Data Article

Data on G-quadruplex topology, and binding ability of G-quadruplex forming sequences found in the promoter region of biomarker proteins and those relations to the presence of nuclear localization signal in the proteins

Jinhee Lee\textsuperscript{a}, Kentaro Teramoto\textsuperscript{b}, Tomomi Yokoyama\textsuperscript{b}, Kinuko Ueno\textsuperscript{b}, Kaori Tsukakoshi\textsuperscript{b}, Koji Sode\textsuperscript{a}, Kazunori Ikebukuro\textsuperscript{b,∗}

\textsuperscript{a} Joint Department of Biomedical Engineering, The University of North Carolina at Chapel Hill and North Carolina State University, Chapel Hill, NC 27599, USA

\textsuperscript{b} Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei, Tokyo 184-8588, Japan

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 4 January 2021
Revised 24 March 2021
Accepted 29 March 2021
Available online 1 April 2021

\textbf{Keywords:}
G-quadruplex
Aptamer
Nuclear localization signal
Promoter region
Biomarker protein

\textbf{A B S T R A C T}

Aptamer is a nucleic acid ligand which specifically binds to its target molecule. Previously, we have designed an identification method of aptamer called “G-quadruplex (G4) promoter-derived aptamer selection (G4PAS)” [1]. In G4PAS procedure, putative G4 forming sequences (PQS) were explored in a promoter region of a target protein in human gene through computational analysis, and evaluated binding ability towards the gene product encoded in the downstream of the promoter. We investigated the topology of the obtained PQSs by circular dichroism measurement, as well as their binding ability against its target protein by surface plasmon resonance measurement and gel-shift assay. Additionally, the presence of nuclear localization signal in the target protein was predicted in silico. This data set summarized all the PQS sequences, their biochemical characteristics, and the presence of nuclear localization signal to address the possibility of binding of these PQS region to the target proteins.

∗ Corresponding author.
\textit{E-mail address:} ikebu@cc.tuat.ac.jp (K. Ikebukuro).

https://doi.org/10.1016/j.dib.2021.107028
2352-3409/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
in vivo. Those data should contribute to increase the success rate of G4PAS. Moreover, considering the G4 motifs in genomic DNA are suggested to be involved in vivo gene regulation [2,3], this data set is also potentially beneficial for the cell biology field.

© 2021 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

**Specifications Table**

| Subject | Biotechnology |
|---------|---------------|
| Specific subject area | Biochemistry, nucleic acid ligand (aptamer) |
| Type of data | Table |
| How data were acquired | Gel-shift assay, Circular dichroism spectroscopy (J-820 spectropolarimeter, JASCO), Surface plasmon resonance measurement (Biacore T200, GE Healthcare), In silico Prediction (NLSdb; https://rostlab.org/services/nlsdb/) and cNLS Mapper; http://nls-mapper.ibk.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi |
| Data format | Raw and analyzed data |
| Parameters for data collection | Known biomarker proteins were chosen as the target, and G-quadruplex-forming DNA sequences were picked up from a genomic region around the transcription start site of the proteins the criterion of G2>, N1->G2>, N1->G2>, N1->G2<, where “G” is guanine base and “N” can be any bases. The binding between the DNA sequences towards the target protein, and the topology of the G-quadruplex-structure were performed with or without 100 mM KCl in Tris-based buffer (pH 7.4) at 25 °C. The search of G-quadruplex-forming sequence in genomic DNA, and the nuclear localization signal prediction in the target proteins were performed by web tools (NLSdb and cNLS Mapper). The binding between the G-quadruplex-forming DNA and the target proteins was investigated by gel-shift assay, surface plasmon resonance measurement. The topology of G-quadruplex-forming sequence was analyzed by Circular dichroism spectroscopy. |
| Description of data collection | Raw data |
| Data source location | Institution: Tokyo University of Agriculture and Technology City/Town/Region: Koganei city, Tokyo Country: Japan Secondary data Primary data sources: Circular dichroism spectrum data http://doi.org/10.17632/5xthrvbspc.3#folder-5980050f-9d75-4675-9ce6-3a25df689c2b Surface plasmon resonance measurement data http://doi.org/10.17632/5xthrvbspc.3#folder-cc350bd1-f6d5-4b10-9f9b-22a647b38ae2 |
| Data accessibility | With the article Repository name: Mendeley Data Direct URL to data: https://data.mendeley.com/datasets/5xthrvbspc/3 Related research article | W. Yoshida, T. Saito, T. Yokoyama, S. Ferri, K. Ikebukuro, Aptamer selection based on G4-forming promoter region. PLoS ONE, 8(6) (2013) e65497. http://doi.org/10.1371/journal.pone.0065497 |

**Value of the Data**

- This data set summarizes the biochemical characteristics (topology of G-quadruplex and presence of nuclear localization signal) as well as the binding of aptamer obtained by
G4PAS method and helps to improve the performance of aptamer selection based on G4PAS method.
• This data can help all who wish to obtain aptamer by G4PAS method.
• This data can be used for further studies aiming to investigate G-quadruplex motif-mediated in vivo gene regulation.

1. Data Description

1.1. G-quadruplex-forming sequences found in the promoter regions

Genomic sequences around the transcription start site of each target proteins have been obtained using the UCSC genome browser (https://genome.ucsc.edu/), and G-quadruplex-forming sequences were identified by the QGRS mapper (http://bioinformatics.ramapo.edu/QGRS/index.php) [4]. All the DNA sequences are listed in Table 1 and deposited in Mendeley data repository [7].

1.2. Nuclear localization signal identification in the target proteins

NLSdb [5] (https://rostlab.org/services/nlsdb/), and cNLS Mapper [6] (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) were used for the prediction of nuclear localization signal. The amino acid sequence of each target protein including its isomers were subjected to the prediction and all the results are shown in Table 1 and deposited in Mendeley data repository [7].

1.3. Binding assay of the extracted G4 forming oligonucleotide towards the target protein

The binding of the identified G-quadruplex-forming sequences towards its target protein was investigated by surface plasmon resonance (SPR) measurement and gel-shift assay. For the SPR assay, each target protein was immobilized on the chip by amine coupling and synthesized PQSs were injected to observe SPR signal. The SPR sensorgrams are indicated as Figs. 1 to 9, and all the raw SPR response data were deposited in Mendeley data repository [7] as well as present in supplementary material. The $K_D$ value was determined based on the sensorgram and shown in Table 1. For the gel-shift assay, each PQS was folded by heat treatment (95 °C for 5 min and gradually cooled down to 25 °C over 30 min) and, 500 nM (final concentration: f.c.) of PQS was mixed with 1 μM (f.c.) of each target protein. After 30 min of incubation, the samples were used for electrophoresed in a 12% polyacrylamide gel. The bands were visualized by FITC fluorescence. The results of gel-shift assay were indicated as Figs. 10 to 18 and summarized in Table 1.

1.4. Circular dichroism measurement for the assessment of G4 topology of each pqs

The G4 topology of each PQS was investigated by CD spectrum. G4 forming oligonucleotide is known to show specific peak pattern, i.e., parallel G4 shows a positive peak at around 260 nm and a negative peak at around 240 nm, and anti-parallel G4 shows a positive peak at around 290 nm and a negative peak at around 260 nm. The spectra were measured either with or without 100 mM of potassium ion, which stabilize certain G4 structure. The CD spectra of each PQS are shown as Figs. 19 to 29. All the raw CD spectrum data were deposited in Mendeley data repository [7] as well as present in supplementary material.
Table 1
Summary of G-quadruplex-forming sequences and its biochemical characterizations. The binding assay results of RB1, c-KIT, VEGFA, PDGFA were referred from the reference [1]. The results of HGF and HBEGF PQS are partially published in the reference [8].

| Target | NLS by NLSdb | NLS by cNLS Mapper | Name   | Sequence (5′ → 3′)                  | Result of gel-shift assay | K_D (M) by SPR | G4 topology  |
|--------|--------------|-------------------|--------|-----------------------------------|--------------------------|----------------|--------------|
| RB1    | Yes          | Yes               | RB1-PQS | CGGGGGGTTT TTGGGGGC              | Bound [1]                 | 4.4 × 10^{-7} [1] | parallel     |
| c-KIT  | No           | No                | c-KIT-PQS1 | CGGGGGGCGGCGC GAGGGAGGGG         | Not bound [1]             | –              | parallel     |
| VEGFA  | No           | Yes               | VEGFA-PQS | CGGGGGGGGGGGGGG GGGGGGGGGGGGGG  | Bound [1]                 | 1.7 × 10^{-7} [1] | parallel     |
| PDGFA  | Yes          | Yes               | PDGFAA-PQS | GAGGGGGGGGGGGGGG GGGGGGGGGGGGGGG | Bound [1]                 | 6.3 × 10^{-9} [1] | parallel     |
| HGF    | No           | No                | HGF-PQS1 | GGGTTGGAGTGTTGA GGGGAGTTGAGG    | –                        | 7.3 × 10^{-8} [8] | parallel     |
| HGF-PQS2 |             |                   | GGAGAGGTGGAA GGGTTGGAGG        | Not bound                | Not apparent     |
| HGF-PQS3 |             |                   | GGGGGATGGCA GTGGGGAGCAG        | Not bound                | Hybrid or mixture |
| HGF-PQS4 |             |                   | GGGGTGGTTGG             | Not bound                | Not apparent     |
| HGF-PQS5 |             |                   | GGGGGTGGGGG             | Not bound                | Not apparent     |
| HGF-PQS6 |             |                   | GGGGGTGGGGG             | Not bound                | Parallel         |
| HGF-PQS7 |             |                   | GGGGGTGGGGG             | Not bound                | Hybrid or mixture |
| HGF-PQS8 |             |                   | GGGGGTGGGGG             | Not bound                | Not apparent     |
| HB-EGF | Yes          | No                | HBEGF-PQS1 | GGGGAAAGGA GGGGCTGGGG          | Not bound                | Hybrid or mixture |
| HBEGF-PQS2 |            |                   | GGGGCTGGGG GTGGGGAGG       | Not bound                | Not apparent     |
| HBEGF-PQS3 |            |                   | GGGGGTCAG         | Not bound                | Hybrid or mixture |
| HBEGF-PQS4 |            |                   | GGGGGGCA         | Not bound                | Not apparent     |
| HBEGF-PQS5 |            |                   | GGGGGGCGG GTGGGGAGG   | Not bound                | Not apparent     |
| HBEGF-PQS6 |            |                   | GGGGGGCGG GTGGGGAGG   | Not bound                | Not apparent     |
| HBEGF-PQS7 |            |                   | GGGGGGCGG GTGGGGAGG   | Not bound                | Not apparent     |
| aFGF   | No           | Yes               | aFGF-PQS1 | GGGGCGGCGGCGGCGG             | Not bound                | Not apparent     |
| aFGF-PQS2 |            |                   | GGAGGGGTCAGGCAGAAG | Not bound                | –                |

(continued on next page)
Table 1 (continued)

| Target | NLS by NLSdb | NLS by cNLS Mapper | Name | Sequence (5' → 3') | Result of gel-shift assay | $K_D$ (M) by SPR | G4 topology |
|--------|--------------|-------------------|------|-----------------|-------------------------|------------------|-------------|
| aFGF-PQS3 |              |                   | GGAGACGGTA | - | Not bound | - |
|          |              |                   | GGGCAAGTGG | - | Not bound | - |
| aFGF-PQS4 |              |                   | GGTGCCTGGGTATGG | - | Not bound | - |
| aFGF-PQS5 |              |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| aFGF-PQS6 |              |                   | GGGAGAGGGAGGAGGG | - | Not bound | - |
| aFGF-PQS7 |              |                   | GGTGGGGGGGG | - | Not bound | - |
| aFGF-PQS8 |              |                   | GGTGGGA | - | Not bound | - |
|          |              |                   | CTGGCCAGGG | - | Not bound | - |
| aFGF-PQS9 |              |                   | GCCAGAGAACAG | - | Not bound | - |
| aFGF-PQS10 |             |                   | GCCCTGGAGT | - | Not bound | - |
| aFGF-PQS11 |             |                   | CTGGCCGAGG | - | Not bound | - |
| aFGF-PQS12 |             |                   | GGGAGAGGGAGGAGGG | - | Not bound | - |
| aFGF-PQS13 |             |                   | GCCATGGTAT | - | Not bound | - |
| aFGF-PQS14 |             |                   | CTGGCCAGGGAGG | - | Not bound | - |
| aFGF-PQS15 |             |                   | GCCCTGGAGA | - | Not bound | - |
| aFGF-PQS16 |             |                   | ACTCCTGGAGG | - | Not bound | - |
| aFGF-PQS17 |             |                   | GAATGACAGGCAGA | - | Not bound | - |
| aFGF-PQS18 |             |                   | GGCAATGGTACGAGG | - | Not bound | - |
| aFGF-PQS19 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| aFGF-PQS20 |             |                   | GCCATGGTAT | - | Not bound | - |
| aFGF-PQS21 |             |                   | CTGGCCAGGGAGG | - | Not bound | - |
| aFGF-PQS22 |             |                   | GGACCTGGAGC | - | Not bound | - |
| aFGF-PQS23 |             |                   | ACTCCTGGAGG | - | Not bound | - |
| bFGF-PQS1 |              |                   | GGGCCAGGGAGG | - | Not bound | - |
| bFGF-PQS2 |              |                   | GGCTGCGGCTGGG | - | Not bound | - |
| bFGF-PQS3 |              |                   | GGGAACTGCTGGG | - | Not bound | - |
| bFGF-PQS4 |              |                   | GCCCTGGAGT | - | Not bound | - |
| bFGF-PQS5 |              |                   | GGGGAAGGGGGGGGGGG | - | Not bound | - |
| bFGF-PQS6 |              |                   | GGGAGAGGGAGGAGGG | - | Not bound | - |
| bFGF-PQS7 |              |                   | GGGCCAGGGG | - | Not bound | - |
| bFGF-PQS8 |              |                   | GGGAGAGGGAGGAGGG | - | Not bound | - |
| bFGF-PQS9 |              |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| bFGF-PQS10 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| bFGF-PQS11 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| bFGF-PQS12 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| bFGF-PQS13 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |
| bFGF-PQS14 |             |                   | GCCACTGGAGGAATGG | - | Not bound | - |

(continued on next page)
| Target | NLS by NLSdb | NLS by cNLS Mapper | Name | Sequence (5′ → 3′) | Result of gel-shift assay | Result of gel-shift assay | Kᵤ (M) by SPR | G4 topology |
|--------|--------------|-------------------|------|-------------------|--------------------------|--------------------------|-----------------|-------------|
|        |              |                   | bFGF-PQS15 | GGGCACGGC | – | Not bound | – |
|        |              |                   | bFGF-PQS16 | GGGAGCCGGC | – | Not bound | – |
|        |              |                   | bFGF-PQS17 | GGGGGGAGGGGAGAGG | – | Not bound | – |
|        |              |                   | bFGF-PQS18 | GCCCGCGCGCA | – | Not bound | – |
|        | No | No | AFP-PQS1 | GGGACTATCTGATCT | Not bound | – | – |
|        | No | No | PSA-PQS1 | GGGTTCAGGGCCAGG | Not bound | – | – |
|        | No | No | PSA-PQS2 | GGGAGACGGGGCTGGG | Not bound | – | – |
|        | No | No | PSA-PQS3 | GGGGACCGGAGG | Not bound | – | – |
|        | No | No | PSA-PQS4 | GGGCAAATGAGGAGGAGG | Not bound | – | – |
|        | No | No | PSA-PQS5 | GGGTTCAGGGCCAGG | Not bound | – | – |
|        | No | No | PSA-PQS6 | GGGGACCGGAGG | Not bound | – | – |
|        | No | No | CRP-PQS1 | GGGATCTGGAG | Not bound | – | – |
|        | No | No | CRP-PQS2 | GGGGACTGCTGGG | Bound | Not bound | – |
|        | No | No | HER2-PQS1 | GGGCCCTGGG | Not bound | – | – |
|        | No | No | HER2-PQS2 | GGGTCTGCTGGG | Not bound | – | – |
|        | No | No | HER2-PQS3 | GGGGGGAGG | Not bound | – | – |
|        | No | No | HER2-PQS4 | GGGCTGCTGGG | Not bound | – | – |
|        | No | No | HER2-PQS5 | GGGAGG | Not bound | – | – |
|        | No | No | HER2-PQS6 | GGGGCGGCGGGTGG | Not bound | – | – |
|        | No | No | NSE-PQS1 | GGGAGAAGGAGG | Not bound | – | – |
|        | No | No | NSE-PQS2 | GGGAAAGGAGG | Not bound | – | – |
|        | No | No | NSE-PQS3 | GGGCGGGGAA | Not bound | – | – |
|        | No | No | NSE-PQS4 | GGGCGGGGAGG | Not bound | – | – |
|        | No | No | NSE-PQS5 | GGGTGAGATGGGA | Not bound | – | – |
|        | No | No | PDGFBB-PQS1 | GGGGCCCCGGG | – | 3.0 × 10⁻⁸ | parallel |
|        | No | No | PDGFBB-PQS2 | GGGTTCCGGG | – | 5.0 × 10⁻⁸ | parallel |
|        | No | No | PDGFBB-PQS3 | GGGGGGGGGG | – | 5.2 × 10⁻⁸ | parallel |
|        | No | No | PDGFBB-PQS4 | GGGGCTGGGGA | – | 4.4 × 10⁻⁸ | parallel |
| Target | NLS by NLSdb | NLS by cNLS Mapper | Name | Sequence (5′ → 3′) | Result of gel-shift assay | K_D (M) by SPR | G4 topology |
|--------|--------------|-------------------|------|-------------------|--------------------------|----------------|-------------|
| PDGFBB-PQS5 | | | | GGGGGCCCAGGG | – | 6.7 × 10⁻⁸ | parallel |
| PDGFBB-PQS6 | | | | GGGGCCCAGGTA | – | 5.5 × 10⁻⁸ | parallel |
| PDGFBB-PQS7 | | | | GGGGGCCCGGG | – | 8.5 × 10⁻⁸ | parallel |
| PDGFBB-PQS8 | | | | GGGGCACTCAGGGTAGG | – | 1.5 × 10⁻⁷ | hybrid or mixture |
| Annexin 2 | No | No | Annexin2-PQS1 | GGACCCTGCGG | Bound | – | hybrid or mixture |
| | | | Annexin2-PQS2 | GGCGCCTGGCGC | Bound | – | anti-parallel |
| | | | Annexin2-PQS3 | GAGCCTGCGG | Bound | – | parallel |
| | | | Annexin2-PQS4 | GGGCCGCGGG | Not bound | – | parallel |
| | | | Annexin2-PQS5 | GGGGAGGAGGG | Not bound | – | hybrid or mixture |
| | | | Annexin2-PQS6 | GGGCGCGCGGG | Bound | – | parallel |
| | | | Annexin2-PQS7 | GGGGCCAGGG | Bound | – | parallel |
| ApoE4 | No | No | ApoE4-PQS1 | GGGCCCGG | Not bound | – | parallel |
| | | | ApoE4-PQS2 | GCTGGCCGAGG | Not bound | – | hybrid or mixture |
| | | | ApoE4-PQS3 | GCCCTGCTGG | Not bound | – | parallel |
| | | | ApoE4-PQS4 | GAGGCGGCGCC | Not bound | – | parallel |
| | | | ApoE4-PQS5 | GATGCGGACGG | Not bound | – | hybrid or mixture |
| | | | ApoE4-PQS6 | GACGCGGG | Not bound | – | parallel |
| | | | ApoE4-PQS7 | GTGGCTGGTCA | Not bound | – | anti-parallel |
| | | | ApoE4-PQS8 | GGGCTGGG | Not bound | – | parallel |
| CS protein | No | No | CS protein-PQS1 | GGGGCGGGAGG | Not bound | – | – |
| | | | CS protein-PQS2 | GALLAAAGGG | Not bound | – | – |
| | | | CS protein-PQS3 | GGGCCCGGA | – | Not bound | hybrid or mixture |
| | | | CS protein-PQS4 | GGGGCGGGAGG | – | Not bound | hybrid or mixture |
| PLGF | No | No | PLGF-PQS1 | GGGCCCGGA | – | Not bound | hybrid or mixture |
| | | | PLGF-PQS2 | GGGGCGGGAGG | – | Not bound | hybrid or mixture |
| | | | PLGF-PQS3 | GGGCCCGGGAGG | – | Not bound | hybrid or mixture |
| | | | PLGF-PQS4 | GGGGCGGGAGG | – | Not bound | hybrid or mixture |
| TNF-α | No | No | TNFα-PQS1 | GGGTGGCAGGG | Not bound | – | parallel |
| | | | TNFα-PQS2 | GGGGATGGGAGG | Not bound | – | parallel |
| | | | TNFα-PQS3 | GAGGCGGAGG | Not bound | – | parallel |
| | | | TNFα-PQS4 | GAGGCGGAGG | Not bound | – | parallel |

- : Not investigated.
Fig. 1. SPR sensorgram for the $K_D$ determination of HGF-PQSs.

2. Experimental Design, Materials and Methods

2.1. Materials

All non-labelled and FITC-labelled DNA oligonucleotides were purchased from Eurofins Genomics (Tokyo, Japan) with HPLC purification and stored in TE buffer (10 mM Tris–HCl, 0.1 mM EDTA; pH 8.0) at the concentration of 100 μM. VEGFA (VEGF165 and VEGF121) and recombinant human PDGF-AA, PDGF-BB and PLGF were purchased from R&D Systems (Minneapolis, MN, USA). Recombinant human RB1 and the intracellular domain of recombinant human c-KIT (corresponding to amino acids 544–976) were purchased from Abcam (Cambridge, UK). The extracellular domain of recombinant human c-KIT (corresponding to amino acids 1–516) was purchased from Sino Biological (Beijing, China). ApoE4, Annexin2, CS protein, TNF-α, were purchased from MP Biomedicals (Irvine, CA, USA), AbD Serotec (Kidlington, UK), ProSpec (Rehovot, Israel), and Cell Signaling Technology (Danvers, MA, USA) respectively. 6X Loading Buffer was purchased from TAKARA BIO INC. (Shiga, Japan). Acrylamide, $N,N'$-methylenebisacrylamide, ammonium persulfate, $N,N',N''$-Tetramethylethlenediamine (TEMED), HEPES, and Tris(hydroxymethyl)aminomethane were purchased from FUJIFILM Wako Pure Chemical Corporation (Osaka, Japan). Hydrochloric acid, sodium acetate, sodium hydroxide, sodium hydrogen phosphate, potassium dihydrogen phosphate, sodium chloride, potassium chloride, methanol, acetic acid, and boric acid were purchased from Kanto Chemical Co., Inc. (Tokyo, Japan). Ethylenediaminetetraacetic acid (EDTA) was purchased from Dojindo Molecular Technologies, Inc. (Kumamoto, Japan).
Fig. 2. SPR sensogram for the $K_D$ determination of HBEGF-PQSs.

2.2. Nuclear localization signal (NLS) search

For the NLS prediction, all the amino acid sequences of target proteins including its isoforms were obtained from UniProt (https://www.uniprot.org). The obtained sequences were subjected to NLS prediction by web tools - NLSdb (https://rostlab.org/services/nlsdb/) [5] and cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) [6]. Prediction by cNLS Mapper were carried out with the cut-off score of 4.0 within the entire region of protein sequence.

2.3. G-quadruplex-forming sequence search

Genomic DNA sequences 1 kbp upstream and 1 kbp downstream from the transcription start site of a target protein-coding region were extracted using the UCSC genome browser (https:
Fig. 3. SPR sensorgram for the $K_D$ determination of aFGF-PQSs.
Putative G-quadruplex-forming sequences within the genomic DNA sequences were extracted using the QGRS mapper (http://bioinformatics.ramapo.edu/QGRS/index.php) [4] with the criterion of $G_2 < N_{1-7} G_2 < N_{1-7} G_2 < N_{1-7} G_2 <$, where “G” is guanine base and “N” can be any bases.
Fig. 5. SPR sensorgram for the $K_D$ determination of CRP-PQSs.

Fig. 6. SPR sensorgram for the $K_D$ determination of PDGFBB-PQSs.

Fig. 7. SPR sensorgram for the $K_D$ determination of ApoE4-PQSs.
Fig. 8. SPR sensorgram for the $K_D$ determination of CS protein-PQSs.

Fig. 9. SPR sensorgram for the $K_D$ determination of PLGF-PQSs.

Fig. 10. Result of gel-shift assay of AFP-PQS.
2.4. Surface plasmon resonance (SPR) measurement

SPR measurement was carried out using a Biacore T200 instrument (GE Healthcare, Buckinghamshire, UK). Each protein was immobilized on a sensor chip CM5 (GE Healthcare) by an amine coupling in appropriate buffer considering the isoelectric point; VEGF165 immobilization
buffer (10 mM acetate; pH 6.0), HGF immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 6.5), HBEGF immobilization buffer (10 mM acetate; pH 5.0), aFGF immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 7.0), bFGF immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 8.0), PDGF-AA immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 7.0), PDGF-BB immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 7.0), ApoE4 immobilization buffer (10 mM acetate; pH 4.0), CS protein immobilization buffer (10 mM acetate; pH 4.5), or PLGF immobilization buffer (10 mM HEPES, 150 mM NaCl, 5 mM KCl; pH 6.5) were used for the corresponding biomarker protein immobilization. When RU reached certain value (Approximately 7500 RU for VEGF165, 1900 RU for PDGF-AA, 4000 RU for PDGF-BB, 1300 RU for ApoE4, 1200 RU for CS protein).
Fig. 15. Result of gel-shift assay of Annexin2-PQSs.

Fig. 16. Result of gel-shift assay of ApoE4-PQSs.
for HGF, 5000 RU for HBEGF, 3000 RU for aFGF, 3000 RU for bFGF, 1150 RU for CRP, 700 RU for PDGF-BB, 900 RU for ApoE4, 900 RU for CS protein, or 1200 RU for PLGF) the chip was used for the binding analysis.

For binding, oligonucleotides were diluted in TBS buffer (10 mM Tris–HCl, 150 mM NaCl, 100 mM KCl; pH 7.4) and heated to 95 °C for 5 min and then cooled to 25 °C gradually over 30 min. The heat-treated oligonucleotides were further diluted to various concentrations using TBS buffer, and were injected into the target protein immobilized sensor chip and SPR signals were measured. The signal of the reference cell, which was treated by the amine-coupling reagent with ethanolamine without protein immobilization, was subtracted from that of the protein-immobilized cell. In all measurements, the DNA association time was 120 s, dissociation time was 120 s, and flow rate was 30 μL/min at 25 °C. TBS buffer was used as the running buffer and 1 M NaCl for the dissociation. K_D was calculated by applying curve fitting using BIAevaluation software (GE Healthcare, Buckinghamshire, UK).

Fig. 17. Result of gel-shift assay of CS protein-PQSs.
Fig. 18. Result of gel-shift assay of TNFα-PQSs.

Fig. 19. CD spectrum of RB1-PQS.
2.5. Circular dichroism (CD) spectroscopy analysis

DNA oligonucleotide samples were diluted to 2 μM in Tris buffer (10 mM Tris–HCl, 150 mM NaCl; pH 7.4) or TBS buffer (10 mM Tris–HCl, 150 mM NaCl, 100 mM KCl; pH 7.4), and were heated to 95 °C for 5 min and then gradually cooled to 25 °C over 30 min. 50 μL of the prepared sample was added into a quartz cell; Micro cell 50 μL 10 mm Path UV (Agilent Technologies, CA), and CD spectra were measured in the range of 220–320 nm using a J-820 spectropolarimeter (JASCO, Tokyo, Japan) with the optical path of 10 mm at 20 °C.

2.6. Gel-shift assay

FITC-labelled oligonucleotides were diluted to 1 μM in TBS buffer (10 mM Tris–HCl, 150 mM NaCl, 100 mM KCl; pH 7.4) and heated to 95 °C for 5 min and then cooled down to 25 °C gradually. The heat-treated oligonucleotides and target proteins were mixed in TBS at the final concentration of 500 nM and 1 μM, respectively. The mixed samples were incubated with shaking (1200 rpm) for 30 min at 25 °C with High Speed Shaker ASCM-1 (AS ONE CORPORATION, Osaka, Japan). The prepared sample was mixed with loading buffer (6% glycerol, 5 mM EDTA, 0.008% bromophenol blue, 0.0058% xylene cyanol), and electrophoresed in 12% polyacrylamide gel in TBE buffer (90 mM Tris, 90 mM Boric acid, 2 mM EDTA, pH 8.16), followed by scanning the gel using Typhoon8600 (GE Healthcare, Chicago, IL, USA).
Fig. 22. CD spectrum of c-KIT-PQSs.

Fig. 23. CD spectrum of HBEGF-PQSs.
Fig. 24. CD spectrum of HGF-PQSs.

Fig. 25. CD spectrum of PDGFBB-PQSs.
Annexin2-PQS1  
Annexin2-PQS2  
Annexin2-PQS3  
Annexin2-PQS4

Annexin2-PQS5  
Annexin2-PQS6  
Annexin2-PQS7

ApoE4-PQS1  
ApoE4-PQS2  
ApoE4-PQS3  
ApoE4-PQS4

ApoE4-PQS5  
ApoE4-PQS6  
ApoE4-PQS7  
ApoE4-PQS8

PLGF-PQS1  
PLGF-PQS2  
PLGF-PQS3  
PLGF-PQS4

TNFα-PQS1  
TNFα-PQS2  
TNFα-PQS3  
TNFα-PQS4

Fig. 26. CD spectrum of Annexin2-PQSs.

Fig. 27. CD spectrum of ApoE4-PQSs.

Fig. 28. CD spectrum of PLGF-PQSs.

Fig. 29. CD spectrum of TNFα-PQSs.
CRedit Author Statement

Jinhee Lee: Investigation, Visualization, Writing - Original Draft; Kentaro Teramoto: Investigation; Tomomi Yokoyama: Investigation; Kinuko Ueno: Investigation; Kaori Tsukakoshi: Supervision; Koji Sode: Supervision; Kazunori Ikebukuro: Conceptualization, Project administration, Supervision, Validation, Writing - Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107028.

References

[1] W. Yoshida, T. Saito, T. Yokoyama, S. Ferri, K. Ikebukuro, Aptamer selection based on G4-forming promoter region, PLoS ONE 8 (6) (2013) e65497, doi: 10.1371/journal.pone.0065497.
[2] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, Trends Cell Biol. 19 (8) (2009) 414–422, doi:10.1016/j.tcb.2009.05.002.
[3] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, The regulation and functions of DNA and RNA G-quadruplexes, Nature Rev. Mol. Cell Biol. 21 (2020) 259–474, doi:10.1038/s41580-020-0236-x.
[4] O. Kikin, L. D’Antonio, P.S. Bagga, QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, Nucleic Acids Res., 34 (2006) W676–W682, doi:10.1093/nar/gkl253.
[5] R. Nair, P. Carter, B. Rost, NLSdb: database of nuclear localization signals, Nucleic. Acids Res. 31 (2003) 397–399, doi:10.1093/nar/gkg001.
[6] S. Kosugi, M. Hasebe, M. Tomita, H. Yanagawa, Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs, Proc. Natl. Acad. Sci. USA 106 (25) (2009) 10171–10176 23, doi:10.1073/pnas.0900604106.
[7] J. Lee, K. Teramoto, T. Yokoyama, K. Ueno, K. Tsukakoshi, K. Sode, K. Ikebukuro, Data on G-quadruplex topology, and binding ability of G-quadruplex forming sequences found in the promoter region of biomarker proteins and those relations to the presence of nuclear localization signal in the proteins, Mendeley Data V3 (2021), doi:10.17632/5xthvrbsp.3.
[8] Y. Yokoyama, K. Tsukakoshi, W. Yoshida, T. Saito, K. Teramoto, N. Savory, K. Abe, Ikebukuro K, Development of HGF-binding aptamers with the combination of G4 promoter-derived aptamer selection and in silico maturation, Biotechnol. Bioeng. 114 (10) (2017) 2196–2203, doi:10.1002/bit.26354.