

# Chemical-Guided SHAPE Sequencing (cgSHAPE-seq) Informs the Binding Site of RNA-Degrading Chimeras Targeting SARS-CoV-2 5' Untranslated Region

Corresponding Author: Dr Jingxin Wang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Tang and co-authors tested the binding affinities of several previously synthesized coumarin derivatives with SARS-CoV-2 5'UTR. They then further synthesized FAI-C30 for performing RNA structure probing. They replaced the FAI with RLR and used C64/C47/C48/C65 to cleave the virus. Overall, the methods are not very novel. The application of using this FAI-C30 is not clear. The biological impact is also not clear.

1. C30 binding remodels the SARS-CoV-2 5'UTR structure. Thus, FAI modification is stronger. The FAI probing alone could provide structural changes with and without C30. The application of C30-FAI is not clear. Is it for all the C30 binding sites across transcriptome?
2. The connection between RLR and FAI is not clear. Why use C64/65/47/48? Why not use C30? It feels like two separate studies.
3. LASER-MAP from Robert Spitale's group is very similar to C30-FAI. Did the authors compare C30-FAI with LASER-MAP? What are the advantages of C30-FAI?
4. The authors should explore the biological impacts of applying C30-FAI and C64/65-RLR.

Reviewer #2

(Remarks to the Author)

In this manuscript, Tang, Hegde, et al. describe the development of chemical-guided SHAPE-seq (cgSHAPE-seq) for identifying ligand binding sites within RNA. In cgSHAPE-seq, an acylating moiety is attached to a ligand so that it will preferentially react with flexible nucleotides that are proximal to the ligand binding site. Modified nucleotides are then detected by mutational profiling and high-throughput DNA sequencing. After using cgSHAPE-seq to identify the binding site of a coumarin derivative (C30) within SL5 of the SARS-CoV-2 genome, the authors used C30 as the basis of a RIBOTAC, which efficiently degraded SARS-CoV-2 RNA and inhibited SARS-CoV-2 virus. The authors' method is a creative use of SHAPE chemistry and will likely be useful for the development and characterization of RNA-targeting small molecules.

Comments:

1. Was there a specific reason for using ProtoScript II reverse transcriptase for error-prone reverse transcription rather than SuperScript II (which was used in the original SHAPE-MaP procedure)? ProtoScript II clearly worked, but I am curious why it was chosen.
2. I thought the first paragraph of the discussion describing the limitations of cgSHAPE was excellent. However, I'm not sure that the denatured RNA control in Extended Data Fig. 2 supports the conclusion that "FAI-N3 has a higher reactivity bias towards G and against U". It's possible that there are context-dependent effects on the detection of adducts using mutational profiling, which is what the denatured control usually controls for. i.e., even if all nucleotides were equally modified, the efficiency of adduct detection might not be the same in all sequence contexts.
3. I think that it might be useful to expand the discussion by adding some guidance for performing cgSHAPE-seq experiments. For example, would it be useful to perform standard SHAPE probing alongside cgSHAPE-seq to determine the

full set of nucleotides that are SHAPE reactive in the folded RNA? How should functional assays be used to validate a putative ligand binding site following cgSHAPE-seq analysis? What are the design considerations for appending an acylating moiety to a ligand (location, linker length, etc.).

4. It be useful to mention that G174 was detected as highly reactive by NAI probing (Manfredonia et al, 2020).

5. Is it also possible that the slightly elevated mutations rates from A131 to G149 (and elsewhere) could be due to transient contacts with SL5 that bring the reactive nucleotides into close proximity with the probe? Although not essential, it might be interesting to try cgSHAPE-seq experiments with other sub-fragments of the SARS-CoV2 5' UTR that do not contain SL5 to see if these signatures persist.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have made lots of efforts in revising the manuscript. The new figures were significantly improved.

Further comments are as follows:

1. Rather than a novel method, it feels like a combination of three methods together. The authors previously developed the coumarin derivatives that could bind to RNAs. Then, the authors performed SHAPE-MaP following the C30 binding to determine the binding site for cooperating with the RLR. The authors could strengthen the novelty of the method in their manuscript.
2. The better performance of anti-viral RIBOTAC is very nice. Still, it may also be a trade-off of the specificity where the general binding property of coumarin derivatives to RNAs could cleave other SLs since the structural changes were limited to individual nucleotides such as G174. Some controls and discussions are necessary. It is also noticeable that the mutation rate differential profiles of C30-FAI have relatively high changes on other SLs.
3. The authors emphasize the cgSHAPE-seq is a new generalizable chemical probing method. However, the method seems very specific for SARS-CoV-2 5' UTR. It is very nice that the authors provide many details regarding the experimental procedure. However, it is unclear how to use cgSHAPE-seq to design anti-viral RIBOTAC on an RNA of interest. If others start using cgSHAPE-seq, do they need to test all the coumarin derivatives on the RNAs of interest? Or they could start from the C30-FAI probing on their RNAs of interest. Some extension on how to generalize the methods is critical.

Reviewer #2

(Remarks to the Author)

The authors have addressed all of my comments.

Reviewer #3

(Remarks to the Author)

The authors describe enhanced covalent modification of the 2'-OH of RNA using a chemical guided SHAPE strategy. The authors tether a non-covalent ligand that improves RNA modification via proximity labeling. This technique improved the RNA-binding ability of cgSHAPE compounds compared to the established SHAPE chemistry. The method and results are interesting, but I have a major concern that the authors performed all experiments with a cgSHAPE compound this is well below established purity standards (>95% is common).

The authors provide evidence for covalent modification of RNA, but the use of an impure probe substantially reduces confidence in results and the subsequent claims. This chemistry will likely be complementary to SHAPE, but due to the impurity of the cgSHAPE probe there is a chance that the method may produce irreproducible results when disclosed to the community. I suggest the authors purify C30-FAI (>95% is recommended) and repeat studies to verify that the results are exclusively from this probe and not a 3:1 mixture of C30-FAI:C30-FCA. Until this is corrected, I am not able to support this manuscript for publication.

Additional comments:

1. The chirality for 2-5A is incorrect throughout the manuscript and in the SI. For example, on page 14 of the SI, the stereo- and regiochemistry of 2-5A-N3 does not match the product, C48. This is not impossible given the complexity of chemical reactions, but it is extremely unlikely in a [3+2] azide-alkyne cycloaddition reaction. The stereo- and regiochemistry of 2-5A and all following products should be drawn accurately throughout the manuscript and SI.
2. The trifluoromethyl moiety is easily characterized by <sup>19</sup>F NMR and via the quartet with a J = 250 Hz in the <sup>13</sup>C NMR. Neither of these are in the SI.
3. High-resolution mass spectrometry data should be included for all compound characterization.
4. Where is the NMR characterization of C30-FAI? Does C30-FAI hydrolyze under HPLC conditions? This would be answered with <sup>1</sup>H NMR data. This is the cgSHAPE probe used to establish the novelty of this work. A maximum of 75% purity is concerning. Was the potency and all other covalent binding-related data adjusted based on 75% purity? How are the authors certain that the 25% C30-FCA in each experiment does not play a role in covalent binding?
5. The HPLC data is poorly integrated. For example, the integration of C30-NM suggests ">99%", but a large portion of the

peak is not integrated and the shoulder to the right, which is likely an impurity, is barely integrated. The authors should properly integrate HPLC data to provide readers with an accurate purity. In another example, C30-FAI has the two major peaks integrated, but there are 10 smaller peaks that are not integrated. Proper integration of this HPLC data will likely bring the purity of the cgSHAPE probe closer to 70%.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed the comments.

Reviewer #3

(Remarks to the Author)

While the chemistry isn't particularly novel, the authors are able to use SHAPE chemistry as an irreversible covalent handle, enabling C30-FAI to exhibit nice selectivity for G174 in SL5 which is important proof-of-concept for selective covalent modification of RNA using their method. However, my original concern and critique remains. I don't believe 91% purity for C30-FAI is acceptable. The fact that there's a lot of impurity in C30-FAI makes all of the biological conclusions ambiguous.

In addition, while the HPLC analysis is correct, there is still a lot of imidazole in the <sup>1</sup>H NMR that they did not integrate, so I'm unable to calculate <sup>1</sup>H NMR purity. In this case, <sup>1</sup>H NMR purity is much more important than HPLC analysis. Imidazole is not a good chromophore and will not absorb nearly as well as the C30-FAI molecule that has two strong chromophores, one of which is coumarin. The imidazole peaks in the <sup>1</sup>H NMR of C30-FAI should be integrated to more accurately assess purity.

Given the ongoing issue with C30-FAI purity, I am unable to support publication of this manuscript in the current form.

Reviewer #4

(Remarks to the Author)

In this manuscript, Tang et al. use a SHAPE technique to map the binding site of RNA guiding chimeras derived from a coumarin scaffold. These compounds bind to the SL5 portion of the SARS-CoV-2 genome, and degrade the genome, ostensibly by recruiting RNase L. This is an interesting paper that leverages multiple cutting edge techniques and presents a series of interesting probe molecules/analyses. Further, the manuscript has been through one round of review already, and I have been asked specifically to comment on the issue of probe purity and the impact of that purity on the results presented in the paper. Comments:

1. The mutation rate (or more accurately, delta-delta-mutation rate) at G174 is ~2% in Figure 2. This is not particularly high for an acylimidazole-based SHAPE reagent. But, the authors discuss this in the paper and present solid dose-dependent SHAPE assays to support their reasonable argument that the increase in modifications specifically at this locus are a compelling argument for G174 as the binding site, arguing that the specific increase at this locus is driven by recognition of the compound.
2. The predominant contaminant observed in NMR is either imidazole, as discussed by the authors and reviewer 3, or possibly contaminating amounts of CDI itself. In both <sup>1</sup>H and <sup>13</sup>C NMR spectra provided in the revised manuscript, "imidazole" peaks are broadened (likely a consequence of protonation/partial protonation). Given the size of these peaks I am concerned, along with Reviewer 3, that the actual purity of C30-FAI is far lower than the 91% HPLC purity reported by the authors. Further, since the peaks are broadened, they are shifted slightly from literature values for peaks for either CDI or imidazole itself so an unambiguous assignment/quantification is difficult.
3. I am concerned that if the contaminant is CDI, itself a reactive acylating agent, it could be the source of some of the altered mutation rate that is observed. Presumably, FAI-N3 does not contain any contaminants (spectra for FAI-N3 are not provided). However, a control assessing the effects of CDI itself is also not provided. Previously, and in unpublished work, our laboratory has seen contamination issues similar to this interfere with SHAPE assays.
4. Hyper-reactivity/increased mutation rate of G174 is observed with multiple diazirines and a chlorambucil mustard. This is somewhat surprising given the different nature of crosslinkers (likely to react with bases) and SHAPE reagents (likely to react with 2'-OH). Is G174 just a hyper-reactive site? I am not sure this can be ruled out based on data shown.

Because of these points, it is critical to use a highly pure sample of C30-FAI for the analyses.

Version 3:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

The authors developed a more efficient route to produce C30-FAI and confirmed the acyl imidazole C30-FAI is indeed the active agent that binds RNA, therefore, I approve this manuscript for publication.

**Open Access** This Peer Review File 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.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File 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 <https://creativecommons.org/licenses/by/4.0/>

Reviewers' comments:

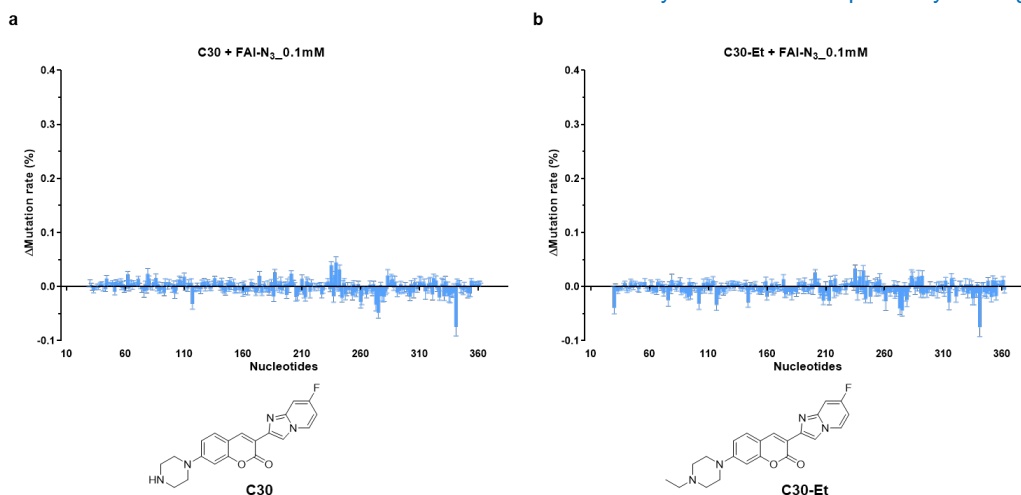
Reviewer #1 (Remarks to the Author):

Tang and co-authors tested the binding affinities of several previously synthesized coumarin derivatives with SARS-CoV-2 5'UTR. They then further synthesized FAI-C30 for performing RNA structure probing. They replaced the FAI with RLR and used C64/C47/C48/C65 to cleave the virus. Overall, the methods are not very novel. The application of using this FAI-C30 is not clear. The biological impact is also not clear.

1. C30 binding remodels the SARS-CoV-2 5'UTR structure. Thus, FAI modification is stronger. The FAI probing alone could provide structural changes with and without C30. The application of C30-FAI is not clear. Is it for all the C30 binding sites across transcriptome?

The reviewer commented that “C30 binding remodels the SARS-CoV-2 5'UTR structure. Thus, FAI modification is stronger. The FAI probing alone could provide structural changes with and without C30”.

We thank the reviewer's consideration. The reviewer might misinterpret our target identification work as simply combining the treatment of C30 and FAI, and attribute the SHAPE signal increase to C30-induced SL5 RNA structure remodel. In fact, our target identification approach involved a novel chemical probe that covalently linked C30 with FAI, instead of simple mixing. We have compelling evidence that demonstrated the SHAPE signal increase was due to significantly enhanced proximity reactivity instead of structure remodeling (Fig. 2d). Although the target is different from the current study, we previously demonstrated that when SHAPE agents (e.g., NAI) and small molecules are used simultaneously as free molecules, SHAPE activity change in the presence of small molecules cannot directly indicate the binding sites<sup>1</sup> (PMID: 29712837). In order to further address the reviewer's concern, we co-treated C30 and FAI at 0.1 mM (an optimal cgSHAPE concentration from our current study) under the same condition as we did for cgSHAPE, and we only observed background noise. To eliminate the possibility that C30 would react with FAI directly through the piperazine, we also prepared C30-Et, where the nucleophilic reaction site was blocked. The SHAPE profile is very similar to the results obtained from C30 + FAI samples and only the background noise was observed (see figure below). The results are consistent with our expectation because 0.1 mM is far below the working concentration of FAI. The result also verifies the modification of G174 by C30-FAI is C30 proximity/binding-guided.



The reviewer also commented that “The application of C30-FAI is not clear. Is it for all the C30 binding sites across transcriptome?”.

In this study, C30-FAI is a chemical probe, and its application is to locate the binding site of C30 within the targeted SARS-CoV-2 5' UTR. This method can also be applied to identify the binding sites of other RNA-binding small molecules.

2. The connection between RLR and FAI is not clear. Why use C64/65/47/48? Why not use C30? It feels like two separate studies.

An RNase L recruiter (RLR) and an RNA binding moiety are both necessary components for RNA-degrading chimeras. C30 is the RNA-binding ligand and cannot be used as a degrader alone. Recent studies<sup>2</sup> (i.e., PMID: 37225982) demonstrated that most RNA binders alone do not have biological consequences. On the other hand,

C64/65/47/48 are conjugates of C30 with RLRs, i.e., functional RNA degrading chimeras. As we highlighted in the title of the manuscript, cgSHAPE-seq (using C30-FAI) informs the binding site of RNA-degrading chimeras (C30-RLR).

3. LASER-MAP from Robert Spitale's group is very similar to C30-FAI. Did the authors compare C30-FAI with LASER-MAP? What are the advantages of C30-FAI?

We thank the reviewer's suggestion. However, we believe that the LASER-MaP and C30-FAI are for completely different purposes. Although they both modify RNAs and involve mutational profiling, LASER-MaP is primarily used for RNA structure elucidation with selectivity for base-paired purines A and G which are solvent accessible. cgSHAPE, on the other hand, is used for identification of RNA binding sites, based on SHAPE chemistry which acylates the 2'-OH of ribose close to the bound nucleotides.

To address the reviewer's concern about comparison study with other approaches, we compared cgSHAPE with two other methods, PEARL-seq and Chem-CLIP. For PEARL-seq two photocrosslinking probes were synthesized, and one nitrogen mustard probe was synthesized for Chem-CLIP. We found that only one of the PEARL-seq probes successfully identified G174, and cgSHAPE-seq is superior to PEARL-seq in signal-to-noise ratio. The results and discussions were added in the new section "Comparison of cgSHAPE-seq and other sequencing-based RNA ligand localization methods". The technical details were documented in SI.

4. The authors should explore the biological impacts of applying C30-FAI and C64/65-RLR.

We used C30-FAI to identify the binding site of C30 within the RNA target of interest in the context of total RNA. We have shown that C64/65 can cause the degradation of SARS-CoV-2 5' UTR RNA in an in vitro assay, a cell reporter assay, and a live virus infection assay in A549 cells (Fig. 4).

Reviewer #2 (Remarks to the Author):

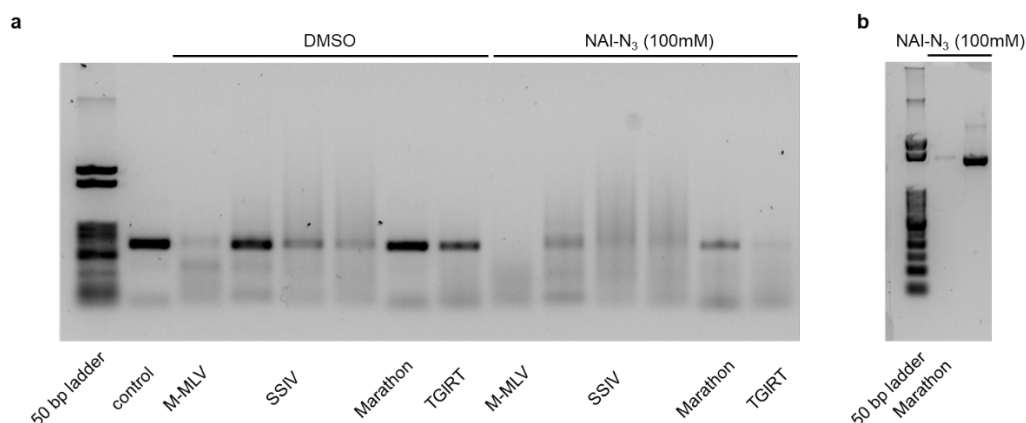
In this manuscript, Tang, Hegde, et al. describe the development of chemical-guided SHAPE-seq (cgSHAPE-seq) for identifying ligand binding sites within RNA. In cgSHAPE-seq, an acylating moiety is attached to a ligand so that it will preferentially react with flexible nucleotides that are proximal to the ligand binding site. Modified nucleotides are then detected by mutational profiling and high-throughput DNA sequencing. After using cgSHAPE-seq to identify the binding site of a coumarin derivative (C30) within SL5 of the SARS-CoV-2 genome, the authors used C30 as the basis of a RIBOTAC, which efficiently degraded SARS-CoV-2 RNA and inhibited SARS-CoV-2 virus. The authors' method is a creative use of SHAPE chemistry and will likely be useful for the development and characterization of RNA-targeting small molecules.

We thank the reviewer's positive response and constructive suggestions.

Comments:

1. Was there a specific reason for using ProtoScript II reverse transcriptase for error-prone reverse transcription rather than SuperScript II (which was used in the original SHAPE-MaP procedure)? ProtoScript II clearly worked, but I am curious why it was chosen.

We compared several commercially reverse transcriptases in the presence of  $Mn^{2+}$  for mutational profiling. Specifically, we examined their abilities to read through an RNA sequence which was extensively modified using 100 mM NAI- $N_3$  under denaturing conditions. As shown in Fig. S2a, ProtoScript II and Marathon are the only two reverse transcriptases that provide the correct amplicon size. We finally chose ProtoScript II because it performed slightly better than Marathon with longer RNA templates (> 800 nt) (Fig. S2b). We have not used SSII but included Promega M-MLV and Thermo Fisher SSIV in this survey.



**Fig. S2. Comparison of reverse transcriptases for mutational profiling.** RNA was diluted in denaturing buffer (90% formamide, 5 mM EDTA), heated to 80 °C for 2 minutes and snap-cooled on ice. DMSO or NAI-N<sub>3</sub> was added and mixed well. The solution was heated at 80 °C for 5 minutes and cooled on ice. RNA was recovered using Qiagen RNeasy kit. RT reaction buffer: 50 mM Tris-HCl (pH 7.4), 75 mM KCl, 10 mM DTT, 3 mM MnCl<sub>2</sub>. Manual suggested optimal reaction temperature of each enzyme was used (42 °C incubation 1 h for M-MLV, ProtoScript II and Marathon, 50 °C incubation 1 h for SSIV and Maxima H minus). For TGIRT, reaction buffer: 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 10 mM DTT, 1.5 mM MgCl<sub>2</sub>. 50 °C incubation 1 h.

2. I thought the first paragraph of the discussion describing the limitations of cgSHAPE was excellent. However, I'm not sure that the denatured RNA control in Extended Data Fig. 2 supports the conclusion that "FAI-N<sub>3</sub> has a higher reactivity bias towards G and against U". It's possible that there are context-dependent effects on the detection of adducts using mutational profiling, which is what the denatured control usually controls for. i.e., even if all nucleotides were equally modified, the efficiency of adduct detection might not be the same in all sequence contexts.

We agree with the reviewer that the sequence context might change the nucleotide modification bias of SHAPE and one cannot make a conclusion based on results from one sequence. There has been some excellent research done by the Weeks lab and others to investigate the SHAPE selectivity<sup>3</sup> (PMID: 31117385). We deleted the Extended Data Fig. 2 and related discussions in the revised manuscript.

3. I think that it might be useful to expand the discussion by adding some guidance for performing cgSHAPE-seq experiments. For example, would it be useful to perform standard SHAPE probing alongside cgSHAPE-seq to determine the full set of nucleotides that are SHAPE reactive in the folded RNA? How should functional assays be used to validate a putative ligand binding site following cgSHAPE-seq analysis? What are the design considerations for appending an acylating moiety to a ligand (location, linker length, etc.).

We thank the reviewer for constructive suggestions. We added more discussions on the guidance of cgSHAPE-seq experiments to the revised manuscript:

"In cgSHAPE-seq, it is critical to employ a concentration at which the predominant source of SHAPE activity is the proximity-enhanced activity (Fig. 2d). Specifically, a standard SHAPE probing should be performed parallelly with an identical probe concentration as that of the cgSHAPE probe to determine the full set of SHAPE reactivity in the folded RNA. At the optimal cgSHAPE probe concentration (i.e., 0.02–0.1 mM in this study), the standard SHAPE activity should be minimal (typically < 0.1% as demonstrated in Fig. 2d)."

"In addition, the putative binding sites determined by cgSHAPE-seq require extensive validation. In the current study, we used mutagenesis and in vitro binding assay to validate the binding site (Fig. 3). We envision that reverse genetics with a mutated binding site coupled with relevant functional assays can also be used for validation purposes if the RNA binder has well-defined biological consequences."

"Multiple crucial factors merit attention when designing cgSHAPE probes: (1) the cgSHAPE probe should be considerably stable in water solution to allow sufficient target engagement and to avoid hydrolysis by moisture in the air during the experiment preparation steps. In this current study, we chose FAI moiety with a reported half-life of ~73 min. We attempted to use C30-NAI conjugate but observed that the probe is considerably less stable than C30-FAI (data not shown). (2) The conjugation site of acylation moiety should be solvent accessible and not impede ligand binding. In this current study, the C30-FAI precursor and C30-D probe both showed significant binding to the RNA

target, albeit 3–7 fold less binding affinity (Fig. S1 and S3). (3) The functional group should be compatible with acylating moiety.”

4. It be useful to mention that G174 was detected as highly reactive by NAI probing (Manfredonia et al, 2020).

We added this statement to the revised manuscript:

It is worth noting that a previous study showed that G174 was highly reactive in a regular SHAPE experiment using another acylation agent, NAI, at 100 mM. This concentration is at least 1,000 times higher than the optimal cgSHAPE probe concentration determined in our study (0.02–0.1 mM), and therefore, the cgSHAPE signal is unlikely attributed to the structure-induced SHAPE background (also see Fig. 2d).

5. Is it also possible that the slightly elevated mutations rates from A131 to G149 (and elsewhere) could be due to transient contacts with SL5 that bring the reactive nucleotides into close proximity with the probe? Although not essential, it might be interesting to try cgSHAPE-seq experiments with other sub-fragments of the SARS-CoV2 5' UTR that do not contain SL5 to see if these signatures persist.

We agree with the reviewer that the slightly elevated mutation rates from A131 to G149 (and elsewhere) may be due to the transient contacts with SL5. However, to our knowledge there is no gold standard experimental method to profile transient small molecule-RNA interactions, making it difficult to benchmark or validate. We are currently collaborating with computational biologist, Dr. Yinglong Miao to understand transient interactions for other RNA targets. For this reason, we decided not to include more experiments to profile transient interactions in the current study.

#### References

1. Wang, J., Schultz, P. G. & Johnson, K. A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4604–E4612 (2018).
2. Tong, Y. *et al.* Programming inactive RNA-binding small molecules into bioactive degraders. *Nat.* **2023** 6187963 **618**, 169–179 (2023).
3. Busan, S., Weidmann, C. A., Sengupta, A. & Weeks, K. M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **58**, 2655–2664 (2019).



Reviewer #1 (Remarks to the Author):

The authors have made lots of efforts in revising the manuscript. The new figures were significantly improved.

Further comments are as follows:

1. Rather than a novel method, it feels like a combination of three methods together. The authors previously developed the coumarin derivatives that could bind to RNAs. Then, the authors performed SHAPE-MaP following the C30 binding to determine the binding site for cooperating with the RLR. The authors could strengthen the novelty of the method in their manuscript.

We appreciate the reviewer's insightful feedback and the opportunity to clarify the novel contributions of our study:

1. C34 exhibits the highest known binding affinity to the SARS-CoV-2 5' UTR, setting a new standard for synthetic ligands in the RIBOTAC context.
2. We introduce cgSHAPE-seq, a pioneering approach that not only locates where RNA ligands bind but can be transformed into RIBOTACs by replacing the acylating moiety into RNase recruiter moiety.
3. To our knowledge, our RIBOTAC C64 is the first cell-permeable RNA-degrading chimera that is active in live virus infected cells.

2. The better performance of anti-viral RIBOTAC is very nice. Still, it may also be a trade-off of the specificity where the general binding property of coumarin derivatives to RNAs could cleave other SLs since the structural changes were limited to individual nucleotides such as G174. Some controls and discussions are necessary. It is also noticeable that the mutation rate differential profiles of C30-FAI have relatively high changes on other SLs.

We acknowledge the reviewer's remarks regarding the specificity trade-off inherent in anti-viral RIBOTAC application. Specificity/selectivity is indeed an important concern in almost all RNA-targeting strategies. We hypothesized that relatively high changes in other SLs and flexible region may be due to transient or non-specific interactions. We have addressed this in the Discussion section, noting that: "In our cgSHAPE-seq result, apart from G174, we also observed a cluster of nucleotides from A131 to G149 showing slightly higher mutation rates than others (Fig. 2b and 2c). This can be potentially caused by transient or nonspecific binding of C30 with flexible sequence."

3. The authors emphasize the cgSHAPE-seq is a new generalizable chemical probing method. However, the method seems very specific for SARS-CoV-2 5' UTR. It is very nice that the authors provide many details regarding the experimental procedure. However, it is unclear how to use cgSHAPE-seq to design anti-viral RIBOTAC on an RNA of interest. If others start using cgSHAPE-seq, do they need to test all the coumarin derivatives on the RNAs of interest? Or they could start from the C30-FAI probing on their RNAs of interest. Some extension on how to generalize the methods is critical.

cgSHAPE-seq is indeed a versatile method intended for identifying the binding sites of small molecules on their target RNAs. While our coumarin derivatives have demonstrated a high binding affinity specifically to the SARS-CoV-2 5' UTR, it is important to note that different RNAs may require different ligands. For researchers looking to design anti-viral RIBOTACs against other RNA targets, the recommended approach would involve an initial screening phase to discover appropriate ligands that

bind effectively to the RNA of interest. Following this, researchers can employ the cgSHAPE-seq method, utilizing custom-designed X-FAI probes to pinpoint precise binding sites. This process will yield critical data that can guide the design of RIBOTAC molecules for a wide array of RNA targets.

In addition, we have provided a guidance in design of cgSHAPE probe in Discussion:” Multiple crucial factors merit attention when designing cgSHAPE probes: (1) the cgSHAPE probe should be considerably stable in water solution to allow sufficient target engagement and to avoid hydrolysis by moisture in the air during the experiment preparation steps. In this current study, we chose FAI moiety with a reported half-life of ~73 min. We attempted to use C30-NAI conjugate but observed that the probe is considerably less stable than C30-FAI. (2) The conjugation site of acylation moiety should be solvent accessible and not impede ligand binding. In this current study, the C30-FAI precursor for cgSHAPE and C30-D probe for PEARL-seq both showed significant binding to the RNA target, albeit 3–7 fold less binding affinity (Supplementary Information, Fig. S1 and S3). (3) The functional group should be compatible with acylating moiety. For example, the FAI-based probes used in this report would not be compatible with nucleophilic RNA ligands due to self-reaction. cgSHAPE-seq potentially has a limitation with bias towards certain types of small molecule-RNA interactions. As shown in conventional SHAPE, FAI moiety has a higher reactivity towards unpaired RNA nucleotides. Although most of the reported RNA ligands target the unpaired region, cgSHAPE-seq may be less reactive for ligands that bind to the double-stranded RNA grooves.”

Reviewer #2 (Remarks to the Author):

The authors have addressed all of my comments.  
We thank the reviewer’s supportive comments.

Reviewer #3 (Remarks to the Author):

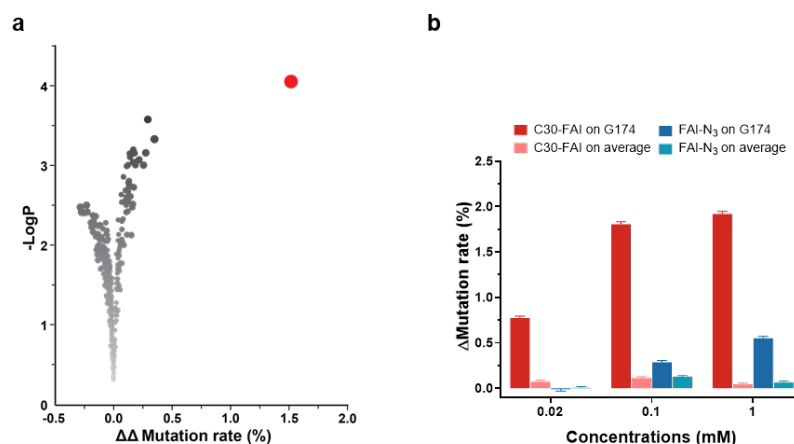
The authors describe enhanced covalent modification of the 2’-OH of RNA using a chemical guided SHAPE strategy. The authors tether a non-covalent ligand that improves RNA modification via proximity labeling. This technique improved the RNA-binding ability of cgSHAPE compounds compared to the established SHAPE chemistry. The method and results are interesting, but I have a major concern that the authors performed all experiments with a cgSHAPE compound this is well below established purity standards (>95% is common).

The authors provide evidence for covalent modification of RNA, but the use of an impure probe substantially reduces confidence in results and the subsequent claims. This chemistry will likely be complementary to SHAPE, but due to the impurity of the cgSHAPE probe there is a chance that the method may produce irreproducible results when disclosed to the community. I suggest the authors purify C30-FAI (>95% is recommended) and repeat studies to verify that the results are exclusively from this probe and not a 3:1 mixture of C30-FAI:C30-FCA. Until this is corrected, I am not able to support this manuscript for publication.

We appreciate the reviewer’s critical assessment of the C30-FAI purity in our study. In response to the reviewer’s concerns, we have refined our synthesis process and significantly enhance the purity of C30-FAI to 91%.

The cgSHAPE probe C30-FAI was synthesized by reacting C30-FCA CDI in DMSO. As the product is susceptible to hydrolysis and difficult to purify in regular chemistry labs, we utilized the crude product freshly for RNA modification without purification. Similar procedures have been reported for other SHAPE analogs with NAI or other acylating moieties (Nat. Chem. 15, 1374–1383 (2023)). We also would like to highlight that quick hydrolysis of FAI into FCA in RNA probing is one of the key advantages to avoid over-labeling (Nat. Protoc. 1, 1610–1616 (2006)). The original crude product of the probe contains 75% C30-FAI and 25% C30-FCA. To address the reviewer’s concern, we attempted different reaction conditions to improve the yield of C30-FAI. Ultimately, we optimized the reaction by further drying the starting materials (C30-FCA and CDI) and using anhydrous DMSO-<sub>d6</sub> as solvent. We also increased the amount of CDI to 1.5 equivalent of C30-FCA. A recent paper used 1.4 equivalent CDI to synthesize the RNA 2’-OH acylation agents. And the excess amount of CDI (would be finally hydrolyzed into imidazole) did not have negative impact in chemical probing (Nat. Chem. 15, 1374–1383 (2023)). These methodological improvements have successfully increased the purity of C30-FAI to 91%, as confirmed by HPLC analysis (details provided in the Supplementary Information).

We then used the purer C30-FAI in cgSHAPE. Although the sequencing depth in this repeat was lower than in our initial report due to unknown technical issues in sequencing, the results remained quite consistent, supporting the reproducibility of our findings. Based on these reproducible results, we decided to maintain the original figures in the manuscript.



**a**, Scatter plot of  $-\text{LogP}$  vs  $\Delta\Delta$ mutation rate at 0.1 mM ( $\Delta\Delta$ mutation rate [(C30-FAI – DMSO) – (FAI-N<sub>3</sub> – DMSO)] indicates the proximity-based differential acylation). **b**, Comparison of the  $\Delta$ mutation rates ([probe – DMSO]) of G174 and on average in RNAs treated with different concentrations of C30-FAI or FAI-N<sub>3</sub>. Three data points (C75, G216, G285) were removed as outliers as they had an abnormally high mutation rate (Z-score > 4.0) in DMSO-treated samples.

Additional comments:

1. The chirality for 2-5A is incorrect throughout the manuscript and in the SI. For example, on page 14 of the SI, the stereo- and regiochemistry of 2-5A-N3 does not match the product, C48. This is not impossible given the complexity of chemical reactions, but it is extremely unlikely in a [3+2] azide-alkyne cycloaddition reaction. The stereo- and regiochemistry of 2-5A and all following products should be drawn accurately throughout the manuscript and SI.

We express our gratitude to the reviewer for bringing this mistake to our attention. We have revised the structures of 2-5A and all related compounds (main text: Fig 4a, Extended Data Fig 6c. SI: synthetic schemes for C48 and C65).

2. The trifluoromethyl moiety is easily characterized by  $^{19}\text{F}$  NMR and via the quartet with a  $J = 250$  Hz in the  $^{13}\text{C}$  NMR. Neither of these are in the SI.

We re-measured the  $^{13}\text{C}$  NMR and added the  $^{19}\text{F}$  NMR for compound C36. We also added the  $^{19}\text{F}$  NMR for compounds C30 and C31. This information has been added to the SI.

$^{13}\text{C}$  NMR (126 MHz,  $\text{DMSO-}d_6$ )  $\delta$  159.3, 155.0, 152.7, 141.8, 139.6, 139.5, 130.0, 129.0, 126.1 (q,  $J = 33.1$  Hz), 123.5 (q,  $J = 270$  Hz), 114.1, 114.1, 113.6 (q,  $J = 5.7$  Hz), 112.2, 110.8, 107.9, 100.3, 54.9, 44.0.

$^{19}\text{F}$  NMR (471 MHz,  $\text{DMSO-}d_6$ )  $\delta$  -61.99.

For C30:

$^{19}\text{F}$  NMR (471 MHz,  $\text{DMSO-}d_6$ )  $\delta$  -112.80.

For C31:

$^{19}\text{F}$  NMR (471 MHz,  $\text{DMSO-}d_6$ )  $\delta$  -141.87.

3. High-resolution mass spectrometry data should be included for all compound characterization.

We measured the high-resolution mass spectrometry for all the final compounds and added the HRMS method and results in SI except for C48 and C65. C48 and C65 are 2-5A conjugates, and the signals are insufficient to conclude an accurate mass.

4. Where is the NMR characterization of C30-FAI? Does C30-FAI hydrolyze under HPLC conditions? This would be answered with  $^1\text{H}$  NMR data. This is the cgSHAPE probe used to establish the novelty of this work. A maximum of 75% purity is concerning. Was the potency and all other covalent binding-related data adjusted based on 75% purity? How are the authors certain that the 25% C30-FCA in each experiment does not play a role in covalent binding?

We thank the reviewer for bringing the lack of characterization to our attention. We added the  $^1\text{H}$  and  $^{13}\text{C}$  NMR for C30-FAI (see revised manuscript and SI). As the reviewer questioned, C30-FAI does hydrolyze under HPLC conditions, we analyzed the HPLC purity of C30-FAI by reacting the material with 11 equivalents of  $\beta$ -mercaptoethanol to form a stable adduct (see HPLC analysis of C30-FAI in SI).

C30-FCA is quickly formed by hydrolysis in aqueous solution during chemical probing of RNA. This has been viewed as a protection mechanism to avoid over-labeling (Nat. Protoc. 1, 1610–1616 (2006)). We have improved the purity of C30-FAI into 91% and adjusted the concentration of newly synthesized cgSHAPE based on this purity. Consistent results were obtained between cgSHAPE experiments with 75% and 91% probes.

5. The HPLC data is poorly integrated. For example, the integration of C30-NM suggests “>99%”, but a large portion of the peak is not integrated and the shoulder to the right, which is likely an impurity, is barely integrated. The authors should properly integrate HPLC data to provide readers with an accurate

purity. In another example, C30-FAI has the two major peaks integrated, but there are 10 smaller peaks that are not integrated. Proper integration of this HPLC data will likely bring the purity of the cgSHAPE probe closer to 70%.

As the reviewer suggested, we repeated the HPLC integration for C48 and C65 and revised the purity. We repeated HPLC analysis for C30-NM using a new column which gives better separation between the product and the impurity. For C30-FAI, please refer to our response to the comments above. We carefully performed the integration and determined the purity as 91%.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors have addressed the comments.

Reviewer #3 (Remarks to the Author):

While the chemistry isn't particularly novel, the authors are able to use SHAPE chemistry as an irreversible covalent handle, enabling C30-FAI to exhibit nice selectivity for G174 in SL5 which is important proof-of-concept for selective covalent modification of RNA using their method. However, my original concern and critique remains. I don't believe 91% purity for C30-FAI is acceptable. The fact that there's a lot of impurity in C30-FAI makes all of the biological conclusions ambiguous.

In addition, while the HPLC analysis is correct, there is still a lot of imidazole in the <sup>1</sup>H NMR that they did not integrate, so I'm unable to calculate <sup>1</sup>H NMR purity. In this case, <sup>1</sup>H NMR purity is much more important than HPLC analysis. Imidazole is not a good chromophore and will not absorb nearly as well as the C30-FAI molecule that has two strong chromophores, one of which is coumarin. The imidazole peaks in the <sup>1</sup>H NMR of C30-FAI should be integrated to more accurately assess purity.

Given the ongoing issue with C30-FAI purity, I am unable to support publication of this manuscript in the current form.

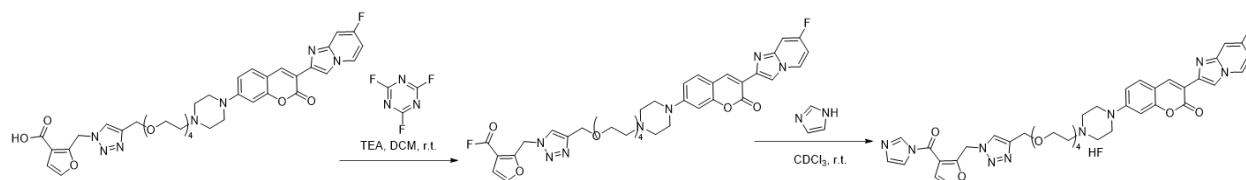
We appreciate the reviewer's detailed feedback regarding the purity of C30-FAI and its potential impact on our biological conclusions. We recognize the concern about imidazole contamination and its complications in the chemical probing process. In response, we have undertaken substantial efforts to improve the purity of C30-FAI.

After exploring various methods, we ultimately developed a new synthetic route designed to enhance the yield and purity of C30-FAI while avoiding the use of excess amount of CDI. Specifically, we transformed C30-FCA (carboxylic acid precursor) into its corresponding acid fluoride. This intermediate was then titrated with 1.0 equivalent of 1H-imidazole to obtain the final product as a pure salt. We have repeated this protocol multiple times to confirm the robustness of the reaction and product purity.

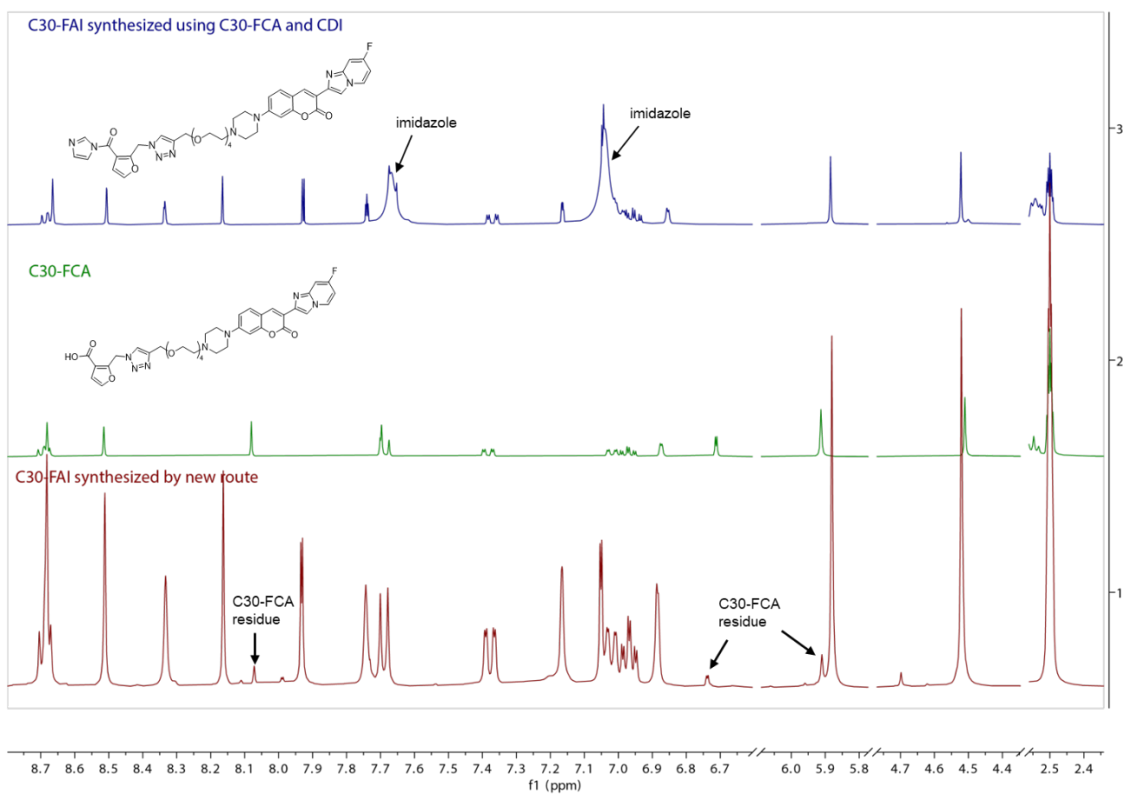
NMR characterization of the newly synthesized C30-FAI confirms the absence of imidazole contamination and a significantly improved NMR spectrum (see figure below), addressing the reviewer's concern. The <sup>1</sup>H NMR analysis indicates a ~91% purity for C30-FAI, with the

only minor impurity being 9% hydrolysis product C30-FCA, which is an expected byproduct due to the extensive characterization process. Importantly, this minor impurity should not interfere with the chemical probing process.

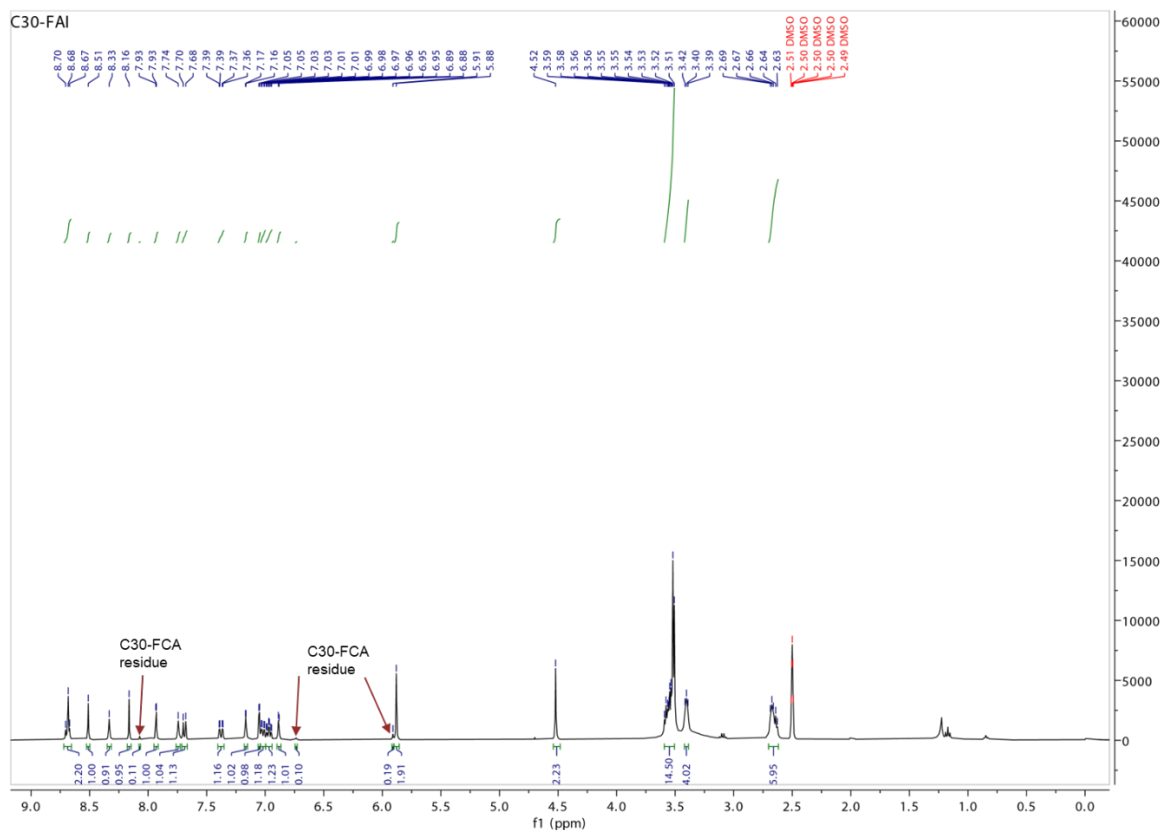
We believe that with a 91% NMR purity and the absence of imidazole (or CDI), the probe is suitable for RNA mutational profiling. We have conducted RNA mutational profiling using this new batch of C30-FAI and obtained results almost identical to those obtained from the previous batch. The updated figures reflecting these new sequencing results are included in the revised manuscript.



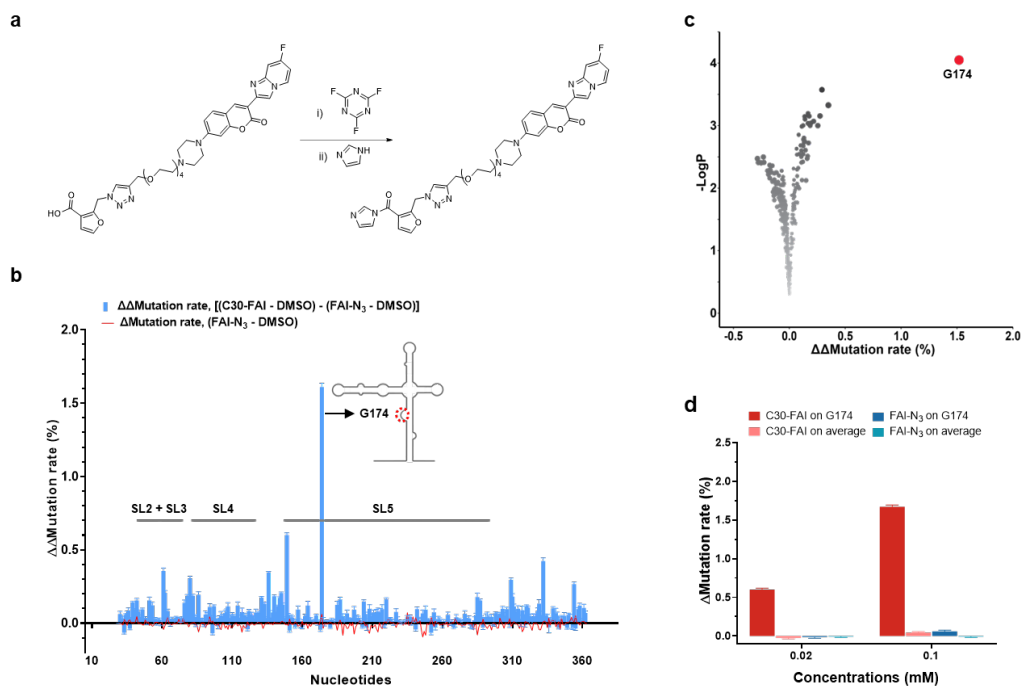
### New route to synthesize C30-FAI without using CDI



<sup>1</sup>H NMR alignment of C30-FCA (middle) and two batches of C30-FAI (new: bottom)



$^1\text{H}$  NMR integration of the new C30-FAI in DMSO- $d_6$ , the C30-FCA residue peaks were marked, the ratio of C30-FCA residue to C30-FAI is  $\sim 1:10$





Mutational profiling results obtained by using the new C30-FAI are almost identical to our previous results.

Reviewer #4 (Remarks to the Author):

In this manuscript, Tang et al. use a SHAPE technique to map the binding site of RNA guiding chimeras derived from a coumarin scaffold. These compounds bind to the SL5 portion of the SARS-CoV-2 genome, and degrade the genome, ostensibly by recruiting RNase L. This is an interesting paper that leverages multiple cutting-edge techniques and presents a series of interesting probe molecules/analyses. Further, the manuscript has been through one round of review already, and I have been asked specifically to comment on the issue of probe purity and the impact of that purity on the results presented in the paper. Comments:

1. The mutation rate (or more accurately, delta-delta-mutation rate) at G174 is ~2% in Figure 2. This is not particularly high for an acylimidazole-based SHAPE reagent. But, the authors discuss this in the paper and present solid dose-dependent SHAPE assays to support their reasonable argument that the increase in modifications specifically at this locus are a compelling argument for G174 as the binding site, arguing that the specific increase at this locus is driven by recognition of the compound.

We appreciate the reviewer's comment. Typically, in SHAPE experiments, the working concentration of SHAPE reagents is around 10 mM for in vitro assays (e.g., 1M7, NMIA) and 100 mM for in vivo assays (e.g., NAI), which can lead to high RNA modification and mutation rates. In our experiment, we used significantly lower concentrations of the cgSHAPE probe, 0.1 mM and 20  $\mu$ M, and still observed a notably high mutation rate at the binding site. In contrast, the SHAPE probe FAI-N<sub>3</sub>, when used at these low concentrations, nearly has negligible RNA modifications with almost undetectable mutation rates compared to the controls. We thus concluded that the relatively high mutation rate at G174 is primarily driven by ligand recognition.

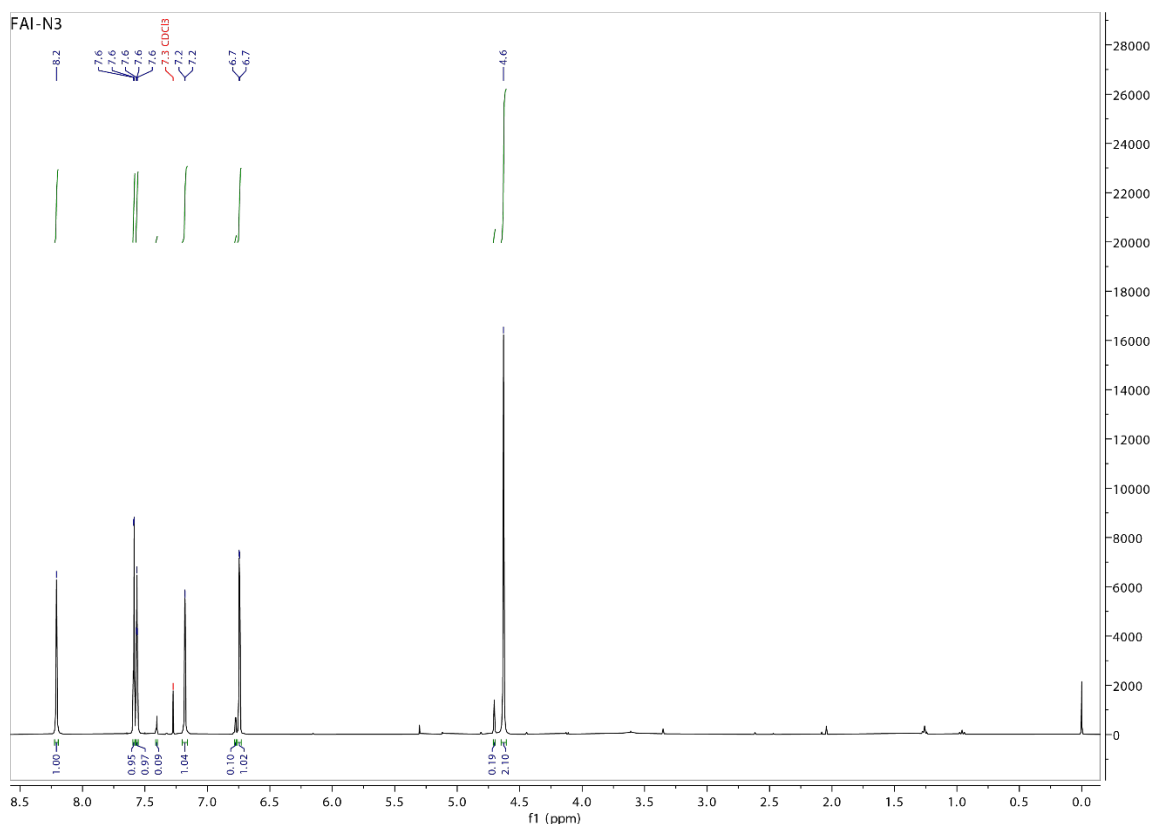
2. The predominant contaminant observed in NMR is either imidazole, as discussed by the authors and reviewer 3, or possibly contaminating amounts of CDI itself. In both <sup>1</sup>H and <sup>13</sup>C NMR spectra provided in the revised manuscript, "imidazole" peaks are broadened (likely a consequence of protonation/partial protonation). Given the size of these peaks I am concerned, along with Reviewer 3, that the actual purity of C30-FAI is far lower than the 91% HPLC purity reported by the authors. Further, since the peaks are broadened, they are

shifted slightly from literature values for peaks for either CDI or imidazole itself so an unambiguous assignment/quantification is difficult.

We thank the reviewer's concern regarding the contamination of imidazole and CDI, and we took the comments very seriously. We checked the old batch of C30-FAI in NMR, and it appeared that the samples only had imidazole as a major impurity, but not CDI. We have tried dozens of conditions to improve the purity and quality of the probe. Finally, we developed a new, reproducible synthetic route for C30-FAI that avoids using CDI (detailed in the response to reviewer 3 as well as in the revised manuscript). NMR characterization of the new C30-FAI confirms the absence of imidazole contamination. We obtained a result that is almost identical to the previous batch of C30-FAI.

3. I am concerned that if the contaminant is CDI, itself a reactive acylating agent, it could be the source of some of the altered mutation rate that is observed. Presumably, FAI-N3 does not contain any contaminants (spectra for FAI-N3 are not provided). However, a control assessing the effects of CDI itself is also not provided. Previously, and in unpublished work, our laboratory has seen contamination issues similar to this interfere with SHAPE assays.

We appreciate the reviewer's comment. The new C30-FAI is free of CDI contamination, as we avoided the use of CDI in its preparation. For FAI-N<sub>3</sub>, we synthesized the compound following the protocol reported in the literature (DOI: 10.1016/j.bmcl.2018.01.042), including a purification step using silica gel column chromatography with 0-50% ethyl acetate in hexanes. The FAI-N3 used in our experiments was a purified compound without CDI or imidazole contamination. The <sup>1</sup>H NMR confirms that FAI-N3 is free of both imidazole and CDI, containing ~8% of the carboxylic acid precursor (expected after extensive purification).



4. Hyper-reactivity/increased mutation rate of G174 is observed with multiple diazirines and a chlorambucil mustard. This is somewhat surprising given the different nature of crosslinkers (likely to react with bases) and SHAPE reagents (likely to react with 2'-OH). Is G174 just a hyper-reactive site? I am not sure this can be ruled out based on data shown. We thank the reviewer for the insightful comments. Diazirine and chlorambucil mustard are two types of probes frequently conjugated to RNA binders for RNA modification and profiling, both of them primarily modify nucleotide bases. The base modifications by C30 diazirine or chlorambucil mustard probes can introduce a bulky group on the base, which can probably be recognized in our reverse transcription condition, resulting in increased mutation rate at the modified bases. Despite this, we observed moderate mutation rates at G174 (less than 0.4%), and the signal-to-noise ratio is relatively low compared to what we observed with C30-FAI. In addition, our Extended Data Figure 3 demonstrated that G174 is not particularly active in the presence of FAI-N3 at 0.2, 0.1, or 1 mM.

On the other hand, we have conducted molecular dynamics simulation studies as part of a new study currently in progress. These simulations reveal that G174 exhibits significant flexibility in the apo structures. The simulations also show that G174 is crucial for ligand

interaction, suggesting that its role as a potential "hot spot" for chemical probing and its involvement in forming a binding pocket might not be mutually exclusive.

In summary, although we cannot rule out if G174 is more reactive than other residues in the chemical probing context, the cgSHAPE signal is primarily driven by the ligand binding, and this should not affect the main conclusions of this manuscript.

Because of these points, it is critical to use a highly pure sample of C30-FAI for the analyses.

We used a new batch of C30-FAI which is free of CDI and imidazole contamination for RNA mutational profiling.