

## **Intrinsically disordered interaction network in an RNA chaperone revealed by native mass spectrometry**

Samantha H. Sarni<sup>1†</sup> (1), Jorjetha Roca<sup>2</sup> (1), Chen Du<sup>1</sup>, Mengxuan Jia<sup>1‡</sup>, Hantian Li<sup>2</sup>, Ana Damjanovic<sup>2</sup>, Ewelina M. Małecka<sup>2</sup>, Vicki H. Wysocki<sup>1\*</sup>, Sarah A. Woodson<sup>2\*</sup>

<sup>1</sup>Department of Chemistry and Biochemistry and Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH, USA;

<sup>2</sup>T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD. USA.

(1) Equal contribution.

\*Corresponding authors: Vicki H. Wysocki; Sarah A. Woodson.

**Email:** [wysocki.11@osu.edu](mailto:wysocki.11@osu.edu) or [swoodson@jhu.edu](mailto:swoodson@jhu.edu)

Current address: <sup>†</sup>Neoleukin Therapeutics, Inc., 188 E Blaine St, Seattle, WA 98102; <sup>‡</sup>Merck, 213 Grand Ave, South San Francisco, CA 94080

### **Author Contributions:**

SHS and MJ designed, performed and analyzed ion mobility mass spectrometry experiments. CD measured the collision cross sections and analyzed data. EMM and JR prepared RNA samples and JR prepared protein samples. HL and AD performed and analyzed molecular dynamics simulations. SHS and JR analyzed data, interpreted the results and wrote the paper. VHW and SAW conceived and directed the project. All authors helped interpret the data and edit the manuscript.

**Keywords:** RNA chaperone, small RNA, intrinsically disordered region, ion mobility mass spectrometry, surface induced dissociation.

## Abstract

RNA-binding proteins contain intrinsically disordered regions whose functions in RNA recognition are poorly understood. The RNA chaperone Hfq is a homohexamer that contains six flexible C-terminal domains (CTDs). The effect of the CTDs on Hfq's integrity and RNA binding has been challenging to study because of their sequence identity and inherent disorder. We used native mass spectrometry (nMS) coupled with surface-induced dissociation (SID) and molecular dynamics (MD) simulations to disentangle the arrangement of the CTDs and their impact on the stability of *E. coli* Hfq with and without RNA. The results show that the CTDs stabilize the Hfq hexamer through multiple interactions with the core and between CTDs. RNA binding perturbs this network of CTD interactions, destabilizing the Hfq ring. This destabilization is partially compensated by binding of RNAs that contact multiple surfaces of Hfq. By contrast, binding of short RNAs that only contact one or two subunits results in net destabilization of the complex. Together, the results show that a network of intrinsically disordered interactions integrate RNA contacts with the six subunits of Hfq. We propose that this CTD network raises the selectivity of RNA binding.

## Significance Statement

Hfq is a protein hexamer necessary for gene regulation by non-coding RNA in bacteria, during infection or under stress. In the cell, Hfq must distinguish its RNA partners from many similar nucleic acids. Mass spectrometry dissociation patterns, together with molecular dynamics simulations, showed that flexible extensions of each Hfq subunit form a dense network that interconnects the entire hexamer. This network is disrupted by RNA binding, but the lost interactions are compensated by RNAs that contact multiple Hfq subunits. By measuring interactions that are too irregular to be counted by other methods, mass spectrometry shows how flexible protein extensions help chaperones like Hfq recognize their RNA partners in the messy interior of the cell.

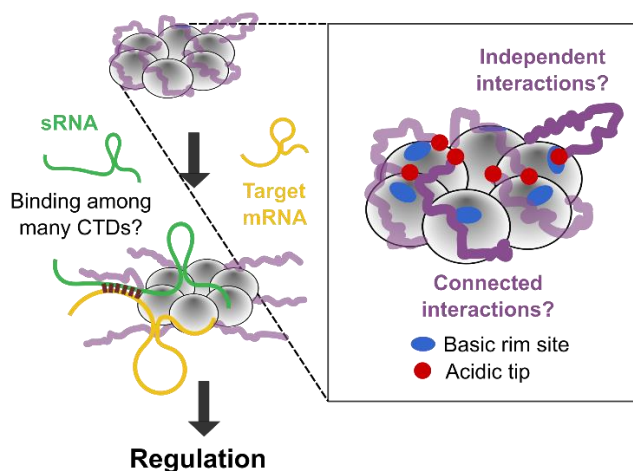
## Main Text

### Introduction

Many RNA-binding proteins (RBPs) contain intrinsically disordered regions (IDRs) (1) with overlapping functions that have been difficult to disentangle. For example, IDRs may augment specific RNA recognition, connect different RNA-binding modules, and enable the assembly of liquid condensates, while also serving as targets for post-translational modification (2–4). The heterogeneous and dynamic structures of IDRs make their interactions especially challenging to quantify, and their functions in most RNA binding proteins remain poorly understood.

Hfq is a bacterial Sm protein that binds small non-coding RNA (sRNA) and chaperones sRNA regulation of complementary mRNAs (5) (Fig. 1). Deletion of Hfq results in pleiotropic effects, including maladaptive responses to stress and decreased virulence (6). The well-folded core of the Hfq hexamer assembles into a symmetric ring that binds U and A-rich sequence motifs in sRNA and mRNA substrates (7). Conserved arginine patches on the outer rim of the hexamer also bind RNA and are essential for its chaperone activity (8, 9).

*E. coli* Hfq also has intrinsically disordered C-terminal domains (CTDs) that extend outward from the core of the hexamer (10). Each monomer containing 102 residues contributes a 37 amino acid (aa) CTD, creating a crowded zone of disordered polypeptide around the protein. This ring-shaped organization, which is unlike disordered regions in other RBPs, raises the possibility that the Hfq CTDs act together rather than individually.



**Fig. 1.** Role of Hfq's CTDs in sRNA regulation. Hfq chaperones the annealing of sRNAs with their target mRNAs, but it is not known how binding of RNAs occurs when the core of Hfq is occluded by many disordered CTDs. Although the acidic tips of the CTDs (red) can interact with basic patches on the rim (blue) (21), the organization and collective behavior of the CTDs is unknown.

Despite their disordered nature, NMR chemical shift perturbations and molecular dynamics simulations determined that the CTDs of *E. coli* Hfq interact with the rim of the hexamer (11). Additionally, unassigned electron density in a crystal

structure of Hfq bound to RydC sRNA suggested that the CTDs make distributed contacts with the protein-RNA surface (12). These results aligned with the earlier observation that the CTDs (residues 65-102) stabilize the Hfq hexamer (13) and contribute to its function (14–19). We found that semi-conserved acidic residues at the C-terminus mimic nucleic acid, competing with RNA for binding to the rim (20–22). Competition with the CTDs can result in preferential dissociation of non-specific RNA, and retention of specific RNA ligands. More recently, it was shown that the bases and the tips of the CTDs interact synergistically with particular Hfq surfaces, leading to different effects depending on the RNA ligand (23).

Because of their intrinsic disorder and sixfold symmetry, how the CTDs organize around Hfq stabilizing the hexamer is still unknown. Additionally, it is not known if each CTD acts locally and independently, or if the six CTDs act together to accommodate or displace an incoming RNA (Fig. 1). Moreover, the energetic contributions of individual CTDs to RNA binding have been almost impossible to quantify.

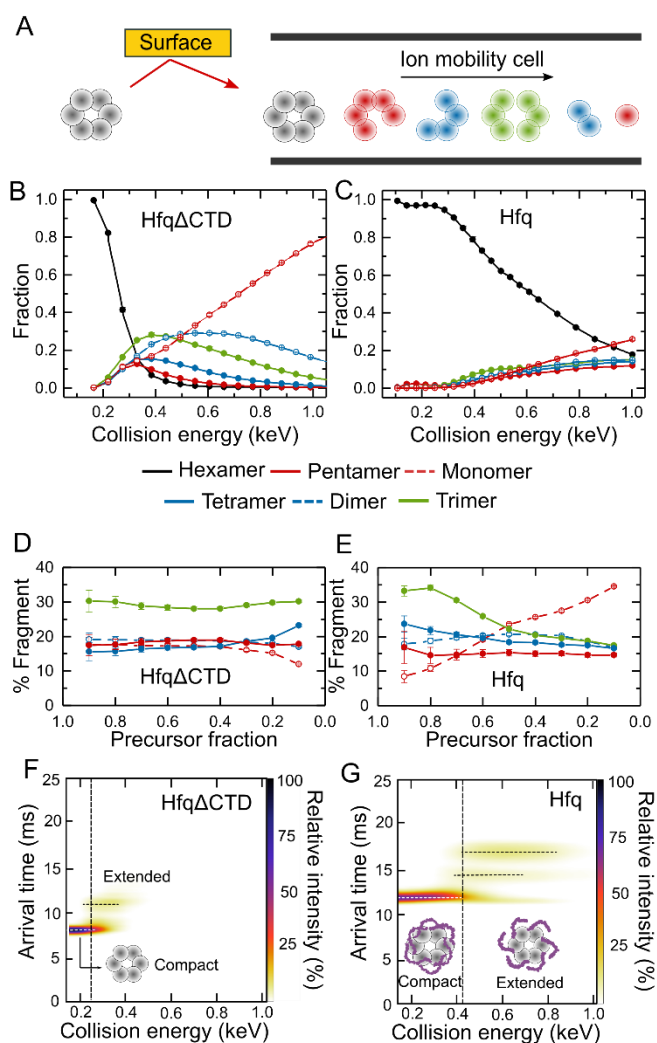
We addressed these challenges by using native mass spectrometry (nMS) coupled with surface-induced dissociation (SID) (Figs. 2A and S1). In nMS, the protein complex is exchanged into a volatile electrolyte allowing transfer of the intact native complex to the gas phase (24). After ionization, collision with a surface (nMS-SID) dissociates the complex into product ions that provide information about the stabilities of the non-covalent interfaces within the complex and their molecular organization (25). This method has been used to characterize the stability, structure and assembly pathways of many protein complexes, including RBPs and membrane proteins (26–28). Yet, despite its promise for discovery, nMS-SID studies of large biomolecular complexes typically require customized instrumentation (Fig. S1).

Here, by using nMS-SID and all-atom molecular dynamics (MD) simulations, we show that the six disordered CTDs of apo Hfq form extensive interactions that energetically connect and stabilize the entire hexamer. When RNA binds any subunit of Hfq, these stabilizing interactions are disrupted throughout the hexamer. Taken together, our results show how disordered regions can integrate RNA-protein interactions across a multi-subunit chaperone.

### Results

**Disordered CTDs stabilize Hfq.** To better understand the interactions of the flexible CTDs, we first used nMS-SID on WT *E. coli* Hfq (102 aa per subunit) and a truncated Hfq lacking the CTDs (Hfq $\Delta$ CTD; 65 aa per subunit). We analyzed the dissociation products (pentamer, tetramer, trimer, dimer, and monomer) obtained at increasing collision energies (CE) with the surface (Fig. 2A). The resulting energy-resolved mass spectra (ERMS) showed that the two proteins have different stabilities and fragmentation patterns (Figs. 2B-C). Hfq $\Delta$ CTD reached 20% fragmentation (80% precursor remaining) at ~220 eV, compared to ~390 eV for WT Hfq (Figs. 2B-C, black lines). This large

difference demonstrated that the wild-type protein is much more stable than the truncated version, in agreement with previous reports (13).



**Fig. 2.** Disordered CTDs stabilize Hfq. (A) Native mass spectrometry coupled with surface-induced dissociation (nMS-SID) dissociates the Hfq hexamer into oligomers that retain the connectivity of the native protein. Fragments are separated according to their arrival time after traversing an ion mobility cell. (B-C) Energy-resolved mass spectrum of (B) Hfq $\Delta$ CTD and (C) Hfq. The collision energies are corrected for the mass of the CTDs ( $m_{\text{Hfq}\Delta\text{CTD}}/m_{\text{Hfq}}$ ). Reported fractions are the sum of the intensities of each dissociation product normalized by the total intensity of all products. Symbols report the average of three replicates. Standard errors are smaller than the symbols for some data points. Solid lines represent a cubic interpolation of the data. (D-E) Percentage of each oligomer (pentamer, tetramer, trimer, dimer or monomer) in the dissociation products, as a function of the remaining precursor fraction for (D) Hfq $\Delta$ CTD and (E) Hfq. Errors are the spread of the ERMS curves, normalized by the total dissociated fraction and converted to a percentage. Colored as in Figs. 2B-C. Solid lines are a visual guide. (F-G) Surface-induced unfolding (SIU) plot of (F) Hfq $\Delta$ CTD and (G) Hfq. The dashed line represents the transition from compact to extended protein and corresponds to  $\sim 0.2$  and  $\sim 0.7$  precursor fraction for each protein, respectively.

The fraction of Hfq $\Delta$ CTD hexamer sharply decreased with modest increases in collision energy, with 90% fragmentation of the precursor at CE  $\sim 350$  eV. In contrast, the amount of WT Hfq dissociation increased gradually over a wide range of collision energies, and  $\sim 20\%$  of the precursor remained intact even at CE = 1000 eV. This gradual response to higher collision energies suggested that the CTDs prevent dissociation of the complex across a wide range of energies. One explanation is that the CTDs form a range of intersubunit interactions that reorganize upon activation. Additionally, the CTDs may be organized differently in each Hfq hexamer, causing dissociation over a continuum of collision energies.

**Disordered CTDs impact the connectivity of Hfq.** The dissociation patterns for Hfq $\Delta$ CTD and Hfq generated fragments in different ratios, indicating a different degree of connectivity between subunits in the two proteins (Figs. 2B-E). We compared the dissociation products of the two proteins at energies that deplete the hexamer precursor equally. Hfq $\Delta$ CTD dissociated into twice as many trimers ( $\sim 30\%$ ), as other fragments ( $\sim 15\text{--}20\%$ ) regardless of the total amount of precursor dissociated (Fig. 2D). This result suggested that the protein interfaces around the ring have a similar probability of dissociating.

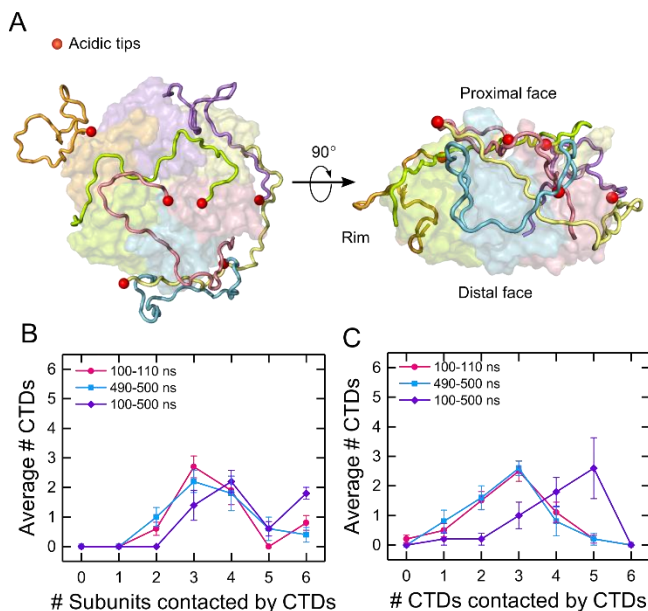
For WT Hfq, the distribution of dissociation products was markedly different from Hfq $\Delta$ CTD (Fig. 2E), indicating that the CTDs contribute to the subunit interfaces, as also reported earlier (13, 29). For all fractions of the precursor, the percentage of pentamers, tetramers and dimers remained similar ( $\sim 10\text{--}20\%$ ). However, the percentage of trimers decreased while the percentage of monomers increased as more precursor was fragmented.

Inspection of ion mobility arrival times revealed that as the collision energy increases, the WT hexamer precursor converts into two complexes that migrate more slowly in the drift chamber, suggesting partial extension or restructuring of the protein hexamer, although we can't conclude that this occurs before the subunits dissociate from one another (surface-induced unfolding (SIU) plots, Figs. 2F-G). We observed only one additional complex of Hfq $\Delta$ CTD upon activation (Fig. 2F), suggesting that one of the extended Hfq complexes comes from restructuring of the core beta sheet. The second extended form was only observed for WT Hfq and likely arises from extension of the CTDs (Fig. 2G). Altogether, the nMS results support a model in which the CTDs are bound to the core and each other in WT Hfq, stabilizing the entire hexamer. At increasing collision energies, the CTDs disentangle, exposing the core and eliminating the stabilizing inter-subunit connections. As a result, the fragile hexamer ruptures into smaller complexes.

**MD simulations reveal a network of CTD interactions on Hfq.** To gain more insights into the organization of the disordered CTDs on Hfq, we performed multiple all-atom molecular dynamics (MD) simulations on the wild-type protein. The simulations were started from 10 previous low-energy Rosetta structures (models top 1-10) (21) and an extended structure started with 4 different initial velocities (extended models 1-4). The



structures obtained after equilibration agreed with the experimentally determined collisional cross section (CCS), with the MD structures being more compact than the initial Rosetta models (Fig. S2 and Table S1). The simulated CTDs adopted a variety of conformations (Figs. 3A and S3). Thus, the diverse organization of the disordered CTDs confers heterogeneity to individual Hfq hexamers, as implied by the ERMS results (Fig. 2C).



**Fig. 3.** Disordered CTDs create a network of interactions with Hfq. (A) (Left) Top and (right) side view of an Hfq model from one molecular dynamics (MD) simulation (top 4, at 500 ns). Subunits are colored individually with each acidic C-terminus shown as a red sphere. (B-C) Average number of CTDs interacting with 0, 1, 2, 3, 4, 5 or 6 (B) subunits or (C) CTDs, for all MD models. Interactions were considered at any time during the first 100-110 ns for all models, the last 490-500 ns for models top 1-5 and for the entire simulation (100-500 ns) for models top 1-5. Symbols are the average and standard error of the mean (SEM) of the various simulations. Solid lines are a visual guide.

The Hfq models showed that each CTD may interact within its monomer and with other subunits, even across the Hfq ring (Figs. S4, 3A). To characterize this behavior, we mapped the interactions of individual CTDs with each of the subunits (core and CTDs) for all the models at different periods of the simulation: (i) after a 100 ns equilibration (100-110 ns, for all models), (ii) at the end of the simulation (490-500 ns for models top 1-5, 190-200 ns for models top 6-10), and (iii) over the entire 400 ns simulation (100-500 ns for models top 1-5) (Figs. S4, S5 and S6). We found that the CTDs interacted with their own core and adjacent subunits, but some CTDs also engaged with as many as 5 other subunits and CTDs. These contacts were not static, with multiple, interconnected interactions observed at different times.

We next asked how common it was for the CTDs to engage in these long-range interactions (Fig. S6). In all models, a CTD interacted with 2–6 subunit cores over a 10 ns timescale, and this degree of connectivity peaked at ~3 CTDs per hexamer (Figs. 3B and S6A). This distribution was similar at the beginning and end

of the simulations, indicating convergence. When averaged over the entire 400 ns simulation, the distribution shifted to larger numbers of subunits (Fig. 3B, purple), suggesting that the CTDs sample different regions of Hfq over time. The CTDs also interacted amongst themselves, with a peak of three CTDs contacting other CTDs when averaged over 10 ns periods at the beginning and the end of the simulation (Figs. 3C and S6B), and five other CTDs when considering the full 400 ns simulation (Fig. 3C, purple). Thus, the disordered and dynamic nature of the CTDs enables a far-ranging network of interactions connecting the whole Hfq hexamer.

### RNA binding stabilizes HfqΔCTD but destabilizes WT Hfq.

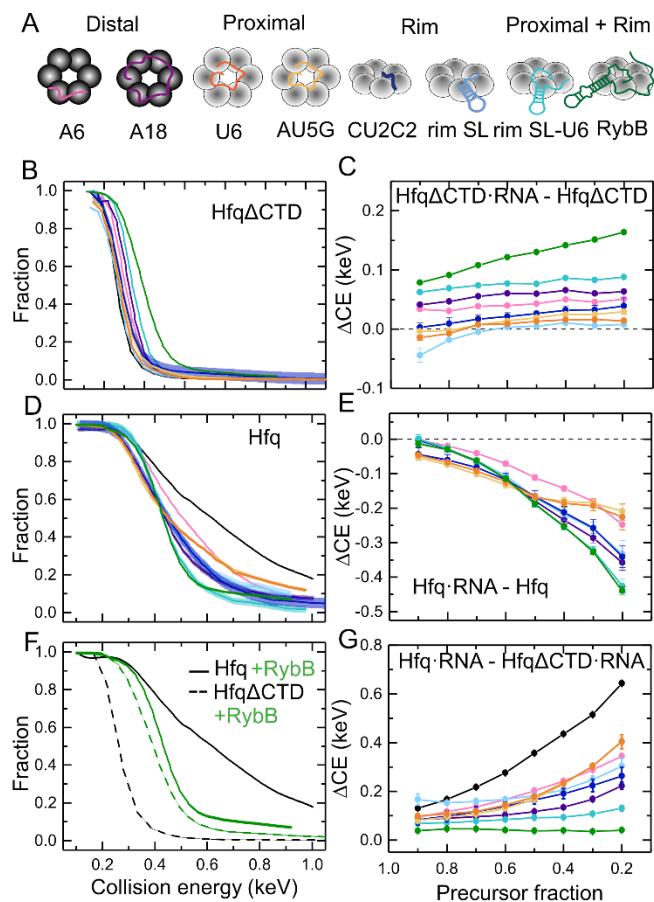
Nucleic acids compete with the CTDs for binding to Hfq's rim (21), and thus, RNA binding has the potential to alter the organization of the CTDs around Hfq. Since perturbation of the CTDs destabilizes the Hfq hexamer (Figs. 2E and 2G), the stability and fragmentation pattern of the Hfq hexamer may report on the impact of RNA binding on the protein. Although the fragmentation pattern doesn't directly reveal the RNA-protein contacts, it provides information on how the bound RNA energetically connects the Hfq subunits.

To determine whether RNA binding perturbs the structure of Hfq, we designed a series of short RNAs that mimic the Hfq binding motifs in natural sRNAs and mRNA targets of Hfq. The designed RNAs interact with different surfaces of the Hfq hexamer (Fig. 4A and S7). rA<sub>6</sub> and rA<sub>18</sub> bind two or six subunits on the distal face of Hfq (30). rU<sub>6</sub> and rAU<sub>5</sub>G mimic the sRNA 3' end and contact 5-6 subunits around the proximal inner surface (31–33). rCU<sub>2</sub>C<sub>2</sub> and rim-SL, which contains rCU<sub>2</sub>C<sub>2</sub> plus a stable stem loop, mimic RNA motifs that bind near a subunit interface within the hexamer, and interact with the arginine patch on the rim (12, 20). Additionally, we studied two larger RNAs that interact with both the proximal face and rim: rim SL-U<sub>6</sub>, designed to mimic an entire sRNA 3' Hfq binding region, and RybB, a 79 nt natural sRNA that contains both the 3' Hfq binding region and a 5' seed region responsible for targeting the mRNA (34, 35).

We exploited nMS to monitor the dissociation of 1:1 RNA:Hfq complexes. The mass-corrected ERMS plots of HfqΔCTD revealed a greater dissociation energy when the protein was bound to RNA, indicating that in the absence of the CTDs, the RNA stabilizes the Hfq hexamer (Figs. 4B and S8). To determine how much stability HfqΔCTD gained from its interactions with the RNA, we calculated the difference in collision energies between HfqΔCTD•RNA and HfqΔCTD at each remaining fraction of precursor (Fig. 4C). We found that all RNAs stabilized HfqΔCTD over a range of precursor fractions. Short RNAs contacting the proximal face (U<sub>6</sub> and AU<sub>5</sub>G) and the rim (CU<sub>2</sub>C<sub>2</sub> and rim-SL) provided minimal extra stability. The longest RNAs tested (RybB and rim-SL-U<sub>6</sub>) provided the most extra stability, followed by RNAs that bind to the distal face (rA<sub>18</sub> and rA<sub>6</sub>).

Unlike HfqΔCTD, RNA binding destabilized WT Hfq, relative to free Hfq. Not only did the RNA complexes fragment at lower collision energies than the unbound protein (Figs. 4D and S9), but the shapes of the ERMS plots differed for the various

RNAs, suggesting that the complexes have different structures. Surprisingly, the difference in collision energies between Hfq•RNA and Hfq showed that even short RNAs and RNAs that bind the distal face destabilized the complexes substantially (Fig. 4E). Thus, in the presence of the CTDs, RNA binding can result in net destabilization of Hfq complexes.



**Fig. 4.** RNA binding stabilizes Hfq $\Delta$ CTD but destabilizes the WT protein. (A) RNAs used in this study (Table S2) interact with different surfaces of Hfq. (B, D) ERMS showing the fraction of precursor for (B) Hfq $\Delta$ CTD and (D) Hfq bound to RNA. Colors correspond to the RNAs in (A). The collision energies are corrected for the mass of the RNAs ( $m_{\text{Protein}}/m_{\text{Protein+RNA}}$ ). The collision energies for Hfq are also corrected for the mass of the CTDs ( $m_{\text{Hfq}\Delta\text{CTD}}/m_{\text{Hfq}}$ ). Solid lines represent a cubic interpolation of the data. The spread on the interpolated line represents the mean of the errors of individual data points. For clarity, data symbols and error bars are not shown. See Figs. S8-S9 for further data on all dissociation products. (F) ERMS of Hfq (solid lines) and Hfq $\Delta$ CTD (dashed lines) with (green) and without (black) bound RybB. (C, E, G) Collision energy difference ( $\Delta$ CE) between (C) Hfq $\Delta$ CTD•RNA – Hfq $\Delta$ CTD, (E) Hfq•RNA – Hfq, and (G) Hfq•RNA – Hfq $\Delta$ CTD•RNA as a function of precursor fraction. Colors are as in (A). Black line depicts Hfq – Hfq $\Delta$ CTD. Solid lines are a visual guide. Errors were calculated from the spread of the ERMS, as explained in SI Materials and Methods.

**RNA binding removes stabilizing CTD interactions.** We hypothesized that RNA binding destabilizes Hfq by perturbing the organization of the disordered CTDs. To determine how much the CTDs stabilize the complexes, we calculated the difference in

collision energy needed to fragment free WT Hfq and Hfq $\Delta$ CTD with and without RNA (Figs. 4F-G). In the absence of RNA, the CTDs significantly stabilized the Hfq hexamer (Fig. 4F-G, black lines). In the presence of RybB sRNA, the complexes had similar stabilities that were intermediate between apo WT Hfq and apo Hfq $\Delta$ CTD (Fig. 4F, green lines). This result suggested that RNA binding opposes the effects of the CTDs while stabilizing the Hfq core.

For the free proteins, the observed stability gap was wider as more precursor was fragmented (Fig. 4G), suggesting a range of complexes with variable strengths of CTD interactions, as already deduced by comparison of the protein's ERMS (Figs. 2B-C). Upon RNA binding, the stability conferred by the CTDs became more uniform as the precursor was dissociated (Fig. 4G, colored lines). Thus, RNA binding results in complexes that dissociate more homogeneously.

All the RNAs tested interfered with stabilization by the CTDs. The effect was strongest for the longest RNAs (rim SL-U<sub>6</sub> and RybB) that presumably displace more CTD contacts with the rim and proximal face of Hfq. However, even the smallest RNAs tested abrogated the stabilizing effect of the CTDs (Fig. 4G). Interestingly, this loss was also substantial for RNAs binding the distal face, even though only the first few residues of the CTDs are close to this surface (10).

Next, we studied how RNA binding affected the dissociation pathways of Hfq and Hfq $\Delta$ CTD, by determining which subcomplexes retained the RNA after hexamer dissociation (Fig. S10A). To minimize contributions from secondary dissociation events, we compared the dissociation products when only 20% of the protein was fragmented (0.8 precursor fraction). Comparison of the dissociation pathways revealed similar fragmentation of RNA complexes with Hfq and with Hfq $\Delta$ CTD (Fig. S10B-C). This result suggested that in WT Hfq, RNA binds to the core and that perturbation of the CTDs likely comes from their displacement from the core and not from direct interactions with the RNA. CTD displacement was maximal for the largest RNA (RybB sRNA), as the ERMS for Hfq and Hfq $\Delta$ CTD almost overlapped (Fig. 4F). The ERMS of shorter RNA complexes were closer to the ERMS of the free proteins, indicating that binding to the core did not require as much displacement of the CTDs (Figs. 4B and 4D). Finally, we observed that the percentage of fragments with bound rim SL was different for each protein, suggesting different modes of binding that could explain the lesser perturbation of the CTDs by this RNA (Figs. S10C and 4G; light blue). The dissociation of Hfq•AU<sub>5</sub>G was also slightly altered, in contrast with U<sub>6</sub>'s, which was identical with or without the CTDs (Fig. S10C; orange and yellow). Given that the fragmentation of both proximal-binding RNAs was very similar for the truncated protein, it is possible that the CTDs help discriminate between optimal proximal-binding RNA motifs.

**Progressive RNA binding displaces the CTDs from Hfq.** The design of our RNAs allowed us to investigate the interplay between Hfq, the disordered CTDs and RNA-protein interactions as an sRNA progressively binds to the protein (Figs. 5A-B and S7B). For this, we compared the stabilities and dissociation

pathways of WT Hfq and Hfq $\Delta$ CTD in the absence and presence of RNA segments mimicking a stepwise binding process (Fig. 5B). On the one hand, we found that as the RNA interacts with more surfaces of Hfq, the RNA conferred stability to the protein core (Fig. 5C; Hfq $\Delta$ CTD). However, RNA binding was accompanied by a loss of favorable CTD interactions (Fig. 5C; CTD) that reduced stability overall (Fig. 5C; Hfq). On the other hand, progressive RNA binding shifted the dissociation products to increasingly larger fragments (Fig. 5D). This dissociation pattern was the same for Hfq and Hfq $\Delta$ CTD, indicating similar RNA interactions with the core in both proteins. Thus, stable interactions between RNA and the core are established as the CTDs are displaced.

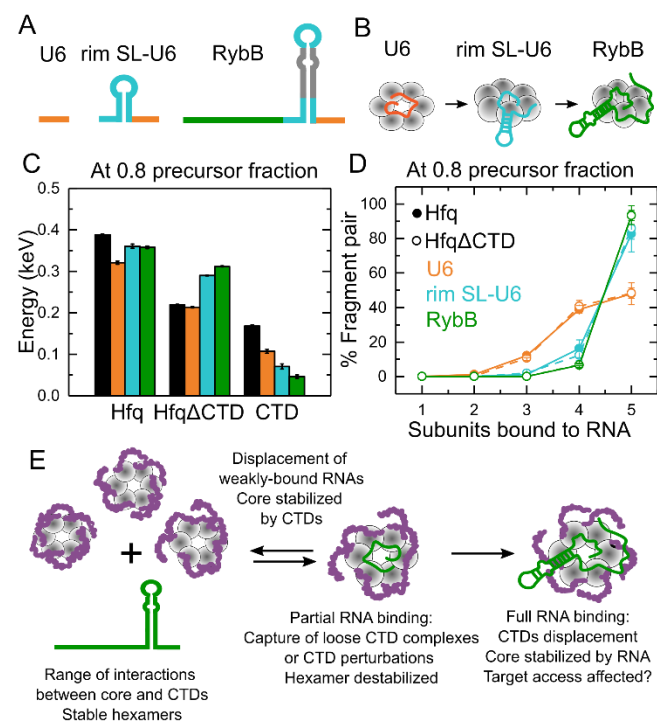


Fig. 5. RNA binding progressively displaces the Hfq CTDs. (A) RNAs mimicking segments of RybB sRNA that progressively interact with more Hfq surface: 3' U tail (orange), terminal stem-loop (cyan), 5' seed (dark green). (B) Cartoon of progressive RNA binding. (C) Stabilities of Hfq, Hfq $\Delta$ CTD (core) and CTDs when bound to the RNAs shown in (A). Colors correspond to the RNAs as shown in (B). The stabilities were determined as the collision energies of Hfq, Hfq $\Delta$ CTD or their difference when 20% of the precursor was fragmented (0.8 precursor fraction). Errors were calculated as in Fig. 4C (see SI Methods). (D) Percentage of fragment pairs after the collision of Hfq (solid lines) or Hfq $\Delta$ CTD (dashed lines) bound to the RNAs in (A) (see also Fig. S10). Errors were calculated as for Figs. 2D-E. (E) Assorted organization of the disordered CTDs results in heterogeneous complexes. RNAs could bind complexes with loose CTD conformations or perturb a CTD. Local CTD disturbances destabilize the hexamer and are transmitted to the rest of the protein, promoting further CTD displacement and stable binding for favorable RNAs. Non-preferred RNAs are displaced by competition with the CTDs (21), reestablishing hexamer stability.

## Discussion

In this work, we employed nMS-SID to analyze Hfq complexes of defined stoichiometry, connectivity and shape. The results demonstrate that the CTDs connect all of the Hfq subunits, stabilizing the hexamer (Fig. 2C, 2G and S10A). Moreover, MD simulations using an improved force-field for IDPs (36) revealed that each CTD can dynamically interact with several other CTDs and folded domains within the hexamer (Figs. 3, S3 and S6). Although RNA binding also stabilizes the folded core of Hfq, RNA disrupts the network of CTD interactions. Only RNAs that form multiple favorable contacts with Hfq offset the loss of stabilization by the CTDs, explaining how the CTDs help Hfq discriminate among different RNAs.

Based on our results, we propose that RNAs stably bind Hfq through a stepwise process (Fig. 5E). An RNA may first take advantage of configurations in which an Hfq region is exposed, or a CTD is loosely folded. The initial interaction with a segment of the RNA perturbs nearby CTDs. Because the CTDs are interconnected, this perturbation propagates to the rest of the protein hexamer, resulting in more CTD displacement and further RNA binding. If binding is favorable, the hexamer is stabilized. However, if binding is not favorable, the hexamer is destabilized, making it more favorable for the CTDs to regain their interactions with the rest of the protein. This search for stability could explain the removal of weakly-bound RNAs (21). Additionally, an integrated organization of the CTDs could facilitate sRNA competition and the search for mRNA targets, as binding of a second RNA will also perturb the CTDs, stimulating sRNA displacement or disassembly of non-cognate ternary complexes. Finally, the varied CTD conformations impart asymmetry to the Hfq homohexamer, that may also contribute to the selection of RNA ligands.

Our nMS-SID results showed that the CTDs make the Hfq hexamer more resistant to dissociation by collision (Fig. 2C). This observation agrees with previous collision-induced dissociation MS experiments showing that *E. coli* Hfq is more stable than Hfq from *V. cholerae*, which has a 22 residue CTD (37). Vincent et al. (37) attributed this stabilization to packing of the CTDs along the intersubunit interfaces. Based on our observation that the CTDs unfold before the subunits dissociate, we propose that Hfq is additionally stabilized by CTD interactions with multiple Hfq subunits that are facilitated by longer CTDs.

Network connectivity can explain why the binding of even short RNAs reduced the CTD's stabilizing contribution (Figs. 4D-G and 5C). Also, since RNA-bound Hfq $\Delta$ CTD and Hfq fragmented similarly (Figs. 5D and S10), stable RNA binding seems to involve contacts with the core and not the CTDs, as previously proposed (12, 38, 39). Finally, destabilization of Hfq but not Hfq $\Delta$ CTD when A<sub>18</sub> RNA is bound (Fig. 4C) supports a reported functional link between the distal RNA binding face and R66 at the start of the CTD (23).

Hfq is a model protein with disordered regions that act synergistically to communicate perturbations among its subunits. This feature is enabled by an architecture of identical monomers, each providing an identical disordered region. The single-stranded DNA binding (SSB) protein, a homotetramer involved in DNA



repair, replication and recombination (40), also contains disordered CTDs with parallels to Hfq: they impart stability to the SSB tetramer (41) but are displaced upon partial DNA engagement, modulating binding and SSB oligomerization (42). Importantly, partial deletion of SSB's CTDs results in impaired activity, indicating a critical role for multiple CTD interactions in cellular function (43). Histone tails are also thought to fold heterogeneously around the DNA (44, 45), and to be disrupted by chaperone binding or post-translational modification (46). It would be interesting to know if other RNA-binding proteins use interconnected IDPs to integrate the molecular interactions within RNA-protein complexes.

## Materials and Methods

**RNA preparation.** RybB RNA was transcribed using phage T7 RNA polymerase followed by 8% polyacrylamide gel purification (8M urea). The remaining short RNAs used in the study were purchased from IDT with HPLC purification. See Table S2 for RNA sequences.

**Protein expression and purification.** Hfq $\Delta$ CTD and Hfq were purified as described before (20).

**Sample preparation for native mass spectrometry (nMS).** Protein (Hfq $\Delta$ CTD and Hfq) and RNA (rA<sub>18</sub>, rim SL-U<sub>6</sub> and RybB) samples were dialyzed overnight into 500 mM ammonium acetate pH 6.8 (99.99%, MilliporeSigma), with eight buffer exchanges (3.5 kDa mass cutoff micro-dialysis, Pierce). This ionic strength prevented precipitation of protein-RNA complexes at the concentrations (10  $\mu$ M hexamer) required for ion mobility MS. The remaining RNA samples were used as supplied by the manufacturer and did not require additional desalting. Protein-RNA complexes were prepared by mixing 1:1 RNA and Hfq $\Delta$ CTD or Hfq to a final concentration of 10  $\mu$ M each. 400 mM ammonium acetate (final concentration) plus triethylammonium acetate (TEAA) (1 M, MilliporeSigma; 100 mM final concentration for charge reduction) were subsequently added to the samples.

**Mass spectrometry.** All samples were introduced into the mass spectrometer using nano-electrospray emitters that were prepared in-house using a Sutter P-97 micropipette puller. All spectra in this work were acquired on a Waters Synapt G2 HDMS instrument (Waters Corporation, Wilmslow, U.K.) modified with a surface-induced dissociated (SID) device between a shortened trap stacked ring ion guide and an ion mobility cell, as described previously (47). SID lenses can be tuned either to transmit ions for MS or to direct the ions onto the surface for collision. Typical voltage settings and instrument parameters used here for transmission mode and SID can be found in the Supporting Information (Tables S3-S4). Energy resolved mass spectra (ERMS) were produced by acquiring data from tandem MS experiments with SID voltage potentials ranging from 15 and 140 V. Each experiment was repeated in technical triplicate. Additional information is provided in SI.

**Analysis of mass spectrometry data.** Ion mobility was used to separate product ions and selection rules for each SID product were made using Waters Corporation Driftscope 2.9 software. The intensity of subcomplexes were extracted from SID spectra with TWIMExtract v1.3 (48). Collision energies were calculated as  $E(eV) = zV_{SID}$ ; where  $z$  is the charge state of the precursor ions and  $V_{SID}$  is the SID voltage. Energy-resolved mass spectra (ERMS) were corrected by the  $m_{Hfq\Delta CTD}/m_{Hfq}$  and  $m_{Protein}/m_{Protein-RNA}$ . Additional information provided in SI.

**MD simulations.** All simulations were performed with the molecular dynamics program OpenMM (49) and CHARMM36m force-field (50). Simulations were started from the PDB ID 1HK9 which included residues 7-68 (29). The starting structures of the missing CTDs were obtained from a) top 10 Rosetta models (15) (top 1-10) and b) 1 structure in which CTDs are fully extended (21). Four simulations with extended CTDs were performed starting with different initial velocities. All protein structures were embedded in a water box and neutralized with 150 mM Na<sup>+</sup>Cl<sup>-</sup>. Additional simulation and setup details are provided in the SI. Following a 100 ns equilibration, the models were run for an additional 400 ns (top 1-5), 100 ns (top 6-10) or 10 ns (extended 1-4) (Table S5). To gain information about the short-term structures and their evolution, we analyzed contacts between CTDs and cores of various subunits during various time intervals of the simulation (Figs. S4-S6). Additional information provided in SI.

## Acknowledgements

The authors thank Benjamin Jones for his assistance with instrument maintenance and automated data acquisition and Andrew Santiago-Frangos for help with initial samples. This work was funded by grants from National Institutes of Health [1P41GM128577-01 to V.H.W.; R35 GM136351-03 to S.A.W.], the Johns Hopkins University Provost's Office [PURA to H.L.]. Part of this research was conducted using computational resources at the Maryland Advanced Research Computing Center.

## References

1. P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169–1175 (2004).
2. D. S. Ottoz, L. E. Berchowitz, The role of disorder in RNA binding affinity and specificity. *Open Biol.* **10**, 1–10 (2020).
3. S. Basu, R. P. Bahadur, A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell. Mol. Life Sci.* **73**, 4075–4084 (2016).
4. P. A. Sharp, A. K. Chakraborty, J. E. Henninger, R. A. Young, RNA in formation and regulation of transcriptional condensates. *RNA* **28**, 52–57 (2022).
5. E. G. H. Wagner, P. Romby, “Small RNAs in Bacteria and Archaea: Who they are, what they do, and how they do it” in *Advances in Genetics*, (Elsevier Ltd, 2015), pp.



- 133–208.
6. P. Sobrero, C. Valverde, The bacterial protein Hfq: Much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* **38**, 276–299 (2012).
  7. A. Santiago-Frangos, S. A. Woodson, Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA* **9**, e1475 (2018).
  8. E. Sauer, S. Schmidt, O. Weichenrieder, Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc. Natl. Acad. Sci.* **109**, 9396–9401 (2012).
  9. S. Panja, D. J. Schu, S. A. Woodson, Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res.* **41**, 7536–7546 (2013).
  10. M. Beich-Frandsen, B. Večerek, B. Sjöblom, U. Bläsi, K. Djinović-Carugo, Structural analysis of full-length Hfq from *Escherichia coli*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 536–540 (2011).
  11. B. Wen, *et al.*, Structural and dynamic properties of the C-terminal region of the *Escherichia coli* RNA chaperone Hfq: integrative experimental and computational studies. *Phys. Chem. Chem. Phys.* **19**, 21152–21164 (2017).
  12. D. Dimastrogiovanni, *et al.*, Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *Elife* **3**, 1–19 (2014).
  13. V. Arluison, *et al.*, The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.* **271**, 1258–1265 (2004).
  14. B. Večerek, L. Rajkowitsch, E. Sonnleitner, R. Schroeder, U. Bläsi, The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res.* **36**, 133–143 (2008).
  15. E. Sonnleitner, I. Moll, U. Bläsi, Functional replacement of the *Escherichia coli* hfq gene by the homologue of *Pseudomonas aeruginosa*. *Microbiology* **148**, 883–891 (2002).
  16. E. Sonnleitner, *et al.*, Functional effects of variants of the RNA Hfq. *Biochem. Biophys. Res. Commun.* **323**, 1017–1023 (2004).
  17. A. S. Olsen, J. Møller-jensen, R. G. Brennan, P. Valentin-hansen, C-Terminally Truncated Derivatives of *Escherichia coli* Hfq Are Proficient in Riboregulation. *J. Mol. Biol.* **404**, 173–182 (2010).
  18. T. B. Updegrove, R. M. Wartell, The influence of *Escherichia coli* Hfq mutations on RNA binding and sRNA-mRNA duplex formation in rpoS riboregulation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1809**, 532–540 (2011).
  19. N. N. Salim, M. A. Faner, J. A. Philip, A. L. Feig, Requirement of upstream Hfq-binding (ARN)<sub>x</sub> elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res.* **40**, 8021–8032 (2012).
  20. A. Santiago-Frangos, K. Kavita, D. J. Schu, S. Gottesman, S. A. Woodson, C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc. Natl. Acad. Sci.* **113**, E6089–E6096 (2016).
  21. A. Santiago-Frangos, J. R. Jeliakov, J. J. Gray, S. A. Woodson, Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *Elife* **6**, 1–25 (2017).
  22. A. Santiago-Frangos, *et al.*, *Caulobacter crescentus* Hfq structure reveals a conserved mechanism of RNA annealing regulation. *Proc. Natl. Acad. Sci.* **116**, 10978–10987 (2019).
  23. K. Kavita, *et al.*, Multiple in vivo roles for the C-terminal domain of the RNA chaperone Hfq. *Nucleic Acids Res.* **50**, 1718–1733 (2022).
  24. A. C. Leney, A. J. R. Heck, Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
  25. S. R. Harvey, *et al.*, Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8143–8148 (2019).
  26. A. Q. Stiving, *et al.*, Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Anal. Chem.* **91**, 190–209 (2019).
  27. M. Zhou, S. Dagan, V. H. Wysocki, Protein subunits released by surface collisions of noncovalent complexes: Nativelike compact structures revealed by ion mobility mass spectrometry. *Angew. Chemie - Int. Ed.* **51**, 4336–4339 (2012).
  28. R. S. Quintyn, J. Yan, V. H. Wysocki, Surface-induced dissociation of homotetramers with D2 symmetry yields their assembly pathways and characterizes the effect of ligand binding. *Chem. Biol.* **22**, 583–592 (2015).
  29. C. Sauter, J. Basquin, D. Suck, Sm-like proteins in Eubacteria: The crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.* **31**, 4091–4098 (2003).
  30. T. M. Link, P. Valentin-Hansen, R. G. Brennan, Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci.* **106**, 19292–19297 (2009).
  31. M. A. Schumacher, R. F. Pearson, T. Møller, P. Valentin-Hansen, R. G. Brennan, Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO J.* **21**, 3546–3556 (2002).
  32. E. Sauer, O. Weichenrieder, Structural basis for RNA 3'-end recognition by Hfq. *Proc. Natl. Acad. Sci.* **108**, 13065–13070 (2011).
  33. W. Wang, *et al.*, Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. *Genes Dev.* **25**, 2106–2117 (2011).
  34. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15**, 1637–1651 (2001).

35. K. Papenfort, *et al.*,  $\sigma$ E-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol. Microbiol.* **62**, 1674–1688 (2006).
36. J. Huang, *et al.*, CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).
37. H. A. Vincent, *et al.*, Characterization of *Vibrio cholerae* Hfq provides novel insights into the role of the Hfq C-terminal region. *J. Mol. Biol.* **420**, 56–69 (2012).
38. M. Beich-Frandsen, *et al.*, Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. *Nucleic Acids Res.* **39**, 4900–4915 (2011).
39. K. E. Robinson, J. Orans, A. R. Kovach, T. M. Link, R. G. Brennan, Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.* **42**, 2736–2749 (2014).
40. R. D. Shereda, A. G. Kozlov, T. M. Lohman, M. M. Cox, J. L. Keck, SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318 (2008).
41. C. E. Mason, *et al.*, *Escherichia coli* single-stranded DNA-binding protein: NanoESI-MS studies of salt-modulated subunit exchange and DNA binding transactions. *J. Am. Soc. Mass Spectrom.* **24**, 274–285 (2013).
42. A. G. Kozlov, M. M. Cox, T. M. Lohman, Regulation of single-stranded DNA binding by the C termini of *Escherichia coli* single-stranded DNA-binding (SSB) protein. *J. Biol. Chem.* **285**, 17246–17252 (2010).
43. E. Antony, *et al.*, Multiple C-terminal tails within a single *E. coli* SSB homotetramer coordinate DNA replication and repair. *J. Mol. Biol.* **425**, 4802–4819 (2013).
44. E. A. Morrison, S. Bowerman, K. L. Sylvers, J. Wereszczynski, C. A. Musselman, The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7**, 1–35 (2018).
45. M. Ghoneim, H. A. Fuchs, C. A. Musselman, Histone Tail Conformations: A Fuzzy Affair with DNA. *Trends Biochem. Sci.* **46**, 564–578 (2021).
46. T. Kouzarides, Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
47. M. Zhou, C. Huang, V. H. Wysocki, Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer. *Anal. Chem.* **84**, 6016–6023 (2012).
48. S. E. Haynes, *et al.*, Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Anal. Chem.* **89**, 5669–5672 (2017).
49. P. Eastman, *et al.*, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, 1–17 (2017).
50. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A Web-Based Graphical User Interface for Computational Chemistry Softwares. *J. Comput. Chem.* **32**, 174–182 (2012).