‘Protein Engineering with Zinc Fingers and TALs’

A thesis submitted towards partial fulfilment of BS-MS Dual Degree Program

(August 2012- April 2013)

By

Soumitra Athavale

(20081053)

Supervisor

Dr Sanjeev Galande, Professor, Centre for Excellence in Epigenetics, IISER Pune.

Department of Biology

Indian Institute of Science Education and Research

Pune-411021

INDIA

Certificate

This is to certify that this dissertation entitled ‘Protein Engineering with Zinc Fingers and TALs’ towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research (IISER), Pune represents original research carried out by Soumitra Athavale at IISER Pune under the supervision of Dr. Sanjeev Galande, Professor, Biology Division, IISER Pune during the academic year 2012-2013.

Dr. Sanjeev Galande
Professor
Biology Division, IISER Pune

Declaration

I hereby declare that the matter embodied in the thesis entitled ‘Protein Engineering with Zinc Fingers and TALs’ are the results of the investigations carried out by me at the Biology Division, IISER Pune under the supervision of Dr. Sanjeev Galande Professor, Biology Division, IISER Pune and the same has not been submitted elsewhere for any other degree

Soumitra Athavale
BS-MS Dual Degree Student
Reg No. 20081053
IISER Pune

Acknowledgments

Working on this project has added to my knowledge and experience in molecular biology and enriched my confidence in posing and tackling research problems. These initial steps into research were guided by an excellent infrastructure and intellectual freedom. I express my deep gratitude to Dr. Sanjeev Galande for providing this atmosphere where I could think independently and fearlessly. Apart from research, I have also learnt from Sanjeev a lot about the finer aspects of communication, peer interactions and professional conduct and for this, I am indebted to him.

It was a pleasure to work alongside members in Sanjeev's group who were always helpful and considerate towards me. In particular, I would like to express my gratitude towards Rahul, Rafeeq, Sunita, Kamal and Dr. Ashwin for stimulating discussions and suggestions. They have been patient even in times when I have been incessantly demanding and have never been hesitant in going out of their way to help me when I got stuck. Thank you guys!

A special thanks to Rahul Jangid who has been like a mentor and teacher for my experimental work. From him, I have learnt a great deal about science and life and working in his company has been an outstanding experience.

Sincere thanks to my parents for their support and patience. Their unconditional love has always been a refuge to fall back on during this time.

Finally, I would like to thank IISER Pune for providing the institutional support and interdisciplinary mindset that has brought me to this important milestone.

Abstract

Design of new, functional proteins is a challenging endeavour due to our limited ability of predicting protein function based solely on its amino acid sequence. We set out to investigate protein design for generation of new proteins encoding functions not represented in nature. Ideas described in this study envisage Zinc Fingers (ZiFs) and TAL domains as ‘programmable recognition units’ to build new proteins with novel functions. This objective is illustrated in the form of two systems: A ‘split-EGFP-ZiF reporter couple’ to tag specific DNA sequences with fluorescence and cell surface expression of ZiFs and TALs for programmable cell adhesion. We cloned constructs with ZiF and TAL fusion proteins to address these objectives and investigated the behaviour of these proteins in vivo in terms of solubility and stability. The Zinc Finger proteins were found to be either insoluble or prone to proteosomal degradation. We showed that such ‘misbehaviour’ can be corrected by fusion with an additional domain that assists in folding. In contrast, TAL domains seem to be a superior alternative to ZiFs in matters of expression and solubility. We have therefore taken steps in establishing few important design principles for these proteins. In light of these results, we conclude by proposing future experiments to establish the utility of these ideas.

List of Figures

List of Tables

Introduction

Zinc Fingers and TAL domains as 'programmable recognition units'

Zinc Fingers are small DNA binding protein motifs with widespread presence in eukaryotic transcription factors (Pabo et. al., 2001). The Cys_2His_2 Zinc Finger motif is about 30 amino acids long and binds to a triplet nucleotide sequence. A single zinc ion that coordinates the Cysteine and Histidine side chains gives the motif its name. The actual DNA recognition is mediated by an alpha helix (called the recognition) helix in the motif. Residues in this helix make specific contacts with the target DNA bases and hence dictate binding specificity (Figure 1).

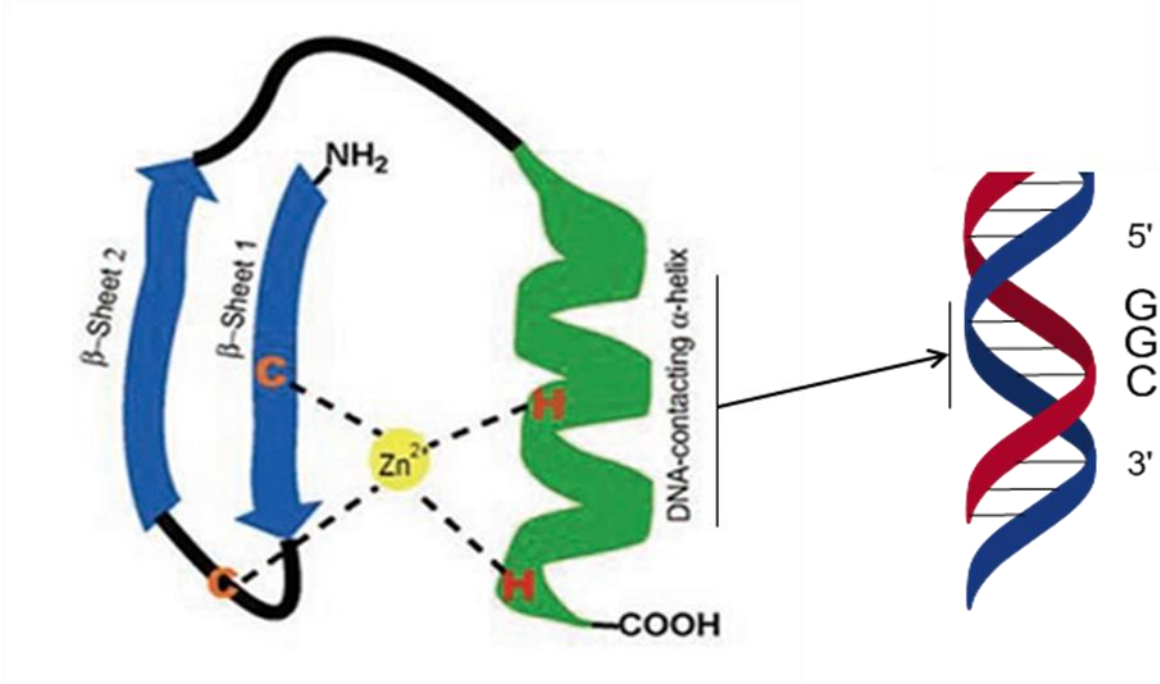


Figure 1: Schematic for binding of a ZiF motif to its target. The ZiF motif has an αββ fold. The β sheets are shown in blue whereas the recognition helix is in green. A central Zinc ion coordinates the Cysteine and Histidine amino acids. Shown alongside the ZiF motif is a hypothetical dsDNA target. The recognition helix makes specific contacts with the target site as depicted by the arrow (adapted from Negi et. al., 2008).

Amino acid mutations in the recognition helix can change the binding specificity of the motif. Through mutagenesis screens, various recognition helix sequences have been identified that will bind specifically to a distinct triplet nucleotide target (Dreier et. al., 2005). In this way, a methodology for novel ZiF motif design to recognize new targets is established. The great utility of ZiF motifs arises from their ability to be

joined in tandem to recognize a larger target sequence (Pabo et. el., 2001). Hence, multiple Fingers can be joined in tandem to recognize a larger DNA sequence. For example, a 4 Finger domain will have a recognition sequence of 12 (4 X 3) bp while a 3 Finger domain will recognize a 9 (3 X 3) bp site (Figure 2).

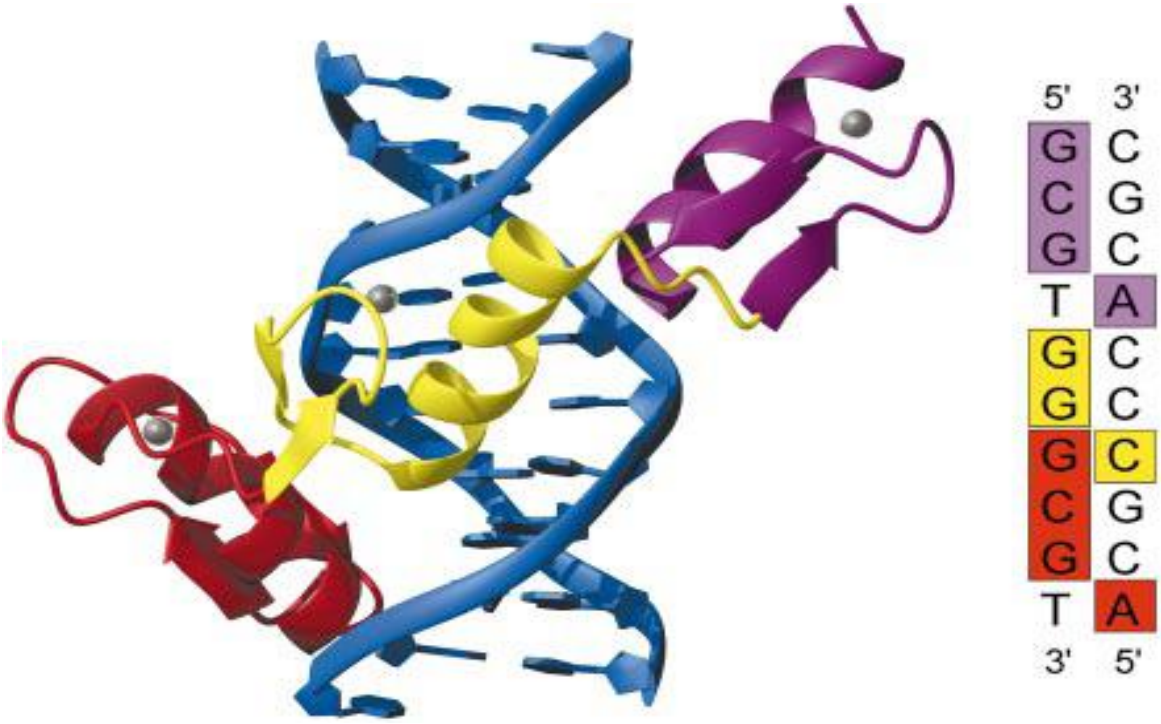


Figure 2: Individual ZiF motifs can be joined in tandem to target a longer DNA sequence. Three ZiF motifs (violet, yellow and red) are joined in tandem to recognize a 9 bp site. The DNA helix is shown in blue. The individual binding sites of the motifs are shown in the sequence alongside. The tandem fusion has a combined binding site of the individual motifs (from Pabo et. al., 2001).

The power of such a modular design enables us in principle to make Zinc Finger domains that will recognize any dsDNA target of our choice. In this study, we concentrate on 3-Finger domains that will bind to a 9 bp DNA target.

Transcription–Activator Like (TAL) effectors are proteins that recognize a specific target DNA sequence. These proteins are composed of tandem repeats (termed as TAL domains) that each recognize a single base pair and in combination, impart the effector protein its target specificity (Bogdanove and Voytas, 2001) (Figure 3). Thus,

conceptually a TAL effector is similar to a ZiF protein due to its modular nature. A single TAL domain is typically 34 amino acids long with specificity for a single DNA base pair. The 12th and 13th residue of the domain are hypervariable (called RVD- repeat variable di residue) and determine the target base (Deng, 2012). Changing these residues can change the binding specificity of the domain. In contrast to a ZiF, the TAL domain specifies a single base pair target defined by the amino acid at its 2 hpervariable positions. In this respect, the 'code' for TAL-DNA binding is extremely straightforward. Unlike the ZiF, no ion or cofactor is required in the TAL domain. However, the single base pair recognition of a TAL domain means that a TAL protein for a given DNA target is much larger (at least three times as much) than the corresponding ZiF protein.

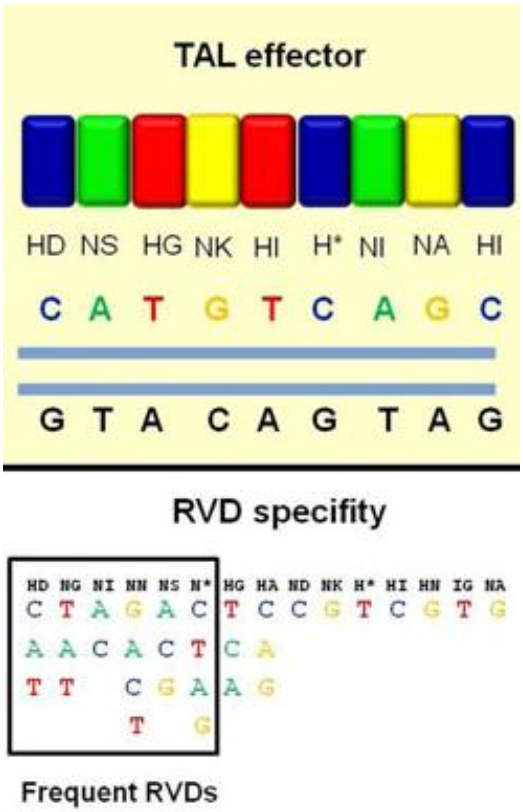


Figure 3: Schematic for TAL-DNA recognition. Individual domains shown in different colours are joined in tandem to give the TAL effector protein. The different colours specify differing target base pair specificity of the individual domains (blue: C, green:A, red:T, yellow: G) The code specifying RVD amino acids and target base is illustrated. Here, 9 individual domains are joined in a modular fashion to generate a TAL protein with a binding site of 9 bp (Bodnar et al., 2012).

In this study, we used custom TAL domains (Life Technologies) that have recognition target sites of 19-21 base pairs.

ZiFs and TALs provide a powerful tool for designing proteins with sequence-specific DNA-binding capabilities. Such logic is unprecedented in terms of protein design. These protein frameworks are unique because rules to modify sequence for predictable functional changes have been laid out. In our study, we have hence chosen these proteins as important constituents to provide a control over programmable DNA-binding.

The split-EGFP system for conditional fluorescence

The fluorescent protein GFP can be fragmented into two polypeptides that are individually not fluorescent (Kerppola, 2009). However, under appropriate conditions, the two halves can associate and reassemble the complete secondary structure of full length GFP. This is termed as fluorescence complementation. The reassembled protein possesses fluorescence characteristics comparable to native, full-length GFP. The fluorophore in full-length GFP results from cyclization of a contiguous Ser, Tyr and Glycine residue chain in the interior of the protein. The fluorophore if artificially synthesized is non fluorescent due to quenching effects by the solvent. The tertiary structure of GFP is extremely crucial for its fluorescence characteristics (Tsien 1998). The β-barrell surrounding the fluorophore shields it from solvent quenching and makes specific contacts with the fluorophore to maximize fluorescence emission. Consequently, GFP half fragments are non fluorescent either due to absence of a fluorophore or due to lack of necessary secondary structure. Upon complementation, the two halves presumably reassemble the secondary structure through supramolecular interactions to give fluorescent GFP. The utility of this system results from an added constraint. Secondary interactions that bring these halves in proximity to facilitate their association are crucial for efficient reassembly (Kerppola, 2009). Thus, rates of complementation for the fragments when fused to interacting partners is many fold higher than spontaneous reassembly. In this sense, the reassembly is conditional (Figure 4B). Fluorescence readout is an indication that the partner proteins are interacting. Multiple ways to fragment the EGFP protein have been reported. In this study, we use the method by Demidov et. al. (2005)

which describes dividing EGFP between amino acids 158 and 159 to get the large and small halves.

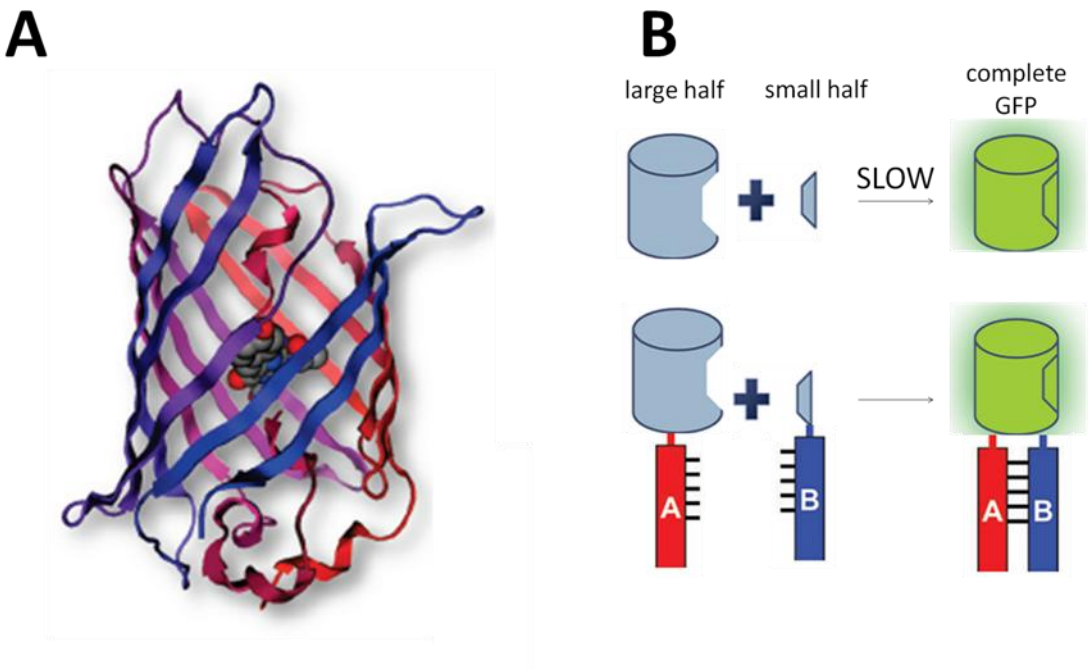


Figure 4: Schematic for GFP complementation. (A) Cartoon showing 3D structure of GFP. The fluorophore is shown as a ball model in the interior of the protein. It is surrounded by the barrelled secondary structure of GFP. (B) Non fluorescent fragments of GFP that can recombine to give back the fluorescent protein are labelled as large and small halves. The interacting partners A and B drive the reassembly by bringing the two halves together. In their absence, complementation rates are slow (adapted from Huang et. al., 2009 and Kerppola, 2009).

Combining programmable DNA-binding with conditional fluorescence.

The DNA-binding ZiFs and TALs along with split-EGFP form the cornerstone of our strategy for designing novel proteins. Using these systems, we will present two strategies that underline the common theme of protein design.

The split-EGFP-ZiF fusion system for in vivo tagging of DNA sequences.

The Objective is to design a GFP-based protein assembly that will bind to a specific DNA sequence and tag it with a fluorescent label in vivo for live monitoring of DNA dynamics.

The system must be able to:

1) Specifically bind to any given DNA sequence in the genome.

2) Induce fluorescence as a result of the binding event. Unbound protein should not fluoresce.

Thus, the assembly will mark the DNA sequence with GFP (This can be a repeat element or part of gene loci or any other sequence of interest). This system combines specific targeting of DNA sequences with Zinc fingers and conditional fluorescence using the split-GFP technology. The strategy involves the construction of split EGFP fragments fused with a three Zinc finger domain. The Zinc finger domain directs binding to the target DNA sequence. Fluorescence is induced when the two separated halves of GFP recombine on binding. This is schematically depicted in Figure 5.

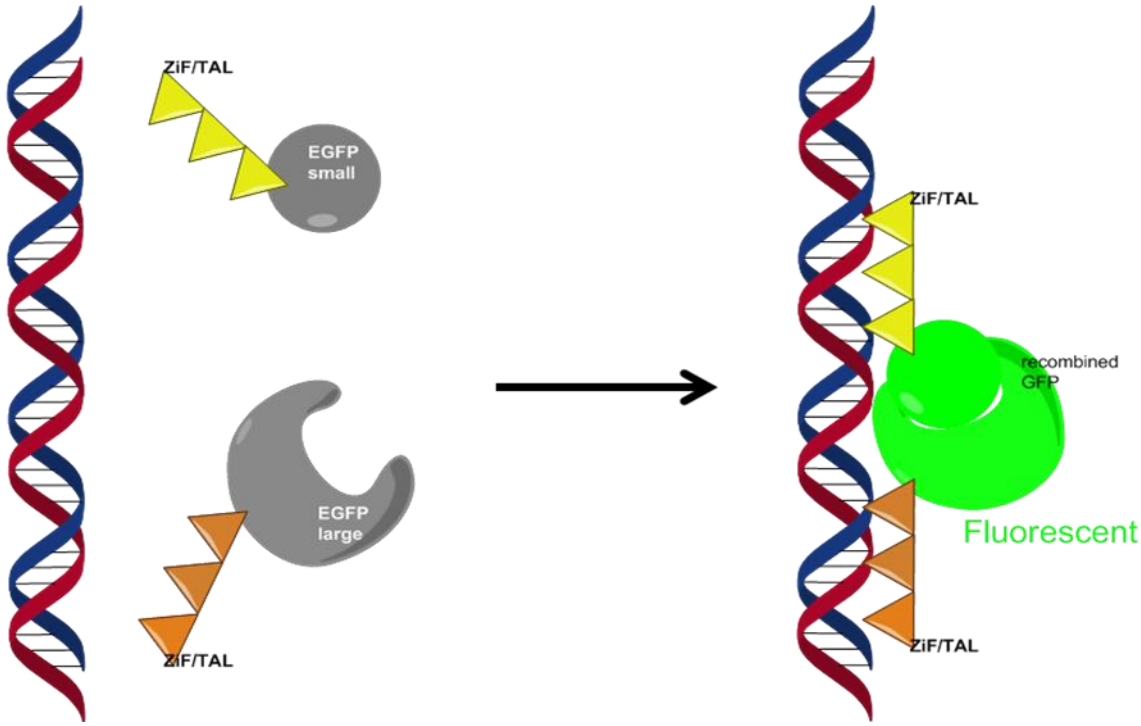


Figure 5: Schematic illustrating conditional fluorescence upon programmable DNA binding. Non-florescent EGFP halves in grey are fused to ZiF or TAL domains (shown as triangles) with predefined binding sites. In presence of the binding site (shown as a stretch of dsDNA) the EGFP halves would be brought into close proximity and reassemble the EGFP fluorescent protein, thus tagging the DNA with a fluorescent mark.

Surface Display Of ZiFs and TALs

We envisioned the possibility of Expressing ZiFs and TALs on mammalian cell surface as controllable DNA binding proteins. The ZiF/TAL and its targets would form an extracellular 'protein-ligand' set where both the protein and its target ligand are completely programmable. The motivation for such an idea comes from the remarkable recent advances in DNA nanotechnology (Seeman, 2010). DNA can be patterned with great precision in complex arrays and 3-dimensional spaces. We wondered if this control over DNA assembly can be extrapolated to cellular assembly by enabling cells to specifically bind to DNA. We propose to use dsDNA oligonucleotides as 'sticky ligands' that will bind to the cell surface and provide a handle for attachment. By providing the requisite DNA-binding protein on the cell surface, we would be able to control which DNA target binds to the cell. A variety of applications can be speculated to arise from this approach (Figure 6). Here, cells shown in yellow have different DNA binding ZiF/TAL protein displayed on their surface that binds to a distinct dsDNA target (shown in green, red and blue). Note that these cells can be identical apart from their distinct surface display protein. Figure 6B illustrates attachment of these cells to a surface coated with the complementary ss oligo. Attachment position in turn is dictated by the oligo pattern on the surface. Specific molecules can be targeted to the cell surface by linking them to the target DNA (Figure 6C). Figure 6D illustrates in vivo cell patterning by using a target oligo which has binding sites for both the red and green protein. In a co culture of cells displaying these proteins on their surface, this 'double target' would link each green cell (cell with the green protein) to a neighbouring red cell (cell with red protein) and vice versa. This would result in a pattern where each green cell is surrounded by and bound to a red cell and vice versa. In this case, the DNA is behaving like an artificial ECM that will dictate relative cell positioning. In summary, surface display of DNA-binding domains can give access to spatial control over cell positioning and provide distinct surface handles with known targets for controlled manipulations.

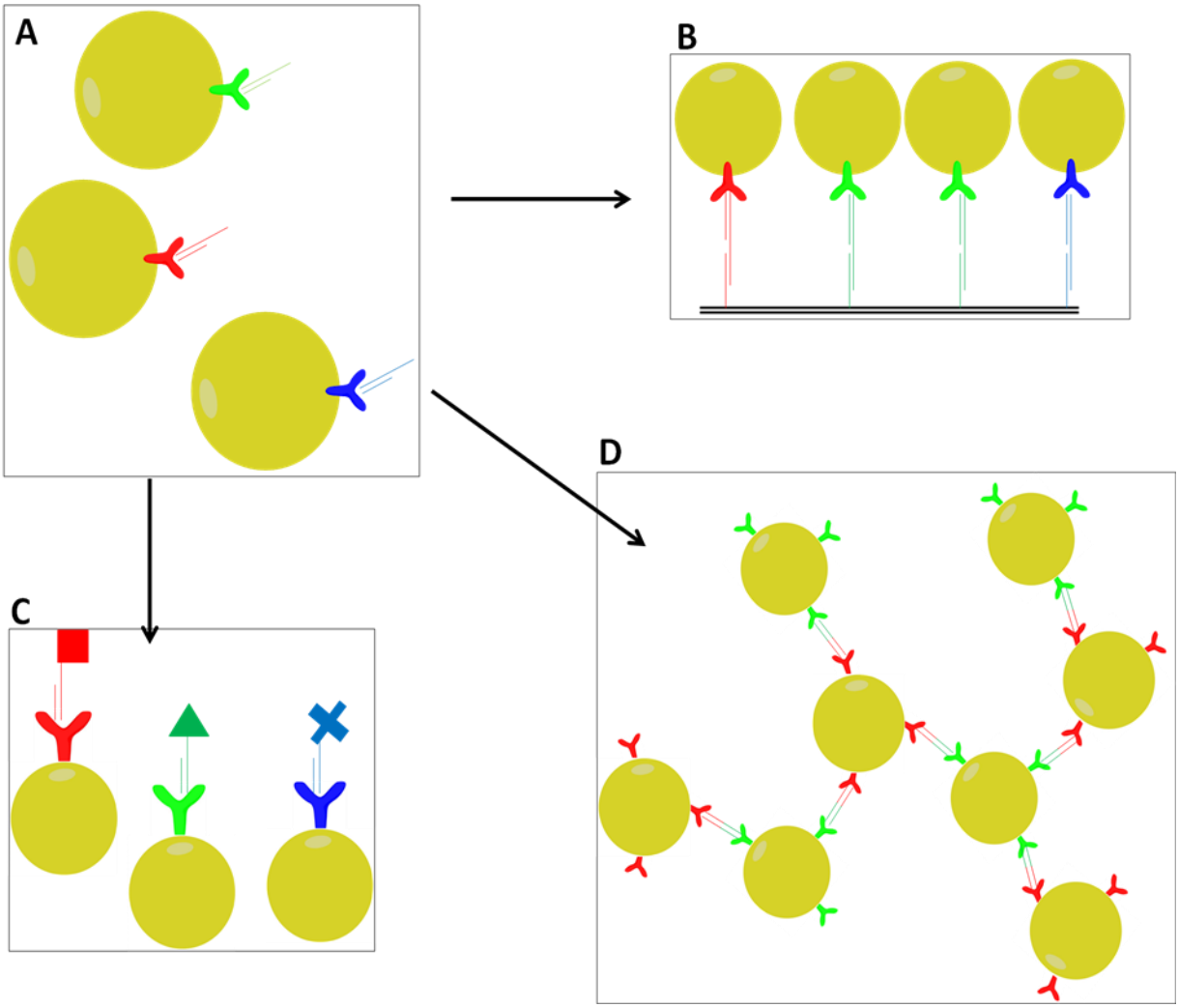


Figure 6: Surface display of ZiF/TALs and their envisioned applications. Cells are depicted in yellow. ZiF/TAL proteins shown as coloured Y shapes on their surface. DNA targets are depicted by lines. The protein and cognate target are shown in same colour (red, green or blue). B) Complementary oligos to the targets can be immobilised with spatial control on a surface (black) This will dictate binding pattern of the respective cells. C) Distinct cargo shown as different shapes (square, triangle, plus) conjugated to the target oligo can be specifically targeted to the cell with the cognate protein. D) A co-culture of 'red' and 'green' cells in presence of a 'double target' will form a pattern where each red cell will be surrounded by a green cell and vice versa.

The approach to address these twin goals of conditional fluorescence and surface display is outlined here (Figure 7). The proteins mentioned have to be cloned appropriately. The proteins can then be purified for an in vitro assay. Likewise, their expression characteristics in vivo must be investigated. Surface expression of DNA-binding proteins must be confirmed. Validation for display proteins is relatively straightforward. Cultured cells with surface expression can be exposed to a fluorescently tagged target oligo. Localization of fluorescence to the cell surface is

the expected readout in case of specific binding. For Fluorescence reconstitution, one must demonstrate fluorescence from the two EGFP half fusion proteins in presence of the target sequence. This can be done in vitro as well as in vivo. In vitro, a mixture of the purified EGFP half fusion proteins alongwith the target oligo must fluoresce. For in vivo validation, target oligo is transfected into cells expressing the two EGFP half fusion proteins and fluorescence readout is assayed.

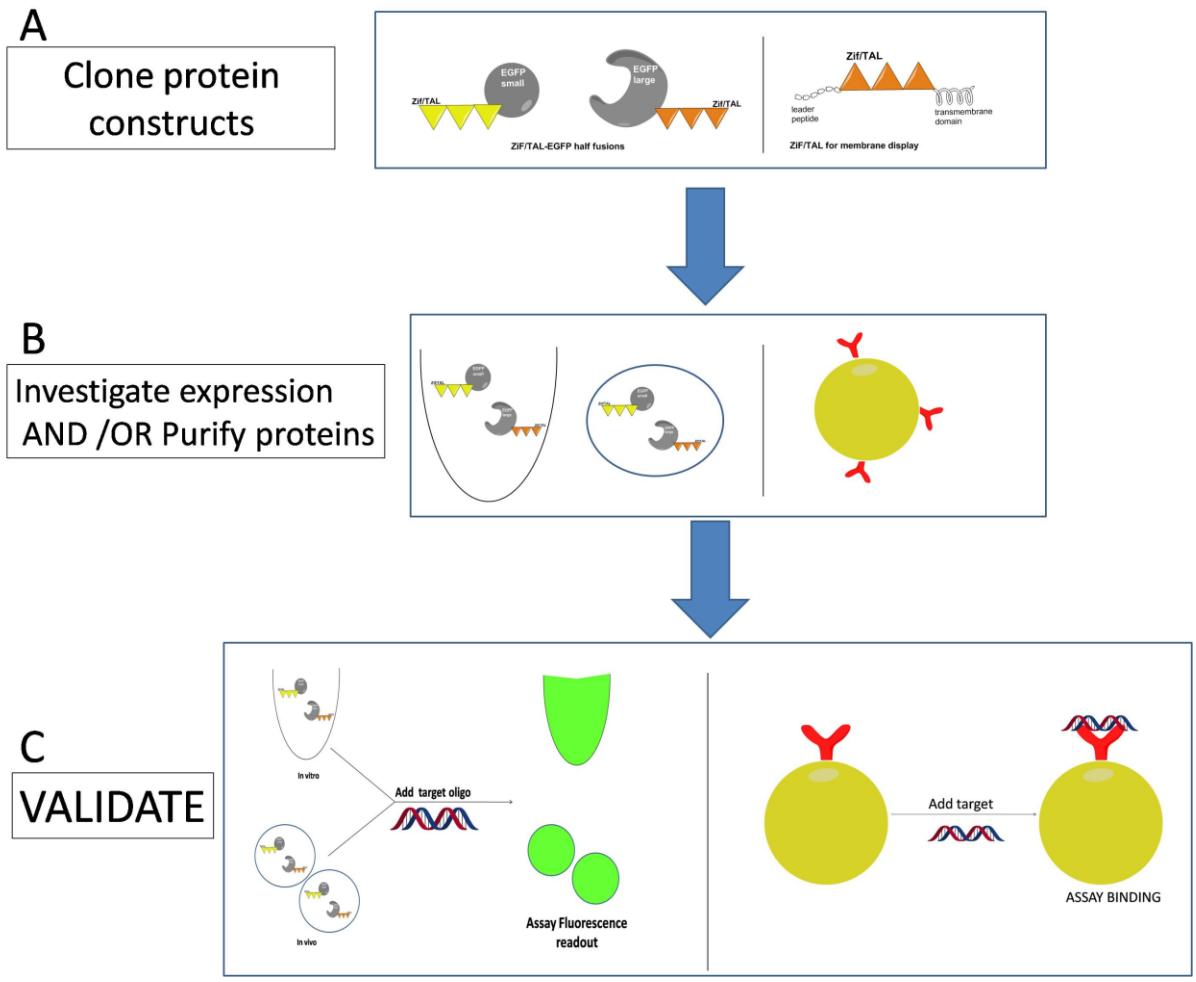


Figure 7: Steps to execute described ideas. In the boxes shown, the left half corresponds to conditional fluorescence and right half to surface display. A) The requisite proteins are cloned. B) The EGFP half proteins proteins are purified and/or their expression characteristics in mammalian cells investigated. For display proteins, surface expression is checked. C) Proof of concept can be provided by demonstrating fluorescence reconstitution in vitro/in vivo and binding of DNA to the surface protein.

Results

Cloning

ZiFs/TALs were fused to EGFP halves and cloned in bacterial/mammalian expression vectors. Likewise, ZiFs/TALs were also cloned in the pDisplay vector (Invitrogen) for mammalian surface display. 45 Constructs have been described in this study. Vector maps have been illustrated in the materials and methods section. Briefly, these can be divided into 8 groups. The type of protein produced and experiment performed is summarised in Table 1.

Table 1 : List of DNA constructs

Protein	Remark	Constructs
ZiF-EGFP half GST tagged	For bacterial expression and protein purification	A12 in pC62-73 B21 in pC62-73 B23 in pC62-73
ZiF-EGFP half HIS tagged	In pTriex vector for bacterial and mammalian expression	A12-HIS in pTriex HIS-B21 in pTriex HIS-B23 in pTriex
ZiF-EGFP half proteins separated by p2A peptide	Mamallian expression. Equal production of the two proteins from a single plasmid.	A12-p2A-B21
ZiF-EGFP half proteins tagged with mCherry and separated by p2A peptide	Production of equal amount of tagged proteins from single plasmid.	A12-mCherry-p2A-mCherry-B21
EGFP half proteins fused to TAL domains	Mamallian expression. EGFP halves fused upstream and downstream of TAL domains	EGFPlarge-TAL1 in pCDNA EGFPlarge-TAL2 in pCDNA EGFPlarge-TAL3 in pCDNA EGFPlarge-TAL4 in pCDNA EGFPsmall-TAL1 in pCDNA EGFPsmall-TAL2 in pCDNA EGFPsmall-TAL3 in pCDNA EGFPsmall-TAL4 in pCDNA TAL1-EGFPlarge in pCDNA

		TAL2-EGFPlarge in pCDNA TAL3-EGFPlarge in pCDNA TAL4-EGFPlarge in pCDNA TAL1-EGFPsmall in pCDNA TAL2-EGFPsmall in pCDNA TAL3-EGFPsmall in pCDNA TAL4-EGFPsmall in pCDNA
ZiF for surface expression, with or without full length EGFP fusion	Cloned in pDisplay vector for targetting to the plasma membrane	ZiF1in pDisplay ZiF2in pDisplay ZiF3in pDisplay EGFP-ZiF1in pDisplay EGFP-ZiF2in pDisplay EGFP-ZiF3in pDisplay ZiF2-EGFP in pDisplay
TAL domains for surface expression	Cloned in pDisplay vector for targetting to the plasma membrane	TAL1 in pDisplay TAL2 in pDisplay TAL3 in pDisplay TAL4 in pDisplay
Others	Constructs used as controls or obtained as intermediates in the cloning steps.	A1 in pC62-73 B2 in pC62-73 p2A in pEGFPN1 A12-p2A in pEGFPN1 A12-p2A-mCherry-B21 TAL1 in pCDNA TAL2 in pCDNA TAL3 in pCDNA TAL4 in pCDNA EGFP in pDisplay

The EGFP halves are designated as A1 (large half, amino acids 1 to 158 of full length EGFP) or B2 (small half, amino acids 159-260 of full length EGFP). 3 Zinc Finger domains namely ZiF1, ZiF2 and ZiF3 were synthesized by a primer overlap extension PCR protocol (see methods). These were cloned as fusion proteins with the EGFP halves. Thus A12 denotes the EGFP large-ZiF2 fusion whereas B21 denotes the ZiF1-EGFP small fusion protein. These abbreviations are used henceforth.

ZiF-EGFP half proteins can be purified with the pC62/Caspase 6 system

The ZiF-EGFP half proteins were cloned in a bacterial expression vector (pC62-73) for GST tagged recombinant protein expression and purification (Purbey et. al. 2006). The presence of a Caspase 6 cleavage site after the GST tag enables downstream removal of the tag to obtain native ZiF-EGFP half proteins. The pC62/Caspase6 system for purification of GST-free recombinant proteins has been described in detail by Purbey et al. and the same protocol was followed in this case. Although a significant fraction of the proteins were found in inclusion bodies (data not shown), the concentrations in the soluble fraction were sufficient to enable purification (Figure 8). The bead bound proteins can be eluted as GST tagged fusions (Figure 8, A and B) or can be cleaved with Caspase 6 (Figure 8, C and D) to give GST free ZiF-EGFP half proteins. The cleavage efficiency for the B21 and B23 proteins was seen to be significantly lower as compared to A12. We hypothesize that an increased concentration and/or longer incubation times with Caspase 6 are needed for complete removal of the GST tag from B21 and B23. Two of the proteins A12 and B23 were also investigated for expression and purification as N- or C-terminal HIS tagged fusions from pTriex constructs (See methods). In contrast to the GST-tagged proteins, A12-HIS failed to express under a variety of conditions. Representative data is shown in Figure 8E. Although HIS-B23 showed good expression, it showed poor solubility (Figure 8E). Supernatant fractions analyzed for presence of soluble protein by binding with Ni-NTA beads showed negligible presence of soluble protein (data not shown). Together, these results indicated that GST-tagged ZiF-EGFP half proteins are amenable to purification using the pC62/Caspase 6 purification system. In contrast the same proteins with a 6XHis tag revealed poor expression and/or solubility.

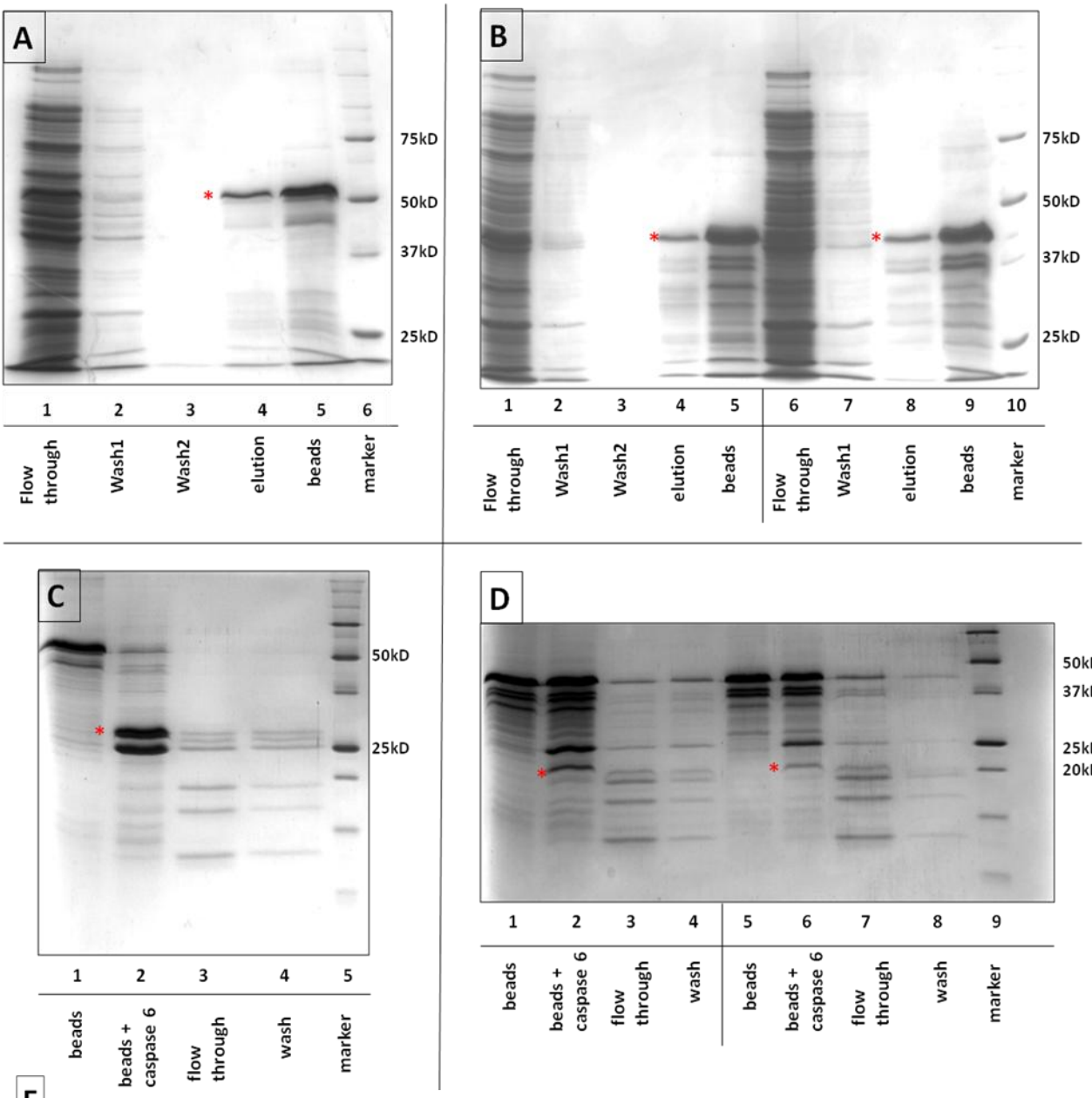

A
75kD
50kD
37kD
25kD
1 2 3 4 5 6
Flow through
Wash1
Wash2
elution
beads
marker
B
75kD
50kD
37kD
25kD
1 2 3 4 5 6 7 8 9 10
Flow through
Wash1
Wash2
elution
beads
Flow through
Wash1
elution
beads
marker
C
50kD
25kD
1 2 3 4 5
beads
beads + caspase 6
flow through
wash
marker
D
50kD
37kD
25kD
20kD
1 2 3 4 5 6 7 8 9
beads
beads + caspase 6
flow through
wash
beads
beads + caspase 6
flow through
wash
marker
E
--40kD
--30kD
--25kD
--20kD
1 2 3 4 5 6 7 8 9
U I U I U I U I
S S P P S S P P

Figure 8: ZiF-EGFP half protein expression with GST tag and 6XHis tag. (A, B) 12.5% SDS Page Gels stained with Commassie brilliant blue showing binding and elution of GST tagged ZiF-EGFP half fusion proteins from glutathione-sepharose beads. (A) GST-A12 (B) GST-B21 (lanes 1to 5) and GST-B23 (lanes 6 to 10). Bands corresponding to the protein of interest are marked with an asterix. Bead-bound protein was eluted with 10 mM reduced glutathione. (C, D) 15% SDS Page Gels stained with Coomassie brilliant blue showing Caspase 6 cleavage of bead-bound GST proteins. Beads were incubated with Caspase 6 in C6 buffer and analyzed on the gel. (C) GST-A12. (D) GST-B21 (lanes 1 to 4), GST-B23 (lanes 5 to 8). Lanes labelled 'beads+caspase 6' show the cleavage products. Cleaved protein is indicated by a star. Free GST band is seen at 25 kDa. (E) 15%SDS-PAGE gel stained with Coomassie brilliant blue to check protein induction for A12-HIS (lanes 1 to 4) and HIS-B23 (lanes 6 to 9). U=uninduced, I=induced, S=Supernatant, P=pellet. HIS-B23 revealed good induction but poor solubility (lane 9, asterix *), whereas a differential induction band was not seen in either the supernatant or pellet for A12-HIS.

The ZiF-EGFP half proteins are insoluble when expressed in vivo and are differentially affected by the proteosome inhibitor MG132

The p2A and pTriex constructs were used for investigating expression of the ZiF-EGFP half proteins in the Human cell line HEK293T. The p2A construct enables expression of proteins A12 and B21 in equal amounts from a single plasmid (Kim et. al., 2011) (see methods). The proteins showed very low solubility when analyzed with a Western blot. Extremely weak bands are seen in the soluble fraction of the lysate and all protein is concentrated in the insoluble, pellet fraction (Figure 9). The absence of protein in the supernatant indicated the possibility of proteosomal degradation of the concerned proteins. To test this, protein expression was analyzed in the presence of the nonspecific proteosomal inhibitor MG132 (Sigma) (Figure 9A lanes 3, 4). A slight difference is seen in the intensity of bands in Lane 1 and Lane 3. However no change is seen in the supernatant fractions (lanes 2 and 4). We concluded that the proteins A12 and B21 are insoluble and are not targeted for degradation. As a comparison, A12-HIS and HIS-B23 (pTriex constructs) were also checked for expression in presence or absence of MG132. A12-HIS shows poor solubility and is unaffected by MG132 treatment (Figure 9B, Lanes 5 to 7). In contrast, B23 is drastically increased in the insoluble fraction on MG132 treatment. Collectively, these results indicated that the ZiF-EGFP half proteins are insoluble in vivo and are differentially targeted for proteosomal degradation.

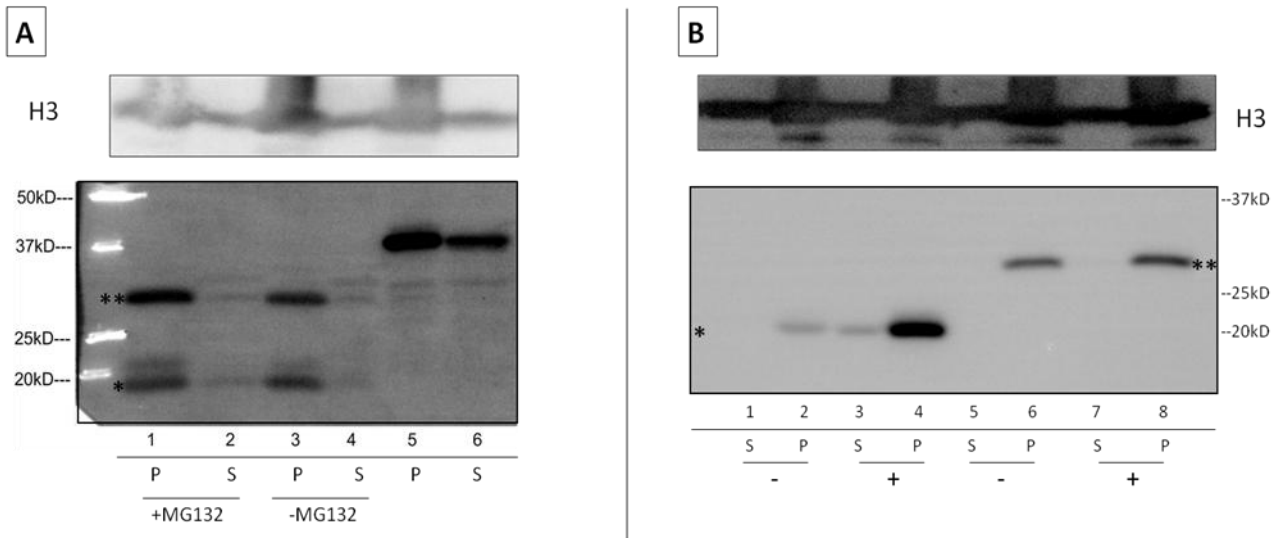


Figure 9: Poor solubility of the ZiF-EGFP half proteins and effect of MG132. (A) Western blot for expression of the ZiF-EGFP half proteins from the p2A vector. Blot was probed with anti-GFP polyclonal antibody. Lanes 1 to 4 depict transfections with the construct A12-p2A-B21 in the presence or absence of the proteosome inhibitor MG132. Lanes 5 and 6 show the expression profile of a control vector (EGFP in pDisplay). After preparation of lysates, the supernatant and pellet both were analyzed. P=pellet, S=Supernatant. ** = A12 protein band. * =B21 protein band. (B) Western blot for expression of the ZiF-EGFP half proteins from the pTriex vector. Blot probed with Anti-6XHis antibody. B23 in pTriex (lanes 1 to 4) and A12 in pTriex (lanes 5 to 8) were analyzed for expression and solubility in the presence (+) or absence (-) of MG132. S=Supernatant, P=pellet. *=A12, **=B23. H3 was used as loading control.

ZiF-EGFPsmall half can be solubilised by fusing it to the fluorescent mCherry protein.

To test whether fusing a well-folding partner protein such as mCherry affected solubility, we used the construct A12-mcherry-p2A-mcherry-B21 to check for protein expression. This construct has both the ZiF-EGFP half proteins tagged with mCherry. Likewise, we also used the construct A12-p2a-mcherry-B21 in which only B21 is tagged. Figure 10 shows the result for both constructs when probed with a polyclonal GFP antibody and, separately with mCherry antibody. The GFP antibody detects both A12 and B21 while the mCherry antibody will detect only mCherry tagged protein. In case of A12-p2a-mcherry-B21, only mcherry-B21 is detected in the soluble fraction whereas untagged A12 is restricted completely to the pellet. However, when both A12 and B21 are tagged with mCherry, A12-mCherry is only slightly enriched in the soluble fraction.

This demonstrated that the insolubility of the two proteins can be partially addressed by mCherry fusion. Solubility of B21 is significantly improved whereas that of A12 is only marginally affected.

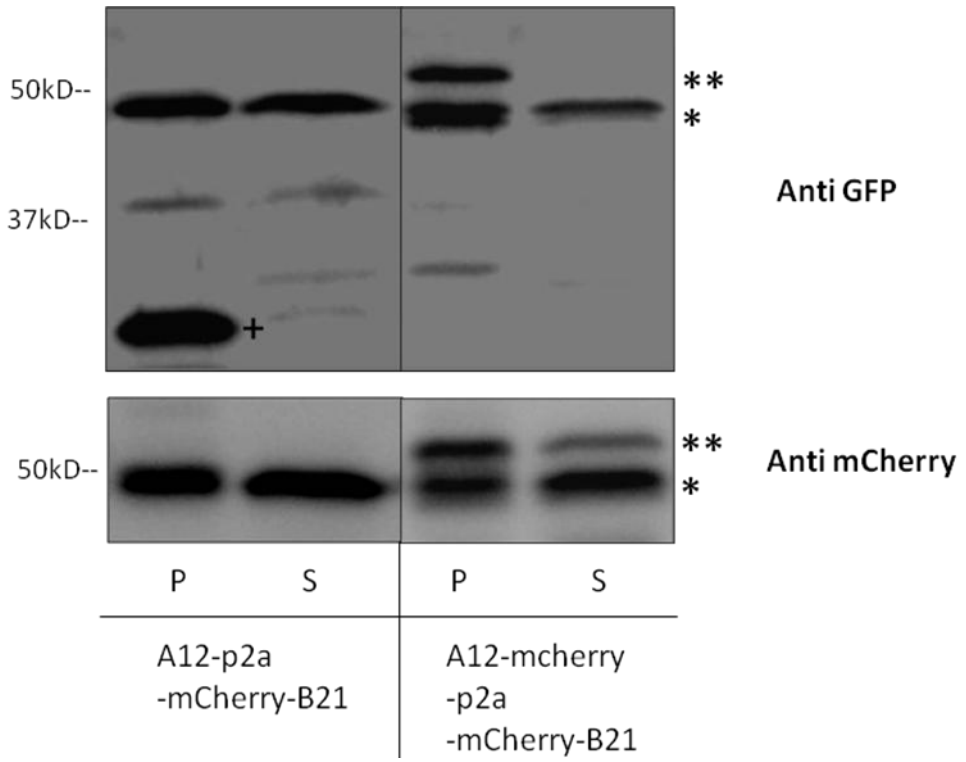


Figure 10: mCherry fusion assists solubility for the ZiF-EGFPsmall half protein. P=pellet, S=Supernatant. *=mCherry-B21, **=A12-mCherry, +=A12. A1 is present only in the pellet fraction whereas mcherry-B21 is also enriched significantly in the supernatant. However, on tagging A12 with mcherry, the increase in the soluble fraction is small as compared to mcherry-B21. Note: All Panels are from a single blot. The blot was first probed with anti-GFP, then stripped and reprobed with anti-mCherry antibody.

The proteins ZiFs in pDisplay are targeted to the proteosome for degradation.

The MCS of the pDisplay vector is located between an upstream secretory sequence and a downstream transmembrane domain sequence (see methods). Consequently, genes cloned in the MCS are expected to be targeted to the plasma membrane. The three Zinc Finger domains Zif1, Zif2 and ZiF3 were cloned in pDisplay for surface expression (see methods for construct details). The constructs were transfected and analyzed for expression by Western blot. Initial experiments (data not shown) yielded no protein expression in either the soluble or insoluble fraction. However in presence

of MG132 (Figure 11) rescued expression of the proteins was seen in the soluble fraction.

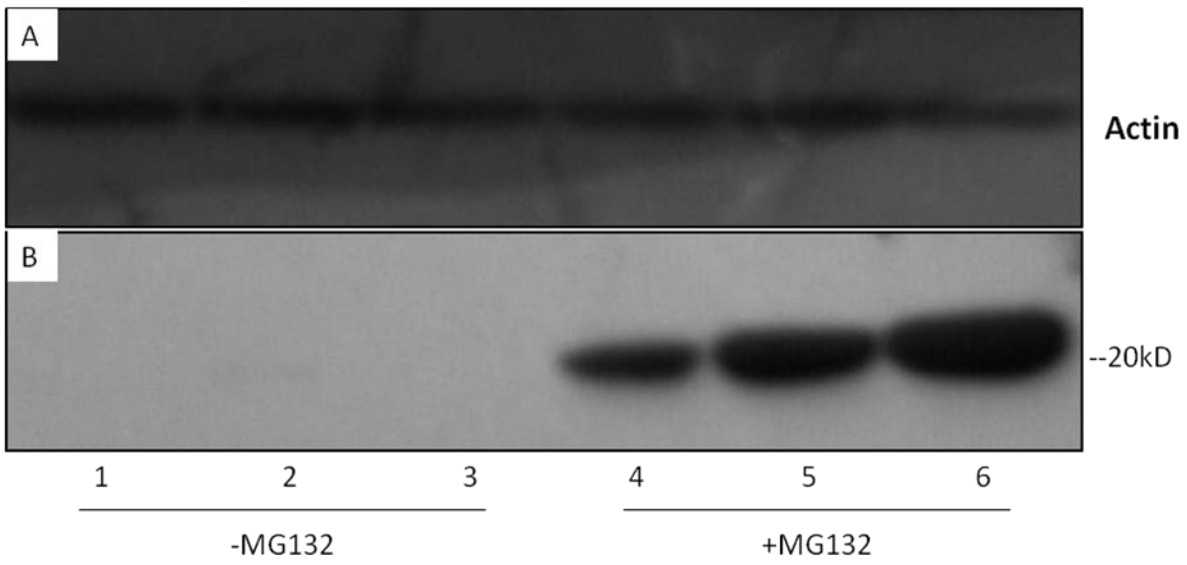


Figure 11: ZiFs in pDisplay undergo proteosomal degradation. The constructs ZiF1 in pDisplay (lane 1 and 4), ZiF2 in pDisplay (lane 2 and 5) and ZiF3 in pDisplay were analyzed for expression in presence or absence of MG132. (A) Actin, loading control. (B) Expression of the ZiF proteins is seen only on treatment with MG132. Blot probed with anti-HA tag antibody.

ZiF in pDisplay proteins can be rescued for expression by tagging them with full length EGFP

Although ZiF proteins in pDisplay are poorly expressed, EGFP in pDisplay shows good expression (see Fig 9A, lane 5). This raises the possibility that tagging the ZiF proteins in pDisplay with full length EGFP may rescue their expression. EGFP expression in cells can be monitored by a fluorescence microscope. EGFP in pDisplay transfected cells show a distinct punctate fluorescence pattern which is also seen in EGFP tagged ZiF in pDisplay constructs (data not shown). This is indicative of a rescue phenomenon. The tagged constructs were transfected and analyzed for protein expression by immunoblotting (Figure 12). Enhanced expression was seen in the tagged constructs.

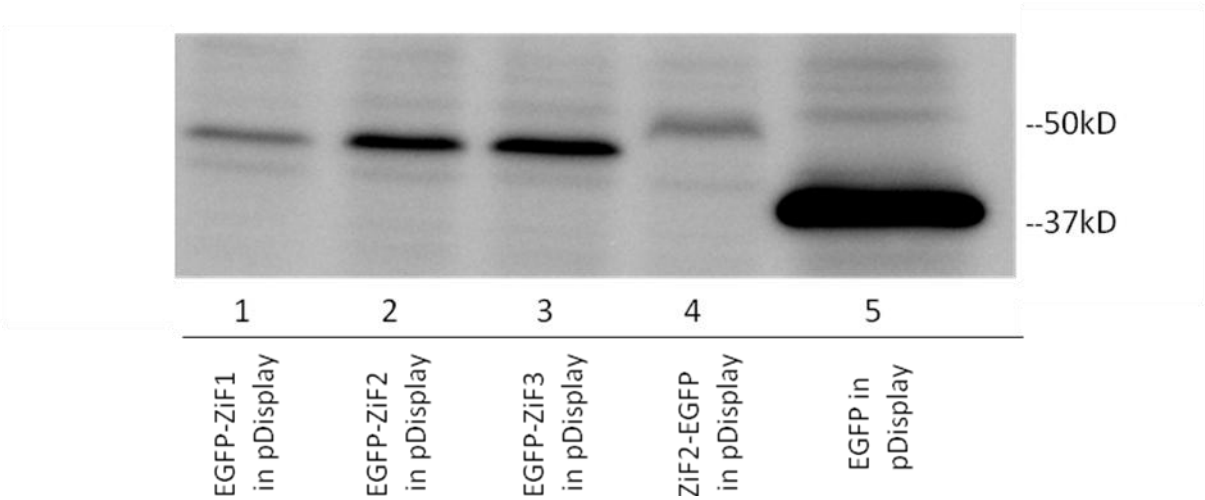


Figure 12: Expression of ZiF proteins in pDisplay tagged with full length EGFP. Blot was probed with anti-HA tag antibody**.**

TAL domain fusion proteins express well in vivo

TAL domains were investigated as alternatives to the Zinc Finger domains. EGFP halves were cloned upstream or downstream of TAL domain (see TAL-pCDNA constructs in methods). The TAL-EGFP half fusion proteins (EGFP halves fused to C terminal of the TAL domain) express well in vivo (Figure 13A). In comparison, the EGFP half-TAL constructs show reduced expression (Figure 13B). An additional band corresponding to the TAL protein is also seen. This is because of two translation initiation sites in the construct. Likewise, TAL domains in pDisplay also show expression unlike their ZiF counterparts. (Figure 13C)

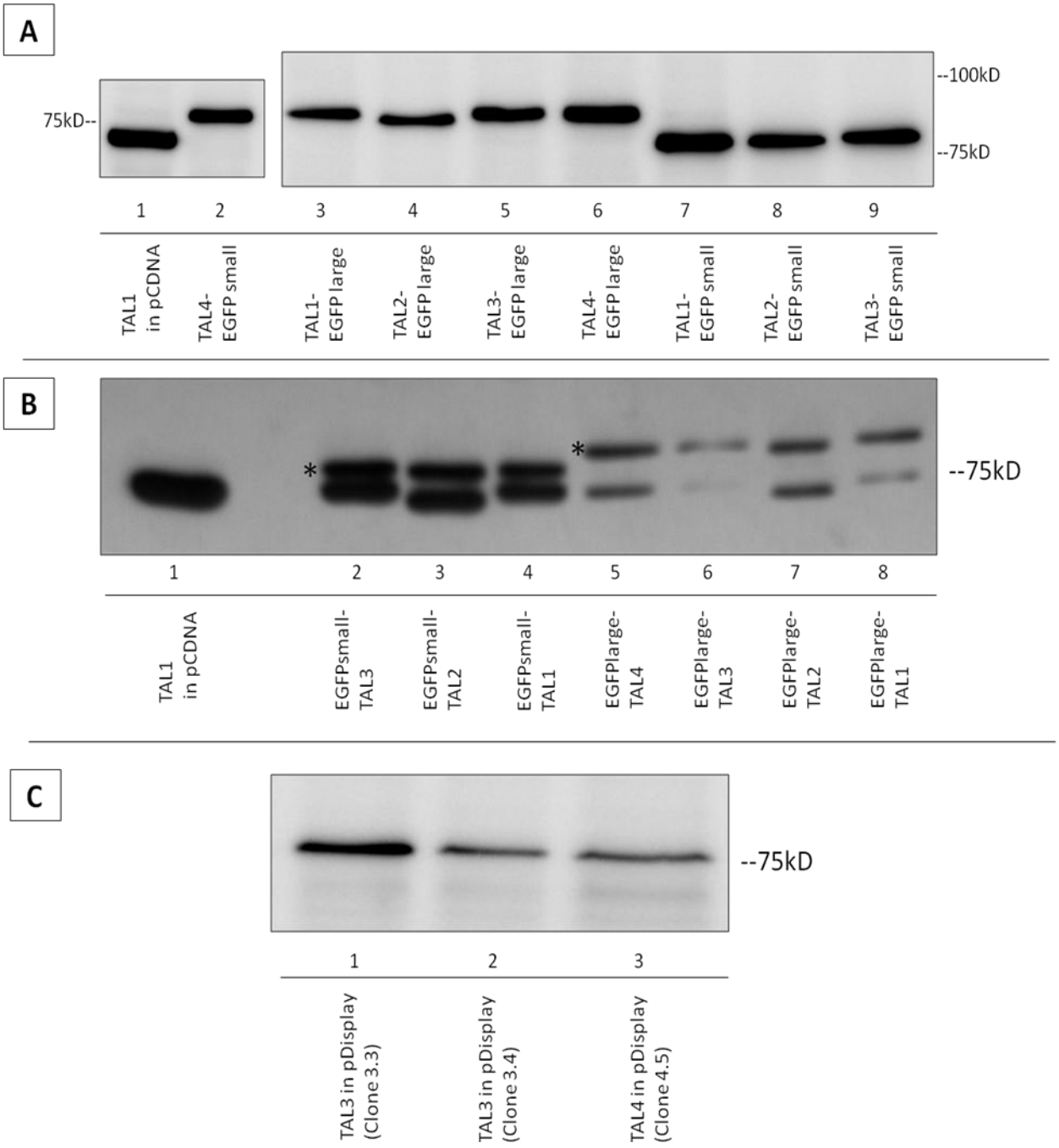


Figure 13: Expression of TAL-EGFP half fusions and TALs in pDisplay. EGFP halves cloned downstream (A) or upstream (B) of TAL domains were tested for expression. Blot probed with anti V5 antibody. The construct used is indicated below the lane. Lane 1 shows the untagged TAL protein for comparison. The GFP half fusions migrate at a higher molecular weight than the free TAL proteins (A, lane 2 to 9. B lanes 2 to 8) For the EGFP halves cloned upstream of TAL (panel B), the lower band corresponds to free TAL protein. The upper band denoting EGFP half-TAL fusion is marked with an asterix. (C) Western Blot showing expression of TAL domains cloned in pDisplay. Blot was probed with anti-HA tag antibody.

Discussion

Considerations regarding EGFP half proteins

The EGFP large half (1-158 amino acids of ful lngth EGFP) and the small half (amino acids 159-260) are non fluorescent polypeptide chains. The large half contains the EGFP fluorophore but lacks the secondary structure of full length EGFP which is necessary to minimize quenching and maximize fluorescence emission (For a detailed discussion, see Demidov et al, 2005). Under appropriate conditions, the two halves can associate and reassemble the complete secondary structure of full length EGFP. The reassembled protein possesses fluorescence characteristics comparable to native, full length EGFP. Here we have chosen the EGFP halves described by Demidov et al. which show fast complementation when fused with complementary oligonucleotides (Figure 14). Importantly, when the secondary interactions (annealed oligos) are disrupted, the recombined protein dissociates to give back the non fluorescent constituent halves. This report describing conditional, fast and reversible association of EGFP halves is an important basis for this study.

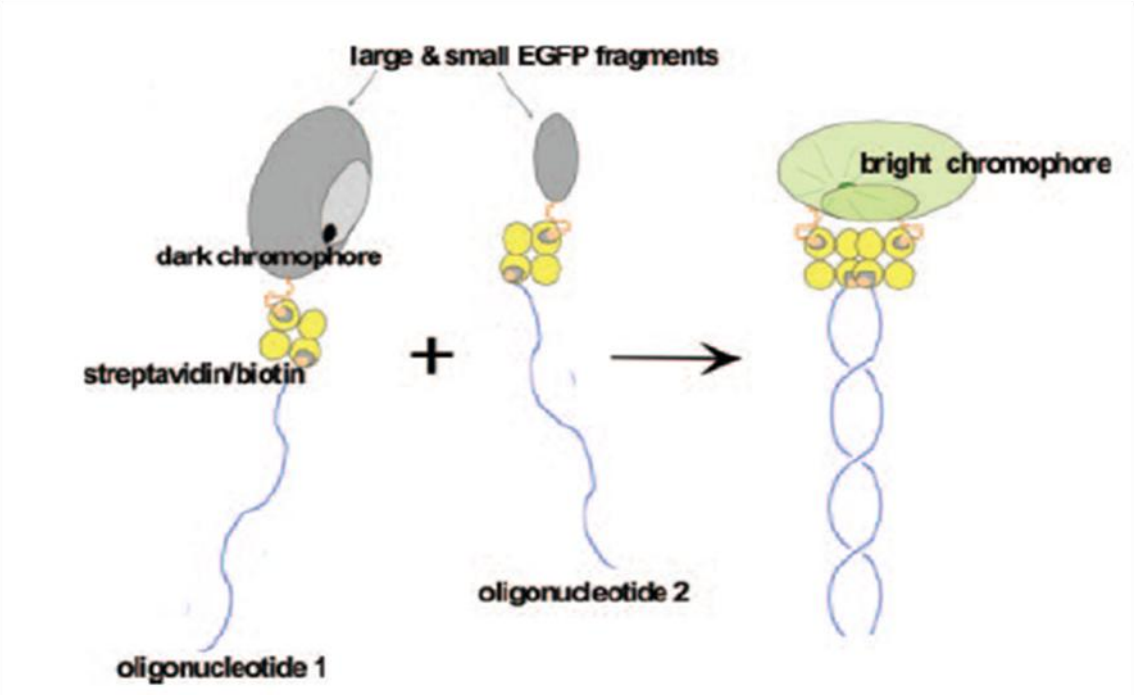


Figure 14: Fluorescence complementation strategy. The two halves of EGFP which are individually non fluorescent are brought together by covalently tagged complementary DNA oligonucleotides (oligonucleotide1 and oligonucleotide2) to facilitate reassembly of a mature, fluorescent EGFP (from Demidov et. al., 2005)

In the Demidov et al. report complementary oligonucleotides were covalently tagged to the C terminal of the large half (A1) and the N terminal of the small half (B2) respectively. From this, it is apparent that the N-terminal of A1 and the C-terminal of B2 are the ends which are important for reassembly. Consequently, our approach involved the fusion of different proteins/DNA binding domains to the C-terminal of A1 and N-terminal of B2 (Figure 15). This ensured that no interfering additions were present at the recombining ends of the two halves. However, a recent report on the two halves by the same group suggests that tagging the C-terminal of B2 also gives a viable partner (Burton et. al., 2007) . Thus it is not entirely clear whether fusion to a particular end is strictly necessary for reassembly. Consequently, while cloning the TAL-EGFP half fusions, we have cloned both the EGFP halves upstream as well as downstream to the TAL protein to explore all possible combinations.

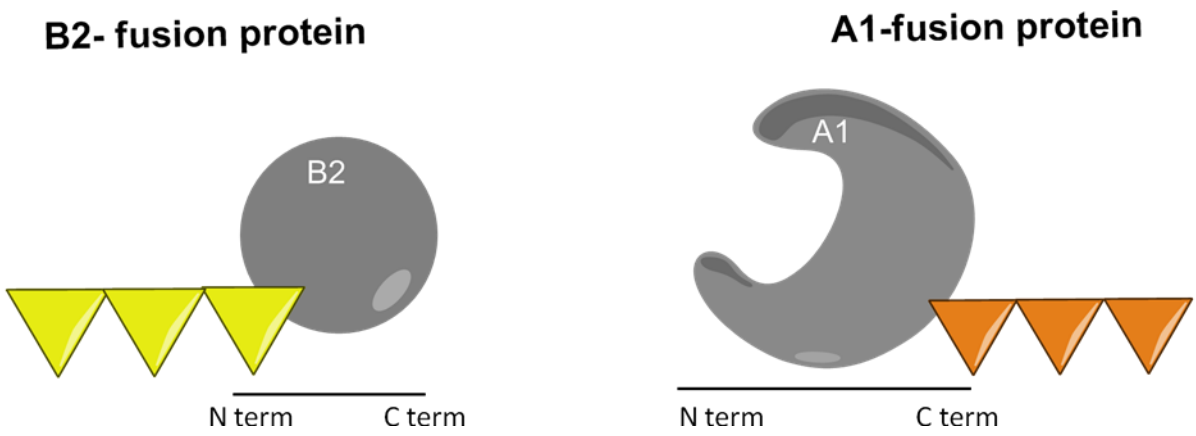


Figure 15: Design of EGFP half fusion proteins. The Large EGFP half (A1) and Small EGFP half (B2) are shown in grey. Triangles denote fusion partners. The fusions are such that the N- terminal of A1 and C-terminal of B2 are untagged and free for complementation.

Specific binding to predetermined dsDNA sequences with Zinc Fingers and TALs

Both Zinc Finger and TAL proteins share a common design principle which enables modular joining of multiple domains with known target recognition sites. The resulting protein has a combined target recognition sequence of the constituent domains. In this study, the Sp1 backbone is used for construction of Zinc Finger

domains (Beerli et. al. 1998). The principle to join multiple fingers in tandem is illustrated in figure 16. The Sp1 framework is constituted by constant regions defined by the backbone and C and N terminal caps. The recognition helix sequence is variable and is the distinguishing feature between individual fingers with different binding sequences. When joining multiple fingers, a 5 amino acid linker sequence (TGEKP) is added between the backbones (Pabo et. al. 2001). Multiple fingers with different recognition helices can be joined this way. The terminal caps are provided only once. An important point to note is that recognition of a 5'-3' target takes place from the C terminal. That is, the C terminal finger recognizes the 5' triplet and so on. Figure 16 illustrates some ZiF domains that can be constructed with these principles.

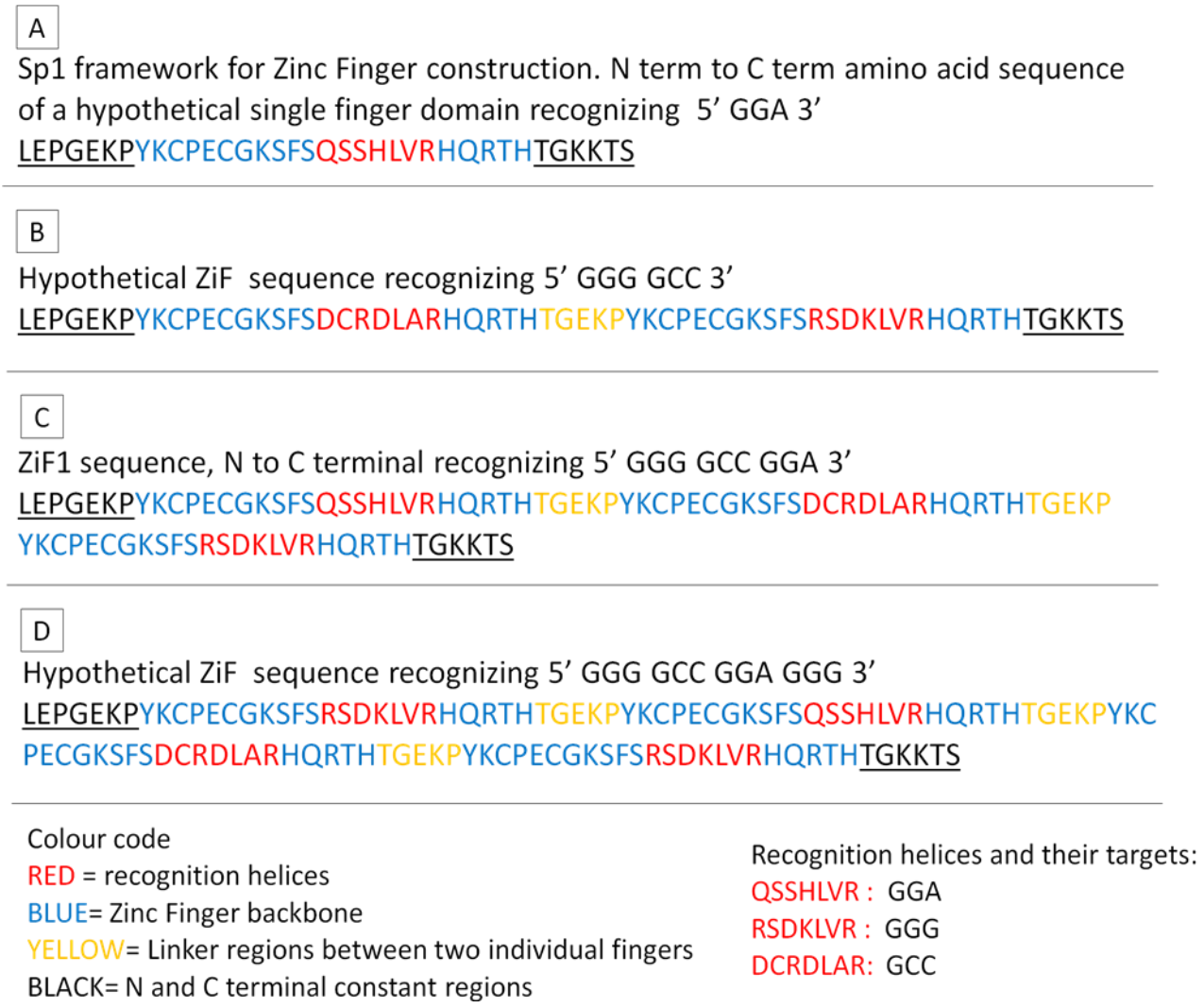


Figure 16: Examples for ZiF domain constructions with Sp1 framework. (A) Generic organisation of a hypothetical single ZiF domain. (B,C,D) The variable recognition helix is changed in distinct fingers.The N and C terminal caps are provided only once. Linker sequences are used between neighbouring fingers.

In this study, Four Zinc Finger domains which have a target recognition sequence of 9bp were used. These were synthesized by an overlap extension PCR protocol (see methods). ZiF1 and ZiF2 have been described earlier and their target association constants determined (Beerli et.al., 1998). ZiF3 and ZiF 4 are novel ZiF proteins. The recognition helices of individual fingers were selected according to reports by the Barbas group (Beerli et. al., 1998, Segal et. al., 1999, Drier et. al., 2001, Drier et. al., 2005.)

Table 2: Details of ZiF domains used in this study

Zinc Finger Domain	Predicted target sequence (dsDNA, 5'-3')	Recognition helices and targets of Individual fingers selected for assembly
ZiF1, K_d=35nM	GGG GCC GGA	QSSHLVR (GGA) DCRDLAR (GCC) RSDKLVR (GGG)
Zif2, K_d=25nM	GCC GCA GTG	RKDSLVR (GTG) QSGDLRR (GCA) DCRDLAR (GCC)
Zif3	ACA ACT ATT	HKNALQN (ATT) THLDLIR (ACT) SPADLTR (ACA)
Zif4	CGT CTA CAG	RADNLTE (CAG) QNSTLTE (CTA) SRRTCRA (CGT)

TAL domains as alternatives to Zinc Fingers

TAL domains are motifs that recognize a single base pair dsDNA target. Tandem assembly of such individual motifs enable synthesis of a TAL protein that has a longer target recognition sequence. Hence they can serve the same objective as custom made Zinc Finger domains. In this study, we explored the utilization of TAL domains in place of Zinc Fingers. Custom made TALEN constructs for targeted gene knockout (Life Technologies) were used as the source of the TAL domains. In these constructs, TAL domains are fused to Fok1 endonuclease. The TAL domain

sequence was excised by restriction digestion with appropriate enzymes and cloned upstream and downstream of the EGFP halves or in the pDisplay vector (see methods). The TAL domains used and their corresponding target recognition sequences are shown in Table 3.

Table 3: Binding sites of TAL domains used in this study

Domain Name	Source of the gene. Parent TALEN construct (Life Technologies)	Recognition sequence, dsDNA, 5' to 3'
TAL1	Tal SB1_Left	TCACTCACATTGTTAGACA
TAL2	Tal SB1_Right	TGAACGAGGCAACTCAGGG
TAL3	Tal SB2_Left	TCTCGCTCCGCCGCTCCAT
TAL4	Tal SB2_Right	TGCCGGTGGGAACTTTGTC

Combining Programmable binding with Conditional fluorescence

By combining Split –EGFP system and Zinc Fingers/TALs we proposed to engineer a new 'protein couple' that will realise the objectives of marking specific DNA sequences with EGFP fluorescence (Figure 5). The recombination of EGFP is driven by their recruitment to proximal binding sites. It is thus clear that unbound protein will not recombine or fluoresce. Consequently, fluorescence is expected ONLY when both the halves are targeted to a predecided target sequence. A 9 bp target for each ZiF domain implies an 18 bp target for the couple. In the context of the human genome, 18 bp is the minimum sequence to give a unique site. The TAL domains have a much longer recognition site and the TAL-EGFP half couple will have a combined target site of upto 40 bp. Some aspects regarding the design of fusion constructs should be noted. An important parameter that will decide the success of the strategy is the distance between the binding sites of the two EGFP half proteins. If the binding sites are too far apart, then the EGFP halves would be separated over a large distance and complementation would fail. It is thus expected that there would

be a range of distances upto which the bound EGFP halves are close enough for efficient complementation. There is no linker region between the EGFP halves and the ZIF/TAL domain This was decided so as to minimize the 'permissible range' of distances between the binding sites. A long linker would enable complementation over a larger range of separation (say for example 0 to 15 bp) whereas in this case, the absence of a linker would constrain the 'permissible range' to a shorter number (say 0 to 5 bp). This would give more stringency for specific complementation.

Expression from the p2A constructs

The p2A self cleaving peptide can give stoichiometric amounts of two proteins coded upstream and downstream to it. The mechanism involved is believed to be a 'ribosome skip step' that separates the two proteins during translation (Kim et al, 2011). The A12-p2A-B21 construct is thus expected to give A12 and B21 in equal quantities from a single vector (Figure 17). Polyclonal antibody to full length EGFP that should detect both the halves of EGFP was used for protein detection by Western blotting. Consequently, we have constructed A12-p2A-mCherry-B21 and A12-mCherry-p2A-mCherry-B21 for stoichiometric production of tagged A12 and B21.

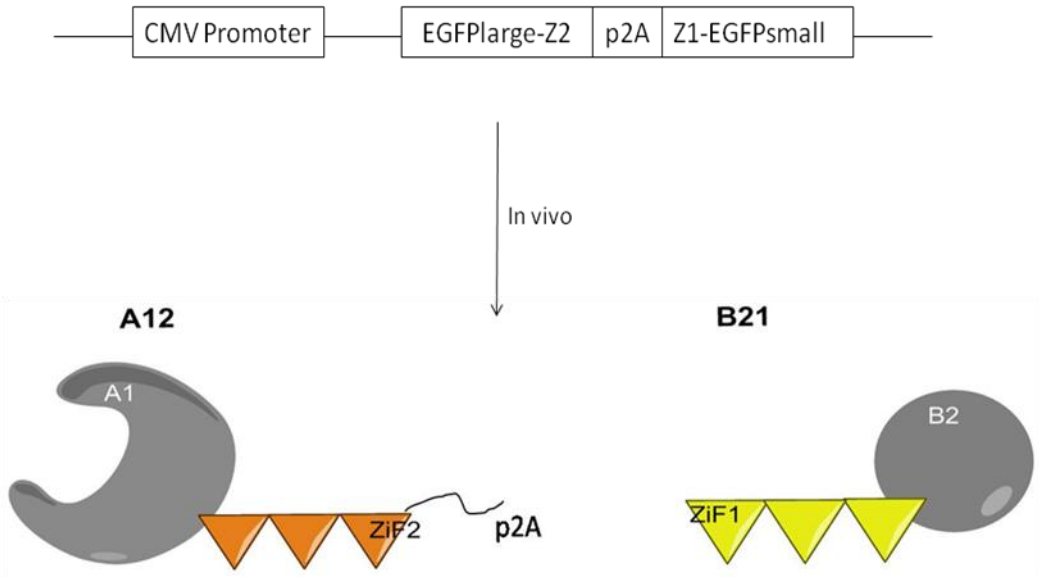


Figure 17: **Schematic for expression from the construct A12-p2A-B21.** Upper half shows reading frame configuration of the fusion construct. In vivo A12 and B21 are expressed in equal amounts and separated during translation. The p2A fragment remains attached to the C terminal of A12.

Cell surface expression of ZiFs/TALs as a tool to achieve programmable cell adhesion

The export of proteins to the cell surface requires targeting sequences for the secretory machinery as well as structural features (like transmembrane domains) for successful incorporation into the plasma membrane. Genes cloned in the pDisplay vector (Invitrogen) have an upstream IgG-k leader sequence and a downstream transmembrane domain. These proteins are hence predicted to be anchored to the extracellular face of the plasma membrane. In addition, presence of an upstream HA tag and a downstream myc tag in the vector sequence facilitates protein detection by immunoblotting/immunofluorescence. The ZiF and TAL domains previously described were cloned in the pDisplay vector and evaluated for surface expression.

Bacterial expression of ZiF-EGFP half proteins

For in vitro testing of fluorescence reconstitution, it is necessary to purify A12 and B21/B23 as proteins without any extra amino acids/tags on the N-terminal of A12 and the C terminal of B21/B23. Thus A12 and B23 were fused to HIS tags at the C-terminus and N-terminus respectively (A12-HIS in pTriex and HIS-B23 in pTriex, see methods) However A12-HIS failed to express in a variety of conditions whereas HIS-B23 was insoluble. Although EGFP half proteins can be extracted and refolded from inclusion bodies, the effect of refolding on ZiF activity is unknown. Consequently, we investigated the pC62/Caspase6 system for protein purification. Here, A12, B21 and B23 are tagged at the N-terminus with GST. The GST tag can be cleaved with Caspase6 to obtain tagless proteins (Purbey at. al. 2006) Figure 8 shows that A12, B21 and B23 are amenable to protein purification using this approach. It must be noted that C6 cleavage efficiencies for B21 and B23 were poor as compared to A12. However, since the GST tag is at the N-terminal, the important C-terminal of B21 and B23 are untagged and as such, a cleavage of GST is not strictly necessary. In contrast, tag removal from A12 is strictly necessary since it is at the N-terminal. We thus demonstrate that untagged A12 and GST tagged B21 and B23 can be purified in soluble form with the pC62/Caspase6 system. In future, we will pursue the purification on a larger scale for an in vitro assay.

An extra domain that shows good folding and solubility characteristics is necessary for expression in vivo

The insolubility of the ZIF-EGFP half protein seen in the bacterial system is also seen in mammalian expression. Likewise, we saw that although EGFP in pDisplay expresses well, its replacement by a ZiF domain results in drastic reduction of soluble expression due to proteosomal degradation. However, our results suggest that both these problems of insolubility and degradation can be solved by fusing these proteins to an appropriate partner. Thus mCherry-B21 showed excellent solubility and expression. Similarly, ZiF-EGFP fusions in pDisplay were rescued from degradation. Intriguingly, A12-mCherry showed only a modest increase in solubility. We chose mCherry and EGFP as fusion partners as these proteins are known to fold well (Tsien 1998) and in addition have the advantage to act as fluorescent reporters. mCherry has an absorption-emission character that is distinct from EGFP and is not expected to interfere with the EGFP signal in case successful complementation occurs. It is interesting to note that insolubility in bacterial systems could be addressed by essentially the same fusion partner strategy, in this case the partner being GST. Poor folding of proteins can be the reason for aggregation and/or degradation. In this case, we propose that the proteins are unstructured on their own and the well folding protein fusion partner (EGFP, mCherry, GST) provides a stable scaffold that organizes these unstructured domains. The final protein thus has a large structured and well folded domain (the fusion partner) which solubilises the protein and rescues it from aggregation or degradation. Our results suggest that inclusion of a well folding domain in the design of ZiF-EGFP half proteins is extremely important for producing soluble and active proteins. Similarly, such a domain should also be included for efficient expression of ZiFs in pDisplay. The low solubility of A12-mCherry indicates that the presence of the EGFP large half at the N terminus is especially deleterious for solubility.

TALs as superior alternatives to ZiFs

When ZiF domains are replaced by TALs, effective soluble expression is seen, circumventing the need for an additional soluble domain. In case of EGFP half-TAL constructs cloned here, there are two Kozak sequences in the ORF that is giving two protein products (Figure 13B). These constructs thus have to be modified by removal of the start codon (or Kozak sequence) of the TAL gene to obtain expression of a

single polypeptide. This is not an issue for the TAL-EGFP half constructs that have a single Kozak sequence and give only the required TAL-EGFP half fusion protein. The TALs used here have a much longer binding target site and are significantly larger in size than their ZiF counterparts. Each TAL domain is typically about 800 amino acids long as compared to the 100 amino acid ZiF domain. The large size can be a factor for efficient folding and soluble expression. It is possible that solubility in ZiF proteins may be achieved by increasing the number of fingers to give a larger ZIF domain. These preliminary results suggest that TAL domains can be a superior alternative to Zinc Fingers in matters of solubility and expression while designing new fusion proteins.

Important inferences from this study are as follows:

1) EGFP-ZiF half fusion proteins are insoluble and must be tagged with a well folding partner protein like mCherry for efficient expression in vivo.

2) ZiF domains in pDisplay are targeted to the proteosome for degradation. The proteins can be rescued from degradation by fusing them with a well folding partner protein like EGFP.

3) TAL domains seem to be superior alternatives to Zinc Fingers for the ideas proposed in this study.

4) The pC62/caspase6 system can be used for purification of soluble EGFP-ZiF half proteins for in vitro studies.

Future perspectives

Conclusions presented here are the foundations for future strategies to attain the objectives mentioned earlier. The insolubility and degradation problems showed by the model proteins were unexpected and were a significant roadblock for further in vivo studies. We have provided solutions to these problems. Work to provide proof of principle is underway. Important issues that need to be addressed are as follows:

1) ZiF-EGFP fusions and TAL domains are efficiently expressed as pDisplay constructs but their localization to the plasma membrane has not been established. Immunostaining and confocal microscopy to check localization is the next necessary experiment. If the proteins are expressed on the plasma membrane, their activity has to be checked with a binding assay with target dsDNA.

2) mCherry-ZiF-EGFP half proteins and TAL-EGFP half fusions will be used as models to test fluorescence reconstitution in presence of target dsDNA sequences.

3) Recombinant ZiF-EGFP half proteins will be purified with the pC62/Caspase6 system for in vitro experiments with target oligos.

Methods and plasmid construction details

Zinc Finger domain gene synthesis by overlap extension PCR

One Zinc Finger domain of 90 amino acids is made up of 3 individual Zinc Fingers. Four different Zinc Finger Domains (ZiF1, ZiF2, ZiF3 and ZiF4) were synthesized by an overlap extension PCR. For each domain, 6 overlapping oligos were used for the synthesis of the complete gene. They are named as oligo1 to oligo 6 (Table 4).

Table 4 : Oligos used for gene synthesis of the ZiF domains

	ZiF1	ZiF2	ZiF3	ZiF4
Oligo1	SG6000	SG6006	SG6012	SG6018
Oligo2	SG6001	SG6007	SG6013	SG6019
Oligo3	SG6002	SG6008	SG6014	SG6020
Oligo4	SG6003	SG6009	SG6015	SG6021
Oligo5	SG6004	SG6010	SG6016	SG6022
Oligo6	SG6005	SG6011	SG6017	SG6023

As an example, the design for ZiF1 is illustrated in Figure 18.

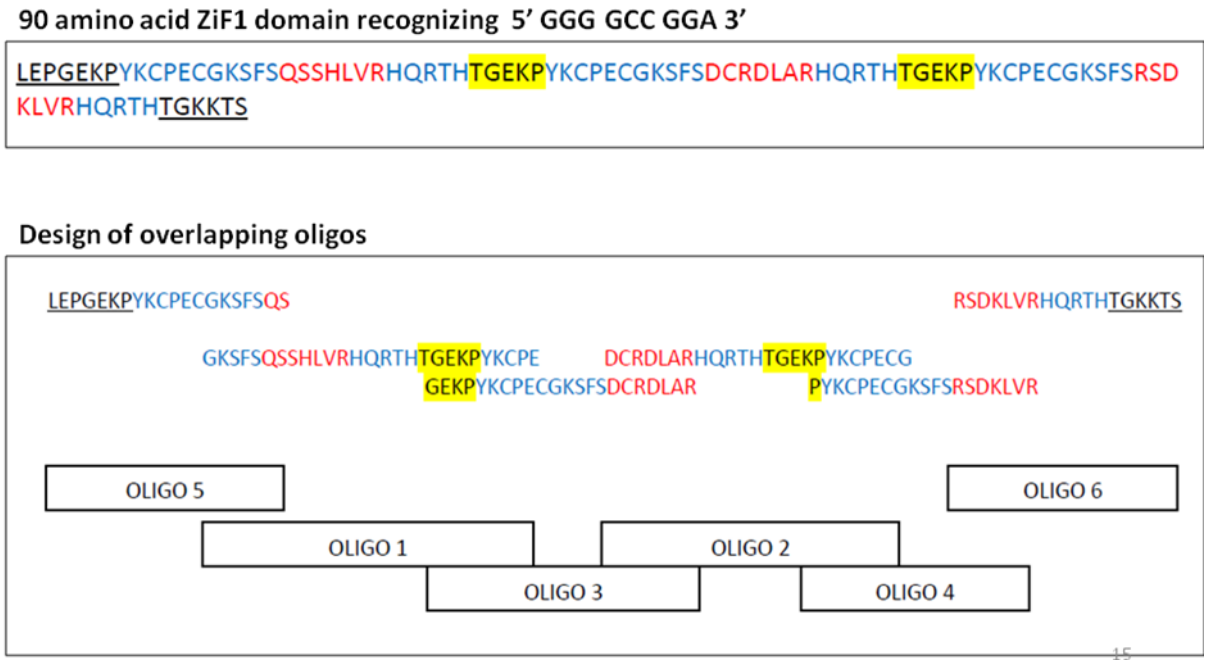


Figure 18: Design of overlapping oligos for ZiF1. The 90 amino acid ZiF1 domains was broken into 6 overlapping regions and oligos were designed corresponding to these (for clarity, amino acid sequence is shown rather than DNA sequence).

The gene synthesis consists of a two step-PCR protocol. In the first step, oligo 2 (0.4 pmol), oligo 3 (0.4 pmol) , oligo 1 (40 pmol) and oligo 4 (40 pmol) are mixed and cycled in a standard 25 μL PCR reaction (30 sec at 94 ^{o}C denaturing; 30 sec at 60 ^{o}C annealing and 30sec at 72 ^{o}C extension; 25 cycles). An aliquot of this reaction mixture is used a template for the second step. It is mixed with oligo5 and oligo 6 (40 pmol each) and cycled in a standard 25 μL PCR reaction (30 sec 94 ^{o}C denaturing; 30 sec 60 ^{o}C annealing and 30 sec 72 ^{o}C extension; 25 cycles). The final product was purified using phenol-chloroform extraction and used for cloning.

Cloning

Standard protocols were followed for Cloning. Briefly, the PCR products were purified by Phenol-Chloroform-Isoamyl (PCI) procedure. This purified insert and vector were digested with the appropriate restriction enzymes (NEB) and purified using Gel extraction kit (QIAGEN). After ligation, the mixture was transformed into DH5-alpha strain of *E. coli* and plated with the required antibiotic resistance. Colonies were screened for positive clones by colony PCR or Miniprep digestions.

Parent vectors used for cloning

pC6-2-73. (Ampicillin resistance)

This is a variant of the pC6-2 vector. The GST tag along with terminal VEID Caspase 6 cleavage site is upstream of the MCS.

Sequence:

.....GCATCTCCAAATCGGACTGGTTCCGCGTGGATCTGTTGAA*ATCGAT*GGA

TCTGTTGAAATCGATGGATCCCCAGGAATTCCCGGGTCGA*CTCGAG*CGGCCGC

ATCGTGACTGACTGACGA

Underlined sequences show the Cla1 and Xho1 restriction sites which have been used for cloning.

Red indicates the C6 site. Blue denotes part of the upstream GST tag.

Other parent vectors used were:

pEGFPN1, pmCherryC1, pTriex 3.3 Neo, pDisplay, pCDNA 3.1 myc/His (-) C.

Total of 45 constructs are described in the study.

For clarity, the list of constructs can be divided into the following groups with a brief description of each type:

1) Those in pC6-2-73: Split EGFP-ZiF proteins tagged with GST and a C6 site for bacterial expression and tagless protein purification
   - A1 in pC62-73
   - B2 in pC62-73
   - A12 in pC62-73
   - B21 in pC62-73
   - B23 in pC62-73

2) Those in pTriex 3.3 Neo: Split EGFP-ZiF proteins with His tag for both mammalian and bacterial expression
   - A12-HIS in ptriex
   - HIS-B21 in ptriex
   - HIS-B23 in ptriex

3) p2A constructs in pEGFPN1: Constructs with p2A self cleaving peptide for generating both the ZiF-EGFP half proteins from a single plasmid in stoichiometric amounts.
   - p2A in pEGFPN1
   - A12-p2A in pEGFPN1
   - A12-p2A-B21
   - A12-p2A-mCherry-B21
   - A12-mCherry-p2A-mCherry-B21

4) TAL in pCDNA constructs: Split-EGFP-TAL constructs. 4 different TAL domains were cloned in pCDNA. EGFP halves were cloned upstream and downstream of each of them.
   - TAL1 in pCDNA
   - TAL2 in pCDNA
   - TAL3 in pCDNA

- TAL4 in pCDNA
- EGFPlarge-TAL1 in pCDNA
- EGFPlarge-TAL2 in pCDNA
- EGFPlarge-TAL3 in pCDNA
- EGFPlarge-TAL4 in pCDNA
- EGFPsmall-TAL1 in pCDNA
- EGFPsmall-TAL2 in pCDNA
- EGFPsmall-TAL3 in pCDNA
- EGFPsmall-TAL4 in pCDNA
- TAL1-EGFPlarge in pCDNA
- TAL2-EGFPlarge in pCDNA
- TAL3-EGFPlarge in pCDNA
- TAL4-EGFPlarge in pCDNA
- TAL1-EGFPsmall in pCDNA
- TAL2-EGFPsmall in pCDNA
- TAL3-EGFPsmall in pCDNA
- TAL4-EGFPsmall in pCDNA

5) pDisplay constructs: Proteins cloned in the pDisplay vector are targeted to the plasma membrane.
   - EGFP in pDisplay
   - ZiF1in pDisplay
   - ZiF2in pDisplay
   - ZiF3in pDisplay
   - EGFP-ZiF1in pDisplay
   - EGFP-ZiF2in pDisplay
   - EGFP-ZiF3in pDisplay
   - ZiF2-EGFP in pDisplay
   - TAL1 in pDisplay
   - TAL2 in pDisplay
   - TAL3 in pDisplay
   - TAL4 in pDisplay

Table 5: List of oligos used for PCR/Cloning

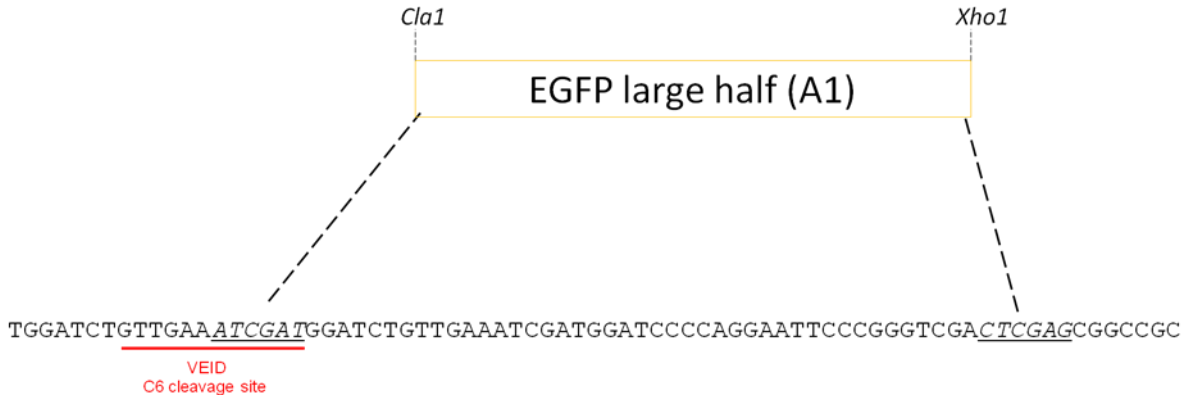
Oligo code	Description	Restriction site	Sequence (5' to 3')	length
SG 5996	egfp 1-158 F Cla1	Cla1	atattaatcgatATGGTGAGCAAGGGCGAGGAG	33
SG 5997	egfp 1-158 R Xho11	Xho1	atatatctcgagCTGCTTGTCGGCCATGATATA	33
SG 5998	egfp 159-239 F Cla1	Cla1	atagcgatcgatAAGAACGGCATCAAGGTGAAC	33
SG 5999	egfp 159-239 R Xho1, stop codon	Xho1	atagcgctcgagtcaCTTGTACAGCTCGTCCATGCC	36
SG 6231	egfp F BGL2	Bgl2	gcgcgcagatctATGGTGAGCAAGGGCGAGGAGCTGTTC	39
SG 6232	egfp R C6 site-Sal1	Sal1	gcgcgcgtcgacatccatttcaacCTTGTACAGCTCGTCCATGCCGAGAGT	51
SG 6233	ZiF F Bgl2	Bgl2	atatat agatct CTG GAG CCC GGG GAG AAG CCC	33
SG 6234	ZiF R Sal1	Sal1	gcgcgcgtcgacACTAGTTTTTTTACCGGTGTG	33
SG 7075	p2A peptide in eGFPN1 Kpn1F Bamh1 R	Kpn1, Bamh1	cgcGGTACCgGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCA GGCTGGAGACGTGGAGGAGAACCCTGGACCTcgGGATCCata	87
SG 7076	p2A peptide in eGFPN1 Kpn1F Bamh1 R	Kpn1, Bamh1	tatGGATCCcgAGGTCCAGGGTTCTCCTCCACGTCTCCAGCCTGCTT CAGCAGGCTGAAGTTAGTAGCTCCGCTTCCcGGTACCgcg	87
SG 7077	ZiF split gfp -A1+2- in eGFPN1 Nhe1F	Nhe1	atagctagcgATGGTGAGCAAGGGCGAGGAG	31
SG 7078	ZiF split gfp -A1+2- in eGFPN1 Kpn1R	Kpn1	gcgggtaccgtACTAGTTTTTTTACCGGTGTG	32
SG 7079	ZiF split GFP-B2+3- in eGFPN1 Bamh1F	Bamh1	ataggatccaCTGGAGCCCGGGGAGAAGCCC	31
SG 7080	ZiF split GFP-B2+3- in eGFPN1 Not1R	Not1	atagcggccgcTCACTTGTACAGCTCGTCCATGCC	35
SG 7197	splitGFP-ZiF- A1+2 Nco1 F	Nco1	gcgc*ccATGG*TGAGCAAGGGCGAGGAG	27
SG 7198	splitGFP-ZiF- A1+2 with Cterm 8XHis tag Not1 R	Not1	atatgcggccgcTCAGTGATGGTGATGGTGATGGTGGTGACTAGTTT TTTTACCGGTGTGAGT	63
SG 7199	splitGFP-ZiF- B2+1 with Nterm 8XHis tag Nco1 F	Nco1	atat*ccatgg*cg*CACCACCATCACCATCACCATCAC*CTGGAGCCCGGG GAGAAGCCC	57
SG6000	ZiF1-pePCR-oligo1		GGT AAG TCC TTC TCT CAG AGC TCT CAC CTG GTG CGC CAC CAG CGT ACC CAC ACG GGT GAA AAA CCG TAT AAA TGC CCA GAG	81
SG6001	ziF1-pePCR-oligo2		GTG ACT GCC GCG ACC TTG CTC GCC ATC AAC GCA CTC ATA CTG GCG AGA AGC CAT ACA AAT GTC CAG AAT GTG GC	74
SG6002	ziF1-pePCR-oligo3		GCG AGC AAG GTC GCG GCA GTC ACT AAA AGA TTT GCC GCA CTC TGG GCA TTT ATA CGG TTT TTC ACC	66
SG6003	ziF1-pePCR-oligo4		ACG CAC CAG CTT GTC AGA GCG GCT GAA AGA CTT GCC ACA TTC TGG ACA TTT GTA TGG C	58
SG6004	ziF1-pePCR-oligo5	Cla1	gcgcgc atcgat CTG GAG CCC GGG GAG AAG CCC TAT AAA TGT CCG GAA TGT GGT AAG TCC TTC TCT CAG AGC	72
SG6005	ziF1-pePCR-oligo6	cla1	gcgcgc atcgat ACT AGT TTT TTT ACC GGT GTG AGT ACG TTG GTG ACG CAC CAG CTT GTC AGA GCG	66
SG6006	ZiF2-pePCR-oligo1		GGT AAG TCC TTC TCT CGC AAG GAC TCT CTG GTG CGC CAC CAG CGT ACC CAC ACG GGT GAA AAA CCG TAT AAA TGC CCA GAG	81
SG6007	ZiF2-pePCR-oligo2		GTC AGT CTG GCG ACC TTC GCC GCC ATC AAC GCA CTC ATA CTG GCG AGA AGC CAT ACA AAT GTC CAG AAT GTG GC	74
SG6008	ZiF2-pePCR-oligo3		GCG GCG AAG GTC GCC AGA CTG ACT AAA AGA TTT GCC GCA CTC TGG GCA TTT ATA CGG TTT TTC ACC	66
SG6009	ZiF2-pePCR-oligo4		GCG AGC AAG GTC GCG GCA GTC GCT GAA AGA CTT GCC ACA TTC TGG ACA TTT GTA TGG C	58
SG6010	ZiF2-pePCR-oligo5	Xho1	gcgcgc ctcgag CTG GAG CCC GGG GAG AAG CCC TAT AAA TGT CCG GAA TGT GGT AAG TCC TTC TCT CGC AAG	72
SG6011	ZiF2-pePCR-oligo6	Xho1	gcgcgc ctcgag TCA ACT AGT TTT TTT ACC GGT GTG AGT ACG TTG GTG GCG AGC AAG GTC GCG GCA GTC	69
SG6012	ZiF3-pePCR-oligo1		GGT AAG TCC TTC TCT CAC AAG AAC GCT CTG CAG AAC CAC CAG CGT ACC CAC ACG GGT GAA AAA CCG TAT AAA TGC CCA GAG	81
SG6013	ZiF3-pePCR-oligo2		GTA CCC ACC TTG ACC TTA TCC GCC ATC AAC GCA CTC ATA CTG GCG AGA AGC CAT ACA AAT GTC CAG AAT GTG GC	74
SG6014	ZiF3-pePCR-oligo3		GCG GAT AAG GTC AAG GTG GGT ACT AAA AGA TTT GCC GCA CTC TGG GCA TTT ATA CGG TTT TTC ACC	66
SG6015	ZiF3-pePCR-oligo4		GCG GGT AAG GTC AGC GGG AGA GCT GAA AGA CTT GCC ACA TTC TGG ACA TTT GTA TGG C	58
SG6016	ZiF3-pePCR-oligo5	Cla1	gcgcgc atcgat CTG GAG CCC GGG GAG AAG CCC TAT AAA TGT CCG GAA TGT GGT AAG TCC TTC TCT CAC AAG	72

SG6017	ZiF3-pePCR-oligo6	Cla1	gcgcgc atcgat ACT AGT TTT TTT ACC GGT GTG AGT ACG TTG GTG GCG GGT AAG GTC AGC GGG AGA	66
SG6018	ZiF4-pePCR-oligo1		GGT AAG TCC TTC TCT CGC GCT GAC AAC CTG ACC GAG CAC CAG CGT ACC CAC ACG GGT GAA AAA CCG TAT AAA TGC CCA GAG	81
SG6019	ZiF4-pePCR-oligo2		GTC AGA ACA GTA CGC TTA CGG AAC ATC AAC GCA CTC ATA CTG GCG AGA AGC CAT ACA AAT GTC CAG AAT GTG GC	74
SG6020	ZiF4-pePCR-oligo3		TTC CGT AAG CGT ACT GTT CTG ACT AAA AGA TTT GCC GCA CTC TGG GCA TTT ATA CGG TTT TTC ACC	66
SG6021	ZiF4-pePCR-oligo4		AGC GCG ACA AGT GCG GCG ACT GCT GAA AGA CTT GCC ACA TTC TGG ACA TTT GTA TGG C	58
SG6022	ZiF4-pePCR-oligo5	Xho1	gcgcgc ctcgag CTG GAG CCC GGG GAG AAG CCC TAT AAA TGT CCG GAA TGT GGT AAG TCC TTC TCT CGC GCT	72
SG6023	ZiF4-pePCR-oligo6	Xho1	gcgcgc ctcgag TCA ACT AGT TTT TTT ACC GGT GTG AGT ACG TTG GTG AGC GCG ACA AGT GCG GCG ACT	69
SG7511	EGFP Reverse to clone into ZiF-pDisplay	Bgl2	gcgcgcagatctCTTGTACAGCTCGTCCATGCC	33
SG7512	EGFP Forward to clone into ZiF-pdisplay	Sal1	gcgcgcgtcgacATGGTGAGCAAGGGCGAGGAG	33
SG7513	large EGFP half reverse to clone into N termTAL-pcDNA	Not1	atatatgcggccgccCTGCTTGTCGGCCATGATATA	36
SG7514	large EGFP half forward to clone into C termTAL-pcDNA	Bamh1	tatataggatccATGGTGAGCAAGGGCGAGGAG	33
SG7515	large EGFP half reverse to clone into C termTAL-pcDNA	Hind3	gcgcgcaagcttCTACTGCTTGTCGGCCATGATATA	36
SG7517	small EGFP half reverse to clone into N termTAL-pcDNA	Not1	atatatgcggccgccCTTGTACAGCTCGTCCATGCC	36
SG7518	small EGFP half forward to clone into C termTAL-pcDNA	Bamh1	gcgctaggatccAAGAACGGCATCAAGGTGAAC	33
SG7519	small EGFP half reverse to clone into C termTAL-pcDNA	Hind3	cgcgcgaagcttCTACTTGTACAGCTCGTCCATGCC	36
SG7520	TAL domain forward to clone in pDisplay	Sal1	atatatgtcgacaGGCGGAGTGGACCTGAGAACA	34
SG7521	TAL domain reverse to clone in pDisplay	Sal1	atatatgtcgacGGCCACTCTGTGAGAGGTCCG	33
SG7522	Cloning mCherry in A12-p2A-B21 after A12 Forward	Kpn1	atatatggtaccgATGGTGAGCAAGGGCGAGGAG	34
SG7523	Cloning mCherry in A12-p2A-B21 after A12 Reverse	Kpn1	atatatggtaccccCTTGTACAGCTCGTCCATGCC	35
SG7524	Cloning mCherry in A12-p2A-B21 before B21 Forward	Bamh1	atatatggatccaATGGTGAGCAAGGGCGAGGAG	34
SG7525	Cloning mCherry in A12-p2A-B21 before B21 Reverse	Bamh1	atatatggatccccCTTGTACAGCTCGTCCATGCC	35
SG7683	EGFPlarge Forward to clone into TAL-pcdna, Apa1	Apa1	tatatagggcccaccATGGTGAGCAAGGGCGAGGAG	36
SG7684	EGFPsmall Forward to clone into TAL-pcdna, Apa1	Apa1	atatatgggcccaccATGAAGAACGGCATCAAGGTGAAC	39

Maps and constructs

All constructs were validated by PCR, restriction digestion and sequencing unless mentioned otherwise.

A1 in pC6-2-73.

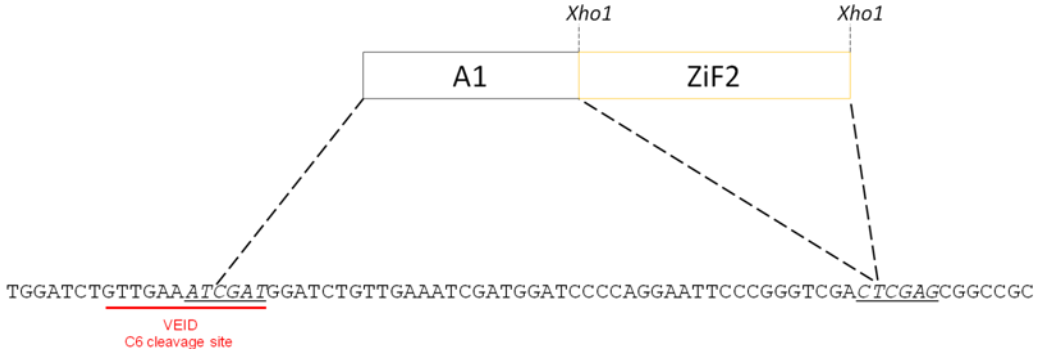


Plasmid used for cloning: pC62-73
Insert: EGFP large half (A1). Generated by PCR with primers SG5996 and SG5997 from the template 'pEGFPN1'
Cloned in the Cla1 and Xho1 sites.

B2 in pC6-2-73

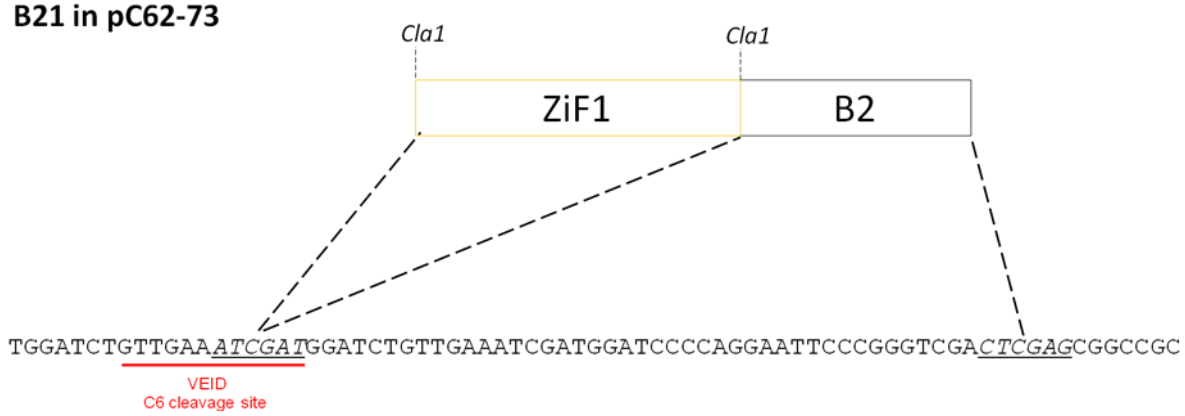
Plasmid used for cloning: pC62-73
Insert: EGFP small half (B2). Generated by PCR with primers SG5998 and SG5999 from the template 'pEGFPN1'
Cloned in the Cla1 and Xho1 sites.

A12 in pC6-2-73



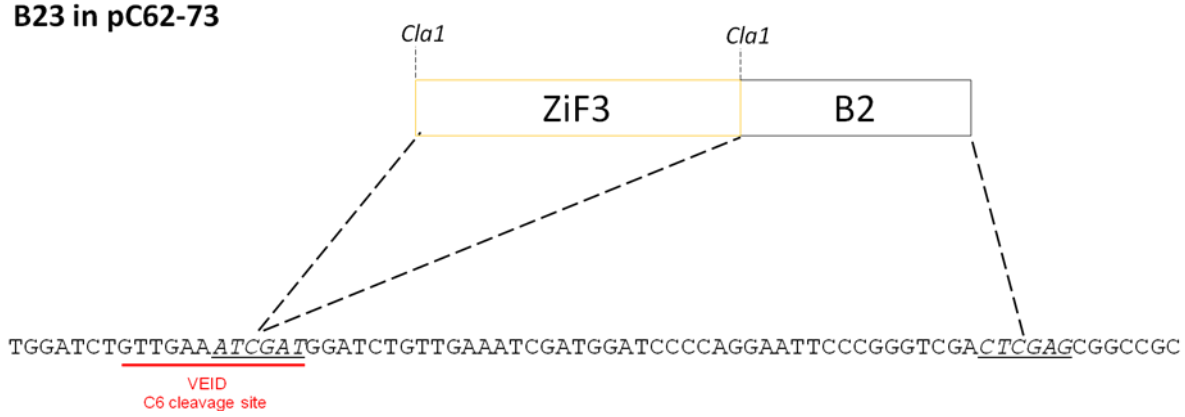
Plasmid used for cloning: A1 in pC62-73
Insert: ZiF2. Assembled by an overlap extension PCR protocol.
Cloned in Xho1 site.

B21 in pC62-73



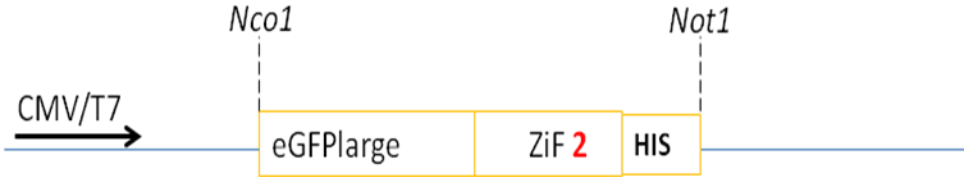
Plasmid used for cloning: B2 in pC62-73
Insert : ZiF1. Generated by overlap extension PCR.
Cloned in Cla1 site

B23 in pC62-73



Plasmid used for cloning: B2 in pC62-73
Insert : ZiF1. Generated by overlap extension PCR.
Cloned in Cla1 site

A12-HIS in ptriex



Plasmid used for cloning: pTriex 3.3 Neo
Insert: EGFPlarge-ZiF2-His (A12-His) Here, 'A1' denotes the EGFP large half and '2' denotes the ZiF2 domain. His denotes 8X HIS tag.
Cloned in the Nco1 and Not1 restriction sites.
For cloning, the insert was PCR amplified from the template 'A12 in pC62'' with primers SG7197 and SG7198.

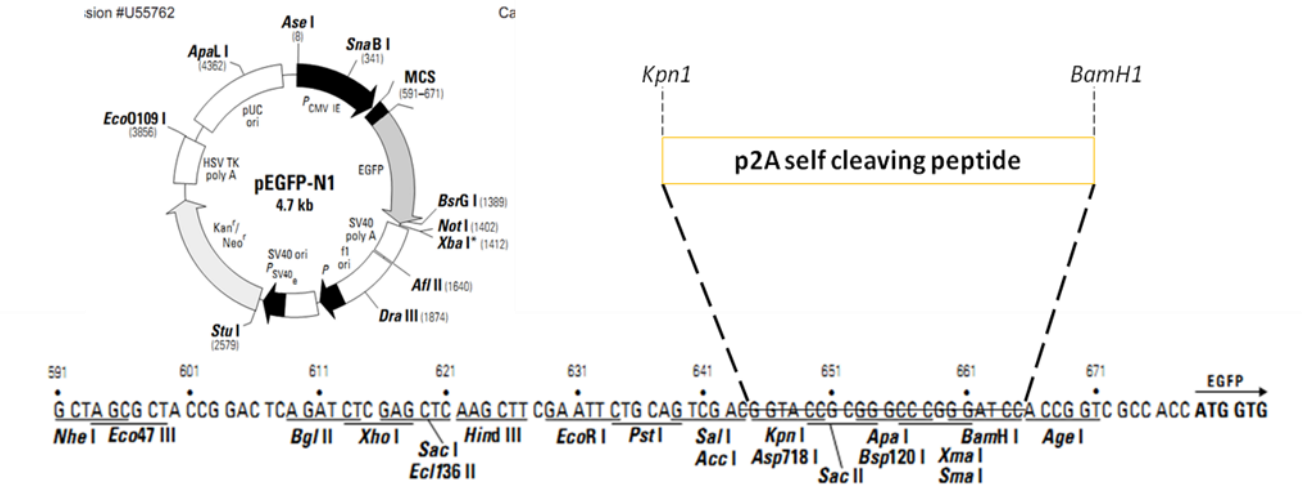
HIS-B21 in ptriex

Plasmid used for cloning: pTriex 3.3 Neo
Insert: HIS-Zif1-EGFPsmall (HIS-B21) Here, 'B2' denotes the EGFP small half and '1' denotes the ZiF1 domain. His denotes 8X HIS tag.
Cloned in the Nco1 and Not1 restriction sites.
For cloning, the insert was PCR amplified from the template 'B21 in pC62'' with primers SG7199 and SG7080.

HIS-B23 in ptriex

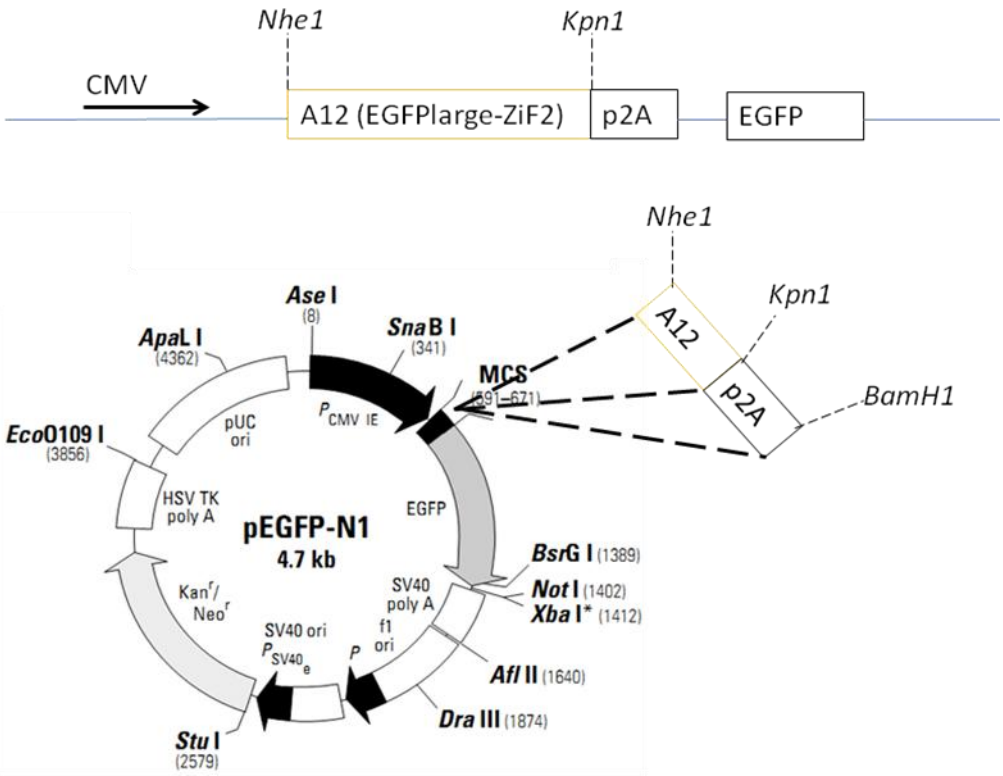
Plasmid used for cloning: pTriex 3.3 Neo
Insert: HIS-Zif3-EGFPsmall (HIS-B23) Here, 'B2' denotes the EGFP small half and '3' denotes the ZiF3 domain. HIS denotes 8X HIS tag.
Cloned in the Nco1 and Not1 restriction sites.
For cloning, the insert was PCR amplified from the template 'B23 in pC62'' with primers SG7199 and SG7080.

p2A5 in pEGFPN1



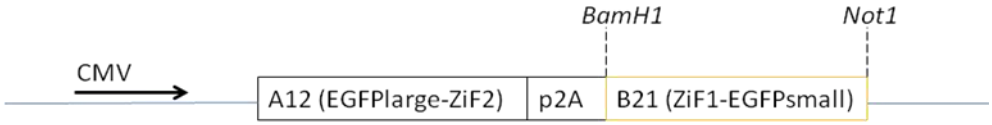
Plasmid used for cloning: pEGFPN1
Insert: p2A self cleaving peptide. The insert is 81bp long and was generated by annealing the oligos SG7075 and SG7076. The annealed product was then digested with Kpn1 and BamH1 and ligated with the cut vector.

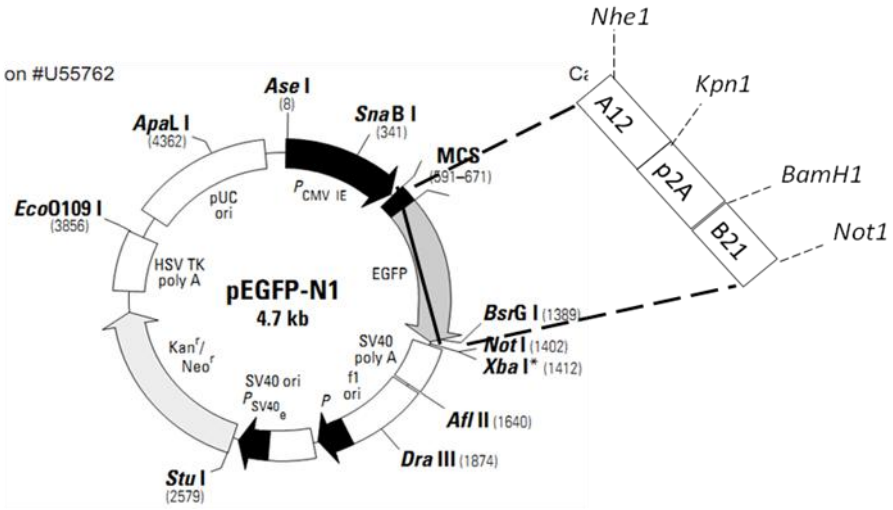
A12-p2A5 in pEGFPN1



Plasmid used for cloning: p2A5 in pEGFPN1
Insert: EGFPlarge half- ZiF2 (A12). Obtained by PCR with primers SG7077 and SG7078 from the template ‘A12 in pC62’’
Cloned in Nhe1 and Kpn1 restriction sites.

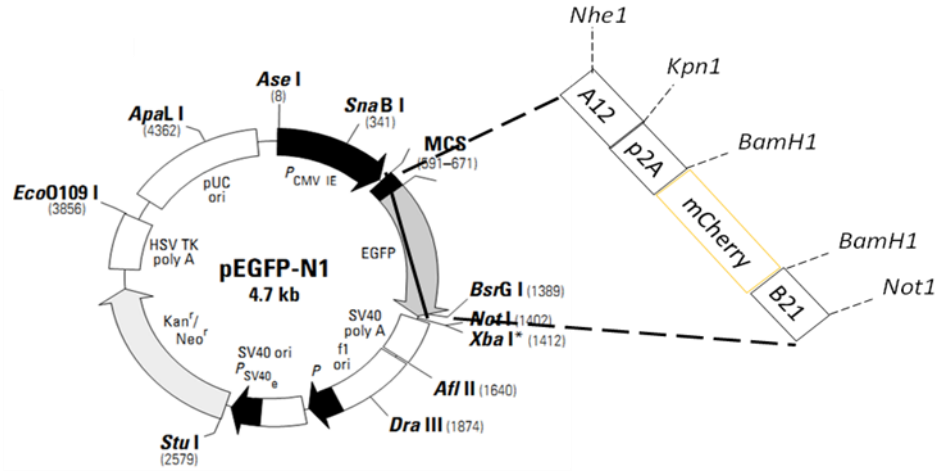
A12-p2A5-B21





Plasmid used for cloning: A12-p2A5 in pEGFPN1
Insert: ZiF1-EGFPsmall half (B21). Obtained by PCR with primers SG7079 and SG 7080 from the template 'B21 in pC62''
Cloned in the BamH1 and Not1 sites. EGFP gene is thus absent in the resulting construct.
Comment: The final construct map is shown above. It encodes the two splitgfp-Zif proteins (A12 and B21) separated by the p2A peptide.

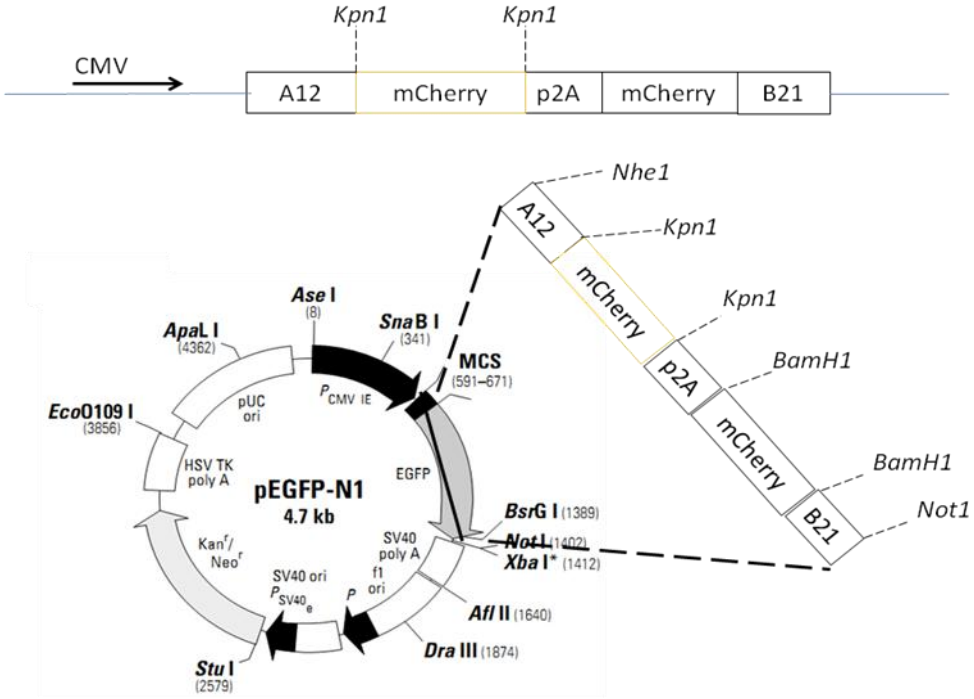
A12-p2a-mCherry-B21



Plasmid used for cloning: A12-p2A-B21

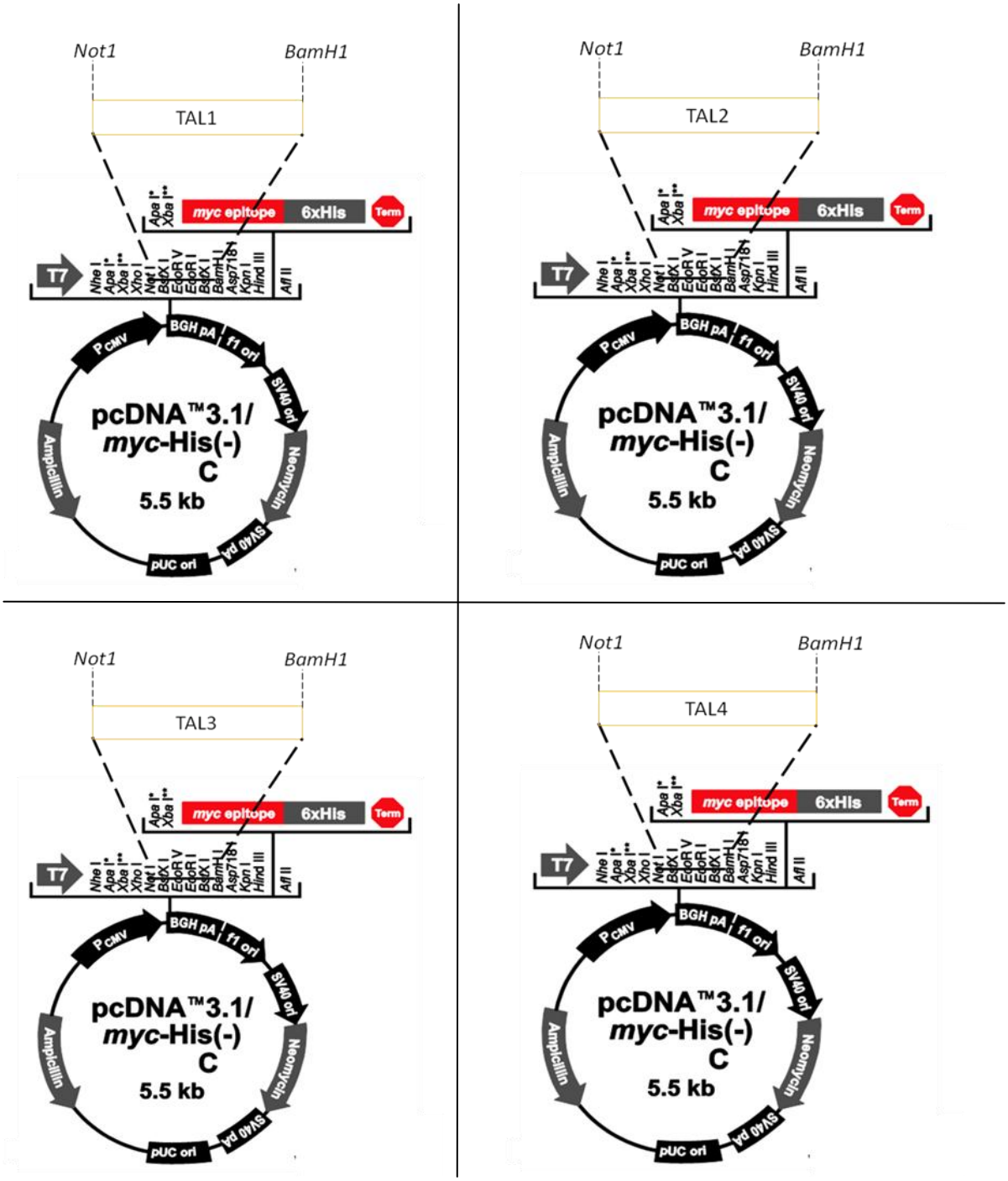
Insert: mCherry. Generated by PCR with primers SG7524 and SG7525 from the template ‘pmCherry C1’
Cloned in BamH1 site.

A12-mcherry-p2A-mcherry-B21



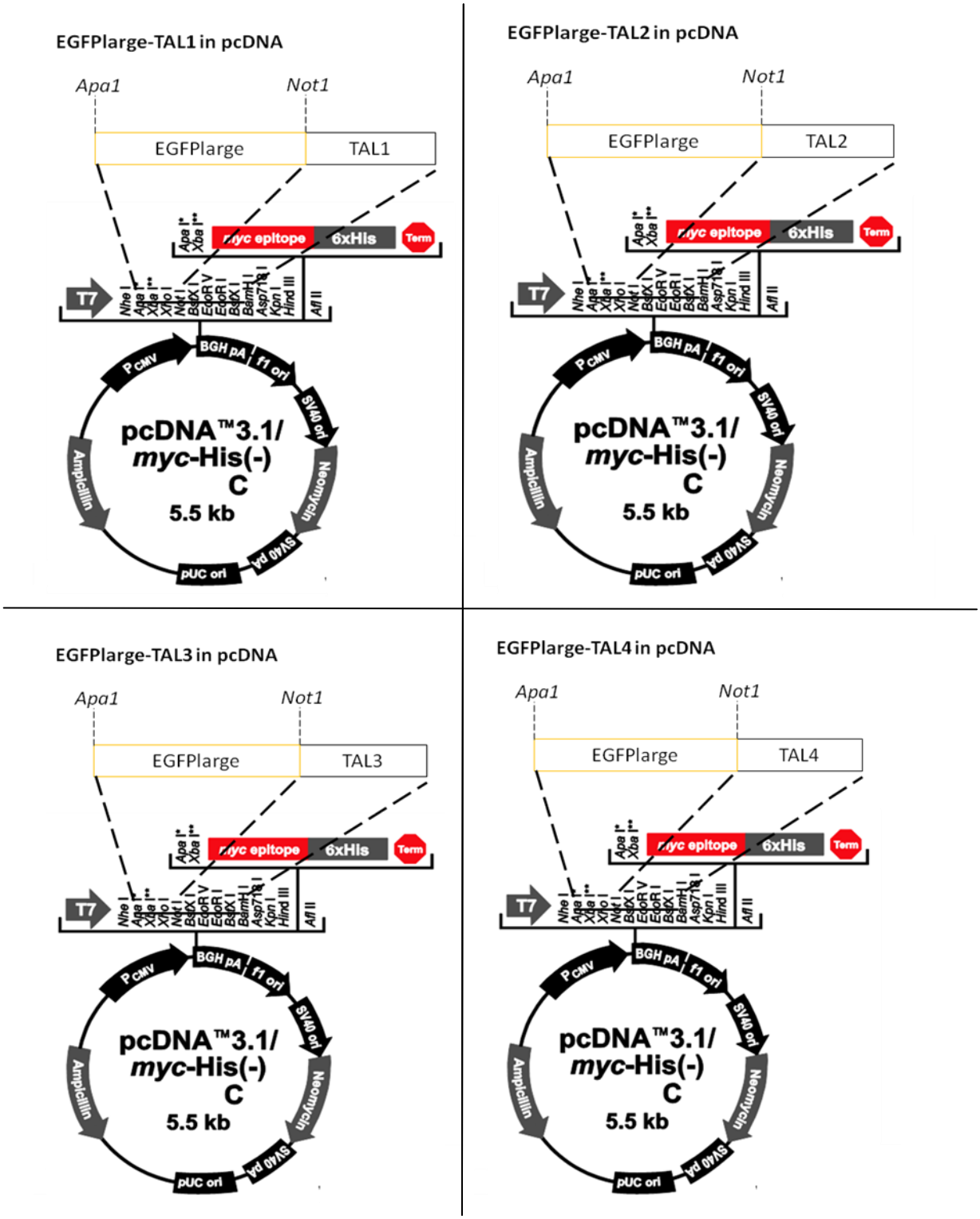
Plasmid used for cloning: A12-p2a-mCherry-B21
Insert: mCherry. Generated by PCR with primers SG7522 and SG7523 from the template ‘pmCherry C1’
Cloned in Kpn1 site.

TALs in-pcDNA



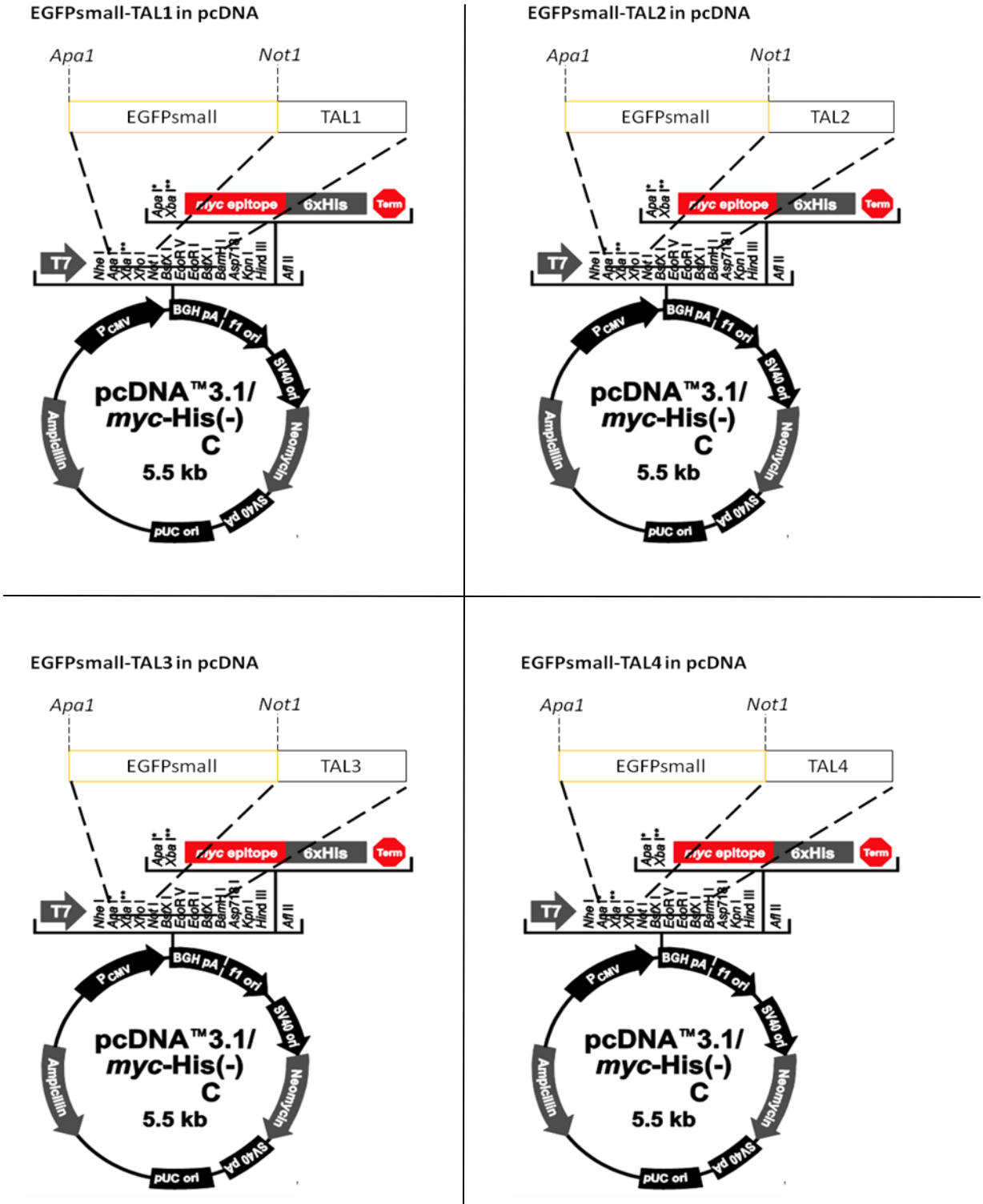
Plasmid used for cloning: **pcDNA3.1 myc-his (-) C**
Inserts (clockwise fron top left): TAL1, TAL2, TAL4 and TAL3. Obtained respectively by digesting 'TAL SB1_left','TAL SB1_right', 'TAL SB2_right' and 'TAL SB2_left' TALEN constructs from Life Technologies with Not1 and BamH1. 'Cut-Paste Cloning' Cloned in the Not1 and Bamh1 sites. Stop codon Absent.
*Note: The cloning is such that the C terminal myc/His tag is NOT in frame with the TAL gene. Consequently the protein product will not have this tag. The TAL gene has its own V5 tag at the N terminal. The start codon is included in the insert. Stop codon comes from the vector.
*Since no PCR step was involved in the cloning, the construct was not sequenced.

EGFPlarge-TAL in pcDNA



Plasmids used for cloning: (clockwise from top left) TAL1 in pcDNA, TAL2 in pcDNA, TAL4 in pcDNA, TAL3 in pcDNA.
Insert: EGFP large half. Generated by PCR with primers SG7683 and SG7513 from the template 'pEGFPN1'. Start codon included.
Cloned into the Apa1 and Not1 sites.

EGFPsmall-TAL in pcDNA



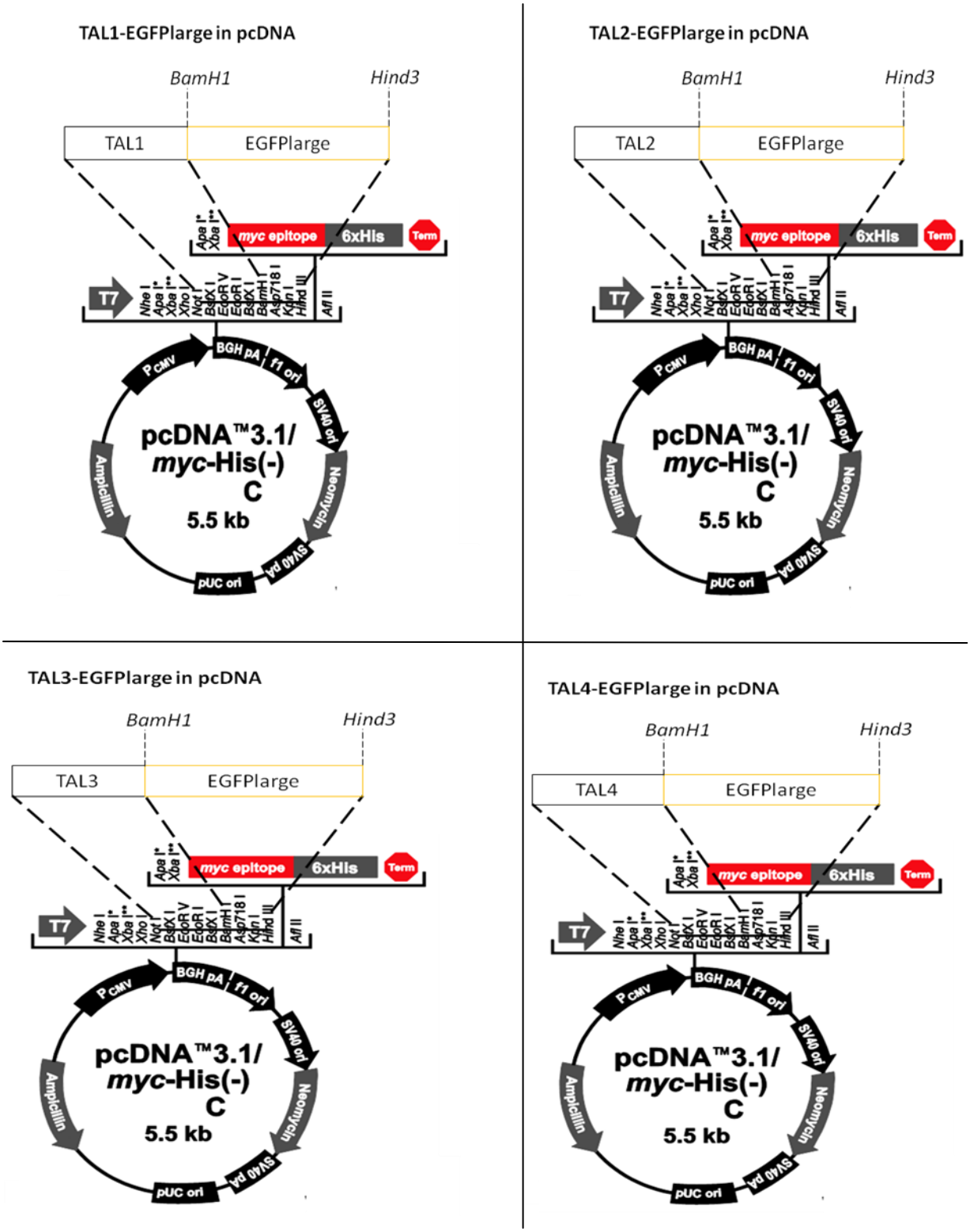
Plasmids used for cloning:

Plasmids used for cloning: (clockwise from top left) TAL1 in pcDNA, TAL2 in pcDNA, TAL4 in pcDNA, TAL3 in pcDNA.

Insert: EGFP small half. Generated by PCR with primers SG7684 and SG7517 from the template ‘pEGFPN1’. Start codon included.

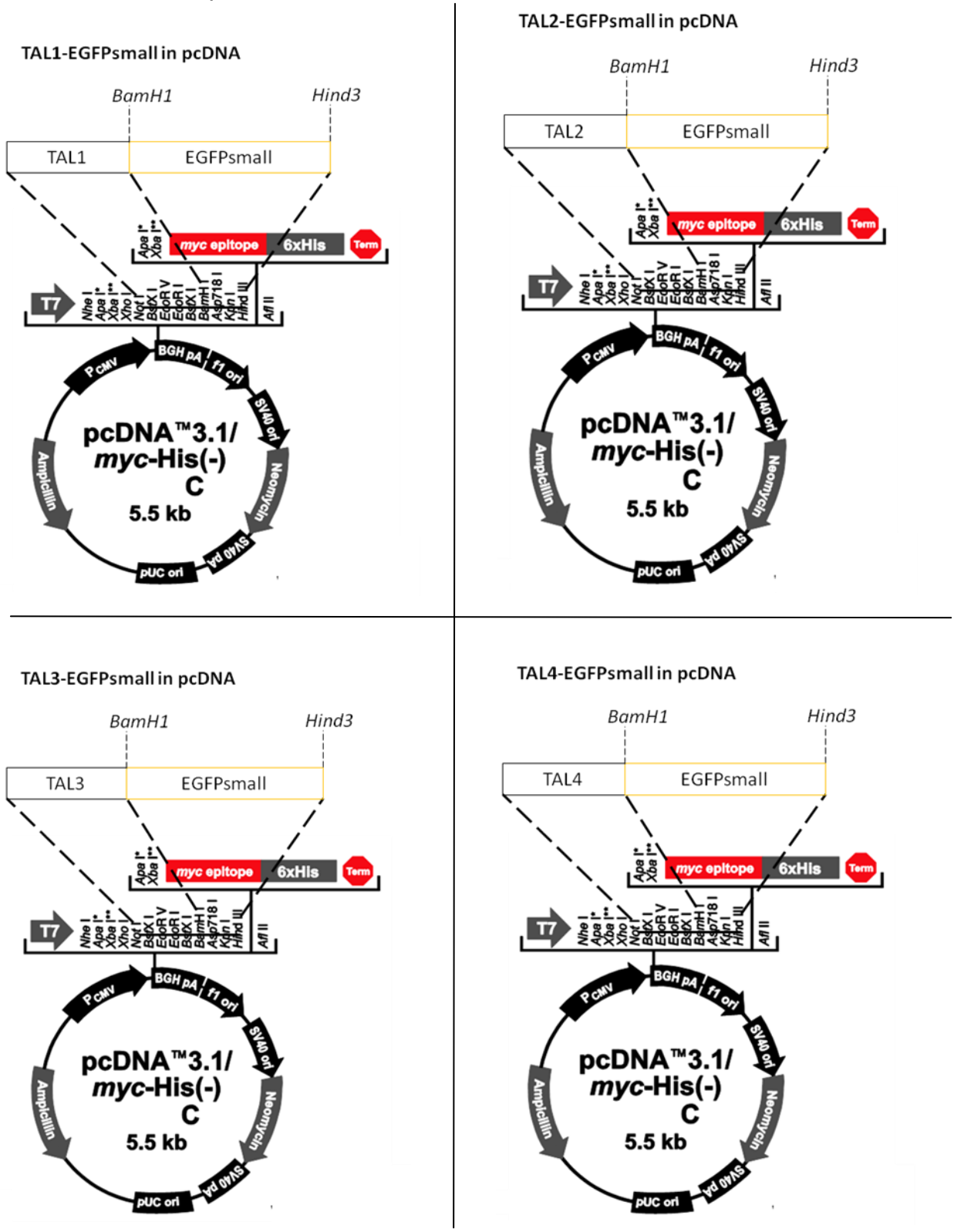
Cloned into the Apa1 and Not1 sites.

TAL-EGFPlarge in pcDNA



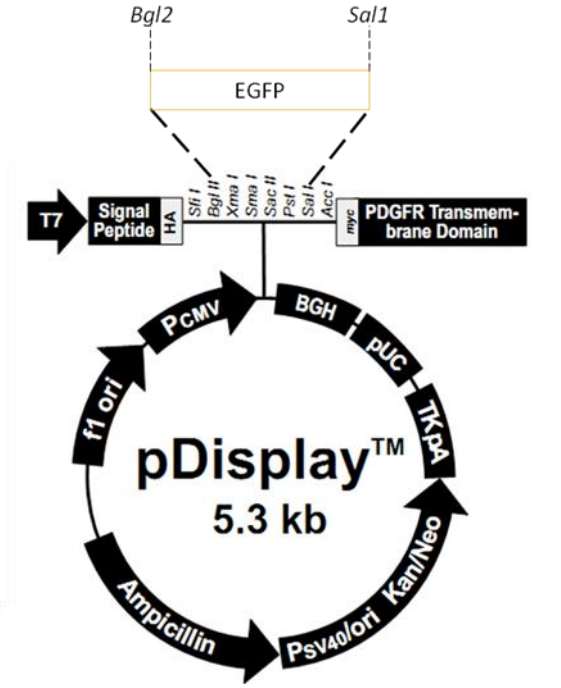
Plasmids used for cloning: (clockwise from top left) TAL1 in pcDNA, TAL2 in pcDNA, TAL4 in pcDNA, TAL3 in pcDNA.
Insert: EGFP large half. Generated by PCR with primers SG7514 and SG7515 from the template 'pEGFPN1'. Stop codon included.
Cloned into the BamH1 and Hind3 sites.

TAL-EGFPsmall in pcDNA



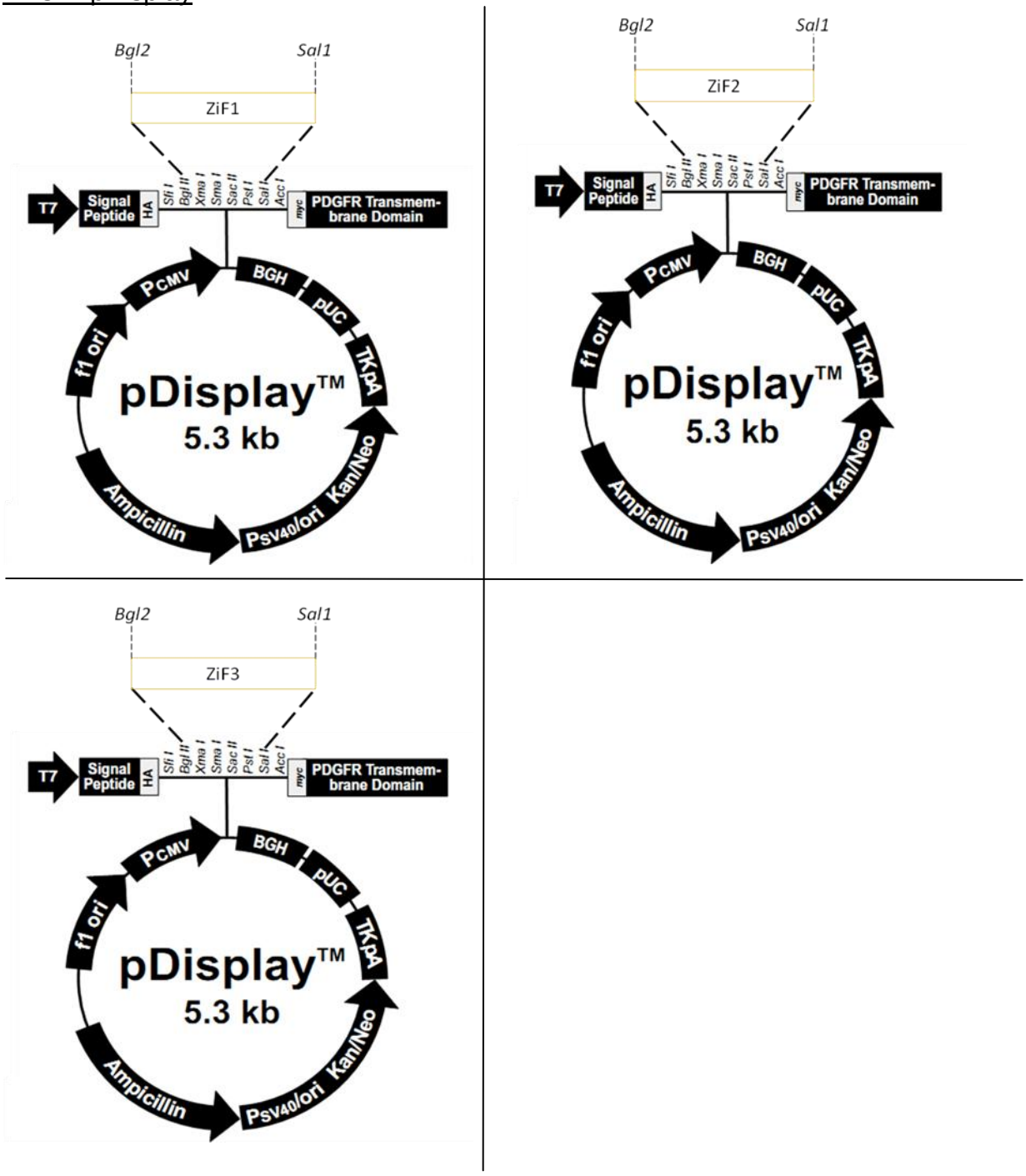
Plasmids used for cloning: (clockwise from top left) TAL1 in pcDNA, TAL2 in pcDNA, TAL4 in pcDNA, TAL3 in pcDNA.
Insert: EGFP small half. Generated by PCR with primers SG7518 and SG7519 from the template ‘pEGFPN1’. Stop codon included.
Cloned into the BamH1 and Hind3 sites.

EGFP in pDisplay



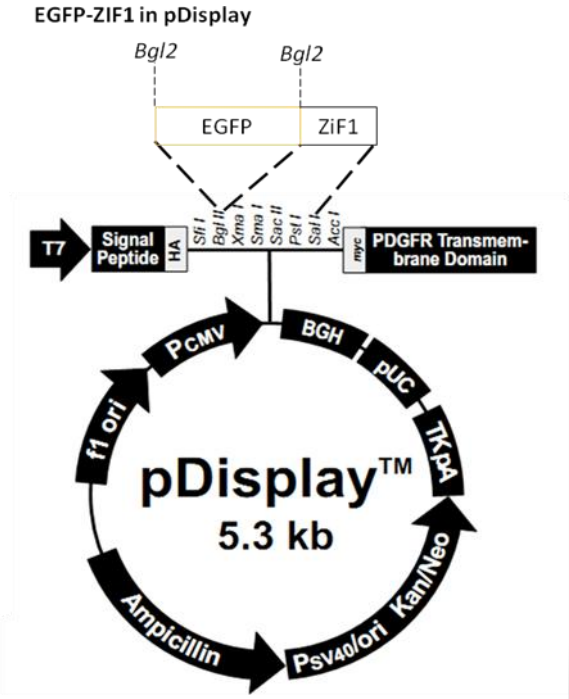
Plasmid used for cloning: pDisplay
Insert: EGFP. Generated by PCR with primers SG6231 and SG 6232 from the template ‘pEGFPN1’
Cloned in Bgl2 and Sal1 sites.
*Note: The Caspase-6 cleavage sequence ‘VEMD’ has been inserted at the C terminal end of the EGFP gene. This was done by incorporating the sequence in the reverse primer used for PCR.

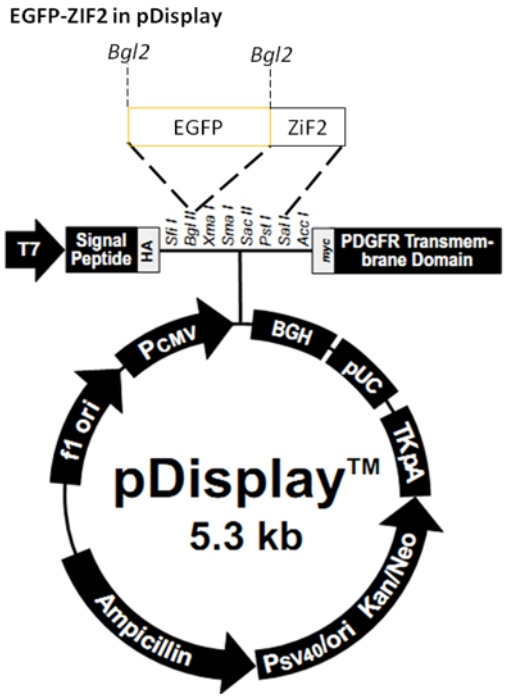
ZiFs in pDisplay

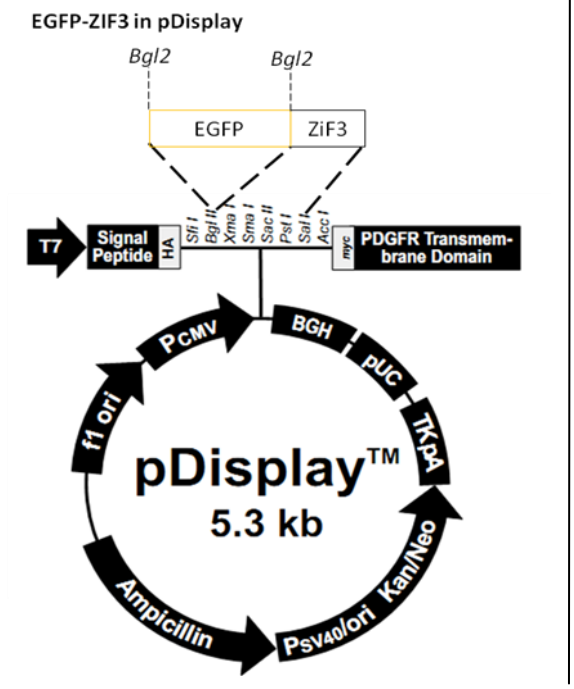


Plasmid used for cloning: pDisplay
Insert (Clockwise from top left): ZiF1, ZiF2 and ZiF3. Generated by PCR with primers SG6233 and SG 6234 from the template 'B21 in pC62', 'A12 in pC62' and B23 in pC62' respectively.
Cloned in Bgl2 and Sal1 sites.
Note: The start and stop codons are already present in the vector sequence.

EGFP-ZIF in pDisplay

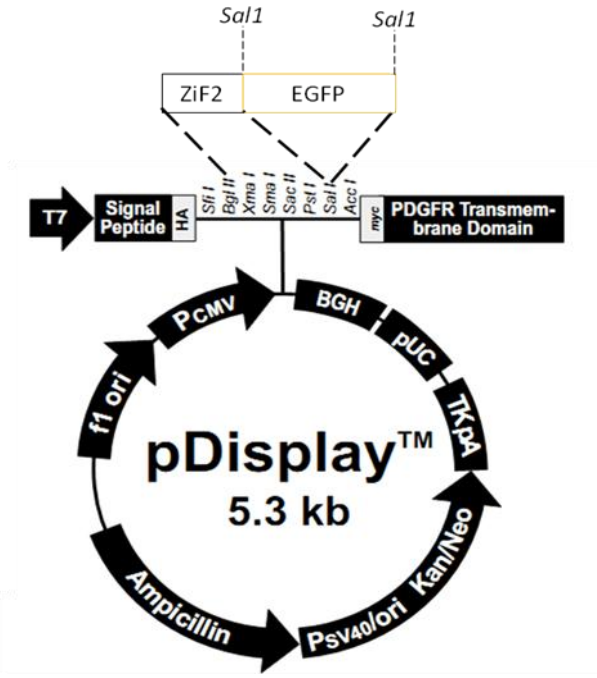






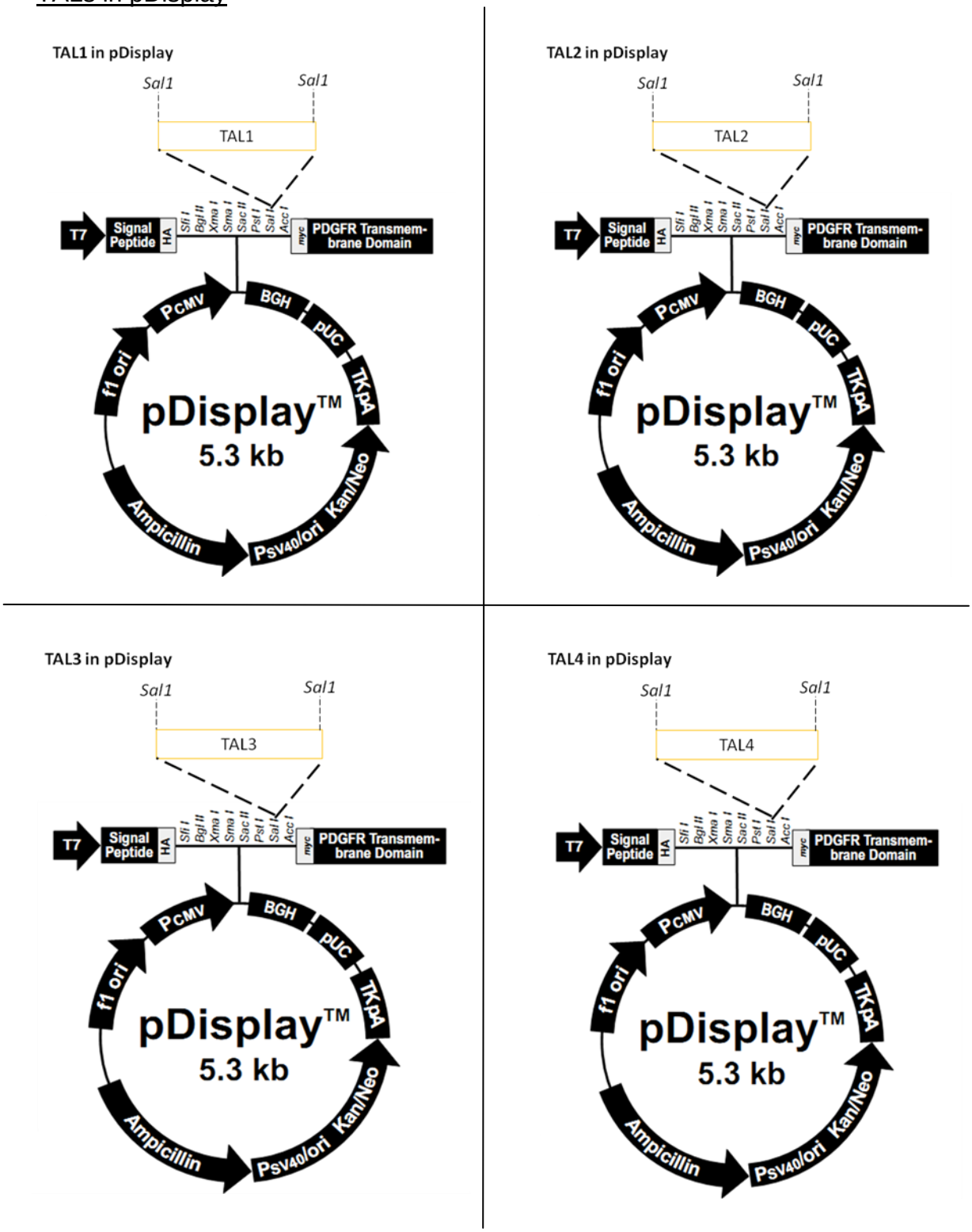
Plasmids used for cloning: (clockwise, from top left) ZiF1 in pDisplay, ZiF2 in pDisplay, ZiF3 in pDisplay.
Insert: EGFP. Obtained by PCR with primers SG6231 and SG7511 using the template 'pEGFPN1'.
Cloned in Bgl2 site.

ZiF2-EGFP in pDisplay



Plasmid used for cloning: ZiF2 in pDisplay
Insert: EGFP. Obtained by PCR with primers SG7512 and SG6232 using the template 'pEGFPN1'.
Cloned in Sal1 site.
*Note: The Caspase-6 cleavage sequence 'VEMD' has been inserted at the C terminal end of the EGFP gene. This was done by incorporating the sequence in the reverse primer used for PCR.

TALs in pDisplay



Plasmid used for cloning: pDisplay
Inserts: (clockwise from top left) TAL1, TAL2, TAL4, TAL3. Amplified by PCR from the templates 'TAL1 in pcDNA', 'TAL2 in pcDNA', 'TAL4 in pcDNA' and 'TAL3 in pcDNA' respectively with the primers SG7520 and SG7521
***Note:** The forward primer (SG 7520) is such that the initial 27 amino acids of the TAL gene are deleted (for clarity, compare the sequence to TAL in pcDNA). This

ensures removal of the NLS sequence of the TAL and the V5 tag. The NLS sequence can interfere with the surface targeting of the TAL-pDisplay construct.
*Note: The coding frame was not sequenced completely due to the length of the TAL gene. However, some of the constructs were validated by expression and Western blot.

Cell Culture and Transfections

HEK 293T cell line was used for all cell culture experiments. Some experiments were also performed in HEK293 cells which gave identical results. Cells were grown at 5%CO_2 in DMEM (Dulbecco's Modified Eagle Medium), with 10% FCS (Fetal Calf Serum) and Antibiotic. Transfections were carried out in either 12 well or 6 well plates. Cells were seeded in the plates so that they would be 60-80% confluent after 24 hours. Transfections were carried out 18-24 hours after seeding with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 2ug plasmid DNA was used for the 12 well plate and 5ug for the 6 well plate.

MG132 treatment

24 hours after transfection, media was replaced with complete media (without antibiotics) containing 10uM MG132 (Sigma). MG132 was used as a 10mM stock solution in DMSO. For control treatment, equivalent amount of pure DMSO was used in the media. Cells were harvested 12 hours later for analysis.

Western blotting

Cells were harvested 36-48 hours after transfection. If both RNA and Protein from the same sample were needed, 70%cells from the well were used for protein and rest 30%for RNA extraction. Protein Lysates were prepared in RIPA buffer. Lysates were cleared by centrifugation at 15000g for 30 minutes. The pellet obtained after centrifugation was resuspended in 50uL RIPA buffer.

The supernatant protein concentration was quantified by BCA kit. For Western blot, approx 30-40ug supernatant protein and about 25uL of resuspended pellet was loaded on the SDS-PAGE gel.

The proteins were transferred on a PVDF membrane for subsequent immunoblotting

Blocking solution used was 5%milk in TST. All Primary and Secondary Antibodies were prepared in 2.5% milk in TST.

Name	Company/Catalogue number	Dilution
GFP(FL) rabbit polyclonal	Santacruz, sc-8334	1:1000
DsRed2(65) mouse monoclonal	Santacruz, sc-101526	1:1000
HA-probe(Y-11) rabbit polyclonal	Santacruz, sc-805	1:1000
His-probe(H-3) mouse monoclonal	Santacruz, sc-8036	1:1000
Myc Tag, mouse monoclonal	ABM, G019	1:1000
V5 Tag, mouse monoclonal	ABM, G189	1:500 – 1:750

Table 6: List of Primary Antibodies and dilutions used

Secondary antibodies were used at a dilution of 1:10000

Protein expression in *E. coli*

GST tagged A12, B21 and B23 and Caspase 6 cleavage

The constructs A12 in pC6-2-73, B21in pC6-2-73 and B23 in pC6-2-73 were used for GST tagged recombinant protein expression in E.Coli XL1-Blue strain. The purification and Caspase 6 cleavage of GST tagged proteins from pC6-2 constructs has been described in detail by Purbey et al. and an identical protocol was followed in this case.

Protein Induction of A12-HIS and HIS-B23

The constructs A12-HIS in pTriex and HIS-B23 in pTriex were transformed in BL21-Rosetta strain. Protein induction was carried out in 5mL cultures at different temperatures for varying times (18^0C, 18hrs; 25^0C, 5 hrs and 37^0C, 2hrs) at 0.5mM IPTG concentration. 10uM $ZnSO_4$ was also used as an additive during induction. After induction, the bacterial pellet was resuspended in lysis buffer (100mM NaCl, 100mM Tris pH8) and sonicated till the lysate appeared clear. The supernatant was cleared by centrifugation. Both the supernatant and pellet fractions were loaded on an SDS-PAGE gel and stained with Coomassie to visualize protein.

References

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas C.F.III. 1998. Toward controlling gene expression at will: Specific regulation ofvthe erbB-2yHER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. PNAS *95*, 14628-14633.

Bodnar, A.M., Bernal, A., Szurek, B., and Lopez, C.E. 2013. Tell Me a Tale of TALEs. Mol Biotechnol. *53,* 228–235.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL Effectors: Customizable proteins for DNA Targeting. Science *333,* 1843-1846.

Burton, M.V., McCullough, R.M., Cantor, C.R., and Broude, N. 2007. RNA visualization in live bacterial cells using fluorescent protein complementation. Nature Methods *4,* 421-427.

Demidov, V.V., Dokholyan, N.V., Witte-Hoffmann, C., Chalasani, P., Yiu, H.W., Ding, F., Yu, Y., Cantor, C.R., and Broude, N.E. 2005. Fast complementation of split fluorescent protein triggered by DNA hybridization. PNAS *103*, 2052-2056.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. 2012. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science *335*, 720-723.

Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Hunber, A., Koksch, B., and Barbas, C.F. III. (2005). Development of Zinc Finger Domains for Recognition of the 5’-CNN-3’ Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. J. Biol. Chem. *280*, 35588-35597.

Drier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F.III. 2001. Development of Zinc Finger Domains for Recognition of the 5’-ANN-3’ Family of DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. J Biol Chem *276*, 29466-29478.

Huang, Y., and Bystroff, C. 2009. Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein. Biochemistry *48*, 929–940.

Kerppola, T. 2009. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment Complementation. Chem. Soc. Rev. *38*, 2876–2886.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., and Choi, S.Y. 2011. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. PLoS ONE *6*, e18556.

Negi, S., Imanishi, M., Matsumoto, M., and Sugiura, Y. (2008). New Redesigned Zinc Finger Proteins: Design Strategy and its Application. Chem. Eur. J. *14*, 3236–3249.

Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and Selection of Novel Cys_2His_2 Zinc Finger Proteins. Annu. Rev. Biochem. *70*, 313–40.

Purbey, P.K., Jayakumar, P.C., Patole, M.S., and Galande, S. 2006. pC6-2/caspase-6 system to purify glutathione-S-transferase-free recombinant fusion proteins expressed in Escherichia coli. Nature Protocols *1*, 1820-1827.

Seeman, N. 2010. Nanomaterials Based on DNA. Annu. Rev. Biochem. *79*, 65–87.

Segal D.J., Drier, B., Beerli, R., and Barbas C.F.III. 1999. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. PNAS *96,* 2758-2763.

Tsien, R.Y. 1998. THE GREEN FLUORESCENT PROTEIN. Annu. Rev. Biochem. *67*, 509–544.