Title

Identification of SLIRP as a G Quadruplex-Binding Protein.

Journal

Journal of the American Chemical Society, 139(36)

Authors

Williams, Preston
Li, Lin
Dong, Xiaoli
et al.

Publication Date

2017-09-13

DOI

10.1021/jacs.7b07563

Peer reviewed

HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2018 September 13.

Published in final edited form as:

J Am Chem Soc. 2017 September 13; 139(36): 12426–12429. doi:10.1021/jacs.7b07563.

Identification of SLIRP as a G Quadruplex-Binding Protein

Preston Williams[†], **Lin Li**[†], **Xiaoli Dong**, and **Yinsheng Wang**[*,iD]

Department of Chemistry, University of California Riverside, Riverside, California 92521-0403, United States

Abstract

The guanine quadruplex (G4) structure in DNA is a secondary structure motif that plays important roles in DNA replication, transcriptional regulation, and maintenance of genomic stability. Here, we employed a quantitative mass spectrometry-based approach to profile the interaction proteomes of three well-defined G4 structures derived from the human telomere and the promoters of *cMYC* and *cKIT* genes. We identified SLIRP as a novel G4-interacting protein. We also demonstrated that the protein could bind directly with G4 DNA with K_d values in the low nanomolar range and revealed that the robust binding of the protein toward G4 DNA requires its RRM domain. We further assessed, by using CRISPR-Cas9-introduced affinity tag and ChIP-Seq analysis, the genome-wide occupancy of SLIRP, and showed that the protein binds preferentially to G-rich DNA sequences that can fold into G4 structures. Together, our results uncovered a novel cellular protein that can interact directly with G4 DNA, which underscored the complex regulatory networks involved in G4 biology.

The DNA guanine quadruplexes (G4) are stable four-stranded, noncanonical structures that can form in regions of the genome with high guanine content.[1,2] Although G4 folding patterns are highly diverse, the structural foundation of all G4s comprises multiple G-tetrads stacked upon one another (Figure 1).[1,2] A G-tetrad is a square planar structure of four guanines that interact with each other through Hoogsteen hydrogen bonding and are further stabilized by a monovalent cation, primarily K^+, located at the center of the four guanines (Figure 1a).[1,2] Although the *in vitro* formation of G4 structure has been known for decades, only recently have its formation and involvement in important biological processes in cells been demonstrated.[1,3]

Computational analysis using the consensus G4 sequence motif of $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ revealed >300 000 motifs in the human genome with potential

[*]**Corresponding Author**. Yinsheng.Wang@ucr.edu.

ORCID

Yinsheng Wang: 0000-0001-5565-283X

[†]These two authors contributed equally to this study.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b07563.

Detailed experimental procedures, representative ESI-MS and MS/MS results, CD spectra, SDS-PAGE for monitoring the purification of SLIRP protein, fluorescence anisotropy data, and ChIP-Seq results (PDF)

The authors declare no competing financial interest.

to fold into G4 structures,[4] and a newly described search algorithm estimated the number of putative G4-forming sequences to be 2–10-fold larger than what was initially predicted.[5] Interestingly, G4 motifs are not evenly distributed throughout the genome, where direct visualization of these motifs with immunofluorescence microscopy using a G4-specific antibody identified many G4-forming hotspots in human cells.[6] Furthermore, Hänsel-Hertsch et al.[7] uncovered, by using a G4-chromatin immunoprecipitation-sequencing (ChIP-seq) approach, approximately 10 000 G4 structures in the human chromatin with high enrichment in many loci of important biological relevance and regulatory functions, including more than 2000 gene promoters and the human telomeres.

Understanding fully the implications of G4 structures in the biological functions of nucleic acids, particularly the roles of G4 structures in gene regulation and human diseases, requires the investigation about how these structures are recognized by cellular proteins. Indeed, many proteins, including, among others, nucleolin, hnRNP A1, hnRNP A2, PARP1, Rif1 and SUB1, were shown to bind to G4 structure.[8–13]

Quantitative proteomics-based interaction screening constitutes a powerful and unbiased approach for uncovering cellular proteins that can bind to modified DNA.[14,15] Here, we employed stable isotope labeling by amino acid in cell culture (SILAC)[16]-based quantitative proteomic method to discover novel G4-binding proteins. To this end, we used three 5′-biotin-labeled G4 probes derived from the G-rich sequences of the human telomere and the promoters of c*KIT* and *cMYC* genes that were previously shown to adopt well-defined G4 folding *in vitro* as baits for pulling down G4-binding proteins.[17–19] We also obtained the corresponding probes where two guanine residues crucial for G4 folding and stability were mutated to thymine or adenine residues, and used these probes as control baits (Table S1). The proper folding of the G4-containing probes was confirmed by circular dichroism (CD) spectroscopy analysis (Figure S1). In this vein, the sequences derived from the promoters of the c*KIT* and c*MYC* genes yield maximum and minimum CD signals at around 260 and 240 nm, respectively (Figure S1), which are characteristic of parallel G4 folding topology.[20] The sequence arising from human telomere (hTel26) displayed maximum CD signals at 270 and 290 nm, which exhibit hydrid-type G4 folding with mixed parallel and antiparallel strands (Figure S1).[17] In contrast, the three mutated control sequences did not display CD signals that manifest G4 folding (Figure S1). To minimize nonspecific protein-beads interactions, we also inserted six thymidine residues between the sequence of interest and the biotin tag (Table S1).

To achieve metabolic labeling of the nuclear proteome, we cultured HeLa cells separately in light or heavy medium, and isolated the nuclear proteins from these cells. Equal amounts of nuclear proteins from the light- and heavy-labeled cells were passed through streptavidin columns that were immobilized with biotin-conjugated G4 DNA and the corresponding mutated sequence, respectively (Figure 1c), which was designated as the forward SILAC experiment. To remove experimental bias, we also conducted reverse SILAC experiment (see Experimental Section).

After incubation with the nuclear protein lysate, the DNA-conjugated beads were washed to remove nonspecific proteins, and the bound proteins were eluted from the beads, digested

with trypsin, and subjected to LC-MS/MS analysis. The LC-MS/ MS results revealed that SLIRP could bind specifically to all three G4 probes, with the SILAC protein ratios being 2.73 ± 1.10, 2.71 ± 0.70, and 3.90 ± 1.25 for G4 sequences derived from *cKIT*, *cMYC*, and hTel26 over their corresponding mutant probes, respectively. Representative LC-MS results for a tryptic peptide derived from SLIRP, SINQPVAFVR, are shown in Figure 2, which clearly showed the stronger binding of SLIRP to the three G4 sequences than the corresponding mutant probes in both forward and reverse SILAC experiments (MS/MS shown in Figure S2). The selective binding of SLIRP toward G4 DNA was also supported by another tryptic peptide derived from SLIRP, i.e., GLGWVQFSSEEGLR (Figure S3).

Considering that the above quantitative proteomics-based interaction screening may also yield proteins that can bind indirectly to G4 DNA via protein–protein interactions, we decided to examine whether SLIRP can bind directly with G4 DNA. To this end, we purified full-length recombinant SLIRP (Figure S4) and measured its binding affinities with G4 DNA and the corresponding mutant DNA using fluorescence anisotropy (the sequences for the fluorescently labeled probes are listed in Table S2). Our results revealed that SLIRP exhibited robust binding to all three G4 foldings with the K_d values for the G4 motifs derived from the promoters of the c*MYC* and c*KIT* genes and the human telomere being 98, 59, and 56 nM, respectively (Figure 3 and Table S3). In line with our proteomic data, the corresponding mutant probes incapable of folding into G4 structures displayed markedly lower binding affinities toward SLIRP, as reflected by the K_d values of 255, 612, and 372 nM, respectively (Figure 3 and Table S3). These results, therefore, demonstrated that SLIRP can bind directly and strongly to all three G4 folding structures. We also observed, from CD measurements, that the presence of SLIRP does not alter the folding of the G4 structures (Figure S5).

SLIRP was initially discovered to be an RNA-binding protein that interacts directly with the STR7 substructure of steroid receptor RNA activator (SRA).[21,22] Similar as other G4-binding proteins (e.g., hnRNP A1, hnRNP A2, and nucleolin), SLIRP possesses an RNA recognition motif (RRM) (Figure S6 shows the sequence alignment for the RRMs derived from these proteins). In addition, Leu62, Arg24, and Arg25 on the RNA binding surface of the RRM of SLIRP were found to be directly involved in this binding, and mutations of these residues to alanines (i.e., the L62A and R24A/R25A mutants) led to pronouncedly decreased interaction between SLIRP and its RNA target.[21] Thus, we next asked whether the interaction between SLIRP and G4 DNA is also modulated by these amino acid residues in the RRM. Our results showed that the L62A mutation or R24A/R25A double mutations led to significant diminutions in binding affinities toward all three G4 sequences, which result in loss of selectively of the two mutant forms of proteins toward G4 DNA over ssDNA, except that some selectivity was still observed for the L62A mutant toward the c*KIT* G4 over the corresponding ssDNA probe (Figure 3, Figure S7, and Table S3). These findings support that the intact RRM domain of SLIRP is required for its recognition of G4 DNA. For comparison, we also measured SLIRP's binding affinity toward truncated STR7 RNA by using fluorescence anisotropy, and it turned out that the binding affinity of the wild-type SLIRP toward the STR7 RNA (with a K_d value of 590 nM) was markedly lower than that toward the G4 motifs (Table S3 and Figure S8).

Having demonstrated the strong and selective binding of SLIRP toward the three distinct G4 folding structures *in vitro*, we next asked whether the protein also binds to G4 sequences in cells by assessing the genome-wide occupancy of SLIRP with ChIP-Seq analysis (Figure 4). To this end, we employed CRISPR-Cas9 genome-editing method to introduce a tandem affinity tag (3 × FLAG, 2 × Strept) to the C-terminus of endogenous SLIRP protein in HEK293T cells, where the successful introduction of the tandem affinity tag was confirmed using Western blot analysis (Figure 4). We then immunoprecipitated endogenous SLIRP and its associated genomic DNA using anti-FLAG M2 beads, and subjected the resulting DNA fragments to next-generation sequencing analysis. Bioinformatic analysis of the sequencing data revealed that many of the fragments pulled down with the tagged SLIRP protein are rich in guanine residues with potential in folding into G4 structures. In particular, we found that 13% and 66% of the total peaks contained the sequence motifs of $GGGN_xGGGN_xGGGN_xGGG$ and $GGN_xGGN_xGGN_xGG$, respectively (Figure 4). Moreover, the ChIP-Seq data clearly revealed strong peaks for SLIRP in binding toward telomeres in a number of chromosomes (Figure S9 and Table S4). This result, therefore, demonstrated that SLIRP can recognize G4 DNA structures in human cells. It is worth noting that, different from *in vitro* binding results, our ChIP-Seq data did not reveal the occupancy of SLIRP to the promoter regions of *cKIT* or *cMYC* gene. This difference could be attributed to the fact that the *in vitro* binding assay was conducted using the G-rich sequence in the absence of the complementary strand, whereas the folding of G-rich strands into G4 structures in HEK293T cells may be inhibited by the presence of the complementary strands (i.e., through duplex formation).

There are several novel findings in the present study. First, we employed a quantitative proteomic method and uncovered SLIRP as a novel cellular protein that can recognize multiple G4 structures, and this approach allowed for rapid and unbiased identification of SLIRP as a novel G4-binding protein without *a priori* knowledge. Second, our work also suggested novel functions of SLIRP. As noted above, SLIRP was initially shown to directly interact with the STR7 substructure in SRA and this interaction involves its RRM.[21] Here we found that mutations of important residues in the RNA-binding surface of RRM, i.e., L62A or R24A/R25A, which were previously found to reduce the binding of the protein to the STR7 substructure,[21] led to greatly diminished binding toward G4 structures. In addition, the binding affinities toward G4 DNA are much greater than that toward the STR7 substructure. Hence, our results suggest that the function of SLIRP may extend far beyond its recognition of STR7 substructure in SRA. More recently, SLIRP was found to form a complex with LRPPRC, which mediates the stability of mitochondrial mRNA.[23–25] Interestingly, mitochondrial DNA was recently shown to fold into G4 structures,[26] suggesting that SLIRP's capability in binding G4 DNA may also contribute, in part, to the protein's function in mitochondria. Although the primary characterized functions of SLIRP are within the scope of mitochondrial biology, not much is known about its role in the nucleus, where the protein also resides.[21] Our ChIP-Seq data revealed that SLIRP binds preferentially to G-rich regions of chromosomal DNA with the potential in folding into G4 structures. Thus, we uncovered a potential new role of SLIRP in the nucleus where it specifically recognizes G4 DNA. With G4 DNA being intimately involved with many

biological functions,[1,3] SLIRP may play a role in many biological processes including transcription and replication.

In summary, we identified, for the first time, SLIRP as a novel G4-binding protein by using an unbiased quantitative proteomic method. We further demonstrated that SLIRP protein can interact directly and selectively with G4 DNA with high affinity *in vitro*, and that the protein preferentially binds to G-rich sequences that can fold into G4 structures in cells. Considering that G-rich sequences in RNA can also fold into G4 structures,[27] it will be important to assess the interaction between SLIRP and G4 structures in RNA in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01 ES019873).

References

1. Bochman ML, Paeschke K, Zakian VA. Nat. Rev. Genet. 2012; 13:770–80. [PubMed: 23032257]
2. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Nucleic Acids Res. 2006; 34:5402–15. [PubMed: 17012276]
3. Gray LT, Vallur AC, Eddy J, Maizels N. Nat. Chem. Biol. 2014; 10:313–8. [PubMed: 24609361]
4. Huppert JL, Balasubramanian S. Nucleic Acids Res. 2005; 33:2908–2916. [PubMed: 15914667]
5. Bedrat A, Lacroix L, Mergny JL. Nucleic Acids Res. 2016; 44:1746–59. [PubMed: 26792894]
6. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Nat. Chem. 2013; 5:182–6. [PubMed: 23422559]
7. Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S. Nat. Genet. 2016; 48:1267–72. [PubMed: 27618450]
8. Gonzalez V, Guo K, Hurley L, Sun D. J. Biol. Chem. 2009; 284:23622–35. [PubMed: 19581307]
9. Gao J, Zybailov BL, Byrd AK, Griffin WC, Chib S, Mackintosh SG, Tackett AJ, Raney KD. Chem. Commun. 2015; 51:7242–4.
10. Kanoh Y, Matsumoto S, Fukatsu R, Kakusho N, Kono N, Renard-Guillet C, Masuda K, Iida K, Nagasawa K, Shirahige K, Masai H. Nat. Struct. Mol. Biol. 2015; 22:889–97. [PubMed: 26436827]
11. Paramasivam M, Membrino A, Cogoi S, Fukuda H, Nakagama H, Xodo LE. Nucleic Acids Res. 2009; 37:2841–53. [PubMed: 19282454]
12. Wang F, Tang ML, Zeng ZX, Wu RY, Xue Y, Hao YH, Pang DW, Zhao Y, Tan Z. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:20413–8. [PubMed: 23184978]
13. Soldatenkov VA, Vetcher AA, Duka T, Ladame S. ACS Chem. Biol. 2008; 3:214–9. [PubMed: 18338862]
14. Mittler G, Butter F, Mann M. Genome Res. 2008; 19:284–293. [PubMed: 19015324]
15. Du Z, Luo Q, Yang L, Bing T, Li X, Guo W, Wu K, Zhao Y, Xiong S, Shangguan D, Wang F. J. Am. Chem. Soc. 2014; 136:2948–51. [PubMed: 24524683]
16. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Mol. Cell. Proteomics. 2002; 1:376–86. [PubMed: 12118079]
17. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Nucleic Acids Res. 2006; 34:2723–35. [PubMed: 16714449]

18. Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. J. Am. Chem. Soc. 2007; 129:4386–92. [PubMed: 17362008]
19. Phan AT, Kuryavyi V, Gaw HY, Patel DJ. Nat. Chem. Biol. 2005; 1:167–73. [PubMed: 16408022]
20. Diveshkumar KV, Sakrikar S, Rosu F, Harikrishna S, Gabelica V, Pradeepkumar PI. Biochemistry. 2016; 55:3571–85. [PubMed: 27226253]
21. Hatchell EC, Colley SM, Beveridge DJ, Epis MR, Stuart LM, Giles KM, Redfern AD, Miles LE, Barker A, MacDonald LM, Arthur PG, Lui JC, Golding JL, McCulloch RK, Metcalf CB, Wilce JA, Wilce MC, Lanz RB, O'Malley BW, Leedman PJ. Mol. Cell. 2006; 22:657–68. [PubMed: 16762838]
22. Lanz RB, Razani B, Goldberg AD, O'Malley BW. Proc. Natl. Acad. Sci. U. S. A. 2002; 99:16081–6. [PubMed: 12444263]
23. Chujo T, Ohira T, Sakaguchi Y, Goshima N, Nomura N, Nagao A, Suzuki T. Nucleic Acids Res. 2012; 40:8033–47. [PubMed: 22661577]
24. Lagouge M, Mourier A, Lee HJ, Spahr H, Wai T, Kukat C, Silva Ramos E, Motori E, Busch JD, Siira S, Kremmer E, Filipovska A, Larsson NG. PLoS Genet. 2015; 11:e1005423. [PubMed: 26247782]
25. Sasarman F, Brunel-Guitton C, Antonicka H, Wai T, Shoubridge EA. Mol. Biol. Cell. 2010; 21:1315–23. [PubMed: 20200222]
26. Huang WC, Tseng TY, Chen YT, Chang CC, Wang ZF, Wang CL, Hsu TN, Li PT, Chen CT, Lin JJ, Lou PJ, Chang TC. Nucleic Acids Res. 2015; 43:10102–10113. [PubMed: 26487635]
27. Cammas A, Millevoi S. Nucleic Acids Res. 2017; 45:1584–1595. [PubMed: 28013268]

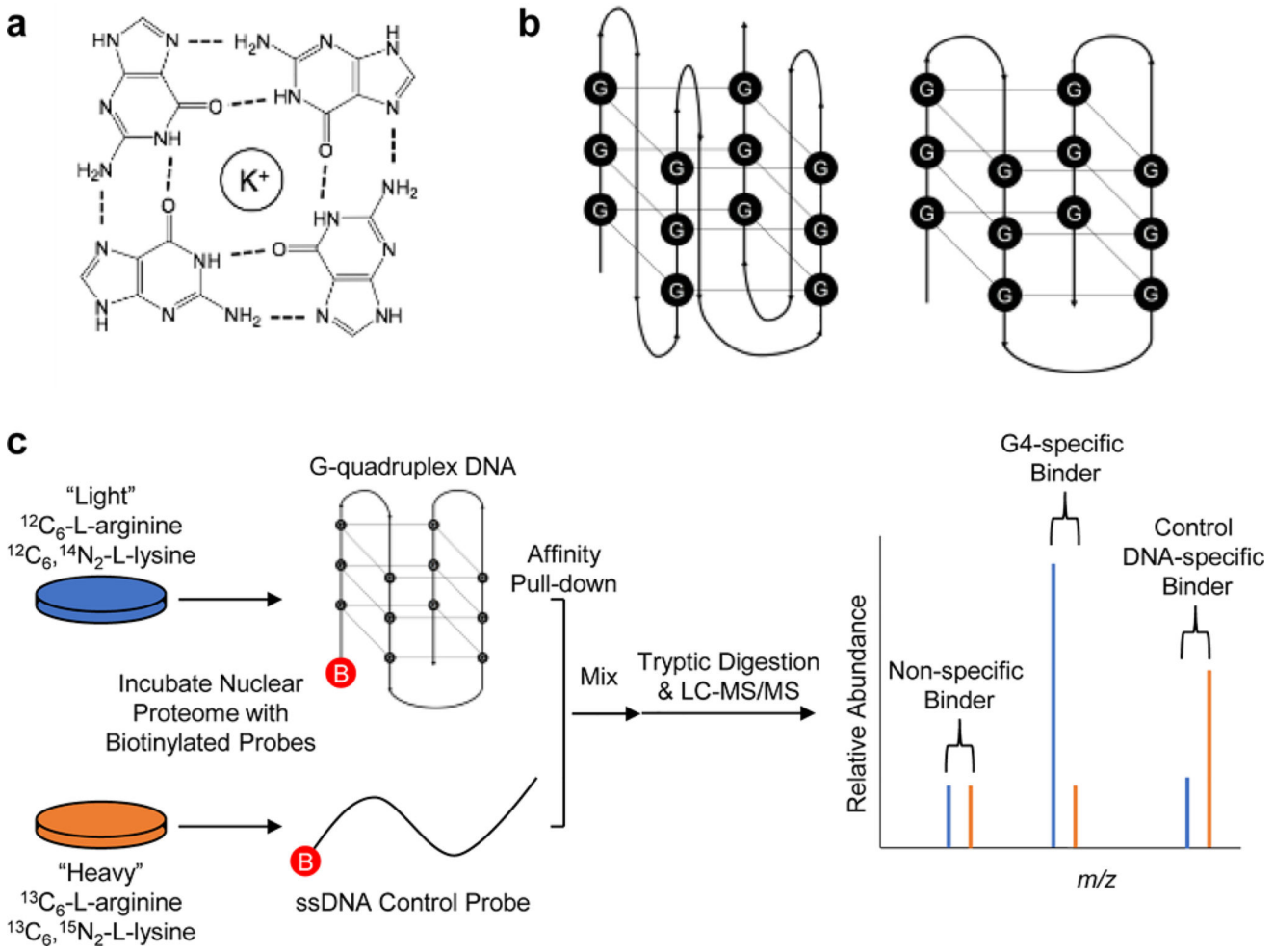

Figure 1.

G-quadruplex structures and the experimental procedures for the identification of novel G-quadruplex-binding proteins. Shown are the G-tetrad structure (a), parallel and antiparallel G-quadruplex foldings (b), and SILAC-based interaction screening for the identification of G quadruplex-binding proteins (c). The 'B' in red circle indicates 5′-biotin labeling.

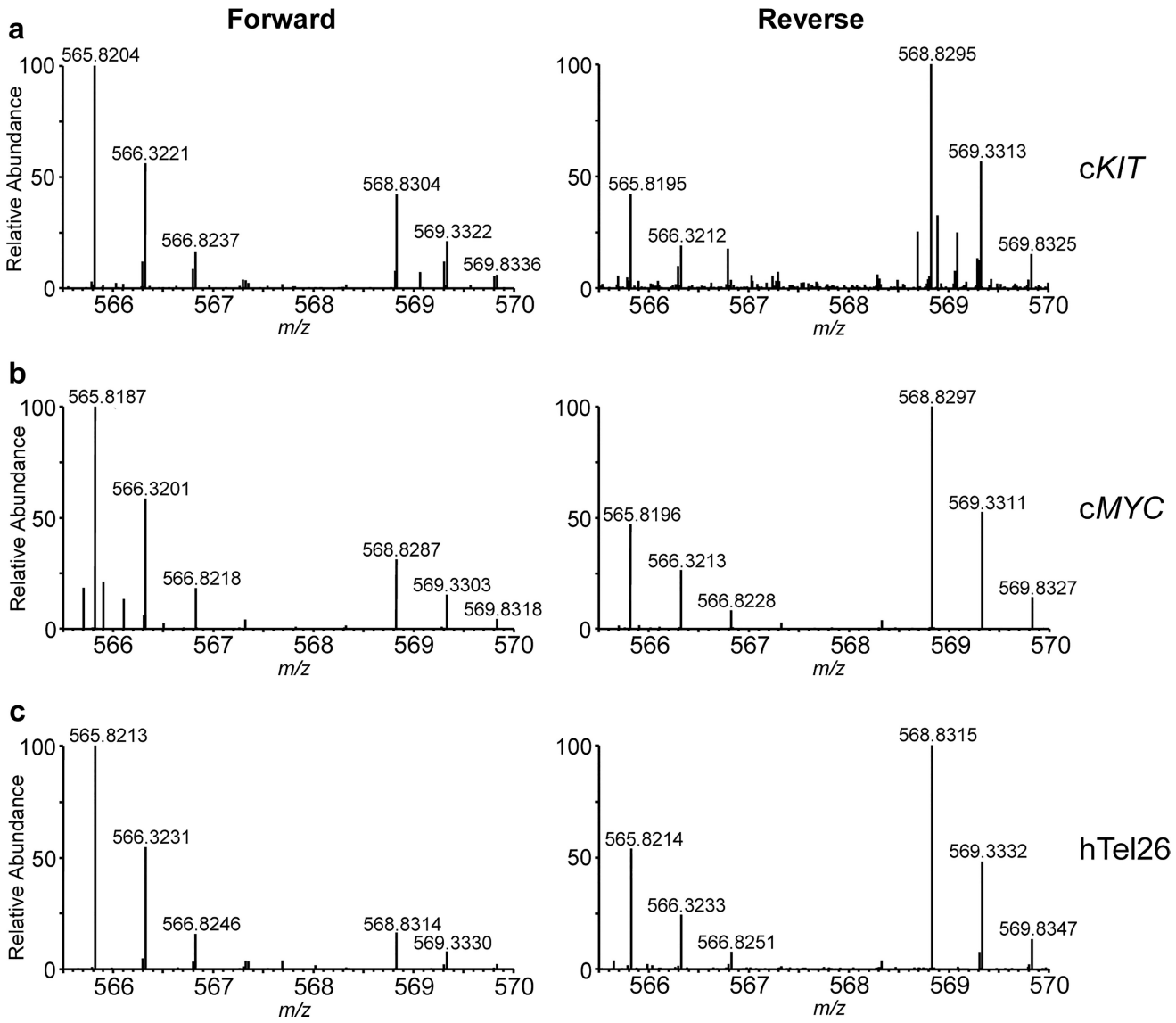

Figure 2.

ESI-MS revealed the preferential binding of SLIRP to G4 structures derived from the promoters of *cKIT* (a) and *cMYC* (b) genes as well as the human telomere (c). Shown are the ESI-MS for the $[M + 2H]^{2+}$ ions of light and heavy arginine-containing peptide SINQPVAFVR with monoisotopic *m/z* values of ~565.8 and 568.8, respectively.

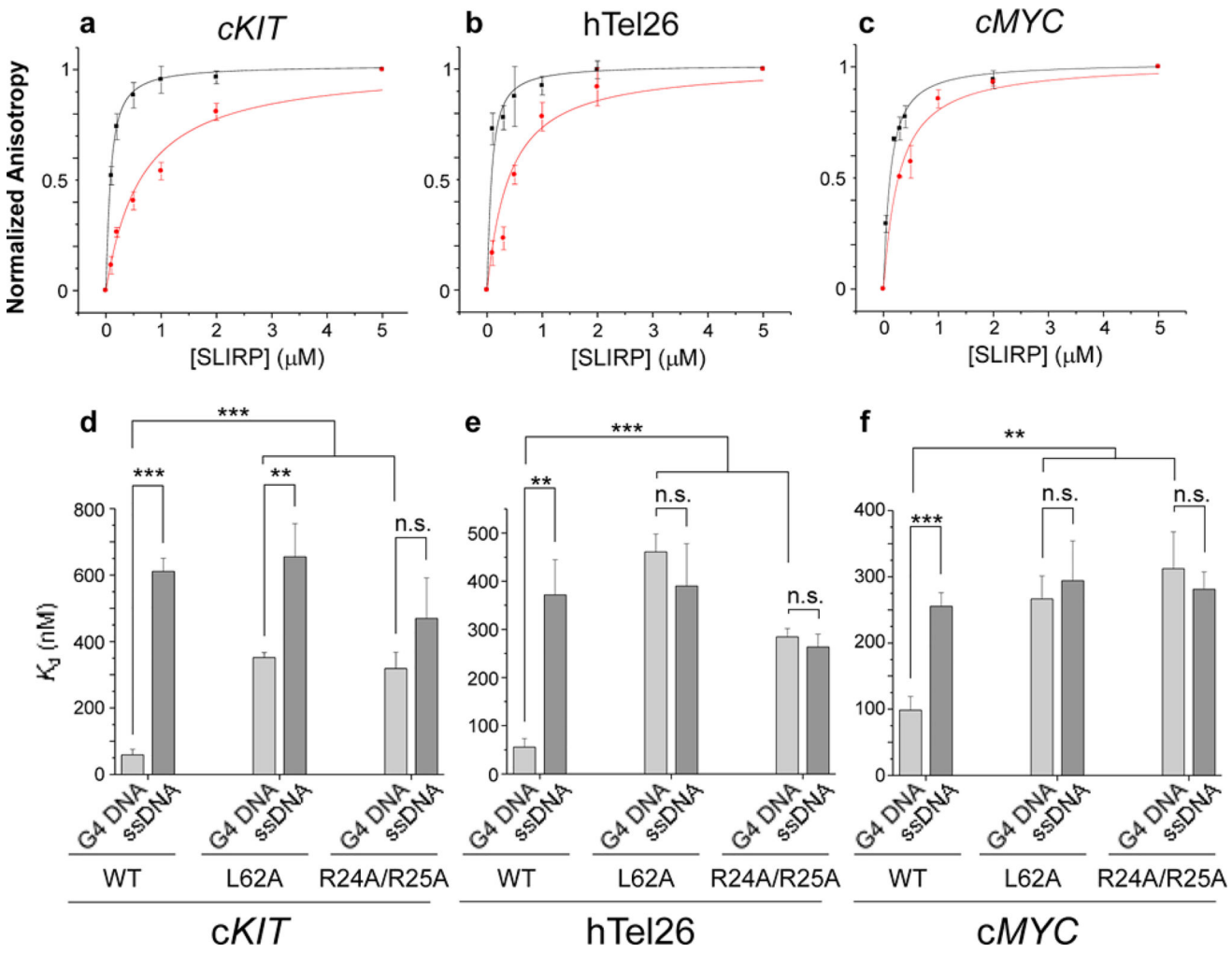

Figure 3.

Fluorescence anisotropy for measuring the K_d values for the binding of wild-type and mutant SLIRP proteins toward G4 structures derived from the promoters of *cKIT* and *cMYC* genes as well as the human telomere (black symbols and curves in panles a–c) and the corresponding mutated sequences that cannot fold into G4 structures (red symbols and curves in panels a–c). The quantification data in panels d–f represent the mean ± SD of results obtained from three separate measurements. **, $p < 0.01$; ***, $p < 0.001$. The p values were calculated using two-tailed, unpaired Student's t-test.

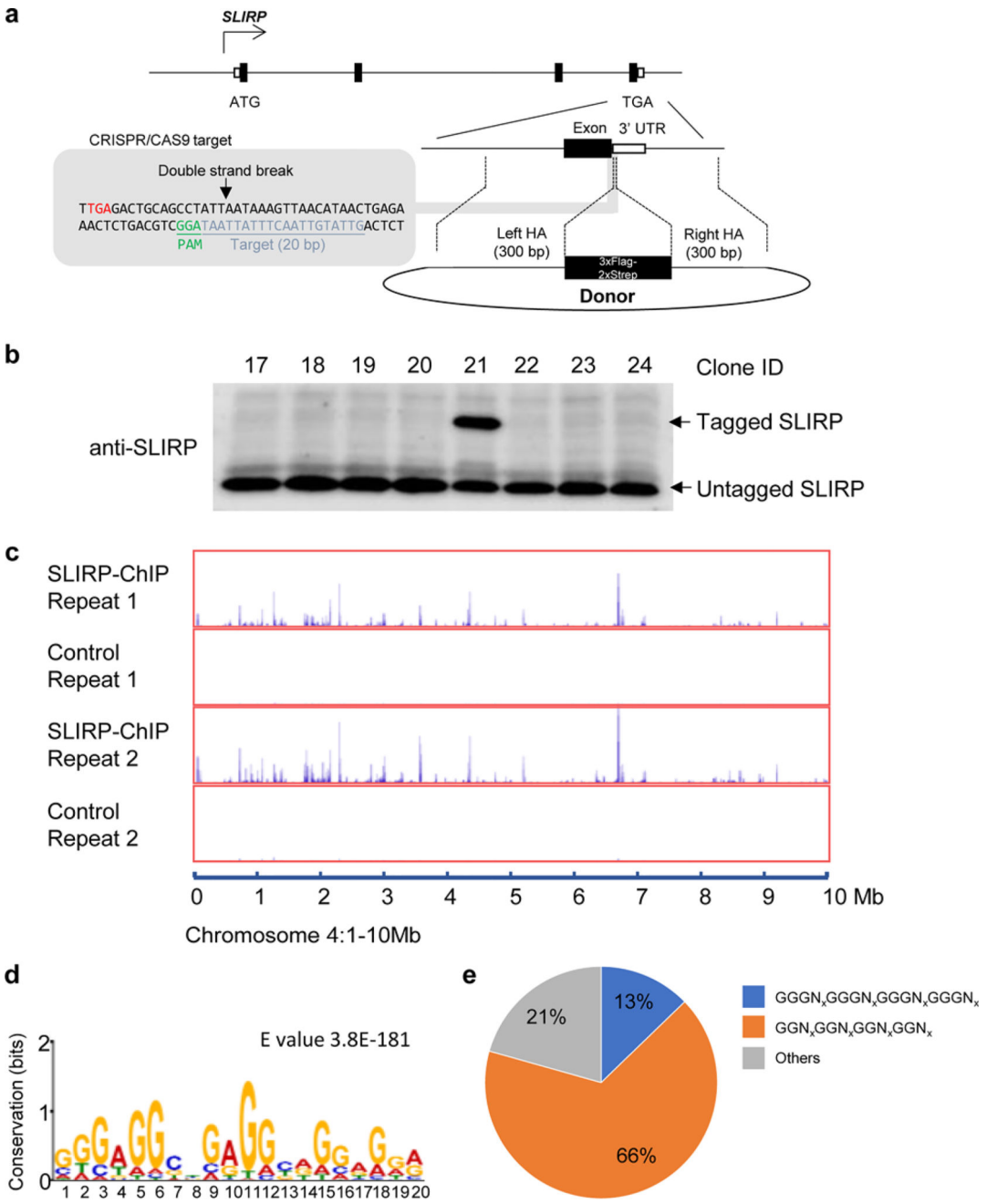

Figure 4.

CRISPR-Cas9-based integration of tandem affinity tag (3 × FLAG, 2 × Strept) to endogenous SLIRP and ChIP-Seq for monitoring the genome-wide occupancy of SLIRP. (a) Design of a CRISPR construct for targeting the endogenous locus of *SLIRP* gene. (b) Western blot revealed the successful incorporation of tandem affinity tag to SLIRP protein in clone 21. (c) Representative data to show the SLIRP peaks on a region of chromosome 4 from two biological replicates and the corresponding ChIP-Seq data obtained from IgG

control. (d) Sequence motif identified from ChIP-Seq reads. (e) Distributions of G4-folding motifs obtained from ChIP-Seq analysis.