

RNA2PS: A sequence-specific coarse-grained RNA model linking structure, thermodynamics and phase separation

Andrés R. Tejedor^{†,1,2} Juan Luengo-Márquez^{†,1} Javier Oller-Iscar^{1,2} Jorge Ramirez³ Alberto Ocana^{4,5} Rosana Collepardo-Guevara^{5,6,7,a)} Pablo Llombart^{1,2,b)} and Jorge R. Espinosa^{1,2,6,5,c)}

¹⁾ *Department of Physical Chemistry, Universidad Complutense de Madrid, Av. Complutense s/n, Madrid 28040, Spain*

²⁾ *Instituto Pluridisciplinar, Universidad Complutense de Madrid, P.^o de Juan XXIII, 1, Moncloa - Aravaca, 28040 Madrid, Spain*

³⁾ *Department of Chemical Engineering, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006, Madrid, Spain.*

⁴⁾ *Experimental Therapeutics Unit, Hospital Clínico San Carlos (HCSC), Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain*

⁵⁾ *PhAsIca Biosciences S.L, Calle Velázquez, 27, 28001 Madrid, Spain*

⁶⁾ *Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*

⁷⁾ *Department of Genetics, University of Cambridge, Cambridge, CB2 3EH*

[†] *These authors contributed equally*

(Dated: 24 April 2026)

RNA plays a central role in the formation and regulation of biomolecular condensates, yet a quantitative understanding of how RNA sequence, structure, and thermodynamics jointly determine phase behaviour, particularly in repeat expansion RNAs, remains incomplete. Here, we introduce RNA2PS, a sequence-specific RNA coarse-grained model for phase-separation that predicts RNA structure from sequence, achieving quantitative agreement for both single-stranded conformations and duplex helical geometry relative to crystallographic PDB structures. RNA2PS represents each nucleotide by two beads that separate the phosphate-ribose backbone from the base. This representation decouples electrostatic interactions from directional base pairing, while explicitly incorporating strand polarity (5' → 3') and local sequence context at the trimer level. Canonical and wobble base pairing are modelled through a multi-body potential with sequence-dependent coordination. Importantly, RNA2PS captures sequence-dependent duplex stability at the nearest-neighbour level and reproduces experimental melting temperatures across a diverse set of sequences. RNA2PS shows that phase separation of trinucleotide repeat RNAs is governed by transient inter-strand duplexes that form reversible cross-links. Competition between intra- and intermolecular base pairing regulates the density of labile RNA-RNA interactions, giving rise to strong sequence- and length-dependent differences in condensation that reproduce cellular RNA foci formation. Overall, RNA2PS provides a near-quantitative predictive framework that links sequence-encoded hybridization thermodynamics to mesoscale condensation of pure RNA sequences.

Keywords: RNA, molecular dynamics simulations, coarse-grained models, biomolecular condensates, phase transitions

I. INTRODUCTION

Biomolecular condensates are a fundamental organizing principle of cellular biochemistry^{1–3}. These dynamic compartments regulate a wide range of essential biological functions, including ribosome biogenesis in the nucleolus^{4,5}, mRNA storage in stress granules^{6–8}, and RNA processing and decay in the P-bodies^{9,10}. While early studies of biomolecular condensates focused primarily on proteins—particularly those containing intrinsically disordered regions (IDRs)—as the main regulators of condensate biophysics, it is now clear that RNA also modulates condensate structure, composition, and material

properties^{11–16}.

In RNA-protein condensates, RNA acts as a key modulator of phase behaviour through both specific and non-specific interactions with RNA-binding proteins^{17,18}. In particular, it can engage both well-defined RNA-binding domains, such as RNA recognition motifs (RRMs), which provide sequence- and structure-specific binding interfaces, and IDRs enriched in positively charged and polar residues that mediate non-specific favourable electrostatic interactions with the RNA backbone^{19,20}. By transforming the intermolecular interactions within RNA-protein condensates, RNA also plays a central role in tuning their material properties^{21–24}. For instance, its incorporation has been shown to fluidize condensates of FUS, hnRNPA1 and positively charged peptides^{21,22,25,26}, while in other contexts it enhances intermolecular connectivity, promoting more viscoelastic or even gel-like behaviour^{23,24,27}.

^{a)} Electronic mail: rc597@cam.ac.uk

^{b)} Electronic mail: pllombart@ucm.es

^{c)} Electronic mail: jorgerene@ucm.es

Beyond its regulatory role in multicomponent protein–RNA assemblies, RNA is capable of undergoing intrinsic phase-separation, without the aid of additional proteins, driven by a combination of electrostatic and sequence-specific interactions^{28–31}. The polyanionic backbone of RNA makes electrostatics a central determinant of biomolecular condensate formation via phase separation—herein condensation—with screening by divalent cations such as Mg^{2+} facilitating assembly by reducing phosphate–phosphate repulsion^{28,32,33}. At the same time, base-dependent interactions—including π – π stacking, hydrogen bonding, and cation– π contacts—introduce an additional layer of specificity, promoting sequence-dependent cohesion and structural heterogeneity within condensates^{34–36}. In particular, RNA repeat expansion sequences provide a well-defined example of this behaviour, where both sequence composition and length govern condensate formation²⁸. Trinucleotide and hexanucleotide repeats such as $(CAG)_n$, $(CUG)_n$, and $(G_4C_2)_n$ undergo phase-separation above a critical repeat length and concentration^{29,30,33}, forming condensates that can evolve from liquid-like droplets to more rigid, gel-like assemblies at high concentrations, a transition directly linked to several neurodegenerative disorders^{32,37–40}.

Molecular dynamics (MD) simulations have become essential to unravel the microscopic interactions underlying biomolecular condensation, offering a level of detail that complements experimental observations across multiple length and time scales^{41–47}. Atomistic MD simulations with explicit solvent have shown to capture the conformational ensemble of RNA molecules, ionic screening, and specific intermolecular contacts with high fidelity^{45,48–51}. Because the high computational cost of atomistic approaches limits their predictive power for biomolecular condensates, coarse-grained models that preserve essential sequence-dependent information are widely used to access the larger system sizes and longer timescales relevant to these systems^{52–58}. Within these sequence-dependent coarse-grained models, a simple representation of RNA is to treat it as a generic polyanionic flexible or semi-flexible polymer that primarily captures charge-driven interactions, while largely neglecting sequence specificity, secondary structure, and base-dependent interactions^{55,59–61}. The recent CALVADOS-RNA model^{55,62} is a sequence-independent, two-bead-per-nucleotide representation of disordered RNA that separates charged backbone interactions from base-mediated short-range interactions, enabling simulations that capture general aspects of RNA–RNA and protein–RNA interactions while remaining agnostic to RNA sequence and persistent secondary structure.

A notable advance in coarse-grained RNA modelling for biomolecular condensates has been the explicit incorporation of sequence-dependent base-pairing interactions in the single-interaction-site (SIS) model^{63,64}. The SIS model represents each nucleotide as a single bead and employs a many-body potential to encode

canonical base pairing, enabling simulations of sequence-dependent RNA condensation under effectively charge-neutral conditions. Simulations are typically performed using low-friction Langevin dynamics to accelerate sampling, which is advantageous for exploring the slow kinetics of RNA clustering. Building on this framework, the many-body base-pairing potential was reformulated with analytical forces and implemented in LAMMPS⁶⁵, significantly improving computational efficiency and enabling systematic investigations of sequence-dependent clustering across RNA repeat sequences. While these SIS-based approaches provide a powerful framework for studying sequence-dependent, protein-free RNA condensation, their accelerated dynamical scheme make them not directly compatible with current sequence-dependent protein coarse-grained models for biomolecular condensates, such as the Mpipi, CALVADOS and HPS families^{52,57,60,61}, without additional reparameterisation, thereby restricting its applicability to pure RNA condensation^{63,65}. At a resolution intermediate between minimal one- or two-bead models and atomistic descriptions, the iConRNA coarse-grained model⁶⁶, which employs multiple beads per nucleotide and explicit ions, has demonstrated improved accuracy in reproducing RNA conformations, folding, and condensation. However, this increased level of detail comes at a significant computational cost, complicating its application for simulations of multicomponent RNA–protein condensates.

Together, these advancements set the stage to develop RNA models that simultaneously capture sequence-dependent interactions and realistic structural behaviour, while remaining computationally efficient and compatible—in terms of resolution, interaction range, and dynamical timescales—with established coarse-grained protein frameworks for biomolecular condensates^{60,67,68}. In this context, we introduce RNA2PS (RNA to Phase Separation), an RNA coarse-grained model for biomolecular phase separation designed to operate at a comparable level of coarse graining, with interaction potentials amenable to integration with protein models such as Mpipi, CALVADOS, and HPS^{52,54,55,57,60,61,69}. RNA2PS is expected to enable applications to multi-component RNA–protein condensates, although this remains to be systematically assessed.

RNA2PS (RNA-to-Phase-Separation) is a coarse-grained model designed to capture the essential structural and thermodynamic properties of RNA within a minimal two-site-per-nucleotide representation. The model considers implicit solvent and incorporates explicit backbone connectivity together with effective base–base interactions that account for π – π stacking, hydrogen bonding, and directional base pairing, enabling the formation of both single-stranded and double-helical conformations in a fully dynamic manner. Inspired by modern sequence-dependent coarse-grained models of DNA^{70–72}, RNA2PS incorporates local sequence specificity by modulating base-pairing interactions according

to the nearest-neighbour context up to the trimer level. This local parameterisation captures the influence of adjacent bases on stacking, steric constraints, and binding competition, thereby encoding cooperative effects that are essential for reproducing both local structural features and emergent collective phase behaviour. The model is quantitatively calibrated against experimental data at multiple levels: it reproduces RNA structural features from crystallographic PDB structures and captures sequence-dependent melting temperatures across a diverse set of sequences, including mismatches. In addition, RNA2PS accurately reproduces RNA condensation propensity across different sequences, including repeat expansion systems, capturing experimental trends in condensation and phase behaviour. In particular, the model predicts critical temperatures for RNA phase-separation within biologically relevant ranges and with significantly improved accuracy compared to previous approaches for RNA repeat condensates. By bridging structural fidelity, thermodynamic accuracy, and computational efficiency, RNA2PS provides a unified framework to investigate RNA-driven phase separation and establishes a foundation for future studies of RNA-containing biomolecular condensates.

II. RESULTS

A. RNA2PS: A two-site coarse-grained model for RNA condensation

We present RNA2PS, a coarse-grained MD model of RNA designed to capture RNA self-assembly and condensation. RNA2PS employs a two-bead-per-nucleotide representation, in which one bead is placed at the geometric centre of the phosphate-ribose group (hereafter phosphate bead, P), while the second is positioned 4 Å along the vector connecting P to the centre of the nitrogenous base (hereafter base bead, B). In contrast to previous models^{63,65}, this representation explicitly decouples electrostatic interactions—localized on the phosphate beads—from the interaction sites responsible for base pairing and cross-stacking, which are carried by the base beads. This separation introduces additional structural flexibility, enabling an accurate description of RNA secondary structures as well as local conformational features of the RNA backbone in single-stranded chains.

In RNA2PS, the intramolecular primary structure is defined through bonded interactions between consecutive phosphate beads (P–P), between phosphate and base beads within the same nucleotide (P–B), and between bases of adjacent nucleotides along the backbone (B–B), effectively encoding the phosphodiester connectivity of the chain (see Section IV and Fig. 1A). Backbone flexibility is primarily controlled by a harmonic angular potential between P beads, parametrised to reproduce the characteristic conformational ensemble of single-stranded RNAs. Within this framework, RNA2PS sup-

ports the formation of both intra- and inter-chain base pairs, including canonical (A–U and C–G) and wobble (G–U) interactions, modelled via a many-body potential⁶³ (Fig. 1B, top). Base-pair formation is further coupled to angular and dihedral constraints that promote double-helical geometry upon binding, in contrast to the more flexible nature of single-stranded RNAs (Fig. 1B, top).

The effective interaction strength of each base pair is determined not only by the identities of the interacting nucleotides, but also by their local sequence context up to the trimer level (see Fig. 1B, bottom). In particular, the interaction energy is modulated by the identity of the nearest neighbours flanking each base, distinguishing between neutral neighbours that do not compete for binding and competing neighbours that introduce alternative pairing configurations (please see Table S1 for the full list of nucleotide types and their interaction parameters). This context dependence encodes an additional layer of effective cooperativity and competition, whereby neighbouring bases can either stabilise or frustrate pairing, thereby promoting collective binding along the RNA strand. Importantly, it also enables RNA2PS to distinguish strand polarity 5′–3′, favouring antiparallel base-pairing geometries in a physically consistent manner. This strategy is conceptually inspired by state-of-the-art sequence-dependent coarse-grained DNA models such as the CGenArate⁷¹ and MADna⁷⁰ models, where local sequence context is incorporated to reproduce structural and mechanical properties with high fidelity. In RNA2PS, this principle is extended to RNA condensation, enabling the model to capture how subtle sequence variations propagate to large-scale conformational and thermodynamic properties while retaining computational efficiency.

In addition to Watson–Crick base pairing, RNA strands can interact through non-canonical pairing⁷³ and base–base stacking interactions⁷⁴. In RNA2PS, stacking interactions between non-adjacent bases are captured using a Wang–Frenkel potential⁷⁵, which depends on the distance between base beads and their biochemical identity. Importantly, this potential also encodes an intrinsic directionality bias that favours correct 5′–3′ alignment between interacting strands, thereby stabilising antiparallel configurations consistent with canonical RNA duplex geometry. This interaction plays a dual role: it drives the condensation of RNA sequences that do not form explicit canonical base pairs, and modulates base pairing by both competing with canonical interactions—through the stabilisation of alternative contacts—and facilitating base-pair formation by promoting favourable spatial proximity, orientation, and registry between interacting bases.

The parametrisation of RNA2PS follows a hierarchical strategy. First, structural parameters are optimised to reproduce fundamental topological features derived from experimentally resolved RNA crystal structures (Fig. 1C), while higher-order structural properties, such

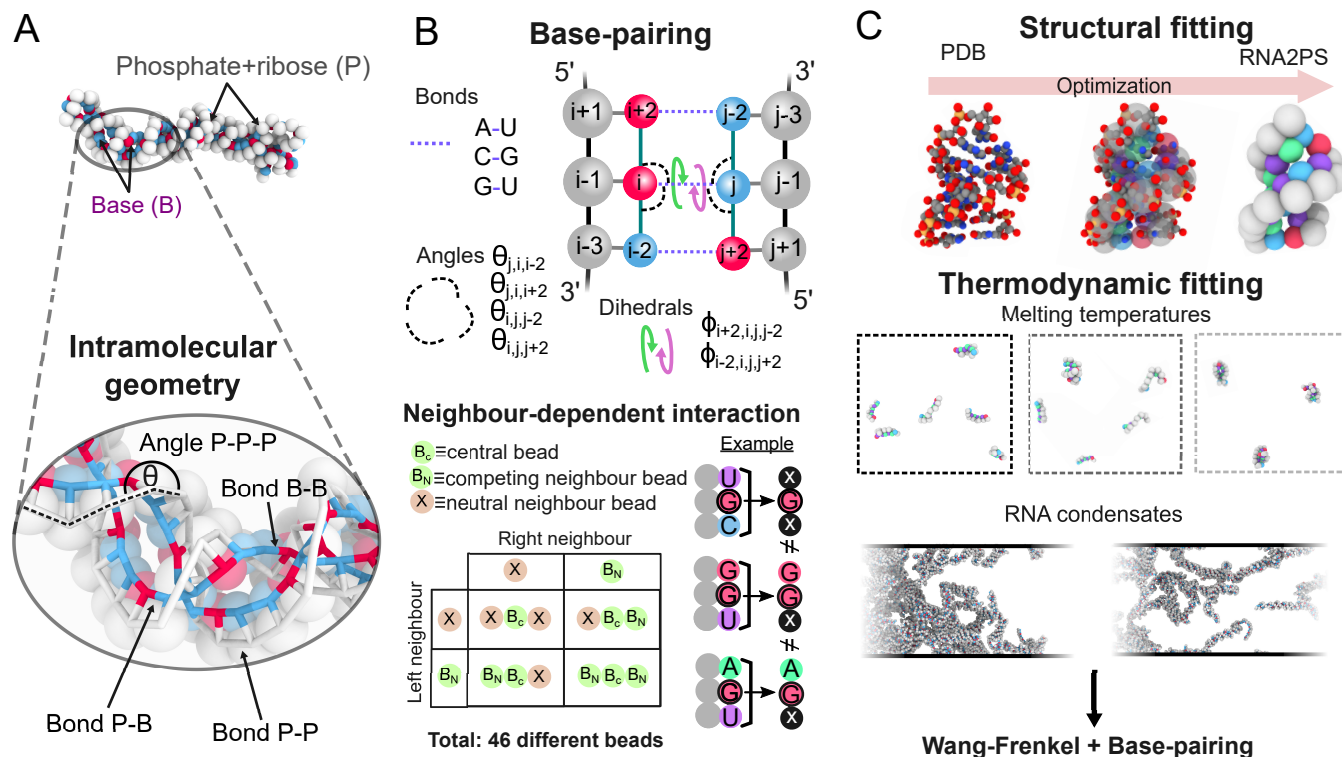


FIG. 1. RNA2PS model to describe RNA phase behaviour with two-bead-per-nucleotide resolution. **A.** (Top) Sketch of the RNA2PS model intramolecular interaction sites to describe the RNA structure. The phosphate+ribose bead (P) is represented in gray and a generic base bead (B) is depicted in blue and red. (Bottom) Zoom on the intramolecular interactions sustaining the geometry of double-stranded RNAs, including the bonds and the angle of the backbone. **B.** (Top) Breakdown of the multi-body potential inducing RNA base-pairing. This includes the different bonds, angles, and dihedrals. (Bottom) Schematic table showing the implemented interaction depending on the nature of the nearest neighbours. We present the different four cases considering a central bead of a triad (B_C) with neutral neighbour (X) and competing neighbour (B_N) combinations. We provide an example with G with no competing neighbours (C, U), with one G neighbour (binding to C and U) and with one A neighbour (binding to U). In Table S1 we provide the full list of interacting sites. **C.** (Top) We use PDB-based experimentally resolved crystal structures (e.g. 4JAB) to fit the geometric parameters of both intramolecular and base-pairing interactions. (Bottom) We fit experimental melting temperatures (e.g. GUAUAUAC) and condensation propensity of RNA repeats (e.g. CAG₄₀) to obtain the interaction strength of the base-pairing multi-body potential and van der Waals (Wang-Frenkel potential) interactions, respectively.

as specific dihedral distributions, emerge without explicit fitting (Section II B). Second, interaction parameters governing base-pairing and stacking (Wang-Frenkel) are tuned to reproduce experimentally measured RNA duplex melting temperatures (Section II C) and the association propensity of RNA repeat sequences that do not form canonical base pairs (Section II D). As discussed in Section II C, this parametrisation enables the model to quantitatively reproduce experimental trends in phase separation of base pairing RNA repeat sequences, as well as the dependence of condensate stability on sequence repeat length. In the present parametrisation and applications (Section II D), electrostatic interactions, modelled via a Yukawa potential, are neglected to mimic high-salt conditions (e.g., in the presence of Mg^{2+}) under which pure RNA phase-separation is observed²⁸. Nevertheless, the modular design of RNA2PS—combining electrostatic and sequence-specific dispersive interactions—makes it readily extensible to the study of RNA-protein mixed

condensates under physiological conditions.

B. Structural parametrisation and validation of RNA2PS for double-stranded helical RNA

The conformational properties of RNA play a central role in shaping its interaction landscape and its propensity to form condensates, with features such as chain flexibility, persistence length, and secondary structure directly influencing intra- vs. intermolecular interactions^{48,51,76}. Although the primary goal of RNA2PS is not high-resolution structural prediction, as in models such as oxRNA^{74,77}, it remains essential to capture a physically realistic spectrum of RNA conformations, ranging from semiflexible single-stranded sequences to double-stranded helical conformations. In this context, recent structure-prediction approaches such as FAR-FAR2⁷⁸, trRosettaRNA⁷⁹, and RoseTTAFoldNA⁸⁰ high-

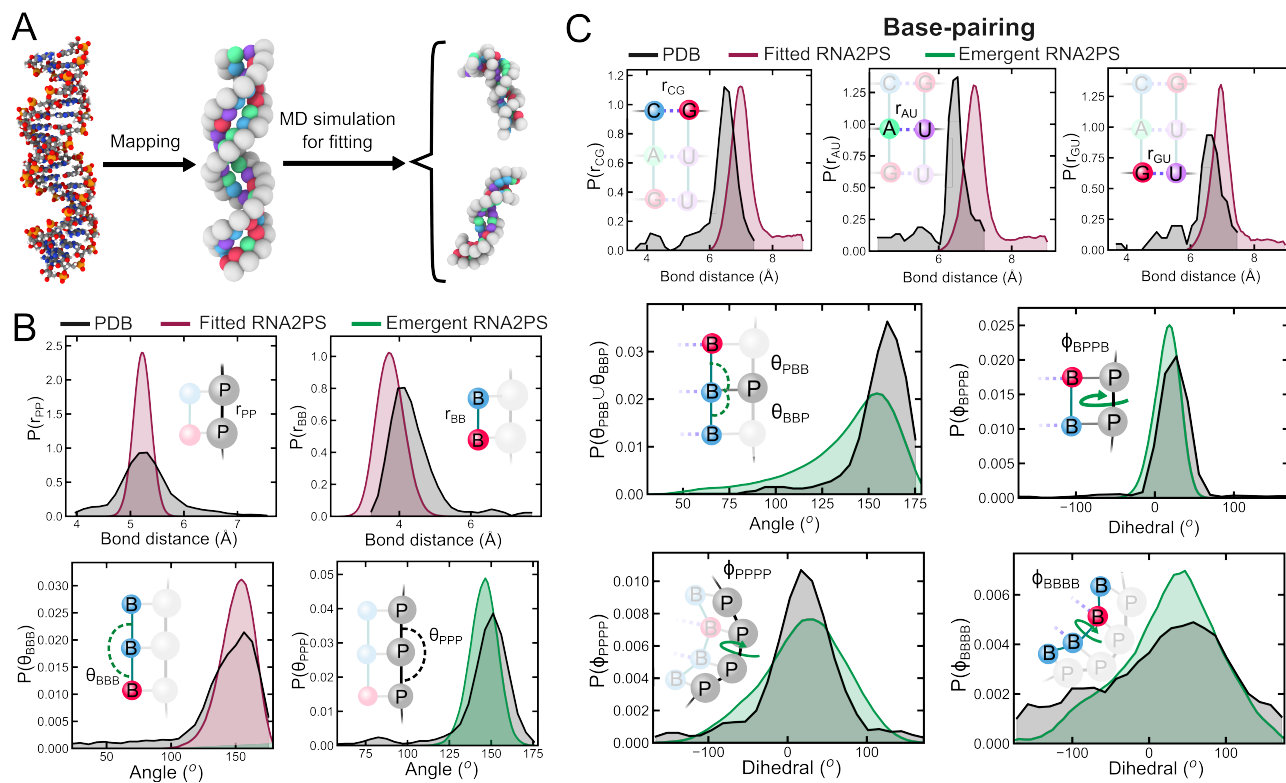


FIG. 2. Structural fitting of the model to crystal structures to recapitulate single- vs. double-stranded RNA conformations. **A.** Schematic representation of the structural fitting starting from the all-atom PDB, mapping to the 2-bead-per-nucleotides model and performing molecular dynamics simulations to obtain the statistics about molecular conformation. **B.** Normalized distributions of the intramolecular structure of RNA2PS including the explicitly parametrised and fitted bonds (P-P and B-B) and angles (B-B-B) in red, and the emergent geometric feature of P-P-P angle in green, compared to their analogue calculated from the PDB structures after mapping in black. **C.** Normalized distributions of local conformational parameters related to the base-pairing for the RNA2PS model. Red histograms represent the structural parameters explicitly included and fitted in RNA2PS (*i.e.* CG, AU and GU bonds), green distributions are obtained for emergent geometric properties from base-pairing in the RNA2PS model (*i.e.* BBP/PBB angles; and BPPB, PPPP, and BBBB dihedrals), and black histograms show the statistics for the homologue parameters computed from the crystal PDBs.

light the central role of base-pairing and stacking interactions in determining RNA structure, as well as their contribution to intermolecular association and condensation.

The RNA2PS model parameters are determined through a structure-based optimization procedure aimed at reproducing experimentally resolved RNA conformations from PDB structures (see Table S1 in the Supplementary Material for the codes of all the PDBs used). The workflow consists of two main steps: a coarse-grained mapping followed by parameter fitting (Fig. 2A). In the mapping stage, each nucleotide is represented by a minimal set of interaction sites, with the P-B bond length fixed at 4 Å to ensure structural stability of double-stranded RNA. This choice yields an average P-P distance of 15 Å along the helix, preserves consistent P-B-B-P geometries, and prevents unphysical strand crossing. In the fitting stage, a diverse set of RNA structures is considered, including canonical and wobble base pairs (*e.g.* 4JAB, 2L2J codes), as well as structural motifs such as hairpins and bulges (*e.g.* 1A4D, 2A43, 1DQF).

These structures are simulated and systematically compared to their corresponding crystallographically conformations, allowing the model interaction parameters to be optimized to reproduce structural features of both single- and double-stranded RNAs.

The resulting intramolecular structure shows that the model accurately reproduces key geometric features of RNA (Fig. 2B). In particular, the r_{PP} and r_{BB} distance distributions, together with the θ_{BBB} angle, correspond to observables that are directly fitted and explicitly defined in the model (see Section IV). In contrast, the θ_{PPP} angle (expected to be 152°) is not explicitly parametrised and instead emerges naturally from the model topology, while still being well reproduced. These structural descriptors are computed over nucleotides in both single-stranded and base-paired regions, ensuring a consistent description across different conformational states. Notably, the r_{BB} distribution obtained from simulations exhibits a slight shift toward more compact conformations compared to the experimental crystallographically resolved structures, consistent with the model design aimed

at promoting double-helical formation. Finally, since the P–B bond length is fixed at 4 Å during the mapping procedure, its distribution is reported from simulations in the SM (Fig. S1).

Regarding base-pairing interactions, we analyse the bond length distributions for canonical (CG and AU) and non-canonical wobble (GU) pairs (Fig. 2C). The resulting distributions exhibit slightly larger equilibrium distances compared to atomistic references, reflecting the dynamic nature of pairing imposed by the many-body potential, which operates beyond the excluded volume defined by van der Waals interactions (modelled here via the WF potential). To ensure a consistent characterisation, statistics are computed from MD simulations only when bases are paired—requiring at least one base to be paired for angle distributions and both bases for dihedral angles. Importantly, several structural features emerge naturally from the model. In particular, the θ_{PBB} (or equivalently θ_{BBP}) angles and the ϕ_{BPPB} dihedral are accurately reproduced, the latter being essential for capturing the correct helical twist of double-stranded RNAs⁵¹. Similarly, the backbone dihedral ϕ_{PPPP} and the base-stacking dihedral ϕ_{BBBB} show good agreement with reference structures, despite their intrinsic flexibility. Notably, RNA2PS explicitly parametrises angular and dihedral terms only for paired bases (see Fig. 1B and Section IV), while the remaining structural correlations emerge implicitly, highlighting the model’s ability to capture key geometrical RNA features with a minimal set of explicit constraints.

C. RNA2PS captures sequence-dependent RNA duplex stability via melting temperatures

Accurately predicting the thermodynamic stability of ds-RNA is essential for describing its conformational behaviour and its propensity to form pure homotypic condensates⁸¹. In RNA, this stability arises from a subtle interplay between base-pairing and stacking interactions, which together govern the formation and disruption of double-helical structures^{82,83}. To calibrate these contributions, RNA2PS is parametrised against experimental melting temperatures (T_m), which provide a direct measure of RNA duplex stability. Specifically, we consider a dataset of 33 ds-RNA sequences of varying lengths, including five sequences containing mismatches^{30,84}.

Fig. 3A shows the time evolution of the fraction of RNA single strands for the representative sequence GUAUAUAC at six different temperatures. At sufficiently high temperatures (dark red curve; 355 K), strands remain predominantly unpaired, whereas at low temperatures (dark blue curve; 280 K) the system undergoes continuous association and dissociation before reaching a well-defined steady state, characterised by an equilibrium between paired and unpaired populations. In this regime, the melting temperature T_m is obtained from the steady-state ensemble as the temperature at which

50% of the molecules are in the single-stranded state (see snapshot in Fig. 3A). The model parameters governing base pairing are then optimized to reproduce this equilibrium behaviour across a diverse set of sequences, ensuring a consistent description of thermodynamic stability across different RNA lengths and compositions^{84,85}. In Fig. 3B, we report the steady-state fraction of single strands as a function of temperature, showing the expected sigmoidal transition between bound and unbound regimes. For this sequence, the RNA2PS model shows an excellent agreement with experimental melting temperature⁸⁶ (blue curve). In contrast, comparison with the SIS model (red curve)⁷ reveals a systematic underestimation of the melting temperature by up to ~ 100 K. This deviation reflects the more permissive interaction landscape of the SIS framework, which, while facilitating structural rearrangements relevant to RNA condensation, reduces its ability to accurately balance association and dissociation equilibria in short sequence duplexes.

The bar plot in Fig. 3C summarises the performance of RNA2PS across the full set of 33 RNA sequences by reporting the deviation between simulated and experimental melting temperatures ($T_{m,\text{sim}} - T_{m,\text{exp}}$). Overall, the model reproduces both the magnitude and sequence dependence of RNA duplex stability, with an average standard deviation of ~ 17.5 K across all sequences, indicating robust transferability across different lengths and compositions. While most deviations are centred around zero, small systematic differences are observed for a subset of mismatch-containing sequences, where local destabilization effects are more pronounced. For comparison, the standard nearest-neighbour (NN) model^{84,85}, explicitly parametrised for short-duplex thermodynamics^{87,88}, achieves slightly lower deviations for short, well-characterised sequences (orange bars), but its performance becomes less consistent outside its optimal parameterisation regime (e.g., the sequences with mismatches marked with asterisks in Fig. 3C). Importantly, those cases are better captured by RNA2PS considering that those sequences were not included in the training of the NN model⁸⁴.

D. Sequence-dependent condensation in RNA repeat expansions

A central objective of the RNA2PS model is to capture RNA condensation behaviour while preserving a consistent description of RNA thermodynamics and structural accuracy across both single- and double-stranded regimes. In particular, trinucleotide repeat expansions provide a stringent benchmark, as they are known to self-assemble into condensates and gel-like states through multivalent RNA–RNA interactions^{28,30}. In RNA2PS, this behaviour emerges from the same interaction framework that governs duplex formation. Base-pairing and stacking interactions, previously calibrated to reproduce melting temperatures across diverse sequences (Fig. 3),

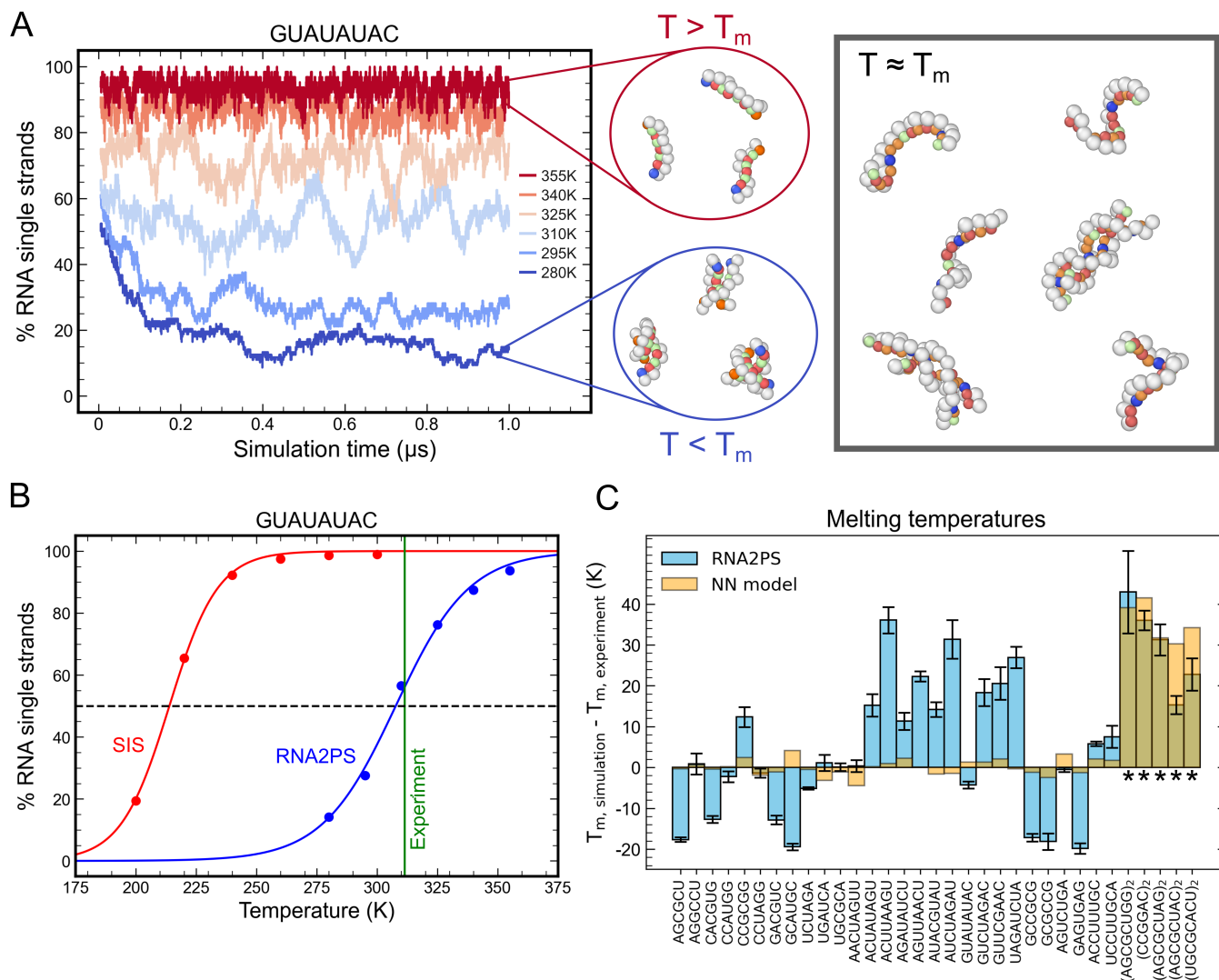


FIG. 3. RNA2PS reproduces sequence effects in melting temperatures (T_m). **A**. Time evolution of the percentage of RNA single strands for the sequence GUAUUAUC at different temperatures in the range 280–355 K. Representative configurations of the RNA sequences below and above the melting temperature T_m are shown as insets. We also provide an illustrative snapshot of our system at $T = T_m$. **B**. Steady-state fraction of RNA single strands as a function of temperature for GUAUUAUC. RNA2PS results are shown in blue, while SIS model predictions are shown in red. The vertical green line indicates the experimental melting temperature, and the horizontal dashed line marks the 50% unpaired threshold used to define T_m . **C**. Deviations between simulated and experimental melting temperatures for 33 RNA sequences perfectly paired⁸⁵ and with mismatches⁸⁴ (marked with asterisks). RNA2PS results are shown in blue, while the nearest-neighbour (NN) model^{84,85} is shown in yellow.

act in concert with sequence-dependent cross-interactions encoded through a Wang–Frenkel potential⁷⁵ (see Methods). The latter depends only on base identity and controls the balance between effective hydrophobicity, as well as attraction vs. repulsion in non-base-pairing repeats. This unified parametrisation enables the RNA2PS model to reproduce experimentally observed phase-separation trends in RNA repeats ($n = 40$)³⁰ without introducing additional fitting parameters for condensation (see Fig. 5C).

To quantify RNA condensation, we perform direct coexistence (DC) simulations of solutions of RNA molecules^{89,90}, which enable the explicit observation of

coexisting condensed and dilute phases. Below the critical temperature, RNA molecules spontaneously form a dense phase stabilized by dynamic intermolecular base pairing, coexisting with a dilute phase of dispersed strands (see Fig. 4A). This behaviour is characterised through density profiles, from which we extract the equilibrium densities of both phases at different temperatures to construct the temperature–density phase diagram (Fig. 4B). The critical temperature T_c is obtained by fitting the coexistence curve to the universal scaling law of density differences and the law of rectilinear diameters⁹¹ (see Section IV). For representative sequences such as CCG₄₀ and CGU₄₀, RNA2PS predicts critical

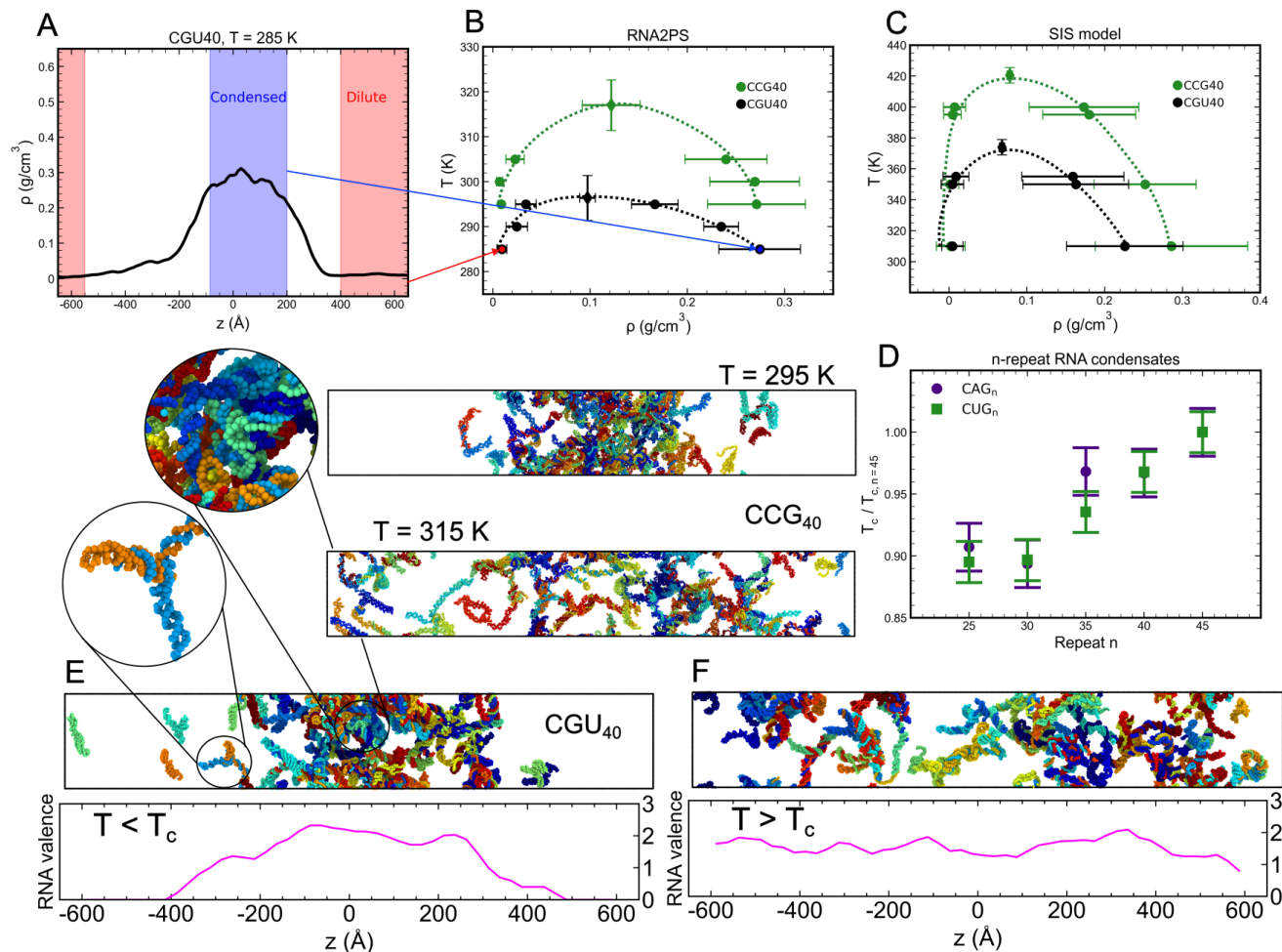


FIG. 4. RNA2PS in describing condensate self-assembly. **A**. Density profile of CGU₄₀ at $T = 285$ K in DC simulation. **B**. Phase diagram in the temperature–density plane for the CCG₄₀ and CGU₄₀ sequences. Critical points are depicted by diamonds. Bottom panels are snapshots of a CCG₄₀ simulation below ($T = 295$ K) and close to ($T = 315$ K) the critical temperature. **C**. Same phase diagrams as in panel B predicted by the SIS model^{63,65}. In both cases, dashed lines are a guide to the eye. **D**. Condensate stability of CAG_{*n*} and CUG_{*n*} repeat sequences increases with repeat length. Normalization constants for the critical solution temperatures of each sequence are $T_{c, CAG45} = (259 \pm 5)$ K and $T_{c, CUG45} = (302 \pm 5)$ K. The full phase diagrams for these sequences can be found in Fig. S1 in the Supplementary Material. **E**. Simulation snapshots of CGU₄₀ below and **F** above the critical solution temperature. Lower panels represent the valency per RNA molecule in that particular frame. Error bars are calculated from panels **B** and **C** fitting to universal scaling law and law of rectilinear diameters (critical point) and half the difference between maximum and minimum density in condensed or dilute phase; and **D** maximum between simulation grid to identify the critical temperature (5 K) and error from fitting to the universal scaling law.

temperatures close to ~ 300 K (Fig. 4B), in good qualitative agreement with experimental trends^{30,92}. In contrast, simulations with the SIS model yield significantly higher critical temperatures (Fig. 4C), indicating an over-stabilized condensed phase. This discrepancy likely arises from the highly cooperative nature of base pairing in the SIS model, where single-bead nucleotides can simultaneously engage in multiple intermolecular interactions across different strands and orientations, effectively artificially enhancing multivalency and leading to excessive stickiness. In the RNA2PS model, this effect is mitigated by the two-bead representation, which geometrically constrains base pairing and disfavors the formation

of multiple simultaneous bonds unless they are coherently aligned along the same strands. As a result, in RNA2PS the balance between inter- and intramolecular interactions is more realistically captured, yielding condensate stabilities within biologically relevant conditions.

A key determinant of RNA condensation is chain length, which directly controls the valency and cooperativity of intermolecular interactions^{28,29}. To probe this effect, we compute the critical temperatures of CAG_{*n*} and CUG_{*n*} condensates for $n = 25, 30, 35, 40,$ and 45 repeats (Fig. 4D). RNA2PS recapitulates this behaviour by reproducing the experimentally observed dependence of condensate stability on n repeat length^{28,30,63}. As shown

in Fig. 4D, the critical temperature is nearly constant for short repeats ($n \lesssim 30$), where intermolecular connectivity is limited, and increases with n as longer chains enable more extensive multivalent interactions⁶³. Remarkably, the normalised critical temperatures of CAG $_n$ and CUG $_n$ collapse onto a common trend, suggesting a general scaling behaviour governed primarily by chain length rather than sequence-specific details. Overall, the normalised critical temperature reflects an approximately linear increase with repeat number, highlighting the progressive stabilisation of condensates with increasing the sequence length.

To further characterise the microscopic organisation of RNA within condensates, we analyse the effective valency of RNA molecules, which provides a direct measure of their intermolecular connectivity and multivalent binding capacity. We define molecular valency as the average number of distinct RNA strands with which a given molecule establishes close contacts, computed along the z -direction of the simulation box to resolve spatial variations across the interface. Fig. 4E,F shows representative configurations of a CGU $_{40}$ condensate slightly below and above the critical temperature, respectively, together with the corresponding valency profiles (bottom panels). Below T_c , RNA molecules in the dense phase exhibit significantly higher valency (e.g. >2) than those at the interface, reflecting the formation of a highly connected network stabilised by intermolecular interactions. This behaviour is consistent with an entropically driven condensation mechanism^{92,93}, in which molecules maximise their number of binding partners in the condensed phase. Moreover, it is consistent with previous patchy particle studies on phase separation showing that valency >2 is required to establish a percolated liquid network in three dimensions capable to stabilize a condensed phase^{94,95}. In contrast, above T_c , the valency becomes nearly uniform across the system (with an average value <2 determined by the supersaturation concentration in the solution), indicating the loss of spatial organization and the emergence of a homogeneous phase. These results demonstrate that RNA2PS not only captures RNA condensation at the mesoscopic level, but also recapitulate the underlying changes in molecular connectivity that drive condensate formation.

In Fig. 5, we benchmark the model against experimental cell foci characterisation of RNA condensation, demonstrating its ability to capture sequence-dependent trends across a broad range of repeat sequences with different composition. Fig. 5A shows the fraction of molecules in the largest cluster (FMLC) for non-base-pair-forming RNA repeats at 293 K. RNA2PS exhibits a strong correlation with the experimentally measured fraction of cells forming foci³⁰, indicating that the model quantitatively captures intrinsic condensation propensities. Notably, this agreement is achieved by tuning only the base-dependent energy scale of the Wang–Frenkel interaction, with all cross-interactions determined via Lorentz–Berthelot mixing rules and negligible contribu-

tions from phosphate beads. In contrast, the SIS model lacks sequence-dependent interactions for non-base-pair-forming sequences, resulting in a nearly flat response that fails to reproduce the experimentally observed variability in phase-separation (Fig. 5B).

We next assess the ability of RNA2PS to predict the stability of condensates formed by base-pair-forming repeats. Fig. 5C shows the correlation between the normalized critical temperature and the experimental propensity to form foci in cells³⁰. Notably, RNA2PS reproduces the overall experimental trends as an emergent property, despite base-pairing interactions being calibrated exclusively against melting temperatures and stacking interactions against non-base-pair-forming sequences. The predicted critical solution temperatures for phase-separation span from (255 ± 5) K to (335 ± 5) K. By comparison, the SIS model yields a broader and systematically shifted distribution (240–440 K) and exhibits a significantly weaker correlation with experimental data (Fig. 5D). These results highlight the ability of RNA2PS to link sequence structural properties, thermodynamics and duplex stabilities, as well as condensate phase behaviour within a unified and transferable modelling framework to be combined with protein coarse-grained models^{55,61,96?}.

III. DISCUSSION

In this work, we present RNA2PS, a coarse-grained RNA model with a minimal two-bead-per-nucleotide resolution that enables a unified description of RNA structure, thermodynamics, and phase behaviour. By separating electrostatic interactions, localized on the phosphate bead (although deactivated in this work to study phase-separation under the presence of small divalent cation salt concentration; e.g., 10 mM MgCl $_2$ ^{28,29}), from directional base-mediated interactions encoded at the base bead, the model introduces a physically grounded decomposition of RNA interactions. This design preserves nucleotide chemical specificity while providing the flexibility required to capture the competition between intra- and intermolecular contacts that underlies RNA folding and condensation. In contrast to previous two-bead approaches⁶², RNA2PS incorporates explicit sequence-dependent base–base interactions and reproduces the formation of double-helical RNAs, enabling molecular-level insight into both isolated RNA structures and collective multi-chain pure RNA condensation.

A key feature of RNA2PS is its ability to reconcile the formation of stable double helices with the conformational flexibility of single disordered strands. In RNA2PS, the geometry of base pairing is encoded through a multi-body interaction, ensuring that double-helical conformations emerge naturally while preventing unphysical overcoordination and simultaneous binding to multiple partners⁶³. At the same time, the spatial separation between backbone and base beads introduces

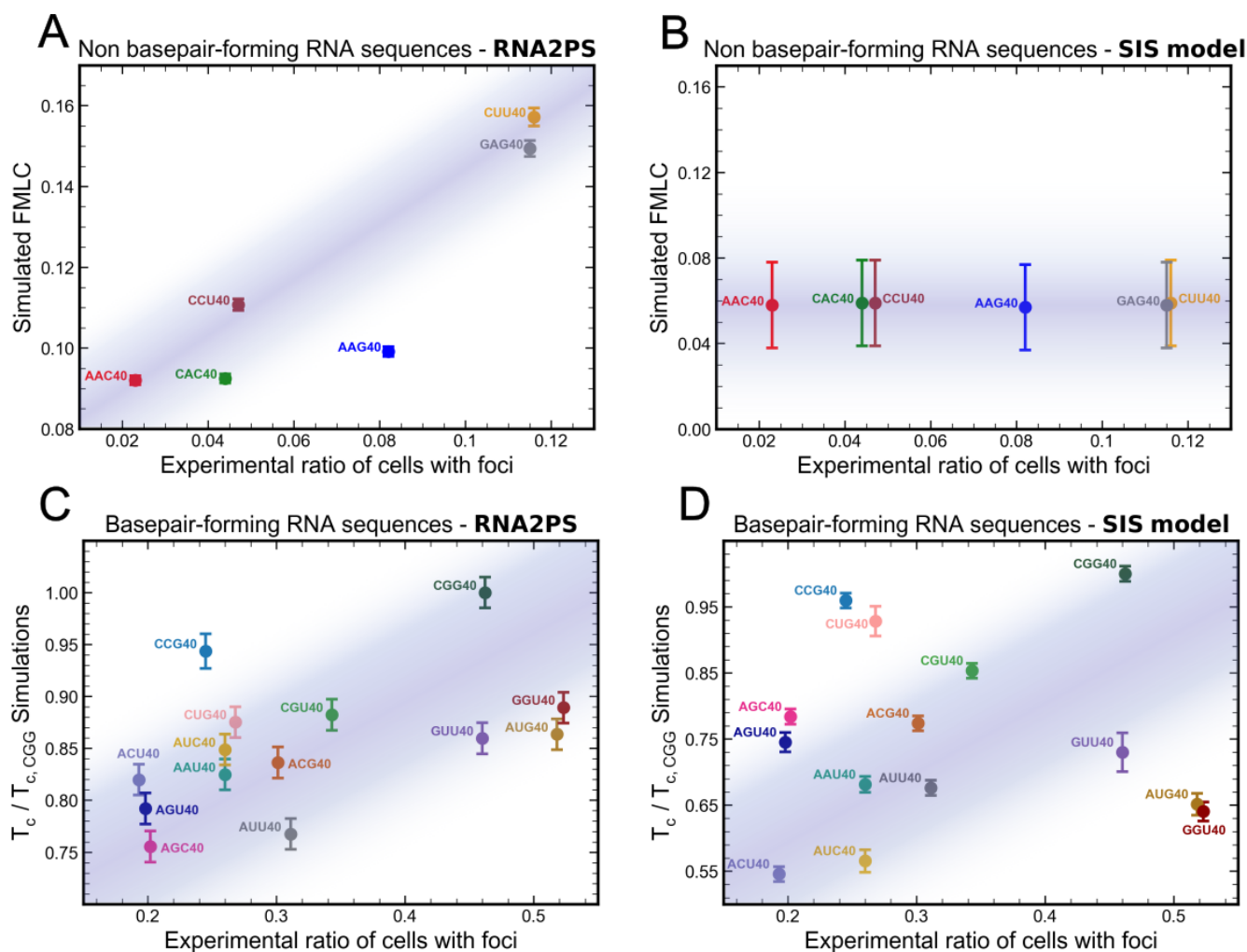


FIG. 5. RNA2PS reproduces experimental trends in sequence-dependent RNA condensation. (A) Fraction of Molecules in Largest Cluster (FMLC) of RNA non basepair-forming repeat sequences (at $T = 293$ K) versus experimental ratio of cells forming foci³⁰. (B) Normalized critical temperatures (T_c) of condensates with basepair-forming RNA repeat sequences correlate with the propensity of cells to express foci. Normalization constant is $T_{c, CGG} = (336 \pm 5)$ K. Bottom panels (C) and (D) correspond to the SIS model^{63,65}. Normalization constant in (D) is $T_{c, CGG} = (438 \pm 5)$ K. Error bars come from (A) and (C) standard error; and (B) and (D) maximum between simulation grid to identify the critical temperature (5 K) and error from fitting to universal scaling law. Phase diagrams for RNA2PS and SIS models can be found in Fig. S2 and Fig. S3 in Supplementary Material, respectively. Shaded regions are a guide to the eye.

intrinsic directionality and excluded-volume constraints, which restrict the relative orientation and proximity of interacting bases. This effectively captures the configurational and entropic cost associated with base pairing and stacking, and limits the number of favourable contacts that each nucleotide can establish. This balance is essential: intramolecular base pairing stabilizes folded conformations, whereas intermolecular interactions promote multivalency and biomolecular condensate formation.

Such interplay is directly reflected in the predicted structural topology of RNA double-strands (Fig. 2) and the thermodynamic behaviour of the RNA2PS model. RNA2PS quantitatively reproduces the melting temperatures of short RNA sequences in dilute conditions (Fig.

3), with a standard deviation of 17.5 K from experimental values, while simultaneously predicting critical temperatures for RNA condensates within biologically relevant ranges (260-335 K; Figs. 4 and 5). Achieving this level of consistency across such distinct properties highlights the model's ability to capture the underlying free-energy landscape governing RNA phase behaviour. In RNA2PS, condensation of RNA solutions in presence of implicit cation divalent ions emerges from a subtle balance between enthalpic gain associated with base pairing and stacking interactions as well as from entropic contributions arising from multivalent intermolecular connectivity. The formation of interchain contacts increases the number of accessible binding configurations at the network level, effectively enhancing the degeneracy of

the condensed state. The two-bead architecture plays a central role in regulating this balance: by geometrically constraining base pairing and limiting the number of simultaneous interactions per nucleotide, it prevents unphysical overcoordination and enforces a finite valency. As a result, the model avoids the overstabilization of the condensed phase predicted by other models⁶³ and preserves a more realistic competition between intra- and intermolecular interactions. This is particularly relevant for repeat-expansion RNAs, where small variations in sequence composition and length lead to measurable differences in condensation propensity³⁰. Importantly, this level of consistency is achieved within a minimal representation, avoiding the need for higher-resolution models⁷⁴, which, although accurate, are often limited by computational cost to large scale systems.

RNA2PS further captures sequence-dependent trends in condensation propensity for both base-pair-forming and non-base-pair-forming RNA repeat sequences (Fig. 5). For base-pair-forming sequences, the predicted critical temperatures follow the expected stability trends^{65,92}, reflecting the interplay between interaction strength, sequence-dependent base pairing energetics, and the number of accessible binding configurations. This framework explains the enhanced stability of sequences such as CCG₄₀ and CGG₄₀, which maximize both base-pairing propensity and multivalency. Moreover, the model highlights the role of non-canonical and dispersive interactions, particularly in sequences containing G–U motifs²⁹, in modulating condensate stability beyond canonical Watson–Crick pairing. Deviations observed for specific sequences in cell foci may indicate additional contributions not captured within an RNA-only description, such as protein-mediated interactions in highly multi-component cellular environments³⁰.

The level of coarse-graining and dynamical properties of RNA2PS are well matched to those of state-of-the-art sequence-dependent coarse-grained protein models commonly used to study biomolecular condensates, such as the Mpipi, CALVADOS, and HPS families^{55,60,61,96,97}. This compatibility will enable the direct simulation of RNA–protein mixtures with explicit RNA base-pairing. In this context, the RNA2PS model, once its cross-interactions with protein force fields are optimized—a task beyond the scope of this work—can provide a robust framework to explore RNA-driven regulation of phase separation^{11,22}, including processes such as stress granule self-assembly⁹⁸ or RNA-mediated modulation of protein condensation²⁰ among others^{25,99}. RNA–protein assemblies are recognized as central to intracellular organization and regulation, and their dysregulation is increasingly linked to multiple diseases^{29,100}, including certain cancers, where aberrant condensate formation can alter transcriptional programs and cellular identity^{101,102}. In this context, tumor-associated condensates provide a particularly relevant example in which changes in RNA– or DNA–protein interactions may directly impact phase behaviour and function. A physically consistent multi-

component model can be therefore essential to dissect the molecular principles governing condensate formation and its dysregulation in pathological states.

IV. MATERIALS AND METHODS.

A. Model force field

We present the RNA2PS, a coarse-grained computational model with two-beads-per-nucleotide resolution to simulate RNA self-condensation. One bead is located at the geometric centre of mass of the phosphate+ribose groups (henceforth phosphate bead, P), and the other at 4 Å in the direction of the vector connecting P to the centre of the nitrogenous base (henceforth base bead, B). The total potential energy of the system is defined as the sum of bonded and non-bonded contributions.

Bead connectivity is enforced through harmonic bond potentials of the form

$$U_{bonds} = \frac{1}{2} \sum_{bonds} k_{bond}(r - r_0)^2, \quad (1)$$

being r the bond distance and r_0 its equilibrium value. Bonds are established between consecutive phosphate and base beads, as well as between bases and phosphates linked by a phosphodiester bond. The rigidity is modelled by means of an angular potential with the form

$$U_{angles} = \frac{1}{2} \sum_{angles} k_{angle}(\theta - \theta_0)^2, \quad (2)$$

where θ is the angle formed by any three consecutive base beads and θ_0 its equilibrium value. Table I contains all intramolecular interaction parameters used in RNA2PS. Base-pairing interactions are described by means of a

Parameter	Value
$r_0(\text{P-P}) / \text{Å}$	5.2
$r_0(\text{B-B}) / \text{Å}$	3.8
$r_0(\text{P-B}) / \text{Å}$	4.0
$k_{bond}(\text{P-P}) / \text{kcal mol}^{-1} \text{Å}^2$	8.8
$k_{bond}(\text{B-B}) / \text{kcal mol}^{-1} \text{Å}^2$	1.5
$k_{bond}(\text{P-B}) / \text{kcal mol}^{-1} \text{Å}^2$	6.95
$\theta_0 / \text{degrees}$	150
$k_{angle} / \text{kcal mol}^{-1}$	6.05

TABLE I. Values of the intramolecular bonded interaction parameters. P and B denotes backbone and nucleotide beads, respectively.

multibody potential⁶³ with the form

$$U_{bp} = U_{bp}^0 \exp \left(U_{bp}^{bond} + U_{bp}^{angle} + U_{bp}^{dihedral} \right), \quad (3)$$

where U_{bp}^0 sets the overall strength of the base-pairing interaction and depends on sequence context, and

$$U_{bp}^{bond} = \frac{1}{2} k_{bp}^{bond} (r_{ij} - r_{bp}^0)$$

$$U_{bp}^{angle} = \frac{1}{2} \sum_{angles} k_{bp}^{angle} (\theta_{ijk} - \theta_{bp}^0)$$

$$U_{bp}^{dihedral} = \sum_{dihedrals} k_{bp}^{dihedral} [1 + \cos(\phi_{ijkl} + \phi_{bp}^0)] ,$$

where U_{bp}^{bond} depends on the distance between the beads forming the base pairing (i and j); U_{bp}^{angle} includes contributions from four angles centred at i and j , each involving the opposing base bead and one of its neighbouring phosphate beads; and $U_{bp}^{dihedral}$ accounts for the two dihedrals angles defined by i , j , and the adjacent base beads along the strand. In addition, K_{bp}^{bond} depends on the specific base identity. Tables II and III summarise all model parameters in Eq. (3). The potential is only applied to bead pairs i and j that (i) have compatible biochemical identities, (ii) are separated by least six nucleotides along the same strand, (iii) lie within a cut-off distance of 8 Å, and (iv) are not located at the termini of the molecule.

	Pair Type		
Parameters	AU	CG	GU
$r_{bp}^0 / \text{Å}$		6.9	
$k_{bp}^{bond} / \text{kcal mol}^{-1} \text{Å}^2$	1.9	1.15	1.80

TABLE II. Values of the intermolecular bonds corresponding to the base pairing parameters.

Cross-stacking interactions are represented by a Wang-Frenkel potential⁷⁵ of the form

$$U(r) = \epsilon \alpha \left(\left[\frac{\sigma}{r} \right]^{2\mu} - 1 \right) \left(\left[\frac{r_c}{r} \right]^{2\mu} - 1 \right)^{2\nu} , \quad (4)$$

where

$$\alpha = 2\nu \left(\frac{r_c}{\sigma} \right)^{2\mu} \left[\frac{1 + 2\nu}{2\nu[(r_c/\sigma)^{2\mu} - 1]} \right]^{2\nu+1} ,$$

and

Parameter	Value
$\theta_{bp}^1 / \text{degrees}$	105.0
$\theta_{bp}^2 / \text{degrees}$	54.0
$k_{bp}^{angle} / \text{kcal mol}^{-1}$	1.5
$\phi_{bp}^1 / \text{degrees}$	105.0
$\phi_{bp}^2 / \text{degrees}$	65.0
$k_{bp}^{dihedral} / \text{kcal mol}^{-1}$	0.5

TABLE III. Values of the intermolecular angles and dihedrals corresponding to the base pairing parameters.

$$r_{min} = r_c \left[\frac{1 + 2\nu}{1 + 2\nu(r_c/\sigma)^{2\nu}} \right]^{1/2\nu} ,$$

where r_c is the cut-off radius, $\nu = 1$, $\mu = 3$ and the energy (ϵ) and distance (σ) parameters are given in Table IV. Cross interactions follow Lorentz-Berthelot mixing rules.

	Bead Type				
Parameters	P	A	U	C	G
$\epsilon / \text{kcal mol}^{-1}$	6.954	6.430	5.958	6.531	6.033
$\sigma / \text{Å}$	0.0100	0.1998	0.5827	0.2498	0.5827

TABLE IV. Wang-Frenkel parameters. R_c is selected for $3\sigma_i$

Finally, electrostatic interactions between phosphate beads are modelled using a Yukawa potential. In this work, all properties are evaluated in the high-salt regime, where the Debye screening length is effectively negligible and electrostatic contributions are strongly screened. Nevertheless, including this term is essential to capture the interplay between RNA and proteins in biomolecular condensates.

B. Simulation of RNA crystal structures

We carried out molecular dynamics simulations to determine the optimal equilibrium parameters for bond, angle and dihedral potentials describing both base-pairing interactions and intramolecular RNA structure. Starting from experimental PDB structures, atoms were mapped onto the RNA2PS representation by placing the backbone bead at the centre of mass of the ribose+phosphate and the base bead along the vector connecting the backbone bead to the centre of mass of the nitrogenous base, at a distance of 4Å. For each sequence, independent simulations were performed at $T=250K$ to mimic the stability of the crystal structure at low temperature. Each simulation run for 1 μs and included 64 replicas of the structure under dilute conditions (~ 10.5 mM) to prevent intermolecular interactions. Bond lengths, angles and dihedrals were extracted from both the mapped PDB structures and the simulation trajectories, and their distributions were compared. This procedure was iteratively refined using a basin-hopping algorithm to minimise the error. In total, a dataset of 25 RNA PDB structures was used.

Molecular dynamics simulations were performed to estimate the melting temperature (T_m) or RNA duplexes. For each sequence, independent simulations were carried out at temperatures ranging from 280 K to 355 K, using 15 K intervals. Each simulation was run for 1 μs , and configurations were saved every 0.1 ns for subsequent analysis. The systems contained 128 RNA strands (64 duplexes) placed in a cubic box of dimensions $200 \times 200 \times 200$ Å³, corresponding to RNA strand concentration of 26.6

mM. From the resulting trajectories, the number of dissociated strands was computed as a function of time and temperature. The fraction of dissociated strands was used to construct thermal denaturation curves, and the melting temperatures were defined as the temperature at which 50% of the strands are dissociated. This value was obtained by fitting the data to a sigmoidal function of the form $f(T) = 1/(1 + e^{(T-T_m)})$, where T_m is the melting temperature. This procedure was applied to a total of 33 RNA sequences, including 5 sequences containing mismatches.

C. Simulation of RNA condensates

We simulate both base-pairing and non-base-pairing condensates using the direct coexistence method^{89,95}. Condensates composed of 100 molecules were prepared in an orthorhombic box with dimensions $210.50 \times 210.50 \times 1376.30 \text{ \AA}^3$, corresponding to an overall density of approximately 0.07 g/cm^3 . The initial configuration was constructed at higher density of 0.40 g/cm^3 to promote condensation. For non-base-pairing systems, simulations were carried out at 293 K over $2 \mu\text{s}$, and the fraction of molecules in the largest cluster was averaged over the final $1.6 \mu\text{s}$. For base-pairing systems, configurations were first heated to 350 K for 1 ns to disrupt base pairing and avoid kinetic trapping. Simulations at the target temperature were then performed for $3 \mu\text{s}$, with only the last $1.2 \mu\text{s}$ used for analysis. In all cases, the integration time step was set to 10 fs , and configuration were saved every 2 ns .

D. Physical observables

We characterise the propensity of non-base-pairing condensates by measuring the fraction of molecules in the largest cluster at $T = 293 \text{ K}$. Clusters are defined as sets of molecules in which each molecule lies within 8 \AA of at least one other molecule in the cluster. The distance between two molecules is defined as the minimum distance between any pair of beads belonging to each molecule.

The stability of base-pairing condensates is assessed by analysing their critical temperatures. We determine the critical point (ρ_c, T_c) of each system using the universal scaling relations near the critical point⁹¹

$$\rho_C - \rho_D = \Delta\rho_0 \left(1 - \frac{T}{T_c}\right)^\beta, \quad (5)$$

where ρ_C and ρ_D are the densities of the condensed and dilute phase, respectively, $\beta \approx 0.325$ is the critical exponent, and $\Delta\rho_0$ is a fitting parameter. In addition, the law of rectilinear diameters⁹¹ is used,

$$\frac{\rho_C + \rho_D}{2} = \rho_c + A(T_c - T), \quad (6)$$

being A is a fitting parameter. The densities ρ_C and ρ_D are obtained by averaging appropriate regions of the density profiles at each temperature. All fits are performed using three data points below the critical temperature, with temperature spacing of $5\text{-}10 \text{ K}$ between consecutive points.

V. DATA AVAILABILITY

LAMMPS implementation files have been made available on Github <https://github.com/pllombard/RNA2PS>.

VI. ACKNOWLEDGEMENTS

A. R. T. acknowledges funding from the MICIU under the Juan de la Cierva fellowship (JDC2024-053759-I). P.L. and J.L.-M. acknowledges funding from the European Union's Horizon 2020 research and innovation program (grant agreement 101160499 to J. R. E). J.R. acknowledges funding from the Spanish ministry of Science and Innovation (project number PID2022-136919NB-C32). R.C.-G. acknowledges funding the UK Research Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC) [EP/Z002028/1], following funding from the European Research Council (ERC) Consolidator Grant "ChromatinDroplets" under the European Union's Horizon Europe research and innovation programme. J. R. E. acknowledges funding from Emmanuel College, the University of Cambridge, the Ramon y Cajal fellowship (RYC2021-030937-I), the Spanish scientific plan and committee for research reference PID2022-136919NA-C33, and the ERC under the European Union's Horizon Europe research and innovation program (grant agreement no. 101160499).

This work has been performed using resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (<http://www.hpc.cam.ac.uk>) funded by EPSRC Tier-2 capital grant EP/P020259/1-CS170. This work has also been performed using resources provided by Archer2 (<https://www.archer2.ac.uk/>) funded by EPSRC Tier-2 capital grant EP/P020259/e829. The authors also thankfully acknowledge RES computational resources provided by Mare Nostrum 5 through the activity FI-2025-3-0065.

VII. REFERENCES

- ¹A. A. Hyman and K. Simons, *Science* **337**, 1047 (2012).
- ²A. A. Hyman, C. A. Weber, and F. Jülicher, *Annual review of cell and developmental biology* **30**, 39 (2014).
- ³S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen, *Cell* **166**, 651 (2016).
- ⁴F. Frottin, F. Schueder, S. Tiwary, R. Gupta, R. Körner, T. Schlichthaerle, J. Cox, R. Jungmann, F. Hartl, and M. Hipp, *Science* **365**, 342 (2019).

- ⁵C. P. Brangwynne, T. J. Mitchison, and A. A. Hyman, *Proceedings of the National Academy of Sciences* **108**, 4334 (2011).
- ⁶J. D. Wurtz and C. F. Lee, *New Journal of Physics* **20**, 045008 (2018).
- ⁷A. Marcelo, R. Koppenol, L. P. de Almeida, C. A. Matos, and C. Nóbrega, *Cell death & disease* **12**, 592 (2021).
- ⁸D. Campos-Melo, Z. C. Hawley, C. A. Droppelmann, and M. J. Strong, *Frontiers in Cell and Developmental Biology* **9**, 621779 (2021).
- ⁹C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoegge, J. Gharakhani, F. Jülicher, and A. A. Hyman, *Science* **324**, 1729 (2009).
- ¹⁰Y. Luo, Z. Na, and S. A. Slavoff, *Biochemistry* **57**, 2424 (2018).
- ¹¹S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Poznaniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillen-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y.-T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman, and S. Alberti, *Science* **360**, 918 (2018).
- ¹²J. R. Mann and C. J. Donnelly, *Neuron* (2021).
- ¹³C. Roden and A. S. Gladfelter, *Nature Reviews Molecular Cell Biology* **22**, 183 (2021).
- ¹⁴B. Ozguney, P. Mohanty, and J. Mittal, *Biophysical Journal* **123**, 3844 (2024).
- ¹⁵S. Tian, H. Kim, H. A. Curnutte, and T. Trcek, *bioRxiv*, 2025 (2025).
- ¹⁶S. Tian, H. Nguyen, Z. Ye, S. Rouskin, D. Thirumalai, and T. Trcek, *bioRxiv*, 2024 (2024).
- ¹⁷A. H. Fox, S. Nakagawa, T. Hirose, and C. S. Bond, *Trends in Biochemical Sciences* **43**, 124 (2018).
- ¹⁸J. C. Schwartz, X. Wang, E. R. Podell, and T. R. Cech, *Cell reports* **5**, 918 (2013).
- ¹⁹B. Yu, B. M. Pettitt, and J. Iwahara, *Accounts of chemical research* **53**, 1802 (2020).
- ²⁰K. Makasewicz, C. Morelli, L. Faltova, and P. Arosio, *bioRxiv*, 2025 (2025).
- ²¹R. S. Fisher and S. Elbaum-Garfinkle, *Nature Communications* **11**, 1 (2020).
- ²²P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, and A. A. Deniz, *Angewandte Chemie* **129**, 11512 (2017).
- ²³A. Vidal Ceballos, A. Geissmann, D. C. Favaro, P. Deshpande, and S. Elbaum-Garfinkle, *Scientific Reports* **15**, 9295 (2025).
- ²⁴H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, and A. S. Gladfelter, *Molecular Cell* **60**, 220 (2015).
- ²⁵M. Linsenmeier, M. Hondele, F. Grigolato, E. Secchi, K. Weis, and P. Arosio, *Nature Communications* **13**, 1 (2022).
- ²⁶A. R. Tejedor, I. Sanchez-Burgos, M. Estevez-Espinosa, A. Garaizar, R. Collepardo-Guevara, J. Ramirez, and J. R. Espinosa, *Nature Communications* **13**, 1 (2022).
- ²⁷B. Van Treeck and R. Parker, *Cell* **174**, 791 (2018).
- ²⁸A. Jain and R. D. Vale, *Nature* **546**, 243 (2017).
- ²⁹P. C. Bevilacqua, A. M. Williams, H.-L. Chou, and S. M. Assmann, *Rna* **28**, 16 (2022).
- ³⁰A. U. Isiktas, A. Eshov, S. Yang, and J. U. Guo, *Cell Reports Methods* **2** (2022).
- ³¹S. Tian, H. A. Curnutte, and T. Trcek, *Molecules* **25**, 3130 (2020).
- ³²J. R. Gatchel and H. Y. Zoghbi, *Nature Reviews Genetics* **6**, 743 (2005).
- ³³G. M. Wadsworth, W. J. Zahurancik, X. Zeng, P. Pullara, L. B. Lai, V. Sidharthan, R. V. Pappu, V. Gopalan, and P. R. Banerjee, *Nature Chemistry* **15**, 1693 (2023).
- ³⁴K. A. Wilson, J. L. Kellie, and S. D. Wetmore, *Nucleic acids research* **42**, 6726 (2014).
- ³⁵A. C. Cheng, W. W. Chen, C. N. Fuhrmann, and A. D. Frankel, *Journal of molecular biology* **327**, 781 (2003).
- ³⁶E. V. Pletneva, A. T. Laederach, D. B. Fulton, and N. M. Kostić, *Journal of the American Chemical Society* **123**, 6232 (2001).
- ³⁷J. K. Tom, P. L. Onuchic, and A. A. Deniz, *The Journal of Physical Chemistry B* **126**, 9715 (2022).
- ³⁸Y. Wu and R. M. Brosh Jr, *The FEBS journal* **277**, 3470 (2010).
- ³⁹A. E. Renton, E. Majounie, A. Waite, J. Simón-Sánchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. Van Swieten, L. Myllykangas, *et al.*, *Neuron* **72**, 257 (2011).
- ⁴⁰N. Sekiyama, K. Takaba, S. Maki-Yonekura, K.-i. Akagi, Y. Ohtani, K. Imamura, T. Terakawa, K. Yamashita, D. Inaoka, K. Yonekura, *et al.*, *Proceedings of the National Academy of Sciences* **119**, e2122523119 (2022).
- ⁴¹S. Torrino, W. M. Oldham, A. R. Tejedor, I. S. Burgos, L. Nasr, N. Rachedi, K. Fraissard, C. Chauvet, C. Sbai, B. P. O'Hara, S. Abélanet, F. Brau, C. Favard, S. Clavel, R. Collepardo-Guevara, J. R. Espinosa, I. Ben-Sahra, and T. Bertero, *Cell* **188**, 447 (2025).
- ⁴²G. L. Dignon, R. B. Best, and J. Mittal, *Annual Review of Physical Chemistry* **71**, 53 (2020).
- ⁴³N. H. Espejo, A. Feito, I. Sanchez-Burgos, A. Garaizar, M. M. Conde, A. Rey, A. Castro, R. Collepardo-Guevara, A. R. Tejedor, and J. R. Espinosa, *bioRxiv*, 2025 (2025).
- ⁴⁴E. Pedraza, D. Hoyos, A. Feito, F. Gámez, I. Sanchez-Burgos, R. Collepardo-Guevara, A. R. Tejedor, and J. R. Espinosa, *bioRxiv* (2025), 10.1101/2025.03.26.645197.
- ⁴⁵A. Feito, E. Pedraza, E. Cuesta, A. Castro, A. Rey, I. Sanchez-Burgos, R. Collepardo-Guevara, A. R. Tejedor, and J. R. Espinosa, *ACS Physical Chemistry Au* (2025).
- ⁴⁶I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, *et al.*, *Nature methods* **13**, 55 (2016).
- ⁴⁷K. Inoue, S. Takada, and T. Terakawa, *Biophysics and physiology* **19**, e190015 (2022).
- ⁴⁸M. Boccacini, Y. Berezovska, G. Bussi, M. Paloni, and A. Barducci, *Proceedings of the National Academy of Sciences* **122**, e2425261122 (2025).
- ⁴⁹S. Bottaro, P. Banáš, J. Sponer, and G. Bussi, *The journal of physical chemistry letters* **7**, 4032 (2016).
- ⁵⁰G. H. Zerze, P. M. Piaggi, and P. G. Debenedetti, *The Journal of Physical Chemistry B* **125**, 13685 (2021).
- ⁵¹A. Plumridge, K. Andresen, and L. Pollack, *Journal of the American Chemical Society* **142**, 109 (2019).
- ⁵²G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, *PLoS Computational Biology* **14**, e1005941 (2018).
- ⁵³A. R. Tejedor, A. Garaizar, J. Ramírez, and J. R. Espinosa, *Biophysical Journal* **120**, 5169 (2021).
- ⁵⁴G. Tesei, T. K. Schulze, R. Crehuet, and K. Lindorff-Larsen, *Proceedings of the National Academy of Sciences* **118**, e2111696118 (2021).
- ⁵⁵G. Tesei and K. Lindorff-Larsen, *Open Research Europe* **2**, 94 (2023).
- ⁵⁶G. H. Zerze, *Journal of Chemical Theory and Computation* **20** (2023).
- ⁵⁷S. Das, Y.-H. Lin, R. M. Vernon, J. D. Forman-Kay, and H. S. Chan, *Proceedings of the National Academy of Sciences* **117**, 28795 (2020).
- ⁵⁸S. Blazquez, I. Sanchez-Burgos, J. Ramirez, T. Higginbotham, M. M. Conde, R. Collepardo-Guevara, A. R. Tejedor, and J. R. Espinosa, *Advanced Science* **10**, 2207742 (2023).
- ⁵⁹R. M. Regy, G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, *Nucleic Acids Research* **48**, 12593 (2020).
- ⁶⁰A. R. Tejedor, A. Aguirre Gonzalez, M. J. Maristany, P. Y. Chew, K. Russell, J. Ramirez, J. R. Espinosa, and R. Collepardo-Guevara, *ACS Central Science* **11**, 302 (2025).
- ⁶¹J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, *Nature Computational Science* **1**, 732 (2021).
- ⁶²I. Yasuda, S. von Bülow, G. Tesei, E. Yamamoto, K. Yasuoka, and K. Lindorff-Larsen, *Journal of chemical theory and computation* **21**, 2766 (2025).
- ⁶³H. T. Nguyen, N. Hori, and D. Thirumalai, *Nature chemistry* **14**, 775 (2022).
- ⁶⁴H. Maity, H. T. Nguyen, N. Hori, and D. Thirumalai, *The*

- Journal of Physical Chemistry Letters **15**, 3820 (2024).
- ⁶⁵D. Aierken and J. A. Joseph, *Journal of Chemical Theory and Computation* **20**, 10209 (2024).
- ⁶⁶S. Li and J. Chen, *Proceedings of the National Academy of Sciences* **122**, e2504583122 (2025).
- ⁶⁷A. Rizuan, S. Rekhi, Y. C. Kim, S. Najafi, J.-E. Shea, and J. Mittal, *Annual Review of Physical Chemistry* **77** (2026).
- ⁶⁸D. Aierken, S. Aland, S. Bo, S. Boeynaems, D. Cai, S. Carra, L. B. Case, H. S. Chan, J. R. Espinosa, T. K. GrandPre, *et al.*, arXiv preprint arXiv:2601.03677 (2026).
- ⁶⁹J. M. Lotthammer, G. M. Ginell, D. Griffith, R. Emenecker, and A. S. Holehouse, *Biophysical Journal* **123**, 43a (2024).
- ⁷⁰S. Assenza and R. Pérez, *Journal of chemical theory and computation* **18**, 3239 (2022).
- ⁷¹D. Farré-Gil, J. P. Arcon, C. A. Laughton, and M. Orozco, *Nucleic Acids Research* **52**, 6791 (2024).
- ⁷²G. da Rosa, L. Grille, V. Calzada, K. Ahmad, J. P. Arcon, F. Battistini, G. Bayarri, T. Bishop, P. Carloni, T. Cheatham Iii, *et al.*, *Biophysical Reviews* **13**, 995 (2021).
- ⁷³J. Das, S. Mukherjee, A. Mitra, and D. Bhattacharyya, *Journal of Biomolecular Structure and Dynamics* **24**, 149 (2006).
- ⁷⁴P. Šulc, F. Romano, T. E. Ouldridge, J. P. Doye, and A. A. Louis, *The Journal of chemical physics* **140** (2014).
- ⁷⁵X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar, and D. Frenkel, *Physical Chemistry Chemical Physics* **22**, 10624 (2020).
- ⁷⁶G. M. Wadsworth, S. Srinivasan, L. B. Lai, M. Datta, V. Gopalan, and P. R. Banerjee, *Molecular cell* **84**, 3692 (2024).
- ⁷⁷J. Bohlin, M. Matthies, E. Poppleton, J. Procyk, A. Mallya, H. Yan, and P. Šulc, *Nature protocols* **17**, 1762 (2022).
- ⁷⁸A. M. Watkins, R. Rangan, and R. Das, *Structure* **28**, 963 (2020).
- ⁷⁹W. Wang, C. Feng, R. Han, Z. Wang, L. Ye, Z. Du, H. Wei, F. Zhang, Z. Peng, and J. Yang, *Nature communications* **14**, 7266 (2023).
- ⁸⁰M. Baek, R. McHugh, I. Anishchenko, H. Jiang, D. Baker, and F. DiMaio, *Nature methods* **21**, 117 (2024).
- ⁸¹S. Majumder, S. Coupe, N. Fakhri, and A. Jain, *Nature Communications* **16**, 4258 (2025).
- ⁸²W. Saenger, in *Principles of nucleic acid structure* (Springer, 1984) pp. 116–158.
- ⁸³M. Petersheim and D. H. Turner, *Biochemistry* **22**, 256 (1983).
- ⁸⁴A. S. O’Toole, S. Miller, N. Haines, M. C. Zink, and M. J. Serra, *Nucleic acids research* **34**, 3338 (2006).
- ⁸⁵T. Xia, J. SantaLucia Jr, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, *Biochemistry* **37**, 14719 (1998).
- ⁸⁶S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner, *Proceedings of the National Academy of Sciences* **83**, 9373 (1986).
- ⁸⁷D. M. Gray, *Biopolymers: Original Research on Biomolecules* **42**, 783 (1997).
- ⁸⁸D. M. Gray, *Biopolymers: Original Research on Biomolecules* **42**, 795 (1997).
- ⁸⁹A. Ladd and L. Woodcock, *Chemical Physics Letters* **51**, 155 (1977).
- ⁹⁰K. S. Sillmore, M. P. Howard, and A. Z. Panagiotopoulos, *Molecular Physics* **115**, 320 (2017).
- ⁹¹J. S. Rowlinson and B. Widom, *Molecular theory of capillarity* (Courier Corporation, 2013).
- ⁹²O. Kimchi, E. M. King, and M. P. Brenner, *Nature Communications* **14**, 332 (2023).
- ⁹³S. Mukherjee and L. V. Schäfer, *Nature Communications* **14**, 5892 (2023).
- ⁹⁴J. R. Espinosa, A. Garaizar, C. Vega, D. Frenkel, and R. Collepardo-Guevara, *The Journal of Chemical Physics* **150**, 224510 (2019).
- ⁹⁵J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, *Proceedings of the National Academy of Sciences* **117**, 13238 (2020).
- ⁹⁶R. M. Regy, J. Thompson, Y. C. Kim, and J. Mittal, *Protein Science* **30**, 1371 (2021).
- ⁹⁷G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, *Proceedings of the National Academy of Sciences* **115**, 9929 (2018).
- ⁹⁸P. Yang, C. Mathieu, R.-M. Kolaitis, P. Zhang, J. Messing, U. Yurtsever, Z. Yang, J. Wu, Y. Li, Q. Pan, *et al.*, *Cell* **181**, 325 (2020).
- ⁹⁹C. Morelli, L. Faltova, U. Capasso Palmiero, K. Makasewicz, M. Papp, R. P. Jacquat, D. Pinotsi, and P. Arosio, *Nature chemistry* **16**, 1052 (2024).
- ¹⁰⁰B. Portz, B. L. Lee, and J. Shorter, *Trends in Biochemical Sciences* **46**, 550 (2021).
- ¹⁰¹A. Boija, I. A. Klein, and R. A. Young, *Cancer Cell* **39**, 174 (2021).
- ¹⁰²I. A. Klein, A. Boija, L. K. Afeyan, S. W. Hawken, M. Fan, A. Dall’Agnese, O. Oksuz, J. E. Henninger, K. Shrinivas, B. R. Sabari, I. Sagi, V. E. Clark, J. M. Platt, M. Kar, P. M. McCall, A. V. Zamudio, J. C. Manteiga, E. L. Coffey, C. H. Li, N. M. Hannett, Y. E. Guo, T.-M. Decker, T. I. Lee, T. Zhang, J.-K. Weng, D. J. Taatjes, A. Chakraborty, P. A. Sharp, Y. T. Chang, A. A. Hyman, N. S. Gray, and R. A. Young, *Science* **368**, 1386 (2020).