

# Cell-free selection of RNA-binding proteins using *in vitro* compartmentalization

Yu Chen<sup>1,\*</sup>, Jana Mandic<sup>2</sup> and Gabriele Varani<sup>1,2</sup>

<sup>1</sup>Department of Chemistry and <sup>2</sup>Department of Biochemistry, University of Washington, Seattle WA, USA

Received July 3, 2008; Revised August 11, 2008; Accepted August 18, 2008

## ABSTRACT

**RNA-binding proteins (RBPs) perform many essential functions in the post-transcriptional control of gene expression. If we were able to engineer RBPs with new specificity, it would also become possible to develop new tools to control and investigate gene expression pathways. Molecular evolution methods such as phage display have been introduced to achieve this goal, but the large interface between these proteins and RNA relative to the size of library that can be constructed limits the efficacy of this method. In order to increase the diversity of libraries used for selection of RBPs, we applied the emulsion-based *in vitro* compartmentalization (IVC) method to select RBPs with defined specificity. A new approach was developed to link genotype and phenotype by fusing the target RBP to zinc finger proteins (ZFPs) that bind to a cognate DNA sequence inserted upstream of the promoter. The expressed fusion protein (ZFP-RBP) binds to its encoding DNA with high affinity via the ZFP target-binding site. After breaking the emulsion, the RBP can be selected based on its affinity for a biotinylated RNA bait. We demonstrate the effectiveness of this method that should enable the selection of RBPs with new specificity or improved affinity.**

## INTRODUCTION

RNA-binding proteins (RBPs) play essential functions in all post-transcriptional regulatory processes, including RNA processing, cellular localization, translation and mRNA decay. These proteins recognize RNA by using relatively few RNA-binding modules that combine to create versatile macromolecular binding surfaces to define their specificity (1–3). In addition to their RNA-binding domains (RBD), they possess auxiliary domains or modules responsible for their biological function. Thus, many times the RNA-binding activity can be separated from the biochemical activity responsible for RNA

degradation, localization etc. This modular architecture suggests that RBPs could be directed to act on non-cognate RNAs by altering the specificity of their RBD, without affecting the functional module significantly if at all. By doing so, it should be possible to control gene expression in a variety of ways using proteins whose specificity has been re-engineered.

The most abundant and most versatile RBP domain is the RNA recognition motif (RRM), also known as RBD or ribonucleoprotein domain (RNP) (1,2). Each RRM contains about 90–100 amino acids with two conserved regions, RNP-1 and RNP-2, and binds to RNA targets with varying sequences and structures. The combination of two or more RRMs allows the continuous recognition of RNA sequences of 8–10 nucleotides with strong binding affinity. In many ways, the RRM plays the same role in RBPs as the zinc-finger motif in DNA-binding proteins. Due to its simple, stable and modular structure and universally functional roles, zinc-finger motifs have been widely used as scaffolds for the construction of customized transcription factors to activate or repress gene expression (4). RRMs could be used in similar ways, as general RNA-binding molecules, if we could re-engineer their specificity by rational or combinatorial methods.

The structural basis of RNA recognition by many RBPs is now established, but specificity remains difficult to rationalize and to exploit towards the discovery of proteins with new activity. Recently a ‘code’ for RNA recognition has been deduced from crystal structures of proteins belonging to the Pumilio (Puf) family of proteins, where each of eight repeats binds to a nucleic acid base in a single-stranded RNA through hydrogen bonds via three amino acid side chains at conserved positions (5,6). This ‘code’ was used to engineer Puf proteins with predictably altered sequence specificity. Zinc fingers have also been used as modules to engineer new RNA recognition activities (7,8), but structures of RNA recognition by zinc fingers are limited compared to the RRM (9,10). Furthermore, available structures show that sequence-specific recognition by zinc finger proteins (ZFPs) to single-stranded RNA is not as straightforward as DNA

\*To whom correspondence should be addressed. Tel: +1 206 543 7113; Fax: +1 206 685 8665; Email: chenyu@u.washington.edu  
Correspondence may also be addressed to G. Varani. Tel: +1 206 543 7113; Fax: +1 206 685 8665; Email: varani@chem.washington.edu

recognition (1), with many contacts originating from the protein's main chain. Altogether, the very versatile RRM proteins provide the most diverse and probably the best opportunity to engineer RBPs to target non-cognate RNA sequences and structures, so as to direct post-transcriptional regulatory events similarly to how re-engineered ZFPs are used to regulate transcription.

Rational methods to design RBPs are yet to be fully developed (11,12). Therefore, we sought to develop an experimental method to evolve RBPs with new specificity. Ultimately, the two approaches will be complementary to each other and provide an improved strategy for the discovery of new RBPs, as demonstrated in many other instances of protein design (13). Phage display is a very well established method for the selection of proteins from diverse expression libraries (14), and it has been applied to study the specificity determinants of RBPs and to clone RBPs from cDNA libraries (15,16). However, the size of the libraries that can be constructed for phage display is limited by the transformation competency of bacterial cells, currently approximately  $10^8$ – $10^9$ /μg of vector DNA. Furthermore, several required *in vivo* steps, particularly the high-efficiency cloning required to construct large expression libraries in bacteria, make this method time-consuming and labor-intensive.

Recently, a totally cell-free selection strategy, called *in vitro* compartmentalization (IVC), has been developed to generate 'artificial cells' for the directed evolution of proteins (17,18). In this approach, individual genes are expressed in minute droplets of emulsified cell-free transcription-translation extract. These water-in-oil droplets can be as small as bacteria, with diameters of about 1 μm and volumes of less than a femtoliter. Formation of the droplets results in co-compartmentalization of the gene and its product, making it possible to select larger gene libraries ( $10^9$ – $10^{11}$  genes) compared to phage display. The cell-free selection strategy also makes it easier to carry out selection over a broad range of temperatures, pH and salt concentrations. However, the IVC method requires direct linkage between genotype and phenotype, and this has never been attempted, as far as we are aware, for RBPs. While various other strategies can be envisioned (19), we considered using high-affinity zinc finger DNA-binding proteins (ZFP-DNA) for genotype-phenotype linkage during selection (20). We reasoned that the highly stable and specific ZFP-DNA interaction would be able to survive the breaking of the emulsion and withstand the subsequent washing steps (20).

In this manuscript, we demonstrate the application of the IVC method to RBPs by using a fusion protein with six zinc-finger peptides linked to the RBD to be selected. The zinc-finger peptides bind with very high affinity and specificity to a cognate DNA sequence upstream of the T7 promoter in the DNA template, thereby allowing the linkage of the genotype with phenotype. We demonstrate the method by selecting an RBP from a model library and recovering both native and mutated U1A proteins with high affinity to a cognate RNA stem-loop. This strategy should allow the selection of RBPs with desired specificity for many functional applications.

## MATERIALS AND METHODS

### Construction of the ZFP-RBP

The cDNAs for the expression of ZIF268 and TFIIIA were ordered from Open Biosystems. Primers ZIF-1 and ZIF-2 as well as TF-1 and TF-2 were used to amplify the first three domains of ZIF268 and of TFIIIA using *Phusion* high-fidelity DNA polymerase (Finnzymes). Primers U1A-1 and U1A-2 were used to amplify the RRM of U1A. For preparing the background library, primers cHuD-1 and cHuD-2 were used to amplify the RRM2 of HuD protein (cHuD). The sequences of all the primers used are listed in Supplementary Table 1. PCR products were purified using a QIAGEN PCR purification kit. Overlap-extension PCR was subsequently performed to combine the three DNA fragments together to make ZIF268(1-3)-TFIIIA(1-3)-U1A, and ZIF268(1-3)-TFIIIA(1-3)-cHuD, which was then cut with EcoRI/NotI. The digested reaction products were gel purified and cloned into EcoRI/NotI-cut pET23a vector (Novagen), yielding constructs pET23a/ZNF-U1A and pET23a/ZNF-cHuD. All constructs were verified by DNA sequencing.

### Preparation of tandem ZNF-binding sites (ZNF-BS)

Primer BS-2 was annealed to oligonucleotide BS-1 encoding two ZNF-BS for ZIF268(1-3)-TFIIIA(1-3) (GGATG GGAGAC-GT-GCGTGGGCG) and extended with Klenow. The double stranded product was then cut with BglIII and BamHI, and subcloned into BglIII-cut pET23a, upstream of the T7 promoter. Clones were analyzed by BglIII/HindIII digestion and the clone yielding the longest fragment was cut again with BglIII and a further ZNF-BS cassette was inserted (20). After two cloning cycles, the plasmid was sequenced and found to contain four copies of the binding sites for ZIF268(1-3)-TFIIIA(1-3).

### Preparation of the *in vitro* expression cassette

Genes encoding the ZNF-U1A and ZNF-cHuD were excised from pET23a vectors by EcoRI/NotI digests and subcloned into pET23a/4BS, yielding two expression constructs pET23a/4BS/ZNF-U1A and pET23a/4BS/ZNF-cHuD, respectively. The corresponding linear expression templates, 4BS/ZNF-U1A and 4BS/ZNF-cHuD were amplified from pET23a/4BS/ZNF-U1A and pET23a/4BS/ZNF-cHuD by PCR using primers pET23-Fwd1\* and pET23-Rev1\* (Supplementary Table 1) using *Phusion* DNA polymerase. Both constructs harbor four tandem copies of the ZNF-BS. All PCR-amplified DNA templates were gel purified using the *Wizard* SV Gel and PCR Clean-Up System (Promega) and quantified by UV reading. The expression cassette fragment lengths were: 4BS/ZNF-U1A, 1750 bp; 4BS/ZNF-cHuD, 1735 bp.

### Preparation of model libraries

All PCR-amplified expression cassettes were diluted to 1 ng/μl final concentration in 0.15 mg/ml of HindIII-digested λ DNA (Ambion) as carrier in order to avoid non-specific adsorption to plastic vials at low concentrations. To achieve the  $1:10^4$  spiking of ZNF-U1A in

ZNF-cHuD, the target construct ZNF-U1A was serially diluted 10-fold per step in competitor construct present at a concentration of 1 ng/ $\mu$ l.

Libraries of U1A protein mutants (Y13, F56) were prepared by overlay extension PCR with primers containing NNS (where N is a mixture of G, A, T and C, and S is a mixture of G and C) at the Y13 and F56 positions of wild-type U1A protein. The codon NNS encodes all 20 amino acids but not the two termination codons UAA and UGA. The final complete templates were amplified by PCR using the same primers pET23-FWD1\* and pET23-Rev1\*. Sequencing was carried out afterwards to check the correctness of the mutations.

### Preparation of RNA-coated beads

Streptavidin-coated M280 Dynabeads (20  $\mu$ l) were washed according to the manufactory's protocol, i.e. twice with the same volume of solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 50 mM NaCl), one time with solution B (DEPC-treated 0.1 M NaCl) and one time with the binding and washing (B&W) buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 1.0 M NaCl). After washing, the beads were re-suspended in B&W buffer and 4  $\mu$ l of biotinylated RNA (10  $\mu$ M, IDT) were added in the ratio of 1 mg beads to 200 pmol of RNA. The mixture was incubated at room temperature for 15 min while slowly rotating the tube. A magnet was used to separate the beads, which were by then coated with the biotinylated RNA. Excess unbound RNA was washed away with three washes in buffer A (Tris-buffered saline, 1% (w/v) bovine serum albumin (BSA), 0.1% (w/v) sonicated salmon sperm DNA, 50  $\mu$ M ZnSO<sub>4</sub>, 2 mM DTT, 0.1% (v/v) Tween-20, Rnase inhibitor 1 u/ $\mu$ l). Magnetic beads were re-suspended in 20  $\mu$ l buffer A and kept on ice until needed.

### In vitro selection of RBPs

The TNT T7 Linear Template Expression kit (Promega) was used to prepare the *in vitro* coupled transcription-translation mixture. Briefly, 50  $\mu$ l of reaction mixes were assembled on ice, including 40  $\mu$ l TNT T7 PCR Quick Master Mix supplemented with 1  $\mu$ l methionine (1 mM) and 500  $\mu$ M of ZnSO<sub>4</sub>. For <sup>35</sup>S-labeled protein, [<sup>35</sup>S]methionine (1 mM) was added instead of unlabeled methionine. The reaction was initiated with PCR DNA template (1 ng/ $\mu$ l) in 0.1 mg/ml of HindIII-cut  $\lambda$  DNA. The reaction mixture was added immediately to 1 ml of oil mix (white mineral oil containing 4.0% ABIL EM 90) (21), while stirring at 1600 rpm in a 14 ml round-bottom tube cooled on ice. After 5 min, the emulsion was transferred to a 1.5 ml Eppendorf tube and incubated at 30°C for 90 min. At the end of the incubation, the aqueous droplets were spun down in a microcentrifuge for 10 min at 8000 rpm. The upper (oil) phase was removed and after a short second spin to remove the last remnants of oil, 100  $\mu$ l of breaking buffer A (Tris-buffered saline, 1% (w/v) bovine serum albumin (BSA), 0.1% (w/v) sonicated salmon sperm DNA, 50  $\mu$ M ZnSO<sub>4</sub>, 2 mM DTT, Rnase inhibitor 1 u/ $\mu$ l) together with 50  $\mu$ g of tRNA were added to the pellet of aqueous phase droplets. The droplets were

broken by hexane-extraction that was repeated at least four times. Residual solvent was removed from the broken emulsion by centrifuging under vacuum for 5 min at 25°C. RNA-coated magnetic beads were added to the hexane-extracted aqueous phase and the mixture was incubated at room temperature for 45 min with slow rotation of the tube. Afterwards, the beads were washed three times with buffer A and two times with TBS buffer. The final separated beads were re-suspended in 20  $\mu$ l of Tris buffer (pH 7.4), and the attached gene population was recovered and amplified for the second round of selection by nested PCR using primers pET23-Fwd2\* and pET23-Rev2\*.

In order to measure the selection efficiency of the targeted gene from the model library, the recovered gene populations from the first and second round of PCR amplification were digested with EcoRI/NotI and cloned back into pET-23a plasmid. Transformations were performed using XL10-gold ultra-competent cells (Stratagene). Colony PCR was carried out the next day using the U1A specific primers U1A-1 and U1A-3 (Supplementary Table 1).

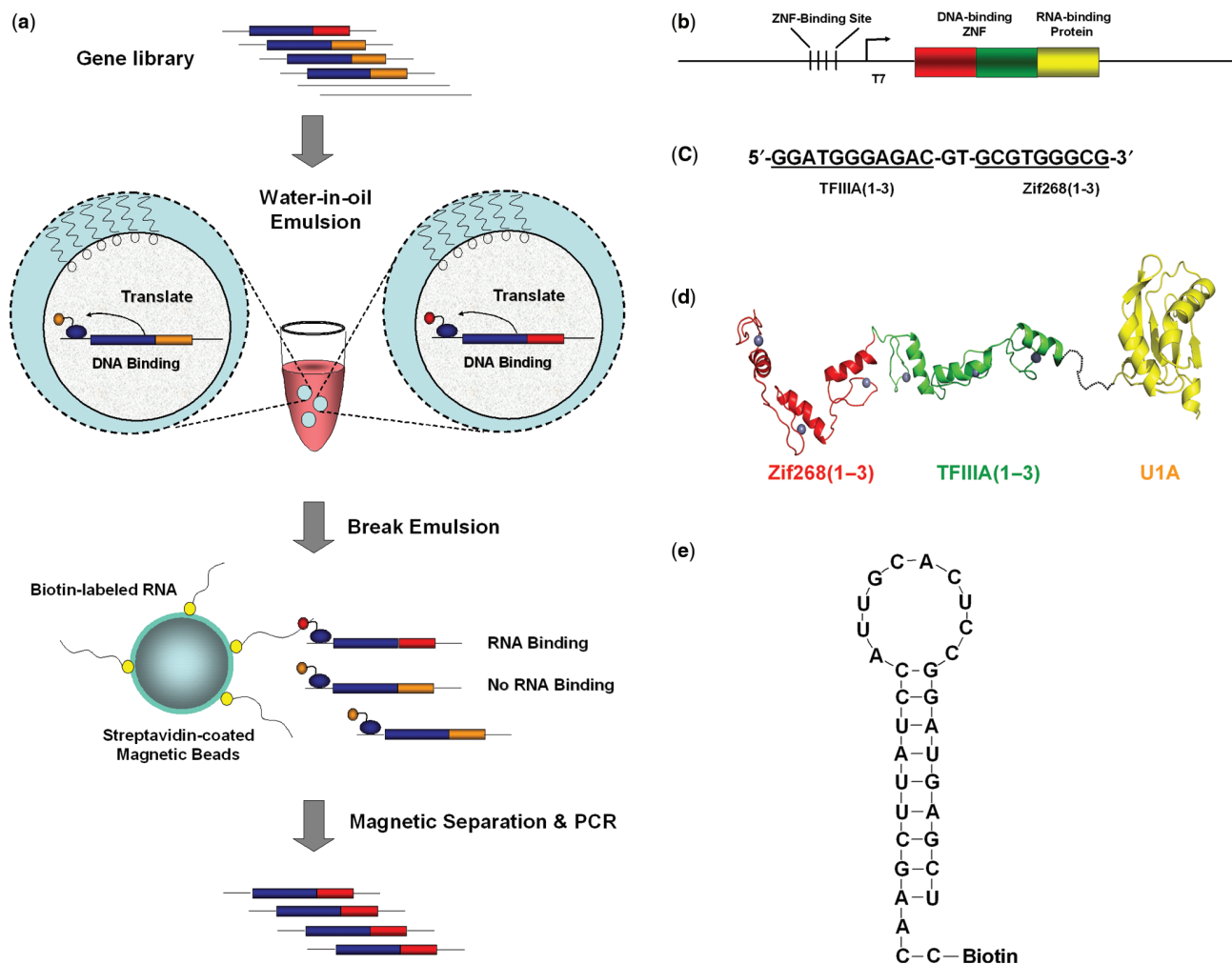
### Gel-shift assays

Selected proteins from IVC experiments were assayed for their DNA- or RNA-binding activity by using <sup>32</sup>P end-labeled synthetic DNA or RNA containing the required binding sequences for ZNF proteins or U1A. The sequences of the DNA and RNA containing the protein-binding sites are shown in Figure 1c and e.

*Determination of the active protein concentration.* To determine the concentration of ZNF-U1A fusion protein produced in the TNT T7 *in vitro* expression system, crude protein samples were directly used in gel-shift assays against a dilution series of the DNA containing appropriate ZNF-BS. Binding site concentration was always well above the estimated dissociation constant ( $K_d$ ) of the protein, but ranged from a higher concentration, at which all available protein binds DNA, to a lower concentration, at which all DNA is bound. Controls were carried out to ensure that labeled DNA was not shifted by the *in vitro* extract in the absence of ZFP. Binding reactions included crude protein samples and varying concentrations of DNA duplex mixed with 20 pM of radiolabeled DNA incubated in binding buffer (20 mM Tris propane, pH 7.4, 100 mM NaCl, 10 mM DTT, 0.1 mg/ml BSA and 0.1% Triton X-100) at room temperature for 1 h. The reaction mixtures then were separated on a 7% native polyacrylamide gel at 4°C for 2 h. Radioactive signals were quantified by PhosphorImager analysis (Molecular Dynamics) to determine the amount of shifted binding site and, from that, the concentration of active ZFP. The concentration of linked U1A protein should be the same as zinc-finger peptide.

*Determination of the binding affinity.* Dissociation constants were determined in parallel to the calculation of active peptide concentration. Serial dilutions of crude protein preparations were made and incubated with 5 pM <sup>32</sup>P-labeled DNA containing ZNF-BS or 10 pM





**Figure 1.** Directed evolution of RBPs using IVC. (a) Schematic representation of the strategy used in selection of RBPs by IVC. RBPs are fused to an N-terminal six zinc-finger DNA-binding poly-peptide which recognizes a cognate sequence presented in four copies upstream of the coding region in the linear DNA templates. The ZNF-RBP genes are compartmentalized in water-in-oil emulsion and allowed to express the corresponding fusion protein by coupled *in vitro* transcription-translation reaction. The expressed chimeric proteins bind to their encoding DNA templates through the zinc-fingers. After breaking the emulsions, streptavidin-coated magnetic beads bound with biotin-labeled RNA were used to capture the RBPs and corresponding encoding DNA; the selected gene expression cassettes are subsequently amplified by PCR. (b) Expression cassettes used in the selection of ZNF-RBP fusion proteins: the six-finger Zif268(1-3)-TFIIIA(1-3) protein is fused with an RBP C-terminus to it. Four copies of the DNA-binding sequence for the ZNF's were inserted upstream of the T7 promoter. (c) DNA-binding sequence for the six-finger ZNF protein. (d) Model of the Zif268(1-3)-TFIIIA(1-3)-U1A chimeric protein based on the structures of the corresponding proteins (30,38,39). (e) The secondary structure of the biotin-labeled U1 RNA used in the IVC experiments.

labeled U1 RNA at room temperature for 1 h. Samples were run on 7% native polyacrylamide gels and the radioactive signals were quantified by PhosphorImager analysis.

## RESULTS AND DISCUSSION

The application of *in vitro* selection methods to the discovery of RBPs with new specificity has lagged considerably behind their application to DNA-binding proteins (4,20). In part, this has occurred because the phage display method is not as well suited for proteins such as the RRM, compared to zinc fingers for example, because of the limited diversity of phage libraries relative to the size of RNA-protein interfaces (1). We wished to explore whether

the IVC method (Figure 1a) could be adapted for RBPs, because of the increased library size and totally cell-free selection strategy that the method provides. In order to apply this method to RBPs, however, there is a need to directly link genotype and phenotype during selection. Here we present an approach to achieve genotype-phenotype linkage derived from a method used in the selection of zinc finger DNA-binding proteins (20), and a new selection scheme to verify the performance of the method.

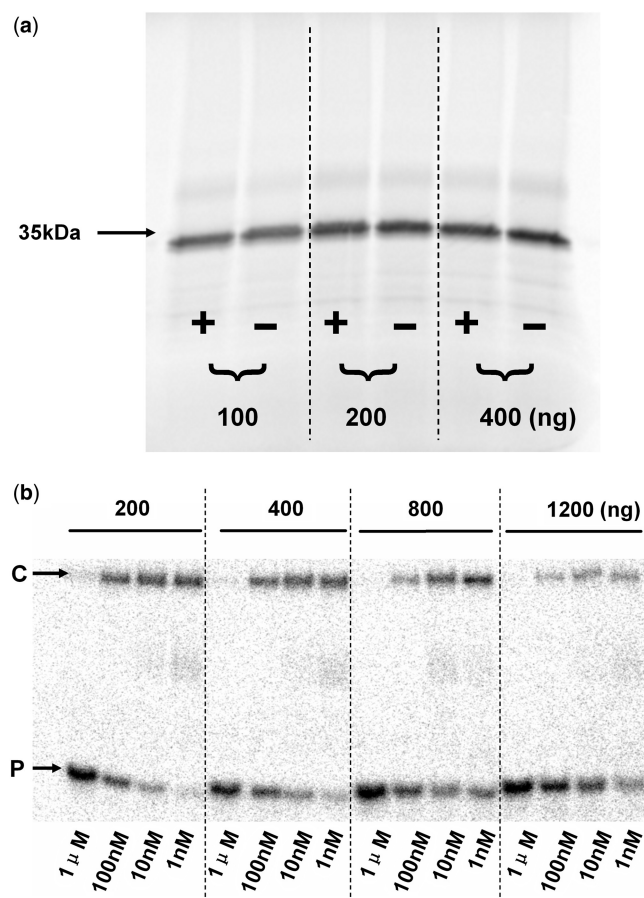
### Linkage of genotype and phenotype during selection of RBPs

In order to link genotype and phenotype, we decided to avoid chemically or enzymatically cross-linking the encoding DNA with the expressed protein because of the

expected low efficiency of such an approach (19). Instead, the RRM we wanted to select (derived from UIA and HuD proteins) were fused downstream of a six-finger DNA-binding domain comprising a head-to-tail fusion of the first three zinc fingers of TFIIIA (TFIIIA(1-3)) to the three zinc fingers of Zif268 protein (Zif268(1-3)) (Figure 1b). We used the extended linker peptide GGGSERP between zinc fingers 3 and 4 of the construct [i.e. between Zif268(1-3) and TFIIIA(1-3)], because this sequence has been reported to significantly increase the DNA binding affinity of a six zinc-finger peptide [Zif268(1-3)-NRE(1-3)], leading to a femtomolar dissociation constant (22). This affinity is well above all known dissociation constants of RRM-containing RBPs, including UIA which has one of the highest RNA-binding affinities known so far ( $K_d$  of about 10–100 pM). The very high DNA-binding affinities is also accompanied by a low  $K_{off}$  rate, i.e. a very long half-life for the complex (22,23). This property is essential for genotype–phenotype linkage because the complex must remain intact for the time required for breaking the emulsions as well as the washing and affinity selection steps. The half life of a high affinity ZNF–DNA complex with  $K_d = 3$  pM is at least 2 h (20). In order to ensure complete binding of the ZNF–RBP fusion protein to its encoding DNA template, four copies of the DNA binding sites for Zif268(1-3)-TFIIIA(1-3) construct were appended in all genes in the library. The presence of multiple copies of the DNA binding sites increases the local concentration of ZNF-binding targets, favoring tight complex formation. However, we did not further investigate the effect of the number of binding sites on the enrichment and fidelity of the selection. Given these considerations, we expected the protein–DNA complex to remain intact under conditions where the protein–RNA complex dissociates and during the time required for all selection steps.

The expression of the chimeric ZNF-UIA protein using TNT T7 *in vitro* transcription/translation system was demonstrated as shown in Figure 2a. In order to visualize the protein,  $^{35}\text{S}$ -methionine was added at the beginning of the reaction. A strong band was observed by SDS PAGE gel at the correct protein size, ca. 300 amino acids (Figure 2a). In order to optimize protein expression levels, we varied the amount of PCR DNA template added in the presence or absence of extra T7 RNA polymerase, which had been suggested to increase the protein production (20). However, protein expression levels appeared to reach a plateau between 200 ng and 400 ng of PCR DNA template, and adding extra T7 RNA polymerase did not increase protein production; in some cases, it even decreased it.

When proteins are selected using *in vitro* evolution methods, multiple sequences with various binding affinities are generally found. Expressing the mutant protein individually and testing the binding affinity for each clone is a bottleneck for the entire project because this step is laborious and time consuming. We sought to reduce the time and labor requirements for this task by adopting a strategy previously used with ZFP design (24), i.e. by using the proteins directly expressed in an *in vitro* transcription/translation system to measure approximate



**Figure 2.** Optimization of the conditions for *in vitro* transcription/translation of ZNF-UIA fusion protein. (a) SDS PAGE gel analysis of the expression of ZNF-UIA protein construct labeled with [ $^{35}\text{S}$ ]methionine with or without the addition of supplemental T7-RNA polymerase (1.5  $\mu\text{l}$  at 1 mg/ml); PCR DNA templates were added in the indicated amounts of 100 ng, 200 ng and 400 ng; (b) Comparison of protein expression levels at different concentrations of PCR DNA template (200 ng, 400 ng, 600 ng and 800 ng) by EMSA. The concentrations of DNA duplexes are indicated at the bottom of each gel. 20 pM of  $^{32}\text{P}$ -labeled DNA was added to each binding reaction. The ZNF-UIA fusion protein expressed directly from *in vitro* transcription/translation reaction binds to its target DNA and causes the upward shift of the DNA band on the gel. The positions of the free probe (P) and complexes (C) are marked.

binding constants. In order to do so, it was first necessary to know the concentration of protein expressed in the *in vitro* transcription/translation reaction. We did so by taking advantage of the high affinity of the six chimeric zinc fingers to their cognate DNA, as shown in Figure 2b. The DNA templates used for expression of the ZNF-UIA protein were amplified with primers excluding the ZFP-BS. The concentrations of target DNA in the final binding reactions were varied from 1  $\mu\text{M}$  to 1 nM. When the intensity of the DNA band shifted upon formation of the protein–DNA complex is equal to the intensity of the free DNA band, then the protein concentration is half of the DNA concentration used in the assay. Since the protein in the crude reaction mixture was mixed with the target DNA in equal volumes, the results show that the concentration of expressed ZNF-UIA protein reached 100 nM when 200 ng and 400 ng DNA templates were added in a

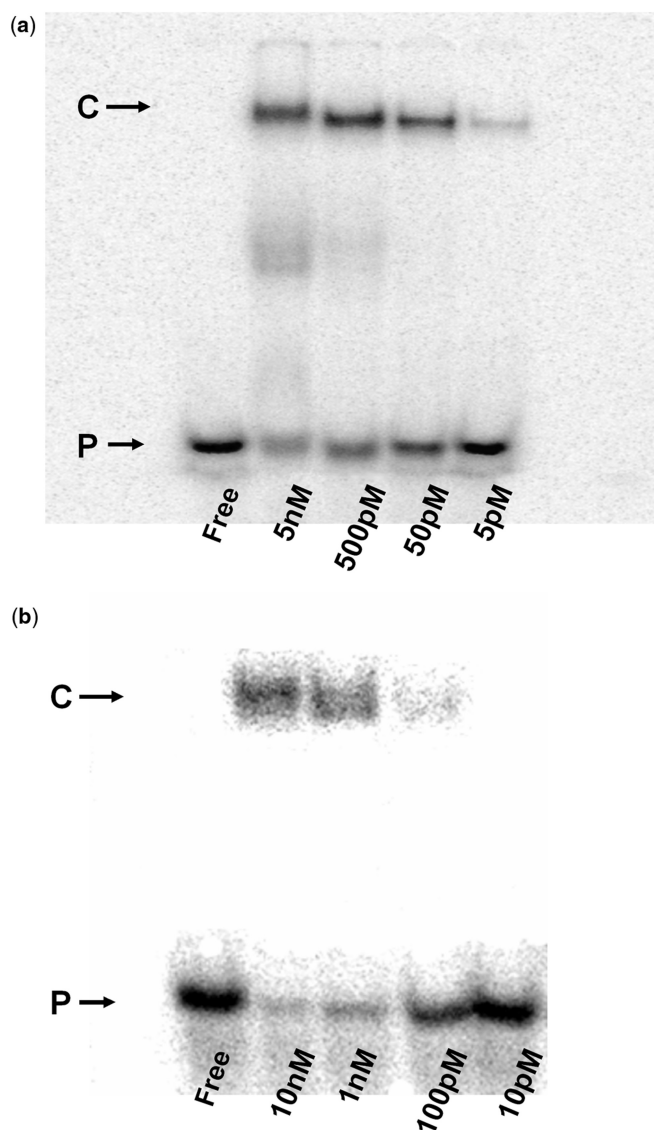
50  $\mu$ l of reaction. Increasing the DNA template concentration further to above 400 ng had negative effects on protein production, consistent with the results obtained with  $^{35}$ S-labeling (Figure 2a).

Binding of the protein construct containing six zinc-fingers to its target DNA has been reported to occur with femtomolar to picomolar dissociation constants (22,24), and U1A protein alone binds to the stem/loop II of the U1 snRNA with picomolar to nanomolar affinity (25,26). However, attaching the two together might affect the binding affinity of each protein. The linker between the poly-zinc-finger and U1A consists of only a short peptide H-P-P-T. The presence of two Pro residues should lead to structural rigidity and the isolation of the linker from the attached nucleic acid binding domains, thus preventing unfavorable interactions between folded domains (27). In fact, two prolines in a row favor the polyproline, or collagen, conformation, which is an extended but not  $\beta$ -sheet structure. We did not add a longer linker because the N-terminus of U1A has a short unstructured sequence pointing away from the RRM domain at its very end (30,31), which can be viewed as part of an extended linker. Gel shift assays show that the chimeric protein has a binding affinity to its cognate DNA of approximately 50 pM (Figure 3a), and to U1 RNA of less than 1 nM (Figure 3b). The DNA-binding affinity for our zinc-finger peptide Zif268(1-3)-TFIIIA(1-3) is lower than the reported femtomolar dissociation constant for a six zinc-finger peptide [Zif268(1-3)-NRE(1-3)] that used the same linker GGGSERP [22]. This difference is probably due to different zinc-finger peptides used, and perhaps to the effect of the attached U1A protein domain. Based on previous studies on poly-zinc-finger proteins (22,24,28), optimizing the length and/or the rigidity of the linker between the poly-zinc-finger protein and U1A might further increase the affinity. We did not do this at this stage because the observed affinities are sufficient for our purposes.

#### Selection of U1A protein-coding gene from a very large background of a related gene

In order to demonstrate that the recovery of genes with desired RNA-binding specificity and affinity could be accomplished efficiently, we performed a model selection experiment. In this control test, we selected a target RBP (U1A) present at very low concentration ( $1 \times 10^{-4}$  ng/ $\mu$ l) from a library containing a large excess of a competitor protein, in this case the second RRM of HuD (cHuD). U1A protein uses its single RRM domain to bind to the loop region of a stem-loop structure with high sequence specificity and affinity (pM to nM) (25,26). In contrast, HuD protein contains three RRMs which bind cooperatively to a single-stranded AU-rich RNA with nM affinity (29). Therefore, although the two RRMs from U1A and HuD have approximately the same structure, they have different RNA binding specificity.

A total of  $10^9$  clones (corresponding to 1.66 fmol) of a  $1:10^4$  mixture of ZNF-U1A:ZNF-cHuD genes were added to a 40  $\mu$ l TNT T7 *in vitro* transcription-translation extract and homogenized into an oil-surfactant mixture to



**Figure 3.** The binding of ZNF-U1A protein produced directly from *in vitro* transcription/translation reaction to both the DNA (a) and the RNA targets (b). Binding was analyzed by EMSA, with decreasing concentrations of proteins equilibrated with 5 pM of  $^{32}$ P-labeled DNA duplex or 10 pM of  $^{32}$ P-labeled U1 RNA. Protein concentrations are indicated at the bottom of gel. The positions of the free probe (P) and complexes (c) are indicated.

create a water-in-oil (w/o) emulsion. Under these conditions, the majority of droplets contain no more than a single gene in addition to all of the molecular machinery needed to express that gene. Once the protein is expressed, DNA binding by the Zif268(1-3)-TFIIIA(1-3) fusion polypeptides should occur within the same droplets (Figure 1a). After breaking the emulsion and following stringent washing, streptavidin-coated *Dynabeads*, to which biotinylated U1 RNA was bound, were used to pull down U1A protein with its attached encoding DNA. Through this scheme, the genotype remains coupled to its phenotype throughout the affinity selection and washing steps, and the selected gene is enriched by PCR in the final step of the protocol. In contrast, the



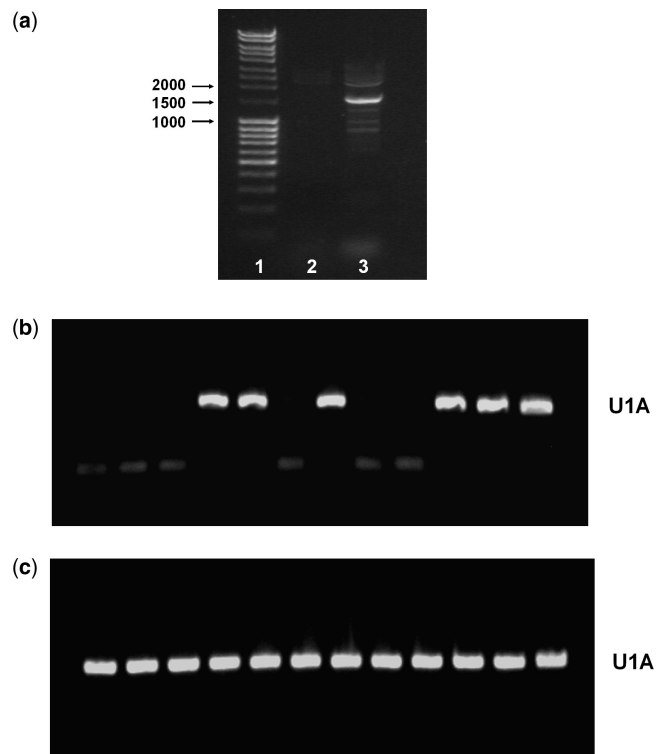
competitor ZNF-cHuD protein would fail to bind to the U1 RNA and therefore the population would be enriched only for U1A protein. A negative control reaction with no ZNF-BS inserted upstream of T7 promoter was carried out in parallel with the selection experiment.

After the first round of selection, a clear band of PCR products was recovered at the correct size of the DNA templates, while the negative control reaction gave no specific product (Figure 4a). This indicates that an interaction between the expressed ZNF fusion protein and the DNA template is required for efficient recovery of the genetic material. Since we do not know whether the amplified PCR product is ZNF-U1A or background ZNF-cHuD, the PCR-amplified DNA templates were gel-purified and cloned back into pET23a plasmid and transformed. Colony PCR showed that about 50% of the genes recovered from the first round of selection are already ZNF-U1A species (Figure 4b), suggesting at least a 1000-fold enrichment from the first round of selection. The ZNF-cHuD species do not bind to the U1A RNA, but some clones remain after the first round of selection probably due to its large excess in the initial library and the non-specific binding to the Dynabeads.

The gel-purified DNA templates that emerged from the first round of selection were then submitted to the next round of selection. DNA recovered after the second round of IVC selection was sub-cloned and colony PCR was again used to determine the percentage of recovery of the ZNF-U1A gene. As shown in Figure 4c, all genes recovered at this stage corresponded to ZNF-U1A, confirming that the target gene had been selected from the library efficiently. Out of the 12 selected colonies that were subjected to colony PCR after the second round of selection, none was found to contain the ZNF-cHuD gene, indicating that non-specific binding of template DNA to the Dynabeads is at undetectable levels after two rounds of selections. The genes recovered after two rounds of selection were further sequenced; the results confirm that the target U1A gene has been selected from the library in all 12 cases.

#### Recovery of high-affinity U1A protein mutants from a randomized U1A library

As a second test of the ability of the method to select RBP with high affinity for a defined RNA target, we probed whether the wild-type U1A sequence as well as other high-affinity mutants could be recovered from a library containing random mutations at two positions. Aromatic residues at three conserved positions on the  $\beta$ -sheet surface of most RRM domains are crucial to the ability of RRM proteins to bind strongly to RNA. U1A protein has two such aromatic residues, one in RNP1 (F56) and one in RNP2 (Y13), which are very important for RNA complex formation and stability (30–32). Crystallographic and NMR studies have shown that these two residues stack with bases of the target RNA (Figure 5), and form a network of interactions that hold the RNA in place against the RRM surface. Various mutational studies have also been carried out to evaluate the contributions of these two amino acids to the

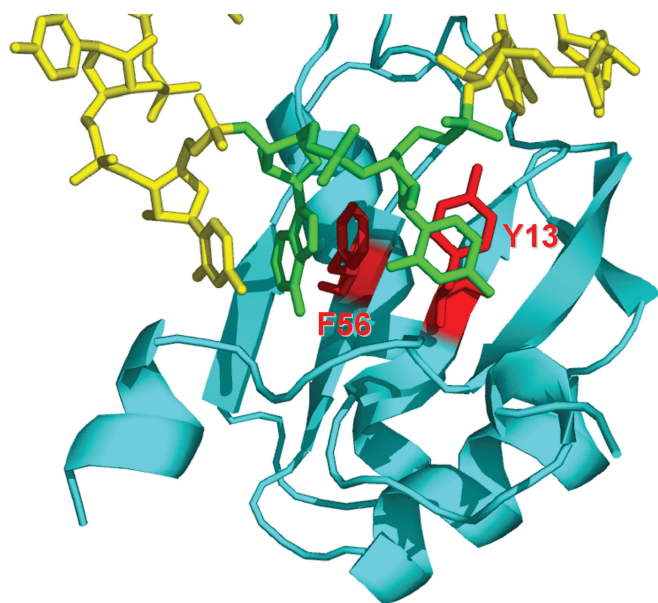


**Figure 4.** Efficient recovery of wild-type U1A protein from a high background of a non-cognate RBP, the second RRM of HuD protein (cHuD). (a) Gel electrophoresis showing the results of the first round of selection after PCR amplification. Lane 1 is the size marker MassRuler™ DNA ladder (Fermentas); lane 2 is the control selection executed with the DNA template without any ZNF-BS; lane 3 is the result of the selection from the model library. (b) Colony PCR showing the results of the first round of IVC selection using U1A-specific primers. Six out of 12 selected colonies contain the U1A gene; the other colonies, presumably containing the cHuD gene, give rise to a non-specific PCR product at low molecular weight. (c) Colony PCR showing the results of the second round of IVC selection using the same U1A-specific primers; all 12 colonies contain the U1A gene.

interaction between U1A protein and RNA (33,34). The results indicate that although the two aromatic amino acids provide little specificity, their interactions contribute significantly to the strength of RNA binding as evidenced by the substantial loss of affinity observed when they are individually mutated. These results make the two amino acids ideal for testing our selection method.

Overlap extension PCR was used to generate the DNA template library that contains the randomized mutations (NNS) at the two positions (Y13 and F56). After two rounds of selection, the selected genes were cloned into pET23 and the identity of 20 selected genes was confirmed by sequencing. Linear PCR DNA templates containing the T7 promoter for each selected gene were then generated and used directly in protein production through the TNT T7 *in vitro* transcription/translation system. The expressed proteins were subsequently used in EMSA binding assays to assess the RNA binding affinity of each selected U1A protein.

From 20 sequenced clones, we recovered 10 distinct sequences, including the wild-type U1A protein and nine mutants. The sequences of wild-type U1A and each



**Figure 5.** Structure of the U1A-RNA interface (pdb code 1URN). The two amino acids (Y13 and F56) randomized in the library are highlighted in red and the RNA bases with which these amino acids stack are in green. The rest of the protein and RNA are colored in cyan and yellow, respectively.

of three combinations [(R13, Y56), (R13, F56) and (R13, W56)] appear twice in the selection. Among the 10 different sequences, seven protein mutants plus the wild-type U1A (eight in total) gave detectable RNA binding in EMSA assays. The results for these eight proteins are summarized in Figure 6. Amino acid 56 is conserved as aromatic in our selection, but can be any of F, Y or W. This result is consistent with previous mutation studies, which indicated that substitution of F56 with other aromatic amino acids led to only a small loss in binding energy (34). This position is occupied by Phe in 74% of all RRM, Tyr in 10% of RRM but only rarely by Trp (35). However, in U1A proteins from different species, Trp occasionally replaces Phe at this position, but Tyr never does (36). The selection experiment did not return His at position 56, although F56H has close to wild-type affinity (34). This result could simply be due to the limited number of clones analyzed in this test.

In contrast to position 56, position 13 is not necessarily aromatic, since Arg in addition to Phe and Tyr is selected as well. Consistent with previous reports (25,33,37), the Y13F substitution and (F13, W56) double mutant had 100-fold lower binding affinity compared to wild-type U1A. However, Arg was never reported to be able to replace Tyr13. Among the Y13R mutants, (R13, W56) has the highest binding affinity at ca. 10 nM. The (R13, Y56) and (R13, F56) pairs both have binding affinity of about 100 nM, similar to (F13, W56) and (F13, F56). Interestingly, the mutant (F13, Y56), which has the same amino acid content as wild-type U1A but in reverse, was not selected. This result is consistent with previous report that the  $K_d$  of this mutant protein is  $10^5$  times higher than wild-type (26), due to the disruption of a hydrogen-bonding network involving Gln54. Among the 10 tested

WT	Mut 1	Mut 2	Mut 3	Mut 4	Mut 5	Mut 6	Mut 7
Y13 F56	Y13 Y56	Y13 W56	R13 W56	R13 Y56	R13 F56	F13 W56	F13 F56
1nM	10nM	1nM	10nM	100nM	100nM	100nM	100nM

**Figure 6.** Binding affinities ( $K_d$ ) determined by EMSA experiments of U1 RNA to wild-type and mutant U1A proteins selected from a randomized library using our IVC method. The sequence of U1 RNA used in the binding assay is shown in Figure 1e, except that the biotin tag was removed.

sequences, only (L13, W56) and (V13, W56) did not show any band shift in the EMSA experiments; presumably their binding affinities are above 100 nM, which is our detection limit in EMSA, as determined by the maximum protein concentration expressed in our *in vitro* transcription/translation reaction. Both Leucine and Valine have a fairly large van der Waals interaction surface (1); they are sometimes found to replace aromatic residues in RRM surfaces to form hydrophobic interactions with the RNA bases, although stacking by aromatic side-chains should provide higher affinity than non-aromatic side-chains. The selection of Leu and Val at position 13, while Trp is not selected, suggests that this position is sterically more restricted than position 56. Increasing the selection stringency by using a higher concentration of bait RNA and adding a competitor RNA (in addition to tRNA) that is closer in structure to the bait will probably further reduce the number of selected low affinity RBPs (16).

## CONCLUSIONS

We have for the first time applied the emulsion-based IVC method to select RBPs with defined specificity. A new approach to link genotype and phenotype was developed based on fusing the target RBP to six ZFP that bind to a cognate DNA sequence inserted upstream of the promoter that drives protein expression. The expressed ZFP-RBP binds to its encoding DNA with high affinity while the fused RBP can be selected based on its affinity for a biotinylated RNA bait. This method is much more efficient, easier to implement and robust than the alternative of physically linking the RBP to its encoding DNA using chemical or enzymatic methods. We demonstrate that the IVC selection method works efficiently by selecting U1A protein from a high background of a different gene, and by selecting native and mutant U1A proteins with high affinity to the U1 RNA from a library randomized at two positions. The effectiveness and simplicity of this method should enable the generation of RBPs with high affinity and specificity for many desired RNA sequences.



## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Institutes of Health–National Cancer Institute and National Institutes of Health–National Institute of General Medical Sciences (G.V.). Funding for open access charge: NIH-NIGMS.

*Conflict of interest statement.* None declared.

## REFERENCES

- Auwater, S.D., Oberstrass, F.C. and Allain, F.H.T. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.*, **34**, 4943–4959.
- Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nature Rev. Mol. Cell Biol.*, **8**, 479–490.
- Chen, Y. and Varani, G. (2005) Protein families and RNA recognition. *FEBS J.*, **272**, 2088–2097.
- Pabo, C.O., Peisach, E. and Grant, R.A. (2001) Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annual Rev. Biochem.*, **70**, 313–340.
- Cheong, C.G. and Hall, T.M.T. (2006) Engineering RNA sequence specificity of Pumilio repeats. *Proc. Natl Acad. Sci.*, **103**, 13635–13639.
- Wang, X., McLachlan, J., Zamore, P.D. and Hall, T.M.T. (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell*, **110**, 501–512.
- McCull, D.J., Honchell, C.D. and Frankel, A.D. (1999) Structure-based design of an RNA-binding zinc finger. *Proc. Natl Acad. Sci.*, **96**, 9521–9526.
- Friesen, W.J. and Darby, M.K. (1998) Specific RNA binding proteins constructed from zinc fingers. *Nat. Stru. Biol.*, **5**, 543–546.
- Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Stru. Mol. Biol.*, **111**, 257–264.
- Lu, D., Searles, M.A. and Klug, A. (2003) Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition. *Nature*, **426**, 96–100.
- Chen, Y., Kortemme, T., Robertson, T. and Varani, G. (2004) A new hydrogen-bonding potential for the design of protein-RNA interactions predicts specific contact and discriminate decoys. *Nucleic Acids Res.*, **32**, 5147–5162.
- Zheng, S., Robertson, T.A. and Varani, G. (2007) An knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J.*, **274**, 6378–6391.
- Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A. et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–195.
- Kehoe, J.W. and Kay, B.K. (2005) Filamentous phage display in the new millennium. *Chem. Rev.*, **105**, 4056–4072.
- Laird-Offringa, I.A. and Belasco, J.G. (1995) Analysis of RNA-binding proteins by in vitro genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc. Natl Acad. Sci.*, **92**, 11859–11863.
- Danner, S. and Belasco, J.G. (2001) T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl Acad. Sci.*, **98**, 12954–12959.
- Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nature Biotech.*, **16**, 652–656.
- Miller, O.J., Bernath, K., Agresti, J.J., Amitai, G., Kelly, B.T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D.S. et al. (2006) Directed evolution by in vitro compartmentalization. *Nature Methods*, **7**, 561–570.
- Griffiths, A.D. and Tawfik, D.S. (2006) Miniaturizing the laboratory in emulsion droplets. *Trends in Biotech.*, **24**, 395–402.
- Sepp, A. and Choo, Y. (2005) Cell-free selection of zinc finger DNA-binding proteins using in vitro compartmentalization. *J. Mol. Biol.*, **354**, 212–219.
- Ghadessy, F.J. and Holliger, P. (2004) A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Eng. Des. Sel.*, **17**, 201–204.
- Kim, J.S. and Pabo, C.O. (1998) Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl Acad. Sci.*, **95**, 2812–2817.
- Moore, M., Klug, A. and Choo, Y. (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl Acad. Sci.*, **98**, 1437–1441.
- Moore, M., Choo, Y. and Klug, A. (2001) Design of polyzinc finger peptides with structured linkers. *Proc. Natl Acad. Sci.*, **98**, 1432–1436.
- Jessen, T.–H., Oubridge, C., Teo, C.H., Pritchard, C. and Nagai, K. (1991) Identification of molecule contacts between the U1A small ribonucleoprotein and U1 RNA. *EMBO J.*, **10**, 3447–3456.
- Kranz, J.K. and Hall, K.B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J. Mol. Biol.*, **285**, 315–231.
- George, R.A. and Heringa, J. (2003) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.*, **15**, 871–879.
- Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F. III (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl Acad. Sci.*, **94**, 5525–5530.
- Park-Lee, S., Kim, S. and Laird-Offringa, I.A. (2003) Characterization of the interaction between neuronal RNA-binding protein HuD and AU-rich RNA. *J. Biol. Chem.*, **278**, 39801–39808.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
- Allain, F.H.T., Howe, P.A., Neuhaus, D. and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5772.
- Avis, J.M., Allain, F.H., Howe, P.W., Varani, G., Nagai, K. and Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
- Law, M.J., Chambers, E.J., Katsamba, P.S., Haworth, I.S. and Laird-Offringa, I.A. (2005) Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction. *Nucleic Acid Res.*, **33**, 2917–2928.
- Shiels, J.C., Tuite, J.B., Nolan, S.J. and Baranger, A.M. (2002) Investigation of a conserved stacking interaction in target site recognition by the U1A protein. *Nucleic Acids Res.*, **30**, 550–558.
- Birney, E., Kumar, S. and Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.
- Simpson, G.G., Clark, G.P., Rothnie, H.M., Boelens, W., Van Venrooij, W. and Brown, J.W.S. (1995) Molecular characterization of the spliceosomal proteins U1A and U2B' from higher plants. *EMBO J.*, **14**, 4540–4550.
- Kranz, J.K., Lu, J. and Hall, K.B. (1996) Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci.*, **5**, 1567–1583.
- Nolte, R.T., Conlin, R.M., Harrison, S.C. and Brown, R.S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl Acad. Sci.*, **95**, 2938–2943.
- Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.8 Å. *Science*, **252**, 809–817.