Analysis of spliceosome assembly and the structure of a regulated intron in *Drosophila in vitro* splicing extracts

Deborah Spikes and Paul M. Bingham*

Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, NY 11794, USA

Received July 14, 1992; Revised and Accepted October 6, 1992

ABSTRACT

We characterize spliceosome assembly in *Drosophila* embryonic nuclear extracts. Further, we show that these extracts contain high levels of a 5' to 3' exoribonuclease activity allowing rapid, convenient protection mapping of 5' splice site and branchpoint sequences. We use this assay to show, for the first time, that a regulated arthropod intron uses a remote branchpoint strikingly similar in structure to those observed previously in regulated vertebrate introns. These results provide new evidence that both regulated and constitutive splicing are similar in detail in vertebrates and arthropods indicating that the powerful genetic systems for analysis of splicing regulation in *Drosophila* are likely to be directly informative for regulated splicing throughout metazoa. In addition, we report formation of a novel class of intron-dependent complexes. Behavior of these complexes indicates that they represent a mutually exclusive, kinetically competing pathway with spliceosome assembly. We propose that this competition represents the basis for a kinetic proofreading mechanism enhancing fidelity of intron recognition. We also discuss possible implications of this model for regulated splicing.

INTRODUCTION

All recognized patterns of regulated metazoan pre-mRNA processing occur in *Drosophila* including regulated exon skipping, alternative exon inclusion, alternative splicing/polyadenylation and regulated blockage of splicing (reviewed in 1–3). Moreover, *Drosophila* is a well developed genetic system in which a number of regulated alternative RNA processing events are amenable to genetic manipulation (4–9). Thus, *Drosophila* is uniquely powerful for analysis, in particular, of regulated splicing. To the extent that arthropods and vertebrates carry out regulated (and constitutive) splicing similarly, *Drosophila* is likely to be one of the definitive experimental systems for study of most aspects of metazoan pre-mRNA processing.

Several earlier observations (for example, 10) suggest that splicing in arthropods and vertebrates is similar; however, the extent of similarity in detail remains to be determined. For example—in spite of the extensive development of genetic systems

for analysis of splicing in *Drosophila*—*in vitro* splicing systems from *Drosophila* permitting assessment of arthropod/vertebrate similarities and permitting biochemical analysis of splicing are still very incompletely defined. Rio (11) reported that *Drosophila* extracts will splice the *Drosophila fushi tarazu (ftz)* intron. However, detailed procedures for extract preparation from *Drosophila* embryos (whose genotype can be controlled) and detailed characterization of behavior of *Drosophila* extracts have yet to be described.

We describe here detailed recipes for production of *in vitro* splicing extracts from *Drosophila* embryos. These procedures use the small amounts of embryos normally produced when genotype is to be controlled and when specialized growth facilities are not available. We use these embryonic extracts to carry out several lines of experimentation generating new evidence that arthropod and vertebrate splicing—both constitutive and regulated—proceed similarly. Further, early steps in intron (exon) recognition in metazoans are still incompletely understood. We report the existence of a novel class of intron-containing complexes. Properties of these complexes indicate that their formation competes with spliceosome assembly suggesting that these new complexes may represent the basis of a kinetic proofreading strategy for intron recognition.

MATERIALS AND METHODS

Collection of embryos

For small scale collections (0.5 to 2.0 gms) embryos are collected on nylon mesh screens stretched over a mound of instant *Drosophila* food (Carolina Biologicals) to which several drops of yeast suspension is added. Food mounds are placed in 100mm square petri plates. Adults (ca. 500) for such collections are placed in 32oz. jars inverted over the food mounds. Collections were carried out at 25°C for 12 hours unless otherwise indicated. Under these conditions a yield of 0.1–0.2 grams of embryos per jar per collection is anticipated. Embryos are eluted from nylon mesh in cold saline (0.7% w/v NaCl), collected in a fluted nylon mesh funnel after filtration through a coarse nylon screen to remove adult body parts and washed with cold saline.

Large scale embryo collections were made from population cages essentially as previously described (12).

Embryos are dechorionated by treatment in 50% bleach, 0.2%

* To whom correspondence should be addressed

triton in saline at room temperature for 90 seconds. Embryos are then collected again on a nylon mesh funnel and washed extensively with cold saline. Embryos are either used immediately for extract preparation or are aliquoted in plastic tubes, flash frozen in liquid nitrogen and stored at -80°C .

Levels of 5' to 3' exoribonuclease activity (RESULTS) are diminished (possibly substantially) in extracts from frozen embryos. All experiments described in which this activity is utilized employed extracts from fresh embryos. Extracts from frozen embryos were used for all other studies described.

Splicing extract preparation

This procedure is adapted from earlier recipes (13,14) and all operations are done on ice. Embryos are ground in a motor-driven teflon homogenizer followed by the loose ball of a dounce homogenizer in buffer A at 0.2 to 1 gm of embryos per 10 mls of buffer A. (Buffer A is 15mM HEPES, pH 7.6, 10mM KCl, 5mM MgCl_2 , 0.1mM EDTA, 0.5mM EGTA, 1mM DTT, 1mM PMSF, 1 microgram/ml leupeptin, 350mM sucrose.) Homogenate was layered over a cushion of buffer B. (Buffer B is buffer A with 800mM sucrose.) The homogenate is spun for 5 minutes at 10krpm in a Sorval HB-4 swinging bucket rotor to pellet nuclei and the supernatant is aspirated off to minimize contamination of the pellet.

The pellet is resuspended with vigorous vortexing in buffer C (0.4mls per gram of embryo starting material). (Buffer C is 20mM HEPES, pH7.6, 420mM NaCl, 1.5 mM MgCl_2 , 0.2mM EDTA, 0.5 mM DTT, 25% v/v glycerol, 1mM PMSF, 1 microgram/ml leupeptin.) The resuspended pellet is transferred to a tightly capped plastic tube and rolled for 30 minutes in the cold. The nuclei are pelleted for 10 minutes at top speed in a microcentrifuge and the supernatant is dialyzed against 3 changes of D buffer (each change 100 volumes for 1 hour with agitation). (D buffer is 20mM HEPES, pH7.6, 100mM KCl, 0.2mM EDTA, 0.5 mM DTT, 20% v/v glycerol, 1mM PMSF, 1 microgram/ml leupeptin.)

Dialyzed extract is cleared by spinning for ten minutes at top microcentrifuge speed. The cleared supernatant is aliquoted, flash frozen in liquid nitrogen and stored at -80°C . Extracts so stored retain substantial activity for at least one year. Extracts typically have total protein concentrations of ca. 15 mg/ml.

In vitro splicing conditions

Splicing reactions were carried out in appropriate multiples of 5 microliters on a room temperature water bath. 5 microliter reactions were made by addition in the following order. RNA in 2.5 microliters of water, 0.5 microliters of a $10\times$ buffer and 2.0 microliters of nuclear extract in D buffer (above).

Our standard $10\times$ buffer is 200mM HEPES (pH 7.6), 25mM MgCl_2 , 20mM ATP and 200 mM creatine phosphate. RNA inputs were typically ca. 10^6 dpm of ^{32}P labeled RNA for radiolabeled experiments or ca. 2ng of cold RNA for spliceosome assembly experiments.

Polyvinyl alcohol as a volume exclusion reagent produces only modest, sporadic enhancement of rates and extents of reaction in our hands and we do not routinely include it.

For analysis of spliced products reactions were terminated as described (15). For analysis on native gels samples were terminated by rapid addition of 1 microliter of 50mM EDTA, 10 mg/ml heparin followed by 1 microliter of glycerol containing xylene cyanol and bromophenol blue. Samples were immediately flash frozen and held in liquid nitrogen for loading.

5' to 3' exoribonuclease digestion

The exoribonuclease in *Drosophila* apparently requires ATP, but is very substantially inhibited by high Mg^{++} concentrations (RESULTS). We thus carry out exonuclease digestions under the following conditions at 22°C : 28mM HEPES, pH7.6, 2.0–2.5mM ATP, 5mM creatine phosphate and 0.1mM MgCl_2 . Under these conditions spliceosome assembly proceeds with similar kinetics to splicing conditions (above) but little or no spliced mRNA and intron lariat accumulate (results not shown).

While the one step procedure described in the preceding paragraph produces generally good results (Figure 3), the following two step procedure was used in the experiments in Figure 4 to maximize spliceosome assembly before exonuclease digestion. RNAs were first incubated for 15 minutes using extracts made as described above except that 5 minute incubation on ice was included between the teflon and dounce homogenization steps and nuclei were initially pelleted for 10 minutes at 5krpm. These conditions produce extracts with lowered exoribonuclease activity but with good spliceosome assembly capability. After this preincubation, samples were diluted fourfold into reactions made with standard extracts (above) and incubated for 45 minutes.

Synthesis of RNA templates

Template RNAs were synthesized as described by Krieg and Melton (16) and capped as described by Hope and Struhl (17). RNAs were labeled with ^{32}P -UTP incorporation in reactions supplemented with 10 micromolar cold UTP to prevent excessive premature termination.

Templates for synthesis of the complete *ftz* intron were derived by linearizing the Sall to EcoRI fragment spanning the *ftz* intron (18) introduced into the SP18 vector. Unless otherwise specified *ftz* template was linearized at the XhoI site in the second *ftz* exon. Template for 3' half *ftz* intron was made by linearizing the AccI to EcoRI *ftz* fragment in SP18 at the *ftz* XhoI site. The template for synthesis of the first *su(w^a)* intron was made by linearizing at the 3' end of the following composite fragment (descriptions in terms of standard coordinate system; 19): *su(w^a)* intervals from 770 to 847 (5' splice site is after base 823) and from 1468 to 1808 (3' splice site is after base 1679) were joined through a 20 base linker to produce an intron sufficiently small for these studies but possessing all sequences necessary to permit splicing *in vivo* (Mims and Bingham, unpublished observations). This segment was introduced into the *Sma*I through *Kpn*I interval of SP18 vector and linearized at the *Kpn*I site at coordinate 1808 for *in vitro* transcription. The *su(w^a)* cDNA segment used in Figure 7 consists of cDNA clone of the fully spliced *su(w^a)* mRNA extending from coordinates 523–823 (first exon) and 1679–1808 (second exon) cloned in the SP6 vector pSPBP4. Linearization with *Kpn*I allows production of a 516 base RNA transcript (including 52 bases of vector sequence) containing exon sequence surrounding the first *su(w^a)* intron.

Native gel electrophoresis

Native gel electrophoresis was carried out as described by Zillman, et al. (20). Gels were run in the cold and prerun for at least 30 minutes before loading. Samples were briefly thawed and applied to the gel. Because of the instability of the micrococcal nuclease-dependent complexes in the presence of heparin and EDTA, samples in the experiments described in

Figure 7 were individually thawed for 10 seconds on a room temperature water bath and applied to the gel under power so that electrophoretic separation began immediately upon sample thawing. Electroblothing of native gels was carried out according to Konarska and Sharp (21).

Denaturing gel electrophoresis and gel extraction

RNA samples were run on conventional urea sequencing gels (22). A sequencing ladder using the -20 primer (GTAAAC-GACGGCCAGTGCCA) on a standard M13mp18 DNA template (USBiochemicals) was used as a molecular weight standard. Sequencing channels were loaded in the order A,T,G,C.

Samples were extracted from native gels for subsequent running on denaturing gels by conventional procedures (23).

Northern transfers (Figure 6) were done from formaldehyde/agarose gels according to conventional procedures (22).

Construction of snRNA probes

The *Drosophila* U5 gene cloned here (below) was introduced into the *EcoRV* site of pBluescriptKS(+); probe was generated by T7 polymerase transcription of *XhoI* linearized template. U1 probe segment is clone pL1 provided by Patrick Lo and Steven Mount and consists of U1 nucleotides 50 to 140 cloned into the pIB124 vector such that probes were synthesized by T7 polymerase transcription of *HindIII* linearized template (24). A 192 base segment consisting of U2 was generated by PCR amplification from plasmid clone 131A (25) and cloned into the *EcoRV* site of the pBluescriptKS(+) vector; probe RNA was generated by T7 polymerase transcription of *XhoI* linearized template. The 115 base U4-internal *AvaI-EcoRV* fragment from clone U4-1 (26) was introduced into the *HincII* site of pBluescriptKS(+); U4 probe was generated by T7 polymerase transcription of *XhoI* linearized template. A 236 base *EcoRI-Sau3A1* fragment containing the entire U6 gene from clone pDU6-2 (26) was introduced into *BamHI-EcoRI* cut pBluescriptKS(+); probe was generated by T3 transcription of *XbaI* linearized template.

Each member of this probe set hybridizes to a single major *Drosophila* RNA species of the expected size when filter transfers of denaturing gels are probed (Figure 6 and results not shown). Thus, a complete set of probes for the major, splicing-associate snRNAs from *Drosophila* is now available.

Cloning of the *Drosophila* U5 gene

A *Drosophila* U5 RNA coding sequence was retrieved by PCR amplification based on the published *Drosophila* U5 RNA sequence (27). Primers synonymous to the 5' end of U5 RNA (ATACTCTGGTTTCTCTTCAATGTCTG) and homologous to the U5 3' end with a cleavage site extension (TTGAATTCTG-GCCCCAGTCTTGCGACAAGAC) were used in a conventional PCR amplification from a *Drosophila* total genomic DNA template.

The U5 homologous sequence obtained was TACTCTGGT-TTCTCTTCAATGTCTGAATAAATCTTTCGCCTTTTACT-AAAGATTTCCGTGGAGAGGAACACTCTAATGAGTCT-AAAATAATTTTGTAGTAGTGTCTTGTGCGCAAGACTGG-GGCCA. The first base of the U5 PCR product was lost in cloning. Further, this sequence differs at a few positions from the published RNA sequence (reviewed in 28). Similar differences between RNA sequences and various individual cloned snRNA genes have been observed previously in the case of *Drosophila* U4 (29).

Micrococcal nuclease digestion

Extracts were made 0.4mM CaCl_2 , 1mM MgCl_2 , 0.5mM ATP and 5mM creatine phosphate. Micrococcal nuclease (MN) was added and digestion was carried out for 10 minutes at room temperature. EGTA was added to 1mM and room temperature incubation was continued for 10 additional minutes. Extracts were held on ice until use (within 1–2 hours). Several attempts to improve the efficiency of MN digestion by preincubation (in the presence of ATP as above) revealed no obvious effect (results not shown); however, this has not been extensively investigated.

RESULTS

Conditions for *in vitro* splicing in *Drosophila* nuclear extracts

Preparation of nuclear extracts from *Drosophila* embryos is described in MATERIALS AND METHODS. The intron from the *Drosophila* *ftz* gene has been shown previously to splice efficiently *in vitro* and the products of this splicing event are well defined (10,11). We therefore used this intron to characterize *Drosophila* embryo extracts.

Reactions were carried out at room temperature (ca. 22°C) which is near optimal for *Drosophila* growth. We investigated the effects of varying Mg^{2+} and K^+ and find optima under our conditions (ATP at 2.0 mM) at around 2.5mM and 40mM, respectively (results not shown). ATP depletion during the splicing reaction appears to be significant in *Drosophila* extracts—addition of creatine phosphate presumably permitting endogenous ATP regeneration is necessary for reaction rates to remain high for long times (Figure 1 and results not shown).

In summary, our standard conditions are 40% (v/v) nuclear extract (producing 40mM KCl); 28mM HEPES, pH 7.6; 2.5mM MgCl_2 ; 2.0mM ATP; 20mM creatine phosphate, 0.08mM

Figure 1. Splicing of the *ftz* intron in *Drosophila* embryo splicing extracts. *Left hand panel:* Reactions were carried out for times indicated in minutes at the top of each channel as described in the text. Deproteinized RNA products were run on a denaturing gel. The positions of the input RNA, spliced mRNA, first exon, lariat-second exon and excised lariat are indicated diagrammatically top to bottom. *Center panel:* Spliceosome assembly in *Drosophila* embryo extracts. Reactions were carried out for the times indicated in minutes above each channel. The lefthand reaction series contains the 3' half of the *ftz* intron and the righthand set contains the intact *ftz* intron. Reactions were fractionated on a native gel (MATERIALS AND METHODS). The positions of the A, C and H complexes discussed in the text are indicated. *Righthand panel:* Products of *ftz* splicing comigrate with the C complex. Radiolabeled *ftz* intron-containing RNA was incubated in a standard splicing reaction for 60 minutes. The complexes formed were fractionated on a native gel and the A, C and H complex regions of the gel (based on autoradiography) were excised. RNA extracted from these three gel segments was fractionated on a denaturing gel in the autoradiograph shown. The positions of the input RNA, spliced mRNA and excised intron lariat are indicated graphically.

Figure 2. snRNA components of spliceosome assembly intermediates in *Drosophila*. Splicing reactions were carried out for the times indicated in minutes above each channel. For the set of ten reactions shown the lefthand five contained approximately half saturating quantities of nonradioactive *ftz* intron-containing RNA (MATERIALS AND METHODS) and the righthand five contained no exogenous RNA. Reactions were fractionated on a native gel and electroblotted onto a membrane for hybridization (MATERIALS AND METHODS). The resulting filter was probed sequentially with the probe indicated below each panel. Superposition of resulting autoradiographs allows extremely precise determination of relationships of species revealed by different probes. Locations of the A, C and H complexes discussed in the text are indicated. U5 and U6 label C complex but not A complex. U2 labels both A and C complex. [Note that two major mobility classes of U2 are seen in extracts without input pre-mRNA—one migrating between A and C complexes and one migrating more rapidly than A and C. Similar U2 labeling patterns are seen in mammalian extracts (21). It is unclear which of these U2 species contributes to A and C complex formation.]

EDTA, 8% (v/v) glycerol. *ftz* RNA inputs of the order of 5 nanograms per 5 microliter standard reaction are half saturating for most extract preparations. At subsaturating inputs products of the first cleavage reaction—intron lariat-second exon (IE2) and first exon (E1)—reach steady state within 10 to 20 minutes (Figure 1). Products of the completed splicing process continue to accumulate for one to two hours (Figure 1).

Characterization of spliceosome assembly in *Drosophila*

We have examined spliceosome assembly in *Drosophila* by analyzing the complexes formed during *in vitro* splicing by native gel electrophoresis (MATERIALS AND METHODS). The gel system we use resolves three major complexes that form on exogenous substrate RNA in *Drosophila* splicing extracts. These complexes appear to correspond to well characterized complexes in mammalian splicing extracts (below).

We refer to the most rapidly moving of these three complexes as H. H complex forms instantaneously upon addition of RNA (Figures 1 and 2). This complex apparently results from rapid association of extract proteins with input RNA. We refer to the two more slowly migrating complexes as A and C (Figures 1 and 2) by analogy with nomenclature used to describe apparently similar vertebrate complexes (see, for example, 21 and 30). The *Drosophila* complex we designate A apparently corresponds to the vertebrate A complex based on the following observations (Figures 1 and 2; results not shown). First, the complex forms rapidly in *in vitro* splicing extracts. Second, a very similar complex is formed both by the intact *ftz* intron and by the 3' half of the *ftz* intron (branchpoint and 3' splice site). Third, this complex contains U2 snRNA as assessed by hybridization, but does not contain U1, U4, U5 or U6 by this assay. Lastly, while A complex contains input RNA it contains no significant amount of spliced products.

We believe that the *Drosophila* C complex corresponds to the

vertebrate C complex on the basis of the following observations (Figures 1 and 2; results not shown). First, C complex assembles more slowly and persists longer than A complex. This indicates a precursor-product relationship between A and C complex. C complex accumulates with kinetics similar to those of accumulation of splicing products. Second, efficient C complex formation apparently requires a 5' splice site in addition to a branchpoint/3' splice site as the 3' half of the *ftz* intron does not efficiently form C complex. Third, C complex contains relatively high levels of excised intron as assessed analyzing RNAs extracted from the C complex region of the gel. [We do not observe high levels of the mature mRNA splice product in the A, C or H complex region of the native gel profile (Figure 1). Thus, mature mRNA may move elsewhere or be heterogeneously distributed in the gel under our conditions.] Lastly, C complex contains U2, U5 and U6 RNAs as assessed by hybridization.

Mammalian spliceosome assembly produces relatively high levels of a complex thought to represent a relatively mature spliceosome that has not completed covalent modification of the substrate. This complex contains U2, U4/6 and U5 and is designated B. We do not detect an obvious species corresponding to B complex here. This could result from comigration of a B-like complex with C complex. Alternatively, the rate of maturation of B to C might be sufficiently high under our conditions that steady state B levels are too low to detect. The failure of U4 probes to detect slowly migrating forms under our conditions (discussed below) precludes our distinguishing between these two possibilities at this point in the analysis.

U5 labels two closely spaced species moving roughly 30% more rapidly than C complex in this gel system (Figure 2). A species comigrating with at least one of these is labeled by U6 (visible on long exposures of the experiment in Figure 2; results not shown). At least one of these complexes is a good candidate for the U4/U5/U6 particle seen in vertebrate systems.

Figure 3. *Drosophila* contains a 5' to 3' exoribonuclease. *Left hand panel:* Capped input RNA is substantially resistant to exoribonuclease degradation. The lefthand three channels are uncapped *ftz* RNA and the righthand three are capped RNA. 0 indicates undegraded input RNA and + indicates one hour treatment under degradation conditions. The uncapped RNA is almost entirely degraded to produce high levels of the 5' splice site and branchpoint exoribonuclease strong stops (positions of stops graphically indicated on diagram of *ftz* RNA). Note that the middle channel of the capped RNA set is underloaded. See text for further details. *Middle panel:* ATP apparently stimulates exoribonuclease degradation. Reactions contained extract, KCl, Pipes and *ftz* RNA at standard inputs (MATERIALS AND METHODS). $MgCl_2$ was added to 0.1mM and ATP at the concentrations indicated above each channel. Reactions were incubated for one hour at 22°C and the products analyzed on a denaturing gel. *Righthand panel:* High $MgCl_2$ concentrations apparently inhibit exoribonuclease degradation. Reactions contained extract, KCl, Pipes and *ftz* RNA at standard inputs (MATERIALS AND METHODS). ATP was added at 0.5mM and $MgCl_2$ at the concentrations indicated above each channel. Reactions were carried out for one hour at 22°C and the products analyzed on denaturing gels.

In summary, our results indicate that spliceosome assembly in *Drosophila* appears to be similar to that observed in other experimental systems and particularly similar to the corresponding processes in vertebrates.

[We consistently fail to detect U4 in any slowly migrating complexes by hybridization to native gel transfers (results not shown). This apparently results from inaccessibility of U4 to hybridization under our native gel conditions as U4 RNA is readily detected on denaturing gels (above). U6 is also apparently not fully efficiently detected by hybridization to native gel transfers in some of its forms as the amount of U6 detected by hybridization in C complex appears to exceed the total amount of U6 in the absence of input RNA (Figure 2). U2, U4, U5 and (to a lesser degree) U6 label sharp, rapidly moving bands on native gels that do not have obvious cognates in vertebrate extracts of which we are aware (Figure 2 and results not shown). In the cases we have tested (U4 and U5) these novel species are RNase sensitive. These species might represent maternally deposited forms of the corresponding snRNAs that have not yet become associated with one or more snRNP proteins. If so, *Drosophila*

Figure 4. Exoribonuclease mapping of the *su(w^u)* first intron branchpoint. At left are shown the results of three independent experiments mapping the *su(w^u)* branchpoint. 0 indicates undegraded RNA and + indicates exoribonuclease degraded samples (MATERIALS AND METHODS). The rightmost *su(w^u)* experiment shows appended molecular weight standard sequencing ladder (MATERIALS AND METHODS). (This image is a composite of two extents of exposure to allow visualization of all relevant features.) Note that a shortened version of the first *su(w^u)* intron is used in these experiments for technical convenience (MATERIALS AND METHODS). At right is an experiments mapping the 5' splice site and branchpoint strong stops on the *ftz* intron with associated molecular weight standard sequencing ladder (MATERIALS AND METHODS). (This image is a composite of two extents of exposure to allow visualization of all relevant features.) The positions of the 5' splice site and branchpoint strong stops on the *su(w^u)* and *ftz* RNAs are indicated graphically. The expected positions of fragments extending from the 3' end of the template through the 5'ss and BP are indicated on the molecular weight standard ladders. Note that the corresponding strong stop species migrate slightly more slowly than these positions indicating stops ca. 10 bases 5' to the corresponding splicing signal (also see Figure 5).

embryos might represent a useful opportunity to study snRNP assembly.]

***Drosophila* contains a 5' to 3' exoribonuclease allowing convenient protection mapping of splicing signals on input RNA**

We noted early in our analysis that specific products of input RNA in addition to the expected splicing products accumulated under some conditions (Figures 3 and 4 and results not shown). Various details of the size of these products suggested that they represented fragments 3' coterminal with the input RNA and thus, likely, represented products of 5' to 3' exonuclease digestion. On this hypothesis the discrete species observed arise when exonuclease digestion is stopped or slowed at specific points on the input RNA molecule. We will refer to these sites of exonuclease stopping or slowing as exonuclease 'strong stops'.

We confirmed this hypothesis in two ways (results not shown). First, *ftz* intron-containing fragments with extended 3' ends generate exonuclease digestion products of correspondingly larger size. Second, when the input *ftz* intron-containing molecule was

Figure 5. Results of exoribonuclease mapping of *ftz* and first *su(w²)* intron splice sites. *su(w²)* is shown at the top and *ftz* at bottom. The sequences surrounding the 5' splice site are diagrammed above the respective structure diagram and the sequences extending from the branchpoint to the 3' splice site below. The positions of the indicated branchpoint sequences are boxed. The approximate positions of the exoribonuclease strong stops are indicated by the short vertical lines appended to each sequence.

3' end-labeled rather than homogeneously labeled the same exonuclease digestion products were observed.

That the activity generating these products is an exonuclease is further corroborated by the relative sensitivity of capped and uncapped input RNAs (Figure 3). The capping procedure used (MATERIALS AND METHODS) generates a population of molecules in which a majority are capped and a significant minority are not. When this material is used as input exonuclease digestion products accumulate to substantially lower levels than when uncapped RNA is used.

Thus, this exoribonuclease has the properties expected of an enzyme dedicated to degradation of the byproducts of pre-mRNA processing. Specifically, this activity should rapidly degrade debranched introns while not degrading capped pre-mRNAs.

This exoribonuclease activity appears not to require loosely bound divalent cations. Extensively dialyzed extracts give high activity that is not stimulated by addition of $MgCl_2$. Indeed, $MgCl_2$ concentrations above 0.5mM consistently inhibit degradation and formation of strong stop species under conditions tested (Figure 3). Note that the use of capped input RNA and $MgCl_2$ concentrations of 2.5mM for *in vitro* splicing (above) results in almost complete suppression of formation of exoribonuclease degradation products.

Generation of exonuclease protection products requires ATP (Figure 3). Increasing ATP concentrations from 0.5mM through 2mM increases degradation of uncapped input RNA and, to some extent, generation of strong stop species suggesting that ATP may be directly necessary for nuclease activity. We emphasize that these experiments in crude extracts cannot establish this unambiguously. ATP may be required indirectly (including in spliceosome assembly, for example; see below) rather than

directly. [At 4mM ATP degradation and formation of strong stop species are reduced relative to 2mM ATP. We note that the ATP is added as a sodium salt and secondary effects of high ATP (such as elevated sodium ion concentration) might account for this observation.]

Measurement of the size of the exonuclease digestion products permits the precise mapping of the sites of strong stops. In RNAs containing the complete *ftz* intron the stops map ca. 10 bases 5' to the 5' splice site and branchpoint, respectively (Figures 4 and 5). In RNAs containing the 3' half of the *ftz* intron the stop maps ca 10 bases 5' to the branchpoint (results not shown). Thus, both classes of *ftz* intron variants that form A-like complexes (above) generate strong stop species. Moreover, complexes comigrating with—presumably identical or similar to—A and C form efficiently under the conditions for exonuclease degradation (results not shown).

Low salt nuclear extracts of mammalian nuclei have *all* components necessary for splicing except for a single, limiting protein necessary to initiate spliceosome assembly (see 31 and references therein). Corresponding extracts from *Drosophila* embryos specifically degrade uncapped RNA similarly to high salt nuclear extracts but do not produce strong stop species (results not shown). This observation strongly suggests that spliceosome assembly is required for strong stop production.

A 5' to 3' exoribonuclease has been observed in vertebrate extracts (32–34). In these cases as well, the degradation products corresponding to the strong stops described above mapped a few bases 5' to the 5' splice and branchpoint, respectively.

In summary, we interpret these results to indicate that *Drosophila* nuclear splicing extracts contain a 5' to 3' exoribonuclease whose degradation of input RNA is slowed or

Figure 6. Degradation of endogenous U1 and U2 snRNAs by the RNaseH activity in *Drosophila* splicing extracts. A short deoxyoligonucleotide homologous to the indicated portion of the U1 or U2 RNAs was introduced and the reaction was incubated under splicing conditions for 30 minutes. Deproteinized RNA products were run on a denaturing gel, transferred to a membrane and probed with the corresponding antisense RNA. The positions of the intact RNAs are indicated by U1 or U2 and the positions of the degradation products are indicated by asterisks. The oligos homologous to the 5' ends of the respective snRNAs cause the slight reduction in size expected from 'trimming' the terminal homologous region. The oligo complementary to the branchpoint-homologous region of U2 directs cleavage of the RNA into two roughly equal size fragments as expected.

stopped at sites on pre-mRNAs at which components of the splicing machinery have assembled. It follows that this exonuclease activity can be used to map 5' splice sites and branchpoints in other introns.

Evidence that a regulated *Drosophila* intron contains a remote branchpoint

Splicing of the first two introns of the *Drosophila* *su(w^a)* gene are blocked by the *su(w^a)* protein to achieve autoregulation of the gene (5,19). Chou, et al. (19) suggested on the basis of the sequence of these regulated introns that they might use remote branchpoints. Remote branchpoints were subsequently observed in several regulated vertebrate introns (see, for example, 34 and 35).

The first two *su(w^a)* introns splice poorly *in vitro* (results not shown). As a result, conventional strategies for branchpoint mapping dependent on *in vitro* formation of intron lariats are not effective. We therefore capitalized on the properties of the 5' to 3' exonuclease in *Drosophila* extracts to map splicing signals in the first *su(w^a)* intron. We observe two strong stops. One maps ca. 10 bases 5' to the 5' splice site further corroborating the reliability of the assay.

However, the second strong stop (presumptive branchpoint strong stop) maps over one hundred bases 5' to the 3' splice site. This indicates either that this strong stop has some other origin or that the branchpoint in the first *su(w^a)* intron is unusually remote from the 3' splice site. Inspection of the sequence of the first *su(w^a)* intron indicates that the latter interpretation is correct (Figure 5). Specifically, ca. 10 bases 3' to the presumptive branchpoint strong stop is an excellent match to the metazoan branchpoint consensus, CACTGAC. This is followed immediately by a 48 base pyrimidine-rich segment—40/48 bases are pyrimidine. These structural features are quite unusual and

Figure 7. Ladder complex formation in micrococcal nuclease treated *Drosophila* splicing extracts. *Left hand panel:* Reactions were carried out under standard splicing conditions for the times indicated in minutes above each channel and fractionated on a native gel. The channel labeled 10+10 indicates a ten minute incubation, followed by heparin-EDTA termination (MATERIALS AND METHODS), followed by incubation for 10 minutes at room temperature before flash freezing and holding for subsequent loading. All other samples were rapidly terminated and frozen immediately as described (MATERIALS AND METHODS). The RNAs used in each experiment are indicated (MATERIALS AND METHODS). Note that the 3' half and intact *ftz* introns and the first *su(w^a)* intron provoke ladder complex formation efficiently. On the other hand the cDNA clone including the first two *su(w^a)* exons does not load efficiently into ladder complexes. *Right hand panel:* Each sample consisted of equal volumes of standard splicing reactions made with MN treated extract and untreated extract. The MN treated extract reaction was preincubated for ten minutes, mixed with untreated extract reactions and time points taken at the times indicated in minutes above each gel channel. For the samples labeled 'MN Preload' *ftz* RNA was added to the MN treated extract reactions at the beginning of the ten minute preincubation. For the samples labeled 'No Preload' *ftz* RNA was added immediately after mixing of preincubated, MN treated and untreated extract reactions. Positions of the A and C complexes are indicated. Note little or no ladder formation in the No Preload samples and little or no A and C complex formation in the MN Preload samples.

are strikingly similar to the remote branchpoints so far mapped in vertebrate systems (see, for example, 34 and 35). Lastly, the AG dinucleotide does not occur on the 94 base interval between the presumptive branchpoint and the 3' splice site as expected if this branchpoint is authentic.

Collectively, these results indicate that the branchpoint in the regulated first *su(w^a)* intron is quite remote and has the highly unusual and presumably diagnostic structure seen previously in remote branchpoints of several regulated vertebrate introns. This is the first report to our knowledge of a presumptive remote branchpoint in an arthropod. These observations indicate that regulated introns in *Drosophila* use remote branchpoints similar or identical to those apparently common among regulated vertebrate splices.

Drosophila extracts contain high levels of a ribonuclease H activity

Several experiments demonstrate that *Drosophila* embryo extracts contain very high levels of an RNaseH activity. First, we observe highly specific cleavage of snRNAs by addition of complementary oligonucleotides (Figure 6). This cleavage occurs in each case at the predicted position in the snRNA. Second, an oligonucleotide directed against the *ftz* RNA directs cleavage of

that RNA on the expected (complementary) interval (results not shown). The rate of *ftz* cleavage under these conditions is extremely high, being essentially complete at the shortest times tested (60 seconds; results not shown).

Similar high endogenous RNaseH activities are observed in vertebrate splicing extracts and have proven useful in various analytical applications (reviewed in 36).

Drosophila nuclear extracts form specific complexes in an intron-dependent fashion after destruction of snRNA with micrococcal nuclease

Early in our studies we confirmed that micrococcal nuclease could be used to efficiently destroy snRNAs in *Drosophila* extracts. These studies revealed an unexpected, potentially important phenomenon.

When control (mock treated) extracts were used in spliceosome assembly experiments they behaved essentially as untreated extracts with only small losses in efficiency (results not shown). However, when micrococcal nuclease digested extracts were used they not only failed to support splicing and spliceosome assembly, they also caused the production of a new class of products—a ladder-like array of multimeric complexes (Figure 7). Both the *ftz* and *su(w^d)* introns produce the multimeric complexes as assessed by native gel electrophoresis. Moreover, a *su(w^d)* cDNA segment—including the exon sequences surrounding the first *su(w^d)* intron—failed to efficiently produce these complexes (Figure 7). Further, a small interval from the 3' region of the *ftz* intron containing the branchpoint and 3' splice site forms a similar (though more widely spaced; see below) multimeric complex array (Figure 7). Collectively, these observations suggest that intron (likely splice site) sequences are required for complex formation. These complexes form with kinetics similar to A complex formation (compare Figures 1, 2 and 7) but disassociate rapidly in the presence of heparin and excess EDTA at room temperature (Figure 7).

Collectively, these results indicate that intron sequences (presumably splice site sequences) in the absence of intact snRNAs direct the formation of novel, specific complexes on RNAs in *Drosophila* extracts. The dependence on intron sequences clearly suggests that these complexes are specific and meaningful. The behavior of the 3' half *ftz* intron is potentially particularly informative here. The half intron forms a ladder of multimeric complexes whose elements correspond to alternating elements in the ladder produced by the intact intron (Figure 7). This indicates that ladder structure is influenced by both 5' and 3' signals (presumably the corresponding splice sites). Apparently both splice site are recognized (independently or otherwise) in initiating or controlling multimeric complex formation.

We have tested the capacity of these multimeric complexes to enter into spliceosome assembly to determine if they might be intermediates (normally transient) in spliceosome assembly. If these complexes are first formed by brief incubation in micrococcal nuclease digested extracts then diluted into splicing reactions with untreated extract they are quite stable and do not enter into spliceosome assembly in the longest tested interval (Figure 7). However, if treated and untreated extracts are mixed in the same proportions *before* RNA is added, little or no multimeric complex is formed and spliceosome assembly occurs largely normally (though at somewhat reduced efficiency) as assessed by native gel electrophoresis (Figure 7). This suggests that multimeric complex formation and spliceosome assembly are competing, mutually exclusive assembly pathways in *in vitro* extracts.

It will be of substantial interest to determine the nature of the proteins responsible for these novel complexes and to assess their relationship to known splicing and hnRNP proteins (DISCUSSION). The sequence requirements for and properties of these complexes strongly suggest that the responsible proteins may be involved in very early stages of intron recognition. We also note that the six ladder elements on the ca. 300 base *ftz* RNA suggests either relatively small segments of RNA contacted by each ladder monomer or multimerization nucleated by the RNA but involving subsequent monomer additions independent of RNA contact.

DISCUSSION

We describe conditions which allow reproducible production of efficient *in vitro* splicing extracts from *Drosophila* embryos. We also describe application of these recipes to the relatively small amounts of starting material that can be conveniently generated without specialized, large scale *Drosophila* growth facilities. *Drosophila* represents a powerful genetic system for the analysis of pre-mRNA processing regulation (INTRODUCTION). The genotype of *Drosophila* embryos can be controlled in general. The availability of convenient, widely usable procedures should permit *in vitro* studies of genetically defined, regulated splices.

We report studies indicating that lessons from analysis of constitutive splicing in *Drosophila* will be directly applicable to understanding splicing in vertebrates. Specifically, we observe that *Drosophila* extracts share a number of properties with vertebrate extracts. We find that spliceosome assembly appears similar in detail in *Drosophila* and vertebrate extracts. Moreover, we find similar, high levels of RNaseH and 5' to 3' exonuclease activities in *Drosophila* and vertebrate extracts.

We show that a regulated arthropod intron—the first *su(w^d)* intron—apparently uses a remote branchpoint. Remote branchpoints are frequently found in regulated vertebrate introns (though their purpose remains unclear in some cases) and the sequence organization of the *su(w^d)* remote branchpoint is strikingly similar to those observed in vertebrates (RESULTS). These observations indicate that lessons from analysis of regulated splicing in *Drosophila* will, likewise, be directly applicable to understanding regulated vertebrate splicing.

Lastly, we show that *Drosophila* splicing extracts have a novel property. They will form ladder-like (multimeric) complexes on input RNAs if snRNAs are destroyed by micrococcal nuclease digestion. These complexes apparently require intron sequences—most likely splicing signals. Details of behavior of complex formation (RESULTS) suggest that information from both the 5' and 3' splicing signals participates in initiation or control of multimeric complex formation. Moreover, addition of intact snRNPs (untreated extract) to nuclease treated extract suppressed multimeric complex formation.

We propose the following interpretation of these observations. Initial intron recognition involves recognition of potential 5' and 3' splice sites (most simply by protein factors). These initial recognition complexes are then templates for two kinetically competing pathways. The first pathway is commitment to spliceosome assembly. The second is incorporation into the multimeric complexes described here. These multimeric complexes, once formed, are refractory to spliceosome assembly. Conversely, commitment to spliceosome assembly prevents stable multimeric complex formation.

This system of competing, mutually exclusive pathways has precisely the properties predicted for a kinetic proofreading

system improving fidelity of intron (exon) recognition. Initial recognition would provide the first round of screening of nascent RNAs for potential splice sites. Competition between commitment to spliceosome assembly and multimeric complex formation would provide a second round of screening improving fidelity. This strategy would be formally equivalent to the well understood kinetic proofreading strategies employed in translation and DNA replication (reviewed in 37). We note that initial recognition on our model might well involve currently recognized proteins such as SF2/ASF (31,38) and U2AF (39). On this hypothesis the novel components remaining to be identified are those allowing formation of the splicing-refractory, multimeric complexes. Our results suggest that these complexes may be sufficiently stable to purify and characterize.

We note that our proposal that multimeric complex formation participates in kinetic proof reading does not exclude a role for this process in regulated splicing. Specifically, in the presence of a kinetically competing pathway, a regulator need produce only relatively modest changes in the rate/efficiency of recognition of a potential splice site by the spliceosome commitment machinery to substantially change the probability of utilization of that site.

ACKNOWLEDGMENTS

We are grateful to a number of colleagues for helpful discussions at various stages of this work. Among these are Zuzana Zachar, Wallace LeSturgeon, Steve Mount, Paul Fisher and Frank Laski. This work was supported by NIH grant GM32003 to PMB.

REFERENCES

- Bingham, P.M., Chou, T.-B., Mims, I. and Zachar, Z. (1988). *Trends Genet.* **4**, 134–138.
- Baker, B.S. (1989). *Nature* **340**, 521–524.
- Maniatis, T. (1991) *Science* **251**, 33–34.
- Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. and McKeown, M. (1987). *Cell* **50**, 739–747.
- Zachar, Z., Chou, T.-B. and Bingham, P.M. (1987a) *EMBO J.* **6**, 4105–4111.
- Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988). *Cell* **55**, 1037–1046.
- Nagoshi, R.N., McKeown, M., Burtis, K. C., Belote, J.M. and Baker, B.S. (1988) *Cell* **53**, 229–236.
- Li, H. and Bingham, P.M. (1991) *Cell* **67**, 335–342.
- Tian, M. and Maniatis, T. (1992) *Science* **256**, 237–240.
- Reed, R. and Maniatis, T. (1985) *Cell* **41**, 95–105.
- Rio, D.C. (1988) *Proc. Nat. Acad. Sci.* **85**, 2904–2908.
- Allis, C.D., Waring, G.L. and Mahowald, A.P. (1977) *Dev. Biol.* **56**, 372–381.
- Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acids. Res.* **11**, 1475–1489.
- Soeller, W.C., Poole, S.J. and Kornberg, T. (1988) *Genes Dev.* **2**, 68–81.
- Parker, C.S. and Topol, J. (1984) *Cell* **36**, 357–369.
- Krieg, P.A. and Melton, D.A. (1987) *Met. Enz.* **155**, 397–415.
- Hope, I.A. and Struhl, K. (1985) *Cell* **43**, 177–188.
- Laughon, A. and Scott, M.P. (1984) *Nature* **310**, 25–31.
- Chou, T.-B., Zachar, Z. and Bingham, P.M. (1987) *EMBO J.* **6**, 4095–4104.
- Zillman, M., Zapp, M. L. and Berget, S.M. (1988) *Mol. Cell Biol.* **8**, 814–821.
- Konarska, M.M. and Sharp, P.A. (1988) *Cell* **49**, 763–774.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Press. Cold Spring Harbor, NY.
- Maxam, A.M. and Gilbert, W. (1980) *Met. Enz.* **65**, 499–560.
- Lo, P.C.H., and Mount, S.M. (1990) *Nucl. Acids Res.* **18**, 6971–6979.
- Alonso, A., Beck, E., Jorcano, J.L. and Hovemann, B. (1984). *Nucl. Acids Res.* **12**, 9543–9550.
- Das, G., Henning, D. and Reddy, R. (1987) *J. Biol. Chem.* **262**, 1187–1193.
- Myslinski, E., Branlant, C., Wieben, E.D. and Pederson, R. (1984) *J. Mol. Biol.* **180**, 927–945.
- Guthrie, C. and Patterson, B. (1988) *Ann. Rev. Genet.* **22**, 387–419.
- Saba, J.A., Busch, H., Wright, W. and Reddy, R. (1986) *J. Biol. Chem.* **261**, 8750–8753.
- Barabino, S.M.L., Blencowe, B.J., Ryder, U., Sproat, B.S., Lamond, A.I. (1990) *Cell* **63**, 293–302.
- Krainer, A.R., Mayeda, A., Kozak, D. and Binns, G. (1991) *Cell* **66**, 383–394.
- Ruskin, B. and Green, M.R. (1985) *Cell* **43**, 131–142.
- Noble, J.C.S., Prives, C. and Manley, J.L. (1986) *Nucl. Acids Res.* **14**, 1219–1235.
- Smith, C.W.J. and Nadal-Ginard, B. (1989) *Cell* **56**, 749–758.
- Helfman, D.M. and Ricci, W.M. (1989). *Nucl. Acids Res.* **17**, 5633–5650.
- Krainer, A.R. and Maniatis, T. (1988) in B.D. Hames and D.M. Glover, eds. *RNA Splicing. in Frontiers in Transcription and Splicing*. (Oxford and Washington, D.C.: IRL press). pp 131–206.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1989). *The Molecular Biology of the Cell*. Garland Publishing, Inc., New York, NY.
- Ge, H., Zuo, P. and Manley, J.L. (1991) *Cell* **66**, 373–382.
- Zamore, P.D. and Green, M.R. (1991) *EMBO J.* **10**, 207–214.