Supplementary Information (SI) for Physical Chemistry Chemical Physics.

NMR-based conformational analysis of DNA G-quadruplex guides mapping essential structure-function relationship in protein chaperoning

Deepraj Negi,[a†] Zijue Huang,[b†] Ahyun Son,[b†] Bharathwaj Sathyamoorthy,[*a] Scott Horowitz[*b]

[a]Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Madhya Pradesh 462066, India.

[b]Department of Chemistry & Biochemistry and the Knoebel Institute for Healthy Aging, University of Denver, Denver, CO, 80208, USA.

[†]Authors contributed equally to this work.
*Corresponding and co-senior authors: scott.horowitz@du.edu, bharathwaj@iiserb.ac.in

Experimental procedures

NMR sample preparation

The Seq576 DNA sequence and the subsequent mutants were all purchased from Integrated DNA Technologies (IDT Inc.) at one micromole scale with standard desalting purification and used without any further purifications. The concentration of single-stranded oligonucleotides was determined by measuring absorbance at 260 nm with a Shimadzu UV-1800 UV-Vis spectrophotometer at 25 ºC using a Helma high-precision cuvette with a sample volume of 70 µL (duplicates performed to ensure data reproducibility). The nearest-neighbor model from the IDT DNA online tool was used to calculate molar absorptivity (https://eu.idtdna.com/calc/analyzer).

The single-strand DNA samples obtained from IDT were diluted to ~8.5 mL with nanopure water. The sample was then heated to 95 °C for 5 minutes, followed by the addition of potassium chloride (KCl) and potassium phosphate stock buffer (pH 7.5) to a final concentration of 50 mM and 10 mM, respectively, to a total volume of ~10 mL (single-stranded DNA concentration of ~50 µM). This mixture was further heated at 95 °C for 5 minutes and set aside at room temperature overnight. The annealed oligonucleotides were then transferred to a 15 mL centrifugal concentrator (3 kDa cut-off, Amicon, Merck Inc.) and spun at a speed of 4000x g for ~50 min at 4 °C to a final volume of ~250 µL (DNA concentration of ~1-1.5 mM). Concentrated samples were made to 300 µL by adding phosphate buffer, D_2O towards field-frequency locking (5% v/v), and trimethylsilylpropanoic standard (TSP, 50 µM) as chemical shift standard. The annealed samples were then transferred to a medium-walled NMR tube (Norell Inc.) for NMR data acquisition.

Sample preparation for chaperone studies

All sequences were resuspended in 10 mM potassium phosphate (pH 7.5) at room temperature, then annealed at 95 °C for 2 minutes followed by cooling from 95 °C to 25 °C at 1 °C per minute while shaking on Eppendorf ThermoMixer C. All stock concentrations were 100 µM, measured with Thermo Scientific NanoDrop UV-Vis Spectrophotometer.

Primary sequences

Sequence Name	Primary sequence (5'- to 3'-end)
Seq576	TGT CGG GC<u>G</u> GG<u>G</u> AGG G<u>G</u>G GG
Seq576^{G9T}	TGT CGG GC**T** GGG AGG GGG GG
Seq576^{G9I}	TGT CGG GC**I** GGG AGG GGG GG
Seq576^{G12T}	TGT CGG GCG GG**T** AGG GGG GG
Seq576^{G12I}	TGT CGG GCG GG**I** AGG GGG GG
Seq576^{G17T}	TGT CGG GCG GGG AGG G**T**G GG
Seq576^{G17I}	TGT CGG GCG GGG AGG G**I**G GG
Seq576^{G17A}	TGT CGG GCG GGG AGG G**A**G GG
Seq576^{G17C}	TGT CGG GCG GGG AGG G**C**G GG
Seq576^{G17AP}	TGT CGG GCG GGG AGG G_G GG
Seq576^{G17isoG}	TGT CGG GCG GGG AGG G**g**G GG
Seq42	AAC GAA AGA ACA TAA TCT CG

Table S1. Abbreviation of each sequence used in this study is provided in column 1, with single point mutations of a guanine (G) to thymine (T), inosine (I), isoguanosine (g), or apurinic site ("_"). Positions 9, 12, and 17 are underlined. Seq42 serves as a non-G4 control of the same length in all bioassays.

NMR data acquisition

All the NMR experiments were acquired on a Bruker Avance 700 MHz spectrometer equipped with an AVANCE III console and a triple resonance $^{1}H/^{13}C/^{15}N$ cryogenically cooled probe with z-axis pulse field gradient coil at 298 K. 1D and 2D [^{1}H-^{13}C] heteronuclear single-quantum coherence (HSQC) and [^{1}H-^{1}H] nuclear Overhauser effect spectroscopy (NOESY) experiments were acquired using the standard Bruker programs. HSQC experiment was acquired with an overall spectral width of 14 and 16 ppm in the ^{13}C and ^{1}H dimension (acquisition time of 64 and 92 ms), respectively, with the NOESY at 22 ppm on both dimensions (acquisition time of 50 and 133 ms in F1 and F2 dimensions, respectively). Extensive spectral aliasing was adopted for the

^{13}C dimension to obtain a high resolution, with the carrier positioned at 81.5 ppm. An inter-scan delay of 1.5/2.5 s was used for the 2D HSQC/NOESY experiments, resulting in an overall measurement time of ~4.5-9.5/~5-10 hours (32-64/4-8 scans per FID). All NMR data were processed using NMRPipe and analyzed using NMRFAM-SPARKY[25, 26].

Circular dichroism of samples used in chaperone assays

CD spectra were measured in a 1 mm pathlength quartz cuvette using a Jasco J-1100 circular dichroism spectrophotometer. All samples were annealed at 20 μM G4s in 10 mM potassium phosphate (pH 7.5). Spectra were taken from 300 to 190 nm at 25 °C with 1 nm intervals, a scanning speed of 50 nm/min, and a data integration time (D.I.T.) of 8 seconds. The final spectra shown are the average of three accumulations. CD spectra of all quadruplexes and mutants show parallel structure topology (Figure S8), as does the preparation for NMR experiments (Figure 1A).

Chaperone activity

Thermal aggregation plate reader assay

Aggregation of 0.5 μM citrate synthase (CS) from porcine heart (Sigma-Aldrich) was measured by light scattering, and recorded as absorbance at 360 nm in a BioTek plate reader at 50 °C using black, clear, flat-bottom, half-area plates (Corning 3880), with shaking and measuring every 40 s for a total of 90 minutes as previously described[3, 9, 12]. An increasing absorbance signal at 360 nm indicates aggregate formation. The CS:DNA ratio was kept at 1:2 in all the assays. All sequences were run in triplicate for reproducibility. Seq42 (Table S1) and Herring testes DNA (Sigma-Aldrich) were used as controls.

Refolding fluorescence assay

TagRFP675 was chemically denatured in 6 M guanidine hydrochloride (GuHCl) overnight. Seq576 was initially diluted in 10 mM phosphate buffer in the quartz fluorometer cuvette, then TagRFP675 was introduced at 1 minute under constant stirring as previously described[10]. The

dilution of the GuHCl into the folding buffer leaves a residual 150 mM GuHCl. The refolding assays were performed with 0.5 μM of TagRFP675, with the protein:DNA ratio was kept at a ratio of 1:2. All trials were run in triplicate for reproducibility. Seq42 was used as a non-G4 control for the refolding assays.

Native PAGE

Native polyacrylamide gel electrophoresis (PAGE) was performed as previously described[12]. Briefly, a 20% Novex Tris-Borate-EDTA (TBE) gel was pre-run in 1x TBE running buffer with 10 mM KCl for an hour before loading and was run at 4 °C for 60-90 mins. Gel was stained with N-methyl mesoporphyrin IX (NMM) in 1x TBE to bind quadruplex first and visualized using SYBR green filter, subsequently gel was stained with SYBR gold for total DNA using SYBR gold filter. Gels were visualized on a BioRad Chemidoc imaging system, and gels were overlayed in ImageJ to confirm whether the higher-order structures were quadruplexes. For quantification of sequences oligomerization on all SYBR gold gels, Seq42 (Table S1) was used as a control for normalization. Correlations were performed using GraphPad Prism 10.

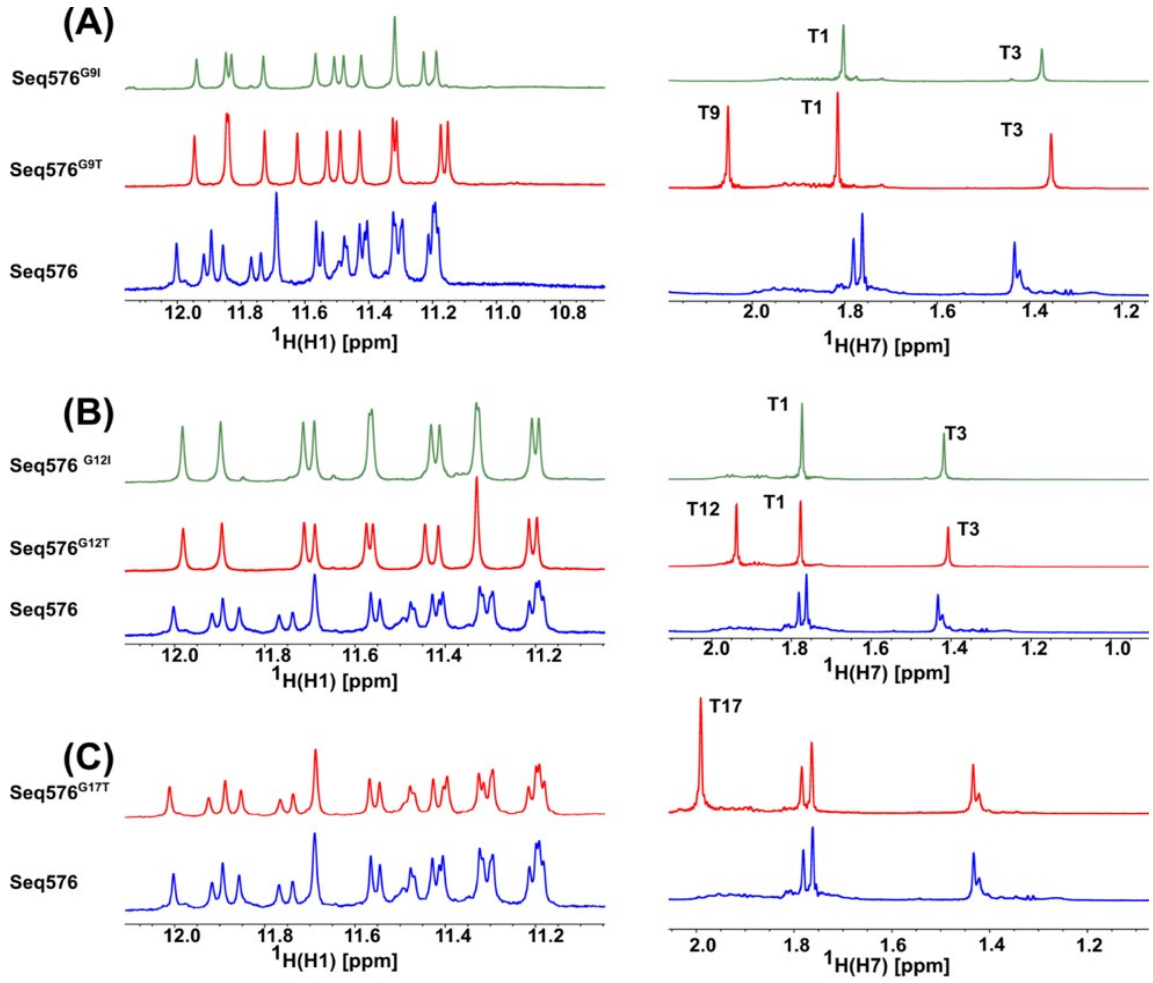

Figure S1. Imino (left) and methyl (right) ^{1}H 1D spectra acquired for Seq576 and its single mutants. (A) Comparison of Seq576 mutants in position 9 with thymine (Seq576^{G9T}) and inosine (Seq576^{G9I}) modifications. (B) Comparison of Seq576 mutants in position 12 with thymine (Seq576^{G12T}) and inosine (Seq576^{G12I}) modifications. (C) Comparison of Seq576 mutants in position 17 with thymine (Seq576^{G17T}) modification.

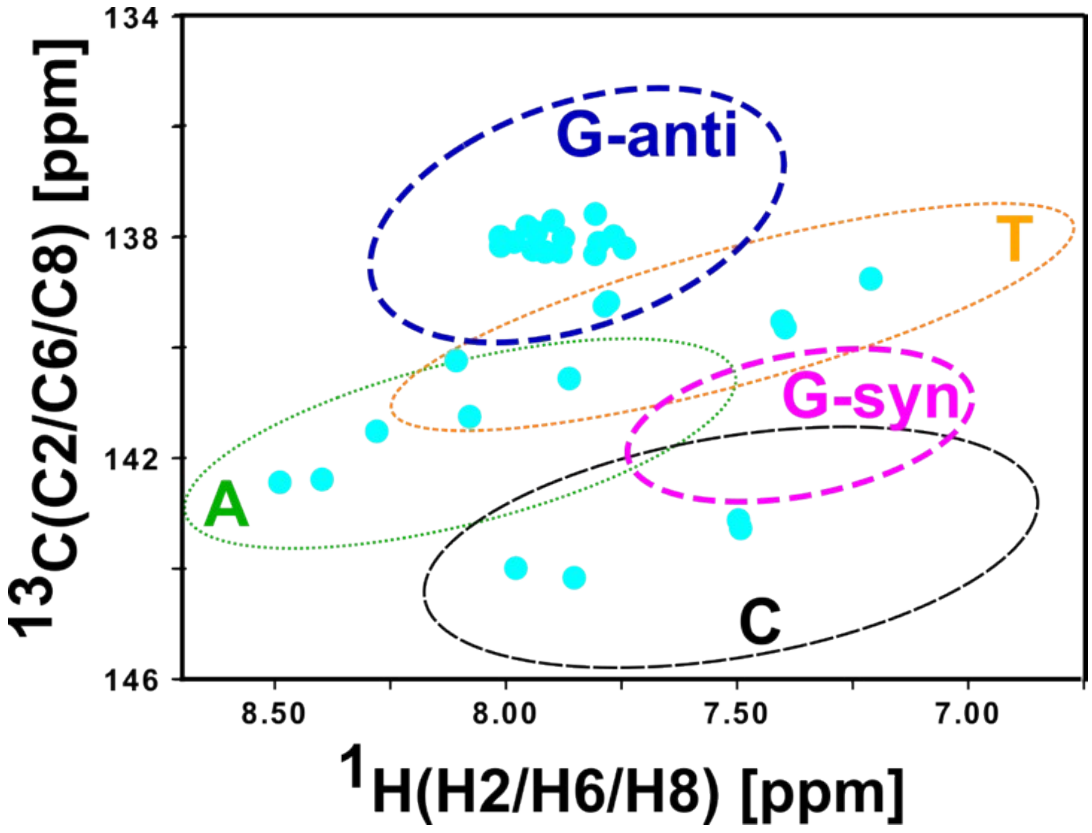

Figure S2. ^{13}C-^{1}H aromatic peaks picked from the HSQC spectrum for Seq576 (represented as cyan-filled circles) overlaid with chemical shift clusters observed for tetrad-forming guanosine nucleotides in anti (G-anti, blue ellipse) and syn (G-syn, magenta ellipse). Loop nucleotides adenine (green ellipse), thymine (orange ellipse), and cytosine (black ellipse) are also represented. Ellipses for each of the clusters represent regions where the corresponding peaks appear for a given nucleotide (refer to reference #10 from the manuscript for more details). Peaks picked from the 2D data (cyan circles) when overlaid on the clusters clearly indicate the absence of G-syn, implying parallel topology for both the states sampled by Seq576. Intranucleotide resonance assignments obtained from the mutants were then used to predict plane-specific assignments using the machine-learning methodology.

Seq576^{G9T}								
S. No.	**H8**	**H1'**	**H2'**	**H2"**	**P1**	**P2**	**P3**	**Predicted plane**
1	7.77	6.22	2.72	2.98	0	1000	0	P2
2	7.79	6.36	2.40	2.59	0	0	1000	P3
3	7.82	6.44	2.55	2.70	0	0	1000	P3
4	7.82	6.44	2.55	2.70	0	0	1000	P3
5	7.86	6.23	2.71	2.85	0	1000	0	P2
6	7.86	6.26	2.51	2.69	0	0	1000	P3
7	7.89	6.14	2.53	2.96	399	601	0	P2
8	7.95	6.25	2.69	2.96	0	1000	0	P2
9	7.96	6.09	2.69	2.86	0	1000	0	P2
10	7.97	6.18	2.50	2.90	999	1	0	P1
11	8.04	6.12	2.79	3.01	1000	0	0	P1
12	8.04	6.18	2.55	2.96	1000	0	0	P1

Seq576^{G12T}								
S. No.	**H8**	**H1'**	**H2'**	**H2"**	**P1**	**P2**	**P3**	**Predicted plane**
1	7.80	6.38	2.42	2.60	0	0	1000	P3
2	7.82	6.45	2.56	2.70	0	0	1000	P3
3	7.84	6.39	2.52	2.67	0	0	1000	P3
4	7.84	6.45	2.56	2.70	0	0	1000	P3
5	7.85	6.25	2.72	2.98	0	1000	0	P2
6	7.85	6.25	2.72	2.98	0	1000	0	P2
7	7.96	6.17	2.78	2.82	0	1000	0	P2
8	7.97	6.17	2.51	2.89	970	30	0	P1
9	7.97	6.11	2.71	2.86	0	1000	0	P2
10	7.98	6.14	2.48	2.91	1000	0	0	P1
11	8.01	6.15	2.64	2.95	1000	0	0	P1
12	8.06	6.15	2.80	3.04	1000	0	0	P1

Table S2. Machine learning based prediction of resonance assignments of Seq576 single mutants. Intranucleotide H8, H1', H2'/H2" assignments (columns 2-5) of tetrad-forming guanosine resonances fed into the machine learning model to predict plane-specific (P1/P2/P3) assignments. The predictions are repeated 1000 times by varying the training dataset (as described previously, reference #10) to assess the reliability of predictions. In the case of Seq576^{G9T} (row number 7), the prediction was found to be incorrectly predicted as plane-2 (P2), which is verified from the NOESY assignments to be part of P1. Apart from that one wrong prediction, all the other plane-specific assignments proved helpful for the sequential assignment process.

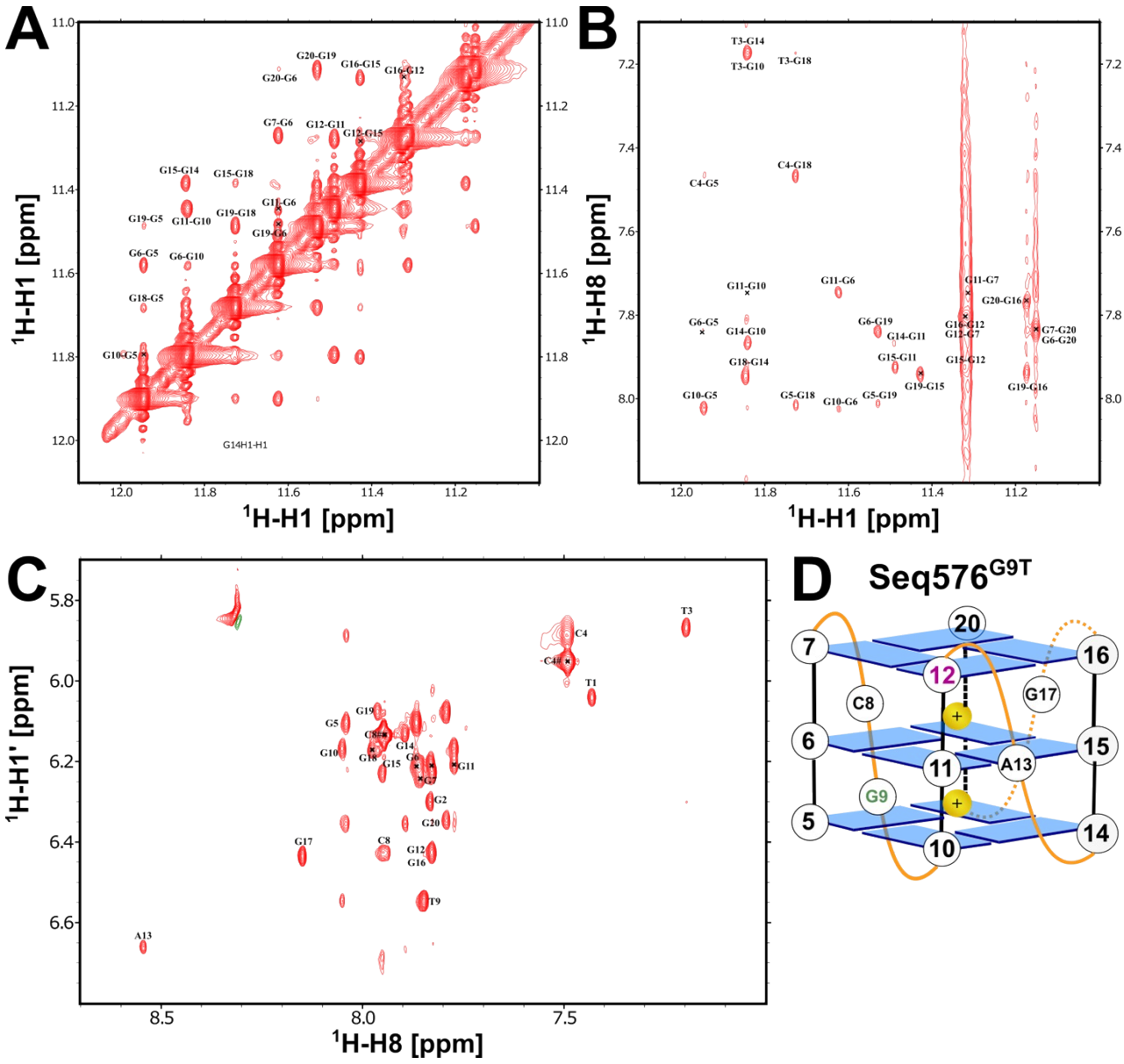

Figure S3. NOESY sequential imino walk assigned for Seq576^{G9T}.

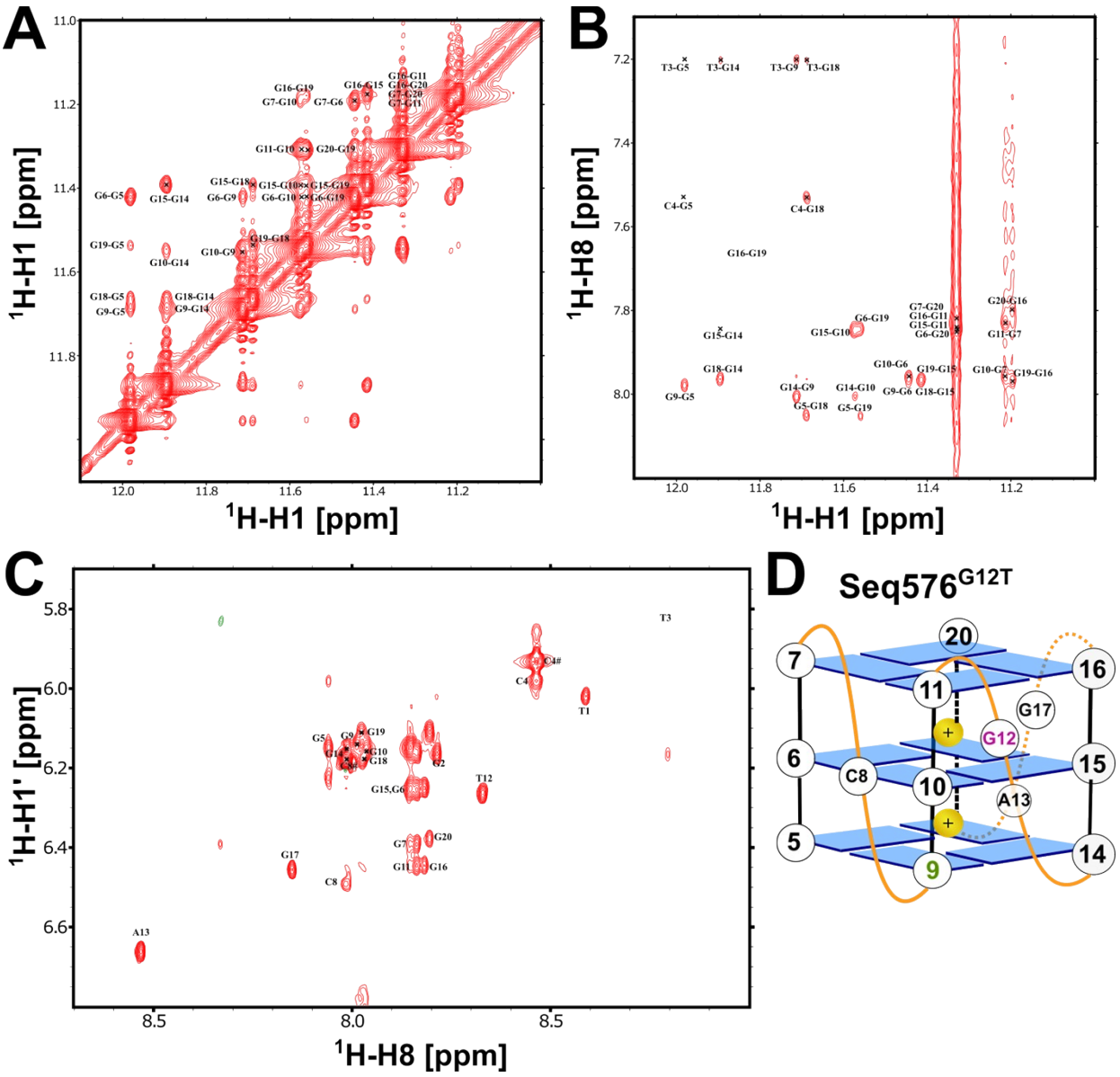

Figure S4. NOESY sequential imino walk assigned for Seq576^{G12T}.

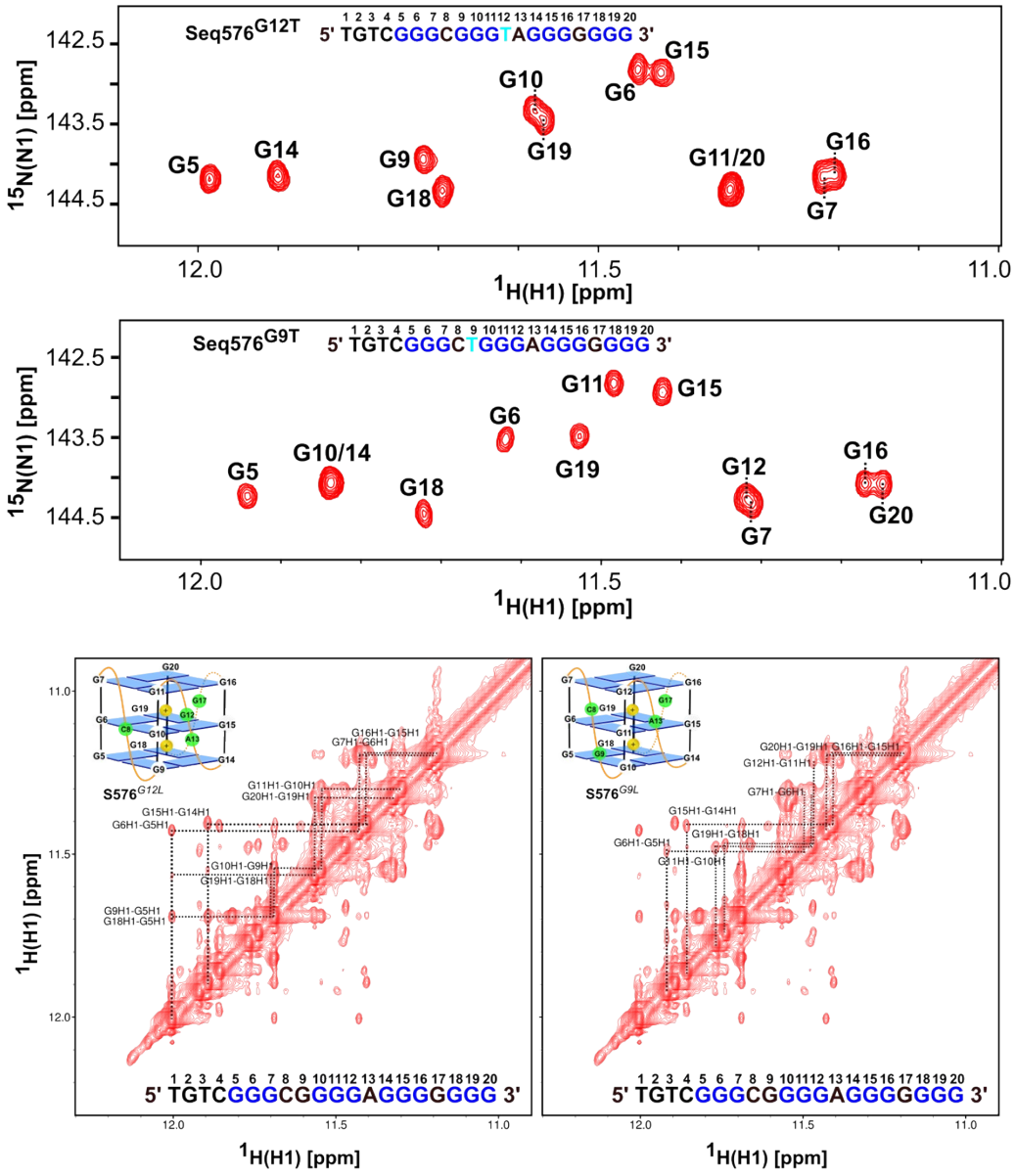

Figure S5. 2D ^{15}N-^{1}H imino correlation and the NOESY based sequential imino walk assigned from single mutants of Seq576: Seq576^{G12T} and Seq576^{G9T}.

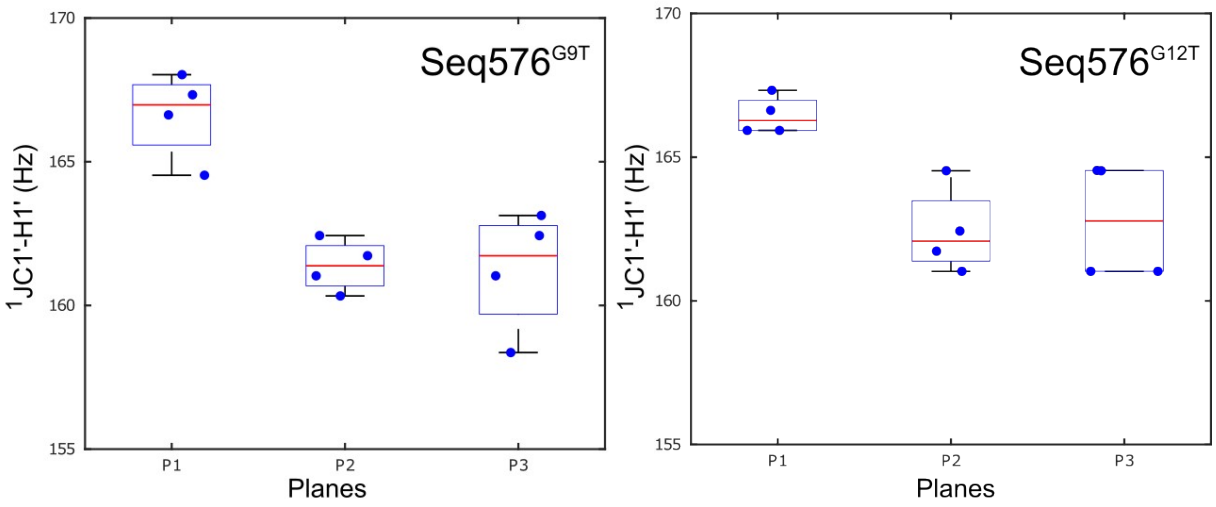

Figure S6. $^{1}J_{C1'–H1'}$ scalar couplings measured for Seq576^{G9T} and Seq576^{G12T}, plotted across their positions in planes 1 (P1), 2 (P2) and 3 (P3). Mean across the measurements is shown as red line, while the box represents the spread of observed values.

Seq576^{G9T}	C1’	H1’	C8	H8	N1	H1	C2	H2	C5	H5	C6	H6	C2'	H2'	H2''
T1	87.5	6.06	-	-	-	-	-	-	-	-	139.6	7.43	39.9	1.87	2.29
G2	85.9	6.32	139.1	7.83	-	-	-	-	-	-	-	-	41.0	2.42	2.46
T3	87.6	5.89	-	-	-	-	-	-	-	-	138.6	7.20	41.4	1.93	2.25
C4	89.4	5.91	-	-	-	-	-	-	98.5	5.97	143.2	7.49	41.9	1.91	2.48
G5	85.0	6.12	137.8	8.04	144.2	11.94	-	-	-	-	-	-	40.4	2.79	3.01
G6	84.8	6.23	137.7	7.87	143.5	11.62	-	-	-	-	-	-	42.9	2.71	2.85
G7	84.9	6.26	138.1	7.86	144.3	11.31	-	-	-	-	-	-	42.9	2.52	2.69
C8	88.6	6.45	-	-	-	-	-	-	99.7	6.15	144.2	7.95	41.9	2.32	2.65
T9	87.7	6.57	-	-	-	-	-	-	-	-	140.5	7.85	39.8	2.63	2.68
G10	85.6	6.19	138.1	8.05	144.1	11.84	-	-	-	-	-	-	42.4	2.55	2.96
G11	84.2	6.23	137.6	7.77	142.8	11.48	-	-	-	-	-	-	42.4	2.72	2.98
G12	85.0	6.45	138.2	7.83	144.2	11.32	-	-	-	-	-	-	41.4	2.55	2.70
A13	86.4	6.68	142.5	8.54			155.5	8.31	-	-	-	-	-	-	-
G14	87.4	6.15	137.7	7.89	144.1	11.84	-	-	-	-	-	-	43.8	2.53	2.96
G15	84.4	6.25	138.3	7.95	142.9	11.42	-	-	-	-	-	-	42.9	2.69	2.96
G16	84.9	6.45	138.1	7.83	144.1	11.17	-	-	-	-	-	-	41.4	2.56	2.70
G17	86.3	6.46	140.2	8.15	-	-	-	-	-	-	-	-	41.5	2.82	2.93
G18	87.1	6.19	138.0	7.98	144.4	11.72	-	-	-	-	-	-	44.1	2.50	2.91
G19	84.2	6.10	138.3	7.96	143.5	11.53	-	-	-	-	-	-	42.7	2.69	2.87
G20	84.8	6.37	137.9	7.79	144.1	11.15	-	-	-	-	-	-	42.8	2.40	2.59

Table S3. Table of ^{1}H, ^{13}C and ^{15}N chemical Shifts for Seq576^{G9T} in the presence of potassium ions (BMRB ID: 53322).

Seq576^{G12T}	C1’	H1’	C8	H8	N1	H1	C2	H2	C5	H5	C6	H6	C2'	H2'	H2''
T1	87.5	6.02	-	-	-	-	-	-	-	-	139.6	7.41	39.9	1.89	2.26
G2	86.1	6.16	139.2	7.78	-	-	-	-	-	-	-	-	40.8	2.41	2.47
T3	87.6	5.85	-	-	-	-	-	-	-	-	138.7	7.20	41.2	2.12	2.32
C4	89.3	5.98	-	-	-	-	-	-	98.5	5.93	143.3	7.53	41.8	1.96	2.41
G5	85.2	6.15	138.0	8.06	144.2	11.99	-	-	-	-	-	-	40.7	2.80	3.04
G6	85.0	6.25	137.7	7.85	142.8	11.45	-	-	-	-	-	-	42.8	2.72	2.98
G7	84.8	6.39	138.1	7.84	144.2	11.22	-	-	-	-	-	-	42.7	2.52	2.67
C8	88.6	6.49	-	-	-	-	-	-	99.5	6.18	144.0	8.01	42.3	2.42	2.75
G9	87.2	6.14	137.9	7.99	143.9	11.72	-	-	-	-	-	-	44.1	2.48	2.91
G10	84.2	6.16	138.4	7.96	143.3	11.58	-	-	-	-	-	-	42.7	2.78	2.82
G11	85.1	6.45	138.1	7.84	144.3	11.34	-	-	-	-	-	-	41.5	2.56	2.70
T12	87.6	6.27	-	-			-	-	-	-	140.0	7.67	41.0	2.25	2.48
A13	87.0	6.66	143.2	8.53			155.6	8.33	-	-	-	-	40.7	2.98	3.08
G14	86.0	6.15	138.1	8.01	144.1	11.90	-	-	-	-	-	-	42.4	2.64	2.95
G15	84.7	6.26	137.9	7.86	142.9	11.42	-	-	-	-	-	-	42.8	2.72	2.98
G16	85.1	6.45	138.1	7.82	144.1	11.21	-	-	-	-	-	-	41.5	2.56	2.70
G17	86.4	6.46	140.2	8.15	-	-	-	-	-	-	-	-	41.6	2.82	2.92
G18	87.1	6.18	137.9	7.97	144.3	11.70	-	-	-	-	-	-	44.1	2.51	2.89
G19	84.2	6.11	138.3	7.98	143.5	11.57	-	-	-	-	-	-	42.7	2.71	2.86
G20	84.9	6.38	138.0	7.81	144.3	11.34	-	-	-	-	-	-	42.9	2.42	2.60

Table S4. Table of ^{1}H, ^{13}C and ^{15}N chemical Shifts for Seq576^{G12T} in the presence of potassium ions (BMRB ID: 53321).

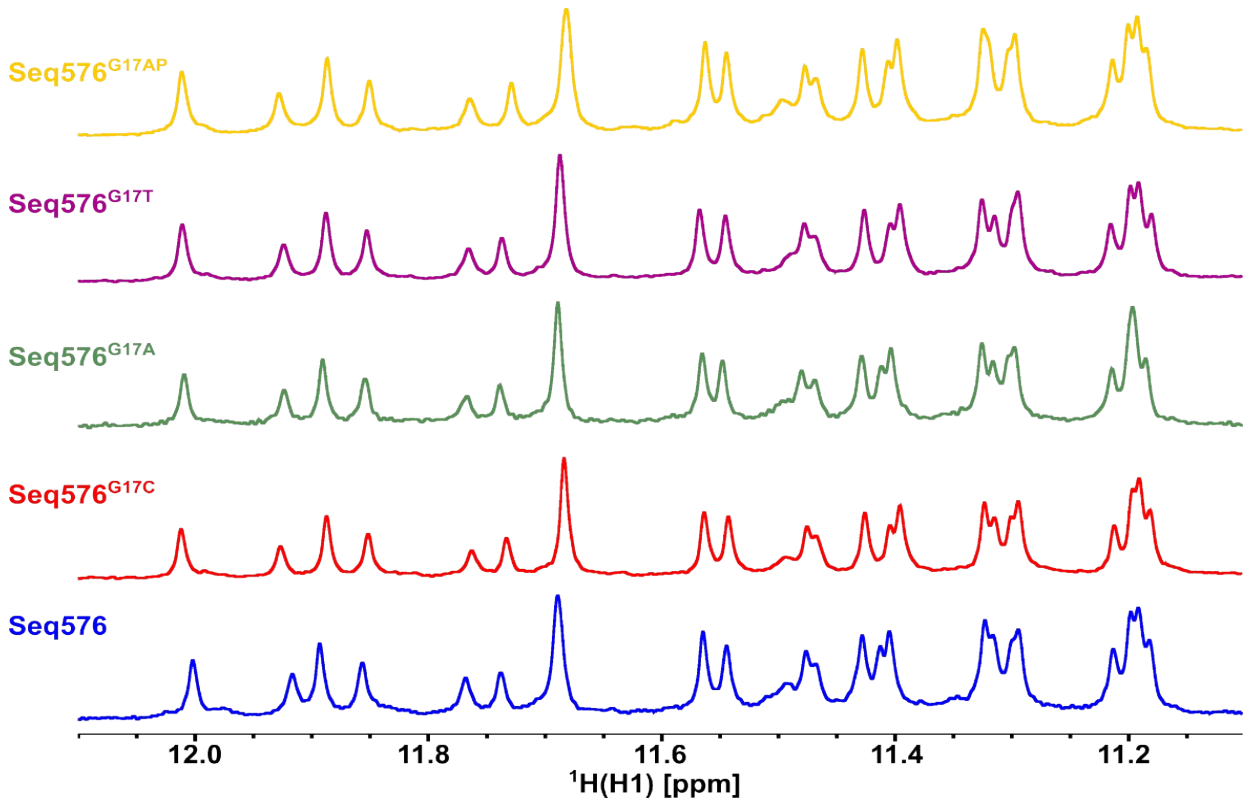

Figure S7. 1D imino proton comparison of Seq576, Seq576^{G17C}, Seq576^{G17A}, Seq576^{G17T} and Seq576^{G17AP}. The blue trace corresponds to the wild-type Seq576, while the red, green, purple, and yellow traces represent the G17C, G17A, G17T, and G17AP mutants of Seq576, respectively.

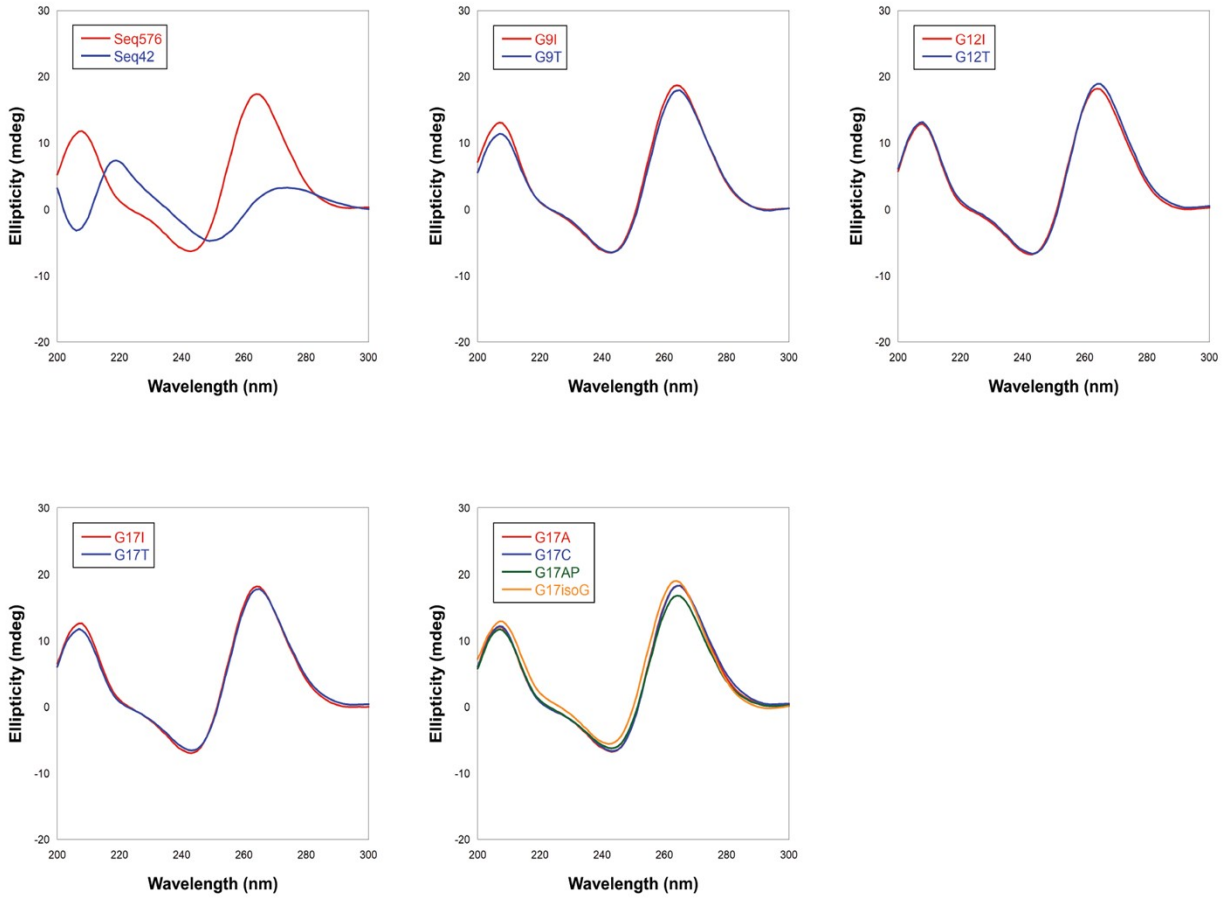

Figure S8. Circular dichroism spectra of wild type 576 and its mutants in 10 mM potassium phosphate buffer (pH 7.5) after annealing. Peaks showing minima at 240 and maxima 260 nm indicate the presence of all parallel G4s. Non-quadruplex Seq 42 is used here as control.

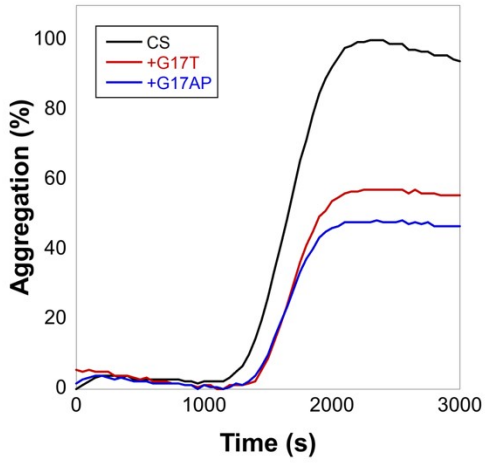

Figure S9. G17T and G17AP show overlapping aggregation rate with T continuing to increase at a similar rate.

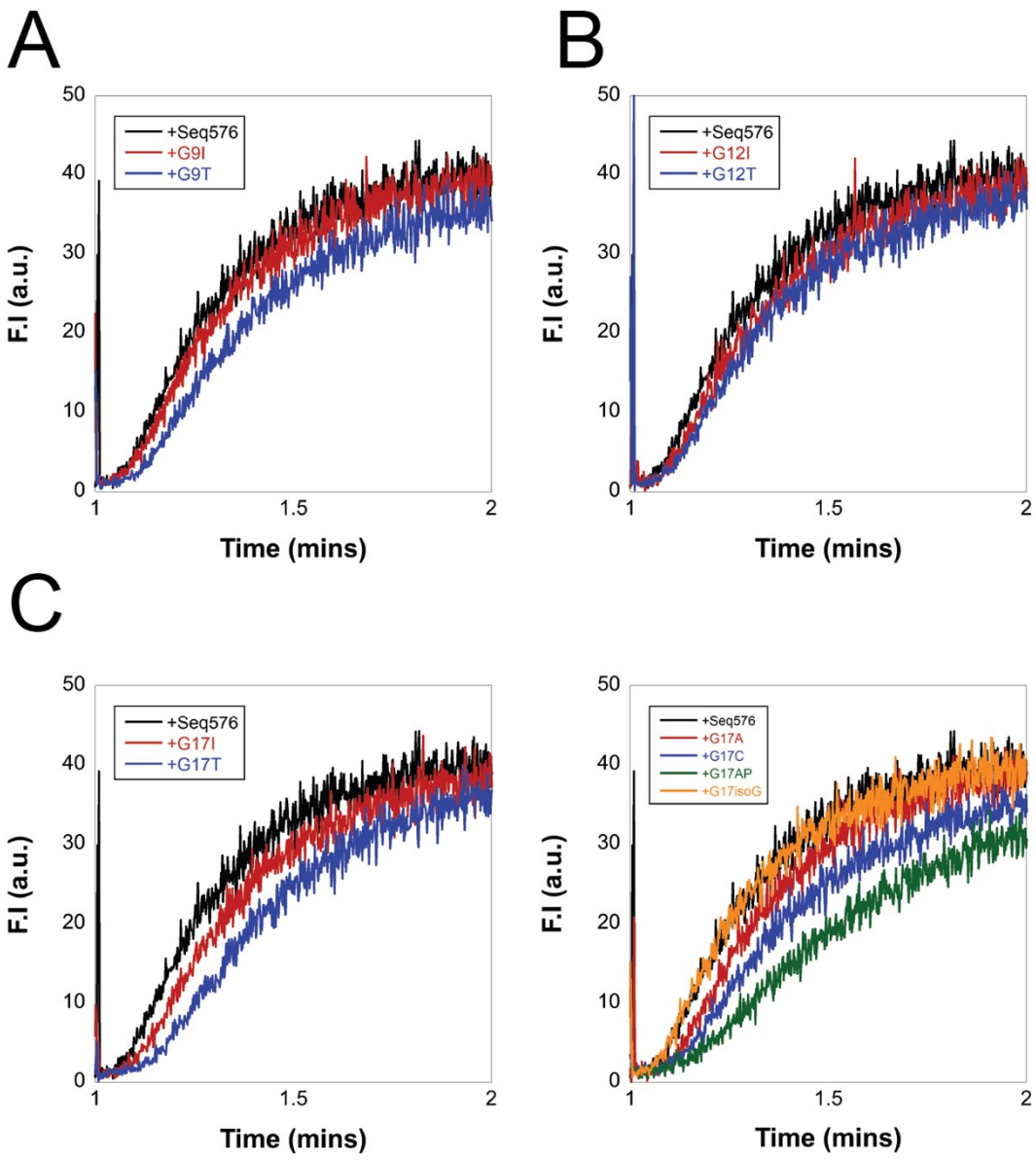

Figure S10. Additional comparisons of refolding kinetics of TagRFP675 in the presence of Seq576 and its mutants shown in Figure 4.

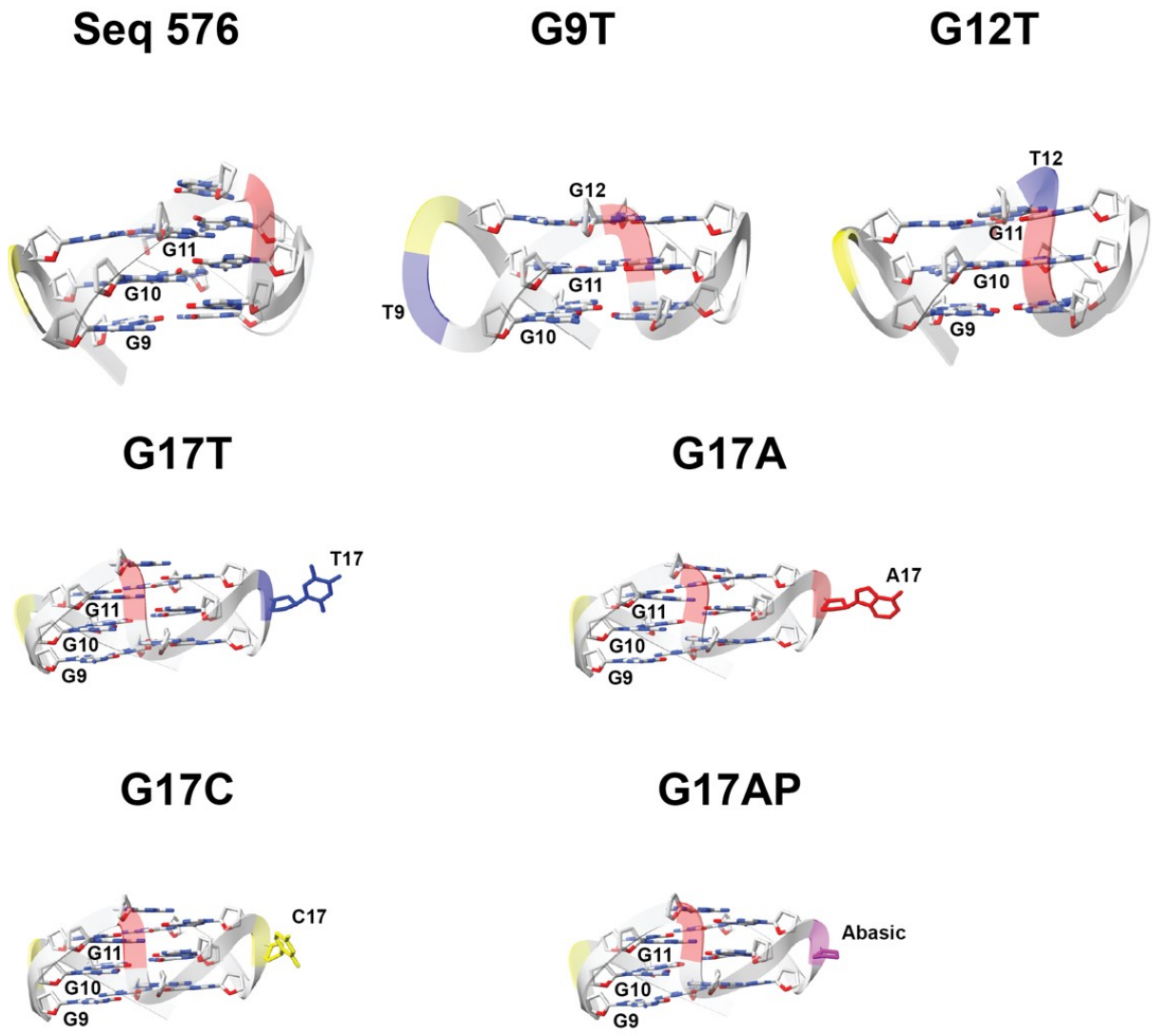

Figure S11. AlphaFold3 structural predictions of Seq576 and its mutants. Bases and the tail (the first 4 bases TGTC) predictions are show more variability than the quadruplex core. Quadruplex cores are not significantly different in all structures at with changes at G17. The tails were cut in all figures for better visualization. Alphafold3 currently does not support inosine and isoguanine modification.

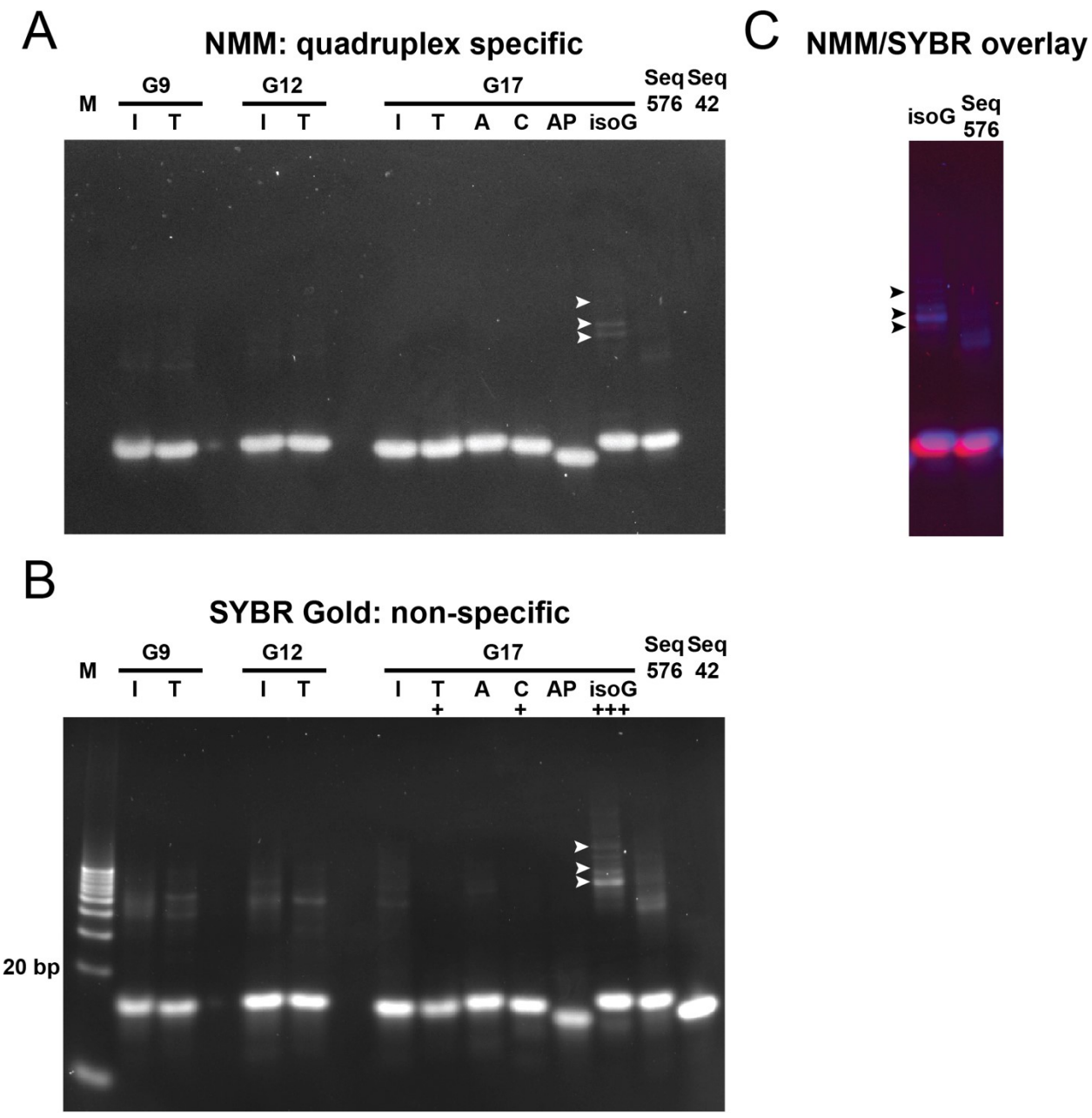

Figure S12. Native PAGE in potassium contained 1xTBE and stained with (A) NMM and (B) SYBR gold. + labeled under G17T and G17C indicate low chaperone activity. And G17isoG with +++ means high chaperone activity. (C) Overlay of both staining (NMM in red and SYBR gold in blue) shows that higher order structure (black arrows) are quadruplexes. Images were merged in ImageJ.

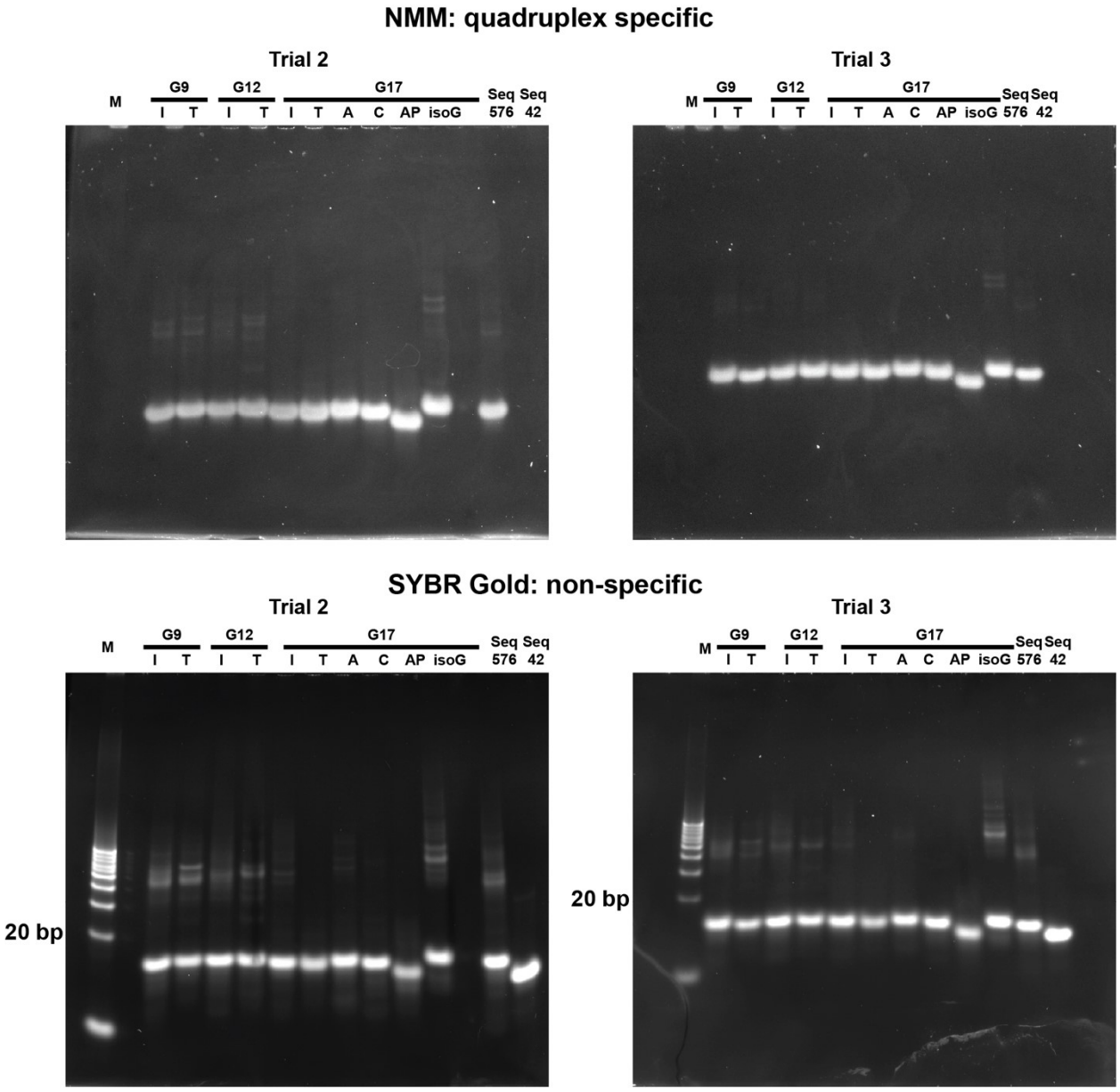

Figure S13. Multiple trials of native PAGE in TBE potassium of Figure 3A-B. stained with NMM (top) and SYBR gold (bottom).

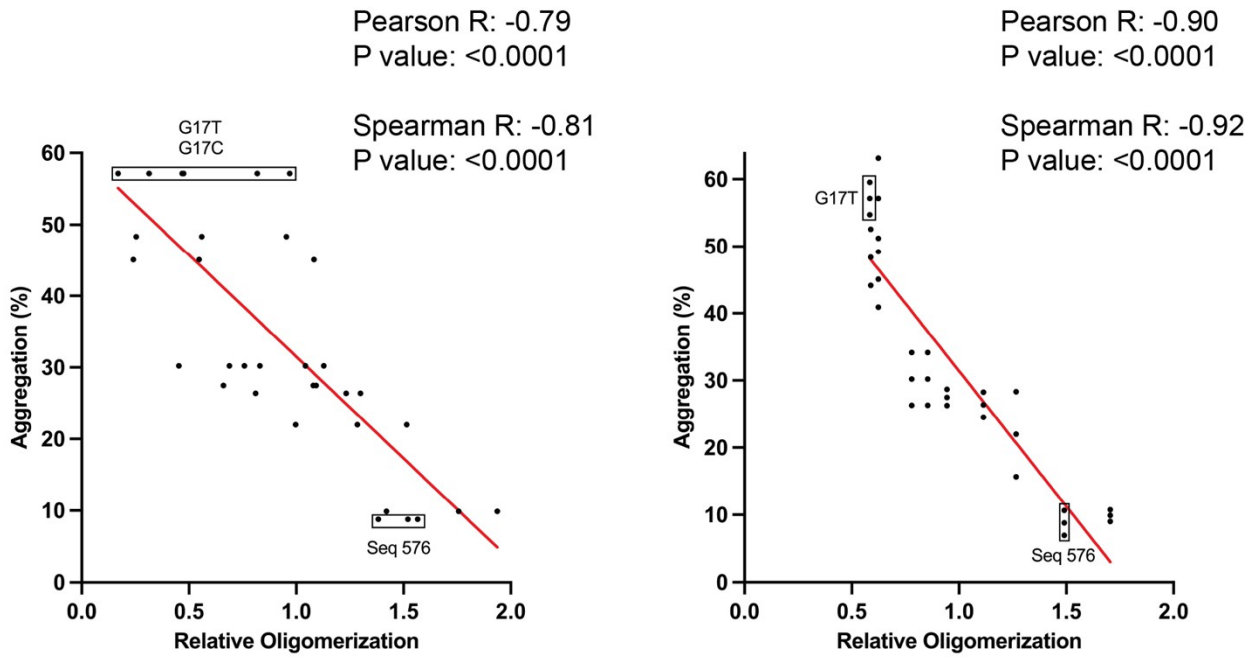

Figure S14. Pearson and Spearman R values comparison for aggregation prevention. Left: the average of aggregation plot against 3 trials of oligomerization. Right: the average of oligomerization plot against 3 trials of aggregation.