7626–7635 Nucleic Acids Research, 2007, Vol. 35, No. 22 Published online 2 November 2007
doi:10.1093/nar/gkm922

Exploring the sequence space of a DNA aptamer using microarrays

Evaldas Katilius*, Carole Flores and Neal W. Woodbury

Center for BioOptical Nanotechnology, The Biodesign Institute and the Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-5201, USA

Received June 18, 2007; Revised September 18, 2007; Accepted October 10, 2007

ABSTRACT

The relationship between sequence and binding properties of an aptamer for immunoglobulin E (IgE) was investigated using custom DNA microarrays. Single, double and some triple mutations of the aptamer sequence were created to evaluate the importance of specific base composition on aptamer binding. The majority of the positions in the aptamer sequence were found to be immutable, with changes at these positions resulting in more than a 100-fold decrease in binding affinity. Improvements in binding were observed by altering the stem region of the aptamer, suggesting that it plays a significant role in binding. Results obtained for the various mutations were used to estimate the information content and the probability of finding a functional aptamer sequence by selection from a random library. For the IgE-binding aptamer, this probability is on the order of 10^{-10} to 10^{-9}. Results obtained for the double and triple mutations also show that there are no compensatory mutations within the space defined by those mutations. Apparently, at least for this particular aptamer, the functional sequence space can be represented as a rugged landscape with sharp peaks defined by highly constrained base compositions. This makes the rational optimization of aptamer sequences using step-wise mutagenesis approaches very challenging.

INTRODUCTION

Aptamers are short, single-stranded nucleic acids which can be selected *in vitro* to bind nearly any target, from small molecules to proteins (1–3). The relative ease of selection, and the fact that the specificity and affinity of aptamers rival that of monoclonal antibodies has led to an increasing number of analytical applications for aptamers (4–6). One such application is the creation of aptamer microarrays that have been used for protein detection with the ultimate goal of proteomic profiling of biological samples for diagnostic purposes (7).

Several approaches for creating low-density arrays (both in terms of number of probes per array and in terms of different aptamers) have been previously described. In one of the earliest reports, an array biosensor utilizing fluorescently labeled DNA- and RNA-based aptamers was used to demonstrate binding of target proteins in complex mixtures by using fluorescence anisotropy changes upon protein target binding to a surface-immobilized aptamer (8). Later reports have described approaches where fluorescently labeled proteins were used to detect binding to the aptamers arranged on a surface in microarray format. For example, DNA-based photoaptamer microarrays were created by immobilizing aptamers on slides using chemical linkage through an amino group present on the 5′ end of the aptamer. These arrays were used to detect and quantify concentrations of up to 17 different target proteins (7). Similarly, DNA aptamers which bind to human immunoglobulin E (IgE) and thrombin were used to create spotted microarrays using 3′-amino-modified sequences (9). Subsequently, more extensive studies including both DNA and RNA aptamers were performed using biotin-modified aptamers which were spotted onto the surface of streptavidin or neutravidin modified slides (10–12). In all of these cases, DNA or RNA aptamers selected employing solution-based SELEX methods were used in a microarray format to demonstrate binding of fluorescently labeled target proteins. Recently, applications of aptamer arrays using different label-free detection modalities have also been demonstrated. For example, surface plasmon resonance (SPR) imaging was used to detect protein binding to RNA aptamer microarrays (13,14). Electrochemical detection of protein target binding to arrays of aptamer-modified gold electrodes have been also demonstrated (15).

The studies described above used aptamer sequences which were chemically synthesized and then deposited on the surface of an array. However, this approach is limited in terms of the number of aptamers per microarray, both because of the requirement that the aptamers be presynthesized and by limitations of robotic

*To whom correspondence should be addressed. Tel: +1 480 727 8790; Fax: +1 480 727 0396; Email: ekatilius@asu.edu

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

printing approaches. *In situ* DNA synthesis technologies, either light-directed synthesis (Affymetrix, NimbleGen) or non-contact printing of nanoliter volumes (Agilent), allow much higher density arrays to be created. It is now possible to obtain large custom microarrays with hundreds of thousands of probes (Agilent, Nimblegen). In the present report, custom DNA microarrays have been designed and used as a means of synthesizing and analyzing variants of an IgE binding aptamer, which has previously been selected using SELEX methodology (16) (Figure 1). This has made it possible to explore the effects of aptamer sequence modification on binding properties and to see if enhancement in binding of surface-bound aptamers can be observed. The IgE-binding aptamer that served as the basis for this study was selected previously using standard, solution-phase SELEX methodology (16), and one question addressed in this work is whether further optimization for use as a surface bound molecular recognition element in a microarray format is possible. Because current technology allows *in situ* synthesis of thousands of distinct sequences in the same microarray, it is straightforward to study the effects of single, double, triple, etc. mutations of a specific aptamer. We have designed a 44 000 feature custom microarray (Agilent) that contains three copies of all possible single- and double-mutations as well as about one-third of all triple mutations of the IgE aptamer. This allows the direct exploration of the topology of aptamer sequence space.

Direct exploration of sequence space utilizing microarrays offers a unique perspective for studying protein–DNA interactions. This has been elegantly demonstrated by protein-binding microarray technology, which provides comprehensive characterization of the *in vitro* DNA-binding specificities of DNA-binding proteins (transcription factors etc.) binding to double-stranded DNA microarrays (17–20). Analysis of DNA-binding specificities using microarrays provides information about the lower affinity sequences, which is largely inaccessible using other methods for sequence space exploration, like selections from doped libraries (21,22). Here we demonstrate that microarray technology can be effectively used for direct exploration of protein-binding aptamer sequence space by providing a complete picture of high-, moderate- and low-binding sequence variants of the IgE-binding aptamer.

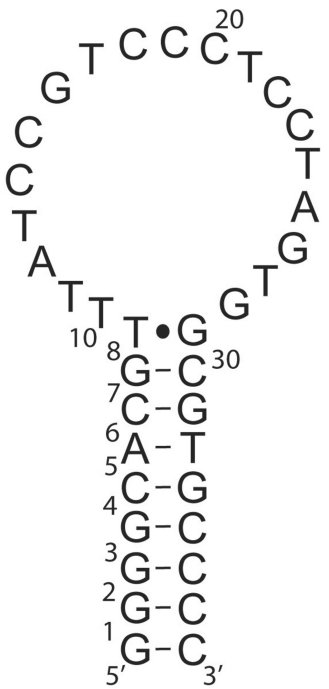


Figure 1. Predicted secondary structure of the IgE-binding aptamer used in this study. Non-Watson-Crick base pairing between T and G is denoted with a dot. The secondary structure shown was determined using the program 'mfold' (30).

MATERIALS AND METHODS

Human IgE was purchased from Athens Research (Athens, GA, USA). For binding experiments, IgE was labeled with Alexa Fluor 647 dye according to the manufacturer's protocol (Invitrogen). Typical labeling resulted in ~6–7 dye molecules per protein molecule.

Custom 44 K DNA microarrays were designed and ordered through the custom microarray program from Agilent. The array design was based on the published IgE aptamer sequence (sequence 17.4, see Figure 1). Microarrays were designed to include triplicates of each probe. Before binding of the fluorescently labeled IgE protein to the array, the microarray surface was blocked using a solution consisting of 0.2% I-Block (Applied Biosystems) and 0.1% Tween-20 in 1× PBS also containing 1 mM $MgCl_2$. Blocking was done for 1 h at room temperature and then the arrays were dried by centrifuging. Protein binding was performed using GeneFrame hybridization chambers. Fluorescently labeled protein was diluted to concentrations ranging from 1 nM to 500 nM in 1% BSA solution in 1× PBSMT (1× PBS (10 mM sodium phosphate buffer pH 7.4, 138 mM NaCl, 2.7 mM KCl) + 0.1% Tween-20 + 1 mM $MgCl_2$). Binding assays were done at 37°C overnight. After incubation, microarrays were washed three times with 1× PBSMT and three times with 1× PBSM (1× PBS + 1 mM $MgCl_2$). Then, the microarrays were dipped in nanopure H_2O to remove any remaining salt and dried by centrifuging (1500 rpm for 5 min using a swinging bucket rotor). Microarrays were imaged using Agilent's microarray scanner at 10 μm resolution. Data were extracted from images using GenePix Pro 6.0 software and results were analyzed using programs written in Matlab or Excel. Results from triplicates of each probe were used to calculate the mean fluorescence signal and standard deviation for each probe. In the majority of cases, the SD was <10% of the mean value. To evaluate reproducibility, incubations at 100 nM protein concentration have been performed three times (on three different arrays), and the array-to-array variability was also within 10% of the mean signals.

RESULTS

DNA sequences on custom DNA arrays from Agilent can be synthesized up to 60 nt in length. Since the IgE-binding aptamer is only 37 nt long (Figure 1), the initial investigation of this aptamer immobilized on the surface of the

microarray focused on establishing the optimum length of the linker separating the aptamer from the surface. Aptamer sequences were designed to contain a linker from 0 to 23 Ts on the 3′ end separating the aptamer from the surface (aptamer attachment to the surface is through the 3′ end, as the DNA strands are synthesized in the 3′–5′ direction). The resulting binding signals with increasing linker length are shown in Figure 2. These results were obtained using 100 nM IgE protein concentration, however, similar trends were observed at other protein concentrations as well. The fluorescence signal increases with increasing length of the linker nearly linearly, indicating that separation of the aptamer from the array surface enhances its ability to bind to the target protein. As the linker approaches about 20 nt in length, its effect becomes less pronounced, though it is unclear if additional linker length might further improve the binding characteristics. All further mutagenesis experiments were performed using aptamer sequences containing 23 T linkers to optimize the binding signal.

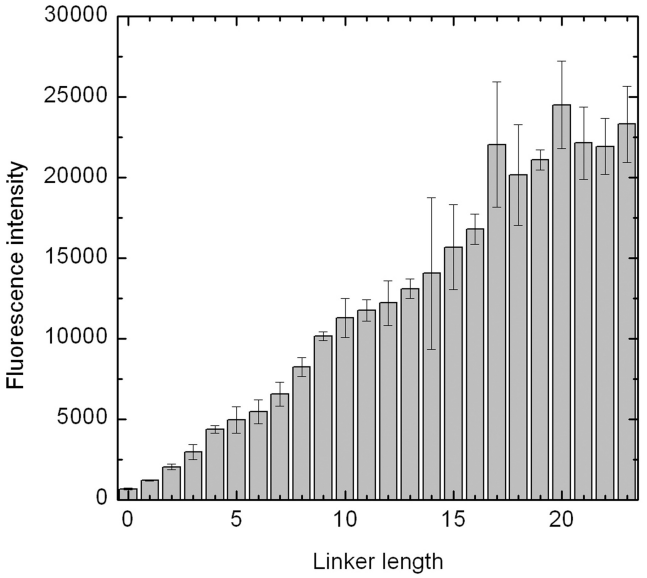


Figure 2. Fluorescence signal dependence on the T linker length. Averages of fluorescence signals from sequences present in triplicate on the array are shown. Error bars represent standard deviations of the signal.

Single IgE aptamer mutations

Sequences corresponding to all single mutations of the IgE aptamer were included on the microarray to study the importance of the specific base at each position. A summary of the results obtained at 100 nM IgE protein is presented in Figure 3. Overall, the results show that the IgE-binding aptamer sequence is extremely sensitive to single mutations, particularly in the loop region. Mutations at most positions in the loop cause nearly complete loss of activity (binding); only six bases can be mutated to other nucleotides while maintaining substantial affinity, and the majority of these mutations still result in decreased fluorescence intensity signals. Fluorescence intensity plotted as a function of protein concentration (Figure 4) shows that at 100 nM protein concentration the signal versus binding curve is still in the linear region suggesting that the fluorescence signal should qualitatively correlate with the relative sequence affinity at this concentration (the correlation between fluorescence signal and binding affinity is addressed in more detail below).

Several interesting, unexpected observations were noted for mutations introduced in the stem region of the aptamer. While changes in the stem generally resulted in decreased signal, the extent of decrease was much less on average than that observed for the deleterious mutations in the loop region. Figures 3 and 4 also show that sequences containing a mismatch mutation at the very end of the stem (either the first or second base pair) exhibit higher signals (higher affinity) compared to the original sequence. The effect of the base pair mismatch is, however, somewhat asymmetric and mutation dependent. Introduction of a mismatch at the second base pair on the 5′ end of the stem seems to have a positive effect, while the introduction of a mutation on the complementary position near the 3′ end of the stem has an adverse effect on binding. The effect of the mutation is also dependent on the specific base substitution.

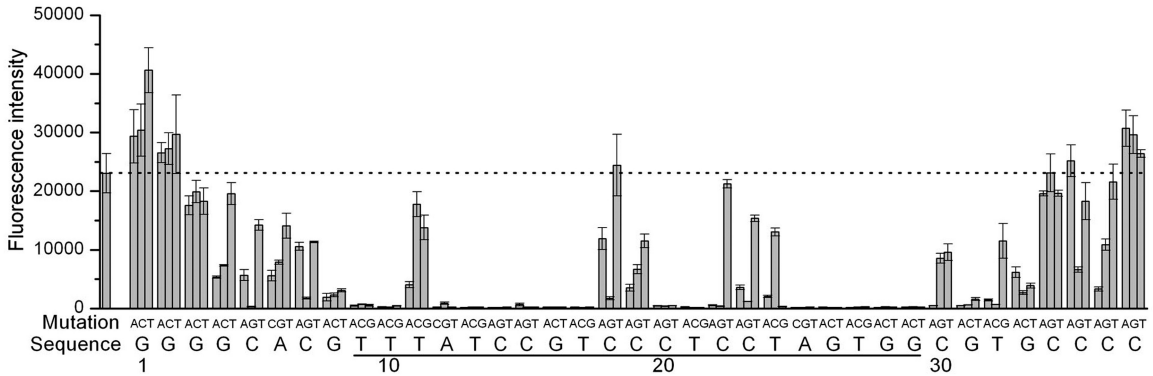


Figure 3. Results of single IgE-binding aptamer mutations. Fluorescence intensity results obtained by allowing 100 nM labeled-protein to bind to all possible sequences containing a single mutation (position shown on the *x*-axis). Error bars represent standard deviations of the signals. The first bar in figure (left side) and the dotted line represent the fluorescence signal intensity obtained for the original aptamer sequence. The loop region of the aptamer sequence (including the non-conventional T–G base pair, see Figure 1) is underlined.

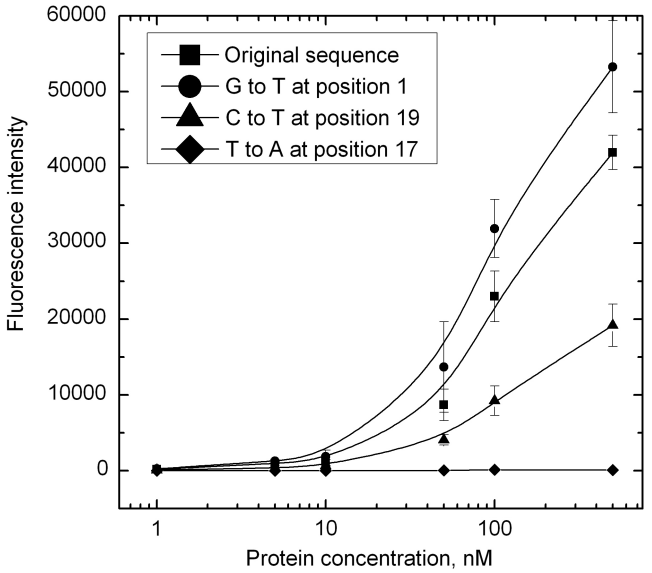


Figure 4. Protein concentration dependence of the fluorescent intensity for the original aptamer sequence (squares) and for three mutations—G mutation to T at position 1 (circles), C mutation to T at position 19 (triangles) and T mutation to A at position 17 (diamonds). Lines in the graph are B-spline interpolations of the data shown only as a guide to the eye. Error bars represent the standard deviations of the signals. Results obtained for the T to A mutation at position 17 are representative of non-specific protein binding to the control random DNA sequences.

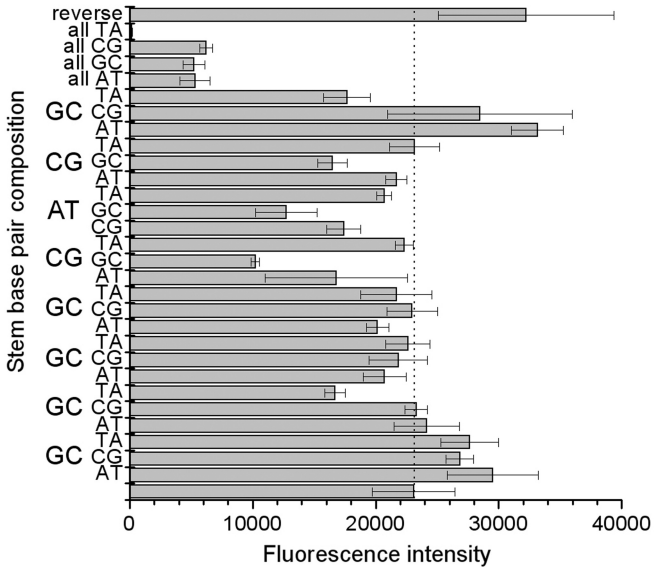


Figure 5. Results of aptamer stem composition analysis. Fluorescence results were obtained for aptamer sequences containing an altered base pair in the stem region or a different composition of the stem altogether (see text and Figure 1 for explanation). 'Reverse' represents the result where the aptamer sequence fragments corresponding to the stem region were switched (5′ versus 3′). The first bar (bottom) in the graph and the dotted line represent the fluorescence signal observed for the original aptamer sequence.

Double mutations

Given that the IgE aptamer is predicted to have a stem-loop secondary structure (Figure 1), we have created two distinct sets of double mutations. First, all possible double stem mutations that conserve the base pairing were investigated to analyze the stem's importance in binding. Another set of all possible double mutations in the loop region (spanning bases 9–29) were created and tested to investigate the effects of these mutations on binding affinity.

The results for the stem mutations are summarized in Figure 5. Overall, these findings suggest that only a small region of the stem is critical for binding. The majority of the base pair-conserving mutations do not have a significant effect on the binding signal. As seen for the single mutations, a higher intensity fluorescence signal is observed for the features containing sequences in which the first base pair in the stem (G–C) is changed to a thermodynamically less stable base pair (A–T or T–A). Interestingly, mutation of the fifth, sixth or seventh base pair in the middle of the stem to GC results in a decrease of the observed signal, suggesting that the particular composition of these base pairs is important for binding to IgE. On the other hand, mutation of the eighth base pair from GC to AT results in a significant increase in the observed fluorescence signal. The specific arrangement of bases also seems to be important as the symmetric reversal of AT to TA in this position results in a decreased signal. Another surprising result is that the aptamer containing the mirror image of the stem (i.e. the sequence on the left switched places with the sequence on the right) shows an increase in the observed fluorescence signal (corresponding to a higher binding affinity) compared to the original sequence. The results of Figure 5 also show that the particular composition of the stem is important, as aptamers with completely mutated stems (all G–C or all C–G or all A–T base pairs) show significant decreases in binding affinity. The results for all T–A base pair stems are most likely biased because of the 23 T linker separating the aptamer from the surface and the run of A's in close proximity at the 3′ end of the stem. In this case, secondary structure calculations predict a configuration that is more stable when the linker sequence folds to form a small all-T loop with the A–T stem. The existence of this stable secondary structure likely interferes with the proper aptamer structure required for binding to IgE.

As the single mutation results described above suggest, little variation is possible in the specific aptamer sequence of the loop region if function is to be maintained. However, in principle, a mutation at one position might be compensated by a mutation at another position. To test this, sequences containing all possible double mutations in the loop region (starting with the non-conventional TG base pair) were included on the microarray to investigate the interplay between mutations within the loop. The results complement the results of the single mutations, showing that most of the positions in the aptamer sequence are critical for binding (Figure 6). Only few positions can be mutated in concert while still maintaining a significant binding affinity. In particular, double mutations including mutations in the positions 11, 18, 19, 22, 23 and 24 retain aptamer activity, although with decreased affinity (Figure 6). These mutations are at the same

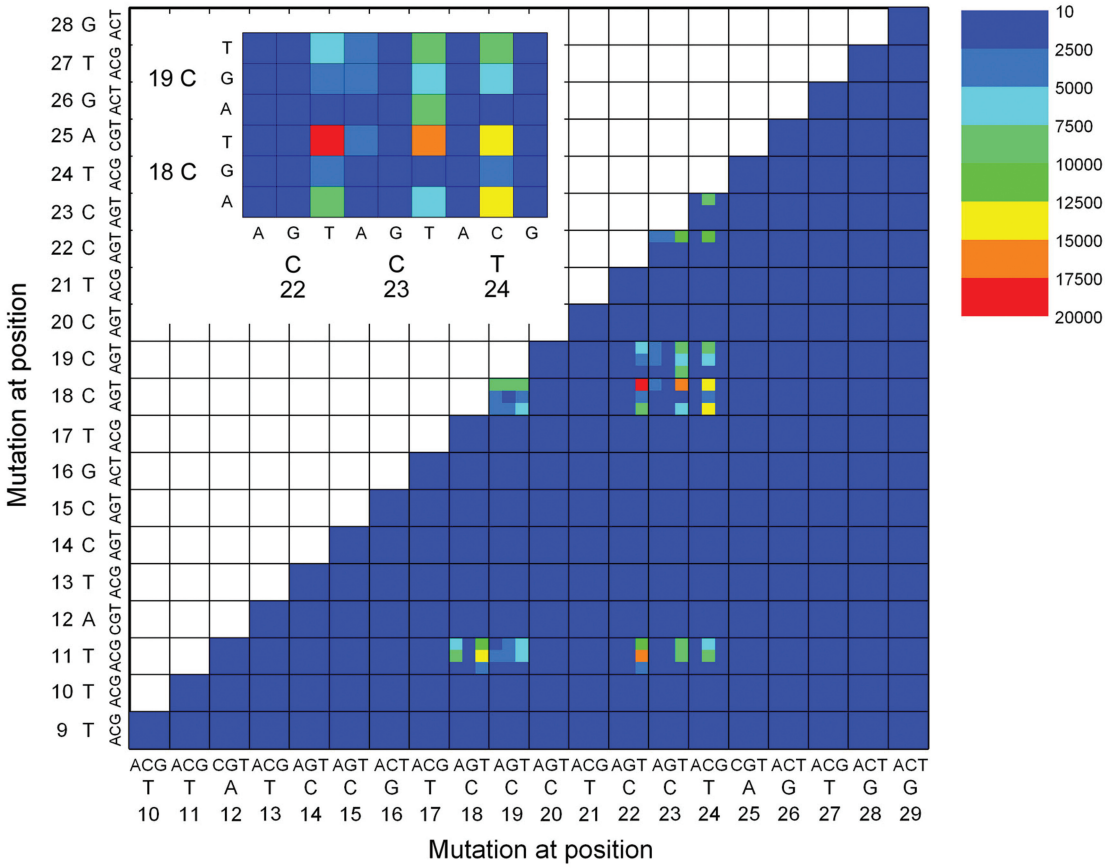


Figure 6. Fluorescence intensity results obtained for different double mutations of the IgE-binding aptamer loop region. Inset shows the results of double mutations at the 18–19 and 22–24 positions in more detail. Fluorescence signal intensities were obtained using 100 nM labeled IgE.

positions where single mutations can be present, suggesting that, rather than being compensating mutations, they are just two mutations at non-critical positions (positions that are not important for forming the tertiary structure or specific contacts with the protein in the bound configuration). Generally, no compensatory activity is observed, i.e. any combination of two mutations results in a fluorescence signal that is less than the signals obtained for each mutation individually. The two notable exceptions are combinations including mutations at position 18. In one case, combining a C to T mutation at position 18 with a C to T mutation at position 22 results in a fluorescence signal that is within a standard deviation of the signals for each of the individual mutations as well as the original aptamer sequence. In another case, mutating C to A at position 18 and T to C at position 24 results in the fluorescence signal that is larger than the signals for either of the individual mutations: ~15 000 counts for the double mutant versus ~12 000 and ~13 000 counts for the individual mutations, respectively (signals at 100 nM protein concentration).

Triple mutations

A complete set of triple mutations of the IgE binding aptamer was not possible within the constraints of the 44 000 feature microarray used for this experiment. Thus, three distinct sets of triple mutations were screened to evaluate at least a limited portion of this sequence space. A set of all possible triple mutations involving purine to purine (A↔G) or pyrimidine to pyrimidine (C↔T) changes in the loop region was created (this set contains 1330 mutations). Another set of triple mutations was created by combining the mutation C to T at position 18 with all other double mutations in the loop region (from base 9 to 29, maintaining T in the position 18). As noticed above, the mutated aptamer sequence containing T instead of C at the position 18 shows a slight increase in the fluorescence signal suggesting higher binding affinity. Combination of this mutation with several other mutations also maintains or increases binding signal (see double mutant results). Thus, a set of triple mutations holding T at position 18 was created to investigate if an improved aptamer could be obtained by addition of other mutations to this one. Lastly, the complete set of all possible triple mutations (35 910 mutations covering loop region from position 9 to 29 in the aptamer sequence) was created by computation and 8913 of these triple mutants were chosen randomly from the complete set as a way of sampling the entire triple mutant sequence space. (This number of mutations was selected to completely fill the microarray; mutated sequences were all present in triplicate.)

Triple purine–purine or pyrimidine–pyrimidine mutations. The results for these mutations essentially match the results of the double mutations. Only combinations of the mutations in positions 11, 18, 19, 22, 23 and 24 are possible while still maintaining significant binding signal (only 20 out of 1330 of mutations). Combinations of mutations at any other positions, even when containing two of the mutations in the above mentioned positions, results in a decrease in fluorescence signal to near background levels, indicating that these sequences do not exhibit significant affinity for the target protein.

Triple mutations containing 18T. Analysis of results for the set of triple mutations containing 18T instead of C also coincide with the results obtained for double and other triple mutations—only mutations in the positions 11, 19, 22, 23 or 24 in addition to a mutation of C to T at position 18 are possible while maintaining reasonable affinity. Only one of the combinations of mutations containing 18T, substitution of A for C at position 19 and T for C at position 23, shows a signal close to that of the original aptamer sequence (~20 000 versus about 23 000 as obtained using 100 nM of labeled target protein). All other combinations of mutations resulted in at least two-fold reduction in signal compared to the original aptamer sequence.

Random triple mutations. Results obtained for nearly 9000 triple mutations which were randomly selected from the set of all possible triple mutations in the loop region confirm the results obtained for the double and other triple mutations. Once again, results show that only combinations of the mutations in positions 11, 18, 19, 22, 23 and 24 still exhibit significant binding signal. Any other combinations containing mutations in other positions result in a decrease of the fluorescence signals to a level close to the background.

Comparison of array data to independently determined K_d values

The original IgE-binding aptamer and three of the variants with either stronger (G to T mutation at position 1), similar (C to T mutation at position 19) or much weaker (T to A mutation at position 17) apparent affinities were synthesized and their K_d values were determined using SPR and/or fluorescence anisotropy methods (results are summarized in Table 1, see also Supplementary Data). SPR experiments were performed using a Biacore instrument, where IgE was immobilized on the surface of a gold chip and different aptamers were flowed over it. The fluorescence anisotropy experiments utilized Texas Red labeled aptamers binding to IgE in solution, similarly to previously published work (23). The fluorescence values observed on the array for this set of aptamer variants using 100 nM labeled IgE corresponds qualitatively to the binding constants determined, suggesting that the level of fluorescence on the array is a reasonable qualitative measure of affinity.

DISCUSSION

Nucleic acid aptamers have been rapidly gaining popularity in various bioanalytical techniques where they have been effectively used to replace antibodies. One of the main reasons for this is the fact that aptamers can be easily synthesized and modified using well-established nucleic acid chemistries. The specific nucleic acid sequence determines how the aptamer is folded, which in turn determines the binding affinity. Thus, investigation of the relationship between sequence, structure and binding affinity is important for a more complete understanding of the biophysical aspects of aptamer–ligand binding. A number of studies of aptamer sequence space have been performed previously utilizing a variety of techniques directed to either chemically modify the aptamer sequences (24–26) or to select the best variants from the doped aptamer libraries (21,22,27). Chemical modification studies, like footprinting using various cleavage agents have been instrumental in shedding some light onto the structural aspects of aptamer binding to proteins, where X-ray or NMR structural data is not available. Chemical cleavage or modification of an aptamer's sequence provides some limited information about the parts of the sequence that are in close proximity (or make contact with) the target and are therefore protected from modification. In combination with the secondary structure predictions, these studies provide some information about the possible structural motifs critical in effective binding of an aptamer to its target (24,25).

Selections of aptamers from doped libraries, where the specific positions or regions in the sequence have been modified to contain an unequal distribution of nucleotides, have been very informative in determining the best consensus sequences for a variety of aptamers as well as determining base distribution, i.e. information content of the sequence (27). However, selection of aptamers from doped libraries has limitations. This process of selection is inherently targeted toward the selection of the highest affinity sequences. Information about these sequences is obtained after cloning and sequencing a limited number

Table 1. Comparison of fluorescence signals to independently measured dissociation constants for different aptamer variants

Aptamer	Fluorescence signal at 100 nM protein	K_D from anisotropy, nM	K_D from SPR, nM	K_a, $M^{-1}s^{-1}$	K_d, s^{-1}
Original sequence	23 000	15 ± 4	4.7	4.3×10^5	0.002
Mutation G to T at position 1	32 000	7 ± 2	4.1	5.6×10^5	0.0023
Mutation C to T at position 19	9200	19 ± 2	7	3×10^5	0.0021
Mutation T to A at position 17	140	450 ± 30	ND	ND	ND

K_a and K_d are association and dissociation rate constants, respectively, fitted from the SPR data using a 1:1 binding model; ND – not determined.

of clones after several rounds of selection (or sometimes in between the rounds). However, no information is usually obtained for the sequences, which do not survive the selection process. Thus, the effects of specific mutations on the affinity of the aptamer can be only indirectly inferred from the consensus sequence distributions. Thus, in general, library selection methods provide only limited information about the degree of affinity over the topology of local sequence space.

We have approached the study of aptamer sequence space using a different method. Instead of performing a mutagenesis and selection study as described previously (27), we have evaluated the properties of a well-characterized IgE-binding DNA aptamer by utilizing custom DNA microarray technology. As mentioned above, several companies currently offer synthesis of custom microarrays containing 60-mer (Agilent) or even 80-mer (Nimblegen) DNA sequences in a microarray format. This length is sufficient for studying most of the known DNA aptamers. Current DNA microarray production technology allows *in situ* synthesis of up to hundreds of thousands of sequences on the microarray surface (a number which is projected to expand to several million in the near future), resulting in a considerable capability to study sequence variations. By specifically defining the sequence at each feature of the microarray it is possible to systematically study the topology of local aptamer sequence space with respect to binding of the target protein. This approach provides more detailed information about the relative affinity of sequences that do not bind the target or bind it at with intermediate affinity. This information is normally largely inaccessible from the results of clone sequencing and consensus sequence analysis, as only the best binding sequences are analyzed after selection.

The relationship between array fluorescence values and affinities

For microarray data to be useful in the analysis of binding as a function of sequence space, fluorescence signals obtained from microarray experiments must correlate to the relative affinity of sequences. It is clear from Figure 4 that measurements in the 100 nM range of the target protein are still within the linear range of the binding curve and thus should be representative of relative binding affinity under the conditions of the measurement. However, the solution phase dissociation constant for the interaction between IgE and this aptamer is ~10 nM as determined by a filter binding assay (16). In contrast, in Figure 4, the apparent K_D for the aptamer on the surface is on the order of several hundred nanomolar (it is difficult to estimate accurately as saturation has not been reached even at 500 nM protein concentration). Previously published results using aptamer microarrays prepared using synthesized aptamers have generally shown binding at protein concentrations more consistent with solution phase dissociation constants (9,11,12). The large apparent K_d for IgE binding to its aptamer on the surface could arise from one of several factors. First, the surface itself can affect aptamer function. It is quite possible that there are interactions between the aptamer and the surface of the slide in the case of the arrays being used here that effectively compete with the IgE interaction (the strong influence of linker length on aptamer affinity in Figure 2 suggests this). Second, the binding protocol was optimized for the best signal to background ratios on these particular arrays. This involved both specific incubation times, temperatures and buffer conditions as well as the presence of blocking agents to decrease non-specific binding. These conditions may favor specificity of binding, but change the binding kinetics or dissociation constant significantly. Past studies of aptamer binding to IgE (9,12) have been performed by incubating the arrays with the target protein at room temperature instead of at 37 C as has been done here. Solution affinities determined for the IgE-binding aptamer using fluorescence anisotropy resulted in K_D values that increased significantly with temperature (23). This effect could also be exaggerated when binding occurs at the surface, depending on the nature of the surface interactions. Finally, the dye labeling of the IgE, particularly with multiple labels per protein at random amines, could change the binding characteristics of the protein as well as increase non-specific binding, as has been demonstrated previously (10).

In order to empirically determine if the fluorescence signals from the microarray measurements corresponded to relative affinities, more traditional methods of evaluating affinity were applied to the original aptamer sequence and several of the variants identified on the microarrays (Table 1). Even though the apparent dissociation constant for the original IgE-binding aptamer on the surface is considerably larger than that previously seen in solution, the relative changes in fluorescence intensities obtained for different sequences correlate very well with the dissociation constants obtained for several different aptamer variants (Table 1 and Supplementary Data) using either anisotropy measurements or SPR analysis. Thus, as a qualitative measure of binding affinity, the fluorescence values determined in the array are apparently valid.

Overall structural requirements

The IgE-binding aptamer has been previously selected from a DNA library containing a sequence of 40 random nucleotides using standard SELEX methods (16). After 15 rounds of selection and the sequencing of 87 clones, it was shown that a highly conserved 21 nt long sequence was responsible for high-affinity binding. It was proposed that this conserved sequence could be folded into a stem-loop secondary structure having an unstable 4 bp stem and a 12 base loop. Extension of the unstable stem by flanking complementary sequences was then shown to stabilize the aptamer structure and improve binding affinity. At the same time, variations in the sequences obtained from the clone sequencing analysis suggested that the sequence of the stem was not critical for binding. Consensus sequence analysis also indicated that only a few positions in the aptamer sequence can be varied (in most cases either C or T were present in the clone sequences at the positions 11, 18, 19, 23 and 24 using the sequence numbering of the D17.4 aptamer) (16).

Overall, the results presented here complement the results from the original selection paper. Analysis of the single-site mutants shows that the specific sequence of the aptamer is critical in determining the aptamer's ability to bind to IgE. Mutations in the positions mentioned above: 11, 18, 19, 23 and 24, are possible while maintaining significant affinity to the target protein. In addition to that, position 22 can also be mutated from C to T without detrimental effects. No other single mutations are possible without significant decreases in affinity. Through systematic analysis of the effects of particular mutations on binding, additional conclusions can be drawn that would not be possible to obtain from library selections without sequencing and analyzing a large number of potential aptamers. For example, consensus sequence analysis in previous work showed that it is possible for either C or T to be present in position 11 and retain binding capability. The results of the current study show that G can also be substituted in this position with nearly the same resulting affinity. Thus, one useful aspect of array analysis is that it can provide a complete picture of all possible equivalent aptamers in local sequence space.

The stem region

Previous studies have suggested that the stem region is not critical for binding (16). Our microarray analysis of the IgE aptamer stem region indicates that some modifications may actually enhance binding (a conclusion that is supported by more traditional affinity measurements for position 1, Table 1). Single-mutation analysis shows that having a mutation in the first two base pairs of the 5′ end of the stem results in an increase in the detected fluorescence signal, implying better binding. This result suggests that having a less energetically stable stem is beneficial for the aptamer function. A similar conclusion is also implied by the results of stem mutagenesis when base pairing in the stem region is conserved. Results presented in Figure 5 show that having a less stable A–T or T–A base pair at the beginning of the stem results in the improvement of detected signal. Based on the consensus sequence obtained from the selection results, it was suggested that the base pair composition in the stem region is important only as a stabilizing structural factor (16). However, the results presented here show that the particular base pair composition of the stem region is important for obtaining the best binding signal. For example, having a stem with all GC or all CG base pairs results in a several-fold decrease in the detected signal, i.e. apparent aptamer affinity. As shown in Figure 5, changes in the base pairs 5 through 8 (which would correspond to the initial stem region just below the loop in the consensus sequence and proposed secondary structure) have quite significant effects on the observed binding signal. Both decreased and increased signals were obtained, implying that optimization of the aptamer sequence in the stem region is possible. Finally, even the symmetry appears to be important; an aptamer sequence containing the reversed stem, wherein the sequences that comprise the 5′ and 3′ stem regions have been exchanged, appear to give rise to a somewhat higher signal than the original sequence. All these results emphasize that the detailed structure of the stem region plays an important role in binding of the target protein.

Information content of aptamer sequence

Previously, the information content of nucleic acid protein-binding sites has been quantified utilizing the Shannon uncertainty measure, which is calculated by determining the probability of finding a specific nucleotide at a particular position (28,29): $H = -\sum P_i \log_2 P_i$, where H is the uncertainty of the particular position, P_i is the probability of finding a specific nucleotide at that position and i represents the four nucleotides (i = A,T,C,G). The probability P_i can be estimated from the frequency of occurrence of the particular nucleotide (i) at a position upon analysis of a number of sequences. Using this approach, the information content of RNA aptamer sequences selected to bind GTP has been calculated (27). In that work, an extensive mutagenesis and selection study was performed to evaluate the effects of mutations in each sequence position as well as to quantify the information content of different aptamers with distinct sequences (also secondary structures) and different affinities for the target molecule. It was concluded that a 10-fold increase in aptamer affinity requires the addition of ~10 additional bits of information, which corresponds to specifying the identity of five additional nucleotide positions. This change in information content also results in about a thousand-fold decrease in abundance of functional sequences within a random set of sequences (27).

The results presented here do not directly provide an estimate of the frequency of a specific nucleotide at a particular position in the aptamer sequence. However, the microarray results for the aptamer sequence variants can be used to estimate the probability of finding a specific nucleotide in the aptamer sequence. For example, Figure 3 shows that aptamer sequences with either C or T at the position 22 show about the same fluorescence intensity signals representing about the same affinity, thus, one can conclude that the probabilities of finding these two nucleotides at that position are about 50% in a functional aptamer, while the probabilities of finding either A or G at this position are essentially zero (the actual probabilities for each nucleotide can be more accurately estimated based on the observed binding signals). From this follows that the information content at this position is about 1 bit. Using this approach, the information content for each position in the loop region of the aptamer sequence was calculated (Figure 7). The information content in the stem region could not be unequivocally evaluated because all possible stem mutations were not tested, making it difficult to assess the importance (and complete information content) of the stem region.

Total information content of the aptamer sequence can be used to evaluate the probability of finding a functional aptamer sequence in random space. The total information content in the loop region of the IgE-binding aptamer is ~29 bits. Thus, the probability of finding a functional aptamer sequence is $2^{-29} \approx 2 \times 10^{-9}$. Taking into account the fact that stem region of the aptamer sequence

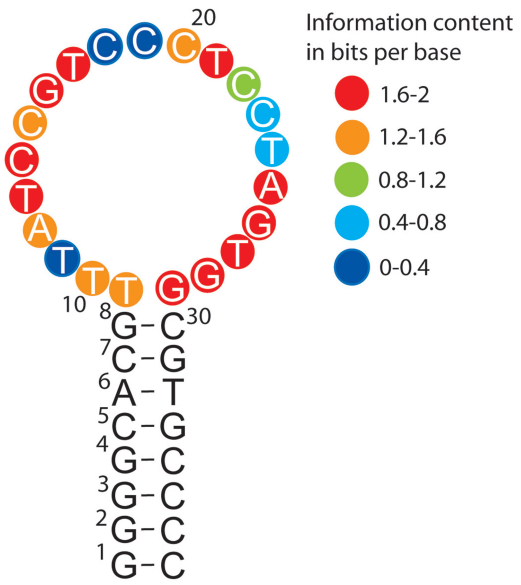


Figure 7. Information content of loop positions in the IgE-binding aptamer. The amount of information is represented as bits per base, positions are color-coded based on the simplified color scheme shown in the figure legend.

probably adds several bits to the total, it is expected that the actual probability is on the order of 10^{-10}.

Topology of functional sequence space

The microarray results presented here also provide a useful picture of the local topology of functional sequence space for the IgE-binding aptamer. One of the striking aspects of Figures 3 and 6 is that very few positions in the sequence can be mutated while maintaining binding affinity similar to the original sequence. Most changes, even of single bases, result in a drop in fluorescence signal essentially to baseline level. The absence of high-affinity compensatory variants in the double and triple mutant studies suggests that the functional sequence space of this aptamer is very constrained. In other words, the local sequence space encompassing the functional aptamer sequences looks more like a rugged landscape containing sharp peaks (Figure 6) rather then a gradually changing surface. This implies that rational optimization of the aptamer functionality through step-wise mutagenesis is essentially impossible. Further studies of other aptamer sequences are required to see how general this observation is.

This study of DNA aptamer sequence space shows yet another application for DNA microarrays. We have demonstrated that microarrays can be effectively used to study a wider range of DNA-protein binding interactions, not limited to proteins which bind to double-stranded DNA as previously demonstrated by protein binding to double-stranded DNA microarrays. It can be anticipated that this methodology will become a useful tool for future investigations of the structure/function properties of aptamers. It can also be expected that in the future microarray technologies can be extended to novel chemistries (expansion in the choice of modified or non-natural nucleotides, different coupling chemistries, etc.) which might provide new information about the protein–DNA interactions and which should prove useful for post-selection modifications and optimizations of aptamers to improve their stability or selectivity under specific conditions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was funded by Arizona University System Technology and Research Initiative Fund and the Biodesign Institute at Arizona State University. We would like to thank Dr Scott Bingham and Dr Jeffery Hock for their help with the Agilent microarray scanner. Funding to pay the Open Access publication charges for this article was provided by Arizona University System Technology and Research Initiative fund.

Conflict of interest statement. None declared.

REFERENCES

1. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
3. Gold,L., Polisky,B., Uhlenbeck,O. and Yarus,M. (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.*, **64**, 763–797.
4. Hamula,C.L.A., Guthrie,J.W., Zhang,H., Li,X. and Le,X.C. (2006) Selection and analytical applications of aptamers. *Trends Anal. Chem.*, **25**, 681–691.
5. Bunka,D.H. and Stockley,P.G. (2006) Aptamers come of age – at last. *Nat. Rev. Microbiol.*, **4**, 588–596.
6. Tombelli,S., Minunni,M. and Mascini,M. (2005) Analytical applications of aptamers. *Biosens. Bioelectron.*, **20**, 2424–2434.
7. Bock,C., Coleman,M., Collins,B., Davis,J., Foulds,G., Gold,L., Greef,C., Heil,J., Heilig,J.S. *et al.* (2004) Photoaptamer arrays applied to multiplexed proteomic analysis. *Proteomics*, **4**, 609–618.
8. McCauley,T.G., Hamaguchi,N. and Stanton,M. (2003) Aptamer-based biosensor arrays for detection and quantification of biological macromolecules. *Anal. Biochem.*, **319**, 244–250.
9. Stadtherr,K., Wolf,H. and Lindner,P. (2005) An aptamer-based protein biochip. *Anal. Chem.*, **77**, 3437–3443.
10. Collett,J.R., Cho,E.J., Lee,J.F., Levy,M., Hood,A.J., Wan,C. and Ellington,A.D. (2005) Functional RNA microarrays for high-throughput screening of antiprotein aptamers. *Anal. Biochem.*, **338**, 113–123.
11. Collett,J.R., Cho,E.J. and Ellington,A.D. (2005) Production and processing of aptamer microarrays. *Methods*, **37**, 4–15.
12. Cho,E.J., Collett,J.R., Szafranska,A.E. and Ellington,A.D. (2006) Optimization of aptamer microarray technology for multiple protein targets. *Analytica. Chimica. Acta*, **564**, 82–90.
13. Li,Y., Lee,H.J. and Corn,R.M. (2006) Fabrication and characterization of RNA aptamer microarrays for the study of protein–aptamer interactions with SPR imaging. *Nucleic Acids Res.*, **34**, 6416–6424.
14. Li,Y., Lee,H.J. and Corn,R.M. (2007) Detection of protein biomarkers using RNA aptamer microarrays and enzymatically amplified surface plasmon resonance imaging. *Anal. Chem.*, **79**, 1082–1088.
15. Xu,D., Xu,D., Yu,X., Liu,Z., He,W. and Ma,Z. (2005) Label-free electrochemical detection for aptamer-based array electrodes. *Anal. Chem.*, **77**, 5107–5113.

16. Wiegand,T.W., Williams,P.B., Dreskin,S.C., Jouvin,M.H., Kinet,J.P. and Tasset,D. (1996) High-affinity oligonucleotide ligands to human IgE inhibit binding to Fc epsilon receptor I. *J. Immunol.*, **157**, 221–230.
17. Bulyk,M.L., Huang,X., Choo,Y. and Church,G.M. (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc. Natl Acad. Sci. USA*, **98**, 7158–7163.
18. Berger,M.F., Philippakis,A.A., Qureshi,A.M., He,F.S., Estep,P.W., III and Bulyk,M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.
19. Berger,M.F. and Bulyk,M.L. (2006) Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods Mol. Biol.*, **338**, 245–260.
20. Bulyk,M.L. (2006) DNA microarray technologies for measuring protein–DNA interactions. *Curr. Opin. Biotechnol.*, **17**, 422–430.
21. Bartel,D.P., Zapp,M.L., Green,M.R. and Szostak,J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, **67**, 529–536.
22. Conrad,R.C., Baskerville,S. and Ellington,A.D. (1995) *In vitro* selection methodologies to probe RNA function and structure. *Mol. Divers*, **1**, 69–78.
23. Gokulrangan,G., Unruh,J.R., Holub,D.F., Ingram,B., Johnson,C.K. and Wilson,G.S. (2005) DNA aptamer-based bioanalysis of IgE by fluorescence anisotropy. *Anal. Chem.*, **77**, 1963–1970.
24. Jensen,K.B., Green,L., MacDougal-Waugh,S. and Tuerk,C. (1994) Characterization of an in vitro-selected RNA ligand to the HIV-1 Rev protein. *J. Mol. Biol.*, **235**, 237–247.
25. Green,L., Waugh,S., Binkley,J.P., Hostomska,Z., Hostomsky,Z. and Tuerk,C. (1995) Comprehensive chemical modification interference and nucleotide substitution analysis of an RNA pseudoknot inhibitor to HIV-1 reverse transcriptase. *J. Mol. Biol.*, **247**, 60–68.
26. Burgstaller,P., Kochoyan,M. and Famulok,M. (1995) Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding. *Nucleic Acids Res.*, **23**, 4769–4776.
27. Carothers,J.M., Oestreich,S.C., Davis,J.H. and Szostak,J.W. (2004) Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, **126**, 5130–5137.
28. Schneider,T.D., Stormo,G.D., Gold,L. and Ehrenfeucht,A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, **188**, 415–431.
29. Stormo,G.D. (1988) Computer methods for analyzing sequence recognition of nucleic acids. *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 241–263.
30. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

Supplementary data

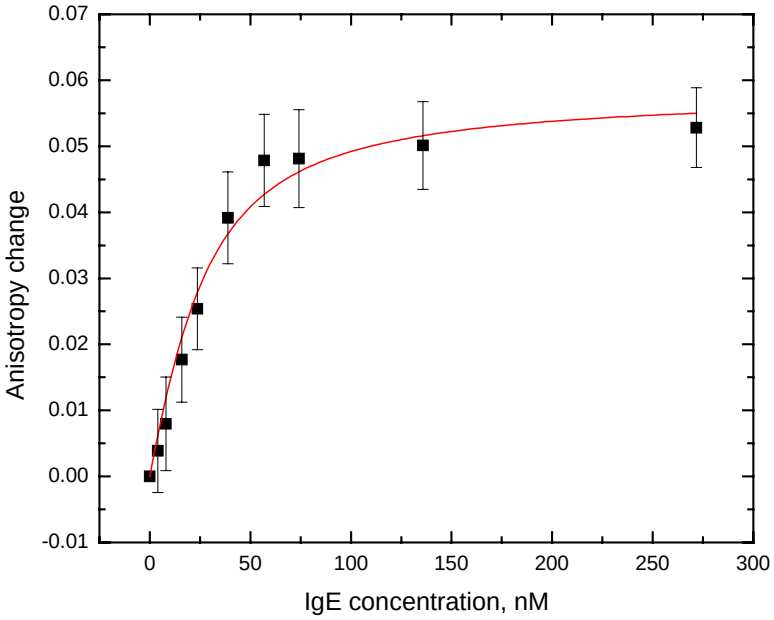
Fluorescence anisotropy measurements

Texas Red fluorescently labeled aptamer sequences: The original sequence (sequence 17.4), and three mutants (G to T mutation at position 1, C mutation to T at position 19 and T mutation to A at position 17) were purchased from IDT DNA Technologies. The labeled DNA molecules had been purified by denaturing poly-acrylamide gel electrophoresis . Stock solutions were made in water and later diluted in 1X PBSM buffer (1X phosphate buffered saline with 1 mM $MgCl_2$) at 20 nM for the fluorescence measurements. The anisotropy measurements were performed using a fluorimeter (Photon Technology International) with the excitation set at 590 nm and emission at 615 nm. Both the excitation and emission monochromators were set at 4 nm spectral resolution. These measurements were carried out using time-based detection, with the fluorescence intensities for vertical excitation/vertical emission and vertical excitation/horizontal emission measured every second. Sixty points at each polarization were aquired, and then mean value for the anisotropy was calculated. The G factor was also evaluated using the same acquisition settings. All measurements were performed using a 50 μl nominal volume microcuvette (Starna Cells). 100 μL solution of 20 nM labeled aptamer was prepared in the cuvette, an initial anisotropy measurement was taken, and then protein was added in small amounts straight to the cuvette. After each addition of protein, the sample was gently mixed and allowed to equilibrate for several minutes before measurements were taken. All measurements were performed at room temperature.

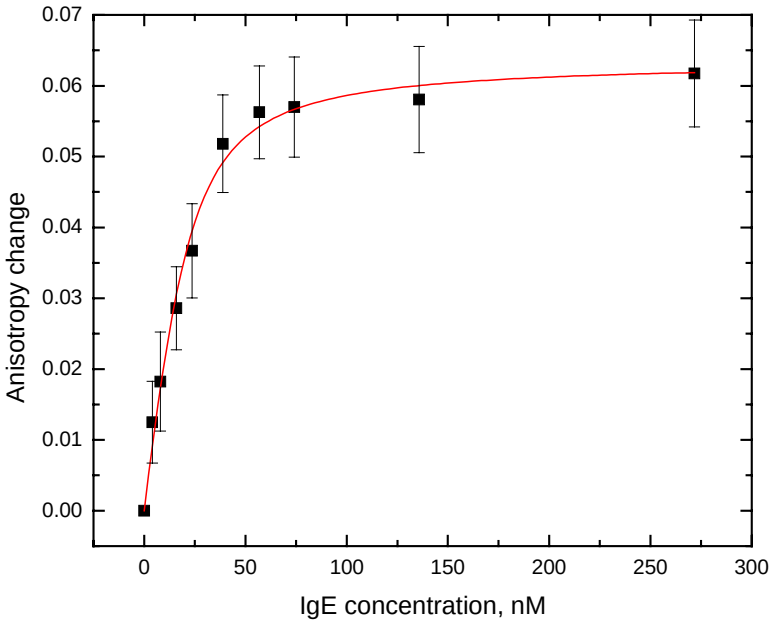
Fitting of the results was performed using the non-linear least squares optimization function in Origin 7.5 (OriginLab). Assuming 1:1 binding model, the anisotropy increase with increasing protein concentration was fitted to the expression:

$$\Delta r = \alpha \cdot \frac{A_0 + P + K_d - \sqrt{\left(A_0 + P + K_d\right)^2 - 4 \cdot A_0 P}}{2}.$$

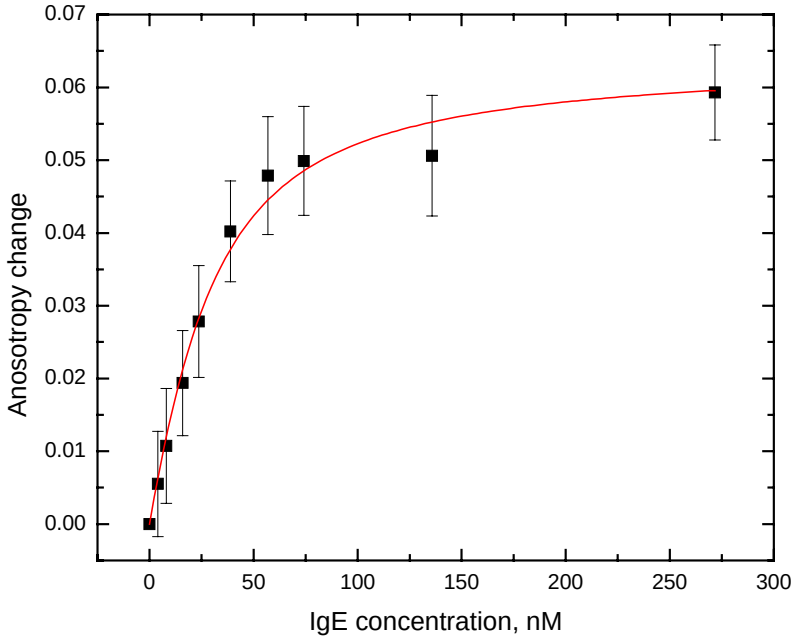
Here Δr is the anisotropy change, α is the parameter proportional to the difference in anisotropies of the aptamer-protein complex and free aptamer in solution, A_0 is the initial concentration of aptamer (20 nM), P is the concentration of protein added, and K_d is the dissociation constant for the aptamer-protein complex dissociation. During the fitting, only α and K_d are varied. Assuming that anisotropy of the aptamer-protein complex does not change as a result of mutation of the aptamer sequence, the same parameter α was used for fitting results from different experiments. Fitting results are summarized in table 1.



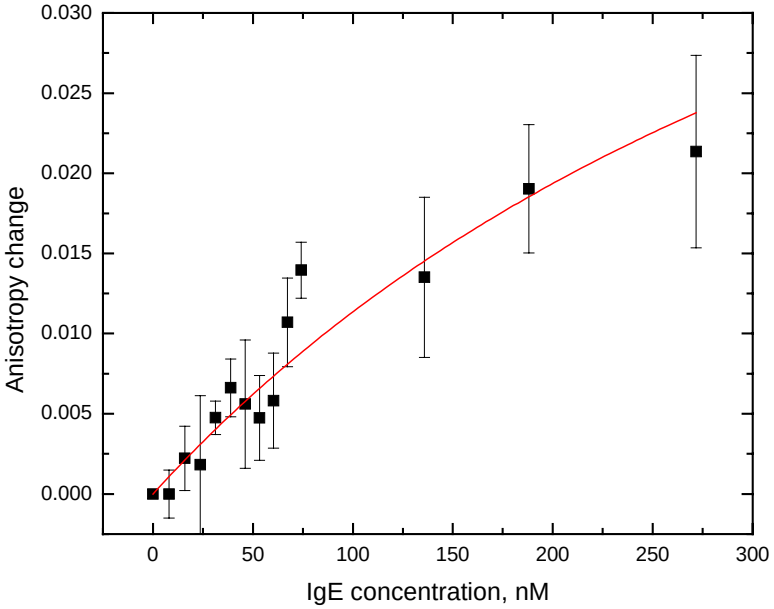
S Figure 1. Anisotropy change due to the aptamer binding to the IgE protein. Aptamer (original sequence) labeled with Texas Red was at 20 nM concentration. Solid squares show the experimental results, red line is the fit as described above.



S Figure 2. Anisotropy change upon binding to IgE protein for the aptamer containing the mutation G to T at position 1. The labeled aptamer concentration was 20 nM. The solid squares show the experimental results; the red line is a fit of the data.



S Figure 3. Anisotropy change upon binding to IgE protein for the aptamer containing the mutation C to T at position 19. The labeled aptamer concentration was 20 nM. The solid squares show the experimental results; the red line is fit of the data.



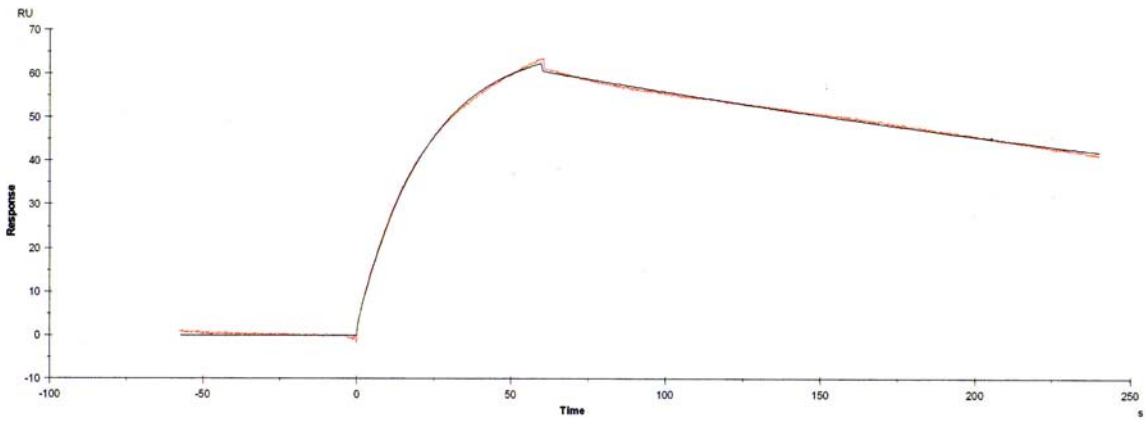
S Figure 4. Anisotropy change upon binding to IgE protein for the aptamer containing the mutation T to A at position 17. The labeled aptamer concentration was 20 nM. The solid squares show the experimental results; the red line is fit of the data.

Surface Plasmon Resonance (SPR) experiments

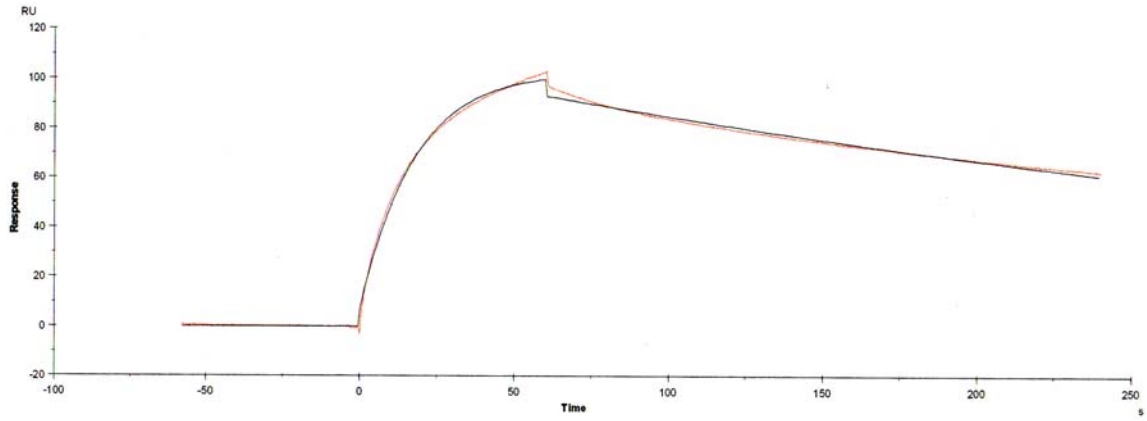
SPR experiments were performed using a Biacore T100 instrument (GE Healthcare). IgE was immobilized on the surface of a C5 chip using the standard carbodiimide- *N*-hydroxysuccinimide (EDC-NHS) coupling protocol. Immobilization resulted in about 14,500 resonance units (RU). The binding of100 nM concentrations of the original IgE-binding aptamer (sequence 17.4), and three variants was then determined. The three variants investigated were the same as those shown in figure 4: the G to T mutation at position 1, the C to T mutation at position 19 and the T to A mutation at position 17 (HPLC purified DNA oligonucleotides were purchased from IDT DNA Technologies). The binding assay was performed in 1X PBSM buffer. Both a no protein control and a buffer-only control were run. The binding was detected in double reference mode; the kinetic traces from the controls were subtracted from the sample traces in order to remove any non-specific binding signals arising from aptamer binding to the dextran matrix or due to refractive index changes upon buffer injection. The results are shown in figures S5 through S8. Fitting of the kinetic data was performed using Biacore data analysis software using a 1:1 binding model. A summary of the results for these fits are presented in Table 1. These experiments were not optimized for a quantitative K_D determination (a short time window and high protein concentration were used for ease of measurement), but simply to give a relative affinity determination for internal verification of the array binding and fluorescence anisotropy measurements. High surface concentration of the protein may lead to the rebinding of the aptamer upon its dissociation, giving rise to artificially low values of K_D. This may be why the SPR based K_D values are somewhat lower in general than the anisotropy values in Table 1.

For both the anisotropy and the SPR measurements, qualitatively consistent results are obtained for the variation in the dissociation constant as a function of the aptamer mutation (Table 1). Note that these measurements give K_D values which are similar to the published values of about 10 nM for the original aptamer. Mutation of the aptamer at position 1 results in faster binding to the protein in the SPR measurement, i.e. a lower K_D, a result seen even more distinctly in the anisotropy measurements. On the other hand, mutation of the aptamer at position 17 results in the complete loss of binding as detected by SPR (a small signal change is observed which is most likely due to non-specific binding of DNA) and a very high K_D from the anisotropy measurements. Mutation of the aptamer at position 19 results in an apparent K_D which is somewhat larger than the apparent K_D of the original sequence, in both the SPR and anisotropy measurements, though in the case of the anisotropy measurements, the change is roughly the same size as the noise.

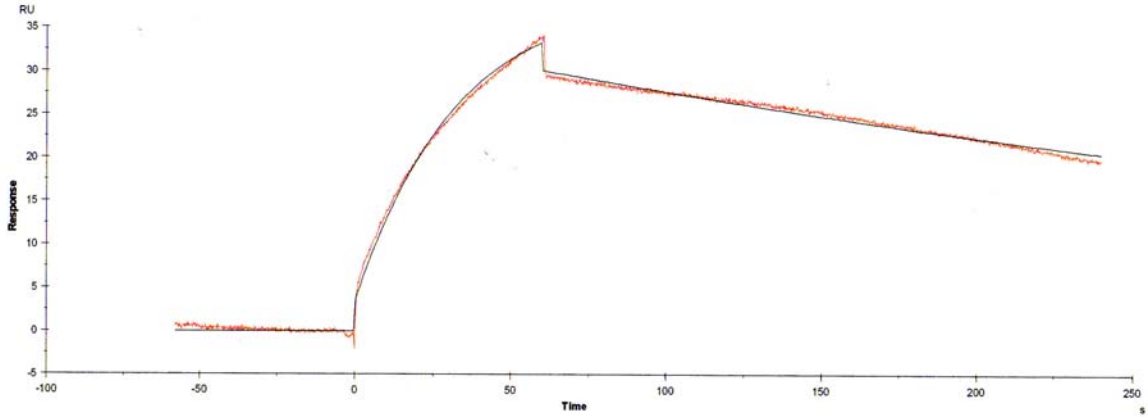
While the microarray results presented in the main text could not be directly interpreted in terms of K_D values for reasons discussed there, the qualitative trend of the fluorescence signals maps precisely with the K_D changes determined using anisotropy and SPR measurements. The G to T mutation at position 1 shows a stronger affinity for the target, C to T at position 19 is roughly two-fold lower in signal intensity and T to A at position 17 is near background fluorescence levels.



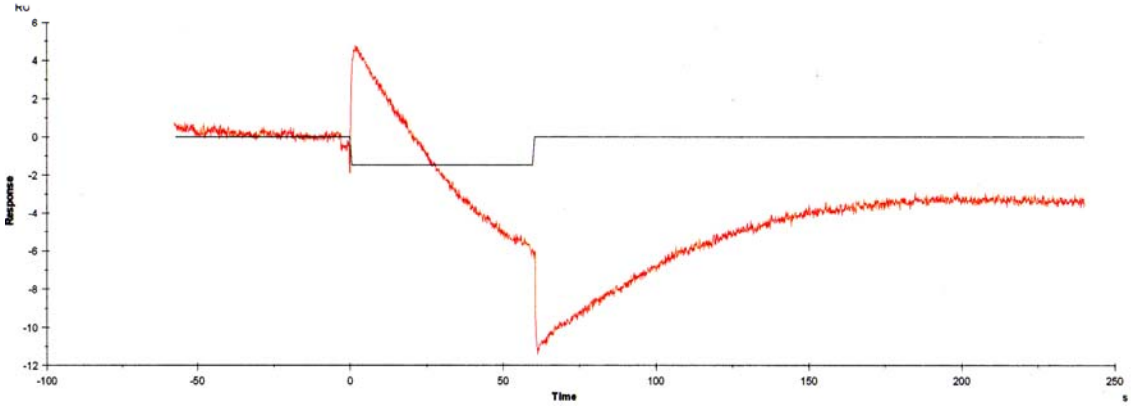
S Figure 5. Kinetic trace for binding of the original aptamer sequence to IgE as measured using SPR. The red line is the experimentally recorded response, and the black line is the fit of the data using a kinetic model assuming 1:1 binding (see text).



S Figure 6. Kinetic trace for binding of the aptamer mutation, G to T at position 1, measured using SPR. The red line is the experimentally recorded response, and the black line is the fit of the data using a kinetic model assuming 1:1 binding (see text).



S Figure 7. Kinetic trace for binding of the aptamer mutation, C to T at position 19, measured using SPR. The red line is the experimentally recorded response, and the black line is the fit of the data using a kinetic model assuming 1:1 binding (see text).



S Figure 8. Kinetic trace for binding of the aptamer mutation, T to A at position 17, measured using SPR. The red line is the experimentally recorded response, and the black like is the fit of the data using a kinetic model assuming 1:1 binding (see text). Notice that for this mutated aptamer the relative change in response is only few Resonance units, which is indicative of non-specific aptamer binding to the target protein.