



**HAL**  
open science

## Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site

Maria Costa, François Michel

### ► To cite this version:

Maria Costa, François Michel. Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO Journal*, 1999, 18, pp.1025 - 1037. 10.1093/emboj/18.4.1025 . hal-03914605

**HAL Id: hal-03914605**

**<https://hal.science/hal-03914605>**

Submitted on 27 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site**

**Maria Costa<sup>1</sup> and François Michel**

Centre de Génétique Moléculaire du C.N.R.S., 91190 Gif-sur-Yvette, France.

<sup>1</sup>Present address: Center for Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

Proofs to: Maria Costa or François Michel

Center for Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

Telephone: (831) 459 2700; Fax: (831) 459 3737

e-mails: maria@ascari.ucsc.edu

michel@mercure.cgm.cnrs-gif.fr

Running title: Footprinting of a group II intron by its 5' exon

## Abstract

Group II self-splicing requires the 5' exon to form base pairs with two stretches of intronic sequence (EBS1 and EBS2) that also bind the DNA target during retrotransposition of the intron. We have used dimethylsulfate modification of bases to obtain footprints of the 5' exon on intron Pl.LSU/2 from the mitochondrion of the alga *Pylaiella littoralis*, as well as on truncated intron derivatives. Aside from the EBS sites, which are part of the same subdomain (ID) of ribozyme secondary structure, three distant adenines become either less or more sensitive to modification in the presence of the exon. Unexpectedly, one of these adenines in subdomain IC1 is footprinted only in the presence of the distal helix of domain V, which is involved in catalysis. While the loss of that footprint is accompanied by a 100-fold decrease in the affinity for the exon, both protection from modification and efficient binding can be restored by a separate domain V transcript, whose binding results in its own, concise footprint on domains I and III. Possible biological implications of the need for the group II active site to be complete in order to observe high-affinity binding of the 5' exon to domain I are discussed.

Keywords: group II intron / intron-exon interactions / methylation probing / ribozyme / 5' exon footprints

## Introduction

Group II self-splicing is a two-step process which is normally initiated by transesterification at the 5' splice site. The resulting reaction intermediate – a 2'-5' branched molecule called a lariat – is the same as in spliceosome-catalyzed nuclear splicing reactions (Peebles *et al.*, 1986; van der Veen *et al.*, 1986). Tight binding of the 5' exon to this lariat intermediate is necessary for efficient coupling of the two reaction steps (Jacquier and Rosbash, 1986) and was shown to rest on extended canonical base pairing between two intron-contained exon-binding sequences called EBS1 and EBS2 and the last 12 or so nucleotides of the 5' exon (Jacquier and Michel, 1987). Pairing of the EBS sequences with the intron-binding (IBS) sections of the 5' exon prior to the first transesterification event also plays a determinant part in selecting the correct 5' splice site (Jacquier and Jacquesson-Breuleux, 1991).

In addition to directing and catalyzing group II self-splicing reactions, the ribozyme component of group II introns participates in the retrotransposition process by which group II introns insert themselves into either intron-less genomic copies of their original insertion site or novel, ectopic sites. Retrotransposition is initiated by (protein-assisted) reverse splicing of the excised intron directly into its DNA target and entails pairing of the sense DNA strand with the EBS sites (Zimmerly *et al.*, 1995; Eskes *et al.*, 1997). Thus, the same intron-contained sequences recognize both insertion sites and 5' splice sites, which makes it possible for newly transposed introns to excise themselves cleanly out of precursor transcripts.

In group I self-splicing introns as well, the 5' exon is recognized by canonical base pairing, but multiple contacts between the ribozyme core and the exon and intron strands of the recognition helix further stabilize the interaction (Pyle and Cech, 1991; Strobel and Cech, 1993; Strobel *et al.*, 1998). Do similar contacts exist in group II introns? Since the base sequence of the EBS-IBS pairings can be changed at will (Jacquier and Michel, 1987; Eskes *et al.*, 1997; Xiang *et al.*, 1998), any interaction with the EBS or IBS bases is unlikely. Moreover, removal of individual ribose 2'OH groups at any but the last and penultimate 5' exon positions is without effect on the  $K_m$  between intron and exon in hydrolysis reactions at the 5' splice site (Griffin *et al.*, 1995). However, the possibility of interactions between the ribozyme core and the intron strands of the EBS-IBS pairings has not been tested. If they were to exist, such contacts could perhaps help to explain why bimolecular setups in which a short

stem-loop structure carrying the EBS1 sequence binds the 5' exon of precursor molecules that lack the EBS1 element are so successful in stimulating splicing (Hetzer *et al.*, 1997).

Any additional contact between one of the six structural domains of the group II ribozyme core and the EBS and IBS sequences might be revealed by direct footprinting of the 5' exon on the intron with base modifying reagents. This approach requires of course the population of intron molecules to be in a uniformly native conformation: fortunately, there now exists at least one group II intron, from the mitochondrial large ribosomal RNA precursor of *Pylaiella littoralis*, which has been shown to meet this requirement by a number of criteria including migration on non-denaturing gels, melting profiles, dimethyl sulfate (DMS) modification and reaction kinetics (Costa *et al.*, 1997b; 1998). We now report DMS probing of the interaction between the PI.LSU/2 intron and its 5' exon. Footprints are observed not only at the two EBS sites, in ribozyme subdomain ID, but at three additional locations. Rather unexpectedly, one of the additional footprints, at position A105 of domain I, is dependent on the presence of the distal helix of the small domain V, which is located in the 3' terminal section of the intron and is believed to be involved in catalysis (reviewed by Michel and Ferat, 1995; Jacquier, 1996; Qin and Pyle, 1998). Removal of the last three base pairs of domain V not only abolishes the A105 footprint, but also increases by about two orders of magnitude the  $K_d$  between the intron and its 5' exon. These findings indicate that contrary to common assumptions, completion of the group II active site is necessary for efficient exon binding. The possible significance of delaying tight exon binding until after the ribozyme component of the intron has been assembled is discussed in the light of the known biochemistry and biology of group II self-splicing.

## Results

### DMS footprinting of the 5' exon on the lariat form of intron PI.LSU/2

Besides the fact that the lariat form of group II introns tends to have greater conformational homogeneity than linear molecules (Jacquier and Jacquesson-Breuleux, 1991; Costa *et al.*, 1997b; 1998), one obvious advantage of using it for structural studies is that it is unreactive by itself. Unfortunately, this advantage is canceled in exon footprinting setups, for incubation of group II lariats with a saturating concentration of their cognate 5' exon results in a partial debranching reaction (Chin and Pyle, 1995; Costa *et al.*, 1997a; Nolte *et al.*, 1998), which rapidly reaches an apparent equilibrium at *ca* 20 % of linear molecules in the case of

intron Pl.LSU/2 (Costa *et al.*, 1998). In order to preserve the integrity of the lariat intron during exon footprinting, we chose to confront Pl.LSU/2 lariat molecules with an inactive analog of the 13-mer 5' exon that we had previously used for debranching experiments (Costa *et al.*, 1998) and to this end, replaced the 3' terminal ribose of the latter molecule by a 3'-deoxy sugar. As expected, incubation of the lariat intron with a concentration (20  $\mu\text{M}$ ) of 3'-deoxy 5' exon analog (3'dE5) that would be largely saturating in the case of the rE5 (3'OH-carrying) 5' exon (Costa *et al.*, 1998, and Fig.1) resulted in no detectable reaction (data not shown). We then checked that the 3'-deoxy molecule acts as a competitive inhibitor of the 5' exon by determining saturation curves for debranching by the rE5 oligonucleotide at different concentrations of 3'dE5 molecules (Fig.1; see Materials and methods for derivation of  $K_d$  from the equilibrium fraction of debranched molecules at different rE5 concentrations). The  $K_I$  for inhibition by 3'dE5 was found to be only slightly higher than the estimated  $K_d$  for binding of the 5' exon to the lariat ( $K_I/K_d = 1.30 \pm 0.15$ ; see legend of Fig. 1). Therefore, the 3'OH group of the 5' exon contributes negligibly to recognition of the exon by the lariat intron and the 3'dE5 molecule should constitute an appropriate substitute for the 5' exon in footprinting experiments.

For exon footprinting, lariat molecules were modified by DMS with or without the 3'dE5 oligonucleotide, which, when present, was at a concentration – 6  $\mu\text{M}$  – which is 15 times higher than the estimated  $K_I$  (0.4  $\mu\text{M}$ ). Modified bases (mostly adenines plus a few cytosines in our experiments; see Fig.3 and Costa *et al.*, 1998) were then identified by gel electrophoresis of primer extension products (Materials and methods). Comparison of extents of modification in presence and absence of 3'dE5 reveals clear footprints (Fig. 2A) at the two known exon-binding sites of group II introns, EBS1 and EBS2 (Fig. 3; Jacquier and Michel, 1987). Extensive protection (more than 80 % on the average) is particularly apparent over all 5 bases (4 A's and one C) of the EBS2 sequence (Fig. 2A; see also Fig. 7A; A204, immediately 5' of EBS2, was also protected, even though it cannot form a Watson-Crick base pair with the nucleotide facing it – an A – in the 5' exon). Although it is only weakly accessible in the absence of the exon (Fig. 2A), the only adenine (A258) in the EBS1 sequence of intron Pl.LSU/2 is also visibly protected, as expected. In addition to protections, increased accessibility of some of the surrounding bases is also observed. Enhancement of modification is particularly visible at position C251, 2 nucleotides upstream of EBS1 in the ID3 loop, and somewhat less reproducibly so, at positions A210 to A212, 3' of the EBS2 sequence (Fig. 2A).

While footprinting of the EBS1 and EBS2 sequences was anything but unexpected, we also observed fully reproducible and nearly complete protection (Fig. 2C) of two bases, A105 and A122, in another domain I subdomain (IC, Fig. 3), which was not known to contact the 5' exon. In the rest of the intron, however, no footprint could be detected, with one exception: enhanced modification in the presence of 3'dE5 was reproducibly observed (3 modification experiments on 2 different lariat preparations) at position A409, in the segment (J2-3) joining domains II and III (Figs. 2B and 3).

### **Domain V is required for efficient binding of the 5' exon by domain I**

Since all sites of protection from DMS modification by the 5' exon are located within domain I, an RNA transcript composed of domain I alone could have been expected to bind efficiently the 3'dE5 oligonucleotide. However, repeated DMS modification with and without 6  $\mu$ M 3'dE5 on a molecule (dI) extending from positions G1 to G343 (Materials and methods) resulted in essentially no detectable footprint (Table I and data not shown).

One trivial reason for our failure to detect exon protection on dI alone could have been the inability of the dI transcript to form a stable three-dimensional structure. However, even though more adenines are accessible in the dI molecule than in domain I of the complete intron (see below), bases A270, A271 and A274 were observed to be protected to the same extent as in the lariat, consistent with the possibility that the  $\alpha$ - $\alpha'$  tertiary interaction, which is internal to domain I, had been preserved. Ultimate proof of the presence of the  $\alpha$ - $\alpha'$  pairing in dI transcripts was obtained by comparing the UV absorbance melting profiles of molecules carrying base substitutions in the  $\alpha$  and  $\alpha'$  segments. As seen in Fig. 4A, the melting profile of dI molecules with a wild-type sequence reveals an early unfolding transition with a  $T_m$  of *ca* 45.7 °C at 2 mM magnesium. Changing the sequence of either  $\alpha$  or  $\alpha'$  abolishes this transition, which is, however, restored by combining the two mutant sequences so as to reconstitute a complete  $\alpha$ - $\alpha'$  pairing. Thermal disruption of the  $\alpha$ - $\alpha'$  helix is therefore part of the initial unfolding of dI transcripts. Moreover, this initial transition must involve a number of other interactions, for the calculated enthalpy of melting the  $\alpha$ - $\alpha'$  double helix (57.1 and 59.9 kcal/mole in molecules with a wild-type and reconstituted  $\alpha$ - $\alpha'$  pairing, respectively; see legend of Fig. 4) is much smaller than the total enthalpy of the transition (between 100 and 115 kcal/mole, see Materials and methods). We also investigated the effects of changing the concentration of salts (data not shown) and observed a marked dependence of  $T_m$  on the

concentration of magnesium, just as expected for a transition corresponding to disruption of tertiary structure (reviewed by Draper, 1996).

How much of the intron sequence must be added back to domain I in order to observe a 5' exon footprint? As expected, bringing in domains II and III (in dI-II-III transcripts, see Materials and methods) results in several additional domain I adenines becoming protected from DMS modification. This is notably the case for A90 and A91 (as may be checked by comparing Fig. 5A with Fig. 1 in Costa *et al.*, 1998). Protection of these two bases, which are known to participate in the  $\theta$ - $\theta'$  tertiary interaction between the IC1 terminal loop and the base of domain II (Costa *et al.*, 1997a), indicates that at least one of the interactions connecting domains I, II and III may be preserved in truncated, dI-II-III molecules. Still, essentially no DMS footprint could be detected in the presence of 6  $\mu$ M 3'dE5 (Table I). In contrast, when dI-to-V transcripts extending all the way from G1 to A2402 (save for the deletion of most of the dIV terminal loop of the original intron; Materials and methods) were modified at the same 3'dE5 concentration, the same pattern of footprints as previously seen on the lariat intron was observed (Table I; only A409 did not change state; that base is already somewhat accessible in the absence of the exon in truncated linear intron molecules, which may be affected in the structure of the central group II 'wheel').

Direct evidence that domain V must be present to observe an exon footprint on domain I at 6  $\mu$ M 3'dE5 was obtained by modifying dI-II-III molecules in the presence of both the 3'dE5 oligonucleotide and an RNA transcript (G2364-A2402; Materials and methods) consisting of domain V and the two short segments that connect it to domains IV and VI. Since the  $K_d$  between domain V and domains I-II-III of the yeast  $\alpha 5\gamma$  intron has been reported to be at most 1  $\mu$ M (Pyle and Green, 1994; Jestin *et al.*, 1997), we expected the dV concentration we used (20  $\mu$ M) to be largely saturating. Comparison with modification of dI-II-III alone (Fig. 5) reveals indeed a double footprint – of the 5' exon on one hand, and of domain V on the other. Except again for A409, all sites previously observed to change status when the lariat intron was modified in the presence of the 3'dE5 exon analog responded in the expected way. Aside from these positions, 11 additional bases are protected or, in the case of A307 and A406, unmasked (Figs. 3 and 5) by the simultaneous presence of domain V and the 5' exon and these are the same nucleotides whose accessibility differs between dI-II-III and dI-to-V transcripts, when these molecules are probed in the absence of the exon (data not shown). Not unexpectedly, the adenines footprinted by domain V include three of the four A's of the known  $\zeta$  receptor for the terminal loop of domain V (Costa and Michel, 1995; Costa *et*



*al.*, 1998) and, more generally, constitute a subset of the ones whose modification was shown to interfere with binding of domain V by domains I-II-III of intron  $\alpha 5\gamma$  of *Saccharomyces cerevisiae* (Jestin *et al.*, 1997; the only exception is A338, see Discussion).

We also investigated whether footprinting of domains I-II-III by the 5' exon could be observed in the absence of domain V, at higher concentrations of the 3'dE5 oligonucleotide. DMS modification in the presence of 36  $\mu\text{M}$  and 80  $\mu\text{M}$  3'dE5 (Table I and Figs. 5 and 6) showed that these conditions ensure partial and extensive footprinting, respectively, of the EBS1, EBS2 and A122 sites of dI-II-III molecules by the 5' exon (partial footprinting of the EBS1 and EBS2 sites of dI transcripts is also observed when modification is performed in the presence of high concentrations (110  $\mu\text{M}$ ) of the 3'dE5 exon – data not shown). However, the footprint remains an incomplete one, for A105 stays unaffected (Fig. 6). Therefore, the presence of domain V is required to observe protection at that position. It is important to note that except for A105 and A409, which remain in the same state, and A165, in the  $\zeta$  receptor motif, whose accessibility increases 2 to 3-fold, the adenines and cytosines that respond to addition of the 3'dE5 molecule are the same ones that sense its presence in the lariat intron. Thus, there is no evidence of the 5' exon inducing any major conformational rearrangement in dI-II-III transcripts.

### **Inefficient 5' exon recognition and incomplete footprints upon deletion of the last three base pairs of domain V**

We attempted to determine which part of domain V is required for efficient binding of the 5' exon to domain I by first substituting, and then truncating the distal extremity of the dV hairpin so as to disrupt one after the other the interactions that bind it to the rest of the molecule. As previously reported (Costa *et al.*, 1998), substitution of the dV GAAA terminal loop by UUCG results in a concise footprint in domain I, at the site of the  $\zeta$  receptor for that loop, and the temperature at which the lariat intron begins to unfold is decreased. However, stability can be restored by raising the concentration of magnesium to 20 mM and comparison of wild-type and dV:UUCG mutant molecules under these conditions reveals no significant change in the  $K_d$  of the lariat intron for the 5' exon. As could have been anticipated from these data, we did not observe any difference in the extent and distribution of footprints on the lariat forms of the wild-type and dV:UUCG introns when the two molecules were modified in parallel in the absence and presence of 6  $\mu\text{M}$  3'dE5 (Table I and data not shown).

Removal of three base pairs at the distal end of domain V of the dV:UUCG molecule (see Fig. 3) results in a primary transcript (dV-3bp) that requires at least 20 mM magnesium to react (Materials and methods). Nevertheless, purified dV-3bp lariats appear no less stable than dV:UUCG ones, since the melting profiles of the two molecules are indistinguishable (compare Fig. 4B with Fig. 3D in Costa *et al.*, 1998; in both mutants, the  $T_m$  of the initial unfolding transition is 6 deg.C below that of the wild-type lariat). In fact, these two mutant molecules have identical DMS modification patterns (not shown), except at A165, in the 5' branch of the  $\zeta$  receptor, which is highly accessible in the dV-3bp lariat, whereas it is only barely modified in the dV:UUCG lariat, and not at all so in wild-type molecules. Even though a UUCG loop cannot interact efficiently with a GAAA receptor (Jaeger *et al.*, 1994; Murphy and Cech, 1994), its presence must offer some degree of steric protection from DMS modification to the N1 atom of A165, which is hydrogen-bonded when facing a GAAA loop (Cate *et al.*, 1996).

Somewhat surprisingly in regard to its structural similarity to the dV:UUCG lariat, the dV-3bp lariat is only weakly footprinted when modified in the presence of 6  $\mu$ M 3'dE5 (Table I). But as is the case for dI-II-III transcripts, extensive footprints can be recovered by using a much higher concentration of the 5' exon analog (in this case, 110  $\mu$ M; see Table I). However, the set of footprints obtained under these conditions is still an incomplete one: as was observed with dI-II-III molecules, the A105 base remains accessible upon exon binding (Fig. 7B), whereas it is completely protected in the wild-type and dV:UUCG lariats (another difference is that the change in accessibility of A409 is less pronounced in the dV-3bp lariat).

Confirmation that the 5' exon has a much lower affinity for the dV-3bp lariat compared to the dV:UUCG lariat was provided by kinetic analyses of the debranching reaction of the former molecule (Fig. 8, Materials and methods and Table II). The  $K_d$  (26  $\mu$ M) is about a 100-fold larger than for wild-type molecules. In addition, there is a marked effect on the catalytic rate constant, which can be estimated to be about 300-fold smaller than the wild-type one by extrapolating the latter to pH 7.0 (see Table II; as already observed for the wild-type and dV:UUCG lariats – Costa *et al.*, 1998 – the final fraction of dV-3bp debranched molecules is independent of the concentration of magnesium and there is only a slight increase in the rate of the debranching reaction between 20 and 50 mM  $MgCl_2$ ).

Finally, we attempted to identify a possible exon partner of the N1 atom of A105 by footprinting the wild-type lariat intron with two exon oligonucleotides with deoxy substitutions of the riboses at positions -1 (2'-1d) and -2 (-2d). Removal of the 2'OH groups of

these sugars has been shown to affect both  $k_{\text{cat}}$  and  $K_{\text{m}}$  for cleavage of oligonucleotide analogs of the exon by a ribozyme composed of domains I and V (Griffin *et al.*, 1995). Our own kinetic analyses (Table II) of the debranching reactions catalyzed by oligoribonucleotides missing one of the four 2'OH groups at the 3' end of the 5' exon agree with Griffin *et al.*'s in that substitutions at positions -1 and -2 are the only ones to affect  $k_{\text{cat}}$ , yet the effects we observe on  $k_{\text{cat}}$  and  $K_{\text{d}}$  are substantially larger than had been reported for the hydrolysis reaction. Nevertheless, when lariat molecules were modified in the presence of presumably saturating concentrations of either the 2'-1d or -2d exon analogs (see Materials and methods), the same degree of protection was observed at A105 as in the EBS2 segment (Table I). Either the N1 position of that base is not contacted by the last or penultimate 2'OH group of the 5' exon or steric hindrance prevents DMS from accessing it.

## Discussion

### Two modes of binding the 5' exon by molecules derived from a group II intron

It has been known for more than ten years that group II introns bind their 5' exon by classical pairing between the last twelve or so bases of the exon and two sequence stretches (EBS1 and EBS2) in secondary structure domain I (Jacquier and Michel, 1987). What is novel in the work presented in this paper is that we now report the existence of two modes of binding the 5' exon by molecules derived from a group II intron. 'Tight binding', as observed in the complete intron, requires not merely the EBS sequences and surrounding structures, but also the small domain V, which, together with domain I, constitutes the minimal group II ribozyme, i.e. the smallest structure compatible with some form of catalytic activity (Koch *et al.*, 1992; Michels and Pyle, 1995). In contrast, 'loose binding' is the mode observed in incomplete transcripts that comprise at least the entire domain I, and can acquire a three-dimensional structure under appropriate ionic conditions, but lack a complete domain V. These two modes of binding can be differentiated from one another not merely by the concentration of 5' exon required to observe DMS footprinting of the two EBS sequences, but by protection of one additional site in domain I (A105) when the exon is tightly bound.

Exactly how large is the difference in exon-intron affinity between 'tight' and 'loose' binding? Even though truncated transcripts missing domain V lack any activity by themselves, we can nevertheless answer this question because of the excellent quantitative agreement between kinetically estimated  $K_{\text{d}}$  values for exon recognition by lariat molecules

and the extent of protection we observe at the EBS sites (such agreement between predicted and observed extents of protection from DMS modification by ligands has been reported in other systems as well – e.g. compare Samaha *et al.*, 1994, and Wallace *et al.*, 1974). Thus, the concentration (6  $\mu\text{M}$ ) of 5' exon analog which ensured essentially complete protection of the EBS sites in the wild-type lariat is well within the range expected from the estimated  $K_I$  (0.4  $\mu\text{M}$ ) of this analog in lariat-debranching reactions. For the dV-3bp lariat, rather little (less than 25 %) protection was observed using the same 6  $\mu\text{M}$  concentration of 5' exon analog, whereas a nearly (about 80 %) complete EBS footprint was obtained at 110  $\mu\text{M}$  3'dE5 (Fig. 7A and Table I). These observations hinted at a  $K_d$  of 25 to 30  $\mu\text{M}$ , an estimate which fits well the kinetically determined  $K_d$  for debranching of the dV-3bp lariat (*ca* 26  $\mu\text{M}$  for the all-ribo 5' exon and, by extrapolation, about 35  $\mu\text{M}$  for its 3'dE5 analog; see Table II and Fig. 1). By taking advantage of this correlation, the  $K_d$  for the complex between dI-II-III transcripts and the 3'dE5 oligonucleotide may be estimated in turn to be about 40  $\mu\text{M}$ , since this is the range of 3'dE5 concentration that results in *ca* 50 % protection of the EBS2 site (Table I). Taken together, these calculations indicate that there is an about 100-fold difference in affinity for the 5' exon between molecules that do and do not include a complete domain V structure.

### **Physical significance of elevated exon dissociation constants**

A necessary condition for any estimate of  $K_d$ , whether based on kinetic measurements or quantitation of protection from chemical modification, to reflect the intrinsic mutual affinity of two molecules is that the two interacting partners possess sufficient conformational homogeneity. Our choice of the Pl.LSU/2 molecule was based on our previous demonstration (Costa *et al.*, 1997b; 1998) that the lariat form of this intron not only forms populations of molecules with a uniform native conformation by a number of different criteria, but does so at typically low magnesium concentrations compared to the ones required to observe optimum activity of other group II introns. At that time the question remained open of whether individual domains or severely mutated forms of the Pl.LSU/2 intron would also be able to fold autonomously in salt solutions. However, a number of observations now indicate that the truncated molecules we have been utilizing throughout this work do have a well-defined and largely native higher-order structure under the conditions used for footprinting experiments and kinetic analyses.

*Domain I.* In the rather closely related yeast  $\alpha 5\gamma$  intron, domain I has been proposed to constitute an independent folding unit based on indirect observations – namely, its ability to bind 5' exon oligonucleotide analogs and the inaccessibility to chemical modification of some bases known to be involved in tertiary interactions (Qin and Pyle, 1997). However the magnesium concentration required to reportedly fold half of the molecules was extremely high (78 mM). In contrast, we now provide direct evidence in the form of melting profiles that an interaction ( $\alpha$ - $\alpha'$ ; Jacquier and Michel, 1987; Harris-Kerr *et al.*, 1993) which is diagnostic for domain I higher-order structure, is present in a substantial fraction of Pl.LSU/2 dI molecules at magnesium concentrations as low as 2 mM. Moreover, parallel probing of dI and lariat molecules with DMS under the conditions used for 5' exon footprinting revealed no detectable difference in the extent of protection of adenines in the  $\alpha$  and  $\alpha'$  segments, nor in fact in any of the secondary structure helices (unpublished results). Therefore, it seems ruled out that gross, global misfolding of dI transcripts could be responsible for our inability to observe a 5' exon footprint at 6  $\mu$ M.

*Domains I-II-III and the footprint of domain V.* In the case of dI-II-III transcripts, evidence of correct folding can be gathered by comparing the autoradiograph in Fig. 5 with the one in Fig. 1 of Costa *et al.* (1998) and also with the DMS accessibility map of the Pl.LSU/2 lariat molecule in the native state (Fig. 3). Complete intron molecules and dI-II-III transcripts have very similar DMS modification patterns over the entire length of domains I and III, except at a few specific sites, which are the ones that change status when modification of dI-II-III transcripts is carried out in the presence of a presumably saturating concentration of a domain V transcript (Fig. 5). It is particularly revealing to compare this concise footprint, which indicates that the structure of domains I and III is largely preformed for recognition of domain V, with the one obtained by Konforti and colleagues using the yeast  $\alpha 5\gamma$  intron (Fig. 1 of Qin and Pyle, 1998). Despite the extensive structural similarity of the two molecules, there are many more bases whose DMS accessibility is affected by presence of domain V in the  $\alpha 5\gamma$  intron (22 instead of 7 for domain I). Even more striking, the residues reportedly footprinted by domain V are dispersed over more than half of domain I secondary structure components, rather than mostly clustered within, and in the immediate vicinity of, the  $\zeta$  receptor for the dV GAAA terminal loop (Costa and Michel, 1995), as we observe. In contrast, and as was to be expected from the general superiority of chemical interference over direct probing when working with poorly, instably and/or inhomogeneously folded molecules, there is excellent agreement between the map published by Jestin *et al.* (1997) of  $\alpha 5\gamma$  domain I residues whose

modification interferes with binding of domain V and our own footprinting of domain V on domain I. Apart from the  $\zeta$  receptor, only the three well-conserved adenines (A171, A181 and A182 in Pl.LSU/2) in the 3-way junction immediately distal of the  $\zeta$  motif appear to be both important for binding dV and in close, if not direct, contact with it (somewhat unexpectedly, the highly conserved A338 residue, which we found to give a clear footprint, is missing from the Jestin *et al.* map, which nevertheless includes a thiophosphate group at the counterpart of position 337). In domain III, however, the two maps diverge, inasmuch as the positions at which we observed a dV footprint (A437, A438 and C439 in the IIIA terminal loop) constitute only a small subset of the ones detected in the a5 $\gamma$  intron by interference. It should be interesting to investigate further the structural role of the IIIA loop, the sequence of which is well conserved in many organellar and bacterial subgroup IIB introns.

*The dV-3bp lariat.* DMS modification can be invoked to rule out misdocking of domain V as a possible source of the decreased affinity of the dV-3bp lariat for the 5' exon. Of the 8 adenines protected by domain V in a wild-type context, only the ones in the  $\zeta$  motif (A165 and A305) are accessible in the dV-3bp lariat, which lacks of course the GAAA counterpart of this tetraloop receptor. Otherwise, the DMS modification pattern of the dV-3bp molecule is strictly identical, in the absence of the 5' exon, to that of the complete intron, and this includes the dV adenine at position 2390 (see Fig. 3), whose protection was previously shown to be indicative of the interaction of domain V with the rest of the molecule (Costa *et al.*, 1998). Therefore, even though truncated, domain V appears to retain an authentic structure and dock correctly into the rest of the intron in this mutant lariat.

### **Structural correlates of high-affinity exon binding**

As previously reported for the a5 $\gamma$  intron (Qin and Pyle, 1997), truncated Pl.LSU/2 molecules can bind the 5' exon as long as they include an intact domain I. Exon binding through the formation of canonical base pairs between the EBS and IBS sequences manifests itself of course by substantial protection from DMS modification of adenines and cytosines within the two EBS elements. At the same time, several of the surrounding bases become more accessible, which probably reflects some degree of conformational rearrangement of the terminal and internal loops that carry the EBS1 and EBS2 sequences upon formation of the EBS-IBS base pairs. Somewhat less expected was footprinting of a much more distant position, A122, which is located in another subdomain of domain I. This base is rather well conserved in close intron relatives of the Pl.LSU/2 molecule (F.M., unpublished

observations). Whether or not it directly contacts the 5' exon, its protection from DMS modification upon exon binding provides further evidence that the higher-order structure of domain I is very similar indeed in complete and truncated molecules.

Tighter exon binding, as observed in setups that include a complete domain V results in one additional site in domain I (A105) becoming protected from DMS methylation. Both A105 and A409 (whose accessibility is enhanced by exon binding; Fig. 2) are highly (about 95 %) conserved bases in group II introns (Michel *et al.*, 1989; F.M., unpublished data), and both of them were already known to lie close to exon-intron junctions: the nucleotide following A105 base pairs with intron position 4, while A407 binds the last intron nucleotide (Jacquier and Michel, 1990). Moreover, A409 itself has been shown to crosslink to the base of domain V, which is widely believed to be part of the active site of group II introns (Podar *et al.*, 1998).

Protection of A105 requires both the 5' exon and domain V to be present. While we ignore which section of the 3'dE5 exon analog (or the intron segments that bind it) is implicated in protecting the A105 base from DMS methylation, we have been fortunate enough to identify the part of domain V that is responsible for this particular footprint. Removal of the last three base pairs of domain V has no measurable effect on the overall thermal stability of the lariat, nor, in the absence of the exon, any detectable influence on the DMS modification pattern of that molecule except at position A165 in the  $\zeta$  receptor motif (in transcripts lacking domain V, this base becomes significantly more accessible at saturating concentrations of the 3'dE5 molecule (see Results), which suggests that it could be involved in the cooperative binding of domain V and the 5' exon by domain I). Nonetheless, the loss of this small segment is sufficient to throw the lariat intron into a loose exon-binding mode and simultaneously abolish the A105 footprint even at saturating exon concentrations, just as is seen in transcripts lacking the entire domain V. If not directly responsible for protecting the A105 base when the exon is bound, the distal helix of domain V would appear to promote a local conformational rearrangement which involves a nucleotide, A105, that is very close to, if not part of, the group II active site, and which also results in the  $K_d$  between exon and intron becoming more favorable by about two orders of magnitude. Failure to make this rearrangement in the ground state could be responsible for the large catalytic penalty of the dV-3bp lariat: while the rate of the dV-3bp debranching reaction is just as highly dependent on pH as seen for the wild-type molecule (Table 2 and our own unpublished results), which is consistent of course with chemistry being rate-limiting, this does not preclude  $k_{cat}$  from

including a conformational term, provided the corresponding equilibrium is rapid enough for relaxation to be faster than reaction.

### **Comparison with other group II introns and possible biological implications**

Our finding that Pl.LSU/2-derived molecules lacking a complete domain V have an about 100-fold decreased affinity for their 5' exon might be regarded as a surprise, since Qin and Pyle (1997) reported that neither domain III, nor domain V contribute any energy to binding of an exon substrate by domain I of the yeast  $\alpha 5\gamma$  intron. Pl.LSU/2 and  $\alpha 5\gamma$  are of course distinct molecules, yet they belong to the same subgroup of closely related group II introns, they share all currently known group II tertiary pairings (except for the apparent absence of the  $\beta$ - $\beta'$  interaction in the former molecule) and also have very similar sequences in those secondary structure segments that are believed to form the group II catalytic core. Still, the two experimental systems present a number of important differences that might account for the apparent contradiction between our and Qin and Pyle's conclusions.

First of all, the substrate used by Qin and Pyle differs from ours in that it extends beyond the intron-exon junction, resulting in a 5-nucleotide overlap with the sequence of their ribozyme (which begins at the first intron nucleotide), so that the nucleotides forming the 5' branch of the  $\varepsilon$ - $\varepsilon'$  helix happen to be duplicated. Then, the  $K_d$ 's they report for intron-exon dissociation – 4 to 20 nM – are clearly incompatible with our own estimates. The reason is, that even though the  $\alpha 5\gamma$  intron is unusually (A+U)-rich, its EBS-IBS pairings are not only longer (13 bp instead of 11 bp), but far richer in G:C base pairs (6 instead of 1) than those of the Pl.LSU/2 molecule. When published enthalpies and entropies for nearest-neighbor interactions between base pairs (Turner *et al.*, 1988) are used to obtain an estimate of the difference in intron-exon binding energy between the two systems, the EBS-IBS pairings of the Pl.LSU/2 molecule are found to be globally weaker by some 7 kcal/mol, which would predict an about  $10^5$ -fold gap in  $K_d$ , instead of the observed 10- to 60-fold difference. The discrepancy between calculated and experimental values is of course much smaller when truncated transcripts are compared, which suggests one possible explanation for the apparent contradiction between our and Qin and Pyle's data: the ground state of the  $\alpha 5\gamma$  intron might be missing the rather subtle structural rearrangement that, in the Pl.LSU/2 molecule, is responsible for tight exon binding. In this respect, it is significant to note that the counterpart of A105 in the  $\alpha 5\gamma$  molecule (A115) is not footprinted by domain V (Fig. 1 in Qin and Pyle, 1998), despite the presence of a covalently attached exon. However, our suggestion of a close



contact between domain V and the 5' exon (or sequences binding it) which might enable the intron catalytic center to 'sense' the presence of the exon and, possibly also, the quality of intron-exon pairings, is not without precedent. In fact, Xiang *et al.* (1998) recently reported that not only do some mismatches in the EBS-IBS pairings of intron  $\alpha 5\gamma$  have unexpectedly large effects on the  $k_{cat}$  for hydrolysis, but at least two of them result in weakening the binding of domain V to the rest of the intron.

What might be the motive(s) for possibly delaying exon assembly until domain V has been synthesized and the intron active center completed? Extrapolation of *in vitro* observations to *in vivo* situations is a dangerous exercise and evolutionary inferences are even riskier. Still, we may at least speculate on what could possibly happen if we were to reinforce the exon-intron pairings in such a way as to allow them to form irreversibly at an early stage in the synthesis of the intron (as might happen, for instance, if we naively attempted to take advantage of the superior aptitude of the Pl.LSU/2 intron at self-assembly and were to engineer it to target any sequence for possible therapeutic purposes; e.g. discussion in Xiang *et al.*, 1998). Locking of the EBS1 and EBS2 loops into long-range tertiary interactions before synthesis of domain I had been completed would be undoubtedly fatal, for assembly of the missing secondary structure components would then require the nascent RNA thread to pass repeatedly through the newly created loop and form as many knots as there were helical turns left to reach the central group II 'wheel'. Even in molecules whose synthesis had proceeded beyond domain I, premature tight binding of the exon could interfere with the assembly of other tertiary structure components and delay or even prevent the final fold (kinetic traps stabilized by native interactions have recently been reported in the *Tetrahymena* group I intron; Treiber *et al.*, 1998). These speculations may seem far-fetched, but it is worth recalling that they could readily be tested, at least *in vitro*, by replacing all or part of the 10 A-U base pairs of the EBS-IBS helices of the Pl.LSU/2 molecule by G-C pairs.

## **Materials and methods**

### **DNA constructs**

The "wild-type" Pl.LSU/2 intron construct used in this work, which lacks all but 56 nt of the terminal loop of domain IV, is the one described in Costa *et al.* (1997b). 'Mutant' versions of this molecule were obtained by subcloning of PCR fragments and were verified by sequencing the entire length of the insert.

## RNA synthesis and purification

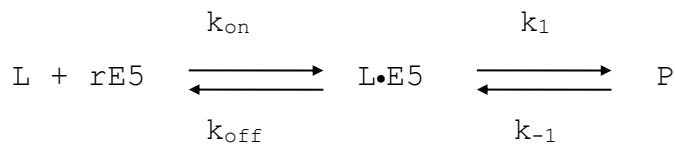
All transcripts were internally labeled by transcription in the presence of [ $\alpha$ - $^{32}$ P] UTP. RNA synthesis and purification were performed as described in Costa *et al.* (1997b). Templates for synthesis of the wild-type, dV:UUCG and dV-3bp precursor transcripts were generated by *Acc65* I digestion of the corresponding plasmids. DNA templates for synthesis of transcripts corresponding to different secondary structure domains or combinations of these were generated by PCR on plasmid constructs followed by digestion with an appropriate restriction enzyme. The following sets of primers were used (the T7 promoter is underlined). Domain V (dV, positions 2364-2402): 5'-GGAGGACTGCAGTAATACGACTCACTATAGTGAGCCGTGTGCGATGA and TTGGGAGGTACCGGATGGGTCTCATCAGAACCGTGCTTGCAC (primer V-rev), digestion by *Bsa*I; domain I (dI, 1-343): 5'-AGCGGTCTAGATAATACGACTCACTATAGTGCGACAAGAAGTTCAGG (primer I-s) and 5'-ATGGGAGGTACCGGTCTCACCAAGTTCCCAGGAAGG, digestion with *Bsa*I; domains I to III (dI-II-III, 1-507): primer I-s and 5'-AGTTGGATAGGTAGACGATCTC, digestion with *Sty*I; domains I to V (dI-to-V, 1-2402): primers I-s and V-rev, digestion with *Bsa*I.

Samples of lariat intron molecules for kinetic analyses, chemical modification and melting experiments were obtained from preparative splicing reactions of precursor transcripts performed under optimal conditions (Costa *et al.*, 1998). Reaction was followed by electrophoresis on 8 M urea/4% polyacrylamide gels and elution of the lariat from the gel (Costa and Michel, 1995). All RNA concentrations were determined from UV absorbance measurements.

## Kinetic analyses

The oligoribonucleotide (rE5) used as 5' exon substitute in lariat debranching reactions has the same sequence as the last 13 residues of the natural PI.LSU/2 5' exon (Costa *et al.*, 1998). Other 5' exon analogs with a 3'H rather than 3'OH terminal group (3'dE5) or with a deoxyribose sugar at one of the last four exon positions (2'-1d, -2d, -3d, -4d) were chemically synthesized on an Applied Biosystems 392 DNA/RNA machine and purified by HPLC on a Nucleopac PA 100-Dionex column (Dionex). Concentrations were determined from UV absorbance measurements, using  $\epsilon_{260}$  values of component nucleotides.

Lariat debranching reactions with 5' exon in excess of ribozyme were set up as described in Costa *et al.* (1998). Unless otherwise stated, conditions were 40 °C, 1 M NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, 0.02 % (w/v) SDS and either 40 mM Na-HEPES pH 7.0 or 40 mM Na-MES pH 5.8. Reactions in presence of the 3'dE5 inhibitor were initiated by mixing at 40 °C equal volumes of (i) a mixture of rE5 and 3'dE5 oligonucleotides at appropriate concentrations in 1X reaction buffer, (ii) renatured (Costa *et al.*, 1998) lariat molecules at a final concentration of 10 nM (final rE5 concentrations ranged from 33 nM to 54 μM). Quantitation of reaction products was as in Costa *et al.* (1998). Kinetic analyses assumed debranching to be a reversible reaction (Chin and Pyle, 1995; Nolte *et al.*, 1998; Costa *et al.*, 1998; L, lariat intron; P, linear intron-5'exon):



Hence the final (equilibrium) fraction of debranched molecules ( $f_{eq}$ ) as a function of the concentration of rE5 ( $f_{eq,max}$  is the final fraction of debranched molecules at saturating exon concentrations,  $K_d = k_{off}/k_{on}$ ):

$$f_{eq} = f_{eq,max} / \{1 + (1 - f_{eq,max}) K_d/[rE5]\} \quad (1)$$

$f_{eq}$  was determined either as indicated in Costa *et al.* (1998), by measuring the molar ratio of debranched over total intron molecules as a function of time and fitting the data to the equation for a monophasic reversible reaction (see also Fig. 8); or, when debranching was too rapid to be followed (e.g. for the wild-type intron at pH 7.0, see Table II), by verifying that prolonged incubation resulted in no further increase in the extent of debranching and taking the average of at least three independent measurements under these conditions.

Reactions in presence of the 3'dE5 exon analog were analyzed by assuming simple competitive inhibition: the 3'dE5 inhibitor, I, is bound by L (dissociation constant  $K_I$ ), but neither by the L•E5 complex, nor by P. Therefore,  $K_d$  is replaced in equation (1) by  $K_{d,I}$  such that

$$K_{d,I} = (1 + I/K_I) K_d \quad (2)$$

In Fig. 1,  $f_{eq}$  can be verified to follow typical saturation behavior with respect to the concentration of the rE5 exon whether or not the 3'dE5 inhibitor was present. Moreover,

$f_{\text{eq,max}}$  does not change significantly in the presence of the inhibitor, which is consistent with the assumption that covalent linkage of the 5' exon to the intron in debranched molecules prevents binding of the free exon or exon analog.]

### **DMS modification**

Typical modification reactions contained 13 nM of RNA destined to be reverse transcribed in 50  $\mu\text{l}$  of 40 mM Na-HEPES (pH 7.6 at 37 °C), 1 M  $\text{NH}_4\text{Cl}$ , 0.02 % (w/v) SDS and 50 mM  $\text{MgCl}_2$  (10 mM  $\text{MgCl}_2$  for most experiments on the wild-type lariat and dI-V transcript). Probing was initiated by addition of DMS (in ethanol) to a final dilution of 1:400 (see figure legends for duration of the modification step). Reactions were performed at 37 °C and stopped by addition of 60  $\mu\text{l}$  of a mixture of 0.6 M  $\beta$ -mercaptoethanol and 0.5 M sodium acetate, 10  $\mu\text{g}$  of bulk yeast tRNA (Boehringer) and 300  $\mu\text{l}$  of ethanol. After precipitation, RNA pellets were washed with 70% ethanol prior to drying. Modified RNAs were resuspended in water at a final concentration of 0.08  $\mu\text{M}$  prior to reverse transcription.

See Text, Table I and Figure legends for concentrations of the non-reactive 3'dE5 exon analog used in exon footprinting experiments (importantly, identical footprinting patterns were obtained using two independent chemical syntheses of the 3'dE5 oligonucleotide). For footprinting with the 2'-1d and -2d exon analogs (see Table I for concentrations), HEPES was replaced by 40 mM Na-MES (pH 6.2 at 37 °C) in order to minimize debranching during incubation. In order to ensure that populations of molecules being probed were folded uniformly and correctly, RNA samples were renatured by addition of buffer at 55 °C, followed by slow-cooling from 55 °C to 37 °C and a further 15-min incubation at the final temperature (see Costa *et al.*, 1998). For exon footprinting experiments, renaturation of RNA molecules was followed by addition of the appropriate amount of 5' exon analog in 1X-concentrated buffer and incubation was carried on for a further 15 min in order to ensure probing under equilibrium binding conditions (it was checked that extending this incubation step up to 1.5 hours did not change the footprinting pattern). Double footprinting of dI-II-III molecules with 6  $\mu\text{M}$  3'dE5 and 20  $\mu\text{M}$  dV transcript (Fig. 5) included the following steps: (i) renaturation of dI-II-III transcripts by addition of buffer and slow cooling from 46 °C to 37 °C; (ii) addition of the 3'dE5 exon analog in 1X-concentrated buffer, followed by a further 15-min incubation; (iii) addition of dV in 1X-concentrated buffer.

### **Reverse transcription**

Reverse transcription of DMS- (or mock-) treated RNA molecules with (5'-<sup>32</sup>P)-labeled complementary DNA primers and subsequent treatment of the samples were performed as described by Costa *et al.* (1998). After denaturation by heating at 80 °C for 10 min, samples were loaded onto denaturing polyacrylamide gels. Fixed and dried gels were quantitated with a PhosphorImager (Molecular Dynamics). Each RNA sample was fully reverse transcribed with the appropriate set of primers at least once (importantly, the pattern of DMS modification revealed by reverse transcription was verified to be the same for differently located DNA primers).

### **Optical melting curves**

Melting curves were generated from 5 µg of RNA in 630 µl of 40 mM HEPES (pH 7.0 at 37 °C), 1M NH<sub>4</sub>Cl, 0.02 % (w/v) SDS and either 2 mM MgCl<sub>2</sub> (for domain I transcripts) or 10 mM MgCl<sub>2</sub> (for lariat molecules). Absorbance at 260 nm was monitored on a Uvikon 933 spectrophotometer (Kontron) equipped with two multiple cell holders containing the RNA samples to be compared. The two holders were simultaneously thermostated by a circulating water bath coupled to a thermoprogrammer. RNA samples in water were renatured by addition of buffer at 55 °C (a temperature slightly above the melting range for tertiary structure), followed by slow-cooling to 20 °C in the melting apparatus. For melting experiments proper, data were collected every 0.05 deg.C from 20 to 80 °C while samples were being heated at a rate of 0.2 deg. C/min. Comparison with data collected during the previous cooling phase made it possible to ascertain that melting of tertiary structure was under equilibrium conditions. Enthalpies of transitions were calculated from the slope at T<sub>m</sub> (e.g. Puglisi and Tinoco, 1989) after deconvolution.

### **Acknowledgements**

We are grateful to Luc Jaeger for his help in oligonucleotide syntheses; to Alain Jacquier and Alexis Nolte for critical reading of the manuscript; and especially to Harry Noller, in whose laboratory the experiments were completed. M.C. was supported by a fellowship from the Fondation pour la Recherche Médicale.

## References

- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
- Chin, K. and Pyle, A.M. (1995) Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA*, **1**, 391-406.
- Costa, M., Christian, E.L. and Michel, F. (1998) Differential chemical probing of a group II self-splicing intron identifies bases involved in tertiary interactions and supports an alternative secondary structure model of domain V. *RNA*, **4**, 1055-1068.
- Costa, M., Deme, E., Jacquier, A. and Michel, F. (1997a) Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.*, **267**, 520-536.
- Costa, M., Fontaine, J.M., Loiseaux-de Goer, S. and Michel, F. (1997b) A group II self-splicing intron from the brown alga *Pylaiella littoralis* is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.*, **274**, 353-64.
- Costa, M. and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276-1285.
- Draper, D.E. (1996) Strategies for RNA folding. *Trends Biochem. Sci.*, **21**, 145-149.
- Eskes, R., Yang, J., Lambowitz, A.M. and Perlman, P.S. (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell*, **88**, 865-874.
- Fersht, A. (1977) *Enzyme structure and mechanism*. W.H. Freeman and Company, New York.
- Griffin, E.A.J., Qin, Z., Michels, W.J.J. and Pyle, A.M. (1995) Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chemistry & Biology*, **2**, 761-770.
- Harris-Kerr, C.L., Zhang, M. and Peebles, C.L. (1993) The phylogenetically predicted base-pairing interaction between alpha and alpha' is required for group II splicing *in vitro*. *Proc. Natl. Acad. Sci. U S A*, **90**, 10658-10662.
- Hetzer, M., Wurzer, G., Schweyen, R.J. and Mueller, M.W. (1997) Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature*, **386**, 417-420.

Inoue,T. and Cech,T.R. (1985) Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc. Natl. Acad. Sci. U S A*, **82**, 648-652.

Jacquier,A. (1996) Group II introns: elaborate ribozymes. *Biochimie*, **78**, 474-487.

Jacquier,A. and Jacquesson-Breuleux,N. (1991) Splice site selection and role of the lariat in a group II intron. *J. Mol. Biol.*, **219**, 415-428.

Jacquier,A. and Michel,F. (1987) Multiple exon-binding sites in class II self-splicing introns. *Cell*, **50**, 17-29.

Jacquier,A. and Michel,F. (1990) Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.*, **213**, 437-447.

Jacquier,A. and Rosbash,M. (1986) Efficient *trans*-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5' exon-intron interaction. *Science*, **234**, 1099-1104.

Jaeger,L., Michel,F. and Westhof,E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.*, **236**, 1271-1276.

Jestin,J.L., Deme,E. and Jacquier,A. (1997) Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.*, **16**, 2945-2954.

Koch,J.L., Boulanger,S.C., Dib-Hajj,S.D., Hebbar,S.K. and Perlman,P.S. (1992) Group II introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.*, **12**, 1950-1958.

Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435-461.

Michel, F., Umesono, K. and Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns -- a review. *Gene*, **82**, 5-30.

Michels,W.J.,Jr. and Pyle,A.M. (1995) Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry*, **34**, 2965-2977.

Murphy,F.L. and Cech,T.R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.*, **236**, 49-63.

Nolte,A., Chanfreau,G. and Jacquier,A. (1998) Influence of substrate structure on *in vitro* ribozyme activity of a group II intron. *RNA*, **4**, 694-708.

Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213-223.

- Podar,M., Zhuo,J., Zhang,M., Franzen,J.S., Perlman,P.S. and Peebles,C.L. (1998) Domain 5 binds near a highly conserved dinucleotide in the joiner linking domains 2 and 3 of a group II intron. *RNA*, **4**, 151-166.
- Puglisi,J.D. and Tinoco,I.,Jr. (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, **180**, 304-25.
- Pyle,A.M. and Cech,T.R. (1991) Ribozyme recognition of RNA by tertiary interactions with specific ribose 2'-OH groups. *Nature*, **350**, 628-631.
- Pyle,A.M. and Green,J.B. (1994) Building a kinetic framework for group II intron ribozyme activity: quantitation of interdomain binding and reaction rate. *Biochemistry*, **33**, 2716-2725.
- Qin,P.Z. and Pyle,A.M. (1997) Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg<sup>2+</sup> ions in the tertiary structure. *Biochemistry*, **36**, 4718-4730.
- Qin,P.Z. and Pyle,A.M. (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.*, **8**, 301-308.
- Samaha,R.R., O'Brien,B., O'Brien,T.W. and Noller,H.F. (1994) Independent *in vitro* assembly of a ribonucleoprotein particle containing the 3' domain of 16S rRNA. *Proc. Natl. Acad. Sci. U S A*, **91**, 7884-7888.
- Strobel,S.A. and Cech,T.R. (1993) Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the *Tetrahymena* ribozyme. *Biochemistry*, **32**, 13593-13604.
- Strobel,S.A., Ortoleva-Donnelly,L., Ryder,S.P., Cate,J.H. and Moncoeur,E. (1998) Complementary sets of non-canonical base pairs mediate RNA helix packing in the group I active site. *Nature Struct. Biol.*, **5**, 60-66.
- Treiber,D.K., Rook,M.S., Zarrinkar,P.P. and Williamson,J.R. (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science*, **279**, 1943-1946.
- Turner,D.H., Sugimoto,N. and Freier,S.M. (1988) RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 167-192.
- van der Veen,R., Arnberg,A.C., van der Horst,G., Bonen,L., Tabak,H.F. and Grivell,L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell*, **44**, 225-234.
- Wallace,B.J., Tai,P.C. and Davis,B.D. (1974) Selective inhibition of initiating ribosomes by spectinomycin. *Proc. Natl. Acad. Sci. U S A*, **71**, 1634-1638.



Xiang,Q., Qin,P.Z., Michels,W.J., Freeland,K. and Pyle,A.M. (1998) Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry*, **37**, 3839-3849.

Zimmerly,S., Guo,H., Eskes,R., Yang,J., Perlman,P.S. and Lambowitz,A.M. (1995) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*, **83**, 529-538.

## Figure Legends

**Figure 1.** Example of competitive inhibition of the PI.LSU/2 lariat-debranching reaction by a 5' exon analog lacking a terminal 3'OH. Eadie-Hofstee plot of the final fraction of debranched molecules ( $f_{eq}$ ) at different concentrations of a 13-mer 5' exon oligoribonucleotide (rE5) in absence (open symbols) or presence (filled symbols) of 0.2  $\mu\text{M}$  of a 3'dE5 exon analog lacking a 3'OH terminal group (see Materials and methods for experimental details). Each point corresponds to the average of three measurements. Values of  $K_d$  ( $0.305 \pm 0.015 \mu\text{M}$ ) and  $K_{d,I}$  ( $0.46 \pm 0.035 \mu\text{M}$ ) in absence and presence of 3'dE5, respectively, were determined by fitting the data to the following equation (see Materials and methods):  $f_{eq} = f_{eq,max} - f_{eq} (1 - f_{eq,max}) K_{d,I} / [rE5]$ . Three determinations of  $K_{d,I}$  at different concentrations of 3'dE5 yielded  $K_I/K_d = 1.30 \pm 0.15 \mu\text{M}$ .

**Figure 2.** Examples of mapping by primer extension of 5' exon DMS footprints on the lariat form of intron PI.LSU/2. DMS modification was performed in the presence of 6  $\mu\text{M}$  of the 13-nt 3'dE5 5' exon analog (see Materials and methods). Reactions were carried out at 37 °C in modification buffer (Materials and methods) containing 10 mM  $\text{MgCl}_2$  for 7 min (+) or 15 min (++). Lane (-), mock-treated samples. G, A, T and C are sequencing lanes run on the same gel as the modified samples; they were generated by reverse transcription of the wild-type lariat intron molecule in the presence of each one of the four ddNTPs. (A) Portion of an autoradiograph showing some of the positions within domain I whose DMS reactivity changes in presence of the 3'dE5 oligonucleotide – relevant adenines are indicated by arrows and brackets (recall that primer extension stops one nucleotide before a modified base; Inoue and Cech, 1985). (B) Portion of an autoradiograph illustrating the conformational change observed within the J2-3 segment upon binding of the 5' exon to the intron. The adenine (A409) whose DMS reactivity is enhanced by 3'dE5 binding is indicated by an arrow. (C) Quantitation of DMS methylation over parts of subdomains IC1 and IC2. Thick grey curve, no DMS control; dotted curve, DMS modification without a 5' exon; thin black curve, DMS modification in the presence of the 3'dE5 molecule.

**Figure 3.** Location of 5' exon and domain V DMS footprints on the secondary structure of intron PI.LSU/2. The Figure is redrawn from Fig. 2A in Costa *et al.* (1998), which showed sites of DMS modification on the native form of the lariat intron (heavy arrows, thin arrows and empty arrowheads indicate extensive, intermediate and weak modification, respectively). EBS1-IBS1, EBS2-IBS2, greek letters, curved arrows and boxing indicate sequences that

participate in known tertiary interactions. The insert at far right shows the structure and probing of domain V in the dV-3bp lariat. Circled minus and plus signs indicate bases that become protected (-) or more accessible (+) in the presence of 6  $\mu\text{M}$  of the 3'dE5 13-nt analog of the 5' exon. Boxed symbols indicate bases (other than the ones that respond to presence of the 5' exon) that become protected or more accessible (+) when a linear dI-II-III molecule is probed in the presence of both the 3'dE5 oligonucleotide and 20  $\mu\text{M}$  of a domain V transcript extending from G2364 to A2402 (Fig. 5 and Materials and methods); black squares, protections that depend on the dV GAAA terminal loop (see Costa *et al.*, 1998); 'X', protections that depend on the last 3 bp of dV (note that A105 is protected only when both domain V and the 5' exon are present); asterisks, protections that depend on the rest of dV.

**Figure 4.** Melting profiles of wild-type and base-substituted lariat and domain I molecules. **A.** Derivative melting profiles of wild-type domain I and variants of it carrying base substitutions in the  $\alpha$  and  $\alpha'$  sequences. Insert shows the  $\alpha$ - $\alpha'$  pairing and mutations introduced. Solid curve (—), wild-type (calculated  $\Delta H^\circ = -57.1$  kcal/mol, see Materials and methods); large dashes (— — —),  $\alpha$  (52CGU:GCA) mutant; small dashes (-----):  $\alpha'$  (271ACG:UGC) mutant; dotted curve (.....):  $\alpha$ - $\alpha'$  (52CGU:GCA, 271ACG:UGC) mutant ( $\Delta H^\circ = -59.9$  kcal/mol). Ordinates are normalized to an absorbance of 1 at 30 °C. **B.** Derivative melting profiles of the wild-type (solid curve: —) and dV-3bp (dotted curve: ..... ) lariats. Ordinates as in A.

**Figure 5.** Mapping by primer extension of changes in the DMS modification pattern of domains I and III upon incubation of a dI-II-III transcript with the 3'dE5 exon analog alone or with both the 3'dE5 oligonucleotide and domain V. **A.** Domain I. Modification (at 50 mM  $\text{MgCl}_2$ , see Materials and methods) was performed at 37 °C for 7 min (+) or 15 min (++); lane (-): mock treated RNA. C, T, A, and G are sequencing lanes run on the same gel as the modified samples; they were generated by reverse transcription of the dI-II-III molecule in the presence of each one of the four ddNTPs. Arrows point to positions that are significantly less or more susceptible to modification in the presence of 80  $\mu\text{M}$  of the 3'dE5 exon analog (central lanes) or 6  $\mu\text{M}$  of 3'dE5 and 20  $\mu\text{M}$  of a dV transcript (right lanes; see Materials and methods). Asterisks indicate positions whose modification is specifically dependent on the presence of domain V. Assignment of bands in the top portion of the gel to the positions indicated was confirmed by reverse transcription with primers located further upstream on the intron sequence (see also Fig. 6). **B.** Domain III. The panel shows phosphorimager scans of the following lanes: 80  $\mu\text{M}$  3'dE5 (thin black curve), 6  $\mu\text{M}$  3'dE5 and 20  $\mu\text{M}$  dV (thick black curve), no-DMS control (thick grey curve). Asterisks indicate sites specifically footprinted by

domain V (the DMS modification pattern of dI-II-III transcripts alone – not shown for clarity – was essentially identical to that observed in presence of 80  $\mu\text{M}$  3'dE5).

**Figure 6.** Protection of position A105 from DMS modification requires both the 5' exon and domain V. A gel similar to the one in Fig. 5A was scanned using a PhosphorImager. Black solid curve, dI-II-III alone; dotted curve, dI-II-III with 6  $\mu\text{M}$  3'dE5 and 20  $\mu\text{M}$  dV; thick grey curve, no-DMS control. Arrows point to changes in the modification pattern caused by the presence of both the 3'dE5 oligonucleotide and domain V (see also Fig. 7B; reproducibility was excellent, except for the stops or pauses at positions 112 and 113). Insert: black solid curve, dI-II-III alone; dotted curve, dI-II-III with 80  $\mu\text{M}$  3'dE5.

**Figure 7.** Incomplete 5' exon footprint on a lariat molecule lacking the last three base pairs of domain V. **A.** Quantitation of DMS methylation over the EBS2 sequence and surrounding nucleotides. Top: wild-type lariat; bottom: dV-3bp lariat. Thick grey curve, no DMS control; thin black curve, DMS modification without a 5' exon; dotted curve, DMS modification with 6  $\mu\text{M}$  (wild-type) or 110  $\mu\text{M}$  (dV-3bp) of the 3'dE5 5' exon analog. **B.** Quantitation of DMS modification over parts of subdomains IC1 and IC2. Top: wild-type lariat; bottom: dV-3bp lariat. Curves and conditions as in A; for DMS modification in the absence of exon, scans from two different experiments are shown. Arrows point to reproducible changes in the modification pattern caused by the presence of the 3'dE5 oligonucleotide.

**Figure 8.** Time course of debranching of the dV-3bp lariat at three 5' exon (rE5) concentrations. Data were fitted with the program KaleidaGraph (Abelbeck Software) by assuming a monophasic reversible reaction (e.g. Fersht, 1977), i.e. fraction debranched =  $[k_{\text{deb}}/(k_{\text{deb}} + k_{\text{br}})] \{1 - \exp[-(k_{\text{deb}} + k_{\text{br}})t]\}$ , with  $k_{\text{deb}}$  and  $k_{\text{br}}$  being the rates of debranching and branching, respectively. Squares and dotted line, 3.35  $\mu\text{M}$  rE5; circles and dashed line, 7.5  $\mu\text{M}$  rE5; diamonds and solid line, 20  $\mu\text{M}$  rE5.

**Table I.** 5' exon footprints on full-length and truncated forms of intron PI.LSU/2.

Molecules probed	5' exon <sup>(a)</sup> ( $\mu$ M)	EBS1 <sup>(b)</sup> footprint	EBS2 footprint	A122 footprint	A105 footprint	A409 footprint
'wild-type' lariat	6.0	yes <sup>(c)</sup>	85 % <sup>(d)</sup>	yes	yes	yes
	15 (2'-1d)	n.d. <sup>(e)</sup>	85 %	yes	yes	n.d.
	110 (-2d)	n.d.	85 %	yes	yes	n.d.
dI	6.0	no	20 %	no	no	absent
dI-II-III	6.0	no	5 %	no	no	(f)
	36	n.d.	45 %	partial	partial	
	80	yes	65 %	yes	no	
dI-to-V	6.0	yes	80 %	yes	yes	(f)
dI-II-III (+dV)	6.0	yes	70 %	yes	yes	(f)
dV:UUCG lariat	6.0	yes	90 %	yes	yes	yes
dV-3bp lariat	6.0	no	20 %	no	no	no
	110	yes	80 %	yes	no	weak

(a) Except otherwise stated, the 3'dE5 oligonucleotide was used as 5' exon (Materials and methods)

(b) The EBS1 exon footprint consists of protection of A258 and enhancement of modification at C251

(c) presence/absence of a footprint was assessed by examining at least two (and up to 8) independent experiments

(d) mean extent of protection over A205 to A208 (values are averages of several determinations; experimental error is about  $\pm 10$  %)

(e) n.d.: not determined

(f) A409 is always accessible in dI-to-V and dI-II-III molecules

**Table II.** Kinetic parameters of debranching reactions.

Lariat	5' exon	$K_d$ ( $\mu\text{M}$ ) <sup>(a)</sup>	$k_{\text{cat}}$ ( $\text{min}^{-1}$ ) <sup>(b)</sup>
wild-type	all-ribo <sup>(c)</sup>	$0.25 \pm 0.015$ $0.19 \pm 0.015$	$1.1 \pm 0.05$ (pH 5.8) n.d. <sup>(d)</sup> (pH 7.0)
dV-3bp <sup>(e)</sup>	all-ribo	$26.5 \pm 7.5$	$0.045 \pm 0.01$ (pH 7.0)
wild-type	2'-1d <sup>(c)</sup>	n.d. $2.2 \pm 0.2$	$0.013 \pm 0.003$ (pH 5.8) $0.155 \pm 0.011$ (pH 7.0) <sup>(f)</sup>
wild-type	-2d <sup>(c)</sup>	$14.5 \pm 5.5$	$0.0022 \pm 0.0002$ (pH 5.8)
wild-type	-3d	$1.3 \pm 0.35$	$0.95 \pm 0.07$ (pH 5.8)
wild-type	-4d	$0.85 \pm 0.3$	$1.3 \pm 0.05$ (pH 5.8)

(a) calculated from the final fraction of debranched molecules ( $f_{\text{eq}}$ ; see Materials and methods)

(b)  $k_{\text{cat}}$  is the catalytic constant of the branching reaction ( $k_{\text{br}}$  in legend to Fig.8), calculated from the observed first-order rate of debranching at saturating exon concentrations and the final fraction of debranched molecules ( $k_{\text{cat}} = k_{\text{obs}} \cdot (1 - f_{\text{eq}})$ )

(c) values are averages of multiple independent determinations (on each of two different syntheses in the case of the 2'-1d exon)

(d) n.d.: not determined

(e) debranching of the dV-3bp lariat was carried out at 50 mM  $\text{MgCl}_2$ ;  $k_{\text{cat}}$  was measured at 20  $\mu\text{M}$  and corrected for saturating exon concentrations

(f) for the 2'-1d exon,  $\log(k_{\text{cat}})$  was verified to vary linearly with pH between 5.8 and 7.0 (data not shown); the slope (0.96) is the one expected for chemistry being rate-limiting (e.g. Pyle and Green, 1994)

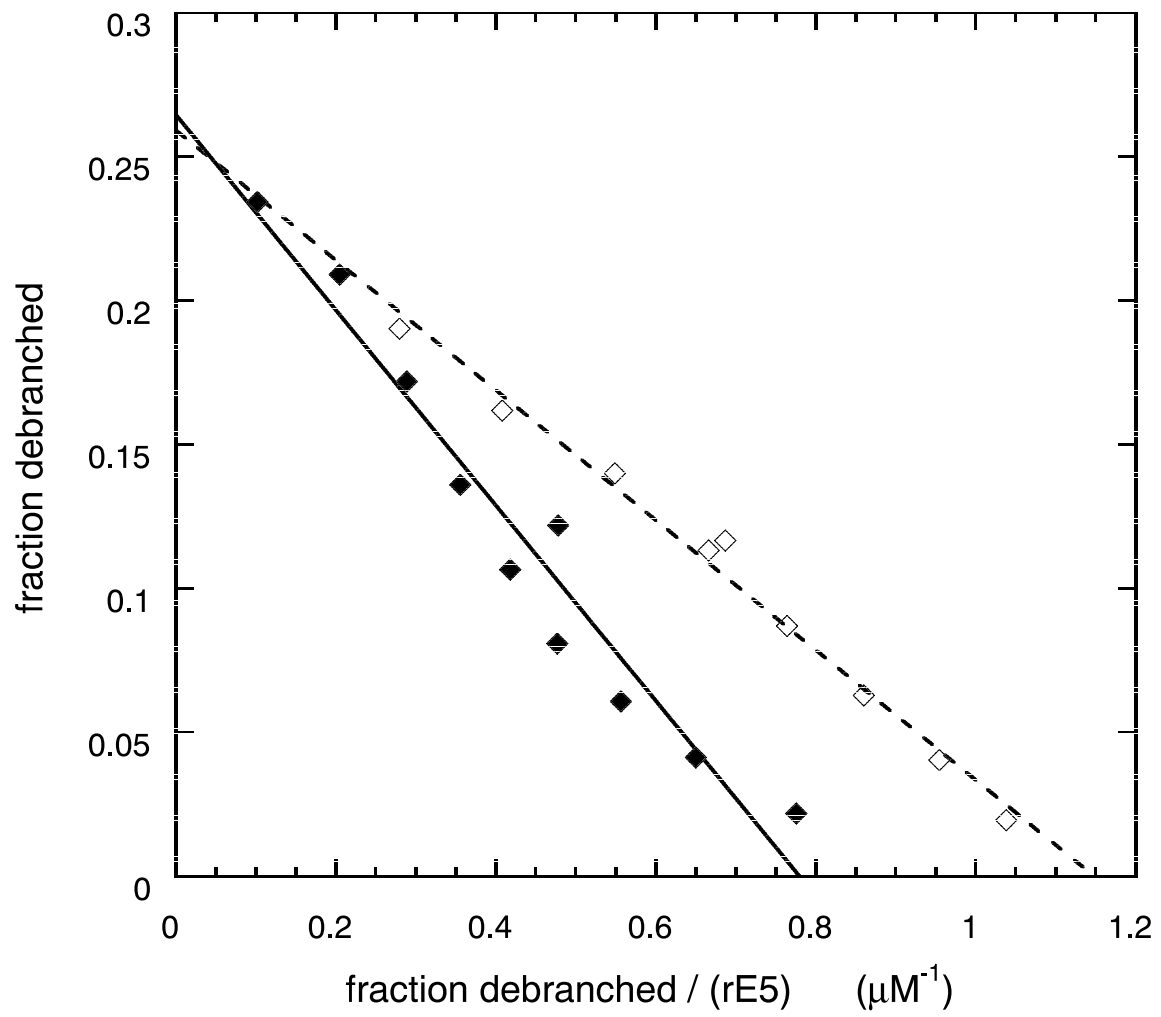
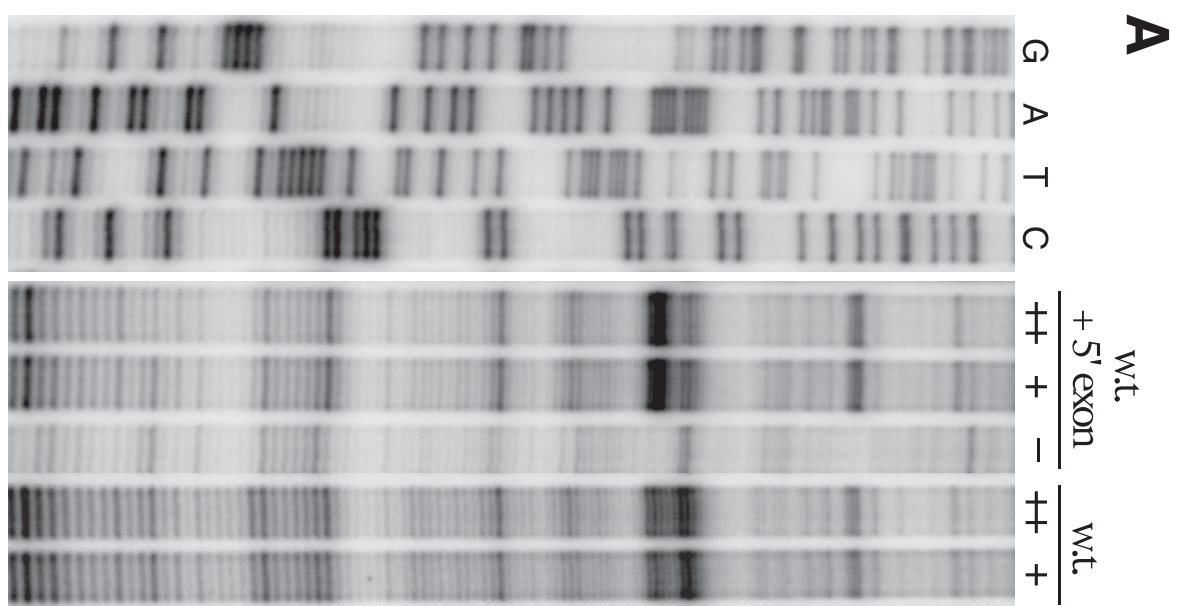
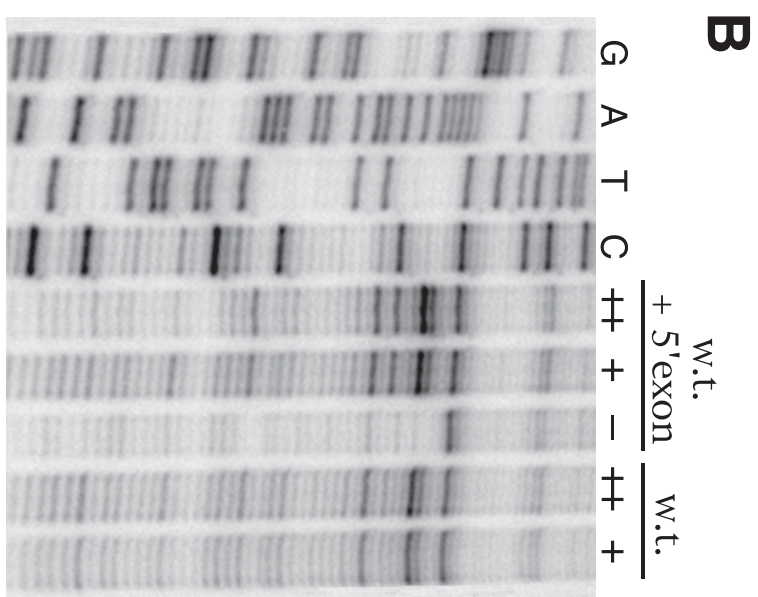


Figure 2 A, B



⇓ EBS2  
⇓ A210-212

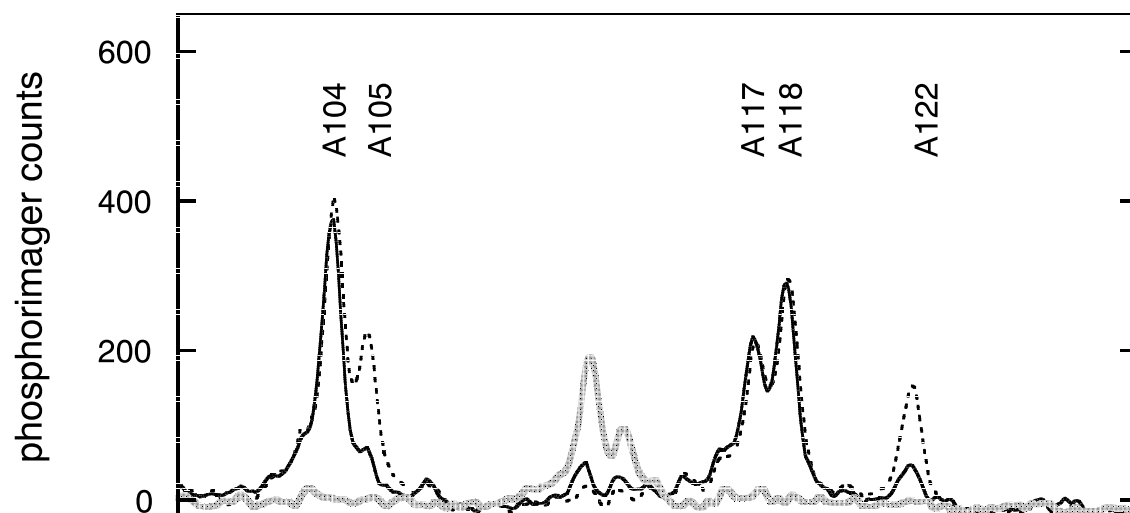
⇓ C251  
⇓ A258

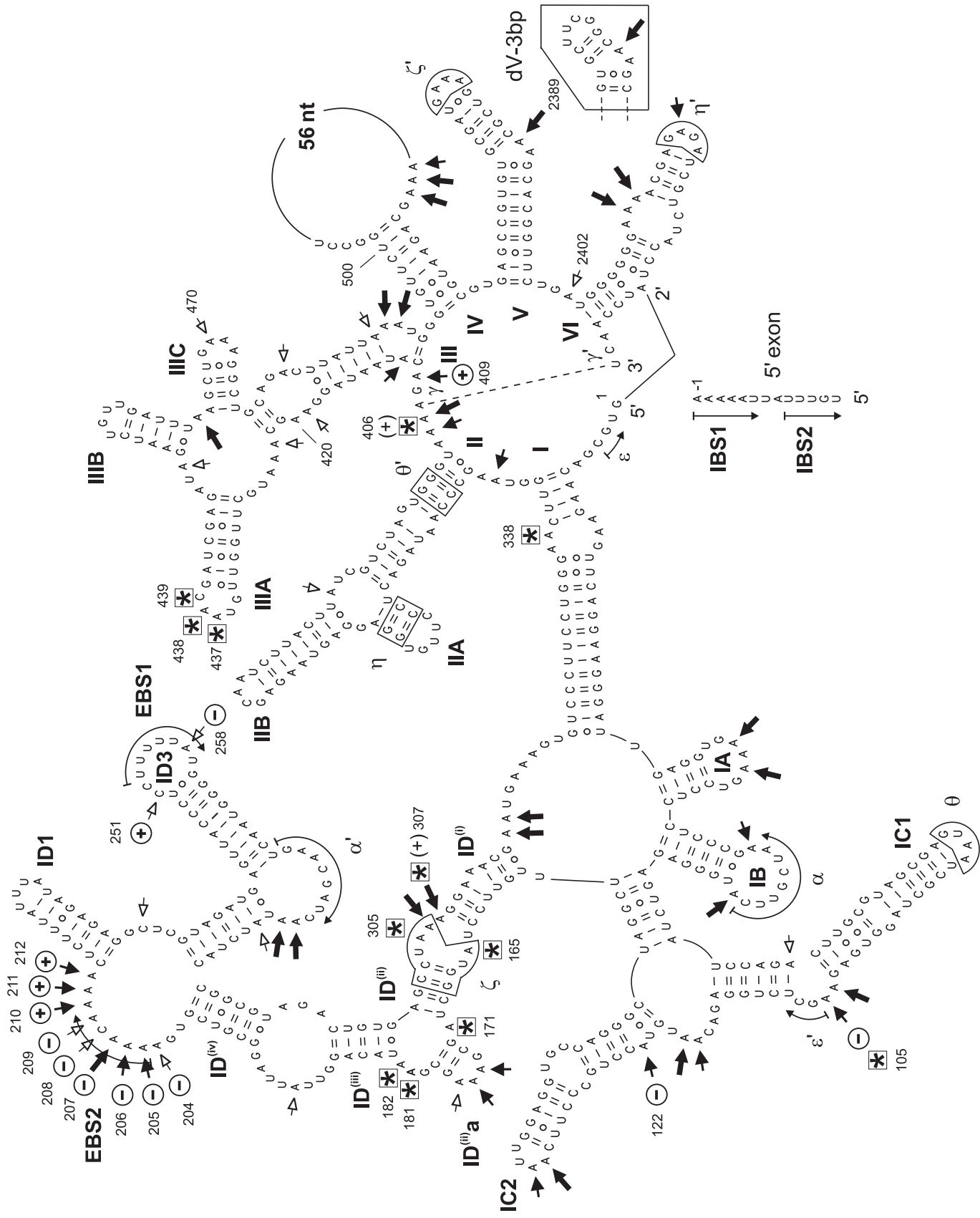


⇓ A409



Figure 2C





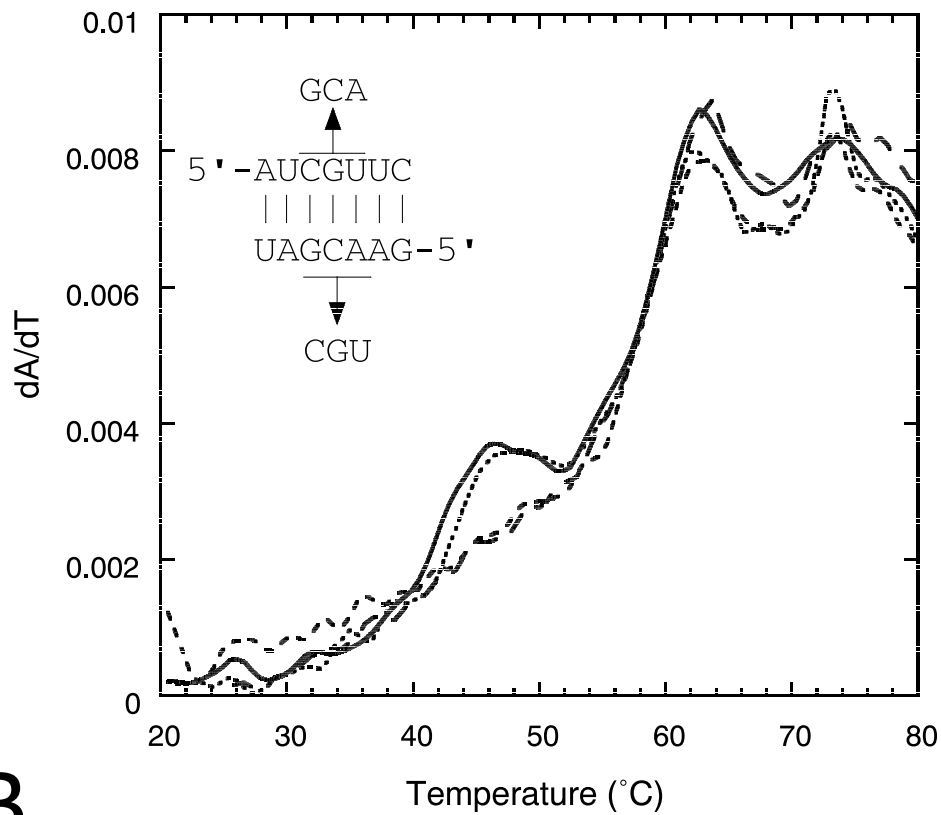
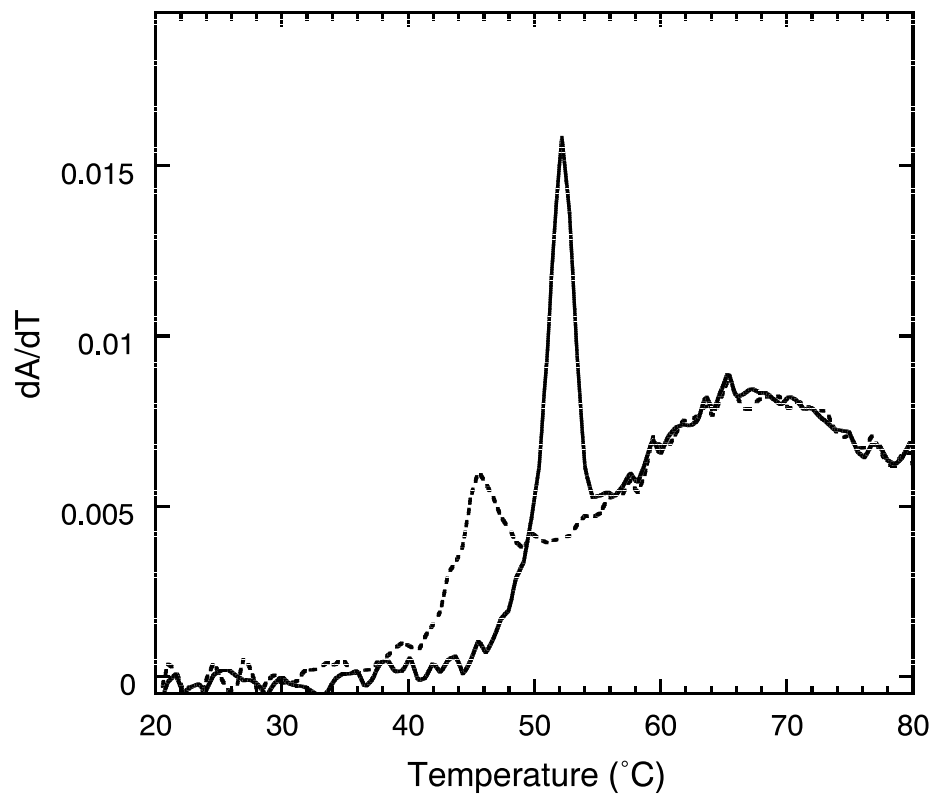
**A****B**

Figure 5A

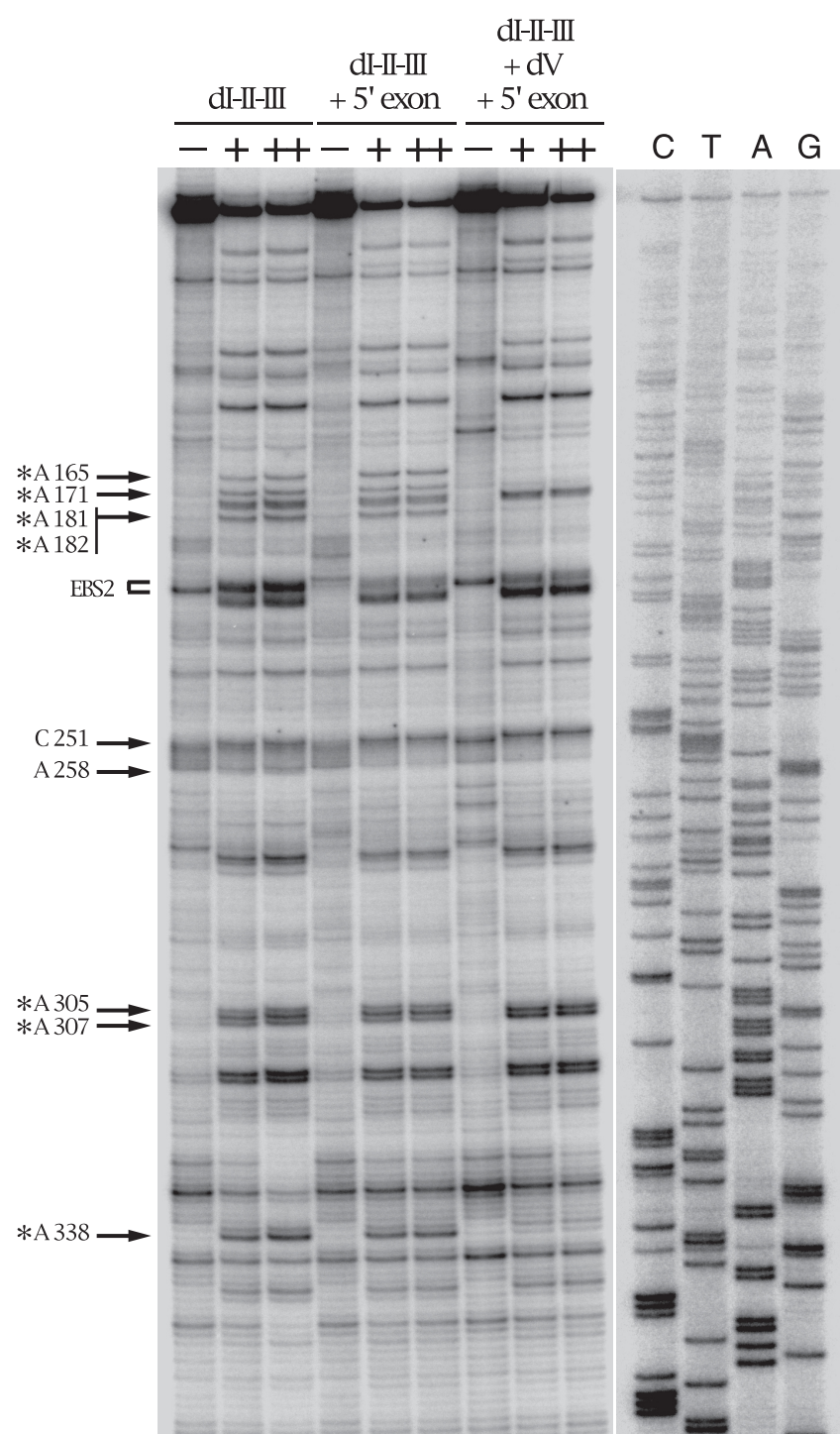
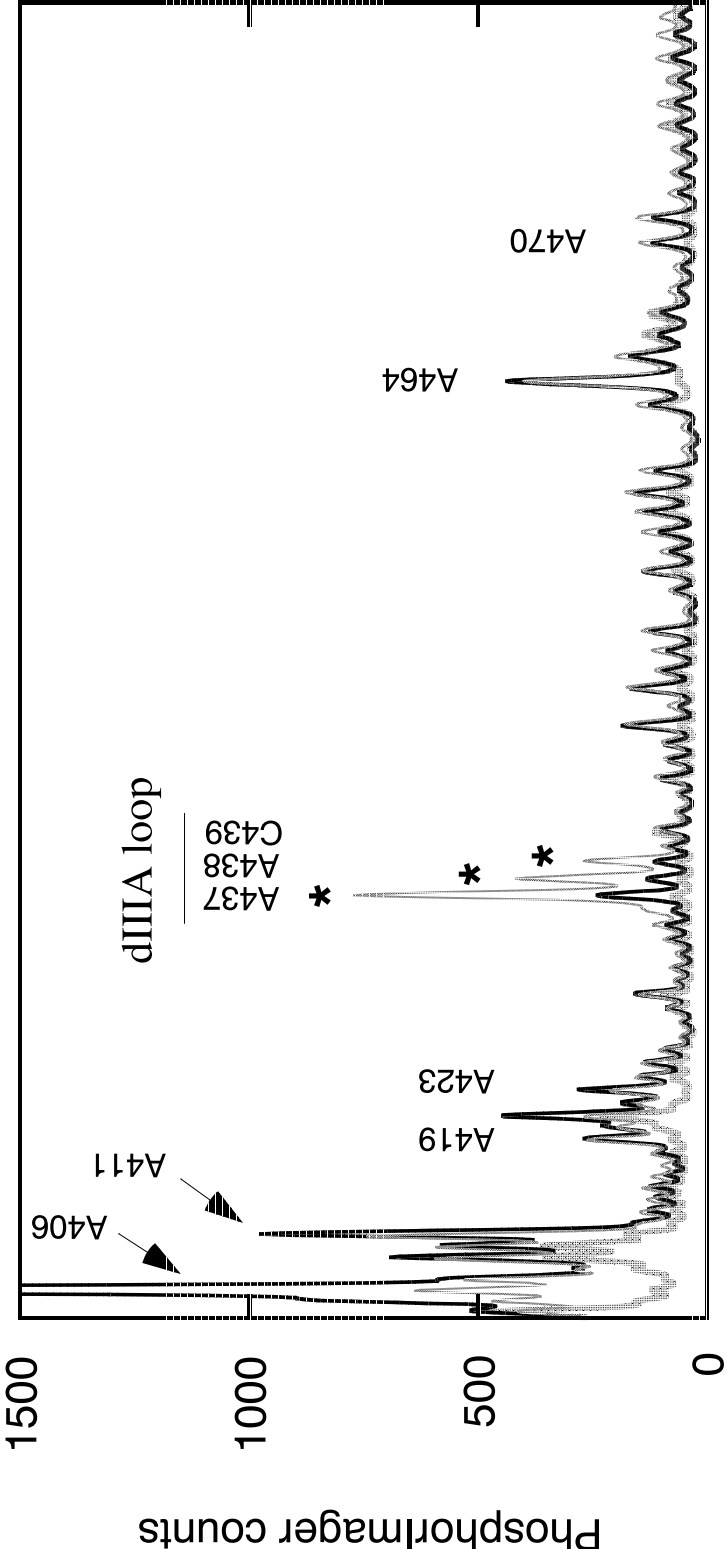


Figure 5B



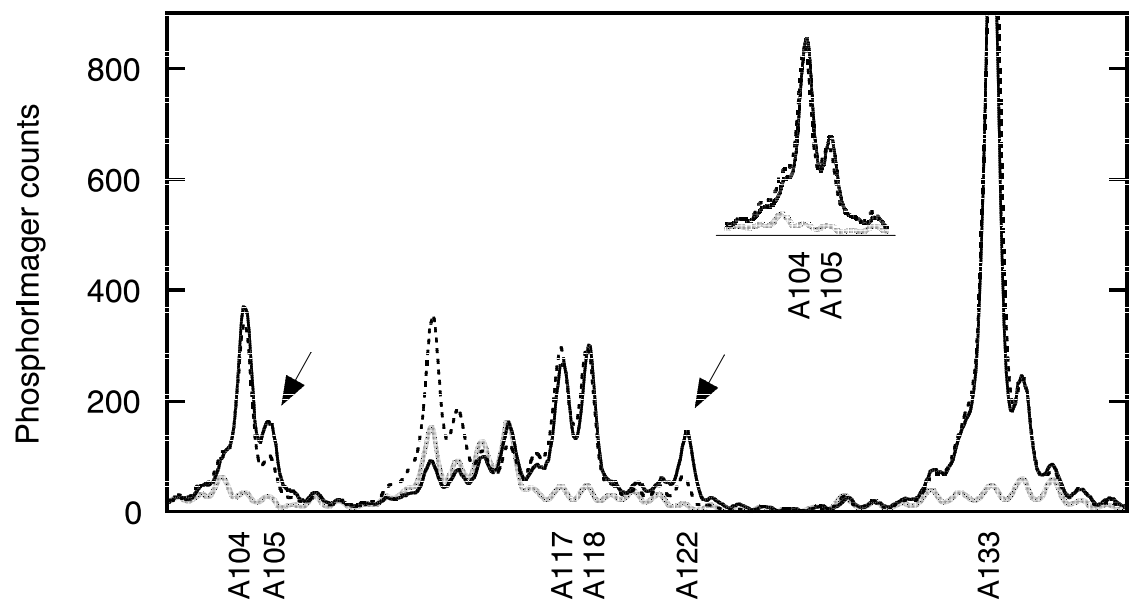


Figure 7A

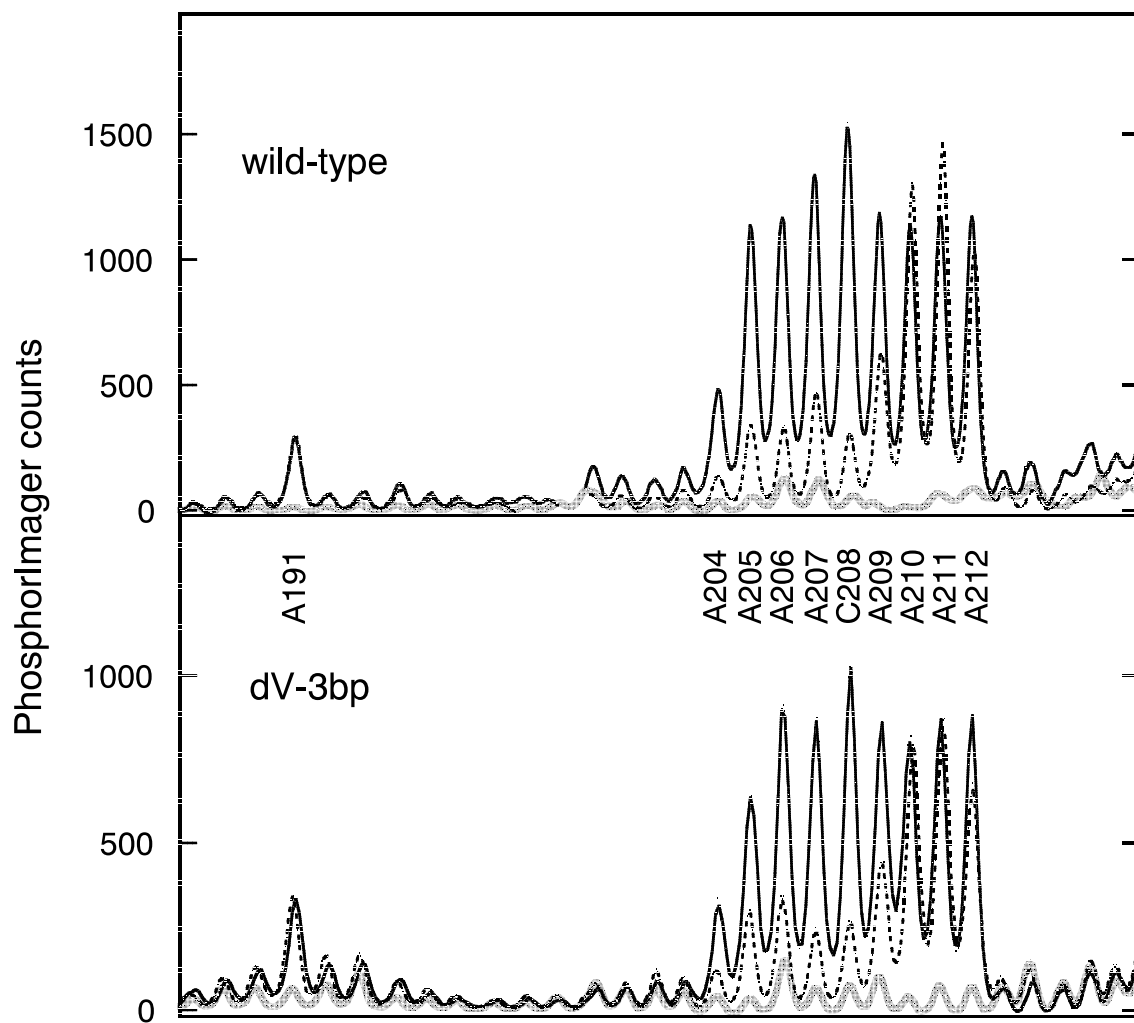


Figure 7B

