

Nucleotide-amino acid π -stacking interactions initiate photo cross-linking in RNA-protein complexes







Anna Knörlein¹, Chris P. Sarnowski ^{2,5}, Tebbe de Vries ³, Moritz Stoltz¹, Michael Götze^{2,6}, Ruedi Aebersold ^{2,4}, Frédéric H.-T. Allain ³, Alexander Leitner ² & Jonathan Hall ¹✉

Photo-induced cross-linking is a mainstay technique to characterize RNA-protein interactions. However, UV-induced cross-linking between RNA and proteins at “zero-distance” is poorly understood. Here, we investigate cross-linking of the RBFOX alternative splicing factor with its hepta-ribonucleotide binding element as a model system. We examine the influence of nucleobase, nucleotide position and amino acid composition using CLIR-MS technology (crosslinking-of-isotope-labelled-RNA-and-tandem-mass-spectrometry), that locates cross-links on RNA and protein with site-specific resolution. Surprisingly, cross-linking occurs only at nucleotides that are π -stacked to phenylalanines. Notably, this π -stacking interaction is also necessary for the amino-acids flanking phenylalanines to partake in UV-cross-linking. We confirmed these observations in several published datasets where cross-linking sites could be mapped to a high resolution structure. We hypothesize that π -stacking to aromatic amino acids activates cross-linking in RNA-protein complexes, whereafter nucleotide and peptide radicals recombine. These findings will facilitate interpretation of cross-linking data from structural studies and from genome-wide datasets generated using CLIP (cross-linking-and-immunoprecipitation) methods.

¹Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zurich, Zurich, Switzerland. ²Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. ³Department of Biology, Institute of Biochemistry, ETH Zurich, Zurich, Switzerland. ⁴Faculty of Science, University of Zurich, Zurich, Switzerland. ⁵Present address: Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland. ⁶Present address: Department of Biology, Chemistry and Pharmacy, Institute of Chemistry and Biochemistry, Free University Berlin, Berlin, Germany. ✉email: jonathan.hall@pharma.ethz.ch

The human genome encodes more than 1500 RNA binding proteins (RBPs) that regulate key processes, including translation, localisation, stability and splicing^{1–3}. 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^{4,5}. These domains recognize short, usually single-stranded regions of 3–8 nucleotides (nt) known collectively as consensus RNA binding elements (RBE)^{6,7} that often contain degenerate positions. Additional binding affinity and selectivity can be generated via supplementary contacts between the RNA and the protein^{5,8}; for example, the RBP FUS has a bipartite binding mode comprising its ZnF domain and its RRM⁹. RNA–protein binding has also been observed with proteins that lack canonical RNA binding domains (RBDs)¹. 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 are responsible for some of its splicing activities^{10,11}.

Many state-of-the-art methods to identify RNA–protein interactions *in vivo* employ RNA–protein cross-linking induced by UV light^{12–15}. For example, by combining UV cross-linking with mass spectrometry approaches, proteins bound to given RNAs can be identified^{16–21}. Conversely, UV cross-linking and immunoprecipitation (CLIP) protocols are commonly used to identify RNA-binding sites for given proteins on a transcriptome-wide scale^{22–26}. Technical advances constantly improve these techniques^{17,27,28}, 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 (RA, AL, FA) 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 protocol²⁸.

The photo-induced reaction between amino acids and ribonucleotides occurs between free radical species at “zero distance”^{29–31}. Reactions involve mainly uridines and guanosines^{18,32,33}, but most amino acids can participate^{17,18}. Nevertheless, cross-links typically only occur at specific positions in the RNA–RBP motif, for which there is currently no mechanistic rationale³⁴. Moreover, it has proven difficult to investigate and identify factors that promote cross-linking, largely because i) the RNA–protein binding site environment, which is critical for cross-linking chemistry, cannot be simulated in simple solvents, and ii) the chemistry usually produces complex product mixtures that are difficult to characterize on a background of protein and nucleic acid UV damage³⁵.

Here, we investigate the structural requirements for the cross-linking of an RNA to its RBP partner. We use the RRM domain of the RBFOX family (FOX_{RRM}) and its RNA consensus binding motif U₁G₂C₃A₄U₅G₆U₇ (FOX_{RBE}) as a model system, exploiting the high affinity of the complex forms and its well-characterized NMR structure³⁶. We introduce ¹³C-labelled ribonucleotides into the FOX_{RBE} heptanucleotide and use CLIR-MS to identify RNA–protein cross-links with site-specific resolution. Cross-linking on the protein clusters at amino acids around two phenylalanines, consistent with previous findings³⁷. However, with few exceptions, it only occurs on the RNA at U₁, G₂ and G₆. We then employ site-specific mutagenesis to probe systematically the influence of nucleobase, nucleotide position and amino acid composition on the cross-linking profile. This reveals that cross-

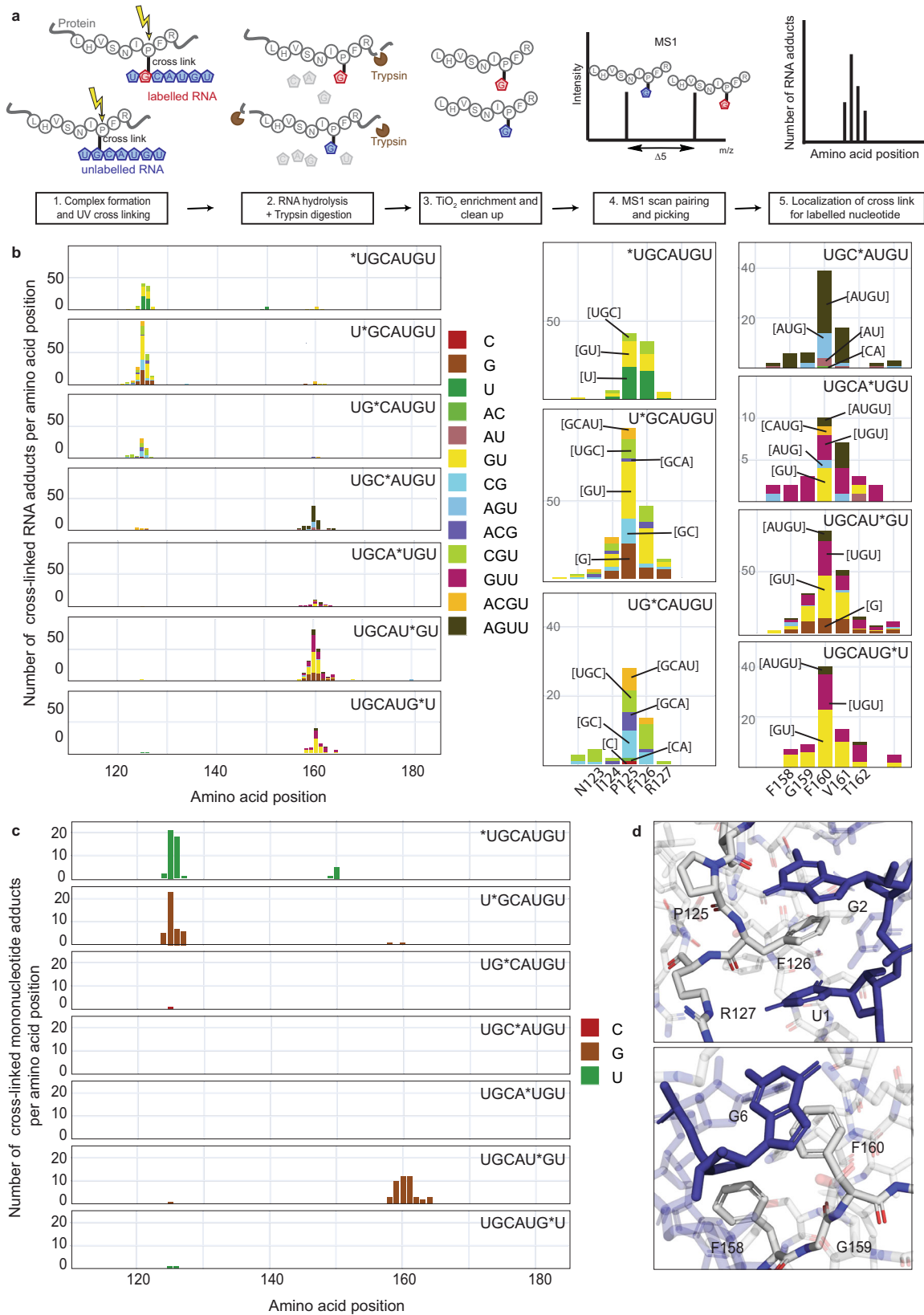
linking only occurs with guanosine or uridine at three of the seven nucleotide positions, and only when bases are stacked to aromatic amino acid side chains. Remarkably, this primary stacking interaction is required for neighbouring amino acids to participate in cross-linking. We identify and confirm the importance of this structural feature in selected published examples from other groups as well with an unbiased analysis of three large datasets, suggesting that it is of primary importance for zero-length cross-linking in native RNA–protein binding sites. Moreover, we expect that this finding will facilitate the interpretation of RNA–protein cross-linking data, especially for non-canonical binding motifs. It will also help guide the design of future cross-linking experiments and will aid the development of new tools for *de novo* motif discovery (see ref. 33).

Results

Optimization of CLIR-MS to identify RNA–protein cross-links with site-specific resolution. The original CLIR-MS protocol (Fig. 1a) employs RNAs with contiguous regions of differentially isotope-labelled nucleotides in the cross-linking step²⁸. After partial RNA and protein digestion, short peptide–oligonucleotide conjugates are identified as matched signal pairs in the precursor ion mass spectrum, which localizes the cross-linked nucleotide to the labelled RNA segment. Overlapping partial sequences then facilitates the localization of the cross-link on the RNA. A drawback of the original implementation of CLIR-MS is the inherent requirement for enzymatic ¹³C/¹⁵N-labelled RNA synthesis (i.e. *in vitro* transcription). This does not allow site-specific nucleotide labelling which is needed to unambiguously assign the reactive nucleotide. A second limitation is the nuclease digestion step, which typically produces short oligonucleotides (i.e., 1–4 nt) and is probably less efficient on nucleotides that are structurally changed by cross-linking¹⁷. In this study, we implemented chemical solutions to help circumvent these problems; we employed ¹³C-labelled phosphoramidites (Fig. S1a) during solid-phase RNA synthesis to incorporate labelled nucleotides site-specifically³⁸; and we switched from RNase digestion to alkaline hydrolysis of RNA, while exercising care not to degrade the protein or the nucleobases. Consequently, the mass analysis of the product mixtures yielded a greater fraction of peptide–mononucleotide adducts, allowing us to better identify nucleotides that are cross-linked (Fig. S1b).

U₁, G₂ and G₆ in FOX_{RBE} cross-link to amino acids centred on phenylalanines in the FOX_{RRM}. We employed a systematic approach in an effort to identify key structural requirements for cross-linking of RBFOX to RNA. We first used ¹³C-labelled versions of FOX_{RBE} to identify all points of reaction between the RNA and the protein. We then synthesized mutated variants of FOX_{RBE} to determine how cross-linking varies with respect to i) the nucleobase, ii) its positions in the RBE, and iii) how it is affected by amino acid composition in the binding site. We were mindful of the fact that mutating sites in the RNA and the protein might alter the mode of (or even abolish) RNA–protein binding, and therefore for each mutant we measured the binding affinity to FOX_{RRM} using surface plasmon resonance spectroscopy (SPR).

We synthesized the seven ¹³C-labelled versions of FOX_{RBE} and confirmed the correct incorporation of the label by liquid-chromatography mass spectrometry (LC-MS) (Fig. S2). We incubated each version of FOX_{RBE} with FOX_{RRM} and performed the CLIR-MS protocol. The 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 Fig. 1b was detected because it contains a ¹³C 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 or U). We noted that all (>99%) of the detected fragments contained at least one uridine or guanine, consistent with literature reports^{15,18,32,33,39} that uracil and guanine mainly participate in

cross-linking. Here, the use of alkaline hydrolysis for RNA digestion proved advantageous, since it digests a larger fraction of the RNA to mononucleotides (see Fig. S1b, c), thereby identifying unambiguously the cross-linking sites. Hence, focusing only on the mononucleotide species in the spectra of Fig. 1b, revealed that cross-linking in the FOX_{RBE} involved almost uniquely U₁, G₂ and

Fig. 1 Photo-induced cross-linking of FOX_{RBE} and FOX_{RRM} occurs only at U₁, G₂ and G₆. **a** Individual steps of the CLIR-MS protocol. **b** CLIR-MS analysis of site-specifically isotope-labelled FOX_{RBE} and FOX_{RRM}. Plots show the number of RNA adducts at each amino acid position. Phenylalanines are at positions 126, 158 and 160. The xQuest software was used to search for cross-linked mono-, di-, tri- and tetra-nucleotide adducts that are present in FOX_{RBE} which are colour coded (see also Supporting Data 1) (*N indicates a ¹³C-labelled nucleotide). The main cross-linking clusters of each FOX_{RBE} variant are shown enlarged on the right side. **c** Data of b) filtered for mononucleotides showing the cross-links to U₁, G₂ and G₆ (see Supporting Data 1). **d** Structure of FOX_{RRM}/FOX_{RBE}³⁶ (PDB ID: 2ERR) showing F126 and U₁/G₂ (upper panel), and F160 and G₆ (lower panel) around which the cross-linking is clustered. Structures were visualized with PyMOL (PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC).

G₆ (Fig. 1c). The absolute 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.

The cross-linking of G₂ and G₆ was consistent with published CLIP data^{33,40} (Fig. 1c). G₂ and G₆ are critical to the binding of FOX_{RBE} to FOX_{RRM}, and their exchange for A₂ or A₆ respectively, greatly reduced protein-binding and cross-linking between the amino acid clusters 126 and 160 and the mutated sites (Fig. S3). Although cross-linking from U₁ was detected in the CLIR-MS experiments, it was hardly observed at U₅ (*vide infra*) or U₇, consistent with our hypothesis that strict structural parameters govern the photo-induced reactions between FOX_{RRM} and FOX_{RBE}. Isolated cross-links were also observed in some of the spectra of Fig. 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 *UGCAUGU and U*GCAUGU (Fig. 1b, top two panels) are consistent with transient (low affinity) binding of U₁G₂ in the binding pocket occupied mainly by U₅G₆. Likewise, cross-links around F126 in the lower panels of Fig. 1b,c may derive from similarly transient contacts with U₅G₆U₇. One cross-link from P125 to cytidine is visible in Fig. 1c. A₄ did not cross-link to FOX_{RRM} (Fig. 1c). Sites of cross-linking at the protein were centered at two phenylalanines (F126 and F160), with a distribution of 1-3 amino acids flanking these positions³⁷. This was confirmed from analysis of the MS/MS spectra in which fragment ions localize the RNA adducts unambiguously on the peptide backbone (Fig. S4).

The cross-links of U₁, G₂ and G₆ aligned well with the NMR structure of FOX_{RRM} bound to FOX_{RBE}³⁶ (Fig. 1d) (PDB ID: 2ERR). The largest number of spectra corresponded to U₁ reacting with P125 and F126, and to a lesser extent with I124 and R127 (Fig. 1c). Similarly, G₂ cross-linked to I124, P125, F126 and R127. Of note, U₁ and G₂ each stack to one face of F126. Hydrogen bonds are also present between the bases of U₁ and G₂, and between R127 and I124, respectively. G₆ reacted with F160 (to which it also stacks), as well as with neighbouring amino acids at positions 158-164; F158 contacts the ribose of G₆. Notably, several close RNA-protein contacts that are visible in the NMR structure (i.e. C₃ interacting with F126 (but not stacking), G₆ stacking with R194 and U₅ stacking to H120)³⁶, 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^{41,42}. However, only three from the seven nucleotides of FOX_{RBE} engaged in efficient cross-linking, despite close contacts between all nucleotides and amino acids in the binding site. Hence, we investigated two obvious parameters that could influence cross-linking: i) the chemical reactivities of the nucleobases and the amino acids, and ii) the relative positioning of the reactive pair. By mutating selected nucleotides and amino acids in the binding pocket, we created a cross-linking structure-activity relationship for the FOX_{RRM}-FOX_{RBE} interaction.

Only uridine cross-links to FOX_{RRM} from position 1 of FOX_{RBE}. We synthesized the three labelled mutants of *NGCAUGU (N = A, G, C; Supporting Table 1), 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_{RRM} using SPR. In this assay, parent UGCAUGU bound strongly to FOX_{RRM} 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_{RBE} contributes to binding by π -stacking to F126 (Fig. 1d).

Next, we incubated the RNAs together with FOX_{RRM} and irradiated the complexes with increasing doses. Work-up and analysis by SDS-PAGE for the three NGCAUGU mutants revealed a new slow-migrating band on the gels, similar to that of the wild-type FOX_{RBE} (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 the mutants cross-linked at the N₁-position, we turned to CLIR-MS (Figs. S5, and S6). CLIR-MS data for per-labelled *N*G*C*A*U*G*U confirmed that the three FOX_{RBE} mutants exhibit the same cross-linking “fingerprint” as wild type FOX_{RBE}, i.e. in the same two amino acid clusters around positions 126 and 160 (Fig. S6a). However, in order to differentiate cross-linking of N₁ to that from G₂ in the 126-cluster, we performed CLIR-MS on the singly labelled sequences (*NGCAUGU). In contrast to U₁, cross-linking hardly occurred at A₁, G₁ or C₁ (Figs. 2c, S5), confirming the high reactivity of uridine in photo-reactions^{18,32}. Nevertheless, it was surprising that G₁ was unreactive given the reactivity of G₂, which may have been due to inappropriate orbital overlap in stacking.

In order to determine systematically the propensity for cross-linking at each position in FOX_{RBE} when a photoreactive nucleotide (i.e. U or G) is present, we performed CLIR-MS on six additional positional FOX_{RBE} mutants. Thus, we exchanged *U for C₃ and A₄ in FOX_{RBE} (UG*UAUGU, UGC*UUGU, resp.), and *G for C₃, A₄, U₅ and U₇ (UG*GAUGU, UGC*GUGU, UGCA*GGU, UGCAUG*G, resp.). In each case, we first confirmed that the mutants bound and cross-linked to FOX_{RRM} using SPR and SDS-PAGE gels (Fig. 2d, e, resp., Fig. S6a). Remarkably, in none of these six examples, did the mutated nucleotides cross-link efficiently to the protein (Fig. 2f). The lack of reactivity at U₃ (in UG*UAUGU) was particularly surprising given the close proximity of C₃ to F126 in the NMR structure.

In summary, while G₂ and G₆ in wild type FOX_{RBE} cross-linked to FOX_{RRM}, guanosine did not cross-link efficiently at any other of the other five locations in the FOX_{RBE}. Similarly, uridine readily cross-linked to FOX_{RRM} from position N₁ - where A, C and G were unreactive - but not from the four other locations in the FOX_{RBE}. 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.

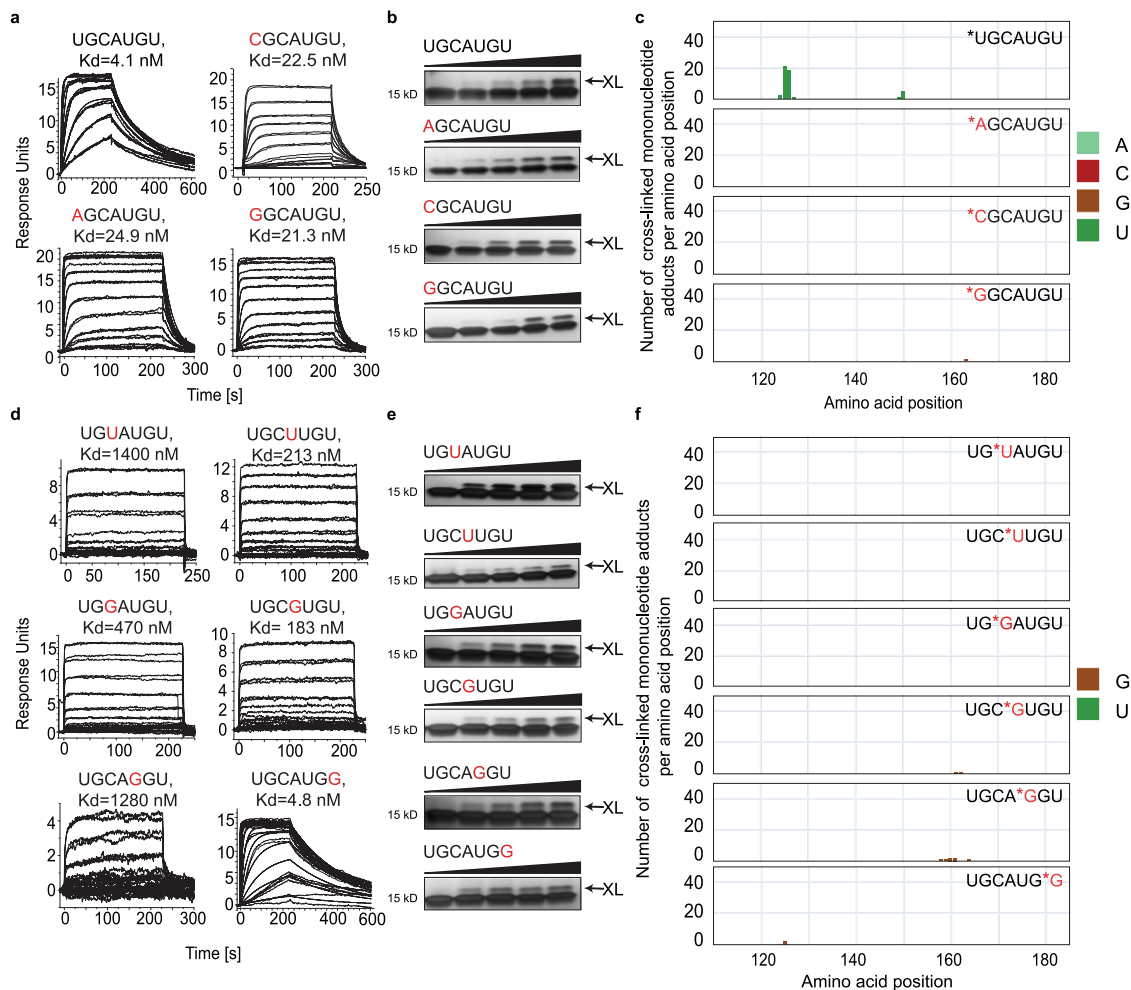
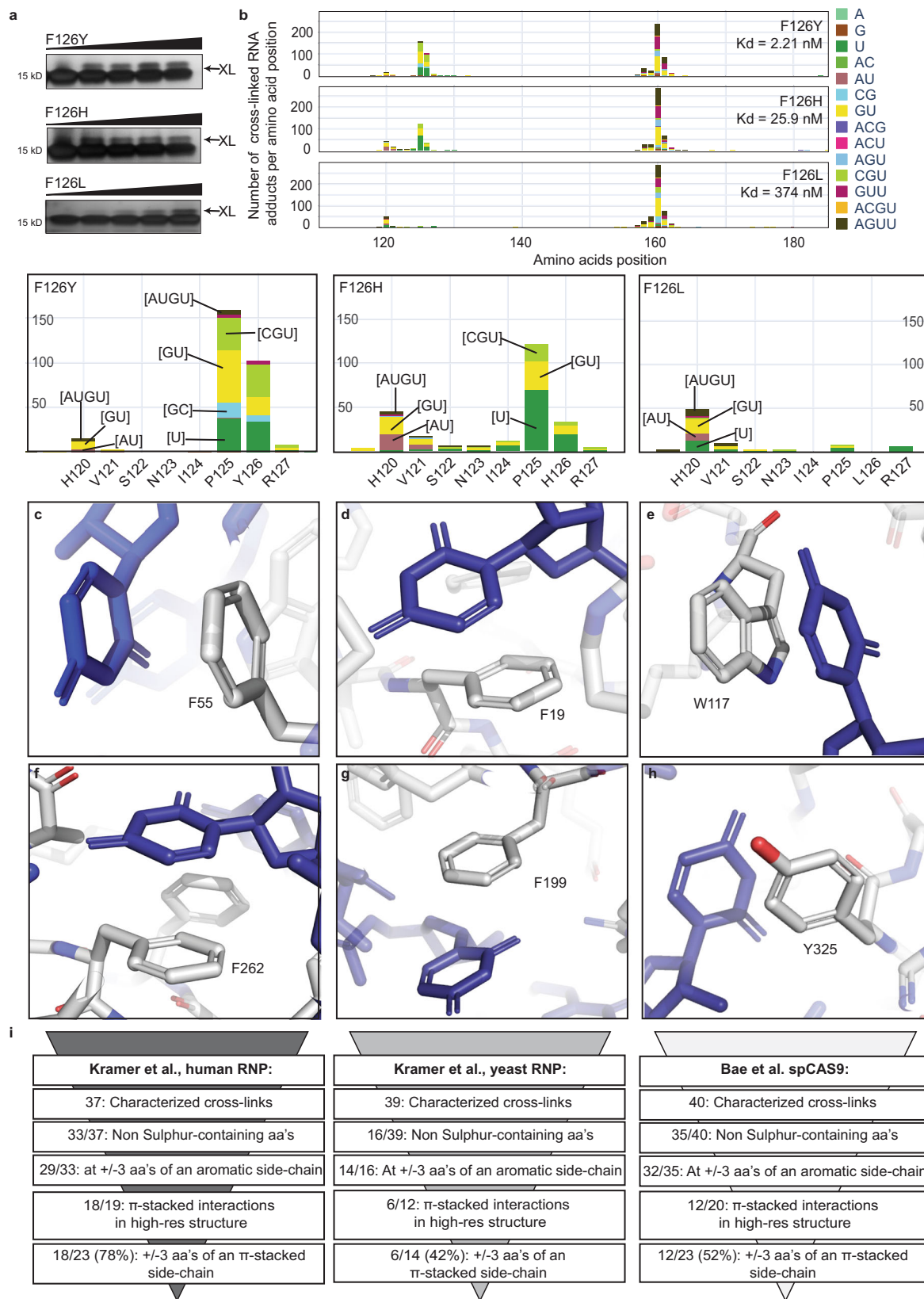


Fig. 2 FOX_{RRM}/FOX_{RBE} cross-linking is nucleotide- and site-specific. **a** SPR sensorgrams of FOX_{RBE} variants bound to FOX_{RRM}. **b** SDS-PAGE gels showing that all N₁-variants undergo cross-linking with FOX_{RRM} with increasing radiation. The cross-linking product band is indicated by “XL” (repeated three times). **c** CLIR-MS plots show that cross-linking occurs at U₁ of FOX_{RBE}, but not with N₁-mutants (the xQuest software was used to search for cross-linked mono-, di-, tri- and tetra-nucleotide adducts and the data was filtered for mononucleotides that are present in FOX_{RBE} mutants which are colour coded (*N indicates a ¹³C-labelled nucleotide, mutated nucleotides are labelled in red) (see Supporting Data 1). **d** SPR sensorgrams show that U or G mutations at N₃, N₄, N₅ and N₇ of FOX_{RBE} attenuate, but do not abolish, FOX_{RRM} binding to FOX_{RBE} variants. **e** SDS-PAGE gels show that FOX_{RBE} variants cross-link to FOX_{RRM} with increasing irradiation (repeated three times). **f** CLIR-MS plots of singly-labelled FOX_{RBE} variants show that protein-RNA cross-linking does not occur with U or G nucleotides located at N₃, N₄, N₅ and N₇ (see Supporting Data 1). (The xQuest software searches for cross-linked all mono-, di-, tri- and tetra-nucleotide that are present in FOX_{RBE} mutants, and the data was filtered for mononucleotides which are colour coded).

Aromatic amino acids play a key role in priming RNA-protein cross-linking reactions. Analysis of the aforementioned CLIR-MS data (Figs. 1c, and S6) provided two important insights: i) on the RNA side, strong cross-linking only occurred with nucleotides that were stacked to aromatic amino acids; and ii) on the protein side, cross-links involved F126 and F160, but also up to three amino-acids up- and downstream of F126 and F160.

We therefore mutated F126 in FOX_{RRM} to histidine, tyrosine and leucine. An effort to perform CLIR-MS on a tryptophan mutant failed because of protein precipitation. We have previously shown using SPR that aromatic amino acids at position 126 are crucial for binding FOX_{RBE} (F126Y: $K_d = 2.21$ nM; F126H: $K_d = 25.9$ nM), although a sterically-fitting aliphatic amino acid such as leucine can partially substitute for the phenylalanine (F126L: $K_d = 374$ nM)³⁶. We irradiated these variants in the presence of FOX_{RBE}. All three protein mutants cross-linked to FOX_{RBE}, as evident from SDS-PAGE (Fig. 3a). Next, we carried out CLIR-MS experiments with uniformly ¹³C-labelled FOX_{RBE}. F126Y and F126H cross-linked to the FOX_{RBE} similarly to FOX_{RRM} (Fig. 3b). The cross-linking profile

was similar for the three complexes at F126 and F160. However, when phenylalanine was exchanged for leucine, binding was weaker and the cross-linking to position 126 was abolished. Notably, cross-linking to the neighbouring amino acids 124-127 was also mostly lost for F126L (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, the F126H mutant appears not to cross-link to G₂, as shown by the absence of G mononucleotides (brown) or CG dinucleotides (turquoise) in Fig. 3b (Supporting Data 1). 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 preference to those of tyrosine or phenylalanine and/or that stacking to the guanosine G₂ is disturbed in this particular complex. Unexpectedly, a H120 cross-link occurred with the three FOX_{RRM} mutants, which was hardly observable in the wild type FOX_{RRM} (Fig. 3b). Analysis of the oligonucleotide fragments in Fig. 3b strongly suggested that the cross-link occurred with U₅. In fact, the NMR structure of FOX_{RRM}-FOX_{RBE} shows that U₅ adopts a stacking arrangement with H120, and thus might have



been expected to cross-link in the wild type FOX_{RBE}-FOX_{RRM} interaction (Fig. S7). Together, the data obtained from these RNA- and protein mutants suggests that π -stacking interactions between aromatic amino acids (e.g. phenylalanine, tyrosine or histidine) and guanosines or uridines are an important prerequisite for their cross-linking, not only to the aromatic side

chains, but also to the flanking amino acids. Clearly, our findings do not speak to all cross-linking reactions in RNA-protein complexes, for instance those involving sulfur-containing amino acids, such as cysteine, which is not present in the FOX_{RRM}, but which is prone to cross-link probably due to the high reactivity of the thiyl radical^{18,30,43}.

Fig. 3 Proximal aromatic amino acids mediate cross-linking in RNA protein complexes. **a** F126 mutants undergo binding and cross-linking with FOX_{RBE} with increasing irradiation, as shown by SDS-PAGE. The upper cross-linked RNA-protein band is indicated by “XL” (repeated three times). **b** CLIR-MS plots show the number of RNA adducts at each amino acid position of FOX_{RRM} mutants (the xQuest software was used to search for cross-linked mono-, di-, tri- and tetra-nucleotide adducts that are present in FOX_{RBE} which are colour coded; see also Supporting Data 1). Binding affinities of FOX_{RBE} are taken from ref. 36. **c** Structure of LIN28 in complex with preEM-let-7f obtained by Nam et al.^{50,51}. (PDB ID: 3TS0). **d** Structure of hnRNP C binding on AUUUUUC obtained by Cienikova et al.^{53,54}. (PDB ID: 2MXV). **e** Structure of the 40 S ribosomal protein S1 in complex with the 18 S rRNA obtained by Ben-Shem et al.^{18,55}. (PDB ID: 4V88). **f, g** Structure of the U2AF binding to poly-U RNA obtained by Mackereth et al.^{18,56}. (PDB ID: 2YH1). **h** Structure of the *Streptococcus pyogenes* Cas9 (spCas9) protein in complex with sgRNA obtained by Jiang et al.⁵⁷. (PDB ID: 4ZT0) (all structures are visualized with PyMOL (PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC)). **i** The proteome-wide cross-linking data of human and yeast RNP from Kramer et al.¹⁸ and the cross-linking data of bacterial protein spCas9 from Bae et al.¹⁷ were manually curated for π -stacking interactions in proximity to characterized, cross-linked (non-S-containing) amino acids. Cross-links were filtered for those falling within of ± 3 residues of an aromatic residue and overlaid with available high-resolution structures to observe possible π -stacking.

RNA-protein cross-linking correlates with π -stacking interactions in other complexes. In order to determine whether these findings apply more broadly to RNA-protein cross-linking, we examined CLIR-MS data from the alternative splicing factor PTB1 in complex with the internal ribosomal entry site (IRES) of encephalomyocarditis virus (EMCV)²⁸. Cross-links mainly clustered around five aromatic amino acids (Y127, Y267, F371, H411, H457), and comprised uridines, as reported by ref. 28. Correlating these observations with the NMR solution structure of PTB1 bound to short poly-pyrimidine sequences (CUCUCU)⁴⁴, confirmed that these amino acids were indeed all π -stacked to pyrimidines (PDB IDs: 2AD9, 2ADB, 2ADC) (Fig. S8), with cross-linking extended for a few amino acids along the protein backbone, flanking the aromatic side chains. Tyrosines Y127, Y267 and histidine H411 π -stack to uridines in CUCUCU, consistent with uridine-containing cross-links in the CLIR-MS dataset from the IRES of EMCV (Fig. S8b). Intriguingly, however, histidine H457 π -stacks to cytidines in the IRES²⁸ and CUCUCU⁴⁴, but produces high numbers of U- and UU-containing cross-links in the CLIR-MS dataset²⁸. Likewise, Cléry *et al.* observed by NMR spectroscopy a π -stacking of cytidine to Y19 in the RRM of SRSF1⁴⁵, whereas Kramer *et al.* found a uridine cross-linked to the same amino acid¹⁸. These observations suggested that C-to-U conversion (i.e. hydrolysis) might occur at π -stacked cytidines during cross-linking or in sample work-up/analysis. Although the cytosine group itself is stable to the conditions used to fragment RNA by base hydrolysis or enzyme digestion (refs. 46,47), the exocyclic amino group of cytidine is more susceptible to hydrolysis when its 5–6 double bond is reduced, i.e. in dihydrocytidine^{48,49} (Fig. S6b). Since cross-linking reactions may produce intermediates or final products in which the cytidine 5–6 carbon-carbon bond is saturated, it is plausible that C-to-U conversion only occurs at π -stacked/cross-linked cytidines. Hence, mindful of the minor differences in the masses of cytidine/uridine-containing fragments, we searched our datasets for supporting evidence of this, using an appropriate set of parameters for the xQuest software. We did not observe significant ¹³C-to-¹³U hydrolysis using CLIR-MS on UG*CAUGU and the FOX_{RBE} mutant *CGCAUGU. However, this might have been because neither of these cytidines underwent efficient cross-linking/ π -stacking to FOX_{RRM} (UG*CAUGU: Fig. 1c, third panel; *CGCAUGU: Fig. S5, second panel). Therefore, we also analyzed additional CLIR-MS data from four fully ¹³C-labelled FOX_{RBE} mutants bearing cytidines at positions N₄, N₅, N₆ and N₇ (Fig. S6c). Indeed, we found that two of the mutants (UGCACGU and UGCAUCU) produced large numbers of cross-links that - consistent with the NMR structure - could only have derived after C-to-U conversion; for example, AUU and AUGU from UGCACGU, bound to H120; and AUU and UU from UGCAUCU, bound to F160 (Fig. S6c–f). Taking together the data from the PTB1 study, that of SRSF1^{18,45} and that of these six

FOX_{RBE} RNAs, we concluded that cytidine likely undergoes partial hydrolysis mainly at positions in an RNA where it π -stacks and cross-links to the protein; for FOX_{RBE}, at positions N₅ and N₆, but not at positions N₃, N₄ and N₇. In contrast to previous assumptions^{15,17}, these findings provide direct mass-spectrometry evidence that cytidine in RNA-protein complexes readily participates in photo-induced cross-linking, especially when it is π -stacked. However, this renders it susceptible to hydrolysis to uridine, which confounds its detection and in some cases may even lead to mis-assignments during RNA-protein modeling.

Next, we sought to confirm the importance of π -stacking to RNA-protein cross-linking in datasets that were generated using techniques other than CLIR-MS. Thus, we searched for structurally well-characterized examples in literature that would speak to the generalization of our findings. A unique strength of the CLIR-MS technique is that in many cases it is possible to simultaneously identify both the precise amino acid and the ribonucleotide in a cross-linked fragment. Indeed, we identified only one published example where cross-linking at both the ribonucleotide and the amino acid were unambiguously defined by isotopic labelling, and where these sites could be mapped to a high resolution structure. In this case, ¹⁸O-RNA labelling and targeted mass spectrometry were used to localise the cross-link of U₁₁ in a let-7 microRNA to a π -stacked phenylalanine (F55) in the LIN28 cold shock domain (Fig. 3c)^{50–52}. On the other hand, we found numerous examples where amino acids involved in stacking interactions (predominantly with uracils) underwent UV-cross-linking, most likely with the same uracil but not unambiguously proven by nucleotide labelling. For example, Panhale *et al.* report a cross-link between F19 and a uridine in hnRNPC, from which the NMR structure with poly-U sequences confirms the π -stacking interaction with F19^{53,54} (Fig. 3d). Kramer *et al.* used a sophisticated workflow to pin-point cross-linking sites on a broad scale from ribonuclear protein complexes (RNPs) isolated from human and yeast cells¹⁸. By correlating their cross-linking data from ribosomal yeast protein S1 with the crystal structure of the protein (PDB ID: [4V88]) (Fig. 3e)⁵⁵, we confirmed that tryptophan W117 π -stacks and cross-links to uridine U₁₇₉₉ from ribosomal S1. Similarly, the same group localised RNA cross-links on the human splicing factor U2AF 65-kDa subunit to amino acids L261, F262 and F199; according to the crystal structure, F262 and F199 both π -stack to uridines in complex with poly-U RNA (Fig. 3f, g) (PDB ID: 2YH1)⁵⁶. Bae *et al.* showed that tyrosines Y325 (Fig. 3h), Y450 and Y1356 in the *Streptococcus pyogenes* Cas9 (spCas9) protein all cross-link with RNA¹⁷; the crystal structure of spCas9 shows that all three residues are π -stacked to uridines or guanosines⁵⁷.

These well-characterized, selected examples already provided supporting evidence for the generality of our findings. However, the aforementioned examples of Kramer *et al.* and Bae *et al.* were

extracted from large well annotated datasets in which, collectively, more than 100 RNA-protein cross-links from a wide variety of RBPs are catalogued. These datasets therefore offered an opportunity to analyze in an unbiased fashion the putative link between π -stacking and cross-linking. The two proteome-wide datasets reported by Kramer et al. each comprise approximately 60 RNA-protein cross-links generated from affinity-captured nuclear pre-mRNAs from human cells¹⁸ and from pre/mRNAs of yeast cells¹⁸. The third dataset reports 84 cross-links to spCAS9, which forms a complex with single guide RNAs¹⁷. We manually annotated each of the three datasets in a systematic fashion in order to determine whether amino acids that undergo cross-linking are located within $+/-3$ positions of an aromatic amino acid side chain (mindful that in a fully random sequence, 20% of the amino acids may be aromatic) and if yes, whether said aromatic side-chains π -stack to nucleobases.

The human RNP dataset¹⁸ details 60 cross-links to approximately 35 proteins, with 37 cross-links that are localized on defined amino acids, in mostly RRM binding domains (Supporting Data 4). From these, 33 cross-links are assigned to non-cysteine and non-methionine amino acids (Fig. 3i; Supporting Data 4), 29 of which are located within three amino acids of an aromatic side chain. High-resolution structures were informative for 19 of these amino acids and, pleasingly, showed that 18 of the aromatic side chains were involved in apparent π -stacking interactions, and one which was not. Taking into account also the four cross-links which are not close to an aromatic amino acid, means that 18/23 (78%) cross-links occur close to a π -stacked aromatic side chain, fully consistent with our findings. In this dataset, neither of the KH domain-bearing proteins carry aromatic amino acids close to cross-links, although both underwent cross-linking to cysteines, demonstrating that cysteine does not follow the pattern, as expected. In contrast, a positive π -stacking/cross-linking association (to an adenine) was present for the cold shock domain of Y-box binding protein, as well as for ribosomal proteins S2, L5, L6 and L34 with distinct domains.

The yeast RNP dataset¹⁸ contains 39 defined cross-links to 52 proteins, containing a variety of domains (Supporting Data 4). Surprisingly, 23/39 cross-links involve cysteines, which the authors suggested might be due to the presence of dithiothreitol (DDT) in the yeast sample which is known to promote cross-links involving cysteine residues^{18,58}. Fourteen cross-links lie within three amino acids of an aromatic side-chain, for which 12 high-resolution structures are available. These show that six cross-links occur at apparent π -stacking interactions, and for two cross-links high-resolution structures are not available (Fig. 3i; Supporting Data 4). Finally, the outcome of cross-linking of the spCAS9 protein to RNA was reported by ref. ¹⁷. Cross-links comprising 40 amino acids were catalogued, of which five were cysteine or methionines and were discarded from further analysis. Of these 35 cross-links, 32 lie within three amino acids of an aromatic side chain and of these, 20 can be studied with the high-resolution structure. Twelve of the cross-links involve apparent π -stacking interactions (Fig. 3i; Supporting Data 4), whereas eleven cross-links are not close to a π -stacking interaction. In summary, this unbiased analysis confirmed the association of cross-linking with π -stacking in a variety of RNA-binding domains for totals of 78, 42 and 52% of the cross-links in studies performed by independent groups in yeast, bacterial and human systems.

Taken together with the aforementioned specific examples from literature and our analysis of the FOX_{RRM} and PTPB1 CLIR-MS data, the data overall strongly supports the importance of π -stacking to the RNA-protein cross-linking chemistry. The absence of a positive correlation for some cross-links may be due to a variety of reasons; e.g. different conditions for protein structures/domain determination in vitro and cross-linking

experiments performed in vivo on protein complexes; or cross-linking reactions that occur as a result of transient interactions (i.e. artifacts). In addition, the lack of structural information for several cross-links in the human RNP and spCAS9 datasets may have prevented an even higher correlation. Finally, it is also apparent that a π -stacking interaction is not a strict requirement for all cross-linking events. Cysteine, which is prone to cross-linking, does not require a π -stacking interaction in order to produce a long-lived, highly reactive radical^{18,30,43}. This is consistent with the lack of aromatic amino acids proximal to cysteine-containing cross-links in the KH domains of proteins in the yeast and human RNP data-sets¹⁸. Thiol-containing molecules present in buffer may also initiate UV-induced cross-linking of proteins and nucleic acids⁵⁸. In addition, recent publications described cross-linking of dsDNA to histones using conventional cross-linking⁵⁹, where π -stacking of the side chain is more difficult to envision because of the double-stranded helical structure; although this may partly explain why double-stranded oligonucleotides are reported to cross-link less efficiently than single-stranded oligonucleotides^{60,61}.

Photo-induced electron transfer in a π -stacked RNA-protein complex may mediate radical reactions of cross-linking.

Free radical reactions of nucleic acids and proteins have been well studied in the context of oxidative damage and electron transfer^{43,62}, but less thoroughly investigated for RNA-protein interactions^{29,41}. However, a description 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 U₁ with F126 (Fig. 4a)³⁵. Hence, photo-induced electron transfer between U₁ and F126 generates a short-lived anion/cation radical pair (exciplex) (Fig. 4b; structures 1 and 2). Subsequent protonation of the uracil radical anion can yield a neutral α -hydroxy radical⁴³, 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 the formation of the direct U₁-F126 cross-link (Fig. 4b; structure 4). An analogous mechanism has been proposed for the reaction between uracils/halogenated uracils and tyrosine derivatives^{31,63}.

Alternatively, the F126 radical, or radical cation, may rearrange to neighboring amino acids in processes mediated by hydrogen atom abstractions⁴³, or via oxidation of amide carbonyls (by the F126 radical cation)⁶⁴, yielding free radicals at peptide α -carbon sites on the protein backbone. Viehe *et al* have proposed that α -carbon radicals are especially stabilized thermodynamically by captodative effects, i.e. simultaneously by electron-withdrawing ($-C=O$) and electron-donating ($-NR_2$) groups⁶⁵ 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 U₁ may form, yielding products that are identified by mass spectral analysis after controlled digestion (e.g. structures 5–7; Fig. 4b). These steps are consistent with the outcome of cross-linking reactions of the F126 mutants. Thus, 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. Several efforts to mimic some of these cross-linking reactions in solutions were unsuccessful, confirming the crucial role played by the local protein binding site environment. Based on the similarity of the cross-linking profiles from U₁ and G₂ (Fig. 1c), it seems intuitively likely that guanosines G₂ and G₆ may follow a similar mechanistic reaction path as U₁. Thus, photo-excitation of the stacked guanine-phenyl ring systems produces free radicals at

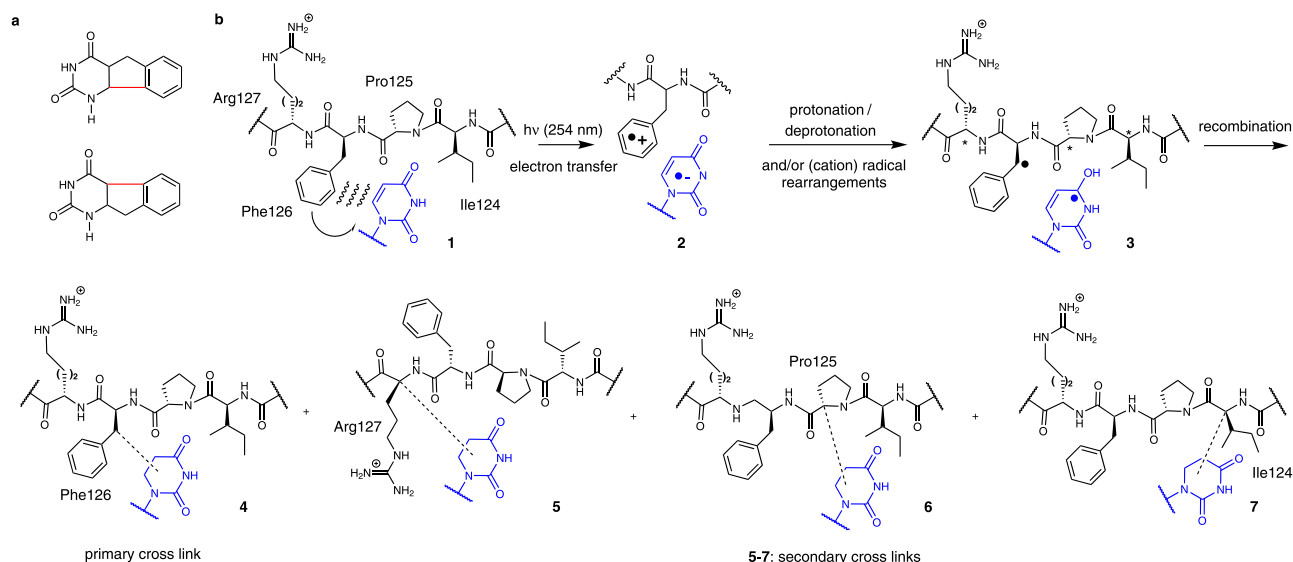


Fig. 4 Proposed mechanism for UV induced RNA-protein cross-linking. **a** UV induced cyclization of 5-benzyluracil and 6-benzyluracil after Sun et al.³⁵. (bonds in red are formed upon cyclisation). **b** Possible mechanism of UV cross-linking between the stacked F126 and U₁ of the FOX_{RRM} (1). Photo-induced 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 U₁ 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 α-carbons (*).

G₂ and G₆, as well as on the peptide backbone around F126 and F160. Recombination yields direct and indirect cross-links, which in the case of G₂ are to the same α-carbon radicals that couple with U₁. 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 π-stacked pyrimidine and purine nucleobases that produce long-lived exciplexes^{66,67}. 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. ^{64,66}), this may not necessarily correlate with the measured 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.

Discussion

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. Therefore, any progress that furthers our understanding of this chemistry is of high value.

The CLIR-MS method²⁸ 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 study, we have broadened the application of CLIR-MS through the use of chemically synthesized ¹³C-labelled RNAs. This enables site-specific incorporation of labelled nucleotides into the RNA. After irradiation of the RNA-protein complex, and controlled digestion to nucleotide-peptide adducts, the locations of cross-linked nucleotides 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 structure³⁶ and studied cross-linking³⁷. Photo-irradiation of the FOX_{RRM}-FOX_{RBE} complex led to key observations with potentially wide-ranging implications: 1) strong cross-linking occurred between U₁, G₂ and G₆ with clusters of amino acids centred around the phenylalanines F126 and F160; 2) very little cross-linking was observed at other uridines in the parent or a mutated FOX_{RBE}; and 3) amino acids that flank F126 and F160 also cross-linked efficiently to U₁, G₂ and G₆, but not to other nucleotides of FOX_{RBE}. Since the NMR structure of FOX_{RRM}-FOX_{RBE}³⁶ shows that U₁ and G₂ π-stack to F126, and that G₆ π-stacks with F160, the data suggested that a π-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 (flanking) cross-linking to U₁/G₂. Other researchers have noted in passing the increased presence of aromatic amino acid side chains in UV cross-linking datasets (see refs. ^{17,18,29,52,68}), but have not to our knowledge recognized its role as a trigger for cross-linking, nor distinguished between direct and indirect cross-link events. We validated our results on the RBFOX system with the correlation of published cross-linking and structural data from CLIR-MS data generated with the PTBP1 protein, and selected examples from LIN28, hnRNPC, U2AF, ribosomal yeast protein S1 and bacterial spCAS9, that were produced using different methods. Our findings were further strengthened by an unbiased analysis of more than 100 cross-links in large-scale data-

sets comprising various RNA-binding domains, where in one case up to 78% of the cross-links showed π -stacking to a proximal aromatic amino acid side chain. It is clear that factors in addition to π -stacking also contribute to cross-linking events in RNA-protein sites, including efficiency of the photo-induced electron transfer between nucleobase and amino acids, the ability to stabilize free radicals, the flexibility of the structure to adopt to the configurations that are required for the radical reactions⁶⁹ and the proximity of reacting radical pairs⁷⁰. Furthermore, our findings do not explain all RNA-protein cross-linking reactions, including those involving cysteine, which is highly photoreactive and prone to cross-link probably due to the high reactivity of the thyl radical^{18,30,43}.

The major findings in this study were enabled by the combination of site-specific labeling with the CLIR-MS protocol, which together provides enhanced knowledge of cross-linking sites at single-nucleotide and amino acid resolution. These included the surprising discovery that cytidine residues which are π -stacked to aromatic residues can undergo partial hydrolysis during photo-induced cross-linking. This observation may explain discordance in some cases between structural- and cross-linking data. Furthermore, the hydrolysis of cytidine should be anticipated in the analysis of CLIR-MS data and may also be relevant to the interpretation of data from CLIP experiments, which is currently an area of intense activity¹⁵.

CLIR-MS technology is inherently flexible and we are exploring further improvements to the method³⁷. However, the method described here requires the use of chemically synthesized, isotope-labelled RNA and is currently restricted to the study of purified individual RNA-protein complexes. Nevertheless, the data produced can aid the interpretation of that from unbiased complex systems. For instance, our findings extended the knowledge on the role of the local environment to cross-linking in the binding site, i.e. beyond the simple proximity of photo-reactive nucleotides and amino acids. This helps at least partly to explain why cross-links occur only at specific sites in an RNA-RBP motif. Furthermore, the localization of π -stacking interactions will aid the interpretation of proteome-wide datasets, for example in cases where proteins lack canonical RNA-binding domains, and in the analysis of CLIP datasets. Thanks to 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^{33,71} for de novo motif discovery.

In a broader sense, the RNA-binding sites of RBPs have garnered attention in the context of disease and drug targeting; for example, the RNA binding site in the intrinsically disordered region of TDP43 contributes to its aggregation in amyotrophic lateral sclerosis (see ref. 17). A fuller understanding of how RNA binding domains in proteins engage with RNAs can support the development of new methods of targeting RBPs via the RNA binding site⁷².

Methods

Protein expression and purification. The FOX_{RRM} and its mutants were expressed in transformed BL21 Codon+ *Escherichia coli* at 37 °C in LB medium with kanamycin and chloramphenicol³⁶. The cells were induced with 1 mM IPTG and after 4 h the cells were harvested by centrifugation. Cells were lysed in 50 mM Na₂HPO₄, 1 M NaCl, pH=8 using a cell cracker and the cell lysate was centrifuged at 17 000 rpm at 4 °C for 30 min. The supernatant was purified using a NiNTA affinity column (Ni-NTA agarose, Qiagen). After washing with buffer 50 mM Na₂HPO₄, 3 M NaCl, pH=8, the protein was eluted with a step gradient of imidazole (40–500 mM). The purest fractions as judged by 5–20 % SDS-PAGE were combined, and the column was repeated. Pure fractions were dialyzed against 5 L 20 mM NaCl, 10 mM NaH₂PO₄, pH=6.5 overnight. The identity of the FOX_{RRM} and its mutants was confirmed using LC-MS/MS measured by top-down analysis and data analysed using ToPIC^{73,74}.

For the biotinylated FOX_{RRM}, the 15-amino acid *E. coli* biotin ligase recognition sequence GLNDIFEAQKIEWHE was introduced between the TEV cleavage site and the gene encoding FOX-1 using standard PCR mutagenesis. *E. coli* protein ligase BirA was cloned, expressed and purified as previously described⁷⁵. The generation of biotinylated FOX_{RRM} was achieved in a 10 ml batch-mode cell-free synthesis reaction which was conducted for 3.5 h in presence of 2 μ M BirA and 400 μ M D-biotin^{75,76}. The proteins were purified as described above and the biotinylated proteins were cleaved overnight at 4 °C with 0.5 mg TEV protease⁷⁶.

RNA synthesis. The synthesis of all oligonucleotides was carried out with the MM12 synthesizer (Bio Automation Inc., Plano, TX) on a 50 nmol scale using 500 Å UnyLinker CPG using standard conditions. Synthesis conditions, purification methods and characterisation (Supplementary Table 1, Fig. S10) are listed in supplementary methods.

Surface Plasmon Resonance Spectroscopy (SPR). The SPR analysis was carried out on the MASS-1 or SPR-2 from Sierra Sensors (Hamburg, DE). For coating, the amine chip was first treated with PBS buffer at a flow rate of 12.5 μ l/min at a pH of 7.5. Next, a solution of 1 M NaCl and 1 M NaOH was injected to all 16 channels for 2 min. Afterwards, 100 μ l of a mixture of 200 mM EDC and 100 mM NHS was added. For coating of the streptavidin, an acetate buffer (10 mM sodium acetate) at a pH of 5.5 was used and a 100 μ l injection resulted in an approx. response of 2500 RU. The running buffer was switched to a HEPES buffer (10 mM HEPES at pH 7.4, 200 mM NaCl, 3.4 mM EDTA, 0.01 % (v/v) Tween 20) before capturing the analyte. Approximately 10 μ l of a 75 nM solution of biotinylated FOX_{RRM} in HEPES buffer was injected only on the second channel resulting in a response of approx. 200 RU. The amount of the injected ligand varied depending on the desired coating. 100 μ l of the analyte was injected at a flow rate of 25 μ l/min with a dissociation time of 480 s. For regeneration, 50 μ l of a 2 M NaCl solution was used. After every injection, a buffer injection was added for double referencing. The binding affinities were determined from kinetic measurements or using steady-state measurements.

Cross-linking and gel electrophoresis. Complexes of FOX_{RRM} and the desired RNA were prepared by mixing both components in equimolar ratios at the desired concentration of 10 μ M in 10 mM sodium phosphate (pH = 6.5) and 50 mM NaCl and incubated for 10 min on ice. 15 μ l of the sample solutions were placed in a 96 well-plate on ice and irradiated at 800 mJ/cm², 1600 mJ/cm², 2400 mJ/cm² and 3200 mJ/cm² at 254 nm in a CL-1000 Ultraviolet Crosslinker (UVP, Cambridge). The samples were then loaded on a 4–20% Tris-Glycine SDS-Gel with a 1xTris/Glycine/SDS running buffer. The gels were stained using the Pierce Silver Stain Kit and uncropped pictures of the gels can be found in the Supplementary Information.

Cross-linking and mass spectrometry. 75 μ g of RNA-protein complexes were made of equimolar mixtures of unlabelled and ¹³C-labelled RNA and irradiated four times with 800 mJ/cm² as described above. Each irradiation step was separated by 1 min for sample cooling. After irradiation, samples were precipitated with 3 volumes of ethanol at –20 °C and 1/10 volumes 3 M sodium acetate (pH 5.2), left at –20 °C for at least 2 h, and centrifuged at 4 °C for 30 min at 13,000 \times g. Resulting pellets were washed by brief vortexing in 80% ethanol at –20 °C, and centrifugation was repeated. For the digestion with alkaline hydrolysis: Pellets were air dried for 10 min, then were resuspended in 50 μ l of 50 mM Tris-HCl (pH 7.9). 1 ml 0.1 M NaOH was added, and the sample incubated at 70 °C for 10 min on a shaking incubator. The sample was neutralized with 105 μ l 1 M HCl, cooled on ice, purified using solid-phase extraction and evaporated to dryness in a vacuum centrifuge. The sample was resuspended in 50 μ l 50 mM Tris-HCl, pH 7.9, 4 M urea and then diluted with 150 μ l 50 mM Tris-HCl, pH 7.9. The exact procedures of the digestion using RNases and trypsin, the enrichment using titanium dioxide affinity chromatography, and LC-MS analysis²⁸ are described in the Supplementary Information. All identified cross-links are listed in Supporting Data 2 and the masses of the RNA adducts and neutral mass losses are given in Supporting Data 3.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE⁷⁷ partner repository with the dataset identifier PXD031381.

The referenced accession codes for the structures in the Protein Data Bank are 2ERR, 2AD9, 2ADB, 2ADC, 3TS0, 2MXY, 4V88, 2YH1, 4ZT0.

Received: 23 July 2021; Accepted: 20 April 2022;

Published online: 17 May 2022

References

- Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
- Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
- Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22**, 185–198 (2021).
- Cléry, A. & Allain, F. H. In *RNA binding proteins* (ed. Zdravko Lorkovic) 137–158 (CRC Press, 2012).
- Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479 (2007).
- Mitchell, S. F. & Parker, R. Principles and Properties of Eukaryotic mRNPs. *Mol. Cell* **54**, 547–558 (2014).
- Steff, R., Skrisovska, L. & Allain, F. H.-T. RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.* **6**, 33–38 (2005).
- Jankowsky, E. & Harris, M. E. Specificity and nonspecificity in RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* **16**, 533 (2015).
- 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. *Mol. Cell* **73**, 490–504.e496 (2019).
- 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. *Nat. Struct. Mol. Biol.* **27**, 901–912 (2020).
- Yeo, G. W. et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16**, 130 (2009).
- Lee, F. C. Y. & Ule, J. Advances in CLIP Technologies for Studies of Protein–RNA Interactions. *Mol. Cell* **69**, 354–369 (2018).
- Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA–protein interactions. *Nat. Methods* **16**, 225–234 (2019).
- Nechay, M. & Kleiner, R. E. High-throughput approaches to profile RNA–protein interactions. *Curr. Opin. Chem. Biol.* **54**, 37–44 (2020).
- Hafner, M. et al. CLIP and complementary methods. *Nat. Rev. Methods Prim.* **1**, 20 (2021).
- Castello, A. et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696–710 (2016).
- 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. *Nat. Struct. Mol. Biol.* **27**, 678–682 (2020).
- Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
- Trendel, J. et al. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391–403.e319 (2019).
- Queiroz, R. M. L. et al. Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* **37**, 169–178 (2019).
- Urdaneta, E. C. et al. Purification of cross-linked RNA–protein complexes by phenol-toluol extraction. *Nat. Commun.* **10**, 990 (2019).
- Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
- Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
- Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
- Ule, J. et al. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212–1215 (2003).
- Zhang, C. & Darnell, R. B. Mapping in vivo protein–RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29**, 607–614 (2011).
- Sharma, D. et al. The kinetic landscape of an RNA-binding protein in cells. *Nature* **591**, 152–156 (2021).
- 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 (2017).
- Williams, K. R. & Konigsberg, W. H. In *Methods in Enzymology* Vol. **208** 516–539 (Academic Press, 1991).
- Jellinek, T. & Johns, R. B. The mechanism of photochemical addition of cysteine to uracil and formation of dihydrouracil. *Photochemistry Photobiol.* **11**, 349–359 (1970).
- Shaw, A. A., Falick, A. M. & Shetlar, M. D. Photoreactions of thymine and thymidine with N-acetyltyrosine. *Biochemistry* **31**, 10976–10983 (1992).
- Sugimoto, Y. et al. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein–RNA interactions. *Genome Biol.* **13**, R67.GB. Abstract (2012).
- Feng, H. et al. Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein–RNA Crosslink Sites. *Mol. Cell* **74**, 1189–1204.e1186 (2019).
- Vieira-Vieira, C. H. & Selbach, M. Opportunities and Challenges in Global Quantification of RNA–Protein Interaction via UV Cross-Linking. *Front. Mol. Biosci.* **8**, 669939 (2021).
- 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. *Org. Lett.* **8**, 681–683 (2006).
- Auweter, S. D. et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
- Götze, M. et al. Single Nucleotide Resolution RNA–Protein Cross-Linking Mass Spectrometry: A Simple Extension of the CLIR-MS Workflow. *Anal. Chem.* **93**, 14626–14634 (2021).
- Wenter, P., Raymond, L., Auweter, S. D., Allain, F. H. & Pitsch, S. Short, synthetic and selectively ¹³C-labeled RNA sequences for the NMR structure determination of protein–RNA complexes. *Nucleic Acids Res.* **34**, e79 (2006).
- Smith, K. C. & Meun, D. H. Kinetics of the photochemical addition of cysteine-35S to polynucleotides and nucleic acids. *Biochemistry* **7**, 1033–1037 (1968).
- Weyn-Vanhenenryck, Sebastien M. et al. HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Rep.* **6**, 1139–1152 (2014).
- Shetlar, M. D. In *Photochemical and Photobiological Reviews: Volume 5* (ed. Kendrick C. Smith) 105–197 (Springer US, 1980).
- Meisenheimer, K. M. & Koch, T. H. Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 101–140 (1997).
- Hawkins, C. L. & Davies, M. J. Generation and propagation of radical reactions on proteins. *Biochimica et Biophysica Acta (BBA) - Bioenerg.* **1504**, 196–219 (2001).
- Oberstrass, F. C. et al. Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science* **309**, 2054–2057 (2005).
- Cléry, A. et al. Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. *Nat. Commun.* **12**, 428 (2021).
- Bock, R. M. In *Methods in Enzymology* Vol. **12** 224–228 (Academic Press, 1967).
- Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29**, 2532–2537 (1990).
- Green, M. & Cohen, S. S. Studies on the biosynthesis of bacterial and viral pyrimidines III. Derivatives of dihydrocytosine. *J. Biol. Chem.* **228**, 601–609 (1957).
- Labet, V. et al. Hydrolytic Deamination of 5,6-Dihydrocytosine in a Protic Medium: A Theoretical Study. *J. Phys. Chem. A* **114**, 1826–1834 (2010).
- 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 (2011).
- Lelyveld, V. S., Bjorkbom, A., Ransley, 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 (2015).
- Ransley, E. et al. Comparative analysis of LIN28-RNA binding sites identified at single nucleotide resolution. *RNA Biol.* **14**, 1756–1765 (2017).
- 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. *J. Am. Chem. Soc.* **136**, 14536–14544 (2014).
- Panhale, A. et al. CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nat. Commun.* **10**, 2682 (2019).
- Ben-Shem, A. et al. The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* **334**, 1524–1529 (2011).
- Mackereth, C. D. et al. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408–411 (2011).
- Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
- Zaman, U. et al. Dithiothreitol (DTT) Acts as a Specific, UV-inducible Cross-linker in Elucidation of Protein–RNA Interactions. *Mol. Cell. Proteom.* **14**, 3196–3210 (2015).
- Stützer, A. et al. Analysis of protein–DNA interactions in chromatin by UV induced cross-linking and mass spectrometry. *Nat. Commun.* **11**, 5250 (2020).
- 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).
- Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip. Rev. RNA* **9**, e1436 (2018).
- Cordes, M. & Giese, B. Electron transfer in peptides and proteins. *Chem. Soc. Rev.* **38**, 892–901 (2009).
- Meisenheimer, K. M., Meisenheimer, P. L. & Koch, T. H. Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods Enzymol.* **318**, 88–104 (2000).

64. Nathanael, J. G. et al. Amide Neighbouring-Group Effects in Peptides: Phenylalanine as Relay Amino Acid in Long-Distance Electron Transfer. *Chembiochem: a Eur. J. Chem. Biol.* **19**, 922–926 (2018).
65. Viehe, H. G., Merényi, R., Stella, L. & Janousek, Z. Capto-dative Substituent Effects in Syntheses with Radicals and Radicophiles [New synthetic methods (32)]. *Angew. Chem. Int. Ed. Engl.* **18**, 917–932 (1979).
66. 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. *Proc. Natl Acad. Sci.* **105**, 10285–10290 (2008).
67. Crespo-Hernández, C. E., Cohen, B. & Kohler, B. Base stacking controls excited-state dynamics in A-T DNA. *Nature* **436**, 1141–1144 (2005).
68. Shchepachev, V. et al. Defining the RNA interactome by total RNA-associated protein purification. *Mol. Syst. Biol.* **15**, e8689 (2019).
69. Bhat, V. et al. Photocrosslinking between nucleic acids and proteins: general discussion. *Faraday Discuss.* **207**, 283–306 (2018).
70. Sato, S. & Nakamura, H. Protein Chemical Labeling Using Biomimetic Radical Chemistry. *Molecules* **24**, 3980 (2019).
71. 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 Res.* **43**, 95–103 (2015).
72. Ghidini, A., Cléry, A., Halloy, F., Allain, F. H. T. & Hall, J. RNA-PROTACs: Degradors of RNA-Binding Proteins. *Angew. Chem. Int. Ed. Engl.* **60**, 3163–3169 (2021).
73. Toby, T. K. et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat. Protoc.* **14**, 119–152 (2019).
74. Kou, Q., Xun, L. & Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **32**, 3495–3497 (2016).
75. Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E. Jr & Wallace, J. C. The C-terminal domain of biotin protein ligase from *E. coli* is required for catalytic activity. *Protein Sci.* **10**, 2608–2617 (2001).
76. Michel, E. & Wüthrich, K. High-yield *Escherichia coli*-based cell-free expression of human proteins. *J. biomolecular NMR* **53**, 43–51 (2012).
77. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2018).

Acknowledgements

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. We thank Phil Becker for his help with the data analysis. 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.A.), the

ERC-2014AdG 670821 from the European Research Council to R.A., the NCCR RNA and Disease of the SNSF (Grant number: 51NF40-182880 to A.L., F.A., J.H.) 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 (A.K., C.S., T.d.V., M.S., M.G., R.A., F.A., A.L., J.H.) interpreted the data. A.K. and J.H. wrote the manuscript together. All authors (A.K., C.S., T.d.V., M.S., M.G., R.A., F.A., A.L., J.H.) contributed to manuscript revisions and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30284-w>.

Correspondence and requests for materials should be addressed to Jonathan Hall.

Peer review information *Nature Communications* thanks Markus Landthaler, Christof Lenz, and Jernej Ule for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022