Evolutionary Diversification of the Sm Family of RNA-Associated Proteins

Douglas G. Scofield and Michael Lynch

Introduction

The Sm family of proteins, encompassing the Sm and Sm-like (Lsm) proteins (Séraphin 1995), are common participants in RNA metabolism in Eubacteria (Valentin-Hansen et al. 2004), Archaea (Salgado-Garrido et al. 1999; Mura et al. 2001), and eukaryotes (Mattaj and Derobertis 1985). Sm proteins primarily occur as small (~9–29 kDa) stand-alone proteins lacking other domains (Anantharaman et al. 2002; for an exception see Pillai et al. 2003) that assemble to form characteristic homomorphic or heteromorphic rings containing six or seven proteins. Members of the family are characterized by the conserved bipartite Sm domain or “Sm fold” which functions, at least in part, in binding to neighboring Sm proteins within such rings (Box 1 and Hermann et al. 1995; Séraphin 1995; Khushial et al. 2005). One highly conserved characteristic of Sm rings is the direct interaction of the central pore of the ring with short uracil-rich stretches of RNA, in both prokaryotes (Box 1 and Töör et al. 2001; Schumacher et al. 2002; Mura, Kozhukhovsky et al. 2003; Thore et al. 2003) and eukaryotes (Branlant et al. 1982; Liautard et al. 1982; Urlaub et al. 2001; Khusial et al. 2005). The Sm family in eukaryotes has undergone considerable diversification, with a variety of heteromorphic Sm rings participating within many RNA-processing pathways and snRNP complexes (Anantharaman et al. 2002; Khusial et al. 2005; Wilusz CJ and Wilusz J 2005). Some multidomain proteins involved in RNA processing carry divergent Sm domains but have not been found within Sm rings (Albrecht and Lengauer 2004; Albrecht et al. 2004; Anantharaman and Aravind 2004; Fleischer et al. 2006; Tritschler et al. 2007); here we focus on Sm protein members of Sm rings.

Though the evolutionary diversification of Sm proteins and Sm proteins in eukaryotes clearly reflects a general increase in RNA-processing complexity in eukaryotes (Anantharaman et al. 2002), there are several dramatic discontinuities in comparison to prokaryotes. These include the formation of heteromorphic, 7-meric Sm rings; the highly stable, static association of Sm rings with snRNA cofactors; the use of a dedicated pathway for assembling Sm rings with snRNA cofactors; and the central participation of Sm–snRNA complexes within heterogeneous small nuclear ribonucleoproteins (snRNPs) having a variety of eukaryote-specific RNA-processing functions.

The existing “diversification–duplication” model for the evolution of the Sm protein family in eukaryotes proposes two primary steps early in the evolution of eukaryotes (Salgado-Garrido et al. 1999). The protein family first diverged to form a single heterogeneous Sm ring, followed by a large-scale duplication that allowed for the assembly of a second but related Sm ring; these two rings are the “canonical” Sm ring at the heart of the U1, U2, U4, and U5 spliceosomal snRNPs and the Lsm ring at the heart of the U6 spliceosomal snRNP. Subsequent smaller duplications led to the creation of other Sm proteins that participate in other Sm rings.

In this study, we consider the pattern of evolutionary diversification of Sm proteins in eukaryotes. We first examine the phylogenetic structure of Sm proteins across a wide variety of prokaryotes and eukaryotes and identify major evolutionary trends in the family. We find that eukaryotic Sm diversity reached much of its present breadth very early in eukaryotic evolution. We then review Sm structure, diversity and function in prokaryotes and eukaryotes, as well as the pathways for Sm ring assembly in eukaryotes. Based on the functional review and the results of the phylogenetic analysis, we partition Sm rings in eukaryotes into two classes: fixed Sm rings and flexible Sm rings, corresponding roughly to the rather informal nomenclature for “Sm-type” and “Lsm-type” rings. Our fixed and flexible Sm ring class designations reflect not only broad differences in the
duration of the Sm ring–RNA cofactor association but also presumed differences in Sm ring–RNA-binding patterns and Sm ring structural stability that underlie differences in Sm ring–RNA associations. Finally, we consider some possible mechanisms for Sm ring evolution in eukaryotes. The analyses presented here provide insights into the evolution of this distinctive, ubiquitous, and critical family of RNA-associated proteins.

Phylogenetic Relationships among Sm Proteins

We examined the evolutionary relationships among Sm proteins across all life by assembling a phylogenetic tree containing 202 Sm proteins from GenBank records. Prokaryotic Sm proteins include those representing Archaea (figures in brackets indicate the number of Sm proteins in the taxa where greater than one): Archaeoglobus fulgidus [2], Thermoplasma volcanium GSS1 [2], Thermoplasma acidophilum DSM 1728 [2], Aeropyrum pernix K1 [2], Pyrobaculum aerophilum IM2 [3], and Methanopyrus kandleri AV19 and Eubacteria: Escherichia coli O157:H7, Staphylococcus aureus ssp. aureus JH1, Clostridium botulinum B strain Eklund 178, Thermobifida fusca YX, α-proteobacterium HTCC2255 [2], and γ-proteobacterium HTCC2207. We also broadly sampled Sm proteins from eukaryotes, including the diplomonad Giardia lamblia, the parabasalid Trichomonas vaginalis G3, the apicomplexan Plasmodium falciparum 3D7, the cryptomonad Guillar dia theta, the amoebozoan Entamoeba histolytica, the chlorarachniophyte Bigelowiella natans, the kinetoplastid Trypanosoma cruzi strain CL Brener, the plant Arabidopsis thaliana, the microsporidian Encephalitozoon cuniculi, the fungus Saccharomyces cerevisiae, and animals, including Drosophila melanogaster, the urochordate Ciona intestinalis, and human. Sm proteins were initially identified by GenBank annotations with the sequence set incrementally expanded via PSI-Blast (Altschul et al. 1997). For species having insufficient GenBank coverage for Sm proteins, the excellent spliceosomal protein sequence database provided by Barbosa-Morais et al. (2006) was used; see their paper for further details of their sources. Multiple sequence alignment was performed using MUSCLE with default settings (Edgar 2004), with a neighbor-joining tree constructed from this alignment using ClustalW with gaps included and 1000 bootstraps (Chenna et al. 2003). Expanded information for the 202 sequences used may be found in supplementary table 1 (Supplementary Material online).

Although the extreme age and high degree of diversification of the Sm protein family has resulted in relatively high uncertainty at more basal nodes within clades, some features are readily apparent in figure 1 and its more detailed version in supplementary figure 1 (Supplementary Material online): 1) the basal positions of archaean and eubacterial Sm proteins, with archaean proteins scattered throughout the tree and eubacterial proteins clustering together in a single clade; 2) the extreme depth of the fundamental splits among individual eukaryotic Sm proteins; and
3) the subsequent evolution within individual Sm protein clades following, with some variation, the course of eukaryotic diversification. In this tree, there is relatively little support for any one configuration of deeper branches above clades for individual Sm proteins. The interrelationships among some Sm protein clades shifts somewhat with different compositions of the data set, though some neighbor clade pairs remain relatively robust, for example, Lsm8 + Lsm1, Lsm5 + SmE, and Lsm3 + SmD2 (data not shown). In this tree, some individual proteins are out of place, such as Lsm8, Lsm6, and SmB in *S. cerevisiae*. This is not surprising, given the great age of individual proteins in this family, but some of these placements may reflect species-specific evolutionary differences in RNA processing, for example, *S. cerevisiae* contains an unusually depauperate set of introns (Fink 1987).

Sm Protein Diversification in Eukaryotes

Eukaryotic Sm proteins have undergone considerable diversification in comparison to their prokaryotic ancestors and play a central role in many eukaryote-specific snRNPs. Nearly all in vivo Sm rings have been found to be 7-merous with each of the 7 members being a distinct Sm protein (Fig. 2A). Exceptions to date include two functional complexes containing 6 distinct Sm proteins: the Sm ring selectively binding U8 snoRNA in *Xenopus* (Tomasevic and Peculis 2002) and the Sm ring binding snR5 RNA in *Saccharomyces* (Fernandez et al. 2004). Eukaryotic Sm proteins are involved in snRNA stability (Mayes et al. 1999; Liu et al. 2004), interaction with nuclear import factors during snRNP maturation and recycling of spliceosomal multi-snRNP complexes (Palacios et al. 1997; Chan et al. 2003; Liu et al. 2004; Narayanan et al. 2004), stabilization of spliceosome–mRNA complexes (Zhang et al. 2001), and modification of a variety of RNA substrates (Pillai et al. 2003).

In this section, we will define and use our fixed and flexible class designations where appropriate. We will consider these class designations more fully in the Discussion.

Fixed Class (Sm-Type) Sm Rings

Fixed class eukaryotic Sm rings form stable, long-term associations with RNA substrates, around which they are assembled via a dedicated pathway. Fixed class Sm rings form largely passive scaffolds around which RNA and protein cofactors assemble in a spatially explicit manner. RNA substrates of fixed Sm rings contain an “Sm site,” which consists of PuAU₄₋₅Gpu flanked by two stem–loop structures (Branlant et al. 1982; Liautard et al. 1982; Urlaub et al. 2001; Khusial et al. 2005). Electron microscopy and UV cross-linking studies of the human U1 snRNP suggest that, rather than encircle one face of the Sm ring as in prokaryotes, the Sm site of U1 snRNA circles through the central pore (Fig. 2B and Stark et al. 2001; Urlaub et al. 2001; Stark and Lührmann 2006). The binding configuration of other fixed Sm ring–RNA associations is currently unknown.

The “canonical” fixed class eukaryotic Sm ring is found at the center of the U1, U2, U4, and U5 spliceosomal snRNPs (Fig. 2A). This Sm ring contains monomers in the order SmD1/SmD2, SmF/SmE/SmG, and SmD3/SmB, where the subgroups indicate spontaneously forming dimers and a trimer that are then assembled to form the complete ring (Kambach, Walke, and Nagai 1999; Kambach, Walke, Young, et al. 1999; Raker et al. 1999). This Sm ring is assembled around its various snRNA substrates in the cytoplasm via the Survival of Motor Neurons (SMN) pathway (Meister et al. 2002; Yong, Wan, and Dreyfuss 2004) with components that selectively bind both Sm proteins (Pu et al. 1999; Friesen et al. 2001) and snRNA (Pellizzoni, Yong, and Dreyfuss 2002; Yong, Pellizzoni, et al. 2002; Golembe, Yong, and Dreyfuss 2005). These Sm-snRNA pre-snRNPs are transported into the nucleus complexed with the SMN protein (Fischer et al. 1991; Palacios et al. 1997; Massenet et al. 2002; Narayanan et al. 2004) and are then matured within Cajal bodies (Jády et al. 2003; Kiss 2004; Cioce and Lamond 2005; Liu, Murphy, et al. 2006; Matera and Shpargel 2006; Stueck and Neugebauer 2006; Tycowski et al. 2006). The high selectivity of the SMN pathway ensures correct association of Sm site-containing snRNA with the appropriate fixed Sm ring, but it is not foolproof. RNAs associated with the primate virus *Herpesvirus saimiri* can outcompete host snRNAs for fixed Sm rings on this same pathway and thereby gain intracellular stability (Golembe, Yong, Battle, et al. 2005).

There are taxa- and tissue-specific variations in the composition of the fixed Sm ring in spliceosomal snRNPs. In trypanosomes, some spliceosomal snRNPs contain variant Sm rings. Trypanosomes’ U2 snRNAs have a divergent Sm site in comparison both to other spliceosomal snRNAs in trypanosomes and to U2 snRNAs in other eukaryotes, and the Sm ring assembled around this Sm site is also divergent (Wang et al. 2006). The SmD3/SmB dimer is replaced by the novel Sm proteins Sm16.5K/Sm15K, the closest sequence homologs of which are SmD3 and SmB, respectively (Wang et al. 2006). In trypanosomes’ U4 snRNP, SmD3 alone is replaced by a novel Sm protein not found in the U2 snRNP (Tkacz et al. 2007). In mammals, there are three variants of SmB (SmB, SmB′, and SmN) that may all substitute for SmB in spliceosomal snRNPs; they are interrelated by alternative splicing and gene duplication (McAllister et al. 1988; McAllister et al. 1989; Chu and Elkon 1991; Griffith et al. 1992; Gray et al. 1999). SmN is most highly expressed in neural tissue, particularly postnatal brain, and its underexpression is associated with Prader–Willi syndrome (Gray et al. 1999; Nicholls and Knepper 2001).

There are other snRNPs that contain fixed Sm rings assembled and matured via the SMN pathway and Cajal bodies. The U7 snRNP (Schümpelri and Pillai 2004) performs 3′-end processing of histone mRNAs and is highly conserved among metazoans. The U7 snRNA contains a divergent Sm site bound by an Sm ring similar in composition to that for the spliceosomal snRNPs but with Sm proteins Lsm10/Lsm11 substituted for SmD1/SmD2, with these new Sm proteins derived from SmD1 (or SmD3) and SmD2, respectively. Lsm11, in contrast to the largely passive role played by all other known Sm proteins found in Sm rings, contains an additional domain that is directly involved in histone mRNA processing (Pillai et al. 2003). The divergent Sm site in U7 snRNA is required for the SMN pathway to assemble the divergent Sm ring (Kolev and
In prokaryotes, Sm proteins form six- or seven-membered homomorphic rings with each member being an identical copy of the protein (Box 2 fig.). For example, Hfq from Escherichia coli forms 6-meric rings (Sauter et al. 2003), whereas the archaean Archaeoglobus fulgidus has two Sm proteins: Sm1, which forms a homomorphic 7-meric ring, and Sm2, which forms a homomorphic ring containing either six (via crystallography, Törö et al. 2002) or seven (via electron microscopy when complexed with RNA, Achsel et al. 2001) Sm2 proteins. Thus, apart from differences that may be artifacts arising from methods of structural determination, the diversity of Sm rings in prokaryotes is determined on a one-to-one basis by the diversity of Sm proteins (Achsel et al. 2001; Törö et al. 2002).

Several solved structures (e.g., Box 2 fig. A and D) show prokaryotic Sm rings complexed with U-rich RNA encircling one face of the central pore of the ring (PDB IDs: 1I5L, Törö et al. 2001; 1LOJ, Mura, Kozhukhovsky et al. 2003; 1M8V, Thore et al. 2003), with one structure showing RNA encircling as well as penetrating the central pore (1KQ2, Schumacher et al. 2002). Strong affinity for U-rich RNA oligomers has been confirmed in the archaean A. fulgidus (Törö et al. 2001) and affinity for A-rich regions, including the poly-A tails of transcripts, has been shown for the Sm protein Hfq in the Eubacteria E. coli (Hajnsdorf and Régnier 2000; Schumacher et al. 2002).
Steitz 2006), and Cajal bodies are involved in U7 snRNP maturation in the nucleus (Stanek and Neugebauer 2006).

The telomerase snRNP replicates chromosome ends in eukaryotes (Collins 2006), and there is evidence suggestive of metazoan telomerase snRNP hosting a fixed Sm ring at its core. Both human and yeast telomerase snRNA contain an Sm site around which Sm proteins assemble, yet curiously, in both humans and yeast only two Sm proteins have yet been identified, SmB and SmD3 in humans (Fu and Collins 2006), SmD1 and SmD3 in yeast (Seto et al. 1999). Given the composition of other Sm rings, it seems reasonable to expect that additional Sm proteins will eventually be found in these complexes. There is a great deal of divergence in maturation pathways among eukaryotic telomerase snRNPs, for example, ciliate telomerase does not contain Sm proteins (Collins 2006). Yeast telomerase snRNPs have been shown to be assembled in the cytoplasm and imported into the nucleus (Ferrezuelo et al. 2002; Teixeira et al. 2002), and in humans, telomerase components are found in Cajal bodies (Fu and Collins 2006), suggesting that maturation pathways in at least some metazoans are similar to those used by spliceosomal and U7 snRNPs.

A wide variety of eukaryotes process mono- and/or polycistronic transcripts via trans-splicing, in which a fragment of a “leader” RNA is spliced onto each cistron from a specialized splice leader (SL) snRNA found within a specialized SL snRNP (Hastings 2005). All SL snRNAs examined to date contain an Sm site (Mandelboim et al. 2003; Zeiner et al. 2004; Zhang et al. 2007) to which is bound an Sm ring that contains many if not all the same Sm proteins in the fixed Sm ring at the center of other spliceosomal snRNPs (Bruzik et al. 1988; Thomas et al. 1988; Palhi et al. 2000; Tkacz et al. 2007). A very curious consequence of trans-splicing is that the life cycle of the SL snRNP containing the SL snRNA is unusually short in comparison to that of other snRNPs anchored by fixed Sm rings (MacMorris et al. 2007). We will consider SL snRNPs further in the Discussion.

Flexible Class (Lsm-Type) Sm Rings

Flexible class eukaryotic Sm rings differ in several respects from fixed class Sm rings, the primary difference being a more fluid association with RNA substrates. Flexible Sm rings can assemble spontaneously and are stable in the absence of RNA (Achsel et al. 1999; Zoric et al. 2005). RNA substrates of flexible Sm rings lack Sm sites but do have U-rich tracts (Will and Lührmann 2001). When attached to an RNA substrate, a flexible Sm ring stabilizes and chaperones RNA as does a fixed Sm ring, yet a flexible Sm ring may be more easily associated and dissociated from RNA in some presumably specific manner. A flexible Sm ring may also play a somewhat more active role, in that its presence or absence may signal a transition in the life cycle of an snRNP or RNA substrate.

A characteristic member of the flexible class of Sm rings is the “Lsm ring” at the center of the U6 spliceosomal snRNP, containing the seven proteins Lsm2 through Lsm8 (Achsel et al. 1999; Mayes et al. 1999). Several aspects of the assembly and maturation of the U6 snRNP and its flexible Sm ring differ from that for other spliceosomal snRNPs. In addition to aggregating spontaneously in the absence of RNA (Achsel et al. 1999; Zoric et al. 2005), the U6-associated flexible ring appears to be transported into the nucleus as an assembled unit (Will and Lührmann 2001). The U6 snRNA is transcribed by RNA pol III and is never exported from the nucleus (Kunkel et al. 1986), and assembly of the final U6 snRNP appears to take place entirely within the nucleus (Will and Lührmann 2001). Though in at least some taxa proteins chaperone pre-U6 snRNP components, for example, the La protein with newly transcribed U6 snRNA in yeast (Xue et al. 2000), both ring assembly and ring–snRNA association steps in U6 snRNP maturation occur without use of the cytoplasmic SMN pathway characteristic of fixed Sm rings.

As details begin to emerge, additional functional differences serve to distinguish the flexible Sm ring at the heart of the U6 snRNP from the fixed Sm ring anchoring the other spliceosomal snRNPs (Karaduman et al. 2006). The flexible Sm ring binds to an U-rich region at the 3′-end of U6 snRNA (Achsel et al. 1999). The extensive base pairing of U4 and U6 snRNAs at the center of the U4/U6 di-snRNP is facilitated by conformational changes in U6 snRNA induced by the ring (Karaduman et al. 2006), and formation of...
the catalytic center of the spliceosome is critically dependent upon its subsequent dissociation from U6 snRNA (Chan et al. 2003). Following dissociation of the spliceosome, the nuclear retention of U6 snRNA and nuclear regeneration of U6 snRNP are both dependent upon the presence of the flexible Sm ring (Verdone et al. 2004; Spiller et al. 2007).

Other flexible Sm rings have similarly transient associations with their RNA substrates. In yeast, a flexible Sm ring containing the proteins Lsm1 through Lsm7 assembles...
in the absence of RNA (Zaric et al. 2005) and functions in cytoplasmic mRNA degradation (He and Parker 2000; Tharun et al. 2000; Bergman et al. 2007). The same flexible Sm ring at the center of U6 snRNP (Lsm2 through Lsm8) also plays a role in mRNA degradation within the nucleus (Kufel et al. 2004).

Flexible Sm rings are also common participants in RNA editing pathways, where the fluidity of their association with RNA substrates may be beneficial. Lsm7, for example, associate with several RNAs: snR5, a box H/ACA snoRNA for guiding site-specific modifications of rRNA (Fernandez et al. 2004); with pre-RNase P RNA (Salgado-Garrido et al. 1999); and perhaps with all rRNA, tRNA, and certain U3 snoRNA precursors (Kufel et al. 2002; Kufel, Allmag, Petfalski, et al. 2003; Kufel, Allmag, Verdone, et al. 2003). A flexible Sm ring composed of a possibly different set of Sm proteins is involved in binding U8 snoRNA which edits rRNA in Xenopus oocytes (Tomasevic and Peculis 2002). It is likely that additional experimental evidence will reveal yet more flexible Sm rings and more variations among species.

Discussion

The basal locations of prokaryotic Sm proteins within the tree provide clear evidence that current eukaryotic Sm proteins are derived from those in prokaryotes (Fig. 1). Since it was first proposed, the diversification–duplication model has been the primary model for the evolution of Sm protein family in eukaryotes (Salgado-Garrido et al. 1999). Our results provide support for some aspects of this model, but there are a number of reasons to believe the picture is considerably more complicated. Consistent with the diversification–duplication model, our phylogenetic analysis revealed that, except for a few more recently derived proteins such as SmY and SmN, the establishment of nearly all existing Sm proteins occurred prior to the last eukaryotic common ancestor (Fig. 1). However, weak support for deeper nodes within the tree makes specific assignment of neighbor relationships among Sm protein clades difficult. A few neighbor clades are robust and largely consistent with the diversification–duplication model (Lsm7 + SmG, Lsm8 + Lsm1, Lsm5 + SmE, Lsm3 + SmD2), but other clades are nested, and neighbor relationships among several clades shift with different data set compositions. It is difficult to reconcile this with the occurrence of a single large-scale duplication event in the family; several partial duplications involving one or a few proteins seem more likely.

The primary novelty in eukaryotic Sm rings is provided by the heteromorphic nature of the Sm ring itself. Although a homomorphic Sm ring as found within prokaryotes can provide spatial specificity for binding and interactions with the pore and faces only within the rotational span of a monomer (Fig. 3A), a heteromorphic Sm ring provides spatial specificity throughout the entire rotational sweep (Fig. 3B). This ensures precise conformational specificity of bound RNA and protein components within the snRNP. This specificity is important for the snRNA, as demonstrated by site-specific pairings of Sm-site uracils in the snRNA with Sm ring pore residues (Hartmuth et al. 1999; Urlaub et al. 2001; Wang et al. 2006) and interactions between snRNA stem–loop secondary structures and faces of the Sm ring (Stark et al. 2001), and for other protein–RNA and protein–protein interactions within the snRNP (Fig. 2B and Urlaub et al. 2000; Dybkov et al. 2006; Stark and Lührmann 2006). The maintenance of consistent three-dimensional structure and snRNA–protein membership of spliceosomal and other snRNPs is certainly dependent upon the heterogeneous nature of the eukaryotic Sm ring. An additional benefit of a heterogeneous Sm ring

![Image](https://example.com/image.png)
is that individual Sm proteins are free to develop functional side chains without repetition around the ring (Pillai et al. 2003).

With this view, substitutions of Sm proteins within Sm rings central to other snRNPs (e.g., the U7 snRNP, Pillai et al. 2001, 2003) reflect variations in spatial and conformational relationships among the differing RNA and protein cofactors and in the interactions of these partners with pore and faces of the Sm ring itself. The use of the “canonical” eukaryotic Sm ring (Fig. 2A and B) in six spliceosomal snRNPs (U1, U2, U4, U5, U11, and U12 of the minor spliceosome) that are themselves heterogeneous in form and function (Collins and Penny 2005; Russell et al. 2006; Will and Lu¨hrmann 2006) indicate that the relationship between snRNP and Sm ring is not exclusive. This suggests a common time of origin for these spliceosomal snRNPs, perhaps coincident with this form of fixed Sm ring.

In support of the necessity of steric specificity in Sm rings, we expect altered spatial conformations of snRNA and proteins within snRNPs anchored by nonstandard Sm rings, that snRNPs anchored by nonstandard Sm rings are more likely to attract novel proteins, that snRNPs containing nonstandard Sm rings may have limited functionality, and that smaller “minimal functional” snRNPs would be formed when anchored by standard Sm rings.

What might have been some of the early events leading from homomorphic Sm rings in prokaryotes to heteromorphic Sm rings in eukaryotes? The initial diversification may not have originated in prokaryotes; none of the proteins in the Archaea we examined with multiple (and self-selective) Sm proteins (P. aerophilum, A. pernin, T. volcanium, T. acidophilum, A. fulgidus) fell near each other in our tree (supplementary Fig. 1, Supplementary Material online). The “seeding” of Sm protein diversity may have occurred via contact between long-isolated Sm proteins that would have lost the “niche-exclusionary” ability to avoid binding with each other’s Sm folds. Such contact could have occurred via lateral transfer among Archaea or between Archaea and Eubacteria or could have coincided with genome clashes and/or gene transfers proposed to have accompanied the endosymbiotic origin of the eukaryotic cell itself (Koonin 2006; Martin and Koonin 2006). The evolved self-affinity necessary for the formation of the original homomorphic rings would still be in place, but competition from a novel Sm fold would have resulted in the formation of heteromorphic Sm rings. Such heteromorphic rings could have remained functional, for two reasons: 1) the novel rings would have continued to bind U-rich RNA, much as this ability is maintained in novel in vitro Sm complexes in yeast (Collins et al. 2003) and 2) because of low rotational specificity in the original homomorphic rings, any protein–protein or protein–RNA interactions required for function likely did not require the entire homomorphic ring and would have been able to occur against the original monomers (Mikułek et al. 2004). This could be tested by examining the in vitro behavior of novel mixtures of Sm proteins from closely and distantly related prokaryotes.

Once established, a heteromorphic Sm ring could have provided a template for further diversification, by allowing for greater steric specificity and thus greater complexity of Sm ring–associated snRNPs. Novel substitutions can occur with comparatively high frequency in heteromorphic Sm rings, as evidenced by clade-, species-, and even tissue-specific substitutions within eukaryotic Sm rings.

Finally, the heptameric structure of Sm rings could have accelerated the process of diversification. Consider a homomorphic heptamer, A7. A single substitution creates a 6 + 1 ring, A6B. For this structure to be stabilized, neighbor–neighbor relationships should evolve some specificity, but 7 is prime; there is no small multiple that eases this transition. In a hexamer, this could occur in pairs (AB)2 or with an additional Sm protein in triplets (ABC)2. Polymerization of such small subunits into novel hexamers and octamers has been observed in vitro for eukaryotic Sm proteins (Zaric et al. 2005). Thus, the most stable seven-member heteromorphic ring may be one that is entirely heteromorphic, with no repeated subunits.

Why do spliceosomal snRNPs contain two distinct Sm rings? The flexible Sm ring at the heart of the U6 spliceosomal snRNP is highly conserved in eukaryotes (Séraphin 1995; Mayes et al. 1999; Salgado-Garrido et al. 1999; Liu et al. 2004). There may be a functional basis derived from the formation of the U4/U6 di-snRNP, which then forms a tri-snRNP with the U5 snRNP in which RNA–RNA, RNA–protein and protein–protein contacts are all important (Vidal et al. 1999; Chan et al. 2003; Karaduman et al. 2006; Liu, Rauhut, et al. 2006).
3) participation of the SL snRNP in RNA associations. In conclusion, the evolutionary diversification of Sm rings coincided with the evolution of eukaryotes. We do not find much specific support for the diversification–duplication model for the development of separate fixed and flexible Sm rings. Fixed Sm rings such as those anchoring the majority of spliceosomal snRNPs and the U7 snRNP are stable, passive, noncatalytic, spatially specific protein scaffolds around which RNA and proteins that are active within snRNPs can organize. Flexible Sm rings share many of these characteristics but can be more freely associated and dissociated from RNA substrates. As further details concerning the diversity of RNA processing in

### Table 1
Characteristics distinguishing fixed from flexible Sm rings

<table>
<thead>
<tr>
<th>Trait</th>
<th>Fixed Sm ring</th>
<th>Flexible Sm ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example ring</td>
<td>SmB/D3/G/E/F/D2/D1 found in U1, U2, U4, U5</td>
<td>Lsm8/4/7/5/6/3/2 found in U6 snRNP</td>
</tr>
<tr>
<td>Alternate rings with Sm protein substitutions?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>RNA has Sm site?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>RNA-binding configuration</td>
<td>Passes through pore (?)</td>
<td>Encircles one face of pore (?)</td>
</tr>
<tr>
<td>Dedicated assembly pathway?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Assembles spontaneously in vitro?</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Dissociates from RNA during normal operation?</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2006). Especially during the formation of the U4/U6 dsnRNP, proper interactions may depend upon each snRNP containing entirely nonoverlapping sets of Sm proteins. The stabilization of these interactions was an early event in the evolution of the spliceosome (Collins and Penny 2005).

**Fixed and Flexible Sm Rings**

We have proposed nomenclature for classes of Sm rings—fixed (abbreviated “Fix”) and flexible (“Flex”) as replacements for the informal classes of Sm-type and Lsm-type, respectively—that is both reflective of the different functional roles played by each class and evocative of the manner in which they are associated and dissociated from their RNA cofactors (table 1). Our nomenclature is also free from potential confusions that may arise when an Sm ring such as that at the center of the metazoan U7 snRNP is composed of both “Sm”-prefixed and “Lsm”-prefixed protein monomers (Pillai et al. 2003). Some of these characteristics are shared by Sm proteins in Archaea, with one ring forming only in the presence of RNA and one forming spontaneously (Achsel et al. 2001). Proteins involved in fixed and flexible rings do not form monophyletic groups (Fig. 1), thus it may be these functional classes for Sm rings, broadly defined and suitably diversified, which have been maintained throughout the evolution of the Sm protein family, rather than the specific protein composition of the rings.

One currently somewhat speculative characteristic that distinguishes fixed from flexible eukaryotic Sm rings is the manner in which each class binds to associated RNA (table 1). The U1 snRNA appears to pass through the central pore of the fixed Sm ring at the center of the U1 snRNP (Fig. 2B, see also Stark et al. 2001; Urlaub et al. 2001; Stark and Lührmann 2006). Such a configuration would likely confer additional stability onto the fixed Sm ring–snRNA association, while increasing the cost of dissociation. In comparison, flexible Sm rings might be more easily dissociated from their RNA cofactors if the RNA circles around one face of the central pore without passing through, as appears to be common in prokaryotic Sm–RNA associations.

The Sm ring that anchors the SL snRNP in organisms that use trans-splicing is an anomalous fixed class ring with rapid association and dissociation from its RNA substrate. The steps required for this are 1) assembly of a fixed Sm ring around SL snRNA, 2) maturation of the SL snRNP, 3) participation of the SL snRNP in trans-splicing, 4) removal of non-Sm proteins from the SL snRNP, and 5) dissociation of the fixed Sm ring from the “spent” SL snRNA. Details of most of these steps are unknown (MacMorris et al. 2007). In trypanosomes, the fixed Sm ring at the center of the SL snRNP appears to be identical in protein composition to the one at the center of the U1, U4, and U5 snRNPs, so some particular feature of the SL snRNP allows this rapid turnover. In trypanosomes and other organisms with trans-splicing, there may be a dedicated pathway that recognizes spent SL snRNPs; though it seems likely to exist, evidence of such a pathway has not yet been found.

SL snRNPs represent a sort of functional intermediate between the longer term associations of fixed Sm rings and the more transient associations typical of flexible Sm rings. For the evolution of trans-splicing, these stumbling blocks would have to be removed. The use of a fixed Sm ring in SL snRNPs may indicate that the SL snRNP is derived from one of the other spliceosomal snRNPs and that trans-splicing is itself derived from ancestral cis-splicing, perhaps multiple times in several eukaryotic lineages (Nilsen 2001; Hastings 2005). The maintenance of the fixed Sm ring at the center of all SL snRNPs may represent simply a contingent characteristic or may indicate the need to have a fixed Sm ring at the heart of an SL snRNP, perhaps to provide greater stability during interactions and reconfigurations involving other spliceosomal snRNPs.

Recently, a number of multidomain RNA-associated proteins found in a wide range of eukaryotes contain divergent but identifiable Sm domains (Albrecht and Lengauer 2004; Albrecht et al. 2004; Tadauchi et al. 2004; Fleischer et al. 2006; Yang et al. 2006; Tritschler et al. 2007). Unlike Lsm11, which contains an additional domain functional in processing of histone mRNAs (Pillai et al. 2003), none of these proteins have been observed in any form of Sm ring. However, they may yet be found in Sm rings or may interact via their Sm domains to form novel multimeric complexes (Fleischer et al. 2006).

In conclusion, the evolutionary diversification of Sm rings coincided with the evolution of eukaryotes. We do not find much specific support for the diversification–duplication model for the development of separate fixed and flexible Sm rings. Fixed Sm rings such as those anchoring the majority of spliceosomal snRNPs and the U7 snRNP are stable, passive, noncatalytic, spatially specific protein scaffolds around which RNA and proteins that are active within snRNPs can organize. Flexible Sm rings share many of these characteristics but can be more freely associated and dissociated from RNA substrates. As further details concerning the diversity of RNA processing in
eukaryotes are found, it is likely that additional forms of both fixed and flexible rings will be discovered.

Supplementary Material

Supplementary Table 1 and Figure 1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by National Science Foundation grants DBI-0434671 to D.G.S. and MCB-0342431 to M.L. D.G.S. thanks D. A. Campbell for helpful discussions and two anonymous reviewers for helpful comments that greatly improved the manuscript.

Literature Cited


Downloaded from https://academic.oup.com/mbe/article-abstract/25/11/2255/1084900 by guest on 15 June 2020
correlates with proper assembly of core U snRNP particles.


Hastings KEM. 2005. SL trans-splicing: easy come or easy go?


Laura Katz, Associate Editor

Accepted July 12, 2008