Sequence specificity of the P6 pairing for splicing of the group I *td* intron of phage T4

Karen Ehrenman^{1,2+}, Renee Schroeder^{1§}, P.Scott Chandry^{1,2} Dwight H.Hall³ and Marlene Belfort^{1,4*}

¹Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, PO Box 509, Albany, NY 12201-0509, ²Albany Medical College, 147 New Scotland Avenue, Albany NY 12208, ³School of Applied Biology, Georgia Institute of Technology, Atlanta, GA 30332 and ⁴School of Public Health Sciences, State University of New York at Albany, Empire State Plaza, Albany, NY 12237, USA

Received August 9, 1989; Revised and Accepted October 10, 1989

ABSTRACT

Seventeen non-directed *td*⁻ (thymidylate synthase-deficient) splicing-defective mutations isolated in phage T4 were localized within the catalytic core of the ribozyme. All of the mutations occur in conserved structural elements that form part of the *td* intron core secondary structure. Remarkably, seven of the seventeen independently isolated mutations clustered in the dinucleotide 5' element (P6[5']) of the putative two-base-pair P6 stem. An analysis of this region was undertaken by site-directed mutagenesis of the plasmid-borne *td* gene, leading to the following findings: First, the short P6 pairing in the *td* secondary structure model was verified with appropriate P6[5'] and P6[3'] compensatory mutations. Second, all P6[5'] and P6[3'] mutants are defective in the first step of splicing, guanosine-dependent 5' splice site cleavage, whereas their activity at the 3' splice site is variable. Third, residual *in vitro* splicing activity of the mutants altered on only one side of the P6 pairing is correlated with the ability to form an alternative two-base-pair P6 stem. Fourth, the degree to which the compensatory mutants have their splicing activity restored is highly condition-dependent. Restoration of phenotype of the compensatory P6[5']:[3'] constructs is weak under stringent *in vitro* conditions as well as *in vivo*. This sequence specificity is consistent with phylogenetic conservation of the P6 pairing elements in group I introns, and suggests either structural constraints on the P6 stem or a dual function of one or both pairing elements.

INTRODUCTION

The bacteriophage group I introns provide a convenient system for studying both the genetic and functional aspects of splicing *in vivo*. The utility of the system lies with the ease of isolating splicing-defective mutations based on phenotypic selections and screening procedures. Additionally, mutations made with cloned intron-containing constructs *in vitro* are readily introduced into their natural cellular context.

Chemically-induced splicing-defective mutations have been isolated from either the cloned *td* gene or directly from the phage. Of the 13 plasmid-borne mutations, all map to phylogenetically conserved elements throughout the catalytic core of the intron (P1 to P9.2) and all reduce 5' splice-site cleavage *in vivo* (1). This collection has revealed one missplicing mutant, in the 5' (exon I) element of P1 (2), and has provided a means to verify the secondary-structure model for the *td* intron, by isolation of non-directed pseudorevertants (Chandry and Belfort, unpublished). On the other hand, the phage mutant collection was used to delimit the two functional domains of the 1016 nt *td* intron (3). These domains comprise the 5' and 3' ends of the intron (ca. 100 and 170 nt, respectively) that fold into the catalytic core, which is split by an open reading frame (ORF) (735 nt) extending out of the P6 pairing.

Characterization of splicing-defective members of the phage collection has revealed a different spectrum of mutations from that in the cloned gene. In contrast to the plasmid-borne mutations, non-directed T4 *td* mutations causing splicing defects were markedly clustered. The most obvious 'hot spot' was in the 5' element of the putative P6 pairing (P6[5']), which contained seven of the seventeen mutations. This observation prompted us to ask whether the two neighboring residues that constitute the hot spot indeed correspond to the dinucleotide P6[5'] element of the short two-base-pair P6 helix of the *td* intron (Fig. 1). This question was of particular interest since the *td* and *sunY* introns of phage T4 are exceptional among the group IA introns with respect to their proposed P6 pairings. In seventeen of nineteen group IA introns, P6 is formed by a 5' G-U and a 3' C-G base pair, whereas the P6 of *td* and *sunY* has a 5' G-C followed by a C-G base pair (Figs 1 and 2; refs 4, 5).

Although site-directed mutagenesis experiments verified the existence of the proposed P6 pairing in *td*, the degree of functional restoration by the compensatory mutants was very low *in vivo* and extremely condition-dependent *in vitro*. This strict sequence specificity in the pairing elements of P6 is consistent with phylogenetic studies that implicate them as part of the highly conserved Q and R sequences of group I introns (5).

MATERIALS AND METHODS

Plasmids, bacterial strains and media

Isolation of phage T4 *td*⁻ derivatives was described by Hall et al. (3). The host strain used for *in vivo* analysis of the P6 plasmid constructs was JM109*thyA* [a *thy*⁻ derivative of JM109 (6) isolated in this laboratory]. For the isolation of single-stranded DNA substituted with uridine, strain CJ236 (*dut*, *ung*, *thi*-1, *recA*, *spoT*1, F', Cam^R) was employed. The P6 mutations were constructed from the 4.28-kb 'wild type' *td* plasmid pTZ*td*ΔK (kindly provided by J.G. Salvo). This plasmid contains a 1.46 kb *EcoRI*-*DraI* fragment, with an intron of 346 nt and intact exon sequences, in the *EcoRI*-*PstI* interval of pTZ18U (USB). The intron deletion is within the ORF, from intron residue 120 to 797, with an *XbaI* site inserted at the deletion junction. Minimal medium enriched with casamino acids but lacking thymine (-THY) was used for selection of the Td⁺ phenotype, whereas the same medium supplemented with trimethoprim and thymine (TTM) was used for selection of Td⁻ clones (1). Rich media and growth conditions for RNA extraction were as previously described (1-3).

Site-directed mutagenesis

Oligonucleotides used for site-directed mutagenesis have the following sequences, with the mutated residues underscored: M1, 5'-TACAATTTAGGACGGGATT-3'; M2, 5'-TACAATTTACCCACGGGATT-3'; M3, 5'-TACAATTTACGACGGGATT-3'; M4, 5'-GTCGTTAGCCATTTATGT-3'; M5, 5'-GTCGTTAGGGATTTATGT-3'; M6, 5'-GTCGTTAGCGATTTA-TGT-3'. Deoxyuridine-substituted single-stranded DNA was isolated following the procedure of Kunkel et al. (7) and Vieira et al. (8). Annealing of the phosphorylated oligonucleotide to the single-stranded DNA, elongation and ligation were according to Williamson et al. (9). Some mutants were screened by hybridization to the mutagenic oligonucleotide, others by plating on selective minimal media, and all mutations were confirmed by dideoxy-sequencing of double-stranded DNA (10). Compensatory mutants M1:4, M2:5 and M3:6 were constructed by cloning plasmid subfragments. The mutations in P6[5'] are contained on a 3.21-kb *XbaI*-*PstI* fragment, and those in P6[3'] on the remaining 1.07-kb *XbaI*-*PstI* fragment of pTZ*td*ΔK. Fragments were gel-purified and ligated to form the compensatory constructs.

In vivo RNA analysis

Primer extension analysis of RNA and sequencing of phage mutants were performed with reverse transcriptase as previously described (3). Where stops in all the lanes obscured the sequence, terminal transferase was added to the reaction (11). For Northern blot analysis, RNA extracted from cells induced with isopropyl- β -D-thiogalactoside (IPTG) was separated on a denaturing 5% acrylamide-7M urea gel and electroblotted onto Hybond (Amersham) membrane. Oligonucleotides used as hybridization probes (12) were complementary to the splice junction (12 nt to each side) and to nt 51–72 of the intron (3).

In vitro RNA analysis

Transcription: Plasmid DNA was digested with *EcoRV*, which cleaves once, in exon II, 1322 nt from the transcriptional start of the T7 promoter. The transcription reaction, which favored the accumulation of pre-mRNA, contained 4 μ g of linearized plasmid DNA, 40 mM Tris-HCl pH 7.5, 3 mM MgCl₂, 0.4 mM spermidine, 0.5 mM DTT, 1 mM of each NTP, 20 units of T7 RNA polymerase (Stratagene), 25 units of placental ribonuclease inhibitor (RNasin, Promega) and 10 μ Ci of [³⁵S]-CTP (850 Ci/mmol, Amersham) in a total volume of 25 μ l. The incubation was for 60 minutes at 30°C. The RNA was then precipitated with ethanol in the presence of 20 μ g of yeast tRNA and 3.5 M NH₄OAc three times to remove unincorporated GTP. Precursor RNA was resuspended in 50 μ l of DEPC-treated water and aliquots were used for splicing assays.

Self-splicing: [³⁵S]-labeled precursor transcript (2×10^5 cpm) was incubated in 10 μ l Buffer A (3 mM MgCl₂, 40 mM Tris-HCl pH 7.5, 0.4 mM spermidine and 100 μ M GTP) for 30 minutes at 42°C. The reaction was stopped by adding 5 μ l of FDM (0.1% xylene cyanol, 0.1% bromophenol blue, 10 mM EDTA and 95% deionized formamide) to the samples, which were then heated for 10 minutes at 65°C and separated on a 5% acrylamide-7M urea gel.

GTP incorporation: Non-radioactive precursor transcripts were used for labeling with radioactive GTP (13). Pre-mRNA (300 nM) was incubated for 30 minutes at 42°C in Buffer B (50 mM NH₄Cl, 3 mM MgCl₂, 1 mM spermidine, 50 mM Tris-HCl pH 7.5, 0.5 mM DTT and 1 unit of RNase Block) (13) or Buffer C (100 mM (NH₄)₂SO₄, 5 mM MgCl₂, and 50 mM HEPES, pH 7.5) (14), with 10 μ Ci of [α -³²P]-GTP (Amersham), in a total volume of 10 μ l. The reaction was stopped by the addition of EDTA to a final concentration of 25 mM. FDM was added to reactions before running on a 5% acrylamide-7 M urea gel.

GpU assay: Unlabeled precursor transcripts (300 nM) were incubated with 100 μ M [³²P]-labeled GpU in either Buffer B or Buffer C (14). GpU was labeled in a mixture containing 2 mM GpU (Sigma), 60 μ Ci [γ -³²P]-ATP (Amersham), 10 units T4 polynucleotide kinase (New England Biolabs), 5 mM DTT, 10 mM MgCl₂, 0.2 mM spermidine and 0.1 M Tris-HCl (pH 8.0) for 90 minutes at 37°C. The kinase was inactivated by heating to 90°C for 3 minutes. Transcripts were incubated with labeled GpU at 37°C for 1 hour, after which reactions were stopped and the samples applied to a gel as described above.

RESULTS

Non-directed mutations cluster in P6

Chemically-induced T4 *td* mutations that mapped within or close to the intron were screened for splicing defects by an RNA dot-blot assay (3). Seventeen independent splicing-defective mutants, nine of which were generated with hydroxylamine (designated by prefix H) and eight with nitrous acid (designated by prefix N), were characterized further by sequence analysis. The sequence changes were determined by primer extension analysis of RNA

Table 1. Summary of chemically induced mutations

Mutation ^a	Base Change	Site ^b	Pairing ^c
H6 ^d	G-A	-2	P1 [5']
N57 ^e	C-T	10	P1 [3']
N56	G-A	78	P6 [5']
H108A	G-A	78	P6 [5']
H8 ^f	C-T	79	P6 [5']
N1	C-T	79	P6 [5']
N9	C-T	79	P6 [5']
N19	C-T	79	P6 [5']
N37	C-T	79	P6 [5']
H9	C-T	870	P7 [5'] ^g
H104A	G-A	914	P3 [3']
H4	C-T	946	P7 [3']
H105A	C-A	946	P7 [3']
N47 ^e	G-A	968	P9 [3']
N49	G-A	968	P9 [3']
H10	G-A	968	P9 [3']
H103A	C-T	982	P9.1 [3']

^aMutagenic agents: N = Nitrous acid, H = Hydroxylamine

^bNumbering corresponds to the coordinates of the intron sequence with the first coded nucleotide = 1.

^cShown in Figure 1.

^dExon I mutation.

^eMutations previously reported (3).

^fThis mutation is td8 of Simon and Tessman (15).

^gBulged nucleotide within P7.

from mutant-infected cells and are summarized in Table 1 and Fig. 1. The linear map (Fig. 1A) confirms that the mutations fall into the two domains that correspond to the secondary structure core (Fig. 1B), which was modeled by phylogenetic comparison (4,16). The distribution of mutations is uneven, with only nine nucleotide positions being represented by the seventeen mutations (Table 1, Fig. 1). Two mutations are in the P1 pairing, one in P3, seven in P6, three in P7 (one of which is in the unpaired residue of this helix), three in P9 and one in P9.1. Two of these mutations, N47 and N57 (Table 1, Fig. 1),

Figure 1. Distribution of non-directed chemically induced *td* mutations. **A.** Linear map. Seventeen of the splicing defective mutations mapped by marker rescue to one of four intervals shown (I, II, III, IV; ref. 3) were sequenced to nine different sites (dots above map). Clustering of mutations at a site in the *td* intron is indicated by a vertical representation of mutations (see Table 1) above the site. The bounds of the ORF are indicated by dotted lines within the intron. **B.** Secondary-structure map. Splicing-defective mutations are circled on the intron secondary structure model proposed by F. Michel. The number of independent mutations found at each site is indicated by black triangles. Phylogenetically conserved regions P, Q, R, and S are shaded (5). The intervening sequence is numbered beginning with the first coded nucleotide. The start and stop codons of the ORF are indicated by three dots each.

have been previously described and shown by primer-extension analysis to be defective in 5' splice-site cleavage (3). By monitoring RNA from phage-infected cells, one representative mutant at each of the other seven loci have similarly been demonstrated to have reduced 5' splice-site activity (data not shown).

Figure 2. Site-directed mutations. A. P6 mutations. The four nucleotides of P6 are boxed, with circles indicating the position(s) of the mutations. M1, M2 and M3 are the 5' mutations at residues 78 and/or 79, and M4, M5, and M6 are the 3' mutations at residues 864 and/or 865. The compensatory mutations are M1:4, M2:5 and M3:6. B. *In vivo* phenotypes. The growth properties of JM109*thyA* harboring either wild-type or mutant plasmids were determined by patching onto differential media according to the template shown. V = Td⁻ vector pTZ18U, WT = Td⁺ wild-type pTZ*tdΔK*, and numbers correspond to mutant designations in A. +THY = minimal media supplemented with thymine (nonselective); -THY = unsupplemented minimal media, selective for TS producers; TTM = minimal media supplemented with both thymine and the folate analog trimethoprim, selective for TS underproducers. All plates were incubated at 37°C overnight.

Generating site-directed mutations in P6

Since seven of the seventeen mutations occurred in the putative 5' element of P6 and none in P6[3'], site-directed mutagenesis was used to verify the P6 pairing and to further analyze its function. On phylogenetic grounds, P6 was proposed to consist of two base pairs, with G78 and C79 pairing with C865 and G864, respectively (Fig. 1; refs 4,16). The site-directed mutations were designed to alter one or both of the nucleotides on either side of the stem, by changing G to C or vice versa. Thus, when compensatory mutations were created, G:C pairings were maintained, with bond strengths comparable, albeit not identical, to the wild type (Fig. 2A).

Oligonucleotide-directed mutations were generated in the single-stranded form of pTZidΔK, which carries a functional *td* fragment that contains a 346-nt intron, deleted within the central non-essential portion. This parental construct, in vector pTZ18U, has the *td* fragment cloned downstream of the tandem *plac* and *pT7* promoters for *in vivo* and *in vitro* expression, respectively.

Verification of the P6 pairing

To assess growth phenotypes, cells harboring the mutant constructs, the wild-type *td* plasmid and a vector control were patched onto solid media that distinguish splicing proficient clones capable of producing the *td* product, thymidylate synthase (TS), from splicing deficient clones that have low TS levels (Fig. 2B, ref. 1). As expected, all of the constructs grew on minimal media in the presence of exogenous thymine, as TS is not required under these conditions. However, in the absence of thymine, none of the unilateral mutants with mutations in either the 5' or 3' elements of P6 (M1 through M6) grew, indicating a splicing-defective TS⁻ phenotype. In contrast, the compensatory mutants (M1:4, M2:5 and M3:6) were able to grow in the absence of thymine, albeit less well than the wild-type. In support of phenotypic suppression being only partial, the compensatory mutants, unlike the wild-type, were able to grow on TTM medium, which is selective for TS-deficient cells. Thus, although this preliminary phenotypic assay supports the presence of the indicated P6 pairing, which appears necessary for accurate splicing of the *td* intron (and the synthesis of TS), the pairing alone seems insufficient for restoring TS production to wild-type levels *in vivo*.

Subsequently, *in vitro* analyses were undertaken. Since intron mutants that are incapable of splicing *in vivo* can often be activated *in vitro* by changing buffer conditions (e.g. Mg⁺⁺, monovalent cations, pH), buffers were chosen to satisfy the following criteria. First, buffers were used that allowed the distinction of subtle differences between the activities of the mutants (Buffers A and B). Second, more stringent conditions were employed in which the activity of the mutants *in vitro* corresponded more closely to that *in vivo* (Buffer C).

To verify P6 *in vitro* and to investigate the self-splicing properties of the mutants, precursor RNAs were incubated in Buffer A at 42°C for increasing times (0, 6, 12 and 20 min). As shown in Fig. 3, the unilateral mutants all exhibited reduced activity relative to wild-type, with M1, M3 and M5 precursors not generating any detectable splice products and M2, M4 and M6 displaying a low level of activity. As will become evident below, the bands observed for M1 are hydrolysis products rather than splice products (Fig. 3B, lane 1). In contrast to the unilateral mutants, compensatory constructs M1:4, M2:5 and M3:6 had splicing substantially restored. For all three compensatory mutants, the level of splice products (mRNA, linear intron, circular intron) and splicing intermediates (exon I and intron-exon II) exceeded that for the respective 5' and 3' unilateral mutants. Whereas compensation was partial for M1:4 (ca. 25% of wild-type activity, with the hydrolysis-

sensitive phenotype of M1 suppressed), M2:5 approached and M3:6 slightly exceeded wild-type activity (Fig. 3C).

To measure activity at the 5' splice site and to distinguish splicing from site-specific hydrolysis, a radioactive GTP incorporation assay was performed in Buffer B (Fig. 4A). Activity at the 5' splice site was evident for exactly those constructs that showed putative splice products in Fig. 3. Labeled intron was not observed for unilateral mutants M1, M3 and M5 (Fig. 4A, lanes 1, 3 and 5, respectively), while weak incorporation was evident for M2, M4 and M6 (lanes 2, 4 and 6, respectively). An intermediate amount of labeled intron and intron-exon II were generated for compensatory mutant M1:4 (lane 7), while for M2:5 (lane 8) and M3:6 (lane 9) these products approached wild-type levels (lane 10). These results indicate that the products seen in Fig. 3 are *bona fide* splice products, rather than hydrolysis products, which appear in a guanosine-independent manner. (The exception here is M1, the hydrolysis-sensitive mutant, which consistently showed many spurious bands [Fig. 3B, lane 1], none of which correspond to a GTP-labeled product). Taken together, the restoration of phenotype of the compensatory mutants *in vivo* and *in vitro* provides confirmation of the proposed P6 pairing.

Figure 3. Self-splicing of P6 mutants. Precursor was generated with T7 polymerase from pTZ Δ ΔK linearized at the *EcoRV* site in exon II. The gel in panel A (5% acrylamide/7M urea) shows a time-course of splicing for M3 (P6[5']), M6 (P6[3']) and M3:6 (compensatory mutant), as well as wild-type (WT). Similar amounts of [35 S]-labeled precursor transcripts (2×10^5 counts) were incubated in Buffer A for 0, 6, 12, and 20 minutes at 42°C. Panel B contains RNA from all of the mutants incubated for 20 minutes at 42°C. Lane 1 = construct M1, lane 2 = M2, lane 3 = M3, lane 4 = M4, lane 5 = M5, lane 6 = M6, lane 7 = M1:4, lane 8 = M2:5, lane 9 = M3:6, lane 10 = wild type. Lane M in both A and B contains a 123 base DNA molecular size ladder. Sizes of the precursor (pre mRNA) ligated exons (ExI-ExII) free exon I (ExI) and intron-exon II (In-ExII) as well as circular intron (C.In) and linear intron (L.In) are shown to the right of the RNA species in B. Doublets corresponding to ExI, ExI-ExII and In-Ex II reflect products generated by use of an efficient cryptic splice site in exon I (2). The relative accumulation of ligated exons is plotted in C. Quantitation of A is represented in the bottom series of graphs (M3, M6). Similar analysis of gels containing the M1, M4, M1:4, WT series and the M2, M5, M2:5, WT series are shown above (plots M1, M4 and M2, M5, respectively). Densitometer tracings from different exposures were quantitated and plotted relative to pre-mRNA at 0 min. Open circles = wild type, filled circles = P6[5'] mutants, squares = P6[3'] mutants, triangles = compensatory mutants.

Sequence specificity of the P6 pairing in vivo

To further investigate the partial plating phenotypes of the compensatory mutants (Fig. 2), RNA extracted from transcriptionally induced cultures was examined by Northern hybridization analysis. With an intron probe, the wild-type construct revealed two discrete bands, one corresponding to linear intron (347 nt), and the other, migrating aberrantly slowly, corresponding to circular intron (345 nt) (Fig. 5A, lane 10). The only other constructs giving these bands were the compensatory mutants, with M1:4 giving the weakest signals, M2:5 slightly more intense signals, and M3:6 giving the most intense signals (Fig. 5, lanes 7, 8 and 9, respectively). However, in all cases the compensation as a sum of the levels of free linear plus circular intron was inefficient, with values <10% of wild-type levels (Fig. 5B). Similar blots were also hybridized to a splice-junction probe,

Figure 4. Analysis of P6 mutants for 5' splice site activity. Equivalent amounts of unlabeled precursors were incubated with [32 P]-GTP for 30 minutes at 42°C in Buffer B (panel A) or Buffer C (panel B). Lanes 1–10 and M are labeled as in Fig. 3B. [5'] designates the P6[5'] mutants, [3'] the P6[3'] mutants, [5']:[3'] the compensatory mutants, and WT the wild type. Bands which were labeled with GTP include linear intron (L.In) and intron-exon II (In-ExII). The heavy bands immediately above L.In and In-Ex II in panel A are likely to reflect use of the cryptic 5' splice site in exon I. The identity of other faint bands is unknown. The L.In and In-ExII bands in panel A were quantitated by densitometry. The levels of the sum of these two species relative to wild type, which was assigned a value of 100, are as follows: lane (1), <1; lane (2), 8; lane (3), <1; lane (4), 6; lane (5), <1; lane (6), 8; lane (7), 26; lane (8), 44; lane (9), 86.

identifying mRNA. Again, the only RNAs that hybridized to the probe were from the compensatory mutants and the wild-type, the latter giving a signal between 7 and 50 times over that for the compensatory mutants (Fig. 5B).

Splicing *in vivo* was also assessed by measuring the production of TS using [3 H]-FdUMP, which binds TS in the presence of methylene tetrahydrofolate (17,18). The only constructs that produced a labeled ternary complex corresponding to TS were the compensatory mutants and the wild-type, with the compensatory mutants producing <10% of wild-type TS levels (data not shown). These results reinforce the conclusion that P6 is necessary for splicing and therefore for TS production, while they corroborate the finding that the nature of the P6 pairing elements have a major impact on splicing proficiency *in vivo*.

Sequence specificity of the P6 pairing in vitro

To test the sequence specificity of the P6 pairing elements *in vitro*, the GTP-incorporation assay was also performed under more stringent buffer conditions (Fig. 4B). Strikingly, only the wild-type intron RNA was labeled in Buffer C and not that of any of the unilateral or compensatory mutants (cf. Figs. 4A and B).

Figure 5. Northern hybridization analysis of the P6 mutants. A. RNA blot. After one hour of transcriptional induction with IPTG, RNA was deproteinized and 5 μ g was separated on a denaturing acrylamide gel, electroblotted and probed with a [32 P]-labeled intron specific oligonucleotide. Lanes and bands (C.In and L.In) are labeled as in Fig. 3B. The high mol wt bands across all lanes are *td* transcripts of undefined length synthesized in the absence of a transcriptional terminator. B. Relative amounts of intron and spliced mRNA. The relative intensities of the linear and circular intron were derived from densitometric tracings of the Northern analyses in A. Similar blots as well as slot blots were probed with splice-junction-specific oligodeoxynucleotide for determination of mRNA levels. Lanes 7, 8, 9 and 10 are numbered as in A.

Activity at the 3' splice site was also tested, by providing a mini-exon I ([32 P]-GpU) (Fig. 6). This assay is analogous to the CpU trans-splicing assay for the *Tetrahymena* rRNA ribozyme (14), with the different buffer conditions for this experiment corresponding to

Figure 6. Analysis of P6 mutants for 3' splice site activity. Equivalent amounts of unlabeled precursor transcripts were incubated with [32 P]-GpU for 1 hour at 37°C in Buffer B (panel A) or Buffer C (panel B). Lanes are labeled as in Fig. 4. The GpU-exon II (ExII) ligation product is indicated. Linear intron (L.In) is labeled by an uncharacterized side-reaction, which is likely to be analogous to CpU-intron ligation in the CpU assay for *Tetrahymena* rRNA 3' splice site activity (14). Quantitation of ExII in panel A by densitometry revealed the following levels of the GpU-exon II band relative to wild type, which was assigned a value of 100: lane (1), <1; lane (2), 55; lane (3), <1; lane (4), 255; lane (5), <1; lane (6), 487; lane (7), 32; lane (8), 57; lane (9), 117.

those used for the GTP incorporation assay (Fig. 4). Mutants M1, M3 and M5, which were inactive in 5' splice site cleavage even under non-stringent Buffer B conditions (Fig. 4A), were also completely defective at the 3' splice site (i.e. [32 P]-GpU ligation to exon II) (Fig. 6A). However, M2, M4 and M6, which were weakly active at the 5' splice site (Fig. 4A), showed strong 3' splice site activity under similar conditions (Fig. 6A), with M4 and M6 exhibiting appreciably higher GpU-exon II levels than either the wild type or the compensatory mutants (Fig. 6A, cf. lanes 4 and 6 with 7 through 10). Although the basis for this variable behavior at the 3' splice site is not known, it suggests that the primary lesion in the P6 mutants is at the 5' splice site. Despite the apparent uncoupling of 5' and 3' splice site activity in some of the unilateral mutants, the degree of phenotypic restoration for the compensatory mutants in the GpU assay parallels that in the GTP incorporation assay under both the more relaxed (cf. Fig. 4A and Fig. 6A) and the stringent

(cf. Fig. 4B and Fig. 6B) buffer conditions. The absence of any 3' splice-site activity in the compensatory mutants under the latter conditions corroborates the sequence specificity of the P6 pairing.

DISCUSSION

A collection of seventeen non-directed phage T4 *td*⁻ mutations that cause defects in the *td* splicing pathway were localized to the core structure of the 1016-nt intron (Table 1, Fig. 1). All but one of the mutations characterized (H9, the bulged residue in P7—to be described elsewhere, R.S. and M.B.), disrupt predicted RNA stems (P1, P3, P6, P7, P9, P9.1) of the *td* intron secondary structure folding (4,16). Of these, seven independently isolated mutations cluster in the dinucleotide P6[5'] element. We further analyzed this mutational hotspot by constructing a set of site-directed mutations (Fig. 2). Six of these constructs disrupt the P6 pairing (M1, M2, M3, M4, M5 and M6) abolishing all detectable splicing *in vivo* (Figs. 2 and 5) and *in vitro* under stringent conditions (Fig. 4 and 6). The compensatory mutations that allow the P6 stem to reform are capable of restoring function, thereby verifying the proposed P6 pairing (Figs. 2–6). The compensatory mutants, however, exhibit only partial function *in vivo* (< 10% of wild-type) and are inactive under stringent *in vitro* conditions, demonstrating that although the P6 pairing is necessary for splicing activity, there is a sequence specificity for P6[5'] and/or P6[3']. Partial function of a P6 compensatory mutant of the *Tetrahymena* ribozyme has also been reported (9).

Estimated free energies for RNA double helical regions do not satisfactorily account for the difference in splicing activity between the wild type and the compensatory mutants. The pairing of the wild type and of M1:4 and M2:5 have the same predicted free energy ($\Delta G = -5.0$ kcal), while M3:6 has a less favorable free energy ($\Delta G = -3.2$ kcal) (19). Yet M3:6 is the most active of the compensatory mutants. Also, in a phylogenetic comparison of P6 stems, a U in the position equivalent to C865 is often observed, for which the ΔG is even less favorable than for M3:6. Thus, there appears to be no correlation between the strength of the P6 stem and activity of the ribozyme. This observation is consistent with a specific structural property of the P6 helix, or, alternatively, with a bifunctional role for P6[5'] and/or P6[3']. These functional constraints appear to be reflected in the high degree of conservation of P6 among the group I introns, with P6[5'] forming the two 3' residues of element Q and P6[3'] forming the two 5' residues of R, Q and R being two of the four conserved group I sequence elements (5).

In rationalizing the sequence specificity of P6, options ranging from constraints on the structure of the stem, to additional roles of the pairing elements are considered. First, the critical location of the P6 stem in the three-dimensional model of the *Tetrahymena* intron is noteworthy. P6 contributes the last two base pairs to helix II, immediately adjacent to junction J6/7 that connects the helices I and II (20). One possibility is that P6 maintains the appropriate juxtaposition of the two helices and that a structural distortion introduced by the P6 compensatory mutants affects the relationship of helices I and II. A second alternative is that the P6 pairing undergoes a conformational change during splicing. The structural dynamics of the P6 compensatory mutants may be altered so as to prevent unwinding and an alternative intramolecular status of P6[5'] and P6[3']. A third possibility, also related to a conformational change, is that P6[5'] and/or P6[3'] are involved in alternate pairings, which are disrupted by the mutated residues. Finally, the P6 pairing elements may be involved in a non-Watson-Crick triple-base pairing interaction, as suggested by mutational analysis of the *sunY* and *Tetrahymena* rRNA introns (F. Michel, A. Ellington,

S. Couture, M. Cherry and J. Szostak, personal communication). Such triple-base pairs may similarly be disrupted by the mutated residues.

The relaxed specificity of the ribozyme under non-stringent *in vitro* conditions allowed the properties of the P6 mutants to be distinguished. Interestingly, mutants M2, M4 and M6 show residual splicing activity (Figs. 3, 4A and 6A), even though the normal P6 pairing is disrupted. We explain this remaining activity by postulating that these mutants are able to form an alternative P6 pairing, using nucleotide C866 to pair with nucleotide G78, as illustrated in Fig. 7. This potential to form an alternative pairing is correlated with low activity at the 5' splice site and high activity at the 3' splice site (Fig. 4A and Fig. 6A). In support of the model, M1, M3 and M5, which are completely splicing defective under all conditions tested, do not have the possibility of undergoing alternative pairing (Figs. 3, 4, 6 and 7). A 'sliding' model has also been proposed to account for the activity of a P6[5'] mutant of the *Tetrahymena* rRNA intron (9). However, a fundamental difference between the two cases exists, in that for the *Tetrahymena* intron, the sliding would introduce a nucleotide between P4 and P6, whereas for the *td* intron, the junction between P6 and P7 (J6/7) would be reduced by one nucleotide. Our results, therefore, suggest that the J6/7 linker region, with a conserved length in nature of 3 nt (5), might maintain partial function *in vitro* with only 2 nt.

Functional analysis of the P6 mutants indicates that the primary lesion is in the first step of splicing, with variable effects at the 3' splice site. Under stringent buffer conditions, all unilateral as well as compensatory mutants have no detectable 5' or 3' splice-site activity (cf. Fig. 4B and 6B). Under more relaxed conditions, mini-exon ligation parallels activity at the 5' splice site for mutants M1, M3, M5 and for the compensatory mutants (cf. Fig. 4A and Fig. 6A). These results suggest an impairment of 3' splice site function in concert with that at the 5' splice site. However, mutants M2, M4 and M6, which have some residual activity at the 5' splice site under non-stringent conditions, seem to be unimpaired at the 3' splice site. That is, those mutants capable of forming the alternative P6 pairing (Fig. 7) have retained the ability to undergo extremely efficient exon ligation, indicating an uncoupling of 3' and 5' splice-site activity (cf. Fig. 4A and 6A). Taken together, these results point to 5' splice-site activity as the primary defect in the P6 variants. In studies verifying the P6 pairing of the *Tetrahymena* rRNA intron, Williamson *et al.* (9) analyzed a compensatory mutant containing a two-base P6[3'] mutation and a two-base P6[5'] mutation. The properties of the four-base compensatory mutant, with a partially restored phenotype, suggested that in this case P6 may have a particular effect on exon ligation, although its influence on 5' splice-site cleavage was unclear. Initially these results appear in conflict to those presented above, which suggest that the primary defect is at the level of the first step of splicing. The apparent differences between the T4 and *Tetrahymena* systems may, however, be accounted for by the dramatic variability that we have observed in 3' splice site activity for the P6 mutants on one hand, and the study of one particular *Tetrahymena* P6 mutant on the other.

The question of why mutations isolated in the phage cluster in P6[5'] is still a matter of speculation. Unusual sensitivity of these two residues to mutation by chemical mutagens

Figure 7. Alternate pairing model for P6 mutants. A sliding model for constructs M2 (A), M4 (B), and M6 (C), all of which retain some degree of *in vitro* splicing activity, is presented. The altered bases may slide to form an alternative pairing, which maintains the integrity of P6. Those mutations in which splicing did not occur (M1 is given as an example in panel D) do not have a feasible alternate pairing. Arrows indicate the shift in predicted pairing, and the circles represent mutated nucleotide(s).

seems unlikely in view of the different spectrum of mutations isolated by similar non-directed mutagenesis of the plasmid, in which case only one of 13 mutations were in P6[5'] (1). The difference in the mutational spectrum between the plasmid and phage systems may have additional bearing on the problem. The T4 *td*⁻ mutants have an intact 735 nt ORF looped out of P6, whereas the plasmid *td*⁻ mutants have a deletion of ca. 630 nt in the ORF. Perhaps the increased length of the ORF, which is not involved in RNA catalysis, places a greater functional constraint on the short P6 pairing. Mutations might then be phenotypically stronger, more readily passing the genetic screens imposed, and thereby causing a mutational bias in the phage. The observation that this bias does not extend to P6[3'] may result from the specificity of the transition mutagens used to generate the mutations. For example, a C to U change at position 865 in P6[3'] would allow a G-U pair with residue 78 in P6[3']. Such a mutation is likely to be phenotypically silent, in view of a G-U pairing occurring naturally at this position in 17 of 19 group IA introns, the *td* and *sunY* introns being the only exceptions (5). Whatever the case may be, these studies have verified the importance of P6 in the functioning of a group I ribozyme, and they underscore a dramatic sequence specificity in the 5' and 3' elements of this pairing, suggesting either a structural constraint on the P6 stem, or a bifunctional role in RNA catalysis.

ACKNOWLEDGEMENTS

We wish to thank our colleagues in the laboratory for their helpful input into this work, particularly Jill Galloway Salvo for useful discussions and William Chan and Debbie Bell Pedersen for moral support to K.E. in the final stages of her Ph.D. research. We are thankful to Jonathan Clyman, Susan Quirk and Jill Galloway Salvo for critical reading of the MS and Carolyn Wieland for expert preparation of the MS. We thank Francois Michel and colleagues for permission to cite their unpublished results. This work was supported by NSF grant DMB8502961 and NIH grant GM39422 to M.B., NIH grant GM36714 to D.H.H., and Fonds zur Forderung der wissenschaftlichen Forschung, Vienna, Austria to R.S.

*To whom correspondence should be addressed

Present addresses: ⁺Fred Hutchinson Cancer Research Center, 1124 Columbia St., Seattle, WA 98104, USA and [§]Institut für Mikrobiologie und Genetik der Universität Wien, Vienna, Austria

REFERENCES

1. Belfort, M., Chandry, P.S. and Pedersen-Lane, J. (1987) Cold Spring Harbor Symp. Quant. Biol., **52**, 181–192.
2. Chandry, P.S. and Belfort, M. (1987) Genes & Devel., **1**, 1028–1037.
3. Hall, D.H., Povinelli, C.M., Ehrenman, K., Pedersen-Lane, J., Chu, F. and Belfort, M. (1987) Cell, **48**, 63–71.
4. Shub, D.A., Gott, J.M., Xu, M.Q., Lang, B.F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. and Belfort, M. (1988) Proc. Natl. Acad. Sci. USA, **85**, 1151–1155.
5. Cech, T.R. (1988) Gene, **73**, 259–271.
6. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene, **33**, 103–119.
7. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Methods Enzymol., **154**, 367–382.
8. Vieira, J. and Messing, J. (1987) Methods Enzymol., **153**, 3–11.
9. Williamson, C.L., Desai, N.M. and Burke, J.M. (1989) Nucl. Acids Res., **17**, 675–689.
10. Bartlet, J.A., Gaillard, Jr., R.K. and Joklik, W.K. (1986) Biotechniques, **4**, 308.
11. DeBorde, D.C., Naeve, C.W., Herlocher, M.L. and Maassab, H.F. (1986) Anal. Biochem., **157**, 275–282.
12. Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) Nucl. Acids Res., **9**, 879–894.
13. Garriga, G. and Lambowitz, A.M. (1984) Cell, **39**, 631–641.

14. Inoue, T., Sullivan, F.X. and Cech, T.R. (1985) *Cell*, **43**, 431–437.
15. Simon, E.H. and Tessman, I. (1963) *Proc. Natl. Acad. Sci. USA*, **50**, 526–532.
16. Michel, F., Jacquier, A. and Dujon, B. (1982) *Biochimie*, **64**, 867–881.
17. Belfort, M., Moelleken, A., Maley, G.F. and Maley, F. (1983) *J. Biol. Chem.*, **258**, 2045–2051.
18. Salvo, J.G.S., Coetzee, T. and Belfort, M. *J. Mol. Biol.* In press.
19. Tinoco, Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.*, **246**, 40–41.
20. Kim, S.-H. and Cech, T.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8788–8792.

**This article, submitted on disc, has been automatically
converted into this typeset format by the publisher.**