OMTN, Volume *23*

Supplemental Information

Expanding the recognition interface of the thrombin-binding aptamer HD1 through modification of residues T3 and T12

Igor Smirnov, Natalia Kolganova, Romualdo Troisi, Filomena Sica, and Edward Timofeev

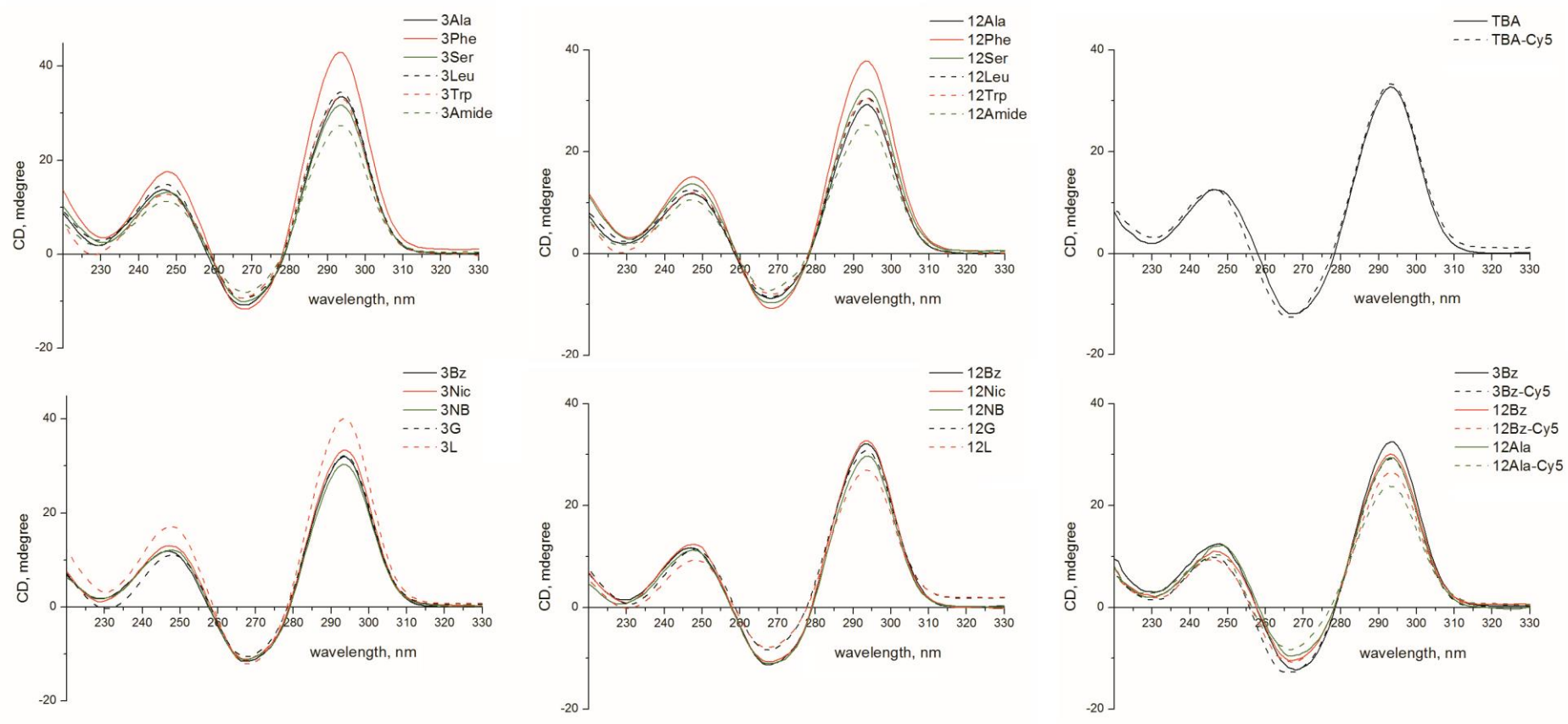


Figure S1. Circular dichroism spectra of modified aptamer variants and selected Cy5-labelled variants at 20 °C in 10 mM sodium cacodylate (pH 7.2) and 100 mM KCl at an aptamer concentration of 5 μM.

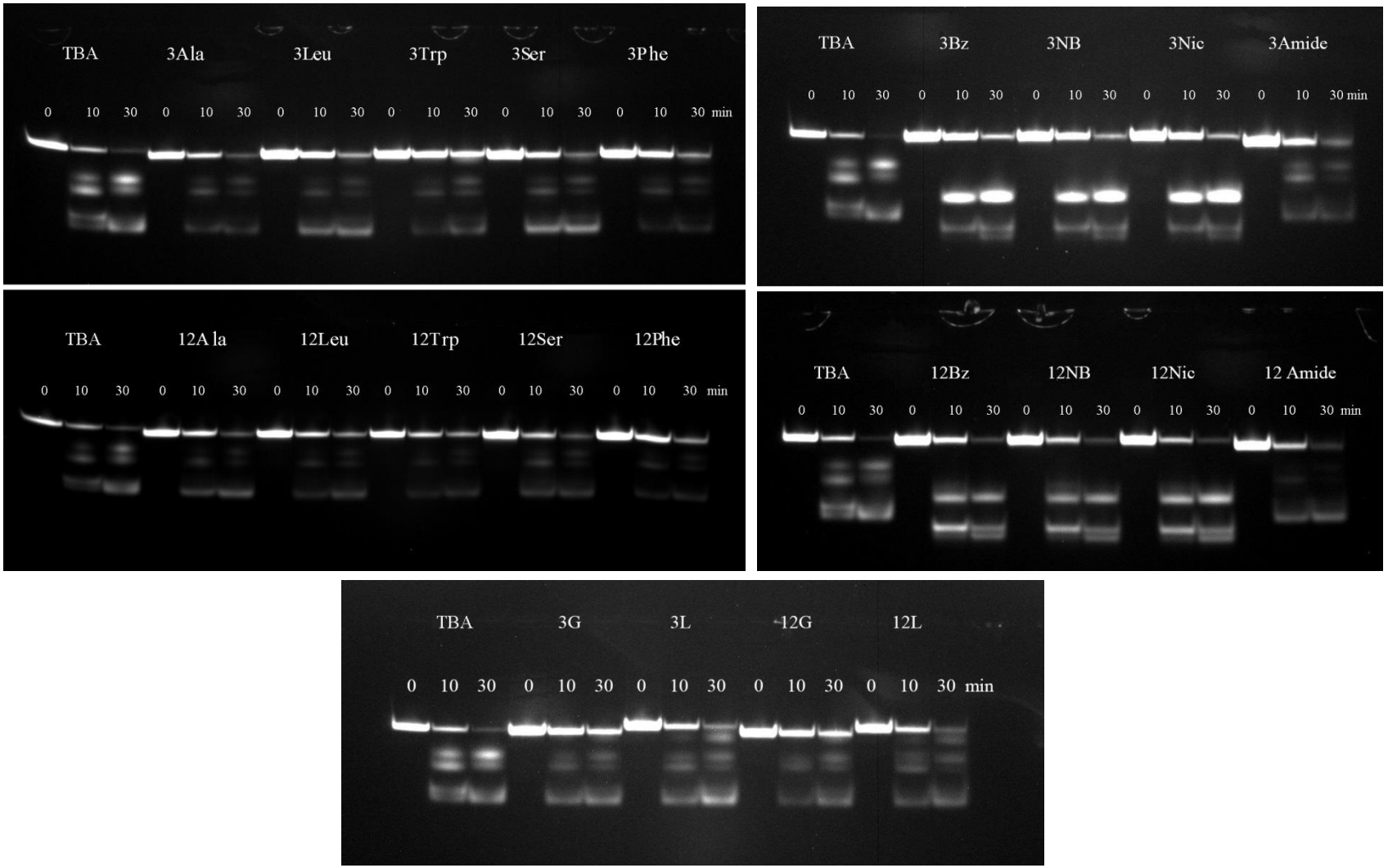


Figure S2. Cleavage of Cy5-labelled TBA (HD1) and modified aptamers by S1 nuclease for 10 and 30 min at 25 °C. The oligonucleotide concentration was 1 μM. Samples were annealed in 1× reaction buffer (Thermo Scientific) before cleavage. Analysis was performed in 20% polyacrylamide gel (19:1) in 1× TBE containing 7 M urea.

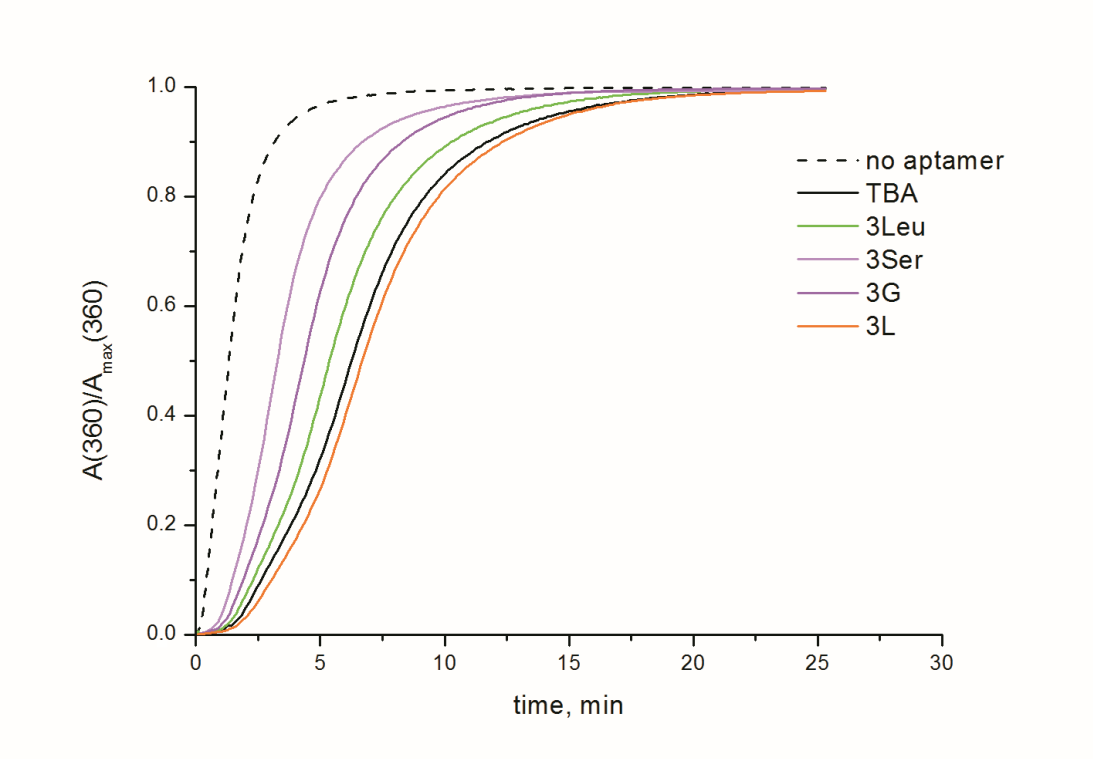


Figure S3. Normalized fibrinogen clotting curves in PBS at 25 °C in the presence of selected aptamers.

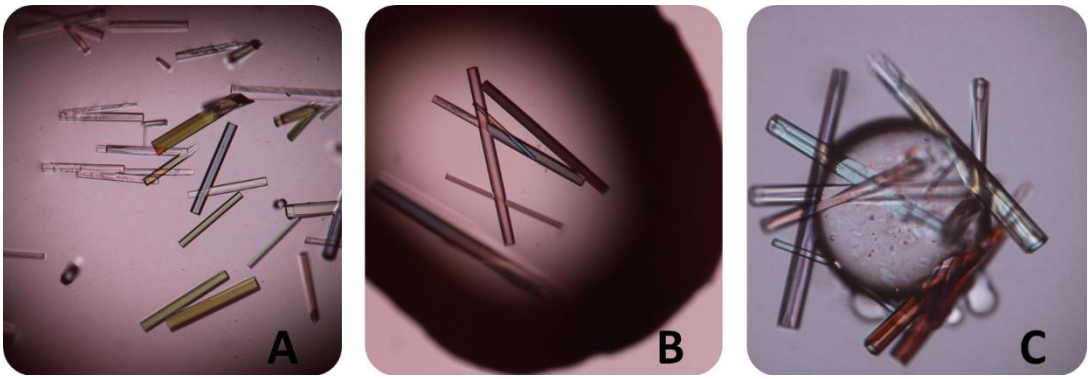


Figure S4. Crystals of the thrombin-3L (A) and thrombin-3G (B) complexes grown in 200 mM KCl, 35% (v/v) pentaerythritol propoxylate and 50 mM HEPES, pH 7.5. Crystals of the thrombin-3Leu (C) complex grown in 18% (v/v) 2-propanol, 18% (w/v) polyethylene glycol 4000 and 100 mM tri-sodium citrate, pH 5.6. The larger crystals are approximately 0.30 mm × 0.05 mm × 0.05 mm.

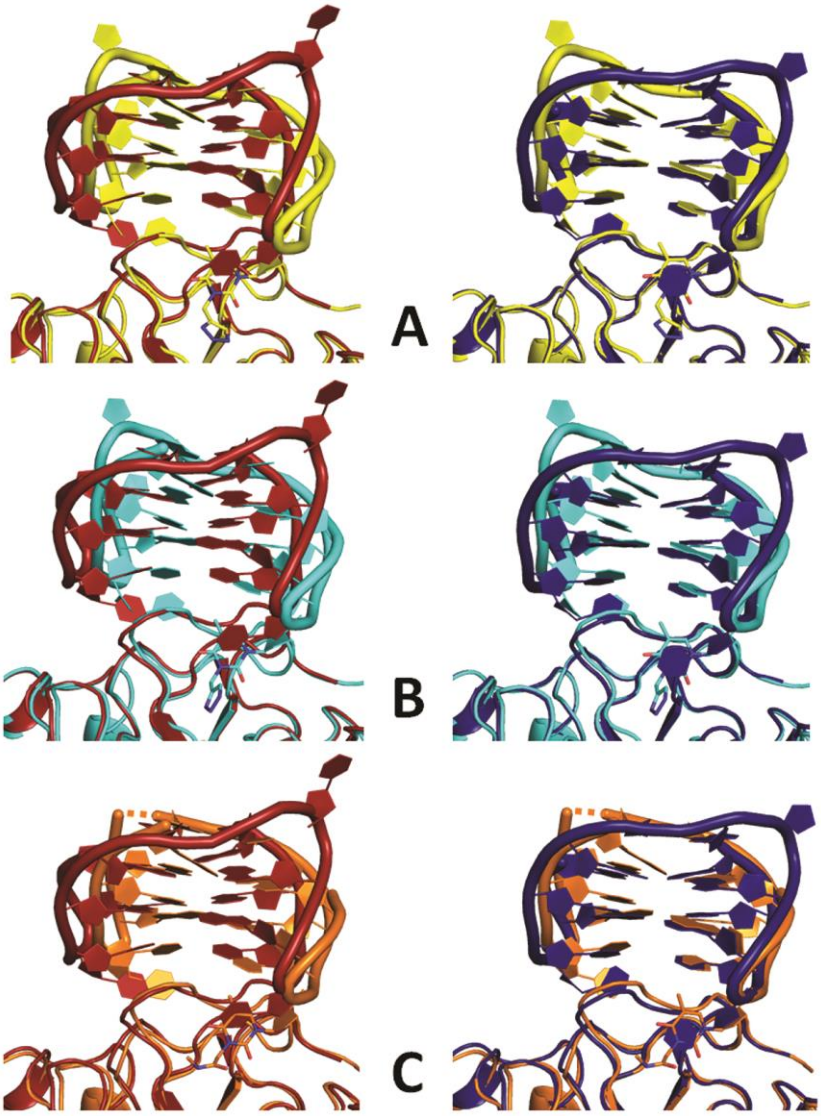


Figure S5. Comparison between the crystal structures of thrombin-TBA-K (dark red, PDB code: 4DII) and thrombin-TBA-Na (blue, PDB code: 4DIH) with those of thrombin-3L (yellow) (A), thrombin-3G (cyan) (B), and thrombin-3Leu (orange) (C), after the superposition of the protein.

Table S1. MALDI MS data for modified TBA variants.

Aptamer variant	$[MH]^+_{calc}$	$[MH]^+_{obs}$ 3T series	$[MH]^+_{obs}$ 12T series
Bz	4931.30	4931.1	4930.2
Nic	4932.29	4931.2	4931.9
NB	4976.30	4975.9	4975.7
Amide	4784.12	4781.7	4783.0
Ala	4855.20	4853.9	4854.3
Leu	4897.28	4895.6	4896.5
Phe	4931.30	4929.3	4925.8
Trp	4970.33	4968.7	4970.4
Ser	4871.20	4871.0	4868.9
G	4970.29	4969.5	4969.0
L	5132.43	5131.2	5130.3

Table S2. Crystallographic statistics for the three thrombin-DNA complexes. Values in brackets refer to the highest resolution shell.

	Thrombin-3L	**Thrombin-3G**	**Thrombin-3Leu**
Crystal data			
Space group	$P3_221$	$P3_221$	$P3_221$
Unit-cell parameters			
a, b, c (Å)	94.92, 94.92, 125.40	94.80, 94.80, 125.33	94.67,94.67, 124.69
α, β, γ (deg)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
V_M ($Å^3$ Da^{-1})	3.90	3.90	3.88
No. of molecules in the asymmetric unit	1	1	1
Solvent content (%)	68.5	68.5	68.3
Data collection			
Resolution limits (Å)	82.20 - 1.58 (1.76 - 1.58)	125.33 - 1.73 (1.94 – 1.73)	81.99 - 2.53 (2.96 - 2.53)
No. of observations	1066050 (28730)	843369 (32671)	130322 (6176)
No. of unique reflections	54291 (2718)	41910 (2097)	6793 (341)
Completeness (%)	95.9 (81.2)	94.4 (72.0)	91.5 (77.8)
<I/σ(I)>	23.2 (2.0)	19.0 (1.9)	10.7 (1.8)
Average multiplicity	19.6 (10.6)	20.1 (15.6)	19.2 (18.1)
$CC_{1/2}$	1.0 (0.8)	1.0 (0.7)	1.0 (0.7)
Refinement			
Resolution limits (Å)	82.20 - 1.58	82.10 - 1.73	81.99 - 2.53
No. of reflections	51616	39863	6123
R_{factor}/R_{free}	0.180/0.206	0.178/0.207	0.245/0.291
No. of atoms	2828	2796	2511
Average B factor ($Å^2$)	27.47	32.95	69.54
RMSD from ideal values			
Bond lengths (Å)	0.015	0.016	0.002
Bond angles (deg)	2.220	2.304	0.764
Ramachandran plot, residues in (%)			
Most favoured region	97.2	97.1	87.0
Additionally allowed region	2.8	2.9	13.0
Generously allowed region	0	0	0
PDB code	6Z8V	6Z8W	6Z8X

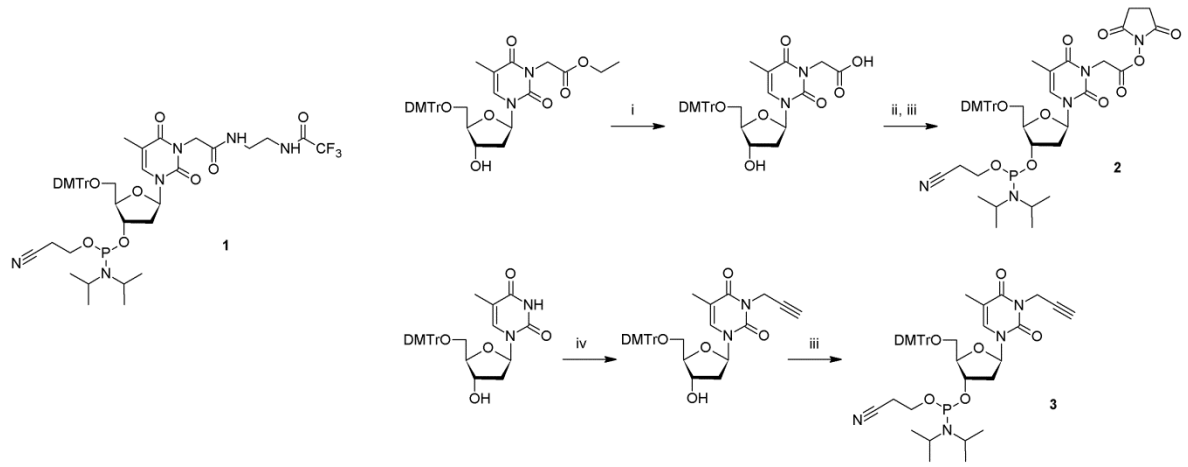
Table S3. RMSD values obtained by the superposition of all atoms of the TBA structures (PDB codes: 4DII and 4DIH) and the TBA3L, TBA3G, and TBA3Leu structures. The nucleotides of the TGT loop were not considered in the analysis because the unmodified and modified TBAs rotate 180° around the G-quadruplex axis.

	TBA-K	**TBA-Na**
TBA-Na	0.67 Å	-
TBA3L	0.80 Å	0.43 Å
TBA3G	0.79 Å	0.44 Å
TBA3Leu	0.69 Å	0.64 Å

Table S4. Interface interactions and areas of the complexes between thrombin and TBA in the presence of potassium ions (PDB code: 4DII) or sodium ions (PDB code: 4DIH), TBA3L, TBA3G, and TBA3Leu, determined by the Contact, CoCoMaps and PISA programs. The threshold distances to select interacting residues were 3.9 Å and 5.0 Å in Contact and CoCoMaps, respectively.

	TBA-K	TBA-Na	TBA3L	TBA3G	TBA3Leu
Contact					
Thy A (in A-region)	Ile24 His71 Ile79 Tyr117	Ile24 Arg75 Glu77 Ile79 Tyr117	Ile24 Arg75 Glu77 Ile79 Tyr117	Ile24 Arg75 Glu77 Ile79 Tyr117	Ile24 His71 Glu77 Ile79 Tyr117
Thy B (in A-region)	Arg75 Arg77A Asn78 Ile79	Arg75 Arg77A Asn78 Ile79	Arg75 Arg77A Asn78 Ile79	Arg75 Arg77A Asn78 Ile79	Arg75 Arg77A Asn78 Ile79
Thy C (in B-region)	Tyr76	Tyr76 Ile82	Tyr76 Ile82	Tyr76 Ile82	Tyr76 Ile82
Thy D (in B-region)	Thr74 Arg75 Tyr76 Arg77A	 Arg75 Tyr76 Arg77A	Thr74 Arg75 Tyr76 Arg77A	Thr74 Arg75 Tyr76 Arg77A	 Arg75 Tyr76 Arg77A
Gua (G-quadruplex)	 Arg75 Arg77A Asn78	Thr74 Arg75 Arg77A	Thr74 Arg75 Arg77A Asn78	Thr74 Arg75 Arg77A Asn78	Thr74 Arg75 Arg77A Asn78
CoCoMaps					
Interface area (Å^2)	540	564	607	610	595
Polar interface area (Å^2)	269	288	329	319	328
Non polar interface area (Å^2)	271	276	278	291	267
No. of interacting residues thrombin	10	11	13	14	12
No. of interacting residues oligonucleotide	8	8	8	8	8
No. of hydrophilic-hydrophobic interaction	3	5	5	5	4
No. of hydrophilic-hydrophilic interaction	18	20	22	23	21
No. of hydrophobic-hydrophobic interaction	0	0	0	0	0
PISA					
Interface area (Å^2)	543	565	606	606	592
No. of H-bonds	11	12	14	13	9

Synthesis of modified phosphoramidites and oligonucleotides



Scheme S1. Chemical structures and synthesis of the phosphoramidite precursors. Reagents and conditions: (i) KOH in aqueous ethanol; (ii) N-hydroxysuccinimide, N,N'-dicyclohexylcarbodiimide, tetrahydrofuran; (iii) 2-cyanoethyl N,N,N′,N′-tetraisopropylphosphorodiamidite, tetrazole, acetonitrile; (iv) allyl bromide, 1,8-diazabicyclo[5,4,0]undec-7-ene, acetonitrile.

The synthesis of N3-substituted thymidine derivatives was performed through highly selective N-alkylation of 5'-O-dimethoxytritylthymidine with ethyl bromoacetate or propargyl bromide in the presence of a strong organic base. A detailed synthetic procedure for compound **1** has been reported earlier.[1] The preparation of phosphoramidites **2** and **3** is outlined in Scheme S1. Thymidine derivatives **1** and **3** were utilized to incorporate amines or propargyl groups into the oligonucleotides for post-synthetic coupling with activated aromatic carboxylic acids or click reactions with carbohydrate azides, respectively. To prepare a panel of amino acid TBA mutants, we used phosphoramidite **2** containing an N-hydroxysuccinimide ester of the N3-carboxymethyl group. The activated carboxyl group was further transformed into a dipeptide residue by coupling with the methyl esters of the respective L-amino acids, i.e., alanine, phenylalanine, serine, leucine, or tryptophan. TBA variants with amino acid side chains were synthesized directly on the column during the course of automated oligonucleotide synthesis. For coupling with NHS-oligonucleotides, we used amino acid ester hydrochlorides in DMF in the presence of DIPEA. Due to the low hydrolytic stability of the activated carboxyl group, the coupling of partially synthesized NHS-oligonucleotides with amino acid derivatives was carried out in the same synthetic cycle before the oxidation step. The efficiency of the reaction varied for different amino acids. As a result, varying amounts of the respective amide by-product (3/12Amide) were always present in a crude oligonucleotide product after deprotection. For the post-synthetic modification of amino-functionalized TBA precursors, we used activated NHS esters of benzoic, p-nitrobenzoic, and nicotinic acids. To generate carbohydrate TBA variants, mono- and disaccharide fragments were attached to N3-propargyl groups through the Cu(I)-catalysed alkyne-azide cycloaddition of 1-β-D-glycopyranosyl azide or 1-β-D-lactopyranosyl azide.

To estimate the binding affinities of non-natural TBA variants, we used fluorescent versions of the modified aptamers with a Cy5-tag at the 3' end. To avoid interference between the post-synthetic modification and the labelling procedure, the Cy5 dye was attached to 3/12Bz, 3/12-NB, or 3/12-Nic mutants by the click protocol. In contrast, for labelling amino acid and carbohydrate series aptamers, we used the NHS-activated dye. Respective modified oligonucleotides with either a 3'-alkyne or 3'-amino group were additionally prepared following the above strategy.

Experimental

Chemical reagents and solvents were purchased from various commercial suppliers and used without further purification. Thin layer chromatography (TLC) was carried out using Kieselgel 260 F aluminum sheets (Merck) and visualized by UV absorption or stained with 2% H_2SO_4 in EtOH with subsequent heating. Column chromatography was performed using silica gel (particle size 0.06-0.2 mm, Sigma-Aldrich). Nuclear magnetic resonance spectra were performed on a Bruker AMX400 spectrometer. Standard reagents for automated oligonucleotide synthesis were purchased from Glen Research. Reactive Cy5 derivatives were purchased from Lumiprobe.

5'-O-(4,4'-dimethoxytrityl)- N3-carboxymethylthymidine.

5'-O-(4,4'-Dimethoxytrityl)-N3-ethoxycarbonylmethylthymidine[1] (1.0 g, 1.6 mmol) was dissolved in 5 mL of a 2 M solution of KOH in aqueous (~50%) ethanol. Hydrolysis was monitored by TLC (DCM/hexane/EtOH/conc.aq. NH_3 9:10:1:0.1). The solution was diluted with ethanol (20 mL), and excess KOH was neutralized by Dowex 50 in pyridinium form. The resin was filtered away and washed twice with ethanol. The solution was evaporated to dryness, and the residue was purified by column chromatography on silica gel (gradient elution from DCM/Et_3N 95:5 to DCM/EtOH/Et_3N 7:2:1) to yield 0.81 g (73%) of the triethylammonium salt of 5'-O-(4,4'-dimethoxytrityl)-N3-carboxymethylthymidine. ^{1}H NMR (400 MHz, DMSO-d6): δ 7.58 and 7.59 (2s, 1H, H-6), 7.21-7.41 and 6.89-6.92 (m, 13H, aromatic DMTr), 6.26 (t, J = 6.8 Hz, 1H, H-1′), 5.33 (s, 1H, OH-3′), 4.33-4.37 (m, 1H, H-3′), 4.19-4.27 (m, 2H, CH_2 at N3), 3.90-3.92 (m, 1H, H-4′), 3.74 (s, 6H, OCH_3), 3.18-3.27 (m, 2H, H-5′), 2.16-2.29 (m, 2H, H-2′), 1.46 (2s, 3H, CH3-5). ^{13}C NMR (100 MHz, DMSO-d6): δ 169.7, 162.2, 158.1, 150.2, 144.6, 135.4, 135.2, 129.7, 127.9, 127.6, 126.7, 113.2, 108.7, 85.9, 85.5, 84.5, 70.4, 63.7, 55.0, 52.0, 45.1, 12.3, 9.5, 7.1; HRMS calcd for $C_{33}H_{34}N_2O_9$ [$(M+Na)^+$]: 625.2157, found: 625.2160.

***5'-O-(4,4'-dimethoxytrityl)-N3-{2-[(2,5-dioxopyrrolidin-1-yl)oxy]-2-oxoethyl}thymidine-3'-O-[(2-cyanoethyl)-N,N-diisopropylphosphoramidite]* (2).**

5'-O-(4,4'-Dimethoxytrityl)-N3-carboxymethylthymidine (0.75 g 1.1 mmol) was treated with N-hydroxysuccinimide (0.13 g, 1.1 mmol) and N,N'-dicyclohexylcarbodiimide (0.24 g, 1.7 mmol) in 3 mL of dry THF under argon at 25 °C for 16 h. N,N'-Dicyclohexylurea was removed by filtration, and the solution was evaporated to generate a white foam, which was dried over P_2O_5 in a desiccator under reduced pressure for 16 h. The foam was dissolved in 2.5 mL of anhydrous MeCN and filtered under argon. Then, tetrazole (75 mg, 1.1 mmol) and 3-5 beads of 4 Å molecular sieves (4-8 mesh) were added to the solution and left under argon for 30 min. 2-Cyanoethyl N,N,N′,N′-tetraisopropylphosphorodiamidite (0.32 mL, 1.1 mmol) was added to the reaction mixture under intensive stirring. The reaction was allowed to proceed for 30 min at 25 °C with monitoring by TLC (DCM/hexane/EtOH 9:20:1). Then, the solution was diluted with ethyl acetate (150 mL) and washed twice with cold water (150 mL). The organic layer was dried over anhydrous Na_2SO_4 and evaporated to generate a white foam. The final crude product was dried over P_2O_5 in a desiccator under reduced pressure for 16 h and then stored under argon at -20 °C before use. Due to the low hydrolytic stability, compound **2** could not be purified and was used in the automatic oligonucleotide synthesis as a crude product containing approximately 50% (by TLC) of the target phosphoramidite. Yield: 0.77 g (80%). ^{1}H NMR (400 MHz, DMSO-d6, selected signals): δ 7.66-7.68 (m, 1H, H-6), 7.19-7.41 and 6.85-6.92 (m, 13H, aromatic DMTr), 6.21-6.32 (m, 1H, H-1′), 4.97 (br. s, 2H, CH_2 at N3), 3.71-3.74 (m, 6H, OCH_3), 2.82 (br. s, 4H, NHS), 1.51-1.57 (m, 3H, CH_3-5), 0.99-1.25 (m, 12H, CH_3 from i-Pr); ^{13}C NMR (100 MHz, DMSO-d6, selected signals): δ 169.6, 158.2, 129.7, 128.9, 127.6, 127.3, 113.2, 112.7, 54.9, 54.7, 34.4, 33.3, 25.4, 24.9, 24.4, 23.9; ^{31}P MNR (162 MHz, DMSO-d6, selected signals): δ 150.5, 150.1; HRMS calcd for $C_{46}H_{54}N_5O_{12}P$ [$(M+Na)^+$]: 922.3399, found: 922.3392.

5'-O-(4,4'-dimethoxytrityl)-N3-propargylthymidine

DBU (1.3 mL, 8.3 mmol) and propargyl bromide (80% soln. in toluene, 0.92 mL, 8.3 mmol) were added to a stirred solution of 5'-O-(4,4'-dimethoxytrityl)-thymidine (1.5 g, 2.8 mmol) in 5 mL of DCM. The reaction was monitored by TLC (DCM/hexane/EtOH 9:10:1). After 30 min, the reaction mixture was diluted with ethyl acetate (200 mL) and washed with saturated aqueous $NaHCO_3$ (150 mL). The organic layer was dried over anhydrous Na_2SO_4 and evaporated. The product was purified by column chromatography on silica gel (gradient

elution from DCM/hexane 1:1 to DCM/EtOH 9:1) to yield 1.41 g (88%) of the title compound. ^{1}H NMR (400 MHz, DMSO-d6): δ 7.60 (2s, 1H, H-6), 7.21-7.40 and 6.88-6.90 (m, 13H, aromatic DMTr), 6.26 (t, J = 6.8 Hz, 1H, H-1′), 5.33 (d, J = 4.6 Hz, 1H, OH-3′), 4.54 (t, J 2.1 Hz, 2H, CH_2 at N3), 4.32-4.36 (m, 1H, H-3′), 3.90-3.93 (m, 1H, H-4′), 3.73 (s, 6H, OCH_3), 3.18-3.26 (m, 2H, H-5′), 3.10 (t, J 2.5 Hz, 1H, CCH), 2.18-2.31 (m, 2H, H-2′), 1.50 (s, 3H, CH_3-5); ^{13}C NMR (100 MHz, DMSO-d6): δ 161.6, 158.1, 149.6, 144.6, 135.4, 135.2, 134.7, 129.7, 127.8, 127.6, 126.7, 113.2, 108.7, 85.8, 85.6, 84.8, 79.0, 72.9, 70.3, 63.6, 55.0, 30.0 12.2; HRMS calcd for $C_{34}H_{34}N_2O_7$ [$(M+Na)^+$]: 605.2258, found: 605.2263.

***5'-O-(4,4'-dimethoxytrityl)-N3-propargylthymidine-3'-O-[(2-cyanoethyl)-N,N-diisopropyl phosphoramidite]* (3).**

5'-O-(4,4'-Dimethoxytrityl)-N3-propargylthymidine (1.3 g, 2.2 mmol), pyridine (0.18 mL, 2.2 mmol), and tetrazole (160 mg, 2.2 mmol) were dissolved in 5 mL of anhydrous MeCN, and 3-5 beads of 4 Å molecular sieves (4-8 mesh) were added to the solution. The mixture was left under argon for 30 min. Then, 2-cyanoethyl N,N,N′,N′-tetraisopropylphosphorodiamidite (0.74 mL, 2.4 mmol) was added under intensive stirring to the solution. The reaction was allowed to proceed for 30 min at room temperature with monitoring by TLC (DCM/hexane/EtOH 9:20:1). Then, the solution was diluted with ethyl acetate (150 mL) and washed twice with cold saturated aqueous $NaHCO_3$ (150 mL) and cold water (150 mL). The organic layer was dried over anhydrous Na_2SO_4 and evaporated to generate a white foam. Yield: 1.36 g (78%). ^{1}H NMR (400 MHz, DMSO-d6): δ 7.61-7.62 (m, 1H, H-6), 7.23-7.41 and 6.87-6.91 (m, 13H, aromatic DMTr), 6.24-6.29 (m, 1H, H-1′), 4.54-4.59 (m, 2H, CH_2 at N3, H-3′), 4.03-4.10 (m, 1H, H-4′), 3.73 and 3.74 (2s, 6H, OCH_3), 3.49-3.80 (m, 4H, CH from i-Pr, $POCH_2$), 3.26-3.30 (m, 2H, H-5′), 3.09 (t, J 2.4 Hz, 1H, CCH), 2.64 and 2.77 (2t, J 5.9 and 6.0, 2H, CH_2CN), 2.32-2.46 (m, 2H, H-2′), 1.53-1.56 (2s, 3H, CH_3-5), 0.99-1.15 (m, 12H, CH_3 from i-Pr); ^{13}C NMR (100 MHz, DMSO-d6): δ 161.6, 158.1, 149.6, 144.5, 135.2, 134.5, 129.6, 127.8, 127.6, 126.8, 113.2, 85.9, 84.8, 78.9, 72.9, 58.3, 58.1, 55.0, 42.6, 42.5, 30.0, 24.3, 24.2, 24.1, 19.7, 12.2; ^{31}P MNR (162 MHz, DMSO-d6): δ 150.5, 150.1; HRMS calcd for $C_{43}H_{51}N_4O_8P$ [$(M+Na)^+$]: 805.3337, found: 805.3315; calcd for $C_{43}H_{51}N_4O_8P$ [$(M+K)^+$]: 821.3076, found: 821.3059.

Synthesis of modified oligonucleotides

DNA oligomers were synthesized using an ABI 3400 DNA/RNA synthesizer. Modifications were introduced at positions 3 or 12 by using N3-modified phosphoramidites. The coupling time was increased to 120 s for the modified units. Amino or propargyl groups at the 3' end of the oligonucleotides were generated by using the 3'-amino-modifier C7 CPG (Glen Research) or alkyne-CPG (Lumiprobe). For a manual on-column coupling of the partially synthesized carboxy-NHS oligonucleotides with amino acid methyl esters, the column was washed with MeCN before the oxidation step, and the respective reaction solution was added to the column via syringe. The solution was prepared by dissolving 100 μmol of amino acid ester hydrochloride in 1 mL of MeCN/DMF (9:1), adding 50 μL DIPEA, and filtering to remove the insoluble components. On-column coupling was allowed to proceed for 10 min. Then, the column was washed with MeCN, and the automated synthesis was resumed. Deprotection of the modified oligomers was carried out with concentrated ammonia for 5 h at 55 °C. Partially deprotected oligonucleotides with amino acid residues, N3-propargyl, or N3-amino groups were purified by reversed-phase HPLC (Hypersil ODS, 5 μm, 4.6 × 250 mm; 10–50% MeCN in 50 mM TEAA for 30 min). After removal of the 5' DMTr group, oligomers were repeatedly purified by reversed-phase HPLC (0-25% MeCN in 50 mM TEAA for 30 min) with the control of fractions by MALDI mass spectrometry and denaturing gel electrophoresis. Cycloaddition of the carbohydrate residue to propargyl-modified oligonucleotides was carried out using 1-β-D-glycopyranosyl azide or 1-β-D-lactopyranosyl azide (Sigma-Aldrich) and a standard click protocol in the presence of TBTA (www.lumiprobe.com). To prepare benzoyl-, nitrobenzoyl-, and nicotinoylamide aptamer derivatives, an amino-oligonucleotide was typically dissolved in 0.2 M aqueous $NaHCO_3$ (50 μL) and mixed with a 0.2 M solution of NHS ester of the respective acid in DMF (100 μL) at 0 °C. The reaction mixture was maintained for 16 h at 4 °C and then precipitated with 2% $LiClO_4$ in acetone. Final purification of the modified aptamers was performed by reversed-phase HPLC with control by MALDI mass spectrometry (Table S1). A few samples were additionally purified by preparative gel electrophoresis in denaturing polyacrylamide gel (19:1, 1×TBE, pH 8.4, 7 M urea). Labelling of the 3' propargyl- or 3'-amino-

oligonucleotides with the azide or NHS ester of Cy5 dye was carried out similarly to the above procedures. All Cy5-labelled aptamers were purified by preparative electrophoresis in a denaturing polyacrylamide gel.

References

1. I. P. Smirnov, N. A. Kolganova, V. A. Vasiliskov, A. V. Chudinov and E.N. Timofeev, Mass-spectrometry analysis of modifications at DNA termini induced by DNA polymerases, 2017, *Sci. Rep.*, **7**(1), 6674.