Improving selectivity of DNA-RNA binding zinc finger using directed evolution

Agata Agnieszka Sulej (asulej@iimcb.gov)
International Institute of Molecular and Cell Biology in Warsaw

Research note

Keywords: Directed evolution, protein engineering, zinc finger, phage display, DNA-RNA hybrids

Posted Date: July 1st, 2019

DOI: https://doi.org/10.21203/rs.2.10756/v1

Version of Record: A version of this preprint was published on December 4th, 2019. See the published version at https://doi.org/10.1186/s13104-019-4833-8.

Abstract

Objective Type C2H2 zinc fingers bind a variety of substrates among which, are specific sequences in the double-stranded DNA. Engineering efforts led to the discovery of a set of general rules that enable obtaining zinc fingers modules that bind to almost any given sequence. The objective of this work was to determine an analogical set of rules for the binding of specific sequences in DNA-RNA hybrid using directed evolution of ZfQQR zinc finger. The target regions for evolution included the amino acid residues that directly interact with the substrate and linkers between modules of the zinc finger. Results The directed evolution was performed using selection based on biopanning of phage-displayed libraries of randomized regions in the ZfQQR zinc finger. The applied strategy of randomization of the middle module of the zinc finger along with input library bias and materials used for biopanning hindered the selection of modules with altered specificity. However, the directed evolution of the linker sequence between modules enabled selection of variants with improved selectivity towards DNA-RNA hybrids in the presence of double-stranded DNA in comparison to the original ZfQQR. This confirms the necessity of linker optimization between modules in zinc finger domains.

Introduction

The C2H2 type zinc fingers are modular domains that specifically recognize and bind bases in the double-stranded DNA (dsDNA), but they can also interact with RNA and proteins [1]. A single domain comprises β-hairpin and α-helix stabilized by coordination of a zinc ion [2]. The helix contains four amino acid residues that form one-to-one contacts with four bases in the dsDNA substrate. A single module binds to three bases on one strand of the dsDNA and to a fourth base on the opposite strand. The recognition code for interaction with all possible DNA sequences was determined paving the way for rational design of zinc fingers with custom specificity [3].

Although, the rules for their interaction with specific bases of the DNA sequence are well established, it was demonstrated that substitution of solely the amino acids directly involved in interaction with bases may not be sufficient to achieve a highly specific zinc finger. Other amino acid residues may also affect affinity for a given sequence i.e. additional interactions, outside the canonical ones, with the substrate [4], the sequence and length of the linker between successive zinc finger [5–8]. Engineering efforts were particularly aimed at expanding the recognition sequence by multimerization of zinc finger domains [9] and their application as targeting modules when fused with effector domains like nucleases, transcription activators and repressors [10]. Efforts were also made to obtain domains that bind other nucleic acids, like structured RNA [11] or specific sequences in DNA-RNA hybrids.

The artificial zinc finger ZfQQR was engineered to bind 5'GGGGAAGAA3' sequence in the DNA strand of the DNA-RNA hybrid [12]. It comprises three zinc finger modules (Zfm), one binds the first trinucleotide 5'GGG3' and two bind the repeated 5'GAA3' (Fig.1A). It was used to construct a fusion enzyme with a ribonuclease H domain, turning a nonspecific processive enzyme into a sequence-specific one [13]. This

opened the possibility of developing molecular tools for precise manipulation of RNA molecules, similarly like restriction enzymes are applied for cleavage of dsDNA.

Here are presented the efforts to determine the recognition code for the zinc fingers that bind specific sequences in DNA-RNA hybrids using directed evolution. It involved randomization of two regions in the ZfQQR, one that encodes amino acid residues that directly recognize the bases in the DNA strand and selection on a panel of substrates, the other randomization and elongation of the linker between modules of the zinc finger domains followed by selection using the original target sequence.

Main Text

Methods

See Additional file 1.

Results and discussion

Directed evolution of ZfQQR variants

Directed evolution of the ZfQQR zinc finger was attempted by creating a library of variants using saturation mutagenesis by codon cassette insertion in target regions. T7 phage display was used in order to obtain phenotype-to-genotype linkage and selection of desired variants from the libraries was performed using biopanning. For this purpose, a modified version of the *zfqqr* gene with unique restriction sites around the target regions was cloned into a T7 phage downstream of the 10B capsid protein. In order to confirm that the T7ZfQQR recombinant phage expresses a functional zinc finger, biopanning was performed using a control mixture of T7 phages with recombinant T7ZfQQR at the ratio 100:1. The control mixture was incubated with immobilized biotinylated substrate that contains the ZfQQR recognition sequence and an empty streptavidin-coated well. Already after the second round of biopanning, the ratio of T7:T7ZfQQR was 5:8 and 95:1 from substrate-immobilized and empty wells, respectively. The enrichment of the initial mixture with T7ZfQQR phages confirmed the functionality of the expressed zinc finger and effectiveness of the selection method.

Three libraries of genes encoding ZfQQR zinc finger variants, Zfm2, L5 and L6, were constructed. The Zfm2 library was designed to select domains with altered sequence specificity by randomization of the residues involved in direct interaction with the substrate (Q56, S58, N59 and K62) in the second zinc finger module (Fig. 1B). The L5 and L6 libraries were designed to enable selection of variants that are more selective towards DNA-RNA hybrids with the target sequence by randomization of the sequence of Zfm2 - Zfm3 linker. In library L5 (Fig.1B), the fragment coding five amino acid residues (TGEKP) was randomized, whereas in the library L6 the randomized fragment was extended to six residues (Fig. 1B). The rationale behind extension of the linker was based on the fact that the structure of the DNA-RNA

hybrid helix is an intermediate between two forms- A with 11 base pairs per turn and B with 10.5 base pairs per turn. It is slightly more packed in comparison to the B form of the dsDNA [14]. A longer and flexible linker might enable the modules to wrap around the DNA-RNA helix and fit better to the compressed structure than a shorter and more rigid one.

In all libraries, the selected codons were replaced by a degenerate NNS codon. After ligation of the library cassettes to the T7ZfQQR construct and *in vitro* packaging 3,3 x 10^5 pfu/ml , 2.4 x 10^5 pfu/ml, 2.1 x 10^6 pfu/ml recombinant phages were obtained for libraries Zfm2, L5 and L6, respectively. Selection of variants from the Zfm2 library was carried out in parallel on a set of 64 biotinylated DNA-RNA hybrid substrates, each having a different possible variation of the three middle nucleotides in the recognition site (Fig. 1A, see Table S1 in Additional file 1). Such approach was aimed at determining the recognition code for binding DNA-RNA hybrids. Libraries L5 and L6 were selected using the original ZfQQR binding sequence. Phage libraries were biopanned for five rounds, the phage titer after each round varied from 10^5 to 10^7. The material after biopanning, input libraries and negative control (phage library Zfm2 biopanned on surface without substrate) were sequenced using MiSeq Illumina sequencer.

On average, 67 thousand reads were obtained with the correct length and sequence flanking the randomized regions for each sample. Distribution of the degenerate NNS sequence in the input Zfm2 library was uneven. The predominant codons encoded mainly P, F, L and V residues accounting for around 50% of reads, whereas the theoretical frequency should be around 25% (see Table S1 in Additional file 2). The most frequently appearing sequence encoded the motif PPPP and was present in 4.5% of all the filtered reads. For the input libraries L5 and L6, no bias in the amino acid distribution was observed (see Table S2 and S3 in Additional file 2).

In the case of variants derived after selection from library Zfm2 and from the negative control, a very similar distribution of amino acids was observed irrespective of the sequence of the substrate used for biopanning or the presence of the substrate during the selection (see Table S1 in Additional file 3). All samples had a similar consensus sequence FVLL (example in Fig.2A) where the consecutive letters of the motif correspond to the residues in the native protein Q56, S58, N59 and K62. Distribution of the amino acid residues in all sequenced samples resembled to a large extent the distribution of the input library Zfm2 (see Table S1 in Additional file 3). The most prominent change observed in the isolated variants was the lowering of frequency of the PPPP motif. Most likely the selection pressure disfavored the presence of a conformationally rigid residue, because of the steric hindrance in the structure of the alpha helix in the zinc finger module [15, 16] . It is also possible that the presence of four proline residues affected the folding of the zinc finger fusion with phage capsid and as a result these variants were eliminated during consecutive biopannings.

The above results may be caused by several factors. The starting material in library Zfm2 was characterized by an uneven distribution potentially causing them to dominate the less frequent ones by occupying the binding sites. Perhaps the selection pressure was insufficient for the applied strategy of randomization of the middle zinc finger module. The binding of the first and third module might have

been strong enough to withstand the selection. This is further supported by the fact that similar sequences were isolated from libraries selected on almost all of the substrates. Additionally, interaction with bases in nucleic acids was rarely observed for this group of amino acid residues [17]. Another possibility is that the lack of binding in the central module promotes binding of the DNA-RNA hybrid structure by eliminating a steric hindrance that may arise from its specific interaction.

Sequencing of the variants derived from the selection of library L5 revealed that the predominant isolated amino acid sequence was TRERN (17% of obtained sequences, see Fig. 2B). For library L6 the sequence NQMMRK (9% of obtained sequences, see Fig. 2C) was most frequently observed. None of the above two amino acid sequences appeared in the results from sequencing of the input libraries, which means that they were present less frequently than 1 in 55162 for the L5 library and 1 in 42323 for the L6 library. What is interesting that in case of the library L5 the sequence NQMRP, partially resembling the one isolated from the L6 library was the fourth most frequently appearing (Fig. 2B).

Binding affinity and selectivity of the isolated variants

The binding affinity of zinc finger variants selected using directed evolution was determined. For the Zfm2 library, the consensus sequence was chosen and variants of the zinc finger containing the Q56F S58V N59L K62L substitutions (termed ZfFVLL) only in the Zfm2 and in both, Zfm2 and Zfm3 (additional substitutions Q28F S30V N31L K34L, termed Zf2xFVLL) were obtained. The most frequently observed amino acid motifs obtained for the libraries L5 and L6 were introduced to the Zfm2-Zfm3 linker (termed ZfTRERN and ZfNQMMRK, respectively) and, additionally to the Zfm1-Zfm2 linker (termed Zf2xTRERN and Zf2xNQMMRK, respectively).

For native ZfQQR and each protein variant, the equilibrium dissociation constant constant was measured using surface plasmon resonance (Fig. 3A). The K_D for the ZfFVLL and Zf2xFVLL zinc fingers was above 5000 nM and could not be measured using this method because the proteins were aggregating in the assay buffer at concentrations above 2 μM. This result of binding analysis and the sequencing results obtained from selection using a panel of 64 substrates, as well as the negative control, support the explanation that the input library bias along with insufficient selection pressure hampered the biopanning. It is most likely that the selected variants result from the background nonspecific binding of phage particles to the streptavidin-coated wells.

The K_D of the ZfTRERN and ZfNQMMRK variants was slightly higher than the ZfQQR (Fig. 3B). However, when the motifs were repeated in the Zfm1-Zfm2 linker, the variants had 10-fold and 40-fold higher K_D than the single motif variants. This result indicates that the engineering of linkers is localization specific and their optimization should be performed separately for each one.

In order to determine if the zinc finger variants were improved in their ability to discriminate between the DNA-RNA hybrids over dsDNA their relative binding to the substrate with the 5'GGGGAAGAA3' sequence in the presence of 100-fold excess of a dsDNA competitor (containing the 5'GGGGAAGAA3' sequence) was

measured using nitrocellulose filter binding assay. All single motif and double motif variants displayed at least 2-fold higher relative binding of the DNA-RNA hybrid than the original ZfQQR (Fig. 3C). Although the variants display a lower K_D than the ZfQQR, their selectivity for DNA-RNA hybrids over dsDNA improved. It might indicate that further optimization of preference for DNA-RNA hybrid vs. dsDNA binding is achievable and that it is distinct from optimization of the sequence selectivity.

Limitations

Sequence bias in the Zfm2 input library resulted in overrepresentation of the P, F, L and V codons. The number of phage particles obtained after *in vitro* packaging was insufficient to represent all the possible codon combinations in the theoretical library. Affinity binding measurements using surface plasmon resonance were done as single experiments.

List Of Abbreviations

dsDNA: double-stranded DNA

Zfm: zinc finger module

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares that she has no competing interests.

Funding

This work was supported by the Iuventus Plus grant (no. IP2012 065672) from the Polish Ministry of Science and Higher Education.

Authors' contributions

AS performed the experiments, analyzed data and wrote the manuscript.

Acknowledgements

The author wishes to thank Krzysztof Skowronek and Elzbieta Purta for advice and correction of the manuscript, and Marcin Sulej for writing a script for NGS data processing.

Bibliography

[1] Iuchi S (2001) Three classes of C2H2 zinc finger proteins. Cell Mol Life Sci 58:625–635

[2] Pavletich NP, Pabo CO (1993) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Angstroms. doi: 10.2210/pdb1zaa/pdb

[3] Sera T, Uranga C (2002) Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. Biochemistry 41:7074–7081

[4] Vandevenne M, Jacques DA, Artuz C, Nguyen CD, Kwan AHY, Segal DJ, Matthews JM, Crossley M, Guss JM, Mackay JP (2013) New insights into DNA recognition by zinc fingers revealed by structural analysis of the oncoprotein ZNF217. J Biol Chem 288:10616–10627

[5] Ryan RF, Darby MK (1998) The role of zinc finger linkers in p43 and TFIIIA binding to 5S rRNA and DNA. Nucleic Acids Res 26:703–709

[6] Moore M, Choo Y, Klug A (2001) Design of polyzinc finger peptides with structured linkers. Proc Natl Acad Sci U S A 98:1432–1436

[7] Imanishi M, Sugiura Y (2002) Artificial DNA-Bending Six-Zinc Finger Peptides with Different Charged Linkers: Distinct Kinetic Properties of DNA Bindings†. Biochemistry 41:1328–1334

[8] van Leeuwen HC, Strating MJ, Rensen M, de Laat W, van der Vliet PC (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. EMBO J 16:2043–2053

[9] Beerli RR, Barbas CF 3rd (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20:135–141

[10] Klug A (2010) The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annual Review of Biochemistry 79:213–231

[11] Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. FEBS Journal 280:3734–3754

[12] Shi Y, Berg JM (1995) Specific DNA-RNA hybrid binding by zinc finger proteins. Science 268:282–284

[13] Sulej AA, Tuszynska I, Skowronek KJ, Nowotny M, Bujnicki JM (2012) Sequence-specific cleavage of the RNA strand in DNA-RNA hybrids by the fusion of ribonuclease H with a zinc finger. Nucleic Acids Res 40:11563–11570

[14] Fedoroff OY, Salazar M, Reid BR (1993) Structure of a DNA : RNA Hybrid Duplex. J Mol Biol 233:509–523

[15] MacArthur MW, Thornton JM (1991) Influence of proline residues on protein conformation. Journal of Molecular Biology 218:397–412

[16] Kim MK, Kang YK (1999) Positional preference of proline in alpha-helices. Protein Sci 8:1492–1499

[17] Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res 29:2860–2874

Additional File Legend

Additional file 1 contains the description of methods, scheme of a fragment of the capsid-*zfqqr* fusion(Figure S1), used oligonucleotides (Table S1) and buffer composition (Table S2).

Additional file 2 contains the theoretical and observed frequencies of amino acid residues in input library Zfm2 (Table S1), L5 (Table S2) and L6 (Table S3).

Additional file 3 contains the number of reads passing filters and frequency sequence logos of randomized regions in the selected variants from Zfm2 library after biopanning on a panel of substrates (Table S1), and from the L5 and L6 library (Table S2).

Figures

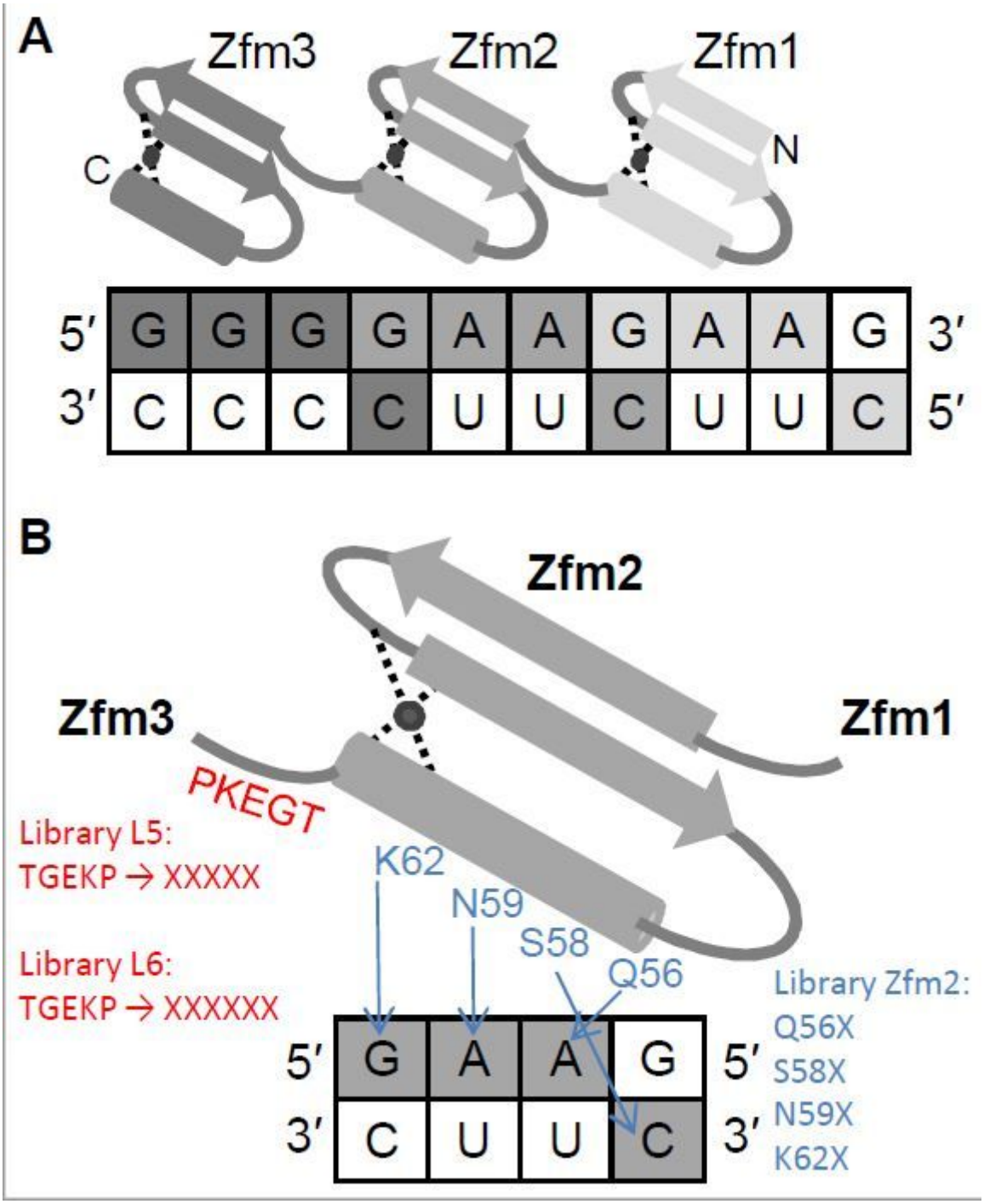
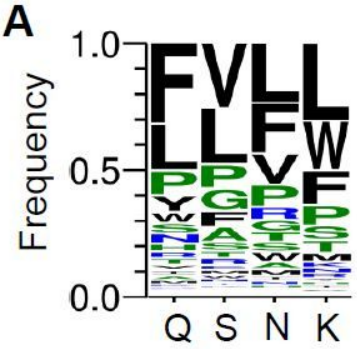


Figure 1

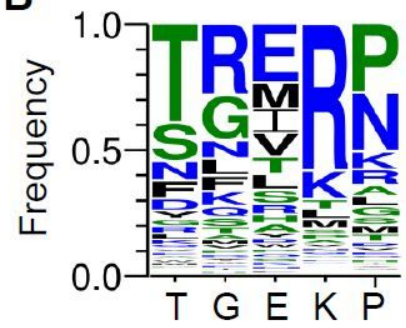
Schematic illustration of interaction between ZfQQR and the recognition sequence in DNA-RNA hybrid substrate. (A) Three zinc finger modules. (B) Zfm2 with linkers. Amino acid residues of the Zfm2-Zfm3 linker and directly contacting bases of the substrate are labelled. In blue the Zfm2 library randomized residues, in red the L5 and L6 libraries randomized residues.

A



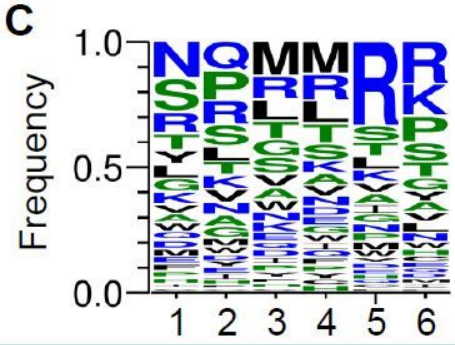
Motif sequence	Number of sequences	Fraction of total number of sequences
LVLL	1532	2.5%
FLLL	1505	2.5%
FLLL	1288	2.1%
FVLL	1096	1.8%
FLFF	1061	1.7%

B



Motif sequence	Number of sequences	Fraction of total number of sequences
TRERN	7147	17.2%
FFIRP	2020	4.9%
DRMRK	1360	3.3%
NQMRP	1004	2.4%
TGHKM	814	2.0%

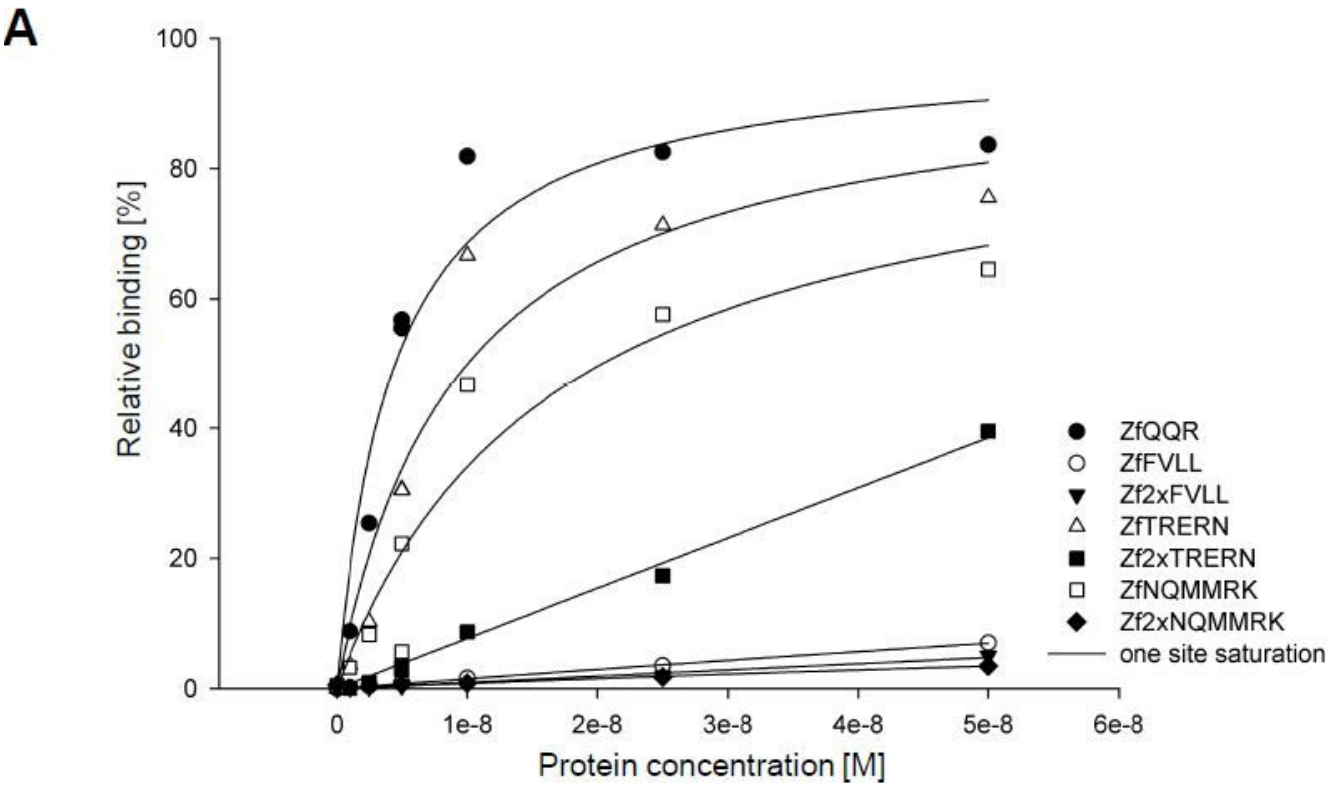
C



Motif sequence	Number of sequences	Fraction of total number of sequences
NQMMRK	2525	9.1%
SPLTRP	517	1.9%
WPMLRR	345	1.2%
SPAERP	338	1.2%
SSGARP	262	0.9%

Figure 2

Sequence logos and five most abundant motifs obtained from sequenced fragments after 5 rounds of phage selection using a substrate containing the 5'GGGGAAGAA3' sequence for: (A) library Zfm2, (B) library L5 and (C) library L6.



B

Name	Equillibrium dissociation constant and standard deviation [nM]
ZfQQR	4.5 ± 1.3
ZfFVLL	> 5000
Zf2xFVLL	> 5000
ZfTRERN	9.9 ± 3.6
Zf2xTRERN	110 ± 15
ZfNQMMRK	30 ± 8
Zf2xNQMMRK	1169 ± 89

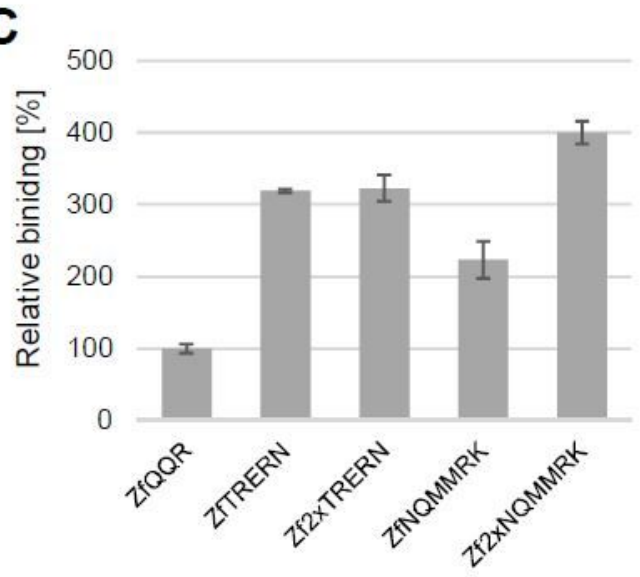


Figure 3

Relative binding of the sequence 5′GGGGAAGAA3′ by ZfQQR and variants. (A) Binding affinity measured using surface plasmon resonance on BIAcore 3000 instrument. B) The equilibrium dissociation constants (KD) were obtained from global fitting of the results of affinity measurement using the ones site saturation model. (C) Relative binding in the presence of dsDNA competitor in 100-fold molar excess. 100% is the binding of ZfQQR to substrate with 5′GGGGAAGAA3′ sequence

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.docx
- supplement2.docx
- supplement3.docx