RNA Chemical Probing Reagents and Protein Amino Acids: a Double-Edged Sword

AUTHORS

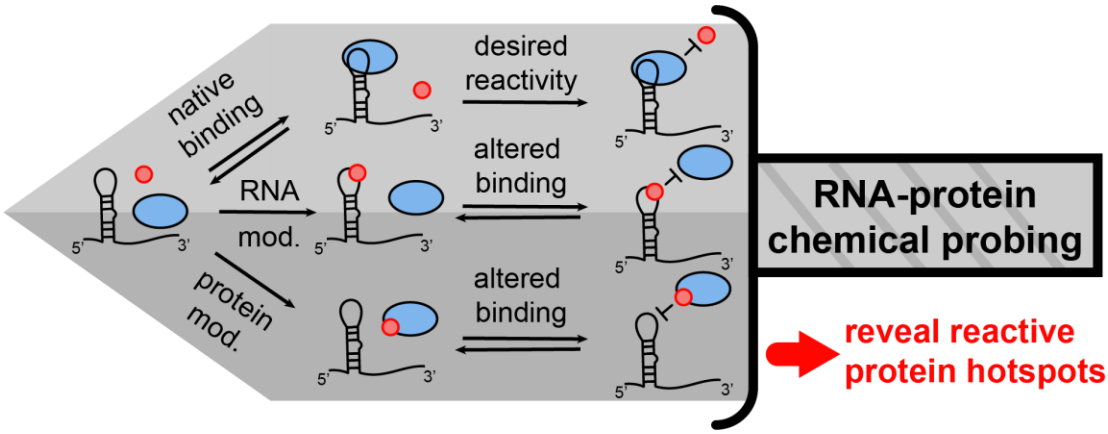
Lucy Fallon[1]†, David Klingler[1]†, Daniel Cohn[1], Liza Marcus[1], and Alisha N Jones[1,*]

[1] Department of Chemistry, New York University, NY, NY 10003, USA

* To whom correspondence should be addressed. Tel: +1 (212) 998-7979; Email: alisha.jones@nyu.edu

† Joint Authors

GRAPHICAL ABSTRACT



ABSTRACT

RNA chemical probing experiments are a broadly used method for revealing the structure of RNA, as well as for identifying protein binding sites. This is beneficial for expanding our understanding of biological processes governed by protein-RNA complex interactions, as well as facilitating the identification of complex inhibiting molecules. The reagents commonly used in chemical probing experiments are highly reactive, methylating or acylating atoms of flexible RNA nucleotides. The highly reactive nature of the chemical probes means that they can also react with nucleophilic amino acid side chains, and subsequently affect protein-RNA binding events. We combine MD simulations, MALDI-MS, and NMR experiments to show that commonly used RNA chemical probes react with protein amino acids and demonstrate that this effect alters protein-RNA binding interactions through binding shift assays. We discuss the implications of this phenomenon in elucidating protein-RNA interactions, and also demonstrate

this as a useful means for identifying surface-exposed protein residues, omitting the need for extensive mutagenesis experiments that are critical for RNA interactions.

INTRODUCTION

RNA structure is best described as an ensemble of various conformations (1). The dynamic nature of RNA makes experimentally characterizing its structure challenging. Biochemical experiments such as selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), and dimethyl sulfate (DMS) chemical probing are used to determine the secondary structure of conformationally dynamic RNA molecules, as well as to identify tertiary structural contacts of RNAs both *in vitro* and *in cells* (2–6). This method has been used to elucidate the secondary structure of both small and large RNA molecules, from micro-RNAs to entire viral genomes (7, 8). Further, chemical probing is often used to identify protein and other ligand binding sites on RNA molecules: regions of the RNA that demonstrate different chemical reactivities in the presence of a binding partner versus in the absence of it are typically interpreted as binding sites (9, 10).

The chemical probes used in SHAPE experiments react rapidly with the 2'-hydroxyl group of the ribose sugar of unpaired or everted nucleotides, covalently modifying the ribose (11). DMS functions similarly: it binds and methylates the N1 position of adenine, and the N3 position of cytosine (12). The covalently modified RNA is then sequenced, using either reverse transcription-stop (RT-stop) or next-generation sequencing (RT-map) methods, which permits identification of the nucleotides that reacted with the chemical probes (7, 12). The reactivities are then used to improve the prediction of RNA structure with *in silico* folding approaches.

One of the benefits of chemical probing experiments is that they can be used to identify sites of protein and other ligand binding in an RNA. For example, in a recent study, SHAPE was used to identify where two RNA binding proteins (RBPs), hnRNP U and hnRNP L, bound to the precursor messenger RNA MALT1 to regulate how it is spliced (13). In another study, DMS probing was used to identify how proteins bind to increase the flexibility of the 7SK noncoding RNA, thereby increasing the accessibility of neighboring binding sites (14). Chemical probing has also been used to assess ligand binding affinity, when carried out at variable ligand concentrations (15–17). These experiments have enabled us to better understand a plethora of biological processes that rely on ligand-RNA interactions.

An aspect of chemical probing protein/ligand mapping experiments that has received little attention is the ability of the chemical probes to interact with the ligand, especially given the on-off rates of the RNA-ligand complex and the rate at which the chemical probes interact. Relative to an RNA-only control (**Figure 1A**), when a protein ligand is present, four scenarios can occur (**Figure 1B-E**). First, binding of the protein to the RNA can prevent the probe from interacting with the RNA molecule (**Figure 1B**). In the second scenario, the probe is bound to the RNA, and prevents the recognition of the RNA by the protein (**Figure 1C**). Most poorly understood is the ability of the chemical probes to interact with the protein (and/or the RNA), preventing complex formation (**Figure 1D, E**). Lastly, combinations of these scenarios may occur. It is important to note that if chemical probing is to be used to investigate binding affinities only one type of interaction, scenario B, is desired.

In this study we combine binding shift assays, NMR, MALDI-MS, and MD simulations to demonstrate that the chemical probes frequently used in RNA chemical probing experiments can bind and react with nucleophilic amino acids that reside on the surface of proteins. Our results show that the covalent modification of amino acids by the chemical probing reagents is a double-edged sword: i) it alters the protein binding surface and obscures the accurate deduction of protein-RNA interactions, and ii) it can be used as a means for identifying surface residues on protein constructs. We discuss the implications of the chemical probes reacting with amino acids in interpreting protein-RNA binding events, thus highlighting the importance of further validating ligand-RNA interactions with alternative methods following chemical probing experiments.

MATERIAL AND METHODS

Protein Expression

Plasmids encoding for SHARP RRM1 and the KH3 domain of hnRNP K (each preceded with an N-terminal 6x histidine tag (his-tag) and a tobacco etch virus (TEV) cleavage site) were transformed into BL21 DE3 *E. coli* chemically competent cells and expressed in M9 minimal media solution supplemented with ^{15}N NH_4Cl. Cells were grown at 37°C; upon reaching an optical density (OD_{600}) of 0.9, they were induced to express protein with 0.5 M isopropyl ß-D-1-thiogalactopyranoside (IPTG) at 18°C. Cells were lysed by sonication and purified using immobilized metal (nickel) affinity chromatography (IMAC). Briefly, the protein was washed in a

buffer containing 50 mM Tris pH 8, 300 mM NaCl, 5 mM beta-mercaptoethanol, and 10 mM imidazole. An on column cleavage of the his-tag was performed by adding 10 mg/mL of TEV protease. The flow through was assessed by sodium dodecyl-sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) followed by size exclusion chromatography (HiLoad 16/60 Superdex 75) in a buffer containing 150 mM NaCl, 5 mM dithiothreitol (DTT), and 25 mM sodium phosphate, pH 6.5. Purified fractions were assessed by SDS PAGE, followed by flash freezing in liquid nitrogen and storage at -80°C until further use.

Protein Modification

Protein samples with a concentration of 200 µM of the respective protein in a buffer (25 mM sodium phosphate, 150 mM NaCl, pH 6.5) were prepared. Stock solutions of 1M7 (a gift from Michael Sattler) in anhydrous DMSO were added for a final concentration of 18 mM (12.5 mM for NMR studies) 1M7 in the reaction mixture. For DMS, suitable amounts of DMS were dissolved in 90 µL of anhydrous ethanol and diluted with 870 µL of water. Aliquots of this solution were then added to 200 µM protein solutions for a final concentration of 20 mM DMS. The mixtures were incubated for 15 minutes at room temperature prior to use.

NMR Analysis

To eliminate contaminations by side products of the modification reactions, the modified proteins were loaded onto a 3 kDa cutoff spin column. Four subsequent washings with fresh NMR buffer (25 mM sodium phosphate, 150 mM NaCl, pH 6.5) restored the initial solute composition. After recovery, the samples were supplemented with 10% D_2O and transferred into 3 mm NMR tubes. Sample concentrations were in the range of 100-120 µM (absorption of 1M7 inhibits concentration determination of 1M7 modified proteins by UV). NMR experiments were conducted on a Bruker 800 MHz Avance NEO equipped with a cryo-probe at 298 K. The 2D ^{1}H-^{15}N correlation spectra were acquired using a SOFAST-HMQC pulse sequence (18). Spectra were processed in Topspin 3.6 and analyzed with CCPN2.5.3 (19). Protein backbone resonance assignments for hnRNP K KH3 have been published elsewhere and could be transferred to our spectra for most backbone resonances without complications (20, 21). For the

RNA titration experiments 5'-$(CCC)_5$-3' RNA oligonucleotides were supplied by Integrated DNA Technologies.

MALDI Mass Spectrometry

All mass spectrometric experiments were conducted on a Bruker Autoflex maX MALDI-TOF instrument. Aliquots of the (modified) protein samples were diluted to approximately 6 µM with water and combined with an α-Cyano-4-Hydroxycinnamic acid matrix using the dried droplet technique. Mass spectra were recorded for SHARP RRM 1 modified with 1M7 and DMS; hnRNP K KH3 modified by 1M7, DMS, NMIA, BzCN, and 2A3 as well as the unmodified proteins.

Molecular Dynamics Simulations

1M7 and DMS were separately docked to the RRM1 domain of SHARP, or the KH3 domain of hnRNP K. A structure for the RRM1 domain of SHARP was modeled using *AlphaFold2* (22), and the KH3 domain of hnRNP K was taken from the X-ray structure (PDB ID 1ZZK) (23); the GB1 solubility tag was removed. Structures for 1M7 and DMS were procured from the CSD (24); partial charges for 1M7 and DMS were assigned for 1M7 and DMS using the *REDDB* (25). The 1M7 and DMS molecules were docked to the RRM1 domain of SHARP using the flexible docking algorithm of UCSF's *Dock6* (26). For each ligand, the top four ranked poses were used.

The four systems were then prepared for simulation: (1) hnRNP K KH3 + DMS, (2) hnRNP K KH3 + 1M7, (3) SHARP RRM1 + DMS, (4) SHARP RRM1 + 1M7. The *tleap* module of *Amber22* (27) was used to solvate the system in a truncated octahedron box of 9,833 *TIP3P* (28) water molecules before neutralizing the total charge with Na+ ions; a 50 mM NaCl buffer was then added. The total system size consisted of 30,901 atoms. The protein was simulated using the *ff14sb* (29) force field, and the ions were modeled with the Joung-Cheatham monovalent ion set in TIP3P (30). The 1M7 and DMS molecules were modeled using *GAFF2* (31). The masses of non-water hydrogen atoms were repartitioned to permit a 4fs timestep in production runs (32). A distance cutoff of 10 Å was applied to nonbonded interactions, with long-range electrostatics calculated using Particle Mesh Ewald (33). A Langevin thermostat (34) was used to maintain the temperature with a collision frequency of 1 ps^{-1}. Simulations were carried out using the GPU variant of the *pmemd* module of *Amber22*.

The systems were relaxed identically using a 10-step equilibration protocol designed to prepare the systems for simulation conditions. The first step included 1,000 steps of steepest descent minimization, followed by 9,000 steps of conjugate gradient minimization with only water molecules and hydrogen atoms unrestrained and 100.0 kcal/(mol*Å^2) Cartesian positional restraints on the rest of the system. The second step involved heating from 100 K to 298.15 K over 1 ns at constant NVT, again with all atoms except hydrogen atoms and water molecules restrained with a 100 kcal/(mol*Å^2) force constant, before maintaining the temperature at 298.15 K for an additional 4 ns. The third step was 1 ns MD simulation at constant NPT with 100 kcal/(mol*Å^2) positional restraints on all atoms except hydrogen and water molecules. The fourth step was 1 ns MD simulation at constant NVT with 100 kcal/(mol*Å^2) restraints on all atoms except hydrogen and water molecules. The fifth step was 1,000 steps of conjugate gradient minimization with only the protein backbone atoms (Cα, C, N) restrained with a 10 kcal/(mol*Å^2) force constant. The sixth step was 1 ns MD simulation at constant NPT with 10 kcal/(mol*Å^2) restraints on the protein backbone, followed by another 1 ns MD at constant NPT with 1 kcal/(mol*Å^2) restraints on the protein backbone, then another 1 ns with 0.1 kcal/(mol*Å^2) restraints, followed by a final 10 ns of unrestrained MD at constant NPT. The four 1M7 and DMS molecules in each system were unrestrained during the final four equilibration steps. The final coordinates and velocities from the last relaxation step were used to seed triplicate independent MD simulations at 298.15 K for each system. Another set of simulations was conducted in an identical manner where the 1M7 molecules were placed randomly in the box instead of being docked to the protein surfaces, and the resulting binding patterns were identical to those presented here (data not shown).

Analysis was carried out using the *cpptraj* (35) module of *Amber22*. The deviation of the simulation structures from the reference structures was monitored by calculating the backbone-RMSD (Cα, C, N; ignoring the first and last 5 residues of the protein). The AlphaFold model was used as the reference structure for SHARP RRM1, and the first model of the NMR structure was used as the reference structure for hnRNP K KH3. The root-mean square fluctuation (RMSF) calculations were performed by fitting the non-terminal backbone atoms of the simulated structures to the reference structure before calculating the per-residue RMSF using all non-hydrogen atoms. The radial distribution function (RDF) of 1M7 and DMS were calculated against each region on the protein to monitor which regions of space DMS and 1M7 occupy the most during the simulation. Structures were visualized using VMD (36).

RESULTS

Amino acid side chains can react with RNA-modifying chemical probes

Several proteinogenic amino acids possess nucleophilic functional groups in their side chains (**Figure 2 and S1**). Some of these have been shown to be post-translationally modified by acylations (37) and methylations (38, 39). The chemical probing reagents 1M7 (1-methyl-7-nitro-isatoic anhydride) and DMS (dimethyl sulfate) are capable of reacting with side chain nucleophiles in a similar manner. 1M7 can potentially acylate the side chains of cysteine (Cys), and histidine (His), as well as the hydroxyl groups of tyrosine (Tyr), serine (Ser), and threonine (Thr) (**Figure 2A**). DMS can potentially methylate Cys, His, Asp, and Glu residues (**Figure 2B**). However, identifying potential modification sites and reactivities for these reagents is substantially more complex in a protein than it is for a peptide or monomer. The nucleophilicity of a side chain in the context of a protein is determined not just by the type of amino acid itself. Neighboring amino acid residues can significantly alter a side chain's nucleophilicity. The catalytic triad is a typical example of an architecture enhancing the nucleophilicity of a specific residue (40). Accessibility for reagents is another major factor to consider when dealing with side chain modifications. Thus, the three dimensional structure of a protein will have an influence on the reactivity of a specific residue when presented with a modifying reagent. In this sense proteins may behave similarly to nucleic acids when presented with chemical probing reagents.

MD simulations reveal 1M7 and DMS occupancy near nucleophilic amino acids

We used MD simulations to qualitatively assess the binding affinity of 1M7 and DMS with two well-known RNA binding proteins: the RNA recognition motif 1 (RRM1) domain of SMRT/HDAC1 Associated Repressor Protein (SHARP) and the KH3 domain of the heterogeneous nuclear ribonucleoprotein K (hnRNP K). SHARP possesses four RNA recognition motifs (RRMs) that have been demonstrated to interact with several RNAs including Xist and SRA (41–43). Xist is responsible for transcriptionally silencing the inactive X chromosome in female placental mammals (44). hnRNP K possesses three RNA binding KH domains; hnRNP K has several cellular roles, including regulation of dosage compensation through an interaction with XIST, and is also implicated in splicing regulation (45, 46).

Both RRM1 of SHARP and the KH3 domain of hnRNP K possess surface and buried amino acids that could potentially interact with 1M7 and DMS. We used MD simulations to determine

where 1M7 and DMS localize on RRM1 of SHARP or the KH3 domain of hnRNP K. Three 1.0 μs MD simulations of the RRM1 domain of SHARP, or the KH3 domain of hnRNP K, were performed with either 1M7 or DMS placed in the simulation box using the *ff14SB* force field in *TIP3P* water, to give a total of four different systems (SHARP + 1M7; SHARP + DMS; hnRNP K + 1M7; hnRNP K + DMS). To ensure the protein structures were stable in the simulations with 1M7 or DMS, we monitored the RMSF of the protein backbone (**Figure S2**). The proteins do not deviate significantly from the reference structures in our simulations, and exhibit the greatest flexibility in the N- and C-terminal residues. The simulations of SHARP RRM1 exhibit additional flexibility in the loop region connecting β-strands 2 and 3 (residues Pro39 - A48 in **Figure S2B**). We conclude that both SHARP and hnRNP K are stable in our MD simulations.

To understand which amino acids the 1M7 or DMS molecules might associate with on the two proteins – and by extension, which amino acids may be accessible for covalent modification by the chemical probes – we monitored the occupancy of the two chemical probes around the proteins throughout the MD simulations. We calculated the radial distribution function (RDF) of 1M7 or DMS against potentially active amino acids, which measures the population of 1M7 or DMS at each distance from the designated amino acids. For both SHARP RRM1 and the KH3 domain of hnRNP K, the 1M7 molecules rapidly associate with aromatic residues (Phe, Tyr, Trp, His) on the protein surfaces. In the simulations of SHARP RRM1 with 1M7, the 1M7 molecules form aromatic stacking interactions with His7, Tyr29, His63, and Tyr78 (**Figure 3A**), which are all potentially reactive with 1M7, and may facilitate an interaction *in vitro*. The KH3 domain of hnRNP K possesses fewer surface aromatic residues than the RRM1 domain of SHARP. Consequently, in the MD simulations of hnRNP K KH3 with 1M7, the 1M7 molecules localize near Tyr75 and Tyr84 residues, as well as solvent-exposed residues on helix-3 such as Gln71, Leu76, and Asn79 (**Figure 3B**). The interactions between 1M7 and the KH3 domain of hnRNP K bring 1M7 adjacent to Glu33 and Ser71. In both sets of simulations containing 1M7, with either SHARP or hnRNP K, the 1M7 molecules associate near potentially catalytically active amino acids.

DMS is smaller than 1M7, is not aromatic, and does not appear to associate with the protein surfaces as rapidly or with as long of a lifetime as 1M7 does in our MD simulations. Instead, DMS is observed to transiently associate with polar or aliphatic residues on the protein surfaces. In the simulations of SHARP RRM1 with DMS, the DMS molecules localize near Tyr29, Asn64, and Asp77, which brings them adjacent to the potentially reactive amino acids His7, His25, and

His63 (**Figure 3C**). In the simulations of the KH3 domain of hnRNP K, the DMS molecules congregate near polar and aliphatic residues: the DMS molecules associate with Gln38, Leu76, and Gln78, bringing them adjacent to His41 and Tyr75 (**Figure 3D**). In simulations of both KH3 of hnRNP K and RRM1 of SHARP, the DMS molecules appear to transiently associate with aliphatic or polar amino acids, which then places them proximal to several His residues.

The two probing reagents, 1M7 and DMS, differ in how they interact with the protein surface in the MD simulations. 1M7 is larger than DMS, and appears to favor interactions with aromatic amino acids, participating in relatively longer binding events (~100 ns) with the protein surface. As a result, 1M7 possesses several distinct high-occupancy binding sites on the protein surface. Conversely, DMS prefers interactions with aliphatic amino acids, and associates and dissociates from the protein surface much more rapidly (~25 ns). As a result, the DMS molecules exhibit a more diffuse occupancy cloud around the protein than 1M7. Although the simulations cannot describe the reactivity of the chemical probes with protein amino acids, the MD simulations suggest that the chemical probing reagents may form favorable non-covalent interactions with the protein surface, bringing them close enough to potentially trigger methylation (DMS) or acylation (1M7) with reactive amino acids.

MALDI-MS reveals covalent modification of proteins

Inspired by our findings from MD simulations, we used MALDI-TOF (matrix assisted laser desorption and ionization-time of flight) mass spectrometry to investigate the potential covalent modification of the hnRNP K KH3 RNA binding domain and the SHARP RRM1 proteins by the chemical probing reagents. To this end, we treated both proteins with 18 mM 1M7 and 20 mM DMS respectively and compared their molecular weight with unmodified proteins. These concentrations mimic the concentrations used in chemical probing experiments. The resulting mass spectra (**Figure 4**) clearly show multiple species with increased molecular weights for both 1M7 modified proteins. This indicates that proteins are indeed covalently modified by 1M7. We are confident that the observed increase in molecular weight is not a result of non-covalent adduct formation since adding the 1M7 hydrolysis product to hnRNP K KH3 did not produce the same pattern in the mass spectrum (**Figure S5**). Considering the resolution of MALDI, the observed differences correspond to the acetylations by the 1M7 chemical probe. The m/z pattern suggests a variety of modification patterns present in the sample. This corresponds to a complex mixture of singly, doubly and, multiply modified proteins. The overall modification rate

is highly significant given that the unmodified protein accounts only for a minor fraction in the mass spectra. In stark contrast, the observed m/z is not changed significantly by DMS for both proteins in this study. For hnRNP K KH3, DMS caused broadening of the m/z peak, likely indicating modification, however, the limited resolution and sensitivity of the MALDI-TOF approach in combination with our protocol may fail to detect the minor change in molecular weight caused by single modifications. To address the inconclusiveness of the results for DMS experiments and to further investigate the modification sites we shifted our focus towards NMR spectroscopic studies as an orthogonal approach. Since both proteins yielded similar mass spectrometric results it was decided to focus on hnRNP K KH3 for the further investigation by NMR.

NMR spectroscopy sheds light on modification sites in hnRNP K KH3

We used NMR to experimentally validate the interaction between 1M7 and DMS with the hnRNP K KH3 protein. As a control, we first carried out a 2D ^{1}H ^{15}N HMQC on the hnRNP K KH3 protein and observed good dispersion of amide resonance peaks, indicating a properly folded protein. We then titrated a 15 nucleotide long polyC RNA oligonucleotide into the hnRNP K KH3 domain and observed line broadening and chemical shift perturbations (CSP) for several residues (**Figure 5A**).

Modification of the protein with 1M7 resulted in severe changes in the observed NMR spectra. While the presence of several distinct resonances and overall good amide backbone distribution indicated the presence of the folded protein, almost all resonances were affected by the modification. Most backbone resonances experience some degree of line broadening or CSP; some are broadened to undetectability or escape assignment by severe chemical shift perturbations. We mapped the most severely affected resonances on the published crystal structure of hnRNP K KH3 to elucidate potential modification sites (**Figure 5D and S3**) (23). As shown, at least some of the residues most affected by 1M7 cluster in a defined region of the protein suggesting a degree of systematic modification by the 1M7 chemical probe. One cluster in particular caught our attention. A multitude of amino acid residues (Arg26, Gly30, Gln34, Lys37, Gln38, Val81, Gln83, Ser85) strongly affected by 1M7 are found in close proximity to Arg35 and Tyr84. Our simulations already suggested a high binding affinity of the 1M7 reagent to this site owed to the aromatic nature of the tyrosine side chain. Interestingly, the crystal structure shows a stacking interaction between the Tyr84 and Arg35 side chain. Considering the

reaction mechanism by which 1M7 modifies biomolecules, this tyrosine-arginine motif should be highly susceptible to 1M7. We speculate that the positively charged arginine side chain stabilizes a transiently deprotonated tyrosine. The resulting phenolic anion is highly nucleophilic and can attack 1M7. Finally, the arginine side chain can provide the required protons for charge balancing. Catalysts promoting deprotonation have been shown to enhance the reactivity of a target RNA residues for 1M7 in a similar manner (11). Our simulations also suggested a high affinity of 1M7 to Tyr75. While Tyr75 is not part of the most severely affected residues it still shows the presence of multiple states indicating potential modifications. Due to the complex mixture of different modification states employing simple 2D ^{1}H ^{15}N backbone correlation spectra alone is not sufficient to elucidate all protein modification sites beyond doubt. However, we believe that this evidence is more than sufficient to confirm the general susceptibility of proteins to 1M7.

Treatment with DMS results in modest line broadening occurring among several residues (**Figure 5C**). Notably, our computational data pointed towards a more transient population of these binding sites. The reactivity of DMS, in the context of RNA chemical probing, is dependent on the pH of the reaction. At pH 8.0, all four nucleotides are susceptible to DMS (47). This inspired us to repeat the DMS reaction at pH 8.0 and compare it to a control conducted at pH 6.5 by MALDI-TOF (**Figure S8**). In these experiments we observed a second peak with a molecular weight difference of 13.2 Da at both pH values. This corresponds roughly to the molecular weight difference introduced by a single methyl group. The broadening of the m/z profiles observed in our experiments indicate methylations caused by DMS; a more resolved mass spectrometry experiment may be required to quantify the methylation.

What was most notable from our results is that a significant overlap exists between residues involved in the binding of polyC RNA to hnRNP K KH3 and residues bound by the 1M7 chemical probe. This suggests that such modifications can potentially impede important binding interactions (**Figure 1D, E**), thus reducing affinity. To investigate this, we turned our attention to binding assays.

A new tool: chemically modified protein alters RNA:protein binding affinity

To understand whether the covalent modifications of protein amino acids by the chemical probes can alter the binding affinity of RBPs for their target RNA, we conducted electrophoretic mobility shift assays (EMSAs) monitoring the protein-RNA interaction in triplicate. After

incubating the KH3 domain of hnRNP K with either 1M7 or DMS, different concentrations of chemically probed protein were added to the polyC RNA. The dissociation constants (K_d) were determined for the unmodified, 1M7-modified, and DMS-modified protein binding events.

The unmodified KH3 domain of hnRNP K binds the polyC RNA with a K_d of 0.10 µM (+/- 0.02 µM; **Figure 6A**). The K_d for the 1M7-modified protein is 0.15 µM (+/- 0.03 µM; **Figure 6B**). Though the dissociation constant for 1M7 is within the overlap of the standard deviations for the unmodified K_d, the data still suggests that the 1M7-induced covalent modifications can affect RNA-binding. Interestingly, the K_d for the DMS-modified KH3 domain of hnRNP K is 0.05 µM (**Figure 6C**) – lower than the unmodified (0.10 µM) or 1M7-modified (0.15 µM) binding affinity. However, the fitted curves for the DMS-modified EMSAs are less congruent with the data points than the unmodified or 1M7-modified EMSAs, reflected by the poor R^2 value for DMS (**Figure 6C**), which may obscure accurate deduction of the DMS-modified binding affinity.

The binding affinities for the unmodified, 1M7-modified, and DMS-modified KH3 domains of hnRNP K suggest that the covalent modifications are capable of altering protein-RNA binding affinities – potentially dampening or enhancing the interaction. This could be instrumental as an alternative to mutating amino acids that are identified and anticipated to facilitate RNA:protein interactions, particularly those that possess nucleophilic functional groups in the side chains (**Figure 2**).

Other chemical probes can also modify amino acids

Encouraged by the results we extended our study and screened other chemical probing reagents and their effect on hnRNP K KH3 by MALDI mass spectrometry (**Figure S7**). These include benzoyl cyanide (BzCN), N-methylisatoic anhydride (NMIA), and (2-Aminopyridin-3-yl)(1*H*-imidazol-1-yl)methanone (2A3). Covalent modification of hnRNP K KH3 could be observed for NMIA and 2A3 with the latter proving especially potent for protein acylation. This is an interesting observation since it suggests that the chemical nature of the chemical probe is more important for its ability of protein modification than its reactivity in the context of RNA SHAPE reactions. BzCN is one of the most reactive probes with the shortest half-life (ms vs min) used in SHAPE but does not modify the protein (48, 49). On the other hand, the long-lived 2A3 reagent does inflict chemical modifications on hnRNP K KH3. For 2A3 we found a pronounced concentration dependence of the modification rate. At higher concentrations of 2A3

only highly modified protein species are detected by MALDI MS (**Figure S7**). Thus, the ability of a chemical probe to modify the protein relates to an intricate balance between sufficient chemical reactivity towards amino acid side chains and a half-life long enough to access reactive sites in the protein. This observation can be used to guide the selection of chemical probing reagents to be used for the investigation of RNA protein complexes.

Remedies to avoid protein modification

It is important to note that potential remedies to this issue may be straightforward but may present additional complications to the investigation of RNA-protein interaction using chemical probing. In our study, the use of DTT in protein buffers - even if only partially - mitigated the modification of hnRNP K KH3 by 1M7 (**Figure S4**). However, the reason for this is the quenching of the chemical probing reaction by DTT (*i.e.* acetylation of DTT instead of the protein). Consequently, DTT does not specifically inhibit or reverse the reaction of the probing reagents with proteins but with all susceptible biomolecules, thus also impeding the desired reactivity. This highlights the care that has to be taken when designing the experimental setup of chemical probing experiments on RNA-protein complexes.

DISCUSSION

An interesting aspect of our study is the fact that our simulations suggest both 1M7 and DMS favorably interact with the surface amino acids of proteins hnRNP K KH3 and SHARP RRM1. The computational data points towards a more transient binding of DMS to the proteins, which suggests DMS may be less reactive to protein amino acids in general relative to 1M7, NMIA, or 2A3. The lower reactivity of DMS for protein amino acids suggested by the MD simulations is supported by the MALDI-MS and NMR experiments, where DMS does not appear to significantly change the m/z pattern or the NMR spectra. In extension of this idea we believe the non-covalent binding of the chemical probes to the protein surface influences the preferred modification sites, but is not the primary driving force behind the chemical modification reaction – *i.e.* not all of the binding events lead to chemical modification. Only if a preferred binding site is located at or near a protein motif prone to acetylation, covalent modification reactions will occur. This explains why some protein residues are shown to have a binding affinity for chemical probes in our simulations but do not experience severe perturbation in NMR spectra upon treatment with chemical probes. Thus, we speculate that by refining our process chemical probes could potentially be used as simple means to reveal nucleophilically reactive and

accessible amino acid side chains in proteins. This technique would, similar to chemical probing of RNA, involve treating a protein of interest with varying concentrations of different reagents. The introduced modifications can then be identified by a method of choice such as high resolution MS/MS approaches yielding the reactive hotspots within the protein.

While not observed in this study, the possibility of non-covalent binding of chemical probing reagents and their hydrolysis products cannot be ruled out based on our data. The potential impact of non-covalent binding largely depends on the chemical nature of the compounds. However, we expect the latter to be more influential since the supply of chemical probe will deplete over time. DMS hydrolysis yields sulfuric acid and methanol. Both compounds, given the right choice of buffer, should not impede binding events between proteins and nucleic acids. 1M7 forms 2-(methylamino)-4-nitrobenzoic acid when reacting with water. This small molecule still possesses a variety of functional groups potentially interacting with the protein. Thus, we believe that non-covalent binding of small molecule side products of a SHAPE reaction can still be relevant in some cases. However, this effect might prove highly elusive to investigation.

The experimental design of chemical probing targeting RNA-protein complexes generally involves the formation of the RNA-protein complex of interest before treatment with chemical probing reagents. While this might look like a potential remedy for the covalent modification of the binding interface on the surface level, we have reasons to believe otherwise. Exposed protein side chains will still be prone to modification. Thus, the actual concentration of chemical probes available for the desired RNA modification will be altered. This can lead to unexpected reactivity profiles and will impede the reproducibility and comparability of the results. Moreover, the initial argument is based on the premise that the complex is highly stable. However, every complex is in a dynamic equilibrium involving constant association and dissociation events. If the half life of unbound protein is sufficient it can still be modified by the chemical probing reagent potentially causing a decrease of affinity over time, or obscuring accurate deduction of binding affinities in general.

While our results clearly show how chemical probing reagents can covalently modify protein side chains, they can also be used to improve chemical probing experiments on RNA-protein complexes. Our data indicates that BzCN does not introduce undesired modifications. Thus, this reagent is well-suited for the investigation of RNA-protein interactions. The selection of a suitable probing reagent for RNA-protein complexes should always take potential interactions of

the probe with the protein surface into consideration. Chemical probes known to readily react with protein side (e.g. 1M7, 2A3) chains should be avoided. In addition, the modification rate will be dependent on the concentration of probing reagent (**Figure S7**). Thus, keeping the concentration of the chemical probes as low as possible will reduce the rate of undesired side reactions. We suggest that each chemical probing reagent is tested for protein side chain interactions in the future before use in RNA-protein complexes.

In summary, we demonstrate that chemical probes that are used to identify single-stranded and flexible nucleotides can also interact with amino acids. Chemical probes such as 1M7 and DMS can attack their respective functional groups in the millisecond to second time range. Many protein RNA interactions occur over timescales from sub seconds to hours (50). Considering that chemical probing protein mapping experiments are carried out by adding protein to the RNA, followed by the addition of the chemical probe, this allows ample time for the probe to interact with the protein, inhibiting its binding. Critically, the binding of a chemical probe to amino acid residues that are known to facilitate interactions with RNA (e.g. tyrosine) can diminish the binding interaction. On a similar note, RNA-binding molecules (e.g. therapeutic agents) can also be modified by a chemical probe, and its interaction with its RNA target inhibited. As has been observed in multiple studies, even at sites of ligand binding, nucleotides can still possess chemical probe reactivity (51–54). Our results confirm our hypothesis that RNA chemical probing reagents may covalently modify protein amino acids, and that the resulting covalent modification of protein amino acids can perturb protein-RNA interactions. As a silver lining, RNA-modifying probes could serve as a useful tool to identify surface residues of proteins and could potentially be used as an alternative to introducing mutations to amino acids on the RNA-binding interface. Future studies will investigate the factors that govern the reactivity of protein amino acids with chemical probing reagents, and the mechanisms by which they perturb protein-RNA interactions.

DATA AVAILABILITY

Data available on request.

SUPPLEMENTARY DATA

Supplementary Data are available.

ACKNOWLEDGEMENTS

The authors would like to thank Emma Gogarnoiu for assistance with figures, Alex Nazzaro for assistance with MALDI experiments, and members of the Jones lab for thoughtful comments and suggestions.

FUNDING

This work was supported by the National Science Foundation (2243667 to A.N.J.).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Vicens,Q. and Kieft,J.S. (2022) Thoughts on how to think (and talk) about RNA structure. *Proc. Natl. Acad. Sci. U. S. A.*, **119**, e2112677119.

2. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.

3. Peattie,D.A. and Gilbert,W. (1980) Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A.*, **77**, 4679–4682.

4. Peattie,D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U. S. A.*, **76**, 1760–1764.

5. Spitale,R.C., Crisalli,P., Flynn,R.A., Torre,E.A., Kool,E.T. and Chang,H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.

6. De Bisschop,G., Allouche,D., Frezza,E., Masquida,B., Ponty,Y., Will,S. and Sargueil,B. (2021) Progress toward SHAPE Constrained Computational Prediction of Tertiary Interactions in RNA Structure. *Noncoding RNA*, **7**.

7. Siegfried,N.A., Busan,S., Rice,G.M., Nelson,J.A.E. and Weeks,K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.

8. Smola,M.J. and Weeks,K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.*, **13**, 1181–1195.

9. Smola,M.J., Calabrese,J.M. and Weeks,K.M. (2015) Detection of RNA–Protein Interactions in Living Cells with SHAPE. *Biochemistry*, **54**, 6867–6875.

10. Jones,A.N., Pisignano,G., Pavelitz,T., White,J., Kinisu,M., Forino,N., Albin,D. and Varani,G. (2020) An evolutionarily conserved RNA structure in the functional core of the lincRNA Cyrano. *RNA*, **26**, 1234–1246.

11. McGinnis,J.L., Dunkle,J.A., Cate,J.H.D. and Weeks,K.M. (2012) The mechanisms of RNA

SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.

12. Zubradt,M., Gupta,P., Persad,S., Lambowitz,A.M., Weissman,J.S. and Rouskin,S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, **14**, 75–82.

13. Jones,A.N., Graß,C., Meininger,I., Geerlof,A., Klostermann,M., Zarnack,K., Krappmann,D. and Sattler,M. (2022) Modulation of pre-mRNA structure by hnRNP proteins regulates alternative splicing of. *Sci Adv*, **8**, eabp9153.

14. Luo,L., Chiu,L.-Y., Sugarman,A., Gupta,P., Rouskin,S. and Tolbert,B.S. (2021) HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions. *J. Mol. Biol.*, **433**, 166885.

15. Martin,S., Blankenship,C., Rausch,J.W. and Sztuba-Solinska,J. (2019) Using SHAPE-MaP to probe small molecule-RNA interactions. *Methods*, **167**, 105–116.

16. Wang,Y., Parmar,S., Schneekloth,J.S. and Tiwary,P. (2022) Interrogating RNA-Small Molecule Interactions with Structure Probing and Artificial Intelligence-Augmented Molecular Simulations. *ACS Cent Sci*, **8**, 741–748.

17. Banijamali,E., Baronti,L., Becker,W., Sajkowska-Kozielewicz,J.J., Huang,T., Palka,C., Kosek,D., Sweetapple,L., Müller,J., Stone,M.D., *et al.* (2023) RNA:RNA interaction in ternary complexes resolved by chemical probing. *RNA*, **29**, 317–329.

18. Schanda,P. and Brutscher,B. (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.*, **127**, 8014–8015.

19. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.

20. Baber,J.L., Libutti,D., Levens,D. and Tjandra,N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.*, **289**, 949–962.

21. Baber,J., Libutti,D., Levens,D. and Tjandra,N. (1999) C-TERMINAL KH DOMAIN OF HNRNP K (KH3). 10.13018/BMR4405.

22. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

23. Backe,P.H., Messias,A.C., Ravelli,R.B.G., Sattler,M. and Cusack,S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure*, **13**, 1055–1067.

24. Groom,C.R., Bruno,I.J., Lightfoot,M.P. and Ward,S.C. (2016) The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater*, **72**, 171–179.

25. Dupradeau,F.Y., Cezard,C. and Lelong,R. (2007) RE DD. B.: a database for RESP and ESP atomic charges, and force field libraries. *Nucleic acids*.

26. Allen,W.J., Balius,T.E., Mukherjee,S., Brozell,S.R., Moustakas,D.T., Lang,P.T., Case,D.A., Kuntz,I.D. and Rizzo,R.C. (2015) DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.*, **36**, 1132–1156.

27. Case,D.A., Metin Aktulga,H., Belfon,K., Ben-Shalom,I.Y., Berryman,J.T., Brozell,S.R., Cerutti,D.S., Cheatham,T.E.,III, Andrés Cisneros,G., Cruzeiro,V.W.D., *et al.* (2023) Amber 2023 University of California, San Francisco.

28. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

29. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.

30. Joung,I.S. and Cheatham,T.E.,3rd (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.

31. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

32. Hopkins,C.W., Le Grand,S., Walker,R.C. and Roitberg,A.E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.*, **11**, 1864–1874.

33. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

34. Davidchack,R.L., Handel,R. and Tretyakov,M.V. (2009) Langevin thermostat for rigid body dynamics. *J. Chem. Phys.*, **130**, 234101.

35. Roe,D.R. and Cheatham,T.E.,3rd (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, **9**, 3084–3095.

36. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–8, 27–8.

37. Shang,S., Liu,J. and Hua,F. (2022) Protein acylation: mechanisms, biological functions and therapeutic targets. *Signal Transduct Target Ther*, **7**, 396.

38. Wang,X., Xie,H., Guo,Q., Cao,D., Ru,W., Zhao,S., Zhu,Z., Zhang,J., Pan,W., Yao,X., *et al.* (2023) Molecular basis for METTL9-mediated N1-histidine methylation. *Cell Discov*, **9**, 38.

39. Murn,J. and Shi,Y. (2017) The winding path of protein methylation research: milestones and new frontiers. *Nat. Rev. Mol. Cell Biol.*, **18**, 517–527.

40. Dodson,G. and Wlodawer,A. (1998) Catalytic triads and their relatives. *Trends Biochem. Sci.*, **23**, 347–352.

41. Arieti,F., Gabus,C., Tambalo,M., Huet,T., Round,A. and Thore,S. (2014) The crystal

structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Res.*, **42**, 6742–6752.

42. Monfort,A., Di Minin,G., Postlmayr,A., Freimann,R., Arieti,F., Thore,S. and Wutz,A. (2015) Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Rep.*, **12**, 554–561.

43. Button,A.C., Hall,S.D., Ashley,E.L. and McHugh,C.A. (2023) Dissection of protein and RNA regions required for SPEN binding to XIST A-repeat RNA. *RNA*, 10.1261/rna.079713.123.

44. Brown,C.J., Hendrich,B.D., Rupert,J.L., Lafrenière,R.G., Xing,Y., Lawrence,J. and Willard,H.F. (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, **71**, 527–542.

45. Pintacuda,G., Wei,G., Roustan,C., Kirmizitas,B.A., Solcan,N., Cerase,A., Castello,A., Mohammed,S., Moindrot,B., Nesterova,T.B., *et al.* (2017) hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell*, **68**, 955–969.e10.

46. Venables,J.P., Koh,C.-S., Froehlich,U., Lapointe,E., Couture,S., Inkel,L., Bramard,A., Paquet,E.R., Watier,V., Durand,M., *et al.* (2008) Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell. Biol.*, **28**, 6033–6043.

47. Mustoe,A.M., Lama,N.N., Irving,P.S., Olson,S.W. and Weeks,K.M. (2019) RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 24574–24582.

48. Mortimer,S.A. and Weeks,K.M. (2008) Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **130**, 16178–16180.

49. Mortimer,S.A. and Weeks,K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.*, **4**, 1413–1421.

50. Licatalosi,D.D., Ye,X. and Jankowsky,E. (2020) Approaches for measuring the dynamics of RNA-protein interactions. *Wiley Interdiscip. Rev. RNA*, **11**, e1565.

51. Stoddard,C.D., Montange,R.K., Hennelly,S.P., Rambo,R.P., Sanbonmatsu,K.Y. and Batey,R.T. (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure*, **18**, 787–797.

52. Garst,A.D., Héroux,A., Rambo,R.P. and Batey,R.T. (2008) Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.*, **283**, 22347–22351.

53. Tyrrell,J., McGinnis,J.L., Weeks,K.M. and Pielak,G.J. (2013) The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry*, **52**, 8777–8785.

54. Watters,K.E., Strobel,E.J., Yu,A.M., Lis,J.T. and Lucks,J.B. (2016) Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.*, **23**, 1124–1131.

TABLE AND FIGURES LEGENDS

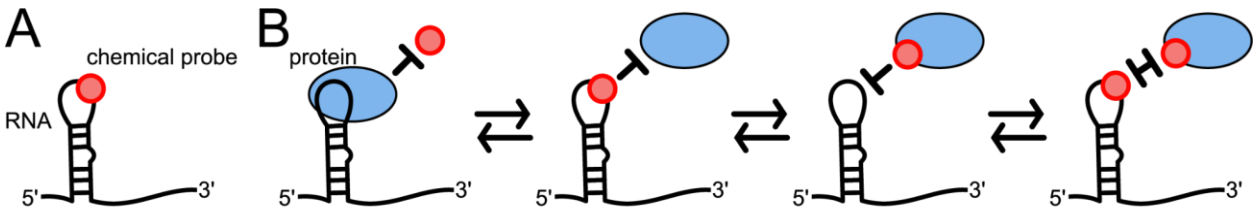


Figure 1. Mechanisms of chemical probing experiments. **A)** Chemical probes modify flexible or single-stranded nucleotides. When protein is present in solution, **B)** the protein can prevent binding by the probe, **C)** the probe can prevent binding of the protein, **D)** the probe can bind to the protein to prevent an interaction, and **E)** the interaction site can be doubly modified, preventing an interaction.

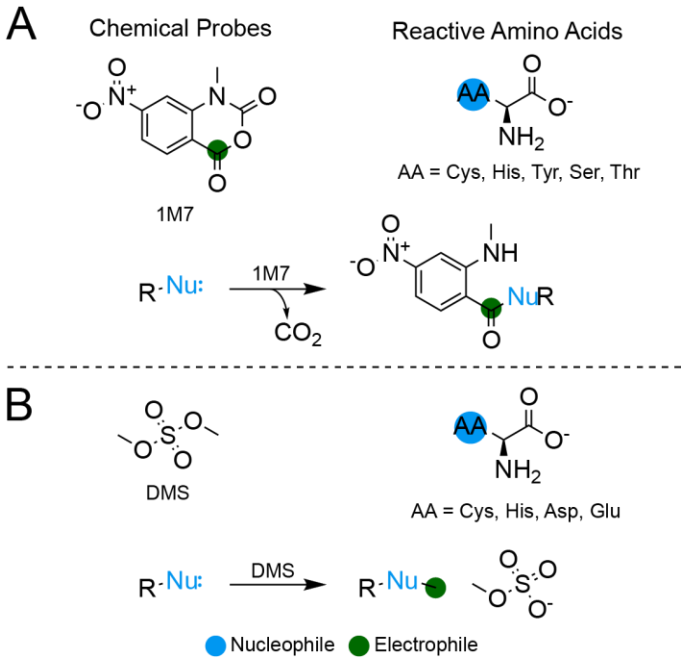


Figure 2. Nucleophilic amino acids that react with RNA-modifying probes. **A)** 1M7 acylates Cys, His, Tyr, Ser, and Thr residues. **B)** DMS methylates Cys, His, Asp, and Glu residues.

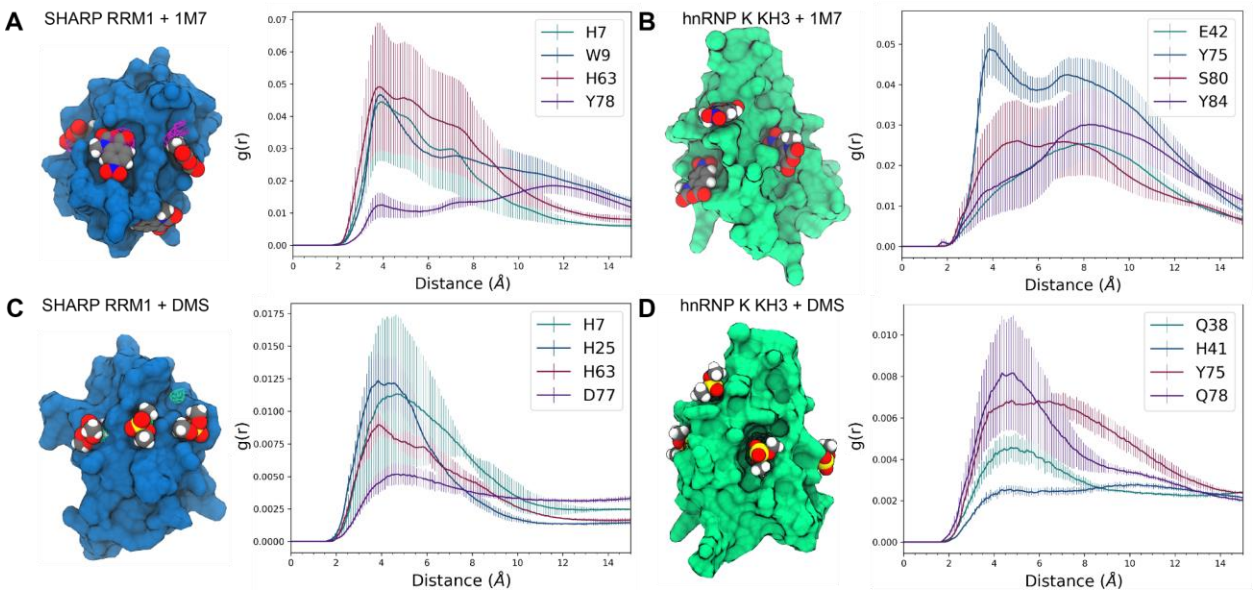


Figure 3. Analysis of the MD simulations of the RRM domain of SHARP with 1M7 and DMS. **A)** Simulation structure of the RRM1 domain of SHARP (blue) with 1M7, and contours representing regions of high 1M7 density (magenta). **B)** Simulation structure of the third KH domain of hnRNP K (green) with 1M7, and contours representing the regions of high 1M7 density (blue). **C)** Simulation structure of the RRM1 domain of SHARP (blue) with DMS, and contours representing regions of high DMS density (green). **D)** Simulation structure of the third KH domain of hnRNP K (green) and contours representing regions of high DMS density (blue). The plots represent the RDF of DMS or 1M7 to specific amino acids, averaged across the three simulation trials and reported with standard deviations.

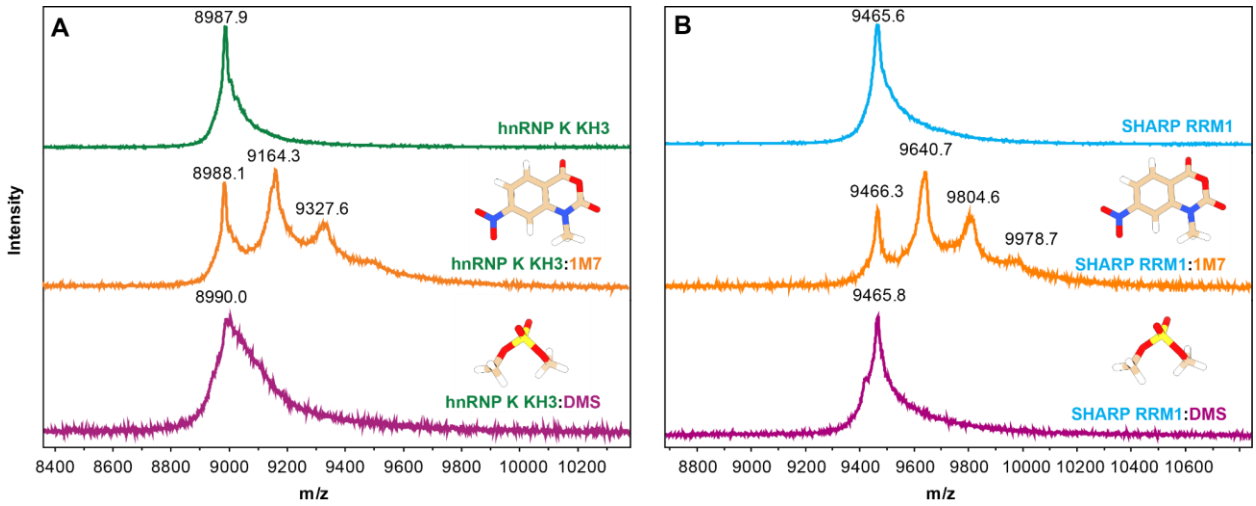


Figure 4. Mass spectrometric investigations of **A)** hnRNP K KH3 and **B)** SHARP RRM1

unmodified (top) and treated with 1M7 (middle) and DMS (bottom) chemical probing reagents respectively.

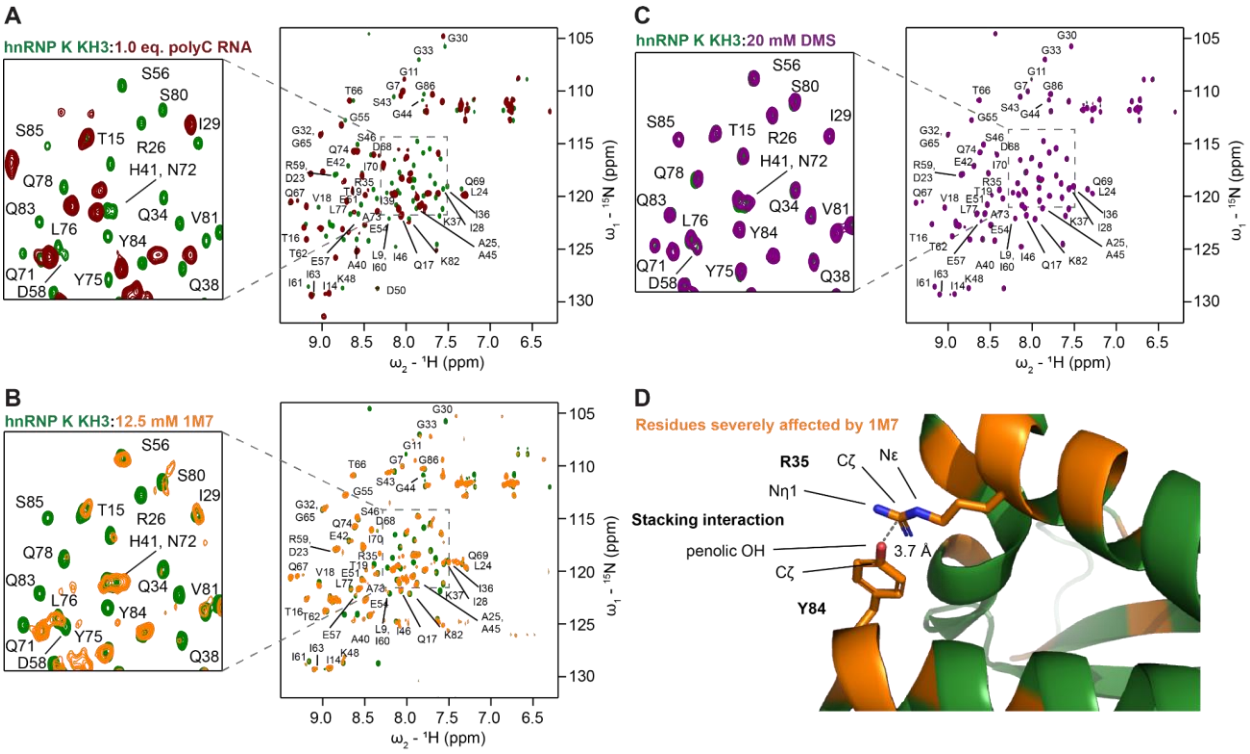


Figure 5. Backbone HMQC spectra recorded at 298K demonstrate binding of hnRNP K KH3 by RNA and interactions with chemical probes. Assignments refer to apo protein since CSPs and line broadening impede transfer of the assignments to bound/modified protein for many residues. Overlays show hnRNP K KH3 backbone spectrum (green) upon addition of **A)** 1.0 eq. of 15 nt polyC RNA, **B)** 12.5 mM 1M7, and **C)** 20 mM DMS. **D)** Residues severely affected by 1M7 (orange) mapped the X-ray structure of hnRNP K KH3. A stacking interaction between Arg35 and Tyr84 potentially catalyzes the reaction. The full structure is shown in **Figure S3**.

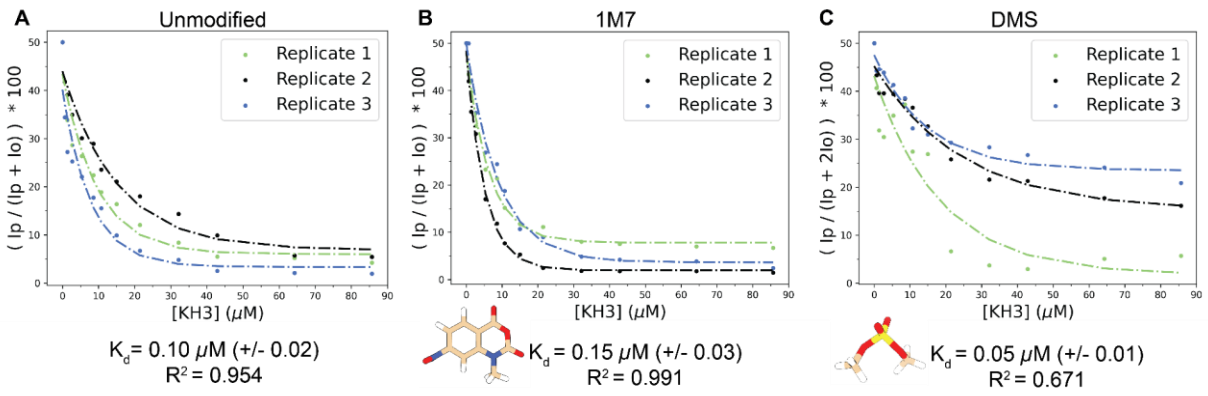


Figure 6. EMSAs measuring the binding affinities between the polyC RNA and **A)** unmodified, **B)** 1M7-modified, and **C)** DMS-modified KH3 domain of hnRNP K. The protein concentrations were changed while the RNA concentration was held fixed, as indicated in the figure.