

In silico post-SELEX screening and experimental characterizations for acquisition of high affinity DNA aptamers against carcinoembryonic antigen

Qiong-Lin Wang, Hui-Fang Cui*, Jiang-Feng Du, Qi-Yan Lv, Xiaojie Song

Department of Bioengineering, School of Life Sciences, Zhengzhou University, 100# Science Avenue, Zhengzhou, 450001, PR China

1. BLI assay principle

In the BLI analysis system, interaction of analytes with the ligands immobilized on a sensor surface forms a monomolecular layer that in turn creates a shift in the interference spectrum of reflected light (Kumaraswamy et al., 2015). The wavelength shift directly reflects the change in the optical thickness of the sensor layer. Any change in the number of molecules bound to the biosensor causes a shift in the interference pattern that can be measured in real-time.

2. CEA crystal structure

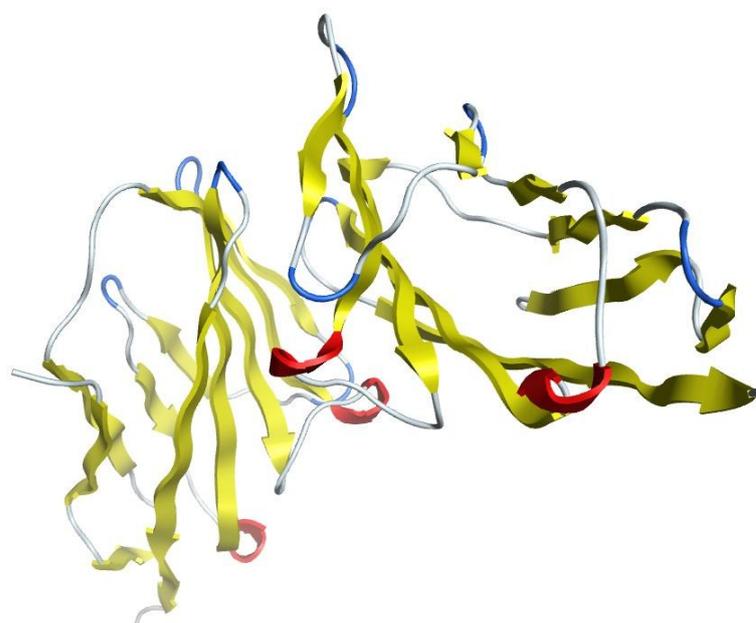
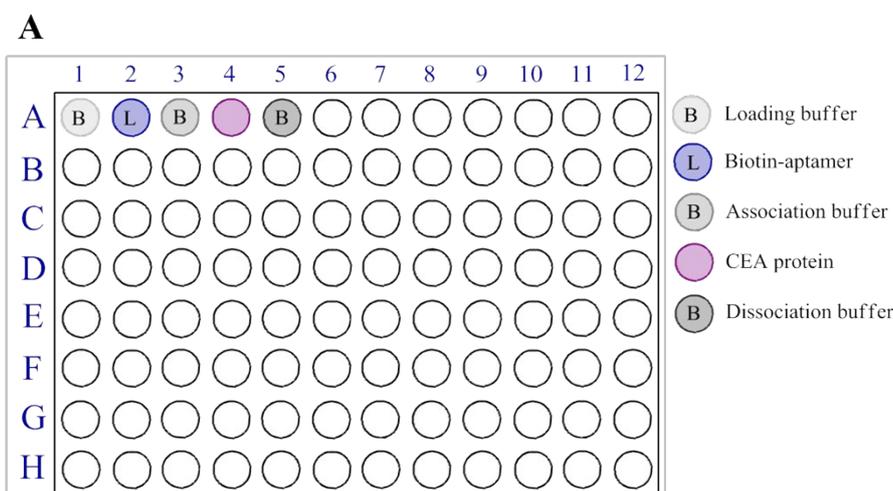


Fig. S1 CEA crystal structure (code: 2QSQ) with a resolution of 1.95 Å.

3. Preliminary experiment for verification of the BLI assay conditions

Before the BLI assay for determination of the affinity of DNA aptamers to CEA, a preliminary experiment was carried out to verify the feasibility of the set BLI assay conditions. The designed plate map was schematically shown in Fig. S2A. Firstly, a SSA sensor tip was pre-wetted in 200 μ L loading buffer (column 1) for 10 min followed by 'baseline 1' step with loading buffer for 300 s. Afterwards, the SSA sensor tip was loaded and immobilized with 200 nM biotinylated parent aptamer in loading buffer (column 2) for 300 s through streptavidin-biotin binding in the 'loading' step, followed by 'baseline 2' step for 300 s in association buffer (column 3). Subsequently, association of 22.73 nM CEA in association buffer (column 4) with the biotinylated aptamers was conducted in the 'association' step for 600 s. Finally, the 'dissociation' step was executed with dissociation buffer (column 5) for 1500 s. All steps were performed on a 96-well plate containing 200 μ L of samples or buffers in each well at 30 $^{\circ}$ C with shaking at 1000 rpm. Fig. S2B shows the sensorgram of wavelength shift (i.e. Binding, nm) versus time, at the above preliminarily set BLI conditions. The loading concentration (i.e. 200 nM) of the DNA aptamer was considered to be appropriately set, because at this condition an obvious signal in the CEA association step was generated. The CEA concentration level (22.73 nM) was decided to be in the detection range for the following BLI assay in determination of the aptamer affinities to CEA, since the BLI signal was obvious but seems not too high to be out of the dynamic range. In addition, the running time for baseline 1, aptamer loading, baseline 2, CEA association, and dissociation for the following BLI affinity assay was decided to be 100 s, 200 s, 100 s, 400 s, and 1500 s, respectively, at which steady states could be reached.



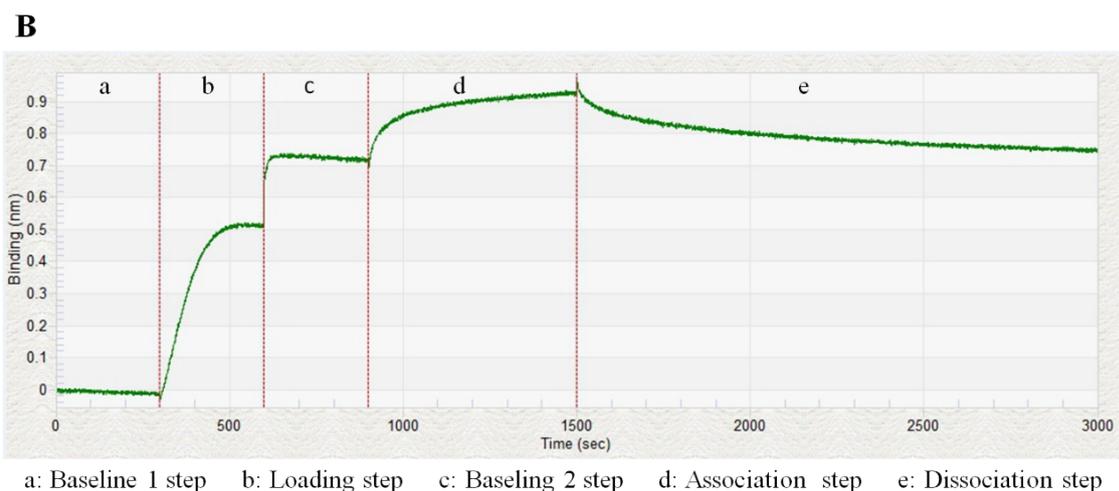


Fig. S2 (A) Plate map diagram for verification of the preliminarily set BLI assay conditions. (B) The sensorgram of wavelength shift (i.e. Binding, nm) versus time, showing all the BLI assay steps with loading of 200 nM parent aptamer and association of 22.73 nM CEA.

4. Determination of affinities of DNA aptamers to CEA

For the BLI affinity assay, the plate map was designed and is schematically shown in Fig. S3. Prior to each assay, five SSA sensor tips were pre-wetted in 200 μ L loading buffer for at least 10 min followed by 'baseline 1' step with loading buffer (column 1) for 100 s. Afterwards, four of the SSA sensor tips were loaded with 200 nM biotinylated aptamers in loading buffer (column 2) for 200 s in the 'loading' step (aptamer P, P-GTG, P-ATG, and GAC-P for sensor A, B, C, and D, respectively). The 'baseline 2' step in the association buffer (column 3) for 100 s was then followed. Subsequently, association of CEA with the biotinylated aptamers was conducted in a series of different concentrations of CEA (4.55, 9.10, 22.73, 45.45, 91.01, and 181.8 nM for column 4, 5, 6, 7, 8, and 9, respectively) in association buffer for 400 s in the 'association' step. After each association step in a CEA solution, the 'dissociation' step in the dissociation buffer (column 10) for 1500 s was followed. The CEA association was performed from low CEA concentration to high CEA concentration in sequence. After each 'dissociation' step, the 'baseline 2' step was repeated right before the next 'association' step with a higher concentration of CEA solution. A control experiment with the sensor E was conducted for comparisons by processing the sensor tip in the corresponding blank buffer solutions. The response data obtained from the reaction surface were exported from the Octet Data Analysis Software 8.2.0.7. The CEA-induced response data in Fig. 3

were obtained by subtracting the signal simultaneously acquired from the control experiment to eliminate buffer-induced interferometry spectrum shift, and then normalized by subtracting the signal acquired from the reference surface (i.e. the starting-point signal of the ‘association’ step).

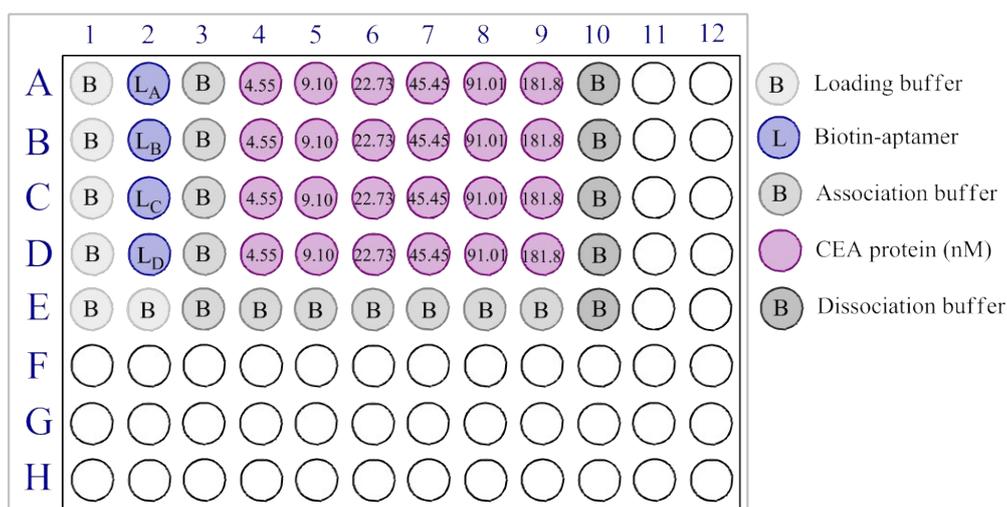


Fig. S3 Plate map diagram for aptamer-CEA BLI assay. L_A, L_B, L_C, and L_D represents P, P-GTG, P-ATG, and GAC-P, respectively).

5. Fabrication of the reported electrochemical CEA aptasensor

For aptasensor fabrication, gold electrodes were polished with slurries of 1.0, 0.3 and 0.05 μm alumina successively, and then rinsed with water. The electrodes were then washed in ultrasonic bath of ethanol, water, a mixture of 25% H_2O_2 and 50 mM KOH, and finally water successively. Immediately after the washed electrodes were blow-dried with N_2 stream, the aptamer was immobilized onto the electrode surface through formation of Au-S bond, by applying a steady potential of +0.4 V at the electrode in 1 mL aptamer solution for 500 s. Prior to the aptamer immobilization, 12 μL of 10 mM TCEP solution was added to 10 μL 20 μM aptamer solution in TE buffer (pH 8.0, containing 10 mM Tris-HCl and 1 mM EDTA), and the mixture was kept at room temperature for 1 h to reduce disulfide bond. After the aptamer immobilized electrode was cleaned with the immobilization buffer, the electrode was passivated by being submerged in 1 mM 6-mercapto-1-hexanol (MCH) at 30 $^\circ\text{C}$ for 1 h. After electrode cleaning, 100 μL standard CEA samples in protein-dissolving buffer were added onto the passivated electrode surface, and the electrode was then kept at 37 $^\circ\text{C}$ for 1 h under gentle shaking. After being rinsed thoroughly with a washing buffer, the electrode surface was covered with 100 μL 0.5 mg mL^{-1} ConA solution in the

ConA activating buffer. The electrodes were sealed and kept at 30 °C for 3 h under gentle shaking. Finally, after being washed with the washing buffer, the electrode surface was dropped with 100 μ L 5 μ g mL⁻¹ HRP solution, and then incubated at 30 °C for 2 h. The resulted sandwich aptasensors were washed with the washing buffer and then immersed into a detection buffer (pH 7.0, 0.1 M PBS containing 0.1 M KCl), in the presence of 2.0 mM H₂Q and 2.5 mM H₂O₂ at room temperature. After 4 min of catalytic reaction, DPV measurement was performed from 0.1 V to -0.3 V at amplitude of 0.05 V and pulse width of 0.01 s.

6. Bioinformatic simulation results

Table S1 illustrates the bioinformatic simulation results of the parent aptamer P and the 14 DNA mutants that have higher ZDOCK scores than P. In Table S1, the dot-bracket notation represents the secondary structures of the DNA sequences in Vienna output format, which were predicted and produced with the Mfold program. When generated with the nucleotide base addition strategy, the DNA mutants were named by putting the added nucleotide bases at the left side or the right side of the letter P (denoted for the parent aptamer), depending on the 5' end or the 3' end, respectively, that was modified. In contrast, when generated with the nucleotide base substitution strategy, the DNA mutants were named by pointing out the position of each modified nucleotide base with Arabic number, and putting the nucleotide base that was substituted and the new nucleotide base at the left side and the right sight of the Arabic number, respectively. The mean ZDOCK scores of the DNA mutants that have higher ZDOCK scores than the P sequence are also listed in Table S1.

Table S1 Bioinformatics simulation results of the parent aptamer P and the DNA mutants that have higher ZDOCK scores than P.

Name	Sequence (5'-3') [¶]	Secondary structure (Dot-bracket notation)	Minimum Δ G of the Secondary structure (kcal/mol)	ZDOCK score with CEA (mean of top 5)
P-ATG	ATA CCA GCT TAT TCA ATT((....))	0.28	1535.11
	<u>ATG</u>			

GAC-P	<u>GAC</u> ATA CCA GCT TAT TCA(.....).	1.35	1466.01
	ATT			
P-GTG	ATA CCA GCT TAT TCA ATT((.....))	0.51	1434.82
	<u>GTG</u>			
TGT-P	<u>TGT</u> ATA CCA GCT TAT TCA	.(.....).....	0.82	1418.62
	ATT			
	ATA CCA GCT TAT TCA ATT(.....)....	1.35	1408.28
	<u>ATC</u>			
A1G A3C C4T C5T	<u>GTC</u> <u>TTA</u> GCT TAT TCA ATT(.....).	1.35	1390.57
CTC-P	<u>CTC</u> ATA CCA GCT TAT TCA(.....).	1.35	1379.19
	ATT			
A3C C4T A11C	<u>ATC</u> <u>TCA</u> GCT <u>TCG</u> TCA <u>GTT</u>((.....))..	0.11	1378.76
T12G A16G				
A3C C5T	<u>ATC</u> <u>CTA</u> GCT TAT TCA ATT(.....).	1.35	1377.40
P-TGA	ATA CCA GCT TAT TCA ATT((.....)).	0.30	1376.80
	<u>TGA</u>			
CCT-P	<u>CCT</u> ATA CCA GCT TAT TCA	..((.....)).....	0.45	1375.78
	ATT			
P-TTG	ATA CCA GCT TAT TCA ATT((.....))	0.02	1369.78
	<u>TTG</u>			
A1C T2C A3T C4T	<u>CCT</u> <u>TTA</u> GCT TAT TCA ATT(.....).	1.35	1369.15
C5T				
T2C	<u>ACA</u> CCA GCT TAT TCA ATT(.....).	1.35	1367.85
Parent aptamer	ATA CCA GCT TAT TCA ATT(.....).	1.35	1366.88

†: Underlined part refers to mutations in the sequence.

Except for the top 3 ZDOCK score DNA mutants (i.e. P-ATG, GAC-P, and P-GTG) and the P sequence, the interaction models with the highest ZDOCK score in docking simulation of other 11 DNA mutants are shown in Fig. S4. The CEA crystal structure (code: 2QSQ), which was used for

the CEA-ZDOCK docking is also shown in Fig. S4. The amino acid residues of CEA in the binding interface are listed in Table S2.

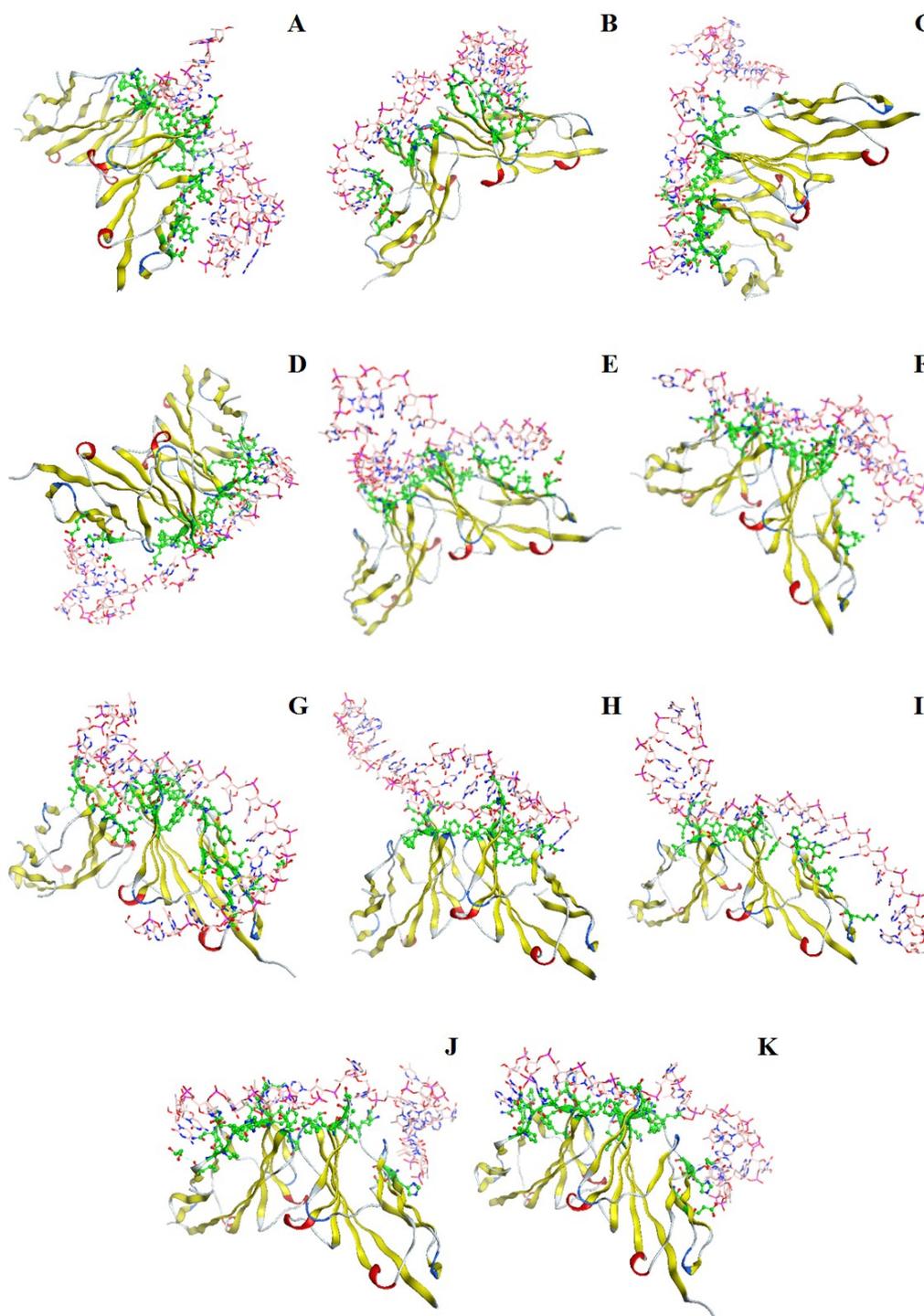


Fig. S4 Structural images of DNA-CEA docking models with the highest ZDOCK docking scores for each DNA mutant sequence. The green tags indicate the binding interface on the receptor (i.e. CEA) structure. The α -helix, β -sheet, turn, and random coil secondary structural states in the CEA structure were labeled in red, yellow, blue, and white, respectively. (A) CEA/TGT-P, (B) CEA/P-

ATC, (C) CEA/A1G A3C C4T C5T, (D) CEA/CTC-P, (E) CEA/A3C C4T A11C T12G A16G, (F) CEA/A3C C5T, (G) CEA/P-TGA, (H) CEA/CCT-P, (I) CEA/P-TTG, (J) CEA/A1C T2C A3T C4T C5T, and (K) CEA/T2C.

Table S2 The amino acid residues of CEA involved in the DNA-CEA binding interface

DNA-CEA Complex	Amino acid residues involved in the binding interface [‡]
CEA/TGT-P	A26, A27, A28, A29, A30, A31, A50, A51, A52, A53, A94, B25, B26, B27, B29, B30, B32, B49, B50, B51, B52, B53, B54, B55, B69, B70, B71, B72, B94
CEA/P-ATC	A36, A39, A41, A43, A44, A81, A82, A83, A84, A85, A86, A107, B2, B4, B5, B10, B11, B38, B86, B102, B104, B106, B108
CEA/A1G A3C C4T C5T	A28, A29, A30, A31, A50, A51, A52, A53, A94, A95, B17, B19, B27, B28, B29, B30, B32, B50, B51, B53, B69, B70, B71, B75, B94
CEA/GTC-P	A27, A28, A29, A30, A50, A51, A52, A53, A94, B29, B30, B32, B50, B51, B52, B53, B69, B70, B71, B72, B94
CEA/A3C C4T A11C T12G A16G	A28, A29, A30, A50, A51, A52, A53, A93, A94, A95, B6, B8, B19, B21, B22, B23, B25, B28, B29, B30, B32, B50, B51, B52, B53, B54, B55, B69, B70, B71, B73, B94
CEA/A3C C5T	A21, A28, A29, A30, A31, A50, A51, A52, A53, A69, A70, A71, A73, A94, B30, B47, B49, B51, B52, B53, B54, B56, B58, B62, B66, B67, B68, B69, B70, B71
CEA/P-TGA	A25, A26, A27, A28, A29, A30, A31, A32, A50, A51, A52, A53, A70, A71, A94, B1, B27, B28, B29, B30, B50, B51, B94, B95
CEA/CCT-P	A17, A19, A25, A26, A27, A28, A29, A30, A31, A32, A50, A51, A52, A53, A67, A69, A70, A71, A75, A94, B28, B29, B30, B50, B51, B52, B53, B94
CEA/P-TTG	A27, A28, A29, A30, A31, A32, A50, A51, A52, A53, A70, A71, A94, B19, B21, B29, B30, B49, B50, B51, B52, B53, B54, B55, B70, B71, B93, B94, B95
CEA/A1C T2C A3T C4T C5T	A1, A6, A23, A24, A27, A28, A29, A30, A50, A51, A52, A53, A93, A94, B25, B27, B28, B29, B30, B32, B50, B51, B52, B53, B69, B70, B71, B72, B94
CEA/T2C	A17, A19, A25, A26, A27, A28, A29, A30, A31, A32, A49, A50, A51, A52, A53, A55, A67, A69, A70, A71, A75, A77, A78, A94, B30, B50, B51, B52, B53, B94, B95, B96

[‡] : The amino acids shaded in red, yellow, blue, and white are in the α -helix, β -sheet, turn, and random coil secondary structural states of the CEA structure, respectively.

7. Bioinformatic simulation of the DNA docking with interference proteins

To simulate the interaction of interference proteins with the selected sequences, firstly the crystal structures of bovine serum albumin (BSA, code: 4F5S), human serum albumin (HAS, code: 3SQJ), γ -globulin (code: 4LLD), human alpha fetoprotein Antigen (AFP, code: 3MRK) and C-reactive protein antigen (CRP, code: 3PVN) were obtained from the Protein Data Bank. Then ZDOCK online server was used to simulate the docking between the interference proteins and the selected DNA aptamer structures. Except that the PDB file of BSA (code: 4F5S) exceeds the size limit of 1.2 MB to be imported with the ZDOCK online server, the interactions of other 4 interference proteins with the selected CEA aptamers P-ATG and GAC-P were simulated. The results showed that CEA exhibited significantly higher mean ZDOCK docking scores with P-ATG and GAC-P than the interference proteins (i.e. HSA, γ -globulin, AFP, and CRP) (Figure S5). The DNA-protein interaction models (i.e. the pose with the highest ZDOCK score) of interference proteins are shown in Fig. S6.

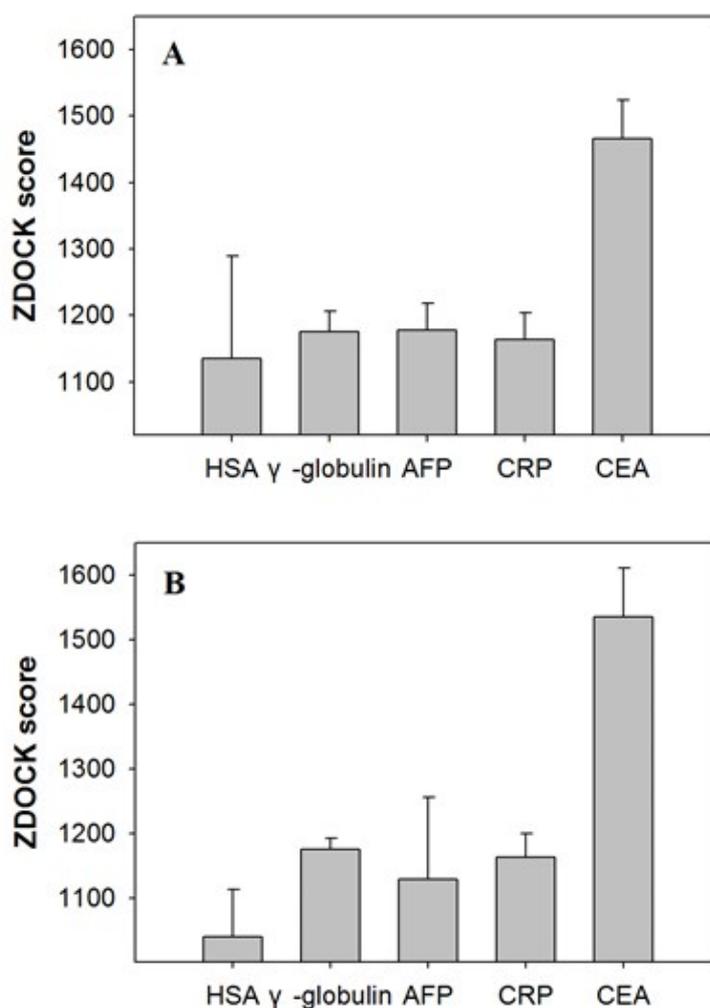


Fig. S5 Histograms of the top 5 ZDOCK binding scores of various interfering proteins including HSA, γ -globulin, AFP and CRP with the aptamer P-ATG (A) and GAC-P (B). $n = 5$.

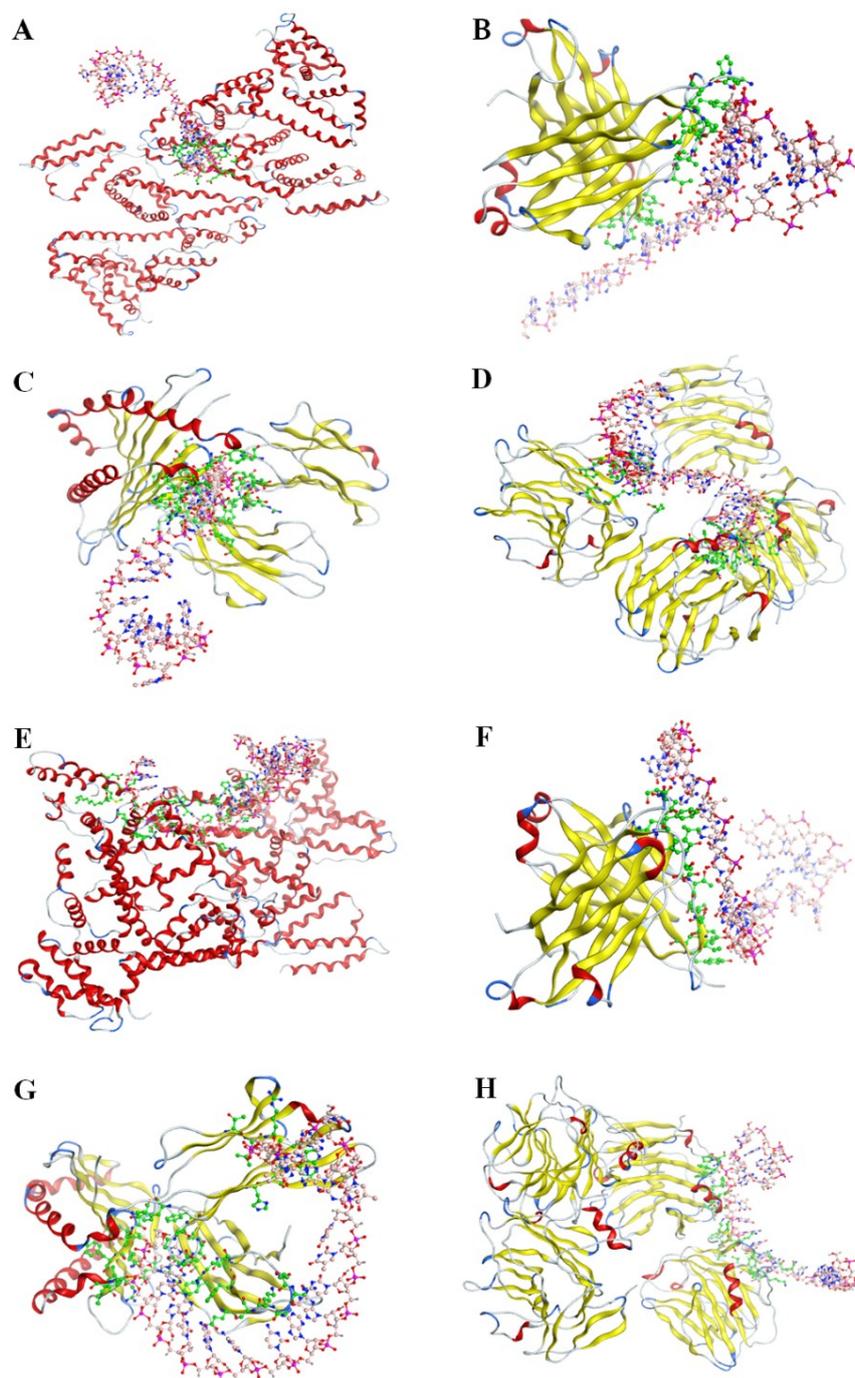


Fig. S6 Structural images of DNA-protein docking models with the highest ZDOCK docking score for each interfering protein with the selected DNA aptamer P-ATG (A, B, C, D) and GAC-P (E, F, G, H). The green tags indicate the binding interface on the receptor (i.e. interfering proteins) structure. The α -helix, β -sheet, turn, and random coil secondary structural states in the interfering

proteins structure were labeled in red, yellow, blue, and white, respectively. (A) HSA/P-ATG, (B) γ -globulin/P-ATG, (C) AFP/P-ATG, (D) CRP/P-ATG, (E) HSA/GAC-P, (F) γ -globulin/GAC-P, (G) AFP/GAC-P, and (H) CRP/GAC-P.

8. Determination of the affinity of the selected DNA aptamers to CEA

The linear regression equations for the DNA mutant P-GTG and the P sequences are $1/\text{Req} = 4.68 + 34.35 \times 1/C_{\text{CEA}}$ ($R = 0.9943$) and $1/\text{Req} = 5.64 + 39.16 \times 1/C_{\text{CEA}}$ ($R = 0.9936$), respectively.

References:

S. Kumaraswamy and R. Tobias, *Methods Mol. Boil.*, 2015, **1278**, 165-182.