Bioconjugates

https://doi.org/10.1038/s41557-022-01116-7

A (cross)link in the chains

Christopher P. Watkins & Ryan A. Flynn

Interactions between proteins and non-proteinaceous biopolymers are essential for life; however, many methods used to characterize these interactions lack precision and display significant biases. Now, a genetically encoded method employing sulfur(VI) fluoride exchange (SuFEx)-based chemical crosslinking has been developed for capturing and analysing protein–RNA and protein–carbohydrate interactions in vivo.

A significant portion of RNA biology and glycobiology is concerned with the following two questions: what RNA or glycan group interacts with any given binding protein? How and with what do these binding proteins discriminate amongst their many ligands? Discovering the answers to these central questions requires tools that determine not just what biopolymer is binding which protein at any time, but also which of the monomers on each biopolymer enable these interactions, even if transient, and where they are specifically located on the relevant binding partner. Transient and weak interactions are responsible for a great deal of biological activity — particular in RNA biology and glycobiology — yet their very nature makes them difficult to study using conventional methods. Typically, crosslinking applications, proximity labelling or proximity-induced enzymatic reactions are employed to provide proof of these interactions. Yet approaches such as ultraviolet (UV) crosslinking have significant limitations. In two articles published in *Nature Chemistry*, two teams led by Lei Wang have now developed a strategy for the in-cell detection of two critical types of non-covalent, transient interaction: those between proteins and RNA[1] and proteins and carbohydrates[2]. These two approaches both make use of highly specific, broadly applicable, proximity-induced chemical crosslinking reactions.

Many unbiased investigations into protein–RNA and protein–carbohydrate interactions utilize UV crosslinking, with or without the aid of photoactivatable crosslinking reagents. Crosslinking using UV usually functions by irradiating samples with low (<300 nm) wavelength light to excite electrons such that they react with proximate molecules. Further enhancements have been made by incorporating photoreactive substrates, such as 4-thiouridine (s^4U)[3] or azide- or diazirine-functionalized sugars[4,5], which allow longer (and less damaging) wavelengths to be utilized, and greatly increase crosslinking efficiency. These methods have provided substantial insight into RNA and glycobiology and the proteins that interact with these biopolymers, to the point of providing proteomic maps of interactions and the discovery of RNA-binding proteins that possess no canonical RNA-binding domains[6], as well as identifying glycan 'reader' proteins and capturing glycan-mediated protein–protein interactions[4,7].

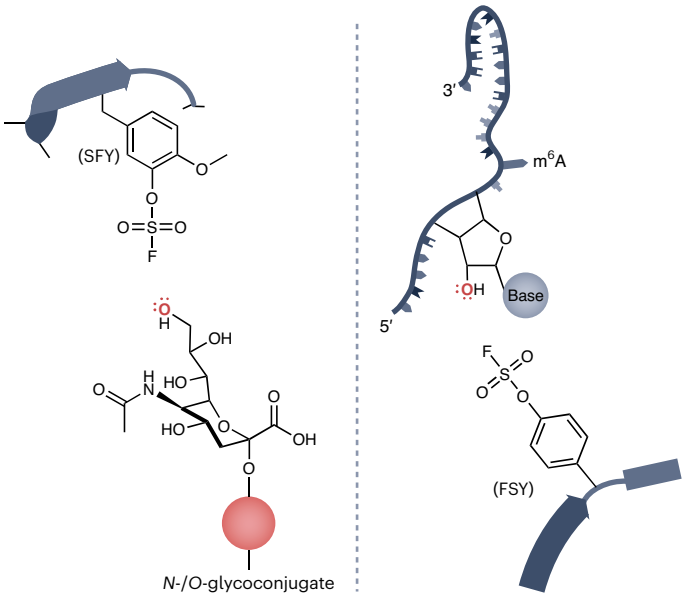

Fig. 1 | Overview of genetically encoded SuFEx residues in proximity to biological hydroxyl groups. Proximity induces nucleophilic attack by the hydroxyl to remove the fluorine, resulting in a covalent crosslink between the biopolymers. Left: SFY-modified protein in complex with a target carbohydrate. Right: FSY-modified protein in complex with target RNA.

The volume and depth of data accumulated using these methods has produced a great deal of biological insight. However, though potent tools, approaches relying on UV crosslinking have significant limitations, because it has well known biases, at both the RNA and protein levels; that is, some nucleotides and residues are much more easily excited by UV light than others, leading to limited reactivity despite their ability to bind[8], and so the level of UV crosslinking does not always equate to the amount or strength of interaction. Furthermore, sub-300-nm wavelengths induce substantial chemical alterations to RNA, including highly damaging mutations to nucleotides; the use of s^4U, while reducing unwanted side-reactions and increasing the crosslinking efficiency, limits the scope of discoverable interactions to those with a uridine nucleoside. Additionally, UV crosslinking generates highly reactive radicals that form covalent bonds with many proximate protein residues, obscuring the residues that are responsible for the interaction(s) with residues or nucleotides that are simply nearby. Glycans possess largely unreactive hydroxyl functional groups, requiring the deployment of metabolic reporters that carry orthogonal, photoactivatable functional groups such as azides and diazirines. Metabolic reporters need to be designed such that they can be used by the endogenous processes that incorporate individual sugars into glycan polymers[9] — no meagre feat. Indeed, such design strategies required the development of the diazirine-bearing carbohydrate monomers, which can suffer

from low rates of incorporation and a limited range of sugars that can be replaced[5]. Furthermore, because metabolic reporters, as well as s^4U, are reliant on the endogenous cellular machinery, their placement is necessarily random, precluding the option of targeting particular interactions with biopolymers of interest.

The strategies developed by Lei Wang and co-workers complement current crosslinking strategies: namely they are targeted, unbiased towards nucleotide or carbohydrate identity, and proximity-induced crosslinking of proteins to RNA and glycans. Their approach, depicted in Fig. 1, utilizes a set of genetically encoded unnatural amino acids (Uaa) bearing an aryl sulfonyl fluoride (FSY or SFY, depending on the chemical nature of the interacting partner). The Uaa was designed to employ the SuFEx reaction, which takes advantage of the biologically poorly reactive hydroxyl groups of carbohydrates and the 2′ OH groups of RNAs as nucleophiles to directly crosslink the specific protein residues to their interaction partner(s) without any apparent bias towards nucleotide identity — although the general applicability still remains to be determined for the varieties of carbohydrate monomers[10]. The strength of this approach lies in its specificity, in other words, a researcher studying a single or a handful of related proteins can identify their native RNA or glycan interacting partners at monomer (residue, nucleotide or carbohydrate)-level resolution. In their article, Sun et al.[1] applied this method to identify sites of m^6A modifications in mRNAs and achieved a resolution of 2–3 nucleotides, an order of magnitude more precise than m^6A-seq approaches utilizing antibodies, and could potentially provide a quantitative measurement of m^6A levels.

Stepping into the realm of potential clinical applications, in their article, Li et al.[2] examined the impact of irreversible covalent bonding of glycan binding proteins to their in trans partners in immune recognition of cancer. Using Siglec-7v — a well-characterized glycan binding protein — as a proof-of-principle, the team observed that the crosslink generated by the sulfonyl fluoride Uaa enhanced the killing potential of natural killer cells. This was fascinating since cancers employ glycans as a shield to evade the innate immune system — meaning that this strategy of site-specific clickable Uaa incorporation may have significant therapeutic potential.

Two potential limitations of this method are that it requires knowledge of the protein structure in complex with an interacting partner in order to practicably select the site(s) of Uaa incorporation. Although many such structures exist, the list is far from comprehensive. Additionally, while many RNA-binding proteins possess structured RNA-binding domains, many others bind RNA using intrinsically unstructured regions, which will be less amenable to this approach. In studies where the Uaa-modified protein(s) is continuously expressed, which may expand the scope and amount of crosslinked material, a second limitation could result from possible perturbation to the cell and the process or pathway under investigation by the Uaa incorporation. In these contexts, the amount of Uaa-modified protein expressed should be carefully evaluated to ensure that the effects of irreversible crosslinking in vivo do not alter the native biology appreciably.

The methods discussed here are likely to have broad applications in future biochemical and biological studies of RNA– and carbohydrate–protein interactions. In particular, this would complement studies that utilize deaminase fusions to study RNA–protein interactions and provide finer detail of the biochemical interactions that underpin the process in question, studies interested in examining protein isoform differences in biological interactions, and discovering and characterizing protein interactions with RNAs lacking poly-A tails. Furthermore, these approaches would probably merge well with other methods such as RNA structure probing, where different mutation signatures could be observed for the crosslink generated by the interaction and the labelled nucleotides. For glycobiology, employing this strategy would not only overcome the difficulties inherent in designing and metabolizing chemically modified sugars, but potentially enable the identification of the glycan patterns recognized by specific glycan binding proteins in vivo. With these methods, we expect to see an expansion and deepening of our knowledge of both RNA biology and glycobiology.

Christopher P. Watkins[1,2] & Ryan A. Flynn[1,2,3] ✉
[1]Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, MA, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [3]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA.
✉e-mail: ryan.flynn@childrens.harvard.edu

Published online: 6 January 2023

References

1. Sun, W. et al. *Nat. Chem.* https://doi.org/10.1038/s41557-022-01038-4 (2022).
2. Li, S., Wang, N., Yu, B., Sun, W. & Wang, L. *Nat. Chem.* https://doi.org/10.1038/s41557-022-01059-z (2022).
3. Hafner, M. et al. *Cell* **141**, 129–141 (2010).
4. Han, S., Collins, B. E., Bengtson, P. & Paulson, J. C. *Nat. Chem. Biol.* **1**, 93–97 (2005).
5. Tanaka, Y. & Kohler, J. J. *J. Am. Chem. Soc.* **130**, 3278–3279 (2008).
6. Castello, A. A. et al. *Cell* **149**, 1393–1406 (2012).
7. Ramya, T. N. C. et al. *Mol. Cell. Proteomics* **9**, 1339–1351 (2010).
8. Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. *Wiley Interdiscip. Rev. RNA* **9**, e1436 (2018).
9. Saxon, E. & Bertozzi, C. R. *Science* **287**, 2007–2010 (2000).
10. Barrow, A. S. et al. *Chem. Soc. Rev.* **48**, 4731–4758 (2019).

Competing interests

The authors declare no competing interests.