


Review

G-Quadruplex-Binding Proteins: Promising Targets for Drug Design

Huiling Shu, Rongxin Zhang, Ke Xiao, Jing Yang and Xiao Sun * 

State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China; 230228565@seu.edu.cn (H.S.); rongxinzhang@outlook.com (R.Z.); kexiao@seu.edu.cn (K.X.); 220195127@seu.edu.cn (J.Y.)

* Correspondence: xsun@seu.edu.cn

Abstract: G-quadruplexes (G4s) are non-canonical secondary nucleic acid structures. Sequences with the potential to form G4s are abundant in regulatory regions of the genome including telomeres, promoters and 5' non-coding regions, indicating they fulfill important genome regulatory functions. Generally, G4s perform various biological functions by interacting with proteins. In recent years, an increasing number of G-quadruplex-binding proteins have been identified with biochemical experiments. G4-binding proteins are involved in vital cellular processes such as telomere maintenance, DNA replication, gene transcription, mRNA processing. Therefore, G4-binding proteins are also associated with various human diseases. An intensive study of G4-protein interactions provides an attractive approach for potential therapeutics and these proteins can be considered as drug targets for novel medical treatment. In this review, we present biological functions and structural properties of G4-binding proteins, and discuss how to exploit G4-protein interactions to develop new therapeutic targets.

Keywords: G-quadruplex; G-quadruplex-binding protein; drug target; G4; diseases; therapeutics



Citation: Shu, H.; Zhang, R.; Xiao, K.; Yang, J.; Sun, X. G-Quadruplex-Binding Proteins: Promising Targets for Drug Design. *Biomolecules* **2022**, *12*, 648. <https://doi.org/10.3390/biom12050648>

Academic Editors: Pavel Ivanov, Prakash Kharel, Anna Varizhuk and Valentina Pirota

Received: 29 March 2022

Accepted: 27 April 2022

Published: 29 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

G-quadruplex (G4) is a polymorphic and stable secondary nucleic acid structure which is mostly formed in single-stranded DNA or RNA guanine-rich regions [1–3]. The basic structural unit of the G4 is the G-quartet, a square planar assembly of four guanines held together by Hoogsteen hydrogen bonding [4]. Stable G4 structure formation is driven by stacks of 2–4 G-quartets in general and monovalent cations such as Na⁺ and K⁺ in the central channel of G4 helix. G4s could be assembled from the same (intramolecular) or distinct (intermolecular) G-rich strands. G4 structures may fold into diverse topologies, which are dictated by the numbers of stacked G-quartets, the orientation and polarity of the nucleic acid strands and the glycosidic conformation of guanine bases in quartets [3–5]. For example, DNA G4 can form a parallel, anti-parallel or mixed structure, while RNA G4 is more inclined to form a parallel structure and it has higher thermal stability [6,7].

Researchers are able to obtain sequence information with the potential to form G4 structures with an algorithmic analysis of G4 motifs such as G_xN₁₋₇G_xN₁₋₇G_xN₁₋₇G_xN₁₋₇ (x ≥ 3 and N can be any base) [8,9]. However, further experimental approaches are required in order to verify whether these sequences truly form G4 structures. Biophysical methods are commonly used, such as circular dichroism (CD), ultraviolet melting, nuclear magnetic resonance (NMR) spectroscopy and X-rays [3,10–12]. Furthermore, a number of genomic techniques have been developed in recent years, including genome-wide DNA polymerase-stop assay followed by high-throughput sequencing (G4-seq), chromatin immunoprecipitation followed by high-throughput sequencing using BG4 (G4 ChIP-seq) or the use of a small artificial protein G4 probe (G4P) [1,3,13–15]. Notably, the latest technique Cleavage Under Targets and Tagmentation (CUT&Tag) for mapping native G4

in mammalian cells can be applied to single-cell G4 detection [16]. In addition, RNA G4 in the human transcriptome can be detected using reverse transcriptase stalling with next-generation sequencing (rG4-seq) and G4-RNA-specific precipitation (G4RP) with sequencing (G4RP-seq) [17,18]. The distribution information of G4s in the genome and transcriptome also lays a solid foundation for subsequent research on the functions of G4s. Sequences with the potential to form G4 structures (pG4) are present in the genomes of all organisms [5,19]. The location of pG4s is not random in that pG4s occur primarily in the functional regions of the genome and they are highly conserved among different species [5]. Significantly, it was found that pG4s are highly enriched in regulatory regions such as telomeres, gene promoters and the border between introns and exons. Furthermore, recent studies have revealed that 90% of human DNA replication origins contain pG4 motifs, and they also colocalize with the region-specifying 5'-untranslated region (UTR) of the encoded mRNAs among 3000 human genes [5,20]. It can be seen that G4 structures correlate closely with the functions of the genome, and play an important role in DNA replication, transcription, translation, and epigenetic modification. The participation of proteins is necessary for the formation of G4 structures and the fulfillment of their biological functions. These proteins that can specifically bind to G4s are generally referred to as G-quadruplex-binding proteins (G4BPs). Firstly, G4BPs could be classified into two main types according to their control mechanisms and functional relationships with G4s, i.e., G4-folding proteins which have an effect on G4 structures and G4-recruited proteins which are functional proteins recruited by G4 (Figure 1). Secondly, G4BPs can also be divided into the following two categories based on the distribution of G4s in the genome: DNA and RNA G4BPs (Figure 2). To be specific, the structural effects of G4-folding proteins on the G4s lie in two aspects. On one hand, these proteins may promote pG4 to form a stable quadruplex structure; on the other hand, they have the ability to unfold G4s [21–23]. Furthermore, the fulfillment of the biological functions of G4 structures is generally accompanied by the recruitment of G4BPs [10]. Telomere-binding proteins can bind to G4s and unfold these structures at telomeres to maintain the length and integrity of telomeres [24,25]; various helicases and DNA repair proteins recruited to the G4 formation sites unwind G4 structures and remove DNA lesions to ensure efficient DNA replication and maintain genome integrity [21,22,26]; epigenetic modulators regulate the methylation of DNA and histones through their interaction with G4s [27,28]; transcription factors bind to G4s at the promoter sites to aid or suppress gene transcription [29,30]. In addition, G4BPs are also involved in multiple biological processes such as post-transcriptional processing, mRNA maturation and translational regulation (Table 1).

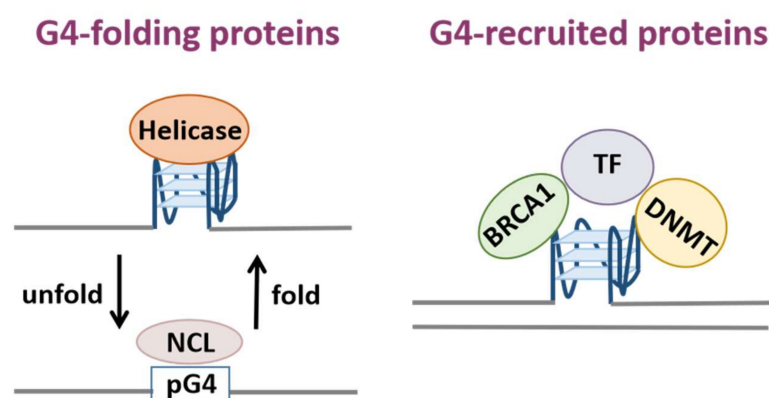
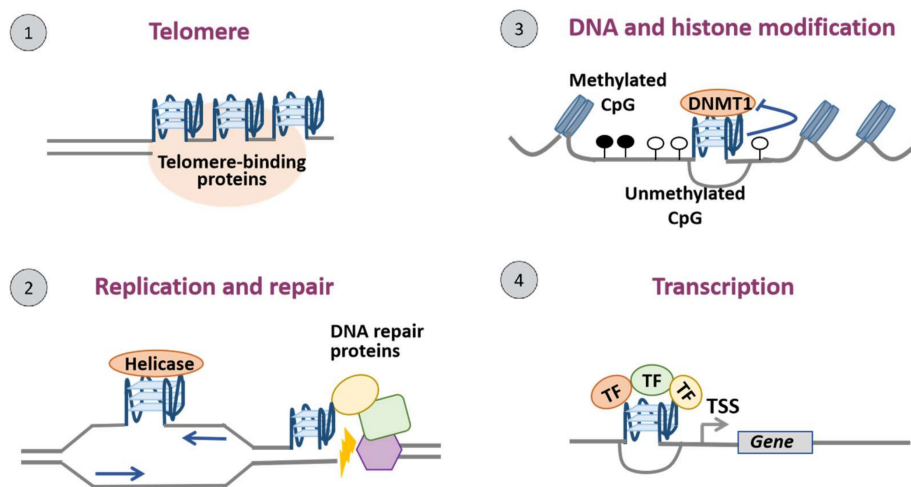


Figure 1. Classification of G-quadruplex-binding proteins. G4BPs can be classified into two categories: G4-folding proteins which can fold or unfold G4s such as nucleolin (NCL) and helicases and G4-recruited proteins which can be recruited by G4 such as transcription factors (TF), DNA repair proteins (BRCA1) and chromatin remodeling proteins (DNA methyltransferase, DNMT).

DNA G4-binding proteins



RNA G4-binding proteins

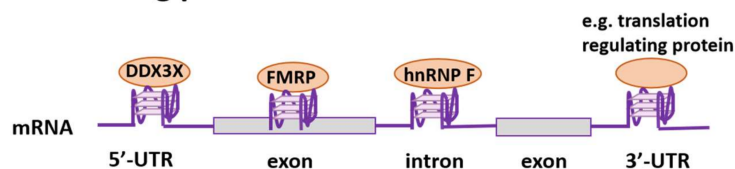


Figure 2. Location and biological functions of G-quadruplex-binding proteins. According to their location, G4BPs can be classified into the following two categories: DNA and RNA G4BPs. In cells, G4BPs perform various biological functions. ① At telomeres, telomere-binding proteins form a ternary complex with the G4 structures of telomeric DNA for telomere maintenance. ② G4s formed during replication present obstacles to the replication machinery and should be resolved by helicases for effective replication. Meanwhile, DNA repair proteins could be recruited to G4 sites to repair DNA double-strand breaks (DSB). ③ In terms of epigenetic control, the inactivation of DNA methyltransferase 1 (DNMT1) when binding G4s results in hypomethylation at CpG islands. ④ In the promoter regions, multiple transcription factors can bind G4s upstream of the transcription start site (TSS) and activate gene transcription. RNA G4BPs regulate mRNA maturation including mRNA export and splicing, and they also function in translational control.

Table 1. In this table, several key G4BPs with different biological functions are summarized. Their location, classification and structural features (if any) are listed.

Location	Biological Role	Protein Name	Classification	Structural Feature	PDB Code	References	
DNA	Telomere maintenance	Shelterin complex	G4-recruited	OB-fold	3EDY	[31–33]	
		CST complex	G4-recruited	OB-fold	5W2L	[34,35]	
	G4 resolution and genome stability maintenance	WRN	G4-folding				[36,37]
		BLM					[38,39]
		FANCI					[40,41]
Transcriptional regulation	NCL	G4-folding	RGG; RRM	2KRR	[42]		
	NM23-H2 p53	G4-recruited			[43] [44]		
Chromatin remodeling and DNA modification		ATRX	G4-recruited			[45]	
		DNMT1				[28]	
RNA	Translational regulation	DDX3X	G4-recruited	RGG	2JGN	[46,47]	
		FMRP		RGG	5DEA	[48]	
		NONO				[49]	

Therefore, the formation, stabilization and resolution of G4 structures *in vivo* should be strictly regulated. A critical step to activate or inactivate physiological or pathological pathways is the recognition and processing of G4s by G4BPs [50]. Analyzing the amino acid composition and the structural properties of G4BPs offers explanation of G4 recognition mechanisms and also provides a structural basis for drug design based on G4BPs [51,52]. The extensive research on spatiotemporal G4BP expression will provide new targets for drug design and pave the way for novel cancer therapy [53,54]. However, there are some difficulties in the analysis of G4-protein recognition. Firstly, the interaction between G4s and proteins *in vitro* may not confirm the existence of such an event *in vivo* due to the plasticity of G4 formation and distinct cellular contexts. Secondly, insufficient three-dimensional structures of G4-protein complexes could be an obstacle to molecular modeling and rational targeting [55]. In this review, we first briefly introduce tools for the detection of G4BPs, then systematically describe DNA and RNA G4BPs and their regulatory roles. We also analyze the structural properties of these proteins in detail and summarize their relationships with multiple diseases. Finally, we discuss how to exploit G4-protein interactions for drug target identification and drug design, thus opening the way to the development of novel therapeutics.

2. G-Quadruplex-Binding Proteins

2.1. Detection of G-Quadruplex-Binding Proteins

The G4 structure is highly dynamic *in vivo* and depends on the cell type and chromatin state [2]. Meanwhile, the formation and unwinding of G4 structures across the whole genome and transcriptome are directly or indirectly regulated by G4BPs, thereby affecting various biological processes [2,3]. Thus, the identification and an in-depth study of G4BPs can provide a full explanation of G4-protein interactions and their biological roles *in vivo*. These studies will further inspire the development of medical applications for these proteins.

G4BPs are most often identified by biochemical experiments. The commonly used methods include affinity chromatography, quantitative methods based on mass spectrometry, and fluorescence energy resonance transfer (FRET) technology. Affinity chromatography is often used in combination with mass spectrometry to separate proteins that bind to specific G4 motifs [56]. For instance, this method was applied to the identification of proteins binding to G4s in the 5' UTR of tumor-associated mRNA [57]. FRET is a spectroscopic technique that provides information about the conformation and dynamics of biomolecules. It has been widely used because this technology can detect whether there is a direct interaction between G4 structures and proteins *in vivo* [58,59]. In addition to performing biochemical experiments, computational analyses could be conducted to identify G4BPs. For example, putative G4 motifs can be predicted at known binding sites of nucleic acid-binding proteins, or the computational modeling of structural features may be exploited to discover new G4BPs [2,51,52,60].

Recently, great advances have been made in the identification of G4BPs. The affinity purification experiments do not take into consideration the native chromatin state, so Shankar Balasubramanian et al. pioneered a co-binding-mediated protein profiling (CMPP) approach for the exploration of DNA G4BPs in living cells [61]. Researchers designed small-molecule ligands that specifically target DNA G4 in cells so that the probes could approach G4BPs with minimal interference with G4-protein interactions and enable labeling by subsequent photoproximity crosslinking [61]. The strategy was employed to identify hundreds of potential G4BPs, and finally *in vitro* experiments confirmed the binding specificity of several candidate proteins. Overall, this approach laid the foundations for the subsequent investigation of new G4BPs.

In conclusion, the detection methods are made up of *in vivo*, *in vitro* and *in silico* approaches. Generally, *in vivo* and *in silico* approaches are employed to identify potential G4BPs while *in vitro* approaches are utilized to confirm G4-protein interactions.

2.2. DNA G-Quadruplex-Binding Proteins

Recent discoveries related to the involvement of DNA G4BPs in the regulation of cellular fundamental functions will be discussed in the following section.

2.2.1. Telomeric G-Quadruplex-Binding Proteins

The human telomeric sequence was one of the first sequences discovered to form G4 structures. Telomeres are nucleoprotein complexes that constitute the ends of eukaryotic chromosomes, which play crucial roles in maintaining the integrity and stability of the genome [10,62]. When specific proteins bind to telomeric DNA, these proteins can prevent not only the degradation of the chromosome ends by nucleases, but also the recognition of them as broken fragments by the DNA repair mechanism [24,25,62].

Telomeric DNA is highly conserved between vertebrates and consists of the identical TTAGGG short repeat sequences with a guanine-rich single-stranded 3' overhang [62]. These repeat sequences have the potential to form a G4 structure. The experiments using G4 specific antibodies and G4 ligands also confirmed the existence of G4 structures at telomeres in vivo [10].

Mammalian telomeric DNA is bound by a protein complex called shelterin, which protects the DNA termini from being considered as damaged and prevents the triggering of the repair mechanism (Figure 3) [31,63,64]. The proteins TRF1 and TRF2 (Telomere Repeat Binding Factor 1 and 2) in shelterin bind to double-stranded telomeric DNA; POT1 (Protection of Telomeres protein 1) binds to 3' overhang of telomeric repeats and regulates the folding and unwinding of the G4 structures with its heterodimeric partner TPP1 (TIN2 Interacting Protein) [10,65–67]. TPP1 connects POT1 to TRF1 and TRF2 via TIN2 (TRF1-interacting Nuclear protein 2) [34]. In addition, the study found that the helicases WRN (Werner syndrome ATP-dependent helicase) and BLM (Bloom syndrome protein) of the RecQ family are recruited to the telomeres and unfold the G4 structures to maintain the integrity of telomeres and ensure telomere replication [68,69]. WRN colocalizes with TRF2 and POT1, and both WRN and BLM can bind to POT1 with high affinity, which indicates that the telomeric DNA-binding proteins are essential for the recruitment of helicases [5].

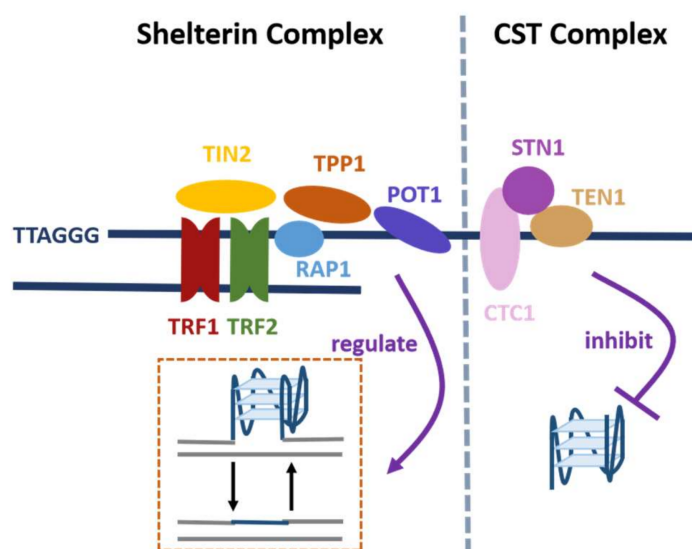


Figure 3. Schematic diagram of the telomere-associated protein complexes shelterin and CST. Shelterin and CST play crucial roles in telomere maintenance. TPP1-POT1 subunit of shelterin regulates the folding and unwinding of G4 structures. CST could resolve and prevent the formation of G4 structures.

Mammalian cells also contain another telomere-associated protein complex called CST (CTC1-STN1-TEN1), which plays a crucial role in efficient telomere replication and in the maintenance of telomere length (Figure 3) [34,70–72]. Human CST is a single-stranded DNA-binding protein complex that helps to solve the genome-wide replication problems [34,71]. For example, GC-rich regions of the genome may induce obstacles to DNA replication, because DNA polymerase may stall at a G4. Experiments confirmed that CST could bind to the G4s and unfold them [34]. G4 structures possibly form on the lagging strand template at telomeres where WRN, BLM and POT1 all participate in G4 removal [5,34,73,74]. However, the presence of CST could make the replication of double-stranded telomeric DNA more effective as this complex unwinds G4 structures more rapidly than POT1 [34].

Other telomere-binding proteins also have similar functions as is the case for these two protein complexes. For example, heterogenous nuclear ribonucleoprotein (hnRNP) A1 and hnRNP A2/B1 form macromolecular complexes with telomere-maintaining factors to regulate telomere length [75]. Given that hnRNP A1 has the ability to unwind the telomeric G4 structures, it can stimulate telomere elongation through the resolution of G4 structures at the end of telomeres [76,77].

2.2.2. G-Quadruplex-Binding Proteins Involved in Replication

The G4 structure has a dual effect on the process of DNA replication. On one hand, the G4 structure has been demonstrated to support the initiation of DNA replication at the replication origin [78]. Furthermore, the G4 structure may prevent the uncoupling of the leading- and lagging-strand polymerases, thereby protecting proper replication [22]. On the other hand, the G4 could hinder the progression of the replication fork and influence DNA synthesis, which may lead to mutations and deletions in the genome. Consequently, helicases usually unfold the G4 structures before replication to maintain genome stability [22].

FANCI (Fanconi anemia complementation group J) is a 5′–3′ DNA helicase, which is involved in various biological processes such as DNA damage repair, G4 resolution, homologous recombination and genome stability maintenance [23]. FANCI can unfold and remove G4 structures for efficient DNA replication while its absence will stop replication at G4s and eventually lead to DNA damage [79]. Studies have shown that FANCI might promote replication at G4s by two independent mechanisms [80]. One mechanism is that FANCI may cooperate with polymerase REV1 to aid replication at the replication fork [81]. REV1 destabilizes the G4 structures so that FANCI can unwind them from the other side of the G4 structures. Second, WRN or BLM may assist FANCI to bind and unfold the G4s from the opposite direction in order to promote replication synergistically [21,80,82].

The helicase Pif1 from yeasts is able to bind and unfold G4 structures to support DNA replication. It is not clear whether Pif1 can play a role unwinding G4 structures on both chains or if it has a binding preference for the G4 structure at a certain chain [22]. However, recent studies have suggested that the ubiquitin ligase complex protein Mms1 is not only a DNA G4-binding protein, but also assists Pif1 to bind to a specific G4 structure located on the lagging strand. It could be observed that the absence of Mms1 leads to a reduction in Pif1 binding and slow replication at G4 motifs, and finally causes G4-dependent genome instability [83].

2.2.3. G-Quadruplex-Binding Proteins Involved in Transcription

It has been found that about 50% of human genes contain G4 motifs near their promoter region, which indicates that G4s play an essential role in the regulation of gene expression [5]. When DNA G4 is located at the first intron downstream of the transcription start site (TSS), it blocks the RNA polymerase and suppresses transcription [29]. However, recent studies have shown that endogenous G4s in promoters are prominent binding sites for multiple transcription factors and are thus invariably linked to high transcription levels [29,30]. Notably, G4s and their associated transcription factors cooperate to shape the cell-specific transcriptome [29,84]. In fact, transcription factors account for a significant part

of the G4BPs. Statistically, there are 14 transcription factors among the 56 DNA G4-binding proteins in the G4IPDB (G4 Interacting Proteins DataBase) [85]. For example, SP1 (Specificity protein 1) is a zinc finger transcription factor, which can bind to the G4 structures on the *c-KIT* promoter and regulate the expression of a variety of housekeeping genes [29]. MAZ (Myc-associated zinc finger) and PARP-1 (Poly [ADP-ribose] polymerase 1) interact with the G4 structures upstream of the transcription start site of *KRAS*, and both of them are activators of *KRAS* [2,5,86].

The G4 motif occurs more frequently in proto-oncogenes and regulatory genes than in housekeeping genes and tumor suppressor genes [5,87,88]. The first reported G4 on the promoter is formed in the nuclease hypersensitivity element III1 (NHE III1) which locates upstream of the P1 promoter of the proto-oncogene *c-MYC* [23,89]. This guanine-rich region controls 85–90% of the transcriptional activation of the gene, and can fold into an intramolecular parallel G4 as a transcriptional repressor element [76]. In addition to *c-MYC*, many genes have been demonstrated to form G4 structures in the promoter regions, such as proto-oncogenes *VEGF* [90], *KRAS* [91], *BCL-2* [92] and *c-KIT* [93]; human platelet-derived growth factor receptor *PDGFR-β* [94]; human telomerase reverse transcriptase *hTERT* [95] and other genes [23]. In particular, the G4s in the promoter regions of the proto-oncogenes have been most intensively studied so far [2].

Nucleolin (NCL) is a multifunctional phosphoprotein that is most abundant in the nucleolus. Nucleolin is mainly associated with ribosome biosynthesis and also involved in chromatin remodeling, transcriptional regulation, G4 binding and apoptosis [76]. Nucleolin can bind to the *c-MYC* G4 with high affinity and promote the formation and stabilization of G4 structures. The luciferase assay results also proved that the overexpression of nucleolin could contribute remarkably to a reduction in *c-MYC*-driven transcription [76]. Another protein NM23-H2 which belongs to the NM23 family of nucleoside diphosphate kinase (NDPK) has a completely different structure effect on G4s from nucleolin. It has a variety of functions, including kinase activity, promoter binding, transcriptional regulation and DNA repair [43]. Experiments have confirmed that NM23-H2 could bind to the *c-MYC* G4 to promote the unfolding of the G4 structure, thereby activating the transcription of *c-MYC* [43].

The tumor suppressor protein p53 functions in apoptosis, DNA repair, cell cycle regulation and aging. As a transcriptional regulator, p53 can inhibit the expression of cell cycle regulatory and growth promoting genes via multiple mechanisms and plays a key role in tumor suppression [44]. Previous studies have found that wild-type p53 (wtp53) and several types of mutant p53 (mutp53) have the ability to selectively bind *c-MYC* and *hTERT* promoter G4s [96], and the C-terminal region of p53 is essential for the recognition of the G4. Accordingly, the interaction between p53 and G4 structures in promoter regions of p53 target genes may play an important role in p53-mediated transcriptional regulation [44].

2.2.4. Other DNA G-Quadruplex-Binding Proteins

Direct evidence has demonstrated that the endogenous human G4 DNA landscape is dynamically shaped by chromatin relaxation or cell status [14,97]. Indeed, several G4BPs also function in chromatin structure regulation and histone modification [98,99]. For example, various epigenetic and chromatin remodeling enzymes bind selectively to DNA G4 [2]. Genomic binding sites of the chromatin remodeling protein ATR-X colocalize with GC-rich tandem repeats and CpG islands (CGI) that have the potential to form G4 structures [100,101].

Guanine-rich sequences are very common around CpG islands, with a high distribution rate of up to 80% [23,102,103]. The presence of G4 structures is closely related to the hypomethylation of CpG islands in the human genome. Studies have revealed that DNMT1 (DNA methyltransferase 1) interacts with these G4 sites, which is consistent with the results observed in biophysical experiments. Specifically, DNMT1 shows a higher binding affinity to G4 compared with double-stranded, single-stranded or hemimethylated DNA [28]. Biochemical analyses demonstrated that G4 structures inhibit the enzymatic

activity of DNMT1, and the formation of G4 also hinders DNMT1 to protect specific CpG islands from methylation and inhibit local methylation [28].

In addition, it has been found that G4s colocalize with CTCF (CCCTC-binding factor) binding sites in CpG islands and interact with CTCF in vitro. G4 is also crucial to the localization of CTCF [104]. CTCF is frequently recruited to CpG islands that are usually hypomethylated. Furthermore, the enrichment of G4s at CpG islands maintains CGI hypomethylation, which may explain the correlation between CpG islands and CTCF [104]. CTCF also functions as a chromatin remodeling factor with the capability of nucleosome repositioning; therefore, G4 can facilitate the binding of CTCF to genomic DNA by recruiting chromatin proteins [99].

2.3. RNA G-Quadruplex-Binding Proteins

It is easier for single-stranded RNA to form G4s in guanine-rich regions, and G4 is also an important structural characteristic of mRNA [10,46]. Recently, in vitro experiments combining high-throughput sequencing with reverse transcriptase stalling at RNA G4s (rG4) have found more than 13,000 loci with the potential to form rG4 structures in the human transcriptome; and immunofluorescence using G4 specific antibodies demonstrated rG4 formation in cells [17,105]. Notably, the highest abundance of rG4 is in functional regions including 5' and 3'-UTR [46]. All these observations of the enrichment of rG4s in functionally important regions suggest that they play crucial roles in transcription termination, alternative splicing, translational regulation, and chromosome integrity maintenance [57,106].

A substantial number of proteins interacting with rG4s have been identified by biochemical experiments, for example hnRNPs, ribosomal proteins and splicing factors [10,57]. Although there are DNA and RNA G4 specific proteins, their binding proteins have a significant overlap due to structural similarities between DNA and RNA G4s [10]. Basically, the discrimination between DNA and RNA G4BPs may depend on their different biological functions. It was found that the fragile X mental retardation protein (FMRP) could bind to the G4s in its own mRNA coding region, thereby regulating its own translation through a negative feedback pathway [107]. Additionally, FMRP is likely to interact with G4s in other mRNAs for translation repression by the recruitment of translation inhibitors, miRNA pathway activation, and direct interaction with ribosomes [57]. FRAXE-associated mental retardation protein FMR2 could also bind to G4s in mRNAs and function in alternative splicing [108].

The rG4 in the region where proto-oncogene *NRAS* 5'-UTR folds into a stable intramolecular parallel G4 structure and it has been demonstrated that it represses translation in vitro [57]. The study revealed that DEAD box helicase DDX3X involved in several pathways of RNA biology could bind to *NRAS* rG4s and the mutations of DDX3X are associated with tumorigenesis, especially medulloblastoma [57]. In addition, some helicases such as DHX36 (DEAH-Box Helicase 36) and DDX21 are able to bind and unfold rG4 structures. Another multifunctional helicase DHX9 shows a binding affinity for several secondary nucleic acid structures including G4s, but it is more inclined to bind RNA substrates. Therefore, helicases with the function of recognition and resolution of rG4s may play essential roles in post-transcriptional biological processes such as mRNA translation, transportation and stability [46].

Although the vast majority of rG4s are present in mRNAs, others are also detected in long non-coding RNAs (lncRNAs) including nuclear paraspeckle assembly transcript 1 (*NEAT1*). *NEAT1* is involved in gene regulation as a scaffold for the assembly of paraspeckles [109]. An upregulation of *NEAT1* could be observed in the majority of solid tumors such as lung cancer, esophageal cancer and hepatocellular carcinoma, and *NEAT1* also plays a critical role in neurodegenerative diseases and viral infection [49,54]. Evidence has shown that nascent *NEAT1* transcripts interact directly with the non-POU domain-containing octamer-binding protein (NONO) through its conserved rG4 motifs. The primary paraspeckle formation is required for the recruitment of NONO to *NEAT1*

transcripts which stabilizes *NEAT1* and lays the foundation for the recruitment of additional protein components to facilitate subsequent steps of assembly and maturation [54].

3. Structural Properties of G-Quadruplex-Binding Proteins

A comprehensive review revealed that the function of a protein is controlled by its amino acid sequence comprising domains and motifs [110]. The types of these domains are the basic elements that constitute the intrinsic properties of a protein such as interacting with nucleic acid sequences, and they also have an impact on the pathways to which it belongs [110]. The G4 recognition of proteins is a multistep process which involves the key domain recognizing the G4 structures through the interaction with adjacent disordered regions [111]. The analyses of known G4BPs suggest that the established or predicted binding regions in G4BPs have certain shared domains or motifs (Figure 4). Several studies indicate that the enrichment of these domains in G4BPs contributes to interacting with G4s [51,111]. The deeper insight into their characteristics will not only make progress in the binding mechanisms of G4-protein interactions, but also provide precise structural targets for subsequent drug design. Recent discoveries in the structural properties of G4BPs will be discussed in the following section.

	NCL	RGGGRGGFGGRRGGRRGGFGGRRGGFGG
RGG/RG motif	hnRNP U	RGGGHRGRGGFNMRGGNFRGGAPGNRGG
	CIRBP	RGGSAGGRGFFRGGRRGRGFSRGG

RRM domain



OB-fold domain



Figure 4. Structural properties of G-quadruplex-binding proteins. RGG/RG motifs are from NCL, hnRNP U and CIRBP. The RRM domain structure is derived from Protein Data Bank with structure code 2KRR (NCL) [112]. RRM domain is an $\alpha\beta$ sandwich structure composed of one four-stranded antiparallel β -sheet and two α -helices packed against the β -sheet. The OB-fold domain structure is derived from Protein Data Bank with structure code 5W2L (CTC1) [113]. OB-fold domain is a β -barrel formed by five antiparallel β -sheets.

3.1. RGG Domain

The RGG (Arginine-Glycine-Glycine) domain, also termed the RGG/RG motif or GAR (glycine-arginine-rich) domain is composed of repeat sequences rich in RGG or RG and is highly conserved in evolution (Figure 4) [110,111]. Researchers have discovered RGG/RG motifs in more than 1000 human proteins which influence transcription, precursor mRNA splicing, DNA damage signaling pathways, mRNA translation, and apoptosis [110]. A recent study analyzed the amino acid composition of 77 human G4-binding proteins [52]. Compared with a random subset of the human proteome and a well-defined group of nucleic acid binding proteins, the study demonstrated a significant enrichment of glycine and arginine and also high abundance in RR, GR and RG in G4BPs. Research was conducted to investigate the presence of a conserved RG-rich motif, which is a typical characteristic of G4BPs [52].

The RGG domain is usually found in G4BPs and it has been shown to mediate G4-protein interactions. For example, hnRNP U contains the RGG domain [62]. The C-terminal region of nucleolin composed of RNA-binding domain (RBD) 3 and 4 and the RGG domain is essential for the recognition of the *c-MYC* NHE III1 sequence and the promotion of G4 formation [60]. In addition, more than half of the newly identified *NRAS* rG4BPs contain the GAR domain which has been proved to be critical for *NRAS* rG4-DDX3X interaction [46].

The short residue gap between RGG repeats in the RGG domain frequently contains aromatic amino acids. The research on the binding mechanisms of the RGG domain revealed that the small segment RGG motif in this domain greatly contributes to the G4 binding affinity. Huang et al. found that the internal arrangement of RGG repeats and gap amino acids are more fundamental to G4-protein interactions than the length of RGG peptides and numbers of RGG repeats [60]. Experiments demonstrated that the peptide 12 with seven RGG repeats could efficiently bind to DNA G4s. Based on the above results, they discovered that the cold-inducible RNA-binding protein (CIRBP) containing peptide 12 could bind G4s both in vitro and in vivo, and this RGG peptide is essential for the G4 recognition of CIRBP [60]. The team provided a great deal of insight into the interaction between the RGG peptide and G4s, and identified a new G4-binding protein based on the exploration of G4-binding RGG motifs. In summary, this approach also adds a new dimension to the discovery of other G4BPs.

3.2. RRM Domain

Several G4BPs, such as hnRNPs, nucleolin, CIRBP, TLS/FUS (translocated in liposarcoma, also known as fused in sarcoma), and EWS (Ewing's sarcoma), have shared structural features, such as RNA recognition motifs (RRM) and RGG domains [42]. RRM, also known as the RNA-binding domain (RBD) or ribonucleoprotein domain (RNP), is one of the most highly conserved nucleic acid binding domains that occurs in approximately 0.5–1% of human genes and folds into an $\alpha\beta$ sandwich structure composed of one four-stranded antiparallel β -sheet and two α -helices packed against the β -sheet (Figure 4) [112,114,115]. Proteins with RRM are implicated in the regulation of transcription, translation, RNA processing, RNA export and stability [27], and they are also common in G4BPs.

The RRM and RGG domains at the C-terminal of nucleolin are necessary to inhibit and induce the formation of the G4 on the *c-MYC* promoter. The RRM in nucleolin can form G4s with guanine-containing single strands, but it unfolds G4s without guanines in the single strands of the 5' and 3' terminals [27]. The RRMs of hnRNP A1 and hnRNP D are able to bind and unfold G4s. The crystal structure of the two RRMs of hnRNP A1 with single-stranded telomeric DNA showed that RRM1 and RRM2 interact directly with d(TAGG) and d(TTAGG), respectively [116]. The RRM of hnRNP D could recognize d(TAG) in d(TTAGGG) determined by NMR [117]. A recent study indicated that a novel G4-binding protein SLIRP (stem-loop interacting RNA binding protein) also contains the RRM domain, which is required for efficient interaction between DNA G4s and SLIRP [118]. Furthermore, the sequence alignment for the RRMs derived from SLIRP and other G4BPs such as hnRNP A1 and nucleolin showed similar amino acid composition of these do-

mains [118]. The findings of these studies shed light on the roles of the RRM domain conserved in many nucleic acid binding proteins and contribute greatly to the exploration of its biological functions.

3.3. OB-Fold Domain

Oligonucleotide/oligosaccharide binding (OB)-fold is a β -barrel structure comprising a five-stranded antiparallel β -sheet, and this barrel is capped by an α -helix located between the third and fourth strands (Figure 4) [119]. The OB-fold structure is highly dynamic, and the dynamic properties enable OB-fold containing proteins to participate in multiple cellular pathways, including the re-initiation of DNA synthesis and the maintenance of genome stability [120].

Replication protein A (RPA) is a single-stranded DNA-binding complex with three subunits which unfolds the G4s and is involved in various biological processes such as DNA replication, repair and recombination. Although both RPA and POT1-TPP1 can bind to telomeric overhangs, RPA is more abundant in cells [10]. The CST complex resembles RPA in that they harbor comparable arrays of OB-folds and possess small subunits with similar structures [34]. Since CST contains multiple OB-folds (one each in STN1 and TEN1, and seven in CTC1), it was estimated that CST could play distinct roles in replication using a dynamic binding mechanism similar to that observed in RPA [34,35,113]. The dynamic properties of RPA binding due to the microscopic dissociation and re-association of individual OB-folds allow RPA to diffuse along the single-stranded DNA and to melt unwanted DNA secondary structures [34]. In addition, POT1 also contains the OB-fold domain, and FRET has shown that it is critical for gradual G4 unfolding [73].

DHX36 can bind DNA and RNA G4 structures with high affinity. It is a multifunctional helicase involved in G4-dependent transcriptional and post-transcriptional regulation, and plays a critical role in heart development, hematopoiesis and embryogenesis in mice [121]. The DHX36-specific motif at the N-terminal of the protein forms a DNA-binding-induced α -helix that together with the OB-fold-like subdomain selectively binds to parallel G4s [121].

4. G-Quadruplex-Binding Proteins as Potential Drug Targets

4.1. Relationship between G-Quadruplex-Binding Proteins and Diseases

Multiple G4-recruited proteins functioning at G4s are critical for genome homeostasis since G4 structures pose a threat to genome stability by hindering efficient replication and inducing DNA damage. Genome instability caused by G4s is linked to several genetic disorders and may further contribute to carcinogenesis [122]. For instance, the elevated expression of G4-resolving helicases such as WRN and BLM has been demonstrated in cancer cells in order for highly proliferating cells to deal with increasing replicative lesions [53,123]. Aside from DNA helicases, a number of polymerases implicated in translesion synthesis (TLS) have the ability to replicate past G4 sites [53].

Similarly, various DNA repair proteins can interact with G4s, suggesting an important role of G4 structures in DNA repair pathways. It has been shown that G4 could positively or negatively affect DNA repair efficiency and the extent of the impact depends on the repair pathway itself [26,122]. Homologous recombination (HR) is critical for DSB (double-strand break) repair during DNA replication. It was proposed that G4 structures need to be processed by several HR factors such as Mre11 and DNA2 to initiate HR-mediated repair [124,125]. In addition, BRAC1 and BRAC2 are key proteins necessary for G4 modulation during HR and cancer cells with a deficient HR protein are sensitive to the treatment of G4-stabilizing ligands which leads to elevated DNA damage and cell death [122]. In the contrast, several studies indicate that G4 structures have positive effects on the nucleotide excision repair (NER) pathway that is frequently activated upon ultraviolet light (UV) irradiation and defects in NER may lead to a high risk of skin cancer development [122,126]. Recent experiments have demonstrated an increase in G4 formation after UV damage while the stabilization of G4s by the protein Zuo1 would further contribute to the recruitment of NER proteins and maintain genome integrity [127].

Additionally, G4BPs which regulate the expression of multiple proto-oncogenes influence gene transcription by stabilizing or unfolding the G4 structure on the promoters, thus they play an important role in cancer initiation and progression. For example, the G4 in *c-MYC* is the target of several G4BPs, such as nucleolin and NM23-H2, which are of great significance in cancer treatment. As a transcription activator of *c-MYC*, the decreased level of NM23-H2 will reduce the *c-MYC* expression, while it is also related to the enhancement of metastatic potential. In various cell types including lung epithelial cancer and A549 cells, decreased expression of *c-MYC* results in cell cycle arrest and apoptosis, and activation of apoptosis contributes greatly to a reduction in metastatic spread [43]. In summary, lower *c-MYC* levels caused by the decreased expression of NM23-H2 in cancer cells would diminish apoptosis of cancer cells and enhance metastatic potential as well [43]. Basically, a comprehensive study of the role that these proteins play in regulating the expression of genes required for tumorigenesis, maintenance and metastasis will enlighten novel cancer therapies by specifically altering their expression [10].

Secondly, several G4BPs are involved in the regulation of neurological diseases. For example, a notably decreased expression of FMRP that binds to G4-containing mRNA and regulates its transport could be observed in Fragile X syndrome, and consequently this reduction also directly affected the translation of several other mRNAs [128]. Another study showed that hnRNP H/F as important components of the cytoplasmic machinery responsible for the structural integrity of rG4 function in rG4-mediated translational control [129]. These proteins are an essential regulatory hub in glioblastoma (GBM) networks and hnRNP H/F overexpression in GBM drives translational control of rG4-containing mRNAs encoding proteins implicated in the maintenance of genome stability and the response to genotoxic damage [129]. Furthermore, an RNA binding protein GLN1 (Guanine Nucleotide-Binding protein-like 1) binds to the G4 structures in the 5'-UTRs of VPS35 and PRKN, and these two genes are related to Parkinson's disease [130]. The target genes of an RNA G4-binding protein DDX3X encode proteins, thereby suggesting roles in the oxidative phosphorylation chain, while severe consequences and several diseases, including Huntington's, Alzheimer's, Parkinson's disease and cancer are caused by dysregulation of the synthesis of oxidative phosphorylation components [46].

Researchers have also discovered that G4BPs play a role in the critical steps of viral infection. The presence of G4 has been confirmed in the genomes of SARS coronavirus [131], human papilloma [132], hepatitis C [133], Zika [134], Ebola virus and herpes simplex virus-1 (HSV-1) [59,135,136]. Human immunodeficiency virus-1 (HIV-1) which is the pathogen of the acquired immune deficiency syndrome (AIDS) is regulated by G4 and its binding proteins. Researchers have identified functionally significant G4s in the Nef coding region and in the unique long terminal repeat (LTR) promoter of *HIV-1* [59]. The G4 formation in the LTR promoter region suppresses the transcription of *HIV-1*. Nucleolin strengthens inhibition capability by inducing G4 formation and stabilizing the G4 structure, while hnRNP A2/B1 activates *HIV-1* transcription by unfolding the LTR G4 structures [59]. Moreover, a recent study has demonstrated direct interaction between the G4 structure in the SARS-CoV-2 RNA genome and viral helicase nsp13. Targeting viral helicase and G4 structure is valuable for potentially inhibiting the SARS-CoV-2 virus [137].

4.2. Advances in G-Quadruplex-Binding Proteins for Drug Design

Since G4BPs are associated with various diseases, understanding the G4-protein interactions could provide new insight into therapeutic interventions when dysregulation of G4 formation and resolution has been considered as a pathogenic cause [138]. Moreover, G4s in live cells may exhibit polymorphism such as several conformations (folded or unfolded) available for ligand binding, such dynamic properties lead to the relatively poor specificity of G4-directed ligands. [139] Thus, researchers start to take into account G4-protein complexes as novel drug targets [55].

Several DNA helicases and polymerases which function in DNA repair pathways are involved in the resolution of G4s to prevent G4-driven genome instability. One clinical

application is to design small-molecule helicase inhibitors for personalized cancer therapy as conventional approaches including chemotherapy and radiation would cause unpleasant side effects such as the cytotoxicity to normal cells and drug resistance [123]. Accordingly, a synthetic lethality approach could be exploited to combat cancer. The efficiency of this new therapy is usually predicated on the combination of a defective genetic background and an inhibitor of the DNA repair protein [53,123,140]. For example, PARP inhibitors have been applied to tumors characterized by a deficiency of BRAC1 or BRAC2 to enhance anticancer effects due to the vital role of the G4BP PARP1 in the DNA repair pathway [123]. Additionally, accumulating evidence shows that a helicase inhibitor in conjunction with G4 structure stabilization can also achieve synthetic lethality. For instance, the previous study demonstrated that cancer cells using a FANCD1-specific helicase inhibitor are highly sensitive to G4-stabilizing ligands [53,141]. Another promising clinical application is the introduction of G4-disrupting small molecules which can unfold G4s to rescue helicase impairment [142].

To lessen the side effects of conventional anticancer drugs, aptamers can be valuable for the development of efficient targeted drug delivery systems due to their high selectivity. Aptamers are short-folded oligonucleotides capable of specifically recognizing target molecules with high affinity. A unique option of aptamer architectures is G4 and G4-based aptamers have potential clinical applications. For example, the aptamer d(GGGT)₄ able to form a 5'-5' dimer of two stacked parallel G4s has shown anti-HIV activity by targeting the HIV-1 integrase (HIV-IN) [143]. Another 26-base G4 AS1411 aptamer which binds nucleolin on the cancer cell surface with high affinity and specificity has been widely used for cancer therapy and diagnosis [55,144]. AS1411 exhibits cancer-selective antiproliferative effects and causes cell death through its interaction with nucleolin [144].

In addition to the development of small-molecule inhibitors and G4-based aptamers, diverse techniques have been applied to target G4BPs. In recent years, proteolysis-targeting chimeras (PROTACs) which hijack the endogenous ubiquitin-proteasome system to degrade target proteins have emerged as a promising therapeutic modality for proteins including those non-druggable ones [138,145]. A novel strategy termed G4-PROTAC is generated from a combination of a G4 warhead and a E3 ligase recruiter. G4-PROTAC is designed for specific degradation of a G4-binding protein (RHAU/DHX36) which has been reported to be highly expressed in tissues of *C9orf72*-linked amyotrophic lateral sclerosis (ALS) patients and thus represents an important therapeutic target [138]. In conclusion, this approach has the prospect of extended application and holds great promise in potential therapeutics against diseases for which an aberrant expression of a G4-binding protein is known to be a major cause [138].

In summary, three kinds of methods can be exploited for drug design. First, G4 aptamers are designed to form a synthetic G4-protein complex for cellular function modulation. Second, small molecules or ligands are used to inhibit target proteins. Third, we could utilize the ubiquitin-proteasome system to degrade specific G4BPs (Figure 5).

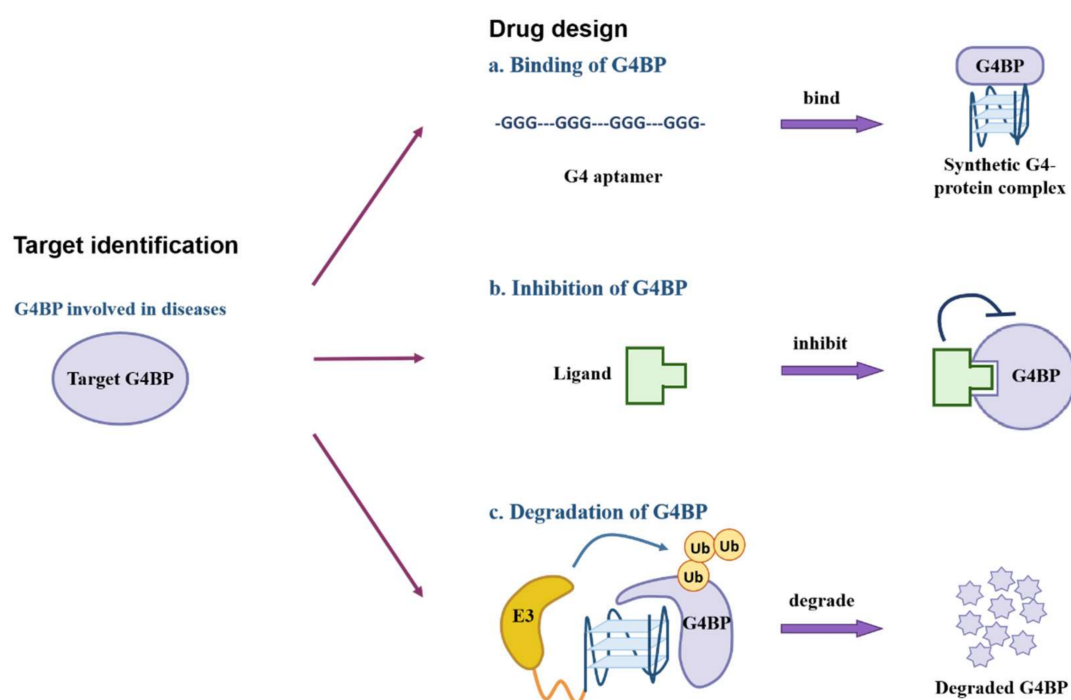


Figure 5. The schematic diagram for G-quadruplex-binding proteins as drug targets. The first step is target identification. After analyzing G4–protein interactions, G4BPs involved in disease-related pathways are chosen as potential drug targets. Secondly, three methods could be exploited for drug design. According to the G4–protein recognition mechanism, G4-based aptamers are designed to target specific proteins and form a synthetic G4–protein complex. Meanwhile, ligands would be used to inhibit G4BPs for synthetic lethality in cells with genetic deficiencies. In addition, E3 ligase could be recruited for ubiquitin-mediated degradation of target proteins.

5. Conclusions and Perspectives

G4BPs are of vital importance in regulating the formation, stabilization and resolution of G4 structures in cells, and this regulatory mechanism has spatial and temporal properties. Generally, there are two kinds of classification methods for G4BPs. Firstly, we can divide these proteins into two main categories as follows: G4-folding proteins and G4-recruited proteins. Then, we can describe the control mechanisms of DNA and RNA G4BPs in detail according to the genomic distribution of G4s. Specifically, DNA G4BPs are mainly implicated in telomere maintenance, transcription, replication and epigenetic modification while RNA G4BPs are involved in mRNA maturation and translational regulation.

In recent years, a great number of new G4BPs have been identified, and researchers have also provided a good deal of insight into shared motifs and domains of these proteins. In terms of application, the targeting of G4BPs will serve as a breakthrough in drug innovation and add new dimensions to medical treatment. Owing to the prevalence of G4 motifs in the regulatory regions of the human genome, G4BPs play a vital role in different cellular pathways related to diseases and the dysregulation of G4BPs leads to tumor formation and viral infection. Nowadays, G4-resolving helicases and polymerases implicated in DNA repair pathways have been considered as a novel kind of drug targeting therapy with high medical value.

Subsequent studies will focus on the molecular mechanism of G4–protein recognition using high-throughput sequencing techniques and computational analyses [60,111]. We expect to have a more comprehensive understanding of the structural properties of G4BPs and the sequence features of their binding sites which lay the solid foundation for the identification of new G4BPs through G4-binding residues [52]. Furthermore, another research hotspot is the change in the regulatory mechanism of G4–protein interactions under normal physiological or abnormal pathological conditions. Since endogenous G4s and G4BPs are

critical in the regulation of various biological processes, we ought to place emphasis on the improvement of current approaches targeting G4BPs [146]. Exogenous G4 aptamers present conformational polymorphisms which limit the characterization of their structural features, hence G4 aptamers are required to make modifications in order to overcome these structural drawbacks [146,147]. Meanwhile, bioinformatic research on G4BPs could unveil a deeper understanding of new target identification that offers clinical potential. The tremendous advances in computational modeling in the past decade have increased the speed and success rate of the discovery of small molecule drug candidates [148]. The DeepMind AlphaFold algorithm is a deep learning-based technology to accurately predict protein structures through amino acid sequences in the absence of experimental data, which provides an important structural basis for drug target discovery, drug design and molecular optimization [149,150]. New drug design targeting proteins requires structure-based virtual screening, so it is important to expand the scope of structure-based modeling.

In conclusion, we expect to build a satisfactory database of G4BPs and collect the sequence-based features obtained from experiments, and then employ homology modeling and deep learning methods to predict the structures of these proteins [151]. On the basis of the relationship between G4BPs and diseases, we could screen potential proteins for drug target discovery and then design ligands and antibodies specific to target G4BPs for medical treatment. The integration of bioinformatics, structural biology, computational chemistry and machine learning will promote the development of novel, safe and effective drugs targeting G4BPs.

Author Contributions: Conceptualization, H.S., R.Z., K.X., J.Y. and X.S.; writing—original draft preparation, H.S.; writing—review and editing, H.S., R.Z., K.X., J.Y. and X.S.; funding acquisition, K.X. and X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No.61972084 and 62002060) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX21_0143).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hansel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **2018**, *13*, 551–564. [[CrossRef](#)] [[PubMed](#)]
2. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [[CrossRef](#)] [[PubMed](#)]
3. 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. [[CrossRef](#)] [[PubMed](#)]
4. Tarsounas, M.; Tijsterman, M. Genomes and G-quadruplexes: For better or for worse. *J. Mol. Biol.* **2013**, *425*, 4782–4789. [[CrossRef](#)] [[PubMed](#)]
5. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627–8637. [[CrossRef](#)]
6. Fay, M.M.; Lyons, S.M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429*, 2127–2147. [[CrossRef](#)]
7. Ding, D.; Wei, C.; Dong, K.; Liu, J.; Stanton, A.; Xu, C.; Min, J.; Hu, J.; Chen, C. LOTUS domain is a novel class of G-rich and G-quadruplex RNA binding domain. *Nucleic Acids Res.* **2020**, *48*, 9262–9272. [[CrossRef](#)]
8. Zhang, R.X.; Liu, Y.J.; Zhang, X.X.; Xiao, K.; Hou, Y.; Liu, H.D.; Sun, X. Detecting and Profiling Endogenous RNA G-Quadruplexes in the Human Transcriptome. *Int. J. Mol. Sci.* **2021**, *22*, 8012. [[CrossRef](#)]
9. Vannutelli, A.; Belhamiti, S.; Garant, J.M.; Ouangraoua, A.; Perreault, J.P. Where are G-quadruplexes located in the human transcriptome? *NAR Genom. Bioinform.* **2020**, *2*, lqaa035. [[CrossRef](#)]
10. Brazda, V.; Haronikova, L.; Liao, J.C.; Fojta, M. DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* **2014**, *15*, 17493–17517. [[CrossRef](#)]

11. Burge, S.; Parkinson, G.N.; Hazel, P.; Todd, A.K.; Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415. [[CrossRef](#)]
12. Mergny, J.L.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37*, 17.1.1–17.1.15. [[CrossRef](#)] [[PubMed](#)]
13. Chambers, V.S.; Marsico, G.; Boutell, J.M.; Di Antonio, M.; Smith, G.P.; Balasubramanian, S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **2015**, *33*, 877–881. [[CrossRef](#)] [[PubMed](#)]
14. Hansel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48*, 1267–1272. [[CrossRef](#)] [[PubMed](#)]
15. Zheng, K.W.; Zhang, J.Y.; He, Y.D.; Gong, J.Y.; Wen, C.J.; Chen, J.N.; Hao, Y.H.; Zhao, Y.; Tan, Z. Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.* **2020**, *48*, 11706–11720. [[CrossRef](#)]
16. Lyu, J.; Shao, R.; Yung, P.Y.K.; Elsasser, S.J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res.* **2022**, *50*, e13. [[CrossRef](#)]
17. Kwok, C.K.; Marsico, G.; Sahakyan, A.B.; Chambers, V.S.; Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **2016**, *13*, 841–844. [[CrossRef](#)]
18. Yang, S.Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A.G.; Wong, J.M.Y.; Monchaud, D. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **2018**, *9*, 4730. [[CrossRef](#)]
19. Huppert, J.L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916. [[CrossRef](#)]
20. Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Res.* **2012**, *40*, 4727–4741. [[CrossRef](#)]
21. Mendoza, O.; Bourdoncle, A.; Boule, J.B.; Brosh, R.M., Jr.; Mergny, J.L. G-quadruplexes and helicases. *Nucleic Acids Res.* **2016**, *44*, 1989–2006. [[CrossRef](#)] [[PubMed](#)]
22. Sauer, M.; Paeschke, K. G-quadruplex unwinding helicases and their function in vivo. *Biochem. Soc. Trans.* **2017**, *45*, 1173–1182. [[CrossRef](#)] [[PubMed](#)]
23. Sun, Z.Y.; Wang, X.N.; Cheng, S.Q.; Su, X.X.; Ou, T.M. Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid(-)Protein Interaction. *Molecules* **2019**, *24*, 396. [[CrossRef](#)] [[PubMed](#)]
24. 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. [[CrossRef](#)]
25. Takahama, K.; Takada, A.; Tada, S.; Shimizu, M.; Sayama, K.; Kurokawa, R.; Oyoshi, T. Regulation of Telomere Length by G-Quadruplex Telomere DNA- and TERRA-Binding Protein TLS/FUS. *Chem. Biol.* **2013**, *20*, 341–350. [[CrossRef](#)]
26. Pavlova, A.V.; Kubareva, E.A.; Monakhova, M.V.; Zvereva, M.I.; Dolinnaya, N.G. Impact of G-Quadruplexes on the Regulation of Genome Integrity, DNA Damage and Repair. *Biomolecules* **2021**, *11*, 1284. [[CrossRef](#)]
27. Oyoshi, T.; Masuzawa, T. Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 39–44. [[CrossRef](#)]
28. Mao, S.Q.; Ghanbarian, A.T.; Spiegel, J.; Martinez Cuesta, S.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **2018**, *25*, 951–957. [[CrossRef](#)]
29. Kim, N. The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **2019**, *26*, 2898–2917. [[CrossRef](#)]
30. Spiegel, J.; Cuesta, S.M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **2021**, *22*, 117. [[CrossRef](#)]
31. Chen, Y. The structural biology of the shelterin complex. *Biol. Chem.* **2019**, *400*, 457–466. [[CrossRef](#)] [[PubMed](#)]
32. Mei, Y.; Deng, Z.; Vladimirova, O.; Gulve, N.; Johnson, F.B.; Drosopoulos, W.C.; Schildkraut, C.L.; Lieberman, P.M. TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. *Sci. Rep.* **2021**, *11*, 1–14. [[CrossRef](#)] [[PubMed](#)]
33. Zhu, J.; Liu, W.; Chen, C.; Zhang, H.; Yue, D.; Li, C.; Zhang, L.; Gao, L.; Huo, Y.; Liu, C.; et al. TPP1 OB-fold domain protein suppresses cell proliferation and induces cell apoptosis by inhibiting telomerase recruitment to telomeres in human lung cancer cells. *J. Cancer Res. Clin. Oncol.* **2019**, *145*, 1509–1519. [[CrossRef](#)] [[PubMed](#)]
34. Bhattacharjee, A.; Wang, Y.; Diao, J.; Price, C.M. Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. *Nucleic Acids Res.* **2017**, *45*, 12311–12324. [[CrossRef](#)]
35. Lim, C.J.; Barbour, A.T.; Zaug, A.J.; Goodrich, K.J.; McKay, A.E.; Wuttke, D.S.; Cech, T.R. The structure of human CST reveals a decameric assembly bound to telomeric DNA. *Science* **2020**, *368*, 1081–1085. [[CrossRef](#)]
36. Karow, J.K.; Wu, L.; Hickson, I.D. RecQ family helicases: Roles in cancer and aging. *Curr. Opin. Genet. Dev.* **2000**, *10*, 32–38. [[CrossRef](#)]
37. Mohaghegh, P.; Karow, J.K.; Brosh, R.M.; Bohr, V.A.; Hickson, I.D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **2001**, *29*, 2843–2849. [[CrossRef](#)]
38. Park, C.J.; Ko, J.; Ryu, K.S.; Choi, B.S. Solution structure of the RecQ C-terminal domain of human Bloom syndrome protein. *J. Biomol. Nmr* **2014**, *58*, 141–147. [[CrossRef](#)]

39. Wu, W.Q.; Hou, X.M.; Li, M.; Dou, S.X.; Xi, X.G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res.* **2015**, *43*, 4614–4626. [[CrossRef](#)]
40. Wu, C.G.; Spies, M. G-quadruplex recognition and remodeling by the FANCD1 helicase. *Nucleic Acids Res.* **2016**, *44*, 8742–8753. [[CrossRef](#)]
41. Sato, K.; Martin-Pintado, N.; Post, H.; Altelaar, M.; Knipscheer, P. Multistep mechanism of G-quadruplex resolution during DNA replication. *Sci. Adv.* **2021**, *7*, eabf8653. [[CrossRef](#)] [[PubMed](#)]
42. Masuzawa, T.; Oyoshi, T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega* **2020**, *5*, 5202–5208. [[CrossRef](#)] [[PubMed](#)]
43. Thakur, R.K.; Kumar, P.; Halder, K.; Verma, A.; Kar, A.; Parent, J.L.; Basundra, R.; Kumar, A.; Chowdhury, S. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **2009**, *37*, 172–183. [[CrossRef](#)] [[PubMed](#)]
44. Petr, M.; Helma, R.; Polaskova, A.; Krejci, A.; Dvorakova, Z.; Kejnovska, I.; Navratilova, L.; Adamik, M.; Vorlickova, M.; Brazdova, M. Wild-type p53 binds to MYC promoter G-quadruplex. *Biosci. Rep.* **2016**, *36*, e00397. [[CrossRef](#)] [[PubMed](#)]
45. Teng, Y.C.; Sundaresan, A.; O'Hara, R.; Gant, V.U.; Li, M.H.; Martire, S.; Warshaw, J.N.; Basu, A.; Banaszynski, L.A. ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nat. Commun.* **2021**, *12*, 1–14. [[CrossRef](#)]
46. Herdy, B.; Mayer, C.; Varshney, D.; Marsico, G.; Murat, P.; Taylor, C.; D'Santos, C.; Tannahill, D.; Balasubramanian, S. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.* **2018**, *46*, 11592–11604. [[CrossRef](#)]
47. Varshney, D.; Cuesta, S.M.; Herdy, B.; Abdullah, U.B.; Tannahill, D.; Balasubramanian, S. RNA G-quadruplex structures control ribosomal protein production. *Sci. Rep.-Uk* **2021**, *11*, 1–11. [[CrossRef](#)]
48. Goering, R.; Hudish, L.I.; Guzman, B.B.; Raj, N.; Bassell, G.J.; Russ, H.A.; Dominguez, D.; Taliaferro, J.M. FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *Elife* **2020**, *9*, e52621. [[CrossRef](#)]
49. Simko, E.A.J.; Liu, H.; Zhang, T.; Velasquez, A.; Teli, S.; Haeusler, A.R.; Wang, J. G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. *Nucleic Acids Res.* **2020**, *48*, 7421–7438. [[CrossRef](#)]
50. Amato, J.; Cerofolini, L.; Brancaccio, D.; Giuntini, S.; Iaccarino, N.; Zizza, P.; Iachettini, S.; Biroccio, A.; Novellino, E.; Rosato, A.; et al. Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein. *Nucleic Acids Res.* **2019**, *47*, 9950–9966. [[CrossRef](#)]
51. McRae, E.K.S.; Booy, E.P.; Padilla-Meier, G.P.; McKenna, S.A. On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. *J. Nucleic Acids* **2017**, *2017*, 9675348. [[CrossRef](#)]
52. Brazda, V.; Cerven, J.; Bartas, M.; Mikyskova, N.; Coufal, J.; Pecinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **2018**, *23*, 2341. [[CrossRef](#)] [[PubMed](#)]
53. Estep, K.N.; Butler, T.J.; Ding, J.; Brosh, R.M. G4-Interacting DNA Helicases and Polymerases: Potential Therapeutic Targets. *Curr. Med. Chem.* **2019**, *26*, 2881–2897. [[CrossRef](#)]
54. Tassinari, M.; Richter, S.N.; Gandellini, P. Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* **2021**, *49*, 3617–3633. [[CrossRef](#)]
55. Sissi, C.; Gatto, B.; Palumbo, M. The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective. *Biochimie* **2011**, *93*, 1219–1230. [[CrossRef](#)] [[PubMed](#)]
56. Matsumoto, K.; Okamoto, K.; Okabe, S.; Fujii, R.; Ueda, K.; Ohashi, K.; Seimiya, H. G-quadruplex-forming nucleic acids interact with splicing factor 3B subunit 2 and suppress innate immune gene expression. *Genes Cells* **2021**, *26*, 65–82. [[CrossRef](#)] [[PubMed](#)]
57. Serikawa, T.; Spanos, C.; von Hacht, A.; Budisa, N.; Rappsilber, J.; Kurreck, J. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* **2018**, *144*, 169–184. [[CrossRef](#)] [[PubMed](#)]
58. Lee, C.Y.; McNERNEY, C.; Myong, S. G-Quadruplex and Protein Binding by Single-Molecule FRET Microscopy. *Methods Mol. Biol.* **2019**, *2035*, 309–322. [[CrossRef](#)]
59. Scalabrin, M.; Frasson, I.; Ruggiero, E.; Perrone, R.; Tosoni, E.; Lago, S.; Tassinari, M.; Palu, G.; Richter, S.N. The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **2017**, *7*, 45244. [[CrossRef](#)]
60. Huang, Z.L.; Dai, J.; Luo, W.H.; Wang, X.G.; Tan, J.H.; Chen, S.B.; Huang, Z.S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.* **2018**, *140*, 17945–17955. [[CrossRef](#)]
61. Zhang, X.; Spiegel, J.; Martinez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **2021**, *13*, 626–633. [[CrossRef](#)] [[PubMed](#)]
62. Izumi, H.; Funai, K. Telomere Function and the G-Quadruplex Formation are Regulated by hnRNP U. *Cells* **2019**, *8*, 390. [[CrossRef](#)] [[PubMed](#)]
63. Stewart, J.A.; Chaiken, M.F.; Wang, F.; Price, C.M. Maintaining the end: Roles of telomere proteins in end-protection, telomere replication and length regulation. *Mutat. Res.-Fund. Mol. M* **2012**, *730*, 12–19. [[CrossRef](#)] [[PubMed](#)]
64. Arnoult, N.; Karlseder, J. Complex interactions between the DNA-damage response and mammalian telomeres. *Nat. Struct. Mol. Biol.* **2015**, *22*, 859–866. [[CrossRef](#)] [[PubMed](#)]

65. Baumann, P.; Cech, T.R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **2001**, *292*, 1171–1175. [[CrossRef](#)]
66. Wang, F.; Podell, E.R.; Zaug, A.J.; Yang, Y.; Baciu, P.; Cech, T.R.; Lei, M. The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **2007**, *445*, 506–510. [[CrossRef](#)]
67. Chaires, J.B.; Gray, R.D.; Dean, W.L.; Monsen, R.; DeLeeuw, L.W.; Stribinskis, V.; Trent, J.O. Human POT1 unfolds G-quadruplexes by conformational selection. *Nucleic Acids Res.* **2020**, *48*, 4976–4991. [[CrossRef](#)]
68. Wu, W.; Rokutanda, N.; Takeuchi, J.; Lai, Y.; Maruyama, R.; Togashi, Y.; Nishikawa, H.; Arai, N.; Miyoshi, Y.; Suzuki, N.; et al. HERC2 Facilitates BLM and WRN Helicase Complex Interaction with RPA to Suppress G-Quadruplex DNA. *Cancer Res.* **2018**, *78*, 6371–6385. [[CrossRef](#)]
69. Budhathoki, J.B.; Ray, S.; Urban, V.; Janscak, P.; Yodh, J.G.; Balci, H. RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res.* **2014**, *42*, 11528–11545. [[CrossRef](#)]
70. Miyake, Y.; Nakamura, M.; Nabetani, A.; Shimamura, S.; Tamura, M.; Yonehara, S.; Saito, M.; Ishikawa, F. RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol. Cell* **2009**, *36*, 193–206. [[CrossRef](#)]
71. Zhang, M.M.; Wang, B.; Li, T.F.; Liu, R.; Xiao, Y.N.; Geng, X.; Li, G.; Liu, Q.; Price, C.M.; Liu, Y.; et al. Mammalian CST averts replication failure by preventing G-quadruplex accumulation. *Nucleic Acids Res.* **2019**, *47*, 5243–5259. [[CrossRef](#)] [[PubMed](#)]
72. Surovtseva, Y.V.; Churikov, D.; Boltz, K.A.; Song, X.Y.; Lamb, J.C.; Warrington, R.; Leehy, K.; Heacock, M.; Price, C.M.; Shippen, D.E. Conserved Telomere Maintenance Component 1 Interacts with STN1 and Maintains Chromosome Ends in Higher Eukaryotes. *Mol. Cell* **2009**, *36*, 207–218. [[CrossRef](#)] [[PubMed](#)]
73. Hwang, H.; Buncher, N.; Opresko, P.L.; Myong, S. POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure* **2012**, *20*, 1872–1880. [[CrossRef](#)] [[PubMed](#)]
74. Leon-Ortiz, A.M.; Svendsen, J.; Boulton, S.J. Metabolism of DNA secondary structures at the eukaryotic replication fork. *DNA Repair* **2014**, *19*, 152–162. [[CrossRef](#)] [[PubMed](#)]
75. Kruger, A.C.; Raarup, M.K.; Nielsen, M.M.; Kristensen, M.; Besenbacher, F.; Kjems, J.; Birkedal, V. Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *Eur. Biophys. J. Biophys.* **2010**, *39*, 1343–1350. [[CrossRef](#)] [[PubMed](#)]
76. Gonzalez, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **2009**, *284*, 23622–23635. [[CrossRef](#)] [[PubMed](#)]
77. Hudson, J.S.; Ding, L.; Le, V.; Lewis, E.; Graves, D. Recognition and Binding of Human Telomeric G-Quadruplex DNA by Unfolding Protein 1. *Biochemistry* **2014**, *53*, 3347–3356. [[CrossRef](#)]
78. Valton, A.L.; Prioleau, M.N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **2016**, *32*, 697–706. [[CrossRef](#)]
79. Castillo Bosch, P.; Segura-Bayona, S.; Koole, W.; van Heteren, J.T.; Dewar, J.M.; Tijsterman, M.; Knipscheer, P. FANCI promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **2014**, *33*, 2521–2533. [[CrossRef](#)]
80. Sarkies, P.; Murat, P.; Phillips, L.G.; Patel, K.J.; Balasubramanian, S.; Sale, J.E. FANCI coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **2012**, *40*, 1485–1498. [[CrossRef](#)]
81. Eddy, S.; Ketkar, A.; Zafar, M.K.; Maddukuri, L.; Choi, J.Y.; Eoff, R.L. Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res.* **2014**, *42*, 3272–3285. [[CrossRef](#)] [[PubMed](#)]
82. Suhasini, A.N.; Brosh, R.M. Fanconi anemia and Bloom’s syndrome crosstalk through FANCI-BLM helicase interaction. *Trends Genet.* **2012**, *28*, 7–13. [[CrossRef](#)] [[PubMed](#)]
83. Schwindt, E.; Paeschke, K. Mms1 is an assistant for regulating G-quadruplex DNA structures. *Curr. Genet.* **2018**, *64*, 535–540. [[CrossRef](#)] [[PubMed](#)]
84. Lago, S.; Nadai, M.; Cernilogar, F.M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S.N. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **2021**, *12*, 3885. [[CrossRef](#)]
85. Mishra, S.K.; Tawani, A.; Mishra, A.; Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **2016**, *6*, 38144. [[CrossRef](#)]
86. Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K.K.; Xodo, L.E. 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.* **2010**, *285*, 22003–22016. [[CrossRef](#)]
87. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [[CrossRef](#)]
88. Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **2006**, *34*, 3887–3896. [[CrossRef](#)]
89. Simonsson, T.; Pecinka, P.; Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **1998**, *26*, 1167–1172. [[CrossRef](#)]
90. Sun, D.; Guo, K.; Rusche, J.J.; Hurley, L.H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **2005**, *33*, 6070–6080. [[CrossRef](#)]

91. Cogo, S.; Xodo, L.E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006**, *34*, 2536–2549. [[CrossRef](#)] [[PubMed](#)]
92. Dexheimer, T.S.; Sun, D.; Hurley, L.H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–5415. [[CrossRef](#)] [[PubMed](#)]
93. Rankin, S.; Reszka, A.P.; Huppert, J.; Zloh, M.; Parkinson, G.N.; Todd, A.K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589. [[CrossRef](#)] [[PubMed](#)]
94. Qin, Y.; Fortin, J.S.; Tye, D.; Gleason-Guzman, M.; Brooks, T.A.; Hurley, L.H. Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. *Biochemistry* **2010**, *49*, 4208–4219. [[CrossRef](#)] [[PubMed](#)]
95. Palumbo, S.L.; Ebbinghaus, S.W.; Hurley, L.H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* **2009**, *131*, 10878–10891. [[CrossRef](#)] [[PubMed](#)]
96. Quante, T.; Otto, B.; Brazdova, M.; Kejnovska, I.; Deppert, W.; Tolstonog, G.V. Mutant p53 is a transcriptional co-factor that binds to G-rich regulatory regions of active genes and generates transcriptional plasticity. *Cell Cycle* **2012**, *11*, 3290–3303. [[CrossRef](#)]
97. Guilbaud, G.; Murat, P.; Recolin, B.; Campbell, B.C.; Maiter, A.; Sale, J.E.; Balasubramanian, S. Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **2017**, *9*, 1110–1117. [[CrossRef](#)]
98. Zyner, K.G.; Simeone, A.; Flynn, S.M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S. G-quadruplex DNA structures in human stem cells and differentiation. *Nat. Commun.* **2022**, *13*, 142. [[CrossRef](#)]
99. Hou, Y.; Li, F.Y.; Zhang, R.X.; Li, S.; Liu, H.D.; Qin, Z.H.S.; Sun, X. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics-U.S.* **2019**, *14*, 894–911. [[CrossRef](#)]
100. Law, M.J.; Lower, K.M.; Voon, H.P.; Hughes, J.R.; Garrick, D.; Viprakasit, V.; Mitson, M.; De Gobbi, M.; Marra, M.; Morris, A.; et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **2010**, *143*, 367–378. [[CrossRef](#)]
101. Wang, Y.; Yang, J.; Wild, A.T.; Wu, W.H.; Shah, R.; Danussi, C.; Riggins, G.J.; Kannan, K.; Sulman, E.P.; Chan, T.A.; et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat. Commun.* **2019**, *10*, 943. [[CrossRef](#)] [[PubMed](#)]
102. Cayrou, C.; Coulombe, P.; Vigneron, A.; Stanojic, S.; Ganier, O.; Peiffer, I.; Rivals, E.; Puy, A.; Laurent-Chabalier, S.; Desprat, R.; et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **2011**, *21*, 1438–1449. [[CrossRef](#)] [[PubMed](#)]
103. Besnard, E.; Babied, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.M.; Lemaitre, J.M. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **2012**, *19*, 837–844. [[CrossRef](#)] [[PubMed](#)]
104. Tikhonova, P.; Pavlova, I.; Isaakova, E.; Tsvetkov, V.; Bogomazova, A.; Vedekhina, T.; Luzhin, A.V.; Sultanov, R.; Severov, V.; Klimina, K.; et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. *Int. J. Mol. Sci.* **2021**, *22*, 7090. [[CrossRef](#)] [[PubMed](#)]
105. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **2014**, *6*, 75–80. [[CrossRef](#)] [[PubMed](#)]
106. Agarwala, P.; Pandey, S.; Maiti, S. The tale of RNA G-quadruplex. *Org. Biomol. Chem.* **2015**, *13*, 5570–5585. [[CrossRef](#)]
107. Schaeffer, C.; Bardoni, B.; Mandel, J.L.; Ehresmann, B.; Ehresmann, C.; Moine, H. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **2001**, *20*, 4803–4813. [[CrossRef](#)]
108. Bensaïd, M.; Melko, M.; Bechara, E.G.; Davidovic, L.; Berretta, A.; Catania, M.V.; Gecz, J.; Lalli, E.; Bardoni, B. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.* **2009**, *37*, 1269–1279. [[CrossRef](#)]
109. Clemson, C.M.; Hutchinson, J.N.; Sara, S.A.; Ensminger, A.W.; Fox, A.H.; Chess, A.; Lawrence, J.B. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **2009**, *33*, 717–726. [[CrossRef](#)]
110. Thandapani, P.; O'Connor, T.R.; Bailey, T.L.; Richard, S. Defining the RGG/RG motif. *Mol. Cell* **2013**, *50*, 613–623. [[CrossRef](#)]
111. Kharel, P.; Becker, G.; Tsvetkov, V.; Ivanov, P. Properties and biological impact of RNA G-quadruplexes: From order to turmoil and back. *Nucleic Acids Res.* **2020**, *48*, 12534–12555. [[CrossRef](#)]
112. Arumugam, S.; Miller, M.C.; Maliekal, J.; Bates, P.J.; Trent, J.O.; Lane, A.N. Solution structure of the RBD1,2 domains from human nucleolin. *J. Biomol. Nmr* **2010**, *47*, 79–83. [[CrossRef](#)] [[PubMed](#)]
113. Shastrula, P.K.; Rice, C.T.; Wang, Z.; Lieberman, P.M.; Skordalakes, E. Structural and functional analysis of an OB-fold in human Ctc1 implicated in telomere maintenance and bone marrow syndromes. *Nucleic Acids Res.* **2018**, *46*, 972–984. [[CrossRef](#)] [[PubMed](#)]
114. Clery, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290–298. [[CrossRef](#)] [[PubMed](#)]
115. Maris, C.; Dominguez, C.; Allain, F.H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118–2131. [[CrossRef](#)] [[PubMed](#)]
116. Ding, J.; Hayashi, M.K.; Zhang, Y.; Manche, L.; Krainer, A.R.; Xu, R.M. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **1999**, *13*, 1102–1115. [[CrossRef](#)] [[PubMed](#)]

117. Enokizono, Y.; Konishi, Y.; Nagata, K.; Ouhashi, K.; Uesugi, S.; Ishikawa, F.; Katahira, M. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J. Biol. Chem.* **2005**, *280*, 18862–18870. [[CrossRef](#)]
118. Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J. Am. Chem. Soc.* **2017**, *139*, 12426–12429. [[CrossRef](#)]
119. Murzin, A.G. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J.* **1993**, *12*, 861–867. [[CrossRef](#)]
120. Nguyen, D.D.; Kim, E.Y.; Sang, P.B.; Chai, W. Roles of OB-Fold Proteins in Replication Stress. *Front. Cell Dev. Biol.* **2020**, *8*, 574466. [[CrossRef](#)]
121. Chen, M.C.; Tippiana, R.; Demeshkina, N.A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D'Amare, A.R. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **2018**, *558*, 465–469. [[CrossRef](#)] [[PubMed](#)]
122. Linke, R.; Limmer, M.; Juranek, S.A.; Heine, A.; Paeschke, K. The Relevance of G-Quadruplexes for DNA Repair. *Int. J. Mol. Sci.* **2021**, *22*, 12599. [[CrossRef](#)] [[PubMed](#)]
123. Brosh, R.M., Jr. DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer* **2013**, *13*, 542–558. [[CrossRef](#)] [[PubMed](#)]
124. Ghosal, G.; Muniyappa, K. *Saccharomyces cerevisiae* Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: Implications for replication of telomeric DNA. *Nucleic Acids Res.* **2005**, *33*, 4692–4703. [[CrossRef](#)]
125. Masuda-Sasa, T.; Polaczek, P.; Peng, X.P.; Chen, L.; Campbell, J.L. 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.* **2008**, *283*, 24359–24373. [[CrossRef](#)]
126. Minetti, C.A.; Sun, J.Y.; Jacobs, D.P.; Kang, I.; Remeta, D.P.; Breslauer, K.J. Impact of bistrand abasic sites and proximate orientation on DNA global structure and duplex energetics. *Biopolymers* **2018**, *109*, e23098. [[CrossRef](#)]
127. De Magis, A.; Gotz, S.; Hajikazemi, M.; Fekete-Szucs, E.; Caterino, M.; Juranek, S.; Paeschke, K. Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nat. Commun.* **2020**, *11*, 1–12. [[CrossRef](#)]
128. Melko, M.; Bardoni, B. The role of G-quadruplex in RNA metabolism: Involvement of FMRP and FMR2P. *Biochimie* **2010**, *92*, 919–926. [[CrossRef](#)]
129. Herviou, P.; Le Bras, M.; Dumas, L.; Hieblot, C.; Gilhodes, J.; Cioci, G.; Hugnot, J.P.; Amedan, A.; Guillonnet, F.; Dassi, E.; et al. hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nat. Commun.* **2020**, *11*, 2661. [[CrossRef](#)]
130. Turcotte, M.A.; Garant, J.M.; Cossette-Roberge, H.; Perreault, J.P. Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease. *RNA Biol.* **2021**, *18*, 1339–1353. [[CrossRef](#)]
131. Tan, J.Z.; Vornrhein, C.; Smart, O.S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J.R.; Schmidt, C.L.; Hilgenfeld, R. The SARS-Unique Domain (SUD) of SARS Coronavirus Contains Two Macrod domains That Bind G-Quadruplexes. *PLoS Pathog.* **2009**, *5*, e1000428. [[CrossRef](#)] [[PubMed](#)]
132. Tluckova, K.; Marusic, M.; Tothova, P.; Bauer, L.; Sket, P.; Plavec, J.; Viglasky, V. Human Papillomavirus G-Quadruplexes. *Biochemistry* **2013**, *52*, 7207–7216. [[CrossRef](#)] [[PubMed](#)]
133. Wang, S.R.; Min, Y.Q.; Wang, J.Q.; Liu, C.X.; Fu, B.S.; Wu, F.; Wu, L.Y.; Qiao, Z.X.; Song, Y.Y.; Xu, G.H.; et al. A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.* **2016**, *2*, e1501535. [[CrossRef](#)] [[PubMed](#)]
134. Fleming, A.M.; Ding, Y.; Alenko, A.; Burrows, C.J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *Acs Infect. Dis.* **2016**, *2*, 674–681. [[CrossRef](#)] [[PubMed](#)]
135. Frasson, I.; Solda, P.; Nadai, M.; Lago, S.; Richter, S.N. Parallel G-quadruplexes recruit the HSV-1 transcription factor ICP4 to promote viral transcription in herpes virus-infected human cells. *Commun. Biol.* **2021**, *4*, 1–13. [[CrossRef](#)]
136. Wang, S.R.; Zhang, Q.Y.; Wang, J.Q.; Ge, X.Y.; Song, Y.Y.; Wang, Y.F.; Li, X.D.; Fu, B.S.; Xu, G.H.; Shu, B.; et al. Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem. Biol.* **2016**, *23*, 1113–1122. [[CrossRef](#)]
137. Ji, D.; Juhas, M.; Tsang, C.M.; Kwok, C.K.; Li, Y.; Zhang, Y. Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief. Bioinform.* **2021**, *22*, 1150–1160. [[CrossRef](#)]
138. Patil, K.M.; Chin, D.; Seah, H.L.; Shi, Q.; Lim, K.W.; Phan, A.T. G4-PROTAC: Targeted degradation of a G-quadruplex binding protein. *Chem. Commun. (Camb)* **2021**, *57*, 12816–12819. [[CrossRef](#)]
139. Aznauryan, M.; Noer, S.L.; Pedersen, C.W.; Mergny, J.L.; Teulade-Fichou, M.P.; Birkedal, V. Ligand Binding to Dynamically Populated G-Quadruplex DNA. *Chembiochem* **2021**, *22*, 1811–1817. [[CrossRef](#)]
140. Datta, A.; Dhar, S.; Awate, S.; Brosh, R.M., Jr. Synthetic Lethal Interactions of RECQ Helicases. *Trends Cancer* **2021**, *7*, 146–161. [[CrossRef](#)]
141. Salvati, E.; Botta, L.; Amato, J.; Di Leva, F.S.; Zizza, P.; Gioiello, A.; Pagano, B.; Graziani, G.; Tarsounas, M.; Randazzo, A.; et al. Lead Discovery of Dual G-Quadruplex Stabilizers and Poly(ADP-ribose) Polymerases (PARPs) Inhibitors: A New Avenue in Anticancer Treatment. *J. Med. Chem.* **2017**, *60*, 3626–3635. [[CrossRef](#)] [[PubMed](#)]
142. Mitteaux, J.; Lejault, P.; Wojciechowski, F.; Joubert, A.; Boudon, J.; Desbois, N.; Gros, C.P.; Hudson, R.H.E.; Boule, J.B.; Granzhan, A.; et al. Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* **2021**, *143*, 12567–12577. [[CrossRef](#)] [[PubMed](#)]

143. Esposito, V.; Pirone, L.; Mayol, L.; Pedone, E.; Virgilio, A.; Galeone, A. Exploring the binding of d(GGGT)₄ to the HIV-1 integrase: An approach to investigate G-quadruplex aptamer/target protein interactions. *Biochimie* **2016**, *127*, 19–22. [[CrossRef](#)] [[PubMed](#)]
144. Vindigni, G.; Raniolo, S.; Iacovelli, F.; Unida, V.; Stolfi, C.; Desideri, A.; Biocca, S. AS1411 Aptamer Linked to DNA Nanostructures Diverts Its Traffic Inside Cancer Cells and Improves Its Therapeutic Efficacy. *Pharmaceutics* **2021**, *13*, 1671. [[CrossRef](#)]
145. Liu, J.; Ma, J.; Liu, Y.; Xia, J.; Li, Y.; Wang, Z.P.; Wei, W. PROTACs: A novel strategy for cancer therapy. *Semin Cancer Biol.* **2020**, *67*, 171–179. [[CrossRef](#)]
146. Ogloblina, A.M.; Khristich, A.N.; Karpechenko, N.Y.; Semina, S.E.; Belitsky, G.A.; Dolinnaya, N.G.; Yakubovskaya, M.G. Multi-targeted effects of G₄-aptamers and their antiproliferative activity against cancer cells. *Biochimie* **2018**, *145*, 163–173. [[CrossRef](#)]
147. Miranda, A.; Santos, T.; Largy, E.; Cruz, C. Locking up the AS1411 Aptamer with a Flanking Duplex: Towards an Improved Nucleolin-Targeting. *Pharm. -Base* **2021**, *14*, 121. [[CrossRef](#)]
148. Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Park, J.S.; Jude, K.M.; Markovic, I.; Kadam, R.U.; Verschueren, K.H.G.; Verstraete, K.; et al. Design of protein binding proteins from target structure alone. *Nature* **2022**. [[CrossRef](#)]
149. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.L.; Zidek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [[CrossRef](#)]
150. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
151. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]