
In the beginning was the U1A protein: a personal reflection

KIYOSHI NAGAI

Structural Studies Division, MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, United Kingdom

When I started my own research group at the MRC Laboratory of Molecular biology in 1987 I was told to initiate a completely new project. I was fascinated by the structures of DNA-protein complexes but this research field was already getting overcrowded. In 1989 the group of Tom Steitz published the structure of glutamyl-tRNA synthetase in complex with its cognate tRNAs and I became very interested in RNA-protein interactions. I came across an interesting review article by Gideon Dreyfuss on “RNA recognition motif (RRM)” and this small protein domain seemed like a good target for structural studies. We first tried to express the RRM of U1-70k based on the *Cell* paper published by Charles Query in Jack Keene’s lab but our first attempt failed as the expressed protein was not well behaved. Our attempt with U1A was successful and we reported the structure of the first RRM in 1990 and Chris Oubridge managed to crystallize and solved the structure of its complex with stem-loop II of U1 snRNA in 1994. Interestingly U2 snRNP contains a protein called U2B’ which is highly homologous to U1A. Iain Mattaj’s group showed that U2B’ binds to stem-loop IV of U2 snRNA with a loop sequence similar to that of the U1A binding site but only in complex with U2A’, a Leu-rich repeat protein. They made chimeras of U1A and U2B’ and showed which amino acid residues are important for the binding of the cognate RNA hairpin and discriminating against the non-cognate RNA hairpin. A student in my group, Stephen Price, crystallized the ternary complex between U2B’, U2A’, and U2 snRNA stem-loop IV and solved its structure. The structure of the U2B’-U2A’-RNA complex and the U1A-RNA complex revealed how these two proteins bind their cognate RNA hairpin in atomic detail and the structures provided important insight into a fascinating mechanism of RNA recognition.

Even though I did not think we could make any meaningful contribution to our understanding of the molecular mechanism of splicing by solving structure of its individual components I decided to work on “the spliceosome.” Joan Steitz and others showed that some auto-immune antibodies bind to protein components of small nuclear ribonucleoprotein particles (snRNPs), large building blocks of the spliceo-

some. Importantly they showed that proteins now known as Sm proteins are present in U1, U2, U4, and U5 snRNPs in common. In those days cDNA sequences were cloned by screening λgt11 (cDNA expression) library with antibodies or by hybridizing degenerate DNA oligo-nucleotides designed on the basis of partial protein sequences. The cDNA sequences of only SmB, SmD, and SmE were known. The group of Reinhard Lührmann discovered that the band on a SDS gel, thought to be SmD, was in fact three proteins (named SmD1, SmD2, and SmD3) running on top of each other. By 1995 they cloned the remaining Sm protein genes and found that the seven Sm proteins exist as three sub-complexes in the absence of snRNAs. The groups of Jean Beggs, Reinhard Lührmann, and Bertrand Séraphin independently discovered that these Sm proteins contain a common sequence motif now known as the Sm motif and published papers back to back in *EMBO Journal* in 1995. Using this knowledge and the cDNA sequences provided by Reinhard Lührmann, my postdocs and student, Christian Kambach, Stephan Walke and Robert Young, managed to express all seven Sm proteins as three sub-complexes in *Escherichia coli* and solved the structure of the SmD1-SmD2 and SmB-SmD3 complexes. By assembling these structures on computer graphics we found that seven Sm proteins could form a ring similar to that observed by negative stain electron microscopy by Lührmann and colleagues. Ten years later Adelaine Leung and Jade Li solved the structure of the U4 snRNP core domain and showed that our earlier model is correct. In the late 1990s Berthold Kastner in Lührmann’s group and I attempted crystallizing U1 snRNP purified from HeLa cell nuclear extract. I was able to experience the whole process of U1 snRNP purification and set up crystallization with Berthold both in Marburg and Cambridge. Most crystallization drops were clear or had heavy precipitates but Berthold found few crystalline-like objects and we took a crystallization tray in a polystyrene box all the way from Marburg to Daresbury synchrotron near Manchester to find disappointingly that they were salt crystals. My group also pursued a recombinant approach and by early 2000s we were able to make all the protein components of U1 snRNP in *E. coli* and reconstitute U1 snRNP. In 2009

Corresponding author: kn@mrc-lmb.cam.ac.uk

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050344.115>. Freely available online through the RNA Open Access option.

© 2015 Nagai This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Daniel Pomeranz Krummel and Chris Oubridge managed to crystallize and obtain an electron density map of a functional core of U1 snRNP at 5.5 Å resolution and we were able to fit known structures into the map and show what U1 snRNP looks like. Berthold and Reinhard, later joined by Markus Wahl and Gert Weber, continued their effort to crystallize native U1 snRNP purified from HeLa cell nuclear extract over ten years and managed to crystallize it after treating it with chymotrypsin to remove floppy parts which prevented crystallization. Their paper describing the structure at 4.5 Å resolution was published in *EMBO Journal* a year later. I really admired their persistence in solving the problem. Chris Oubridge, Yasushi Kondo, and Marike van Roon continued our effort to obtain much better ordered crystals of U1 snRNP but this proved very difficult. Recently they managed to obtain well diffracting crystals of U1 snRNP in two parts and revealed nearly the whole structure of human U1 snRNP in atomic detail. It was particularly rewarding to see how U1 snRNP recognizes the 5' splice site of pre-mRNA.

In the early 2000s we started to invest our effort on other snRNPs and in 2013 we solved the crystal structure of two large U5 snRNP proteins, Prp8 and Brr2. Andy Newman and Chris Norman showed that Prp8 could be cross-linked to critical positions of pre-mRNA, U5 and U6 snRNAs suggesting that Prp8 is in close contact with the catalytic RNA core. Andy Newman and I worked on the expression and purification of these proteins for a long time but when two students, Wojtek Galej and Kelly Nguyen, joined my group they made tremendous progress. Wojtek succeeded in crystallizing a large fragment of Prp8 in complex with a U5 snRNP assembly factor called Aar2 and solved its structure. The group of Jean Beggs showed that Prp8 and Aar2 form a stable complex in yeast cytoplasm and are imported into the nucleus together with Snu114 and the U5 snRNA core domain where Aar2 is replaced by Brr2. Wojtek showed the large domain of Prp8 consists of Reverse Transcriptase-like, linker and type II endonuclease-like domains. This domain architecture

is reminiscent of Group II intron encoded proteins. The Reverse Transcriptase-like domain consists of the finger, palm, and thumb domain as first observed by Tom Steitz and colleagues in DNA polymerase. Dlakic and Mushegian had pointed out a significant sequence similarity between the thumb domain and the thumb/maturase X (Th/X) domain of fungal group II intron reverse transcriptase. This sequence similarity and our structure suggest that Prp8 and Group II intron encoded proteins, which play a role in intron mobility, share a common evolutionary origin. In the mid-1980s Phil Sharp and Tom Cech proposed that nuclear pre-mRNA splicing and Group II intron self-splicing might have a common evolutionary origin because both excise introns via a lariat intron intermediate. Sebastian Fica and others in the lab of Jon Staley and Jo Piccirilli identified ligands for the catalytic divalent ions in the spliceosome and showed that the active site of the spliceosome is structurally very similar to that of a group II self-splicing intron revealed by Anna Marie Pyle and Navtej Toor. Kelly Nguyen solved the structure of yeast Brr2 helicase in complex with the Jab1/MPN domain of Prp8 and the group of Markus Wahl and Reinhard Lührmann solved the structure of human counterpart. Kelly and Wojtek were able to show why Aar2 and Brr2 cannot bind to Prp8 simultaneously based on their structures.

Immediately after Venki Ramakrishnan gave an internal seminar at the LMB to unveil the crystal structure of the 30S ribosomal subunit, Max Perutz and Aaron Klug both said that they did not expect to see the structure of the ribosome in their lifetime. When the RNA Society came into existence 20 years ago only a handful of spliceosomal proteins had been characterized and U1A was the only component whose structure was known. Thanks to genome sequencing and mass-spectrometry we now know the complete protein repertoire of the spliceosome and our structural knowledge is advancing day-by-day. Within the next ten years our understanding of the molecular mechanism of pre-mRNA splicing will probably reach the level of the ribosome.