Nucleic Acids Research, 2024, 1–14
https://doi.org/10.1093/nar/gkae1001
Chemical Biology and Nucleic Acid Chemistry

Profiling of i-motif-binding proteins reveals functional roles of nucleolin in regulation of high-order DNA structures

Yuki Ban[1,†], Yuka Ando[1,†], Yuma Terai[1], Risa Matsumura[1], Keita Nakane[2], Shigenori Iwai[1], Shinichi Sato[2] and Junpei Yamamoto[1,*]

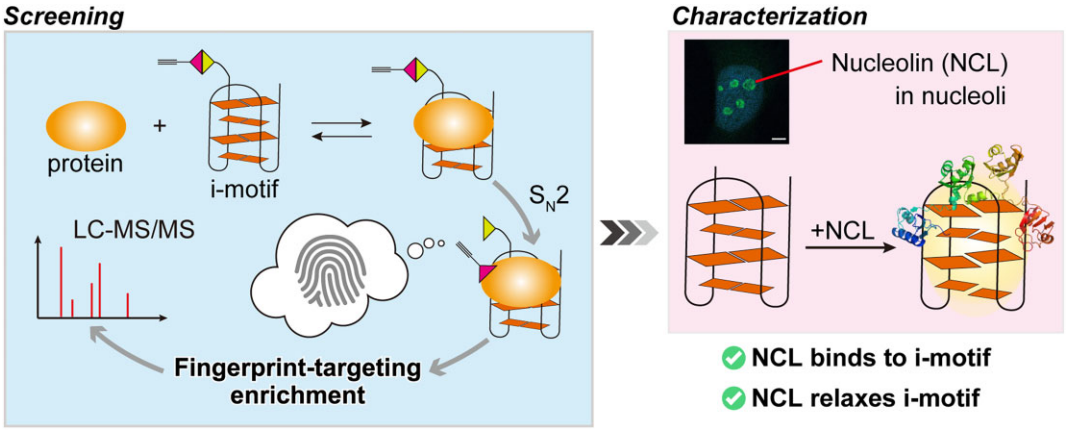
[1]Division of Chemistry, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 6-3 Aramaki aza-Aoba, Aoba-ku, Sendai, Miyagi 980-8578, Japan

[*]To whom correspondence should be addressed. Tel: +81 6 6850 6240; Fax: +81 6 6850 6240; Email: yamamoto.junpei.es@osaka-u.ac.jp
[†]The first two authors should be regarded as Joint First Authors.

Abstract

Non-canonical DNA structures, such as the G-quadruplex (G4) and i-motif (iM), are formed at guanine- and cytosine-rich sequences, respectively, in living cells and involved in regulating various biological processes during the cell cycle. Therefore, the formation and resolution of these non-canonical structures must be dynamically regulated by physiological conditions or factors that can bind G4 and iM structures. Although many G4 binding proteins responsible for tuning the G4 structure have been discovered, the structural regulation of iM by iM-binding proteins remains enigmatic. In this study, we developed a protein-labeling DNA probe bearing an alkyne moiety through a reactive linker, for proximity-labeling of nucleic acid-binding proteins, and searched for new iM-binding proteins. Alkyne-modified proteins in the nuclear extract of HeLa cells were labeled with biotin via a click reaction and then captured with streptavidin-coated magnetic beads. This fingerprint-targeting enrichment, followed by proteome analyses, identified new candidate proteins that potentially bind to the iM structure, in addition to the reported iM-binding proteins. Among the newly identified candidates, we characterized a nucleolar protein, nucleolin, that binds to the iM structure and relaxes it, while nucleolin stabilizes the G4 structure.

Graphical abstract



Introduction

Non-canonical high-order DNA structures are formed in the guanine- and cytosine-rich regions of telomeres and promoter sequences of proto-oncogenes and are called G-quadruplexes (G4) and intercalated motifs (i-motifs, iM), respectively (Figure 1) (1,2). G4 is composed of planar G-tetrads, each containing four guanine bases connected through Hoogsteen hydrogen bonds, and is stabilized by both π-π stacking interactions of the G-tetrad and monovalent cations captured in the center of the G-tetrad. In contrast, iM is formed by intercalations of hemi-protonated cytosine-cytosine base pairs ($C{\cdot}C^+$), thus requiring acidic conditions for its formation *in vitro*. These non-canonical DNA structures are reportedly stabilized under multi-molecular crowding conditions that mimic the intracellular environment (3,4). Although biochemical analyses using G4- and iM-forming oligonucleotides have shown that

Received: June 4, 2024. **Revised:** September 18, 2024. **Editorial Decision:** October 12, 2024. **Accepted:** October 17, 2024

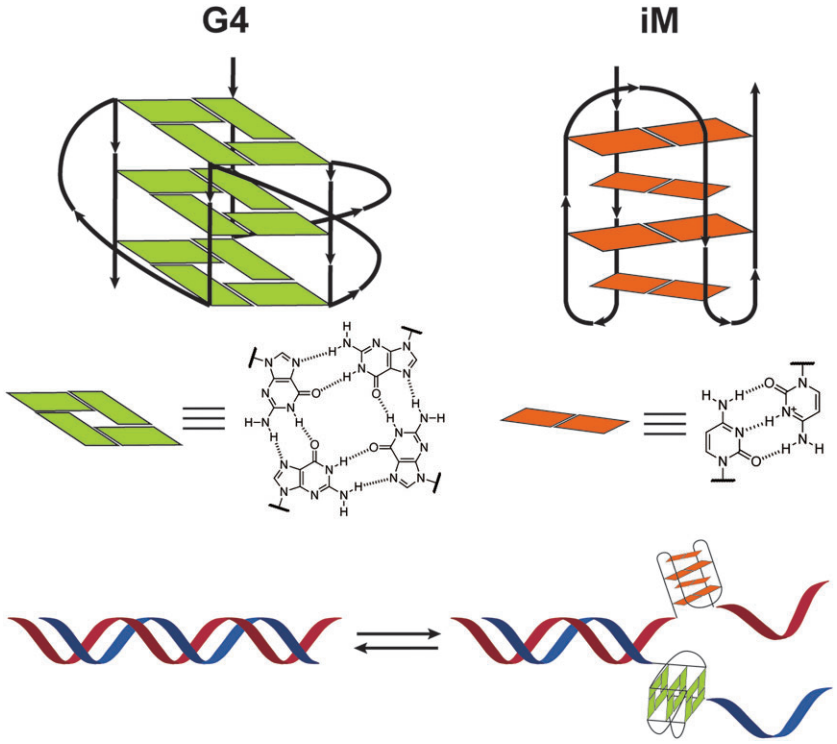


Figure 1. G-quadruplex (G4) and i-motif (iM) DNA structures. The formation and resolution of these structures are regulated by various factors in cells.

G4 and iM formation significantly inhibited telomere elongation, DNA replication, and transcription (5,6), recent cellular investigations using antibodies specific to G4 and iM (BG4 and iMab, respectively) have clarified the functions of these high-order DNA structures in cells. Genome-wide sequencing studies have revealed that G4 and iM are mainly present in open chromatin regions, especially in highly expressed genes for G4 and transcriptionally repressed genes for iM (7,8). Immunostaining of G4 and iM has provided further insights into their cell cycle-dependent formation in the S- and G1-phases, respectively (9–11). These findings comprehensively suggest that G4 and iM formation in cells is dynamic and depends on the cellular conditions, meaning that there are machineries that facilitate the formation and resolution of the high-order DNA structures associated with various biological processes.

Proteins that interact with G4 have been identified over the past two decades. For example, yeast Rap1 was originally discovered as a transcriptional regulator and telomere-binding protein but later recognized as a G4-binding protein (G4BP) (12,13). This protein stabilizes telomeric G4 DNA structures and therefore negatively regulates telomere length. However, several proteins are known to unfold G4. An RNA helicase associated with an AU-rich element, also known as DHX36, was characterized as a G4BP that unwinds G4 using a DHX36-specific motif (14,15). In addition to DHX36, some disease-related DNA helicases, such as WRN, BLM and Pif1, unwind G4 and eliminate its suppressive effects in biological functions (16). Many G4BPs have been reported (17), and some are summarized in the G4 structure-interacting protein database (18), enabling us to search for potential target proteins that interact with G4. In addition, new techniques for searching G4BPs in live cells have recently been developed (19,20), opening a new stage in G4 biology.

Contrary to the well-studied G4BPs, a limited number of iM-binding proteins (iMBPs) have been reported, such as heterogeneous nuclear ribonucleoprotein (hnRNP) LL (21), hnRNP A1 (22), hnRNP K (23), DNA helicases (24) from Homo sapiens and interleukin enhancer binding factor (ILF) from the silkworm *Bombyx mori* (BmILF) (25). The regulation of the iM structure by some of these proteins has been investigated. The hnRNP LL protein was identified as a transcriptional activator for the *BCL2* gene upon binding to iM and converted it into a flexible hairpin structure (21). Further analyses revealed that hnRNP LL transiently induced iM formation in the initial stage of binding to the *BCL2* promoter and then destabilized its structure with time (26). The hnRNP K protein recognizes iM in the *c-MYC* promoter and activates transcription by unfolding the iM structure (27). In contrast, hnRNP K reportedly stabilizes iM formation in the long terminal repeat promoter of the human immunodeficiency virus-1 proviral genome, modulating viral transcription (28). The BmILF protein was found to stabilize the iM structure and promote the transcription of the *POUM2* gene, which is dependent on iM formation (25). Similar to G4BPs, some iMBPs have also been shown to function as regulators of biological processes, by either stabilizing or unfolding the iM structure.

Toward discovery of new iMBPs, proteomic analyses after enrichment processes can be a powerful strategy. With respect to the iMBPs reported to date, however, only hnRNP LL and BmILF were identified by affinity-based pull-down enrichment, in which iMBPs were captured from cell extracts using iM-forming oligonucleotides as bait, and the bound proteins were subjected to mass spectrometry. The other reported proteins were investigated based on the common hypothesis that some G4BPs may also bind to iM and actually did so. The very few successful examples of iMBP identification indicate that solid strategies for screening iMBPs have not yet been

established. The potential difficulties of applying the canonical pull-down enrichment method to iMBPs may be due to the structural transition of the iM oligonucleotides on beads during the washing process, since the iM structure formed by the iM-forming sequences is significantly affected by solution conditions (6). Another possible explanation is that unidentified iMBPs may exhibit weak and/or transient DNA–protein interactions, and thus the affinity-based pull-down enrichment would not be effective. New strategies for screening iMBPs are needed to broaden the scope of iM biology.

Recently, proximity-dependent biotinylation by engineered biotin ligases such as BioID (29) and TurboID (30) has been used to identify interacting partners of proteins of interest (POIs), and proteome analysis after the enrichment of biotinylated proteins with streptavidin beads can provide an interactome overview for the target protein (31). New G4BPs have also been discovered using this strategy, in which the G4-binding motif of DHX36 is fused with miniTurbo to biotinylate proteins proximal to G4 (20). Similarly, proximity-dependent protein modification with clickable azide/alkyne moieties using G4-ligand derivatives successfully enriched proteins near G4, after further modification of the azide/alkyne moieties on the protein with biotin via click chemistry (19,32). Although the application of these strategies will be promising in identifying new iMBPs, the hidden hypothesis that the protein/ligand binding to G4/iM will not significantly inhibit the interactions of other proximal proteins remains untested, especially for iMBPs.

Here, we report the screening of iMBPs from nuclear extracts of HeLa cells, using chemically modified iM-forming oligonucleotides bearing an alkyne moiety through a reactive linker for the proximity-labeling of iMBPs. Upon protein binding to the native iM architecture, nucleophilic amino acid side chains attack the reactive linker, resulting in the transfer of the alkyne moiety to the protein surface. The presence of an alkyne moiety in proteins can therefore be regarded as a fingerprint indicating a history of proximity. Enrichment was achieved by targeting the fingerprint, and subsequent proteomic analyses indicated that many proteins, including known iMBPs, were enriched. We focused on the nucleolar protein nucleolin (NCL) and biochemically characterized its binding and function with iM. The results revealed that NCL indeed bound iM, and interestingly, in contrast to G4, it relaxed the iM structure upon binding. Thus, this study provides a first step for future research in iM biology, in addition to potential applications in the screening of proteins that bind other nucleic acids.

Materials and methods

General

Reagents and solvents were purchased from FUJIFILM Wako Pure Chemical Corporation, Sigma-Aldrich, and Tokyo Chemical Industry unless specified otherwise. Oligonucleotides used in this study were purchased from FASMAC (Supplementary Table S1). High-performance liquid chromatography (HPLC) analyses were performed using a Gilson gradient-type analytical system equipped with a Waters 2998 photodiode-array detector. DNA samples were analyzed and purified using a Waters μBondasphere C18 5 μm 300 Å column (3.9 mm × 150 mm), at a flow rate of 1.0 ml min^{-1} with a linear gradient of acetonitrile in 0.1 M triethylammonium acetate (pH 7), generated over 20 min. NanoLC-MS/MS analysis was performed by an LC-nano-ESI-MS system comprising a quadrupole time-of-flight mass spectrometer (Orbitrap Fusion, Thermo Fisher Scientific) equipped with a nanospray ion source and a nanoLC system (Easy-nLC 1000, Thermo Fisher Scientific). The LC-MS/MS data were acquired in a data-dependent acquisition mode, controlled by the Xcalibur 4.3 software (Thermo Fisher Scientific). The absorption and circular dichroism (CD) spectra were measured using a V730 spectrophotometer and J-805 spectropolarimeter (JASCO), respectively.

Preparation of alkyne-bearing oligonucleotides through tosylate linker

The oligonucleotides iM-WT and iM-Mut (Entries 1 and 2) bearing an amino linker at the 5′ terminus were dissolved in 547.2 μl of 100 mM carbonate/bicarbonate buffer (pH 8.5), and then 60.8 μl of a 0.1 M DMSO solution of alkyne-tosylate-NHS ester **1** (see Supplementary synthetic procedures) was added to the oligonucleotide solutions. The mixtures were incubated overnight at room temperature. After gel filtration on NAP-10 columns (Cytiva) to remove excess reagents, the alkyne-bearing oligonucleotides were purified by HPLC with a 3–20% acetonitrile gradient. The oligonucleotides were quantified from the UV absorption spectra (Supplementary Figure S1). Prior to use for protein labeling, the modified oligonucleotides (40 μg each) were dissolved in 65 μl of 50 mM Tris-acetate buffer (pH 4.1), heated up at 95°C for 10 min, and then cooled to 10°C over 1 h.

Preparation of nuclear extract

HeLa cells were cultured in DMEM (High Glucose) (Fujifilm) supplemented with 10% heat-inactivated fetal bovine serum (Gibco), 1% penicillin-streptomycin (Pen Strep) (Gibco) and 1% MEM Non-Essential Amino Acids (MEM NEAA) (Gibco). The cells were grown at 37°C in a 5% CO_2 atmosphere. A total of 12 dishes (10 cm ϕ) of HeLa cells were washed with ice-cold phosphate-buffered saline (PBS), and then the cells were collected, flash-frozen and kept at –80°C until use. Nuclear extracts of HeLa cells were prepared using an EzSubcell Extract Kit (ATTO), according to the manufacturer's instructions. Fractionation of the cellular components was confirmed by Western blotting (Supplementary Figure S2), and the protein concentration of the nuclear extract was determined by the Bradford assay, using the TaKaRa Bradford Protein Assay Kit (Takara).

Western blotting

Protein samples were mixed with an equal volume of sodium dodecyl sulfate (SDS) buffer (0.25 M Tris-HCl (pH 6.8), 0.5 M dithiothreitol (DTT), 10% SDS, 0.25% bromophenol blue, and 50% glycerol), and denatured at 95°C for 5 min. The samples were loaded onto an Extra PAGE One Precast Gel (Nacalai Tesque), along with either Precision Plus Protein Dual Color Standard markers (Bio-Rad) or ExcelBand All Blue Broad Range Plus Protein Marker (SMOBIO), and electrophoresis was performed. Proteins in the gel were transferred onto an Immobilon-P Transfer Membrane (Merck Millipore) treated with methanol. After blotting, the membrane was successively treated with SuperBlock™ (PBS) Blocking Buffer (Thermo Scientific) at 4°C overnight, and a primary antibody (either acetylated α-Tubulin (6–11B-1) mouse

monoclonal IgG (sc-23950, Santa Cruz Biotechnology), HSP60 (LK1) mouse monoclonal IgG (sc-59567, Santa Cruz Biotechnology), Lamin A/C (E-1) mouse monoclonal IgG (sc-37628, Santa Cruz Biotechnology) or horseradish peroxidase (HRP)-Conjugated Streptavidin (N100, Thermo Scientific)) diluted with PBS-T, at room temperature for 1 h, followed by washing with PBS-T three times. If required, the membrane was further treated with an anti-mouse IgG HRP Conjugate (W402B, Promega) diluted with PBS-T at room temperature for 1 h. The bands were visualized using an ImmunoStar LD Kit (Fujifilm), and chemiluminescent images were obtained using an Image Quant LAS 500 imager (Cytiva).

Sample preparation for nanoLC-MS/MS

Mixtures of lysis buffer (280.5 μl, 50 mM HEPES-KOH (pH 7.9), 50 mM KCl, 1 mM EDTA, 500 μM phenylmethylsulfonyl fluoride, 1 mM DTT, and 0.132% NP-40) and 54.5 μl of nuclear extract (500 μg), in the absence or presence of poly(dI·dC) (40 μg), were incubated at 37°C for 10 min. The abovementioned solution of the alkyne-bearing WT or Mut oligonucleotides was then added to the mixture, and the protein labeling reaction was performed at 37°C for 24 h. After mixing the oligonucleotide and nuclear extract solutions, the pH value of the solution was confirmed to be 6.

Proteins were precipitated using methanol/chloroform treatment. First, the mixture was equally divided (200 μl each) and transferred to 2 ml tubes. Methanol (800 μl) was added to the reaction mixture (200 μl), and the mixture was vortexed for 1 min, followed by the addition of chloroform (200 μl), and vortexing for 1 min. Water (600 μl) was then added, and the mixture was vortexed for 1 min. After centrifugation (15 000 *g* for 3 min), the upper layer was removed by pipetting, and methanol (800 μl) was added. The solution was mixed by inverting the tubes and centrifuged (15 000 *g* for 3 min). The supernatant was removed and the precipitated proteins were dried using a SpeedVac centrifugal concentrator (Thermo Fisher Scientific).

The alkyne-bearing proteins were labeled with biotin via a click reaction. The precipitates of the recovered proteins in each tube were dissolved in 150 μl of phase transfer surfactant (PTS) buffer (100 mM Tris-HCl (pH 8.0), 12 mM sodium deoxycholate and 12 mM sodium N-lauroyl sarcosinate). After sonication for 5 min and heating at 95°C for 10 min, 600 μl of 100 mM Tris-HCl (pH 8.0) was added. The solutions in the two tubes were combined in a 2 ml tube. To the solution, 100 μl of a click mixture (300 μl DMSO, 1.2 ml tert-butanol, 15 μl 170 mM Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine dissolved in DMSO, 500 μl 50 mM $CuSO_4$, 25 μl 10 mM DADPS biotin azide (Fisher Scientific) dissolved in DMSO and 500 μl of 50 mM Tris(2-carboxyethyl)phosphine hydrochloride) was added, followed by vortexing for 1 min. After two more rounds of click mixture addition, the resulting solution was incubated at 37°C overnight. The supernatant of the reaction mixture was concentrated using an Amicon Ultra filter (0.5 ml, MWCO 10 kDa; Merck Millipore) and then combined with the precipitates. After adjusting the volume to 200 μl by adding RIPA buffer (Nacalai Tesque), proteins were recovered by methanol/chloroform precipitation as mentioned above. The protein precipitates were dissolved in 150 μl of PTS buffer, sonicated for 5 min and heated at 95°C for 10 min. At this stage, an aliquot (3 μl) of the protein solution was analyzed by Western blotting, to confirm the presence of the biotin modification on proteins (Supplementary Figure S3).

The biotin-modified proteins were enriched using streptavidin beads. Sera-Mag Magnetic Streptavidin Coated Particles (50 μl, Cytiva) were washed with pre-wash buffer (PTS buffer/100 mM Tris-HCl (pH 8.0) = 1/4, v/v), three times. After removal of the pre-wash buffer, 150 μl of the protein solution and 600 μl of 100 mM Tris-HCl (pH 8.0) were added, and the mixture was incubated at room temperature for 1 h. The beads were successively washed with PTS buffer, washing buffer 1 (20 mM HEPES-OH (pH 7.5), 100 mM NaCl, and 1 M urea), and washing buffer 2 (20 mM HEPES-OH (pH 7.5), 1 M NaCl).

The beads were resuspended with 200 μl of 100 mM Tris-HCl (pH 8.0), and 2 μl of 1 M DTT was added. After 30 min of incubation at room temperature, 10 μl of 1 M iodoacetamide was added for reductive alkylation of the cysteine side chains, and the mixture was further incubated for 30 min. To the solution, 2 μg of trypsin was added for on-bead digestion of the enriched proteins, and the mixture was incubated at 37°C overnight. Detergents in the solution were removed with a HiPPR Detergent Removal Resin Column Kit (Thermo Fisher Scientific). The supernatant was loaded onto resin equilibrated with 100 mM Tris-HCl (pH 8.0), and the peptides were recovered by centrifugation (1500 *g* for 2 min). The peptide solution was then passed through an SPE C-Tip KT200 (C18) (Nikkyo Technos Co., Ltd.) for desalting, and the eluates were evaporated using the SpeedVac centrifugal concentrator. The peptide fragments were dissolved in 20 μl of acetonitrile/trifluoroacetic acid (TFA) aqueous solution (CH_3CN/0.1% TFA aqueous solution = 5/95, v/v), and subjected to nanoLC-MS/MS analysis.

NanoLC-MS/MS acquisition and data analyses

NanoLC-MS/MS analysis was performed using an LC-nano-ESI-MS system comprising a quadrupole, Orbitrap, and ion trap Tribrid mass spectrometer (Orbitrap Fusion; Thermo Fisher Scientific) equipped with a nanospray ion source and nano HPLC system (Easy-nLC 1000; Thermo Fisher Scientific). The trap column used for the nano HPLC was a 2 cm × 75 μm capillary column packed with 3 μm C18-silica particles (Thermo Fisher Scientific), and the separation column was a 12.5 cm × 75 μm capillary column packed with 3 μm C18-silica particles (Nikkyo Technos Co., Ltd.). The micropump (flow rate: 300 nL min^{-1}) gradient method was used as follows: mobile phase A, 0.1% formic acid; mobile phase B, 80% acetonitrile, aqueous 0.1% formic acid; 0–1 min: 0–5% B, 1 − 61 min: 5–40% B, 61–63 min: 40–95% B, 63–80 min: 95% B. The settings for the data-dependent acquisition were as follows: maximum injection time, 0.05 s; full MS (MS1, orbitrap) scan range, 375–1500 m/z, excluding the former target ion for 20 s; and 10 ppm of mass tolerance. The top speed cycle time was set to 3 s and the maximum MS2 scan was acquired within that time. The MS2 (isolation: quadrupole, detection: ion trap) scan maximum injection time and range were 0.035 s, and the m/z range was defined based on the MS1 scan. Measurements were performed in triplicate for each sample and the raw data were deposited in jPOSTrepo (see Data Availability) (33).

Identification and label-free quantification of the proteins were conducted using the Proteome Discoverer 3.0 software embedded with the Sequest algorithm (Thermo Fisher

Scientific). A list of human proteins was obtained from the UniProt database (taxonomy 9606, downloaded on September 13, 2023). Label-free quantification was performed using the workflows shown in Supplementary Figures S4 and S5 for the processing and consensus steps, respectively.

The volcano plot analysis was performed using the Perseus software (version 1.6.15.0). The settings of the parameters were as follows: 'Transform' was 'log2(x)', 'Filter rows based on valid values' was 'Min. number of values: 4', 'Imputation, Replace missing values from normal distribution' was 'Width: 0.3, Down shift: 1.8, Mode: Separately for each column'.

Gene ontology (GO) analysis was performed using Metascape (https://metascape.org; version 3.5, database last update date: May 1, 2024) (34). Proteins found in the $x > 0$ region in the volcano plot; that is, those enriched with the iM-WT oligonucleotide, were extracted and summarized in lists, which were further used as inputs. The output lists containing the proteins annotated with GO terms were manually analyzed, and the number of proteins with the GO terms DNA binding (GO:0003677), RNA binding (GO:0003723) and nucleus (GO:0005634) were counted to prepare Venn diagrams. The proteins in the lists were analyzed using UniProt (https://www.uniprot.org) to link each protein with UniProt-Knowledgebase (KB) keywords.

Preparation of recombinant GST-tagged nucleolin

A codon-optimized human NCL cDNA that encodes the amino acid sequence from Met284 to the stop codon, lacking the N-terminal domain (NTD; Met1–Glu283), ($NCL_{\Delta NTD}$), was purchased from Genewiz (Supplementary Figure S6). The regions encoding $NCL_{\Delta NTD}$ and NCL lacking both the NTD and the C-terminal RGG domain (Pro647–Glu710) ($NCL_{\Delta NTDRGG}$) were amplified by polymerase chain reaction, using PrimeSTAR GXL polymerase (Takara) and the primer sets shown in Supplementary Table S1 (Entries 8 and 9 for $NCL_{\Delta NTD}$, 8 and 10 for $NCL_{\Delta NTDRGG}$). Each amplicon was inserted into the pGEX-4T-1 plasmid linearized at the BamHI/XhoI sites, using an In-Fusion Cloning Kit (Takara). *Escherichia coli* Stellar competent cells (Clontech) were transformed with the reaction mixture, and the obtained plasmids were sequenced and used for protein production. For GST production, the native pGEX-4T-1 plasmid was used.

Escherichia coli BL21(DE3) star competent cells were transformed with the plasmids and grown in 1 l of Luria-Broth (LB) medium containing 50 μg ml^{-1} ampicillin at 37°C, until the turbidity at 660 nm reached 0.5–0.6. For induction, 1-β-isopropylthiogalactopyranoside was added at a final concentration of 100 μM, and the culture was continued at 25°C. After 24 h, cells were harvested by centrifugation (5000 *g* for 15 min), and the pellets were frozen with liquid nitrogen and stored at −80°C. The pellets were thawed on ice, 30 mg of lysozyme was added, and after suspension in lysis buffer (40 ml, consisting of 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM DTT and 5% glycerol, pH7.4), the cells were lysed by sonication. Next, 2.5 ml of 20% Triton X-100 was added, and the mixture was further incubated for 30 min on ice. Cell debris was removed by centrifugation (20 000 *g* for 1 h), and the supernatant was loaded onto a Glutathione Sepharose 4B column. The resin was washed with lysis buffer, and the GST-tagged proteins were eluted with elution buffer (20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 20 mM glutathione and 5% glycerol, pH 7.4). The pooled fractions were loaded onto a HiTrap Heparin HP column (Cytiva) and purified with a linear gradient of 100–1000 mM NaCl in buffer containing 20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT and 5% glycerol, pH 7.4. Fractions containing the target protein were concentrated, and the protein concentration was determined using the Bradford assay.

The GST protein was purified according to the above procedure, but with the omission of the Heparin affinity chromatography.

Electromobility shift assay

The iM- or G4-forming oligonucleotides with the Cy5 fluorophore at the 5′ terminus (Entries 3 and 4, Supplementary Table S1) were dissolved at 1 μM in either 50 mM Tris-acetate buffer at pH 4.1 or 100 mM KCl, respectively, and were heated at 95°C for 10 min, followed by slowly cooling down overnight. The iM oligonucleotides (100 nM) were mixed with 0–500 nM $NCL_{\Delta NTD}$ or 0–1 μM $NCL_{\Delta NTDRGG}$, in buffer containing 10 mM Tris-HCl, 50 mM KCl, 4 mM EDTA, 5 mM $MgCl_2$, 1 mM DTT, 0.05% NP-40 and 2.5% glycerol, pH 6.0 or 7.0, and incubated on ice for 1 h. For the binding of NCLs to G4, the G4 oligonucleotides (100 nM) were mixed with NCLs in buffer containing 10 mM Tris-HCl, 1 mM EDTA, 1 mM DTT and 25 mM KCl, pH 7.4, and incubated on ice for 1 h. The mixtures were then subjected to 7.5% polyacrylamide gel electrophoresis, and the fluorescent bands were visualized using an FLA 7000 image analyzer (Fujifilm). The band intensities were quantified using the Multi-Gauge software version 3.1 (Fujifilm), and the bound fractions were obtained. Experiments were performed in triplicate, and the average bound fractions were plotted against the protein concentrations. Data were fitted using Hill's function.

Primer-extension by DNA polymerase

A primer (5 μM) with the Cy5 fluorophore at the 5′ terminus, d(GCTTCGGCTTAATACGA) (Entry 7 in Supplementary Table S1), was hybridized with either the iM- or G4-forming template oligonucleotide (5 μM, Entries 5 and 6 in Supplementary Table S1, respectively) in 40 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer at pH 6.0 or 7.0. The primer-template oligonucleotides (100 nM) were mixed with $NCL_{\Delta NTD}$ (500 nM) in buffer (50 μl), containing 40 mM MES (pH 6.0 or pH 7.0), 8 mM $MgCl_2$, 50 mM NaCl and 100 μM dNTP mixture, and 1 unit of Klenow fragment of *E. coli* DNA polymerase I lacking the exonuclease activity (KF_{exo-}) was added to initiate the primer-extension reaction. After 1, 5 and 10 min of incubation at 37°C, aliquots (10 μl) were deproteinized with phenol/chloroform/isoamyl alcohol (25/24/1, v/v/v), and the DNA was recovered by ethanol precipitation. The DNA was dissolved in 95% formamide containing 0.5 M EDTA, and the products were analyzed by denaturing polyacrylamide gel electrophoresis on a 10% gel containing 7.5 M urea. Bands were visualized using the FLA 7000 image analyzer. Band intensities were quantified by the Multi-Gauge software. Experiments were performed in triplicate.

Results

Design of proximity-labeling oligonucleotides containing reactive linker

We devised a reaction-based proximity-labeling method for proteins present in close proximity to the iM, using chemi-

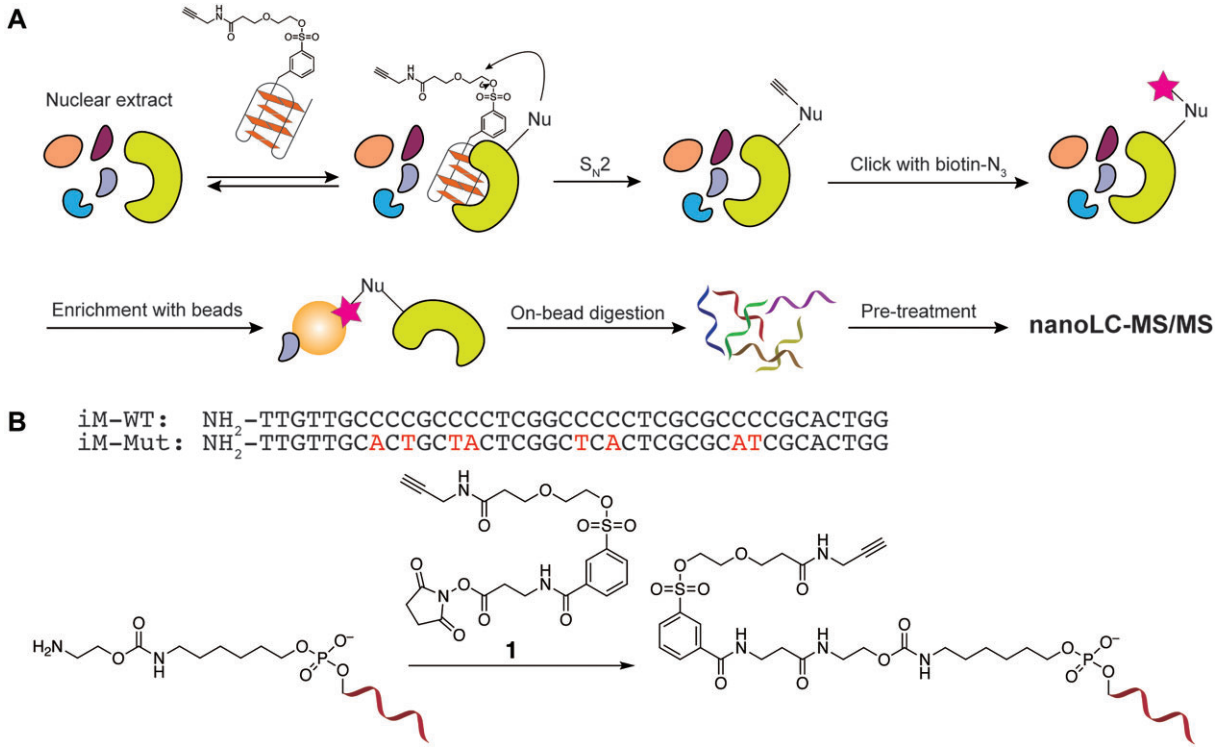


Figure 2. Screening of iMBPs by FiTE. (**A**) Schematic view of FiTE followed by nanoLC-MS/MS. During association/dissociation of the iMBP, the alkyne moiety on the oligonucleotide probe is transferred to the nucleophilic amino acid side chains (–Nu) of proteins located proximal to the probe. After biotinylation via click chemistry, followed by enrichment, peptide fragments of the enriched proteins are analyzed by nanoLC-MS/MS. (**B**) Conjugation of alkyne-tosylate-NHS ester **1** with 5′-amino-modified oligonucleotides. Some cytosine bases in the iM-WT sequence were replaced with adenine and thymine in the iM-Mut sequence.

cally modified oligonucleotides. At the 5′-terminus of the iM-forming oligonucleotide, an alkyne moiety was introduced through a reactive tosylate linker, which has been utilized for ligand-directed protein labeling (35,36), and the modified oligonucleotide was mixed with a cellular nuclear extract. During the association/dissociation of proteins with the oligonucleotide, nucleophilic amino acid side chains near the reactive linker may react with a carbon atom adjacent to the tosylate leaving group via the S_N2 reaction (Figure 2A), leading to the transfer of the alkyne moiety to proteins that are proximal to the iM oligonucleotide. Once this reaction occurs, the alkyne group is retained on the proteins and thus presents an archival record of their proximity to the iM, like a fingerprint. The alkyne moiety on the proteins can be a scaffold for bioorthogonal conjugation via copper-catalyzed alkyne-azide cycloaddition, which typically results in >90% yields (37). After further modification of the alkyne group with biotin via click chemistry, the biotinylated proteins are enriched with streptavidin beads. This fingerprint-targeting enrichment (FiTE), followed by proteomic analyses, facilitates the screening of a wide range of iMBPs without considering the selection pressures and undesirable effects of additional ligands/proteins. However, iMBPs are not the only proteins detected in the FiTE strategy. For example, proteins abundant in the extract will be promiscuously labeled with the alkyne group, because the S_N2 reaction kinetics exclusively relies on the concentrations of both the protein-labeling probe and proteins. The interacting partners of proteins directly bound to iM may also be labeled, depending on their position relative to the tosylate linker, in addition to iMBPs. Based on these features of the FiTE strategy, the efficiency of the alkyne transfer can vary depending on the nature of proteins. Therefore, the experimental and analytical conditions need to be chosen with consideration.

We first synthesized an oligonucleotide modifying reagent, in which a terminal alkyne moiety is connected with an *N*-hydroxysuccimide ester through a reactive *m*-tosylate linker (alkyne-tosylate-NHS ester **1**, see Supplementary synthetic procedure), and attached it to the 5′ terminus of an iM-forming oligonucleotide (Figure 2B and Entry 1 in Supplementary Table S1, iM-WT). For comparison, a mutant oligonucleotide (Entry 2 in Supplementary Table S1, iM-Mut), in which some cytosine bases in the iM-WT sequence were replaced with adenine and thymine (Figure 2B), was conjugated with **1**. The iM-WT sequence is derived from the promoter of the *POUM2* gene in the silkworm *Bombyx mori* (25), and is known to form an iM structure. The capability of iM formation in the iM-WT and iM-Mut oligonucleotides was confirmed by measuring CD spectra (Supplementary Figures S7 and S8), which showed a peak shift of the positive CD signal from 281 to 286 nm upon a pH drop from pH 8 to 4.5, in good agreement with the previous report (25). Note that it is advantageous to use the iM-forming DNA sequence from silkworm instead of the human iM-forming sequence for the screening of human iMBPs, because we can exclude the possibility of capturing the human promoter-specific factors and selectively capture proteins that bind the probe molecule in a structure-specific manner.

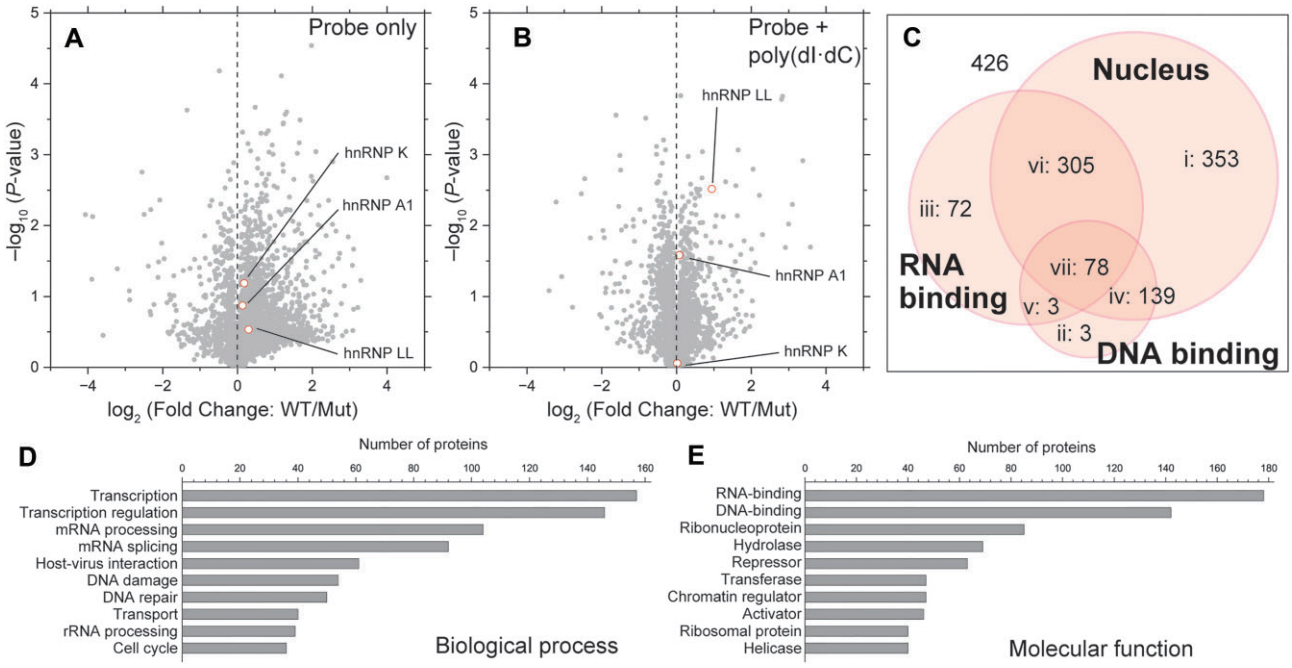


Figure 3. Proteome analyses of the proteins enriched via the FiTE strategy. (**A** and **B**) Volcano plots displaying proteins enriched with the iM-WT and –Mut oligonucleotides in the absence (**A**) or presence (**B**) of poly(dI·dC). The complete list of proteins in the volcano plot is summarized in Supplementary File 1. The known iMBPs are highlighted in the panel. (**C**) Venn diagram of proteins in the x > 0 region of the volcano plot (panel B) based on three GO terms, showing the overlap of annotated GO terms. The number of proteins in each area (i–vii) is shown, and lists of the proteins found in each area are summarized in Supplementary File 2. (**D** and **E**) Distribution of the top 10 UniProtKB keywords for biological processes (**D**) and molecular functions (**E**) of the proteins found at the intersections (areas iv–vii) in the Venn diagram.

Screening of iM-binding proteins

To screen for iMBPs, the iM-WT and Mut oligonucleotides bearing the alkyne-tosylate modification were incubated with the nuclear extracts of HeLa cells at pH 6 for iM formation (Supplementary Figure S8). Afterward, the alkyne-modified proteins were labeled with biotin via click chemistry (Figure 2A), and the presence of biotin on the recovered proteins was confirmed by Western blotting (Supplementary Figure S3), indicating that a nucleophilic reaction between the proteins and the tosylate linker on the oligonucleotide occurred. Biotinylated proteins were enriched and subjected to a shotgun analysis, using nanoLC-MS/MS after on-bead trypsin treatment.

Proteins detected in the iM-WT and iM-Mut samples were analyzed by a volcano plot using fold-changes in the protein abundances and the *P*-values, in order to characterize the distribution of the enriched proteins with the FiTE approach (Figure 3A). Each dot in the x > 0 and x < 0 regions represents a protein enriched more efficiently with the WT and Mut oligonucleotides than with the other oligonucleotide, respectively. Notably, the known iMBPs, hnRNP LL, A1 and K, were found in the x > 0 region (Figure 3A), corroborating that the reaction-based FiTE strategy for screening potential iMBPs was successful. One concern is the closeness of these known iMBPs to the x = 0 line in the volcano plot. It is worth mentioning that the protein labeling by the FiTE strategy is distinct from that in previous reports on the identification of G4BPs using G4 ligands derivatives (19,32). In the FiTE strategy, protein labeling proceeds via an S_N2 reaction and thus exclusively relies on physical collisions between nucleophilic amino acid side chains on the protein and the reactive tosylate linker on the DNA. Even if proteins could bind the i-motif DNA, there are possibilities that the proteins would rarely be labeled because of the absence of the nucleophilic amino acid side chains, and this could have happened with these known iMBPs.

Generally, nucleic acid-binding proteins that bind specifically to unique nucleic acids also exhibit non-specific binding abilities, which involve interactions with the DNA backbone (38). Since our strategy for screening iMBPs relies on the binding capabilities of proteins to the iM structure, such non-specific binding by proteins in the nuclear extract to the iM oligonucleotide would mask the DNA probe and prevent the structure-specific binding of inherent iMBPs. To suppress non-specific binding, double-stranded (ds) DNA is often used as a competitor (39). Therefore, the proximity labeling of nuclear extracts with the iM-WT and -Mut oligonucleotides bearing the alkyne-tosylate linker was performed in the presence of poly(dI·dC). The enriched proteins were analyzed in the same way as described above (Figure 3B). Among the known iMBPs, the fold change for hnRNP LL was shifted to positive with increased significance, as compared to those in Figure 3A, indicating that hnRNP LL binding to the iM structure became prominent upon the addition of the ds competitor. In contrast, the positions of hnRNP A1 and K shifted closer to the x = 0 line, likely due to the competitive binding to poly(dI·dC), although they are still present in the x > 0 region. These results indicate that new iMBPs will also be included in the x > 0 region of the volcano plot for the samples with poly(dI·dC).

To narrow down the potential iMBPs, an enrichment analysis was performed for the 1379 proteins detected in the x > 0 region of Figure 3B. The proteins were annotated with GO terms using Metascape (34). Because some iMBPs are known to be RNA-binding proteins, we exclusively focused on proteins annotated with three GO terms: Nucleus (875 proteins),

Table 1. Selected[a] G4 binding proteins enriched with the iM-WT oligonucleotide

UniProt ID	Gene Symbol	Description	Ref.
O14647	CHD2	Chromodomain-helicase-DNA-binding protein 2	(40)
P06748	NPM1	Nucleophosmin	(41,42)
P09429	HMGB1	High mobility group protein B1	(43,44)
P09651	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	(45)
P11387	TOP1	DNA topoisomerase 1	(46)
P19338	NCL	Nucleolin	(47,48)
P27695	APEX1	DNA-(apurinic or apyrimidinic site) endonuclease	(49)
P51532	SMARCA4	Transcription activator BRG1	(19)
P55265	ADAR	Double-stranded RNA-specific adenosine deaminase	(50)
Q00059	TFAM	Transcription factor A, mitochondrial	(51)
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	(52)
Q14103	HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	(53)
Q16666	IFI16	Gamma-interferon-inducible protein 16	(54)
Q9H2U1	DHX36	ATP-dependent DNA/RNA helicase DHX36	(55)

[a]G4BPs found in area vii in Figure 3C are summarized.

RNA binding (458 proteins) and DNA binding (223 proteins) and prepared a Venn diagram (Figure 3C). A total of 78 proteins belonged to the intersection of the three GO terms (Supplementary Table S2). Interestingly, hnRNP A1 and K also belonged to this intersection, whereas hnRNP LL was found at the intersection between Nucleus and RNA binding. These observations suggest that the previously reported iMBPs were properly narrowed down by the enrichment analysis, thus indicating that new iMBPs are included in the intersections. The distribution of the top UniProtKB keywords for the 525 proteins found in the intersections (areas iv–vii) is summarized in Figure 3D and E. The distributions of the biological processes and the molecular functions of the proteins indicate that a large proportion of these proteins are involved in RNA-related processes. It is noteworthy that various G4BPs were also enriched with this strategy, and in total, 14 characterized G4BPs were included in the intersection of the three GO terms (Table 1).

Although the known iMBPs are properly enriched with the FiTE strategy, further verification of iM DNA structure binding by the other proteins in the list is required. To select a candidate for biochemical characterization, we set reasonable criteria. Previous investigations of iMBPs (22–24) have shown that some G4BPs bind to iM. It is therefore advantageous to select one of the 14 G4BPs observed in this study (Table 1) so that we can compare the affinities and functions of the chosen protein toward the G4 and iM structures. The subcellular location of these proteins suggested by UniProt is mostly the nucleoplasm; however, interestingly, the second most abundant location is the nucleoli, which are the largest non-membrane sub-compartments in the nucleus and function in protein quality control (56). Because this observation reminds us of the involvement of G4 RNA in the formation of liquid-like condensates (57), the characterization of a potential iMBP localized in the nucleoli is of interest. Based on these criteria, we focused on NCL, one of the most abundant proteins in the nucleoli.

Biochemical characterization of nucleolin as a potential iM-binding protein

NCL consists of three domains: an NTD rich in acidic side chains, four RNA-binding domains (RBD1–4) and a C-terminal Arg-Gly-Gly (RGG) repeat (Figure 4A). Since the RBDs and RGG repeat are reportedly involved in binding to G4 (47), recombinant GST-tagged NCL proteins lacking only the NTD and both the NTD and RGG ($NCL_{\Delta NTD}$ and $NCL_{\Delta NTDRGG}$, respectively; Figure 4A and B) were prepared for the biochemical analyses. First, electromobility shift assay (EMSA) (58) was performed for these NCL proteins, using 5′-Cy5 labeled iM- and G4-forming oligonucleotides (Entries 3 and 4 in Supplementary Table S1). As reported previously (47), bands with low mobility were observed for the G4 oligonucleotide with increasing concentrations of $NCL_{\Delta NTD}$ (Figure 4C), confirming that our recombinant NCL proteins properly bind to the G4 structure.

Similarly, bands of the protein–DNA complex were also observed for the iM oligonucleotide at pH 6 (Figure 4D), whereas a sole GST protein lacking the NCL was unable to form the complex (Supplementary Figure S9A), suggesting that NCL proteins bind to the iM structure. To corroborate the observation, the binding of $NCL_{\Delta NTD}$ to the iM oligonucleotide was investigated at pH 7, where iM formation is remarkably suppressed (Supplementary Figure S8D). As a result, the formation of the complex with increasing amounts of $NCL_{\Delta NTD}$ was obviously reduced (Supplementary Figure S9B), as compared with that at pH 6 (Figure 4D). Although $NCL_{\Delta NTD}$ exhibited basal affinity to the non-structured oligonucleotide at higher protein concentrations (Supplementary Figure S9B), these results indicate that $NCL_{\Delta NTD}$ preferentially binds the iM structure. An EMSA using the $NCL_{\Delta NTDRGG}$ protein was also performed at pH 6 (Supplementary Figure S10), and the dissociation constants (K_d) of the NCL proteins for the G4 and iM oligonucleotides were estimated (Table 2). The apparent K_d values of $NCL_{\Delta NTD}$ for the G4 and iM oligonucleotides were 203 and 166 nM, respectively, indicating that NCL binds iM with similar affinity to G4. In the absence of the RGG domain, the affinities for both G4 and iM were reduced approximately by half (Table 2). These results demonstrated that the electrostatic interactions between the positively charged arginine residues in the RGG domain of NCL and the negatively charged DNA backbone reinforce the binding of NCL to these non-canonical DNA structures.

Modulation of iM stability upon NCL binding

NCL reportedly stabilizes the G4 structure upon binding (47); therefore, it is possible that NCL also modulates the structural stability of iM. Both G4 and iM reportedly inhibit DNA replication by the Klenow fragment of DNA polymerase I (lacking the exonuclease activity) in *E. coli* (KF_{exo-}) when these non-canonical DNA structures are formed in the template strand (59). This characteristic is beneficial for qualitatively analyzing whether the structural transition of the non-canonical DNA structures is inducible in the presence of NCL; that is, if NCL destabilizes the non-canonical DNA structures, then KF_{exo-} should be able to continue extending the primer DNA,

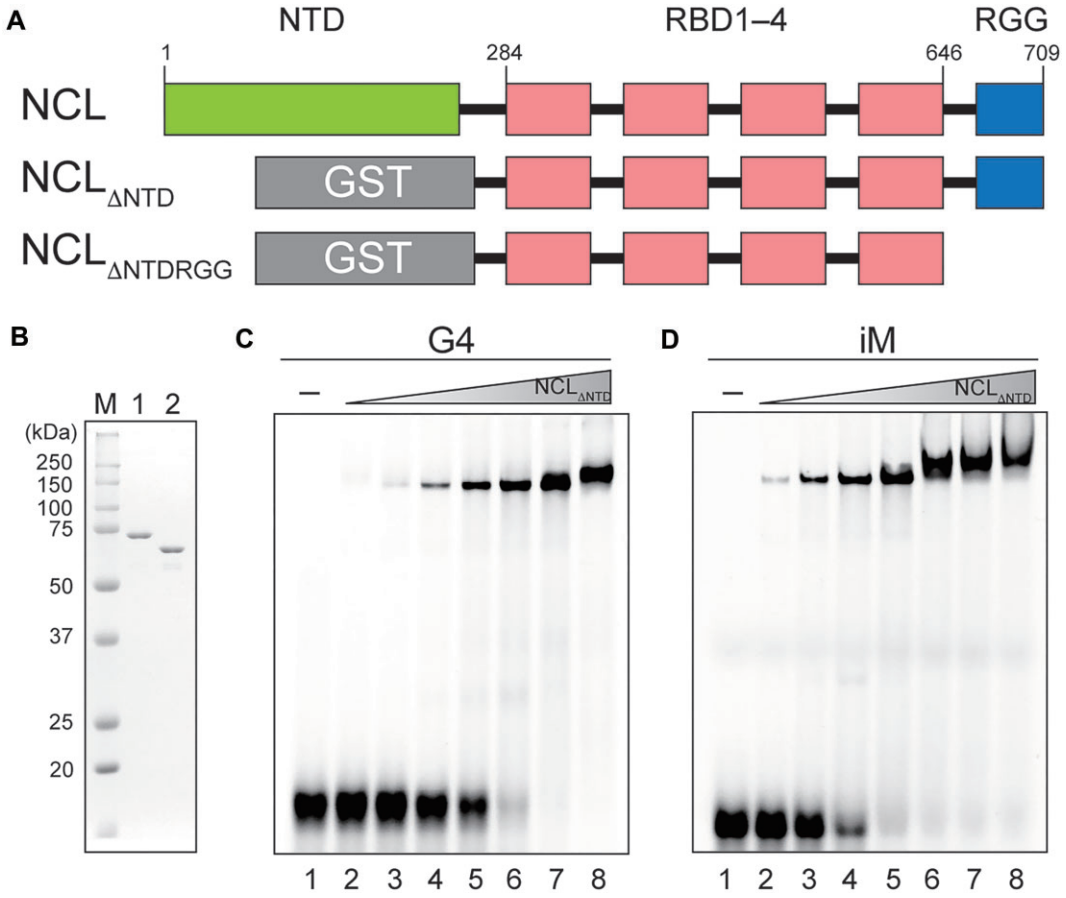


Figure 4. Binding of recombinant NCL proteins to G4 and iM oligonucleotides. (**A**) Binary structure of NCL. The full-length NCL consists of an NTD, four RNA binding domains (RBD1–4) and a C-terminal RGG repeat domain. In this study, two recombinant GST-tagged NCLs, one lacking the NTD (NCL$_{\Delta NTD}$), and the other lacking both the NTD and RGG (NCL$_{\Delta NTDRGG}$), were prepared. (**B**) Purified NCL$_{\Delta NTD}$ (lane 1) and NCL$_{\Delta NTDRGG}$ (lane 2) proteins. The proteins (0.5 μg each) were analyzed by 12% SDS-polyacrylamide gel electrophoresis (PAGE) and stained with Coomassie Brilliant Blue. (**C** and **D**) EMSAs for the binding of NCL$_{\Delta NTD}$ to G4 (**C**) and iM (**D**) oligonucleotides. Each oligonucleotide (100 nM) was incubated on ice with increasing amounts of NCL$_{\Delta NTD}$ (0, 50, 100, 150, 200, 250, 300 and 500 nM in lanes 1–8, respectively) for 1 h, in buffers containing either 10 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 25 mM KCl and pH 7.4 for G4 (**C**) or 10 mM Tris-HCl, 50 mM KCl, 4 mM EDTA, 5 mM $MgCl_2$, 1 mM DTT, 0.05% NP-40, 2.5% glycerol and pH 6.0 for iM (**D**). The mixture was subjected to 7.5% PAGE. Experiments were performed in triplicate, and representative images are shown.

Table 2. Apparent dissociation constants of NCLs

	K_d (nM)	
	G4	iM
NCL$_{\Delta NTD}$	203 ± 5	166 ± 3
NCL$_{\Delta NTDRGG}$	376 ± 19	318 ± 11

but if NCL stabilizes them, then the DNA extension should be inhibited (Figure 5A).

To ascertain the feasibility of the DNA extension assay for detecting the structural regulation of the non-canonical DNA structures by NCL, a DNA extension assay using KF$_{exo-}$ was performed (Supplementary Figure S11A) using the 5′-Cy5 labeled primer (Entry 7 in Supplementary Table S1) hybridized with the G4-forming template (Entry 6 in Supplementary Table S1). In the absence of potassium ions, the formation ratio of the fully extended product (F_{ext}) was 47%, whereas primer extension was strongly inhibited (F_{ext} ∼10%) upon the addition of K^+ (lanes 2 and 3 in Supplementary Figure S11B), suggesting that G4 formation in the template clearly blocks DNA replication. In the presence of either NCL$_{\Delta NTD}$ or NCL$_{\Delta NTDRGG}$ in addition to K^+, the F_{ext} values decreased by ∼4 and ∼6%, respectively (lanes 3–5 in Supplementary Figure S11B and C). This further inhibition was caused by the stabilization of the G4 structure by NCLs, as reported previously (47).

Next, a DNA extension assay was performed using an iM-forming template. Although the iM-forming DNA sequence from silkworm was used for the screening of iMBPs, another iM-forming sequence from the human Hif1 promoter (59) was employed in this assay. Since the iM structure is formed under acidic conditions, its formation in the template can be controlled by performing a DNA extension assay at pH 6 or 7, allowing us to investigate whether NCLs modulate the stability of the iM structure.

In the absence of NCL, KF$_{exo-}$ elongated the primer to give a fully extended product in 5 min at pH 7 (lanes 1–4 in Figure 5B), whereas a band that migrated faster than the fully extended product was observed at 5 min under acidic conditions (lanes 5–8 in Figure 5B). After 10 min, the stalled product was slightly reduced, and a fully extended product was observed. These observations indicated that the primer extension reaction by KF$_{exo-}$ was partially inhibited by the iM structure formed in the template, as observed for the G4 template (Supplementary Figure S11B). In the presence of excess

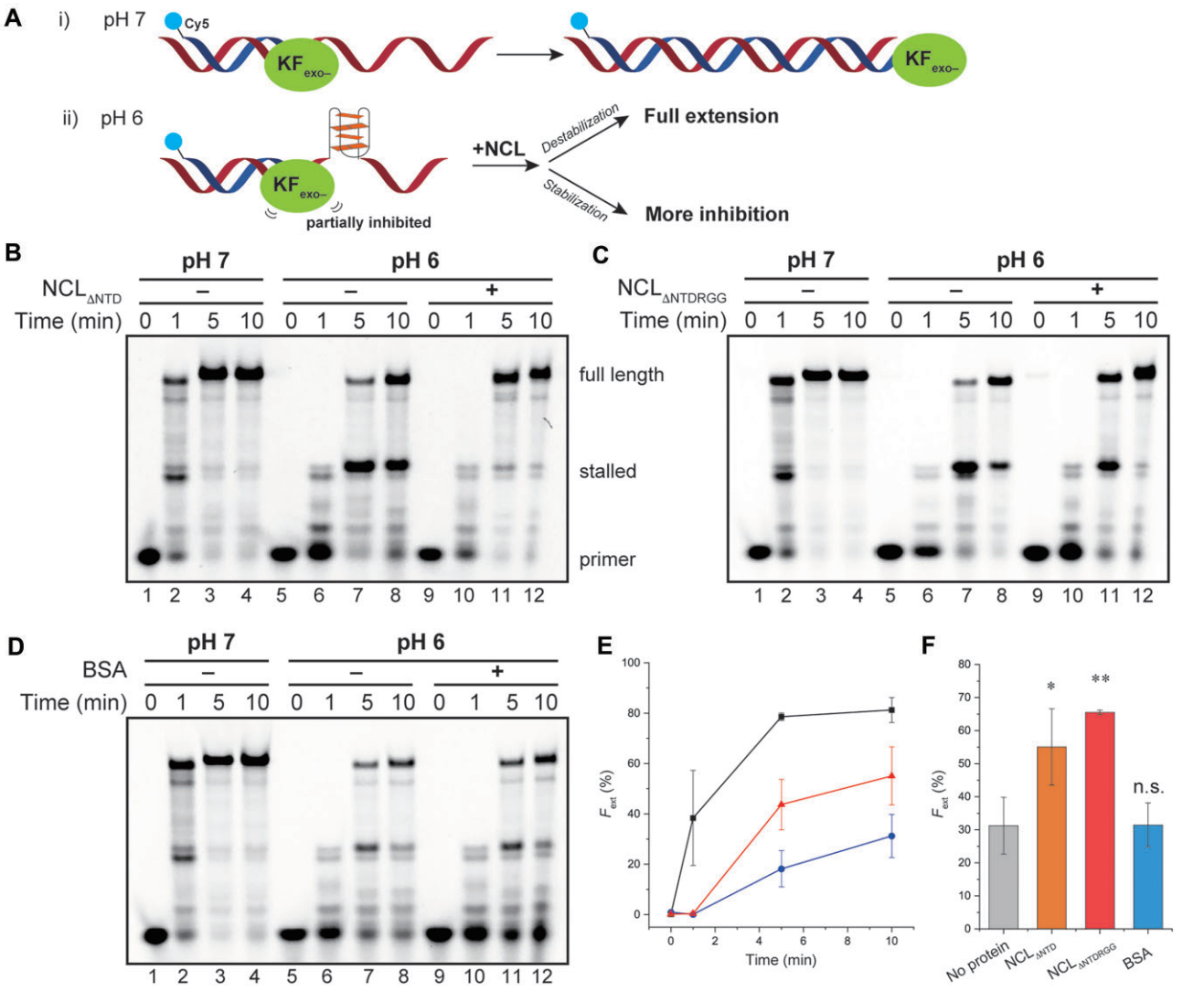


Figure 5. Modulation of iM stability by NCLs. (**A**) Experimental designs of the primer extension assay using KF_{exo-}. At pH 7 (i), KF_{exo-} can produce a full-length product, whereas the extension is partially inhibited at pH 6 (ii) because of iM formation in the template strand. If NCLs stabilize or destabilize the iM structure, the primer extension by KF_{exo-} will be inhibited or facilitated, respectively. (**B–D**) Denaturing PAGE analyses of the primer extension reactions at pH 7 (lanes 1–4) and 6 (lanes 5–12) in the absence (lanes 1–8) or presence (lanes 9–12) of (**B**) $NCL_{\Delta NTD}$, (**C**) $NCL_{\Delta NTDRGG}$, or (**D**) BSA as a negative control. Experiments were performed in triplicate ($n = 3$), in buffer containing 40 mM MES (pH 6.0 or 7.0), 8 mM $MgCl_2$, 50 mM NaCl and 100 μM dNTP mixture, and representative gel images are shown. (**E**) Time course of the formation ratio of the fully extended product (F_{ext}) at pH 7 (squares) or pH 6 in the absence (circles) or presence (triangles) of $NCL_{\Delta NTD}$. Each point and error bar represents mean value ± standard deviation, respectively. The time courses for $NCL_{\Delta NTDRGG}$ and BSA are shown in Supplementary Figure S12. **(F)** Mean F_{ext} values at pH 6 after 10 min of the KF_{exo-} reaction in the absence or presence of the proteins. Error bars represent the standard deviation. Statistical significance was analyzed by a two-sided Student's *t*-test compared with the results in the absence of proteins, where the significance cutoff value was set to 0.05. Asterisks indicate $P < 0.05$ (*) or $P < 0.01$ (**).

$NCL_{\Delta NTD}$ at pH 6 (lanes 9–12 in Figure 5B), the stalled product was clearly reduced and the fully extended product was already prominent at 5 min. Similar results were obtained for $NCL_{\Delta NTDRGG}$ (Figure 5C), whereas no facilitation of the DNA extension reaction was observed in the presence of bovine serum albumin (BSA) (Figure 5D), indicating that this effect was exclusively caused by NCL.

The temporal development of the F_{ext} values for Figure 5B is summarized in Figure 5E and provides clues for elucidating the functional roles of NCL upon iM binding. In the absence of NCL at pH 6, primer extension was clearly retarded due to iM formation, yielding a reduced F_{ext} compared to that at pH 7. In the presence of NCL at pH 6, the F_{ext} values after the >5 min reaction were larger than those in the absence of NCL, indicating that NCL resolves the iM structure. However, no fully extended product was formed after 1 min of the reaction at pH 6, regardless of the presence or absence of NCL, whereas ∼40% of the fully extended product was observed at pH 7. This difference suggests that NCL does not fully resolve the iM structure upon binding; rather, it relaxes the high-order structure and/or switches iM to another architecture that the DNA polymerase can easily resolve during elongation. Comparison of the F_{ext} values after 10 min of reaction at pH 6 (Figure 5F) showed slightly better primer extension by KF_{exo-} in the presence of $NCL_{\Delta NTDRGG}$ than in the presence of $NCL_{\Delta NTD}$, indicating that the RBDs in the NCL are exclusively involved in the regulatory function of the iM structure.

Discussion

In this study, we biochemically identified NCL as an iMBP among the proteins enriched by the FiTE strategy using the iM oligonucleotide; however, it is expected that other iMBPs are also included in the candidate protein list. The known iMBP hnRNP LL reportedly unfolds the *BCL2* iM structure and stretches it to a single-stranded form, using its RBDs (21), in a manner similar to that of NCL. RBDs are abundantly found in various eukaryotic proteins, with similar protein-folding topologies and characteristic motifs in the amino acid sequences (60), as shown in the multiple sequence alignment of the RBDs in human NCL and hnRNPs LL and A1 (Supplementary Figure S13). The fact that the RBDs in NCL and hnRNP LL are commonly involved in iM binding and unfolding provides a potential guideline for other iMBPs; i.e. proteins containing several RBDs can bind iM. A domain search of the proteins found in area vii in Figure 3C, using the UniProt database, indicated that 15 out of 78 proteins (Supplementary Table S2), including NCL and hnRNP A1, contained an RBD. These proteins are potential iMBPs.

A domain search also indicated that proteins containing helicase domains were enriched. Area vii in Figure 3C contains 6 proteins with a helicase domain, and 3 of them are G4BPs (Table 1 and Supplementary Table S2), suggesting that these helicases are also potential iMBPs. Helicases were also enriched in area vi in Figure 3C, the intersection between the two GO terms Nucleus and RNA binding. Except for a few helicases, most were DEAD-box RNA helicases and their variants (DEAH- and DExD-box). Their subcellular locations are heterogeneous, and approximately half are located in nuclear bodies, such as nucleoli and nuclear speckles, which are non-membrane organelles presumably formed via liquid-liquid phase separation (LLPS) (61). Because our proximity-labeling of proteins was performed using nuclear extracts *in vitro*, these nuclear bodies were no longer intact in the extract. However, the biased enrichment of these helicases implies that either the iM itself or some iMBPs may be related to the formation and/or function of these subcellular structures. The enrichment of a nuclear body marker protein, splicing factor, proline- and glutamine-rich (SFPQ) (Supplementary Table S2), corroborated the idea that the iM probe may have been incorporated into the condensed phases.

Although we selected NCL as a potential iMBP based on the known iMBPs as a control to justify the FiTE strategy, another criterion can be adopted to find new iMBPs. Given the characteristics of the FiTE strategy, if POIs detected in the mass spectrometry bind the iM structure non-specifically, then their binding would be competed in the presence of poly(dI·dC), thus reducing the labeling efficiency of the POI. Conversely, if POIs bind the iM structure in a structure-specific manner, then the labeling efficiency of POIs in the presence of poly(dI·dC) would theoretically be higher than in the absence of the competitor, because the POIs have more chances to bind the iM oligonucleotide liberated from non-specifically bound proteins by the competitor. Therefore, a comparison of the volcano plots in the presence and absence of poly(dI·dC) will provide insights into the specificity of binding to the iM structure. Proteins present in the region where the P value is less than 0.05 and the fold-change between the WT and Mut samples is greater than 1.5 ($-\log_{10}(P\text{-value}) > 1.3$ and $\log_2$(fold change: WT/Mut) > 0.584, Supplementary Figure S14A) were selected from Figure 3B, and the $\log_2$(fold change: WT/Mut) value without poly(dI·dC) of each protein was extracted to obtain the ratio of the x values in the presence and absence of poly(dI·dC). The ratio can be considered as a measure of the specificity of iM binding ability. Based on the above criteria, 53 proteins were selected, and their specificity ratios are shown in Supplementary Figure S14B. Among them, eight proteins exhibited specificity ratios greater than that of hnRNP LL, although the cellular locations of some proteins are outside of the nucleus. These are promising candidates of iMBPs or proteins closely related to iMBPs. Especially, BEN domain-containing protein 3 (BEND3) is of interest for future investigations, because it is a transcriptional repressor that associates with a nucleolar remodeling complex (62). Coincidentally, we have characterized the representative nucleolar protein NCL, implying a potential relationship of iM to nucleolar biology.

NCL is mainly localized in the nucleoli, accounting for approximately 10% of the protein content within each nucleolus (63), and is therefore widely utilized as a marker protein. Nucleoli are involved in various essential cellular processes and are the origin of ribosome biogenesis. Ribosomal DNA (rDNA), consisting of tandem repeats of the ribosomal part transcribed by RNA polymerase I (Pol I) and the intergenic spacer (IGS) in nucleolar organizing regions, is recognized as a hotspot of non-canonical DNA structures and the most G4-rich locus in the eukaryotic genome (64), indicating the presence of C-rich sequences in the complementary strand. Indeed, a machine learning-based prediction tool for iM, iM-Seeker (65), suggested that rDNA contains many putative iM sequences in both the sense and antisense strands (Supplementary Figure S15), with a sparser distribution in the IGS than in the ribosomal part. Interestingly, the distribution of the putative iM sequences in rDNA overlapped well with the localization of NCL on rDNA, as investigated by the ChIP-seq analysis (66), implying that cellular NCL may interact with either the G4 or iM structures in rDNA.

Based on our finding that NCL is an iMBP, several plausible functional roles for NCL on rDNA may be considered. The first involves facilitation of ribosomal transcription. Pol I transcription can potentially be inhibited by the presence of putative iM sequences in the ribosome coding regions. NCL binds to putative iM sequences, relaxes the iM structure and facilitates the transcription by Pol I. The knockdown of NCL reportedly repressed Pol I transcription but not Pol II or Pol III transcription (67), which may be attributed to the relaxation of the iM structures in rDNA by NCL.

Another potential role of the iM-bound NCL is as a scaffold for nucleoli formation. Human nucleoli have dynamic architectures that disintegrate in the mitotic phase and reassemble in early telophase (68). Nucleolar dynamics are reportedly coupled with Pol I transcriptional activity, which is downregulated in early M phase and upregulated in telophase. To reinitiate the rDNA transcription in daughter cells, the Pol I machinery and its regulatory factors must be assembled in the ribosome-coding region of rDNA. We propose that NCL binding to iM and/or G4 in rDNA may help recruit other nuclear proteins, including the Pol I machinery, to the nucleolar compartment during its re-assembly. Nucleophosmin 1 (NPM1) is a nucleolar protein that induces LLPS by interacting with arginine-rich peptides (69). Since NCL possesses an RGG domain and reportedly interacts with NPM1 (70,71), the NCL-NPM1 interaction would initiate the nucleolar formation via LLPS and concomitantly locate the rDNA loci in the nucleoli.

It should be mentioned that NPM1 was also enriched with the iM-WT oligonucleotide in this study (Table 1), supporting a possible relationship to iM binding. Nucleoli are composed of subdomains, with each playing a unique role and containing distinct proteins. Although more detailed investigations are required, the potential link between the high-order DNA structures in rDNA and the nucleoli will contribute to a deeper understanding of the cell biology of nucleoli.

Conclusion

Because the alkyne-tosylate-NHS ester **1** synthesized in this study can be attached to any of the amino-modified DNA/RNA oligonucleotides, the proximity labeling of proteins that bind nucleic acids of interest will also be possible. Therefore, the FiTE strategy associated with proteomic analyses is a primary option for the *in vitro* enrichment of proteins that bind a variety of nucleic acids.

The FiTE approach and subsequent proteomic analyses successfully identified new potential iMBPs, and biochemical analyses comprehensively revealed that the RBDs in NCL are involved in binding to and relaxing the iM structure. Recent studies have strongly suggested that iM is formed in living cells and is implicated in various biological processes, in which the primary step should be recognition of the iM structure by iMBPs. Our study provides clues toward elucidating how the downstream proteins are involved in the biological processes and will open the doors to a new stage in iM biology.

Data availability

Raw data from the nanoLC-MS/MS analysis and the search results were deposited in jPOSTrepo (JPST003020) and are available in the ProteomeXchange consortium with the accession number of PXD051188. Other data underlying this article are available within the article and in its online Supplementary Material.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Prof. Keiichiro Suzuki (Institute for Advanced Co-creation Studies, Osaka University) for his kind support in cell culture and Dr. Eiji Hishinuma, Naomi Matsumura, and Chizu Fujiwara (Tohoku University) for their kind help in data collection for nanoLC-MS/MS analysis. Part of this study was supported by the Research Equipment Sharing System of Tohoku University (373·Orbitrap Fusion).

Author contributions: J.Y. conceived and directed the work; Y.B., Y.T., R.M., S.I. and J.Y. developed strategies for FiTE; Y.B., K.N. and S.S. identified proteins using nanoLC-MS/MS; Y.B., Y.A. and Y.T. performed biochemical experiments; Y.B., S.I., S.S. and J.Y. analyzed the data; Y.B. and J.Y. wrote the manuscript; all authors reviewed and approved the final version of the manuscript.

Funding

The Uehara Memorial Foundation [to J.Y.]; Takeda Science Foundation [to J.Y.]; Japan Society for the Promotion of Science (a Grant-in-Aid for Transformative Research Areas (A) Biophysical Chemistry for Material Symbiosis) [21H05503 to S.S.]; JST FOREST Program [JPMJFR2005 to S.S. and JPMJFR2057 to J.Y.]. Sunbor Scholarship and a JSPS Fellowship [to Y.T. and K.N.], respectively. Funding for open access charge: JST FOREST.

Conflict of interest statement

None declared.

References

1. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA*, **48**, 2013–2018.
2. Gehring,K., Leroy,J.L. and Guéron,M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, **363**, 561–565.
3. Miyoshi,D., Nakao,A. and Sugimoto,N. (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry*, **41**, 15017–15024.
4. Cui,J., Waltman,P., Le,V.H. and Lewis,E.A. (2013) The effect of molecular crowding on the stability of Human c-MYC promoter sequence I-motif at neutral pH. *Molecules*, **18**, 12751–12767.
5. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
6. Assi,H.A., Garavís,M., González,C. and Damha,M.J. (2018) i-motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.*, **46**, 8038–8056.
7. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M., *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
8. Zanin,I., Ruggiero,E., Nicoletto,G., Lago,S., Maurizio,I., Gallina,I. and Richter,S.N. (2023) Genome-wide mapping of i-motifs reveals their association with transcription regulation in live human cells. *Nucleic Acids Res.*, **51**, 8309–8321.
9. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
10. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.
11. King,J.J., Irving,K.L., Evans,C.W., Chikhale,R.V., Becker,R., Morris,C.J., Peña Martinez,C.D., Schofield,P., Christ,D., Hurley,L.H., *et al.* (2020) DNA G-quadruplex and i-motif structure formation is interdependent in Human cells. *J. Am. Chem. Soc.*, **142**, 20600–20604.
12. Giraldo,R. and Rhodes,D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.*, **13**, 2411–2420.
13. Traczyk,A., Liew,C.W., Gill,D.J. and Rhodes,D. (2020) Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic Acids Res.*, **48**, 4562–4571.
14. Harrington,C., Lan,Y. and Akman,S.A. (1997) The identification and characterization of a G4-DNA resolvase activity. *J. Biol. Chem.*, **272**, 24631–24636.
15. Chen,M.C., Tippana,R., Demeshkina,N.A., Murat,P., Balasubramanian,S., Myong,S. and Ferré-D'amaré,A.R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, **558**, 465–469.
16. Shu,H., Zhang,R., Xiao,K., Kharel,P., Varizhuk,A., Pirota,V., Shu,H., Zhang,R., Xiao,K., Yang,J., *et al.* (2022) G-quadruplex-binding proteins: promising targets for drug design. *Biomolecules*, **12**, 648.
17. Meier-Stephenson,V. (2022) G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys. Rev.*, **14**, 635–654.

18. Mishra,S.K., Tawani,A., Mishra,A. and Kumar,A. (2016) G4IPDB: a database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.*, **6**, 38144.
19. Zhang,X., Spiegel,J., Martínez Cuesta,S., Adhikari,S. and Balasubramanian,S. (2021) Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.*, **13**, 626–633.
20. Lu,Z., Xie,S., Su,H., Han,S., Huang,H. and Zhou,X. (2024) Identification of G-quadruplex-interacting proteins in living cells using an artificial G4-targeting biotin ligase. *Nucleic Acids Res.*, **52**, e37.
21. Kang,H.J., Kendrick,S., Hecht,S.M. and Hurley,L.H. (2014) The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.*, **136**, 4172–4185.
22. Miglietta,G., Cogoi,S., Pedersen,E.B. and Xodo,L.E. (2015) GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1. *Sci. Rep.*, **5**, 18097.
23. Sutherland,C., Cui,Y., Mao,H. and Hurley,L.H. (2016) A mechanosensor mechanism controls the G-quadruplex/i-motif molecular switch in the MYC promoter NHE III1. *J. Am. Chem. Soc.*, **138**, 14138–14151.
24. Gao,B., Zheng,Y.-T., Su,A.-M., Sun,B., Xi,X.-G. and Hou,X.-M. (2022) Remodeling the conformational dynamics of I-motif DNA by helicases in ATP-independent mode at acidic environment. *iScience*, **25**, 103575.
25. Niu,K., Zhang,X., Deng,H., Wu,F., Ren,Y., Xiang,H., Zheng,S., Liu,L., Huang,L., Zeng,B., *et al.* (2018) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res.*, **46**, 1710–1723.
26. Cui,Y., Koirala,D., Kang,H.J., Dhakal,S., Yangyuoru,P., Hurley,L.H. and Mao,H. (2014) Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL. *Nucleic Acids Res.*, **42**, 5755–5764.
27. Wu,W.Q., Zhang,X., Bai,D., Shan,S.W. and Guo,L.J. (2022) Mechanistic insights into poly(C)-binding protein hnRNP K resolving i-motif DNA secondary structures. *J. Biol. Chem.*, **298**, 102670.
28. Ruggiero,E., Lago,S., Šket,P., Nadai,M., Frasson,I., Plavec,J. and Richter,S.N. (2019) A dynamic i-motif with a duplex stem-loop in the long terminal repeat promoter of the HIV-1 proviral genome modulates viral transcription. *Nucleic Acids Res.*, **47**, 11057–11068.
29. Roux,K.J., Kim,D.I., Raida,M. and Burke,B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, **196**, 801–810.
30. Branon,T.C., Bosch,J.A., Sanchez,A.D., Udeshi,N.D., Svinkina,T., Carr,S.A., Feldman,J.L., Perrimon,N. and Ting,A.Y. (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.*, **36**, 880–887.
31. Guo,J., Guo,S., Lu,S., Gong,J., Wang,L., Ding,L., Chen,Q. and Liu,W. (2023) The development of proximity labeling technology and its applications in mammals, plants, and microorganisms. *Cell Commun. Signal.*, **21**, 269.
32. Masuzawa,T., Sato,S., Niwa,T., Taguchi,H., Nakamura,H. and Oyoshi,T. (2020) G-quadruplex-proximity protein labeling based on peroxidase activity. *Chem. Commun.*, **56**, 11641–11644.
33. Okuda,S., Watanabe,Y., Moriya,Y., Kawano,S., Yamamoto,T., Matsumoto,M., Takami,T., Kobayashi,D., Araki,N., Yoshizawa,A.C., *et al.* (2017) jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res.*, **45**, D1107–D1111.
34. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.
35. Tsukiji,S., Miyagawa,M., Takaoka,Y., Tamura,T. and Hamachi,I. (2009) Ligand-directed tosyl chemistry for protein labeling in vivo. *Nat. Chem. Biol.*, **5**, 341–343.
36. Takaoka,Y., Ojida,A. and Hamachi,I. (2013) Protein organic chemistry and applications for labeling and engineering in live-cell systems. *Angew. Chem. Int. Ed.*, **52**, 4088–4106.
37. Liang,L. and Astruc,D. (2011) The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) 'click' reaction and its applications. An overview. *Coord. Chem. Rev.*, **255**, 2933–2945.
38. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
39. Larouche,K., Bergeron,M.J., Leclerc,S. and Guérin,S.L. (2018) Optimization of competitor poly(dI-dC)•poly(dI-dC) levels is advised in DNA-protein interaction studies involving enriched nuclear proteins. *BioTechniques*, **20**, 439–444.
40. Li,Y., Zhang,X., Gao,Y., Shi,J., Tang,L. and Sui,G. (2018) G-quadruplexes in the BAP1 promoter positively regulate its expression. *Exp. Cell. Res.*, **369**, 147–157.
41. Federici,L., Arcovito,A., Scaglione,G.L., Scaloni,F., Lo Sterzo,C., Di Matteo,A., Falini,B., Giardina,B. and Brunori,M. (2010) Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *J. Biol. Chem.*, **285**, 37138–37149.
42. Chiarella,S., De Cola,A., Scaglione,G.L., Carletti,E., Graziano,V., Barcaroli,D., Lo Sterzo,C., Di Matteo,A., Di Ilio,C., Falini,B., *et al.* (2013) Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Res.*, **41**, 3228–3239.
43. Amato,J., Madanayake,T.W., Iaccarino,N., Novellino,E., Randazzo,A., Hurley,L.H. and Pagano,B. (2018) HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation?. *Chem. Commun.*, **54**, 9442–9445.
44. Amato,J., Cerofolini,L., Brancaccio,D., Giuntini,S., Iaccarino,N., Zizza,P., Iachettini,S., Biroccio,A., Novellino,E., Rosato,A., *et al.* (2019) Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein. *Nucleic Acids Res.*, **47**, 9950–9966.
45. Cogoi,S., Paramasivam,M., Spolaore,B. and Xodo,L.E. (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res.*, **36**, 3765–3780.
46. Keller,J.G., Hymøller,K.M., Thorsager,M.E., Hansen,N.Y., Erlandsen,J.U., Tesauro,C., Simonsen,A.K.W., Andersen,A.B., Petersen,K.V., Holm,L.L., *et al.* (2022) Topoisomerase 1 inhibits MYC promoter activity by inducing G-quadruplex formation. *Nucleic Acids Res.*, **50**, 6332–6342.
47. González,V., Guo,K., Hurley,L. and Sun,D. (2009) Identification and characterization of Nucleolin as a c-myc G-quadruplex-binding. *Protein. J. Biol. Chem.*, **284**, 23622–23635.
48. Khan,Y., Azam,T., Sundar,J.S., Maiti,S. and Ekka,M.K. (2023) Biophysical characterization of nucleolin domains crucial for interaction with telomeric and TERRA G-quadruplexes. *Biochemistry*, **62**, 1249–1261.
49. Pramanik,S., Chen,Y., Song,H., Khutsishvili,I., Marky,L.A., Ray,S., Natarajan,A., Singh,P.K. and Bhakat,K.K. (2022) The human AP-endonuclease 1 (APE1) is a DNA G-quadruplex structure binding protein and regulates KRAS expression in pancreatic ductal adenocarcinoma cells. *Nucleic Acids Res.*, **50**, 3394–3412.
50. Kang,H.J., Le,T.V.T., Kim,K., Hur,J., Kim,K.K. and Park,H.J. (2014) Novel interaction of the Z-DNA binding domain of Human ADAR1 with the oncogenic c-myc promoter G-quadruplex. *J. Mol. Biol.*, **426**, 2594–2604.
51. Lyonnais,S., Tarrés-Soler,A., Rubio-Cosials,A., Cuppari,A., Brito,R., Jaumot,J., Gargallo,R., Vilaseca,M., Silva,C., Granzhan,A., *et al.* (2017) The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci. Rep.*, **7**, 43992.
52. Izumi,H. and Funa,K. (2019) Telomere function and the G-quadruplex formation are regulated by hnRNP U. *Cells*, **8**, 390.

53. Dempsey,L.A., Sun,H., Hanakahi,L.A. and Maizels,N. (1999) G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.*, **274**, 1066–1071.
54. Hároníková,L., Coufal,J., Kejnovská,I., Jagelská,E.B., Fojta,M., Dvoøáková,P., Muller,P., Vojtesek,B. and Brázda,V. (2016) IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation. *PLoS One*, **11**, e0157156.
55. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.
56. Frottin,F., Schueder,F., Tiwary,S., Gupta,R., Körner,R., Schlichthaerle,T., Cox,J., Jungmann,R., Hartl,F.U. and Hipp,M.S. (2019) The nucleolus functions as a phase-separated protein quality control compartment. *Science*, **365**, 342–347.
57. Pavlova,I., Iudin,M., Surdina,A., Severov,V. and Varizhuk,A. (2023) G-quadruplexes in nuclear biomolecular condensates. *Genes*, **14**, 1076.
58. Carey,J. (1991) Gel retardation. *Methods Enzymol.*, **208**, 103–117.
59. Takahashi,S., Brazier,J.A. and Sugimoto,N. (2017) Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc. Natl. Acad. Sci. USA*, **114**, 9605–9610.
60. Maris,C., Dominguez,C. and Allain,F.H.T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
61. Wang,B., Zhang,L., Dai,T., Qin,Z., Lu,H., Zhang,L. and Zhou,F. (2021) Liquid–liquid phase separation in human health and diseases. *Signal Transduct. Target. Ther.*, **6**, 290.
62. Khan,A., Giri,S., Wang,Y., Chakraborty,A., Ghosh,A.K., Anantharaman,A., Aggarwal,V., Sathyan,K.M., Ha,T., Prasanth,K.V., *et al.*2015) BEND3 represses rDNA transcription by stabilizing a NoRC component via USP21 deubiquitinase. *Proc. Natl. Acad. Sci. USA*, **112**, 8338–8343.
63. Tajrishi,M.M., Tuteja,R. and Tuteja,N. (2011) Nucleolin: the most abundant multifunctional phosphoprotein of nucleolus. *Commun. Integr. Biol.*, **4**, 267–275.
64. Smirnov,E., Molínová,P., Chmúrčiaková,N., Vacík,T. and Cmarko,D. (2023) Non-canonical DNA structures in the human ribosomal DNA. *Histochem. Cell Biol.*, **160**, 499–515.
65. Yang,B., Guneri,D., Yu,H., Wr Ight,E.P., Enqian Chen,W., Waller,Z.A.E. and Ding,Y. (2024) Prediction of DNA i-motifs via machine learning. *Nucleic Acids Res.*, **52**, 2188–2197.
66. Cong,R., Das,S., Ugrinova,I., Kumar,S., Mongelard,F., Wong,J. and Bouvet,P. (2012) Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. *Nucleic Acids Res.*, **40**, 9441–9454.
67. Rickards,B., Flint,S.J., Cole,M.D. and LeRoy,G. (2007) Nucleolin is required for RNA polymerase I transcription in vivo. *Mol. Cell. Biol.*, **27**, 937–948.
68. Németh,A. and Grummt,I. (2018) Dynamic regulation of nucleolar architecture. *Curr. Opin. Cell Biol.*, **52**, 105–111.
69. Mitrea,D.M., Cika,J.A., Guy,C.S., Ban,D., Banerjee,P.R., Stanley,C.B., Nourse,A., Deniz,A.A. and Kriwacki,R.W. (2016) Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife*, **5**, e13571.
70. Li,Y.P., Busch,R.K., Valdez,B.C. and Busch,H. (1996) C23 Interacts with B23, A putative nucleolar-localization-signal-binding protein. *Eur. J. Biochem.*, **237**, 153–158.
71. Liu,H.T. and Yung,B.Y.M. (1999) In vivo interaction of nucleophosmin/B23 and protein C23 during cell cycle progression in HeLa cells. *Cancer Lett.*, **144**, 45–54.

Received: June 4, 2024. **Revised:** September 18, 2024. **Editorial Decision:** October 12, 2024. **Accepted:** October 17, 2024