

Unveiling the unusual i-motif-derived architecture of a DNA aptamer exhibiting high affinity for influenza A virus

Vladimir Tsvetkov^{1,2,3†}, Bartomeu Mir^{4†}, Ruziya Alieva⁵, Alexander Arutyunyan⁵, Ilya Oleynikov⁵, Roman Novikov⁶, Elizaveta Boravleva⁷, Polina Kamzееva⁸, Timofei Zatsеpin⁵, Andrey Aralov^{8*}, Carlos González^{4*}, Elena Zavyalova^{5*}

¹ Center for Mathematical Modeling in Drug Development, Sechenov First Moscow State Medical University, Moscow Russia

² Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia

³ A.V. Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences, Moscow, Russia

⁴ Instituto de Química Física ‘Rocasolano’, CSIC, Madrid, Spain

⁵ Lomonosov Moscow State University, Moscow, Russia

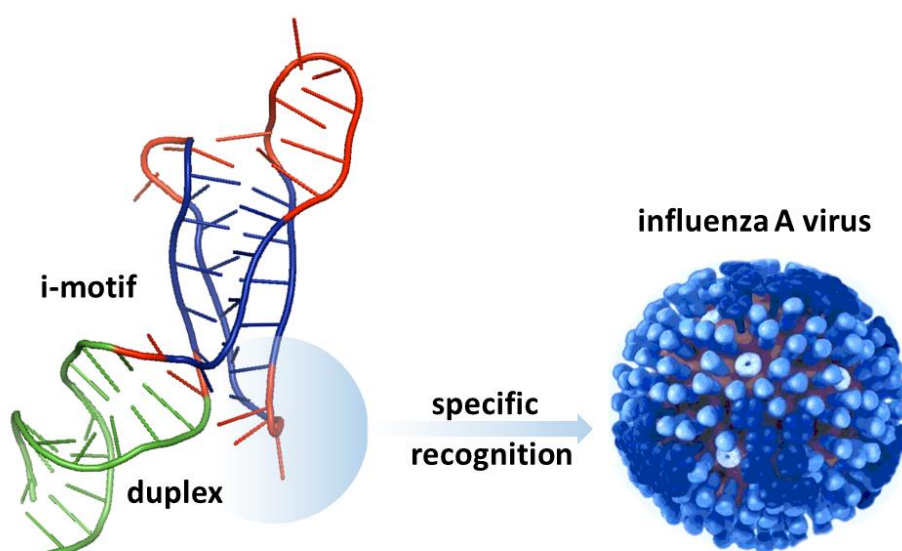
⁶ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

⁷ Chumakov Federal Scientific Center for Research and Development of Immunobiological Products (Institute of Poliomyelitis), Russian Academy of Sciences, Moscow, Russia

⁸ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

† - these authors contribute equally

*corresponding author: baruh238@mail.ru (AA), cgonzalez@iqfr.csic.es (CG) and zlenka2006@gmail.com (EZ)



Abstract

Non-canonical nucleic acid structures have been shown the capacity to selectively interact with proteins, thereby exerting influence over various intracellular processes. Among these structures, both natural and artificial G-quadruplexes have been extensively studied in relation to their structure-activity relationships. In contrast, the role of i-motifs remains incompletely evaluated. In this study the artificial aptamer BV42 possessing a high affinity for hemagglutinin of influenza A virus was proven to contain the i-motif structure even at neutral pH. However, conformational heterogeneity of BV42 poses challenges for in-depth structural investigations. Using molecular dynamics simulations and introducing chemical modifications for molecular probing, a putative binding site in the aptamer was suggested. These findings have enabled us to redesign the aptamer, eliminating the conformational diversity associated with the i-motif while preserving its binding affinity. Subsequent validation through NMR spectroscopy confirms the presence of the i-motif/duplex junction with the 3-cytosine loop located near the junction. This study elucidates a distinctive instance of an unusual nucleic acid architecture involved in molecular recognition, thereby shedding light on the structural peculiarities of functional i-motifs.

Introduction

Nucleic acid aptamers are structured oligonucleotides that exhibit high affinity and specificity towards target molecules. Aptamers can be chemically synthesized and modified at specific sites with various functional groups [1]. The conformation of aptamers is susceptible to reversible thermal denaturation and regulation of their activity through the presence of toehold oligonucleotide strands. These unique properties make aptamers valuable tools in the fields of therapeutics and diagnostics [2-6].

Optimization of affinity and selectivity of aptamers is a critical aspect in the development of aptamers both during and after the selection process known as SELEX (systematic evolution of ligands by exponential enrichment). Through sequence maturation and incorporation of non-natural nucleotides that can expand the nucleic acid alphabet and introduce additional contacts to increase stability of aptamer-target complex reducing dissociation constants (K_D) to values as low as 2-20 pM [7-11]. Moreover, structural optimization also plays a pivotal role in enhancing the fraction of favorable aptamer conformations and further lowering the K_D values of aptamer-protein complexes 10-100-fold [12,13].

Typically, aptamers contain nucleic acid hairpins with indels that form bulges, non-canonical base pairs, or other structural irregularities, creating a unique surface for target interaction. Furthermore, non-canonical secondary structures of nucleic acids, such as triplexes and G-quadruplexes (G4s), contribute significantly in target recognition in several cases (Figure 1A-C). While aptamers with confirmed i-motif (iM) structures have not been previously described, a few proteins are known to bind natural iMs [14-16].

The iM structure is formed by two parallel duplexes intercalated with antiparallel orientation and stabilized by hemi-protonated cytosine-cytosine⁺ (C:C⁺) base pairs (Figure 1D,E). Natural iMs can be formed under slightly acidic conditions by oligonucleotides containing cytosine repeats. The transition pH values (pH_i), which represent the pH at which 50% of the oligonucleotide is

folded into the iM structure, typically range from 5 to 6 [17-19]. However, iMs with extended cytosine tracts can remain stable even at neutral pH, with a pH_i as high as 7.9 [20,21].

Recently, Musafia et al. reported several DNA aptamers that bind to the hemagglutinin protein of the influenza virus [22]. These aptamers contain runs of cytosines that have the potential to fold into iMs, along with complementary sequences that form a duplex structure between 5'- and 3'-ends (Figure 1F). The leading aptamer, BV42, was shown to inhibit the binding of the influenza A virus to host cells and subsequent infection at low-nanomolar concentrations. Notably, all biological experiments were conducted in a phosphate-buffered saline (PBS) solution with neutral pH [22]. Given that BV42 has a pH_i of pH 7.1 [23], the necessity of the iM structure for the functioning of this aptamer was called into question.

This study aims to elucidate the essential role of the iM structure of BV42 in hemagglutinin binding and identify the putative binding site through combined *in vitro* and *in silico* approaches. In response to the inherent structural polymorphism of BV42, we developed and applied an approach to limit conformational heterogeneity without affecting the binding site, resulting in an aptamer with both high affinity for hemagglutinin and structural homogeneity. BV42 structure was successfully unraveled with NMR methods, revealing a junction between i-motif and duplex regions, thereby delineating an unconventional aptamer architecture critical for protein recognition.

Results

The i-Motif is necessary for the aptamer function

Typical iMs are assembled under slightly acidic conditions at pH 5-6 [17-19]. However, BV42 forms a stable iM even at neutral pH as confirmed by circular dichroism (CD) spectroscopy. CD spectra of BV42 at pH 6.0-7.0 have a positive band with a maximum at 285 nm and a negative band with a minimum at 263 nm (Figure 2A and Supplementary Figure 1) that are characteristic for iMs [24]. The sample with pH 8.0 has a CD spectrum of an unstructured oligonucleotide (Supplementary Figure 1). pH_i value was 7.3, in agreement with pH_i of iMs with extended C-runs, e.g., iM from HIF-1 α promoter region with C3-C5-C4-C4 cytosine runs ($pK_a=7.2$) [20] or an artificial iM with C4-C3-C4-C3 cytosine runs ($pK_a=7.9$) [21].

According to CD spectroscopy the melting temperature (T_m) of the iM structure decreases monotonously with increasing pH (Supplementary Table 1). The i-motif was stable even at neutral pH having T_m of 43.7°C, 32.4°C, and 26.1°C at 6.5, 7.0, and 7.3 pH, respectively (Supplementary Figures 2 and 3). Similar melting temperatures were obtained using UV spectroscopy (Supplementary Figure 4, Supplementary Table 1). This iM structure has unusually high thermal stability at neutral pH; similar melting temperatures were shown for an iM from the PDGF-A promoter region with C2-C4-C5-C5-C5-C13 cytosine runs [20]. The duplex part is assembled independently, but the melting curves often do not allow discriminating a two-stage process resulting in an apparent increase of T_m in a UV melting experiment. Melting of the duplex part is clearly seen at pH 8.0 with $T_m=61.2\pm0.1^\circ\text{C}$ (Supplementary Figure 5, Supplementary Table 1). To sum up, BV42 preserves a structure with iM and duplex modules from slightly acidic pH up to neutral pH.

Next, affinity of BV42 to hemagglutinin was studied at different pH to elucidate whether the iM structure is responsible for protein recognition. Recombinant hemagglutinin from subtype H7N1 of influenza A virus was immobilized on the sensor, and a complex formation between the aptamer and the protein in pH 6.0 (folded iM) or 8.0 (unfolded iM) PBS buffer was monitored in real time mode using biolayer interferometry. BV42 binds hemagglutinin with high affinity at

pH 6.0 (Figure 2B). Estimated K_D of the complex was 1.4 ± 0.4 nM at pH 6.0. On the contrary, K_D was 125-times higher at pH 8.0 ($K_D = 176 \pm 9$ nM, Supplementary Figure 6A). A similar experiment was performed with RHA0385, a G-quadruplex-based aptamer, binding to hemagglutinin [25], to prove that the conformation of hemagglutinin was not disrupted at pH 8.0. RHA0385 binds hemagglutinin at pH 6.0 and 8.0 with nearly the same affinity (Supplementary Figure 6B). Thus, the iM structure is required for the functional activity of BV42.

The binding affinity of the aptamer to influenza A virus was also shown to be pH-dependent. BV42 bound influenza A virus of subtype H7N1 with a dissociation constant of 0.010 ± 0.002 pM at pH 6.0 (Extended Data Figure 1A) having no affinity to either the off-target virus of Newcastle disease at pH 6.0 or the influenza A virus of subtype H7N1 at pH 8.0 (Extended Data Figure 1B). The extremely high affinity to virions is in good agreement with affinity to the recombinant protein keeping in mind that one viral particle bears nearly 1 000 molecules of hemagglutinin on its surface as well as that the multipoint binding to aptamer-functionalized surface provides an additional avidity effect [26]. This experiment proves that the iM structure is required for the biological activity of the aptamer.

Assembly of parallel cytosine duplexes stabilized by Ag^+ ions

Silver cations (Ag^+) have been previously considered as an alternative to protons for iM stabilization [27,28] working at pH 7-8, where most iMs do not exist. However, more recent data revealed that Ag^+ stabilizes parallel duplexes made up of cytosine runs without further intercalation into the iM structure [29,30]. This alternative structure was assembled for BV42 at pH 7-8 by adding $AgNO_3$. CD spectra confirmed the change of conformation of both folded iM (pH 7.0 and 7.3) and unfolded iM into a cytosine parallel duplex (Extended Data Figure 2A-C) having a characteristic intense negative maximum at 267 nm and a slight positive maximum at 290 nm at equimolar concentration of Ag^+ [30].

The affinity of this alternative structure was even lower than that of unfolded iM (Extended Data Figure 2D, Supplementary Table 2). The association rate constant increased for the alternative structure, but the dissociation rate constant increased simultaneously. These data can be interpreted as the structured state of cytosine runs being necessary for hemagglutinin binding along with suboptimal conformation of the Ag^+ -BV42 structure for stability of the aptamer-hemagglutinin complex. Cytosine duplexes alone are not enough for hemagglutinin binding; their organization into an iM is necessary.

BV42 is a mix of several conformers

The structure of BV42 was modeled using molecular dynamics tools with parameters previously optimized for iMs [31]. Molecular dynamics predicted several conformers with the same duplex and different iMs fragments (Figure 2C-E, Extended Data Figures 3,4, Supplementary Figures 7-10). Six conformers have close free energies (Supplementary Figure 11), which means that BV42 is represented as a set of coexisting iMs with the same duplex module. All predicted variants have ten C:C⁺ pairs and a conserved AAGAAGAA loop, whereas the lengths of the 2nd and 3rd loops made up from C_n were variable due to sliding of the C₂₂-run. Stability of C:C⁺ pairs was estimated by the geometry of the pairs, namely, the distances between the mass centers of cytosines and the angles between the normals to the planes of the nucleobases. Deviations from the mean for these values are shown in Extended Data Figures 3,4 and Supplementary Figures 7-10. Analysis of these data revealed that the conformer with one cytosine in the 2nd loop has destabilized C11:C30⁺ and C28:C45⁺ pairs that are located near the 2nd loop and duplex module. The C11:C30⁺ pair was totally broken, whereas the C28:C45⁺ pair is still coplanar (Extended

Data Figure 3). Thus, the conformer with the C₁ loop does not exist, the iM is partially unfolded and the free energy of this conformer is 13-31 kcal/mol higher compared to other conformers. The conformer with two cytosines in the 2nd loop has only a destabilized C11:C30⁺ pair, and this pair is still coplanar (Supplementary Figure 7). The conformers with 3 to 6 cytosines in the 2nd loop have ten C:C⁺ pairs in the iM core and the same free energy of the structure (Extended Data Figure 4, Supplementary Figures 8-10). Increasing the length of the 2nd loop beyond 3 nucleotides increased the possibility of hydrogen bonding between the loop and the duplex module (Supplementary Figure 12, Supplementary Table 3) and the number of conformations of the loop. The model with six cytosines in the 2nd loop has the least free energy, but other conformers differs by 6-8 kcal/mol, which is less than the energy of a single hydrogen bond.

Notably, the last pair in the duplex (A-T near the iM) was revealed to be broken during the modeling (see Figure 2D and a supplementary pbd file with the structure of the aptamer with 3 cytosines in the 2nd loop). The iM structuring disturbs the adjacent duplex. Next, coaxiality of iM and duplex part was tested. Duplex and iM were coaxial in the initial structure, but the angle between the two vectors decreased down to 139° immediately and fluctuated within ±15° (Extended Data Figure 5). Thus, the junction between two modules is rather flexible.

The conformational polymorphism was confirmed with ¹H NMR spectroscopy (Extended Data Figure 6). The spectra have characteristic peaks for the duplex module chemical shift 12.3-13.6 ppm and for the iM module at 15.0-15.6 ppm. However, iM peaks were not resolved forming a single broad peak that is characteristic for a mix of several conformers. Both modules were unfolded simultaneously during heating (Extended Data Figure 6). Modification of the BV42 aptamer is necessary to identify the conformer with affinity to hemagglutinin.

Phenoxazine (tC^o) modification for probing the iM structure

A cytosine analogue with an extended aromatic system, 1,3-diaza-2-oxophenoxazine (tC^o, Supplementary Figure 13), is known to maintain the cytosine H-bonding pattern, minimally affecting, if at all, iM thermal stability, p*H*_i and the folding pathway. tC^o can be used as a fluorescent analogue of cytosine, enabling the monitoring of individual residues within iMs [32]. Fluorescence quenching is characteristic for tC^o due to electron coupling interactions with surrounding protonated dC residues. Modification in the middle of a cytosine run (tC^o:C⁺ pair sandwiched between two neighboring C:C⁺ pairs) provides maximal fluorescence quenching. Modification at the end of a cytosine run (tC^o:C⁺ pair stacked only with one neighboring C:C⁺ pair) provides 1.5-times lower fluorescence quenching; whereas tC^o in the loop (no tC^o:C⁺ pair) leads to slight quenching of the initial fluorescence intensity, presumably due to weak stacking with the neighboring C:C⁺ pair and/or loop nucleobases [32].

The modification was inserted into the BV42 aptamer to elucidate the preferred conformations of iM and local stability of C:C⁺ pairs. A set of modified oligonucleotides (Supplementary Table 4) was studied by CD, UV and fluorescent spectroscopies and biolayer interferometry. The first C₅-run (C11-C15), terminal cytosines of the second C₅-run (C24 and C28) and nucleotides in the region of the putative 2nd C-loop (C29, C30, C31, and C32) were modified individually to identify the cytosines involved in iM formation.

In accordance with the previous report [32], iM stability was affected minimally by a single introduction of tC^o. Melting temperatures in CD and UV melting experiments varied slightly (up to ±10% of the value for the unmodified aptamer) confirming that tC^o is an excellent non-perturbing fluorescent cytosine surrogate for probing individual residues in the BV42 aptamer (Supplementary Tables 5 and 6, Supplementary Figures 14-17).

Based on position-dependent fluorescence quenching, the C11, C15, C24 and C32 nucleotides are involved in the formation of outer C:C⁺ pairs, whereas the C12, C13 and C14 nucleotides are inside the iM structure at pH 6.0 and 6.5 (Supplementary Table 7). In turn, C-sliding may be responsible for partial relocation of the C28 residue from the outer hemiprotonated pair inside the C run, accompanied by a displacement of C29 and C30 residues from the 2nd loop into the iM core and a change in the loop length. Finally, the C31 nucleotide is positioned in loop regardless of C sliding. Cytosines from outer C:C⁺ pairs (C11, C15 and C24) quit the iM at pH 7.0 leaving C12, C13, C14, C28 and C29 in the shortened iM version. The C28 and C32 nucleotides retain their place in the iM at pH 7.0, that can be possible in the case of total remodeling of the iM during folding and indicates, in addition to conformers existing at pH 6.0, the presence of extra iM variants at pH 7.0. Thus, the fluorescence quenching experiments confirm iM folding and its structural polymorphism, as well as allow mapping some of the residues within the iM structure.

CD and UV melting experiments provide information about overall iM stability. In contrast, fluorescence melting experiments give insight into the local stability of the molecular environment involved in the interactions with fluorescent tC^o residue. In this case, melting temperatures are drastically dependent on the site of modification [32,33] and may strongly deviate from those obtained by CD and UV melting (Table 1, Supplementary Figures 17-19). Fluorescence melting showed negligible differences in the local tC^o environment at high iM stability (pH 6.0). Only tC^o in the BV42°C11 aptamer quickly dropped out of the stacking interactions with the neighbors. As the pH value increased (pH 6.0→7.3), the highest decrease in melting temperature was observed for the C11, C15, C24 and C30 modifications (Table 1), that may indicate their location in the outer hemiprotonated pairs in accordance with the proposed structure.

Affinity studies of tC^o-modified aptamers to hemagglutinin could provide insight into the structural moieties involved in target recognition. The affinity of all tC^o-modified aptamers to the recombinant hemagglutinin was close to that of unmodified BV42 (Table 2, Supplementary Figure 20-24). The most significant differences were observed for aptamers BV42°C28 and BV42°C31. BV42°C28 has the highest dissociation constant, 3.4 times higher than K_D of BV42. The decrease in affinity was caused by a 3-fold decrease of the association rate constant. BV42°C31 has the lowest dissociation constant with 2-fold decrease compared to BV42. The increase in affinity was achieved by stabilization of the complex due to a 1.5 times lower dissociation rate constant of BV42°C31 compared with BV42. Both modified positions are located near or within the 2nd loop; C28 is a terminal cytosine in the iM core, whereas C31 is located in the 2nd loop (Figure 2F). The deviations in the affinity of C28- and C31-modified aptamers suggest the participation of these residues in the binding to hemagglutinin.

Molecular dynamics confirmed the localization of tC^o at the 28th and 31th positions in the iMs. tC^o at 28th position forms a stable C:C⁺ pair with C45 (Supplementary Figure 25, Supplementary Table 8). Overall iM stability was diminished by 1.3%, in agreement with the slight decrease in thermal stability (Supplementary Tables 5 and 6). On the contrary, tC^o at the 31th position tends to be in the loop instead of a C:C⁺ pair, having pair occupancy below 50% (Extended Data Figure 7, Supplementary Figure 26, Supplementary Table 8). Again, the overall stability of iM was diminished by 1.2%, in agreement with the slight decrease in thermal stability (Supplementary Tables 5 and 6).

To conclude this part, molecular probing has confirmed the putative iM structure and located the hemagglutinin recognition site near the 2nd loop.

Structural studies. Redesigning of iM-based aptamer

As mentioned above, ^1H NMR spectra of BV42, consistent with computational predictions, indicate the simultaneous formation of iM and a duplex with Watson-Crick base pairs. DNA molecules, in which i-motif coexists with B-DNA moieties, have been proposed and recently observed at neutral pH [34]. To date, the only high-resolution structure of a DNA molecule with an i-motif and a B-DNA part, forming a so-called iM/duplex junction, has been determined by Serrano-Chacón [34]. The structure remains stable at neutral pH, in part due to the presence of a minor groove G:T:G:T tetrad capping one edge of the i-motif. Minor groove tetrads can induce significant shifts in the pK_a of the neighboring C:C⁺ base pair, leading to strong stabilization at neutral pH. The presence of a minor groove tetrad at one side of the iM moiety not only stabilizes the structure but also eliminates potential effects of sliding of a C-tract with respect to others, resulting in better quality of NMR spectra.

We used this approach to redesign the BV42 aptamer to obtain more comprehensive structural information in the absence of C-sliding. Based on the affinity data of BV42°C10 and BV42°C13, we explored various DNA constructs while preserving the 2nd loop of the iM and replacing the 1st and 3rd loops (AAGAAGAA and C19-C22) with short sequences capable of forming a G:T:G:T tetrad with two T2 loops (Supplementary Figure 27). Also, the AT pairs were replaced with GC-pair at the duplex-iM interface in order to stabilize the duplex part in NMR experiments. We compared several constructs with a different number of cytosines in the 2nd loop (1, 3 or 5 cytosines, respectively). The sequences studied (CapTr5C1, CapTr5C3 and CapTr5C5) are shown in Supplementary Table 4.

The NMR spectra of CapTr5C1 and CapTr5C3 exhibited narrower signals than the parent aptamer BV42, suggesting reduced conformational heterogeneity compared to BV42. The imino region spectra are consistent with the formation of C:C⁺, G:C and A:T Watson-Crick base pairs, and G:T mismatches at neutral and acidic pH (Supplementary Figure 28). Although the size of these systems makes the spectra too complex for complete sequential assignments, the cross-peaks patterns in the NOESY experiments (Supplementary Figures 29 and 30) confirm the formation of duplex and i-motif structures, in agreement with the designed structures (Supplementary Figure 27). In both cases, exchangeable protons signals are visible at high temperature, indicating that the structures are very stable (Supplementary Figure 28). This high stability is confirmed by CD and UV melting experiments (Supplementary Figure 33, Extended Data Figure 8, and Supplementary Table 9).

Interestingly, CapTr5C1 and CapTr5C3 have the same affinity for hemagglutinin as the parent aptamer BV42 ($K_D=1.4\text{--}1.5\text{ nM}$). In turn, CapTr5C5 exhibited extremely low affinity for hemagglutinin with $K_D>100\text{ nM}$ (Table 3, Supplementary Figure 31). The affinity of several other variants was also tested (Table 3, Supplementary Figure 32). CapTr5C2 and CapTr5C4, with two or four cytosines in the 2nd loop, respectively, have 5-fold and 10-fold lower affinity than BV42. Thus, the length of the 2nd loop is important for target recognition. Models of the aptamer with a 1-nucleotide loop are not stable, as can be clearly seen from molecular dynamics calculations for the BV42 conformer and pH dependence for CapTr5C1 (Extended Data Figure 3). This is consistent with the NMR spectra of CapTr5C1. The low-intensity imino signals observed at low temperature at 15.0 and 15.9 ppm (Supplementary Figure 28) suggest the presence of minor i-motif conformations, most probably affecting the 2nd loop, since the signals from the G:T:G:T tetrad and the Watson-Crick base pairs do not show any indication of multiple species. These signals are not observed in CapTr5C3, suggesting that a one-residue loop is

suboptimal, leading to distortion of the structure. The higher molar circular dichroism observed for CapTr5C3 vs. CapTr5C1 is also in line with this conclusion.

Our approach, consisting of replacing the loops responsible for structural heterogeneity with a G:T:G:T tetrad at the side of the i-motif opposite to the duplex region, provided shorter oligonucleotides with good affinity and stability properties. However, these DNA constructs are still too large to yield NMR spectra of sufficient quality for their complete assignment. To gain more insight into the structure of these aptamers, we explored shorter constructs with a reduced number of C:C⁺ pairs in the i-motif module and eliminated the terminal AT base pairs in the duplex region. The sequences CapTr4C3 and CapTr2C3 were studied using NMR methods (Figure 3A, Supplementary Figure 34). ¹H NMR spectra present similar features to the larger constructs previously studied (Figure 3B, Supplementary Figure 35). In the case of CapTr2C3, signals of the C:C⁺ pair disappear at a relatively lower temperature compared to previous constructs, indicating that the i-motif part of the structures is less stable. Therefore, we focused on the sequence CapTr4C3.

NMR spectra of CapTr4C3 are shown in Figure 3, Extended Data Figure 9, Supplementary Figure 36-39. Complete assignment of the duplex region could be achieved following standard strategies. Characteristic sequential amino-methyl cross-peaks from 5' – TC – 3' steps of the duplex region served as a starting point to identify spin systems corresponding to T4 and C5. From this point onwards, sequential aromatic-aromatic and sugar-aromatic contacts, together with cross-peaks between exchangeable protons involved in Watson-Crick base pairing, allowed sequential assignment of the B-DNA moiety. Formation of the G8-C36 base pair could not be confirmed, indicating its low stability due to its proximity to the interface. Assignment of the i-motif region was more complicated due to strong overlapping of the cytosine amino signals, but it could be partially achieved. Key contacts between hemiprotonated C residues confirmed the presence of a C:C⁺ stack capped by G residues involved in the formation of the G:T:G:T tetrad. The 5' – C12-G13-T14 – 3' and 5' – C27-G28-T29 – 3' tracks were easily followed by sequential Me/H2'/H2''-H6/H8 and by GH1-CH3⁺ cross-peaks. G13:G28 minor groove interaction was confirmed on the basis of H1'-H1' cross-peaks. GH1-TH1 cross-peaks were detected at slightly acidic conditions (pH 6), confirming the formation of the G:T:G:T tetrad. H42/H41 protons could not be related to aromatic/sugar protons due to large overlap of the signals. However, CH3⁺-CH3⁺ cross-peaks led us to sequentially assign six stacked C:C⁺ bp as C12-C27⁺ (15.34 ppm), C17-C32⁺ (15.69 ppm), C11-C26⁺ (15.49 ppm), C18-C33⁺ (15.50 ppm), C10-C25⁺ (15.64 ppm) and C19-C34⁺ (15.52 ppm). This stacking order was confirmed by characteristic H3⁺-H2'/H2'' and H42-H2'/H2'' spatial cross-peaks between cytosines through the major grooves of the structure. The assignment list is shown in Supplementary Table 10. Protons of the residues in the G:T:G:T tetrad and in most of the i-motif regions could be assigned. However, the protons in the C-loop and in the two cytosines close to the duplex region (C20 and C35) were not detected. These results correlate well with the experiments on phenoxazine-modified BV42 whereas the C:C⁺ pairs near the duplex part disappeared under the increase of pH from 6 to 7 (Supplementary Table 7).

The lack of exchangeable signals for these residues indicates that they are not involved in base pairs. 234 experimental distance constraints plus additional distances derived from confirmed base pairs (Supplementary Table 11) were used to determine the structure, as described in the Materials and Methods section. The ensemble of superimposed structures is shown in Extended Data Figure 10. The duplex and most of the i-motif regions are well-defined when superimposed independently (see Supplementary Table 10). However, due to the lack of distance constraints in

the interface and in the C-loop regions, the relative position of the i-motif and the duplex is not well-defined (Extended Data Figure 10).

Despite the lack of definition in the interface region, the solution structure of CapTr4C3 at pH 7 (Extended Data Figure 10) confirms many of the features suggested by the NMR spectra of previous constructs. The complementary region adopts a well-defined B-form helical structure (Figure 3). Six C:C⁺ base pairs are clearly intercalated, forming a standard i-motif. Cytosines C9, C20, C24 and C35 are not base-paired, as well as the three cytosines in the 2nd loop. The GC base pair in the duplex close to the C-rich region is probably distorted, but the presence of a guanine imino proton nearly at 13 ppm suggests that these residues are base-paired. This highly flexible interface region contrasts with the well-defined interface observed in the three-dimensional structure of 7O5E [34]. It must be noted that the topology of the two interfaces is different since the complementary region in CapTr4C3 occurs between two terminal segments, whereas in 7O5E it occurs in a loop region connecting two C-tracts.

Although the two redesigned aptamers, CapTr5C1 and CapTr5C3, can recognize hemagglutinin with high affinity, the latter presents better thermal and pH stability than that of BV42, with a p*H*_i of 7.3 (Extended Data Figure 8) and the highest melting temperatures among the aptamers studied (Supplementary Table 9, Supplementary Figure 33). In contrast, the loop bearing three thymine residues (CapTr5T3) had a 2-fold diminished affinity, reflecting the loss of 1-2 polar contacts in the complex. The suboptimal loops decreased the association rate constants, affecting the recognition process. Moreover, shortening the C-run length by one or two cytosine residues (CapTr4C1 and CapTr4C3 vs. CapTr3C1) results in a 4-5-fold and 37-fold decrease in affinity, respectively (Table 3). However, for three cytosine runs, iM stability was diminished significantly, complicating the interpretation of the results (Supplementary Figure 40). On contrary, CapTr4C3 has stable iM with melting temperature 57.4±0.6°C at pH 6.0 that is even higher than *T*_m of BV42 under the same conditions (Supplementary Figure 41). Similarly, pH stability of the redesigned aptamer is high, namely, p*H*_i=7.5 that is higher than that of BV42 (p*H*_i=7.3, Supplementary Figure 40). The effect of C-run shortening is obvious at neutral pH providing 3°C lower *T*_m at pH 7.0 (Supplementary Figure 42). Thus, the redesigned version is a good model of BV42 conformer with 3 cytosines in the 2nd loop.

Discussion

iMs are known to be possible regulatory elements in genomes affecting both transcription and replication [16,35]. The folded iM structures have been identified in the genome by immunoassay [36]. However, there are a few proteins known to bind genomic iMs up to date [14-16]. The structural features of iMs that are recognized by proteins are still unknown. Recently, an artificial C-rich oligonucleotide was designed to bind hemagglutinin of influenza A virus and inhibit cell binding of the virus and virus entry into the host cell [22]. Our data reveal that BV42 contains an iM module that is important for the protein recognition (Figure 2). It binds both recombinant and viral hemagglutinin (Figure 2 and Extended Data Figure 1) in folded state at pH 6.0 with *K*_D 1.4 nM and 10 fM, respectively; whereas affinity in the unfolded state at pH 8.0 was negligible (Supplementary Figure 6A). In addition, alteration of aptamer conformation by Ag⁺ ions leads to the absence of the functional activity (Supplementary Table 2).

The aptamer is conformationally heterogeneous due to sliding of the C₂₂-run during the folding process. Molecular dynamics predicted the most likely conformations with the 2nd loop bearing 3-6 cytosine residues (Extended Data Figures 3 and 4, Supplementary Figures 7-10). The 1-nucleotide loop destabilizes the adjacent C:C⁺ pair, increasing the actual length of the loop. The conformational heterogeneity led to unresolvable ¹H NMR spectra (Extended Data Figure 6).

The tC^o nucleotide was introduced into BV42 to map the cytosine residues that tend to be included in iM. This modification was originally designed as a flanks-dependent duplex stabilizer [37,38], having the ability to change fluorescence when intercalating into the duplex [39,40]. Maintenance of the cytosine H-bonding pattern and non-disturbing geometry of the modification are responsible for the minimal impact, if any, on thermal stability, pK_i and the folding pathway and allows the use of tC^o as a fluorescent surrogate of cytosine to monitor individual residues within iMs [32]. The single-nucleotide modification caused slight rearrangement of the iM, namely, introduction of modified nucleotide into the core. Overall stability was minimally affected independent from the tC^o position in iMs (Supplementary Tables 5 and 6). On contrary, fluorescence melting revealed significant difference in the stability of tC^o:C⁺ pairs depending on the modification site and pH of folding (Table 1). These experiments support that iM arrangement is different at pH 6.0 and 7.0. Thus, the structure of the aptamer is too complicated for unambiguous structural assignment.

Serrano-Chacón et al. proposed an alternative way for iM stabilization at neutral pH, where the G:T:G:T tetrad formed a cap that restricted C-run sliding [34]. This approach was used to redesign BV42. Affinity experiments on tC^o-modified BV42 variants showed significant changes in K_D for the two modified aptamers (Table 2) with tC^o near the 2nd loop (C29-C31), which suggests that the loop participates in hemagglutinin recognition. This loop was kept in the new sequences but its length was varied from 1 to 5 nucleotides. The G:T:G:T cap was introduced instead of AAGAAGAA and C37-C40 loops. The designed sequences were shown to form iM structures at pH 6.0 (Extended Data Figure 8). However, appropriate affinity was maintained only for two variants, namely, CapTr5C1 and CapTr5C3 with 1 and 3 cytosine residues in the 2nd loop, respectively. These two oligonucleotides form remarkably stable structures at neutral pH. According to molecular dynamics simulations, the CapTr5C1 loop is unstable and, most probably, the structure losses one outer C:C⁺ pair, leading to the restoration of 3-nucleotide loop arrangement. A three-cytosine loop might be required for the proper configuration of the recognition site (Extended Data Figure 3), since replacing cytosine residues in the 2nd loop with thymines while maintaining the 3-nucleotide length led to 2-fold loss in affinity. Attempts to shorten the iM core resulted in 5-fold and 37-fold decreases in affinity for C₄ and C₃ iM runs, respectively (Table 3). Thus, the high affinity to hemagglutinin is determined by a combination of an extended stable iM core and a cytosine 3-nucleotide loop.

The NMR data and molecular dynamics have shown that duplex and iM modules coexist both in BV42 and its redesigned truncated version. Surprisingly, the modules are not in stacking, having a flexible junction between the modules. As the putative site of hemagglutinin recognition is within the 2nd loop near the junction, flexibility could play a significant role in induced fit of the aptamer conformation during hemagglutinin binding.

Conclusions

In this study, we present the first demonstration of i-motif structures in aptamers playing a pivotal role in protein/aptamer recognition. Specifically, the iM-duplex structure operates as an interface for hemagglutinin binding, with the short cytosine loop positioned prominently. Molecular dynamics elicited several conformers with comparable free energies for the structures, impeding comprehensive structural investigations. Subsequent aptamer redesign furnished a new configuration characterized by a conformational homogeneity and a high affinity for hemagglutinin. The unique i-motif/duplex architecture with a flexible junction was elucidated with NMR. These findings represent a significant achievement exemplifying the rational design

of aptamers and highlighting the functional capability of DNA i-motifs in specific protein recognition under physiological conditions.

Materials and Methods Online

Reagents

Inorganic salts, acids, alkaline and tris were purchased from AppliChem GmbH (Darmstadt, Germany). 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC-HCl) from Roth (Karlsruhe, Germany) and sodium salt of N-hydroxysulfosuccinimide (s-NHS) from Chem-Impex Int'l (Wood Dale, IL, USA) were used. Glutaric aldehyde was purchased from Ruschim (Moscow, Russia). Recombinant hemagglutinin (HA1 subunit) from A/Anhui/1/2013 (H7N9) strain (ab190421, Abcam, Burlingame, CA, USA) was used. HEPES-HCl buffer (1 M) with pH 7.5 and AgNO₃ were purchased from AppliChem GmbH (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, MO, USA), respectively.

Buffers with different pH were prepared based on conventional phosphate buffered saline (PBS, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl pH 7.35) with addition of NaOH or HCl solutions. The target pH values were pH 6.0, 6.5, 7.0, 7.3 and 8.0. pH values were controlled with ST20 pH-meter (Ohaus Corporation, Parsippany, New Jersey, United States). All solutions were prepared using ultrapure water produced by Millipore (Merck Millipore, Burlington, Massachusetts, United States).

Oligonucleotides and sample preparation

Oligonucleotides were synthesized using commercially available reagents by a solid-phase phosphoramidite method, followed by purification with high performance liquid chromatography. The tC^o phosphoramidite was prepared according to the reported procedure [41]. The sequences and sites of modification are provided in Supplementary Table 4. Typically, 1 or 2 μ M solution of the aptamer in PBS buffer (pH 6.0, 6.5, 7.0, 7.3 or 8.0) was heated at 95°C for 5 minutes. The solutions were used after gradual cooling to room temperature. For the affinity experiments, the solutions were diluted with PBS buffer with the same pH. For the experiments with Ag⁺ ions, the AgNO₃ solution was added to 1.3 μ M BV42 aptamer solution in PBS buffer (pH 7.0, 7.3 or 8.0), the sample was incubated for 5 minutes at room temperature. The final concentrations of AgNO₃ were in the range of 0.6-20 μ M.

Circular dichroism and UV spectroscopy

1-2 μ M aptamer solutions in PBS buffer with pH 6.0-8.0 were placed into quartz cuvettes with 1 cm path. CD and UV spectra were acquired using Chirascan CD spectrometer (Applied Photophysics, Leatherhead, Great Britain) and MOS-500 spectrometer (BioLogic, Seyssinet-Pariset, France) equipped with a thermoelectric temperature regulator. The spectra were acquired in the wavelength range of 230–360 nm. The spectrum of the buffer was subtracted as the baseline. The samples were heated with a mean ramp of 1.0°C/min. The melting experiments were conducted in the temperature range of 10–85°C. The melting temperatures were derived from the temperature dependencies of molar CD intensity at the wavelength of 285 nm and UV absorption at 260 nm. The curves were approximated with the Boltzmann function using OriginPro software (OriginLab, Northampton, MA, USA), and the parameter x0 corresponded to the value of the melting temperature.

NMR

¹H NMR spectra of BV42 were recorded using Bruker AVANCE III HD 300 and Bruker AVANCE III HD 400 spectrometers with 300.1 and 400.1 MHz, respectively (Bruker, MA, USA). ¹H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). ¹H NMR spectra of samples in H₂O (90%) / D₂O (10%) buffer (phosphate-citrate buffer with 100 mM Na₂HPO₄ and 50 mM citric acid with pH adjusted to 4.7 or 6.0) were recorded in the temperature range from 5°C to 80°C using the WATERGATE-W5 pulsed-field gradient pulse sequence (zgpgw5 from Bruker library) for complete H₂O-signal clear suppression. A relaxation delay of 2.0 sec and an acquisition time of 1.5 sec were used for all experiments. Starting solutions of DNA aptamer were annealed in all cases for 5 minutes at 95°C and cooled at 5°C temperature. All samples were diluted by D₂O to the final content of 10%.

Samples for NMR experiments were dissolved in 9:1 H₂O/D₂O (10 mM sodium phosphate buffer). Experiments were carried out in neutral or slightly acidic conditions (pH 6 to 7). The pH was adjusted by adding aliquots of concentrated solution of either DCl or NaOD. NMR spectra were acquired on Bruker spectrometers operating at 600 or 800 MHz, equipped with cryoprobes and processed with TOPSPIN software (Bruker, MA, USA). NOESY spectra in 9:1 H₂O/D₂O were acquired with mixing times of 100 and 200 ms. TOCSY spectra were recorded with 80 ms mixing time.

Fluorescence melting experiment

2 μM tC^o-modified aptamer solutions in PBS buffer with pH 6.0-8.0 were placed in a 96-well black non-binding microplate with clear bottom (Greiner Bio-One GmbH, Frickenhausen, Germany). The fluorescence spectra were obtained using CLARIOstar (BMG Labtech, Ortenberg, Germany). The excitation wavelength was 370 nm, the emission wavelengths were in the range of 410-550 nm. The melting experiment was performed by gradual increase in temperature from 20°C to 65°C with a mean ramp of 0.7°C/min.

Viruses

Influenza viruses and allantoic fluids were provided by the Chumakov Federal Scientific Center for Research and Development of Immune and Biological Products of the Russian Academy of Sciences. Influenza A virus (strain A/chicken/Rostock/45/1934, H7N1) and Newcastle disease virus were studied. Viral stocks were propagated in the allantoic cavity of 10-day-old embryonated specific pathogen-free chicken eggs. The eggs were incubated at 37°C for 48 h after infection, cooled to 4°C, kept for 16 h and then harvested. The study design was approved by the Ethics Committee of the Chumakov Institute of Poliomyelitis and Viral Encephalitis, Moscow, Russia (Approval #4 from 2 December 2014). Viruses were inactivated by the addition of 0.05% (v/v) glutaric aldehyde, preserved by the addition of 0.03% (w/v) NaN₃ and stored at +4°C.

The activity of the viruses was determined by a hemagglutination assay according to Killian et al. [42]. 50 μl of solutions of the viruses diluted two times step-by-step in PBS was placed into a V-shaped 96-well microtiter plate in a volume of 50 μl. Then, 50 μl of 0.5% chicken red blood cells in PBS were added to the well. The plate was kept in the refrigerator at 4°C for 1 h. Then, the hemagglutination titer was estimated as the maximal dilution of the virus that did not cause the precipitation of red blood cells; this well contained 1 HAU of the virus in the probe. Influenza A virus had hemagglutination activity of 80 000 HAU/mL, and Newcastle disease virus had hemagglutination activity of 2 000 HAU/mL.

The concentrations of viruses were determined with Nanoparticle Tracking Analysis (NTA). Nanoparticle Tracking Analysis (NTA) was conducted using a ZetaView® PMX420-QUATT instrument (Particle Metrix GmbH, Germany), while the data were analyzed by ZetaView NTA software. Operating instructions of the manufacturer were followed before calibrating the instrument with a known concentration of 100 nm polystyrene nanoparticles (Applied Microspheres B.V., Netherlands). The standards were suspended in particle-free water, whereas the investigated samples were diluted 1:100 with pH 7.3 PBS. Particles were counted and size-distributed at 10 cycles of 11 frames per cycle under sensitivity of 65 and a shutter value of 100. Influenza A virus contained $6 \cdot 10^9$ VP/mL, and Newcastle disease virus contained $4 \cdot 10^{10}$ VP/mL.

Affinity assays

The affinity of the aptamers to recombinant hemagglutinin and influenza viruses was estimated using biolayer interferometry (Blitz, ForteBIO, Menlo Park, CA, USA) at 20°C. Samples were placed into black 0.5 mL tubes (Sigma-Aldrich, St. Louis, MO, USA) in a 220 µL volume.

To study the affinity to recombinant hemagglutinin, biosensors intended for the amine coupling reaction (Octet AR2G biosensors, ForteBio, Menlo Park, CA, USA) were activated for 5 min in a solution of 200 mM EDC-HCl and 100 mM s-NHS. Then, hemagglutinin was loaded from a 10 µL drop of 15 µg/mL solution in PBS with pH 6.0 for 10 min. The sensor was washed and blocked with 100 mM HEPES-HCl buffer with pH 7.5 for 3 min. After signal stabilization in the PBS with pH 6.0 or 8.0, the association step was conducted with aptamer solution in the concentration 1000, 500, 250 or 125 nM in the same PBS buffer. Then, the dissociation step was conducted in the same PBS buffer. Both the association and the dissociation stages were monitored for 200 s. Regeneration of the sensors was conducted by incubation in PBS buffer with pH 8.0 for 5 min. The design of the experiment provided zero reference signal, so the sensorgrams were processed without subtraction with baseline correction only. Each experiment was performed twice.

To study the affinity to influenza viruses, 5'-biotinylated aptamers were immobilized onto streptavidin-coated biosensors (Octet SAX biosensors, ForteBIO, Menlo Park, CA, USA) by incubating the sensor in 1 µM aptamer solution in PBS pH 6.0 or 8.0 for 5 min. Viral particles of A/chicken/Rostock/45/1934 (H7N1) strain or Newcastle disease virus were prepared in the PBS pH 6.0 or 8.0 by diluting the virus 5, 10 or 20 times. After signal stabilization in the PBS with pH 6.0-8.0, the association step was conducted with virus solution in the same PBS buffer. Then the dissociation step was conducted in the same PBS buffer. Both the association and the dissociation stages were monitored for 200 s. Regeneration of the sensors was conducted by incubation in PBS buffer with pH 8.0 for 5 min. The design of the experiment provided zero reference signal, so the sensorgrams were processed without subtraction with baseline correction only. Each experiment was performed twice.

Sensorgram processing was performed using OriginPro software (OriginLab, Northampton, MA, USA) using the Langmuir 1:1 binding model recommended for affinity estimations [43]. Values of the association (k_{on}) and dissociation (k_{off}) rate constants were calculated; then, equilibrium dissociation constants K_D were calculated as the ratio of the two rate constants ($K_D = k_{off}/k_{on}$).

Molecular modeling

The aptamer models were built based on the strategy described in the work [31] using Sybil-X software (Certara, USA). At the first stage, duplex and IM models were created. The iM's core comprising ten cytosine-cytosine pairs was obtained based on the reported tetramolecular iM models (PDB 1YBL). Then they were joined to each other, then loops were added to the IM. At

each stage of docking, MM-optimization was carried out using the SYBYL X and the Powell method with the following settings: partial charges and parameters for interatomic interactions were from the Amber7ff02 force field, a nonbonded cut-off distance was set to 8 Å, a distance-dependent dielectric function was applied, the number of iterations was equal to 500, the simplex method was used in the initial optimization, and the energy gradient convergence criterion of 0.05 kcal/mol/Å was used. Partial charges on 1,3-diaza-2-oxophenoxazine atoms were calculated according to the following scheme. First, calculation of electron density distribution was performed by DFT using the hybrid meta functional M06-2X [47] and 6-311G+(d,p) Pople basis sets. Then, the Merz-Singh-Kollman scheme was applied to the electron density distribution obtained to calculate the grid for the electrostatic potential fitting with the following parameters: (6/41=10) - the number of surfaces around the atoms and (6/42=17) - the density of test points on these surfaces. The RESP (Restrained ElectroStatic Potential) method was applied to the fitting of the grid obtained in the previous step to calculate partial atomic charges. All quantum mechanics simulations were performed using Gaussian 16 software (Gaussian, Inc., Wallingford, CT, USA) [48].

The models were verified by MD simulations using Amber 20 software (AMBER Software, San Francisco, CA, USA) [49]. The influence of the solvent was simulated using the OPC3 water model [50]. Rectangular box and periodic boundary conditions were used in the simulation. The space between the models and the periodic box wall was at least 15 Å. Potassium ions were used to neutralize the negative charge of the DNA backbone. The parameters needed for interatomic energy calculation were taken from the force fields OL15 [51,52] for the DNA and gaff2 for 1,3-diaza-2-oxophenoxazine. At the beginning of the computation, the models with thrombin were optimized in two stages. First, the location of the solvent molecules was optimized by using 1000 steps (500 steps of steepest descent followed by 500 steps of conjugate gradient). At this stage, the mobility of all solute atoms was restrained by a force constant of $500 \text{ kcal} \times \text{mol}^{-1} \times \text{Å}^{-2}$. In the second stage, the optimization was performed without restrictions using 2500 steps (1000 steps of steepest descent and 1500 steps of conjugate gradient). Then, gradual heating to 300 K was carried out for 20 ps. To avoid spontaneous fluctuations at this point, weak harmonic restraints were applied with a force constant of $10 \text{ kcal} \times \text{mol}^{-1} \times \text{Å}^{-2}$ for all atoms other than the solvent ones. The SHAKE algorithm was applied to constrain hydrogen-containing bonds, which allowed the use of a 2 fs time step. Scaling of 1–4 nonbonded van der Waals and electrostatic interactions was performed by using the standard Amber values. The cut-off distance for nonbonded interactions was set to 10 Å, and the long-range electrostatics were calculated using the particle mesh Ewald method. The MD simulations in the production phase were performed using constant temperature ($T = 300 \text{ K}$) and constant pressure ($p = 1 \text{ atm}$) over 80 ns. To control the temperature, a Langevin thermostat was used with a collision frequency of 1 ps^{-1} . The energy of interaction of bases of cytosines in the in C-C pairs was estimated using the `lie` command from the `cpptraj` module with a dielectric constant of 2. Free energy was calculated as the sum of the electrostatic energies (E_q), Van der Waals energies (E_{VDW}), energy of solvation and deformation energy of valence bonds, valence and dihedral angles (U). The energy of solvation was calculated as the sum of the polar and nonpolar contributions. The polar contribution (E_{GB}) was computed using the Generalized Born (GB) method and the algorithm developed by Onufriev et al. for calculating the effective Born radii [53]. The non-polar contribution to the solvation energy (E_{surf}), which includes solute-solvent van der Waals interactions and the free energy of cavity formation in solvent, was estimated from solvent-accessible surface area (SASA). The plots of geometrical parameters and energy of interaction vs. time were smoothed using the moving average method (span = 5).

Acknowledgments

The NTA experiments were performed using the equipment of the MSU Shared Research Equipment Center “Technologies for obtaining new nanostructured materials and their complex study” and purchased by MSU in the frame of the Equipment Renovation Program (National Project "Science"). The circular dichroism experiments were performed using the equipment of the MSU Shared Research Equipment Center “Subdiffraction microscopy and spectroscopy” and purchased by MSU in the frame of the Equipment Renovation Program (National Project "Science"). Salary support from the Government project 121031300037-7 is acknowledged.

The authors are grateful to A. Korshun for a fruitful discussion of the manuscript.

References

- [1] Mayer G. The chemical biology of aptamers. *Angew Chem Int Ed Engl.* 2009, 48 (15), 2672-89.
- [2] Chen J., Zhou J., Peng Y., et al. Aptamers: A prospective tool for infectious diseases diagnosis. *J Clin Lab Anal.* 2022, 36(11), e24725.
- [3] Zhuo Z., Yu Y., Wang M., et al. Recent advances in SELEX technology and aptamer applications in biomedicine. *Int J Mol Sci.* 2017, 18 (10), 2142.
- [4] Xu Y, Jiang X, Zhou Y, et al. systematic evolution of ligands by exponential enrichment technologies and aptamer-based applications: recent progress and challenges in precision medicine of infectious diseases. *Front Bioeng Biotechnol.* 2021, 9, 704077.
- [5] Chandola C., Kalme S., Casteleijn M.G. et al. Application of aptamers in diagnostics, drug-delivery and imaging. *J Biosci.* 2016, 41, 535–561.
- [6] Zhou W., Huang P.J.J., Ding J., Liu J. Aptamer-based biosensors for biomedical diagnostics. *Analyst.* 2014, 139, 2627-2640.
- [7] Antipova O.M., Zavyalova E.G., Golovin A.V. et al. Advances in the application of modified nucleotides in SELEX technology. *Biochemistry (Moscow).* 2018, 83, 10, 1161–1172.
- [8] Duo J., Chiriac C., Huang R. Y.-C., et al. Slow off-rate modified aptamer (SOMAmer) as a novel reagent in immunoassay development for accurate soluble glypican-3 quantification in clinical samples. *Anal Chem.* 2018, 90, 8, 5162–5170.
- [9] Abeydeera N.D., Egli M., Cox N., et al. Evoking picomolar binding in RNA by a single phosphorodithioate linkage. *Nucleic Acids Res.* 2016, 44 (17): 8052-8064.
- [10] Davies D.R., Gelinas A.D., Zhang C., et al. Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc Natl Acad Sci USA.* 2012, 109(49), 19971–19976.
- [11] Yatime L., Maasch C., Hoehlig K., et al. Structural basis for the targeting of complement anaphylatoxin C5a using a mixed L-RNA/L-DNA aptamer. *Nat Commun.* 2015, 6, 6481.
- [12] Zavyalova E.G., Legatova V.A., Alieva R.S., et al. Putative mechanisms underlying high inhibitory activities of bimodular DNA aptamers to thrombin. *Biomolecules.* 2019, 9, 2, 41.

- [13] Zavyalova E., Tagiltsev G., Reshetnikov R., et al. Cation coordination alters the conformation of a thrombin-binding G-quadruplex DNA aptamer that affects inhibition of thrombin. *Nucleic acid therapeutics*. 2016, 26, 5, 299–308.
- [14] Niu K., Zhang X., Deng H., et al. BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res*. 2018, 46(4), 1710–1723.
- [15] Bai X., Talukder P., Daskalova S.M., et al. Enhanced Binding Affinity for an i-Motif DNA Substrate Exhibited by a Protein Containing Nucleobase Amino Acids. *J Am Chem Soc*. 2017, 139, 13, 4611–4614.
- [16] Kang H.-J., Kendrick S., Hecht S.M., Hurley L.H. 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*. 2014, 136, 11, 4172–4185.
- [17] Abou Assi H., Garavís M., González C., Damha M.J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res*. 2018, 46(16), 8038–8056.
- [18] Day H.A., Pavlou P., Waller Z.A. i-Motif DNA: structure, stability and targeting with ligands. *Bioorg Med Chem*. 2014, 22(16), 4407–4418.
- [19] Abdelhamid M.A.S., Waller Z.A.E. Tricky topology: persistence of folded human telomeric i-motif DNA at ambient temperature and neutral pH. *Front Chem*. 2020, 8, 40.
- [20] Brazier J.A., Shah A., Brown G.D. i-Motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes. *Chem Commun*. 2012, 48, 10739–10741.
- [21] Fujii T., Sugimoto N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys Chem. Chem. Phys*. 2015, 17, 16719–16722.
- [22] Musafia B., Oren-Banaroya R., Noiman S. Designing anti-influenza aptamers: novel quantitative structure activity relationship approach gives insights into aptamer – virus interaction. *PLOS ONE*. 2014, 9(5): e97696.
- [23] Zavyalova E., Kopylov A. G-quadruplexes and i-motifs as scaffolds for molecular engineering of DNA aptamers. In *G-quadruplex structures, formation and roles in biology* (2016), Nova Publishers New York, pp. 53–80.
- [24] Kypr J., Kejnovská I., Renciuk D., Vorlíčková M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res*. 2009, 37(6), 1713–1725.
- [25] Bizyaeva A.A., Bunin D.A., Moiseenko V.L., et al. The functional role of loops and flanking sequences of G-quadruplex aptamer to the hemagglutinin of influenza A virus. *Int J Mol Sci*. 2021, 22, 2409.
- [26] Einav T., Gentles L.E., Bloom J. D. SnapShot: influenza by the numbers. *Cell* 2020, 182, 532.
- [27] Day H.A., Huguin C., Waller Z.A.E. Silver cations fold i-motif at neutral pH. *Chem Commun*. 2013, 49, 7696.
- [28] Shi Y., Sun H., Xiang J., et al. i-Motif-modulated fluorescence detection of silver(I) with an ultrahigh specificity. *Anal Chim Acta*. 2015, 857, 79–84.

- [29] Huard D.J.E., Demissie A., Kim D., et al. Atomic structure of a fluorescent Ag₈ cluster templated by a multistranded DNA scaffold. *J Am Chem Soc.* 2019, 141, 29, 11465–11470.
- [30] Kohl F.R., Zhang Y., Charnay A.P., et al. Ultrafast excited state dynamics of silver ion-mediated cytosine–cytosine base pairs in metallo-DNA. *J Chem Phys.* 2020, 153, 105104.
- [31] Tsvetkov V.B. Modeling of possible quadruplexes and i-motifs formed during DNA contacts: strategy, classification, most probable shapes, origami based on quadruplexes. *bioRxiv* 2022.10.10.511558.
- [32] Reilly S.M., Lyons D.F., Wingate S.E., et al. Folding and hydrodynamics of a DNA i-motif from the c-MYC promoter determined by fluorescent cytidine analogs. *Biophys J.* 2014, 107(7), 1703–1711.
- [33] Bielecka P., Juskowiak B. Fluorescent sensor for pH monitoring based on an i-motif - Switching aptamer containing a tricyclic cytosine analogue (tC). *Molecules* 2015, 20, 18511–18525.
- [34] Serrano-Chacón I., Mir B., Escaja N., González C. Structure of i-Motif/Duplex Junctions at Neutral pH. *J Am Chem Soc.* 2021, 143, 33, 12919–12923.
- [35] Kendrick S., Muranyi A., Gokhale V., et al. Simultaneous drug targeting of the promoter MYC G-quadruplex and BCL2 i-motif in diffuse large B-cell lymphoma delays tumor growth. *J Med Chem.* 2017, 60(15), 6587–6597.
- [36] Brown S.L., Kendrick S. The i-motif as a molecular target: more than a complementary DNA secondary structure. *Pharmaceuticals (Basel).* 2021, 14(2), 96.
- [37] Lin K.-Y., Jones R.J., Matteucci M. Tricyclic 2'-deoxycytidine analogs: syntheses and incorporation into oligodeoxynucleotides which have enhanced binding to complementary RNA. *J Am Chem Soc.* 1995, 117, 13, 3873–3874.
- [38] Ortega J.-A., Blas J.R., Orozco M., et al. Binding affinities of oligonucleotides and PNAs containing phenoxazine and G-clamp cytosine analogues are unusually sequence-dependent. *Org Lett.* 2007, 9, 22, 4503–4506.
- [39] Sandin P., Börjesson K., Li H., et al. Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue, *Nucleic Acids Research.* 2008, 36(1), 157–167.
- [40] Gardarsson H., Kale A.S., Sigurdsson S.T. Structure–function relationships of phenoxazine nucleosides for identification of mismatches in duplex DNA by fluorescence spectroscopy. *ChemBioChem.* 2011, 12(4), 567–575.
- [41] Schönrath I., Tsvetkov V.B., Zatsépin T.S. et al. Silver(I)-mediated base pairing in parallel-stranded DNA involving the luminescent cytosine analog 1,3-diaza-2-oxophenoxazine. *J Biol Inorg Chem.* 2019, 24, 693–702.
- [42] Killian, M. L. Hemagglutination assay for influenza virus, in animal influenza virus. Editor E. Spackman (New York, NY: Humana Press), 1161, 3–9. *Methods in Molecular Biology (Methods and Protocols)*, 2014.
- [43] Masson L., Mazza A., De Crescenzo G. Determination of affinity and kinetic rate constants using surface plasmon resonance. *Methods Mol Biol.* 2000, 145, 189–201.

- [44] Kato K., Ikeda H., Miyakawa S., et al. Structural basis for specific inhibition of autotaxin by a DNA aptamer. *Nat Struct Mol Biol.* 2016, 23(5), 395–401.
- [45] Jarvis T.C., Davies D.R., Hisaminato A., et al. Non-helical DNA triplex forms a unique aptamer scaffold for high affinity recognition of nerve growth factor. *Structure.* 2015, 23(7), 1293–1304.
- [46] Russo Krauss I., Spiridonova V., Pica A., et al. Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res.* 2016, 44(2), 983–991.
- [47] Zhao, Y., Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account.* 2008, 120; 215–241.
- [48] Gaussian 16, Revision C.01. Frisch M. J., Trucks G. W., Schlegel H. B. et al.. Gaussian, Inc., Wallingford CT, 2016.
- [49] Case D.A., Belfon K., Ben-Shalom I.Y. et al. (2020). AMBER 2020, University of California, San Francisco.
- [50] Izadi S., Onufriev A.V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.*, 2016, 145, 074501.
- [51] Zgarbová M., Luque F.J., Šponer J. et al. Toward improved description of DNA backbone: Revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.*, 2013, 9, 2339–2354.
- [52] Zgarbová M., Sponer J., Otyepka M. et al. Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theor. and Comp.*, 2015, 12, 5723–5736.
- [53] Onufriev A., Case D.A., Bashford D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* 2002, 23, 1297-1304.

Figures

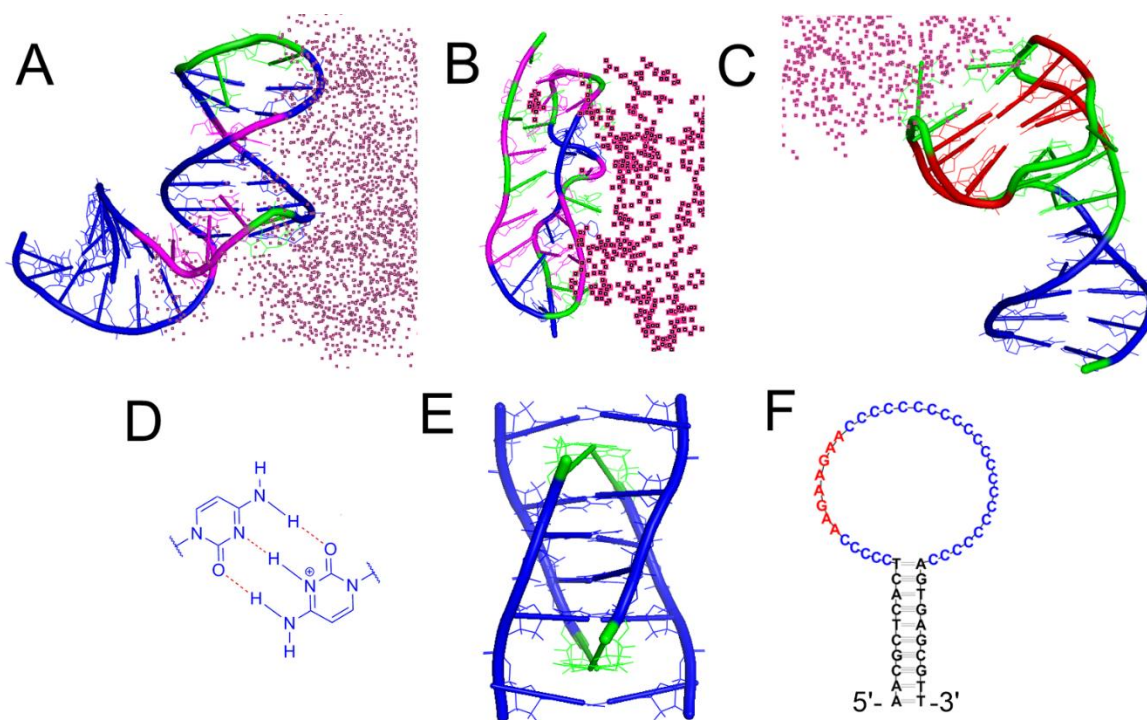


Figure 1. Examples of aptamer structures in complexes with target proteins. The non-canonical secondary structures of the DNA aptamers provide unique surfaces for target recognition. A – hairpin-based aptamer RB011 with autotaxin (PDB id 5hrt) [44]; B – triplex-based aptamer SL1049 with nerve growth factor (PDB id 4zbn) [45]; C – G-quadruplex-based aptamer RE31 with thrombin (PDB id 5cmx) [46]. The proteins are shown as arrays of red dots; aptamers are colored structure-dependently, namely, canonical Watson-Crick pairs are shown in blue, non-canonical base pairs are shown in magenta, guanines from G-quadruplexes are shown in red, unpaired nucleotides are shown in green. Another non-canonical DNA structure, iM, is composed of hemi-protonated cytosine pairs (D) that form intercalated DNA duplexes (E). The strands in the duplexes are in parallel orientation; 5'-ends of DNA are colored in green. The sequence of DNA aptamer, BV42, with a long C run that might form the iM structure (F).

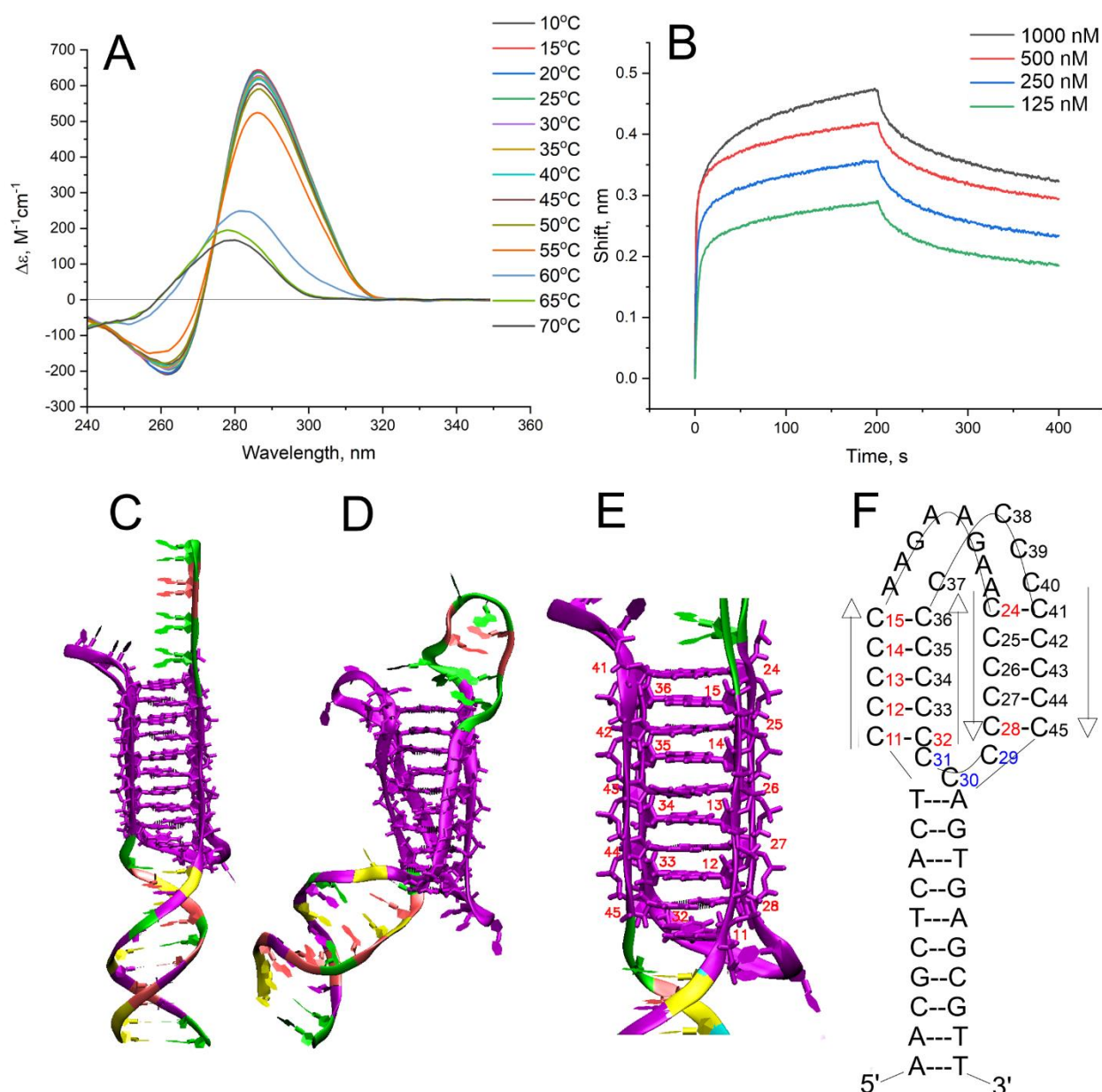


Figure 2. Structure and affinity of aptamer BV42. Temperature dependence of CD spectra in PBS buffer at pH 6.0 (A). Aptamer binding to immobilized recombinant hemagglutinin measured by biolayer interferometry at pH 6.0 (B). The conformers of the BV42 aptamer with three cytosines in the 2nd loop at the start (C) and at the end (D) of the molecular dynamics simulation. The iM structure is shown separately (E). A schematic representation of the stable conformer with sites labelled by tC^o modification (F); red color indicates modified cytosines in the iM core; blue color indicates modified cytosines in the loop; arrows indicate the polarity of the strand from 5'-end to 3'-end. The modified aptamers contained a single modification at each indicated site.

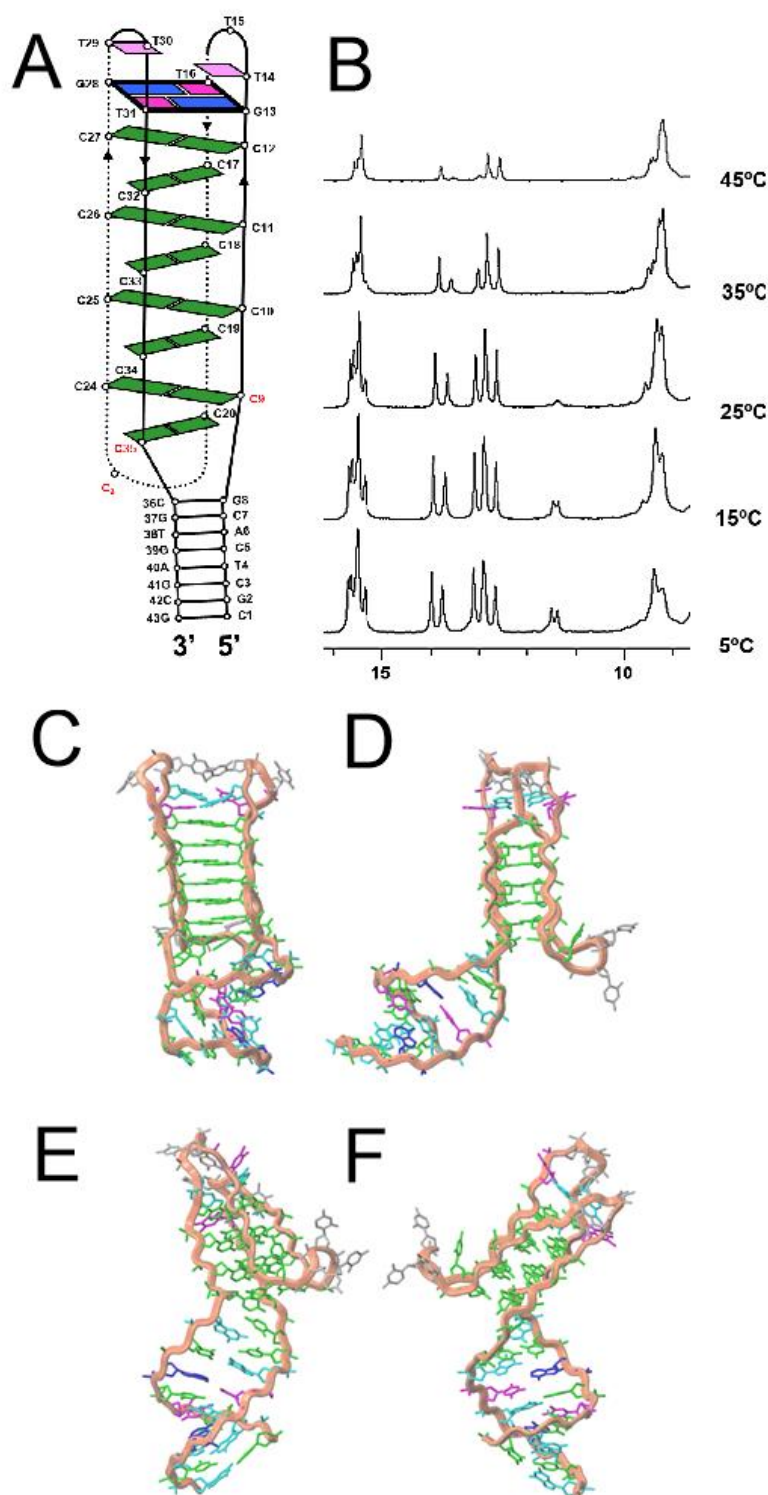


Figure 3. Structure of the redesigned CapTr4C3 aptamer and NMR spectra confirming the proposed structure and its conformational homogeneity. Scheme of the putative structure of CapTr4C3. Cytosines labeled in red may not be forming C:C⁺ base pairs. (A). ¹H NMR spectra at various temperatures recorded for an 800 μ M solution of CapTr4C3 in 10 mM phosphate buffer with pH 7 (B). Views of a representative structure of CapTr4C3: view from the major groove of the i-motif moiety (C); view from the i-motif minor groove (D); view from the minor groove side of the duplex moiety (E); view from the duplex major groove (F). Color code: thymine are shown in magenta, cytosines are shown in green, guanines are shown in cyan and disordered loops residues are shown in gray.

Tables

Table 1. Melting temperatures of the tC⁰-modified BV42 aptamers at pH 6.0-7.3 as determined by fluorescent spectroscopy.

Aptamer	pH 6.0	pH 6.5	pH 7.0	pH 7.3
BV42°C11	44.9±0.5	32.0±0.8	<20	<20
BV42°C12	53.0±0.7	44.0±0.3	40.9±1.0	33±4
BV42°C13	>65	47.9±0.7	34.1±0.6	28.8±0.6
BV42°C14	>65	50.4±0.3	37.0±0.3	34.3±0.5
BV42°C15	59.0±0.2	45.0±0.3	28.7±1.6	<20
BV42°C24	>65	55.1±0.7	35.7±0.4	<20
BV42°C28	62±10	56.9±1.0	43.5±0.6	32.7±0.4
BV42°C29	>65	52.6±0.5	45±3	36.5±0.5
BV42°C30	>65	47.5±0.7	34.1±0.3	<20
BV42°C31	>65	66±2	52.5±1.9	37.2±0.3
BV42°C32	>65	57.4±0.9	42.2±0.3	37±2

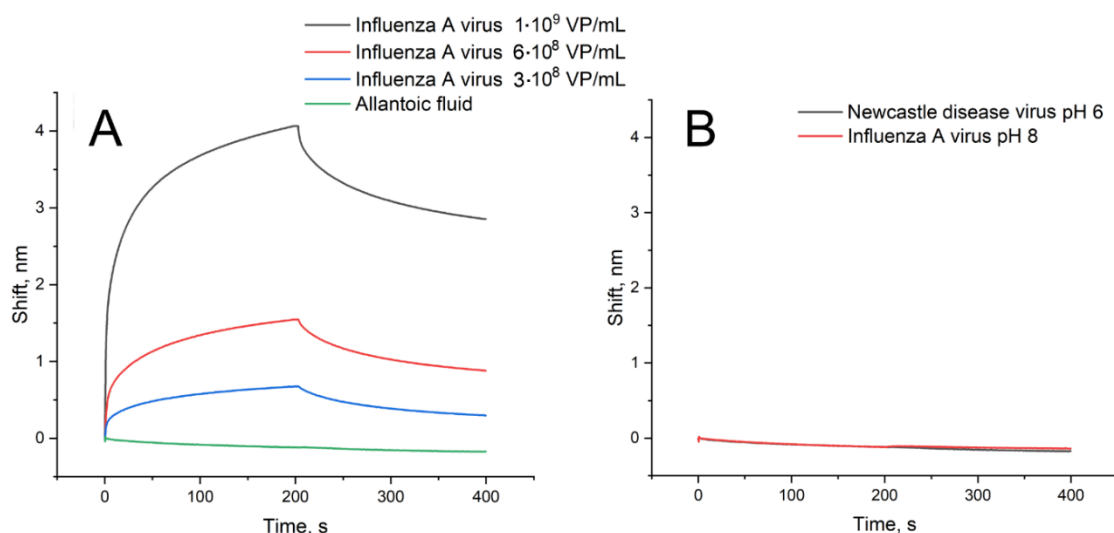
Table 2. Affinity of tC⁰-modified BV42 aptamers to recombinant hemagglutinin. The data were obtained at pH 6.0 and 8.0. K_D is the equilibrium dissociation constant; k_{ass} is the association rate constant; k_{diss} is the dissociation rate constant.

Aptamer	K _D , nM (pH 6.0)	k _{ass} , μM ⁻¹ ·s ⁻¹ (pH 6.0)	k _{diss} · 10 ³ , s ⁻¹ (pH 6.0)	K _D , nM (pH 8.0)
BV42	1.4±0.4	1.3±0.4	1.5±0.2	176±9
BV42°C11	2.2±0.5	0.61±0.16 ↓	1.37±0.12	90±10
BV42°C12	1.6±0.6	1.2±0.6	1.3±0.2	>100
BV42°C13	1.1±0.3	1.1±0.4	1.13±0.12 ↓	>100
BV42°C14	2.1±0.5	0.9±0.2	1.7±0.3	>100
BV42°C15	2.8±0.4 ↑	1.25±0.11	3.0±1.2 ↑	>100
BV42°C24	1.7±0.4	0.56±0.13 ↓	1.3±0.3	>100
BV42°C28	4.8±0.9 ↑	0.45±0.10 ↓	2.2±0.3 ↑	80±20
BV42°C29	1.0±0.4	1.9±0.8	1.68±0.15	80±30
BV42°C31	0.7±0.2 ↓	1.4±0.3	0.90±0.06 ↓	70±30
BV42°C32	0.9±0.3	1.0±0.2	1.06±0.12 ↓	70±15

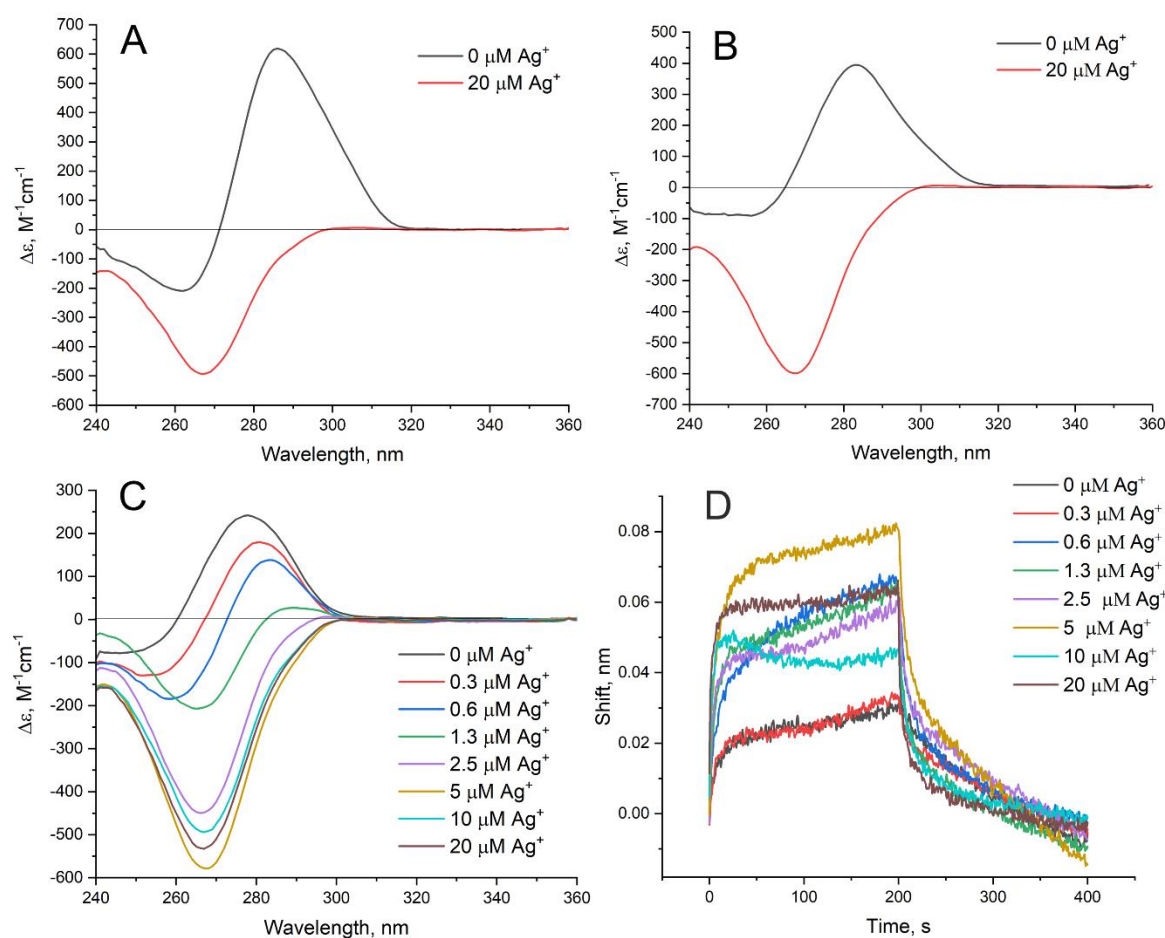
Table 3. Affinity of the aptamer candidates to recombinant hemagglutinin. The data were obtained at pH 6.0. K_D is the equilibrium dissociation constant; k_{ass} is the association rate constant; k_{diss} is the dissociation rate constant.

Aptamer	K_D , nM	k_{ass} , $\mu\text{M}^{-1}\cdot\text{s}^{-1}$	$k_{\text{diss}} \cdot 10^3$, s^{-1}
BV42	1.4±0.4	1.3±0.4	1.5±0.2
CapTr5C1	1.4±0.2	2.2±0.8	2.9±0.7
CapTr5C2	7±2 ↑	0.39±0.08 ↓	2.5±0.4
CapTr5C3	1.5±0.4	1.7±0.4	2.5±0.2
CapTr5C4	13±5 ↑↑	0.24±0.09 ↓	2.9±0.2
CapTr5C5	>100 ↑↑↑	n.d.	n.d.
CapTr5T3	2.8±0.7 ↑	0.7±0.2	1.8±0.2
CapTr4C1	8±2 ↑	0.37±0.09 ↓	2.8±0.2
CapTr4C3	5.0±1.5 ↑	0.38±0.09 ↓	1.8±0.2
CapTr3C1	52±17 ↑↑	0.12±0.04 ↓↓	5.9±0.3 ↑

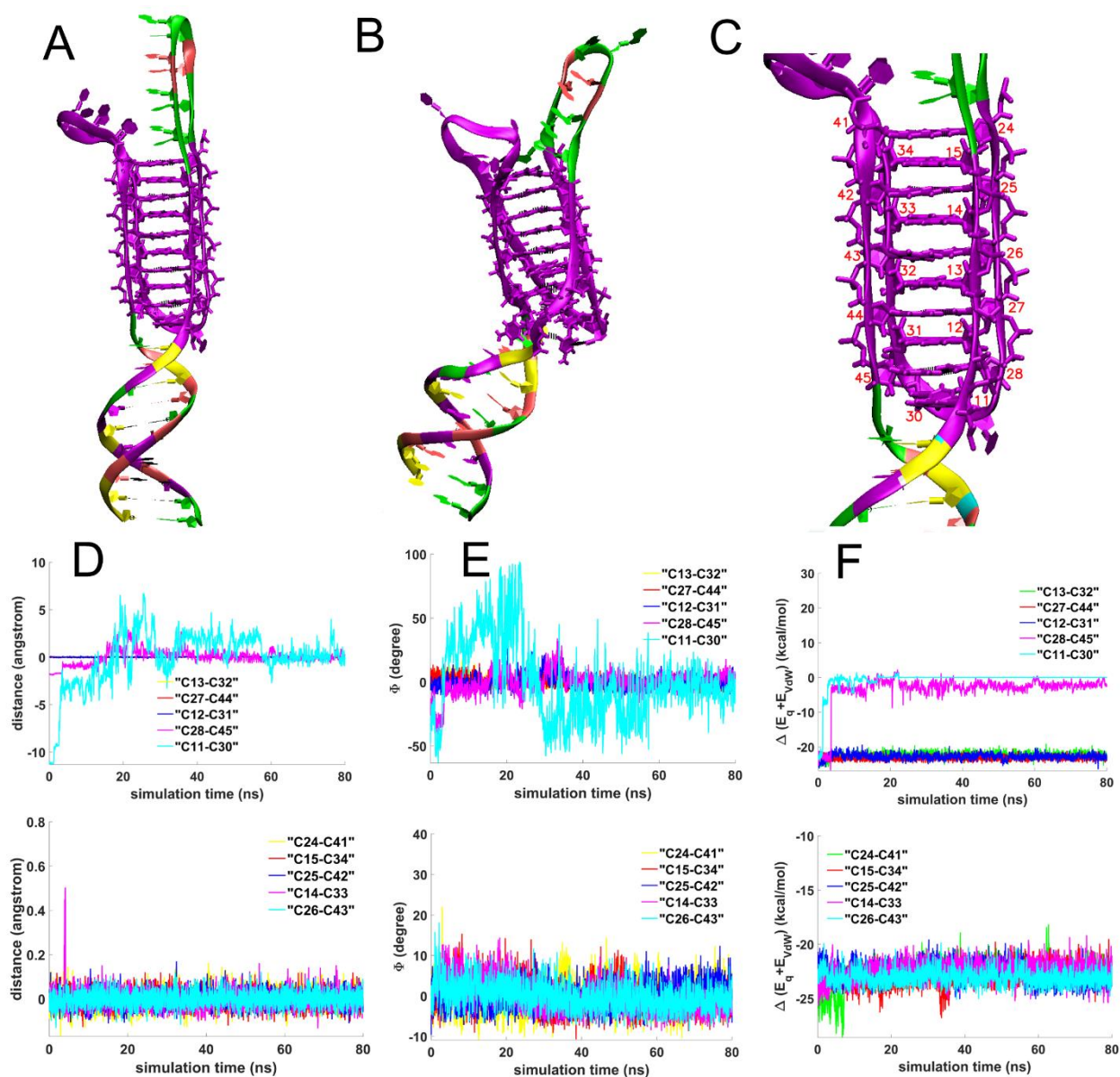
Extended data figures



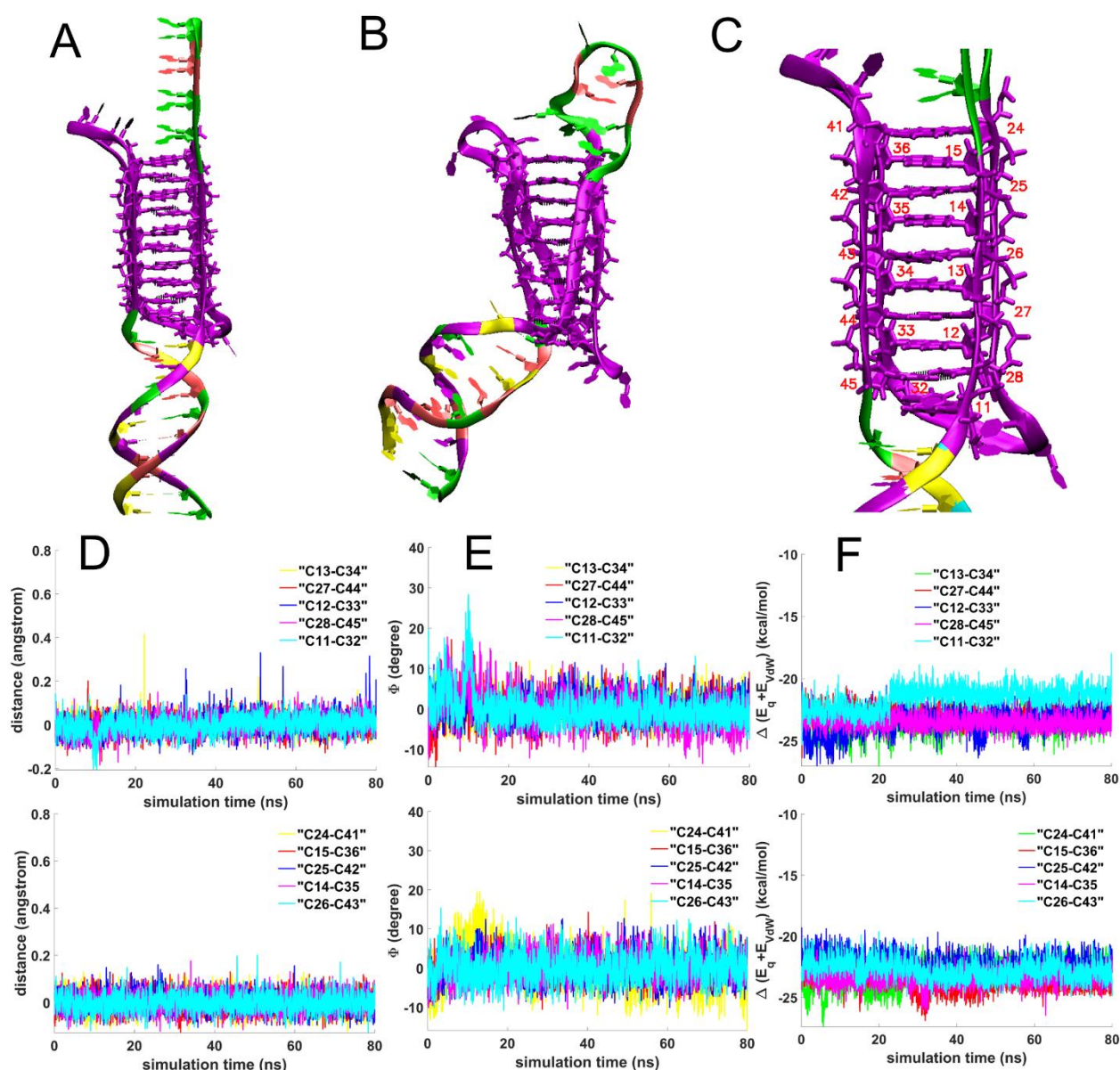
Extended Data Figure 1. Virus binding to immobilized BV42 aptamer measured by biolayer interferometry. Influenza A virus and allantoic fluid binding at pH 6.0 (A). Newcastle disease virus at pH 6.0 and influenza A virus at pH 8.0; both viruses were in concentrations of $2 \cdot 10^9$ VP/mL (B).



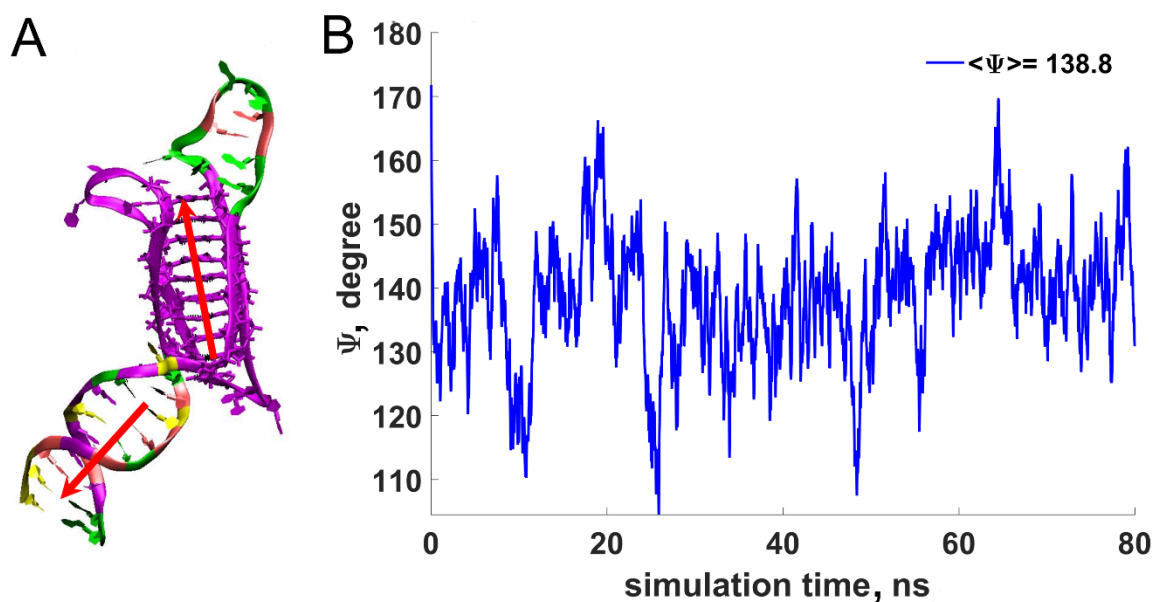
Extended Data Figure 2. CD spectra of BV42 in PBS buffer before and after incubation with AgNO₃ at pH 7.0 (A), 7.3 (B), 8.0 (C). Aptamer concentration was 1.3 μ M. Subset D shows the binding of Ag⁺-stabilized BV42 to immobilized recombinant hemagglutinin measured by biolayer interferometry at pH 8.0.



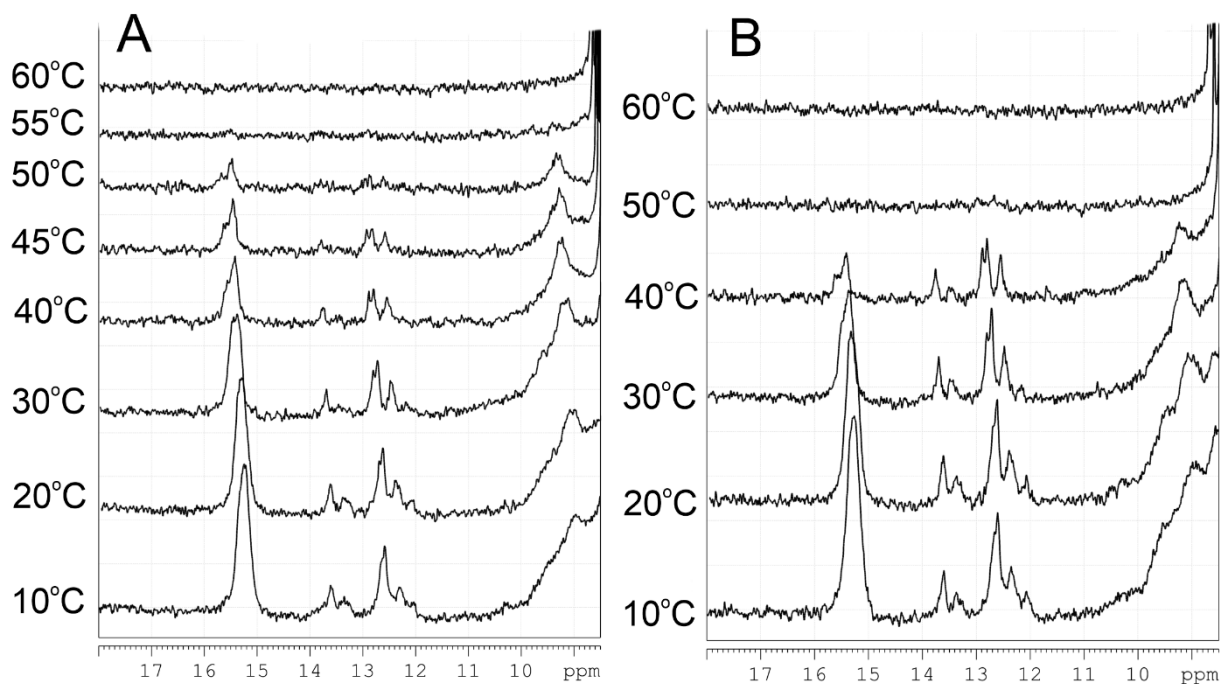
Extended Data Figure 3. Molecular dynamics simulation results of BV42 with one cytosine in the 2nd loop. Conformers of BV42 at the start (A) and at the end (B) of the molecular dynamics simulation. The enlarged iM structure is shown in subset (C). Distances between mass centers of the cytosines (D); angles between normals to the cytosines (E); sum of electrostatic (E_q) and Van der Waals (E_{vdw}) contributions to interaction energy between cytosines (F).



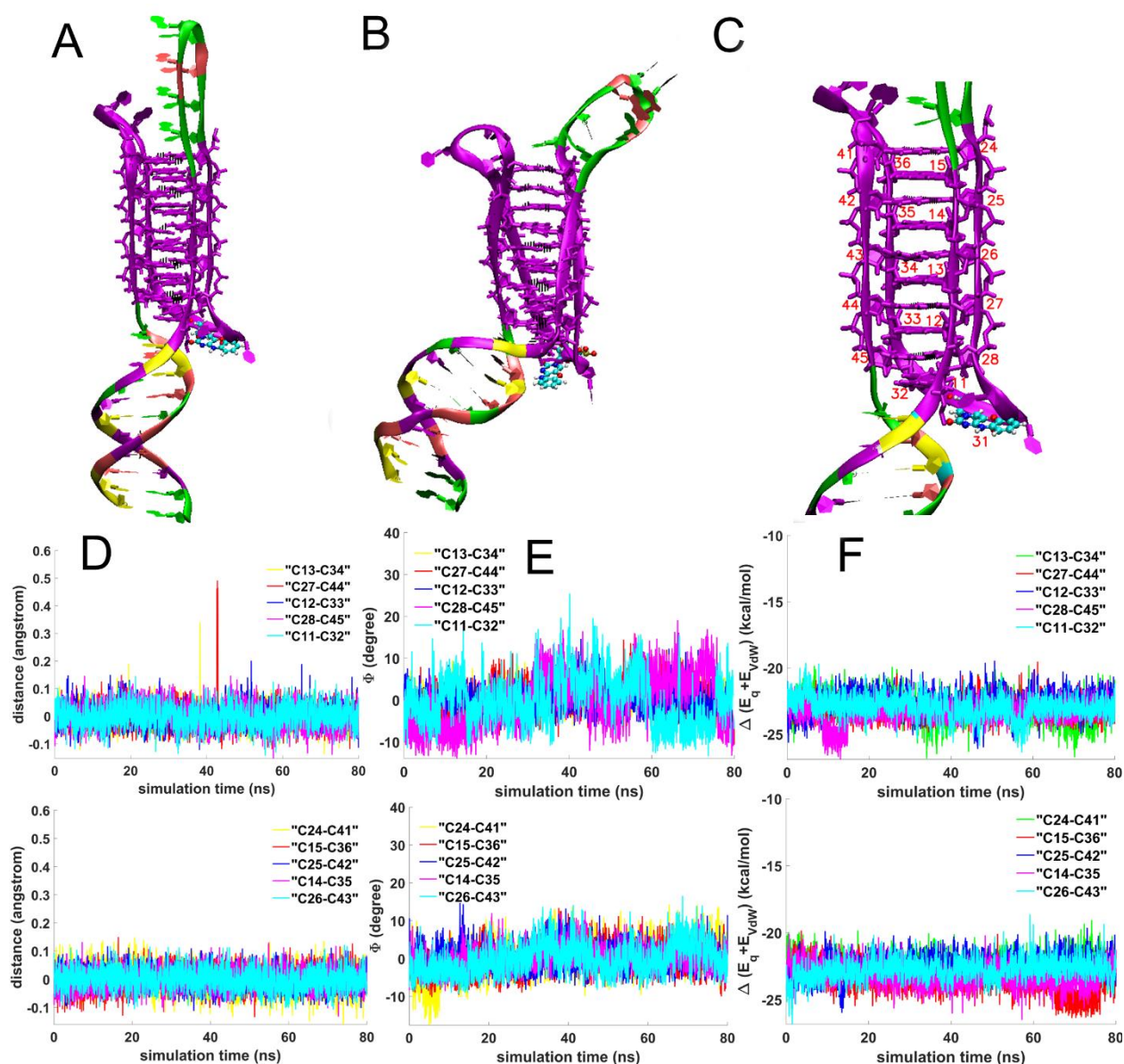
Extended Data Figure 4. Molecular dynamics simulation results of BV42 with three cytosines in the 2nd loop. Conformers of BV42 at the start (A) and at the end (B) of molecular dynamics simulation. The enlarged iM structure is shown in subset (C). Distances between mass centers of the cytosine bases (D); angles between normals to the cytosine bases (E); sum of electrostatic (E_q) and Van der Waals (E_{vdw}) of contributions to interaction energy between cytosine bases (F).



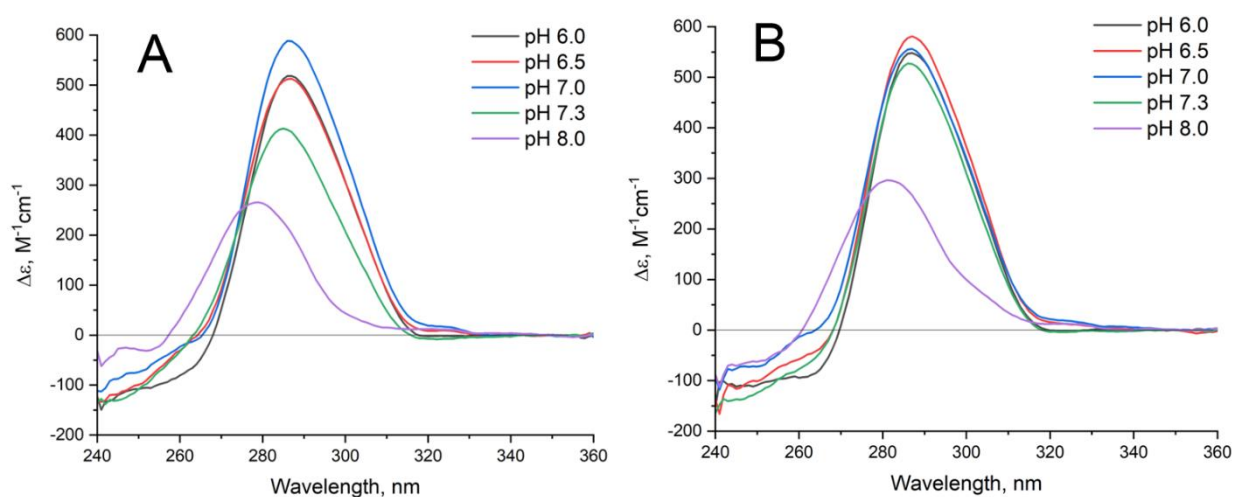
Extended Data Figure 5. Changes in coaxiality of duplex and iM modules derived from the molecular dynamics experiment with BV42 with three cytosines in the 2nd loop. The vectors studied are shown in (A). The angle between two vectors decreased down to 139° immediately and fluctuated within $\pm 15^\circ$.



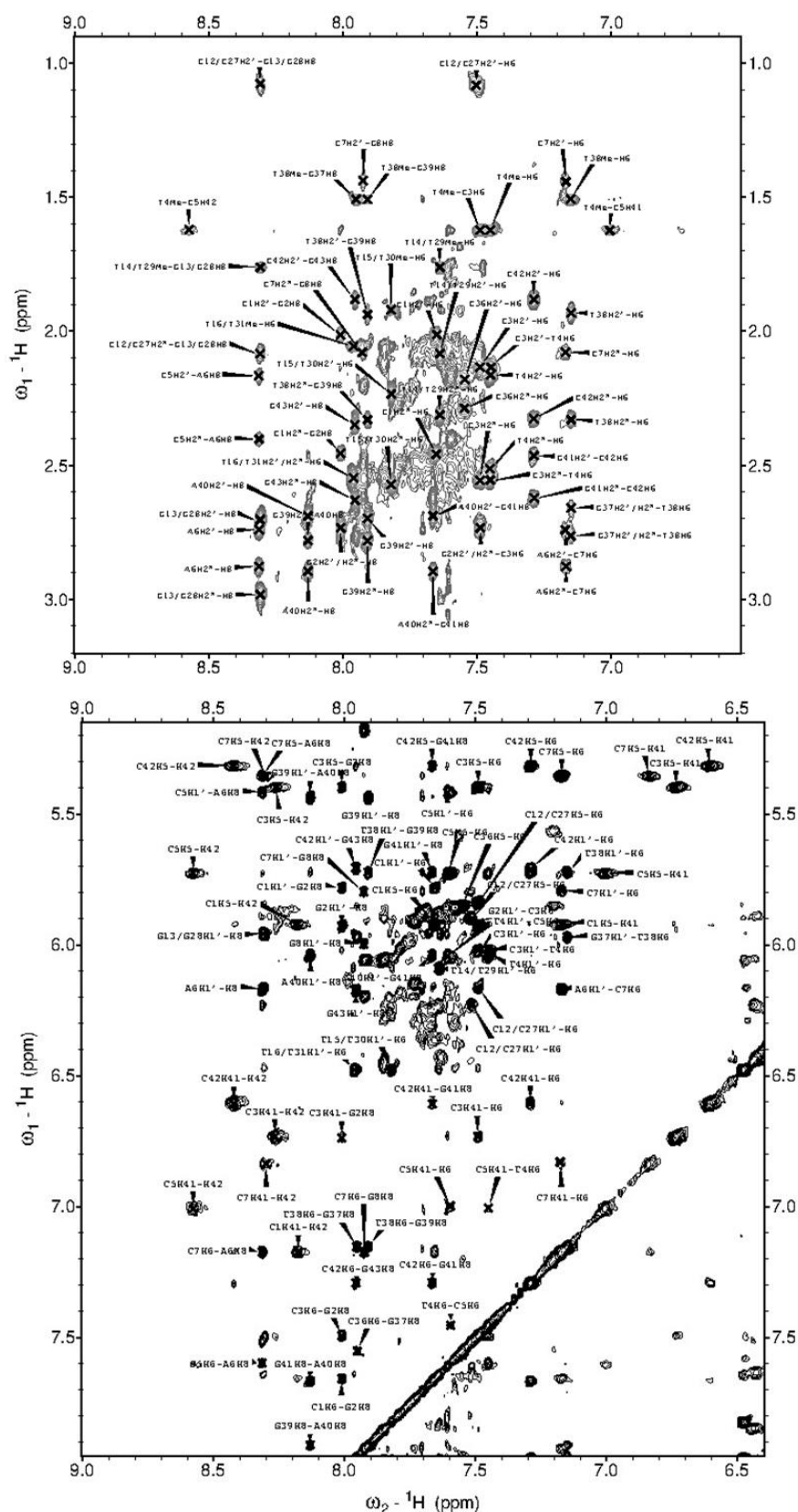
Extended Data Figure 6. ¹H NMR spectra of BV42 at different temperatures in 0.15 M phosphate-citrate buffers with pH 4.7 (A) and 6.0 (B).



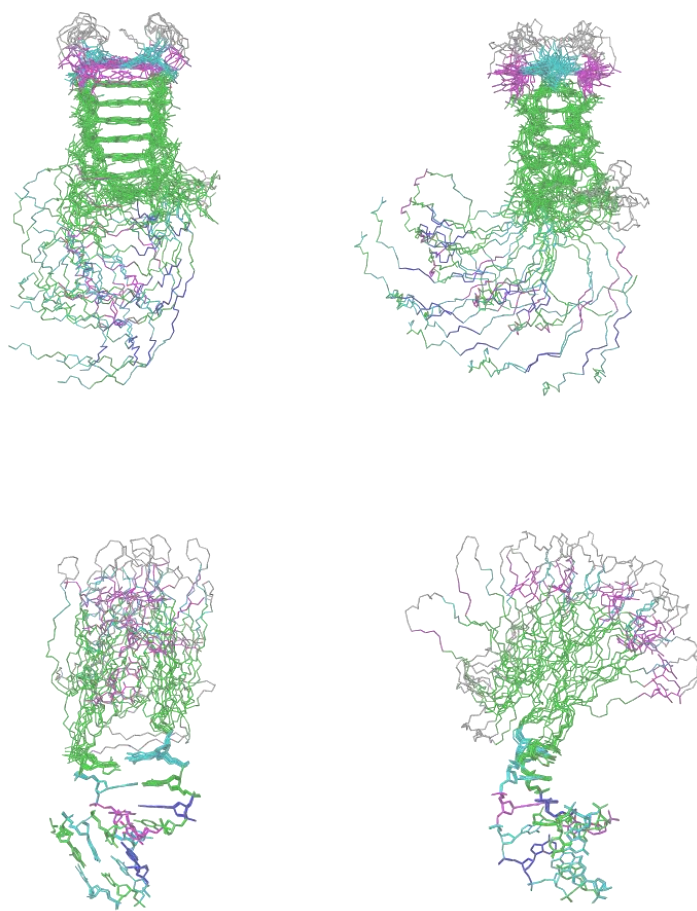
Extended Data Figure 7. Molecular dynamics simulation results of BV42°C31 with tC° in the 2nd loop. Conformers of BV42 at the start (A) and at the end (B) of the molecular dynamics simulation. The enlarged iM structure is shown in subset (C). Distances between mass centers of the cytosine bases (D); angles between normals to the cytosine bases (E); sum of electrostatic (E_q) and Van der Waals (E_{vdw}) of contributions to interaction energy between cytosine bases (F).



Extended Data Figure 8. CD spectra of CapTr5C1 (A) and CapTr5C3 (B) in PBS buffer in the pH range from 6.0 to 8.0.



Extended Data Figure 9. Non-exchangeable proton regions of the NOESY spectrum (200ms) of CapTr4C3 at pH 7 and 5°C. H₂O/D₂O 90:10, 10 mM phosphate buffer, aptamer concentration was 800 μ M.



Extended Data Figure 10. Two views of the ensemble of 10 resulting structures of CapTr4C3. Top: Superpositions considering only cytosines forming C:C⁺ pairs in the i-motif moiety. Bottom: Superpositions considering only residues involved in the duplex moiety. Color code: Thymine are shown in magenta, cytosines are shown in green, guanines are shown in cyan, adenines are shown in violet and disordered loops residues are shown in gray. Only the backbone is shown for residues not considered in the superposition.