

Published in final edited form as:

J Biol Chem. 2025 April 08; 301(5): 108491. doi:10.1016/j.jbc.2025.108491.

The ribonucleoprotein hnRNPA1 mediates binding to RNA and DNA telomeric G-quadruplexes through an RGG-rich region

Sangeetha Balasubramanian^{#1}, Irawati Roy^{#1}, Rajeswari Appadurai^{1,2}, Anand Srivastava^{1,*}

¹Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, India.

²Department of Biology, Indian Institute of Science Education and Research, Tirupati, Andhra Pradesh, India.

[#] These authors contributed equally to this work.

Abstract

hnRNPA1, a protein from the heterogeneous-nuclear ribonucleoprotein family, mediates cellular processes such as RNA metabolism and DNA telomere maintenance. Besides the folded RNA recognition motifs, hnRNPA1 has a ~135 amino-acids long low-complexity domain (LCD) consisting of an RGG-rich region and a prion-like domain (PrLD). Biochemical data suggest that the RGG-rich region modulates the recognition of G-quadruplexes (GQs) in the telomeric repeats. Here, we utilize an in-house developed replica exchange technique (REHT) to generate the heterogeneous conformational ensemble of hnRNPA1-RGG and explore its functional significance in telomere maintenance. Single chain statistics and abundance of structural motifs, as well as consistency with experimentally reported structural data suggest faithful recapitulation of local interactions. We also introduce a protocol to generate functionally significant IDP-nucleic acid complex structures that corroborate well with the experimental knowledge of their binding. We find that RGG-box preferentially binds to the grooves and loops of GQs providing specificity towards certain GQ structures with its sequence and secondary structures. Turn-like structures expose Phe and promote stacking with the G-tetrads, while Tyr and Asn residues form essential hydrogen bonds and electrostatic interactions. Several of these residues were also identified as important by the earlier reported HSQC chemical shift data. Our binding and simulation studies also reveal that a minor population of the RGG-box can perturb telomeric GQs structure, which likely expedites the unfolding activities of hnRNPA1-UP1 at the telomeric end.

Biomolecular interactions between proteins and nucleic acids, particularly higher-order structures of nucleic acids like G-quadruplexes (GQ), have a critical role in gene regulation, telomere maintenance, genome stability, and other cellular processes (1). GQs are non-

This is an open access article under the CC BY license <https://creativecommons.org/licenses/by/4.0/>

*For correspondence: Anand Srivastava, anand@iisc.ac.in.

Author contributions

S. B. and I. R. writing—original draft; S. B., I. R., R. A., and A. S. visualization; S. B., I. R., and A. S. validation; S. B. and I. R. software; S. B., I. R., and R. A. methodology; S. B., I. R., R. A., and A. S. investigation; S. B., I. R., R. A., and A. S. formal analysis; S. B., I. R., and A. S. data curation; S. B., I. R., R. A., and A. S. conceptualization; R. A. and A. S. writing—review & editing; R. A. and A. S. project administration; A. S. supervision, A. S. resources, A. S. funding acquisition.

Conflicts of interest

The authors declare that they have no conflicts of interest with the contents of this article.

canonical super-secondary structures formed by G-rich repeats, (*e.g.* d(GGG[T/U][T/U]A)_n in human telomeres) where guanines arrange in a square planar configuration to form G-quartets, stabilized by hydrogen bonds and cation coordination at the core. These quartets stack to form a stable three-layered GQ structure connected by loops. While the DNA GQs can exist in diverse topologies ranging from parallel or antiparallel strands of tetramolecular, bimolecular, and unimolecular forms based on the flanking sequence, cation type, and the interacting partners, the RNA GQs can adopt only a parallel structure coordinated by K⁺ ions (2).

Recognition of GQs by GQ-binding proteins tightly regulates downstream processes by stabilizing or destabilizing the GQ structures. For instance, the GQ binding proteins such as FUS, EWS, and Nucleolin are shown to stabilize the folded telomeric DNA and RNA GQs (3, 4). On the other hand, heterogenous nuclear ribonucleoprotein A1(hnRNPA1) and DEAD-box helicases unfold the GQs (5). The architecture of GQ binding proteins often consists of one or more structured domains such as RRM (RNA recognition motif) domains, OB-fold domains, and disordered RGG domains composed of Arg-Gly dipeptide or Arg-Gly-Gly tripeptide repeats. Besides the role of structured domains, the role of disordered RGG domains in GQ recognition has been widely recognized in recent times. GQ-binding proteins like helicases, hnRNP, FET family proteins, nucleolin, and FMRP among others, have one or more RRM domains and RGG-rich regions that function concertedly or individually to stabilize or destabilize the G-quadruplexes (5). Particularly, the RGG domains of FUS and EWS are shown to promote the folding of telomeric DNA and RNA GQs leading to telomere shortening and histone methylation at the telomeres (3, 4). In the case of nucleolin, both RRM and RGG domains are required to form a GQ in the c-myc promoter sequence and the RGG domain is important to stabilize this GQ (6). In the case of hnRNPA1, the RRM can function independently to unfold GQ (7–10). However, the presence of the RGG domain enhances the unfolding activity, which indicates the role of the RGG-box beyond that of a biomolecular recognition adaptor (11, 12).

The RGG-box of diverse GQ binding proteins is believed to exhibit distinct specificity based on its sequence composition. The RGG of translocated in liposarcoma (TLS) protein has been shown to bind both RNA and DNA GQ (13). However, mutating the Tyr residues in the RGG domain to Phe leads to specific binding with RNA GQ by recognizing the ribose 2'-OH exposed in the loops. Moreover, the Tyr-only RGG and Phe-only RGG constructs exhibit specific binding to telomeric DNA and RNA GQs, respectively (14). These studies also suggest that the specific recognition of GQs by RGG modulates histone methylation at the telomeres. With the objective to design the most consensus GQ-binding motif, Huang *et al.* (15) identified a 20-mer peptide (RGRGRGRGGSGGSGG RGRG) with seven RG/RGG repeats having a high affinity to parallel GQs. Mutational study of this peptide clearly high-lights the role of amino acid composition and their position for GQ binding affinity. The role of Arg for GQ binding has been investigated previously and its importance is attributed to the multivalent behavior (16). Apart from their ability to form hydrogen bonds and π -stacking interactions with the nucleobases, the positively charged guanidine group is involved in electrostatic interactions with the negatively charged phosphate backbone of nucleic acids. Both Arg and Phe were shown as crucial for GQ binding due to their ability to form electrostatic as well as pi-stacking interactions. Gly

residues in the RGG motif provide the necessary flexibility that allows adaptive binding to several GQ topologies. In general, GQ-binding proteins are rich in Glycine, Arginine, Aspartic acid, Asparagine, and Valine (17). Although some biochemical insights into RGG-GQ binding are available, the structural and mechanistic principles governing this interaction remain unclear due to its disordered nature. Recently, a similar motif, the F/YGG motif, has been identified as a nucleic-acid binding motif with an intrinsically disordered nature (18).

In this study, we aim to uncover high-resolution insights into the mechanism of RGG-mediated GQ binding and its impact on GQ structure, using the RGG domain of hnRNPA1 as an example. hnRNPA1 is a member of a complex and diverse group of proteins called hnRNPs that collectively play an important role in processing heterogeneous nuclear RNAs into mature mRNAs. This abundant nuclear protein modulates many RNA metabolism processes, including transcription (19–21), splicing (21–24), RNA stability, export (25–27), and translation (28–30). The role of hnRNPA1 is not just limited to mRNA bio-genesis but extends to other processes such as microRNA processing and telomere maintenance (31). hnRNPA1 promotes the capping of telomeres by direct binding with DNA-dependent protein kinase catalytic subunit, DNA-PKcs (32). Phosphorylation of hnRNP promotes Shelterin complex formation by recruiting POT1 (protection of telomeres 1) protein to telomeres which then associates with Shelterin complex-forming proteins (33). It also modulates the binding of telomerase regulatory proteins like RPA (replication protein A) (34). Apart from the proteins, hnRNP A1 can also bind to the single strand and G-quadruplex folds of both telomeric DNA and RNA (35). Telomerase activity requires access to single-stranded G-rich 3'-overhangs, however, these regions are known to form G-quadruplex structures. hnRNP A1 is shown to bind to the telomeric DNA GQs and unfold them for telomerase activity. Similarly, sequestration of hnRNP A1 by TERRA can lead to telomerase inhibition. The binding of hnRNP A1-TERRA has been shown to regulate TERRA-mediated inhibition of telomerase in a TERRA-concentration-dependent manner (11, 12, 36).

Structurally (Fig. 1>A), the N-terminus of hnRNPA1 contains two well-folded RNA recognition motifs (RRMs) connected by a small linker region. Jointly these two motifs constitute the unwinding protein 1 (UP1) (37), which mediates interactions with target RNAs. UP1 domain is shown to bind to and stabilize telomeric ssDNA and also unfold telomeric GQ (9, 31, 38–40). In addition, a flexible Glycine rich C-terminal region consisting of Arg-Gly-Gly tripeptide repeats (RGG), interspersed with aromatic (Phe, Tyr) residues to form RGG boxes (Fig. 1B), provide both protein and RNA binding capabilities to hnRNPA1 (37). Although the RGG domain of hnRNP A1 is called the “RGG-rich” region, this region contains five FG/FGG motifs and six Asparagines in the vicinity of Phe. Downstream from the RGG-box region, the C-terminal domain harbors a prion-like domain (PrLD) and a nuclear shuttling sequence called the M9 sequence (41). PrLD has been shown to induce protein-protein interaction, leading to stress granule assembly and pathological protein aggregation (41). The RGG and PrLD domains together constitute the disordered low-complexity domain (LCD) of hnRNPA1.

Several lines of evidence have demonstrated that both UP1 and RGG-box domains are involved in GQ unfolding activity at the telomeric end of DNA (11, 42). It has been shown previously that the GQ unfolding rate of the UP1 domain ($k_{\text{obs}} = 0.63 \pm 0.01$) is lower

than UP1+RGG ($k_{\text{obs}} = 1.93 \pm 0.01$) (11). RGG-box might act as an adaptor for the UP1 by bringing it closer to the GQ, which leads to an efficient binding of UP1 followed by the unfolding of the GQ structure (11). Finally, UP1 can form a high-affinity complex with the single-stranded DNA and stabilizes the extended form. It is also possible that the RGG domain might induce perturbations in the GQ structure leading to faster and more efficient binding of UP1 domain, which has a preference for unfolded DNA/RNA. Despite tremendous strides made in the understanding of hnRNPA1 interactions with GQ using biochemical studies, high-resolution structural and dynamic insights into hnRNPA1-GQ interaction and telomere maintenance are still missing. In our work, we utilize the integrative molecular modeling approach that combines atomic-resolution molecular simulations with available ensemble-averaged data from NMR to faithfully reconstruct the atomic-resolution ensemble of hnRNPA1-RGG and use the structural data to investigate the functional role of RGG in telomere maintenance.

In this work, we have focused on and explored the structural heterogeneity and associated biological activities of hnRNPA1 with the help of our advanced sampling (43) and large-scale biomolecular simulations of hnRNPA1-RGG and its RNA/DNA GQ complexes. The article is organized as follows. In the Results and Discussion section following this Introduction section, we report the data from our extensive docking and simulations of RGG-region interacting with multiple RNA and DNA GQ topologies. Our simulations provide molecular-level insights into the reported role of the RGG-region in telomere maintenance. We discuss in detail how the RGG-region binds differently to telomeric DNA and telomeric RNA GQs and we also demonstrate how a very small population of hnRNPA1-RGG conformers are capable of perturbing the GQs. After the Results section, we briefly summarize our observations and findings in the Conclusions section, which is followed by the Materials and Methods where we provide details of all the methods that we have used. Additional information is included in the Supporting Information (SI) file. All our data including the simulated trajectories and input files for our simulations, all analyses-related data and codes are publicly available <https://figshare.com/s/4eb0701bdde1af2c2711>.

Results and discussion

Atomic resolution structural insights into the RGG conformation ensemble: conformations consistent with NMR data show a prevalence of β -motifs

We obtained a heterogeneous ensemble of the RGG domain (197–249 amino acids) (Fig. 1A) to gain conformational insights by harnessing a Hamiltonian and temperature-scaled hybrid replica exchange technique called Replica Exchange with Hybrid Tempering (REHT) (43–45). Please refer to the Methods section for more details. We calculated the α - $\text{C}\beta$, $^{13}\text{C}\text{O}$, and $^{15}\text{N}^{\text{H}}$ secondary chemical shifts using the conformation ensemble generated from our REHT simulation. The residue-wise comparison between the calculated data from the simulation and the experimentally available chemical shift data showed a considerable match (46) as shown in Figure 1, C–E. Upon performing a basic secondary structure analysis across the time evolution of our REHT ensemble, we observed a significant propensity to form transient secondary structures, preferentially β -sheet, and extended β -structures at different regions of RGG, as shown in Fig. S1 in SI. In a separate study, we have recently

explored the abundance of transient and non-transient beta-sheet structures in the RGG domain of hnRNPA1 and low-complexity domains of other prion-like proteins (47). In general, we find that the overall percentage of β -structures in each frame is higher when compared to α -helices (Grey in Fig. 1*F*). To verify this, we also used the chemical shift data of the experiment and simulation ensemble to predict the secondary structure propensities per residue. The secondary structure prediction (SSP) (48) scores of both experimental and simulated data re-emphasize our secondary structure findings of the ensemble data by showing a higher β -sheet population compared to that of α -helix (Fig. 1*G*). Altogether, the secondary chemical shifts and secondary structure analysis suggest a heterogeneous ensemble with an overall inclination to form local contacts through transient secondary structure formation.

At this point, it may be pertinent to discuss the prevalent ambiguity regarding secondary structure prediction from chemical shift data using conformational ensembles generated from molecular simulations. First of all, at the experimental interpretation level itself, predicting the secondary structure of intrinsically disordered proteins (IDPs) from chemical shift data is often potentially ambiguous due to several key reasons. Some of these reasons are the lack of a well-defined secondary or tertiary structure under the physiological conditions of IDPs and IDRs, the chemical shift averaging over many conformations, the high dependence of the chemical shift of an IDP on its local environment, lack of suitable reference database and last but not the least, the weak secondary structure propensities. The weak and transient secondary structure propensities exhibited by IDPs often do not produce chemical shift deviations that are strong enough to be reliably detected and interpreted using conventional secondary structure prediction algorithms. Due to this specific reason, in our particular case of the RGG box, we have observed outcomes from SSP analysis that have to be interpreted in light of inherent ambiguity in making strong deductions from the experimental data itself. The use of $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and ^{13}CO gave a higher percentage of α -helix propensity, whereas the use of $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^1\text{H}_\alpha$ produced a higher β -sheet propensity. Due to the effect of the weak secondary structure propensity, the reliability of the prediction gives way to obvious ambiguity. Nevertheless, in this work, we followed the recommended set of raw chemical shift data according to the SSP prescription of IDP (which are $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^1\text{H}_\alpha$) and were able to qualitatively match the result with the outcome of secondary structure calculation from our simulation data set using DSSP.

Functional insights from RGG conformational ensemble: role of RGG box in telomere maintenance

In 2018, Ghosh and Singh compared the 15N-1H HSQC data of the UP1+RGG-box domain and the isolated RGG-box domain (11). A superposition of the NMR spectra of these two constructs showed that RGG occupies the same conformational space with or without the existence of UP1. This in turn sheds light on the modular nature of the UP1-RGG construct where the tethered RGG-box behaves as a structurally independent domain from the folded UP1 domain. This also opens up the possibility that the full-length LCD (RGG + PrLD) could be modular. Bioinformatics studies on sequence space have suggested distinct compositional boundaries within IDRs —revealing a possible non-random modular organization in the sequence space (49). Fortuitously, NMR studies from

completely different groups have reported the HSQC spectra of the RGG domain (11) and the full-length LCD with the RGG domain and PrLD domain together as one chain (46). In order to explore the possibility that the full-length LCD of the hnRNPA1 could be modular, we superposed these HSQC spectra and it is clearly seen that the chemical shifts of RGG overlap in the two systems (Fig. S2A in SI). The NMR data clearly portrays the modularity of hnRNPA1 and signifies that the RGG domain, in spite of being disordered, can function independent of the other regions of hnRNPA1 such as UP1 and PrLD. This structural and functional modularity motivated us to dig deeper into the functionality of the RGG domain, in particular, our focus was to understand the functional relevance of the RGG box domain in telomere maintenance. hnRNPA1 expresses a concentration-dependent binding with telomeric RNA and DNA; in particular, it localizes at the GQ structural motif. Studies have hypothesized that the RGG-box binds to GQ, and recruits UP1 which in turn destabilizes the GQ (11, 12). We explore this hypothesis of RGG-GQ interactions, but the heterogeneity of the RGG ensemble complicates the selection of a reference RGG structure for the study. We used our recently developed machine learning-based clustering technique, which employs t-distributed stochastic neighbor embedding (t-SNE), to group the huge heterogeneous ensemble into 50 homogeneous clusters (50) (see Fig. S2B in SI). Our tSNE-based clustering algorithm for IDPs combines t-distributed stochastic neighbor embedding projection and kMeans clustering technique to group the heterogeneous data into homogenous subgroups. Through our Silhouette score, we can see that each cluster is evidently highly homogeneous, whereas the conformations between clusters are heterogeneous. The algorithm and parameters used to perform tSNE clustering is discussed in the Methods section. In a recent study, we have modeled the interaction of the RRM-RGG domains of FUS with a stem-loop RNA by all-atom molecular dynamics simulations (51). This study provided useful insights on the interaction of RGG with RNA and tSNE was used to characterize the conformational ensemble. We selected a central conformation from each cluster to represent the different states sampled by the RGG in solution and further used them to study the functional interactions. Being an IDP, RGG might interact with other partners *via* several mechanisms like “folding upon binding” or fuzzy interactions (52–54). In computational studies, first-pass docking methods generally provide insight into the formation of “encounter complexes”. To identify these encounter complexes and the probable basis for GQ recognition by RGG, we performed docking studies of RGG with telomeric DNA as well as RNA quadruplexes using HADDOCK. Further, the docked structures were then subjected to molecular simulations to capture the dynamics of the complex. A flowchart depicting the protocol adopted to generate the initial RGG-GQ encounter complexes is shown in Fig. 2.

Encounter complexes of telomeric RNA GQ are “end stacking” and telomeric DNA GQ are “groove binding”

We docked each of the 50 conformations of RGGs with seven different unimolecular, bimolecular, and tetramolecular GQs of telomeric RNA (named TERRA6, TERRA12, and TERRA24, respectively) and DNA (named tel6, tel12, tel22h1, and tel22h2). The systems under consideration are listed in Table 1, and the HADDOCK scores for all complexes are plotted in Fig. S3 in SI. Preliminary visual inspection and analysis of three-dimensional occupancy maps (drawn on an isosurface with a 6 Å distance cutoff) reveal the most

preferred binding sites on the 7GQs by the 50 RGG conformations (Fig. S3 in SI). The maps also clearly distinguish the mode of binding of RGG with RNA and DNA GQs. The RGG dominantly encounters the solvent-exposed external G-quartets of TERRA RNA in an “end stacking” mode. In contrast, it binds mostly in the grooves and loops of DNA GQs.

Interestingly, the binding modes influence the relative binding strength of the complexes. The calculated HADDOCK scores for the complexes in end stacking mode, such as those involving TERRA6, TERRA12, and TERRA24, are relatively lower in the range of –85 to –65. Whereas in groove binding mode, as in tel6, tel12, and tel22 complexes, the binding is slightly better (in the range of –100 to –75), with tel12 showing the strongest binding owing to the mixed mode of interactions. Another significant observation is that the HADDOCK scores for complexes involving tetra-molecular GQs (TERRA6, tel6) without propeller loops are weaker than those for GQs with propeller loops (such as TERRA 12, TERRA24, tel12, and tel22). These results corroborate well with experimental findings that the RGG primarily interacts with the GQ loops (11, 12). Our atomic resolution complex structures further provide deeper insights into the residue-wise contributions to the inter-molecular binding. In Tables S1 and S2, we provide a detailed list of inter-molecular interactions for the three strongest RGG-GQ complexes, screened by their HADDOCK scores. Intriguingly, in all these complexes, we observe over-whelming contributions of Arg and Gly from the RGG repeat motifs to Hydrogen bonding and electrostatic interactions, emphasizing the importance of RGG repeats in nucleic acid recognition. Also, the interspersing aromatic residues such as Tyr237, Tyr244, and several Phe residues are mediating π -stacking interactions. We also note a larger contribution of Ser residues in the DNA complexes.

Both DNA and RNA GQ with loops and overhangs have stronger binding consistent with recent biochemical observations

All the encounter complexes obtained from HADDOCK were simulated for a period of 50 ns each (a total of 17.5 μ s of all-atom simulations) to monitor the variations in RGG binding orientation and its induced changes in the GQ structural stability. These simulation trajectories clearly revealed the dynamic changes in the binding modes of RGG with the GQs and the type of stabilizing interactions. Figure 3 depicts the inter-molecular interactions between the amino acids in RGG and each base of GQ. These interactions are normalized with the individual amino acid composition and indicate the participation of each amino acid type toward the stability of the complex. When an amino acid in RGG resides within 4 Å of the central Guanine O6 atom, it is considered to explore the end-stacking mode. Groove binding mode is identified when an amino acid of RGG forms contact with any of the atoms exposed in the grooves, like “N2”, “C2”, “N3”, “C4”, “N6”, “C8” or “N9” atoms of Guanine. Similarly, the loop binding mode of RGG is described by the interaction with TTA, irrespective of its backbone, sugar or base. The three binding modes are not exclusive, and a given conformation of RGG might exhibit all three binding modes simultaneously by different regions of RGG. The encounter complexes from docking studies clearly show the end stacking mode of TERRA12, TERRA24 and tel12, and groove or loop binding modes in other GQ complexes (Fig. S4 in SI shows occupancy maps before simulation). However, after simulations, the binding modes vary, and groove or loop binding modes are preferred in all GQ complexes (Fig. 4).

Simulated ensemble of RGG-GQ complexes show interactions consistent with NMR experiments

In the previous section, we delineated the RGG binding sites in GQ and the preference of each amino acid type to interact with specific binding sites. Following this, we explore the lifetime of all RGG-GQ interactions by generating contact matrices (Figs. 5 and 6). The lifetime of RGG-GQ interactions in TERRA complexes is higher than in DNA GQ complexes. Preliminary inspection of these contact maps clearly shows that the N- and C-termini of RGG do not exhibit strong interactions. Previous NMR studies have highlighted the role of Tyr244 in binding GQs. Other residues of RGG like Ser197, Ser231, Arg232, Gly233, Tyr244, Asn245, and Gly246 were also reported to show chemical shift perturbations upon binding with GQ (11, 12, 46). Our contact analysis shows that these residues interact strongly with GQs, along with several other residues, particularly in the core of RGG (between Phe216 to Arg232). Altogether, by using our protocol, we were able to generate protein-nucleic acid complex structures that match well with experimental data and also provide additional atomic-level insights into the structural features.

RGGs can perturb both RNA and DNA GQs

Among the 350 simulated complexes, visual inspection shows that few of the GQ complexes were unstable. Hence, before analyzing these unstable complexes, we ensured the stability of the GQs by simulating their unbound conformations. The human telomeric repeat sequence has three consecutive guanines ($d(TTAGGG)_n$), and consequently, each GQ topology chosen in our study has three G-quartets. Each G-quartet is stabilized by 2 Hoogsteen H-bonds between adjacent guanines and a total of 8 H-bonds per quartet. Collectively, the three quartet GQs are stabilized by a minimum of 24 H-bonds. Our apo-state simulations show that all seven GQ structures were stable for the entire simulation period of 300 ns as seen by the time evolution of their G-quartet hydrogen bonds (Fig. S5 in SI). Hence, any changes in the stability of GQs in the RGG bound complexes are considered as the effect of RGG binding. During simulation, the RGG remains dynamic and experiences changes in its conformation as well as its binding with GQ. Considering the stability of apo GQ structures and the dynamic behavior of RGG, if a 50 ns simulation time induces any local disruptions in the GQ structure, it can be interpreted as a direct effect of RGG binding. The average number of hydrogen bonds between RGG and GQ as well as within the G-quartets were calculated in all the simulated complexes and a scatter plot correlating these is shown in Figure 7. A majority of the GQs possess 22 ± 2 hydrogen bonds and hence, they are stable. Loss of even two hydrogen bonds in each of the three G-quartets would elicit destabilization in the overall GQ structure. Therefore, we used an arbitrary cut-off of 15 hydrogen bonds or less (of the total 24) within the G-quartets to consider it as unstable. Figure 7 shows that at least one among each of the three types of TERRA complexes is unstable. Similarly, in the case of the telomeric DNA (Fig. 7), three among the 50 tel6 and one among the 50 tel22h2 complexes are unstable. Snapshots of these complex structures after the 50 ns molecular dynamics (MD) simulation are also shown in Figure 7.

The binding mode of RGG defines the mechanism of RGG-induced perturbation of GQs

About 2% (8 out of the 350 complexes) of the simulated RGG-GQ complexes show perturbation in the GQ structure upon binding with RGG. Interesting mechanistic insights into RGG-induced GQ destabilization were observed by analyzing the trajectories of these unstable complexes (a detailed discussion is included in Annexure 1 in SI). Instabilities in GQ can be induced by two major mechanisms like perturbing the core-bound ions or the grooves. One of the TERRA12 complexes which is unstable with only 15 G-quartet H-bonds, binds RGG in the end stacking mode (labeled RIV in Fig. 7). By visualizing this trajectory, we clearly observed the expulsion of K^+ ion by RGG (Fig. 8A). A π -stacking interaction by Phe222 stabilizes the G12, while the Gly219 is positioned to intrude into the core of GQ, leading to the repositioning and expulsion of a K^+ ion in the core. We have shown in the previous sections that the RGG domain has a propensity to form beta-structures and turn-like structures. Our simulations show that these turn-like structures can stack over the G-quartets *via* π -stacking interaction between Phe and Guanine bases and cause perturbations in the core ions. To validate this RGG structure-induced GQ perturbation hypothesis, we trimmed the RGG to retain only the 205 to 219 region in one of the TERRA24 complexes and simulated this truncated RGG-GQ complex. During the simulation, the peptide continues to interact with the GQ *via* Phe-Guanine stacking and induces the displacement of K^+ ions from the GQ core. Interestingly, the Phe residues responsible for π -stacking interaction are a part of the FGG motif of hnRNPA1.

The tetramolecular structures are particularly easy to unfold by interactions at the grooves. In this mode of binding, the side chains of RGG amino acids are oriented to interact with the backbone and nucleobases of GQs, straining the GQ to unfold. In one of the tel6 complexes (labeled DIII in Fig. 7), the sidechain of Asn249, interacts directly with an oxygen atom in the deoxyribose sugar of G5 and N3 atom in the Watson-crick face of G4 (Fig. 8B). These interactions are due to the protrusion of the Asn sidechain into the groove between two parallel strands of DNA, thereby disintegrating the tetramolecular GQ through the grooves. The ability of Asn to form bifurcated H-bonds is well-known and this is the major reason for Asn-induced GQ perturbation. We mutated this Asn into Ala (to prevent H-bond formation) or Arg (to mimic bifurcated H-bonds) and simulated the two complexes to reinforce the role of Asn in perturbing the GQ structure. In the case of Ala mutation, the tel6-GQ remained stable, whereas in the case of Arg mutation, the tel6-GQ was perturbed indicating that the interaction of these amino acids at the grooves of GQ might cause perturbations. It is important to note that these Asn residues are in the vicinity of the Phe residues forming the FGG motif. The stability of complexation due to the stacking of Phe over the G-quartets might allow the Asn to establish multivalent H-bonds in the GQ grooves, leading to GQ perturbation. Altogether, our observations indicate that the orientation of amino acid sidechains and their ability to form β -structured motifs induces destabilization in the GQ structure irrespective of their binding modes, strengthening the possibility of an RGG-induced mechanism for destabilizing GQs.

It has been shown previously that the GQ unfolding rate of the UP1 domain ($k_{\text{obs}} = 0.63 \pm 0.01$) is lower than UP1+RGG ($k_{\text{obs}} = 1.93 \pm 0.01$) (11). Our study supports this observation and provides a structural basis for the earlier biochemical observation. Since we show that

the individual RGG domain is capable of causing perturbations in the GQ topology, it could be a possible mechanism of catalyzing the unfolding activity of UP1. The observation suggests that RGG not only acts as an adaptor for recognizing GQ but also could play a crucial role in expediting the GQ unfolding.

Conclusion

IDRs and IDPs are shown to exhibit a variety of functions by exploiting their inherent disorderliness. They form unique molecular signatures to specifically cater to a definite function. These molecular or structural signatures can be the formation of a particular secondary structure upon binding with a folded partner (55) or retain the disordered state to recognize the folded partner (56–59), and in few cases, express non-specific interactions with another disordered molecule (53). Such diversity in their structures is necessary to fulfil their diverse functions by recognizing different biomolecules. One such IDR is the RGG-rich domain, which has the ability to bind proteins and nucleic acids. The nucleic acid binding property of RGG-rich regions is well known, and the RGGs of several functionally important IDPs like hnRNPA1, FMRP, and FUS express either stabilizing or destabilizing effects on GQ motifs (16, 60). It is intriguing that the RGG domains of different proteins show differential binding with G-quadruplexes, and some RGG like that of TDP-43, are able to selectively bind to the parallel form of telomeric G-quadruplex. In regard to the sequence properties, Glycine residues are proposed to interact with the nucleic acid based on the activity of the RGGGGR peptide of FMRP (61). Also, binding studies on FMRP RGG-box with an RNA duplex-quadruplex junction revealed a well-formed β -turn in RGG is responsible for its binding to the RNA (62). A similar involvement of β -spiral structure in FUS is reported as necessary for binding a GQ motif (63). Though the importance of sequence and secondary structures like β -motifs to bind nucleic acids, particularly GQ motifs, is known, a high-resolution mechanistic detail is still lacking. Our study is an attempt to faithfully capture a high-resolution conformational landscape explored by the RGG domain of hnRNPA1 through advanced simulations and further explore its functional significance.

Our hybrid replica exchange method has been able to successfully generate the conformational states attained by RGG in a solution that also matches well with the chemical shifts of an NMR ensemble. Beyond matching the chemical shifts, our generated conformation ensemble showed secondary structures similar to their propensities in the experimental ensemble. The conformation ensemble generated in our study shows a high propensity for β -turns and transient but recurring β -sheets. Our docking study highlights that these turn-like structures can stack on the G-quartets and they can be accommodated well in the grooves of GQ. Although the RGG domain of hnRNP A1 is called the “RGG-rich” region, this region is richer in FG/FGG motifs. Our study shows an important contribution for these Phe and Asn to facilitate the localization of RGG to GQ (stacking of Phe over the G-quartets) and induce perturbation (by bifurcated H-bonds at the GQ grooves) in the GQ structure. The individual RGGs were able to perturb the GQs, although less frequently (~8 events out of the total 350 complexes generated). This perturbation of GQ by the RGG of hnRNP A1 should be attributed to a combination of events like the stacking of Phe due to

their exposure in a turn-like structure, and the consequent orientation of Asn sidechain at the GQ grooves.

The RGG of hnRNPA1 is reported to enhance the unfolding of intramolecular human telomeric GQ by UP1 during telomere maintenance. The primary recognition site of RGG is reported as the bases forming the propeller loops of unimolecular GQs. The enhanced GQ unfolding by UP1 in the presence of RGG can be attributed to two different possibilities. In the first case, the RGG might act as the initial recognition factor, binding stably with the GQs and recruits UP1. Along with UP1 recruitment, the RGG can also induce slight perturbations in the GQ structure, thereby facilitating an easier and efficient binding of UP1. Based on our extensive docking and simulations, we propose the possibility of RGG-induced GQ perturbation. The concentration of hnRNPA1 and GQs, along with physiological conditions, play a vital role in molecular recognition since the RGG conformations and GQ topologies are highly dependent on these factors. Altogether, in this study, we have presented a protocol for generating a high-resolution conformation ensemble of an IDP (the RGG domain of hnRNPA1), capturing the heterogeneity and exploring their functional significance.

Methods

Conformer generation of RGG

The initial unfolded structure of the RGG domain was obtained from Iterative Threading ASSEmbling Refinement (I-TASSER) which is a widely used server for automated protein structure prediction and structure-based functional annotation (64–66). Using the amino acid sequence as a starting point, I-TASSER constructs 3D structural models. The generated structure of the RGG domain is a collapsed coiled-coil structure. The usage of I-TASSER is limited to generating initial coordinates of the protein and further conformation sampling was carried out using molecular dynamics (MD) simulations.

For IDPs, computational methods can be used to generate a conformation ensemble to study their functions. Some of the commonly used, less computationally demanding methods are flexible-meccano (67), BUNCH (68), and IDP-ConformerGenerator (69), among others. These methods use backbone dihedrals of non-secondary structural elements compiled from the existing protein structures, folded or disordered. Based on prior knowledge, propensities for secondary structures or long-range interactions can also be included as criteria in these models. However, the most commonly used conformer generation methods are based on molecular dynamics or Monte Carlo simulations. In general, coarse-grained models (70, 71) are used to reduce the time complexity over all-atom models, however, these force fields have several limitations to sample IDPs. We have developed an advanced sampling technique called replica exchange with hybrid tempering (REHT) to generate faithful conformations of IDPs (43). This method has been used to generate all-atom models and has been tested extensively to generate conformation ensembles of several IDPs like Histatin five and α -synuclein. We have used this REHT method to generate the conformation ensemble of RGG in this study.

REHT simulation

Unlike the commonly used temperature-based replica exchange, our hybrid approach involves the exchange of both the temperature as well as the Hamiltonian across the replicas. The protein was solvated using the 3-site rigid TIP3P water model in a cubic box with a minimum distance of 1 nm from the surface of the protein. The systems were also neutralized to maintain a physiological concentration of NaCl (0.15 M). The Charmm36 m (72) forcefield was used to model the protein. The simulations were carried out using Gromacs-2016.5 patched with Plumed-2.4.1 (73–75). Initially, the modeled protein was energy minimized using the steepest descent algorithm for 50,000 steps to avoid any poor contacts. The energy-minimized structure was then thermalized and equilibrated sequentially in NVT and NPT ensembles for 1 ns.

Next, the protein and the solvent were coupled separately to the target temperatures using the Nose-Hoover thermostat, and the final production simulation was performed in the NVT ensemble. A cut-off of 1 nm was used for calculating the electrostatic and vdw interactions and Particle Mesh Ewald was used for long-range electrostatics. To integrate the equations of motions, the Leap-frog integrator with the time step of 2 fs was used. LINCS algorithm was used to constrain all the hydrogen atoms. Exchanges were attempted at every 1 ps interval. In this approach, the Hamiltonian of the solute particle in the different replicas is scaled down to 0.6 to 1 while also raising the temperature of the systems simultaneously up to a maximum of 500 K. This way, as a function of both lambda scaling and explicit thermostat conditions, a very high temperature is ensured to be realized on the protein solute to sufficiently overcome the energy barriers. The solvent is heated up mildly in order for the rapid reorientation of the hydration shell upon conformational change of the protein. 20 replicas were used to simulate the RGG in a temperature range of 300 K to 500 K for a cumulative period of 1 μ s and an acceptance probability of 0.3. Post-processing analyses of the trajectories were performed with the Gromacs analysis tools and the trajectories were visualized using VMD.

Chemical shift calculation

We have used SPARTA+(version 2.90) in conjunction with a software package called Mdtraj (76, 77). SPARTA + employs a well-trained neural network algorithm to construct quantitative relations between chemical shifts and protein structures, including backbone and side-chain conformation, H-bonding, electric fields, and ring-current effects. Its single-level feed-forward multilayer artificial neural network (ANN) is capable of identifying the dependence of $^{15}\text{N}^{\text{H}}$, ^{13}CO , $^{13}\text{C}_{\alpha}$, $^{13}\text{C}_{\beta}$, $^1\text{H}_{\alpha}$, $^1\text{H}^{\text{N}}$ chemical shifts on the local structural and dynamic information as well as amino acid type, and those of its immediate neighbors. Mdtraj is a Python software package that is designed to analyze trajectory data generated from MD simulations.

We have compared our predicted data with the experimentally calculated chemical shift submitted in the Biological Magnetic Resonance Data Bank (BMRB entry 50017) (46, 78). For comparison purposes, we have used secondary chemical shift data by removing random coil chemical shifts from the raw chemical shift. For C_{α} - C_{β} calculation, the secondary

chemical shifts of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ are subtracted to do away with the referencing error, if there is any.

Secondary structure propensity calculation

We have used the secondary structure prediction algorithm (SSP) developed by Prof. Julie D. Forman-Kay's group (48). We have used $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^1\text{H}_\alpha$ chemical shifts for this purpose as recommended by the algorithm developers in the context of intrinsically disordered proteins. For random coil and secondary structure chemical shifts and standard deviations, we have used RefDB. All the data have been re-referenced.

t-SNE clustering

We have chosen a nonlinear dimensionality reduction method called t-distributed Stochastic Neighbor Embedding (t-SNE) in order to separate high-dimensional heterogeneous IDP ensembles into meaningful clusters. An important aspect of the t-SNE algorithm is the perplexity value, a tuneable parameter in t-SNE that balances the information between the local and global features of the dataset under investigation. Diligent choice of perplexity value for a given data is important to be able to most discretely divide the data into unambiguous clusters. In our work, we have noticed that for very low perplexity values, the clustering is extremely diffusive in nature and for a high value of perplexity parameter, the entire dataset seems to be treated as a single cluster. The performance of t-SNE is fairly robust to changes in the perplexity. In this work, we have explored several perplexity value options ranging from 50 to 2000. We have chosen the values that most discretely divide the data points into separate clusters. To aid the result of t-SNE clustering in further partitioning, we have employed a widely used centroid-based technique called K-means clustering which is one of the favored and standard unsupervised machine learning algorithms. We have used the SciKit tool for these analyses which is an open source library for machine learning in Python (79).

Structures of RGG and GQ

The human telomeric DNA (htel) and RNA (TERRA) GQs, formed by the telomeric repeat d(GGG[T/U][T/U]A)_n, is made by the stacking of three G-quartets connected by TTA or UUA loops. The Guanine bases in a quartet are stabilized by hydrogen bonds between their Watson-Crick and Hoogsteen edges forming a square planar arrangement. The central-facing carbonyl oxygen atoms of guanines exert electrostatic repulsion, which is shielded by the coordination of cations at the core of GQs. Polymorphism is a common feature in DNA GQs, where different GQ topologies are observed for a given sequence under different physiological conditions like the available cation types, neighboring G-rich sequence, and their interacting partners (80, 81). The human telomeric DNA GQs exhibit structural polymorphism with a wide range of topologies varying from tetramolecular or bimolecular all-parallel quadruplex to unimolecular all-parallel, anti-parallel as well as hybrid forms. Also, the variation is evident with the presence of different cations like Na⁺ and K⁺. However, due to the presence of ribose sugar, TERRA GQs can adopt only a parallel GQ coordinated by K⁺ ions (2).

The 50 conformations of RGG obtained by clustering the REHT trajectories were docked with the human telomeric DNA and RNA GQ (also known as GQ). Unimolecular, bimolecular, and tetramolecular forms of both DNA and RNA quadruplexes were modeled in this study (Table 1). The tetramolecular TERRA (or TERRA6) was modeled using the solution structure with PDB ID: 6GE1 (82), while the unimolecular TERRA24 was modeled based on the telomeric DNA quadruplex with PDB ID: 1KF1 (83). The NMR structure of bimolecular telomeric RNA (or TERRA12) is available in the Protein Databank with the ID 2KBP (84). The structures of all forms of telomeric DNA quadruplexes were taken directly from the Protein Databank with IDs 1NP9 (tetramolecular or tel6) (85), 1K8P (bimolecular tel12) (83), 2HY9 (unimolecular form one tel22h1) (86) and 2JPZ (unimolecular form two tel22h2) (87). The unimolecular telomeric DNA in K_+ ion solution adopts two different hybrid conformations that are equally populated and both the structures (differentiated as tel22h1 and tel22h2) were taken for our docking studies. The numbers behind TERRA* or tel* indicate the length of each RNA or DNA strand forming the GQ. Additional bases were added to the 5' or 3' ends of the GQ structures to match the sequences used in ITC studies by Ghosh *et al.* (11, 12). Two K_+ ions were placed manually between the three G-quartets of all 7 GQs. The modeled structures were simulated for a period of 300 ns and the structures extracted after 10 ns simulation were used as starting structures for docking studies.

HADDOCK docking protocol

HADDOCK (88) utilizes a data-driven approach for molecular docking. The stand-alone version of HADDOCK 2.4 was used to drive the docking of RGG with telomeric RNA and DNA GQs. Since the RGG conformations are randomly coiled structures of ~53 amino acids, all residues are highly solvent-exposed. To perform an unbiased docking study, the distance between the center of mass of the protein and GQ alone was used as a restraint criterion. The docking process is performed in three steps: rigid-body docking, semi-flexible refinement, and water refinement. About 10,000 complexes were generated during the first stage of rigid-body docking by using the standard HADDOCK protocol. Among these, 400 lowest-energy structures were selected for subsequent semi-flexible simulated annealing and explicit solvent (water) refinement, to optimize the interface. HADDOCK scoring was performed according to the weighted sum (HADDOCK score) of different energy terms, including the van der Waals energy, electrostatic energy, distance restraints energy, inter-vector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, binding energy, desolvation energy, and buried surface area. The final structures were clustered using the fraction of common contacts (FCC) with a cutoff of 0.6 and a minimum cluster size of 4.

Simulation of RGG-GQ complexes

For interaction studies of RGG-GQ complexes (listed in Table 1), we performed molecular dynamics simulation using Gromacs-2020.4 (89). We used a99SBdisp forcefield of Paul Robustelli for RGG, OL15 force field for DNA quadruplexes, and OL3 force field for RNA quadruplexes (90–92). The RGG-GQ complex systems (obtained from HADDOCK) were solvated with TIP3P water in a periodic box extending up to 1 nm and 1.2 nm in all directions respectively. The systems were neutralized and additional ions were added to mimic a salt concentration of 0.15 M. The short-range interactions were truncated with

a cut-off distance of 1 nm. Electrostatic interactions were treated by particle-mesh Ewald with a real space cut-off value of 1 nm. Bonds containing hydrogens were constrained using the LINCS algorithm. The solvated and neutralized systems were energy minimized using the Steepest Descent algorithm followed by an equilibration of 5 ns and subsequently production runs. The temperature and pressure of the systems were maintained at 310 K and 1 atm using the nose-hoover thermostat and Parrinello-Rahman barostat in an NPT ensemble. Post-processing analyses were performed with the Gromacs analysis tools, and CPPTRAJ module of AmberTools20 and the trajectories were visualized using Visual Molecular Dynamics (93), UCSF Chimera v1.13. RGG-GQ interaction matrices were consolidated by calculating all heavy-atom contacts within 0.4 nm that are present for at least 1% of the total simulation period for each GQ (50 complexes * 50 ns = 2.5 μ s).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

A. S. thanks Prof Mahavir Singh at the Molecular Biophysics Unit for numerous discussions during the course of the work. IR would like to acknowledge financial support from the Ministry of Education, Government of India. R. A. would like to thank the early career fellowship from the DBT-Wellcome Trust India Alliance (Grant number: IA/E/18/1/504308). S.B. would like to thank the National Supercomputing Mission grant for financial support. AS acknowledges the financial support from the Indian Institute of Science-Bangalore and the high-performance computing (HPC) facility “Beagle” setup from grants by a partnership between the Department of Biotechnology of India and the Indian Institute of Science (IISc-DBT partnership programme). AS also thanks the DST for the National Supercomputing Mission grants (DST/NSM/R\&D-HPC-Applications/2021/03.10 and DST/NSM/R\&D-HPC-Applications/Extension Grant/2023/27) for the HPC support. FIST program sponsored by the Department of Science and Technology and UGC, Centre for Advanced Studies and Ministry of Human Resource Development, India. AS would also like to thank the Teams Science Grant from the DBT-Wellcome Trust India Alliance (Grant number: IA/TSG/21/1/600245) and the DBT National Network Project (NNP) grant (BT/PR40323/BTIS/137/78/2023 grants. This work was initiated through the Matrics grants (MTR/2023/001040) from the Science and Engineering Board (SERB), India and both authors thanks SERB for this support.

Abbreviations

FCC	fraction of common contacts
GQ	G-quadruplexes
IDPs	intrinsically disordered proteins
LCD	low-complexity domain
POT1	protection of telomeres 1
PrLD	prion-like domain
REHT	replica exchange with hybrid tempering
RGG	Arg-Gly-Gly tripeptide repeats
TLS	translocated in liposarcoma
UP1	unwinding protein 1.

Data availability

Input files needed to initiate molecular simulations and full trajectory data of all simulations considered in this work are available on our server for download. Server data can be publicly accessed *via* the Figshare repository: <https://figshare.com/s/4eb0701bdde1af2c2711c>.

References

1. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol.* 2020; 21: 459–474. DOI: 10.1038/s41580-020-0236-x [PubMed: 32313204]
2. Lyu K, Chow EY-C, Mou X, Chan T-F, Kwok CK. RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* 2021; 49: 5426–5450. DOI: 10.1093/nar/gkab187 [PubMed: 33772593]
3. Takahama K, Kino K, Arai S, Kurokawa R, Oyoshi T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA-and RNA-binding protein. *FEBS J.* 2011; 278: 988–998. [PubMed: 21244633]
4. Takahama K, Takada A, Tada S, Shimizu M, Sayama K, Kurokawa R, et al. Regulation of telomere length by G-quadruplex telomere DNA-and TERRA-binding protein TLS/FUS. *Chem Biol.* 2013; 20: 341–350. [PubMed: 23521792]
5. Shu H, Zhang R, Xiao K, Yang J, Sun X. G-Quadruplex-Binding proteins: promising targets for drug design. *Biomolecules.* 2022; 12: 648. doi: 10.3390/biom12050648 [PubMed: 35625576]
6. González V, Hurley LH. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry.* 2010; 49: 9706–9714. DOI: 10.1021/bi100509s [PubMed: 20932061]
7. Takimoto M, Tomonaga T, Matunis M, Avigan M, Krutzsch H, Dreyfuss G, et al. Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. *J Biol Chem.* 1993; 268: 18249–18258. [PubMed: 8349701]
8. Paramasivam M, Membrino A, Cogoi S, Fukuda H, Nakagama H, Xodo LE. Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.* 2009; 37: 2841–2853. DOI: 10.1093/nar/gkp138 [PubMed: 19282454]
9. Fukuda H, Katahira M, Tsuchiya N, Enokizono Y, Sugimura T, Nagao M, et al. Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc Natl Acad Sci.* 2002; 99: 12685–12690. [PubMed: 12235355]
10. Ling X, Yao Y, Ding L, Ma J. The mechanism of UP1 binding and unfolding of human telomeric DNA G-quadruplex. *Biochim Biophys Acta (BBA) - Gene Regul Mech.* 2023; 1866 194985 [PubMed: 37717939]
11. Ghosh M, Singh M. RGG-box in hnRNP A1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res.* 2018; 46: 10246–10261. DOI: 10.1093/nar/gky854 [PubMed: 30247678]
12. Ghosh M, Singh M. Structure specific recognition of telomeric repeats containing RNA by the RGG-box of hnRNP A1. *Nucleic Acids Res.* 2020; 48: 4492–4506. DOI: 10.1093/nar/gkaa134 [PubMed: 32128583]
13. Takahama K, Oyoshi T. Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. *J Am Chem Soc.* 2013; 135: 18016–18019. [PubMed: 24251952]
14. Takahama K, Miyawaki A, Shitara T, Mitsuya K, Morikawa M, Hagihara M, et al. G-quadruplex DNA- and RNA-specific-binding proteins engineered from the RGG domain of TLS/FUS. *ACS Chem Biol.* 2015; 10: 2564–2569. [PubMed: 26360301]
15. Huang Z-L, Dai J, Luo W-H, Wang X-G, Tan J-H, Chen S-B, et al. Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J Am Chem Soc.* 2018; 140: 17945–17955. [PubMed: 30517002]

16. Chong PA, Vernon RM, Forman-Kay JD. RGG/RG motif regions in RNA binding and phase separation. *J Mol Biol.* 2018; 430: 4650–4665. [PubMed: 29913160]
17. Brázda V, Červe J, Bartas M, Mikysková N, Coufal J, Pevínka P. The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules.* 2018; 23: 2341 doi: 10.3390/molecules23092341 [PubMed: 30216987]
18. Van Lindt J, Lazar T, Pakravan D, Demulder M, Meszaros A, Van Den Bosch L, et al. F/YGG-motif is an intrinsically disordered nucleic-acid binding motif. *RNA Biol.* 2022; 19: 622–635. DOI: 10.1080/15476286.2022.2066336 [PubMed: 35491929]
19. Li H-P, Zhang X, Duncan R, Comai L, Lai MMC. Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription-regulatory region of mouse hepatitis virus RNA. *Proc Natl Acad Sci.* 1997; 94: 9544–9549. DOI: 10.1073/pnas.94.18.9544 [PubMed: 9275159]
20. Lau JS, Baumeister P, Kim E, Roy B, Hsieh T-Y, Lai M, et al. Heterogeneous nuclear ribonucleoproteins as regulators of gene expression through interactions with the human thymidine kinase promoter. *J Cell Biochem.* 2000; 79: 395–406. [PubMed: 10972977]
21. Feng J, Zhou J, Lin Y, Huang W. hnRNP A1 in RNA metabolism regulation and as a potential therapeutic target. *Front Pharmacol.* 2022; 13: 986409 doi: 10.3389/fphar.2022.986409 [PubMed: 36339596]
22. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell.* 1992; 68: 365–375. [PubMed: 1531115]
23. Swanson MS, Dreyfuss G. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J.* 1988; 7: 3519–3529. DOI: 10.1002/j.1460-2075.1988.tb03228.x [PubMed: 3208740]
24. Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem.* 1993; 62: 289–321. [PubMed: 8352591]
25. Nakielný S, Dreyfuss G. Transport of proteins and RNAs in and out of the nucleus. *Cell.* 1999; 99: 677–690. [PubMed: 10619422]
26. Mili S, Shu HJ, Zhao Y, Piñol-Roma S. Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. *Mol Cell Biol.* 2001; 21: 7307–7319. DOI: 10.1128/MCB.21.21.7307-7319.2001 [PubMed: 11585913]
27. Jean-Philippe J, Paz S, Caputi M. hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci.* 2013; 14: 18999–19024. DOI: 10.3390/ijms140918999 [PubMed: 24065100]
28. Bonnal S, Pileur F, Orsini C, Parker F, Pujol F, Prats A-C, et al. Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. *J Biol Chem.* 2005; 280: 4144–4153. [PubMed: 15525641]
29. Jo OD, Martin J, Bernath A, Masri J, Lichtenstein A, Gera J. Heterogeneous nuclear ribonucleoprotein A1 regulates cyclin D1 and c-myc internal ribosome entry site function through Akt signaling. *J Biol Chem.* 2008; 283: 23274–23287. DOI: 10.1074/jbc.M801185200 [PubMed: 18562319]
30. Clarke JP, Thibault PA, Salapa HE, Levin MC. A comprehensive analysis of the role of hnRNP A1 function and dysfunction in the pathogenesis of neurodegenerative disease. *Front Mol Biosciences.* 2021; 8: 1–19. DOI: 10.3389/fmolb.2021.659610 [PubMed: 33912591]
31. Zhang QS, Manche L, Xu R-M, Krainer AR. hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA.* 2006; 12: 1116–1128. DOI: 10.1261/rna.58806 [PubMed: 16603717]
32. Sui J, Lin Y-F, Xu K, Lee K-J, Wang D, Chen BPC. DNA-PKcs phosphorylates hnRNP-A1 to facilitate the RPA-to-POT1 switch and telomere capping after replication. *Nucleic Acids Res.* 2015; 43: 5971–5983. DOI: 10.1093/nar/gkv539 [PubMed: 25999341]
33. Choi YH, Lim J-K, Jeong M-W, Kim K-T. HnRNP A1 phosphorylated by VRK1 stimulates telomerase and its binding to telomeric DNA sequence. *Nucleic Acids Res.* 2012; 40: 8499–8518. DOI: 10.1093/nar/gks634 [PubMed: 22740652]
34. Flynn RL, Centore RC, O'Sullivan RJ, Rai R, Tse A, Songyang Z, et al. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature.* 2011; 471: 532–536. DOI: 10.1038/nature09772 [PubMed: 21399625]

35. Fiset S, Chabot B. hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res.* 2001; 29: 2268–2275. DOI: 10.1093/nar/29.11.2268 [PubMed: 11376145]
36. Redon S, Zemp I, Lingner J. A three-state model for the regulation of telomerase by TERRA and hnRNPA1. *Nucleic Acids Res.* 2013; 41: 9117–9128. DOI: 10.1093/nar/gkt695 [PubMed: 23935072]
37. Bekenstein U, Soreq H. Heterogeneous nuclear ribonucleo-protein A1 in health and neurodegenerative disease: from structural insights to post-transcriptional regulatory roles. *Mol Cell Neurosci.* 2013; 56: 436–446. [PubMed: 23247072]
38. LaBranche H, Dupuis S, Ben-David Y, Bani M-R, Wellinger RJ, Chabot B. Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet.* 1998; 19: 199–202. [PubMed: 9620782]
39. Dallaire F, Dupuis S, Fiset S, Chabot B. Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J Biol Chem.* 2000; 275: 14509–14516. [PubMed: 10799534]
40. Thibault PA, Ganesan A, Kalyanammoorthy S, Clarke J-PWE, Salapa HE, Levin MC. hnRNP A/B proteins: an encyclopedic assessment of their roles in homeostasis and disease. *Biology.* 2021; 10: 712. doi: 10.3390/biology10080712 [PubMed: 34439945]
41. Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature.* 2013; 495: 467–473. DOI: 10.1038/nature11922 [PubMed: 23455423]
42. Liu X, Ishizuka T, Bao H-L, Wada K, Takeda Y, Iida K, et al. Structure-dependent binding of hnRNPA1 to telomere RNA. *J Am Chem Soc.* 2017; 139: 7533–7539. DOI: 10.1021/jacs.7b01599 [PubMed: 28510424]
43. Appadurai R, Nagesh J, Srivastava A. High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme. *Nat Commun.* 2021; 12: 1–11. DOI: 10.1038/s41467-021-21105-7 [PubMed: 33397941]
44. Natarajan C, Srivastava A. Efficiently determining membrane-bound conformations of peripheral membrane proteins using replica exchange with hybrid tempering. *Eur Phys J Spec Top.* 2024; 233: 3039–3051.
45. Bowers SR, Lockhart C, Klimov DK. Replica exchange with hybrid tempering efficiently samples PGLa peptide binding to anionic bilayer. *J Chem Theor Comput.* 2023; 19: 6532–6550. [PubMed: 37676235]
46. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science.* 2020; 367: 694–699. DOI: 10.1126/science.aaw8653 [PubMed: 32029630]
47. Roy I, Appadurai R, Srivastava A. Unambiguous assignment of kinked beta sheets leads to insights into molecular grammar of reversibility in biomolecular condensates. *bioRxiv.* 2024; doi: 10.1101/2024.12.05.627008
48. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: implications for fibrillation. *Protein Sci.* 2006; 15: 2795–2804. DOI: 10.1110/ps.062465306 [PubMed: 17088319]
49. McConnell BS, Parker MW. Protein intrinsically disordered regions have a non-random, modular architecture. *Bioinformatics.* 2023; 39: btad732 doi: 10.1093/bioinformatics/btad732 [PubMed: 38039154]
50. Appadurai R, Koneru JK, Bonomi M, Robustelli P, Srivastava A. Clustering heterogeneous conformational ensembles of intrinsically disordered proteins with t-distributed stochastic neighbor embedding. *J Chem Theor Comput.* 2023; 19: 4711–4727. DOI: 10.1021/acs.jctc.3c00224 [PubMed: 37338049]
51. Balasubramanian S, Maharana S, Srivastava A. “Boundary residues” between the folded RNA recognition motif and disordered RGG domains are critical for FUS–RNA binding. *J Biol Chem.* 2023; 299: 105392 doi: 10.1016/j.jbc.2023.105392 [PubMed: 37890778]

52. Morris OM, Torpey JH, Isaacson RL. Intrinsically disordered proteins: modes of binding with emphasis on disordered domains. *Open Biol.* 2021; 11 210222 doi: 10.1098/rsob.210222 [PubMed: 34610267]
53. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature.* 2018; 555: 61–66. DOI: 10.1038/nature25762 [PubMed: 29466338]
54. Sisk TR, Robustelli P. Folding-upon-binding pathways of an intrinsically disordered protein from a deep Markov state model. *Proc Natl Acad Sci.* 2024; 121 e2313360121 doi: 10.1073/pnas.2313360121 [PubMed: 38294935]
55. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol.* 2009; 19: 31–38. DOI: 10.1016/j.sbi.2008.12.003 [PubMed: 19157855]
56. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, Thomas PJ, et al. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat Struct Mol Biol.* 2007; 14: 738–745. DOI: 10.1038/nsmb1278 [PubMed: 17660831]
57. Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci.* 2008; 105: 17772–17777. DOI: 10.1073/pnas.0809222105 [PubMed: 19008353]
58. Hendus-Altenburger R, Pedraz-Cuesta E, Olesen CW, Papaleo E, Schnell JA, Hopper JTS, et al. The human Na⁺/H⁺ exchanger 1 is a membrane scaffold protein for extracellular signal-regulated kinase 2. *BMC Biol.* 2016; 14: 1–17. DOI: 10.1186/s12915-016-0252-7 [PubMed: 26728391]
59. Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, et al. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell.* 2015; 163: 734–745. DOI: 10.1016/j.cell.2015.09.047 [PubMed: 26456112]
60. Ishiguro A, Katayama A, Ishihama A. Different recognition modes of G-quadruplex RNA between two ALS/FTLD-linked proteins TDP-43 and FUS. *FEBS Lett.* 2021; 595: 310–323. [PubMed: 33269497]
61. Phan AT, Kuryavii V, Darnell JC, Serganov A, Majumdar A, Ilin S, et al. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol.* 2011; 18: 796–804. DOI: 10.1038/nsmb.2064 [PubMed: 21642970]
62. Vasilyev N, Polonskaia A, Darnell JC, Darnell RB, Patel DJ, Serganov A. Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc Natl Acad Sci.* 2015; 112: E5391–E5400. DOI: 10.1073/pnas.1515737112 [PubMed: 26374839]
63. Yagi R, Miyazaki T, Oyoshi T. G-quadruplex binding ability of TLS/FUS depends on the i^2 -spiral structure of the RGG domain. *Nucleic Acids Res.* 2018; 46: 5894–5901. DOI: 10.1093/nar/gky391 [PubMed: 29800261]
64. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015; 12: 7–8. DOI: 10.1038/nmeth.3213 [PubMed: 25549265]
65. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008; 9: 1–8. DOI: 10.1186/1471-2105-9-40 [PubMed: 18173834]
66. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010; 5: 725–738. DOI: 10.1038/nprot.2010.5 [PubMed: 20360767]
67. Ozenne V, Bauer F, Salmon L, Huang J, Jensen MR, Segard S, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 2012; 28: 1463–1470. [PubMed: 22613562]
68. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical J.* 2005; 89: 1237–1250. DOI: 10.1529/biophysj.105.064154 [PubMed: 15923225]
69. Teixeira JMC, Liu ZH, Namini A, Li J, Vernon RM, Krzeminski M, et al. IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states. *The J Phys Chem A.* 2022; 126: 5985–6003. DOI: 10.1021/acs.jpca.2c03726 [PubMed: 36030416]

70. Shea J-E, Best RB, Mittal J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Curr Opin Struct Biol.* 2021; 67: 219–225. DOI: 10.1016/j.sbi.2020.12.012 [PubMed: 33545530]
71. Latham AP, Zhang B. Unifying coarse-grained force fields for folded and disordered proteins. *Curr Opin Struct Biol.* 2022; 72: 63–70. DOI: 10.1016/j.sbi.2021.08.006 [PubMed: 34536913]
72. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017; 14: 71–73. DOI: 10.1038/nmeth.4067 [PubMed: 27819658]
73. Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods.* 2019; 16: 670–673. [PubMed: 31363226]
74. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Computer Phys Commun.* 2009; 180: 1961–1972.
75. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. Plumed 2: new feathers for an old bird. *Computer Phys Commun.* 2014; 185: 604–613.
76. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical J.* 2015; 109: 1528–1532. DOI: 10.1016/j.bpj.2015.08.015 [PubMed: 26488642]
77. Shen Y, Bax A. SPARTA₊: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR.* 2010; 48: 13–22. DOI: 10.1007/s10858-010-9433-9 [PubMed: 20628786]
78. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. BioMagResBank. *Nucleic Acids Res.* 2007; 36: D402–D408. DOI: 10.1093/nar/gkm957 [PubMed: 17984079]
79. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Machine Learn Res.* 2011; 12: 2825–2830.
80. Dai J, Carver M, Yang D. Polymorphism of human telomeric quadruplex structures. *Biochimie.* 2008; 90: 1172–1183. DOI: 10.1016/j.biochi.2008.02.026 [PubMed: 18373984]
81. Phan AT. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* 2010; 277: 1107–1117. [PubMed: 19951353]
82. Andrałojc W, Małgowska M, Sarzyńska J, Pasternak K, Szpotkowski K, Kierzek R, et al. Unraveling the structural basis for the exceptional stability of RNA G-quadruplexes capped by a uridine tetrad at the 3' terminus. *RNA.* 2019; 25: 121–134. DOI: 10.1261/rna.068163.118 [PubMed: 30341177]
83. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature.* 2002; 417: 876–880. [PubMed: 12050675]
84. Martadinata H, Phan AT. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K₊ solution. *J Am Chem Soc.* 2009; 131: 2570–2578. [PubMed: 19183046]
85. Gavathiotis E, Searle MS. Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)₄ containing the human telomeric repeat: evidence for A-tetrad formation from NMR and molecular dynamics simulations. *Org Biomol Chem.* 2003; 1: 1650–1656. [PubMed: 12926351]
86. Dai J, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang D. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.* 2007; 35: 2440–2450. DOI: 10.1093/nar/gkm009 [PubMed: 17395643]
87. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K₊ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* 2007; 35: 4927–4940. DOI: 10.1093/nar/gkm522 [PubMed: 17626043]
88. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a Protein–Protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003; 125: 1731–1737. [PubMed: 12580598]
89. Lindahl, Abraham, Hess, van der Spoel. GROMACS 2020.4 Manual. 2020; doi: 10.5281/zenodo.4054996

90. Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Cheatham TEI, et al. Refinement of the Cornell Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theor Comput.* 2011; 7: 2886–2902. DOI: 10.1021/ct200162x [PubMed: 21921995]
91. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci.* 2018; 115: E4758–E4766. DOI: 10.1073/pnas.1800690115 [PubMed: 29735687]
92. Zgarbová M, Šponer J, Otyepka M, Cheatham TEI, Galindo-Murillo R, Jurev ka P. Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J Chem Theor Comput.* 2015; 11: 5723–5736. [PubMed: 26588601]
93. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol graphics.* 1996; 14: 33–38. [PubMed: 8744570]

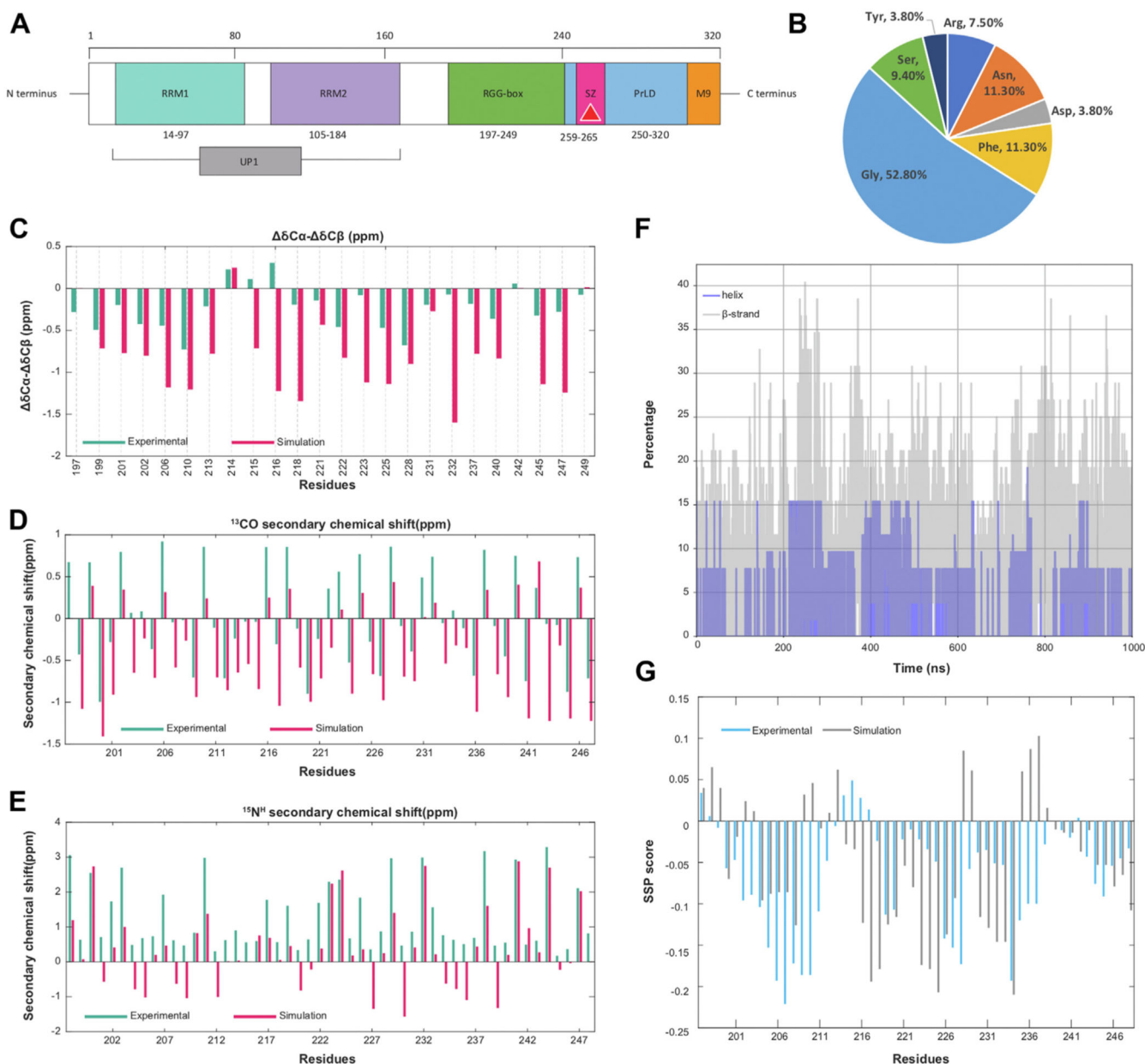


Figure 1. Sequence and structural insights into RGG-box domain ensemble.

A, domain architecture of hnRNPA1. The *Teal* and *Purple* colored domains are the N terminal-ordered RNA recognition motifs. Together, these two domains are called UP1. The steric-zipper motif (259–265) is colored in *Magenta*. The *Green* and the *Blue* domains are intrinsically disordered RGG-box domain and PrLD, respectively. The *Orange* colored domain is the nuclear shuttling sequence M9. **B**, sequence composition of RGG-box motif. **C–E**, the $\text{C}\alpha$ - $\text{C}\beta$, ^{13}CO , and $^{15}\text{N}^{\text{H}}$ secondary chemical shift plots for RGG, respectively. *Teal* is experimental data. *Dark pink* indicates simulation data. The $\text{C}\alpha$ - $\text{C}\beta$ chemical shifts are calculated and plotted for the residues containing $\text{C}\beta$ atoms. **F**, comparison of α -helix (*gray*) and β -sheet (*blue*) content of RGG during the simulation. **G**, predicted secondary structure propensity scores from chemical shift data of experiments (*blue*) and

simulation (*gray*) using the SSP algorithm. SSP score above 0 indicates helical propensity and below 0 indicates β -sheet propensity.



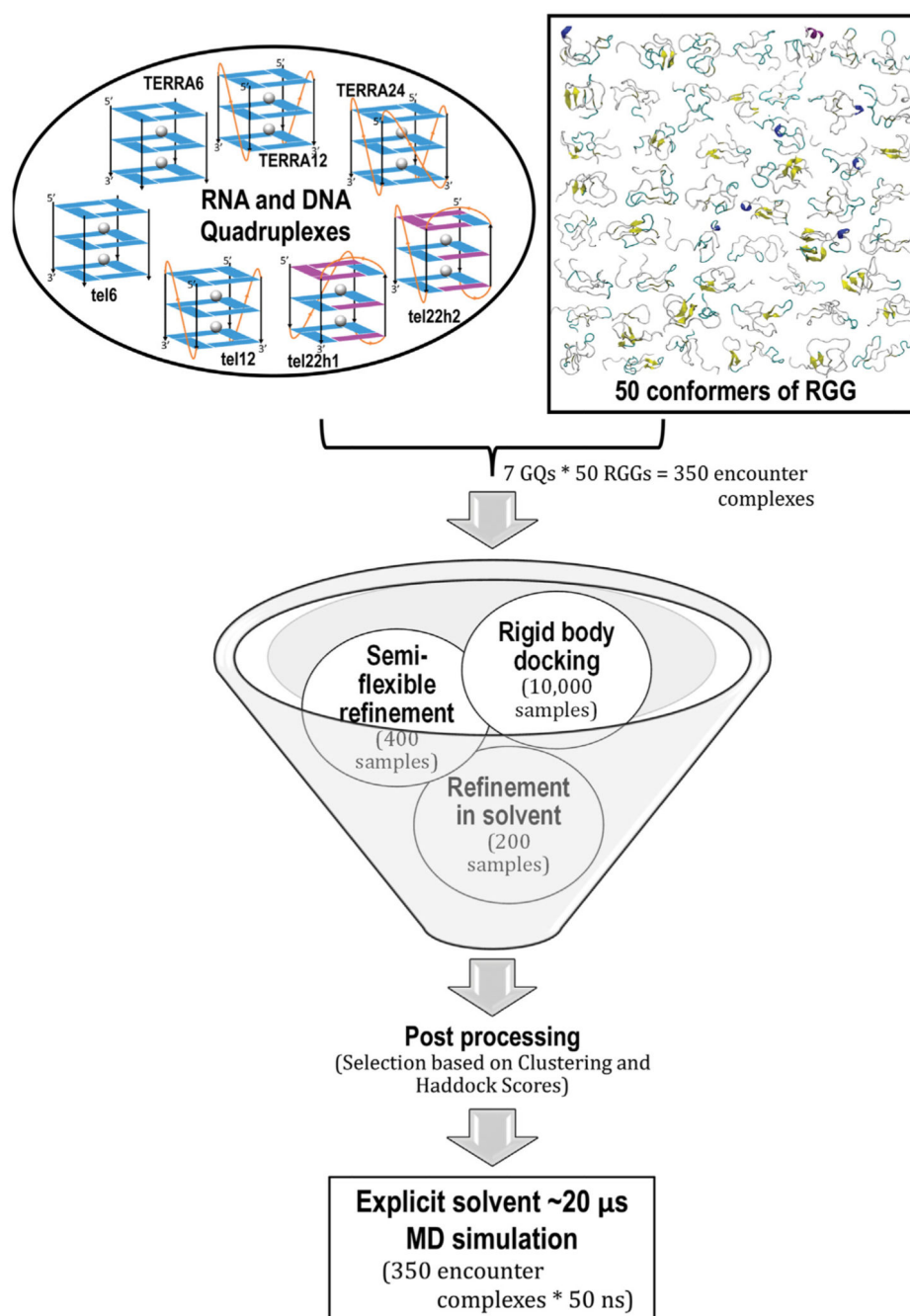


Figure 2. Flowchart of the protocol adopted to generate RGG-GQ complexes.

Seven different telomeric DNA and RNA GQ structures chosen from the literature were docked with each of the 50 conformers of RGG obtained by t-SNE clustering of the generated ensemble. HADDOCK protocol with a center of mass distance criteria was used to reduce the sample space from 10,000 initial random geometries for each complex. The 350 (7 GQs with 50 RGGs) encounter complexes were further simulated for 50 ns each resulting in a 17.5 μs ensemble to explore the mode of binding, stability, and conformational dynamics of the RGG-GQ binding process.



Figure 3. Binding modes of RGG with GQ.

The normalized number of interactions between each amino acid type in RGG (y-axis) with every base of GQ (x-axis) was calculated from the simulated complexes. The interactions were calculated based on the distance between heavy atoms and the number of interactions was normalized on the total number of each amino acid type. Rows represent the seven GQ systems, while the columns represent end-stacking, groove, and loop-binding modes. The contact probabilities are color-coded on a *yellow to black* scale where darker colors indicate a higher probability of interactions.

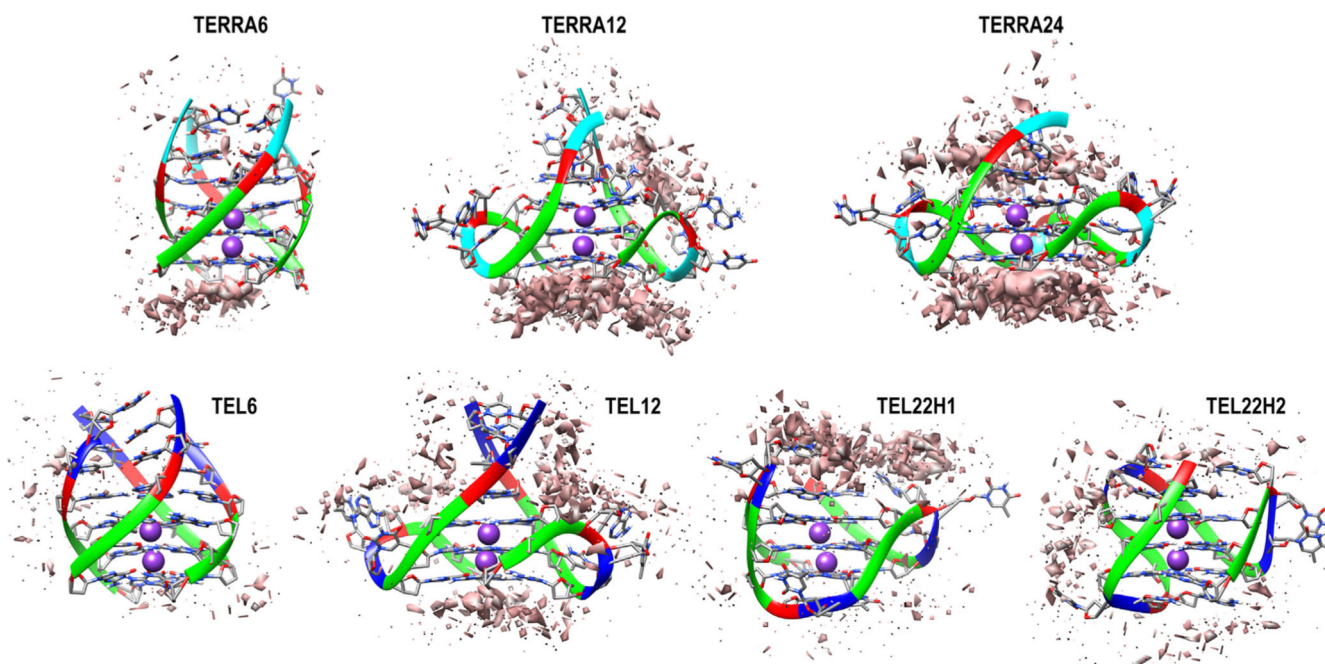


Figure 4. Occupancy map depicting the extent of RGG-GQ interaction after the 50 ns simulation.

A high density of interactions at specific sites on GQ is shown as an isosurface mimicking the first solvation shell (*orange* surface). The interacting sites and density vary considerably from the complexes before simulation (see Fig. S3 in SI for occupancy maps before simulation). The GQ backbone is shown as ribbons while the bases are displayed as sticks. The bases are colored as *Green*: Guanine, *Red*: Adenine, *Blue*: Thymine, and *Cyan*: Uracil.

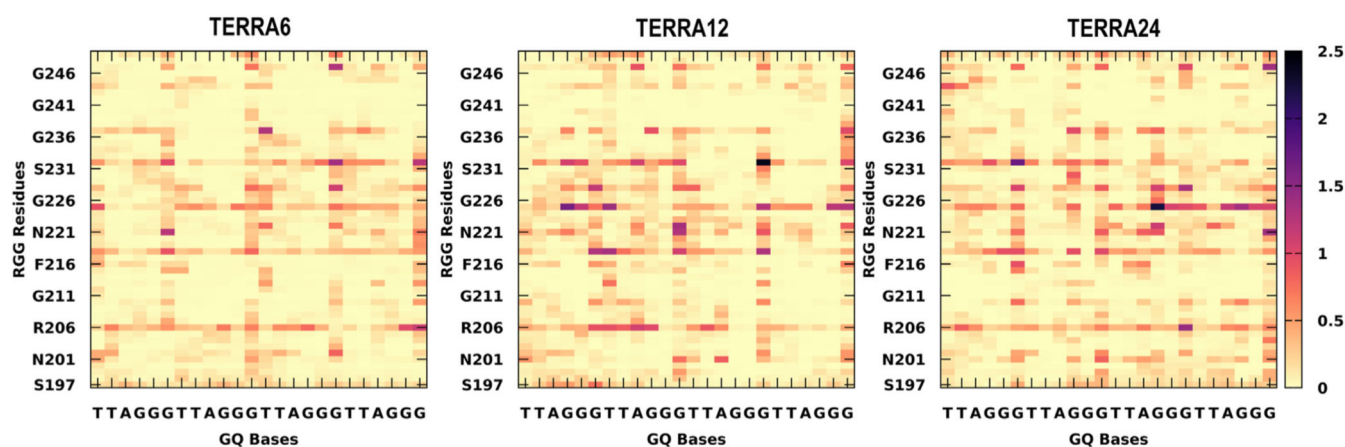


Figure 5. Residue-wise contact maps representing telomeric RNA-GQ interactions.

The x-axis indicates the GQ bases and the RGG sequence on the y-axis. The color scale varies from *yellow* to *black* and indicates a longer lifetime of interactions. The contacts were calculated with a distance cutoff of 4 Å.

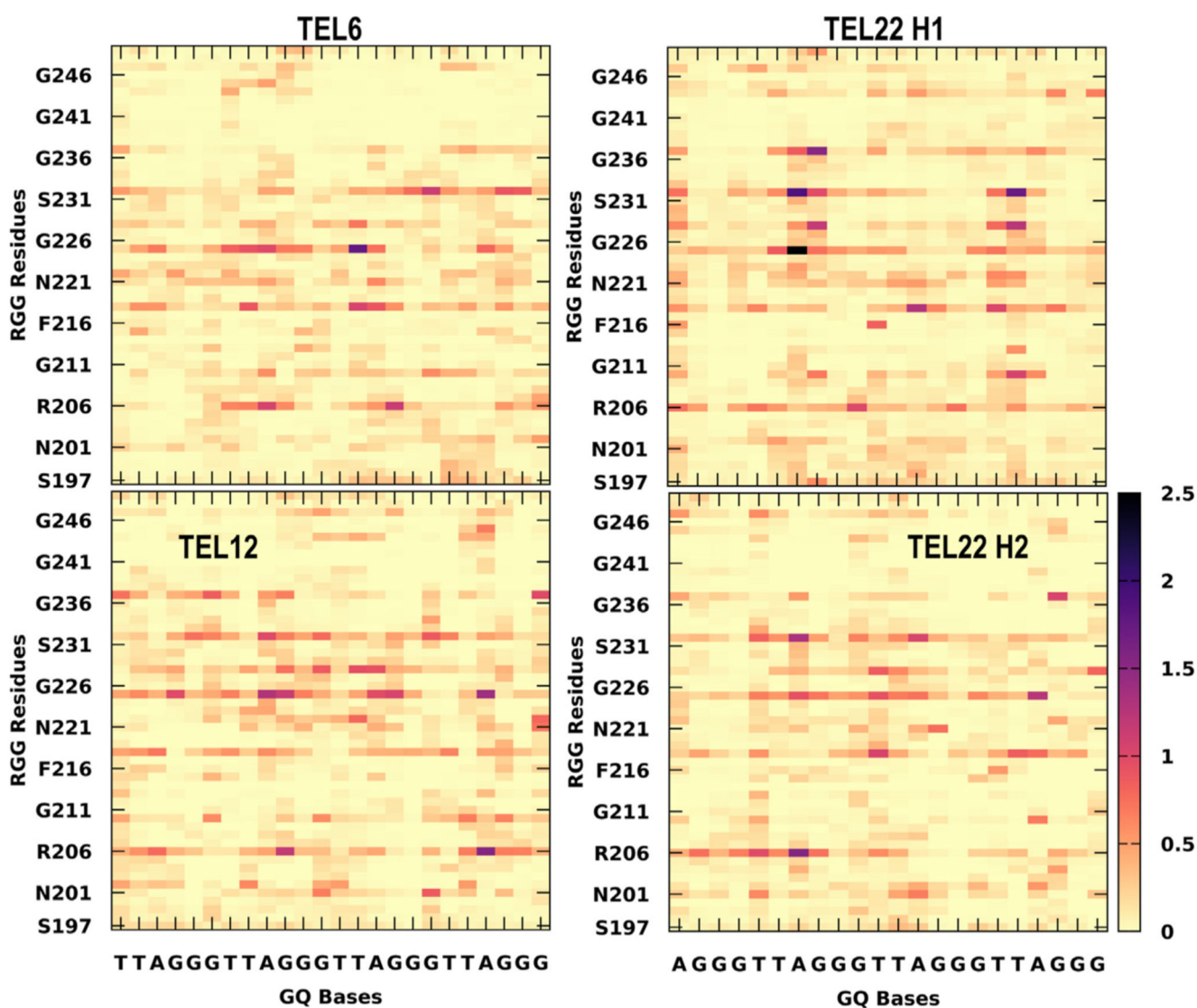


Figure 6. Residue-wise contact maps representing telomeric DNA-GQ interactions. The x-axis indicates the GQ bases and the RGG sequence on the y-axis. The color scale varies from *yellow* to *black* and indicates a longer lifetime of interactions. The contacts were calculated with a distance cutoff of 4 Å.

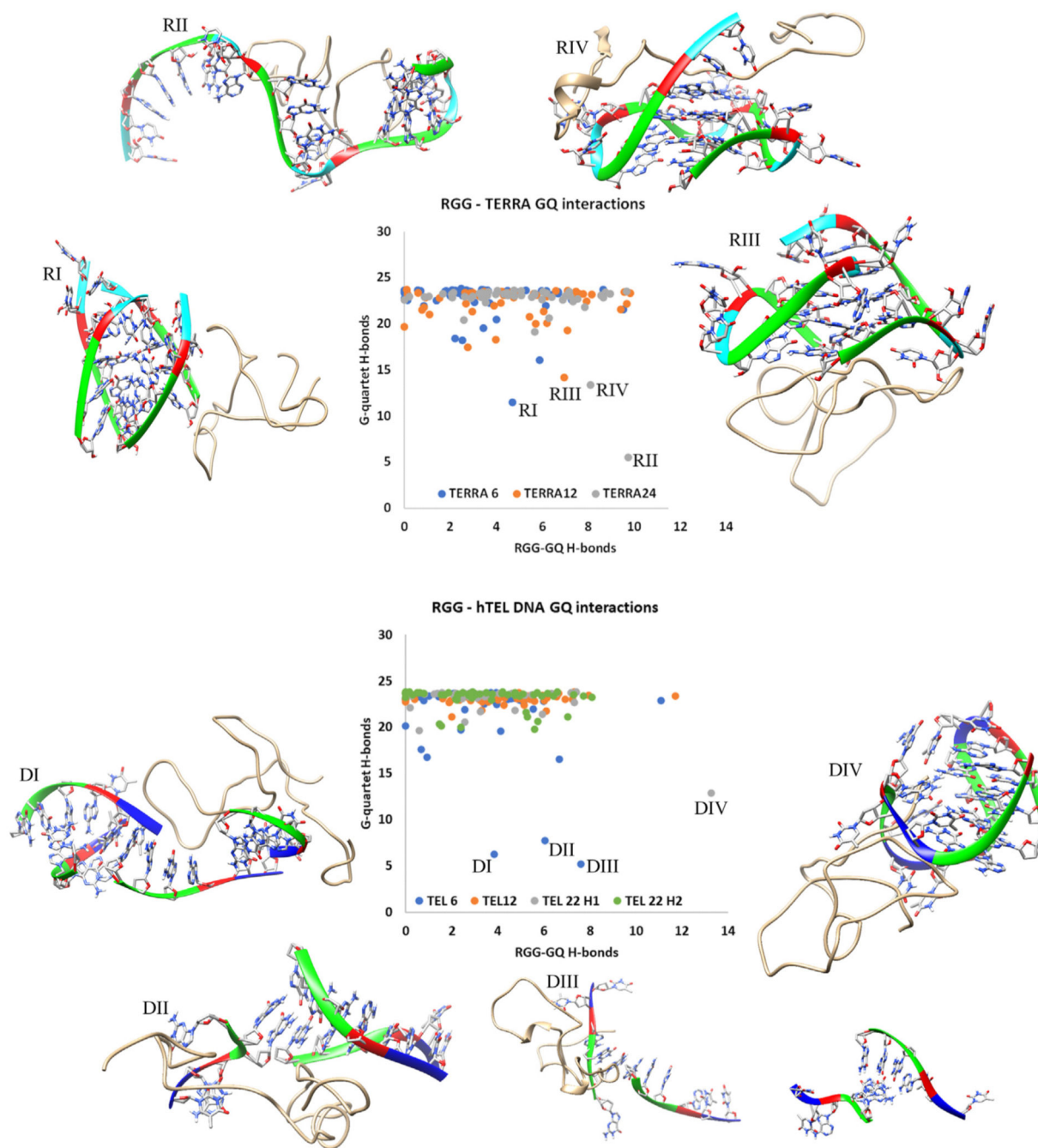


Figure 7. Scatter plot between the number of intermolecular RGG-GQ (RNA and DNA) hydrogen bonds and intramolecular G-quartet hydrogen bonds highlighting the complex stability.

The unfolding complexes are labeled and the structures after 50 ns MD are shown.

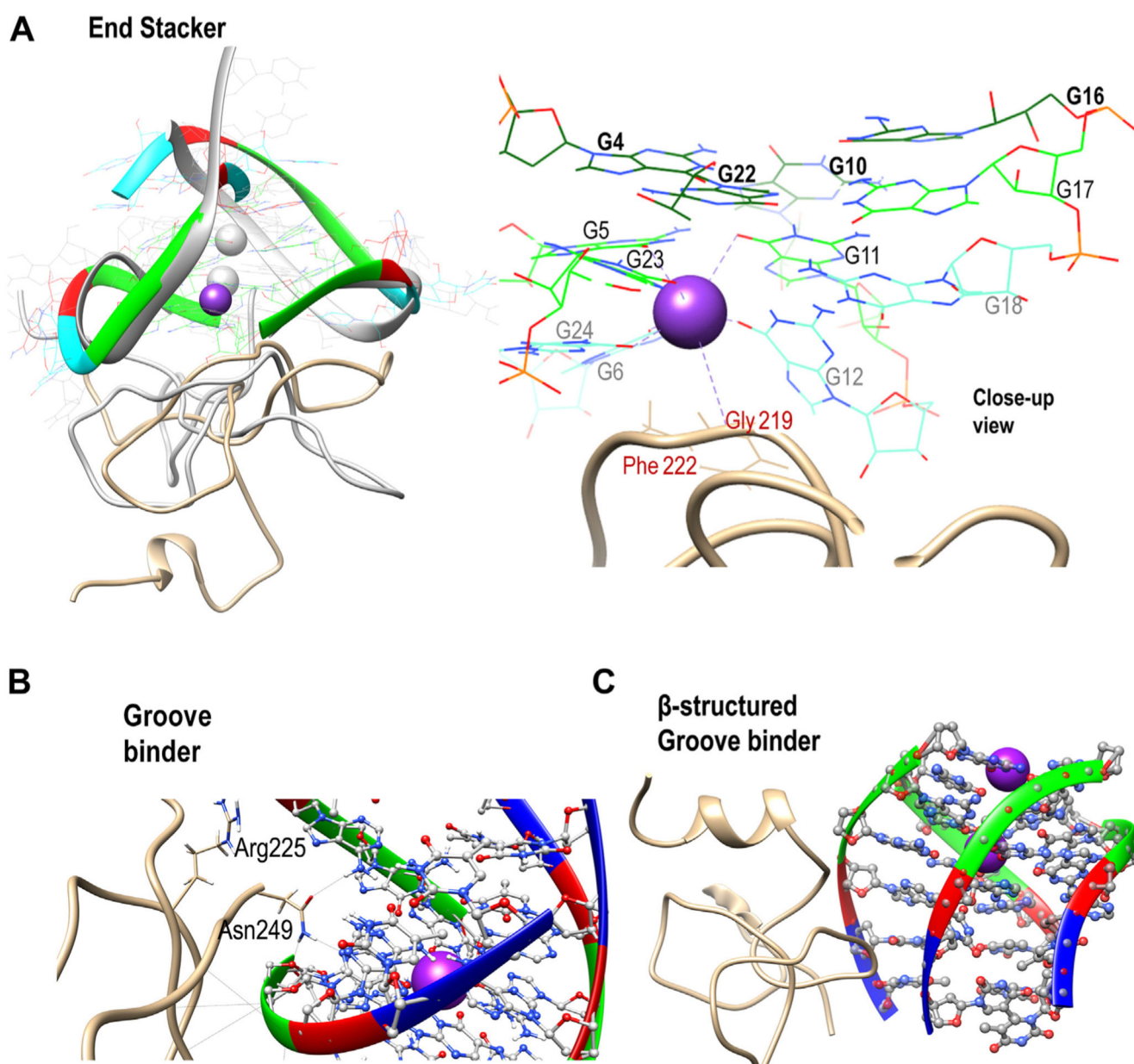


Figure 8. RGG-induced unfolding mechanism.

A, the end stacking mode of TERRA12 and (B) the groove binding mode of tel6 are shown as examples. The superposition of the initial structure (in *grey*) and 50 ns simulated conformations of TERRA12 are shown to clearly portray the loss of coordinated K⁺ ions due to RGG. C, depicts the unfolding of tel6 by a β-structured motif binding in the groove. The RGG backbone is shown as ribbons while the sidechains are depicted in wire representation. The K⁺ ion is shown as a *purple* sphere and its coordination shell is identified with dotted lines. Phosphate: *Yellow*, Carbon: *gray* (RGG)/*brown* (GQ), Nitrogen: *Blue*, Oxygen: *red*, Hydrogen: *white* and K⁺ ion: *purple*.

Table 1 Telomeric RNA and DNA structures modeled in this study

Systems	PDB ID	Simulated systems		Comments
		Apo (ns)	With the 50 conformers of RGG (ns)	
TERRA6	based on 6GE1	300	50	tetramolecular RNA
TERRA12	2KBP	300	50	bimolecular RNA
TERRA24	based on 1KF1	300	50	unimolecular RNA
Te16	1NB9	300	50	tetramolecular DNA
tel12	1K8P	300	50	bimolecular DNA
tel22h1	2HY9	300	50	unimolecular DNA hybrid 1
tel22h2	2JPZ	300	50	unimolecular DNA hybrid 2

For every GQ, there are fifty complexes.

