

Integrative solution structure of PTBP1-IRES complex reveals strong compaction and ordering with residual conformational flexibility

Received: 16 September 2022

Accepted: 27 September 2023

Published online: 13 October 2023

 Check for updates

Georg Dorn^{1,6}, Christoph Gmeiner^{2,6}, Tebbe de Vries^{1,6}, Emil Dedic¹, Mihajlo Novakovic¹, Fred F. Damberger¹, Christophe Maris¹, Esteban Finol¹, Chris P. Sarnowski³, Joachim Kohlbrecher⁴, Timothy J. Welsh², Sreenath Bolisetty⁵, Raffaele Mezzenga⁵, Ruedi Aebersold³, Alexander Leitner³, Maxim Yulikov²✉, Gunnar Jeschke²✉ & Frédéric H.-T. Allain¹✉

RNA-binding proteins (RBPs) are crucial regulators of gene expression, often composed of defined domains interspersed with flexible, intrinsically disordered regions. Determining the structure of ribonucleoprotein (RNP) complexes involving such RBPs necessitates integrative structural modeling due to their lack of a single stable state. In this study, we integrate magnetic resonance, mass spectrometry, and small-angle scattering data to determine the solution structure of the polypyrimidine-tract binding protein 1 (PTBP1/hnRNP I) bound to an RNA fragment from the internal ribosome entry site (IRES) of the encephalomyocarditis virus (EMCV). This binding, essential for enhancing the translation of viral RNA, leads to a complex structure that demonstrates RNA and protein compaction, while maintaining pronounced conformational flexibility. Acting as an RNA chaperone, PTBP1 orchestrates the IRES RNA into a few distinct conformations, exposing the RNA stems outward. This conformational diversity is likely common among RNP structures and functionally important. Our approach enables atomic-level characterization of heterogeneous RNP structures.

Gene expression is critically regulated by protein-RNA interactions. RNA-binding proteins (RBPs) usually contain multiple RNA-binding domains (RBDs), among which the RNA recognition motif (RRM) is the most abundant domain type¹. Typically, RBDs are flanked and linked by intrinsically disordered regions (IDR) of various lengths and

sequences². Recently, these flanking IDRs have been studied extensively in the context of liquid-liquid phase separation^{3–5}. For example, the low complexity domain of the ALS-associated protein FUS has not only been connected to functional phase-separated states but is also associated with plaque formation that causes neurotoxicity^{6,7}. The

¹Institute of Biochemistry, Department of Biology, ETH Zürich, Zürich, Switzerland. ²Laboratory of Physical Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, Zürich, Switzerland. ³Institute of Molecular Systems Biology, Department of Biology, ETH Zürich, Zürich, Switzerland. ⁴Laboratory for Neutron Scattering and Imaging, Paul Scherrer Institut, Villigen, Switzerland. ⁵Laboratory of Food & Soft Materials, Institute of Food, Nutrition and Health, Department for Health Sciences and Technology, ETH Zürich, Zürich, Switzerland. ⁶These authors contributed equally: Georg Dorn, Christoph Gmeiner, Tebbe de Vries. ✉e-mail: maxim.yulikov@phys.chem.ethz.ch; gunnar.jeschke@phys.chem.ethz.ch; allain@bc.biol.ethz.ch

IDRs that link RBDs potentially have diverse biological functions and they unavoidably cause flexibility and heterogeneity of the overall protein conformation. However, cooperative binding of the RBDs to RNA-binding sites may lead to an increased conformational order of the protein. This raises questions about the extent to which a disorder-to-order transition is achieved, and whether both flexibility and rigidity are important for the biological function of the protein-RNA complex.

A plethora of structures of isolated RBDs are available from nuclear magnetic resonance (NMR) and X-ray crystallography (see examples in refs. 8–14). They generally represent well-defined, three-dimensional folds following the Anfinsen dogma¹⁵. However, the presence of structurally undefined flexible regions connecting these subunits and the associated conformational heterogeneity hinders the structure determination of full-length RBP and their complexes by using exclusively these classical techniques, as well as cryo-EM^{16–18}. Thus, current access to ribonucleoprotein (RNP) structures is limited to protein-RNA machineries with stable intermediates such as ribosomes and spliceosomal particles or to protein-RNA complexes of mainly single and tandem RBDs bound to optimized RNA sequences^{19–21}. Yet, the biological function of protein-RNA complexes may depend on the adoption of multiple conformations, which may be important for binding distinct targets^{14,22}. Integrative structural biology approaches that combine multiple methods overcome the restrictions of classical techniques in the characterization of such heterogeneous structures¹⁸. However, integrating a variety of structural constraints with different levels of precision and length, as well as, the aforementioned features of the protein-RNA complexes, also necessitates a shift in the concept of structure determination during the modeling stage.

The abundant 57 kDa RBP polypyrimidine tract protein 1 (PTBPI, also hnRNP I) is a general regulator of cellular mRNA metabolism. In the context of splicing regulation, PTBPI acts either as a splicing repressor, potentially by binding competition with other factors, or as a splicing activator in a position-specific manner^{23–27}. PTBPI also determines the localization, stabilization and polyadenylation of many target RNAs and is involved in the translational regulation of cellular and viral mRNAs as a prototypical internal ribosome entry site (IRES) trans-acting factor (ITAF)²⁸. IRES elements allow non-canonical translation initiation in a 5'-cap independent manner. These large, highly structured RNA sequences are present in the 5' untranslated regions (UTRs) of particular cellular mRNAs and of the genomes of positive-strand RNA viruses. IRES sequences enable translation during global repression of canonical translation due to cell stress. For instance, during viral infections, canonical translation is arrested and non-canonical translation activates the antiviral response and programmed cell death to prevent further viral spread. In this context, PTBPI acts as an ITAF and favors the translation of important pro-apoptotic factors^{29–31}. Similarly, RNA viruses have evolved PTBPI-responsive IRES elements in their 5' UTRs to hijack the antiviral response and favor the translation of their viral proteins^{32,33}. PTBPI is among the most frequently found ITAFs and it is an extensively characterized enhancer of the IRES-mediated translation in picornaviruses, including poliovirus (PV), human rhino virus (HRV), hepatitis C virus (HCV), foot and mouth disease virus (FMDV) and several cardiomyocytosis viruses such Theiler's murine encephalomyelitis virus (TMEV) and encephalomyocarditis virus (EMCV)³⁴.

Structurally, PTBPI consists of four RRM, a flexible N-terminus, which contains both nuclear localization and nuclear export signals, and IDR linkers connecting RRM1 with RRM2, as well as, RRM2 with RRM3, respectively^{8,35,36} (Fig. 1a). The nuclear magnetic resonance (NMR) solution structures showed that each of the four RRM adopts the classical RRM topology of $\beta\alpha\beta\alpha\beta$, which is extended by a fifth β -strand in RRM2 and RRM3, and that each domain binds to specific cytosine/uridine-rich sequences within single-stranded RNA^{10,37}. It has been proposed that PTBPI could act as an RNA chaperone since RRM3

and RRM4 stably interact, thereby spatially restricting the orientation of the cognate RNA-binding sites^{8,35}. Recent NMR studies of the individual RRM1 and RRM2 bound to UCUUU pentaloops illustrated the capability of PTBPI to bind to structured RNA targets and identified an α -helix at the C-terminus of RRM1 as of importance for sensing RNA secondary structure^{38,39}. However, the interplay of the four domains and their assembly on a natural RNA target remain unknown. Understanding the RNA-binding mode of PTBPI is important, as misregulation of its expression has been involved in disease promotion, including colorectal cancer invasion, breast cancer cell growth and Parkinson's disease^{40–42}.

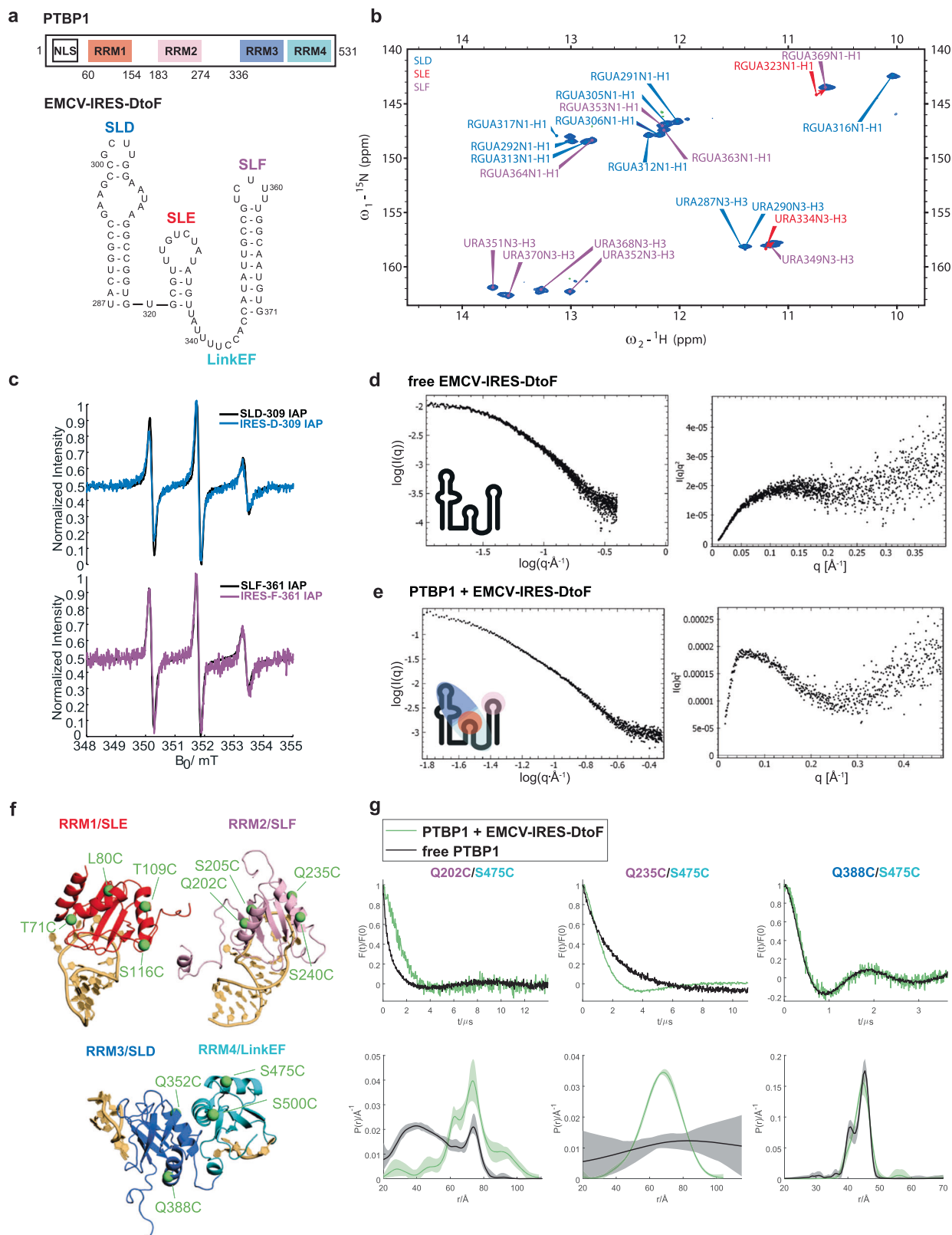
Previous studies of PTBPI in complex with the IRES of EMCV (834 nucleotides, nts) suggested a protein-RNA binding ratio of 2:1 with binding of one molecule of PTBPI to the IRES stem-loops (SLs) D-F (herein referred to as EMCV-IRES-DtoF, Fig. 1a) and of a second molecule to SLs H-L⁴³. Mutation of the RNA interface of the individual RRM reduced IRES activity for RRM1, RRM3 and RRM4 and abrogated IRES activity for RRM2 and RRM4⁴⁴. Interestingly, mutation of the RNA interface of an individual RRM did not change the binding of the other RRM. Recently, we revised the interaction pattern of PTBPI with EMCV-IRES-DtoF and proposed three-dimensional models of the individual PTBPI RRM in complex with parts of the RNA³⁹.

Here, we applied an integrative structural biology approach combining electron paramagnetic resonance (EPR), NMR, small-angle neutron and X-ray scattering (SANS/SAXS), and cross-linking of segmentally isotope-labeled RNA with tandem mass spectrometry (CLIR-MS/MS), to study the structure of the full-length PTBPI in complex with an RNA derived from the IRES of EMCV (EMCV-IRES-DtoF, 84 nts). PTBPI and EMCV-IRES-DtoF are highly dynamic molecules in their isolated states. Upon complex formation, they assemble into a compact and ordered structure while still sampling a fairly large conformational space. This conformational variety can only be described by a structural ensemble. As confirmed using EPR and bicistronic reporter assays, binding of RRM4 to the single-stranded RNA linker between SLE and SLF is crucial for a stable complex assembly as well as for IRES activity, consistent with an RNA chaperone role of PTBPI in enhancing IRES-mediated translation.

Results

PTBPI and the IRES RNA are flexible molecules in solution

As a first step toward the structural characterization of PTBPI in complex with EMCV-IRES-DtoF, we analyzed the RNA sequence conservation within cardiomyocytosis (Supplementary Table 1) and the structural conformations of the protein and the RNA in their isolated states using NMR, EPR and SANS/SAXS (Supplementary Fig. 1). The bioinformatics analysis of the EMCV-IRES-DtoF RNA (Supplementary Fig. 1c, d) suggested evolutionary conservation of the stem-loops D and F (herein referred to as SLD and SLF, Fig. 1a), while the stem of stem-loop E (SLE) appeared to be preserved only in EMCV-related viruses. The PTBPI binding sites for all four domains³⁹ in the RNA (pyrimidine-tracts) are also conserved in cardiomyocytosis (Supplementary Fig. 1c, d). Using NMR, the predicted secondary structure of EMCV-IRES-DtoF was validated through the detection of imino proton signals, which report on the base pairing within the RNA SLs. All expected imino protons of stem-loops D, E and F were assigned by assignment transfer from the individual SLs to EMCV-IRES-DtoF (Fig. 1b, Supplementary Fig. 1a, b). The imino proton signals of the individual SLD (except G317) and SLF overlapped perfectly with their corresponding signals in the EMCV-IRES-DtoF RNA, indicating the same fold. Imino proton resonances for the G323-U334 base-pair of SLE were detectable, even though the U334 signal was slightly shifted compared to the isolated SLE that embedded a longer stem (Fig. 1b, Supplementary Fig. 1b). This indicated that the four base-pairs containing stem (321-324/333-336) exists in the EMCV-IRES-DtoF construct, and that, overall, the predicted RNA secondary structure forms in solution, in agreement with its evolutionary



conservation (Supplementary Fig. 1c, d). By continuous wave (CW) EPR experiments, we observed that paramagnetic spin labels (iodoacetamido proxyl, IAP) attached to isolated SLs reveal the same EPR line shapes compared to corresponding positions within the entire EMCV-IRES-DtoF, reflecting that the tumbling rate of the spin labels between individual SLs and the full-length RNA construct remained unaffected (Fig. 1c). The small peak-to-peak line widths in these spectra, as well as

the nearly symmetric EPR lines, showed that spin labels are in the fast tumbling regime when they are attached to RNA. The spectra are comparable to the previously reported CW-EPR spectra in solution for the individual SLs of the EMCV IRES and exhibit narrower lines than for spin labels attached to single RRMs, which are larger and tumble more slowly⁴⁵. Thus, the spectra indicate substantial relative tumbling of individual SLs with respect to each other in free EMCV IRES. We

Fig. 1 | PTBP1 and EMCV-IRES-DtoF are flexible in the free state and rigidify within the complex. **a** Domain scheme of PTBP1 and an overview of the EMCV-IRES-DtoF RNA construct used in this study. **b** 2D ^1H - ^{15}N TROSY spectrum of EMCV-IRES-DtoF imino cross-peak signals confirm base-pairing. Imino cross-peaks of G323 and U334 of SLE shifted as indicated by red arrows. **c** CW-EPR spectra of SLD and SLF (black) and EMCV-IRES-DtoF (magenta) with spin label attached to nucleotide 309 and 361, respectively. The sharpness of the signals does not change, implying a similar flexibility of the label in the single SL RNA as in the whole EMCV-IRES-DtoF. **d** The SAXS data measured on free EMCV-IRES-DtoF shows in the Kratky plot (right) the characteristics of a flexible and unstructured molecule. **e** The SAXS curve of the PTBP1-EMCV-IRES-DtoF complex shows in the Kratky plot (right) characteristics of an overall structured molecule with reduced flexibility compared to the free RNA shown in (d). **f** Overview of spin-label attachment sites on the

individual RRM of PTBP1 used in this study. Additionally, spin labeling sites were placed in the linker region between RRM1 and RRM2 (152 and 156) and RRM2 and RRM34 (288, 315, 327), respectively. **g** Selected DEER-EPR measurements using spin labels attached to RRM2 (202, 235), RRM3 (388) and RRM4 (475). In the free state (black), signals decay smoothly (top panel), reflecting a broad distance distribution (bottom panel) with the exception of RRM3-RRM4, which interact stably and, thus, lead to an oscillatory decay corresponding to a narrow distance distribution. Upon complex formation (green), the mean distance between RRM2 and RRM4 can increase (202/475) or decrease (235/475), while there is no significant change upon binding for RRM3/RRM4 (388/475). These distance distributions reflect a more ordered and rigidified arrangement of the RNA within the complex. Semi-transparent areas in distance distribution plots correspond to 95% confidence intervals. The data underlying panels (b–e, g) are provided as Source Data.

therefore conclude that in its free state, EMCV-IRES-DtoF is a highly flexible molecule. This conclusion is supported by SAXS measurements recorded on the free EMCV-IRES-DtoF, which resulted in a Kratky plot showing a plateau and no maximum, which is characteristic for highly flexible molecules⁴⁶ (Fig. 1d). The experimental maximal distance D_{max} is 129.00 Å, the radius of gyration (R_g , see also Supplementary Table 2) is 33.41 Å. This is larger than the value of 24.1 Å calculated according to Hyeon et al.⁴⁷ for folded RNA and thus suggests a structure that is less compacted than a folded RNA with tertiary contacts. Taken together, NMR, EPR and SAXS data show that the RNA in its free form adopts the predicted secondary structure with a large degree of overall conformational flexibility.

Acquiring distance information by EPR requires site-directed spin labeling at two sites within the protein or RNA. For proteins, we exploited methanethiosulfonate spin labeling (MTSSL) or IAP spin labeling and used the four-pulse double electron-electron resonance pulse (DEER) experiment for measuring electron spin-spin distance distributions^{45,48}. The inter-domain DEER measurements between spin-labeled sites located in RRM2 and RRM34 revealed broad distance distributions (Fig. 1f, g). This implies the absence of any defined placement of RRM2 with respect to RRM34, consistent with an independent tumbling of RRM2 (as well as RRM1) as noted in previous NMR studies⁸. In contrast, the distance distribution between spin labels attached to RRM3 and RRM4 reflects the well-defined mutual arrangement of the two RRM by hydrophobic interactions between the α -helices of the RRM as described earlier³⁵. Analysis of the small-angle neutron scattering (SANS) curves suggests an elongated shape for the full-length PTBP1, as already described earlier by others⁴⁹. Interestingly, ensemble modeling of free PTBP1 using EOM⁵⁰ resulted in 14 unique models that (1) were not all extended and (2) did not show any orientational preferences between the rigid bodies RRM1, RRM2 and RRM34. The R_g and D_{max} are 46.20 Å and 168.20 Å (Supplementary Table 2).

In summary, the NMR, EPR and SANS/SAXS data show that the free forms of both the RNA and protein are best described by rigid, well-defined bodies (RRMs or SLs) connected by flexible linkers that allow fluctuating spatial arrangements.

PTBP1 and RNA form a compact complex with residual dynamics

Complex formation between PTBP1 and EMCV-IRES-DtoF causes drastic changes in the structural arrangement of both components. Inter-domain distance distribution measurements by DEER showed clear oscillations in the dipolar evolution functions, corresponding to significantly narrower distance distributions compared to the respective free states (Fig. 1g). This clearly indicates a preferred domain-to-domain orientation within the complex. Considering that all domains bind distinct sites on the RNA³⁹, this implies that the RNA structure is also stabilized in a preferred conformation. Importantly, the local secondary structure of the RNA remains unchanged upon PTBP1 binding since the imino proton signals were identical compared to the free state (Supplementary Fig. 1e). In total, we measured a set of 35

DEER distance distributions between 17 spin labeling sites on doubly-labeled PTBP1 in complex with native EMCV-IRES-DtoF (Supplementary Fig. 2, Supplementary Table 3). Interestingly, not all selected spin-label pairs in PTBP1 resulted in narrow distance distributions in the complex, and some spin pairs exhibited standard deviations of distances of up to 28.1 Å. This exceeds the combined conformational flexibility of the two labels of at most 12 Å for MTSSL and 15 Å for IAP by far and thus reflects the preserved flexibility of the complex. Therefore, although the PTBP1/EMCV-IRES-DtoF complex undergoes rigidification and compaction, it also retains certain dynamics and is likely to adopt multiple conformations. In other words, the disorder-to-order transition upon binding is incomplete. Interestingly, DEER measurements between residues located in the RRM and the protein linkers revealed that the distance distributions between RRM1 or RRM2, and their connecting linker (Link12) are narrower than between RRM2 or RRM3, and Link23 (Supplementary Fig. 2, Supplementary Table 3). Hence, Link12 seems to be more restrained in the complex state and less flexible than Link23. The results from EPR and NMR spectroscopy are in line with SAXS data on PTBP1/EMCV-IRES-DtoF and the free RNA. The experimental R_g and D_{max} are 45.03 Å and 138.75 Å, respectively, which are thus smaller than for the unbound protein, indicating compaction. The Kratky plot of the complex compared to the free RNA (Fig. 1e) reflects a more structured molecule⁴⁶ confirming that the IRES RNA, which contributes dominantly to the scattering, is rearranged upon PTBP1 binding.

Structure determination of the PTBP1-EMCV-IRES-DtoF complex

Due to the substantial flexibility of the PTBP1-EMCV-IRES-DtoF complex, its structure must be represented by an ensemble of conformers. Several approaches have been developed for ensemble modeling based on experimental restraints^{50–58}, with some of them specifically geared toward utilizing distance distribution information from EPR^{59–65}. Recent hybrid approaches have incorporated experimental data into molecular dynamics simulations to bias the simulations and generate a raw ensemble through enhanced sampling of the relevant regions of conformational space^{66,67}. In the subsequent steps of this process, refining the ensemble may involve reweighting to improve the fit of the experimental data to generate the final ensemble^{66,68}. These sophisticated approaches draw upon Bayesian statistics to effectively balance the significance of various types of information that may not be entirely consistent. However, given the size and complexity of the PTBP1-EMCV-IRES-DtoF complex, the application of these theoretically well-rooted yet intricate approaches encounters challenges of a technical and computational nature.

Therefore, we have developed a simplified and computationally efficient structure determination protocol that incorporates both short-range distance restraints from NMR and CLIR-MS/MS and inter-domain, long-range distances from EPR and SANS/SAXS to calculate the structure of the PTBP1/EMCV-IRES-DtoF complex. This approach builds upon our previously introduced methodology^{60,61}. Here, we essentially defined rigid bodies comprising RRM/RNA sub-complexes,

which were positioned with respect to each other by EPR long-range distance restraints using the simulated annealing structure calculation algorithm CYANA⁶⁹. In this step, we used lower and upper bounds for the distances derived from the mean values and standard deviations of the distance distributions reported in Supplementary Table 3. Hence, we also relied on the enhanced biased sampling of the relevant regions of conformational space^{58,66}, which, in our case, becomes possible at the level of single conformers by having access to distance distributions rather than only the mean values of distances^{62,65}. We note that the standard deviation serves as a measure for the width of the distance distribution, encompassing the combined disorder of the protein backbone and spin-label sidechain. The mean value of the distribution is defined with much higher precision than this standard deviation may imply. The structures of conformers in the initial ensemble were refined by YASARA, and the entire ensemble was fitted against EPR and SAS data (Supplementary Fig. 3). In the subsequent ensemble reweighting step, we utilized the complete distance distribution information, including the shape of the distributions, which may deviate from a Gaussian shape. We accomplished this by maximizing the mean overlap between simulated and experimental distance distributions. This overlap quantifies the shared area between the two distributions. Given that both distributions are normalized to the unit area, this parameter ranges from 0 and 1, where 0 signifies complete disagreement and 1 indicates perfect agreement.

The following observations allowed us to confidently define that the PTBPI/EMCV-IRES-DtoF complex consists of three rigid building blocks connected by flexible peptide and RNA linkers. First, through a combination of NMR and CLIR-MS/MS data for each of the four RRM of PTBPI, we could place their location to a unique RNA part of the EMCV-IRES-DtoF³⁹. The NMR spectra of the full-length PTBPI bound to EMCV-IRES-DtoF overlap with those of the individual domains bound to their respective RNA targets, suggesting no apparent interactions between the individual domains themselves and with the inter-domain linkers or the N-terminus. This analysis also confirmed that the structures of individual RRMs bound to their RNA targets are preserved in the full-length protein. Second, as a basis for performing site-directed spin labeling (SDSL) and EPR-based distance measurements on the full-length PTBPI protein, we demonstrated that the EPR-based distance distributions within each isolated RRM are consistent with the NMR-based structures⁴⁵. Importantly, the distance distributions between the labeled positions within each RRM of the full-length protein were identical to those measured on the isolated RRMs (Supplementary Fig. 2). Finally, as shown above, the complex formation does not alter the secondary structure of the EMCV-IRES-DtoF RNA (Supplementary Fig. 1e).

Accordingly, we defined RRM1 bound to SLE, RRM2 bound to SLF and RRM34 bound to SLD (RRM3) and to four nucleotides of the LinkEF (RRM4) as rigid bodies. The structural details of these sub-complexes have been determined previously using a hybrid approach of NMR, CLIR-MS/MS and available structural data³⁹. Uncertainties in the RRM1/SLE model were resolved by additional NMR analysis, including the assignment of intermolecular nuclear Overhauser effects (NOEs) to confirm the binding register (Supplementary Fig. 4). Structures of the individual RRMs and SLs were used as input for the structure calculation in CYANA and kept rigid by fixing the backbone torsions within the corresponding segments throughout the calculation. To form RNA stems, we used hydrogen-bond restraints while loops and bulges were maintained flexible. Intermolecular restraints between RRMs and SLs were incorporated to obtain arrangements as in the previously determined structures. In addition to the 35 long-range, protein-protein EPR distance distributions between these three individual sub-complexes, we additionally recorded SAXS data as well as a SANS contrast variation series. The EPR-derived long-range distances were incorporated into CYANA structure calculations as upper and lower limit constraints to position the rigid bodies with respect to each other. Using CYANA, we

calculated a raw ensemble of 20,000 models, of which the 2000 models with the lowest target function were selected (for details, see “Methods” and Supplementary Fig. 3). Subsequently, this ensemble was filtered for structural integrity and the individual conformers were refined using YASARA. We refer to the ensemble of 766 conformers obtained at this point as the raw ensemble.

To enhance the fit with experimental restraints, we employed integrative ensemble reweighting, where populations are assigned to individual conformers. Simultaneously, we contracted the ensemble by removing conformers with very small populations, while maintaining an accurate representation of the experimental data. This leads to a more concise representation of the ensemble. The process of ensemble reweighting was performed using MMMx⁶⁴, resulting in a final ensemble of 25 conformers (Supplementary Table 4). Each of these conformers is annotated with a distinct probability that ensures a well-balanced fit to the 35 EPR distance distributions between protein sites and the three small-angle scattering curves (Fig. 2a). We refer to this reweighted and contracted ensemble as the main ensemble, and we will elaborate on its structural characteristics in a subsequent paragraph. Altogether, the main ensemble achieves a fit to all the distance distributions, with an overlap ranging between 0.396 and 0.845 for the experimental and back-calculated distance distributions, and a geometric average of 0.662 (Supplementary Fig. 5). A loss of merit of 0.332 upon integrating distance distribution and small-angle scattering restraints indicates that both sets of restraints are largely, but not entirely, consistent with each other (Supplementary Figs. 3 and 6a–f). The good agreement of the back-calculated SAXS and SANS curves in the small-angle range demonstrates the accuracy of our model in fitting the size of the complex.

We conducted a series of tests to assess the robustness of the main ensemble. In the first test, we excluded the 25 conformers of the main ensemble from the raw ensemble and then repeated the ensemble reweighting and contraction with only the remaining 741 conformers. The resulting validation ensemble comprises 30 conformers and demonstrates a smaller loss of merit (0.266) compared to the main ensemble. Furthermore, a geometric average distance distribution overlap of 0.683 is achieved, slightly surpassing that of the main ensemble (Supplementary Fig. 7a, Supplementary Table 5). Among the small-angle scattering curves, the fit quality for the 4 m SANS curve slightly improves, while there is a slight decline in fit quality for the SAXS curve and a moderate decline for the 1.2 m SANS curve. When all conformers are superimposed on RRM34, pseudo-electron densities of the main and this validation ensembles exhibit 79% overlap with each other (Supplementary Table 6).

For the second test, we implemented a new ensemble reweighting algorithm into MMMx. With this algorithm, we fitted the populations of all 761 conformers in the raw ensemble simultaneously by non-negative linear least squares (NNLLSQ ensemble). After the fitting process, we eliminated conformers with less than 0.1% of the total population, resulting in an ensemble composed of 32 conformers. This approach involved fitting to background-corrected DEER traces rather than distance distributions and utilizing least-squares instead of minimizing χ^2 for the small-angle scattering curves. Nevertheless, we assessed the fit quality using the same figures of merit employed for the main ensemble (Supplementary Table 5). The pseudo-electron density of the NNLLSQ ensemble exhibits 73% overlap with the main ensemble (Supplementary Table 6). The NNLLSQ ensemble is visualized in Supplementary Fig. 7b.

In a third test, we maximized distance distribution overlap while ignoring the small-angle scattering curves. This DEER-only ensemble, visualized in Supplementary Fig. 7c, features 62 conformers. Its pseudo-electron density exhibits 68% overlap with the main ensemble (Supplementary Table 6).

Comparison of the fit quality of all these ensembles with the raw ensemble (Supplementary Table 5) reveals that ensemble reweighting

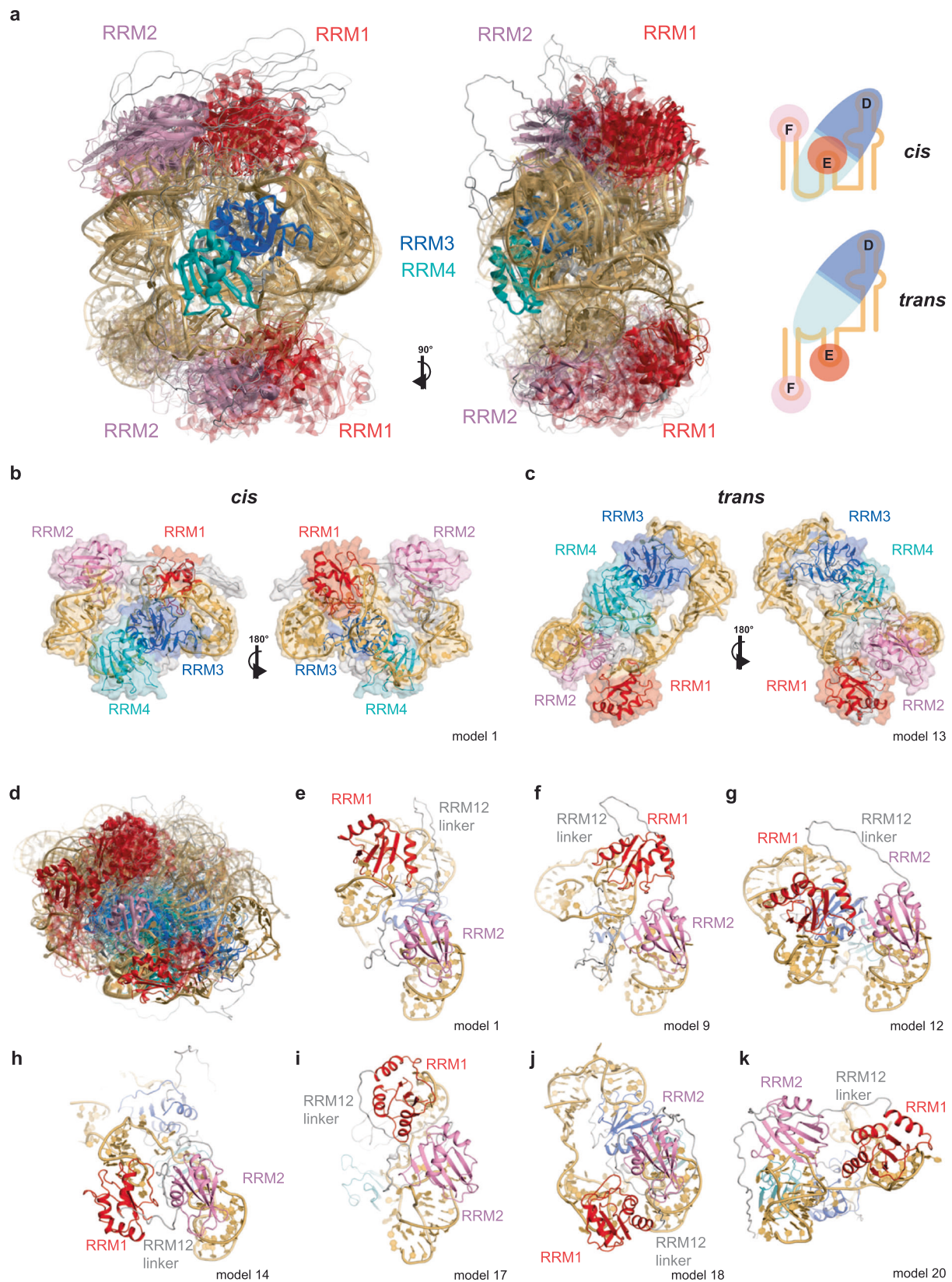


Fig. 2 | PTBP1 and EMCV-IRES-DtoF form a compacted complex with pronounced conformational flexibility. **a** Structural ensemble of the PTBP1/EMCV-IRES-DtoF complex in a conformer population-weighted visualization (opacity corresponds to the population of each conformer). The 25 models of the ensemble were superimposed on RRM34. Two views related by a vertical rotation of 90° are shown. SLE and SLF can be positioned in *cis* or *trans* with respect to SLD, illustrated

in schemes on the right. **b, c** Conformers representative of the two subclasses (*cis* and *trans*) in two views. **d** Overlay of the ensemble on RRM2 illustrating the conformational flexibility of RRM1 with respect to RRM2. **e–k** Examples of conformers showing different ways RRM1 and RRM2 can be spatially close within the PTBP1/EMCV-IRES-DtoF complex.

and contraction generally improve figures of merit. The raw ensemble provides already a good overlap with the distance distributions, as the distance restraints were considered in its generation, but very poor agreement with the small-angle scattering curves. This latter agreement improves substantially in the main and validation ensemble, albeit at the expense of some deterioration of overlap with the distance distributions. This suggests a certain level of inconsistency between DEER data and small-angle scattering curves, which might be caused by the requirement to perform DEER measurements under cryogenic conditions. Below, we will discuss the features of the main ensemble, keeping in mind the uncertainty implied by the variation between the main, validation, and NNLSQ ensembles.

In a final validation fit, we computed a model of only the protein in the context of the complex by ignoring topological restraints imposed by the RNA. In this, we could still use the SANS curves, since the solvent matches RNA neutron scattering, as well as the EPR distance restraints, since all our labeling sites are on PTBP1. We generated the raw ensemble by the RigiFlex approach as implemented in MMMx⁶⁴ and performed ensemble reweighting as for the main and validation ensembles, except for skipping the SAXS curve. This protein-only ensemble is visualized in Supplementary Fig. 8a. While the protein-only ensemble cannot be compared quantitatively with the other ensembles, it reveals that the general arrangement of the RRM, including the presence of the cis and trans sub-ensembles discussed below, is encoded by the DEER restraints and the SANS curves alone. Indeed, this arrangement even persists when skipping the SANS curves. This finding implies that the RRM arrangement found in the present study does not depend on the previous CLIR-MS/MS results³⁹, which are, however, required for modeling the RNA.

Finally, we tested how the fit quality and size of the reweighted ensemble depend on the size of the raw ensemble (Supplementary Fig. 9). We observed that the overlap of experimental and back-calculated distance distributions reasonably converges for raw ensembles with a size of more than 600 conformers. The mean χ^2 of the three small-angle scattering curves is also converged for this size of the raw ensemble, whereas the size of the reweighted ensemble varies slightly (21 to 26 conformers).

As seen in Supplementary Fig. 5, the overlap between experimental and fitted distance distributions is generally smaller than expected from only the uncertainty of the experimental distributions. For these broad distributions, additional uncertainty from spin-label rotamer simulations is expected to be a minor contribution. In particular, some of the simulated distributions are clearly bimodal, reflecting the cis and trans sub-ensembles discussed below, whereas the corresponding experimental distributions are so broad that bimodality cannot be resolved. To some extent, this is reflected in the broader conformer distribution of the DEER-only ensemble (Supplementary Fig. 7c). However, even this ensemble exhibits bimodality in conformer distribution. This bimodality is more pronounced in the integrative model that also considers the small-angle scattering data, which may indicate that the shock-freezing required for the DEER measurements affects conformer distribution to some extent. For individual data sets with particularly poor overlap, such as 205/388 and 235/388, we cannot safely exclude that the label itself biases the conformer distribution. However, in the DEER-only ensemble, overlap improves from 0.397 to 0.692 for the pair 205/388 and from 0.449 to 0.605 for the pair 235/388, indicating that inconsistency between distance distributions and small-angle scattering data is the main reason for the poor fit of these distributions by the main ensemble. In the following, we discuss structural features of the main ensemble that are reasonably robust with respect to fit procedure and restraints used.

A structure with several conformations exposing the RNA

The pairwise root mean square distance (rmsd) between the 25 conformers of the PTBP1/EMCV-IRES-DtoF main ensemble varies from 2 to

30 Å over the entire complex (Supplementary Fig. 6g, h), reflecting the wide distance distributions measured for many spin pairs. Despite this apparent structural heterogeneity, the structural ensemble of the complex reveals very interesting features. The best way to illustrate the conformational space covered by this structural ensemble is to superimpose each conformer onto RRM34 (Fig. 2a). This representation reveals that RRM1 and RRM2 localize primarily on two sides with respect to RRM34. Based on the position of RRM1 and RRM2 relative to RRM34, the structural ensemble can be divided into two groups with a major form (13/25 conformers, with a $60 \pm 3\%$ population by amounting the individual probabilities) and a minor form (12/25 conformers, with a $40 \pm 3\%$ population). In the major form, SLF and SLE are in cis with respect to SLD (Fig. 2b), while in the minor form, the two SLs are in trans with respect to SLD (Fig. 2c). In the cis conformation, RRM1 and RRM3 are close to each other, while in the trans conformation, they are more distant. Importantly, despite being separated by a flexible linker, RRM1 and RRM2 are always within close proximity. Yet, RRM1 is not situated at a fixed position relative to RRM2 but occupies multiple sites with distinct orientations (Fig. 2d). Some of the determined RRM1 and RRM2 orientations can be explained energetically by the conformation of the inter-domain linker, molecular contacts involving the domains and the linker or the two domains and protein-RNA intermolecular interactions (see below) (Fig. 2e–k). Independent structural ensembles generated by either removal of the 25 conformers that were included in the best-fitted integrative ensemble (Supplementary Fig. 7a) or fitting only against DEER data (Supplementary Fig. 7c), resulted overall in similar conformational ensembles composed of these two cis/trans families.

PTBP1 RRM34 interacts with EMCV-IRES-DtoF using the same interface and, remarkably, the same directionality, which we had previously identified when studying its binding to single-stranded RNA⁷⁰. Indeed, RRM3 binds the RNA at the 5'-end (SLD) and RRM4 further downstream (LinKEF) using the interaction surface near the small helix of the linker. This is seen in both the major and minor forms of the complex. However, only in the minor form, the trans arrangement brings the intervening RNA near the RRM34 inter-domain linker enabling additional protein-RNA interactions (Fig. 2c). As independent support for these contacts, we could identify protein-RNA cross-links between this region of the linker and the RNA (Supplementary Fig. 10). The detection of these protein-RNA cross-links to U/UU validate the contacts found in this subpopulation of conformers (Supplementary Fig. 10e). Strikingly, hydroxyl radical cleavage data⁴³ for the linker residue N432 support its position close to SLE (Supplementary Fig. 11). The two main conformations originate from RRM4 binding to the linker between SLE and SLF, thereby restricting the mobility of the SLs, but still allowing SLD and SLE to be positioned in cis and trans as not all nucleotides of the LinKEF sequence are bound by RRM4. In the trans conformation, the RRM34 inter-domain linker seems to contribute to the stabilization of this form. Additional “weak contacts” between RRM1 and RRM2 then help further stabilize most conformers of the complex (see below).

Importantly, despite the very strong compaction seen upon the complex formation of the RNA and the protein (Supplementary Fig. 8b), we still observe residual conformational flexibility (Fig. 2a). As indicated above, an unexpected and interesting feature of the PTBP1/EMCV-IRES-DtoF complex is the spatial proximity of RRM1 and RRM2, which is induced upon RNA binding. RRM1 and RRM2 do not strongly interact, resulting in RRM1 being in proximity to RRM2 but not at a fixed position (Fig. 2d). This proximity is facilitated by the 10 residues C-terminal to RRM1 that fold into an α 3-helix upon binding to SLE, reducing the overall length and dynamics of the linker³⁸. Although the DEER data is not sufficiently resolved to discuss features at atomic length scales, and we have only one distance distribution restraint corresponding to the linker between RRM1 and RRM2, the entropy of this linker is somewhat reduced by the relative arrangement of these

two RRM. However, we cannot rule out the possibility that the following observations are influenced by the refinement of the conformer structures in YASARA using the AMBER force field. Close examination of the 25 conformers revealed that RRM1 may favor conformations placing it close to RRM2 because the inter-domain linker is very hydrophobic and folds back as a hairpin, resulting in RRM1 facing the β -sheet of RRM2 (seen in models 1–7) (Fig. 2e). In other cases (models 11–12) (Fig. 2g), contacts between the α 3-helix of RRM1 and the inter-domain linker between RRM2 and RRM3 brings RRM1 close to RRM2. In another subset (models 8–10), the α 2- β 4 loop of RRM1 interacts with the β 3- α 2 loop of RRM2, and the inter-domain linker makes contacts with the α 1- α 2 surface of RRM1, resulting in the β -sheets of both RRMs facing in the same direction (Fig. 2f). In a fourth subset (models 13–15), the RRM1-RRM2 proximity is mediated by the RNA, since SLE is sandwiched between both RRMs (Fig. 2h). In a fifth subset (models 16–17), α 2 of RRM1 mediates contacts to the top part of RRM2 (Fig. 2i) and in one case (model 18), the two domains are arranged back-to-back (Fig. 2j). In a final subset (models 20–25), the two domains do not form any contacts and the inter-domain linker is elongated interacting with the RRMs (Fig. 2k). This illustrates the complexity of the conformational landscape of this type of protein-RNA complex, where weak interactions between folded domains, protein linkers and RNA are competing, resulting in a globally compact but still very dynamic structural ensemble.

Overall, a key aspect of this PTBPI-IRES structural ensemble is that protein binding results in a compacted RNP structure with reduced (compared to the free protein and RNA), but still pronounced conformational flexibility. Importantly, the four RRMs lie in the interior of the structure and expose the three conserved RNA stems to the outside (Fig. 2a–c). This conformation could explain the proposed chaperone role of PTBPI, since exposing these RNA stems presents the IRES structure for subsequent interactions with the ribosome or translation initiation factors as seen in IRES-ribosome structures^{71–74} (see “Discussion”).

RRM4-LinkEF interaction is crucial for IRES activity

The RRM4-LinkEF interaction is central to the complex structure as it stabilizes the longest RNA linker (LinkEF) and thereby participates in the compaction of the complex. We therefore investigated the effect on the structural integrity of the complex by interfering with the RRM4-LinkEF interaction to test its functional importance (Fig. 3). Replacement of the pyrimidine tract in LinkEF with purines (Lmut in Fig. 3b) resulted in a complex with significantly broadened distance distribution between spin labels on RRM1 and RRM34 compared to the non-mutated complex (Fig. 3a), while the RNA secondary structure remained unchanged (Supplementary Fig. 1f). Similarly, mutation of key amino acids for RNA-binding of RRM4 to alanine (H457A, R523A, K528A; referred to as RRM4ko) resulted in broader distance distributions between RRM4 and RRM2, and between RRM1 and RRM2, compared to non-mutated PTBPI (Supplementary Fig. 12). These experiments confirm that RRM4 is crucial for the stabilization of the complex.

To explore if altering the binding of the different RRMs would affect the biological functionality, we investigated the influence of each individual RRM-binding site on IRES activity using a bicistronic luciferase reporter assay in HEK-293T cells⁷⁵. In this assay, Renilla Luciferase is translated 5'-cap-dependent and Firefly Luciferase is translated under the control of the EMCV-IRES, both from the same transcript. Thus, the ratio of Firefly Luciferase activity to Renilla Luciferase activity is a direct measure of IRES activity⁷⁶. We first compared the activity of the EMCV-IRES wild-type (WT) (Fig. 3b, c) with a deletion mutant that lacks the entire EMCV-IRES-DtoF sequence (“DtoF deleted”). The deletion showed about a fifth of the WT activity (down to $19 \pm 1.0\%$), in line with previous reports that this part of the IRES is not essential but important for stimulating the IRES activity³³. Mutating

the individual RRM-binding sites on SLD, SLE or SLF by exchange of pyrimidines to purines led to a strong reduction down to $56 \pm 2.7\%$, $46 \pm 1.8\%$ or $42 \pm 1.8\%$, respectively (Fig. 3b, c). Strikingly, mutating pyrimidines in the RRM4-binding site (LinkEF) to purines reduced the activity to $18 \pm 0.4\%$, similar to the deletion of the complete EMCV-IRES-DtoF. This strongly suggests that structural stabilization and compaction by RRM4 is crucial for EMCV-IRES activity.

Together these experiments revealed rather surprisingly that a flexible RNA linker, which connects two SLs, has an essential role in the biological activity of an RNA molecule. PTBPI binds to the flexible RNA linker and remodels the overall RNA structure. The resulting structure and the ordering of the RNA structure upon PTBPI binding are crucial for the IRES-mediated translation initiation, potentially by facilitating the recruitment of the ribosome or interaction with translation initiation factors^{28,72–74}. The overall conformation that exposes the RNA stems and the remaining conformational dynamics may allow the RNP to present a subset of conformations that can be recognized by the translation machinery through conformational capture. In contrast, without PTBPI binding, the inherent flexibility due to the lack of linker binding loosens the complex arrangement and therefore impedes translation initiation.

Discussion

Here, we present insights into the relationship between structure, dynamics and function of a multi-RBD protein-RNA complex by modeling its ensemble structure, which can lead to multimodal function. Since we took an approach to determine the structure, which integrates five structural methods and aims for the best fit among the different structural constraints, two questions may arise: How accurate is the structural ensemble, and what does it tell us about the biological function? Reassuringly, the presented structural ensemble is consistent with the hydroxyl radical cleavage that was performed by others⁴³, and thus this data can be seen as an independent support for the structural conformers (Supplementary Fig. 11). Additionally, the identification of the trans conformation helped us explain, retrospectively, the cross-links we identified earlier³⁹ between the linker of RRM34 and the RNA (Supplementary Fig. 10). Functionally, our results rationalize the proposed roles of PTBPI as RNA chaperone and ITAF in the context of IRES-mediated translation initiation. In recent years, a few structures of IRES elements bound to ribosomes have been determined^{71–74,77}, however, in all these cases, the IRES RNA is stabilized exclusively by pseudoknots and RNA tertiary interactions, although most viral and cellular IRESs require protein ITAFs for optimal translation activity. Our structural work provides evidence that PTBPI strongly compacts the RNA and stabilizes the RNA fold into discrete conformations. As a result of PTBPI binding, the RNA helices of SLD, E and F are exposed to the solvent for further interactions with the translation machinery, consistent with what was found in protein-free IRESs bound to ribosomes^{71–74,77}. Interestingly, in particular, the sequence of the solvent-exposed SLD containing a purine-rich internal loop is conserved, suggesting that the structure and the sequence of the exposed IRES element may be recognized by the ribosome or additional factors (Supplementary Fig. 1). In the context of EMCV-IRES, the SL J-K are bound by the HEAT-1 domain of eukaryotic initiation factor 4G (eIF4G), which require a pre-organization to position protein-recognition⁷⁸. Interestingly, a second PTBPI molecule has been shown to bind to this region of the EMCV-IRES and thereby may contribute to this organization, which has been shown to be common in type 2 picornavirus IRESs^{43,79}. The compaction and stabilization effect was not necessarily expected because PTBPI itself is a very dynamic protein containing four folded domains separated by flexible linkers, and previous studies on the shape of the free protein have established the view that PTBPI is present in solution as an elongated particle^{36,49}. A number of features unique to PTBPI make the protein able to fold EMCV-IRES-DtoF into a structured RNA: First, the four RRMs separated by flexible linkers allow binding to separate RNA

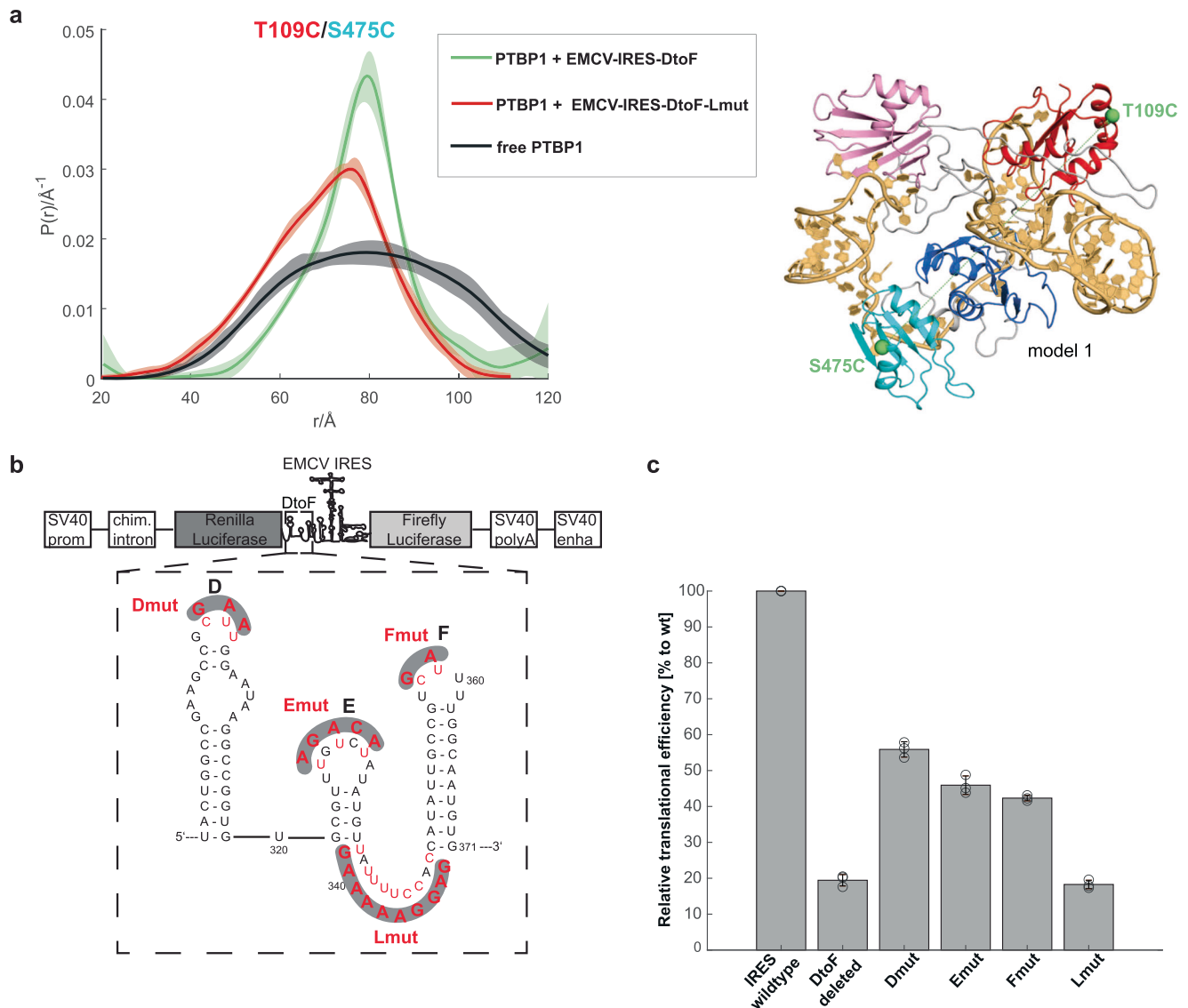


Fig. 3 | RRM4 binding is crucial for complex stabilization and translation enhancement. **a** Removal of the pyrimidine stretch in LinKEF (same mutation as shown in **(b)** for the Lmut Luciferase Reporter Assay construct) interferes with RNA binding of RRM4 and broadens the distance distribution measured between spin labels attached to RRM1 and RRM34 significantly. Compared to free PTBP1, the distance distribution narrows upon binding of EMCV-IRES-DtoF, but the mean distance does not shift substantially. This effect is less pronounced with the Lmut. Semi-transparent areas in distance distribution plots correspond to 95% confidence intervals. **b** Constructs for the Luciferase Reporter Assay (top) with the deleted or

mutated (pyrimidine to purine) sequence of EMCV-IRES-DtoF (bottom, only one mutation at a time). **c** Luciferase Reporter Assay given as IRES-mediated translation efficiency with respect to the WT-IRES. Remarkably, the removal of the RRM4-binding site (Lmut) results in the same activity loss as the deletion of the whole EMCV-IRES-DtoF sequence, which links the importance of RRM4 binding for complex stabilization and rigidification with translational efficiency of the IRES. The error bars represent the standard deviations of the three biological replicates. The value for each biological replicate was determined as a mean of three technical replicates. The data underlying panels **(a, c)** are provided as Source Data.

sequences within the IRES. The observed RNA structure stabilization through PTBP1 binding explains also previous data showing that mutations of individual RRM4 do not disrupt RNA binding, but still result in a strong reduction of translation efficiency⁴⁴. Second, the tight interaction between RRM3 and RRM4 brings distant RNA sequences into close proximity. Our structural and functional study identifies the interaction of RRM4/LinKEF as a major element for structuring the core of PTBP1/EMCV-IRES-DtoF. An increase in flexibility of the complex upon mutation of the RRM4-binding site in LinKEF coincides with a dramatic reduction in translation efficiency. Third, the binding of RRM1 to SLE results in a shortening of the linker between RRM1 and RRM2³⁸. Fourth, there is a close interaction between RRM1 and RRM2 that is mediated by several sets of interactions and not a unique one. This last feature is primarily responsible for the residual conformational

flexibility seen in the final structural ensemble. Conformational flexibility was also seen in protein-free IRES structures^{71–74,77}, so it seems to be a common structural feature of IRESs. This flexibility may help with the initial recognition by the ribosome and also for the IRES to adapt its structure during the different phases of translation that are accompanied by the structural dynamics of the ribosome. In agreement with this, recent in vivo single-molecule experiments showed that EMCV-IRES transitions between translationally active and inactive RNA states⁸⁰. Such RNA remodeling capability of PTBP1 is also likely to be important for splicing regulation, consistent with reports supporting the role of RRM4, and the interaction with RRM3 for splicing activity^{70,81–83}. Thus, compaction and stabilization of the RNP complex is not restricted to IRES-mediated translation initiation, but a general mode of action of PTBP1.

The structural details presented here also provide general insights into how RBPs interact with RNAs. To our knowledge, the only other structure of a multi-RRM containing protein with more than two RRMs in complex with a natural RNA is the crystal structure of the spliceosome-specific RNA chaperone Prp24 bound to U6 snRNA⁸⁴. Both Prp24 and the U6 snRNA experience large structural rearrangements upon complex formation, and the crystal structure has a rigid interlocked topology. In this case, only three of the four RRMs interact with RNA (RRM2-4), and RRM2 interacts extensively with RRM1 and RRM4, while the inter-domain linker between RRM3 and RRM4 folds into two α -helices upon RNA binding. Therefore, PTBP1 and Prp24 use completely different modes of action, although both act as RNA chaperones. Importantly, while Prp24 has a single RNA-binding substrate, PTBP1 has thousands of RNA-binding targets being involved in almost all post-transcriptional gene regulatory processes⁸⁵. Therefore, it may not be surprising that the complex formation of PTBP1 results in a discrete number of complex structures and not a single one. Moreover, it is likely that it would be energetically more favorable for an RNP complex to have a higher degree of flexibility and thereby maintain a higher conformational entropy compared to a rigid structure, especially considering the large entropic cost associated with the binding between two very flexible molecules. The preservation of conformational dynamics in the bound state may also facilitate the recruitment of additional binding partners and allow them to exert additional influence on the RNAs conformational landscape, required to attain functional conformations.

We expect that the approach presented here will prove valuable for future structural characterizations of RNP complexes involving the numerous RBPs in the human genome that, similarly to PTBP1, are composed of multiple RBDs separated by flexible linkers. RNA binding may result in several, substantially different conformations and could be vital to their function. Many RNP machineries rely on major structural rearrangements, as seen in the different steps catalyzed by the spliceosome or the ribosome^{19,20}. Our approach that combines short and long-range distance constraints from NMR, MS and SAS, as well as, distance distributions from EPR, allows us to determine a structural ensemble covering the full conformational space covered by such dynamic RNPs. Distance distributions from EPR experiments inform on the width of conformational distributions and thus represent a basis for integrative structure modeling of large systems, provided that three-dimensional models of the structurally well-defined building blocks already exist. The recent advances in AI-driven structure predictions will aid the availability of such building blocks⁸⁶. Further work is required in perfecting integrative ensemble modeling and, in particular, in providing uncertainty estimates for broad ensembles. Integrating NMR and EPR data, protein-RNA cross-linking restraints, and SAS curves holds much promise as a strategy to determine ensembles of rigid-body arrangements and even protein-RNA condensates, to provide insights into the relationship between structure, dynamics and function of RBPs.

Methods

Site-directed mutagenesis

Positions for site-directed spin labeling (SDSL) on the protein were chosen based on simulating the spin-label attachment to existing NMR solution structures of individual RRMs of PTBP1⁸ using the software package Multiscale Modeling of Macromolecular systems (MMM)⁸⁷. For inter-domain EPR distance measurements, we chose amino acids in the RRMs that are predominantly located within the α -helices, with geometrical similarities and related chemical properties compared to cysteine. Positions in the peptide linker regions were selected close to the C-terminus of RRM1 and RRM2, close to the N-terminus of RRM3 or in the center of the linker. Site-directed mutagenesis was performed using quick-change PCR with specific primers listed in Supplementary Table 7 and pET28a-PTBP1(1-531) plasmid reported earlier as template.

Cys250 and Cys251 were mutated to serines in all constructs. Cys23 was mutated to serine in all protein constructs for EPR measurements.

Protein expression and purification

Protein expression and purification of PTBP1 and individual RRMs were performed as described previously^{39,45}. For distance measurements >5.5–6 nm, expression was performed in 95–99% D₂O using minimal M9 medium followed by the same purification protocol. Samples for NMR and SANS experiments were finally transferred to 10 mM NaPO₄, pH 6.5, 20 mM NaCl buffer with 1–2 mM DTT by size-exclusion chromatography and dialysis.

RNA preparation

In vitro transcription of EMCV-IRES-DtoF and of individual SLs was performed as described previously^{39,45}. For spin labeling of RNA, respective RNA segments were removed by RNase H cleavage and chemically synthesized (Dharmacon) thio-uridinated RNA segments were ligated after spin labeling similarly as described before^{45,60,88}.

Site-directed spin labeling of protein and RNA

Spin labeling protocols used for all PTBP1 mutants have been described elsewhere⁴⁵. For SDSL of PTBP1, we used predominantly MTSSL ((I-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate, Toronto Research Chemicals) or IAP (3-(2-iodoacetamido)-proxyl, Sigma Aldrich). For RNA SDSL, different uridine positions in the loop or linker regions in EMCV-IRES-DtoF were selected and the respective short oligonucleotides containing 4-thiouridine were spin-labeled and ligated to the full-length RNA following earlier published protocols^{45,60}.

PTBP1/EMCV-IRES-DtoF complex formation

Complex formation and purification were performed essentially as described before³⁹. In brief, EMCV-IRES-DtoF was diluted in all cases to approximately 4 μ M in 5 mL low salt buffer (10 mM NaPO₄, 20 mM NaCl, pH 6.5). Concentrated doubly-labeled PTBP1 was mixed with the RNA in a molar ratio of 1:1.2 and the complex was then purified by size-exclusion chromatography using a Superdex 75 column (Cytiva). Complex samples were then buffer-exchanged into D₂O and concentrated to approximately 50–100 μ M. For DEER distance measurements, complex samples were diluted 1:1 (v/v) with d₈-glycerol (Sigma Aldrich), and 30 μ L sample solution was filled into 3 mm quartz tubes. Complex formation was ensured by native RNA polyacrylamide gels.

Bicistronic reporter assays

Bicistronic pRemcvF plasmids were a generous gift from Prof. Dr. A. Willis (MRC Toxicology Unit, Leicester, UK)⁸⁹. EMCV-IRES mutants were generated using standard site-directed mutagenesis procedures. HEK-293T cells were derived from a lab stock and maintained in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal calf serum (GibcoBRL) and antibiotics. For IRES activity assays, a 24-well plate was seeded with 500 μ L cell suspension (approx. 350,000 cells/mL). After 24 h growth, the medium was exchanged to medium without antibiotics. Then, cells were transfected with 500 ng plasmid/well and a final volume of Lipofectamin 2000 (Thermo Fisher) of 1 μ L/well in a total volume of 100 μ L OPTI-MEM (Gibco) medium/well. Cells were incubated for 24 h and lysed using the Dual-Luciferase Assay lysis buffer supplemented with Roche Complete Protease Inhibitor (Roche, incubation of 10 min on a shaker). Lysates were cleared by centrifugation (16,000 \times g, 4 °C). Then, 5 μ L of lysate/well was pipetted into a 96-well ELISA plate and 25 μ L of Firefly-Luciferase substrate solution was added and measured after 15 s shaking in the plate reader. Then, 25 μ L of Stop-N-Go solution was added and samples were measured to assess the Renilla-Luciferase activity. All IRES assays were performed in technical triplicates of biological triplicates and measured on a Berthold MicrolumatPlus luminometer. IRES activity is determined by normalizing the Firefly-

Luciferase activity to the Renilla-Luciferase activity, whereas the activity of the WT-plasmid is set to 100%.

CLIR-MS/MS

Mass spectrometry data from previous CLIR-MS/MS analysis of the PTBPI-IRES complex³⁹ was reanalyzed with an optimized parameter set containing more neutral loss products, to increase the number of identified protein-RNA cross-links. Mass spectrometry data files (Thermo Fisher RAW format) corresponding to the uniformly labeled EMCV-IRES RNA experiment were retrieved from ProteomeXchange via the PRIDE partner repository (PXD005566) and converted to mzXML using msconvert.exe (ProteoWizard msConvert v.3.0.9393c)⁹⁰. The files were searched using RNxQuest⁹¹, an add-on to xQuest (both available for download from https://gitlab.ethz.ch/leitner_lab/)^{92,93}, against a database containing only the target PTBPI protein sequence, with expected light-heavy RNA adducts defined as mono-links. Adducts with lengths of 1–4 nucleotides were considered. In addition to the whole RNA adduct, the following neutral losses were considered: -H₂O, -H₄O, -HPO₃, -H₃PO₃, -H₂, -HPO₃+H₂O, -HPO₂, -H₄O₂, -H₅PO₅. Further details about the expanded parameter set can be found elsewhere⁹⁴. All amino acids were set as cross-linkable. Delta masses for each adduct were defined depending on the nucleotide sequence composition, as described previously³⁹. Further parameters used for xQuest searching: delta mass tolerance: +/- 15 ppm, retention time tolerance: 60 s, enzyme = Trypsin, maximum missed cleavages = 2, MS1 mass tolerance = 10 ppm, MS2 mass tolerance = 0.2 Da. Identifications with an Id.Score > 20 (according to the scoring scheme described previously⁹²) were considered. Further processing and visualization were performed using the scripts contained in the RNxQuest package. The reanalyzed mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository⁹⁵ with the dataset identifier PXD034894.

NMR spectroscopy

All NMR experiments were recorded on Bruker Avance III 500, 600, 700 or 900 MHz spectrometers equipped with cryo-probes and on a Bruker Avance III 750 MHz spectrometer with a room temperature probe. NMR protein spectra (free proteins and protein-RNA complexes) were acquired at 313 and 323 K, imino ¹H-¹H-NOESY spectra were recorded at 278 or 283 K. Imino-¹H¹⁵N-TROSY spectra were recorded at 278 and 293 K. We processed spectra with Topspin 2.1 or Topspin 3.0 and analyzed in Sparky 3.0⁹⁶. ¹H-¹H imino assignments of the RNA were achieved by standard methods^{97,98}. Imino resonances of SLE were assigned using conventional jump-return NOESY with 200 ms mixing time and facilitated with SMT experiments⁹⁹ at 278 and 275 K. Non-exchangeable protons were assigned using sequential walk in NOESY spectra recorded at 298 K and in 100% D₂O NMR buffer (10 mM NaPO₄, pH 6.5, 20 mM NaCl buffer with 1 mM DTT). Assignment of other hydrogens and heteronuclei was obtained using intra-nucleotide assignment experiments, including aromatic HSQC, CT HSQC of sugar resonances, TROSY HCCH COSY for adenine H2-H8 assignment, HCCH COSY for H5-H6 correlations of pyrimidines, and finally HCN and HCNCH experiments. Assignments of protein resonances were obtained using conventional triple resonance experiments, HCCH-TOCSY and 3D NOESY-HSQC. Intermolecular NOEs were identified in ¹³C,F3-filtered HSQC-NOESY experiments using 100–150 ms NOE mixing time and L-PROSY NOESY recorded with 10 loops and 40 ms per loop^{100,101}. Throughout the study, simple titration experiments were recorded using 0.2 mM protein/RNA samples, assignments experiments for molecules in the free state were performed with 1 mM samples, and the protein-RNA complexes were at approximately 0.7–0.8 mM of each molecule.

EPR spectroscopy

Labeling efficiencies and spin-label tumbling regimes of spin-labeled protein and RNA samples were determined by CW-EPR spectroscopy at

ambient temperature. Experiments were performed at X band (9.5 GHz) using a Bruker Elexsys E500 spectrometer with a Bruker super-high Q resonator ER4122SHQ and spectra were recorded with a field modulation of 100 kHz, a modulation amplitude of 1 G, a time constant of 10.24 ms, a conversion time of 40.96 ms and an attenuation of 25 dB of 200 mW incident microwave power. Samples were filled into glass capillaries (BLAUBRAND®) with a diameter of 1.5/0.9 mm (outer/inner). Labeling efficiencies were determined by digital double integration of the EPR spectra and comparing them to free nitroxide radical IAP of standard concentration. Changes in the spin-label mobility could be detected by amplitude reduction of the low- and high-field components of the nitroxide spectra, which indicated successful spin labeling.

To obtain EPR distance restraints, four-pulse DEER experiments⁴⁸ were performed at a home-built high-power Q-band spectrometer (35 GHz) with 200 W microwave power, a Bruker ElexSys acquisition system (E580) and a home-built TE001 pulse probe¹⁰². Temperature stabilization during all measurements was ensured by a He-flow cryostat (ER 4118CF, Oxford Instruments) and a temperature control system (ITC 503, Oxford Instruments). All free protein and complex measurements were carried out at 50 K because this temperature corresponds to the optimal conditions for nitroxide radicals with respect to their longitudinal and transverse relaxation. Pump pulses were always set to 12 ns and detection pulses to 16 ns. While a setup with all 12 ns pulses may provide better sensitivity, it requires a very good resonator mode and thus careful sample volume and sample position adjustment. In practice, we found it more convenient to slightly compromise the bandwidth of the detection pulses by increasing their length from 12 ns to 16 ns. The first inter-pulse delay τ_1 was set to 400 ns, τ_2 was set according to the expected distance and its required DEER trace length for suitable background correction. Time increment t was also adapted to the length of the τ_2 as described earlier⁴⁵. The pump pulse was always applied on the maximum field position of the nitroxide spectrum, whereas detection pulses were applied by an offset of approximately 100 MHz. Samples that were diluted 1:1 (v/v) with d₈-glycerol (Sigma Aldrich) were shock-frozen with liquid nitrogen and measured in this state.

Analysis of DEER data

For processing all DEER data we used the software package DeerAnalysis¹⁰³, version 2022, in automated mode. When an artifact arose from the overlap between the excitation bands of the pump and detection pulses^{103,104}, up to 15% of the data was cut at the end. For background correction, we used a dimensionality of 3, corresponding to a homogeneous three-dimensional distribution of complexes in the sample as expected for soluble proteins. Automated comparative analysis computes distance distributions by neural network analysis with DeerNet¹⁰⁵ and Tikhonov regularization¹⁰⁶ with DeerLab in a single step with background correction¹⁰⁷ and provides 95% confidence intervals for both distributions. Unless otherwise indicated, we report the mean of the two distributions and confidence intervals that include uncertainty due to model bias. The mean distances ($\langle r \rangle$) and standard deviations ($\sigma(r)$) of the distance distributions were implemented as upper and lower limits in CYANA modeling (see below).

Small-angle scattering experiments

Small-angle neutron scattering experiments were recorded at the SANS-I and SANS-II facility, Swiss Spallation Neutron Source, SINQ, Paul Scherrer Institute, Switzerland. Scattering density matching points for RNA and protein with respect to the D₂O/H₂O ratio of the buffer (10 mM NaPO₄, pH 6.5, 20 mM NaCl, 2 mM DTT) were determined by contrast variation and extrapolation of I₀ to be 66% D₂O for RNA, and 42% or 112% for protonated or deuterated protein, respectively. Free protein and RNA were recorded at a wavelength of the neutron beam of 6 Å, a collimation of 6 m and a detector distance of 2 m and 6 m. SAS data collection parameters are summarized in Supplementary Table 8.

For both, SANS and SAXS experiments, complexes were reconstituted as described previously³⁹ and dialyzed for 24 h against suitable buffers. Each reference cuvette was filled with the corresponding dialysis buffer. Complexes were measured at SANS-I at a wavelength of 4.5 Å and either a collimation of 11 m and a detector distance of 11 m or a collimation of 3 m and a detector distance of 2 m. Scattered neutrons were detected using a two-dimensional 96 × 96 cm² detector with a pixel size of 0.75 cm. The cross section of the collimator was 50 × 50 cm² and the effective sample diameter was 0.8 × 1.5 cm². At SANS-II, the selected wavelength was 4.9 Å and collimation and detector distance were 2 m and 1.2 m or 4 m and 4 m, respectively. The detector had a diameter of 64 cm with 128 × 128 pixels.

Reduction and analysis of SANS data were performed with the program BerSans¹⁰⁸ and visualized using Primus QT¹⁰⁹ from the ATSAS program¹¹⁰. Beamline-specific correction factors for data reduction are 1.338 and 1.284 for data recorded at a wavelength of 4.5 and 6 Å, respectively. SAXS experiments were performed on the liquid samples using a Rigaku MicroMax-002 microfocussed beam (4 kW, 45 kV, 0.88 mA). The Cu K α radiation ($\lambda_{\text{CuK}\alpha} = 1.5418$ Å) was collimated by three pinholes (0.4, 0.3, and 0.8 mm) collimators. The scattered X-ray intensity was detected by a two-dimensional Triton-200 X-ray detector (20 cm diameter, 200 mm resolution) for SAXS. SAXS detectors cover an effective scattering vector range of $0.1 \text{ nm}^{-1} < q < 4 \text{ nm}^{-1}$, where q is the scattering wave vector defined as $q = 4\pi \sin \theta / \lambda_{\text{CuK}\alpha}$, with a scattering angle of 2θ . The SAXS samples were measured in 2 mm quartz X-ray capillaries (purchased from Hilgenberg).

Analysis of experimental data (R_g , D_{max} , Porod Volumes) was done using the ATSAS Suite, in particular Primus and functions therein¹¹¹.

Ensemble modeling with EOM

The SANS curve measured for protonated PTBPI at 0% D₂O was used for ensemble modeling with EOM. For this, the first models of each NMR structure of the free PTBPI-domains RRM1, RRM2 and RRM34 (pdb codes: 1SJK, 1SJR and 2EVZ, respectively)^{10,35} were renumbered according to the PTBPI amino acid sequence and used as rigid bodies, RRM34 (amino acids 324–531) was set as “fixed”. Sequence ranges 56–141, 177–284 and 324–531 were defined as rigid bodies, all other as “disordered”.

Modeling with CYANA

For the integrative structure modeling of PTBPI/EMCV-IRES-DtoF (Supplementary Fig. 3), we used the previously determined CLIR-MS/MS derived models of PTBPI RRM1, RRM3 and RRM4, as well as, the NMR-derived solution structure of RRM2 in complex with the respective RNA sites as rigid bodies³⁹. However, as CYANA does not allow to load protein-RNA complexes as input structures, protein and RNA were loaded separately. Based on the imino-¹H/¹H-NOESY signals and structure prediction by McFold/McSym, we modeled RNA stems using standard A-form RNA-helix angles and Watson-Crick base-pair restraints. We merged the protein models (only the first state of the bundle) and RNA models in one PDB file (numbering was adjusted to fit all models on a single chain as required for CYANA) and used the regularized macro of CYANA to generate a similar (but not identical) bundle of PTBPI input structures (RMSD < 0.3 Å). Subsequently, the definition of rigid bodies in CYANA for protein domains was achieved using the command “angle fix”. To preserve the RNA secondary structure, Watson-Crick base-pair restraints were enforced in the stem, while RNA loops and bulges were kept flexible. These regularized rigid bodies were used as input for a CYANA calculation. The linkers between these rigid domains were generated with random starting conformations by CYANA. In these calculations, intermolecular restraints that were adapted from previous work³⁹ were used to define protein-RNA contacts as observed in the individual sub-complexes.

To include EPR distance distributions in CYANA, we generated RRM models with 50 rotamers of the corresponding spin-labeled

cysteine and calculated the geometric center of the radical positions. This average position was represented by a dummy glycine-C α that was fixed by three or four strong restraints to the domain or linker-peptide, respectively (upper limits and lower limits in CYANA, ± 0.05 Å of geometric center). A final set of 35 EPR distance restraints were implemented as upper and lower limits corresponding to the mean plus or minus one standard deviation of the distance distribution, respectively. Conformers are not penalized for their deviation from the mean distance as long as the distance is between the lower and upper limits.

As CYANA does not include electrostatics, we used artificial lower limits of 6 Å between RNA-phosphate groups with corresponding upper limits set to 200 Å. The latter distance is larger than D_{max} determined by SAS experiments and, thus, does not artificially compact the RNA.

Using these input rigid bodies (protein domains and RNA fragments) in combination with the short and long-range intermolecular restraints, we executed a simulated annealing protocol with 30,000 steps and calculated 20,000 models (20 × 1000 with different seeds). Out of these 20,000 models, the 10% energy best structures (lowest target function) were selected for further refinement. This filtering does not further bias the raw ensemble toward the EPR data.

Structure refinement

All pseudoatoms from the CYANA output were removed, it was split into two chains (protein and RNA) and the RNA was renumbered by a customized script. Afterward, models with flexible peptide linkers that are threaded through RNA or contain C4'-O4' bond lengths in RNA sugars longer than 1.6 Å were discarded. These tests discard between 2 and 8 models per run (out of 100 models). The models, which passed both tests, were refined by YASARA version 21.8.27 calling YASARA Structure with the standard script used by the YASARA Minimization Server and included in the MMMx distribution on GitHub as minimization_server.mcr. The resulting structure remains close to the input structure while removing clashes, and fixing wrong bond lengths, wrong bond angles, and wrong torsion angles.

The individual output models of YASARA have different protonation states of RNA, however, models in a PDB file must have the same number of atoms. Therefore, all protons that did not exist in all structures of the first run were removed to obtain consistent models with minimal protonation. Models with fewer protons than the first model, which usually indicates problems in RNA geometry, were discarded. Once again, models were checked for flexible peptide linkers threaded through RNA. This is necessary because some CYANA models were very strained, and their structures changed so strongly during refinement that such threading occurred even if it did not exist in the input model of the YASARA refinement. In the case of threading after YASARA refinement, models were discarded. Similarly, sugar bonds were checked again. When trying to remove strain, YASARA sometimes breaks a C4'-O4' bond, because this is less costly in terms of energy than the alternatives. All models with altered sugar bonds were discarded. The set of all YASARA-refined CYANA models that passed these tests contained 766 conformers and furnished the raw ensemble for ensemble reweighting.

Ensemble reweighting

Ensemble reweighting was performed with MMMx commit b330052. In total, 35 EPR (distance distribution) restraints, the SAXS curve and two SANS curves (66% D₂O, 1.2 and 4 m detector distance) were used for refinement. An additional ensemble was refined only with the EPR restraints. In these refinements, the whole distance distribution, including its shape, is fitted, rather than only the mean and standard deviation.

Fitting was done in blocks of 100 conformers from the raw ensemble of 766 conformers with an algorithm searching for the global minimum by adjusting their populations⁶⁴. In each such fit, all

conformers with less than 1% of the population of the most populated conformer were discarded. Many populations were driven to exactly zero. Thus, after fitting one block of 100 conformers, a preliminary ensemble of much less than 100 conformers, together with corresponding populations, was obtained. This preliminary ensemble was topped up to obtain a new block of 100 conformers, by adding conformers from the raw ensemble that were not part of the first block of 100 conformers. The new block was fitted to obtain another preliminary ensemble with less than 100 conformers. This step was repeated multiple times, using up all conformers of the raw ensemble, with the last block usually containing less than 100 models.

In each block, the following fitting steps were performed. First, the block was fit only to the EPR restraints. This provides the best overlap deficiency⁶⁴ with the conformers of this block by assigning populations (e.g., OD1). Overlap for a single distance distribution restraint corresponds to the common area below the experimental and simulated distance distributions, whereas the total area of each distribution is normalized to 1. It ranges between 0 for completely distinct and 1 for identical distributions. Overlap deficiency for the set of all distance distribution restraints ranges between 0 and 1, with 1 meaning no overlap for at least one restraint and 0 meaning perfect overlap for all restraints. Second, the block was fit to only the three small-angle scattering (SAS) curves by minimizing the sum of the χ^2 values for all SAS curves. This provides the best sum of χ^2 values that can be achieved for the conformers in this block by assigning populations (e.g., SCS1). Finally, the block was fit to EPR and SAS restraints simultaneously. In this fitting step, conformers of the block are selected, and populations are determined. Fit quality of EPR and SAS restraints are balanced. For each trial set of populations, the mean overlap deficiency (OD2) and the sum of χ^2 values for the SAS curves (SCS2) are computed. Because this cannot be better than OD1 in this block, $OD2 \geq OD1$. Likewise, it cannot be better than SCS1, i.e., $SCS2 \geq SCS1$. Hence, $(OD2/OD1 + SCS2/SCS1)/2 \geq 1$. The minimum of 1 can only occur if the same set of populations provides the best fit for EPR restraints as well as the best fit for SAS restraints. If the restraints are inconsistent or the set of input conformers is not good enough, the value will be substantially larger than 1. In order to get a transparent measure, we defined the loss of merit, $(OD2/OD1 + SCS2/SCS1)/2 - 1$. The loss of merit is zero, if both restraint sets are best fitted by the same set of conformers and their populations. If it is larger than one, the two figures of merit (overlap deficiency and sum of χ^2 values) have more than doubled. If this happens, the subsets of restraints are inconsistent or the set of conformers is very poor. Values for the loss of merit of 0.2 to 0.4 are quite normal¹¹². The loss of merit for the PTBPI/EMCV-IRESDtoF ensemble refined with both DEER and SAS data is 0.332 and thus indicates overall consistent restraint sets and supports the integrity of the CYANA raw ensemble.

In a second approach to ensemble reweighting, we determined all conformer populations simultaneously by non-negative linear least squares (NNLSQ). In this case, we fitted to small-angle scattering curves disregarding the dependence of noise on the scattering vector and to background-corrected DEER traces, which are output by automated comparative DEER analysis. The approach allows for an additional polynomial background correction, which we performed up to order 2. This approach was newly implemented into MMMx. It turned out that only a small fraction of the populations (48 out of 766) deviated significantly from zero (Supplementary Fig. 13). The NNLSQ approach balances the loss of fit quality between individual data sets, whereas the non-linear approach balances it between groups of similar restraints (distance distributions and small-angle scattering curves). Since we have 35 distance distributions, but only 3 scattering curves, NNLSQ favors overlap with the distance distributions at the expense of increased χ^2 of the small-angle scattering curve fits. In this work, we prefer the non-linear approach that balances between groups of

restraints for the ensembles specified as the main ensemble. MMMx is freely available at <https://github.com/gieschke/MMMx>.

Ensemble comparison and assessment of ensemble quality

Radii of gyration for all ensembles (Supplementary Table 9) and overlap of pseudo-electron density for all pairs of ensembles were computed with the EnsembleAnalysis module of MMMx commit 330052. MolProbity was downloaded from GitHub (<https://github.com/rlabduke/MolProbity>, accessed Dec 21, 2022). The command line tool online analysis was used to provide quality measures for the individual conformers. Quality measures for the ensemble, as reported in Supplementary Table 10, were obtained by population-weighted averaging over the ensemble with the newly introduced MMMx function `rd_MolProbity_results.m`. MMMx is freely available at <https://github.com/gieschke/MMMx>.

Reporting summary

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

Data availability

As the integrative structure modeling approach is not listed as an accepted experimental method for structure generation by the wwPDB, the structural ensembles have been deposited in Zenodo (<https://doi.org/10.5281/zenodo.7798970>). Further data supporting the findings of this study are available in the Supplementary Information. All the DEER, NMR, and SAS data are publicly available through the Zenodo repository (<https://doi.org/10.5281/zenodo.8379933>)¹¹³. Additionally, the MMMx script for ensemble fitting was uploaded to Zenodo under the same DOI. The reanalyzed mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier [PXD034894](https://doi.org/10.5281/zenodo.8380079). PDB codes of previously published structures used in this study are [1SJQ](https://doi.org/10.5281/zenodo.8380079), [1SJR](https://doi.org/10.5281/zenodo.8380079) and [2EVZ](https://doi.org/10.5281/zenodo.8380079). Other source data are provided with this paper as Source Data file. Source data are provided with this paper.

Code availability

The MMMx suite used for structure refinement, ensemble reweighting, and ensemble comparison is available through the Zenodo at <https://doi.org/10.5281/zenodo.8380079>.

References

1. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
2. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
3. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
4. Kedersha, N., Ivanov, P. & Anderson, P. Stress granules and cell signaling: more than just a passing phase? *Trends Biochem. Sci.* **38**, 494–506 (2013).
5. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillogenesis. *Cell* **163**, 123–133 (2015).
6. Murakami, T. et al. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678–690 (2015).
7. Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).

8. Oberstrass, F. C. et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–2057 (2005).
9. Joshi, A. et al. Crystallographic analysis of polypyrimidine tract-binding protein-raver1 interactions involved in regulation of alternative splicing. *Structure* **19**, 1816–1825 (2011).
10. Simpson, P. J. et al. Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure* **12**, 1631–1643 (2004).
11. Cieniková, Z., Jayne, S., Damberger, F. F., Allain, F. H. T. & Maris, C. Evidence for cooperative tandem binding of hnRNP C RRMs in mRNA processing. *RNA* **21**, 1931–1942 (2015).
12. Phelan, M. M. et al. The structure and selectivity of the SR protein SRSF2 RRM domain with RNA. *Nucleic Acids Res.* **40**, 3232–3244 (2012).
13. Ding, J. et al. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115 (1999).
14. Maslah, G. et al. Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains. *EMBO J.* **37**, e97089 (2018).
15. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
16. Ma, H., Jia, X., Zhang, K. & Su, Z. Cryo-EM advances in RNA structure determination. *Signal Transduct. Target. Ther.* **7**, 58 (2022).
17. Daubner, G. M., Cléry, A. & Allain, F. H.-T. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.* **23**, 100–108 (2013).
18. Schlundt, A., Tants, J. N. & Sattler, M. Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods* **118–119**, 119–136 (2017).
19. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 1–30 (2020).
20. Jobe, A., Liu, Z., Gutierrez-Vargas, C. & Frank, J. New insights into ribosome structure and function. *Cold Spring Harb. Perspect. Biol.* **11**, a032615 (2019).
21. Corley, M., Burns, M. C. & Yeo, G. W. How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **78**, 9–29 (2020).
22. Duss, O. et al. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* **509**, 588–592 (2014).
23. Gromak, N., Matlin, A. J., Cooper, T. A. & Smith, C. W. J. Antagonistic regulation of α -actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA* **9**, 443–456 (2003).
24. de Vries, T. et al. Sequence-specific RNA recognition by an RGG motif connects U1 and U2 snRNP for spliceosome assembly. *Proc. Natl Acad. Sci. USA* **119**, e2114092119 (2022).
25. Romanelli, M. G., Diani, E. & Lievens, P. M. J. New insights into functional roles of the polypyrimidine tract-binding protein. *Int. J. Mol. Sci.* **14**, 22906–22932 (2013).
26. Llorian, M. et al. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* **17**, 1114–1123 (2010).
27. Izquierdo, J. M. et al. Regulation of fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell* **19**, 475–484 (2005).
28. Sawicka, K., Bushell, M., Spriggs, K. A. & Willis, A. E. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem. Soc. Trans.* **36**, 641–647 (2008).
29. Spriggs, K. A., Stoneley, M., Bushell, M. & Willis, A. E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol. Cell* **100**, 27–38 (2008).
30. Bushell, M. et al. Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. *Mol. Cell* **23**, 401–412 (2006).
31. King, H. A. et al. Remodelling of a polypyrimidine tract-binding protein complex during apoptosis activates cellular IRESs. *Cell Death Differ.* **21**, 161–171 (2014).
32. Witherell, G. W., Gil, A. & Wimmer, E. Interaction of polypyrimidine tract binding protein with the encephalomyocarditis virus mRNA internal ribosomal entry site. *Biochemistry* **32**, 8268–8275 (1993).
33. Kaminski, A., Hunt, S. L., Patton, J. G. & Jackson, R. J. Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *RNA* **1**, 924–938 (1995).
34. Belsham, G. J. & Sonenberg, N. Picornavirus RNA translation: roles for cellular proteins. *Trends Microbiol.* **8**, 330–335 (2000).
35. Vitali, F. et al. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **25**, 150–162 (2006).
36. Auweter, S. D. & Allain, F. H. T. Structure-function relationships of the polypyrimidine tract binding protein. *Cell. Mol. Life Sci.* **65**, 516–527 (2008).
37. Conte, M. R. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132–3141 (2000).
38. Maris, C. et al. A transient α -helix in the N-terminal RNA recognition motif of polypyrimidine tract binding protein senses RNA secondary structure. *Nucleic Acids Res.* **48**, 4521–4537 (2020).
39. Dorn, G. et al. Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat. Methods* **14**, 487–490 (2017).
40. Wang, Z. N. et al. High expression of PTBP1 promote invasion of colorectal cancer by alternative splicing of cortactin. *Oncotarget* **8**, 36185–36202 (2017).
41. He, X. et al. Involvement of polypyrimidine tract-binding protein (PTBP1) in maintaining breast cancer cell growth and malignant properties. *Oncogenesis* **3**, e84 (2014).
42. Qian, H. et al. Reversing a model of Parkinson’s disease with in situ converted nigral neurons. *Nature* **582**, 550–556 (2020).
43. Kafasla, P. et al. Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation. *Mol. Cell* **34**, 556–568 (2009).
44. Kafasla, P., Lin, H., Curry, S. & Jackson, R. J. Activation of picornaviral IRESs by PTB shows differential dependence on each PTB RNA-binding domain. *RNA* **17**, 1120–1131 (2011).
45. Gmeiner, C., Dorn, G., Allain, F. H. T., Jeschke, G. & Yulikov, M. Spin labelling for integrative structure modelling: a case study of the polypyrimidine-tract binding protein 1 domains in complexes with short RNAs. *Phys. Chem. Chem. Phys.* **19**, 28360–28380 (2017).
46. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
47. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *J. Chem. Phys.* **125**, 194905 (2006).
48. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.* **142**, 331–340 (2000).
49. Petoukhov, M. V. et al. Conformation of polypyrimidine tract binding protein in solution. *Structure* **14**, 1021–1027 (2006).
50. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217 (2015).
51. Bernadó, P. et al. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl Acad. Sci. USA* **102**, 17002–17007 (2005).
52. Cho, M.-K. et al. Amino acid bulkiness defines the local conformations and dynamics of natively unfolded α -synuclein and tau. *J. Am. Chem. Soc.* **129**, 3032–3033 (2007).

53. Alber, F. et al. Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
54. Wells, M. et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl Acad. Sci. USA* **15**, 5762–5767 (2008).
55. Salmon, L. et al. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 8407–8418 (2010).
56. Ozenne, V. et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470 (2012).
57. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
58. Bonomi, M., Heller, G. T., Camilloni, C. & Vendruscolo, M. Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **42**, 106–116 (2017).
59. Boura, E. et al. Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc. Natl Acad. Sci. USA* **108**, 9437–9442 (2011).
60. Duss, O., Yulikov, M., Jeschke, G. & Allain, F. H. T. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nat. Commun.* **5**, 3669 (2014).
61. Duss, O., Yulikov, M., Allain, F. H. T. & Jeschke, G. Combining NMR and EPR to determine structures of large RNAs and protein-RNA complexes in solution. *Methods Enzymol.* **558**, 279–331 (2015).
62. Jeschke, G. Ensemble models of proteins and protein domains based on distance distribution restraints. *Proteins* **84**, 544–560 (2016).
63. del Alamo, D., Jagessar, K. L., Meiler, J. & Mchaourab, H. S. Methodology for rigorous modeling of protein conformational changes by Rosetta using DEER distance restraints. *PLoS Comput. Biol.* **17**, e1009107 (2021).
64. Jeschke, G. & Esteban-Hofer, L. Integrative ensemble modeling of proteins and their complexes with distance distribution restraints. *Methods Enzymol.* **666**, 145–169 (2022).
65. Jeschke, G. Integration of nanometer-range label-to-label distances and their distributions into modelling approaches. *Biomolecules* **12**, 1369 (2022).
66. Hummer, G. & Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **143**, 243150 (2015).
67. Bonomi, M., Camilloni, C., Cavalli, A. & Vendruscolo, M. Metainference: a Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2**, e1501177 (2023).
68. Orioli, S., Larsen, A. H., Bottaro, S. & Lindorff-Larsen, K. How to learn from inconsistencies: integrating molecular simulations with experimental data. *Progress in Molecular Biology and Translational Science*, Vol. 170 (eds Strodel, B. & Barz, B.) Ch. 3, 123–176 (Academic Press, 2020).
69. Güntert, P. Automated NMR structure calculation with CYANA. *Protein NMR Techniques* (ed. Downing, A. K.) 353–378 (Humana Press, 2004).
70. Lamichhane, R. et al. RNA looping by PTB: evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc. Natl Acad. Sci. USA* **107**, 4105–4110 (2010).
71. Quade, N., Boehringer, D., Leibundgut, M., Van Den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat. Commun.* **6**, 1–9 (2015).
72. Colussi, T. M. et al. Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature* **519**, 110–113 (2015).
73. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. W. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831 (2014).
74. Leppek, K. et al. Gene- and species-specific Hox mRNA translation by ribosome expansion segments. *Mol. Cell* **80**, 980–995.e13 (2020).
75. Venkatesan, A., Sharma, R. & Dasgupta, A. Cell cycle regulation of hepatitis C and encephalomyocarditis virus internal ribosome entry site-mediated translation in human embryonic kidney 293 cells. *Virus Res.* **94**, 85–95 (2003).
76. van den Akker, G. G. H. et al. Current practice in bicistronic ires reporter use: a systematic review. *Int. J. Mol. Sci.* **22**, 5193 (2021).
77. Chamond, N., Deforges, J., Ulryck, N. & Sargueil, B. 40S recruitment in the absence of eIF4G/4A by EMCV IRES refines the model for translation initiation on the archetype of Type II IRESs. *Nucleic Acids Res.* **42**, 10373–10384 (2014).
78. Imai, S., Kumar, P., Hellen, C. U. T., D’Souza, V. M. & Wagner, G. An accurately preorganized IRES RNA structure enables eIF4G capture for initiation of viral translation. *Nat. Struct. Mol. Biol.* **23**, 859–864 (2016).
79. Yu, Y., Abaeva, I. S., Marintchev, A., Pestova, T. V. & Hellen, C. U. T. Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors. *Nucleic Acids Res.* **39**, 4851–4865 (2011).
80. Koch, A., Aguilera, L., Morisaki, T., Munsky, B. & Stasevich, T. J. Quantifying the dynamics of IRES and cap translation with single-molecule resolution in live cells. *Nat. Struct. Mol. Biol.* **27**, 1095–1104 (2020).
81. Taylor, K. & Sobczak, K. Intrinsic regulatory role of RNA structural arrangement in alternative splicing control. *Int. J. Mol. Sci.* **21**, 5161 (2020).
82. Liu, H., Zhang, W., Reed, R. B., Liu, W. & Grabowski, P. J. Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein. *RNA* **8**, 137–149 (2002).
83. Mickleburgh, I. et al. The organization of RNA contacts by PTB for regulation of FAS splicing. *Nucleic Acids Res.* **42**, 8605–8620 (2014).
84. Montemayor, E. J. et al. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat. Struct. Mol. Biol.* **21**, 544–551 (2014).
85. Xue, Y. et al. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* **36**, 996–1006 (2009).
86. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
87. Jeschke, G. MMM: a toolbox for integrative structure modeling. *Protein Sci.* **27**, 76–85 (2018).
88. Duss, O., Maris, C., Von Schroetter, C. & Allain, F. H. T. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res.* **38**, e188 (2010).
89. Stoneley, M. et al. Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. *Nucleic Acids Res.* **28**, 687–694 (2000).
90. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
91. Sarnowski, C. P., Götze, M. & Leitner, A. RNxQuest: an extension to the xQuest pipeline enabling analysis of protein–RNA cross-linking/mass spectrometry data. *J. Proteome Res.* **22**, 3368–3382 (2023).
92. Walzthoeni, T. et al. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9**, 901–903 (2012).

93. Rinner, O. et al. Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318 (2008).
94. Sarnowski, C. P. et al. Sensitive detection and structural characterisation of UV-induced cross-links in protein-RNA complexes using CLIR-MS. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.31.486537> (2022).
95. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
96. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
97. Aeschbacher, T. et al. Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41**, e172 (2013).
98. Varani, G., Aboul-Ela, F. & Allain, F. H. T. NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.* **29**, 51–127 (1996).
99. Novakovic, M. et al. Magnetization transfer to enhance NOE cross-peaks among labile protons: applications to imino-imino sequential walks in SARS-CoV-2-derived RNAs. *Angew. Chem. Int. Ed. Engl.* **133**, 11991–11998 (2021).
100. Novakovic, M. et al. The incorporation of labile protons into multidimensional NMR analyses: glycan structures revisited. *J. Am. Chem. Soc.* **143**, 8935–8948 (2021).
101. Novakovic, M., Cousin, S. F., Jaroszewicz, M. J., Rosenzweig, R. & Frydman, L. Looped-PROjected Spectroscopy (L-PROSY): a simple approach to enhance backbone/sidechain cross-peaks in ¹H NMR. *J. Magn. Reson.* **294**, 169–180 (2018).
102. Polyhach, Y. et al. High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Phys. Chem. Chem. Phys.* **14**, 10762–10773 (2012).
103. Jeschke, G. et al. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).
104. Jeschke, G. Interpretation of dipolar EPR data in terms of protein structure. *Structural Information from Spin-Labels and Intrinsic Paramagnetic Centres in the Biosciences* (eds Timmel, C. R. & Harmer, J. R.) 83–120 (Springer, 2013).
105. Keeley, J. et al. Neural networks in pulsed dipolar spectroscopy: a practical guide. *J. Magn. Reson.* **338**, 107186 (2022).
106. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.* **172**, 279–295 (2005).
107. Schiemann, O. et al. Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules. *J. Am. Chem. Soc.* **143**, 17875–17890 (2021).
108. Keiderling, U. The new 'BerSANS-PC' software for reduction and treatment of small angle neutron scattering data. *Appl. Phys. A Mater. Sci. Process* **74**, 1455–1457 (2002).
109. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
110. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
111. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40**, s223–s228 (2007).
112. Ritsch, I. et al. Phase separation of heterogeneous nuclear ribonucleoprotein A1 upon specific RNA-binding observed by magnetic resonance. *Angew. Chem. Int. Ed. Engl.* **61**, e202204311 (2022).
113. Dorn, G. et al. Integrative structure determination of PTBP1-viral IRES complex in solution. Zenodo <https://doi.org/10.5281/ZENODO.8379933> (2022).

Acknowledgements

This work was supported by the SNF projects 31003A-149921 and 31003A-170130 and the SNF-NCCR RNA and Disease (to F.H.-T.A.). C.G., M.Y. and G.J. thank ETH Zürich for the financial support (Grant No. 0-2066-14). G.J. and F.H.-T.A. also thank the Sinergia grant CRSII5-170976 for financial support. G.D. was supported by a PhD fellowship from the Boehringer Ingelheim Fond. T.d.V. and C.P.S. were funded by an ETH Research Grant (grant ETH-24 16 2). T.J.W. would like to thank the ThinkSwiss Scholarship for funding.

Author contributions

G.D., C.G., M.Y., G.J. and F.H.-T.A. designed the project. G.D., C.G., T.d.V., C.M. and T.J.W. prepared protein and RNA samples. G.D., M.N. and F.F.D. performed NMR data collection and analysis. C.G., M.Y., T.J.W. and G.J. performed EPR data collection and analysis. G.D., E.D., S.B., R.M. and J.K. performed SANS/SAXS data collection and analysis. T.d.V., C.P.S., R.A. and A.L. performed re-analysis of CLIR-MS/MS data. G.D., T.d.V., F.H.-T.A. and G.J. performed integrative structural modeling and analysis. E.F. performed bioinformatics work. T.d.V., F.H.-T.A., G.D., C.G. and G.J. wrote the manuscript with contributions and critical feedback from all the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42012-z>.

Correspondence and requests for materials should be addressed to Maxim Yulikov, Gunnar Jeschke or Frédéric H.-T. Allain.

Peer review information *Nature Communications* thanks David Cafiso and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023