Protein Design

Modular Assembly of RanBP2-Type Zinc Finger Domains to Target Single-Stranded RNA**

*Mitchell R. O'Connell, Marylène Vandevenne, Cuong D. Nguyen, Jacqueline M. Matthews, Roland Gamsjaeger, David J. Segal, and Joel P. Mackay**

The many functions of RNA are only now beginning to be clearly appreciated. For example, elements within messenger RNAs can control the levels of the encoded protein and modulate the function of that protein through alternative splicing. Non-coding RNAs have also emerged as wide-ranging players in eukaryotic biology.[1]

RNA-binding proteins (RBPs) with tunable specificity would allow the interrogation and modulation of cellular RNAs.[2] Several approaches to RBP design have already been taken. For example, libraries of classical zinc fingers yielded variants that recognized double-stranded RNA (dsRNA), although the sequence specificity of these proteins has not been extensively assessed.[3] More recently, the binding specificity of Pumilio repeat domains, which recognize up to eight nucleotides of single stranded RNA (ssRNA), has been manipulated by rational design and the resulting proteins were used in a variety of cellular assays.[4]

The successful design of designer DNA-binding zinc-finger (ZF) proteins[5] has been predicated on several key properties of these domains. First, the molecular basis for the interaction of classical ZFs with DNA is well understood.[6] Second, the interaction is driven by sidechain interactions, allowing sequence specificity to be tuned through mutagenesis. Third, the classical ZF is a small (ca. 30 residues), stable fold that is very tolerant to mutagenesis,[7] allowing many sequence variants to be created that have distinct specificities.[8] Finally, classical ZF domains generally recognize DNA in a modular fashion—that is, a single ZF recognizes three base pairs of double stranded DNA in a manner that is largely independent of the surrounding ZFs.[6] To design proteins that target ssRNA in a similar manner, we have focused on a structurally distinct family of
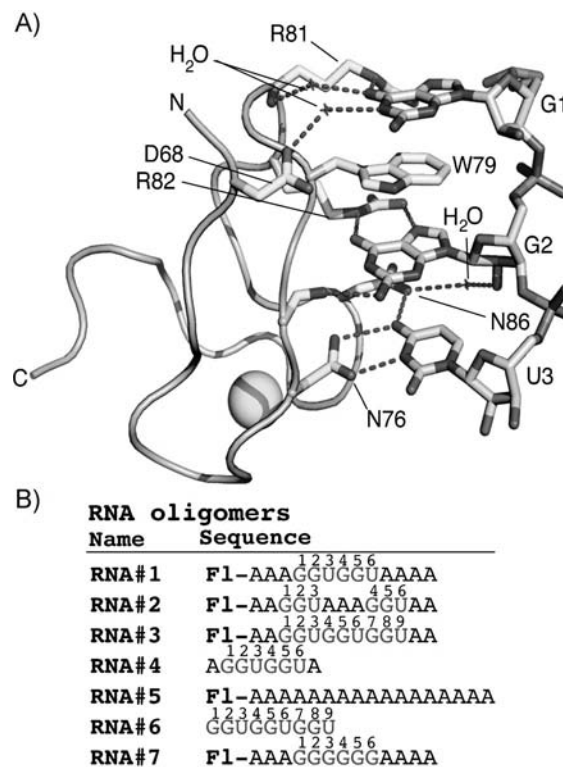
[*] M. R. O'Connell, Dr. M. Vandevenne, C. D. Nguyen,
Prof. J. M. Matthews, Dr. R. Gamsjaeger,[+] Prof. J. P. Mackay
School of Molecular Bioscience, University of Sydney
Sydney, N.S.W 2006 (Australia)
E-mail: j.mackay@sydney.edu.au

Prof. D. J. Segal
Genome Center and Department of Biochemistry and Molecular
Medicine, University of California, Davis
Davis, CA 95616 (USA)

[+] Present address: School of Science and Health, University of
Western Sydney (Australia)

Figure 1. Structure of the second ZF of ZRANB2 (F2) bound to ssRNA and sequences used in this work. A) X-ray crystal structure of F2 bound to ssRNA containing the sequence GGU (from ref. [9b]; PDB code 2K1P). Hydrogen bonds are indicated as dashed lines. B) Sequences of RNA oligonucleotides used in this work. GGX motifs are colored gray and numbered according to nucleotide position.

ZFs—the RanBP2-type ZFs (Figure 1 A).[9] These domains are present in approximately 20 human proteins, and are found throughout Eukarya. We have previously shown that each of the two RanBP2-type ZFs from the human protein ZRANB2 binds to ssRNA with an affinity of ca. $1 \times 10^6 \, \text{M}^{-1}$ and with specificity for a core GGU motif.[9b] The two domains, separated by a ca. 25-residue linker (Figure S1), can recognize double GGU motifs that are separated by an intervening sequence of variable length.

The crystal structure of finger 2 (F2) from ZRANB2 bound to ssRNA[9b] revealed that binding is primarily mediated by sidechain hydrogen bonds (Figure 1 A). Two arginine sidechains form hydrogen bonds to each of the two guanines in the GGU motif, and recognition of the uridine is specified by hydrogen bonds from two asparagine sidechains. A tryptophan is sandwiched by the two guanines, providing

additional affinity and shape complementarity. The domain is stable to mutagenesis,[9b] and sequence diversity among RanB2-type ZFs (see Supporting Information, Figure S2) underscores the mutability of the fold and suggests that other RanBP2-type ZFs target different RNA sequences. Consistent with this idea, we have shown that the RanBP2-type ZF from RBM5 specifically recognizes a GGG, rather than a GGU, motif.[10]

We therefore propose that RanBP2-type ZFs might constitute a suitable scaffold for the design of ssRNA-binding proteins (ssRBPs). Here we demonstrate that we can assemble a tandem array of three RanBP2-type ZFs and that this engineered protein can recognize a nine-nucleotide ssRNA target. We also show that we can manipulate the RNA-binding specificity of a RanBP2-type ZF array by substituting ZRANB2 ZFs with the ZF from RBM5. Taken together, these data demonstrate that RanBP2-type ZFs can be assembled in a modular fashion to target contiguous ssRNA sequences, opening up possibilities for the implementation of RanBP2-type ZF arrays as addressable RNA-binding molecules.

One of the fundamental considerations in the assembly of multidomain proteins is the composition and length of the inter-domain linker. Although longer linkers can allow consecutive RNA-binding domains to take up optimal orientations for RNA binding, they can increase the entropic cost of binding and are more likely to be subject to proteolytic degradation.[11] The additional flexibility can also reduce binding specificity by allowing the protein to bind sites with different nucleotide spacings (see, for example, ref. [8b]).

We have shown previously that the 25-residue linker in ZRANB2 is flexible, even upon RNA binding,[9b] suggesting that it is unlikely to contribute to recognition. To assess the influence of linker length on the RNA-binding properties of RanBP2-type ZF arrays, we designed deletion mutants of the ZRANB2 double finger ($F12_{WT}$; amino acids 1–95 of human ZRANB2) in which up to 24 amino acids were removed from the linker (Figure S1). We used fluorescence anisotropy titrations to test the ability of each mutant to bind a 13-nucleotide (13-nt) ssRNA containing a GGUGGU motif (RNA #1; Figure 1 A). Figure 2 B shows that removal of 20 residues from the linker had no effect on the affinity of the protein for a GGUGGU motif; association constants were $2.3 \pm 0.5 \times 10^7\,\text{M}^{-1}$ and $2.6 \pm 0.2 \times 10^7\,\text{M}^{-1}$ for $F12_{WT}$ and the $F12_{\Delta45\text{-}64}$ truncation, respectively. Isothermal titration calorimetry measurements (Figure S3) corroborated this finding. Small (ca. 20%) reductions in both ΔH and ΔS (both of which are negative) were associated with shortening the linker. In contrast, the removal of 21 or 24 residues reduced the affinity to the extent that these constructs bound with an affinity similar to that of a single ZF, suggesting that only one of the two ZFs was able to contact the RNA.

To test the effect of linker length on protein stability, we incubated purified $F12_{WT}$ and $F12_{\Delta45\text{-}64}$ alone and in the presence of the protease thrombin at 37°C for up to 96 h and examined their integrity by SDS-PAGE (Figure 2 C). After 96 h in the absence of thrombin, $F12_{\Delta45\text{-}64}$ was unchanged while $F12_{WT}$ had undergone significant proteolysis. In the presence of thrombin, only $F12_{\Delta45\text{-}64}$ remained uncleaved after
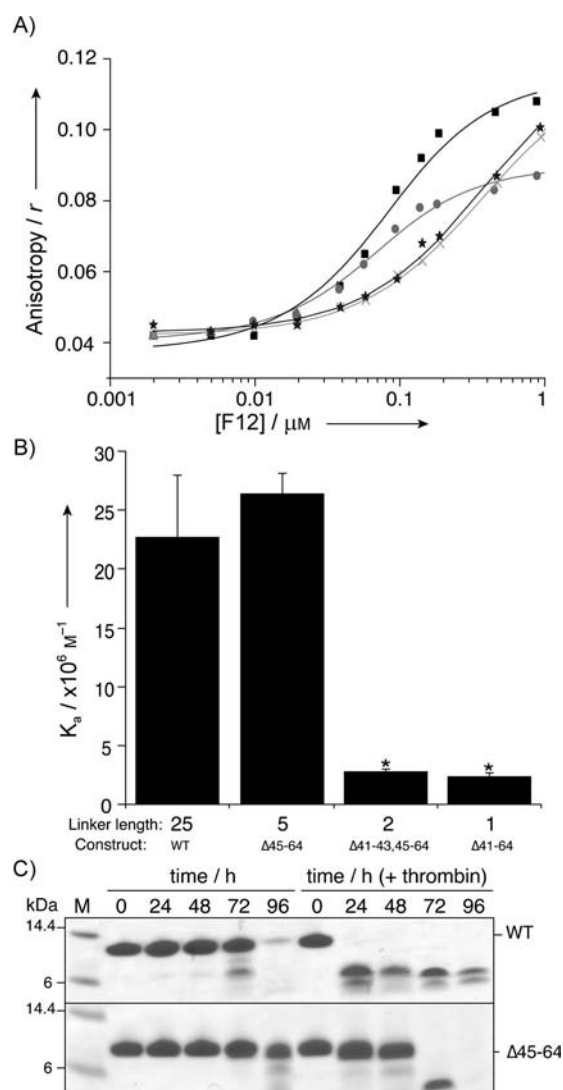

Figure 2. Reduction of the $F12_{WT}$ interdomain linker by 20 residues does not affect ssRNA binding but improves proteolytic stability. A) Fluorescence anisotropy titration data showing the binding of $F12_{WT}$ (■), $F12_{\Delta45\text{-}64}$ (●), $F12_{\Delta41\text{-}43, 45\text{-}64}$ (×), and $F12_{\Delta41\text{-}64}$ (★) to a 5′-fluorescein-tagged 13-nt ssRNA oligonucleotide (RNA #3 in Figure 1 B). Fits to a 1:1 binding isotherm are shown. B) Association constants for the indicated interactions. Each value is the average of three measurements; error bars indicate standard deviations from the mean. The asterisk (*) indicates a statistically significant difference ($p < 0.01$) with respect to the $F12_{WT}$:GGUGGU interaction. C) SDS-PAGE illustrating differences in stability between $F12_{WT}$ (upper panel) and $F12_{\Delta45\text{-}64}$ (lower panel). Each polypeptide was incubated at 37°C alone (left) or in the presence of thrombin (right).

48 h. Together, these data show that $F12_{\Delta45\text{-}64}$ has improved proteolytic stability compared to $F12_{WT}$.

To determine whether tandem arrays comprising three RanBP2-type ZFs can be constructed, we added a second copy of ZRANB2-F2 to the $F12_{\Delta45\text{-}64}$ construct using the same shortened linker, creating the three-ZF F122 protein (Figure 3 A). A ^{15}N-HSQC spectrum of this protein indicated that each of the three ZFs was correctly folded (Figure S3). Signals from the second and third ZFs were almost identical in position, indicating that these two domains take up the same
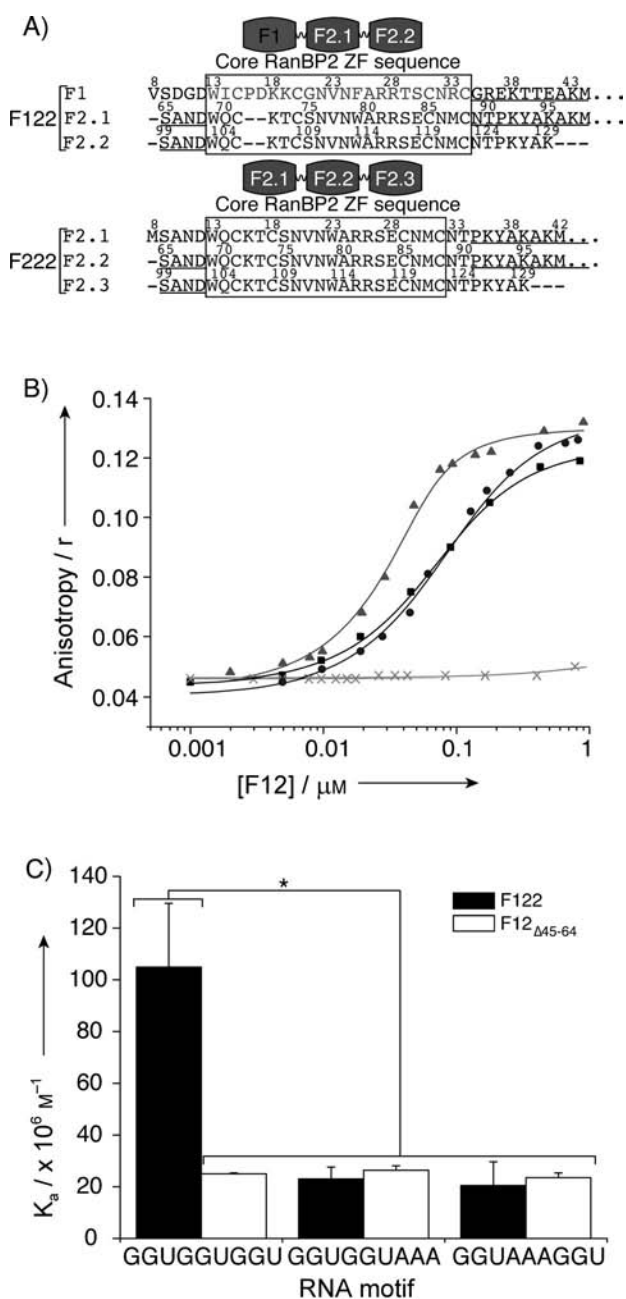
Figure 3. An engineered three-ZF protein binds specifically to a 9-nt target. A) Schematic of F122, showing the core (boxed) and linker (underlined) amino acid sequences of the three ZFs. Residues are numbered according to their positions in $F12_{WT}$. B) Fluorescence anisotropy titration for the binding of F122 to 13-nt ssRNA containing a GGUGGU (■), GGUAAAGGU (●) or GGUGGUGGU (▲) motif (RNAs #1,2,5; Figure 1B). Binding to polyA (RNA #5, Figure 1B) is shown as a control (×). C) Association constants for the indicated interactions. Each value is the average of three measurements; error bars indicate standard deviations from the mean. The asterisk (*) indicates a statistically significant difference ($p < 0.006$) with respect to the F122:GGUGGUGGU interaction.

conformation and that the three-ZF protein can be considered to be three modular domains.

F122 binds to a GGUGGUGGU sequence with an affinity of $1.05 \pm 0.25 \times 10^8 \, \text{m}^{-1}$, which is significantly higher than its affinity for either GGUGGUAAA or GGUAAAGGU ($2.3 \pm 0.5 \times 10^7 \, \text{m}^{-1}$ and $2.0 \pm 0.9 \times 10^7 \, \text{m}^{-1}$, respectively; $p < 0.006$ in both cases). For comparison, $F12_{\Delta45\text{-}64}$ binds to GGUGGUGGU, GGUGGUAAA and GGUAAAGGU with affinities of $2.5 \pm 0.03 \times 10^7 \, \text{m}^{-1}$, $2.6 \pm 0.2 \times 10^7 \, \text{m}^{-1}$ and $2.3 \pm 0.2 \times 10^7 \, \text{m}^{-1}$, respectively (Figure 3C). These data indicate that all three ZFs in the F122 construct are able to make sequence-specific contacts with the RNA—but only in the case where the RNA contains three cognate binding sites. If the RNA contains only two GGU motifs, F122 binds with an affinity that is indistinguishable from the affinity of F12 for the same RNA. It is also notable that the affinity of F122 for GGUGGUGGU is comparable to that of the canonical DNA-binding protein Zif268 for its 9 bp DNA target.[12]

We also engineered a construct containing three copies of F2 (F222, Figure 3A) and measured its affinity for the same ssRNA containing three GGU motifs. Nine sequence differences exist between the two ZFs, although the RNA contact surfaces are essentially identical. The affinity, $0.93 \pm 0.21 \times 10^8 \, \text{m}^{-1}$, was indistinguishable from the affinity of F122 for that sequence, providing further evidence for the modular nature of RanBP2-type ZFs.

To determine the stoichiometry of the F122:GGUG-GUGGU ssRNA interaction, we subjected the complex to multiangle laser light scattering (MALLS) analysis. F122:ssRNA mixtures in 1:1 or 2:1 molar ratios (Figure 4A, middle and lower panel, respectively) yielded data consistent with the formation of the expected 1:1 complex ($M_{\text{exp.}} = 16.2$ kDa, $M_{\text{theor.}} = 14.7$ kDa).

To further characterize the binding mode of F122, we recorded a ^{15}N-HSQC spectrum in the absence and presence of GGUGGUGGU (RNA #6; Figure 1B). We note that a complete RNA titration was not possible due to significant precipitation of F122 when partial saturated with RNA. Resonances corresponding to all three ZFs underwent substantial shifts in the presence of RNA (Figure 4B). No resonances from the unbound state remained, demonstrating that all three ZFs contact the RNA. Furthermore, the pattern of chemical shift changes was mirrored that observed in titrations with both F12 and F2[9b] suggesting that all three ZFs bind the RNA target in an identical and modular manner. Unfortunately, imino protons from the RNA are not sufficiently exchange-protected to be observed in the 1D ^1H NMR spectrum.

We previously showed that the RanBP2-type ZF from RBM5 displays a slight (ca. 1.5-fold) preference for GGG over GGU sequences, and a 2- and 8-fold preference for GGG over GGA and GGC, respectively.[10] We therefore asked whether this ZF could be assembled into a two-finger array that could specifically target GGGGGG sequences. As shown in Figure 5, R11, a protein comprising two RBM5 ZFs binds to GGGGGG with an affinity of $1.6 \pm 0.24 \times 10^7 \, \text{m}^{-1}$. This affinity is similar to that of the ZRANB2 $F12_{\Delta45\text{-}64}$ construct for its preferred GGUGGU target sequence. In contrast, R11 binds to a GGUGGU sequence with an affinity of $0.57 \pm 0.16 \times 10^7 \, \text{m}^{-1}$, and therefore exhibits a 3-fold preference for GGGGGG over GGUGGU.
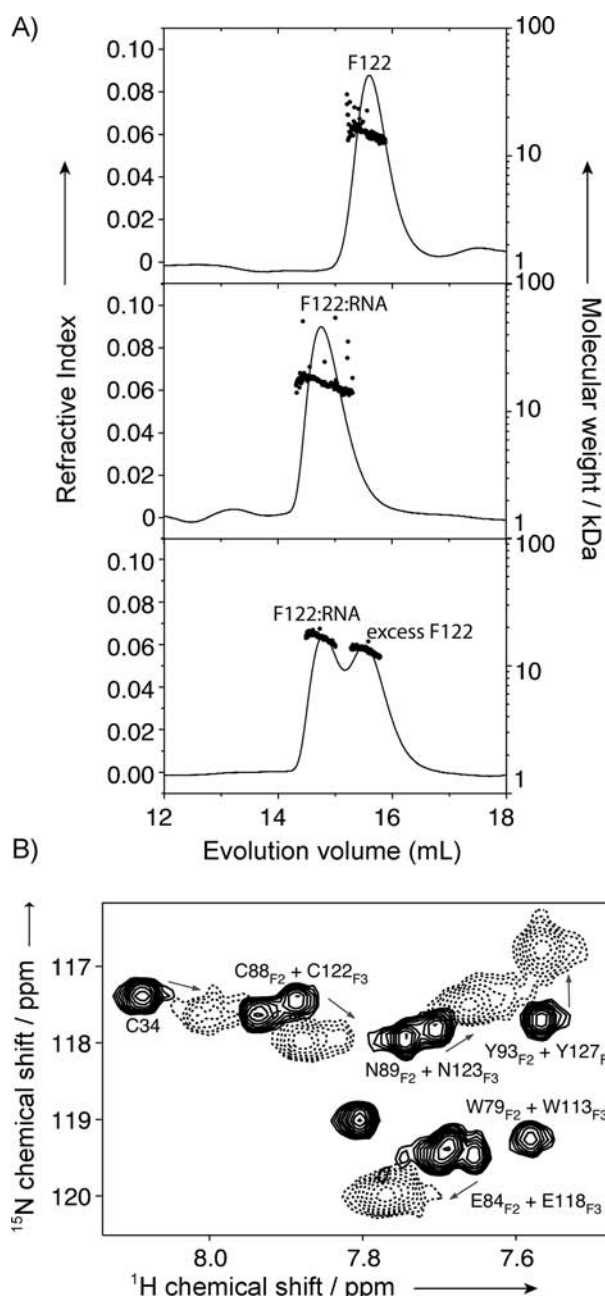
A)

B)

Figure 4. F122 binds RNA with a 1:1 stoichiometry via all three ZFs. A) Size exclusion and MALLS data showing that F122 binds to GGUGGUGGU ssRNA with a 1:1 stoichiometry. Each panel shows a size exclusion chromatography trace monitored by refractive index and multiangle laser light-scattering. Upper panel: F122 alone; middle panel: 1:1 mixture of F122 and ssRNA; lower panel: 2:1 mixture of F122 and ssRNA. B) Section of a ^{15}N-HSQC spectrum of F122 alone (solid line) and as a 1:1 mixture with GGUGGUGGU RNA (dashed line). Assignments and directions of movement are indicated.

One approach to the creation of ssRBPs with tunable sequence specificity involves arrays of multiple RNA-binding domains. In this scenario, the constituent domains should be small, stable and should recognize RNA predominantly through sidechain-mediated interactions. We show here that RanBP2-type zinc fingers fulfil these criteria. We can create
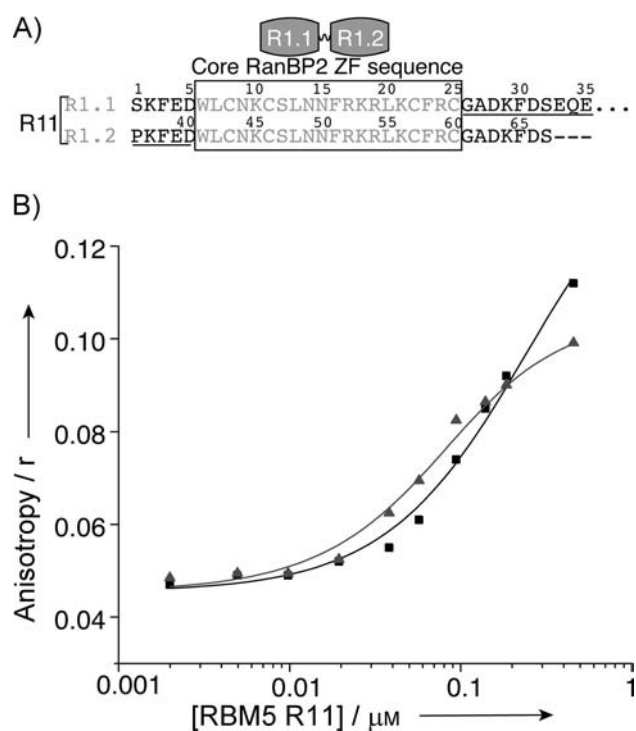
A)

B)

Figure 5. An engineered two-RBM5 RanBP2-type ZF protein binds specifically to GGGGGG. A) Schematic of the RBM5 R11 construct, showing the core (boxed) and linker (underlined) amino acid sequences of the two RBM5 ZFs used in the design of this construct. B) Fluorescence anisotropy data for the binding of R11 to 13-nt ssRNAs containing either a GGGGGG (▲) or GGUGGU (■) motif (RNAs #1 and #7, Figure 1B).

an array of three RNA-binding RanBP2-type ZFs, a topology that does not exist in any naturally occurring protein, and demonstrate that this protein can recognize a tripartite RNA sequence in a sequence-specific manner. The design of this protein incorporated substantially shortened interdomain linkers, which increased the stability of the protein, and NMR data showed clearly that 1) the three-ZF protein was properly folded and 2) all three domains participated in native-like interactions with their RNA targets. The modularity of these engineered proteins is further demonstrated by our ability to create F222 and RBM5-based proteins that bind RNA with the expected affinities and specificities. Finally, the latter protein was designed and demonstrated to preferentially bind to the sequence GGGGGG; this is the first indication that engineered zinc fingers might be able to serve as addressable RNA-binding proteins. Our data provide a framework from which combinatorial approaches such as phage display might be used to create a library of RanBP2-type ZFs with tailored binding specificities.

[1] J. S. Mattick, I. V. Makunin, *Hum. Mol. Genet.* **2006**, *15*, R17–R29.

[2] J. P. Mackay, J. Font, D. J. Segal, *Nat. Struct. Mol. Biol.* **2011**, *18*, 256–261.

[3] a) P. Blancafort, S. V. Steinberg, B. Paquin, R. Klinck, J. K. Scott, R. Cedergren, *Chem. Biol.* **1999**, *6*, 585–597; b) W. J. Friesen, M. K. Darby, *Nat. Struct. Biol.* **1998**, *5*, 543–546; c) W. J. Friesen, M. K. Darby, *J. Biol. Chem.* **1997**, *272*, 10994–10997.

[4] a) J. Tilsner, O. Linnik, N. M. Christensen, K. Bell, I. M. Roberts, C. Lacomme, K. J. Oparka, *Plant J.* **2009**, *57*, 758–770; b) Y. Wang, C. G. Cheong, T. M. T. Hall, Z. F. Wang, *Nat. Methods* **2009**, *6*, 825–863; c) T. Ozawa, Y. Natori, M. Sato, Y. Umezawa, *Nat. Methods* **2007**, *4*, 413–419; d) A. Filipovska, M. F. M. Razif, K. K. A. Nygård, O. Rackham, *Nat. Chem. Biol.* **2011**, *7*, 425–427.

[5] a) C. O. Pabo, E. Peisach, R. A. Grant, *Annu. Rev. Biochem.* **2001**, *70*, 313–340; b) R. R. Beerli, C. F. Barbas 3rd, *Nat. Biotechnol.* **2002**, *20*, 135–141.

[6] S. A. Wolfe, L. Nekludova, C. O. Pabo, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183–212.

[7] S. F. Michael, V. J. Kilfoil, M. H. Schmidt, B. T. Amann, J. M. Berg, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4796–4800.

[8] a) B. Dreier, R. P. Fuller, D. J. Segal, C. V. Lund, P. Blancafort, A. Huber, B. Koksch, C. F. Barbas, *J. Biol. Chem.* **2005**, *280*, 35588–35597; b) B. Dreier, R. R. Beerli, D. J. Segal, J. D. Flippin, C. F. Barbas, *J. Biol. Chem.* **2001**, *276*, 29466–29478; c) B. Dreier, D. J. Segal, C. F. Barbas, *J. Mol. Biol.* **2000**, *303*, 489–502; d) D. J. Segal, R. R. Beerli, P. Blancafort, B. Dreier, K. Effertz, A. Huber, B. Koksch, C. V. Lund, L. Magnenat, D. Valente, C. F. Barbas, *Biochemistry* **2003**, *42*, 2137–2148.

[9] a) C. A. Plambeck, A. H. Y. Kwan, D. J. Adams, B. J. Westman, L. van der Weyden, R. L. Medcalf, B. J. Morris, J. P. Mackay, *J. Biol. Chem.* **2003**, *278*, 22805–22811; b) F. E. Loughlin, R. E. Mansfield, P. M. Vaz, A. P. McGrath, S. Setiyaputra, R. Gamsjaeger, E. S. Chen, B. J. Morris, J. M. Guss, J. P. Mackay, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5581–5586; c) B. Wang, S. L. Alam, H. H. Meyer, M. Payne, T. L. Stemmler, D. R. Davis, W. I. Sundquist, *J. Biol. Chem.* **2003**, *278*, 20225–20234.

[10] C. D. Nguyen, R. E. Mansfield, W. Leung, P. M. Vaz, F. E. Loughlin, R. P. Grant, J. P. Mackay, *J. Mol. Biol.* **2011**, *407*, 273–283.

[11] Y. Shamoo, N. Abdul-Manan, K. R. Williams, *Nucleic Acids Res.* **1995**, *23*, 725–728.

[12] N. P. Pavletich, C. O. Pabo, *Science* **1991**, *252*, 809–817.