Carbodiimide reagents for the chemical probing of RNA structure in cells

PETER Y. WANG,1,2 ALEC N. SEXTON,1,2 WILLIAM J. CULLIGAN,1,2,3 and MATTHEW D. SIMON1,2

1Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA
2Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, USA
3Department of Cell Biology, Yale University, New Haven, Connecticut 06511, USA

ABSTRACT

Deciphering the conformations of RNAs in their cellular environment allows identification of RNA elements with potentially functional roles within biological contexts. Insight into the conformation of RNA in cells has been achieved using chemical probes that were developed to react specifically with flexible RNA nucleotides, or the Watson–Crick face of single-stranded nucleotides. The most widely used probes are either selective SHAPE (2′-hydroxyl acylation and primer extension) reagents that probe nucleotide flexibility, or dimethyl sulfate (DMS), which probes the base-pairing at adenine and cytosine but is unable to interrogate guanine or uracil. The constitutively charged carbodiimide N-cyclohexyl-N′-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMC) is widely used for probing G and U nucleotides, but has not been established for probing RNA in cells. Here, we report the use of a smaller and conditionally charged reagent, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), as a chemical probe of RNA conformation, and the first reagent validated for structure probing of unpaired G and U nucleotides in intact cells. We showed that EDC demonstrates similar reactivity to CMC when probing transcripts in vitro. We found that EDC specifically reacted with accessible nucleotides in the 7SK noncoding RNA in intact cells. We probed structured regions within the Xist IncRNA with EDC and integrated these data with DMS probing data. Together, EDC and DMS allowed us to refine predicted structure models for the 3′ extension of repeat C within Xist. These results highlight how complementing DMS probing experiments with EDC allows the analysis of Watson–Crick base-pairing at all four nucleotides of RNAs in their cellular context.

Keywords: RNA structure; carbodiimide; chemical probing; in-cell

INTRODUCTION

RNA relies on Watson–Crick base-pairing to adopt intricate structures that are often essential to biological function. RNA conformation can be inferred using chemical modifiers that react based on the accessibility or flexibility of the folded RNA (Kwok et al. 2015; Choudhary et al. 2017). For instance, dimethyl sulfate (DMS) is used to probe the Watson–Crick faces of A and C nucleotides, N-cyclohexyl-N′-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMC) is used to probe those of U and G, and kethoxal those of G, while SHAPE (2′-hydroxyl acylation and primer extension) reagents react with the ribose sugar preferentially in regions with greater flexibility (Litt and Hancock 1967; Incarnato et al. 2014; Smola et al. 2015b; Lin et al. 2018). A common strategy of RNA chemical probing studies that greatly improves structure prediction is combining data from multiple chemical reagents to obtain a more comprehensive reactivity profile (Maenner et al. 2010; Novikova et al. 2012; Incarnato et al. 2014). Such strategies have been used to predict structures for a wide range of RNAs. These approaches have been especially important for understanding noncoding RNAs such as ribosomal RNAs (Lempereur et al. 1985; Moazed et al. 1986), spliceosomes (Black and Pinto 1989; Murphy and Cech 1993), and 7SK (Wassarman and Steitz 1991). The development of more sensitive and scalable approaches to RNA probing have made these experiments particularly useful for larger RNAs such as the long noncoding RNA (lncRNA) Xist (Maenner et al. 2010; Fang et al. 2015; Smola et al. 2016; Liu et al. 2017; Pintacuda et al. 2017), which are not readily amenable to structure determination by other means.

© 2019 Wang et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Xist is an ~18 kb lncRNA responsible for the recruitment of effector proteins that silence one X chromosome in mammalian females (Gendrel and Heard 2014). Xist contains many predicted structural regions, including repetitive elements that are critical for downstream function, and likely mediate interactions with proteins involved in the formation of heterochromatin (Maenner et al. 2010; Moinodrot and Brockdorff 2016; Liu et al. 2017). We previously reported the first structural predictions derived from probing the entire transcript of Xist in cells with DMS (Fang et al. 2015). These models continue to be refined using complementary approaches including SHAPE analysis in cell extracts (Smola et al. 2016) and comprehensive analysis of Xist fragments in vitro (Liu et al. 2017). Ideally, respective regions of Xist could be probed directly in cells with a suite of complementary reagents to develop higher quality models. The limited number of probing reagents effective in cells makes the probing of base-pairing states of lncRNAs such as Xist a principle barrier to understanding their structure and biology.

The chemical probing of RNA structure in vivo and in cells has been limited by the ability of reagents to permeate into cells and modify RNAs. Insight gained from the few probing reagents established for in-cell probing motivates the further development of other complementary reagents. DMS and SHAPE reagents have been demonstrated to be effective in cells (Spitale et al. 2013; Smola et al. 2016; Zubradt et al. 2017), and glyoxal and its derivatives have recently been added to this list (Mitchell et al. 2018). Thanks to recent advances in adapting chemical probing experiments to high-throughput sequencing, these in-cell experiments can be performed in either a targeted fashion for a specific RNA or transcriptome-wide (Ding et al. 2014; Rouskin et al. 2014; Fang et al. 2015; Sexton et al. 2017; Zubradt et al. 2017). Overall, while SHAPE has been established for in-cell probing of nucleotide flexibility, DMS for probing the Watson–Crick base-pairing state of only A and C, and glyoxal for G nucleotides, there are currently no well-validated methods to examine the Watson–Crick base-pairing of U nucleotides of RNAs.

This information gap prompted us to seek a reagent similar to CMC but more compatible for use directly in cells. The G/U reactivity of CMC is particularly attractive for its complementarity to that of DMS. The in vitro reactivity of CMC relies on its carbodiimide group, which reacts with uracil and guanine to form a covalent guanidinium link (Fig. 1A,B; Gilham 1962). We were interested in whether a different carbodiimide could retain the G/U reactivity of CMC yet have chemical properties better suited for the probing of RNA structure in cells.

Here, we report the development of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) for RNA probing both in vitro and in cells. We find that EDC probes the secondary structure of RNA with the same high fidelity, reproducibility, and signal strength as CMC in the context of high-throughput sequencing assays in vitro. Combining EDC chemical probing data with that of DMS allows the refinement of structure predictions of structured RNAs in cells, as we demonstrate for a subdomain within the Xist RNA.

RESULTS

Development of EDC as a probe with reactivity similar to CMC

While CMC has been extensively used for chemical probing in vitro, it is the only well-established carbodiimide probe of RNA structure. We reasoned that EDC might be an ideal alternative probe, as it is a commercially available carbodiimide with a lower molecular weight. Further, we predicted EDC would have higher cell permeability than CMC because of the lack of a quaternary ammonium ion, which causes CMC to have a constitutive positive charge. To ascertain whether the reactivity of EDC with RNA is comparable to CMC, we used a gel shift assay with a fluorescently labeled RNA oligonucleotide, where each adduct on the same RNA molecule causes a corresponding gel shift due to the net increase in positive charge from the adducts, allowing a discrete measure of the number of carbodiimide modifications per length of RNA. We used the assay to analyze the modification frequency across different ranges of temperature, reaction time, and concentrations for CMC and EDC. Both CMC and EDC reacted with RNA, exhibiting similar incremental modification of RNA, and forming a clear ladder pattern of modified oligonucleotide products (Fig. 1C–E). The highest total number of modified bases did not exceed the number of accessible G and U nucleotides in the RNA, suggesting no detectable unexpected side reactivity to A or C nucleotides. Both CMC and EDC demonstrated increased modification frequency with increased reaction time (Fig. 1C), temperature (Fig. 1D), and carbodiimide concentration (Fig. 1E). Under matched conditions, CMC displayed higher reactivity than EDC (Fig. 1C–E).

By making the simplifying assumption that each carbodiimide reacts with each G or U nucleobase in an identical and independent rate, the distribution of the number of modifications per RNA can be modeled as a binomial distribution allowing the calculation of an apparent second order rate constant under each reaction condition (Materials and Methods; Fig. 1C–E). We found the distribution of bands matched a binomial distribution as expected (Supplemental Fig. S1). Using the model under the concentration series, we found the apparent rate constant of EDC to be approximately 23-fold lower than that of CMC (Fig. 1E). While it is unsurprising that EDC is less electrophilic than CMC, we reasoned that the difference in rates would be unlikely to negatively impact RNA probing experiments. Respective concentrations of CMC and EDC were then
chosen to achieve a similar degree of modification frequency, as well as no more than one modification per 100 U and G nucleobases, to adapt each reagent for chemical probing experiments (Fig. 1E). These observations together suggest that EDC might be a viable alternative to CMC to probe RNA conformation.

The RNA probing specificity of EDC agrees with that of CMC in vitro

To test if EDC modifies the same accessible G and U nucleotides as CMC in vitro, we probed an in vitro-transcribed segment of Xist (nt 4701–5001), which is within a structured domain of Xist that we have studied previously (3′ extension of Xist repeat C, nt 4658–5090, NR_001463.3; hereafter Xist4658–5090) (Fang et al. 2015). This region of Xist is of interest because targeting this region with antisense PNA or LNA oligonucleotides in cells causes Xist to be displaced from the chromatin (Beletskii et al. 2001; Sarma et al. 2010; Simon et al. 2013), suggesting a role for the region in tethering Xist to the inactive X chromosome. We analyzed treated RNA with targeted reverse transcription, high-throughput sequencing, and a previously established analysis pipeline (Sexton et al. 2017) to obtain reverse transcription termination probabilities. These probabilities allowed us to compare the CMC and EDC reactivity profiles.

Plotting the termination probabilities along the RNA for CMC and EDC data sets revealed distinct peaks of high modification at various nucleotides for both CMC- and EDC-treated samples, as expected for RNA probing data (Fig. 2A). We observed a reactivity preference for U nucleotides and limited G-reactivity in the CMC data set, which is in agreement with previous reports using CMC as a chemical probe in vitro (Fig. 2B). We also observed a similar U-preference in the termination probabilities of EDC samples and G reactivity to a lesser extent (Fig. 2B). Thus, EDC shares the distinct hallmark of CMC’s RNA modification profile. Furthermore, the normalized termination rates induced by CMC and EDC show strong correlation when comparing data from U and G nucleotides, implying that EDC not only reacts at the same locations

FIGURE 1. RNA modification chemistry of carbodiimides and its characterization. (A) Chemical structures of CMC and EDC with carbodiimide highlighted in blue. (B) Reaction of carbodiimide modification of uracil at N3. The reaction with guanine occurs similarly at N1. Under mildly basic conditions, the deprotonated N in the nucleobase attacks the electrophilic C in carbodiimide, resulting in the formation of a guanidinium adduct. (C–E) Electrophoretic mobility shift assay for RNA modification by carbodiimides at varying reaction conditions. The shared bottom band corresponds to unmodified RNA oligonucleotide, and each band above corresponds to each incremental number of modifications per oligonucleotide. $k_{app}$ denotes approximate apparent second order rate constants calculated using a binomial distribution (Materials and Methods). Reactions were assayed at varying reaction times (C), temperatures (D), and carbodiimide concentrations (E).
as CMC but also reproduces the relative quantitative reactivity profile of CMC (Fig. 2C). Nonetheless, at the chosen conditions, EDC exhibited around twofold lower reactivity values across the sequence compared to CMC, leading to a slightly higher background reactivity reflected by some termination probability at A and C nucleotides (Fig. 2A).

Comparison of background carbodiimide signal (at A and C) from this and other data sets demonstrates that even with variability in background modification, carbodiimide background rates are similar to those of the commonly used structure probe DMS (Supplemental Fig. S2). Since CMC has long been an established robust chemical

---

**FIGURE 2.** In vitro validation and comparison of CMC and EDC for RNA chemical probing using a subregion in Xist<sub>4658-5090</sub> (nt 4701–5001). (A) P(stop) values obtained for CMC and EDC probing. (B) Cumulative distribution of P(stop) values for all nucleotides probed, showing bias for U in both data sets. (C) Correlation of normalized P(stop) values between CMC and EDC data sets. Respective P(stop) values are normalized at the 95th percentile as described (Materials and Methods). (D) Reactive G and U nucleotides identified in CMC and EDC data sets as mapped to the predicted secondary structure model of the region. Asterisk denotes conflicts with the structure model, and a blue asterisk denotes a conflict consistent between CMC and EDC.
probe of RNA structure, the marked similarity of EDC’s reactivity profile with that of CMC strongly suggests that EDC can also function as a robust structure probe.

High-throughput sequencing data generated by DMS probing has been shown to contain information not only in modification-induced stops but also mutations. These different readouts both report on the locations of chemical adducts and can provide complementary information (Novoa et al. 2017; Sexton et al. 2017; Yu et al. 2018). Therefore, we analyzed the mutation rates in our carbodiimide data sets. Neither CMC nor EDC data sets yielded a significant increase in mutation rates under our reverse transcription conditions, presumably due to the larger adduct size and the position on the hydrogen bonding face of the nucleobase compared with other reagents that give useful mutational information (Supplemental Fig. S3). Nonetheless, with continued improvement of mutational profiling pipelines for RNA probing experiments (Zubratt et al. 2017; Busan and Weeks 2018), it is possible that the right combination of reverse transcriptase and conditions will allow analysis of adduct-induced mutations, as has recently been reported when using CMC to map the locations of pseudouridine (Zhou et al. 2018). It is possible that the smaller adduct formed by EDC relative to CMC will prove beneficial for future mutational analyses.

To verify that EDC is a bona fide chemical probe for RNA structure, we identified and mapped reactive nucleotides in each data set to the previously predicted model of the region. CMC reacts with nucleotides at or flanking single-stranded RNA or nucleotides involved in a G-U wobble-pair. In both CMC and EDC data sets, U and G nucleotides identified as highly or moderately reactive are located at expected locations in the structure model (Fig. 2D). The few exceptions (3 of 27) are all moderately reactive nucleotides, and are generally consistent between CMC and EDC data sets (Fig. 2D). Therefore, CMC and EDC both react at the expected U and G nucleotides in vitro, leading to similar reverse transcription termination events. These results demonstrate that EDC behaves as a chemical probe for RNA in vitro.

EDC is a robust chemical probe of RNA conformation in cells

Having established EDC as a functional RNA structure probe analog to CMC in vitro, we tested if EDC could be used to probe RNA in cells. We treated intact cells with a range of concentrations of either EDC or CMC, where a higher range of concentrations was used for EDC due to its higher solubility and lower reactivity in vitro. We performed targeted analysis of the 7SK noncoding RNA in mouse embryonic fibroblast (MEF) cells. The 7SK RNA has a well-documented, conserved secondary structure (Marz et al. 2009). The first major stem–loop in 7SK has been extensively studied with in vitro chemical probing (Wassarman and Steitz 1991; Brogie and Price 2017), NMR (Bourbigot et al. 2016), and X-ray crystallography (Martinez-Zapien et al. 2017), revealing its multiple U-bulges and loops with single-stranded U, thus providing a point of comparison for exposed bases that should be detectable by chemical probing.

Probing of cells with EDC led to distinct EDC-dependent termination events across all concentrations used (Fig. 3A). Furthermore, examination of the cumulative distributions of modification probabilities showed the same strong U-preference as was observed in vitro, consistent with the expected carbodiimide reactivity (Fig. 3B). Data from the lowest concentrations of EDC (100 mM) demonstrated slightly weaker correlation with those from higher concentrations (Supplemental Fig. S4A), while samples treated with the highest concentration of EDC (1.5 M) showed markedly higher signal at U nucleotides and also more signal at G nucleotides (Fig. 3A,B).

In contrast to EDC, CMC-treated samples at lower concentrations (10 mM, 30 mM) displayed minimal signal (Fig. 3A), in line with previous expectations. To our surprise, however, at the highest concentration of CMC attempted (90 mM), we observed peaks of high termination frequency that appear similar to those observed at the lower EDC concentration experiments (Fig. 3A). The probability distribution has the same U-preference observed in EDC data sets (Fig. 3B), and the reactivity profile also correlates well with EDC (Supplemental Fig. S3A).

All three EDC data sets and the 90 mM CMC data set accurately identified nucleotides at allowed open positions on the secondary structure, with very few conflicts (≤2 of ~25; Fig. 3C; Supplemental Fig. S4B). In particular, crystallography and NMR identified U40, U41, U63, and U72 to be unpaired and lacking tertiary interactions (Bourbigot et al. 2016; Martinez-Zapien et al. 2017). These were strongly captured in all of our EDC and 90 mM CMC probing results (Fig. 3A,C). We were also able to observe mild reactivity at U68 which demonstrates flexible base-pairing in molecular dynamics simulations (Martinez-Zapien et al. 2017), NMR (Lebars et al. 2010; Bourbigot et al. 2016), and SHAPE probing (Lebars et al. 2010). These together suggest that EDC, as well as high concentrations of CMC, can be used for the probing of RNA conformation in intact cells.

We wondered if the carbodiimide-reactive G residues would show overlap with those identified by glyoxal treatment. We treated cells with glyoxal based on published protocols for a targeted analysis of 7SK (Mitchell et al. 2018). Both carbodiimide and glyoxal data sets identified single-stranded G bases, and we found partial overlap in which specific nucleotides were modified (Fig. 3A; Supplemental Fig. S5A,B). The discrepancies are likely due to differences in the properties of carbodiimides and glyoxal, and suggest that complementing glyoxal and carbodiimide data sets could help identify more unpaired G nucleotides in RNA structures.
FIGURE 3. In-cell carbodiimide probing of 7SK RNA structure. (A) Bar plots showing \( P(\text{stop}) \) values obtained by in-cell probing with three concentrations each of CMC or EDC. The 1500 mM EDC plot includes some bars that are above the maximum value shown to better display the range of values. (B) Cumulative distribution of \( P(\text{stop}) \) values showing clear U-preference across data sets. (C) Reactive nucleotides found in 300 mM EDC samples as mapped onto the known secondary structure of 7SK, with conflicts marks as previously described. (D) The difference in normalized reactivity for the first major stem–loop of 7SK (nt 24–87), where positive values denote reduced reactivity in cells than in vitro, and negative values increased reactivity. Uridines discussed in the text are labeled. Reactivity difference values were mapped to the secondary structure of the stem–loop on the right. The \((\text{GAUC})_2\) motif recognized by HEXIM is bolded for reference.
Comparing EDC probing in vitro and in cells

Chemical probing data from cellular and noncellular contexts can be compared to gain knowledge about differences in signal due to folding or contacts with proteins. With SHAPE, comparisons between in-cell and ex vivo probing have been used to reveal structural perturbations and protein binding of RNAs in cells (Smola et al. 2015a). Our in vitro and in-cell EDC data provided a similar opportunity. We normalized and compared the reactivity patterns of EDC probing of the human 7SK transcript in vitro with in-cell values of the mouse 7SK (nearly identical in sequence) to identify context-dependent differences in base accessibility (Fig. 3D). We focused our analysis on the first major stem–loop (nt 24–87) of the 7SK RNA, whose binding proteins, such as HEXIM, and structural changes upon protein binding are well-documented in literature. A previous report describing specific photocrosslinking of U30 to HEXIM suggested the proximity of bound HEXIM at the nucleobase (Bélanger et al. 2009). Interestingly, we found U30 to be less reactive in cells than in vitro (Fig. 3D). In addition, previous NMR results suggested the opening of the (GAUC)₂ HEXIM motif and the hydrogen bonding of the flanking U40, U41, and especially U63, upon HEXIM binding (Lebars et al. 2010). We observed an increase in reactivity at U63 in cell, while we observed a slight increase at U40 and none at U41, indicating a change in base conformation or interactions (Fig. 3D).

Combining EDC with DMS data improves the conformational modeling of Xist IncRNA

With EDC as a validated and robust structure probe for U and G nucleotides in cells, we reasoned that it could be used to complement DMS data and refine RNA structure predictions via additional modeling constraints. This opportunity allowed us to use carbodiimide reagents to revisit the conformations of Xist₄₆₅₈–₅₀₉₀, a recently identified structured region of the Xist IncRNA, which has been predicted to contain many highly structured regions implicated for function (Fang et al. 2015; Smola et al. 2016).

Xist₄₆₅₈–₅₀₉₀ is a region that is conserved between human and mouse and predicted to exhibit secondary structure, but the conformational model proposed based on DMS probing of Xist in cells (Fang et al. 2015) conflicted in some regions with subsequent ex vivo SHAPE probing experiments (Supplemental Fig. S6A,B; Smola et al. 2016). These discrepancies could be due to differences in Xist conformation in cells versus in extracts, or due to protein footprints in cells that might create false negative signals and interfere with structural prediction, though their impact should be minimized by using only positive probing signals. If either were the case, additional data from in-cell probing would be expected to largely agree with the model from in-cell DMS probing. Alternatively, if the previous in-cell model was under-constrained, data from in-cell EDC probing could help refine the conformational prediction. Therefore, we generated additional data from in-cell EDC probing. Cellular RNA was probed with EDC (300 mM) and targeted reverse transcription was used to analyze this region of Xist. The reactivity data was integrated with the data from in-cell DMS probing (Fig. 4A; Fang et al. 2015). We also included mutational probability readout information from DMS probing in addition to termination probabilities, which has recently been shown to allow a more comprehensive interrogation of the base-pairing state of A and C nucleotides in DMS probing data (Fig. 4A; Novoa et al. 2017; Sexton et al. 2017; Yu et al. 2018). These data were used to generate a revised model for the region (Fig. 4B). Probing with 90 mM CMC and 1.5 M EDC identified similar sets of reactive nucleotides that closely agreed with the structure, corroborating the model and further demonstrating agreement across different carbodiimide probing experiments (Supplemental Fig. S7).

The revised model preserves all the stem–loops shared in the Fang et al. (2015) in-cell model and the Smola et al. ex vivo model (Fig. 4B; Supplemental Fig. S5A,B). In regions that showed discrepancies between the two previously reported models, the new model features stem–loops that share some characteristics with both models (Fig. 4B). In particular, when DMS and EDC data were integrated, part of the 4860–4908 stem–loop was brought into agreement with the Smola et al. model (Fig. 4B), while its end overlaps with sites of predicted protein binding by SHAPE analysis and CLIP (Supplemental Fig. S8; Smola et al. 2016). The base-pairing of 4956–5004 with 4699–4707 found only in the Fang et al. model was preserved in the new model (Fig. 4B). This stem was previously shown to be targeted by the antisense oligonucleotide LNA-4978, which could abolish Xist localization, thus maintaining the suggested structure-function relationship (Fang et al. 2015). Finally, the hairpin at 4911–4979 in the revised model does not share any features with either previous model (Fig. 4B). Thus, EDC probing data can complement existing reagents that work in cells to provide complementary data to inform structure models of challenging RNAs.

DISCUSSION

The upswing in interest in cellular RNA structures has motivated a wide range of studies examining both transcriptome-wide and RNA-specific conformations using chemical probing directly in cells. Historically, biochemically enriched and in vitro-transcribed RNAs were probed with several complementary reagents to develop high-quality conformational models. While adapting these experiments to a sequencing platform has advantages that have expanded the throughput and scope of RNA chemical probing experiments, RNA probing reagents that work in cells have been limiting. Indeed, most studies probing...
RNA in cells have used data from only a single chemical probe to develop conformational models. Recent advances with in-cell probing using SHAPE reagents have demonstrated the power of probing nucleotide flexibility in cells (Spitale et al. 2013; McGinnis and Weeks 2014; Mustoe et al. 2018). In-cell probing of Watson–Crick pairing, on the other hand, has largely been limited to the specificity of DMS reactivity toward only A and C nucleobases in RNA (Kubota et al. 2015). The recent addition of glyoxal to the toolkit expands in-cell probing to include G (Mitchell et al. 2018). By establishing carbodiimides for chemical probing in cells, it is now possible to interrogate the RNA base-pairing of all four canonical nucleotides including U nucleobases.

CMC was originally developed as a water-soluble reagent to activate carboxylic acids for amide bond formation (Sheehan and Hlavka 1956). The utility of CMC as a mild carbodiimide-based reagent that modifies G and U nucleotides was noted over half a century ago (Gilham 1962). It was shown soon after to react preferably at nonhydrogen bonded bases, establishing its potential as a chemical probe of RNA conformation (Augusti-Tocco et al. 2018).

**FIGURE 4.** Revised structure model for Xist4458–5090. (A) Bar plots showing three chemical probing data sets used to constrain the structure prediction, namely P(stop) and P(mut) from in-cell DMS probing from the Fang et al. (2015) data set, and P(stop) from in-cell probing with EDC (300 mM). (B) The new revised model of Xist4458–5090, showing reactive nucleotides identified via in-cell chemical probing, and base-pairing agreement with the Fang et al. (2015) in-cell model and the Smola et al. (2016) ex vivo model.
and Brown 1965; Metz and Brown 1969). Aside from the reactivity of the carbodiimide, it was unclear to us whether or not other properties of CMC were important for successful probing. While the constitutive positive charge on CMC may enhance its reactivity, we found that an alternative carbodiimide, EDC, provides similar nucleotide specificity. Though EDC is a commonly used water-soluble crosslinking reagent at low pH (Sehgal and Vijay 1994), there have never been reports of its use as a chemical probe for RNA structure. While this study suggests many other carbodiimides will also provide the desired reactivity in cells, the availability, solubility, and size of EDC satisfied all our criteria for an effective in-cell probing reagent. Comparing EDC with the previously used carbodiimide CMC, EDC consistently demonstrated less reactivity toward RNA than CMC in our in vitro experiments. Nonetheless, in-cell experiments with EDC showed comparable reactivity to CMC at similar concentrations, possibly due to its greater ability to permeate into cells, contributing to its improved performance for in-cell probing. While we show here that CMC is an effective in-cell probe, we note that reproducing the strong signal from higher concentrations of EDC with CMC would require concentrations beyond its saturation point in water, whereas EDC was readily soluble in water without the need for organic solvents even at the highest concentration we attempted, implying that EDC could allow a much wider range of concentrations for the optimal chemical probing of RNA in cells and potentially provide stronger chemical probing signals. Indeed, EDC probing directly in cells helped us refine the modeling of a region within the Xist IncRNA, demonstrating how this approach can be applied to understand the structures of long and moderately expressed RNAs. We also demonstrate how these reagents can be used in other ways, such as comparing the conformational differences at U nucleobases of 7SK in its cellular (protein-bound) context and in vitro. Similarly, our revised model of Xist7455-L5090 provides an opportunity for comparisons with protein binding data and analyses such as ΔSHAPE (Smola et al. 2016). More generally, we anticipate many applications that have been developed for SHAPE and DMS probing will also be compatible with EDC probing. Therefore, the chemistry of carbodiimides such as EDC complements the current set of in-cell-compatible chemical probes for RNA structure, and completes our ability to probe the Watson–Crick face of all four nucleotides, thereby allowing a more complete understanding of the RNA conformations underlying RNA biology.

**MATERIALS AND METHODS**

**RNA gel shift assay**

A Cy3-labeled 17-mer RNA oligonucleotide (Cy3-5'-GGGGGCUU CAAAUCCCUC-3', including nine total G/Us; Dharmacon) at a final concentration of 100 nM with 1:1000 final volume of SUPERase I RNase inhibitor (Thermo Fisher) was mixed with corresponding concentrations of freshly dissolved CMC or EDC in borate reaction buffer (50 mM Na2B4O7 pH 8.0, 100 mM KCl, 5 mM MgCl2) to give a final total volume of 20 µL. The reaction was warmed up to 37°C at various time lengths and temperatures, then quenched with sodium acetate (pH 5.2) to a final concentration of 0.3 M. Aliquots of the reaction mix were then run on 20% urea-PAGE at 250 V for 90 min in darkness, and subsequently scanned for the Cy3 fluorophore. Gel images were analyzed using ImageJ and R to integrate each incremental band by approximate integration over time, 

\[
\frac{d[\text{modified G or U}]}{dt} = k_{app}[\text{G or U}][\text{carbodiimide}],
\]

and by approximate integration over time, 

\[
9[\text{RNA}] P_{mod} = k_{app} \cdot 9[\text{RNA}][\text{carbodiimide}] \Delta t.
\]

**In vitro probing of RNA**

RNA transcribed in vitro (2 µL at 2.5 µM) was first incubated in borate reaction buffer for 10 min at 37°C at a final volume of 20 µL, then cooled briefly to room temperature. Then, 10 µL of 30 mM (3 × final) CMC, 300 mM (3 × final) EDC, or dH2O control solution was added. The reaction was vortexed briefly to mix and incubated at 25°C for 20 min, before quenching with sodium acetate (5 µL at 2M, pH 5.2), followed by ethanol precipitation of the RNA.

**Cell culture and in-cell probing of RNA**

MEF cells were cultured in DMEM with 10% FBS (Thermo Fisher) and penicillin/streptomycin to 90% confluency in 15 cm plates, in duplicate for each treatment sample. Cells were washed with PBS and then incubated with respective reagent solutions. For CMC and EDC treatment, cells were incubated with 10 mL of respective concentrations of reagents (CMC: 10 mM, 30 mM, 90 mM; EDC: 100 mM, 300 mM, 1500 mM) in borate reaction buffer for 30 min at 25°C. For glyoxal treatment, cells were incubated with 10 mL of 50 mM glyoxal solution in reaction buffer (50 mM HEPES pH 8.0, 50 mM KCl, 0.5 mM MgCl2) for 15 min at 25°C (Mitchell et al. 2018). No-treatment control plates were incubated with PBS. After treatment, cells were lifted using Cell Scraper (Corning) and collected in PBS by centrifugation (3 min, 1000 rpm, 4°C), and the supernatant was discarded. The RNA was extracted using TRIzol (Thermo Fisher) according to manufacturer’s instructions. The extracted RNA was treated with RNase (1 h, 37°C; Promega), and subsequently purified with phenol/chloroform extraction and ethanol precipitation.

![Image]

*www.rnajournal.org* 143
Reverse transcription

Reverse transcription was carried out using either 2 µg (7SK) or 10 µg (Xist) of extracted RNA, and 1 pmol total of primers. The RNA/primer mix was brought up to 12 µL and incubated at 70°C for 5 min, then 4°C for 2 min for primer annealing. Next, 3 µL of 5X FS buffer (Life Technologies) was added for a 10 min incubation at 55°C. Finally, 5 µL of reverse transcription mix (1 µL 5X FS buffer, 1 µL 10 mM dNTPs, 1 µL 100 mM DTT, 0.2 µL RNase OUT [Thermo Fisher], 0.5 µL SuperScript III [Thermo Fisher], 1.3 µL dH2O) was added and mixed before incubation at 55°C for 45 min, followed by heat inactivation at 85°C for 5 min. After reverse transcription, RNA was degraded by adding 0.4 µL of 5 M NaOH and incubating at 95°C for 3 min, before neutralization with 0.4 µL of 10 M acetic acid and cooling on ice.

Sequencing library preparation

The cDNA from reverse transcription was purified with AMPure XP beads (Beckman Coulter) by adding and mixing 1 volume of bead solution and 3 volumes of PEG solution (2.5 M NaCl, 20% PEG 8000) at room temperature. The mixture was let to incubate at room temperature for 15 min, before immobilizing the beads on a magnetic rack for 10 min. The beads were washed with fresh 80% ethanol twice before being dried for 10 min and eluted in 16 µL of dH2O. 3′ adaptor ligation was carried out using T4 RNA Ligase I by mixing 5 µL of cDNA with 1 µL of 20 µM 3′ adaptor (5′-Phos/N/NNAAGTGGAAAGACGTCGTGA/3′Bio/), 10 µL 80% ethanol twice before being dried for 10 min and eluted in 16 µL of dH2O. 3′ adaptor ligation was carried out using T4 RNA Ligase I by mixing 5 µL of cDNA with 1 µL of 20 µM 3′ adaptor (5′-Phos/N/NNAAGTGGAAAGACGTCGTGA/3′Bio/), 10 µL 50% PEG 8000, 1 µL 1 mM ATP, 1 µL T4 RNA Ligase I buffer, and incubating at 25°C for 16 h. Ligated DNA was purified with 1 volume of AMPure XP beads, and eluted in 16 µL dH2O. Products were amplified with eight cycles of PCR using Phusion (New England Biolabs), and primers for the 5′ and 3′ adaptor sequences (5′-CA GACGTTGCTCTTCGGATC-3′; 5′-CTACAGACGCTCTTCGG ATCT-3′) with cycles of 98°C, 20 sec; 64°C, 20 sec; 72°C, 90 sec. Products were purified as before with 1 volume of AMPure XP beads. Forward and reverse primers for Illumina TruSeq sequencing were added by 4–8 more cycles of PCR with Phusion, followed by purification with 1 volume of AMPure XP beads as previously described. Multiplexed sequencing libraries were mixed for 2 × 75 paired-end sequencing on Illumina HiSeq 2500 at the Yale Center for Genome Analysis. Sequencing data are available at the Gene Expression Omnibus (GEO) repository, accession GSE118309.

Determination and analysis of reactive nucleotides

FASTQ sequencing reads were trimmed with Cutadapt to remove Illumina adaptor sequences, and aligned using Bowtie2 local alignment to respective RNA target sequences. Reverse transcription stop and mutations were identified, screened, and tallied using previously published software (Sexton et al. 2017). Reverse transcription stop and mutation probabilities were calculated as

\[
P_{\text{raw}}(\text{stop}) = \frac{n_{\text{stop}}}{n_{\text{stop}} + n_{\text{read-through}}},
\]

\[
P_{\text{raw}}(\text{mut}) = \frac{n_{\text{mut}}}{n_{\text{read-depth}}}
\]

Treated modification probabilities were subtracted with untreated probabilities:

\[
P(\text{stop}) = P_{\text{treated}}(\text{stop}) - P_{\text{no treatment}}(\text{stop}),
\]

\[
P(\text{mut}) = P_{\text{treated}}(\text{mut}) - P_{\text{no treatment}}(\text{mut}).
\]

Only values from more than 100 read-throughs or read-depths are considered. All values were averaged between the two biological replicates. To reconcile different reactivity profiles across different reagents and reaction conditions, probability values were normalized using a 90% Winsorizing strategy modified from that previously used to normalize data across DMS and CMC (Incamnato et al. 2014). DMS data sets were normalized to A and C nucleotides, CMC and EDC to U nucleotides, and glyoxal to G nucleotides. High- and moderate-reactivity nucleotides were then called by setting appropriate thresholds for each reagent and condition.

Structure modeling

Secondary structure models were generated using RNAstructure v6.0 (Reuter and Mathews 2010) via free energy minimization, using an approach previously used for DMS probing (Fang et al. 2015). Highly reactive nucleotides shown via P(stop) in the EDC probing data, and both P(stop) and P(mut) in the previously published DMS probing data (Fang et al. 2015; Sexton et al. 2017) were identified as modifiable nucleotides and used as constraints in the software package.

Comparison of in-cell and in vitro probing data

The 95th percentile-normalized P(stop) values for the three titrations of in-cell EDC probing were averaged, and then subtracted from the 95th percentile-normalized P(stop) values from in vitro probing to obtain relative modification values across all U nucleotides in 7SK. Positive values suggest decreased reactivity in cellular context, and negative values denote increased reactivity.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank Michael Rutenberg-Schoenberg and members of the Simon laboratory for helpful discussions. This work was supported by the Rosenfeld Science Scholars Program (P.Y.W.); the Anderson Fellowship (A.N.S.); National Institutes of Health T32GM007223 and T32GM007205 (W.J.C.); National Institutes of Health New Innovator Award DP2 HD083992-01 (M.D.S.); and a Searle scholarship (M.D.S.).

Received June 1, 2018; accepted October 18, 2018.

REFERENCES

Augusti-Tocco G, Brown GL. 1965. Reaction of N-cyclohexyl, N-β(4-methylmorpholinium) ethyl carbodiimide iodide with nucleic acids and polynucleotides. Nature 206: 683. doi:10.1038/206683a0
Wang et al.

Sarma K, Levasseur P, Aristarkhov A, Lee JT. 2010. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. Proc Natl Acad Sci 107: 22196–22201. doi:10.1073/pnas.1009785107

Sehgal D, Vijay IK. 1994. A method for the high efficiency of water-soluble carbodiimide-mediated amidation. Anal Biochem 218: 87–91. doi:10.1006/abio.1994.1144

Sexton AN, Wang PY, Rutenberg-Schoenberg M, Simon MD. 2017. Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. Biochemistry 56: 4713–4721. doi:10.1021/acs.biochem.7b00323

Sheehan JC, Hlavka JJ. 1956. The use of water-soluble and basic carbodiimides in peptide synthesis. J Org Chem 21: 439–441. doi:10.1021/jo01110a017

Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maiar VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504: 465–469. doi:10.1038/nature12719

Smola MJ, Calabrese JM, Weeks KM. 2015a. Detection of RNA–protein interactions in living cells with SHAPE. Biochemistry 54: 6867–6875. doi:10.1021/acs.biochem.5b00977

Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM. 2015b. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nat Protoc 10: 1643–1669. doi:10.1038/nprot.2015.103

Smola MJ, Christy TW, Inoue K, Nicholson CO, Friedersdorf M, Keene JD, Lee DM, Calabrese JM, Weeks KM. 2016. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist IncRNA in living cells. Proc Natl Acad Sci 113: 10322–10327. doi:10.1073/pnas.1600081113

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. Nat Chem Biol 9: 18–20. doi:10.1038/nchembio.1131

Wassarman DA, Steitz JA. 1991. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. Mol Cell Biol 11: 3432–3445. doi:10.1128/MCB.11.7.3432

Yu AM, Evans ME, Lucks JB. 2018. Estimating RNA structure chemical probing reactivities from reverse transcriptase stops and mutations. bioRxiv doi:10.1101/292532

Zhou KI, Clark WC, Pan DW, Eckwahl MJ, Dai Q, Pan T. 2018. Pseudouridines have context-dependent mutation and stop rates in high-throughput sequencing. RNA Biol 15: 892–900. doi:10.1080/15476286.2018.1462654

Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat Methods 14: 75–82. doi:10.1038/nmeth.4057