Experimental report

27/05/2024

Proposal: 8-03-1062 **Council:** 10/2022

Title: Structural analysis of 3' splice site recognition by U2AF/SF1 during early spliceosome assembly

Research area: Biology

This proposal is a new proposal

Main proposer: Clara MORGUET

Experimental team: Frank GABEL
Clara MORGUET
Matthias Johannes Gerhard BRANDL
Santiago MARTINEZ LUMBRERAS

Local contacts: Anne MARTEL

Samples: C4943 H7728 N1422 O1646 S51 P40 Zn3

Instrument	Requested days	Allocated days	From	To
D22	2	2	02/06/2023	04/06/2023

Abstract:

In metazoan organisms, the proteome diversity is greatly expanded by alternative splicing of pre-mRNA. In the early steps of spliceosome assembly, the exon and intron boundaries need to be carefully selected, facilitated by RNA binding proteins that recognize conserved cis-regulatory sequences in the pre-mRNA. These include the branch point site recognised by SF1, followed by the poly-pyrimidine tract (PY-tract) and the 3¿ splice site (3¿-SS), which are recognised by the U2 auxiliary factor (U2AF) heterodimer. Additional UHM/ULM interactions stabilize the ternary complex of SF1 and the U2AF heterodimer. Details about the overall architecture of the human 3¿-SS recognition complex are yet to be unveiled, which is the major goal of our project. For this purpose, we recently reconstituted in vitro the ternary 3¿-SS recognition complex, composed of recombinant SF1, U2AF2 and U2AF1 proteins and the 3¿ SS RNA. By NMR spectroscopy and SAXS we could show that the protein-RNA complex is stable and functional. Studying this complex by SANS would enable us to obtain individual data of the proteins/RNA and would allow structural characterization in combination with our SAXS and NMR studies.

Background and Aim of the project

In metazoan organisms, the proteome diversity is greatly expanded by alternative splicing of pre-mRNA. This complex process consists of the excision of non-coding regions (introns) and ligation of the coding parts (exons) to give rise to the final mRNA sequence for translation into proteins. In the early steps of spliceosome assembly, the exon and intron boundaries need to be carefully selected, facilitated by different factors that recognise conserved cis-regulatory sequences in the pre-mRNA. These include the branch point site recognised by SF1[1,2], followed by the poly-pyrimidine tract (PY-tract) and the 3' splice site (3'-SS), which are recognised by the U2 auxiliary factor (U2AF) heterodimer, specifically U2AF2 RRM 1,2 [3,4] and U2AF1 zinc fingers (ZnF) [5-8], respectively. Additional UHM/ULM interactions stabilise the ternary complex of SF1 and the U2AF heterodimer [9–11]. More interesting, it has been shown that for weak 3'-SS, i.e. PY-tracts with reduced pyrimidine content, the presence of U2AF1 is especially important for 3'-SS definition [4,6]. However, details about the overall architecture of the human 3'-SS recognition complex are yet to be unveiled, which is the major goal of our project. For this purpose, we recently reconstituted in vitro the ternary 3'-SS recognition complex, composed of recombinant SF1, U2AF2 and U2AF1 proteins and the 3'-SS RNA (Figure 1). We are studying this complex by NMR spectroscopy and small-angle X-ray scattering (SAXS). Rigid body modelling and fitting of structure models to the SAXS measurements result in good fits for the entire complex. However, the arrangement of single parts is still very heterogeneous and requires additional structural information, which can be achieved by contrast-matching experiments in small-angle neutron scattering (SANS).

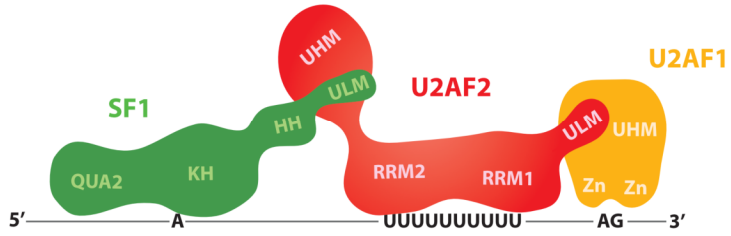

Figure 1 Scheme of the ternary SF1/U2AF complex assembled on 3' splice site

Sample details and experimental setup

We addressed this question by measuring the protein-RNA complex with different deuteration schemes in buffers containing different amounts to vary contrast and to match out unlabelled protein (42 % D_2O), RNA or labelled protein (100% D_2O). Samples which had low contrast compared to the unlabelled sample were measured for longer time periods to accumulate statistics and gain a better overall intensity after buffer substraction and data reduction. Due to the longer acquisition times than expected and issues at the beamline on the first measurement day, we did not measure all samples in all D_2O concentrations but focused on samples that would provide us the information required for more effective model selection (Table 1).

All recombinant proteins used for this experiment were expressed in either LB (unlabeled protein) or M9 minimal medium in 72, 73 or 100 % D_2O (labelled protein). Cells were lysed by frenchpress (G. Heinemann DIGI-F-frenchpress with HTU-800 Lysis cell) and purified by affinity chromatography and size exclusion chromatography. The 27 nt RNA containing optimal binding sites for all three proteins was purchased from Horizon (Dharmacon).

Table 1 Samples measured at beamline D22 with informations on protein deuteration, D_2O concentration in the used buffers and acquisition time.

Nr	Sample	Amount of D2O in buffer	Acquisition time
1	SF1 + U2AF1/2 + 3'-SS RNA	0 %	2 hours
2	SF1 + U2AF1/2 + 3'-SS RNA	42 %	1 hour
3	SF1 + U2AF1/2 + 3'-SS RNA	100 %	1 hour
4	SF1 + **U2AF1/2 (100 % deuterated)** + 3'-SS RNA	0 %	1 hour
5	SF1 + **U2AF1/2 (100 % deuterated)** + 3'-SS RNA	42 %	2 hours
6	SF1 + **U2AF1/2 (100 % deuterated)** + 3'-SS RNA	100 %	1 hour
7	**SF1 (100 % deuterated)** + U2AF1/2 + 3'-SS RNA	0 %	1 hour
8	**SF1 (100 % deuterated)** + U2AF1/2 + 3'-SS RNA	42 %	2 hours
9	**SF1 (100 % deuterated)** + U2AF1/2 + 3'-SS RNA	70 %	4 hours
10	**SF1 (100 % deuterated)** + U2AF1/2 + 3'-SS RNA	100 %	1 hour
11	SF1 + **U2AF1/2 (70 % deuterated)** + 3'-SS RNA	100 %	4 hours
12	**SF1 (70 % deuterated)** + U2AF1/2 + 3'-SS RNA	100 %	4 hours

We performed SEC-SANS experiments for all samples with a Cytiva Superdex Increase 200 5/150 GL column, which was run at a flow of 0.15 ml/min. This enabled us to remove aggregates, excess of RNA and disassembled complex. During acquisition, the flow was stopped for the measurement time listed in the sample table. We loaded 70 µl of concentrated sample (15-20 mg/ml) on the size exclusion column and measured at the D22 beamline with a maximum q-range of 0.5 $Å^{-1}$, at a wavelength of 6 Å. We used 5.6 m collimation and 5.6 m detector distance to cover the Guinier range for Rg determination as well as a wide-angle range to perform solvent subtraction and obtain short-range information (as advised by our local contact Anne Martel). For planning the sample order, we received support from Frank Gabel, who also gave advice for later data processing and analysis; data reduction was performed as shown by Anne Martel.

Results

After data reduction and buffer subtraction by the In-house GRASP software variant from ILL, the derived curves were further analysed using the ATSAS software package (EMBL, Hamburg). Some curves were excluded from further analysis due to low signal-to-noise levels and a resulting poor intensity of the scattering curve. Of the remaining curves, we plotted the P(r) with a fixed Dmax, which was determined based on the SAXS data curve recorded in-house at the SAXS benchtop device of the SFB1035 in Munich, BNMRZ, and performed by Matthias Brandl. Here, we excluded all curves that had a significantly higher Dmax and, therefore, a descent that was too steep when fixing the Dmax (Figure 2). From the remaining data, we determined Rg from the Guinier fit, obtained the distance distribution and Kratky plots. We used the curves with the best signal-to-noise for structural modelling.

We performed rigid body modelling using available crystal and NMR structures of the different regions of the complex, which were published in the past years; leaving linkers as flexible parts. To account for all possible conformations in our system, we calculated a pool of 18,000 structures. Next, we fitted each model to SAXS data using CRYSOL from the ATSAS software package and against SANS data using Pepsi-SANS (ILL, Grenoble). Already, single models agreed well with SAXS and SANS data, and we did not gain marginal improvement of the fits by linear combination of single curves. Consequently, we decided not to form an ensemble of structure models but to filter against the experimental data to get a set of structures that explain our data best. Cutoff for the SAXS and SANS curves were determined by optical inspection of fits, giving

different Chi^2 values. After first filtering the structure models by SAXS data and the four best SANS curves, we obtained a pool of 140 structure models that agreed well with all selected datasets which were more restricted than filtering only by SAXS (ca. 2000 fitting models). Validation of our best-fitting structure model by available NMR measurements (PRE data, published or paper in preparation) is ongoing.

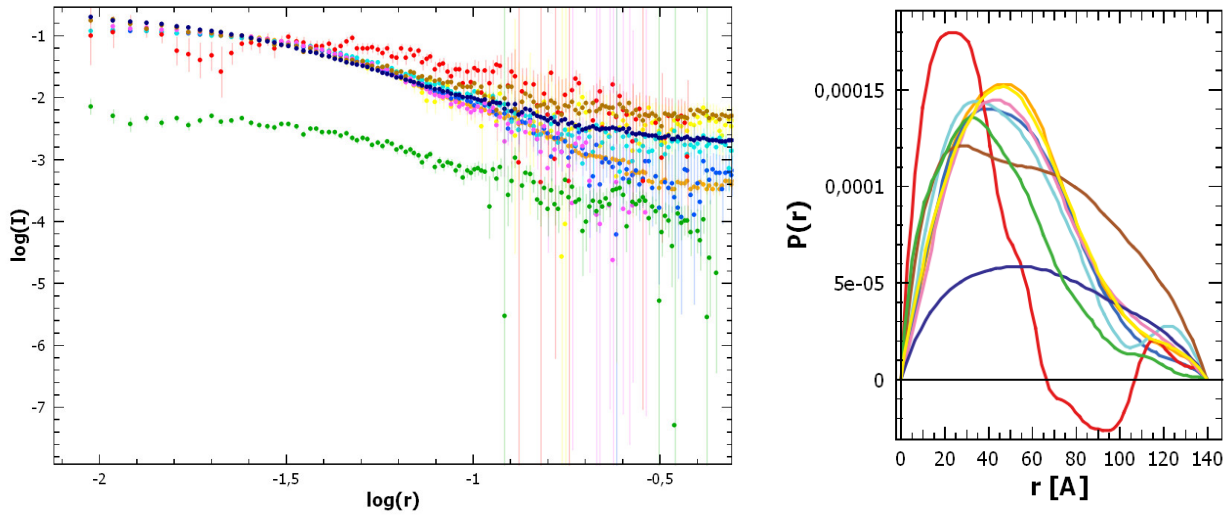

Figure 2 Scattering curves and P(r) curves of all measured samples. Curves: 1 – yellow, 3 – orange, 4 – medium blue, 5 – pale blue, 6 – red, 7 – pink, 10 – brown, 11 – dark blue, 12 –green. Curves 2, 8 and 9 did not show scattering after buffer subtraction.

References

[1] J. A. Berglund, Cell. 1997.

[2] Z. Liu, Science. 2001.

[3] A. A. Agrawal, Nat. Commun. 2016.

[4] C. D. Mackereth, Nature. 2011.

[5] D. A. R. Zorio, Nature. 1999.

[6] S. Wu, Nature. 1999.

[7] L. Merendino, Nature. 1999.

[8] H. Yoshida, Genes Dev. 2015.

[9] C. L. Kielkopf, Cell. 2001.

[10] L. Corsini, Nat. Struct. Mol. Biol. 2007.

[11] P. Selenko, Mol. Cell. 2003.