

# Structural requirements for photo-induced RNA-protein cross-linking

Anna Knörlein<sup>1</sup>, Chris P. Sarnowski<sup>2</sup>, Tebbe de Vries<sup>3</sup>, Moritz Stoltz<sup>1</sup>, Michael Götze<sup>2</sup>, Ruedi Aebersold<sup>2,4</sup>, Frédéric H.-T. Allain<sup>3</sup>, Alexander Leitner<sup>2</sup>, Jonathan Hall<sup>1\*</sup>

<sup>1</sup>Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zurich, Zurich, Switzerland

<sup>2</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

<sup>3</sup>Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich, Switzerland

<sup>4</sup>Faculty of Science, University of Zurich, Zurich, Switzerland

## Abstract

Understanding structure-function relationships of RNA-binding proteins requires knowledge of how they bind RNAs *in vivo*. RNA-protein interactions are studied using light-induced cross-linking at “zero-distance”, yielding nucleotide/amino-acid adducts for mass-spectrometry (MS)-based characterization. However, prerequisites for cross-linking are poorly understood, limiting interpretation of cross-linking data. Here, we report novel insights on cross-linking requirements from studying RBFox-RRM domain bound to <sup>13</sup>C-labeled variants of its heptaribonucleotide binding element as a model. We probed the influence of nucleotide identity, sequence position and amino-acid composition using tandem-MS to assign cross-links at site-specific resolution. We observed cross-linking at three nucleotides, which were stacked onto phenylalanines. Surprisingly, this stacking was required for neighbouring amino-acids to cross-link, and is apparent in published RNA-protein datasets. We hypothesize that  $\pi$ -stacking activates cross-linking *via* electron transfer, whereafter nucleotide- and peptide radicals, possibly stabilized by *capto-dative* effects, recombine. These findings should facilitate interpretation of cross-linking data from structural studies and genome-wide datasets.

## Introduction

The human genome encodes more than 1500 RNA binding proteins (RBPs) that regulate key processes, including translation, localisation, stability and splicing<sup>1-3</sup>. In order to understand fully the structure-function relationship of an RBP, it is necessary to identify to which RNAs it binds *in vivo*, and how non-covalent interactions occur in the binding site. RNA-protein binding occurs at conserved RNA binding domains, such as RNA recognition motifs (RRM), heterogeneous nuclear ribonucleoprotein (hnRNP) K-homology domains and zinc finger (ZnF) domains<sup>4,5</sup>. These domains recognize short, usually single-stranded regions of 3-8 nucleotides (nt) known collectively as consensus RNA binding elements (RBE)<sup>6,7</sup> that often contain degenerate positions. Additional binding affinity and selectivity can be generated *via* supplementary contacts between the RNA and the protein<sup>5,8</sup>; for example, the RBP FUS has a bipartite binding mode comprising its ZnF domain and its RRM<sup>9</sup>. RNA-protein binding

has also been observed with proteins that lack canonical RNA binding domains (RBDs)<sup>1</sup>. Taken together, these features render difficult the prediction of an RBP's substrates based only on a computational search for its consensus RBE. Indeed, recent studies of the RBFOX protein family showed that only one half of the isolated RNA targets contain the RBFOX consensus binding motif and that other motifs presumably bear responsibility for some of its splicing activities<sup>10-12</sup>.

Many state-of-the-art methods to identify RNA-protein interactions *in vivo* employ RNA-protein cross-linking induced by UV light<sup>13-16</sup>. For example, by combining UV cross-linking with mass spectrometry approaches, proteins bound to given RNAs can be identified<sup>17-24</sup>. Conversely, UV cross-linking and immunoprecipitation (CLIP) and related protocols are commonly used to identify RNA-binding sites for given proteins on a transcriptome-wide scale<sup>25-30</sup>. Technical advances constantly improve these techniques<sup>18,31,32</sup>, however, a long-standing challenge in structure/mechanism-oriented studies is to identify the points of cross-linking on *both* the RNA and the protein with site-specific resolution. Recently, we introduced cross-linking of segmentally isotope-labelled RNA and tandem mass spectrometry (CLIR-MS), which identifies the sites of amino acid/ribonucleotide cross-links in a single protocol<sup>32</sup>. In CLIR-MS, RNA regions that are suspected to interact with a protein are synthesized in isotopically labelled light- and heavy variants, so that nucleotides involved in cross-linking events appear as peak doublets in the resulting mass spectrum. By focusing on peak doublets during data analysis, cross-linked amino acids and nucleotides are reliably identified after controlled degradation<sup>32</sup>. The photo-induced formation of covalent bonds between amino acids and ribonucleotides occurs between free radical species at "zero distance"<sup>33-35</sup>. Reactions are biased towards uridines and guanosines<sup>19,36,37</sup>, but most amino acids can participate<sup>18,19</sup>. However, cross-links typically only occur at specific positions in the RNA-RBP motif, for which there is currently no mechanistic rationale<sup>38</sup>. Moreover, it has proven very difficult to gain a deeper understanding of the factors that promote cross-linking, at least partly because the RNA-protein binding site environment, which is critical for cross-linking chemistry, cannot be created in simple solvents. Furthermore, cross-linking reactions usually produce complex product mixtures that are difficult to separate and characterize on a background of possible protein and nucleic acid UV damage<sup>39</sup>.

Here, we have investigated in systematic fashion the structural requirements for the cross-linking of an RNA to its RBP partner. We used the RRM domain of the RBFOX family (FOX<sub>RRM</sub>) and its RNA consensus binding motif U<sub>1</sub>G<sub>2</sub>C<sub>3</sub>A<sub>4</sub>U<sub>5</sub>G<sub>6</sub>U<sub>7</sub> (FOX<sub>RBE</sub>) as a model system because the complex forms with high affinity, the RRM is a highly abundant RNA binding domain and its structure has been well characterized<sup>40</sup>. We introduced <sup>13</sup>C-labelled ribonucleotides systematically into the FOX<sub>RBE</sub> heptanucleotide using solid-phase synthesis and used CLIR-MS to identify RNA-protein cross-links with site-specific resolution. High-yielding cross-linking was localised at two clusters of amino acids around two phenylalanines, consistent with previous findings (Götze et al, submitted). However, with few exceptions, it only occurred at U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub>. We then employed site-specific mutagenesis of both the RNA and the protein to probe the influence of nucleotide species, its position in the RNA sequence and amino acid composition on cross-linking. For each RNA-protein pair, we first confirmed that the same binding mode was maintained as in the wild-type interaction. The analysis revealed that: i) high-yielding cross-linking only occurred at three of the seven nucleotide positions and involved guanosine or uridine, and that guanine and uracil were not interchangeable; ii) only nucleobases that stacked

with the aromatic amino acid side chains reacted strongly; and, surprisingly, iii) the primary stacking interaction was a requirement for neighbouring amino acids to react. We confirmed the importance of this primary stacking feature in other published studies for which cross-linking and structural data are available. The analysis strongly suggested that a stacking interaction is required to activate a nucleotide for free-radical-based cross-linking reactions in native RNA-protein complexes. We expect that this finding will facilitate the interpretation of RNA-protein cross-linking data. Moreover, it will aid the development of new tools for *de novo* motif discovery (see refs in<sup>37</sup>) especially for non-canonical binding motifs, and it will help guide the design of future cross-linking experiments and methods.

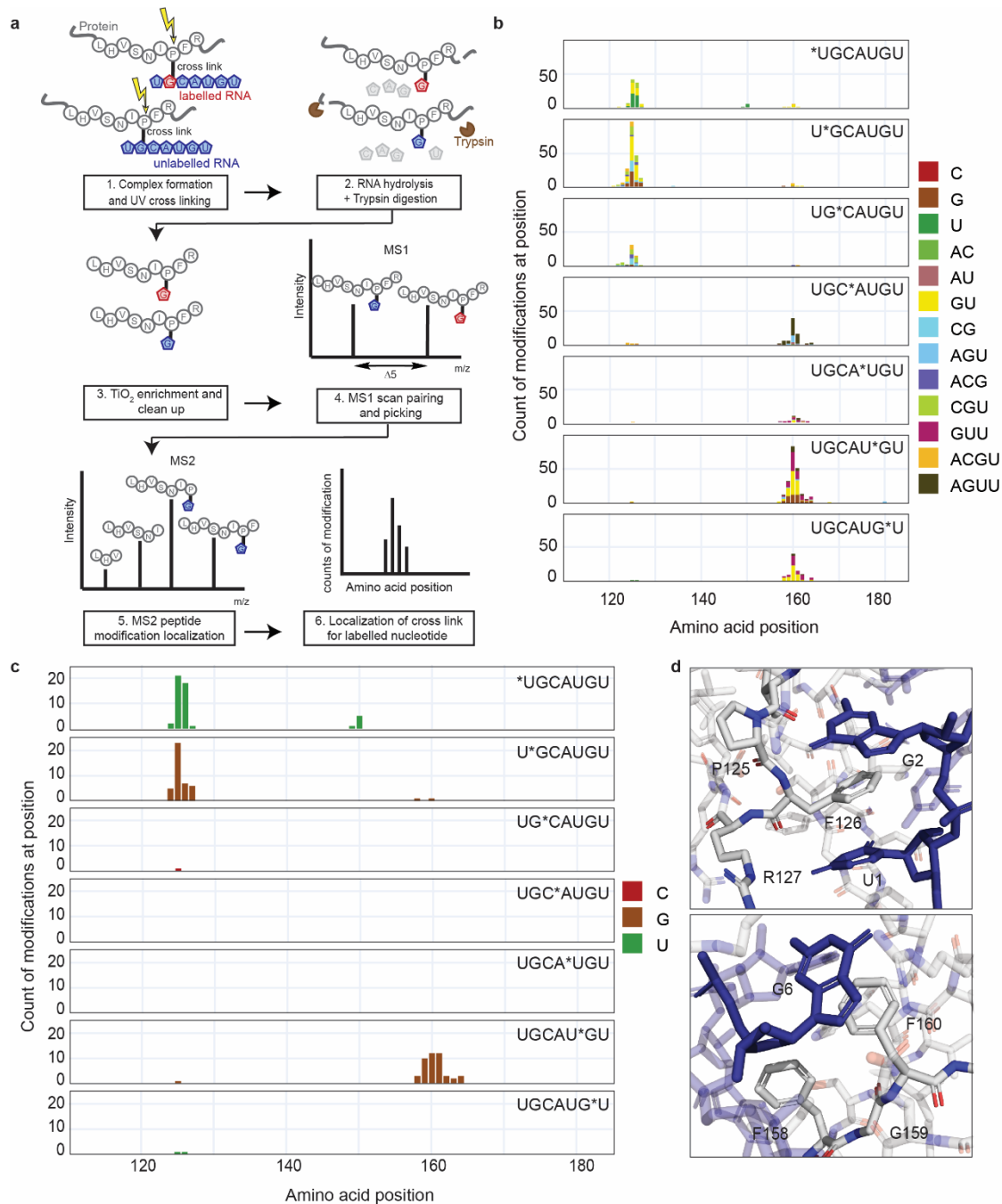
## Results

### Optimization of CLIR-MS to identify RNA-protein cross-links with site-specific resolution

Recently we (RA, AL, FA) introduced the CLIR-MS technique which identifies RNA-protein cross-links on both protein and RNA<sup>32</sup> (**Fig. 1a**). The original CLIR-MS protocol employs RNAs with contiguous regions of differentially isotope-labelled nucleotides in the cross-linking step. After partial RNA and protein digestion, peptide-oligonucleotide conjugates are identified as matched signal pairs in the precursor ion mass-spectrum. This facilitates localisation of the cross-linked nucleotide to the labelled RNA segment. One drawback of the original implementation of CLIR-MS is the inherent requirement for enzymatic <sup>13</sup>C/<sup>15</sup>N-labelled RNA synthesis (i.e. *in vitro* transcription). This limits the minimal length and the number and positions of labelled segments that can be included within an RNA of interest. A second limitation is the nuclease digestion step, which typically results in short oligonucleotides (i.e., 1-4 nt). This means that the specific base involved in the cross-link usually cannot be called with confidence, although longer overlapping partial sequences facilitate the localization of the interacting regions on the RNA. In this study, we sought chemical solutions to both of these problems. First, we switched from RNase digestion to alkaline hydrolysis of RNA (**Fig. S1a**), while exercising care not to degrade the FOX<sub>RRM</sub>. Hence, the mass analysis of the product mixtures detected a much greater fraction of mono-nucleotide adducts, allowing us to define ambiguously nucleotides that are cross-linked (**Fig. S1b**). Second, we employed <sup>13</sup>C-labelled phosphoramidites during solid-phase synthesis of RNAs to incorporate <sup>13</sup>C-labelled nucleotides site-specifically<sup>41</sup>(**Fig. S2a**).

### U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub> in FOX<sub>RBE</sub> cross-link to amino acids centred on phenylalanines in the FOX<sub>RRM</sub>

We employed a systematic approach in an effort to identify key structural requirements for RNA-protein cross-linking. We first used <sup>13</sup>C-labelled versions of FOX<sub>RBE</sub> in a CLIR-MS protocol to identify all points of reaction between the RNA and the protein. We then synthesized mutated <sup>13</sup>C-labelled variants of FOX<sub>RBE</sub> to determine how cross-linking varies with respect to i) the type of the individual nucleobase, ii) their positions in the RBE, and iii) how it is affected by amino acid composition in the FOX<sub>RRM</sub> binding site. We were mindful of the fact that mutating key sites in the RNA and the protein might alter the mode of (or even abolish) RNA-protein binding, and therefore for each variant we measured the binding affinity to FOX<sub>RRM</sub> using surface plasmon resonance spectroscopy (SPR).



**Figure 1: Photo-induced cross-linking of FOX<sub>RBE</sub> and FOX<sub>RRM</sub> occurs only at U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub>.**

**a)** Schematic representation of the CLIR-MS protocol. **b)** Analysis of the cross-linking of segmentally isotope-labelled FOX<sub>RBE</sub> to FOX<sub>RRM</sub>. Phenylalanines occur at positions 126, 158 and 160. The same amount for each sample, was cross linked, hydrolyzed with 0.1 M NaOH at 70°C and trypsin-digested overnight. The clean-up, enrichment and LC-MS/MS analysis was performed according to the standard CLIR-MS protocol. The xQuest search was performed for the masses of mono- to tetranucleotides with a defined mass shift of 5 Da (see **Table S1**). (\*N indicates a <sup>13</sup>C-labelled nucleotide). **c)** CLIR-MS results of **b)** filtered for mononucleotides showing the site-specific reactivity at U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub> (see **Table S1**). **d)** FOX<sub>RRM</sub>/FOX<sub>RBE</sub> structure showing two cross-linking clusters centered on F126 and U<sub>1</sub> and G<sub>2</sub> (upper panel), and the cluster at F160 and G<sub>6</sub> (lower panel) based on the structure of Auweter *et al.*<sup>40</sup>. Structures were visualized with PyMOL (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC).

We synthesized the seven <sup>13</sup>C-labelled isotopic versions of FOX<sub>RBE</sub> and confirmed the correct incorporation of the label by liquid-chromatography mass spectrometry (LC-MS) (**Fig. S2b-c**). We incubated each version of FOX<sub>RBE</sub> with FOX<sub>RRM</sub> and performed the CLIR-MS protocol. Mass analysis

identified short oligonucleotide fragments cross-linked to peptides in clusters close to F126 and F160 (**Fig. 1b**). Each oligonucleotide signal in the spectrum of Figure 1b was detected because it contained a  $^{13}\text{C}$ -labelled ribose. However, other than for mononucleotides, the actual site of cross-linking in the fragment could not be called; for example, the tetranucleotide fragment containing A, C, G and U might have cross-linked at any of the four bases (A, G, C, U). We noted that all (>99%) of the fragments contained at least one uridine or guanosine, consistent with literature reports<sup>16,19,36,37,42</sup> that uracil and guanine mainly participate in cross-linking. Here, the use of alkaline hydrolysis for RNA digestion proved advantageous, since it caused a larger fraction of the RNA to be fully fragmented to mononucleotides (**Fig. S1b**), thereby facilitating unambiguous assignment of cross-linking sites. Hence, focusing only on the mononucleotide species in the spectra of Figure 1b, revealed that cross-linking in the  $\text{FOX}_{\text{RBE}}$  involved almost uniquely  $\text{U}_1$ ,  $\text{G}_2$  and  $\text{G}_6$  (**Fig. 1c**). The numbers of cross-links were in a similar range for the three nucleotides, although numbers of cross-links cannot be confidently compared between different experiments using the current CLIR-MS protocols. Analysis of cross-link products comprising the oligonucleotides in light of the data of the mononucleotides, revealed that the oligonucleotide species shown in Figure 1b with few exceptions derived from cross-linking with only one of these three nucleotides (to  $\text{U}_1$ ,  $\text{G}_2$  or  $\text{G}_6$ ). For example, when  $\text{A}_4$  is labeled, no mononucleotide are detected but small amounts of the tetranucleotides  $\text{U}_1\text{G}_2\text{C}_3\text{A}_4$  and  $\text{A}_4\text{U}_5\text{G}_6\text{U}_7$  are found cross-linked to F126 and F160, respectively (**Fig. 1b**). This results from  $\text{U}_1/\text{G}_2$  and  $\text{G}_6$  cross-linking to F126 and F160, respectively. The cross-linking of  $\text{G}_2$  and  $\text{G}_6$  was consistent with published CLIP data<sup>37,43</sup> (**Fig. 1c**). Although cross-linking from  $\text{U}_1$  was detected in the CLIR-MS experiments, cross-linking was hardly observed to  $\text{U}_5$  (*vide infra*) or  $\text{U}_7$ , consistent with the hypothesis that strict structural parameters govern the photo-induced reactions between  $\text{FOX}_{\text{RRM}}$  and  $\text{FOX}_{\text{RBE}}$ . Very small numbers of isolated cross-links were also observed in some of the spectra of Figure 1b-c. Although low numbers of cross-links must be considered with caution, their locations suggested that in several cases they were not artifacts. In particular, the cross-links at F160 seen with  $^*\text{UGCAUGU}$  and  $\text{U}^*\text{GCAUGU}$  (**Fig. 1b**, top two panels) are consistent with transient (low affinity) binding of  $\text{U}_1\text{G}_2$  in the binding pocket occupied mainly by  $\text{U}_5\text{G}_6$ . Likewise, cross-links around F126 in the lower panels of Figure 1b/c may derive from similarly transient contacts with  $\text{U}_5\text{G}_6\text{U}_7$ . One cross-link from P125 to cytidine is visible in Figure 1c.

Sites of cross-linking at the protein appear clearly centred at two phenylalanines (F126 and F160), with a distribution of 1-3 amino acids flanking these sites (Götze et al. submitted). This was confirmed from study of the MS2 spectra in which fragment ions present localise the RNA adducts unambiguously on the peptide backbone (**Fig. S3**). Exchange of  $\text{G}_2$  or  $\text{G}_6$  for  $\text{A}_2$  or  $\text{A}_6$  greatly reduced overall cross linking of the mutated sites to the amino-acid clusters 126 and 160, respectively (**Fig. S4**). The cross-links of  $\text{U}_1$ ,  $\text{G}_2$  and  $\text{G}_6$  aligned well with the NMR structure of  $\text{FOX}_{\text{RRM}}$  bound to  $\text{FOX}_{\text{RBE}}$ <sup>40</sup> (**Fig. 1d**) (PDB ID: 2ERR). The largest number of spectra corresponded to  $\text{U}_1$  reacting with P125 and F126 and to a lesser extent with I124 and R127 (**Fig. 1c**). Similarly,  $\text{G}_2$  cross-linked to P125, I124, F126 and R127.  $\text{U}_1$  and  $\text{G}_2$  each stack on one face of F126. Hydrogen-bonds also occur between the bases of  $\text{U}_1$  and  $\text{G}_2$ , and between R127 and I124, respectively.  $\text{G}_6$  reacted with F160, to which it also stacks, as well as with neighbouring amino acids at positions 158-164. F158 contacts the ribose of  $\text{G}_6$ . Notably, several close RNA-protein contacts visible in the NMR structure, such as  $\text{C}_3$

interacting with F126 (but not stacking), G<sub>6</sub> stacking with R194 and U<sub>5</sub> stacking to H120<sup>40</sup>, did not produce extensive cross-linking.

The current understanding of RNA-protein cross-linking is that close contact between nucleotides and amino acids is the main pre-requisite for a cross-linking event<sup>44,45</sup>. However, only three from the seven nucleotides of FOX<sub>RBE</sub> engaged in efficient cross-linking, despite close contact between all nucleotides and amino acids in the binding site. In an effort to resolve this apparent contradiction, we investigated two obvious parameters that could influence cross-linking: the chemical reactivities of the nucleotides and the amino acids, as well as their relative positioning of the reactive pair in space. By mutating selected nucleotides and amino acids in the binding pocket, we created a cross-linking structure-activity relationship for the FOX<sub>RRM</sub>-FOX<sub>RBE</sub> interaction.

### Only uridine cross-links to FOX<sub>RRM</sub> from position 1 of FOX<sub>RBE</sub>

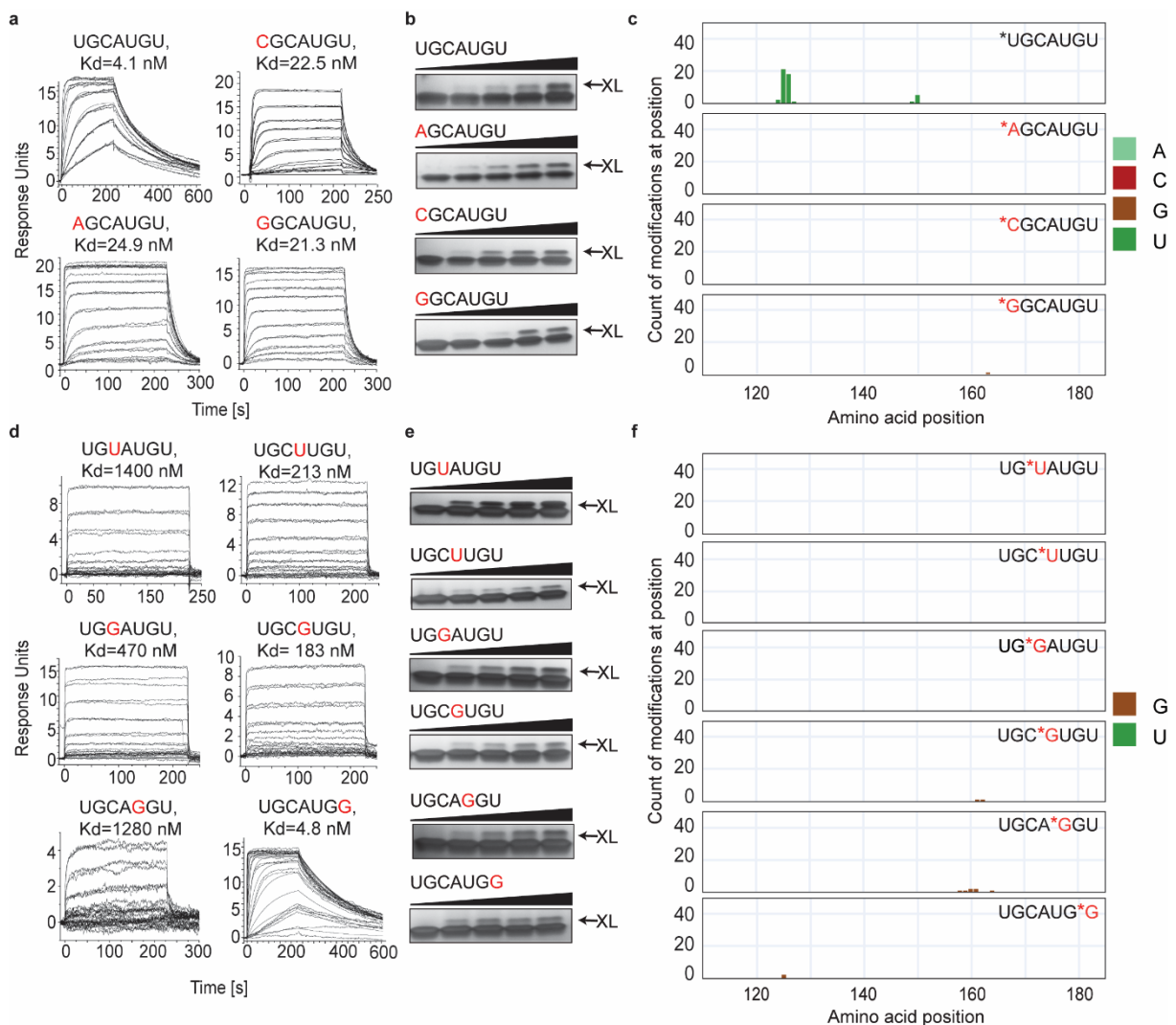
We synthesized the three labelled mutants of \*NGCAUGU (N=A, G, C; **Table S2**), as well as the corresponding per-labelled control sequences \*N\*G\*C\*A\*U\*G\*U. We first confirmed that the NGCAUGU variants bound to FOX<sub>RRM</sub> using SPR. In this assay, parent UGCAUGU bound strongly to FOX<sub>RRM</sub> with a  $K_d$ =4.1 nM. Substitution of the 5'-uridine reduced the strength of the interaction by 4-6 fold for the three variants (AGCAUGU:  $K_d$  =24.9 nM; CGCAUGU:  $K_d$  =22.5 nM; GGCAUGU:  $K_d$  =21.3 nM) (**Fig. 2a**). This was consistent with the NMR structure showing that the 5'-uridine of FOX<sub>RBE</sub> contributes to binding by  $\pi$ -stacking to F126 (**Fig. S4**)<sup>12,40</sup>.

Next, we incubated the RNAs together with FOX<sub>RRM</sub> and irradiated the complexes at 254 nm at increasing energies. Work-up and analysis by SDS-PAGE for the three NGCAUGU sequence mutants revealed a new slow-migrating band on the gels, similar to that of the wild-type FOX<sub>RBE</sub> (N=U), consistent with RNA-protein cross-linking (**Fig. 2b**). The appearance of a band on an SDS-PAGE confirms that cross-linking occurs, but it does not identify the site of cross-linking nor the composition of the product. In order to determine whether any of the mutants cross-linked at the N<sub>1</sub>-position, we turned to CLIR-MS. CLIR-MS data for per-labelled \*N\*G\*C\*A\*U\*G\*U confirmed that the three FOX<sub>RBE</sub> mutants exhibit the same cross-linking "fingerprint" as wild type FOX<sub>RBE</sub>, i.e. in the same two amino acid clusters around positions 126 and 160 (**Fig. S6**). However, in order to differentiate cross-linking of N<sub>1</sub> to that from G<sub>2</sub> in the 126-cluster, we performed CLIR-MS on the singly labelled sequences (\*NGCAUGU). In contrast to U<sub>1</sub>, cross-linking hardly occurred at A<sub>1</sub>, G<sub>1</sub> or C<sub>1</sub> (**Fig. 2c**), consistent with the systematic presence of G<sub>2</sub> in the di- or tri-nucleotides found cross-linked to F126 confirming that only G<sub>2</sub> in the mutants can cross-link efficiently to the phenylalanine (**Fig. S5**). This also confirmed the strong bias for uridine photo-reaction<sup>19,36</sup>, though it was surprising that G<sub>1</sub> was unreactive, given the reactivity of G<sub>2</sub>.

In order to determine systematically the propensity for cross-linking at each site in FOX<sub>RBE</sub> when a more photo-reactive nucleotide (i.e. U or G) is present, we performed CLIR-MS on six additional positional FOX<sub>RBE</sub> mutants. Thus, we exchanged \*U for C<sub>3</sub> and A<sub>4</sub> in FOX<sub>RBE</sub> (UG\*UAUGU, UGC\*UUGU, resp.), and \*G for C<sub>3</sub>, A<sub>4</sub>, U<sub>5</sub> and U<sub>7</sub> (UG\*GAUGU, UGC\*GUGU, UGCA\*GGU, UGCAUG\*G, resp.). In each case, we confirmed that the mutants bound and cross-linked to FOX<sub>RRM</sub> using SPR and SDS-PAGE gels (**Fig. 2d** and **Fig. 2e**, resp.). Remarkably, in none of these six

examples, did the mutated nucleotides cross-link efficiently to the protein (**Fig. 2f**). The lack of reactivity at U<sub>3</sub> (in mutant UG\*UAUGU) was particularly surprising given the close proximity of C<sub>3</sub> to F126 in the NMR structure.

In summary, while G<sub>2</sub> and G<sub>6</sub> in wild type FOX<sub>RBE</sub> cross-linked to FOX<sub>RRM</sub>, guanosine did not cross-link efficiently at any other of the other five locations in the FOX<sub>RBE</sub>. Similarly, uridine readily cross-linked to FOX<sub>RRM</sub> from position N<sub>1</sub> - where A, C and G were unreactive - but not from the four other locations in the FOX<sub>RBE</sub>. Taken together, the data from this controlled model study confirmed that RNA-protein cross-linking events have strict requirements, beyond simply the proximity of a reactive nucleotide and a reactive amino acid.



**Figure 2: FOX<sub>RRM</sub>/FOX<sub>RBE</sub> cross-linking is nucleotide- and site-specific.**

**a)** SPR sensograms of FOX<sub>RBE</sub> variants bound to FOX<sub>RRM</sub>. Data were measured in duplicates in a 1:1 dilution series at at least six concentrations starting at 2  $\mu$ M. The curve for UGCAUGU was fitted using a 1:1 Langmuir binding model including mass-transport limitations. The K<sub>d</sub> of the remaining sequences were analysed by a steady state affinity fit curve; wild-type FOX<sub>RBE</sub> shows the highest affinity for FOX<sub>RRM</sub>. **b)** All N<sub>1</sub>-variants undergo cross-linking with FOX<sub>RRM</sub> with increasing radiation dose. The cross-linking product band is indicated on SDS-PAGE gels by "XL". **c)** CLIR-MS plots show that cross-linking occurs at U<sub>1</sub> of FOX<sub>RBE</sub>, but not with N<sub>1</sub>-mutants. The clean-up, enrichment and LC-MS/MS analysis was performed according to the CLIR-MS protocol. The xQuest search was carried out for the masses of mono- to tetranucleotides with a defined mass shift of 5 Da. The data was filtered for mononucleotides (\*N indicates a <sup>13</sup>C-labelled nucleotide, the mutated nucleotide is labelled in red) (see **Table S1**). **d)** SPR traces show that U or G mutations at N<sub>3</sub>, N<sub>4</sub>, N<sub>5</sub> and N<sub>7</sub> of FOX<sub>RBE</sub> attenuate but do not abolish, FOX<sub>RRM</sub>

binding to FOX<sub>RBE</sub> variants. The curve for UGCAUGG was fitted using a 1:1 Langmuir binding model including mass-transport limitations. The K<sub>d</sub> of the remaining sequences were analysed by a steady state affinity fit curve. **e)** SDS-PAGE gels show that FOX<sub>RBE</sub> variants cross-link to FOX<sub>RRM</sub> with increasing radiation dose. **f)** CLIR-MS analysis of singly-labelled FOX<sub>RBE</sub> variants show that protein-RNA cross-linking does not occur with U or G nucleotides located at N<sub>3</sub>, N<sub>4</sub>, N<sub>5</sub> and N<sub>7</sub> (see **Table S1**). The clean-up, enrichment and LC-MS/MS analysis was performed according to the CLIR-MS protocol. The xQuest search was carried out for the masses of mono- to tetranucleotides with a defined mass shift of 5 Da. The data was filtered for mononucleotides (\*N indicates a <sup>13</sup>C-labelled nucleotide).

### **Aromatic amino acids play a key role in RNA-protein cross-linking**

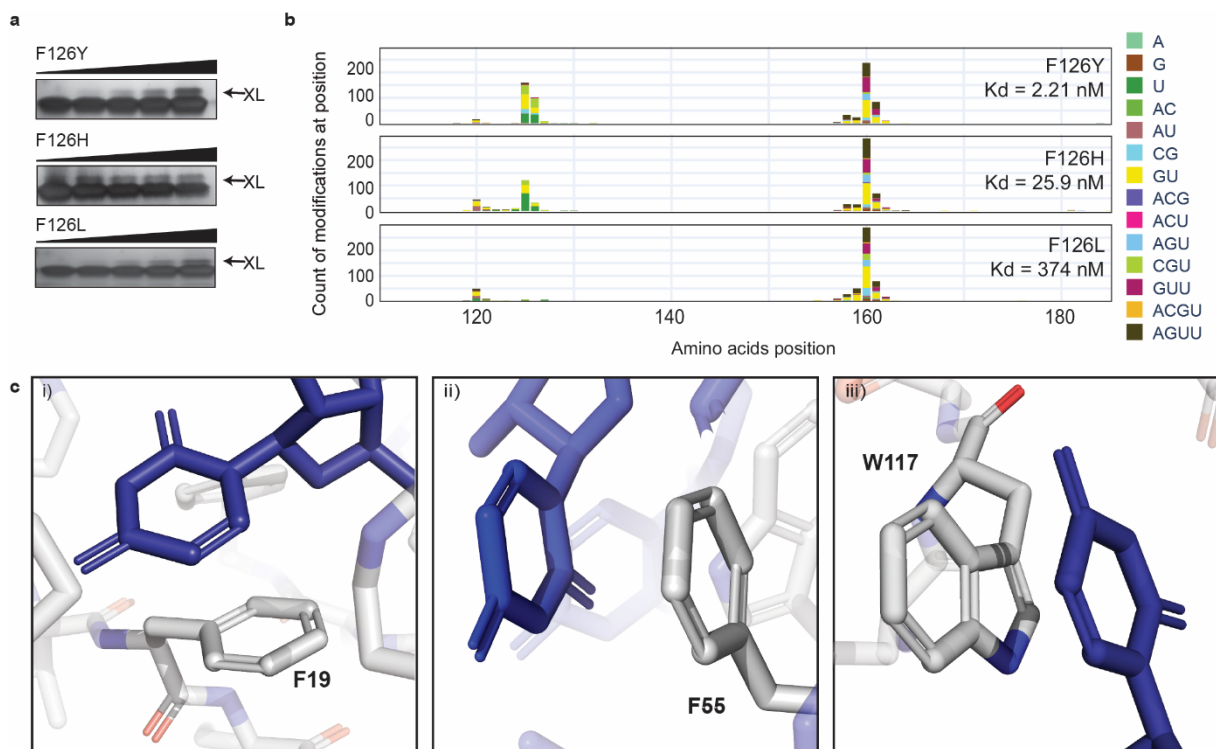
Analysis of the aforementioned CLIR-MS data (**Fig. 1c**, **Fig. S6**) had provided two important insights: i) on the RNA side, strong cross-linking only occurred with nucleotides that were stacked to aromatic amino acids (F126, F160, **Fig. 1d**); and ii) on the protein side, whereas the cross-links involved F126 and F160, cross-links were also detected at one to three amino-acids up- and downstream of F126 and F160 (positions 124-127, 158-164).

We therefore mutated F126 in FOX<sub>RRM</sub> to histidine, tyrosine and leucine. We have previously shown using SPR that aromatic amino acids at position 126 are crucial for binding FOX<sub>RBE</sub> (F126Y: K<sub>d</sub> =2.21 nM; F126H: K<sub>d</sub> =25.9 nM), although a sterically-fitting aliphatic amino acid such as leucine can partially substitute for the phenylalanine (F126L: K<sub>d</sub> =374 nM)<sup>40</sup>. We irradiated these variants in the presence of FOX<sub>RBE</sub>. All three mutants cross-linked to FOX<sub>RBE</sub>, as evident from SDS-PAGE (**Fig. 3a**). In order to pinpoint the sites of cross-linking, we carried out CLIR-MS experiments with uniformly <sup>13</sup>C-labelled FOX<sub>RBE</sub> (**Fig. 3b**). F126Y and F126H mutants cross-linked to the FOX<sub>RBE</sub> similarly to FOX<sub>RRM</sub>. The cross-linking profile was similar for the three complexes at F160, however, cross-linking of the mutants was concentrated on P125 and F126, whereas, when phenylalanine was exchanged for leucine, cross-linking to F126 was abolished. Notably, cross-linking to the neighbouring amino acids 124-127 was also mostly lost (**Fig. 3b**), confirming the primary role of the aromatic side chain in mediating the cross-linking reactions with flanking amino acids at positions 124, 125 and 127. Interestingly, in comparison to FOX<sub>RRM</sub> wild-type, the F126H mutant appears not to cross-link to G<sub>2</sub>, consistent with the absence of G mononucleotides (brown) or CG dinucleotides (turquoise) in Figure 3b (**Table S1**). Although we do not have supporting data, nor know of any precedence in literature, it is plausible that the histidine has a different cross-linking preferences to tyrosine or phenylalanine and/or that stacking to the guanosine is changed in this particular binding site. Unexpectedly, a H120 cross-link occurred with the three FOX<sub>RRM</sub> mutants, which was hardly observed in the wild type FOX<sub>RRM</sub> (**Fig. 3b**, **Fig. S6**). Analysis of the oligonucleotide cross-links in Figure 3b strongly suggested that the cross-link occurred with U<sub>5</sub>. In fact, the NMR structure of FOX<sub>RRM</sub>-FOX<sub>RBE</sub> shows that U<sub>5</sub> adopts a stacking arrangement with H120, and thus might have been expected to cross-link in the wild type FOX<sub>RBE</sub>-FOX<sub>RRM</sub> interaction (**Fig. S7**). Taking these findings together, we propose that  $\pi$ -stacking interactions between aromatic amino acids (e.g. phenylalanine, tyrosine or histidine) and guanosines or uridines are an important pre-requisite for their cross-linking to aromatic side chains, and also to the flanking amino acids.

### **RNA-protein cross-linking correlates with $\pi$ -stacking interactions in other complexes**



In order to determine whether these findings apply more broadly to RNA-protein cross-linking, we analysed data from four published structures in which the required resolution of nucleotide/amino acid cross-linking was available. For example, Panhale *et al.* localised a cross-link in hnRNPC around a uridine  $\pi$ -stacked with F19<sup>46,47</sup> (**Fig. 3c**). In a novel approach, Lelyveld *et al.* used <sup>18</sup>O-RNA labelling and targeted mass spectrometry to localise the cross-link of U<sub>11</sub> in a let-7 microRNA precursor to a  $\pi$ -stacked phenylalanine (F55) in the Lin28 cold shock domain (**Fig. 3c**)<sup>48-50</sup>. Kramer *et al.* used a combination of cross-linking and mass spectrometry to pin-point cross-linking sites on amino acids from yeast ribosomal protein S1<sup>19</sup>. The data was correlated with published crystal structures of the protein, which showed that tryptophan (W117) participates in a  $\pi$ -stacking interaction with uridine U<sub>1799</sub> (PDB 4V88) (**Fig. 3c**)<sup>19,51</sup>. Finally, our own CLIR-MS data from PTBP1 in complex with internal ribosomal entry site (IRES) of encephalomyocarditis virus (EMCV)<sup>32</sup> provided additional evidence for the role of aromatic amino acids in cross-linking reactions. Cross-links occur mainly at four aromatic amino acids (Y127, Y267, H411, H457), and mostly with uridines. Cross-comparison with the NMR solution structure of PTBP1 in complex with short polypyrimidine sequences shows that these amino acids are  $\pi$ -stacked to nucleotides (PDB IDs: 2AD9, 2ADB, 2ADC) (**Fig. S8**)<sup>52</sup>. In most cases, there is cross-linking not only to the stacked amino acids but also to flanking amino acids, such as Q412. Interestingly, one of these  $\pi$ -stacked interactions (H457) appear to be with cytosine. Although other groups have described that cytosine participates at a low level in cross-linking<sup>37,42,53</sup>, we did not observe more than minor amounts in mass spectral analyses in this study<sup>18,19,47</sup>.



**Figure 3: Proximal aromatic amino acids mediate cross-linking in RNA protein complexes.**

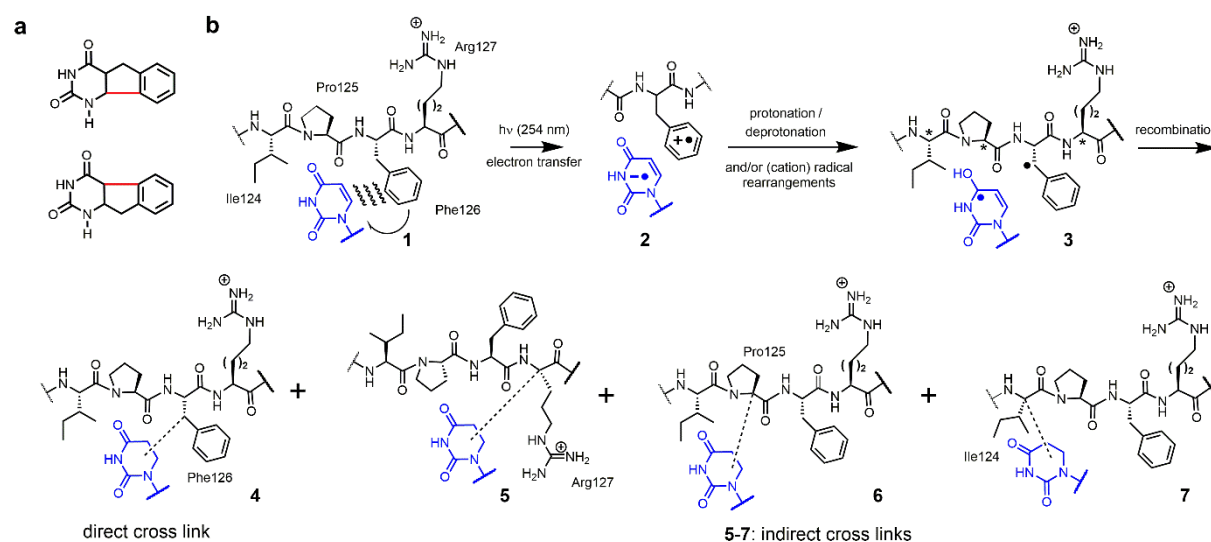
**a)** F126 mutants undergo binding and cross-linking with FOX<sub>RBE</sub> with increasing doses of irradiation. The cross-linking band is indicated by XL. SDS-PAGE gel of RNA-protein cross-link to determine cross-linking reactivity with FOX<sub>RBE</sub>. The upper band (XL) indicates the complex of the protein cross-linked to RNA. **b)** CLIR-MS analysis of uniformly labelled UGCAUGU to FOX<sub>RRM</sub> F126 mutants. The digestion, enrichment and LC-MS/MS analysis was performed according to the standard CLIR-MS protocol. xQuest search was performed for mono- to tetra-nucleotides with a defined mass shift of 5 to 20 Da (see **Table S1**). Binding affinities of FOX<sub>RBE</sub> are taken from ref<sup>40</sup>. **c)** Illustration of the localised cross-link with the solution structure i) for hnRNP C

binding on AUUUUUC obtained by Cienikova *et al.*<sup>46,47</sup> (PDB 2MXY). ii) for Lin28A in complex with preEM-let-7f obtained by Nam *et al.*<sup>49,50</sup> (PDB 3TS0). iii) for 40S ribosomal protein S1 in complex with the 18S rRNA obtained by Ben-Shem *et al.*<sup>19,51</sup> (PDB 4V88) visualized with PyMOL (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC).

### Photo-induced electron transfer in a $\pi$ -stacked RNA-protein complex may mediate radical reactions of cross-linking

The site-specific resolution provided by the CLIR-MS technique, and in several published structures, has highlighted the role of  $\pi$ -stacking interactions as a pre-requisite for cross-linking of the FOX<sub>RRM</sub>-FOX<sub>RBE</sub> interaction. Thus, U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub> yielded “direct” cross-links with the two phenylalanines with which they stack, and “indirect” cross-links with 1-3 amino acids flanking the phenylalanines. Most notably, nucleobases that do not stack to aromatic side chains, did not partake in cross-linking.

Free radical reactions of nucleic acids and proteins have been well studied in the context of oxidative damage and electron transfer<sup>54,55</sup>, but less thoroughly investigated for RNA-protein interactions<sup>33,44</sup>. However, a study of the photo-induced intramolecular cyclization of 5-benzyluracil and 6-benzyluracil *via* benzyl and uracil radical intermediates suggests a plausible model for the cross-linking of U<sub>1</sub> with F126 (**Fig. 4a**)<sup>39</sup>. Hence, photo-induced electron transfer in the U<sub>1</sub>-F126 complex generates a short-lived anion/cation radical pair (exciplex) (**Fig. 4b; 1 and 2**). Subsequent protonation of the uracil radical anion can yield a neutral  $\alpha$ -hydroxy radical<sup>55</sup>, whereas ready deprotonation of the F126 radical cation will produce a stabilized benzylic radical. In the absence of oxygen, the major fate of these free radicals is recombination with formation of the direct U<sub>1</sub>-F126 cross-link (**Fig. 4b; 4**). An analogous mechanism has been proposed for the reaction between free uracils/halogenated uracils and tyrosine derivatives<sup>35,56</sup>.



**Figure 4: Proposed mechanism for UV induced RNA protein cross-linking.**

**a**) UV induced cyclization of 5-benzyluracil and 6-benzyluracil after Sun *et al.*<sup>39</sup> (bonds in red are formed upon cyclisation). **b**) Possible mechanism of UV cross-linking between the stacked F126 and U<sub>1</sub> of the FOX<sub>RRM</sub> (1). Photo-induced electron transfer leads to a radical ion pair (2). After protonation/deprotonation steps, the radicals on the benzylic position of F126 and C4 of uridine (3) recombine to yield direct cross-links (4). Indirect cross-links between U<sub>1</sub> and R127 (5), P125 (6) or I124 (7) may form when the radical cation of F126 oxidizes amide carbonyls from flanking amino acids, which rearrange to radicals stabilized by captodative effects at the  $\alpha$ -carbons (\*).

Alternatively, the F126 radical, or radical cation, may rearrange to neighbouring amino acids in processes mediated by hydrogen atom abstractions<sup>55</sup>, or *via* oxidation of amide carbonyls (by the F126 radical cation)<sup>57</sup>, yielding free radicals at peptide  $\alpha$ -carbon sites on the protein backbone. Viehe *et al* have proposed that  $\alpha$ -carbon radicals are especially stabilized thermodynamically by *capto-dative* effects, i.e. simultaneously by electron-withdrawing ( $-\text{C}=\text{O}$ ) and electron-donating ( $-\text{NR}_2$ ) groups<sup>58</sup> and, furthermore, that they readily combine with other radicals. Hence, depending on the lifetimes and the locations of these radicals on the protein backbone, “indirect” cross-links to  $\text{U}_1$  may form, yielding products that are identified by mass spectral analysis after controlled RNA- and protein-degradation (e.g. **5-7**; **Fig. 4b**). These steps are consistent with the outcome of cross-linking reactions of F126 mutants, i.e. the exchange of phenylalanine for histidine and tyrosine produced similar direct and indirect cross-links, whereas leucine was mostly inactive since its aliphatic side chain cannot partake in the initial electron transfer.

Based on the similarity of the cross-linking profiles from  $\text{U}_1$  and  $\text{G}_2$  (**Fig. 1c**), it seems intuitively likely that guanosines  $\text{G}_2$  and  $\text{G}_6$  may follow a similar mechanistic reaction path as  $\text{U}_1$ . Thus, photo-excitation of the stacked guanine-phenyl ring systems produces free radicals at  $\text{G}_2$  and  $\text{G}_6$ , as well as on the peptide backbone around F126 and F160. Recombination yields direct and indirect cross-links, which in the case of  $\text{G}_2$  are to the same  $\alpha$ -carbon radicals that couple with  $\text{U}_1$ . The nature of the initial exciplex formed from electron-transfer in a stacked guanosine-phenylalanine is unclear, and we were unable to identify a literature precedent for such a mechanism. However, well-cited studies have shown photo-induced electron transfer between  $\pi$ -stacked pyrimidine and purine nucleobases that produce long-lived exciplexes<sup>59,60</sup>. Electron transfer between an amino acid and a nucleotide might be expected to occur in the direction that yields the lowest-energy exciplex. However, due to the special environment of an RNA-protein binding site (see discussions in refs<sup>57,59</sup>), this may not necessarily correlate with the redox potentials of isolated nucleotides or aromatic amino acid side chains.

Together, our observations demonstrate the importance of local environment to cross-linking in the RNA-protein binding site, and at least partly explain why cross-links occur only at specific positions in an RNA-RBP motif.

## Conclusion

For a complete understanding of the roles that RBPs play in cellular processes, it is necessary to understand at the atomic level how RNA binding domains in proteins engage with RNAs. RNA-protein interactions are generally characterized in two main ways *in vivo*: isolating proteins and sequencing the bound RNAs (CLIP methods), and identifying proteins bound to RNAs, for example, by mass spectrometry. Most of these approaches rely upon photo-induced cross-linking, which provides direct evidence of binding under native conditions. However, presently, native cross-linking-based methods suffer from two drawbacks: i) it is challenging to identify simultaneously sites of cross-linking on the RNA and protein, ii) cross-linking in an RNA-RBP motif typically proceeds inefficiently and in an unpredictable fashion.

Recently, we introduced the CLIR-MS method<sup>32</sup>. This technique employs isotope-labelled RNAs to resolve amino acid/ribonucleotide cross-links in a single protocol, whereby segments of labelled RNA are produced by *in vitro* transcription prior to ligation-assembly into a full length RNA. In this new study, we have broadened the application of CLIR-MS through the use of chemically synthesized <sup>13</sup>C-labelled RNAs. This enables site-specific incorporation of labelled nucleotides into the RNA. After irradiation, sample work-up and controlled degradation to nucleotide-peptide adducts, the locations of cross-linked nucleotides in the RNA sequence are pinpointed site-specifically. We demonstrated this methodological advance with a study of the interaction of the RRM domain of the RBFOX family bound to its consensus binding element (U)GCAUGU, for which we have previously determined an NMR structure<sup>40</sup> and studied cross-linking (Götze et al. *submitted*).

Photo-irradiation of the FOX<sub>RRM</sub>-FOX<sub>RBE</sub> complex led to key observations with potentially wide-ranging consequences: 1) strong cross-linking occurred between U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub> with clusters of amino acids centered around the phenylalanines F126 and F160; 2) very little cross-linking was observed at other uridines in the parent or a mutated FOX<sub>RBE</sub> (C<sub>3</sub>/U<sub>3</sub>, U<sub>5</sub> and U<sub>7</sub>); and 3) amino acids that flank F126 and F160 also cross-linked efficiently to U<sub>1</sub>, G<sub>2</sub> and G<sub>6</sub>, but not to other nucleotides of FOX<sub>RBE</sub>. Since the NMR structure of FOX<sub>RRM</sub>-FOX<sub>RBE</sub><sup>40</sup> shows that U<sub>1</sub> and G<sub>2</sub> stack to F126, and that G<sub>6</sub> stacks with F160, the data suggested that a stacking interaction is a requirement for cross-linking events in an RNA-protein interaction, at least for this RRM domain. Indeed, other aromatic side chains could substitute for F126 in cross-linking, but incorporation of leucine abolished direct and almost all indirect cross-linking to U<sub>1</sub>/G<sub>2</sub>. Other researchers have noted in passing the increased presence of aromatic amino acid side chains in UV cross-linking data sets (see refs<sup>18,19,33,48,61</sup>), but have not to our knowledge recognized its role as a trigger for cross-linking, nor distinguished between direct and indirect cross-link events. It is clear that additional factors may also contribute to cross-linking events in RNA-protein sites, including: efficiency of the photo-induced electron transfer, the ability to stabilize free radicals, the flexibility of the system to adopt to the configurations that are required for the radical reactions<sup>62</sup> and the proximity of reactive pairs<sup>63</sup> (Sarnowski *et al*; *in preparation*). Furthermore, our findings cannot explain all RNA-protein cross-linking reactions, at least those involving sulphur-containing amino acids, such as cysteine, which are highly photoreactive and prone to cross-link probably due to the high reactivity of the thiyl radical<sup>19,34,55</sup>. However, our results may partly explain why double-stranded RNA cross-links so poorly to proteins<sup>64</sup>, since in a double-stranded RNA  $\pi$ -stacking between aromatic amino acids and nucleobases is hindered. Indeed, double-stranded RNA-binding motifs (dsRBM) recognize their substrates mainly by shape and contacts with the sugar-phosphate backbone<sup>7,65</sup>.

The findings in this work were enabled by the combination of site-specific labeling with the CLIR-MS protocol, which provides enhanced knowledge of cross-linking sites at single nucleotide and amino acid resolution. This technology is inherently flexible, and we are exploring further improvements to the method (Götze et al, *submitted*). Our findings here demonstrate the importance of local environment to cross-linking in the binding site, i.e. beyond the simple proximity of photo-reactive nucleotides and amino acids, thereby helping to explain at least in part why cross-links occur only at specific sites in an RNA-RBP motif. Because of the inherent variations in the ways that RBPs recognize their RNA targets, predictive modeling of RBP selectivity is extremely challenging. Our

findings can be implemented into the development of new tools<sup>37,66</sup> for *de novo* motif discovery. In a broader sense, we expect our findings also to be of value in understanding RNA-protein interactions from the analysis of CLIP data sets, which is currently an area of intense research<sup>16</sup>.

### Acknowledgments

We thank Bernd Giese and Gunnar Jeschke for very helpful discussions. We thank Timo Hagen for the help with SPR, Miriam Vuk for the help with the RNA synthesis and Erich Michel for the production of the biotinylated FOX RRM. This work was supported in parts by grants from the ETH Scientific Equipment program (to R. A.), the European Union Grant ULTRA-DD (FP7-JTI 115766 to R.A.), Strategic Focus Area for the ETH Domain “Personalized Health and Related Technologies” (TechTransfer Project PHRT-503 to A.L. and F.H.-T.A.), the ERC-2014AdG 670821 from the European Research Council to R.A., the NCCR RNA and Disease of the SNSF and the ETH Zurich (Research Grant ETH-24 16-2 to R.A., A.L., F.A., J.H.) and Sinergia grant of the SNSF (CRSII3\_127454) to J.H.

### Author contributions

A.K. and T.d.V. expressed and purified the RBFOX RRM for the study. M.S. performed the SPR studies. A.K. produced chemically synthesized RNA and performed cross-linking of protein-RNA complexes. A.K. and C.S. performed sample preparation for mass spectrometry and analysed the data using xQuest. All authors interpreted the data. A.K. and J.H. wrote the manuscript together. All authors contributed to manuscript revisions and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### References:

- 1 Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**, 327-341, doi:10.1038/nrm.2017.130 (2018).
- 2 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nature Reviews Genetics* **15**, 829-845, doi:10.1038/nrg3813 (2014).
- 3 Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. RNA-binding proteins in human genetic disease. *Nature Reviews Genetics* **22**, 185-198, doi:10.1038/s41576-020-00302-y (2021).
- 4 Cléry, A. & Allain, F. H. in *RNA binding proteins* (ed Zdravko Lorkovic) 137-158 (CRC Press, 2012).
- 5 Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology* **8**, 479, doi:10.1038/nrm2178 (2007).
- 6 Mitchell, S. F. & Parker, R. Principles and Properties of Eukaryotic mRNPs. *Molecular Cell* **54**, 547-558, doi:<https://doi.org/10.1016/j.molcel.2014.04.033> (2014).
- 7 Stefl, R., Skrisovska, L. & Allain, F. H.-T. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO reports* **6**, 33-38, doi:<https://doi.org/10.1038/sj.embor.7400325> (2005).
- 8 Jankowsky, E. & Harris, M. E. Specificity and nonspecificity in RNA–protein interactions. *Nature Reviews Molecular Cell Biology* **16**, 533, doi:10.1038/nrm4032 (2015).

- 9 Loughlin, F. E. *et al.* The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Molecular Cell* **73**, 490-504.e496, doi:<https://doi.org/10.1016/j.molcel.2018.11.012> (2019).
- 10 Begg, B. E., Jens, M., Wang, P. Y., Minor, C. M. & Burge, C. B. Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nature structural & molecular biology* **27**, 901-912, doi:10.1038/s41594-020-0475-8 (2020).
- 11 Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature Structural & Molecular Biology* **16**, 130, doi:10.1038/nsmb.1545  
<https://www.nature.com/articles/nsmb.1545#supplementary-information> (2009).
- 12 Lambert, N. *et al.* RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Molecular Cell* **54**, 887-900, doi:10.1016/j.molcel.2014.04.016 (2014).
- 13 Lee, F. C. Y. & Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell* **69**, 354-369, doi:<https://doi.org/10.1016/j.molcel.2018.01.005> (2018).
- 14 Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein interactions. *Nature Methods* **16**, 225-234, doi:10.1038/s41592-019-0330-1 (2019).
- 15 Nechay, M. & Kleiner, R. E. High-throughput approaches to profile RNA-protein interactions. *Current Opinion in Chemical Biology* **54**, 37-44, doi:<https://doi.org/10.1016/j.cbpa.2019.11.002> (2020).
- 16 Hafner, M. *et al.* CLIP and complementary methods. *Nature Reviews Methods Primers* **1**, 20, doi:10.1038/s43586-021-00018-1 (2021).
- 17 Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell* **63**, 696-710, doi:10.1016/j.molcel.2016.06.029 (2016).
- 18 Bae, J. W., Kwon, S. C., Na, Y., Kim, V. N. & Kim, J.-S. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nature Structural & Molecular Biology* **27**, 678-682, doi:10.1038/s41594-020-0436-2 (2020).
- 19 Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11**, 1064-1070, doi:10.1038/nmeth.3092 (2014).
- 20 Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403.e319, doi:<https://doi.org/10.1016/j.cell.2018.11.004> (2019).
- 21 Queiroz, R. M. L. *et al.* Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology* **37**, 169-178, doi:10.1038/s41587-018-0001-2 (2019).
- 22 Urdaneta, E. C. *et al.* Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nature Communications* **10**, 990, doi:10.1038/s41467-019-08942-3 (2019).
- 23 Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393-1406, doi:<https://doi.org/10.1016/j.cell.2012.04.031> (2012).
- 24 Beckmann, B. M. *et al.* The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications* **6**, 10127, doi:10.1038/ncomms10127 (2015).
- 25 Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719, doi:10.1038/s41586-020-2077-3 (2020).
- 26 Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology* **17**, 909-915, doi:10.1038/nsmb.1838 (2010).
- 27 Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464, doi:10.1038/nature07488  
<https://www.nature.com/articles/nature07488#supplementary-information> (2008).
- 28 Ule, J. *et al.* CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212-1215, doi:10.1126/science.1090095 (2003).



- 29 Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-514, doi:10.1038/nmeth.3810 (2016).
- 30 Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**, 607-614, doi:10.1038/nbt.1873 (2011).
- 31 Sharma, D. *et al.* The kinetic landscape of an RNA-binding protein in cells. *Nature*, doi:10.1038/s41586-021-03222-x (2021).
- 32 Dorn, G. *et al.* Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat Methods* **14**, 487-490, doi:10.1038/nmeth.4235 (2017).
- 33 Williams, K. R. & Konigsberg, W. H. in *Methods in Enzymology* Vol. 208 516-539 (Academic Press, 1991).
- 34 Jellinek, T. & Johns, R. B. The mechanism of photochemical addition of cysteine to uracil and formation of dihydrouracil. *Photochemistry and Photobiology* **11**, 349-359, doi:doi:10.1111/j.1751-1097.1970.tb06007.x (1970).
- 35 Shaw, A. A., Falick, A. M. & Shetlar, M. D. Photoreactions of thymine and thymidine with N-acetyltirosine. *Biochemistry* **31**, 10976-10983, doi:10.1021/bi00160a006 (1992).
- 36 Sugimoto, Y. *et al.* Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biology* **13**, doi:10.1186/gb-2012-13-8-r67 (2012).
- 37 Feng, H. *et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Molecular Cell* **74**, 1189-1204.e1186, doi:<https://doi.org/10.1016/j.molcel.2019.02.002> (2019).
- 38 Vieira-Vieira, C. H. & Selbach, M. Opportunities and Challenges in Global Quantification of RNA-Protein Interaction via UV Cross-Linking. *Frontiers in Molecular Biosciences* **8**, doi:10.3389/fmolb.2021.669939 (2021).
- 39 Sun, G., Fecko, C. J., Nicewonger, R. B., Webb, W. W. & Begley, T. P. DNA-Protein Cross-Linking: Model Systems for Pyrimidine-Aromatic Amino Acid Cross-Linking. *Organic Letters* **8**, 681-683, doi:10.1021/ol052876m (2006).
- 40 Auweter, S. D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *The EMBO journal* **25**, 163-173, doi:10.1038/sj.emboj.7600918 (2006).
- 41 Wenter, P., Reymond, L., Auweter, S. D., Allain, F. H. & Pitsch, S. Short, synthetic and selectively <sup>13</sup>C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes. *Nucleic Acids Res* **34**, e79, doi:10.1093/nar/gkl427 (2006).
- 42 Smith, K. C. & Meun, D. H. Kinetics of the photochemical addition of cysteine-35S to polynucleotides and nucleic acids. *Biochemistry* **7**, 1033-1037 (1968).
- 43 Weyn-Vanhentenryck, Sebastien M. *et al.* HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Reports* **6**, 1139-1152, doi:10.1016/j.celrep.2014.02.005 (2014).
- 44 Shetlar, M. D. in *Photochemical and Photobiological Reviews: Volume 5* (ed Kendric C. Smith) 105-197 (Springer US, 1980).
- 45 Meisenheimer, K. M. & Koch, T. H. Photocross-linking of nucleic acids to associated proteins. *Critical Reviews in Biochemistry and Molecular Biology* **32**, 101-140, doi:10.3109/10409239709108550 (1997).
- 46 Cieniková, Z., Damberger, F. F., Hall, J., Allain, F. H. T. & Maris, C. Structural and Mechanistic Insights into Poly(uridine) Tract Recognition by the hnRNP C RNA Recognition Motif. *Journal of the American Chemical Society* **136**, 14536-14544, doi:10.1021/ja507690d (2014).
- 47 Panhale, A. *et al.* CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nature Communications* **10**, 2682, doi:10.1038/s41467-019-10585-3 (2019).
- 48 Ransey, E. *et al.* Comparative analysis of LIN28-RNA binding sites identified at single nucleotide resolution. *RNA Biol* **14**, 1756-1765, doi:10.1080/15476286.2017.1356566 (2017).
- 49 Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080-1091, doi:10.1016/j.cell.2011.10.020 (2011).

- 50 Lelyveld, V. S., Bjorkbom, A., Ransey, E. M., Sliz, P. & Szostak, J. W. Pinpointing RNA-Protein Cross-Links with Site-Specific Stable Isotope-Labeled Oligonucleotides. *J Am Chem Soc* **137**, 15378-15381, doi:10.1021/jacs.5b10596 (2015).
- 51 Ben-Shem, A. *et al.* The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* **334**, 1524-1529, doi:10.1126/science.1212642 (2011).
- 52 Oberstrass, F. C. *et al.* Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science* **309**, 2054-2057, doi:10.1126/science.1114066 (2005).
- 53 Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R. & von Hippel, P. H. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *Journal of Biological Chemistry* **261**, 3512-3518 (1986).
- 54 Cordes, M. & Giese, B. Electron transfer in peptides and proteins. *Chemical Society Reviews* **38**, 892-901, doi:10.1039/B805743P (2009).
- 55 Hawkins, C. L. & Davies, M. J. Generation and propagation of radical reactions on proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1504**, 196-219, doi:[https://doi.org/10.1016/S0005-2728\(00\)00252-8](https://doi.org/10.1016/S0005-2728(00)00252-8) (2001).
- 56 Meisenheimer, K. M., Meisenheimer, P. L. & Koch, T. H. Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods in enzymology* **318**, 88-104 (2000).
- 57 Nathanael, J. G. *et al.* Amide Neighbouring-Group Effects in Peptides: Phenylalanine as Relay Amino Acid in Long-Distance Electron Transfer. *Chembiochem* **19**, 922-926, doi:10.1002/cbic.201800098 (2018).
- 58 Viehe, H. G., Merényi, R., Stella, L. & Janousek, Z. Capped Substituent Effects in Syntheses with Radicals and Radicophiles [New synthetic methods (32)]. *Angewandte Chemie International Edition in English* **18**, 917-932, doi:<https://doi.org/10.1002/anie.197909171> (1979).
- 59 Takaya, T., Su, C., de La Harpe, K., Crespo-Hernández, C. E. & Kohler, B. UV excitation of single DNA and RNA strands produces high yields of exciplex states between two stacked bases. *Proceedings of the National Academy of Sciences* **105**, 10285-10290, doi:10.1073/pnas.0802079105 (2008).
- 60 Crespo-Hernández, C. E., Cohen, B. & Kohler, B. Base stacking controls excited-state dynamics in A-T DNA. *Nature* **436**, 1141-1144, doi:10.1038/nature03933 (2005).
- 61 Shchepachev, V. *et al.* Defining the RNA interactome by total RNA-associated protein purification. *Molecular Systems Biology* **15**, e8689, doi:<https://doi.org/10.15252/msb.20188689> (2019).
- 62 Bhat, V. *et al.* Photocrosslinking between nucleic acids and proteins: general discussion. *Faraday Discussions* **207**, 283-306, doi:10.1039/C8FD90005A (2018).
- 63 Sato, S. & Nakamura, H. Protein Chemical Labeling Using Biomimetic Radical Chemistry. *Molecules* **24**, doi:10.3390/molecules24213980 (2019).
- 64 Liu, Z. R., Wilkie, A. M., Clemens, M. J. & Smith, C. W. Detection of double-stranded RNA-protein interactions by methylene blue-mediated photo-crosslinking. *RNA* **2**, 611-621 (1996).
- 65 Masliah, G., Barraud, P. & Allain, F. H. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci* **70**, 1875-1895, doi:10.1007/s00018-012-1119-x (2013).
- 66 Bahrami-Samani, E., Penalva, L. O., Smith, A. D. & Uren, P. J. Leveraging cross-link modification events in CLIP-seq for motif discovery. *Nucleic Acids Research* **43**, 95-103, doi:10.1093/nar/gku1288 (2015).