The EMBO Journal vol.9 no.8 pp.2555 – 2561, 1990

Unexpected flexibility in an evolutionarily conserved protein – RNA interaction: genetic analysis of the Sm binding site

Michele Haltiner Jones and Christine Guthrie

Department of Biochemistry and Biophysics, University of California, San Francisco, CA, 94143, USA

Communicated by J.A.Steitz

Human autoantibodies of the Sm specificity recognize a conserved set of proteins found in the U class small nuclear ribonucleoproteins (U snRNPs), key *trans*-acting factors involved in the splicing of mRNA precursors. The Sm protein binding site in U snRNAs is unusual because of its single-stranded nature and its simple sequence motif ($AU_{5-6}GPu$). Here we use genetics to probe this specific protein – RNA interaction by saturation mutagenesis of the Sm binding site of the *Saccharomyces cerevisiae* U5 snRNA. The assay system used to analyze these mutations takes advantage of a conditionally expressed U5 gene which does not support growth under non-permissive conditions; U5 genes containing Sm site mutations were tested for their ability to complement this lethal phenotype. Our results indicate that the Sm binding site is remarkably tolerant to mutation despite its high degree of conservation, suggesting that relatively few or redundant specific contacts can determine recognition of single-stranded RNA by protein. A complementary biochemical analysis of these mutants demonstrates that integrity of the Sm site is necessary for snRNP stability *in vivo* and *in vitro*.

Key words: RNA – protein interaction/Sm site/yeast snRNPs

Introduction

The best characterized *trans*-acting factors involved in pre-mRNA splicing are a class of small nuclear ribonucleoprotein particles (U snRNPs) that contain either U1, U2, U4/U6, or U5 snRNAs (for review, see Steitz *et al.*, 1988; Guthrie and Patterson, 1988). It is known from work in a number of laboratories that U snRNAs undergo a complex pathway of biogenesis before performing their roles in splicing (reviewed in Mattaj, 1988). According to the current model, U snRNAs (with the exception of U6) are transcribed by RNA polymerase II, transported to the cytoplasm, assembled with the highly conserved family of Sm proteins and other snRNP proteins, modified internally and at the 5′ end, and transported back to the nucleus. The structure of mammalian U snRNPs has been studied extensively using α-Sm autoantibodies which are produced by patients with certain autoimmune diseases (Pettersson *et al.*, 1984; for review, see Luhrmann, 1988). Seven polypeptides (B – G) of approximate mol. wts 29 kd (B′), 28 kd (B), 16 kd (D), 15.5 kd (D′), 12 kd (E), 11 kd (F) and 9 kd (G) are shared among the U snRNP particles; in addition, each of the U snRNPs has unique polypeptides. The binding of the common proteins is not required for, but probably stabilizes binding of the snRNP specific proteins (Mattaj and DeRobertis, 1985).

The binding site on the snRNAs for one (or more) of the common Sm proteins has been identified by both nuclease protection experiments (Liautard *et al.*, 1982) and deletion analysis (for review, see Mattaj, 1988). Core snRNPs containing proteins D, E, F and G can be partially purified by their ability to withstand CsCl equilibrium density centrifugation (Liautard *et al.*, 1982). Extensive digestion of core snRNPs with micrococcal nuclease leaves a protected 25 base ribonucleotide which contains a sequence now called the Sm site. This motif is highly conserved among the U snRNAs across a large number of eukaryotic species (consensus = $PuAU_{3-4}NUGPu$, derived from Guthrie and Patterson, 1988), and is found in a single-stranded region based on secondary structure prediction and nuclease sensitivity (for review, see Reddy and Busch, 1988). Although it is not clear which protein(s) directly contacts the RNA, UV crosslinking experiments of purified snRNP particles have implicated the F protein in this role (Woppman *et al.*, 1989).

Using another approach, Mattaj and coworkers have performed deletion analysis on the human U2 snRNA, using as an assay immunoprecipitability with a α-Sm antibodies after microinjection into *Xenopus* oocytes (Mattaj and De Robertis, 1985). The ability of the mutant snRNAs to be capped by trimethylguanosine (TMG; a characteristic of snRNAs) and to be correctly localized to the nucleus was also analyzed (*ibid.*; Mattaj, 1986). Deletion of a 12 base sequence containing the Sm site prevents α-Sm immunoprecipitation, TMG capping and localization to the nucleus. Furthermore, a synthetic, unrelated transcript engineered to contain an Sm site can be immunoprecipitated and capped with TMG, albeit inefficiently, suggesting that the Sm site is not only necessary but may be sufficient for these activities (*ibid.*). However, the artificial transcript is not localized to the nucleus, indicating some additional requirement for this process. These experiments demonstrate that the Sm site is critical in the assembly of U snRNAs into RNP particles and in their localization to the nucleus, both of which are required for their role in pre-mRNA splicing.

Recently, the yeast U snRNP homologues (referred to as yU1, yU2, yU4/U6 and yU5) have been identified, although they are much less abundant than in higher eukaryotes (for review, see Guthrie, 1988). Each contains (varying degrees of) sequence and structural homology to its mammalian counterpart. In addition, all can be immunoprecipitated using human α-Sm antibodies (Siliciano *et al.*, 1987; Tollervey and Mattaj, 1987; our unpublished data) and each contains a sequence similar to the metazoan Sm consensus site (Guthrie and Patterson, 1988), except U6 which is immunoprecipitated by virtue of association with U4. Thus both the antigenic epitope(s) of the Sm proteins and the RNA recognition motif appear to be highly conserved in evolution. Yeast Sm proteins have not been directly identified, probably because of their low abundance.

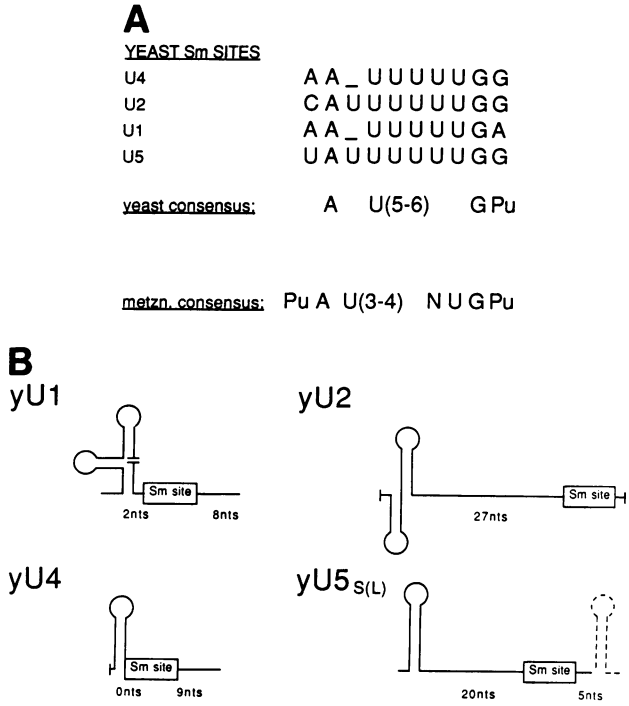

Fig. 1. Sm sites in yeast snRNAs. (**A**) Nucleotide sequence of individual Sm sites. For U1 and U2 in which there are several potential Sm sites, the site was chosen on the basis of its position in the molecule after alignment with a phylogenetic base of the corresponding molecule. Alignment and metazoan consensus from Guthrie and Patterson (1988). N = any nucleotide; Pu = purine. (**B**) Position of Sm sites in individual snRNAs relative to proposed stems or to end of molecule (numbers below molecule), only showing relevant regions. A perpendicular line at the end or middle of the molecule indicates sequence has been omitted. The long form of U5 is shown with a dashed line.

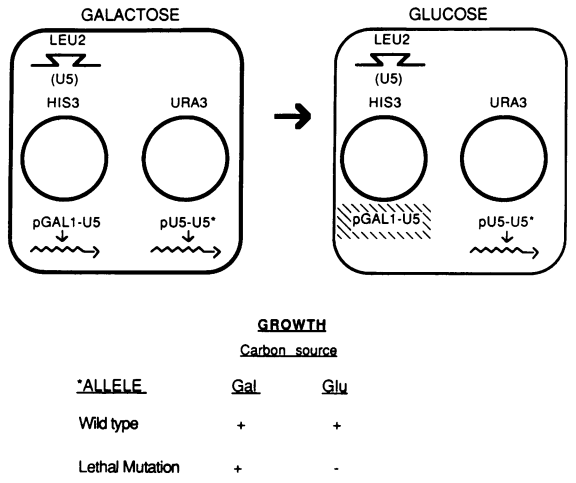

Fig. 2. Scheme for analyzing biological phenotype of Sm site mutations. *LEU2* = chromosomal disruption of the yeast U5 gene; *HIS3* plasmid contains the U5 gene under control of *GAL1* promoter (pGAL1-U5); *URA3* plasmid contains a wild-type or mutant version of U5 with its own promoter (pU5-U5*). Shading indicates repression of transcription in glucose. + or − represents growth on indicated carbon source.

In contrast to mammalian snRNAs, which are encoded by multicopy gene families (Dahlberg and Lund, 1988), yeast snRNAs are encoded by single copy genes (Guthrie, 1988), making them ideal candidates for a genetic analysis of the Sm site and its associated proteins. To determine which nucleotides in the highly conserved Sm domain are essential for function, we have generated all possible single point mutations and a variety of multiple point mutations, deletions and insertions in the Sm site of yeast U5 snRNA. The assay system used to analyze the mutations takes advantage of a conditionally expressed U5 gene which does not support growth under non-permissive conditions (Patterson and Guthrie, 1987); mutant U5 genes were tested for their ability to complement this lethal phenotype. Here we describe the effects of Sm site mutations on yeast viability, on the steady state level of U5 snRNA *in vivo*, and on U5 snRNP stability *in vitro*.

Results

Sm site mutations affect yeast growth

The Sm sites in yeast spliceosomal snRNAs are shown in Figure 1A, together with the derived consensus sequence $AU_{(5\text{-}6)}GPu$. The Sm site lies in a single-stranded region (D.Frank, unpublished data), which in three cases (yU1, yU4 and $yU5_s$) is at the very 3′ end of the molecule (Figure 1B). The consensus is very similar although not identical to that for higher eukaryotes, which is $PuAU_{3\text{-}4}NUGPu$ (Figure 1A). To learn the minimum requirements for Sm site function, we performed oligonucleotide directed mutagenesis of the Sm site in yU5 snRNA. In yeast, there exist two forms of U5 (l, s) transcribed from a single gene (215 and 185 nucleotides, differing at the 3′ terminus; see Figure 1B) (Patterson and Guthrie, 1987). While the functional significance of the two species is unknown, cells that synthesize only the short form, which terminates 5 nucleotides after the Sm site, are fully viable (B.Patterson and C.Guthrie, unpublished observations). The mutations were made in a construct that produces mainly the short form of U5.

The effect of Sm site mutations on viability was assayed using a U5 gene whose synthesis can be repressed under certain growth conditions (Figure 2; Patterson and Guthrie, 1987). A haploid strain bearing a chromosomal deletion of the gene encoding U5 (*LEU2* gene substitution) and a *HIS3* plasmid with the wild-type U5 gene under control of the yeast *GAL1* promoter (pGAL1-U5) can grow on galactose medium, in which the promoter is induced, but not on glucose medium, in which the promoter is repressed. This strain was transformed with a *URA3* plasmid bearing either a wild-type U5 gene or a U5 gene containing an Sm site mutation (pU5-U5*). Transformants expressing the *URA3* gene (Ura^+) were streaked to glucose medium to determine the ability of the mutant yU5 gene to complement the chromosomal deletion. We first showed that a complete deletion or a substitution of the Sm site with the sequence UCGGGGGGUU is lethal in yeast (data not shown).

Since the length of the uridine stretch is somewhat variable in all naturally occurring Sm sites (Figure 1A), the first set of specific mutations was designed to determine the minimum and maximum number of U residues that can be tolerated. As shown in Figure 3A, an Sm site containing only three Us is lethal. The minimum length which still provides viability is four Us. However, as assayed by growth in glucose after a shift from galactose, the 4U mutant site

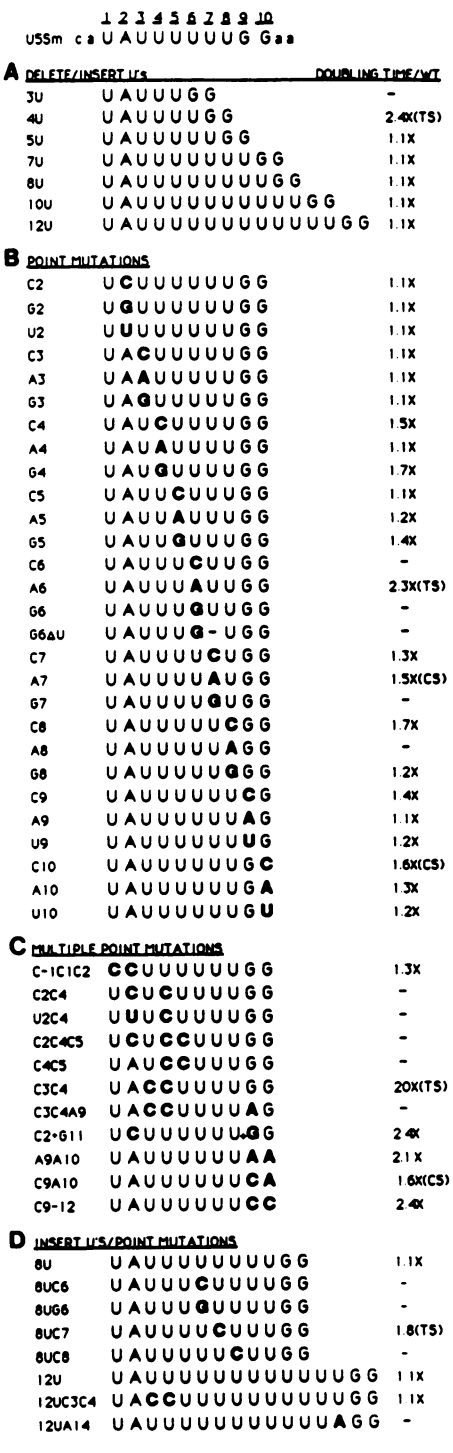

Fig. 3. Biological phenotypes of mutations in yU5 snRNA Sm site. The wild-type site is shown at the top with flanking nucleotides in lowercase letters. Point mutations are in bold print. C9−12 = C9C10C11C12. TS = temperature sensitive; CS = cold sensitive. Doubling times were determined by measuring the optical density of yeast cells grown in liquid culture at 30°C.

confers an increased doubling time (2.4×) compared with that of wild-type (1×). Surprisingly, an Sm site containing as many as 12 Us confers virtually no growth defect. Since the maximum number of Us in naturally occurring Sm sites is six, it is not clear how the additional Us are accommodated within a single site. It is possible that the Sm site is comprised of two adjacent half-sites and that the additional Us are looped out between them. Alternatively, although we have assumed that the 5′ and 3′ borders of the Sm site are the adenine (A2) and guanine (G10) residues respectively, it is possible that one functional border of the site may lie within the uridine stretch.

To explore these possibilities and to define the site further, we sought to determine the respective contributions of each nucleotide by constructing all possible single nucleotide substitutions in the wild-type U5 Sm site. The growth phenotypes of these mutants (Figure 3B) demonstrate that the internal region of the Sm site (nucleotides 6−8) is quite sensitive to mutation; certain single point mutations are lethal, e.g. see C6, G6, G7, A8. However, at several positions in the site, different nucleotide substitutions affect growth quite differently; for example, while A8 is lethal, G8 has an insignificant effect on growth (these data are summarized in Figure 6). Indeed, there is no single location in positions 4−10 at which all three substitutions confer the same phenotype. The C6 mutation was shown to be lethal in the context of the yU2 snRNA as well as in yU5 (data not shown), showing that at least one mutation has the same phenotype in two different Sm sites.

In contrast to the center nucleotides, the borders of the site (nucleotides 2−5 and 9, 10) are relatively insensitive to point mutation. To evaluate the significance of slight effects of border position mutations, we constructed multiple mutations at these sites to test if the effects are synergistic (Figure 3C). All but one of these combinations results in severe growth defects and many are lethal (e.g. see C2C4, C4C5, etc.). In one case (C2+G11), the insertion of a single G at the 3′ border of the site exacerbates the effect of the C2 mutation.

The lethal phenotypes of three point mutations in the 8U background (8UC6, 8UG6 and 8UC8, Figure 3D) argue against the hypothesis that the additional internal Us in the expanded sites are simply a non-functional 'spacer' between two binding sites. Since mutations at position 2 in the normal length site have no effect on their own (Figure 3B), we favored the idea that the protein was recognizing the 7U-12U mutants by binding to the 3′ half of the expanded sites (equivalent to the point mutation U2). To test this hypothesis, we made two mutations in the 12U site (Figure 3D) that should disrupt function if the site is bound at the 3′, or alternatively at the 5′ end: 12UA14 (like A8 from 3′ end) and 12UC3C4 (like C3C4 from 5′ end). In the expanded site, 'A8' is lethal just as it is in the normal length site. In contrast, 'C3C4' has no effect, suggesting that the 5′ half of the expanded site is not making essential contacts with the Sm protein.

All mutations were tested for their ability to support growth at temperatures that are not optimal for yeast growth (18°, 25° and 37°C) compared to that at 30°C. Mutations 4U, A6, A7, C10, C9A10 and 8UC7 have several fold increased doubling times, either at 37°C (TS = temperature sensitive), or at 18°C (CS = cold sensitive) (Figure 3). C3C4 is lethal at 37°C.

Our results also demonstrate that Sm sites that function in other organisms do not necessarily function in yeast. For example, two mutations, G6ΔU and C6, which are lethal in yeast (Figure 3B), represent naturally occurring Sm sites in human U1 and broad bean U2, respectively. This suggests

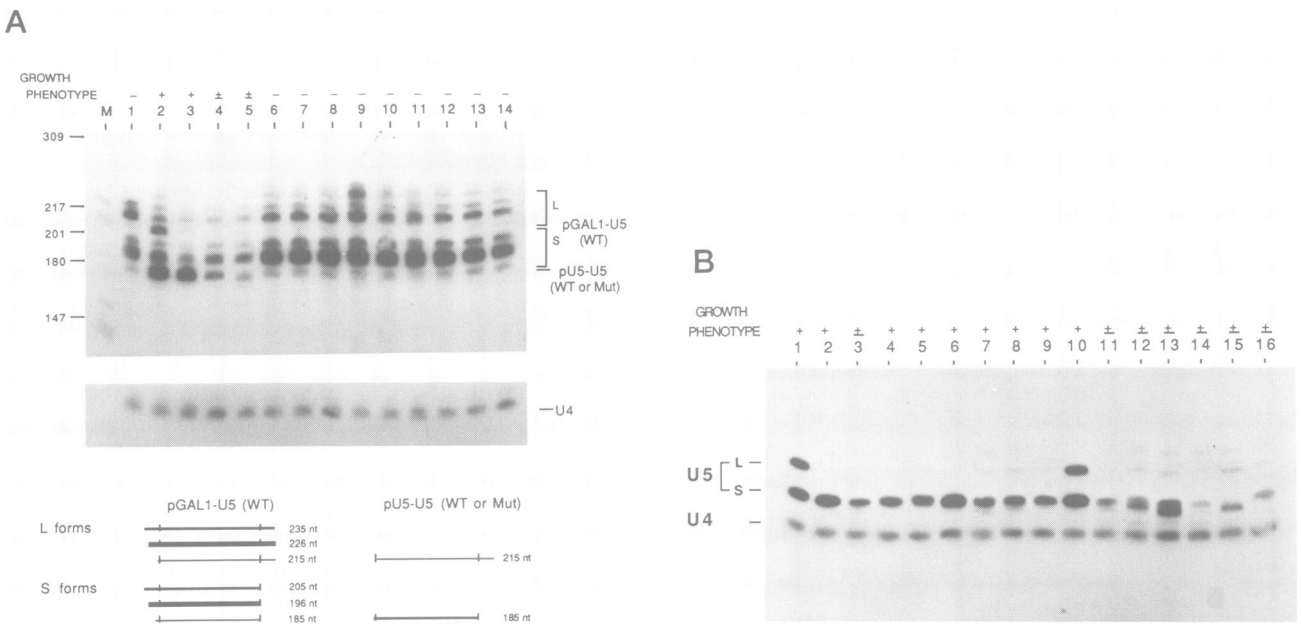

Fig. 4. Northern analyses of steady state levels of Sm mutant snRNAs. (**A**) Lower panel shows sizes and amounts (thickness of line) of expected U5 transcripts (5′−3′) from the *GAL1* promoter and the U5 promoter. All cells contained a wild-type U5 gene from the *GAL1* promoter (pGAL1-U5) and either a wild-type or mutant U5 gene from its own promoter (pU5-U5), as follows: (1) only pGAL1-U5; (2) WT, l + s, a construct that produces both l and s forms; (3) WT,s, a construct that produces mainly the s form; (4) A5; (5) 4U; (6) CS4A9; (7) C6; (8) G6δU; (9) G7; (10) A8; (11) 3U; (12) 8UC6; (13) 8UG6; (14) 8UC8. M = pBR×*Hpa*II markers. Approximate growth phenotypes from Figure 3 are shown above lanes. L, S = long and short forms of U5, respectively. (**B**) cells contained either a wild-type gene or a mutant U5 gene as follows: (1) WT,l + s; (2) WT,s; (3) A5; (4) 8U; (5) 10U: (6) C3; (7) A9; (8) A4; (9) C2; (10) C4; (11) C-1C1C2; (12) 8UC7; (13) 4U; (14) C9-12; (15) C3C4, 25°C; (16) C2+G11. Approximate growth phenotypes from Figure 3 are shown above lanes. Note that for the Northern analysis, C3C4 was grown at 25°C instead of 30°C (doubling time = 3× wild-type).

that, although probably very similar, the detailed aspects of snRNA−Sm protein interaction are not identical across eukaryotes.

Sm mutations affect the steady state levels of U5 snRNA

To test the hypothesis that the above mutations weaken binding of Sm proteins to U5 snRNA, we monitored the amount of steady mutant snRNA, reasoning that the mutant U5 snRNAs may be more susceptible to degradation if they are bound less tightly by these core proteins.

In the case of lethal mutations, cells were grown under conditions in which the pGAL1-U5 gene is expressed (galactose) in order to complement the chromosomal U5 deletion and plasmid-borne Sm mutants. Northern analysis of RNA from the pGAL1-U5 strain transformed with either wild-type or mutant U5 genes is shown in Figure 4A; the bottom panel shows the expected products from the pGAL1-U5 construct (both long and short forms) and from the transformed wild-type or mutant U5 gene (pU5-U5, mostly short form). Note that there are three major sites of U5 transcription initiation from the *GAL1* promoter, the strongest being at −12 relative to the wild-type U5 start site and giving rise to the 226 nucleotide (nt) and 196 nt products (a strain bearing only the pGAL1-U5 gene is shown in lane 1). In cells additionally transformed with a wild-type U5 gene (pU5-U5), there is relatively more steady state U5 derived from its own promoter than from the *GAL1* promoter (lane 3, compare lower two bands in the 'S' series). In contrast, in cells transformed with U5 genes containing lethal Sm mutations (lanes 6−14, compare lower two bands) there is virtually no steady state mutant U5 snRNA (after subtracting the background from the pGAL1-U5 snRNA, lane 1). All other lethal U5-Sm mutations show a similar phenotype (data not shown).

In two cases, cells expressing pGAL1-U5 were transformed with U5 genes containing intermediate mutations (A5 and 4U, lanes 4 and 5, respectively). For these, there is nearly an equimolar ratio of the pGAL1-U5 snRNA and the mutant U5 snRNA. Thus, the steady state amount of mutant U5 snRNA compared with the pGAL1-U5 RNA is directly correlated with the growth rate. This experiment suggests that the U5 specific proteins are probably limiting relative to U5 snRNA; although an approximately equal amount of RNA is present in each lane (see U4 internal control), there is substantially more pGAL1-U5 snRNA in lanes 6−14 than lanes 3−5, probably because the mutant U5 from the transformed gene in the former cases does not successfully compete for snRNP proteins. In light of the competition model, however, we do not have an explanation for the reduced total amount of U5 snRNA in lanes 4 and 5. A related observation to ours has been made previously in a system in which the transformation of human U1 snRNA genes into mouse cells caused a reduction in endogenous mouse U1 snRNA, while the total level of U1 snRNA remained approximately constant (Mangin *et al.*, 1985).

For all mutations that had either a wild-type phenotype or an intermediate growth phenotype, RNA was prepared from cells that contained only the mutated U5 gene, or a wild-type control gene. The U5 RNA level was compared with a U4 RNA internal control. Northern analysis of RNA from these cells (Figure 4B and data not shown) demonstrates: (i) for every mutation that causes a growth defect, there is a decreased level of the (full length) mutant snRNA relative to the wild-type snRNA, suggesting that a major contributing factor to the growth defect is a limiting amount of

U5 snRNA. C4 is an exception as it does not show reduced RNA levels yet causes a growth defect. In addition, among the mutants the magnitude of the growth defect does not always correlate exactly with the RNA level; together these observations suggest that some mutant snRNPs may also be relatively less functional in splicing. (ii) Some of the mutant U5 snRNAs (lanes 12, 13 and 15) contain significant amounts of a smaller product, the size of which is consistent with degradation from the 3′ end to the Sm site. (iii) Finally, some of the mutations, especially those that cause growth defects, appear to increase the ratio of long/short forms of U5; the molecular basis of this observation is not understood.

In summary, the Northern results are consistent with a scenario in which U5 snRNAs containing mutations that cause growth defects are recognized less well by the Sm proteins during and after snRNP assembly and therefore are more accessible to degradative nucleases. This results in a decreased steady state level of mutant U5 snRNA relative to wild-type. Because Northern analysis measures steady state RNA, a formal, although unlikely, possibility is that the Sm mutations affect synthesis of U5 snRNA rather than stability.

Sm-mutant U5 snRNPs are less stable in vitro than wild-type U5 snRNPs

As a more definitive test of our hypothesis that U5-Sm site mutations affect growth by decreasing the affinity of Sm proteins for mutant U5, thus leading to limiting levels of U5 snRNA in the cell, we sought to measure directly the stability of mutant snRNPs relative to wild-type snRNPs. As mentioned, purification of mammalian snRNPs on cesium gradients results in core snRNP particles containing only the proteins associated with the Sm site (D, E, F and G) (Liautard *et al.*, 1982): therefore, we used this procedure analytically as a measure of Sm site binding. We first showed that, like mammalian snRNPs, yeast snRNPs are maintained in highly salt-resistant particles after Cs_2SO_4 equilibrium density centrifugation (data not shown); they band at a characteristic density of 1.4 g/ml (Liautard *et al.*, 1982) whereas free RNA migrates at 1.6 g/ml. These particles contain Sm protein(s) because they are immunoprecipitable by α-Sm antibodies (data not shown).

We next tested whether three U5 snRNPs which contain mutations that cause growth defects (C3C4, C2+G11 and 8UC7, see Figure 3), are stable to centrifugation in cesium; in each case, the mutant U5 snRNP migrates in a position identical to a wild-type control (data not shown). We then assayed their relative stability to denaturants. Cesium gradients were run in the presence of 7 M urea, and three fractions, from the top (1.4 g/ml), middle (1.5 g/ml) and bottom (1.6 g/ml), were collected (Figure 5). Under these conditions, wild-type U5 snRNPs are partially denatured, i.e. the majority of the snRNA molecules still migrates at 1.4 g/ml and a small amount migrates as free RNA at 1.6 g/ml (panel 1). In contrast, the three Sm mutant snRNPs (panels 2–4) migrate as free RNA at 1.6 g/ml and thus are completely denatured under these conditions. The internal control (U4 snRNP) migrates the same as wild-type U5 snRNP in all four gradients. Therefore, of three Sm mutant snRNPs tested, all are relatively less stable than wild-type snRNPs, strongly supporting the hypothesis that Sm mutations reduce the affinity of Sm protein binding, and consistent with their relatively low *in vivo* levels, and with the growth defects conferred by these mutations.

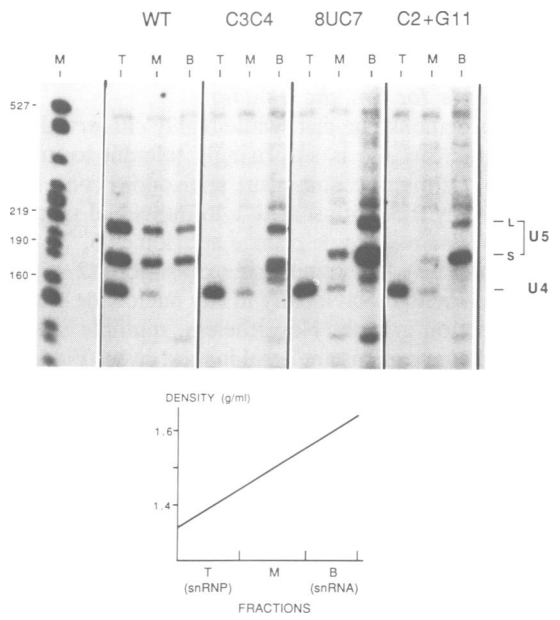

Fig. 5. Northern analysis of urea–cesium gradient fractions of wild-type and mutant U5 snRNPs. Lower panel shows the density of the fractions across the gradient and the expected migration of wild-type snRNP particles and free snRNA in a cesium gradient. T = top; M = middle; B = bottom.

Discussion

We sought to determine genetically which nucleotides in the highly conserved Sm domain are essential for function. We have generated all possible single point mutations and a variety of multiple point mutations, deletions and insertions in the Sm site of yeast U5 snRNA. We have evaluated the effects of these mutations by three criteria: (i) stability of the U5 snRNP particle during centrifugation in Cs_2SO_4 in the presence of denaturants; (ii) the steady state level of U5 snRNA *in vivo*, and (iii) yeast growth rate. Cesium gradient centrifugation provides a sensitive assay for the Sm protein–RNA interaction as it has been demonstrated that mammalian snRNPs purified in this manner retain only the snRNP proteins associated with the Sm site, and migrate at a characteristic density (Liautard *et al.*, 1982). We have shown that the density of the yeast snRNPs is similar to that of the mammalian core particles under these conditions and that they contain Sm proteins as assayed by immunoprecipitability with α-Sm antibodies (data not shown); therefore the snRNP stability assay should primarily reflect the strength of association of the Sm proteins with the U5 snRNA. It is likely that the steady state *in vivo* snRNA levels, assayed by Northern analysis, reflect the ability of the snRNA to be protected by Sm and other snRNP proteins from degradative nucleases; this may occur both during assembly (and hence reflect Sm protein association) and after assembly (reflecting both dissociation and association). The results of these three assays show a strong correlation, suggesting that growth rate is a valid indication of snRNP association/dissociation values, although it may also reflect

the efficiency of other steps of snRNP biogenesis and function.

Requirements for Sm site binding

Our results demonstrate that, while highly conserved evolutionarily, the Sm site is surprisingly tolerant to mutation (Figure 3). Of the nine consensus nucleotides conserved in yeast U5, only three are sensitive to individual mutations; i.e. at least one nucleotide change at uridines 6, 7 and 8 is lethal (Figure 3B). The remaining six positions (2–5, 9 and 10) can be mutated individually to any nucleotide with little if any effect on growth. Nevertheless, multiple mutations involving these positions are lethal in most cases (Figure 3C). This suggests that, although positions 2–5 and 9, 10 may represent redundant or more minor contacts, together they are critical for binding.

Interestingly, while uridines 6–8 represent important contacts, the degree of the defect is highly dependent on the identity of the mutant nucleotide and is different for each position (Figure 3B, summarized in Figure 6). For example, although C6, G6, G7 and A8 are lethal (smallest font), mutations A6, A7 and C8 have only intermediate effects on doubling time (1.5–2.3× wild-type, shown in medium hollow front) and C7 and G8 affect doubling time very little (1.2–1.3× wild-type, shown in largest font). This implies that each of the uridines is recognized by one or more unique determinants. We expect *a priori* that the Sm protein is more likely to recognize primary than secondary structure since poly-uridylic acid is known to be the least structured of the homopolymers (Saenger, 1984). However, a comparison of the more neutral mutations at each position to the original base does not yield obvious rules as to the nature of the primary recognition determinant. Although we consider it unlikely, we cannot rule out the possibility that the effect of certain substitutions may be due to the introduction of inhibitory structures rather than to removal of determinants required for recognition.

Once bound, the protein may act to maintain a particular single-stranded conformation in this region of the RNA. Precedents exist for proteins that affect the structure of nucleic acids to which they are bound. For example, although poly(U) can assume a stable secondary structure at low temperatures, it is converted stoichiometrically by ribosomal protein S1 to a completely denatured form (Bear *et al.*, 1976). In another case, the binding of glutamine tRNA synthetase to its cognate tRNA causes the usually stacked nucleotides in the anticodon loop to be separated into three different binding pockets of the protein (Rould *et al.*, 1989).

Comparison to other RNA–protein interactions

Comparison of the Sm site with other protein binding sites in RNA emphasizes its apparent lack of secondary structure and relatively few sequence specific contacts. For example, the binding site that has been most extensively studied is the stem–loop region on bacteriophage R17 RNA which is bound by its translational repressor (Carey *et al.*, 1983). As tested in an *in vitro* filter binding assay, four of the five loop nucleotides are essential for binding and the integrity of the 7 bp stem must also be maintained. The binding site for the 70K protein on U1 snRNA also consists of a stem–loop structure with eight of ten loop bases being essential for binding *in vitro* (Surowy *et al.*, 1989). The small number (three) of critical contacts in the Sm site is especially remarkable considering that six additional residues within the site are evolutionarily well conserved, and raises the paradox of apparent dispensibility despite conservation. Additional functional studies of protein binding sites on RNA will be necessary to determine whether this is a recurrent theme. In any case, it is likely that a mutational disadvantage suffered by an organism which is slight over the period of the assay may make a sufficient difference over evolutionary time to account for the observed conservation.

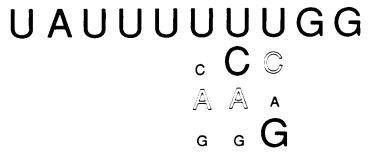

Fig. 6. Summary of phenotypes of mutations at positions 6–8. Wild-type U5 Sm site with nucleotide changes made in positions 6–8 shown below. The font size reflects the doubling time relative to wild-type conferred by the mutation, the largest font indicating a doubling time of 1.2–1.3× wild-type, the medium hollow font indicating a doubling time of 1.5–2.3× wild-type, and the smallest font indicating a lethal phenotype.

Function of Sm site binding

Our results suggest that correct binding of the Sm protein(s) is essential for stabilizing the snRNA. The steady state amount of snRNA derived from Sm mutant U5 genes that cause growth defects is reduced several fold relative to wild-type U5 snRNA. Moreover, there is a complete absence of snRNA derived from Sm mutant U5 genes that contain lethal mutations. The reduced levels are most likely due to an increased susceptibility of the mutant snRNAs to degradation during and after assembly, as we have shown that the snRNP particles containing mutant Sm sites are less stable than wild-type snRNPs *in vitro* to Cs_2SO_4 centrifugation. A contributing factor to the decreased stability may be a defect in the binding of the U5 specific proteins as well as the core proteins. It has been demonstrated that binding of Sm core proteins is not required for, but most likely significantly stabilizes binding of the snRNP specfic proteins for U1 and U2. Mattaj and coworkers showed that deletion of the Sm site inhibits the binding *in vivo* of U2 specific proteins as well as Sm proteins (Mattaj and De Robertis, 1985). However, binding of at least one U2 specific protein can precede binding of the core proteins in HeLa cells under conditions of viral infection (Fresco *et al.*, 1987). Also, *in vitro* binding studies of the U1 specific proteins 70K and C show that although binding to U1 does not absolutely require the Sm site (Hamm *et al.*, 1987; Surowy *et al.*, 1989), it does appear to be stabilized in some assays by the presence of this domain (Hamm *et al.*, 1987; Patton and Pederson, 1988).

Perspectives

The low abundance of yeast Sm snRNP proteins has made traditional biochemical approaches to their purification and cloning difficult (unpublished observations). As an alternative approach, with the bank of Sm site mutations described in this manuscript, we can now use those with conditional phenotypes to isolate second site suppressors which should include the Sm snRNP proteins. It is clear from our results that a function of the protein(s) interacting at the Sm site

is to preserve the integrity of the snRNP particle. From work in other systems, we know that the Sm site is at least indirectly required for trimethyl guanosine capping and nuclear localization of the snRNAs. Key issues that remain to be addressed include determining which of the many Sm core proteins identified in mammalian systems are involved in each of these diverse reactions and why such a large number of Sm proteins are required. Moreover, genetic access to the Sm proteins will allow a test of whether the core proteins also play direct roles in splicing. The recent demonstration of tissue specific variants of the Sm B′ protein suggest the possibility that core proteins may influence regulated splicing events (Li *et al.*, 1989).

Materials and methods

Materials

Restriction enzymes were obtained from New England Biolabs. [^{32}P]ATP was purchased from ICN. HyBond hybridization membranes were obtained from Amersham. Sequenase and cesium sulfate was purchased from United States Biochemical Corporation. Bluescript vectors were purchased from Stratagene. Oligonucleotides were synthesized by the Biomolecular Resource Center, UCSF. pSE360 was a kind gift from S.Elledge and contains a *URA3* gene and sequences to allow replication in yeast. The yU5 gene was cloned into the polylinker sequence.

Plasmid construction

pGAL1-yU5(*HIS3*) has been described (Patterson and Guthrie, 1987). yU5-Sm(*URA3*) was constructed by putting the *Bam*HI−*Pvu*II fragment of snr7-BI$^+$ into the *Bam*HI−*Sma*I fragment of pSE360 (S.Elledge, personal communication). snr7$^-$B$^+$ was constructed by putting an *Eco*RI−*Bam*HI fragment containing the yeast U5 gene into Bluescript$^+$ vector.

Strains

The haploid strain in which the mutants were screened is a derivative of YPH399 (genotype:a, ura3-52, trp1-289, his3-532, lys2-801, ade2-101, leu2$^-$) and contained the plasmid pGAL1-yU5(*HIS3*) and a chromosomal replacement of the U5 gene with the *LEU2* selectable marker.

Media

YEP-glucose and synthetic media are described in Sherman *et al.*, 1974. YEP-galactose is identical to YEP-glucose except that galactose is the carbon source. Cells were grown at 30°C unless otherwise stated.

Oligonucleotide directed mutagenesis

Mutagenesis was performed using a dut$^-$ung$^-$ protocol (Kunkel *et al.*, 1987) on the plasmid snr7$^-$BI$^+$. Mutagenesis reactions were transformed into DG98 *Escherichia coli* strains. Transformants were screened for mutations by sequencing using a modified version of the dideozy method (Sanger *et al.*, 1977).

Northern analysis and cesium sulfate centrifugation

Northern analysis was performed as described by Patterson and Guthrie (1987) using an oligonucleotide described therein (annealing to snr7). Total RNA was prepared as described by Domdey *et al.* (1984). For cesium gradients, splicing extracts were prepared as described by Lin *et al.* (1985) and centrifuged in 1 m gradients containing 1.55 g/ml cesium sulfate (with or without 7 M urea) using a TLA 100.2 rotor in an TL100 ultracentrifuge at 90 000 r.p.m. until they had reached equilibrium (14 h). Fractions were collected, the densities analyzed and the samples were deproteinized and subjected to Northern analysis.

Acknowledgements

We thank D.Frank, H.Madhani, S.Wolin and K.Strub for critical reading of the manuscript and P.Walter, O.Uhlenbeck, H.Noller and members of the Guthrie laboratory, especially D.Frank, B.Patterson and J.Milligan for valuable discussion. We are grateful for the excellent technical assistance of Lucita Esperas. The research was supported by a grant from the Leukemia Society (M.H.J.) by PHS Award 5 T32 CA09270-14 (M.H.J.), and by National Institutes of Health grant GM 21119 (C.G.) and National Science Foundation grant PCM 8303667 (C.G.).

References

Bear,D.G., Ng,R., Van Derveer,D., Johnson,N.P., Thomas,G., Schleich,T. and Noller,H.F. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 1824−1828.

Carey,J., Cameron,V., de Haseth,P.L. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2601−2610.

Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp. 38−70.

Domdey,H., Apostol,B., Lin,R.J., Newman,A., Brody,E. and Abelson,J. (1984) *Cell*, **39**, 611−621.

Fresco,L.D., Kurilla,M.G. and Keene,J.D. (1987) *Mol. Cell. Biol.*, **7**, 1148−1155.

Guthrie,C. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp. 196−211.

Guthrie,C. and Patterson,B. (1988) *Ann. Rev. Gen.*, **22**, 387−420.

Hamm,J., Kazmaier,M. and Mattaj,I.W. (1987) *EMBO J.*, 3479−3485.

Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367−382.

Liautard,J.-P., Sri-Widada,J., Brunel,C. and Jeanteur,P. (1982) *J. Mol. Biol.*, **162**, 623−643.

Li,S., Klein,E.S., Russo,A.F., Simmons,D.M. and Rosenfeld,M.G. (1989) *Proc. Natl. Acad. Sci. USA.*, **86**, 9778−9782.

Lin,R.-J., Newman,A.J., Cheng,S.-C. and Abelson,J. (1985) *J.Biol. Chem.*, **260**., 14780−14792.

Luhrmann,R., (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp. 71−99.

Mangin,M., Ares,M. and Weiner,A. (1985) *Science*, **229**, 272−275.

Mattaj,I.W. (1986) *Cell*, **46**, 905−911.

Mattaj,I.W. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp. 100-114.

Mattaj,I.W. and De Robertis,E. (1985) *Cell*, **40**, 111−118.

Patterson,B. and Guthrie,C. (1987) *Cell*, **49**, 613−624.

Patton,J.R., and Pederson,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 747−751.

Petterson,I., Hinterberger,M., Mimori,T., Gottlieb,E. and Steitz,J.A. (1984) *J. Biol. Chem.*, **259**, 5907−5914.

Reddy,R. and Busch,H. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp. 1−37.

Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) *Science*, **246**, 1135−1142.

Saenger,W. (1984) In Cantor,C.R. (ed.), *Principles of Nucleic Acid Structure.* Springer-Verlag, New York, p.310.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

Sherman,F., Fink,G.R. and Lawrence,C.W. (1974, revised annually) *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Siliciano,P.G., Jones,M.H. and Guthrie,C. (1987) *Science*, **237**, 1484−1487.

Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag, New York, pp 115−154.

Surowy,C.S., van Santen,V.L., Scheib-Wixted,S.M. and Spritz,R.A. (1989) *Mol. Cell. Biol.*, **9**, 4179−4186.

Tollervey,D. and Mattaj,I.W. (1987) *EMBO J.*, **6**, 469−476.

Woppmann,A., Rinke,J. and Luhrmann,R. (1989) *Nucleic Acids Res.*, **16**. 10985−11004.

Received on May 8, 1990