G-Quadruplex-Protein Interactions: Screening, Characterization and Regulation

Yicong Dai[a,b], Xucong Teng[a,b], Qiushuang Zhang[a,b], Hongwei Hou[c,*], Jinghong Li[a,b,*]

a Department of Chemistry, Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology, Tsinghua University, Beijing 100084, China

b Center for BioAnalytical Chemistry, Hefei National Laboratory of Physical Science at Microscale, University of Science and Technology of China, Hefei 230026, China

c China National Tobacco Quality Supervision and Test Center, Zhengzhou 450001, China

*Correspondence: jhli@mail.tsinghua.edu.cn (J.H. Li), houhw@ztri.com.cn (H.W. Hou)

Abstract

G-quadruplex (G4) is a kind of peculiar nucleic acid secondary structure formed by DNA or RNA, which is considered as a fundamental feature of the genome. Many proteins can specifically bind to G4 structures. There is increasing evidence that G4-protein interactions involve in the regulation of important cellular processes such as DNA replication, transcription, RNA splicing and translation. Besides, G4-protein interactions have been proved to be potential targets for disease treatment. In order to unravel the detailed bioregulatory mechanisms of G4-binding proteins (G4BPs), biochemical methods for detecting G4-protein interactions with high specificity and sensitivity are highly demanded. Here, we review recent advances in the screening and validation of new G4BPs and bioregulatory tools of G4BPs.

Keywords: G-quadruplex binding protein, high-throughput screening, characterization, regulation, nucleic acid secondary structure

The importance of G-quadruplex-protein interactions

G-quadruplex (G4) is a peculiar nucleic acid secondary structure formed by stacks of **G-tetrads** (see **Glossary**). As numerous G4 structures are present in the genome and transcriptome of living cells, the biological functions of G4s have attracted extensive attention in the past decade[1-7]. It has been increasingly clear that G4s play an important role in telomere maintenance[8, 9], DNA replication[10], chromatin remodeling[11, 12], transcription[13], RNA splicing[14], and translation[15-17]. G4s participate in biological processes through two major mechanisms. One is as a module for recruitment of functional G4-binding proteins (G4BPs), such as transcription factors. The other is to hinder the extension of the DNA polymerase, RNA polymerase and ribosome on the nucleic acid strand. So far, about one hundred functional proteins have been confirmed to interact with G4 (**Box 1**). These G4BPs are closely related to diseases such as tumors, ageing and viral infections, and are considered to be promising therapeutic targets[18-20] (**Box 2**). However, there are many questions about G4BPs remain unclear. For instance, there is little understanding on how G4-binding transcription factors specifically recognize promoters instead of binding to all G4-containing promoter regions. Moreover, the mechanism by which G4-unfolding proteins and G4-recruited proteins coordinate with each other is still unclear.

To unravel the detailed mechanism of G4BP-regulated biological processes, the detection of G4-protein interactions with high specificity and sensitivity is a prerequisite. In the past decade, many biochemical tools for G4-protein interactions have been established. These methods greatly facilitated the screening, identification and structural analysis of G4-protein interactions, as well as the regulation of G4BP-mediatied bioprocesses. In this review, we summarize the recent development on how to screen and validate new G4BPs. Then, we discuss the bio-regulation tools of G4BPs. Finally, we propose improvements that are still needed in the future development of G4-protein interaction biochemical methods.

Methods for the discovery of new G4BP

How many proteins have interactions with G4s, is still unclear. Discovering new G4BP can greatly help to understand the underlying regulatory mechanisms related to intracellular G4s. The screening of G4BP relies on the development of recognizing and labeling technologies

for specifically targeting G4-binding proteins, as well as high-throughput proteomic technologies.

Affinity enrichment methods

The affinity enrichment method can be utilized to enrich for G4BPs in cell lysates and identify G4BPs accompanied with Western-blotting. In this method, G4-forming DNA or RNA strands are conjugated to magnetic beads or agarose gels to form a G4-affinity enrichment probe. Then the probe is incubated with the cell lysate to capture and extract G4BPs. Taking advantages of proteomic mass spectrometry, the enriched proteins can be identified and analyzed (**Table 1**).

In this assay, it is important to exclude the proteins that only have non-specific binding capacity to single-stranded nucleic acids but have no specific binding capacity to G4 structures. To achieve this goal, the common strategy is to use nucleic acid strands that do not form G4 structures as controls, such as magnetic beads or agarose gels conjugated with Gs-As mutant G4 sequences and nucleic acid hairpins, as well as bare magnetic beads or agarose gels[21-25] (**Figure 1A**). After LC-MS/MS analysis and **label-free quantitative analysis**, a G4-specific binding protein is identified when the difference in protein signal between the G4 group and the control group is more than 2-fold and the p-value is less than 0.05. Sometimes more stringent filtering conditions are also used to improve the confidence level. Based on this strategy, earlier work could only identify 10-20 potential G4-binding proteins with high confidence[21, 22, 24], possibly limited by the performance of the mass spectrometer. Recently, Herdy et al. used this strategy to identify G4BPs targeting NRAS mRNA 5'-UTR G4, and found 80 high-confidence G4BPs, including five newly identified G4BPs, DDX3X, DDX5, DDX17, GRSF1 and NSUN5[25]. A complex but accurate approach to exclude non-specific nucleic acid binding proteins was previously taken by Gonza´lez et al[26] (**Figure 1B**). They first applied the extracted protein mixture to the G4-MUT affinity column that contains a mutation that destabilizes the c-MYC G4 structure. The G4-MUT **flow-through** protein mixture was applied to the G4 affinity column. In this way, proteins that bound to G4 but not to G4-MUT were identified. Although they identified only 10 high-confidence proteins, most of these proteins were verified to be indeed G4-binding proteins.

In addition, Wang's group used a forward stable isotope labeling by amino acid in cell culture (SILAC)-based quantitative proteomic method in order to quantitatively compare the protein signal difference between the G4 group and the control group[27-29] (**Figure 1C**). The HeLa cells were cultured separately in light or heavy medium, and equal amounts of nuclear extracts from the light- and heavy-labeled cells were applied to the G4 affinity column and the corresponding G4-MUT affinity column, respectively. Taking advantages of the SILAC-based method, they identified a new G4BP, SLIRP[27]. This method enables accurate quantification, but is complex, costly, time-consuming, and requires sophisticated bioinformatic analysis as compared to label-free quantitative strategy.

The affinity enrichment method coupled with protein mass spectrometry for the discovery of new G4BPs offers great advantages. On the one hand, the affinity enrichment method can efficiently isolate and enrich G4-binding proteins from cell lysates with complex composition. On the other hand, due to the modification of unique G4 sequences on magnetic beads or agarose gels, this method allows the identification of G4BPs that bind to a specific G4 sequence and the analysis of the preference of G4BPs to bind to different G4 sequences. For example, the binding preference of several G4BPs to BCL2, NRAS, MMP16 and ARPC2 mRNA G4s have been well evaluated[28]. However, the affinity enrichment method also has many shortcomings. First, the interaction state of proteins to synthesized nucleic acids in cell lysates is not exactly the same as it in living cells. In living cells, the binding site of the protein or the G4 structure may be occupied by other biomolecules. The in vitro affinity enrichment method can only isolate and enrich G4BPs in cell lysates, but cannot reflect the interactions between G4BP and G4 structures in living cells under physiological environment. Therefore, protein-G4 interactions discovered using affinity enrichment methods may mislead the study of intracellular biological processes. Second, a large amount of nucleases present in the lysate of some cells can degrade the G4 probe to some extent, and thus interfere with affinity enrichment. In particular, in the affinity enrichment analysis of RNA G4BP, the interference of RNases could not be completely eliminated even after the addition of RNase inhibitors. Finally, low-abundance G4BP is difficult to be effectively identified by affinity enrichment methods. Low-abundance proteins have lower enrichment efficiency and

have poorer signal quality in MS analysis due to interference from high-abundance protein signals, thus are filtered out in the data quality control step.

In situ intracellular labeling methods

In order to characterize the interactome between G4BPs to G4 structures in the physiological state, a **G4 ligand**-mediated **photo-crosslinking** protein assay has been developed to identify and label G4BPs in living cells (**Figure 1D**). This assay utilizes a cell-permeable photo-crosslinked G4 probe. The probe consists of three functional parts: the G4 ligand for targeting and recognition of G4 structures, a photoreactive alphatic diazirine group for labeling the target interacting proteins, and the alkyne group for conjugating azide-biotin with protein-probe complex through click reaction. After incubating in the cell culture medium for several hours, the photo-crosslinked G4 probe can bind to the G4 structures in living cells. The probe is then cross-linked to the target proteins by 365 nm light irradiation to form a protein-probe complex, and the biotin tag is conjugated to the protein-probe complex by a click reaction catalyzed by monovalent copper ions. The protein-probe complex was enriched and purified by streptavidin magnetic beads and finally identified by label-free mass spectrometry. Potential G4BPs can be identified by comparing the results of the experimental group with those of the control group (using a photo-crosslinked probe without G4 ligand). Taking advantages of this method, Zhang et al. identified more than 200 high-confidence proteins, and validated four new G4BPs from them[30]. In the same year, Su et al. used a similar approach to study G4BP in SV589 and MM231 cell lines and validated 8 new G4BPs[31].

G4 ligand-mediated photo-crosslinking protein assay can well reflect the interaction between G4BP and G4 structures in the intracellular microenvironment. Therefore, this method can provide more accurate information on the G4-protein interactome for studying the detailed mechanisms of G4-mediated biological processes. In addition, this method is simple, time effective, and efficient in labeling both high-abundance and low-abundance proteins, as compared to the affinity enrichment method. However, this method still has some disadvantages. The G4 ligand-based photo-crosslinking probe can bind to all G4 motifs in living cells, including DNA G4s and RNA G4s, and in theory labels all G4BPs at the same

time. Because there are hundreds of thousands of DNA G4s and tens of thousands of RNA G4s in human cells, this method cannot reveal which G4BPs are performing biological functions at the unique G4 structure in a specific gene, which are very important for uncovering the regulation mechanism of G4-mediated biological processes. Moreover, it is controversial whether G4 ligands compete with G4BPs in binding G4 structures. It has been reported that G4 ligands and G4BPs can bind to G4 structures simultaneously[32], but at high concentration G4 ligands can interfere with the binding of G4BP to the G4 structure[33], serving as a G4BP inhibitor. Also, G4 ligands can induce unfolded G4 sequences to form G4 structures[34], thus altering the original state of intracellular G4-protein interactions.

Computational prediction methods

In recent years, deep learning methods for predicting protein structures have developed very rapidly[35], and are expected to be used to discover the G4 structure binding ability of known proteins. Pioneering work has analyzed the common features of G4BP using bioinformatics methods and identified RG-rich regions as the key structural domains of G4BP[36]. Because there are few reports on the structural data of G4BPs, the bioinformatical prediction can only be carried out through amino acid motif analysis, which restricts the accuracy of the prediction results. As more and more three-dimensional structures of G4BP in the state of interaction with G4 structures are resolved, a breakthrough will be made in the discovery of new G4BPs by deep learning prediction methods.

Future perspective of G4BP screening

To address the shortcomings of current methods for G4BP recognition and proteomic analysis, there is still a need to develop new biochemical methods for high-throughput detection of G4BP. Importantly, it is necessary to develop high-sensitive biochemical methods for labeling G4BPs that bind to a unique G4 motif in living cells, providing the possibility to study the function and mechanism of a G4 structure of a specific gene. Yang’s group developed a quinoline derivative PEQ that selectively recognize MYC promoter G4[37]. PEQ has a free amino group that can be functionalized with a diazirine group, thus

has the potential to be applied to G4 ligand-mediated photo-crosslinking methods for probing G4BP in the MYC promoter region.

Methods for the characterization of G4BP

G4BP affinity enrichment and in situ labeling methods combined with proteomic analysis methods provide a powerful tool for the discovery of new potential G4BPs. However, the identified proteins have the potential to be false positives. And, since intracellular proteins often form protein complexes, it is possible that these identified proteins only form complexes with other G4BPs and do not directly interact with G4 structures. So far, all high-throughput screening efforts for G4BPs have identified a total of more than 500 potential G4BPs, of which only about 20% have been reported to have the ability to bind G4 structures, and the rest of the proteins still need further validation. Therefore, a series of biochemical methods have been developed for G4-protein interactions in order to validate whether the potential G4BPs do interact specifically with G4 structures[38], as well as to analyze the binding constants and characterize the structural basis of the interactions (**Table 2**).

In vitro characterization of G4-protein interactions

The traditional methods for characterizing G4-protein interactions are electrophoretic mobility shift assay (EMSA) and dimethyl sulfate (DMS) footprinting assay[26, 39]. The validation of G4-unfolding proteins is detected by the alteration of the folding state of the G4 structure, which can be achieved by the fluorescence resonance energy transfer (FRET) method[40]. In this method, two fluorophores are conjugated at each end of the G4 nucleic acid strand, and the dynamic process in which the G4-unfolding protein disrupts the G4 structure can be revealed by analyzing the intensity of the FRET signal. The above methods can provide useful information about the binding ability of G4BP, but are severely limited by the availability of high-purity proteins. Therefore, methods that do not rely on in vitro expression of proteins are much needed. The affinity enrichment methods used for screening G4BPs can also be used for in vitro characterization of G4-protein interactions[30]. The interaction of the target protein with G4 can be confirmed by detecting the presence of the target protein in the protein mixture after G4 affinity enrichment based on the Western-

blotting method. This method is relatively simple, while it relies on expensive antibodies and the specificity of the antibodies is sometimes difficult to guarantee. It is worth noting that during the affinity enrichment of cell lysates, it is possible to capture proteins that bind to the G4-forming nucleic acid through indirect interactions, which cannot be considered as G4BPs.

In contrast to the above methods, single-molecule analysis methods can characterize the kinetics of G4-protein interactions. For this purpose, single-molecule force spectroscopy and zero-mode waveguide (ZMW) methods have been used for the study of G4-protein interactions. Single-molecule force spectroscopy, such as optical tweezers, magnetic tweezers, is to measure the relationship between the stretch distance and mechanical tension of individual biomolecules in real time. By measuring the change in mechanical tension of G4 structures before and after G4BP binding, the effect of G4BP binding on G4 structures, such as stabilizing or unfolding of G4 structures, can be studied at the single-molecule level[41-43]. ZMW method uses microchambers with an attoliter volume to capture biomolecules, thus ensuring that only one biomolecule is analyzed at a time. For instance, Patra et al. combined pulsed-interleaved dual-color fluorescence cross-correlation spectroscopy with the ZMW technique to study the process of binding and dissociation of the RGG motif from the G4 structure[44].

The determination of the affinity constant of G4BP to the G4 structure is another important objective, which can be achieved by enzyme-linked immunosorbent assay (ELISA) [32], surface plasma resonance (SPR) [45], micro-scale thermophoresis (MST) and other methods. The most commonly used method to detect the G4-protein binding constant is ELISA. For instance, Zhang et al. used the ELISA method to verify that several proteins identified by high-throughput screening methods do have specific G4-binding capacity[30]. MST quantifies the binding constants of biomolecular interactions by detecting the directional movement of biomolecules in a micro-scale temperature gradient. Kazemier et al. used the MST method to measure the interaction between the G4 antibody 1H6 and a series of nucleic acid molecules and found a strong interaction between the protein and the thymidine-rich single stranded DNA, implying a severe non-specificity of this antibody[46].

In addition, high-resolution nuclear magnetic resonance (NMR) and X-ray techniques can characterize detailed structural information of G4-protein interactions and help to explore the structural basis of G4BP-mediated biological processes. Three-dimensional structures of G4-protein complexes provided by NMR and X-ray are very essential to molecular modeling and rational targeting. These methods have been discussed in detail[47]. For example, Chen et al. used a solid-state NMR method to resolve the structural features of the DHX36 protein during binding and unfolding parallel G4 structures[48]. This study found that binding of DHX36 to the G4 structure leads to rearrangement of the helicase cover while pulling on the single-stranded DNA tail, ultimately driving G4 unfolding one residue at a time.

Intracellular characterization of G4-protein interactions

The above biophysical and biochemical methods are limited to in vitro studies, however, due to the distinct cellular context and dynamic G4 folding, the interactions detected in in vitro analysis may not actually occur in living cells. Therefore, in situ intracellular analysis that can reveal the state of G4-protein interactions in the cellular microenvironment is demanding. Until now, methods for in situ analysis of G4-protein interactions in cells are relatively few.

Our group developed a module assembled multifunctional probes assay (MAMPA) for single-cell imaging of the RNA G4 located in an individual gene, and found that the occupancy of the RNA G4 by G4BPs in an individual gene can be detected by the MAMPA method, combining with the conventional rolling circle amplification (RCA) RNA imaging method[49]. In this method, two probes recognize the G4 motif and the adjacent specific RNA sequence separately, and then trigger RCA by click chemistry. Since the cells were fixed with 4% paraformaldehyde before MAMPA imaging, the G4 structures bound by G4BPs in the cells could not be recognized by the probes, and thus the corresponding fluorescent spots could not be obtained. Meanwhile, the intracellular abundance of the RNA containing the target G4 structure can be detected by the RCA method targeting other regions of the RNA molecule, and is not affected by the binding of G4BPs. Therefore, the occupation of G4BP at the RNA G4 motif of a specific gene can be derived from the number of MAMPA spots and the number of RCA spots. However, this method can only provide the

proportion of the RNA G4 structure occupied by G4BPs, but cannot identify which G4BP binds to that G4 motif.

Future perspective of G4BP characterization

Many in vitro methods for G4-protein interaction have been developed based on some traditional nucleic acid-protein interaction assay. These techniques are easy to conduct, cost effective, and can provide meaningful information about G4-protein interactions. However, it is becoming increasingly clear that the rules of G4-protein interactions are sophisticated, various and context dependent. Therefore, the field calls for new in vivo methods with high sensitivity and specificity to investigate the G4-protein interaction dynamics and their biological regulation in cellular milieu.

There are several key points for the improvement of biochemical methods for intracellular G4-protein interactions. First, interference of G4 structures or G4-protein interactions by the probes should be avoided. Currently, commonly used intracellular G4 probes are small molecule G4 ligands. These G4 ligands can stabilize the G4 structure and induce the folding of unfolded G-rich sequences to form G4 structures, thus destroying the native G4 pool[34]. And, some of the G4 ligands also interfere with G4-protein interactions[50]. How to address these problems is a great challenge. Second, it is important to specifically detect the interaction between an individual protein and a G4 motif in an individual gene. In this way, the functions and regulatory mechanisms of G4-protein interactions in specific biological pathways can be revealed, providing the basis for G4BP-based disease diagnosis and treatment. The key challenge is to design new G4 ligands that specifically recognize individual G4 motifs. The arrayed deep cavitand hosts offer the possibility to address this challenge[51]. These water-soluble, self-folding deep cavitands can enable the discrimination between DNA strands with highly similar G4 structures and can be easily functionalized with different fluorescent dyes. The CRISPR technology[52], which recognizes nucleic acid sequences with high specificity, is also expected to be adapted for the recognition of nucleic acid secondary structures. Third, enabling high-throughput analysis of G4-protein interactions in living cells is an important demand and challenge for the field. Similar to DNA-encoded compound library (DEL) technology[53], bio-orthogonal

labeling of G4BP with nucleic acid tags, combined with high-throughput sequencing, is expected to achieve high-throughput analysis of the intracellular G4-protein interactome.

Regulation of G4-protein interactions

The molecular tools for regulating G4-protein interaction in living cells can contribute to the study of the mechanisms of G4-related biological processes, and to the therapeutics targeting G4BPs. Current approaches to the regulation of G4-protein interactions can be classified into three main categories, small molecule antagonists, G4 aptamers and targeted degradation of G4BPs.

Small molecule antagonists of G4BPs

It has been reported that low concentrations of G4 ligands do not interfere with G4-protein interactions, but form a triple co-binding complex. However, high concentrations of G4 ligands compete with G4BP to bind G4 structures, thus affecting the interaction of G4BP with G4. In vitro experiments have demonstrated that many G4 ligands, such as carboxyPDS and QUMA-1, disrupt the interaction of G4BP with G4 structures[50]. In living cells, G4 ligands can regulate biological processes by interfering with G4-protein interactions[33] (**Figure 2A**). For instance, Telomestatin can disrupt the binding of TRF2 to the telomere in HeLa cells, leading to telomere disfunction[54]. Another G4 binding molecule, TMPyP4, inhibits the RecQ helicase to unwind DNA G4 structure[55]. Similarly, N-methyl mesoporphyrin IX (NMM) can form an NMM-G4 complex and subsequently trap the helicase such as BLM and Sgs1p[55]. In addition, Drygin et al. found that the G4 ligand, CX-3543, disrupts the formation of nucleolin-rDNA G4 complexes, which inhibits RNA polymerase I-mediated transcription, leading to apoptosis of cancer cells[56]. Therefore, small molecule antagonists of G4BP have broad prospects in disease therapeutics, but the detailed mechanism still needs further study.

G4 aptamers

In addition to G4 ligands, L-RNA aptamers can also regulate G4-protein interactions through a competitive mechanism (**Figure 2B**). L-RNA aptamers are a class of RNA aptamers

consisting of L-nucleotides that are mirror images of natural D-RNAs. Due to their different chirality from natural nucleic acid molecules, they cannot be recognized by intracellular nucleases and therefore have better stability and longer half-life in vivo. Therefore, L-RNA aptamer is considered as a potential nucleic acid drug. Kwok's group developed the RNA G-quadruplex-SELEX technique to efficiently screen for L-RNA aptamers that bind to a specific RNA G4 structure[50, 57]. The identified aptamer Ap3-7 was shown to block the binding of the G4BP RHAU53 to TERRA rG4, while the identified aptamer L-Apt.4-1c was shown to block the binding of the G4BP nucleolin to hTERC rG4, as demonstrated by EMSA experiments. Further, they found that the stability of L-Apt.4-1c in the physiological state could be improved without losing the binding ability and specificity after the tail end of L-Apt.4-1c was ligated into a loop by click reaction[58]. Moreover, Zhao et al. developed an L-RNA aptamer, L-Apt.8f, that specifically binds to APP 3'-UTR rG4, and verified that this aptamer can be used for translation regulation of reporter genes in living cells[59].

Targeted Degradation of G4BPs

PROTAC (proteolysis-targeting chimeras) is a technique that uses small molecules to regulate the degradation of target proteins by the ubiquitin-proteasome system (UPS)[60]. This strategy can also be applied to the regulation of G4-protein interactions. Patil et al. developed a G4BP degradation method, G4-PROTAC, for the degradation of the G4BP DHX36[61] (**Figure 2C**). G4-PROTAC uses a small molecule consisting of two parts, a G4-forming DNA and an E3 ubiquitin ligase recruiter. The G4-forming DNA is used to recruit DHX36 protein, while the E3 ubiquitin ligase recruiter, similar to other PROTAC techniques, is used to recruit E3 and E2 ubiquitin ligases to degrade the target protein. The results showed that G4-PROTAC can efficiently degrade DHX36 in living cells. The G4-PROTAC strategy has a great potential in the regulation of G4-protein interactions. The G4 structure for recruiting G4BPs consists of specific nucleic acid sequences. Since G4BPs have a preference on the binding of different G4-forming sequences, a finely designed G4-PROTAC molecule has the potential to specifically regulate the G4-protein interactions on the G4 structure of an individual gene.

Future perspective of G4BP regulation

Intracellular regulation of G4-protein interactions has been achieved based on competitive inhibition mechanisms and targeted degradation mechanisms. However, the specificity of these regulatory tools is still insufficient and further development and optimization are urgently needed[51]. G4 ligands based on competitive inhibition mechanisms do not have the ability to discriminate between different G4-forming sequences, while L-RNA aptamer, although showing a stronger binding preference for certain G4 structures (cycL-Apt.4-1c has the specificity to target hTELO-rG4), still can bind to other G4 structures. Recently, Liu et al. used phage display method to screen for bicyclic peptide G4 ligands that specifically bind to G4 structures[62]. Similar to the recognition of nucleic acids by proteins, the targeted recognition of G4 by bicyclic peptide ligands is based on multivalent interactions in the binding cavity. Bicyclic peptide G4 ligands are expected to be able to, modulate G4-protein interactions on an individual G4 structure of a specific gene, but future works are demanded to screen the bicyclic peptide. There are also new advances in small molecule G4 ligands targeting an individual G4 structure. Recently, Yang's group has developed a small molecule ligand targeting MYC-G4, which provides a molecular basis for G4-protein interaction regulation in a specific gene region[37]. In addition, regulating G4-protein interactions on a specific G4 structure in living cells is more easily achieved by G4-PROTEC, but its effect on the biological pathways remains unclear. In the G4-PROTEC method, if the G4-forming nucleic acid sequence is used as a bait for recruiting G4BPs, the efficient delivery and long-time stability of the nucleic acid-E3 ligase ligand complex is a challenge. If a small molecule ligand is used as a bait, the G4-protein interactions may be affected by the binding of the ligand to G4s, thus reducing the efficiency of ubiquitination of G4BPs. These issues need to be addressed in future studies. With the continuous improvement of G4-protein interaction regulatory tools, the regulation of G4-protein interactions targeting the G4 structure of disease-related genes is likely to provide new ideas for disease treatment[63].

Conclusion Remarks

In the past decade, the study of biochemical methods for G4BPs has made tremendous progress, achieving unprecedented throughput, sensitivity and resolution, and greatly contributing to the biological study of G4BPs. It is becoming increasingly clear that G4-

protein interactions should be considered as a fundamental regulatory mechanism of cellular biological processes. G4BP is implicated in many important biological processes, particularly transcription, RNA splicing and maintenance of genomic stability. The key challenge for the future is to elucidate the regulatory network of the G4-protein interactome, especially at telomeres, promoters and RNA splicing sites, and validate the specific mechanisms of their involvement in biological regulation. To achieve these goals, G4BP biochemical methods still need to be improved and developed (see **Outstanding questions**), as has been discussed in detail in this review. The emergence of more powerful biochemical tools, including the discovery, identification, and regulation of G4BP, will lead to a clearer understanding of the role of G4-protein interactions in disease, providing new ideas for the treatment of tumors and viral infections, and ultimately their clinical application.

Glossary

Flow-through: The components of the sample after passing through the chromatographic column.

G-tetrad: The four guanines are arranged into a square planar configuration through the Hoogsteen hydrogen bonds between adjacent guanines, which is called G-tetrad.

G-quadruplex ligands: Aromatic planar compounds that can specifically bind to DNA and/or RNA G-quadruplex.

Label-free quantitative analysis：A quantitative analysis method of protein based on the correlation between protein abundance and the intensity of peptide peak in the primary spectrum or the number of secondary spectrum of mass spectrometry data.

Photo-crosslinking: A photochemical reaction process in which photosensitive groups produce free radicals under light and covalently cross link with adjacent biomacromolecules.

Box 1. Overview of G4-binding proteins

In 2000, Mergny's group reported that human DNA topoisomerase I (TOP1) can bind to G4 structures[64] (**Figure IA**), opening the research on G4BPs. Subsequently, G4BP has attracted increasing interest. New G4BPs have been reported every year, especially in the last 5 years (**Figure IB**). G4BP is emerging as a promising field in chemical biology.

There are two main mechanisms by which G4BPs interact and function with intracellular G4s[65]. One is to unfold the G4 structure after binding to it. For instance, telomere repeat sequences can form stable G4 structures and impede telomere maintenance, thus some G4-unfolding proteins located in telomere region are essential to maintain telomere homeostasis and integrity[66]. Based on a similar mechanism, some G4-binding DNA helicases help to remove DNA lesions caused by G4 structures and promote the correct replication of G4-containing DNA regions[67]. The other interaction mechanism is to be recruited to specific functional regions of the DNA or RNA strand by binding to the G4 structure. The most important example is G4-binding transcription factors. Both bioinformatic prediction analysis and G4 ChIP-seq proved that G4s are enriched in the promoter regions of human genome, and these G4s serve as transcription factor binding hubs to mold cell-type specific transcriptome[13, 68].

We summarized all experimentally identified G4BPs in **Table S1**. Among the 105 validated G4BPs, approximately one third were G4-unfolding proteins and the others were G4-recruited proteins (**Figure IC**). These proteins are involved in numerous biological processes, mainly including transcription, RNA splicing and translation (**Figure ID**). The biological processes related to G4BPs have been reviewed in detail elsewhere[33, 65].

Box 2. The relationship between G4-binding proteins and diseases

Since G4BP is involved in the formation of G4 structures and mediates biological functions based on G4 structures, dysregulation of these proteins is closely associated with many diseases such as tumors, viral infections and neurodegenerative diseases[69]. Therefore, the extensive research on G4BP will contribute to the discovery of new therapeutic targets[70].

Many G4-recruited proteins play key roles in DNA replication, DNA repair and chromosome homeostasis, because G4 structures can cause DNA lesions and replication errors. If these G4BPs function abnormally, the genomic instability caused by G4 can accumulate to form oncogenic gene variations[71]. For example, the proto-oncogene p53 binds to the G4 structure in the promoter region[72], which in turn regulates the expression of cell cycle-related genes, thereby promoting the proliferation of tumor cells with high p53 expression. In addition, the high expression of the G4-unfolding helicase WLN and BLM in tumor cells allows the DNA lesions caused by G4 to be cleared, which is related to the unlimited proliferative capacity of tumor cells[73-76].

The presence of G4 structures has been demonstrated in the DNA or RNA of many kinds of viruses, such as human immunodeficiency virus (HIV)[77], human papillomavirus (HPV)[78], herpes simplex virus (HSV)[79], hepatitis C virus (HCV)[80], Epstein-Barr virus (EBV)[81] and SARS-COV-2[82]. G4BPs also have important functions in the life cycle of these viruses, which has been summarized in detail[63]. For example, the mRNA of Epstein−Barr nuclear antigen 1 (EBNA1) contains 13 repeated G4 structures, which can interact with linking region 1 (LR1) and linking region 2 (LR2) to mediate viral DNA replication[83]. It has been demonstrated that disrupting this interaction by G4 ligand BRACO-19 can effectively inhibit EBV replication. The nucleocapsid protein of HIV-1 unfolds the G4 structures in the viral genomic RNA, promoting the initiation of viral reverse transcription[84]. Targeting the G4-protein interaction in viruses is valuable for potentially antiviral therapeutics[63].

Finally, G4BP is also closely associated with neurodegenerative diseases. For example, the mRNA of genes associated with Parkinson’s disease contains stable G4 structures in the 5’ UTR regions[85]. These G4s are bound with guanine nucleotide-binding protein-like 1 (GNL1), which may inhibit the translation of the corresponding mRNA. In addition, G4-unfolding protein FMRP has an obvious low expression in Fragile X syndrome[86, 87], which are further demonstrated to interfere with the transport and translation of some RNAs.

Acknowledgements

This work was financially supported by National Key Research and Development Program of China (No. 2021YFA1200104), National Natural Science Foundation of China (No. 22027807, No. 22034004), the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDB36000000) and Vanke Special Fund for Public Health and Health Discipline Development (No. 2022Z82WKJ003).

Declaration of interests

The authors declare no competing financial interests.

References

1 Millevoi, S., *et al.* (2012) G-quadruplexes in RNA biology. *Wiley Interdiscip Rev RNA* 3, 495-507
2 Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* 43, 8627-8637
3 Fay, M.M., *et al.* (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J Mol Biol* 429, 2127-2147
4 Hansel-Hertsch, R., *et al.* (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 18, 279-284
5 Dumas, L., *et al.* (2021) G-Quadruplexes in RNA Biology: Recent Advances and Future Directions. *Trends Biochem Sci* 46, 270-283
6 Robinson, J., *et al.* (2021) DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res* 49, 8419-8431
7 Varshney, D., *et al.* (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* 21, 459-474
8 Collie, G., *et al.* (2009) Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem Commun (Camb)*, 7482-7484
9 Yang, S.Y., *et al.* (2021) G-quadruplexes mark alternative lengthening of telomeres. *NAR Cancer* 3, zcab031
10 Kanoh, Y., *et al.* (2015) Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat Struct Mol Biol* 22, 889-897
11 Guilbaud, G., *et al.* (2017) Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat Chem* 9, 1110-1117
12 Li, L., *et al.* (2021) YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* 17, 161-168
13 Lago, S., *et al.* (2021) Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nature communications* 12, 3885
14 Georgakopoulos-Soares, I., *et al.* (2022) Alternative splicing modulation by G-quadruplexes. *Nature communications* 13, 2404

15 Wolfe, A.L., *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* 513, 65-70
16 Arora, A., *et al.* (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* 14, 1290-1296
17 Kumari, S., *et al.* (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* 3, 218-221
18 Gebauer, F., *et al.* (2021) RNA-binding proteins in human genetic disease. *Nat Rev Genet* 22, 185-198
19 Sissi, C., *et al.* (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie* 93, 1219-1230
20 Qiu, J., *et al.* (2015) Biological Function and Medicinal Research Significance of G-Quadruplex Interactive Proteins. *Curr Top Med Chem* 15, 1971-1987
21 Zhang, T., *et al.* (2012) Capture and identification of proteins that bind to a GGA-rich sequence from the ERBB2 gene promoter region. *Anal Bioanal Chem* 404, 1867-1876
22 Mori, K., *et al.* (2013) hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol* 125, 413-423
23 von Hacht, A., *et al.* (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* 42, 6630-6644
24 Pagano, B., *et al.* (2015) Identification of novel interactors of human telomeric G-quadruplex DNA. *Chem Commun (Camb)* 51, 2964-2967
25 Herdy, B., *et al.* (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res* 46, 11592-11604
26 Gonzalez, V., *et al.* (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J Biol Chem* 284, 23622-23635
27 Williams, P., *et al.* (2017) Identification of SLIRP as a G Quadruplex-Binding Protein. *J Am Chem Soc* 139, 12426-12429
28 Serikawa, T., *et al.* (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* 144, 169-184
29 Gao, Z., *et al.* (2021) A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* 20, 4919-4924
30 Zhang, X., *et al.* (2021) Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat Chem* 13, 626-633
31 Su, H., *et al.* (2021) Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J Am Chem Soc* 143, 1917-1923
32 Biffi, G., *et al.* (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* 5, 182-186
33 Sun, Z.Y., *et al.* (2019) Developing Novel G-Quadruplex Ligands: from Interaction with Nucleic Acids to Interfering with Nucleic Acid(-)Protein Interaction. *Molecules* 24

34 Di Antonio, M., *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem* 12, 832-837
35 Jumper, J., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589
36 Brazda, V., *et al.* (2018) The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* 23
37 Dickerhoff, J., *et al.* (2021) Structural recognition of the MYC promoter G-quadruplex by a quinoline derivative: insights into molecular targeting of parallel G-quadruplexes. *Nucleic Acids Res* 49, 5905-5915
38 Riccardi, C., *et al.* (2022) Fighting the Huntington's Disease with a G-Quadruplex-Forming Aptamer Specifically Binding to Mutant Huntingtin Protein: Biophysical Characterization, In Vitro and In Vivo Studies. *International journal of molecular sciences* 23
39 Chen, L.Y., *et al.* (2012) The human CST complex is a terminator of telomerase activity. *Nature* 488, 540-544
40 Mendoza, O., *et al.* (2015) A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands. *Nucleic Acids Res* 43, e71
41 Cheng, Y., *et al.* (2021) Characterization of G-Quadruplexes Folding/Unfolding Dynamics and Interactions with Proteins from Single-Molecule Force Spectroscopy. *Biomolecules* 11
42 Patrick, E.M., *et al.* (2020) Observation of processive telomerase catalysis using high-resolution optical tweezers. *Nat Chem Biol* 16, 801-809
43 You, H., *et al.* (2017) RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res* 45, 206-214
44 Patra, S., *et al.* (2021) Fast interaction dynamics of G-quadruplex and RGG-rich peptides unveiled in zero-mode waveguides. *Nucleic Acids Res* 49, 12348-12357
45 Lago, S., *et al.* (2018) Surface Plasmon Resonance kinetic analysis of the interaction between G-quadruplex nucleic acids and an anti-G-quadruplex monoclonal antibody. *Biochim Biophys Acta Gen Subj* 1862, 1276-1282
46 Kazemier, H.G., *et al.* (2017) Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. *Nucleic Acids Res* 45, 5913-5919
47 McRae, E.K.S., *et al.* (2017) On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. *J Nucleic Acids* 2017, 9675348
48 Chen, M.C., *et al.* (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* 558, 465-469
49 Dai, Y., *et al.* (2022) Single-Cell Visualization of Monogenic RNA G-quadruplex and Occupied G-quadruplex Ratio through a Module-Assembled Multifunctional Probes Assay (MAMPA). *Angewandte Chemie* 61, e202111132
50 Umar, M.I. and Kwok, C.K. (2020) Specific suppression of D-RNA G-quadruplex-protein interaction with an L-RNA aptamer. *Nucleic Acids Res* 48, 10125-10141
51 Zhong, W. and Hooley, R.J. (2022) Combining Excellent Selectivity with Broad Target Scope: Biosensing with Arrayed Deep Cavitand Hosts. *Acc Chem Res* 55, 1035-1046
52 Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting

genomes. *Nature biotechnology* 32, 347-355
53 Petersen, L.K., *et al.* (2021) Screening of DNA-Encoded Small Molecule Libraries inside a Living Cell. *J Am Chem Soc* 143, 2751-2756
54 Tahara, H., *et al.* (2006) G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene* 25, 1955-1966
55 Wu, X. and Maizels, N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res* 29, 1765-1771
56 Drygin, D., *et al.* (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res* 69, 7653-7661
57 Umar, M.I., *et al.* (2022) Development of RNA G-quadruplex (rG4)-targeting L-RNA aptamers by rG4-SELEX. *Nat Protoc* 17, 1385-1414
58 Ji, D., *et al.* (2021) Circular L-RNA aptamer promotes target recognition and controls gene activity. *Nucleic Acids Res* 49, 7280-7291
59 Zhao, H., *et al.* (2022) Novel L-RNA Aptamer Controls APP Gene Expression in Cells by Targeting RNA G-Quadruplex Structure. *ACS applied materials & interfaces* 14, 30582-30594
60 Li, K. and Crews, C.M. (2022) PROTACs: past, present and future. *Chemical Society reviews* 51, 5214-5236
61 Patil, K.M., *et al.* (2021) G4-PROTAC: targeted degradation of a G-quadruplex binding protein. *Chem Commun (Camb)* 57, 12816-12819
62 Liu, K.C., *et al.* (2020) Affinity-Selected Bicyclic Peptide G-Quadruplex Ligands Mimic a Protein-like Binding Mechanism. *J Am Chem Soc* 142, 8367-8373
63 Zhai, L.Y., *et al.* (2022) Targeting the RNA G-Quadruplex and Protein Interactome for Antiviral Therapy. *J Med Chem* 65, 10161-10182
64 Arimondo, P.B., *et al.* (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res* 28, 4832-4838
65 Brazda, V., *et al.* (2014) DNA and RNA quadruplex-binding proteins. *International journal of molecular sciences* 15, 17493-17517
66 Mendoza, O., *et al.* (2016) G-quadruplexes and helicases. *Nucleic Acids Res* 44, 1989-2006
67 Alavi, S., *et al.* (2021) G-quadruplex binding protein Rif1, a key regulator of replication timing. *J Biochem* 169, 1-14
68 Spiegel, J., *et al.* (2021) G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol* 22, 117
69 Shu, H., *et al.* (2022) G-Quadruplex-Binding Proteins: Promising Targets for Drug Design. *Biomolecules* 12
70 Estep, K.N., *et al.* (2019) G4-Interacting DNA Helicases and Polymerases: Potential Therapeutic Targets. *Curr Med Chem* 26, 2881-2897
71 Cammas, A. and Millevoi, S. (2017) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res* 45, 1584-1595
72 Petr, M., *et al.* (2016) Wild-type p53 binds to MYC promoter G-quadruplex. *Biosci Rep* 36
73 Budhathoki, J.B., *et al.* (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence

of ATP. *Nucleic Acids Res* 42, 11528-11545
74 Lee, S., *et al.* (2020) Recognition and Unfolding of c-MYC and Telomeric G-Quadruplex DNAs by the RecQ C-Terminal Domain of Human Bloom Syndrome Helicase. *ACS Omega* 5, 14513-14522
75 Wu, W.Q., *et al.* (2015) BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* 43, 4614-4626
76 Tang, W., *et al.* (2016) The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum Mol Genet* 25, 2060-2069
77 Piekna-Przybylska, D., *et al.* (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* 53, 2581-2593
78 Tluckova, K., *et al.* (2013) Human papillomavirus G-quadruplexes. *Biochemistry* 52, 7207-7216
79 Artusi, S., *et al.* (2016) Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res* 44, 10343-10353
80 Wang, S.R., *et al.* (2016) A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci Adv* 2, e1501535
81 Murat, P., *et al.* (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat Chem Biol* 10, 358-364
82 Zhao, C., *et al.* (2021) Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angewandte Chemie* 60, 432-438
83 Norseen, J., *et al.* (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J Virol* 83, 10336-10346
84 Butovskaya, E., *et al.* (2019) HIV-1 Nucleocapsid Protein Unfolds Stable RNA G-Quadruplexes in the Viral Genome and Is Inhibited by G-Quadruplex Ligands. *ACS Infect Dis* 5, 2127-2135
85 Turcotte, M.A., *et al.* (2021) Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease. *RNA Biol* 18, 1339-1353
86 Melko, M. and Bardoni, B. (2010) The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. *Biochimie* 92, 919-926
87 Goering, R., *et al.* (2020) FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *Elife* 9
88 Alavi, S., *et al.* (2022) SPR Analysis of SUMO-Murine Rap1-Interacting Factor 1 C-Terminal Domain Interaction with G4. *Biosensors (Basel)* 12
89 Ji, D., *et al.* (2021) Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief Bioinform* 22, 1150-1160
90 Dai, Y.X., *et al.* (2022) Structural mechanism underpinning Thermus oshimai Pif1-mediated G-quadruplex unfolding. *EMBO Rep* 23, e53874

Figure I G-quadruplex and G4-binding proteins. A, Structure of a G-quartet formed by the Hoogsteen Hydrogen-bonded guanines and schematic representation of a G-quadruplex. **B**, The number of validated G4BPs in every year from 1999 to 2022. **C**, Pie

chart of two main mechanisms by which G4BPs interact and function with intracellular G4s. **D**, Histogram of biological process in which G4BPs involved.

Figure 1 Bioanalytical methods for discovering new G4BPs. A, Schematic diagram of affinity enrichment methods for screening G4BPs. **B**, Schematic diagram of the approach to exclude nonspecific nucleic acid binding proteins with a G4-mut affinity column and a G4 affinity column. **C**, Schematic diagram of stable isotope labeling by amino acid in cell culture (SILAC)-based quantitative proteomic method to screening G4BPs. **D**, Schematic diagram of G4 ligand-mediated photo-crosslinking protein assay for labeling and identifying G4BPs.

Figure 2 The molecular tools for regulating G4-protein interactions in living cells. A, Structures of G4 ligands that can inhibit G4-protein interactions, including Telomestatin, TMPyP4, *N*-methyl mesoporphyrin Ⅸ(NMM), and CX-3543. **B**, Schematic diagram of L-Apt.4-1c-induced interference with hTERC rG4-RHAU53 peptide interactions. **C**, Schematic diagram showing the mechanism of RHAU degradation by G4-PROTAC.

Table 1 Affinity enrichment methods in the discovery of G4BPs.

Strategy for excluding non-specific nucleic acid binding proteins	**G4 type**	**Bait G4 sequence**	**Protein source**	**Identified proteins**	**Reference**
The cell-extracted proteins were first passed through a non-G4 single-stranded DNA affinity column, then the flow-through protein mixture was applied to G4 DNA affinity column.	DNA G4	human c-MYC G4	HeLa whole-cell extract	10	[26]
The identified proteins were filtered according to Fold Change > 2 (at least a		ERBB2 promoter G4	MCF-7 cell nuclear extract	10	[21]

twofold stronger binding to the G4 strand than to the non-G4 strand).		26-mer truncation of human telomeric DNA sequence	HCT116 cell nuclear extract	19	[24]
Forward stable isotope labeling by amino acid in cell culture (SILAC)-based quantitative proteomic method.		c-KIT, c-MYC and human telomere G4 sequence	HeLa cell nuclear extract	1	[27]
		c-KIT, c-MYC and human telomere G4 sequence		81	[29]
The identified proteins were filtered according to Fold Change > 2 (at least a twofold stronger binding to the G4 strand than to the non-G4 strand).	RNA G4	GGGGCC repeat RNA	HEK293 cell nuclear extract	20	[22]
		MMP16 and ARPC2 mRNA G4	HEK293 and HeLa cell nuclear extract	25	[23]
		human NRAS mRNA G4	HeLa cell cytosolic extract	80	[25]
Forward stable isotope labeling by amino acid in cell culture (SILAC)-based quantitative proteomic method.		BCL2, NRAS, MMP16 and ARPC2 mRNA G4	HEK293 cell nuclear extract	12	[28]

Table 2 Bioanalytical methods for characterizing G4-binding proteins

Method	Information of G4-binding proteins	Features and limitations	Reference
In vitro biophysical methods			

Electrophoretic mobility shift assay (EMSA)	The formation of G4-protein complexes	● Inexpensive, simple to perform; ● Need high-purity G4BPs; ● Limited to in vitro studies.	[26, 31, 39]
Dimethyl sulfate (DMS) foorprinting assay	The formation of G4-protein complexes	● Inexpensive, simple to perform; ● Need high-purity G4BPs; ● Limited to in vitro studies.	[26, 39]
Fluorescence resonance energy transfer (FRET)	The disruption of G4 structures mediated by G4-unfolding proteins	● High sensitivity in the investigation of G4-unfolding proteins; ● Need high-purity G4BPs and fluorophore conjugated nucleic acid strands; ● Limited to in vitro studies.	[40]
Affinity enrichment	The formation of G4-protein complexes	● Not dependent on purified G4BPs; ● Dependent on high quality antibodies; ● May have false positive caused by indirect interactions.	[30]
Single-molecule methods			
Optical tweezer	Kinetics of G4-protein interactions	● Detection G4-protein interaction at single-molecule level; ● Need high-purity G4BPs; ● Limited to in vitro studies.	[42]
Magnetic tweezer	Kinetics of G4-protein interactions	● Detection G4-protein interaction at single-molecule level; ● Need high-purity G4BPs; ● Limited to in vitro studies.	[43]
Zero-mode waveguide (ZMW)	Kinetics of G4-protein interactions	● Detection G4-protein interaction at single-molecule level; ● Need high-purity G4BPs; ● Limited to in vitro studies.	[44]
Methods for detecting the affinity constant of G4BP			
Enzyme-linked immunosorbent assay (ELISA)	The binding constant of G4-protein complexes	● Dependent on high quality antibodies;	[32]
Surface plasma resonance (SPR)	The binding constant of G4-protein	● Need high-purity G4BPs;	[45, 88]

	complexes		
Micro-scale thermophoresis (MST)	The binding constant of G4-protein complexes	• Need high-purity G4BPs;	[46, 89]
Nuclear magnetic resonanace (NMR)	3D structure of G4-protein complexes	• Provide exact 3D structure of G4BPs; • Need high-purity G4BPs;	[48]
X-Ray Crystallography	3D structure of G4-protein complexes	• Provide exact 3D structure of G4BPs; • Need high-purity G4BPs;	[90]
Intracellular characterization method			
Module assembled multifunctional probes assay (MAMPA)	Occupied G4 Ratio of an individual G4 motif in a specific RNA	• Single-cell imaging of RNA G4 located in an individual gene; • Low throughput.	[49]

G-Quadruplex-Protein Interactions: Screening, Characterization and Regulation

Yicong Dai[a,b], Xucong Teng[a,b], Qiushuang Zhang[a,b], Hongwei Hou[c,*], Jinghong Li[a,b,*]

a Department of Chemistry, Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology, Tsinghua University, Beijing 100084, China

b Center for BioAnalytical Chemistry, Hefei National Laboratory of Physical Science at Microscale, University of Science and Technology of China, Hefei 230026, China

c China National Tobacco Quality Supervision and Test Center, Zhengzhou 450001, China

*Correspondence: jhli@mail.tsinghua.edu.cn (J.H. Li), houhw@ztri.com.cn (H.W. Hou)

Outstanding questions

1. In order to study the function and mechanism of a G4 structure of a specific gene, can we develop high-sensitive biochemical methods for labeling G4BPs that bind to a unique G4 motif in living cells?
2. What methods can be used to identify G4PBs without disruptions on G4-protein interaction state?
3. Can deep learning prediction methods help to discover new G4BPs?
4. Can we achieve high-throughput analysis of the G4-protein interactome in living cells?
5. How can we achieve intracellular regulation of G4-protein interactions on an individual G4 structure of a specific gene?
6. What kind of regulatory tools can be used to modulate G4-protein interactions with efficient delivery and long-time stability, which has potential to be applied in disease treatment?

G-Quadruplex-Protein Interactions: Screening, Characterization and Regulation

Yicong Dai[a,b], Xucong Teng[a,b], Qiushuang Zhang[a,b], Hongwei Hou[c,*], Jinghong Li[a,b,*]

a Department of Chemistry, Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology, Tsinghua University, Beijing 100084, China

b Center for BioAnalytical Chemistry, Hefei National Laboratory of Physical Science at Microscale, University of Science and Technology of China, Hefei 230026, China

c China National Tobacco Quality Supervision and Test Center, Zhengzhou 450001, China

*Correspondence: jhli@mail.tsinghua.edu.cn (J.H. Li), houhw@ztri.com.cn (H.W. Hou)

Highlights

- Affinity enrichment methods and G4 ligand-mediated photo-crosslinking methods can be used to screen new G4BPs with high throughput and accuracy.
- Many in vitro methods for the validation of G4BP have been developed based on some traditional nucleic acid-protein interaction assay, but methods for investigating the G4-protein interactions in cellular milieu are still under demand.
- Small molecule antagonists of G4BP and G4 aptamers are powerful tools for the regulation of G4BP in living cells.
- PROTAC (proteolysis-targeting chimeras) can now be used in the degradation of G4BP.

Table S1. Experimentally validated G4-binding proteins.

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
AFF2	G4-recruited	RNA splicing	2009	[1]
APE1	G4-recruited	DNA oxidation and DNA repair	2020	[2-4]
ATRX	G4-recruited	Chromatin remodeling	2010	[5, 6]
BLM	G4-unfolding	DNA replication	2014	[7-9]
BRCA2	G4-recruited	Telomere maintenance	2022	[10]
BRIP1	G4-unfolding	DNA unwinding	2012	[11-13]
CNBP	G4-recruited	Transcription	2008	[14]
CTC1	G4-unfolding	DNA replication	2019	[15]
DDX1	G4-unfolding	DNA unwinding	2021	[16]
DDX17	G4-unfolding	RNA splicing	2014	[17, 18]
DDX21	G4-unfolding	Translation	2017	[19]
DDX24	G4-unfolding	RNA unwinding	2021	[16]
DDX3X	G4-unfolding	DNA unwinding	2018	[17]
DDX5	G4-unfolding	Transcription	2014	[17]
DHX36	G4-unfolding	Translation	2018	[20, 21]
DHX9	G4-unfolding	Translation	2018	[20]
DNAJC2	G4-recruited	DNA repair	2020	[22]
DNA2	G4-unfolding	DNA replication	2008	[23]
DNMT1	G4-recruited	DNA methylation	2016	[24, 25]

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
DNMT3A	G4-recruited	DNA methylation	2016	[25]
DNMT3B	G4-recruited	DNA methylation	2016	[25]
EFHD2	G4-recruited	N. D.	2014	[26]
EIF4G1	G4-recruited	Translation	2020	[27]
EIF4G2	G4-recruited	Translation	2020	[27]
EIF4G3	G4-recruited	Translation	2020	[27]
ERCC2	G4-recruited	DNA repair	2014	[28]
ERCC3	G4-recruited	DNA repair	2014	[28]
EWSR1	G4-recruited	Transcription	2011	[29]
EZR	G4-recruited	Transcription	2021	[30]
FMR1	G4-recruited	RNA splicing	2008	[31-33]
FUS	G4-recruited	Transcription and telomere maintenance	2013	[34-36]
FXR1	G4-recruited	Translation	2018	[17]
FXR2	G4-recruited	Translation	2018	[17]
GNL2	G4-recruited	RNA metabolism	2021	[37]
G3BP1	G4-recruited	N. D.	2021	[30]
GDI2	G4-recruited	N. D.	2021	[30]
GRSF1	G4-recruited	RNA splicing	2018	[17]
HMGA1	G4-recruited	Transcription	2021	[30]
HMGB1	G4-recruited	Telomere maintenance	2015	[38, 39]

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
HNRNPA1	G4-unfolding	RNA splicing	2003	[40-43]
HNRNPA2B1	G4-unfolding	RNA splicing	2007	[44]
HNRNPA3	G4-unfolding	RNA splicing	2013	[45]
HNRNPF	G4-recruited	RNA splicing	2017	[46]
HNRNPH1	G4-unfolding	RNA splicing	2022	[47]
HNRNPL	G4-recruited	N. D.	2013	[45]
HNRNPQ1	G4-unfolding	Translation	2016	[48]
HNRNPU	G4-unfolding	Telomere maintenance	2019	[49]
IFI16	G4-recruited	DNA repair	2016	[50]
IGF1	G4-recruited	Transcription	2019	[51]
IGF2	G4-recruited	Transcription	2006	[52]
IGF2BP1	G4-recruited	N. D.	2013	[45]
ILF2	G4-recruited	Transcription	2013	[45]
ILF3	G4-recruited	Transcription	2013	[45]
KHSRP	G4-recruited	Telomere maintenance	2015	[39]
LMNB1	G4-recruited	Telomere maintenance	2015	[39]
MAZ	G4-recruited	Transcription	2013	[53]
METTL14	G4-recruited	RNA methylation	2022	[54]
MYH9	G4-recruited	N. D.	2009	[55]
NCL	G4-recruited	Transcription	2009	[55]

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
NEIL1	G4-recruited	DNA repair	2021	[56]
NME2	G4-recruited	Transcription	2017	[57]
NOA1	G4-recruited	Mitochondrial activity	2013	[58]
NONO	G4-recruited	Transcription	2020	[59]
NPM1	G4-recruited	RNA splicing	2022	[60]
NRF1	G4-recruited	Transcription	2021	[36]
NSUN5	G4-recruited	RNA methylation	2018	[17, 61]
PARP1	G4-recruited	Telomere maintenance	2010	[62]
PCNA	G4-recruited	N. D.	2021	[30]
PDGFA	G4-recruited	Cell signalling	2015	[63]
PDGFB	G4-recruited	Cell signalling	2015	[63]
PIF1	G4-unfolding	DNA unwinding	2001	[64]
POT1	G4-unfolding	DNA unwinding	2005	[65-68]
PURA	G4-recruited	N. D.	2012	[69]
RACK1	G4-recruited	N. D.	2021	[30]
RBM22	G4-recruited	N. D.	2021	[16]
RBM4	G4-recruited	Transcription	2018	[61, 70]
RECQL5	G4-recruited	DNA unwinding	2016	[71]
REV1	G4-unfolding	DNA replication	2014	[72]
RIF1	G4-unfolding	Telomere maintenance	2018	[73, 74]

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
RPA1	G4-unfolding	Telomere maintenance	2006	[75]
SERBP1	G4-recruited	N. D.	2021	[30]
SFPQ	G4-recruited	N. D.	2013	[45]
SLIRP	G4-recruited	N. D.	2017	[76]
SMARCA4	G4-recruited	Transcription	2021	[16]
SP2	G4-recruited	Transcription	2021	[36]
SPATS2L	G4-recruited	N. D.	2021	[30]
SRSF1	G4-recruited	RNA splicing	2014	[26]
STN1	G4-unfolding	DNA replication	2019	[15]
SUB1	G4-unfolding	DNA unwinding	2017	[77]
SYNCRIP	G4-unfolding	RNA splicing	2005	[78]
TARDBP	G4-recruited	RNA transport	2016	[79]
TDRD5	G4-recruited	RNA metabolism	2020	[80]
TDRD7	G4-recruited	RNA metabolism	2020	[80]
TEN1	G4-unfolding	DNA replication	2019	[15]
TERF2	G4-recruited	Telomere maintenance and transcription	2012	[81]
TOP1	G4-unfolding	DNA replication	2000	[82, 83]
TP53	G4-recruited	Transcription	2016	[84]
U2AF2	G4-recruited	RNA splicing	2014	[26]
UHRF1	G4-recruited	N. D.	2021	[16]

Gene Symbol	G4-Protein Interaction Mechanism	G4BP-Involved Biological Process	Reported Year	Reference
VEGFA	G4-recruited	N. D.	2015	[63]
VIM	G4-recruited	Chromatin remodeling	2022	[85]
WRN	G4-unfolding	DNA unwinding	2016	[86]
XRCC5	G4-recruited	DNA repair	2012	[69]
XRCC6	G4-recruited	DNA repair	2012	[69]
YY1	G4-recruited	Chromatin Remodeling	2021	[87]

N.D., not determined.

References

1 Bensaid, M.*, et al.* (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res* 37, 1269-1279

2 Howpay Manage, S.A.*, et al.* (2022) Cysteine Oxidation to Sulfenic Acid in APE1 Aids G-Quadruplex Binding While Compromising DNA Repair. *ACS Chem Biol* 17, 2583-2594

3 Pramanik, S.*, et al.* (2022) The human AP-endonuclease 1 (APE1) is a DNA G-quadruplex structure binding protein and regulates KRAS expression in pancreatic ductal adenocarcinoma cells. *Nucleic Acids Res* 50, 3394-3412

4 Roychoudhury, S.*, et al.* (2020) Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. *Proc Natl Acad Sci U S A* 117, 11409-11420

5 Law, M.J.*, et al.* (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* 143, 367-378

6 Teng, Y.C.*, et al.* (2021) ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nature communications* 12, 3887

7 Budhathoki, J.B.*, et al.* (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res* 42, 11528-11545

8 Lee, S.*, et al.* (2020) Recognition and Unfolding of c-MYC and Telomeric G-Quadruplex DNAs by the RecQ C-Terminal Domain of Human Bloom Syndrome Helicase. *ACS Omega* 5, 14513-14522

9 Wu, W.Q.*, et al.* (2015) BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* 43, 4614-4626

10 Lee, J.*, et al.* (2022) Dynamic interaction of BRCA2 with telomeric G-quadruplexes underlies telomere replication homeostasis. *Nature communications* 13, 3396

11 Castillo Bosch, P.*, et al.* (2014) FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J* 33, 2521-2533

12 Sarkies, P.*, et al.* (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-

quadruplex DNA. *Nucleic Acids Res* 40, 1485-1498
13 Wu, C.G. and Spies, M. (2016) G-quadruplex recognition and remodeling by the FANCJ helicase. *Nucleic Acids Res* 44, 8742-8753
14 Armas, P.*, et al.* (2008) Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. *J Cell Biochem* 103, 1013-1036
15 Zhang, M.*, et al.* (2019) Mammalian CST averts replication failure by preventing G-quadruplex accumulation. *Nucleic Acids Res* 47, 5243-5259
16 Zhang, X.*, et al.* (2021) Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat Chem* 13, 626-633
17 Herdy, B.*, et al.* (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res* 46, 11592-11604
18 Dardenne, E.*, et al.* (2014) RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep* 7, 1900-1913
19 McRae, E.K.S.*, et al.* (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res* 45, 6656-6668
20 Murat, P.*, et al.* (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol* 19, 229
21 Chen, M.C.*, et al.* (2015) Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res* 43, 2223-2231
22 De Magis, A.*, et al.* (2020) Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nature communications* 11, 3907
23 Masuda-Sasa, T.*, et al.* (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *J Biol Chem* 283, 24359-24373
24 Mao, S.Q.*, et al.* (2018) DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* 25, 951-957
25 Cree, S.L.*, et al.* (2016) DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett* 590, 2870-2883
26 von Hacht, A.*, et al.* (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* 42, 6630-6644
27 Lyons, S.M.*, et al.* (2020) eIF4G has intrinsic G-quadruplex binding activity that is required for tiRNA function. *Nucleic Acids Res* 48, 6223-6233
28 Gray, L.T.*, et al.* (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol* 10, 313-318
29 Takahama, K.*, et al.* (2011) Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J* 278, 988-998
30 Su, H.*, et al.* (2021) Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J Am Chem Soc* 143, 1917-1923
31 Bole, M.*, et al.* (2008) Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA. *Mol Biosyst* 4, 1212-1219
32 Melko, M. and Bardoni, B. (2010) The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. *Biochimie* 92, 919-926
33 Goering, R.*, et al.* (2020) FMRP promotes RNA localization to neuronal projections through

interactions between its RGG domain and G-quadruplex RNA sequences. *Elife* 9
34 Takahama, K. and Oyoshi, T. (2013) 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* 135, 18016-18019
35 Takahama, K.*, et al.* (2013) Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem Biol* 20, 341-350
36 Spiegel, J.*, et al.* (2021) G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol* 22, 117
37 Turcotte, M.A.*, et al.* (2021) Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease. *RNA Biol* 18, 1339-1353
38 Amato, J.*, et al.* (2019) Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein. *Nucleic Acids Res* 47, 9950-9966
39 Pagano, B.*, et al.* (2015) Identification of novel interactors of human telomeric G-quadruplex DNA. *Chem Commun (Camb)* 51, 2964-2967
40 Enokizono, Y.*, et al.* (2003) Destruction of quadruplex by proteins, and its biological implications in replication and telomere maintenance. *Nucleic Acids Res Suppl* 3, 231-232
41 Kruger, A.C.*, et al.* (2010) Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *Eur Biophys J* 39, 1343-1350
42 Zhang, Q.S.*, et al.* (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* 12, 1116-1128
43 Hudson, J.S.*, et al.* (2014) Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1. *Biochemistry* 53, 3347-3356
44 Khateb, S.*, et al.* (2007) The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res* 35, 5775-5788
45 Mori, K.*, et al.* (2013) hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol* 125, 413-423
46 Huang, H.*, et al.* (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev* 31, 2296-2309
47 Vo, T.*, et al.* (2022) HNRNPH1 destabilizes the G-quadruplex structures formed by G-rich RNA sequences that regulate the alternative splicing of an oncogenic fusion transcript. *Nucleic Acids Res* 50, 6474-6496
48 Williams, K.R.*, et al.* (2016) hnRNP-Q1 represses nascent axon growth in cortical neurons by inhibiting Gap-43 mRNA translation. *Mol Biol Cell* 27, 518-534
49 Izumi, H. and Funa, K. (2019) Telomere Function and the G-Quadruplex Formation are Regulated by hnRNP U. *Cells* 8, 11
50 Haronikova, L.*, et al.* (2016) IFI16 Preferentially Binds to DNA with Quadruplex Structure and Enhances DNA Quadruplex Formation. *PLoS One* 11, e0157156
51 Chen, H.*, et al.* (2019) Insulin-like growth factor type I selectively binds to G-quadruplex structures. *Biochim Biophys Acta Gen Subj* 1863, 31-38
52 Connor, A.C.*, et al.* (2006) Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J Am Chem Soc* 128, 4986-4991
53 Cogoi, S.*, et al.* (2013) MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice.

Nucleic Acids Res 41, 4049-4064
54 Yoshida, A.*, et al.* (2022) Recognition of G-quadruplex RNA by a crucial RNA methyltransferase component, METTL14. *Nucleic Acids Res* 50, 449-457
55 Gonzalez, V.*, et al.* (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J Biol Chem* 284, 23622-23635
56 Ferino, A. and Xodo, L.E. (2021) Effect of DNA Glycosylases OGG1 and Neil1 on Oxidized G-Rich Motif in the KRAS Promoter. *International journal of molecular sciences* 22
57 Saha, D.*, et al.* (2017) Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J Biol Chem* 292, 15205-15215
58 Al-Furoukh, N.*, et al.* (2013) Binding to G-quadruplex RNA activates the mitochondrial GTPase NOA1. *Biochim Biophys Acta* 1833, 2933-2942
59 Simko, E.A.J.*, et al.* (2020) G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Res* 48, 7421-7438
60 Ghosh, A.*, et al.* (2022) Alternative splicing modulation mediated by G-quadruplex structures in MALAT1 lncRNA. *Nucleic Acids Res* 50, 378-396
61 Serikawa, T.*, et al.* (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* 144, 169-184
62 Cogoi, S.*, et al.* (2010) The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J Biol Chem* 285, 22003-22016
63 Saito, T.*, et al.* (2015) Identification of RNA Oligonucleotides Binding to Several Proteins from Potential G-Quadruplex Forming Regions in Transcribed Pre-mRNA. *Molecules* 20, 20832-20840
64 Bessler, J.B.*, et al.* (2001) The Pif1p subfamily of helicases: region-specific DNA helicases? *Trends Cell Biol* 11, 60-65
65 He, Q.*, et al.* (2014) G-quadruplex-mediated regulation of telomere binding protein POT1 gene expression. *Biochim Biophys Acta* 1840, 2222-2233
66 Zaug, A.J.*, et al.* (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A* 102, 10864-10869
67 Hwang, H.*, et al.* (2012) The Dynamic Interplay Between Telomere-Binding Proteins POT1 and TPP1 in G-Quadruplex DNA. *Biophysical Journal* 102, 15A-15A
68 Chaires, J.B.*, et al.* (2020) Human POT1 unfolds G-quadruplexes by conformational selection. *Nucleic Acids Res* 48, 4976-4991
69 Zhang, T.*, et al.* (2012) Capture and identification of proteins that bind to a GGA-rich sequence from the ERBB2 gene promoter region. *Anal Bioanal Chem* 404, 1867-1876
70 Niu, K.*, et al.* (2019) Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res* 47, 7306-7320
71 Budhathoki, J.B.*, et al.* (2016) A Comparative Study of G-Quadruplex Unfolding and DNA Reeling Activities of Human RECQ5 Helicase. *Biophys J* 110, 2585-2596
72 Eddy, S.*, et al.* (2014) Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res* 42, 3272-3285
73 Alavi, S.*, et al.* (2021) G-quadruplex binding protein Rif1, a key regulator of replication timing. *J Biochem* 169, 1-14
74 Moriyama, K.*, et al.* (2018) Oligomer formation and G-quadruplex binding by purified murine Rif1 protein, a key organizer of higher-order chromatin architecture. *J Biol Chem* 293, 3607-3624

75 Salas, T.R., *et al.* (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res* 34, 4857-4865
76 Williams, P., *et al.* (2017) Identification of SLIRP as a G Quadruplex-Binding Protein. *J Am Chem Soc* 139, 12426-12429
77 Lopez, C.R., *et al.* (2017) Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA. *Nucleic Acids Res* 45, 5850-5862
78 Enokizono, Y., *et al.* (2005) Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J Biol Chem* 280, 18862-18870
79 Ishiguro, A., *et al.* (2016) TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells* 21, 466-481
80 Ding, D., *et al.* (2020) LOTUS domain is a novel class of G-rich and G-quadruplex RNA binding domain. *Nucleic Acids Res* 48, 9262-9272
81 Biffi, G., *et al.* (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J Am Chem Soc* 134, 11974-11976
82 Arimondo, P.B., *et al.* (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res* 28, 4832-4838
83 Keller, J.G., *et al.* (2022) Topoisomerase 1 inhibits MYC promoter activity by inducing G-quadruplex formation. *Nucleic Acids Res* 50, 6332-6342
84 Petr, M., *et al.* (2016) Wild-type p53 binds to MYC promoter G-quadruplex. *Biosci Rep* 36
85 Ceschi, S., *et al.* (2022) Vimentin binds to G-quadruplex repeats found at telomeres and gene promoters. *Nucleic Acids Res* 50, 1370-1381
86 Tang, W., *et al.* (2016) The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum Mol Genet* 25, 2060-2069
87 Li, L., *et al.* (2021) YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* 17, 161-168

Figure I of Box 1

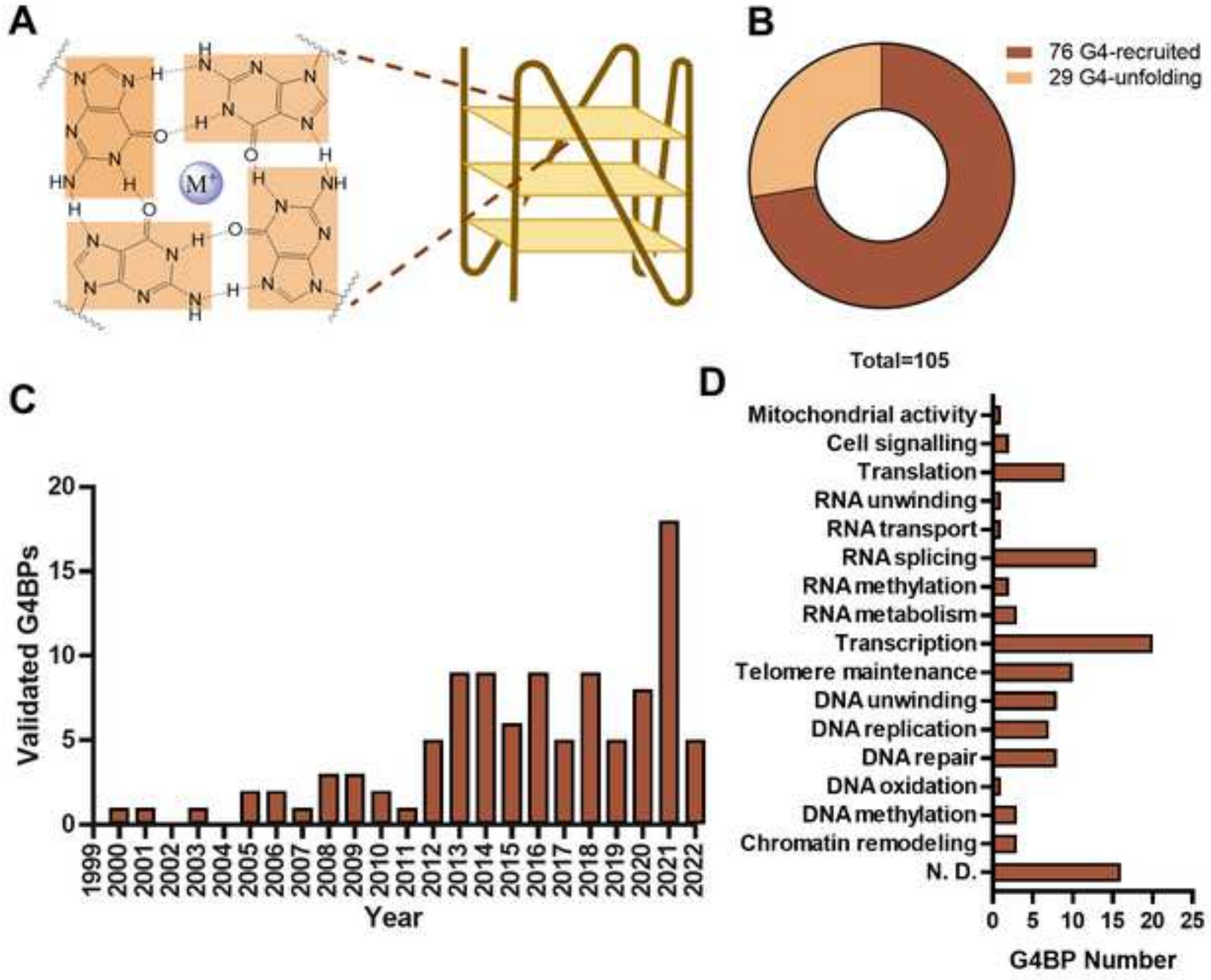


Figure 1

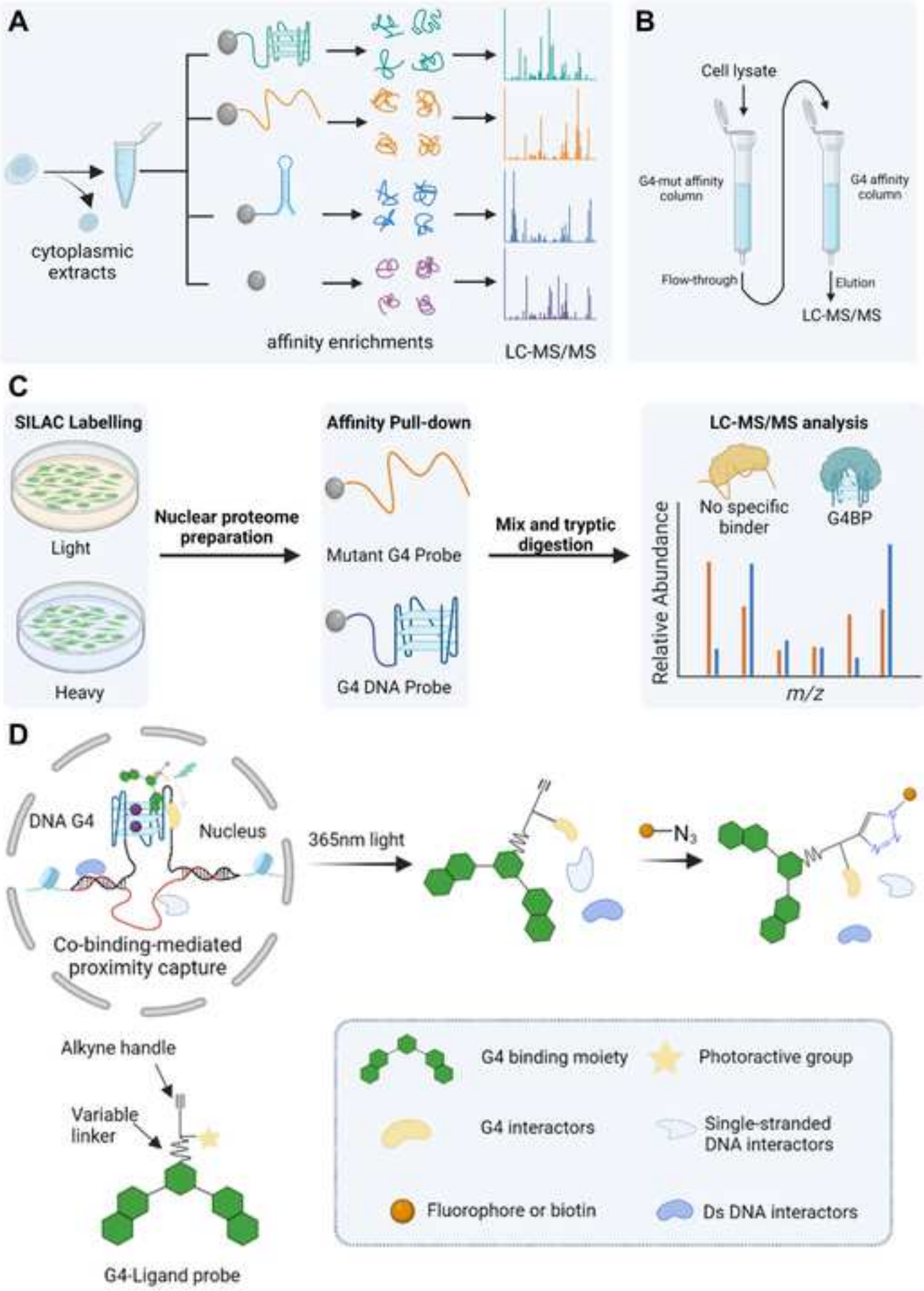
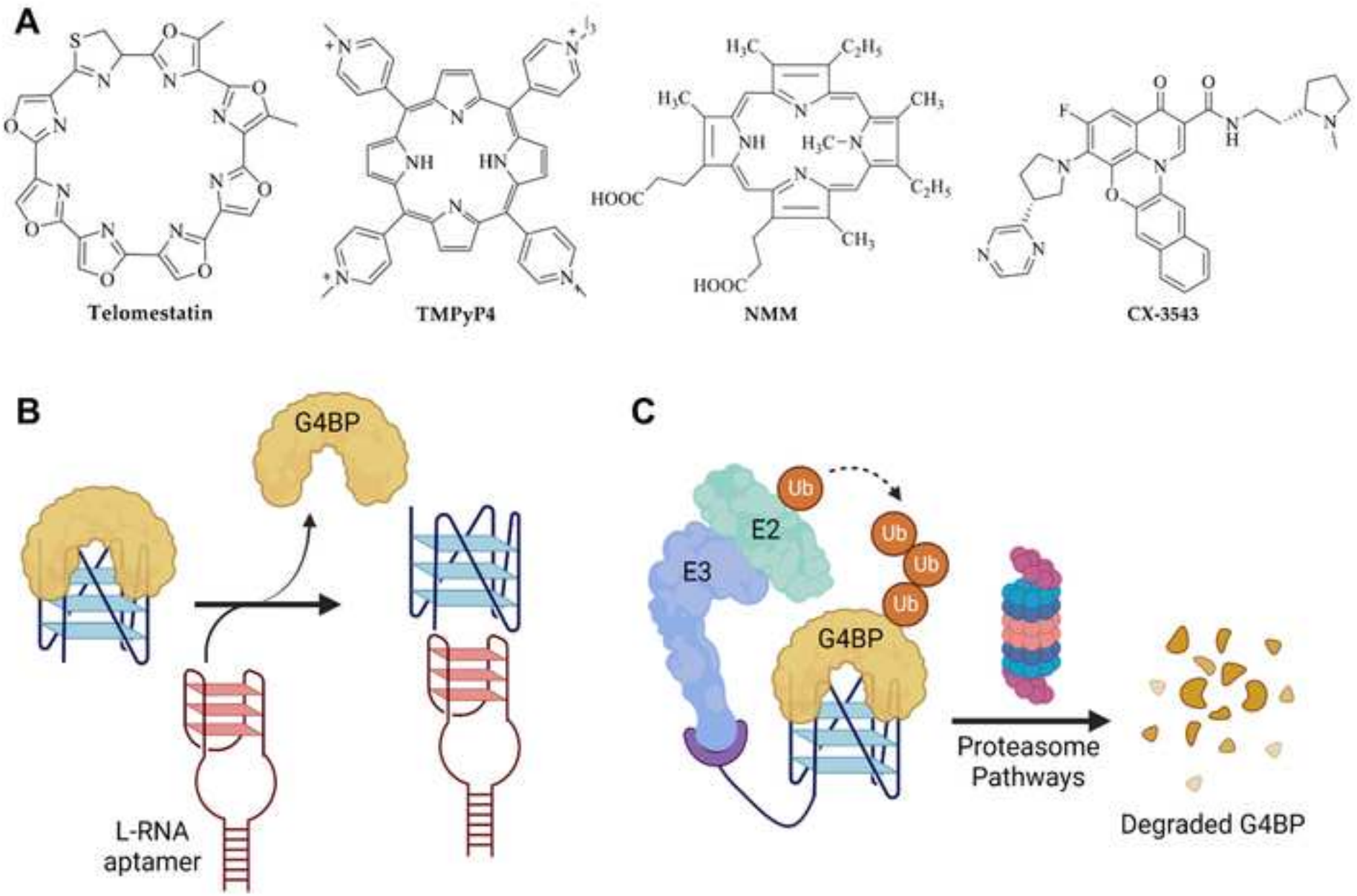

A
cytoplasmic extracts
affinity enrichments
LC-MS/MS
B
Cell lysate
G4-mut affinity column
G4 affinity column
Flow-through
Elution
LC-MS/MS
C
SILAC Labelling
Light
Heavy
Nuclear proteome preparation
Affinity Pull-down
Mutant G4 Probe
G4 DNA Probe
Mix and tryptic digestion
LC-MS/MS analysis
No specific binder
G4BP
Relative Abundance
m/z
D
DNA G4
Nucleus
Co-binding-mediated proximity capture
365nm light
N3
Alkyne handle
Variable linker
G4-Ligand probe
G4 binding moiety
Photoractive group
G4 interactors
Single-stranded DNA interactors
Fluorophore or biotin
Ds DNA interactors

Figure 2


A
Telomestatin
TMPyP4
NMM
CX-3543
B
G4BP
L-RNA
aptamer
C
Ub
E2
E3
G4BP
Proteasome
Pathways
Degraded G4BP