 90 °C. The extent of chemical modification at each site was determined by primer extension analysis (Inoue & Cech, 1985; Moazed et al., 1986) using five oligonucleotide primers for each of the precursors. The resulting cDNAs terminate one nucleotide before a base that was chemically modified (Hagenbuchle et al., 1978; Youvan & Hearst, 1979).

An example of one chemical modification experiment is shown in Figure 2. Each nucleotide was judged for its chemical reactivity as one of six possible values: (1) unreactive under all conditions, (2) enhanced reactivity in the native compared to the denatured state, (3) unchanged in reactivity in the native and denatured state, (4) unreactive in the native state at 4 °C compared to the denatured state, (5) unreactive in the native state at 37 °C compared to the denatured state, or (6) the nucleotide could not be scored because of high background of hydrolysis. It was possible to score 85% of the bases in both pre-mRNA molecules for their chemical reactivity according to these categories. The positions that were scored were distributed evenly over the entire length of the RNA such that there is no bias in the data for any particular base or region. The 3′-terminal 40 nucleotides of each molecule were exceptions to this analysis because they could not be assayed by the reverse transcription experiment. The reactivity over these molecules is summarized in the histograms

A Human Beta Globin Pre-mRNA : Chemical Probing Data

```
        10        20        30        40        50
         |         |         |         |         |
GGGAAAGCUUGCUUACAUUUGCUUCUGACACAACUGUGUUCACUAGCAAC
+++000+0+++++++++*000^0000′J^0^*+++*0+^*0000000000

        60        70        80        90       100
         |         |         |         |         |
CUCAAACAGACACCAUGGUGCACCUGACUCCUGAGGAGAAGUCUGCCGUU
000***00++0+00 +++++00+0-++000+***+0*+0**  0 ++0++

       110       120       130       140       150
         |         |         |         |         |
ACUGCCCUGUGGGGCAAGGUGAACGUGGAUGAAGUUGGUGGUGAGGCCCU
+0  000***+00- ++**+*00000++0+0++0--0+00000000*00+

       160       170       180       190       200
         |         |         |         |         |
GGGCAGGUUGGUAUCAAGGUUACAAGACAGGUUUAAGGAGACCAAUAGAA
0**  0000++*++0+0+++*+++0+0000000+++++-+00000*0+++

       210       220       230       240       250
         |         |         |         |         |
ACUGGGCAUGUGGAGACAGAGAAGACUCUUGGGUUUCUGAUAGGCACUGA
+0++++00*+*00+*+0++******+*0 0++*+++++^00++0000000+

       260       270       280       290       300
         |         |         |         |         |
CUCUCUCUGCCUAUUGGUCUAUUUUCCCACCCUUAGGCUGCUGGUGGUCU
0+0+000000  ***00+ +*+000*****^000000++0++00+0+0 +

       310       320       330       340       350
         |         |         |         |         |
ACCCUUGGACCCAGAGGUUCUUUGAGUCCUUUGGGGAUCUGUCCACUCCU
*000000000 0^0^+***0000^*^000*0000++*****++00+0-*0+

       360       370       380       390       400
         |         |         |         |         |
GAUGCUGUUAUGGGCAACCCUAAGGUGAAGGCUCAUGGCAAGAAAGUGCU
+*++*0++*****0+0**+00+0000++++*00++++0*0++++++++0+0

       410       420       430       440       450
         |         |         |         |         |
CGGUGCCUUUAGUGAUGGCCUGGCUCACCUGGACAACCUCAAGGGCACCU
+0000 + 00*******+0  000 +**0 +0+****+000+**00

       460       470       480       490
         |         |         |         |
UUGCCACACUGAGUGAGCUGCACUGUGACAAGCUGCACGUGGAUC
```

B Rabbit Beta-globin Pre-mRNA : Chemical Probing Data

```
        10        20        30        40        50
         |         |         |         |         |
GAAUACAAGCUCUGCUGCUUACACUUGCUUUUGACACAACUGUGUUUACU
 ++++++++0+0++0++0+++000000000000+0+0+00*00+++*+0+

        60        70        80        90       100
         |         |         |         |         |
UGCAAUCCCCCAAAACAGACAGAAUGGUGCAUCUGUCCAGUGAGGAGAAG
+^0+++00000++++000+0+0+0*****0*0 0^000+0+0+00++00+

       110       120       130       140       150
         |         |         |         |         |
UCUGCGGUCACUGCCCUGUGGGGCAAGGUGAAUGUGGAAGAAGUUGGUGG
+++0+0++000++---****+0+ ++*+0000+0+00++++++++^^0++

       160       170       180       190       200
         |         |         |         |         |
UGAGGCCCUGGGCAGGUUGGUAUCCUUUUUACAGCACAACUUAAUGAGAC
++0000++000+++++++00++*++******+0000+++++*++++0+0*0

       210       220       230       240       250
         |         |         |         |         |
AGAUAGAAACUGGUCUUGUAGAAACAGAGUAGUCGCCUGCUUUUCUGCCA
+0+++*+++*0***+-++0*0+00+000000++00^^^00+0++*0+000+

       260       270       280       290       300
         |         |         |         |         |
GGUGCUGACUUCUCUCCCCUGGGCUGUUUUCAUUUUCUCAGGCUGCUGGU
00+***00000 0 000000000 0+*****++++++^000++-+00000+

       310       320       330       340       350
         |         |         |         |         |
UGUCUACCCAUGGACCCAGAGGUUCUUCGAGUCCUUUGGGGACCUGUCCU
+*+++00000*00*0000000+**0*****0+^^+++++ + 0000+00+

       360       370       380       390       400
         |         |         |         |         |
CUGCACAUGCUGUUAUGAGCAAUCCUAAGGUGAAGGCUCAUGGCAAGAAG
0+0+++++0000+++00+00++000*++***0+++^000-+++++++000

       410       420       430       440       450
         |         |         |         |         |
GUGCUGGCUGCCUUCAGUGAGGGUCUGAAUCACCUGGACAACCUCAAAGG
0^0+++0+000 ++ 0++*0^0++0+++0+00^++++0000000000*

       460       470       480       490       500
         |         |         |         |         |
CACCUUUGCUAAGCUGAGUGAACUGCACUGUGACAAGCUGCACGUGGAUC
000 +++
```

FIGURE 4: Chemical reactivities of the nucleotides in the human and rabbit pre-mRNA. Panels A and B show the complete sequences of the human and rabbit pre-mRNA and their reactivities. Noted are nucleotides that were chemically unreactive under all conditions (−), nucleotides that were hyperreactive in the native state compared to the denatured state (*), nucleotides that were as reactive in the native state as in the denatured state (+), nucleotides that were unreactive at 4 °C in the native state compared to the denatured state (0), and nucleotides that were unreactive at 37 °C in the native state compared to the denatured state (^). A number of positions in each molecule could not be scored due to a high frequency of hydrolysis even in control samples.

of Figure 3. The reactivities at each nucleotide are shown in the list in Figure 4.

Both protection and enhancement of chemical reactivity occur throughout the human and rabbit β-globin RNA molecules, indicating that there must be secondary structures in both molecules that give rise to the specific pattern. However, there are no long stretches of protected nucleotides that would indicate the presence of extensive stable complementarities. This indicates that structures will be composed of short complementary regions interspersed with single-stranded segments. The absence of obvious base-paired regions raises the possibility that the molecules could contain a combination of short-range and long-range interactions and that they could have multiple conformations. To help determine the arrangements of the RNAs, an alternate structural analysis by intramolecular cross-linking was undertaken.

Psoralen Cross-Linking To Determine the Overall Structure. Cross-linking experiments were performed to determine the identity of base-paired interactions within these molecules. To determine first if there were thermally stable long-range interactions, psoralen (AMT) monoadducts were attached to the pre-mRNAs (Wollenzien et al., 1987), and the RNAs were then resuspended in assembly buffer, equilibrated at different temperatures, and cross-linked at those temperatures. The pattern of cross-linking, shown in Figure 5, indicates that the RNA molecules directly cross-linked at 4 °C and the RNA/monoadduct molecules cross-linked at 4 °C through 30

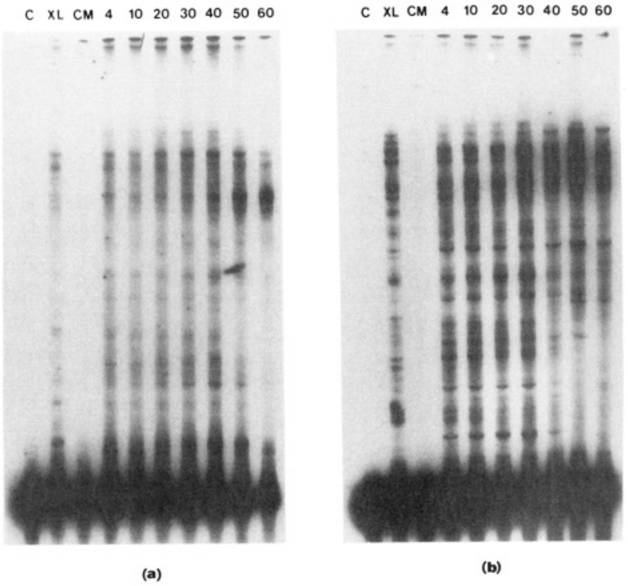

FIGURE 5: Psoralen cross-linking patterns of pre-mRNAs in assembly buffer at different temperatures. Approximately 0.15 μg of human pre-mRNA (a) or rabbit pre-mRNA (b) was loaded in each lane of a 5% polyacrylamide/8.3 M urea gel and electrophoresed for 24 h at 36 V/cm. Lane C contains un-cross-linked RNA; lane XL contains RNA directly cross-linked at 4 °C; lane CM contains RNA with AMT monoadducts; lanes 4–60 contain monoadduct RNAs that were subsequently converted to cross-links at the indicated temperatures.

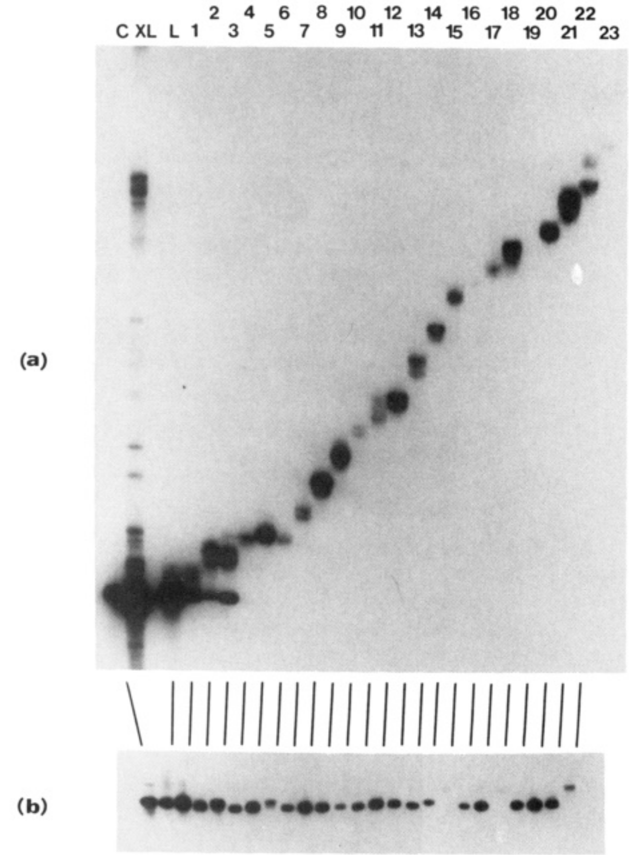

FIGURE 6: Characterization of fractionated, cross-linked human pre-mRNA molecules. The RNA was cross-linked in assembly buffer at 4 °C and then fractionated by two gel electrophoresis steps as described under Materials and Methods. The fractions were then electrophoresed on polyacrylamide under denaturing conditions (lanes 1–23, panel a) or on nondenaturing agarose gels (lanes 1–23, panel b). Lane C contains linear RNA, lane XL contains total unfractionated cross-linked RNA, and lane L contains the RNA fraction that was treated with psoralen and electrophoresed with the same mobility as linear RNA. Note that all of the fractionated samples on the agarose gel electrophoresed with the mobility of monomer-sized human pre-mRNA except fractions 16, 19, and 23 which have the mobilities of dimers.

°C have very similar banding patterns. This suggests that both the human and rabbit pre-mRNAs can fold into definite structures. Above 30–40 °C, some bands start to disappear, and others become more intense, indicating changes in the thermal stability of certain regions involving psoralen binding sites.

The structures of the RNA molecules cross-linked both at 4 °C and at 60 °C in assembly buffer were determined by reverse transcription of fractionated cross-linked molecules. For purification of individual cross-linked RNA species, a fractionation system was developed that resulted in sufficiently pure RNA species such that the location of cross-links could be determined. The purity of the fractions was determined by reelectrophoresing samples on denaturing and nondenaturing gels (Figure 6). The locations of the cross-links in each fraction were determined by a series of primer extension experiments since a psoralen adduct is an absolute stop for reverse transcriptase (Wollenzien, 1988; Ericson & Wollenzien, 1988). Assignments of cross-links to particular reverse transcription stops were based on three criteria. First, only the stops that were identified by different primers in the same cross-linked RNA fraction could be considered as being the 5′ and 3′ ends of a psoralen cross-link, since one reverse transcription reaction could not detect both parts of one cross-link. Second, psoralen reacts within secondary structures at opposite adjacent pyrimidine bases (Thompson & Hearst, 1983; Garrett-Wheeler et al., 1984), so potential pairs of sites were examined to determine if they contained this arrangement. The final criterion was based on the mobility of the purified cross-linked RNA fraction on 5% polyacrylamide denaturing gels. The fractionation system separates the RNA molecules on the basis of the size of the loop formed by the psoralen cross-link. The relative migration on the denaturing gel is inversely related to the size of the loop. Plots of the relative migration of the cross-linked RNAs in a fraction versus the loop size of the putative cross-links proved useful in making assignments in fractions which contained more than one cross-linked species (data not shown).

An example of the identification of a cross-link between nucleotides U273 and U376 in the human β-globin pre-mRNA is shown in Figure 7. In this example, each of the 23 samples that resulted from fractionation of the cross-linked RNA were reverse-transcribed with the primer that initiates at position 429 (panel a) or with the primer that initiates at position 348 (panel b). Reverse transcription stops at U377 and at U274 suggested the possibility that fraction 15 contained molecules in which U376 and U273 were cross-linked together. A base-paired interaction can be made around the site of these nucleotides (see Figure 11); the size of the loop, 102 nucleotides, is consistent with its electrophoretic mobility ($R_f = 0.45$).

Cross-linking data for both the human and rabbit β-globin pre-mRNAs cross-linked at 4 and 60 °C are illustrated in Figure 8. Each line represents the results from one fraction. Reverse transcription stops in each fraction are shown as small lines at the corresponding nucleotide position, and nucleotides that have been assigned as being cross-linked together are connected by brackets. The fraction indicated by "MA" in each drawing shows the locations of the reverse transcription stops that occur in all fractions of the RNA. These stops may be due to psoralen monoadducts; alternately, they may represent the 3′ end of cross-links that have loop sizes less than about 25 nucleotides, since the fractionation system does not separate molecules with loop sizes of 25 nucleotides or less from linear RNA molecules. These data are plotted in two dimensional histograms, shown in Figure 9. In these plots, each point indicates the 5′ and 3′ position of a cross-link on the *x* and *y* axes.

These data indicate that both RNA molecules have two major structural domains. One domain, stable only at lower temperatures, is located in the first exon. In the molecules used for these experiments, the first 15 nucleotides of the human and the first 20 nucleotides of the rabbit pre-mRNA are encoded by vector and nontranscribed genomic sequences. These extra nucleotides do not alter the structure in this region extensively, since only one cross-link, between U20 and U159, in the rabbit pre-mRNA would not exist in the authentic pre-mRNA transcribed in vivo.

The second domain includes cross-links between the intron and the 5′ end of the second exon. However, the exact location of the second domain differs for the two pre-mRNAs. The human pre-mRNA has a long extended structure involving the central sequences of exon 2 and mainly the 3′ half of the intron, at 4 and 60 °C. In the rabbit molecule, more complex structures exist, as indicated by the pattern of cross-links in the intron and 3′ exon. Although structures containing connected complementary regions can be proposed, the existence of different cross-links that share some of the same nucleotides indicates that several alternative structures would have to exist for this region of the rabbit pre-mRNA. On close inspection, many of these cross-links do not occur at significant com-

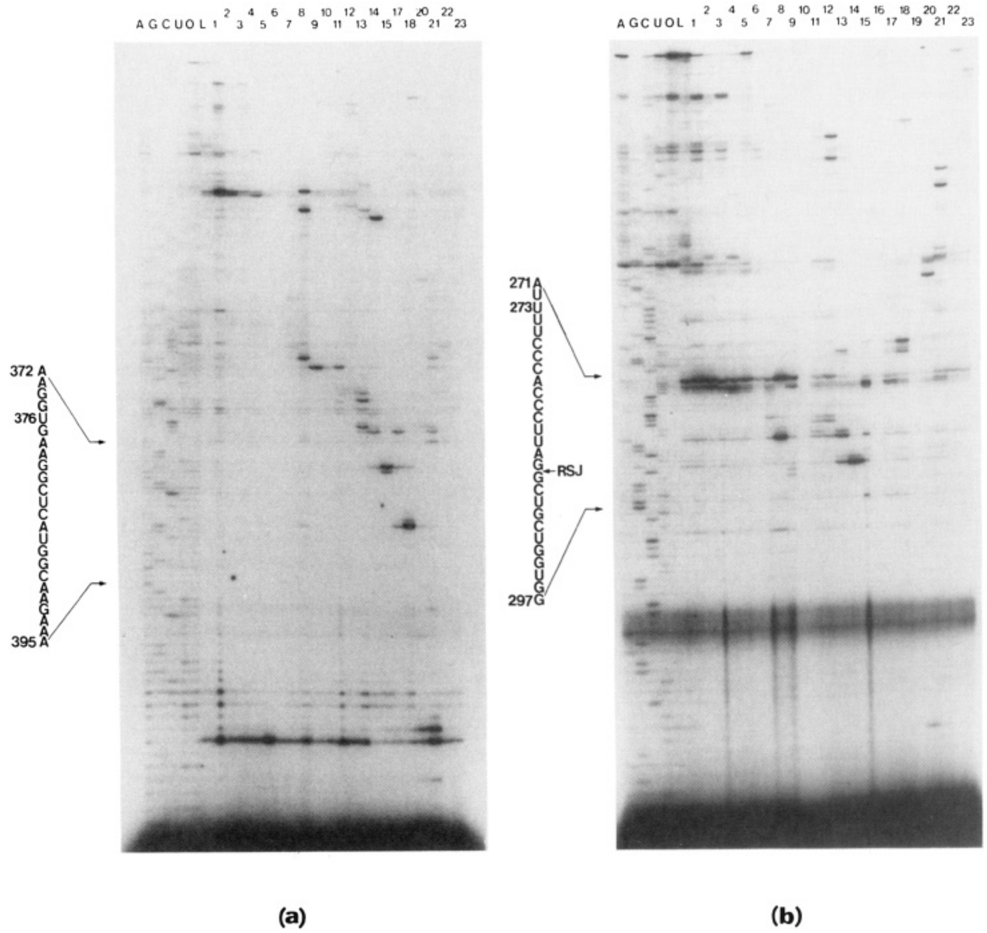
A G C U O L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 20 21 22 23
372 A
A
G
G
376 U
G
A
A
G
G
C
U
C
A
U
G
G
C
A
A
G
A
A
395 A
A G C U O L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
271 A
U
273 U
U
U
C
C
C
A
C
C
C
U
U
A
G
RSJ
G
C
U
G
C
U
G
G
U
G
297 G
(a)
(b)

FIGURE 7: Reverse transcription analysis of the fractionated, cross-linked human pre-mRNA. Samples containing RNA from fractions 1–23, shown in Figure 5, were extended from primers that hybridize to nucleotides 348–366 (panel b). Lanes A, G, C, and U contain sequencing reactions, lane O contains extension products of linear template RNA as a control, and lane L contains extension products from fractions L, purified from the cross-linking reaction. Lane L is a control for reverse transcription stops caused by AMT monoadducts. Lanes 1–23 are the primer extension reactions on the indicated fractions. A reverse transcription stop at U 376 in fraction 15 is shown in panel a. The other end of the cross-link in this fraction, U 274, is shown in panel b.

plementary regions, and they are not consistent with the chemical probing data (see below). In an attempt to understand the reason for the behavior of the rabbit pre-mRNA, we examined the nucleotide and dinucleotide composition of both molecules. The rabbit pre-mRNA has a higher frequency of UpU and UpC dinucleotides than the human pre-mRNA with a large difference occurring in the 5′ half of the intron (35% in the rabbit versus 18% in the human pre-mRNA). These types of sequences are favorable for psoralen reaction as determined by the pattern of monoadduct formation. This difference in dinucleotide frequency may explain the increase in psoralen cross-linking in this region of the rabbit RNA.

As a verification for the structural determinations using psoralen as a photochemical probe, we determined the locations of UV-induced cross-links in the human pre-mRNA. Cross-linking was performed in assembly buffer at 4 °C, and the RNA was fractionated and analyzed as described above. Some of the UV-induced cross-links and psoralen-induced cross-links occur in similar areas of the molecule (Figures 8C and 9). Two UV cross-links were identical with two psoralen cross-links. Although we have determined a smaller number of the UV cross-links, the cross-linking pattern shows the same general arrangement as in the psoralen data set. We interpret these data to indicate that psoralen has not altered the overall structure of the human pre-mRNA.

Comparison of the Chemical Probing and Cross-Linking Experiments. The chemical modification data and the cross-linking data for both molecules were compared on two-dimensional plots, shown in Figure 10. The small marks on these plots indicate sites in the molecule (pairs of nucleotides whose positions are indicated on the *x* and *y* axes) that are candidates for being base-paired. Nucleotide pairs at these sites are unreactive to the chemical reagents under native conditions and are complementary. An additional criterion was usually imposed on the data to screen out random complementary protected base pairs: a mark was drawn only if the next potential base pair adjacent to the first also was protected and was complementary. In another analysis, we required three adjacent base pairs for a mark to be drawn. Since it is likely that secondary structure interactions will be longer than two or three base pairs, these restrictions screen out noise from the random matching of protected bases. It must be emphasized that these plots show all potential interactions that could occur given the pattern of chemical reactivity.

When the cross-linking data were superimposed on the chemical probing data, there were overlaps indicating the locations of stable long-range structures within the molecule Figure 10C,D). At some potential base-pairing sites, there are multiple psoralen cross-links that confirm the interaction. As expected, many of the potential secondary structure interactions are not associated with cross-links. To determine if these potential interactions could have been cross-linked if they existed under native conditions, 20 potential base-pairing interactions that were not close to experimentally detected cross-links were examined. All of these potential sites, except

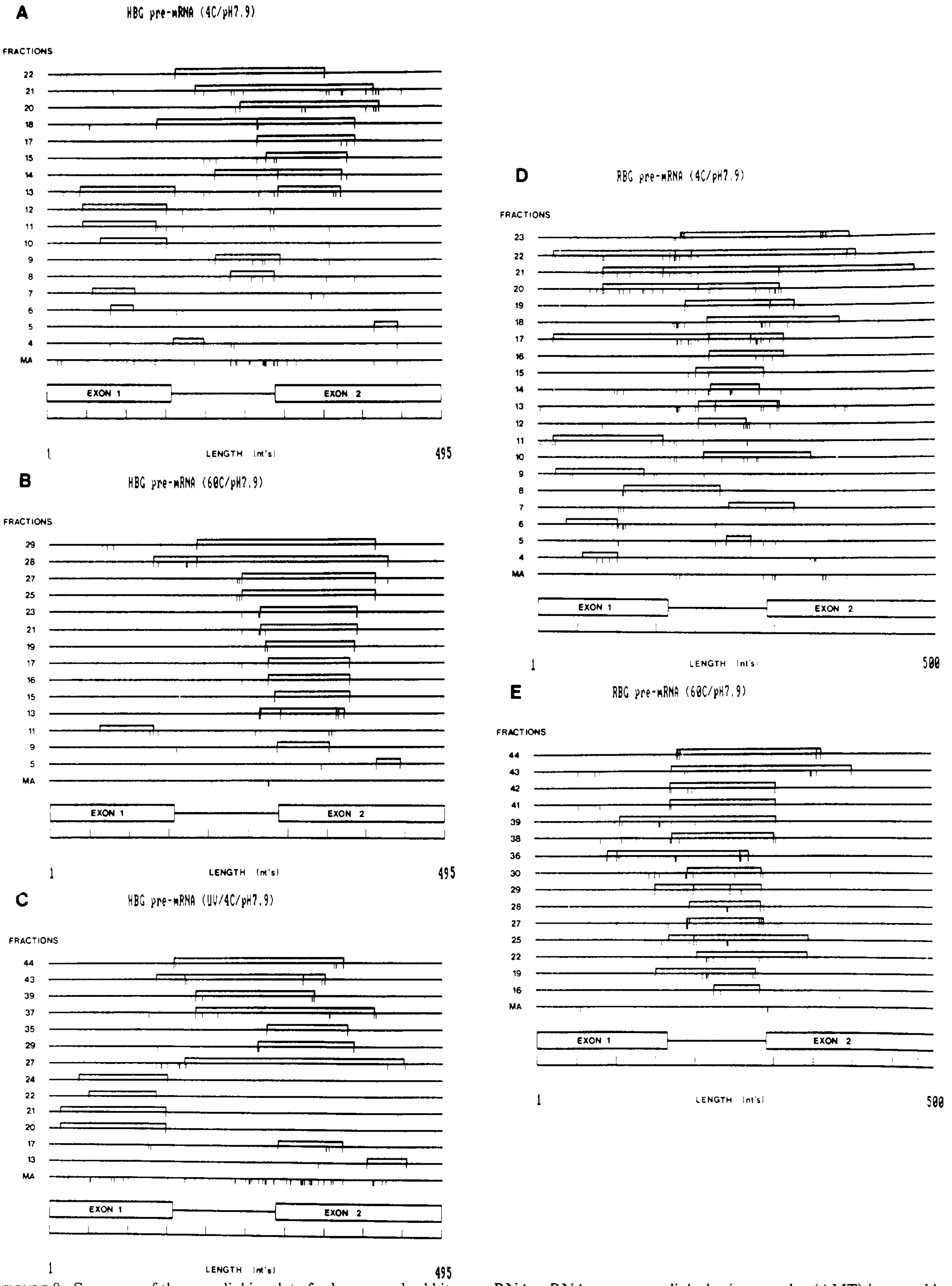

FIGURE 8: Summary of the cross-linking data for human and rabbit pre-mRNAs. RNAs were cross-linked using psoralen (AMT) in assembly buffer at 4 °C (A and B) or at 60 °C (B and E). The UV-induced cross-links located in human pre-mRNA in assembly buffer at 4 °C are shown in panel C. The locations of the reverse transcription stops in each fraction are shown as the short lines pointing down from the center line of each fraction, and the stops that are assigned as being cross-links are represented as a bar connecting the reverse transcription stops above the center line. Psoralen monoadducts (or UV-induced stopping points in panel C) that are in every fraction are shown on the bottommost line of each graph.

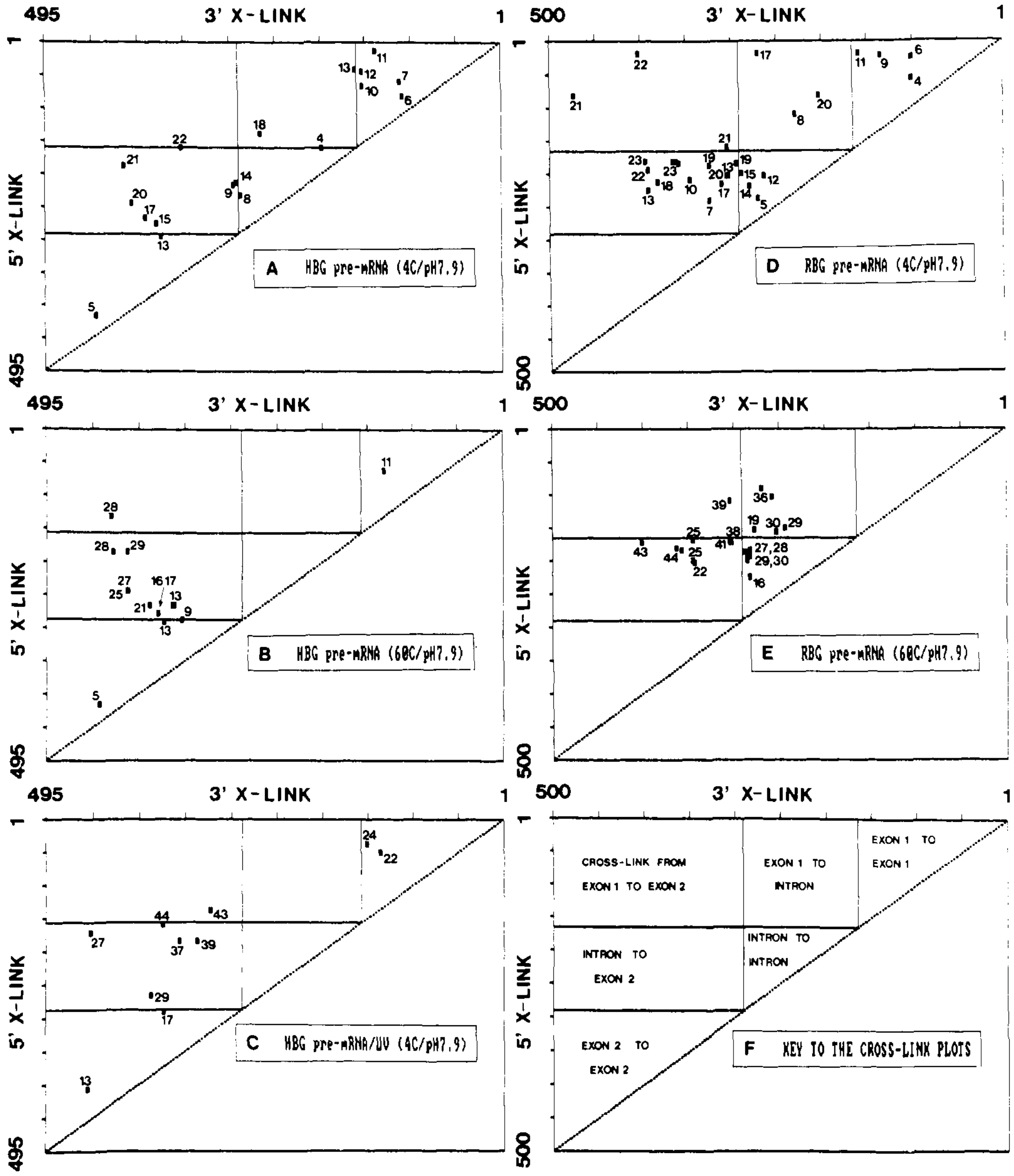

FIGURE 9: Two-dimensional plots of cross-linking sites in the human and rabbit pre-mRNA. For each cross-linked pair of nucleotides indicated in Figure 8, the residue number of the 3′-nucleotide is plotted versus the residue number of the 5′-nucleotide. The arrangements of panels A–E are the same as in Figure 8; panel F indicates the correlation between the sections of the plots and the segments of the RNA in intramolecular interactions.

three, contain opposite adjacent uridine/uridine residues or opposite adjacent uridine/cytidine residues in, or immediately adjacent to, the potential base-paired interaction. These sites are similar in pyrimidine configuration to the sites that were regularly cros-linked. The three other potential base-paired interactions contained sites, two nucleotides away, that also would have been cross-linked; cross-linking at such sites has also been seen in the experimental data. Therefore, in nearly all of the cases, the absence of cross-linking indicates that the potential base-paired interactions are only fortuitous matches that do not actually exist in the RNA.

In the data shown in Figure 10C,D, not all the cross-links are associated with potential base-pairing interactions. These cross-links may represent regions of the RNA that can form alternative structures which would not be protected in all molecules during the chemical probing experiments. Alternately, the cross-links may occur at the sites of weak interactions that also would not result in chemical protection. Notably, a cluster of psoralen cross-links involving the 5′ half of the intron in the rabbit pre-mRNA are not associated with potential base pairs (Figure 10C). As discussed in the preceding section, it is likely that in this region, the unusual reactivity of the sequence with psoralen is responsible for these cross-links.

Since the gel separation system does not separate cross-linked molecules that have loops smaller than about 25 nucleotides in size from linear molecules, the cross-linking experiments have not provided data for short-range interactions. The thin line in the plots in panels C and D in Figure 10 denotes the limitation of resolution of cross-links that can be determined by the fractionation system. It is evidence from these plots that both RNA molecules have potential structures in this size range.

Prediction of the Secondary Structure of the Pre-mRNA Molecules. An RNA folding program, PCFOLD (Zuker & Sankoff, 1986), has been used to help construct secondary structure models for the human and rabbit pre-mRNAs that are consistent with as much of the chemical modification data and cross-linking data as possible (Figure 11). This was done by predicting the secondary structure of segments of the RNA

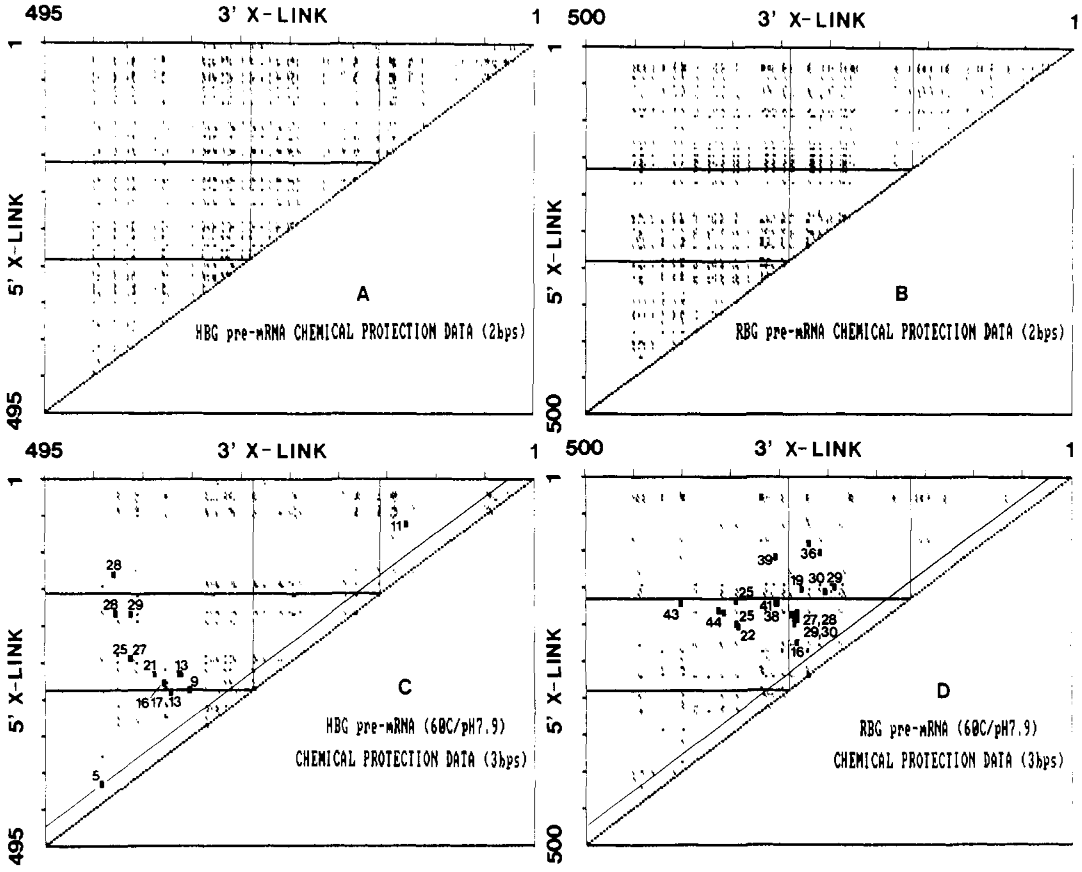

FIGURE 10: Two-dimensional analysis of chemical protection data and cross-linking data. The chemical protection for each molecule was processed to illustrate the potential sites of base pairs. The axes of the two-dimensional plots are the residue number of each nucleotide in the molecule. In panels A and B, a mark is made if two adjacent nucleotides (their positions indicated on the top axis) are unreactive to the chemical reagents and are complementary to two adjacent nucleotides, also unreactive to chemical reagents (positions indicated on the vertical axis). Panels C and D show the same data plotted on the two-dimensional histogram except that three adjacent complementary protected nucleotides were needed for a mark to be drawn. In panels C and D, the psoralen cross-links in each molecule at 60 °C are also indicated. Data for human pre-mRNA are shown in panels A and C; data for rabbit pre-mRNA are shown panels in B and D. The thin line above and parallel to the diagonal represents the limit of resolution of the cross-linking experiments.

without constraints and then examining them to determine if the structures satisfied the experimentally determined parameters. We have also imposed the experimentally determined constraints on the folding as it was being done, by using the auxillary function of the PCFOLD program. In some regions of the RNA molecules, more than one secondary structure was equally consistent with the experimental data, and in these instances, alternate structures are indicated. In the models in Figure 11, cross-links are indicated on the secondary structure models and in the alternate versions of some regions. Regions for which the computer-predicted structures are not consistent with the experimentally determined data have been left single stranded.

Examination of the Structure Surrounding the First Intervening Sequence in the Complete Human β-Globin Pre-mRNA. We wished to determine if the structures in the first exon, first intervening sequence, and second exon would be altered in a pre-mRNA molecule that also contained the second intervening sequence and the third exon. In vitro transcription was used to synthesize a 2076-nucleotide molecule that contained the entire pre-mRNA transcript. This molecule was subjected to chemical modifications as described above, and oligonucleotide primers 176–193 and 332–348 were chosen to examine the pattern of modification in the regions surrounding the 5′ splice site and 3′ splice site of the first intervening sequence. The pattern of reactive and unreactive nucleotides was indistinguishable from that seen in the short version of the pre-mRNA (results not shown). Therefore, we conclude that the presence of the second intervening sequence and third exon does not influence the structure surrounding the first intervening sequence. Due to the large size of this molecule, it was not feasible to confirm this conclusion by a cross-linking experiment.

Discussion

Chemical probing experiments and psoralen cross-linking experiments were done on human and rabbit pre-mRNA in solution. Taken together, the experiments place strong contraints on possible secondary structure models for these two molecules. A close examination of each of the sets of experimental data reveals that we have not been able to incorporate every detail of them into the structural models. This suggests that each experimental approach individually has its limitations when used on this type of large RNA molecule that does not contain highly stable secondary and tertiary structures such as those found in ribosomal RNA or transfer RNA. However, with these two experimental approaches used together, it is possible to arrive at a secondary structure model for each molecule that contains details about some local structures and also information about the overall arrangement of the molecule. The pattern of chemical modification in a complete human β-globin pre-mRNA molecule was the same as in the truncated molecule reported here; therefore, it is likely that the RNA structure surrounding the first intervening sequence is formed independently from the rest of the molecule. Since all of these experiments were done on purified RNA

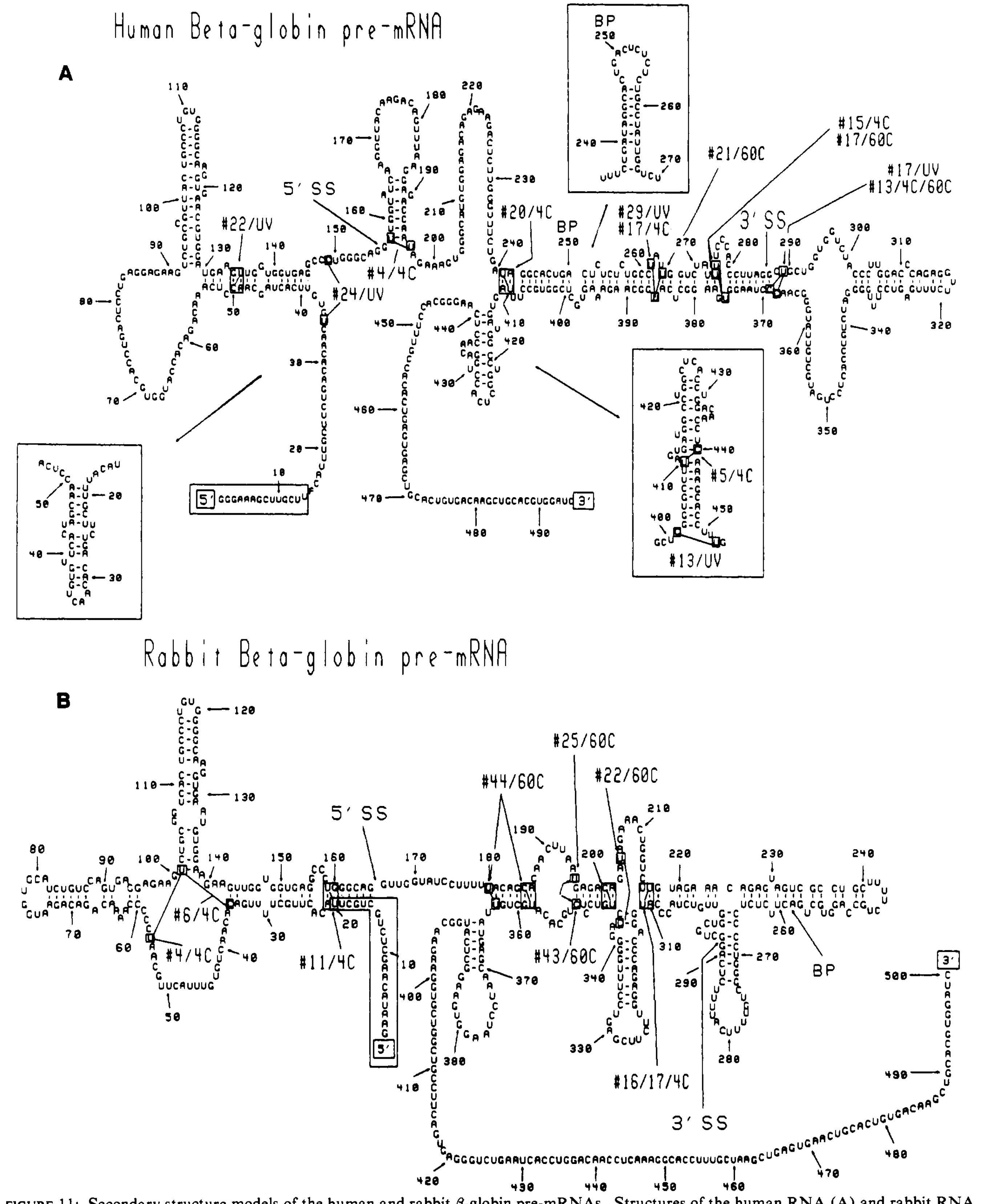

FIGURE 11: Secondary structure models of the human and rabbit β-globin pre-mRNAs. Structures of the human RNA (A) and rabbit RNA (B) consistent with constraints determined by the chemical probing and cross-linking data are shown. Both molecules are numbered from their 5′ termini, including the section of the RNA that results from transcribing vector sequences. The 5′ splice site, branch point, and 3′ splice site are indicated by 5′ SS, BP, and 3′ SS in each molecule. The boxes surrounding the 5′ termini indicate sequences encoded by the vector and nontranscribed 5′ genomic sequences that are not part of the in vivo transcribed RNA. Psoralen and UV cross-linking sites are indicated by connected, boxed nucleotides, the fraction in which they were determined, and the temperature of the cross-linking experiment; UV cross-links are indicated by "UV". For the human RNA, three alternative structures are shown in boxes that are consistent with the structural data. Although the cross-linking data for the rabbit pre-mRNA suggest that several alternative structures exist including contacts between exon 1 and the intron and between the exons, there is poor correlation between possible secondary structures and the chemical probing data. This is also true for several other cross-links in the human pre-mRNA. Thus, other alternative structures could exist for these RNAs, but they are not stable enough to be seen by both the chemical probing and cross-linking data. Regions of the RNAs for which we are not sure of the structures have been left single stranded.

molecules, the models pertain to the RNA molecules in solution without any components that would normally interact with a precursor mRNA in vivo; we have no information yet to indicate the structure these two molecules have in ribonucleoprotein complexes.

The experimental data indicate that both RNA molecules can be modeled into two domains as shown in Figure 11. One domain is located in the first exon; the other encompasses the intron and the 5′ end of exon 2 and is stable at a higher temperature. An important feature of this arrangement is that neigher RNA molecule shows any significant interaction between the 5′ and 3′ splice sites through secondary structures. This indicates that the information required for initial splice site recognition is likely to be contained in the regions surrounding the splice sites and that communication between the 5′ and 3′ splices sites must be accomplished at some other level of organization during spliceosome formation. This is consistent with the observations that partial ribonucleoprotein particles can be assembled on human β-globin pre-mRNAs containing either a 5′ or a 3′ splice site (Bindereif & Green, 1987). In addition, it is also known that the 3′ splice site is recognized by factors (U2 snRNP, U2AF, and others) independent of other factors that bind to the 5′ splice site (Bindereif & Green, 1987; Kramer, 1988; Konarska & Sharp, 1987). In light of these results, it is not surprising that chimeric precursor molecules can be made that have a 5′ splice site from one gene and a 3′ splice site from another and still splice efficiently both in vitro and in vivo (Chu & Sharp, 1981; Chabot et al., 1985; Munroe, 1988).

The most likely candidates for functionally important secondary structures are short-range interactions. Osheim et al. (1985) used electron microscopy to examine the structure of nascent RNA transcripts for *Drosophila melanogaster* chorion proteins containing introns 91 and 228 nucleotides in length. They found that the RNA rapidly became associated with protein components of the nucleus to form hnRNP particles and that the location and behavior of the particles was consistent with their involvement in splicing (Osheim et al., 1985). Since this assembly process occurred before transcription was complete, it is unlikely that the pre-mRNA could form long-range secondary structure. A rapid association of the pre-mRNA with nuclear splicing components is also consistent with the observation that in RNA molecules that have been engineered to contain complementary sequences, these complementarities influence splice site selection only if they are not placed too far away from one another (Eperon et al., 1988).

Several short-range structures are seen in the secondary structure models in Figure 11. Of these, only one structure is very similar between the two molecules; this is the stem/loop structure (nucleotides 92–129 in the human molecule and nucleotides 101–138 in the rabbit molecule) about 25 nucleotides from the 5′ splice site. Both molecules contain significant secondary structure around the 3′ splice site. The human pre-mRNA contains a stable stem/loop structure (nucleotides 302–335 in Figure 11) in this region. This stem/loop structure is at the end of a large interrupted hairpin that includes part of the intron and the 5′ end of exon 2. Two alternate short-range structures are also present in this region of the human RNA. The first (nucleotides 237–267) surrounds A250, the nucleotide that is used as the branch point in the lariat intermediate (Padgett et al., 1984). This same structure has been proposed previously for this section of the RNA by Hall et al. (1988). The second alternate stem/loop structure occurs in exon 2 in nucleotides 402–449. None of these structures are present in the rabbit pre-mRNA. Instead, there are three stem/loop structures surrounding the 3′ splice site that form a branched arrangement. The first (nucleotides 226–265) contains in its duplex region, A258, the rabbit branch point nucleotide (Zeitlin & Efstratiadis, 1984). The second (nucleotides 266–297) includes the 3′ splice site in its duplex region. The third occurs between nucleotides 312 and 344 in the second exon. This third stem/loop structure is similar in size and location but not in sequence to the one found in the human pre-mRNA at nucleotides 302–335. Thus, both molecules have stable secondary structures in this region, but the details of the structures are nearly totally different.

The location of the splice sites within the human β-globin pre-mRNA has been shown to have an effect on splice site usage during in vitro splicing reactions (Reed & Maniatis, 1986; Nelson & Green, 1988). Reed and Maniatis (1986) constructed pre-mRNA that contained a normal splice site and a duplicated test splice site containing different amounts of adjacent exonic sequence. They determined the pattern of splicing to investigate if the exon sequences modulated the use of the adjacent splice site: for both the 5′ splice site and the 3′ splice site, a certain minimum amount of adjacent exon sequence was needed for maximum splice site usage. Nelson and Green (1988) inserted a synthetic 3′ splice site into different regions of a human β-globin pre-mRNA which was deleted to eliminate the normal polypyrimidine tract and AG acceptor site of the first intron. The synthetic site was used to a greater or lesser extent depending upon its position in exon 2 (Nelson & Green, 1988). There is some correspondence between the regions that modulate splice site usage and the location of the secondary structure elements. The stem/loop in nucleotides 91–129 is in the interval identified by Reed and Maniatis as being important for maximal 5′ splice site usage. The sequence from nucleotides 369 to 402, that forms the complementary strand to the region containing the branch point and 3′ splice acceptor site, is in the interval determined to be important for maximal 3′ splice site usage. In the experiments of Nelson and Green (1988), the synthetic 3′ splice site was utilized if it was installed to the 5′ side of the region that contains the stem/loop at nucleotides 302–335; it was inefficiently used if it was installed within or to the 3′ side of this interval. However, these correspondences may be coincidental since we do not know yet whether there is any secondary structure in the pre-mRNA at the time that splicing factors engage it and determine the splicing sites. The structures we have presented do suggest sites in the pre-mRNA that will be the subject for future mutagenesis experiments and will be examined to determine the nature of factor–RNA interactions that occur during spliceosome assembly and in the complete form of the spliceosome.

Phylogenetic conservation is a powerful indicator of functionally important structures (Fox & Woese, 1975). These two β-globin pre-mRNAs were chosen for examination in expectation that common structures, if they were functionally important, would be detected. There are similarities in the human and rabbit β-globin pre-mRNA in regions adjacent to the splice sites; however, except for one stem/loop structure, there is little conservation in the exact structure in these two molecules. Therefore, the specific secondary structures themselves could not be necessary to influence splice site selection, but rather it would have to be some feature of the general distribution of structures in these pre-mRNAs that was important. This would indicate a rather loose requirement for RNA secondary structures that help designate correct splice sites, if in fact the initial steps of splice site selection occur on naked RNA not yet associated with protein or other factors. This may be a general feature of splice junctions bounded by coding exon sequences since it could only be an ancillary function of the genetic code to generate RNA secondary structures that influence the identification of the splice site.

Acknowledgments

We thank Drs. Argiris Efstratiadis and Cynthia Hemenway for critically reading a preliminary version of the manuscript. Dr. Michael Zuker is thanked for providing the program PCFOLD and for discussions of its use.

References

Bindereif, A., & Green, M. (1987) *EMBO J. 6*, 2415–2424.

Chabot, B., Black, D., LeMaster, D., & Steitz, J. (1985) *Science 230*, 1344–1349.

Chu, G., & Sharp, P. A. (1981) *Nature 289*, 378–382.

Davanloo, P., Rosenberg, A. H., Dunn, J., & Studier, F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035–2039.

Eperon, L., Graham, I., Griffiths, A., & Eperon, I. (1988) *Cell 54*, 393–401.

Ericson, G., & Wollenzien, P. L. (1988) *Anal. Biochem. 174*, 215–223.

Fox, G., & Woese, C. R. (1975) *Nature 256*, 505–507.

Frendewey, D., & Keller, W. (1985) Cell *42*, 355–367.

Furdon, P. J., & Kole, R. (1988) *Mol. Cell. Biol. 8*, 860–866.

Garrett-Wheeler, E., Lockard, R., & Kumar, A. (1984) *Nucleic Acids Res. 12*, 3405–3423.

Grabowski, P., Seiler, S., & Sharp, P. (1985) *Cell 42*, 345–353.

Hagenbuchle, O., Santer, M., Steitz, J., & Mans, R. (1978) *Cell 13*, 551–563.

Hall, K. B., Green, M. R., & Redfield, A. G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 704–708.

Inoue, I., & Cech, T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 648–652.

Khoury, G., Gruss, P., Dhar, R., & Lai, C. (1979) *Cell 18*, 85–92.

Konarska, M., & Sharp, P. (1987) *Cell 49*, 763–774.

Krainer, A. R., Maniatis, T., Ruskin, B., & Green, M. (1984) *Cell 36*, 993–1005.

Kramer, A. (1987) *J. Mol. Biol. 196*, 559–573.

Kramer, A. (1988) *Genes Dev. 2*, 1155–1167.

Lockard, R. E., & Kumar, A. (1981) *Nucleic Acids Res. 9*, 5125–5140.

Moazed, D., Stern, S., & Noller, H. (1986) *J. Mol. Biol. 187*, 399–416.

Munroe, S. (1988) *EMBO J. 7*, 2523–2532.

Nelson, K., & Green, M. (1988) *Genes Dev. 2*, 319–329.

Noller, H., Stern, S., Moazed, D., Powers, T., Svensson, P., & Changchien, L.-M. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 695–708.

Osheim, Y. N., Miller, O. L., Jr., & Beyer, A. L. (1985) *Cell 43*, 143–151.

Padgett, R. A., Knoarska, M. A., Grabowski, P. J., Hardy, S. F., & Sharp, P. A. (1984) *Science 225*, 898–903.

Parent, A., Zeitlin, S., & Efstratiadis, A. (1987) *J. Biol. Chem. 262*, 11284–11291.

Reed, R., & Maniatis, T. (1986) *Cell 46*, 681–690.

Romby, P., Moras, D., Dumas, P., Ebel, J. P., & Giege, R. (1987) *J. Mol. Biol. 195*, 193–204.

Ruskin, B., Zamore, P., & Green, M. (1988) *Cell 52*, 207–219.

Sharp, P. (1987) *Science 235*, 766–771.

Solnick, D., & Lee, S. (1987) *Mol. Cell. Biol. 7*, 3194–3198.

Somasekhar, M. B., & Mertz, J. E. (1985) *Nucleic Acids Res. 13*, 5591–5609.

Swanson, M. S., & Dreyfus, G. (1988) *EMBO J. 7*, 3519–3529.

Thompson, J. F., & Hearst, J. E. (1983) *Cell 32*, 1355–1365.

Wollenzien, P. (1988) *Methods Enzymol. 164*, 319–329.

Wollenzien, P., Goswami, P., Teare, J., Szeberenyi, J., & Goldenberg, C. (1987) *Nucleic Acids Res. 15*, 9279–9297.

Youvan, D., & Hearst, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3751–3754.

Zeitlin, S., & Efstratiadis, A. (1984) *Cell 39*, 589–602.

Zuker, M., & Sankoff, D. (1986) *Bull. Math. Biol. 46*, 591–621.