RNA chaperones are proteins that aid in the folding of nucleic acids, but remarkably, many of these proteins are intrinsically disordered. How can these proteins function without a well-defined three-dimensional structure? Here, we address this question by studying the hepatitis C virus core protein, a chaperone that promotes viral genome dimerization. Using single-molecule fluorescence spectroscopy, we find that this positively charged disordered protein facilitates the formation of compact nucleic acid conformations by acting as a flexible macromolecular counterion that locally screens repulsive electrostatic interactions with an efficiency equivalent to molar salt concentrations. The resulting compaction can bias unfolded nucleic acids towards folding, resulting in faster folding kinetics. This potentially widespread mechanism is supported by molecular simulations that rationalize the experimental findings by describing the chaperone as an unstructured polyelectrolyte.
Ever since the discovery of transfer RNA and ribosomal RNA, the number of functions attributed to structured nucleic acids has continuously increased. Due to the polyanionic nature of nucleic acids, formation of these structures is strongly influenced by electrostatics. Additionally, as in the case of proteins, non-native interactions can impede the molecular search for biologically active conformations and result in kinetic traps. The folding problem for nucleic acids is exacerbated by the presence of nucleic acids prevents investigation of this process at the high sample concentrations required for many ensemble techniques, such as circular dichroism or NMR spectroscopy. Therefore, we use an integrative combination of single-molecule Förster resonance energy transfer (FRET) spectroscopy and molecular simulations to elucidate the mechanism of chaperone-assisted folding. We find that NCD facilitates the folding of nucleic acids by acting as a macromolecular counterion that locally screens charge repulsion with extreme efficiency. Note that in this work, the terms RNA chaperone and chaperone are rather than the more restrictive definition of "proteins that aid in the folding of RNA" rather than the more restrictive definition of "proteins that aid in the process of RNA folding by preventing misfolding or by resolving misfolded species", following the common classification of NCD as an RNA chaperone.

**Results**

**NCD promotes folding via electrostatic interactions.** The capacity for NCD to chaperone hairpin formation is apparent from confocal single-molecule FRET experiments on freely diffusing nucleic acids designed to form a 7-bp stem-loop structure (Fig. 1a). These hairpin molecules were site-specifically labeled near the 3′ and 5′ ends with donor and acceptor dyes, respectively, allowing the formation of the stem-loop (i.e., folding) to be monitored via a change in the highly distance-dependent transfer efficiency. In the absence of NCD and at near-physiological concentrations of monovalent cations, the hairpin primarily occupies unfolded conformations, which result in a low mean transfer efficiency. With increasing concentration of NCD (Fig. 1b, c), the hairpin more frequently occupies high-transfer-efficiency folded conformations. The low-nanomolar apparent dissociation constant, \( k_{d} \), of this chaperone-induced transition (Fig. 1c) highlights the high affinity of NCD for the DNA hairpin. Similar results were observed for experiments with an RNA hairpin (Supplementary Fig. 2a). Furthermore, experiments on an oligonucleotide with identical length and a sequence that is unable to form a hairpin reveal that the chaperone also binds and compacts this folding-incompetent nucleic acid (Supplementary Fig. 2b).

Given the large and opposite structural charge of NCD and the DNA (−25/−3) and the RNA (−60) (Fig. 1a), their association is expected to be electrostatically driven. Indeed, the stability of the complex is extremely sensitive to the addition NaCl, with a strong concomitant increase in \( K_{d} \) (Fig. 1c, d). This effect can be interpreted in terms of the release of ~14 counterions when NCD and the DNA associate (Supplementary Fig. 3), a substantially larger number than previously reported for other nucleic acid-binding proteins. Notably, the addition of physiological concentrations of Mg\(^{2+}\) does not alter NCD’s ability to promote hairpin formation (Supplementary Fig. 4a). In low-salt conditions, multiple chaperone molecules can bind a single hairpin (Supplementary Fig. 4b), presumably because of the charge imbalance between NCD and the hairpin. However, in the following sections, we focus on the one-to-one complex (see Methods section). To summarize, the strong electrostatic contribution to binding suggests that the interaction is generic for different kinds and sequences of nucleic acids and rationalizes our finding that NCD has the capacity to bind both DNA and RNA as well as folded and unfolded molecules.

Chaperone binding accelerates folding. To observe chaperone-assisted folding directly, we recorded fluorescence time traces of individual surface-immobilized 5′-3′ FRET-labeled hairpin molecules using confocal fluorescence spectroscopy. In the absence of NCD (Fig. 2a), the average dwell time in the low-transfer-efficiency unfolded (U) conformation, \( \langle \tau_{U} \rangle = 60 \pm 20 \text{ ms} \), is more than twice as long as the average dwell time in the high-transfer-efficiency folded (F) conformation, \( \langle \tau_{F} \rangle = 26 \pm 8 \text{ ms} \). The resulting transfer efficiency histograms resemble those from free-diffusion measurements under identical conditions (Fig. 1b; red), indicating that immobilization does not perturb the system. With 150 nM unlabeled NCD (Fig. 2b), a concentration where the hairpin is bound to chaperone for more than 95% of the time (Fig. 1c), the fluorescence time traces look very different from those in the absence of NCD: The transfer efficiency of the hairpin is almost always high because NCD favors the chaperone-bound folded conformation of the hairpin (FC). However, there are frequent short excursions to a second conformation with an average dwell time of \( \langle \tau_{FC} \rangle = 3 \pm 1 \text{ ms} \), which give rise to a small population at a transfer efficiency of ~0.5 (Fig. 2b, d). This small population is also visible in the free-diffusion measurements under identical conditions (Fig. 1b; cyan), and its transfer efficiency is close to that of the folding-incompetent hairpin bound to NCD (Supplementary Fig. 2b). These observations suggest that the chaperone-bound unfolded hairpin (UC) is transiently populated at equilibrium, and that it is more compact than the chaperone-free unfolded hairpin (U). The existence of the UC population was further confirmed by recurrence analysis of independent free-diffusion measurements (Supplementary Fig. 5).
In both the absence (Fig. 2c) and presence (Fig. 2d) of NCD, a quantitative two-state kinetic analysis (see Methods section) of the dwell times yielded rate constants consistent with the donor-acceptor fluorescence cross-correlation functions. A comparison of the rate constants in the two distinct kinetic regimes reveals that hairpin folding is accelerated more than 20-fold when NCD is bound, with no significant change in the unfolding rate constant. Again, these results agree with a recurrence analysis of corresponding free-diffusion experiments (Supplementary Fig. 5), providing further support that immobilization does not alter the behavior of the hairpin or NCD.

The full kinetics of chaperone-assisted folding are most obvious at NCD concentrations near \( K_d \), where NCD is bound roughly half of the time. Under these conditions (Fig. 2e), fluorescence time traces show single molecules toggling between the two distinct kinetic regimes (Fig. 2a, b) as the chaperone binds and dissociates on the timescale of seconds. The single-molecule experiments report on chaperone binding primarily via changes in folding/unfolding kinetics, as the transfer efficiencies of F and FC are very similar. Therefore, we analyzed the fluorescence time traces using a four-state hidden-Markov model (Fig. 2f) (see Methods section). Maximum-likelihood (MLH) analysis shows that NCD binds to the unfolded and folded hairpin with similar rate constants of \( \sim 10^8 \text{M}^{-1}\text{s}^{-1} \), but that it dissociates an order of magnitude more slowly from the folded than from the unfolded hairpin (Fig. 2f; gray). The resulting binding and dissociation rate constants, as well as the \( K_d \) for NCD and the unfolded (U) hairpin were consistent with those observed for surface-immobilized folding-incompetent hairpins (Supplementary Fig. 6), where these rate constants can be measured independently, confirming both the accuracy of the MLH analysis and the applicability of the four-state model (Fig. 2f). In summary, our results suggest that NCD aids in nucleic acid folding by increasing the free energy of the unfolded relative to the folded state, which leads to a concomitant reduction in the folding free-energy barrier and shifts the equilibrium towards folded conformations. It is worth mentioning that this mechanism is not expected to specifically prevent misfolding.

A disordered protein-nucleic acid complex. What are the conformational properties of this protein-nucleic acid complex, and how do they explain the molecular mechanism by which NCD accelerates folding? There is increasing evidence that positively charged intrinsically disordered proteins can form high-affinity complexes with nucleic acids and remain disordered and highly dynamic in the bound state, including NCD in complex with viral RNA hairpins, suggesting that NCD performs its chaperoning function as a disordered protein. To test whether it indeed remains disordered and correspondingly dynamic after binding to the hairpin, we labeled a two-cysteine variant of NCD (S2C-T65C) with donor and acceptor dyes (Fig. 3a) and used nanosecond fluorescence correlation spectroscopy to probe its reconfiguration dynamics in the presence and absence of the unlabeled hairpin and its folding-incompetent counterpart (Fig. 3b). In all cases, the resulting correlation functions reveal pronounced fluorescence fluctuations of donor and acceptor emission on the 50-ns timescale, as expected for the long-range...
distance dynamics characteristic of intrinsically disordered proteins, supporting the notion that NCD remains disordered while bound to nucleic acids.

To further test the presence of disorder and map out the structural properties of the two binding partners, we placed FRET dyes at different positions within the chaperone and the hairpin (Fig. 3a, c). To isolate the UC complex, analogous experiments were conducted with the folding-incompetent DNA. The intramolecular transfer efficiencies of NCD (S2C-S130C, S2C-T65C, and T65C-S130C) increase when the UC complex is formed; the same is true for the intramolecular transfer efficiencies of the folding-incompetent DNA (U) (5′-3′, 3′-MID, and 5′-MID) relative to its chaperone-bound counterpart (UC). These results indicate that both the chaperone and the unfolded nucleic acid become more compact when the complex is formed (Fig. 3c), as expected from the mutual screening of the repulsive charges within each chain.

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10356-0

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10356-0 | www.nature.com/naturecommunications
Fig. 2 Chaperone binding accelerates hairpin folding. a Representative donor (green) and acceptor (magenta) fluorescence time traces depicting unassisted folding of the surface-immobilized 5'-3' FRET-labeled hairpin. The uncorrected transfer efficiency, \( \hat{E} \), (black) and the most likely state trajectory (gray) based on the Viterbi algorithm are shown below. b Analogous to a, but depicting chaperone-assisted hairpin folding with saturating concentrations of NCD (150 nM) to ensure that hairpin molecules are almost always chaperone-associated. c Survival probability plots of folded- (upper left) and unfolded-state (lower left) dwell times in the absence of chaperone, fit with single-exponential decays to determine the unfolding and folding rate constants, \( k_u \) and \( k_f \), respectively; \( \langle k_u \rangle \) and \( \langle k_f \rangle \) represent mean values obtained from 54 single-molecule measurements. The donor-acceptor fluorescence cross-correlation function (right) decorrelates with a rate constant that is consistent with the sum of \( k_u \) and \( k_f \) from the dwell-time analysis. d Analogous to c, but for chaperone-assisted folding; \( \langle k_u \rangle \) and \( \langle k_f \rangle \) represent mean values obtained from 26 single-molecule measurements. The black lines in c and d are fits with single exponentials with 99% confidence bands shown in gray. e Representative fluorescence time trace measured at a chaperone concentration near 300 nM. Transitions between chaperone-bound (cyan shading) and chaperone-free (red shading) kinetic regimes arise from binding and dissociation of NCD. f Kinetic 4-state model for chaperone-assisted folding. The folded, F, and unfolded, U, conformations of the hairpin freely interconvert in the absence of chaperone, C, with an equilibrium constant that favors U. When the chaperone is bound, the unfolded, UC, and folded, FC, conformations of the hairpin still interconvert, but with an equilibrium constant that favors F. g Trajectory from coarse-grained molecular dynamics simulation showing binding (cyan shading) and dissociation (red shading) of the chaperone to the hairpin concomitant with the formation of intermolecular contacts (gray). The folding and unfolding of the hairpin, monitored via native contacts (black), is shifted to favor hairpin formation when the chaperone is bound. Structural representations are taken from the simulations (see Supplementary Movies 1 and 2). Source data for mean values in c and d are provided as a Source Data File.

Fig. 3 Structural properties of chaperone-hairpin complex. a Molecular surface representations of the binding partners colored by electrostatic potential (red, negative; blue, positive). Arrows indicate the approximate location of labeling sites. b Donor-donor (light green), acceptor-acceptor (magenta), and donor-acceptor (gray) fluorescence correlation functions for FRET-labeled NCD S2C-T65C in isolation (top), bound to the hairpin (middle), and bound to the folding-incompetent hairpin (bottom). Solid black lines are global fits of the three respective correlation functions, which are used to determine the reconfiguration times, \( \tau_r \), of the chaperone. c Mean transfer efficiencies for all labeling pairs from the mapping experiment (triangles) at 30 mM PB pH 7.0 and the coarse-grained molecular simulations (circles). Mean values are provided in the Source Data file. Labeling positions in the chaperone and the hairpin are shown in cyan and red, respectively. The uncertainties associated with transfer efficiency values are -0.01 based on multiple simulations and -0.03 for the experiments (error dominated by instrument calibration). d Molecular representations of the disordered nucleic acid-chaperone complex from the simulations (see Supplementary Movies 1 and 2); color code as in Fig. 1a. e Correlation between experimental and simulated mean transfer efficiencies. The dashed gray line is the identity line, the solid black line is a linear regression.
The RNA and DNA hairpins used in this report represent a structural motif that is ubiquitous in folded nucleic acids, and therefore the functional mechanism we propose is expected to be applicable to a wide range of chaperone-assisted conformational transitions in nucleic acids. The extent of its applicability certainly merits further investigation. For instance, it has been proposed that hepatitis C virus genome dimerization is initiated by forming intermolecular base pairs between the unpaired nucleotides in the stem-loop structure of the dimer linkage sequence (DLS; Supplementary Fig. 1), giving rise to what is commonly referred to as RNA kissing loops. It is plausible that NCD promotes kissing loop formation, and therefore genome dimerization, via the charge screening mechanism we describe for chaperone-assisted hairpin formation: NCD will bind to the DLS, serving as a macromolecular counterion and reducing the effective charge of the nucleoprotein complex, which reduces the repulsion between the two interaction sites (i.e., the unpaired nucleotides in the loops of the DLS). As a result, the rate constant for kissing loop formation would be expected to increase, and the equilibrium would shift to favor genome dimerization after binding of NCD.

Intrinsically disordered RNA chaperones that bind nucleic acids with high affinity thus represent an efficient and genetically controllable strategy for altering the conformations and dynamics of functional nucleic acids (both DNA and RNA) in a cellular environment, whereas factors such as cellular salt concentrations can only be regulated within narrow bounds. These types of interactions are a likely reason why nucleic acids exposed to cellular extracts are bound to a variety of proteins and why the structural and dynamic properties of RNA in cellular environments tend to differ from those in vitro. Furthermore, the lack of sequence and structural specificity of the macromolecular counterion mechanism described here suggests that other positively charged intrinsically disordered viral proteins, such as NCP7 and Ta244,45 from HIV-1, as well as the core protein from the West Nile virus, might employ a similar mode of action. In
addition to modulating the affinity of nucleic-acid-binding proteins, positively charged intrinsically disordered regions may also represent a way to append this type of chaperone activity to transcription factors, nucleocapsid proteins, and histone-like proteins.

Although macromolecular charge screening is likely to be a widely applicable mechanism of intrinsically disordered RNA chaperones, there is of course a range of complementary mechanisms. For example, the structurally more well-defined Moloney murine leukemia virus nucleocapsid protein preferentially interacts with guanosine residues, preventing them from forming non-native contacts that would otherwise impede formation of the native state. The hexameric bacterial RNA chaperone Hfq makes use of multiple intrinsically disordered tails to facilitate release of RNA target molecules, while other RNA chaperones with ATP-dependent helicase activity unwind their substrates, leading to an iterative annealing mechanism. Despite the wide range of mechanisms employed by RNA chaperones, many of them are likely to complement the macromolecular counterion mechanism we propose here and may therefore cooperate in facilitating conformational transitions for a wide range of DNA and RNA molecules.

Methods

Preparation and labeling of nucleic acids. Both the hairpin (HP) and the non-complementary folding-incompetent hairpin (ncHP) were purchased as chemically synthesized deoxyoligonucleotides (Integrated DNA Technologies). One or two of the following chemical moieties were introduced during synthesis to facilitate dye attachment: a 5′ dithiol at position 0 (i.e., the phosphate preceding the first cytosine) for maleimide labeling, an amino-modified C6 diT towards the 3′ end (position 48) of the hairpin for succinimidyl ester labeling. In total, four different oligonucleotides were used to generate the seven different labeling permutations for HP; the same is true for ncHP (Supplementary Table 1). The RNA-based HP had a sequence similar to the DNA HP with similar synthetic modifications (Supplementary Table 2).

Prior to labeling, the synthetically incorporated thiol groups were reduced using 5 mM tris(2-carboxyethyl)phosphine. Either N-hydroxysuccinimide-functionalized or maleimide-functionalized donor (Cy3B; GE Healthcare) and acceptor (CF660R; Biotium) dyes were site-specifically incorporated at the desired positions in the oligonucleotides, which contained either primary amines or thiols, using the manufacturers’ suggested protocols. The labeled constructs were purified on a reversed-phase C18 column (Dr. Maisch) using a high-performance liquid chromatography (HPLC) system (Agilent 1100 series). Samples were lyophilized overnight, then dissolved in 50 mM sodium phosphate buffer (PB) pH 7.0 and stored at −80 °C until use.

Preparation and labeling of NCD. Gene variants of a clinical isolate of NCD (GenBank: CAE46584.1) were cloned into a pET-47b (+) vector, transfected into BL21 (DE3) E.coli cells (New England Biolabs). After 3 h of growth at 37 °C, recombinant expression was induced via 0.5 mM isopropyl β-D-1-thiogalactopyranoside for a total of for 3 h at 37 °C. In the sequence, the terminal (134) tryptophan residue was removed to prevent dye quenching; the only natural cysteine at position 98 was changed to methionine; and either one or two cysteine residues were introduced using site-directed mutagenesis at the desired labeling positions (S2, T65, and S130), resulting in 6 different amino acid sequences (Supplementary Table 3). The correct mass of all proteins was confirmed via electrospray ionization mass spectrometry (Functional Genomics Center Zurich). See Supplementary Table 7 for the sequences of all primers used in this work.

Immunobilized metal-ion affinity chromatography and HPLC were used to purify the recombinantly expressed proteins. The proteins were then stoichiometrically labeled with maleimide-functionalized donor and/or acceptor dyes and purified on a reversed phase C18 column (Dr. Maisch) using HPLC to remove double donor and double acceptor species. Samples were lyophilized overnight, then dissolved in 50 mM sodium phosphate buffer (PB) pH 7.0 and stored at −80 °C until use.

Free-diffusion single-molecule FRET spectroscopy. Free-diffusion experiments were performed on a modified version of a confocal single-molecule instrument (MicroTime 200, PicoQuant) using pulsed interleaved excitation and time-correlated single photon counting. Direct excitation of the donor was accomplished using 100 μW (measured at the back aperture of the microscope objective) of filtered light (520 ± 5 nm) from a pulsed (29 MHz) continuum laser (SC400, Fluorham), which also triggered the 635-nm pulsed diode laser (PicoQuant) that directly excited (100 μW) the acceptor. Prior to being detected on one of four
avalanche photodiodes (AQQRH-series, PerkinElmer), photons were separated by polarization and color using a polarizing beam splitter cube (PicoQuant) and two dichroic mirrors (Chroma), respectively. All photon arrival time records were binned at 0.5 ms. The background-corrected numbers of acceptor \( N_X \) and donor \( N_D \) photons resulting from direct donor excitation were used to calculate the uncorrected transfer efficiency, \( E = \frac{N_X}{N_S + N_D} \), for each burst, and the total numbers of photons resulting from donor \( (N_X^D) \) and acceptor \( (N_X^A) \) direct excitation were used to determine the apparent fluorescence stoichiometry ratio, \( S = \frac{N_X^D}{N_X^A} \). Only bursts resulting from molecules with active donor and acceptor dyes (0.3 < S ≈ 0.7) were used to construct \( E \) histograms. Correction factors for cross talk, direct excitation, as well as differences in excitation and detection efficiencies were calculated from the mean \( E \) and \( S \) values of the three populations present in the sample (active FRET pair at \( S = 0.5 \), active donor only at \( S = 1.0 \), and active acceptor only at \( S = 0.0 \)) and used to calculate corrected values for \( E \), the transfer efficiency, and \( S \), the stoichiometry ratio. For intermolecular FRET experiments conducted in low-salt buffer, we occasionally observed 1:2 hairpin-chaperone complexes (Supplementary Fig. 4b). In these cases, we only consider the transfer efficiency of the 1:1 species (at \( S = 0.5 \)) because we were primarily interested in the conformational properties of the 1:1 complex and because interpreting transfer efficiencies in the presence of multiple acceptors can be difficult. For intramolecular FRET experiments on the FRET-labeled hairpin conducted in low-salt buffer, we work at sub-saturating concentrations of chaperone to minimize the interference of these 1:2 complexes with our transfer efficiency measurements.

Recurrence analysis. For recurrence analysis, experiments were conducted on a custom-built confocal single-molecule fluorescence instrument, where 100 μW of light from a continuous-wave 532-nm laser (LaserBoxx 532, Oxxius) was used to excite the donor, and 25 μW of light from a 5-MHz 635 nm pulsed laser diode (PicoQuant) was used to excite the acceptor. Photons were detected using four single-photon avalanche diodes (AQQRH-series, PerkinElmer) and recorded in a time-correlated single-photon counting module (HydraHarpp, PicoQuant). For the analysis, photon arrival times were binned in 250-μs bins. Only bins containing fluorescence from both donor and acceptor dyes were considered as bursts for recurrence analysis. Additional details regarding recurrence analysis can be found elsewhere. All data analysis was conducted using a custom WSTP add-on for Mathematica (Wolfram).

Surface-immobilization experiments. All surface-immobilization experiments of FRET-labeled HP were conducted on the MicroTime 200 (PicoQuant) used for the free diffusion measurements, but 1 μW (measured at the back aperture of the microscope objective) of light from a continuous-wave 532-nm laser (LaserBoxx 532, Oxxius) was used to excite the donor since pulsed excitation led to lower photostability. Biotin-neutravidin binding was used to non-covalently immobilize the donor and double-acceptor populations. Correlation functions were generated from 10-hour measurements using 100 μW of 488 nm continuous-wave excitation (measured at the back aperture of the microscope objective) to directly excite the donor. All measurements contained approximately 50 pM FRET-labeled NCD in low-salt buffer. Only photons arising from the FRET-labeled population (i.e., \( E > 0.2 \)) were considered for the correlation analysis. Lag times ranging between \( t = 150 \) ns and \( t = 1500 \) ns from the acceptor autocorrelation function, \( g_2(\tau) \), the donor autocorrelation function, \( g_1(\tau) \), and the donor-acceptor crosscorrelation function, \( g_{12}(\tau) \), were globally fit to a model with two exponential components for each correlation:

\[
g_{12}(\tau) = a_1 + A_{AB} e^{-(\tau/	au_{AB})} + a_2 + A_{CD} e^{-(\tau/	au_{CD})},
\]

where the amplitude \( a_1 \) depends on the mean number of molecules in the confocal volume, triplet blinking, and the background signal. Small differences in the temporal alignment of the donor and acceptor channels are accounted for with the time offset, \( \tau_0 \). In this model, \( \tau_{AB} \) is a fast (< 5 ns), uncorrelated component (i.e., \( A_{AB} = 0 \)) for photon antibunching (AB), and \( \tau_{CD} \) is a slower (> 10 ns) component shared by all correlations and corresponding to chain dynamics (CD). This slower component is anti-correlated in the donor-acceptor (DA) crosscorrelation function (i.e., \( A_{CD} < 0 \)), as expected for fluctuations in transfer efficiency due to distance dynamics. The reconfiguration time \( (\tau_r) \) we report accounts for the fact that \( \tau_{CD} \) does not represent the direct reconfiguration from one state to the other and is instead an additional decay function; additional details regarding the nanosecond correlation spectroscopy are presented elsewhere. All data analysis was conducted using a custom WSTP add-on for Mathematica (Wolfram).

Analysis of binding equilibria. All binding equilibria were analyzed using a 2-state Hill-type binding isotherm with the following functional form:

\[
f(x) = f_0 + \frac{f_1 - f_0}{x + \frac{x_0}{y_0}}.
\]

where \( f \) is the experimental observable (e.g., \( E \) or fraction bound), and \( x \) the experimental variable (e.g., NaCl or NCD concentration). The fit parameters \( f_0 \) and \( f_1 \) correspond to the asymptotic limits of \( f(x) \) at low and high concentration, respectively. The fit parameter \( y \) represents the midpoint (e.g., apparent dissociation constant) of the transition from \( f_0 \) to \( f_1 \) and is defined as \( \frac{x_0}{y_0} = \frac{1}{f_1 - f_0} \). Finally, the fit parameter \( f \) corresponds to the sensitivity of the transition to concentration and is defined as \( \frac{df}{df} / \frac{dx}{dx} \), where \( K \) represents the apparent equilibrium constant. When the experimental variable is salt (e.g., NaCl), as in the case of washing in the Supplementary Fig. 3, the dissociation constant is defined as \( K = \frac{[x_0]}{[x]} \).
Coarse-grained representation and energy function. In the coarse-grained model, each residue of the protein NCD was represented by a single bead (of type C) located at the position of the alpha-carbon atom for all simulations involving DNA. This parameter was further fixed for all simulations involving DNA. The energy function was given by a standard force field expression:

\[ E = \sum_{\langle i,j \rangle} k_{ij} (r_{ij} - \sigma_{ij})^2 + \sum_{\langle i,j \rangle} \epsilon_{ij} \sigma_{ij} \left( \sigma_{ij}^{12} - \sigma_{ij}^{6} \right) + \sum_{\langle i,j \rangle} \epsilon_{ij} \sigma_{ij} \left( \frac{1}{r_{ij}} - \frac{1}{\sigma_{ij}} \right) \]

(4)

The first two terms of Eq. 4 account for the covalently bonded structure of the chain, with \( r_{ij} \) being the distance between atoms \( i \) and \( j \), \( k_{ij} \) is the harmonic bond force constant, \( \sigma_{ij} \) the angle between bonded atoms \( i, j, \) and \( k \), and \( \epsilon_{ij} \) is the harmonic angle force constant. The ideal values of the bond lengths and angles (\( r_{ij0} \) and \( \sigma_{ij0} \), respectively) and the force constants are given in Supplementary Table 4. The third term is a stacking energy that is defined for all base pairs (B) that are adjacent in sequence, irrespective of whether they are involved in forming duplex structure. The fourth and fifth terms of Eq. 4 correspond, respectively, to a generic non-bonded term for atom pairs not in contact in the native DNA duplex and not involved in stacking, and a Go-like structure-based term for those atom pairs in contact in the DNA duplex. The generic non-bonded term is a function of the parameters \( \epsilon_{ij} \) and \( \sigma_{ij} \) which depend only on the types of beads interacting (i.e., protein bead C, or P, R or B bead of DNA). The corresponding interaction matrix is given in Supplementary Table 5. The Go-like term is defined for beads which form base pairs in the duplex structure, and was only used for the DNA sequence capable of forming a duplex. The set of native interaction parameters \( \epsilon_{ij} \) and \( \sigma_{ij} \) for each pair of base-paired residues is given in Supplementary Table 6. Note that, by symmetry, analogous native contacts are implicitly defined with \( i \) and \( j \) permuted. The native contact distances are based on idealized B-form Watson-Crick DNA. The \( \epsilon_{ij} \) and \( \sigma_{ij} \) are, respectively, the permittivity of free space and the relative dielectric of the medium, taken to be 80. Note that the only sequence-specific features of the protein in the model are the residue charges. For the DNA, the only sequence-dependent feature is the presence of the complementary strands in the foldable sequence, which are included via the structure-based term.

Comparing simulations and experiments. Transfer efficiencies were calculated from the coarse-grained molecular dynamics simulations using the distribution of inter-residue distances. Because the reconfiguration dynamics of the complex occur on a 50-ns time scale, we assume that the distribution of inter-residue distances, \( P(r) \), is sampled on a time scale that is much slower than the fluorescence lifetime of the donor (e.g., 3 ns) and that the rotational correlation time of the dye is much faster than the fluorescence lifetime of the donor. This gives rise to the following expression for the mean transfer efficiency \( E = \int P(r) \, dr \), where \( E(r) = 1 + (\Gamma(r) R_{0})^{-1} \) and \( R_{0} = 59 \) Å is the Förster radius of the Cy3B-C660R FRET pair.

Model parameter optimization. The bonded parameters (bonds, angles) were based, in the case of the protein, on an extended random-coil structure, and in the case of the DNA, on an idealized Watson-Crick duplex structure. These parameters were fixed and not subject to optimization. The non-bonded radius for all protein and DNA beads was also fixed to 0.6 nm. This value is approximately consistent with the volume of a typical amino acid, ribose or base. The single protein-protein non-bonded energy parameter was adjusted to match the FRET efficiencies of the protein in isolation and was fixed to that value in all subsequent simulations. The obtained value of 0.15 \( k_{B}T \) is almost identical to the 0.16 \( k_{B}T \) used for protein-protein interactions in a closely related model we used to interpret recent FRET experiments on the interaction of two highly charged proteins. For the DNA, the non-bonded interaction energy \( \epsilon_{ij} \) was set to a small value for most beads, except for adjacent bases, which are known to form favorable base stacking interactions. The single stacking pair energy parameter between sequence-neighboring bases was optimized in order to reproduce the FRET efficiencies of the isolated, folding-inefficient DNA (without native interactions defined). This led to an average stacking free energy of \(-8.79 \text{ kJ mol}^{-1}\), which is consistent with other estimates of the free energy gain by base stacking. This parameter was then fixed for all simulations involving DNA. The last parameter optimized to match FRET data was a single protein-DNA interaction energy, common to all DNA beads, which was determined to match the intermolecular FRET data for the complex of NCD with the unfolded DNA. No adjustable parameters were added for the folded DNA. For comparison with experimental FRET efficiencies of the folded DNA in isolation and in complex with the NCD, the strength of each duplex native contact was set to a sufficiently large value (5.0 \( k_{B}T \)) that the DNA was essentially always folded. For simulations of equilibrium folding dynamics, a smaller native contact energy of 3.279 \( k_{B}T \) was used.

Simulation details. For most simulations, unbiased Langevin dynamics were run at 300 K with Gromacs version 5.1.4 or 4.0.2 (www.gromacs.org), using a time step of 10 fs and a friction coefficient of 0.1 ps \(^{-1}\). All bond lengths were constrained to their initial values. A cubic box of side 40 nm was used for all simulations of equilibrium folding dynamics of the DNA. Folding kinetics were estimated from long equilibration simulations of the NCD and DNA together. A four-state kinetic model comprising the states U, F, UC, and FC (Fig. 2f) was fitted to the simulation data by assigning transitions using a transition-based or core state assignment. The trajectory was initially unassigned until it reached one of the core states. Then it only changed assignment once it reached a different core state. The core sets were defined as Core_U = Unfolded ∩ Bound, Core_F = Folded ∩ Bound, Core_UC = Folded ∩ Unbound, Core_FC = Unfolded ∩ Unbound. The folded and unfolded sets were defined as \( (x, Q(x) > 0.7) \) and \( (x, Q(x) < 0.2) \), respectively, with \( Q(x) \), the fraction of native hairpin contacts for conformation \( x \), defined as:

\[ Q(x) = \sum_{\langle i,j \rangle} 1 \exp \left( -\frac{1}{\beta} \left( r_{ij} - \sigma_{ij} \right) \right) \]

Here, the sum runs over DNA native contact pairs \( (i, j) \) defined as in Supplementary Table 6, and the constants \( \beta \) (set to 30 \( \text{nm}^{-1} \)) and \( \gamma \) (set to 1.4), respectively, control the steepness of the switching function and account for fluctuations within the native state. The Bound core set was defined by \( (x, d_{un}(x) < 0.7 \text{ nm} \land n_{m}(x) < 0.4) \), where \( d_{un}(x) \) is the minimum of all pair distances between protein and DNA beads in conformation \( x \), and \( n_{m}(x) \) is the number of intermolecular contacts (i.e., number of protein-DNA atom pairs less than 1.0 nm apart). Similarly, the unbound set was defined as \( (x, d_{un}(x) > 3 \text{ nm} \land n_{m}(x) < 0.4) \). All transitions with free energies of transitions between states were determined using a lifetime-based method.

To mimic the confining effect of the chaperone on the DNA, a bias potential was built by fitting a smooth curve \( V_{bias} \) to the difference between the free energy surface projected onto the radius of gyration, \( r_{G} \), and the isolated DNA and the DNA in complex with the chaperone (note that we follow the convention here of using lower case \( r_{G} \) to refer to the radius of gyration of an individual conformation, and uppercase \( R_{G} \) for the ensemble average). The exact bias used was (with \( V_{bias} \) in units of \( k_{B}T \) and \( r_{G} \) in nm):

\[ V_{bias}(r_{G}) = \begin{cases} -49.302 + 27.525r_{G} - 4.7448r_{G}^{2} + 0.27657r_{G}^{3}, & \text{for } r_{G} < 5.7186 \\ 2.4204 + 0.39127r_{G}, & \text{otherwise} \end{cases} \]

which is continuous and differentiable everywhere. This bias was then applied as an external potential to the radius of gyration of the DNA without chaperone using PLUMED\(^{68-69} \) (www.plumed.org). The potential was implemented using cubic splines with control points defined every 0.005 nm using the above equation. The folding equilibrium and kinetics of the DNA in the presence and absence of this external potential were compared to the corresponding kinetics in the presence of the NCD. A similar lifetime-based kinetic model was used in this case, with only two states defined using folded and unfolded as core sets.

Salt-induced compaction of unfolded hairpin. The salt-induced compaction of the unfolded DNA is fit using a worm-like chain model\(^{66-71} \), which describes the end-to-end distance \( (r) \) distribution of a polymer chain in terms of the persistence length \( (l_{p}) \), the contour length \( (L) \), and a normalization constant \( (C) \):

\[ P(r) = 4\pi C^{2} \frac{\xi}{\left(1-(r/L)^{2}\right)^{5/2}} \exp \left( -\frac{3l_{p}}{4L} \left(1-\frac{r}{L}\right)^{2} \right) \]

(7)

Odijk-Skolnick-Fixman (OSF) theory\(^{22,23} \) describes how the electrostatic component of the persistence length \( \xi_{es} \) of a polyelectrolyte depends on the ions of the solution via \( \xi_{es} \propto \frac{C_{i}^{1/2}}{\gamma_{i}^{4/3}} \), where \( \gamma_{i} \) is the Bjerrum length, \( \kappa^{-1} \) is the Debye length, and \( A \) is the effective segment length in the absence of electrotaxis. The total effective persistence length is thus \( l_{p} = l_{p0} + A \), and the contour length of the chain is \( L = N \lambda_{sp} \), where \( N \) is the number of effective chain segments. Since only the values of \( N \) and \( A \) are known for our unfolded DNA, they are treated as fit parameters, while any given combination of \( N \) and \( A \) gives rise to a unique salt dependence of \( (r) \). Although the best fit gives rise to a chain with \( N = 210 \pm 10 \) small segments \( (A = 0.05 \pm 0.01 \) nm), the product of
these two values is 10.5 nm, which is a reasonable contour length for our single-stranded nucleic acid.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. The source data underlying mean values and gel images in Figure panels 1c, d, 2a–d, 3c, e, and 4c, d in addition to Supplementary Figure panels 1c, 2a, b, 4a, 5b, 6a, b, and 10b, c, f are provided as a Source Data file.

**Code availability**

Example input files for running the coarse-grained model in Gromacs (www.gromacs.org) are provided in a compressed archive (initial coordinates, topology files, run input and auxiliary files, and a summary README file) as Supplementary Software (Simulation_Models.tgz). A custom WSPF add-on for Mathematica (Welfram Research) used for the analysis of single-molecule fluorescence data is available upon request.

Received: 29 January 2019 Accepted: 6 May 2019

**Published online: 05 June 2019**

**References**

1. Breaker, R. R. & Joyce, G. F. The expanding view of RNA and DNA function. *Chemistry & Biol.* 21, 1059–1063 (2014).
2. Thirumalai, D. & Hyeon, C. RNA and protein folding: common themes and variations. *Biochemistry* 44, 4957–4970 (2005).
3. Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270, 20871–20874 (1995).
4. Chen, S. J. & Dill, K. A. RNA folding energy landscapes. *Proc. Natl Acad. Sci. USA* 97, 646–651 (2000).
5. Rauchfuss, L. et al. RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* 4, 118–130 (2007).
6. Ditzler, M. A., Rueda, D., Mo, J., Hakansson, K. & Walter, N. G. A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* 36, 7088–7099 (2008).
7. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 463, 681–684 (2010).
8. Woodson, S. A. Taming free energy landscapes with RNA chaperones. *RNA Biol.* 7, 677–686 (2010).
9. Hyeon, C. & Thirumalai, D. Generalized iterative annealing model for the action of RNA chaperones. *J. Chem. Phys.* 139, 121924 (2013).
10. Tijerina, P. & Russell, R. in Russell, Rick. *Biophysics of RNA Folding* (Springer, New York, 2013) 205–230.
11. Woodson, S. A., Panja, S. & Santiago-Frangos, A. Proteins that chaperone RNA regulation. *Microbiol. Spectr.* (2018).
12. Yang, J., Xia, H., Qian, Q. & Zhou, X. RNA chaperones encoded by RNA. *Science* 341, 34–36 (2013).
13. Zuniga, S., Solà, I., Cruz, J. L. & Enjuanes, L. Role of RNA chaperones in virus replication. *Virus Res.* 139, 253–266 (2009).
14. Grohman, J. K. et al. A guanosine-centric mechanism for RNA chaperone function. *Science* 340, 190–195 (2013).
15. Ivanyi-Nagy, R., Laverence, J. P., Gabus, C., Ficheux, D. & Darlix, J. L. RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic Acids Res.* 36, 712–725 (2008).
16. Longhi, S. Structural disorder in viral proteins. *Protein Pept. Lett.* 17, 930–931 (2010).
17. Xue, B. et al. Structural disorder in viral proteins. *Chem. Rev.* 114, 6961–6980 (2014).
18. Dervignaud, J. B. et al. Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein. *Biochem. Biophys. Res. Commun.* 378, 27–31 (2009).
19. Holmstrom, E. D., Nettels, D. & Schuler, B. Conformational plasticity of hepatitis C virus core protein enables RNA-induced formation of nucleocapsid-like particles. *J. Mol. Biol.* 430, 2453–2467 (2018).
20. Thomas, D. L. Global control of hepatitis C: where challenge meets opportunity. *Nat. Med.* 19, 850–858 (2013).
21. Ivanyi-Nagy, R. et al. Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res.* 34, 2618–2633 (2006).
22. Majeau, N. et al. The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. *J. Gen. Virol.* 85, 971–981 (2004).
52. Müller, B. K., Zaychikov, E., Bräuchle, C. & Lamb, D. C. Pulsed interleaved excitation. Biophys. J. 89, 3508–3522 (2005).
53. Kudryavtsev, V. et al. Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements. ChemPhysChem 13, 1060–1078 (2012).
54. Lee, N. K. et al. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys. J. 88, 2939–2953 (2005).
55. Holmstrom, E. D. et al. Accurate transfer efficiencies, distance distributions, and ensembles of unfolded and intrinsically disordered proteins from single-molecule FRET. Methods Enzymol. 611, 287–325 (2018).
56. Vogelsang, J. et al. A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. Angew. Chem. Int. Ed. Engl. 47, 5465–5469 (2008).
57. Krainer, G., Hartmann, A. & Schlierf, M. farFRET: extending the range in single-molecule FRET experiments beyond 10 nm. Nano Lett. 15, 5826–5829 (2015).
58. Gopich, I. V. & Szabo, A. Decoding the pattern of photon colors in single-molecule FRET. J. Phys. Chem. B 113, 10965–10973 (2009).
59. Nettels, D., Gopich, I. V., Hoffmann, A. & Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc. Natl. Acad. Sci. USA 104, 2655–2660 (2007).
60. Nettels, D., Hoffmann, A. & Schuler, B. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. J. Phys. Chem. B 112, 6137–6146 (2008).
61. Cantor, C. & Schimmel, P. Biophysical Chemistry Part III: The Behavior of Biological Macromolecules. First edn, 859–866 (W.H. Freeman and Company, New York, 1980).
62. Record, M. T. Jr., Anderson, C. F. & Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q. Rev. Biophys. 11, 103–178 (1978).
63. Schuler, B. Single-molecule FRET of protein structure and dynamics—a primer. J. Nanobiotechnol. 11(Suppl 1), 52 (2013).
64. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. 34, 564–574 (2006).
65. Abraham, M. J. M. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1, 19–25 (2015).
66. Buchete, N. V. & Hummer, G. Coarse master equations for peptide folding dynamics. J. Phys. Chem. B 112, 6057–6069 (2008).
67. Tribello, G. A., Bonomi, M., Branduardi, A., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. Comput. Phys. Commun. 185, 604–613 (2014).
68. Bonomi, M. et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. Comput. Phys. Commun. 180, 1961–1972 (2009).
69. O’Brien, E. P., Morrison, G., Brooks, B. R. & Thirumalai, D. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? J. Chem. Phys. 130, 124903 (2009).
70. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys. J. 86, 2530–2537 (2004).
71. Chen, H. et al. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. Proc. Natl. Acad. Sci. USA 109, 799–804 (2012).
72. Odijk, T. Polyelectrolytes near the rod limit. J. Polym. Sci. Part B 15, 477–483 (1977).
73. Skolnick, J. & Fixman, M. Electrostatic persistence length of a wormlike polyelectrolyte. Macromolecules 10, 944–948 (1977).

Acknowledgements

The authors thank Franziska Zosel for helpful discussions regarding data analysis and the Functional Genomics Center Zurich for expert mass spectrometry analysis. Support for this work was provided by the Swiss National Science Foundation (to B.S.) and the European Molecular Biology Organization (to E.D.H., ATLF 471-2015). R.B.B. was supported by the Intramural Research Program of the NIDDK at the National Institutes of Health. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Author contributions

The project was conceptualized and supervised by E.D.H., R.B.B. and B.S. Data curation and analysis was carried out by E.D.H., Z.L and R.B.B. Research funds were acquired by E.D.H., R.B.B. and B.S. Scientific investigations were conducted by E.D.H., Z.L and R.B.B., with methodological developments provided by E.D.H., D.N., R.B.B. and B.S. Administrative aspects of the project were overseen by E.D.H. and B.S. The project resources were provided by R.B.B. and B.S. Software development was carried out by E.D.H., D.N. and R.B.B. Data validation was overseen by E.D.H., R.B.B. and B.S. Visualization of data was performed by E.D.H. and R.B.B. The original draft was written by E.D.H. with edits and revisions provided by E.D.H., D.N., R.B.B. and B.S.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10356-0.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks Philip Bevilacqua, Carlo Camilloni and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019