Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis

Adelaine K. W. Leung[1], **Kiyoshi Nagai**, and **Jade Li**
MRC Laboratory of Molecular Biology Hills Road, Cambridge CB2 0QH England

Abstract

The spliceosome is a dynamic macromolecular machine that assembles on pre-mRNA substrates and catalyses the excision of non-coding intervening sequences (introns)[1-3]. Four of the five major components of the spliceosome, U1, U2, U4 and U5 snRNPs, contain seven Sm proteins (SmB/B', SmD1, SmD2, SmD3, SmE, SmF and SmG) in common[4,5]. Following export of the U1, U2, U4 and U5 snRNAs to the cytoplasm[6,7], the seven Sm proteins chaperoned by the survival of motor neurons (SMN) complex assemble around a single-stranded, U-rich sequence called the Sm site in each snRNA, to form the core domain of the respective snRNP particle[8,9]. Core domain formation is a prerequisite for re-import into the nucleus[10], where these snRNPs mature via addition of their particle-specific proteins. Here we present a crystal structure of the U4 snRNP core domain at 3.6 Å resolution, detailing how the Sm site heptad (AUUUUUG) binds inside the central hole of the heptameric ring of Sm proteins, interacting one-to-one with SmE-SmG-SmD3-SmB-SmD1-SmD2-SmF. An irregular backbone conformation of the Sm site sequence combined with the asymmetric structure of the heteromeric protein ring allows each base to interact in a distinct manner with four key residues at equivalent positions in the L3 and L5 loops of the Sm fold. A comparison of this structure with the U1 snRNP at 5.5 Å resolution[11,12] reveals snRNA-dependent structural changes outside the Sm fold, which may facilitate the binding of particle-specific proteins that is crucial to biogenesis of spliceosomal snRNPs.

Proteins in the Sm family are characterized by Sm1 and Sm2 motifs joined by a variable linker[13-15] (Fig. S1). SmB/B', SmD1 and SmD3 contain extended C-termini whereas SmD2 and SmE contain extended N-termini. The Sm-fold consists of an N-terminal α-helix and a five-stranded antiparallel β-sheet containing the Sm motifs and folded upon itself[16] (Fig. S2). The subunit interfaces in the SmD1-SmD2 and SmD3-SmB hetero-dimers suggest that the seven Sm proteins are assembled in a ring[16] in the snRNP core domain[17,18], and this has been confirmed by the crystal structure of the U1 snRNP at 5.5 Å resolution[11,12]. Crystal structure of *A. fulgidus* Lsm-1 homo-heptamer in complex with penta-uridylate showed how Lsm-1 provides U-specificity[19,20]. However, crystal structures of U1 snRNP at 5.5 or 4.4 Å resolution have left the side chain interactions between the Sm site and Sm proteins unresolved[11,12,21]. Hence, it remains unknown how the heteromeric Sm proteins combine to specifically recognize the Sm site sequence.

We solved the structure of the U4 snRNP core domain (Table S1) assembled on a fragment of U4 snRNA (Fig. S3 and S4), which crystallised[22] with 12 copies in the asymmetric unit (Fig. S5). The seven Sm proteins form a ring with a relatively flat face over which the N-terminal helices lie, and a tapered face carrying the L4 loops (Fig. 1a). U4 snRNA threads through the central hole lined by loops L3, L5 and L2 (Fig. 1c). The Sm site sequence is bound around the inner wall near the rim on the flat face. Its phosphates are exposed in the hole (Fig. 1b) revealing an irregular backbone conformation (Fig. 2), and its bases projected into the Sm proteins bound by their loops L3 and L5 also vary in its orientation from nearly parallel to nearly perpendicular to the ring plane (Fig 2 and S6). The 5′ flanking adenine (A118) is outside the hole, and the 5′ stem (U4 SL-II) makes little contact with the Sm proteins (Fig. 1a). The 3′ flanking nucleotides (A126-C127) traverse the hole with phosphates contacting SmD1 and SmB, and bases contacting L2 of SmD2 and SmF and L5 of SmF. As the 3′ stem (U4 SL-III) emerges on the tapered face with the helical axis roughly 60° to the plane of the ring, the first base pair (C127:G144) comes into contact with Trp-25F in loop L2. The 3′ unpaired nucleotide (G145) is wedged between SmE and SmF. Along the 3′ stem the phosphates on both strands interact with basic residues from L2 and L4 loops of all the Sm proteins except SmG, particularly those from the Lysine-rich, long L4 loops of SmD2 and SmB (Fig. S7). These interactions can support the association between Sm proteins and different snRNAs during core domain assembly[23],[24].

Small shape differences over the β–sheet of different Sm proteins (Table S2), attributable to the size differences of the conserved inward pointing residues between the corresponding Sm1 and Sm2 motifs (Fig. S1), cause the Sm protein ring to be asymmetric. The L3 and L5 loops are held at different heights and orientations relative to the plane of the ring (Fig. S8), allowing their key residues to contact each nucleotide uniquely. The nucleotide-binding loops L3 and L5 have the consensus sequences of Asp-hydrophilic-aromatic-Met-Asn (residues L3.1-L3.5) and Arg-Gly-(acidic/Asn) (residues L5.1-L5.3) (Table S3). In the co-crystal structure of *A. fulgidus* Lsm-1 with penta-uridylate, U specificity is achieved by sandwiching the uridine base between the side chains of His-37(L3.3) and Arg-63(L5.1), and hydrogen-bonding of its N3 and O4 atoms respectively with Oδ1 and Nδ2 of the invariant Asn-39(L3.5)[19]. In the U4 core domain (Fig. 3), however, the base stacking with the aromatic residue (L3.3) present in five of the Sm proteins, and the interaction with the Arg/Lys (L5.1) vary along the heptad and from the Lsm-1 example (Fig. 3 and S9), due to the irregular RNA conformation around the central hole (Fig. 2 and S6) and the sequence variations in L3 and L5 (Table S3). We have inferred hydrogen-bonding interactions around the Sm site from the residue positions, corroborated by unambiguous orientation of base planes and aromatic side chains in the electron density, and the non-crystallographic symmetry (ncs) agreement in the conformation of the conserved contact residues (Fig. 3).

At the first position of the heptad an adenine is required, since replacement by G abolished core domain assembly with a U4 oligonucleotide, and replacement by U destabilized the assembly[23]. The A119 base is stacked with Tyr-53E(L3.3) and hydrogen-bonded at N3 to Lys-80E(L5.1), at N6 to Asp-51E(L3.1) and at N1 to Asn-55E(L3.5) (Fig. 3a), demonstrating A-specificity of its binding pocket. A salt bridge between its phosphate and Arg-61D2, which is hydrogen-bonded to Asp-37F (Fig. 3g), stabilizes A119 binding to the ring. In U1 snRNP and the reconstituted U4 core domain, the N7 atom of this adenine is unexpectedly nucleophilic in becoming methylated by dimethylsulphate. It indicates a perturbation of the π-electron distribution over its double ring system[25], which could result from multiple hydrogen bonds to the base.

The U120 base is sandwiched between the side chains of Phe-37G and Arg-63G and hydrogen-bonded at O4 with Nδ2 of Asn-39G (Fig. 3b). Similarly, the U122 base is stacked with His-37B and hydrogen-bonded at O4 to Asn-39B (Fig. 3d), except that a different base

orientation relative to L3 requires His-37B to adopt a different rotamer to achieve stacking (Fig. S9). These interactions are similar to the Lsm-1 complex[19] and account for the cross-linking of the first and third U to L3 residues of SmG and SmB, respectively[26]. SmD3 and SmD1 lack the aromatic residue at L3.3. SmD3 displays U specificity in a novel mode distinct from LSm-1[19]: U121 is hydrogen-bonded on O4 to both Asn-38D3 and Asn-40D3, on N3 to Asn-40D3 and on O2 to the peptide amide of Ser-66D3, besides being stacked with Arg-64D3 (Fig. 3c). U123 forms no stacking interaction (Fig. 3e). Its base is within hydrogen-bonding distance from the side chain of the invariant Asn-37D1, which is positioned to form the conserved buttressing hydrogen bonds with Asp-33D1[16,19,20]. This configuration is consistent with U123 adopting an enol tautomer that would present O2(H) and N3 as the hydrogen-bond donor and acceptor, respectively, to Oδ1 and Nδ2 of Asn-37D1. Consequently U123 is accommodated without U-specific base contacts, which explains the lack of preference for U at this position (Fig. S3e). In human U1 snRNA G replaces U, and in U1 snRNA of other species all four bases are tolerated (http//:rfam.sanger.ac.uk)[27]. U124 is hydrogen-bonded on O2, not O4, to the invariant Asn-64D2 (Fig. 3f), and therefore our structure cannot fully account for the preference for U at this position. His-62D2 while stacking with U124 also interacts edge-to-face with the 5′-flanking Adenine, A118 (not shown) in some ncs copies. The G125 base is not intimately associated with SmF, as it is stacked only on one edge between Tyr-39F and Arg-65F, and too distant for hydrogen-bonding with the invariant Asn-41F (Fig. 3g). Cys-66F, which is absolutely conserved in SmF, replaces the Gly in L5 without causing a clash. The absence of G-specific base contacts explains why replacing this G with A had no effect on Sm protein binding[23], and it is consistent with replacement of this G (Fig. S3e) by other bases in U4 and U5 snRNAs of different species[27,28]. Thus, the last nucleotide of the Sm site heptad acts as a transition to the variable 3′ stem. In the bound heptad the phosphate groups of U120, U121 and U122 are in close proximity. Their negative charge density is likely stabilized by electrostatically held magnesium ions of indeterminate positions[29] or by chelated cations unresolvable at our resolution.

In mammalian U4 and U5 snRNAs the Sm site heptad and the 3′ stem are linked by a single nucleotide, whereas in U1 and U2 snRNA they are linked by five nucleotides[27,28]. Structural comparison between U1 snRNP[11,12] and the U4 core domain shows that the snRNAs emerge from the central hole in similar positions (red circle in Fig. 4a and 4b) but with such different stem orientations that their 3′ termini fall on opposite sides of the hole. In U1 snRNP, the N-terminal 70 residues of U1-70K wrap around the tapered face of the core domain by skirting around the RNA stalk passing SmD3-SmG-SmE-SmF-SmD2[11,12]. Mapping the U1-70K N-terminal fragment onto the U4 core domain reveals that the 3′ strand of U4 SL-III would obstruct its polypeptide path (Fig. 4c) and hence prevent its binding to the snRNP. The N-terminal 97-residue fragment of U1-70K is sufficient to bind to the U1 core domain but fails to bind to the U5 core domain[30]. In U5 snRNA the 3′ stem is linked to the Sm site as closely as in U4 (Fig. S3c and S3d) and could exclude U1-70K analogously.

The N-terminal extension of SmD2 and C-terminal extension of SmD1 are disordered in the absence of RNA[16]. In U1 snRNP, SmD2 forms an extra helix (helix 0) at the N-terminus that points into the minor groove of RNA helix H (Fig. 4a). This anchors the SmD2 helix 1, whose C-terminus interacts with U1-70K[11,12,21]. In the U4 core domain, the C-terminal extension of SmD1 forms helix 3, while the N-terminal extension of SmD2 forms helix 0 (Fig. 4b) in the ncs copies where the loop between SmD2 helices 0 and 1 interacts with the backbone of the 5′ stem (Fig. 1a). The U4 SmD1 helix 3 interacting with the U4 SmD2 helix 1 in its snRNA-dependent orientation might also obstruct U1-70K (Fig. 4b). In U1 snRNP, the SmB helix 1 interacts with the backbone of stem II[11]. These snRNA-dependent structural differences on the flat face of the core domain may, in addition to the snRNA

itself, provide selectivity for the cognate particle-specific proteins and play a crucial role in snRNP biogenesis.

METHODS SUMMARY

The U4 snRNP core domain was reconstituted from the seven Sm proteins and a variant of human U4 snRNA[16],[22]. Crystals in space group $P3_1$ with 12 complexes per asymmetric unit were grown by vapour diffusion and diffracted X-rays anisotropically to 3.45 Å resolution. Initial phases were determined by the MAD method using SeMet substitution within Sm sub-complexes. The structure containing 8101 protein and RNA residues has been refined under 12-fold ncs restraints at 66.2-3.6 Å resolution to R_{free} of 32.1% with excellent geometry (Table S1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Burge, CB.; Tuschl, T.; Sharp, PA. The RNA World II. Gesteland, RR.; Cech, TR.; Atkins, JF., editors. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1999. p. 525-560.

2. Will, CL.; Lührmann, R. The RNA World. Gesteland, RF.; Cech, TR.; Atkins, JF., editors. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2006. p. 369-400.

3. Yu, Y-T.; Scharl, EC.; Smith, CM.; Steitz, JA. The RNA World II. Gesteland, RR.; Cech, TR.; Atkins, JF., editors. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1999. p. 487-524.

4. Hinterberger M, Pettersson I, Steitz JA. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. J. Biol. Chem. 1983; 258:2604–2613. [PubMed: 6185498]

5. Bringmann P, Lührmann R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J. 1986; 5:3509–3516. [PubMed: 2951249]

6. Mattaj IW. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell. 1986; 46:905–911. [PubMed: 2944599]

7. Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell. 2000; 101:187–198. [PubMed: 10786834]

8. Meister G, Eggert C, Fischer U. SMN-mediated assembly of RNPs: a complex story. Trends Cell Biol. 2002; 12:472–478. [PubMed: 12441251]

9. Pellizzoni L, Yong J, Dreyfuss G. Essential role for the SMN complex in the specificity of snRNP assembly. Science. 2002; 298:1775–1779. [PubMed: 12459587]

10. Fischer U, Sumpter V, Sekine M, Satoh T, Lührmann R. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. EMBO J. 1993; 12:573–83. [PubMed: 7679989]

11. Pomeranz Krummel DA, Oubridge C, Leung AKW, Li J, Nagai K. Crystal structure of the human spliceosomal U1 snRNP at 5.5 Å resolution. Nature. 2009; 458:475–480. [PubMed: 19325628]

12. Oubridge C, Krummel D. A. Pomeranz, Leung AKW, Li J, Nagai K. Interpreting a low resolution map of human U1 snRNP using anomalous scatterers. Structure. 2009; 17:930–938. [PubMed: 19604473]

13. Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, Brahms H, Lührmann R. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. EMBO J. 1995; 14:2076–2088. [PubMed: 7744013]

14. Seraphin B. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. EMBO J. 1995; 14:2089–2098. [PubMed: 7744014]

15. Cooper M, Johnston LH, Beggs JD. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. EMBO J. 1995; 14:2066–2075. [PubMed: 7744012]

16. Kambach C, Walke S, Young R, Avis JM, de la Fortelle M, Raker VA, Lührmann R, Li J, Nagai K. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell. 1999; 96:375–387. [PubMed: 10025403]

17. Kastner B, Lührmann R. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5′ RNA terminus. EMBO J. 1989; 8:277–286. [PubMed: 2469573]

18. Kastner B, Bach M, Lührmann R. Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. Proc. Natl. Acad. Sci. U S A. 1990; 87:1710–1714. [PubMed: 2137927]

19. Törö I, Thore S, Mayer C, Basquin J, Séraphin B, Suck D. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J. 2001; 20:2293–2303. [PubMed: 11331594]

20. Thore S, Mayer C, Sauter C, Weeks S, Suck D. Crystal Structures of the Pyrococcus abyssi Sm Core and Its Complex with RNA. Common features of RNA binding in Archaea and Eukarya. J. Biol. Chem. 2003; 278:1239–1247. [PubMed: 12409299]

21. Weber G, Trowitzsch S, Kastner B, Lührmann R, Wahl MC. Functional organization of the Sm core in the crystal structure of human U1 snRNP. EMBO J. 2010; 29:4172–4184. [PubMed: 21113136]

22. Leung AKW, Kambach C, Kondo Y, Kampmann M, Jinek M, Nagai K. Use of RNA tertiary interaction modules for the crystallization of the spliceosomal snRNP core domain. J. Mol. Biol. 2010; 402:154–164. [PubMed: 20643141]

23. Raker VA, Hartmuth K, Kastner B, Lührmann R. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. Mol. Cell Biol. 1999; 19:6554–8565. [PubMed: 10490595]

24. McConnell TS, Lokken RP, Steitz JA. Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. RNA. 2003; 9:193–201. [PubMed: 12554862]

25. Hartmuth K, Raker VA, Huber J, Branlant C, Lührmann R. An unusual chemical reactivity of Sm site adenosines strongly correlates with proper assembly of core U snRNP particles. J. Mol. Biol. 1999; 285:133–147. [PubMed: 9878394]

26. Urlaub H, Raker VA, Kostka S, Lührmann R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. EMBO J. 2001; 20:187–196. [PubMed: 11226169]

27. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. Rfam: updates to the RNA families database. Nucleic Acids Res. 2009; 37:D137–140. http://rfam.sanger.ac.uk; http://janelia.org.

28. Guthrie C, Patterson B. Spliceosomal snRNAs. RNA. 2003; 9:193–201. [PubMed: 12554862]

29. Draper DE. RNA folding: thermodynamic and molecular descriptions of the roles of ions. Biophys. J. 2008; 95:5489–5495. [PubMed: 18835912]

30. Nelissen RL, Will CL, van Venrooij WJ, Lührmann R. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. EMBO J. 1994; 13:4113–4125. [PubMed: 8076607]
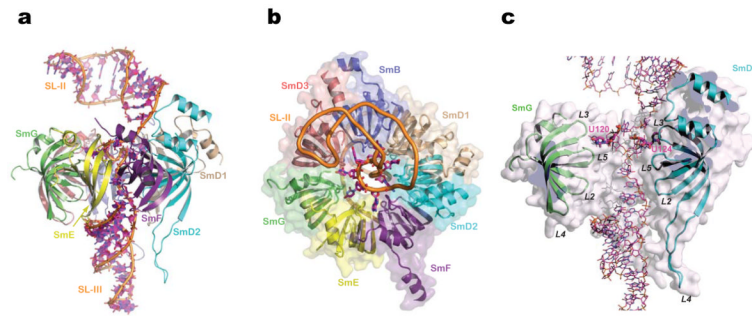
Figure 1. Overall structure of the U4 snRNP core domain
a, Side view of the core domain showing the ring with its flat face up and tapered face down. **b**, View from the flat face of the ring. The N- and C-terminal extensions of the Sm fold interact between SmD3 and SmB, and between SmD1 and SmD2. **c**, the heptameric ring is cleaved along a plane (dark blue patches) through SmG and SmD2, leaving the five subunits SmG-SmD3-SmB-SmD1-SmD2 that bind the penta-uridylate to form the protein envelope in the background. Loops L3, L5 and L2 of the Sm fold line the walls of the funnel shaped hole, whereas L4 is exposed on the tapered face. The bases of the Sm site nucleotides, such as U120 and U124, are bound between L3 and L5 near the rim on the flat face.
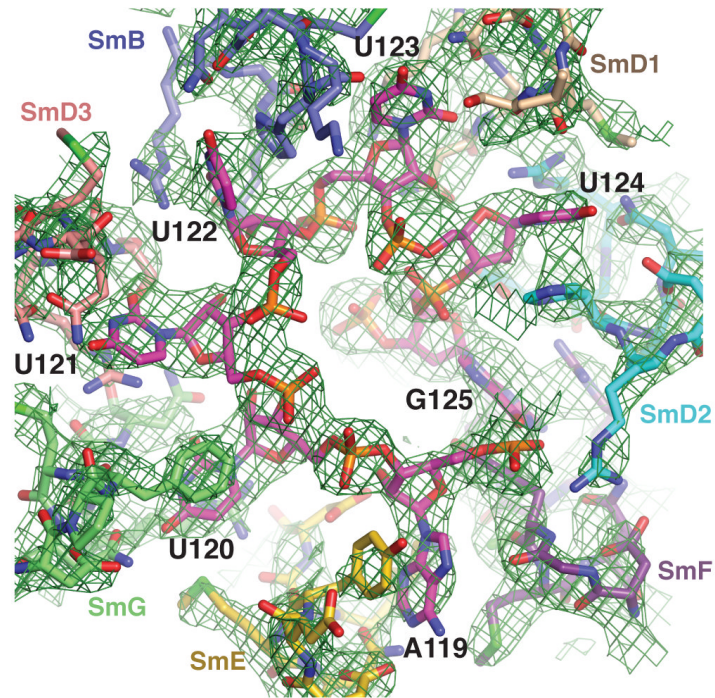
Figure 2. The Sm site RNA binds asymmetrically in the central hole of the heptamer ring
The Sm site heptad sequence is shown together with Sm proteins in the central hole of the core domain. Carbon atoms are colour-coded by chain: SmE (yellow), SmG (green), SmD3 (salmon), SmB (light blue), SmD1 (tan), SmD2 (cyan), SmF (purple), RNA (magenta). Nitrogen, oxygen, sulphur and phosphorus atoms are in blue, red, green and orange, respectively. A sharpened (B=−15 Å^2) ncs-averaged electron density map is contoured at 8σ.
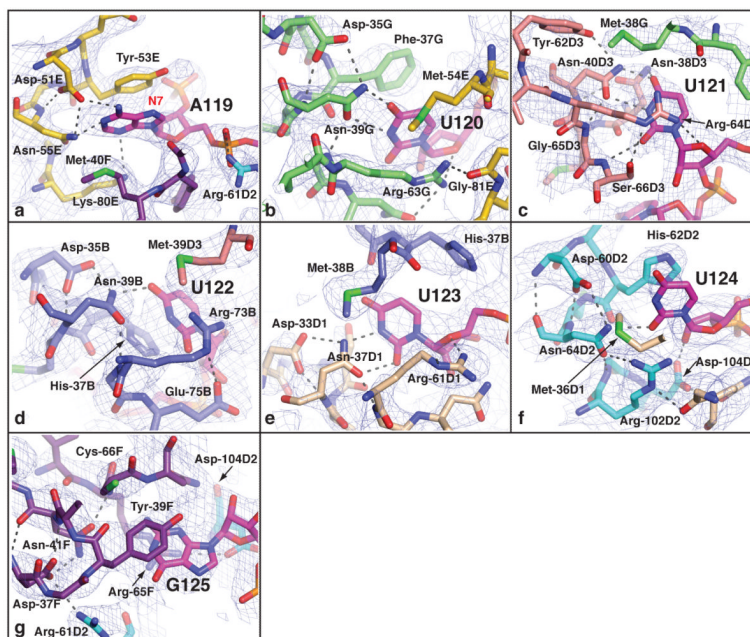
Figure 3. Interactions between the U4 Sm site heptad nucleotides and the Sm proteins
a, A119. **b**, U120. **c**, U121. **d**,U122. **e**, U123. **f**, U124. **g**, G125. Shown in dashed lines are the hydrogen-bonding interactions inferred from the residue configurations. Similarities with the RNA-free hetero-dimers[16] and the Lsm-1 complex[19] are present: the invariant Asn(L3.5) is buttressed by hydrogen bonds with the side chain of Asp(L3.1) and peptide amide of Gly(L5.2) in six cases (**a-b, d-f**), and with Glu-36D3(L3.1) via Tyr-62D3 and with the peptide amide of Gly-65D3(L5.2) in SmD3 (**c**); Met(L3.4) contributes van der Waals contacts to the base bound by a neighbouring Sm protein in six cases (**a-f**), and Gly(L5.2) is conserved in six cases (**a-f**) where a side chain would clash with the base contacting residues. The 2Fo-Fc map is shown sharpened with B=−15 Å^2 and contoured at ~1.5σ. Atom colours are as in Figure 2.
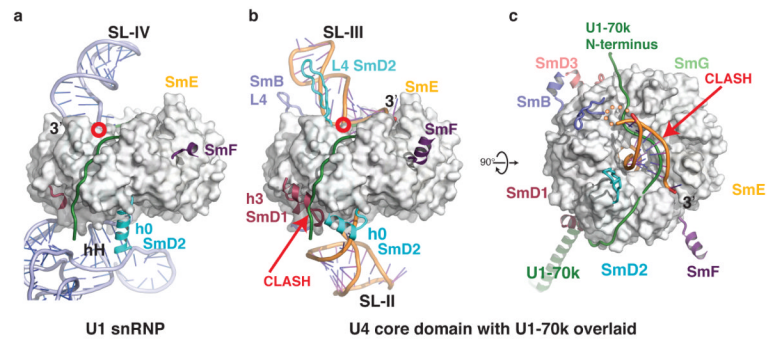
Figure 4. snRNA-dependent structural changes of the U1 and U4 core domains
a, U1 snRNP[11] and **b**, U4 core domain in the same orientation; **c**, the U4 core domain in tapered-face view. The N-terminal fragment of U1-70K (green)[11,12] is overlaid onto the U4 core domain (**b** and **c**), and the Sm folds common to both structures are masked with a white envelope. The red circle indicates where U1 and U4 snRNAs come out from the central hole. In U4, the L4 loops of SmB and SmD2 contact the backbone of the 3′ stem (**b**). In U1[11,21], SmD2 helix h0 points into the minor groove of RNA helix H (hH) (**a**), but in U4, SmD2 h0 is orientated almost orthogonal to this, with its N-terminus and the C-terminus of the masked SmD2 h1 both pointing at SmD1 helix h3 (**b**). The latter is positioned to obstruct the path of U1-70K (arrow). Moreover the first seven base pairs of the 3′-stem of U4 snRNA (orange) would clash on its 3′-strand with U1-70K (arrow) (**c**).