

A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics

Thomas J. Magliery¹ and Lynne Regan^{1,2,3}

¹Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, CT 06520-8114 and ²Department of Chemistry, Yale University, New Haven, CT, USA

³To whom correspondence should be addressed.
E-mail lynne.regan@yale.edu

Combinatorial methodologies have revolutionized studies in biomolecular function, but they have so far proven less useful for understanding macromolecular structure and stability. This is largely because of the difficulty of screening libraries of molecules for biophysical properties, and the difficulty of interpreting structural effects in complicated molecules. Here, we report a novel, robust, cell-based screen for function of the four-helix bundle protein, Rop. By expression of green fluorescent protein from a ColE1 plasmid, the screen reports the copy number of the plasmid, which is modulated in *Escherichia coli* by Rop. We have engineered the screen so that the fluorescent phenotype can correspond to either Rop activity or lack thereof. We have used the screen to demonstrate with systematically constructed Rop core variants that not all molecules that bind small stem-loop RNAs *in vitro* are active *in vivo*. Rop is well understood from structural work and systematic mutations, which makes it possible to construct rational, targeted libraries. This screen makes it possible to rapidly interrogate such libraries effectively for proper protein folding and stability. In addition to its intended utility for combinatorial experiments in biophysics, the screen will allow further dissection of the mechanism of Rop-mediated plasmid copy number regulation *in vivo*.

Keywords: ColE1 origin/combinatorial/four-helix bundle/Rop/screen

Introduction

Many diseases are caused by mutations that reduce the stability or compromise the folding of key proteins. Examples include some of the pathologically relevant mutations of the tumor suppressor p53 (Bullock and Fersht, 2001), the F508 deletion in CFTR which results in cystic fibrosis (Boucher, 2002), and various amyloidogenic mutations in PrP that are associated with familial spongiform encephalopathies or increased susceptibility to transmissible forms (Prusiner, 1998). Also, many proteins of potential therapeutic importance are insufficiently stable for reliable administration (Bishop *et al.*, 2001). Both for such biomedical reasons and in the pursuit of a better fundamental understanding of protein stability, there have been many studies of the effects of mutations on protein stability. Nevertheless, we still have only a qualitative understanding of the effects of point mutation or more substantial core repacking (Richards, 1997). Certainly, there is as yet no computational model that can faithfully predict the

thermodynamic effects of mutation on protein stability (Guerois *et al.*, 2002).

Without doubt, part of the reason for our inability to predict the effects of mutation is that our notions of the underpinnings of protein stability come from a relatively small number of mutations to any particular protein architecture. (By ‘small number’, we mean relative to the size of sequence space, which is 20^N for N sites, and thus is vast.) Also, the effects of mutation on protein stability are obscured by the fact that most protein architectures are complicated, making it hard to understand the results of site-directed mutations. One particularly fruitful method for gleaning insight into the basis of protein stability has been the systematic modification of model proteins, such as lysozyme (Matthews, 1995), λ repressor (Sauer *et al.*, 1990), barnase (Fersht, 1993) and Rop (Munson *et al.*, 1996). These proteins have typically been chosen for the ease of preparation and crystallization (e.g. lysozyme) or the ability to select for binding even if further biophysical characterization is more difficult or protein architecture is relatively complex (e.g. λ repressor). In general, the scope of systematic studies is limited by the fact that it is not easy to screen large numbers of protein variants directly for biophysical properties.

It is sometimes possible to screen libraries of mutants of a protein for function, using visible phenotypes (e.g. chromogenicity), or metabolic or antibiotic resistance selections, that link the function of the protein mutant to a cell-based property (fluorescence, survival, etc.). We believe that a general solution to the problem of ‘combinatorial biophysics’ can be achieved by combining the use of a well behaved, simple model protein of regular architecture with a functional assay. Structural characterization and systematic mutations of a model protein can be used to guide the construction of targeted libraries of protein variants, and a screen can then interrogate those libraries for active molecules. Since one can know from the outset that one is mutating structurally important residues (e.g. hydrophobic core residues), one can effectively screen for structure or stability. Such approaches have been applied in elegant combinatorial experiments on λ repressor (Lim and Sauer, 1989, 1991), and have also recently been exploited in other contexts. For example, chorismate mutase catalytic activity has been used as a reporter of structural mutations (Taylor *et al.*, 2001), and phage-displayed IgG-binding domains of protein L (Gu *et al.*, 1995) and protein G (Distefano *et al.*, 2002) have been used to screen for structured variants of the parent proteins by binding to immobilized IgG.

Rop is an antiparallel, homodimeric four-helix bundle protein that has been studied extremely thoroughly: the crystal (Banner *et al.*, 1987) and solution structures (Eberle *et al.*, 1991) are known, and systematic mutations have been made to the RNA-binding (Predki *et al.*, 1995), hydrophobic core (Munson *et al.*, 1994a, 1996) and turn residues (Nagi and Regan, 1997; Nagi *et al.*, 1999). Moreover, Rop has an activity: the modulation of the copy number of ColE1 plasmids in

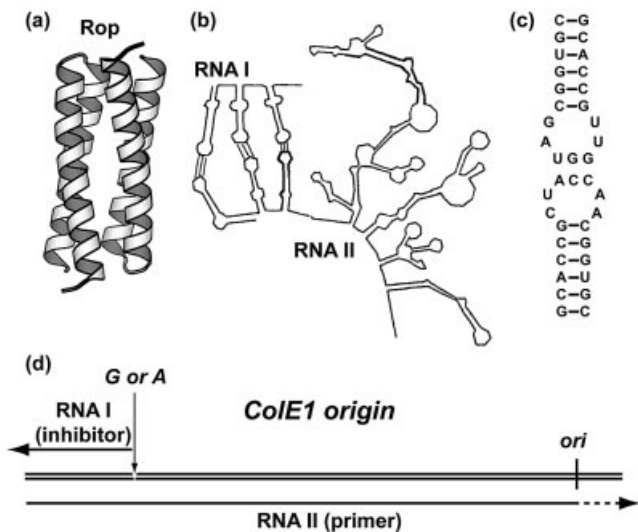


Fig. 1. The ColE1 origin. (a) X-ray crystal structure of Rop (Banner *et al.*, 1987) rendered in MOLSCRIPT (Kraulis, 1991) from 1ROP. (b) Schematic representation of the interaction between RNA I (108 nt) and RNA II (550 nt after processing) (Cesareni *et al.*, 1991). (c) Examples of small stem-loop RNAs derived from the complex in (b), used in gel-shift assays for Rop function (Predki *et al.*, 1995). (d) Organization of the ColE1 origin. RNA I and RNA II are divergently transcribed, allowing the inhibitor RNA to bind in a complementary fashion to the priming RNA. The position of the G/A polymorphism between pBR322 and pUC19 is noted.

Escherichia coli by facilitation of the binding of the inhibitory RNA I to the priming RNA II of the ColE1 origin (Figure 1) (Tomizawa and Som, 1984). The simplicity of the architecture and the depth of characterization make it possible to construct rational, targeted libraries of Rop variants. However, no simple *in vivo* screen for Rop function has been reported. Castagnoli *et al.* (1994) previously fused Rop to the DNA-binding domain of the λ repressor, creating a complex but selectable system wherein sufficient Rop homodimerization resulted in immunity to λ infection. However, there is no way to ‘tune’ this selection, and selections make it difficult to analyze the ‘negatives’, because the cells containing these proteins are dead. Also, Cesareni *et al.* (1982, 1984) fused β -galactosidase to the first 110 nt of RNA II, which results in diminution of β -galactosidase activity in response to Rop activity. However, this is a negative screen for Rop function, and it is not entirely clear how this activity is related to Rop’s ability to reduce plasmid copy number. An electrophoretic mobility shift assay for the binding of Rop to ‘kiss complexes’ composed of isolated stem-loops from RNA I and RNA II has been developed (Eguchi and Tomizawa, 1990, 1991; Gregorian and Crothers, 1995). Although this assay allowed some details of the protein–RNA interaction to be elucidated (Predki *et al.*, 1995; Munson *et al.*, 1996; Lee and Crothers, 1998), it is a decidedly low-throughput approach and an *in vitro* simplification of the actual binding problem that Rop faces *in vivo*.

Here, we demonstrate a novel cell-based screen for Rop function that makes it possible to screen large numbers of Rop variants quickly and reliably. The screen relies on modulation of the expression of a reporter molecule [here, green fluorescent protein (GFP)] as a result of modulation of the copy number of the ColE1 plasmid from which GFP is expressed. By altering the expression scheme for GFP, both ‘positive’ and ‘negative’ screens are possible, wherein either Rop activity or

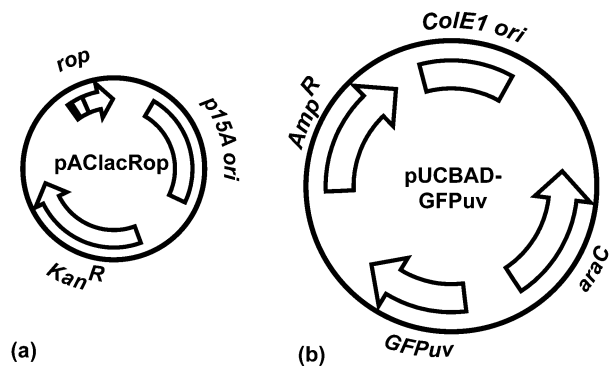


Fig. 2. Vector maps. (a) The pAClacRop plasmid expresses Rop from a synthetic *lac* promoter. The p15A origin and kanamycin resistance marker make it compatible with the pUC...GFPuv vectors. Rop is replaced with a short linker in pAClacLink and the gene for CAT in pAClacCm. (b) The pUCBADGFPuv plasmid expresses GFPuv (Cramer *et al.*, 1996) from the arabinose promoter. It bears the pUC19 version of the ColE1 origin and the ampicillin resistance marker. In pUCLacGFPuv, the arabinose promoter and *araC* gene are replaced by the *lac* promoter from pUC19.

inactivity can be reported by cellular fluorescence. By screening Rop variants that were previously subjected to gel-shift assay for activity *in vitro*, we show that binding to the small RNA stem-loops is not sufficient for activity *in vivo*. We also have engineered a Rop expression vector that makes it possible to readily clone libraries of Rop variants for screening. Currently, we are using this screen to interrogate libraries of Rop core variants in a rigorous, statistical manner, an approach that we believe will significantly expand the database of knowledge from which potential functions can be created for modeling protein stability (T.J.Magliery and L.Regan, manuscript in preparation).

Materials and methods

Construction of the reporter plasmids

Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase, Deep Vent DNA polymerase and Klenow DNA polymerase were acquired from New England Biolabs (Beverly, MA). DNA sequencing and deoxyribonucleotide synthesis were carried out by the Yale University W.M. Keck core facility. The vector pBAD-GFPuv (Cramer *et al.*, 1996) has the ColE1 origin which is also present in pBR322, in which the nucleotide immediately before the +1 position of RNA I is a guanine. In contrast, the ColE1 origin of pUC19 has an adenine in this position. The plasmid pUCBADGFPuv (Figure 2) was constructed by digesting both pBAD-GFPuv and pUC19 with *Alw*NI and *Bgl*II and ligating the smaller, origin-containing fragment from pUC19 into the pBAD-GFPuv vector. The presence of the –1 adenine was confirmed by sequencing. The plasmid pUCLacGFPuv was created by digesting pUCBADGFPuv with *Cla*I and *Nhe*I to excise the arabinose promoter and *araC* gene. The *lac* promoter from pUC19 was amplified by PCR with *Cla*I and *Nhe*I ends using the synthetic DNA oligonucleotides 5'-aataatcatgatcgcaagcgaattggtgtg-3' and 5'-aataatgctagccatagctgttctctgtgtaaattg-3'. The PCR product was digested appropriately and ligated into the digested pUCBADGFPuv vector.

Construction of the Rop screening plasmids

For ease of future library manipulations and to prevent complicating effects of autoregulation of Rop’s template, a

second compatible plasmid was engineered for Rop expression in the screen. The plasmid pACYC177lacRop was created by digesting pACYC177 with *Bam*HI and *Ban*I and ligating a fragment with a synthetic *lac* promoter and *rop* gene (created by overlap PCR and digested with the same restriction enzymes). The insert was created by Klenow extension of the deoxyribonucleotides 5'-aaggaaggatcctttacacattatgctccgctcgtatgtgtgtggaattgtgagcggataa-3' and 5'-ggtttttctgtttgtcatatgtgttctgtgtgaattgttatccgctcacaattccaca-3' and PCR amplification of *rop* from pBR322 using the deoxyribonucleotides 5'-catatgaccaaacaggaaaaaacc-3' and 5'-aaggaaggcactcagagg-tttaccgctc-3'. These two products were mixed and subjected to PCR amplification using the latter *rop* amplification primer and 5'-aaggaaggatcctttacac-3'. The vector used for screening, pAClacRop (Figure 2), was additionally modified to contain a synthetic T7 promoter before the *lac* promoter by insertion between the *Bst*EII and *Bam*HI sites of pACYC177lacrop. The inserted sequence was created by mixture and phosphorylation of the deoxyribonucleotides 5'-gtcacctaatagcactactatag-3' and 5'-gatcctatagtgagtcgtatttag-3'. The T7 promoter had no apparent effect on *lac*-driven Rop expression in DH10B *E.coli*; however, no expression of Rop was detected upon IPTG induction when pAClacrop was transformed into BL21(DE3) *E.coli*.

The negative control plasmid pAClacLink was created by digestion of pAClacRop with *Nde*I and *Ban*I and ligation of a phosphorylated fragment created by admixture of the deoxyribonucleotides 5'-tatggcatgacgtcgcgatg-3' and 5'-gcaccatg-cgacgtcatgccca-3'. This created an *Aat*II-cleavable linker in the Rop expression cassette. This linker was exchanged for the *Cm^R* gene from pACYC184 by PCR amplification of the gene for CAT using the deoxyribonucleotides 5'-aataataatcatatgagaaaaaatcactggatatacc-3' and 5'-aataatggcaccctacgccccgcctgccactcatcg-3', creating the vector pAClacCm.

Plasmids for expression of Rop variants Ala₂Leu₂-4 (Rop23), Ala₂Leu₂-8-rev (Rop29), Ala₂Met₂-8 (Rop39) and Leu₄-8 (Rop47) were created by PCR amplification of the appropriate genes from plasmids created by Munson (Munson *et al.*, 1994a,b, 1996) using deoxyribonucleotides that encode *Nde*I and *Ban*I-cleavable ends. These deoxyribonucleotides vary slightly with the Rop sequence, but, for example, those used to amplify Rop23 are 5'-aataataatcatatggggaccaaaccagga-gaaaacc-3' and 5'-aataatggcaccctcagaggtttaccgctc-3'. These digested PCR products were ligated into appropriately digested pAClacLink. The Ala₂Leu₂-2 variant was created synthetically by Klenow extension of the deoxyribonucleotides 5'-accaaa-caggaaaaaccgccttaacatggccccgctttc-tgagaagccaggcgtaacgcttctggagaaactcaacgagctggacgcggatg-3' and 5'-accgctcaccgaaacgcgcgaggcagctgcggtaaagctcatccg-cgtggtcgtgcagcgattcacagatatctgctgttcatccgctccagctcg-3' followed by PCR amplification with 5'-cacaggaacacatgatgaccaaaccaggaaaaaacc-3' and 5'-cagtgaggcactcagaggtttaccgctcaccgaaacc-3'. The resulting product was digested and ligated as with the other Rop variants.

Screening Rop variants

Positive and negative screening strains were created by transforming electrocompetent DH10B *E.coli* with pUCBADGFPuv or pUClacGFPuv, respectively. DH10B is a robust cloning strain from which high-efficiency electrocompetent cells can easily be prepared by standard methods. Electrocompetent cells of these strains were then transformed with the appropriate pAClac plasmid (e.g. pAClacRop for

positive control or pAClacLink for negative control). Cells were plated on LB agar supplemented with 100 µg/ml ampicillin, 35 µg/ml kanamycin, 100 µM IPTG and 0.0005% arabinose and grown for 16 h at 42°C. (The IPTG and arabinose can optionally be omitted with the pUClacGFPuv-based negative screening strain.)

To examine the copy number of the ColE1 plasmids directly, a 5 ml culture of the doubly transformed *E.coli* of interest was grown overnight at 37 or 42°C in LB/Amp/Kan and 1 ml of that culture was subjected to alkaline lysis miniprep (Qiagen). The isolated DNA was digested with *Aat*II and *Xma*I (which makes a single cut in the pUC...GFPuv vectors, a single cut in the pAClacRop vector and two cuts in the pAClacLink vector) and analyzed by agarose gel electrophoresis.

Plasmid sequences

The sequences of the pUC...GFPuv and pAClac... vectors will be available on the Regan Group Website Publications page (<http://www.csb.yale.edu/people/regan/publications.html>).

Results

Principle of the screen

Rop modulates the copy number of ColE1 plasmids by facilitating the binding of the inhibitory RNA I to the priming RNA II of the ColE1 origin. Thus, RNA II is prevented from functioning as a primer for DNA replication, and plasmid copy number is reduced (Figure 1) (Cesareni *et al.*, 1991). We hypothesized that a cell-based screen for Rop function could be created by expressing GFP from a ColE1 plasmid. In the presence of active Rop, the copy number of ColE1 plasmid diminishes, and the amount of GFP (and hence cellular fluorescence) is expected to decrease (Figure 3a). The vector pUClacGFPuv contains the pUC19 version of the ColE1 origin (which has a point mutation in RNA II), the gene for ampicillin resistance, and an expression cassette for a GFP variant under the control of the *lac* promoter (see Figure 2 for vector schematics). The pUC19 variant of the ColE1 origin was effective for our screen, as opposed to the version found in plasmids such as pBR322, because the pUC19 origin confers a much larger difference in copy number with and without Rop, especially at 42°C (data not shown). This RNA II point mutation was previously known to confer an extremely high copy number at 42°C in the absence of Rop (Lin-Chao *et al.*, 1992; Lahijani *et al.*, 1996).

We also constructed the plasmids pAClacRop, pAClacLink and pAClacCm, which express, respectively, the gene for wild-type Rop (wt Rop) (positive control), a short linker (negative control) and the gene for chloramphenicol resistance (a long linker for ease of cloning libraries of Rop variants). These plasmids all contain an expression cassette under the control of a synthetic *lac* promoter, the gene for kanamycin resistance and the p15A origin from pACYC177, which is not regulated by Rop and allows for co-maintenance with ColE1 plasmids.

A screening strain was created by transforming pUClacGFPuv into DH10B *E.coli*, and competent cells of this strain were transformed with pAClacRop or pAClacLink. On solid LB medium, cells bearing pAClacLink (i.e. no Rop) were found to be strongly fluorescent, but those bearing pAClacRop (i.e. wt Rop) were only very weakly fluorescent, after 16 h at 42°C (Figure 3b). Thus, cellular fluorescence corresponds to the lack of active Rop under these conditions.

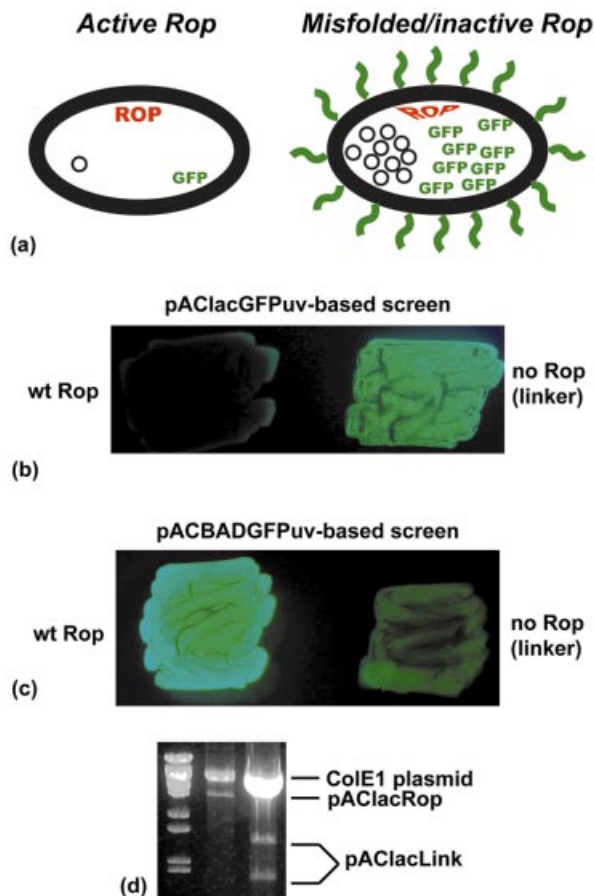


Fig. 3. Screening for Rop function. (a) The principle of the negative screen is that active Rop lowers the copy number of the ColE1 plasmid from which GFP is expressed, reducing cellular fluorescence. (b) Negative screen. DH10B *E.coli* transformed with pUClacGFPuv and pAClacRop (left) or pAClacLink (right) were grown for 16 h at 42°C on LB agar supplemented with kanamycin and ampicillin. Cells are visualized under long-wave UV light. (c) Positive screen. DH10B *E.coli* transformed with pUCBADGFPuv and pAClacRop (left) or pAClacLink (right) were grown for 16 h at 42°C on LB agar supplemented with 0.0005% arabinose, 100 μ M IPTG, kanamycin and ampicillin. Cells are visualized under long-wave UV light. (d) Plasmid minipreps from the positive screen. A 1% agarose gel of *Xma*I/*Aat*II-cleaved DNA minipreps from 200 μ l of liquid culture grown in LB supplemented as with the plates above. Lane 1, λ BstEII marker (NEB); lane 2, pAClacRop/pUCBADGFPuv DH10B (bright cells); lane 3, pAClacLink/pUCBADGFPuv DH10B (dim cells). The double scission of the smaller plasmid in lane 3 confirms that it is pAClacLink.

Positive screening with P_{BAD} -GFP on a ColE1 plasmid

A drawback to the version of the screen described above is that cells bearing active Rop are dim, while those lacking Rop activity are fluorescent (i.e. it is a negative screen). However, one can imagine trivial reasons that might lead to diminution of fluorescence in a library experiment (such as spontaneous mutations of the GFP), and it would generally be convenient to have a screen in which brightly fluorescent cells correspond to Rop activity.

We found that this was possible under certain conditions by expressing GFPuv under the control of the arabinose promoter (P_{BAD}) on the ColE1 plasmid. The plasmid pUCBADGFPuv is identical to pUClacGFPuv, except that the *lac* promoter was replaced with the arabinose promoter, including the *araC* gene, which is divergently transcribed from the GFPuv gene. Again,

a screening strain (the ‘positive screening strain’) was created by transforming DH10B *E.coli* with pUCBADGFPuv, and competent cells of this strain were transformed with pAClacRop or pAClacLink. Surprisingly, on solid LB medium containing 0.0005% arabinose and 100 μ M IPTG, cells grown at 42°C overnight were fluorescent when bearing wt Rop, but dim when bearing no Rop (Figure 3c). Thus, cellular fluorescence in this format of the screen corresponds to the presence of active Rop. At lower concentrations of arabinose, none of the cells were fluorescent; at higher concentrations, all of the cells were fluorescent (data not shown).

To confirm that the fluorescence phenotypes were the result of changes in plasmid copy number, we extracted and roughly quantified the plasmid DNA in the cells. Active Rop visibly reduces the amount of ColE1 plasmid that is isolated in alkaline lysis miniprep (Figure 3d). We examined cells grown in solution under conditions identical to those on agar plates, which verified that the bright, Rop-containing cells contained less of the pUCBADGFPuv ColE1 plasmid. In addition, SDS-PAGE analysis of lysates of these same cells showed that the bright, Rop-containing cells contain more GFP than the dim, linker-containing cells, and that it is mostly in the soluble fraction (data not shown).

We believe that this phenotypic reversal is the result of the mechanism of the arabinose promoter, wherein the AraC regulator acts as a repressor in the absence of arabinose but as an activator in its presence (Schleif, 2000). We hypothesize that in the absence of Rop, the high copy number of the ColE1 plasmid requires more arabinose for AraC-mediated activation of GFP expression, rendering the cells dim under these conditions. Regardless of the exact details of the mechanism, this system allows us to screen for active variants of Rop by looking for brightly fluorescent cells under these conditions.

Assay of systematically repacked Rop variants

Previously in this laboratory, Munson *et al.* (1994a, 1996) examined the gross structural effects of repacking the hydrophobic core of Rop with combinations of Ala, Leu and Met. Briefly, the hydrophobic core of Rop can be considered to contain eight layers composed of two amino acids from each monomer. Owing to the symmetry of the homodimer, a change in one layer concomitantly alters the symmetrically associated layer at the other end of the four-helix bundle. Thus, engineering the fourth layer to contain Ala₂Leu₂ simultaneously causes the fifth layer to be Ala₂Leu₂. This arrangement is called Ala₂Leu₂-2. Moreover, the core is generally composed of alternating ‘small’ and ‘large’ residues, except for in the penultimate layers at each end, which are said to be reversed. Rop variants in which the next-to-last layers are reversed are denoted ‘rev’ (Figure 4).

We examined five of these Rop variants for Rop activity by cloning the variants into the pAClacLink vector, transforming into the positive selection strain and screening as described above. Three of the variants, Ala₂Leu₂-2, Ala₂Leu₂-4 and Ala₂Leu₂-8-rev, were previously found (Munson *et al.*, 1996) to have predominantly dimeric structure, CD spectra similar to wt Rop, higher T_m values than wt Rop and the ability to bind to small RNA hairpins similar to loops in RNA I and RNA II with affinities approaching that of wt Rop (as judged by gel-shift assay). However, by screen or examination of copy number by plasmid prep, the Ala₂Leu₂-8-rev variant is totally inactive, the Ala₂Leu₂-4 variant may be slightly active, and the Ala₂Leu₂-2 variant is strongly active. As expected, the Ala₂Met₂-8 variant

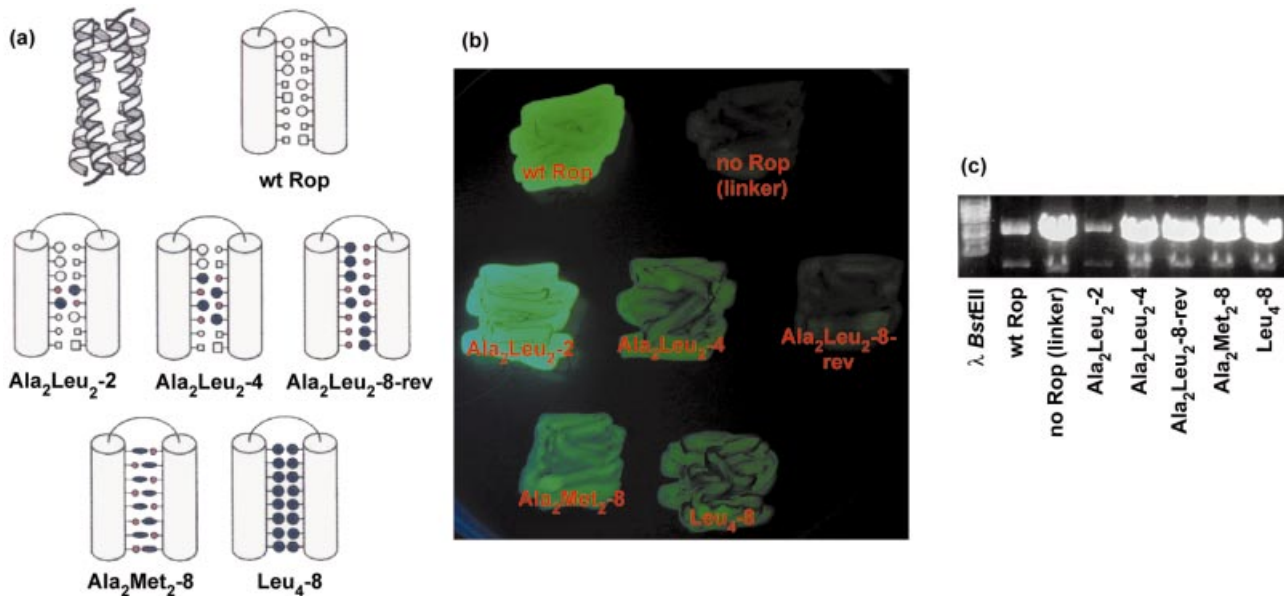


Fig. 4. Systematically repacked Rop variants. (a) Schematic representation of the five repacked Rop variants previously designed by Munson *et al.* (1994a, 1996). Large closed circles represent Leu, small closed circles represent Ala, and closed ovals represent Met. The nomenclature represents the composition of each layer of the core (e.g. Ala₂Leu₂), the number of layers that are repacked (i.e. 2, 4 or 8), and whether or not the residue size alternation is maintained or reversed (i.e. rev) in the penultimate layers. (b) Positive screen (growth and visualization as in Figure 3) of the repacked Rop variants. Owing to the position of the UV lamp in the lower-left corner of the plate (necessary for photography), the Ala₂Met₂-8 variant appears brighter than it does when the lamp is directly above the plate. (c) DNA minipreps of the same cells cleaved with *Xma*I. Only Ala₂Leu₂-2 has strong activity, as evidenced by cellular fluorescence or ColE1 plasmid amount in the miniprep. In contrast, Munson *et al.* found that all three of these Ala₂Leu₂-X variants bound model stem-loop RNAs with dissociation constants comparable with wt Rop.

(which was found to be weakly dimeric, to have a CD spectrum similar to wt Rop, to have a depressed T_m , and was not observed to bind RNA) and the Leu₄-8 variant (apparently tetrameric, CD spectrum indicative of high helical content, non-cooperative melting transition, no RNA binding) were inactive in the cell-based assay (Figure 4).

Evidently, *in vivo* Rop function requires more than just the ability to bind the small stem-loop hairpins derived from the much larger RNA I and RNA II molecules. Munson *et al.* (1996) concluded that those Rop variants that bound the RNA hairpins were the most Rop-like in structure. However, neither the solution nor the crystal structures of these variants was ascertained, and it is therefore not known how Rop-like the variants are at atomic resolution. Also, the repacked Rop variants exhibited extremely different folding kinetics for dimerization, and this may be a factor contributing to the difference between *in vivo* activity and *in vitro* binding. Apparently, merely maintaining the presence of the RNA binding residues in a Rop variant that forms a stable, four-helix bundle is insufficient for activity. Consequently, those Rop variants that pass the screen are likely to be extremely Rop-like in structure and thermodynamic properties, possessing not only the overall Rop fold but also atomic-level details that are required for function.

Preparing for libraries: the pAClacCm vector

In order to interrogate libraries of Rop variants with the screen, one must efficiently clone genes for those variants into the pAClac vector between *Nde*I and *Ban*I sites. However, we found that *Nde*I cuts the pAClacLink vector very inefficiently, requiring in excess of 400-fold more enzyme than anticipated (Figure 5). Since the linker in pAClacLink is very small, it is impossible to resolve singly and doubly cut vector on an

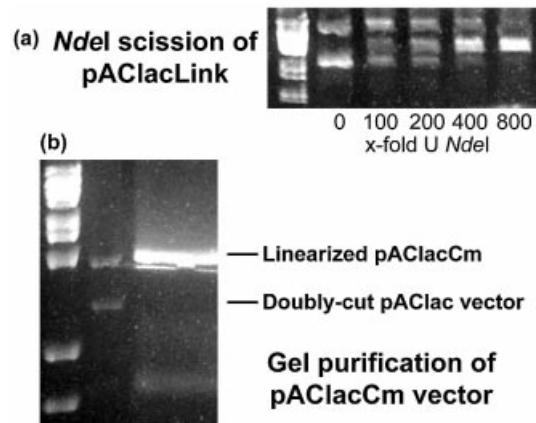


Fig. 5. Poor *Nde*I scission of pAClac and gel separation of pAClacCm/*Nde*I/*Ban*I scission products. (a) Scission of 50 ng of pAClacLink plasmid for 1 h at 37°C with *Nde*I. Lane 1, λ BstEII marker; lanes 2–6, scission with 0, 5, 10, 20 and 40 U of *Nde*I. The ratio of units of enzyme to amount of DNA is shown below each lane (typically, 1× is sufficient under these conditions). The plasmid preparation, which contains slow-running nicked DNA and fast-running supercoiled DNA, can be seen to resolve to a single, intermediately running linearized species after 400–800-fold scission. A normal amount of *Ban*I is sufficient for complete linearization (data not shown). (b) Agarose gel of *Nde*I/*Ban*I digested pAClacCm plasmid. Lane 1, λ BstEII marker; lane 2, singly (slow) and doubly cut (fast) material can easily be resolved on a 1% agarose gel; lane 3, a portion of a preparative-scale lane from which the doubly cut material was removed by electrophoresis onto a dialysis membrane.

agarose gel, and this leads to high background upon ligation and transformation. To combat this problem, we cloned the *Cm^R* gene for chloramphenicol acetyltransferase (CAT, from pACYC184) under the control of the *lac* promoter in

pAClacLink. Since the gene for CAT is ~650 bp, it is possible to separate singly and doubly cut vector (Figure 5), and the background rate of recircularization is <15% using calf intestinal alkaline phosphatase treatment and gel purification (data not shown).

This vector also allows us to get some idea of the expression level of a gene expressed from the pAClac vector under the synthetic *lac* promoter. The plasmid pACYC184, which has a p15A origin and CAT gene under the control of its native promoter from Tn9, allows survival of bacteria to between 35 and 70 µg/ml chloramphenicol. However, pAClacCm confers resistance to only 2–4 µg/ml chloramphenicol. Thus, the synthetic *lac* promoter is weaker than the Tn9 CAT promoter in pACYC184.

Discussion

How the screen works

The principle of this cell-based screen for Rop function is the relatively simple idea that increasing plasmid copy number would increase the expression of a gene transcribed from that plasmid. It is worth noting that this implies, at least with these promoters and in the relevant copy number regime, that the amount of gene expression is related to the amount of DNA template. This seems to contrast with Rop expression from another plasmid created in this laboratory, pMR101, a low copy p15A plasmid used to over-express Rop from the T7lac promoter (Munson *et al.*, 1994b). The T7lac promoter is a very strong promoter used to over-express proteins from plasmids of low to high copy number. The exceptionally high level of Rop expression from pMR101 implies that the low copy number of the template is not limiting, and it may be that there is only a relationship between template copy number and expression level with relatively weak promoters.

In fact, the cellular fluorescence is only related to the copy number in the positive (P_{BAD} -GFPuv) screen in a very narrow range of arabinose concentrations (near 0.0005%). At lower concentrations of arabinose the cells are dim, and at higher concentrations of arabinose, the cells are fluorescent, regardless of Rop activity. (By ‘cells’, we mean colonies or bulk culture; we did not examine individual cells microscopically.) More remarkable, of course, is that at this highly tuned arabinose concentration, cells with less ColE1 plasmid (i.e. active Rop resulting in less DNA template) make more GFP and are more fluorescent. This is especially peculiar in light of the Hu group’s recent observation that the relationship between arabinose concentration and P_{BAD} expression is probably determined by how many cells express protein, not how much each one expresses (Siegele and Hu, 1997). Since the AraC level also presumably changes in response to Rop activity (it is expressed from the ColE1 plasmid), this implies that the probability of switching from repression to activation is controlled both by the concentration of arabinose and the concentration of AraC. Regardless, the effect is fortuitous: using the *lac* and P_{BAD} promoters, we can screen for Rop activity either negatively or positively.

The dynamic range of the screen is limited by the difference in copy number of the ColE1 plasmid in the presence of wt Rop versus an inactive variant. Thus, maximizing this difference is critical to achieving the best level of signal. This required both using the pUC19 version of the ColE1 origin (instead of the pBR322 version) and screening at 42°C. The pUC19 origin differs from the pBR322 origin at a single nucleotide in RNA

II, immediately before the RNA I inhibitor template, resulting in a higher copy number, and a much higher copy number at 42°C. Interestingly, Rop largely ameliorates the loss of plasmid copy number control at both 37 and 42°C, resulting in a larger copy number difference at the higher temperature. It is conceivable that a small library of randomized nucleotides near the beginning of the RNA I template might yield a ColE1 variant with even more desirable properties, and indeed it is known that other mutations in this region alter ColE1 copy number (Fitzwater *et al.*, 1988).

The screen as an assay for Rop function

This *in vivo* screen has some advantages over the gel-shift assay that has recently been used to interrogate Rop function. It is technically simpler, since one need not purify the Rop variant or run a gel to know the activity. Moreover, one can directly relate the fluorescence level to plasmid copy number by extraction of the plasmid DNA from the cells used in the screen. Perhaps most importantly, the gel shift assay uses only small RNA stem-loops, while the screen involves the actual RNA substrates and demands functionality rather than merely binding. The result of this is clear: it is possible to find Rop variants with a Rop-like fold, that bind the small RNA stem-loops, that are nevertheless not active. This suggests that Rop function may indeed be more complicated than merely binding stem-loop RNAs, perhaps involving the observed ability to aid in the conversion of an unstable RNA I:RNA II complex into a more stable one (Tomizawa, 1990). While the model stem-loop complexes have been useful for elucidating structural details of Rop:RNA complexes, this screen may be useful in better elucidating Rop’s molecular mechanism. It should be noted that the screen is a *de facto* screen for Rop expression, as well, and it is conceivable that a negative phenotype from some Rop variants might be as a result of the lack of protein expression. However, all of the variants used in this study are produced in large quantities under conditions of T7 over-expression (in other expression plasmids), and we have no reason to believe that lack of expression is a problem here.

The gel-shift assay does offer the ability, at least in principle, to measure the dissociation constant of the Rop:RNA interaction. Interestingly, we have preliminarily observed that Rop variants with randomized core residues produce intermediate levels of fluorescence in some cases (T.J.Magliery, unpublished observations). It would be very interesting to investigate the relationship between the fluorescence level and parameters like the K_D with respect to RNA binding or Rop dimerization.

Screening a library: technical considerations

When one is interested in the phenotype conferred by all of the molecules that are transformed from a particular cloning reaction, the efficiency of each step becomes a serious issue. With selections, ‘background’ (or ‘noise’ from erroneous cloning) is eliminated if the phenotype is null; with a screen, however, such background clones hamper our ability to derive useful information from screen ‘negatives’. Here, for example, extremely poor *NdeI* scission of the Rop expression vector required a means of physically separating singly cut vector from vector with the whole *NdeI*/*BanI* linker excised. The introduction of the Cm^R gene both allows for successful isolation of the empty vector and allows us to get a rough idea of the Rop expression level in the pAClac construct.

Rop is probably expressed at exceedingly low levels in its wild-type context in pBR322. Its cryptic promoter is likely to be quite weak, and its initiation codon is GUG (instead of the

usual AUG), which is seen in only 14% of ORFs in the *E. coli* genome and requires decoding by tRNA^{fMet}'s CAU anticodon (i.e. a G:U wobble-pair is required in the first codon position). Here, we have replaced the GUG start codon with an AUG codon, but the synthetic *lac* promoter nonetheless results in a very small amount of protein. One of the original reasons chloramphenicol resistance was engineered into vectors was because, unlike β -lactamase, CAT is produced in quantities too low to observe on a gel in whole lysate (Martinez *et al.*, 1988). This level of expression from pACYC184, for example, corresponds to a MIC of between 35 and 70 $\mu\text{g/ml}$ for chloramphenicol. The MIC associated with pAClacCm is at least an order of magnitude lower than this, $\sim 2\text{--}4 \mu\text{g/ml}$, suggesting that the Rop variants are expressed at extremely low levels from pAClac. This is of some significance, since Rop presumably must act as a homodimer. If the monomer is present at vanishingly low levels in cytosol, which has a very high overall protein concentration, the screen therefore demands a relatively strong and specific interaction.

A new tool for combinatorial experiments in biophysics

We have demonstrated here the utility of a novel cell-based screen for the function of a relatively simple and well understood four-helix bundle protein. In combination with our ability to generate interesting libraries of Rop variants, this new tool gives us the ability to effectively sort extremely Rop-like, stable molecules from unstructured or unstable molecules. Moreover, Rop is easy to purify and examine *in vitro* by biophysical methods like CD, NMR and X-ray crystallography. We have begun screening libraries in which the two central layers of the Rop core are completely randomized, and we have found that there are multiple patterns that lead to stable proteins, including those deduced from systematic work and other novel modes (T.J.Magliery and L.Regan, manuscript in preparation). Using multiple libraries and screening technologies, we will be able to generate statistical models that will give further insight into the fundamental basis of protein stability and thereby allow the creation of more general and accurate potential functions for the computational analysis of protein variants.

Acknowledgements

The authors thank Jon Ness and Pim Stemmer (Maxygen) and Neal Cariello (GlaxoSmithKline) for supplying the pBAD-GFPuv vector. T.J.M. is an NIH Postdoctoral Fellow (GM065750-02). This work was supported by a grant to L.R. from the NIH (GM49146-09).

References

- Banner,D.W., Kokkinidis,M. and Tsernoglou,D. (1987) *J. Mol. Biol.*, **196**, 657–675.
- Bishop,B., Koay,D.C., Sartorelli,A.C. and Regan,L. (2001) *J. Biol. Chem.*, **276**, 33465–33470.
- Boucher,R.C. (2002) *Adv. Drug Deliv. Rev.*, **54**, 1359–1371.
- Bullock,A.N. and Fersht,A.R. (2001) *Nat. Rev. Cancer*, **1**, 68–76.
- Castagnoli,L., Vetricani,C. and Cesareni,G. (1994) *J. Mol. Biol.*, **237**, 378–387.
- Cesareni,G., Muesing,M.A. and Polisky,B. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 6313–6317.
- Cesareni,G., Cornelissen,M., Lacatena,R.M. and Castagnoli,L. (1984) *EMBO J.*, **3**, 1365–1369.
- Cesareni,G., Helmer-Citterich,M. and Castagnoli,L. (1991) *Trends Genet.*, **7**, 230–235.
- Crameri,A., Whitehorn,E.A., Tate,E. and Stemmer,W.P. (1996) *Nat. Biotechnol.*, **14**, 315–319.
- Distefano,M.D., Zhong,A. and Cochran,A.G. (2002) *J. Mol. Biol.*, **322**, 179–188.
- Eberle,W., Pastore,A., Sander,C. and Rosch,P. (1991) *J. Biomol. NMR*, **1**, 71–82.

- Eguchi,Y. and Tomizawa,J. (1990) *Cell*, **60**, 199–209.
- Eguchi,Y. and Tomizawa,J. (1991) *J. Mol. Biol.*, **220**, 831–842.
- Fersht,A.R. (1993) *FEBS Lett.*, **325**, 5–16.
- Fitzwater,T., Zhang,X.Y., Elble,R. and Polisky,B. (1988) *EMBO J.*, **7**, 3289–3297.
- Gregorian,R.S., Jr, and Crothers,D.M. (1995) *J. Mol. Biol.*, **248**, 968–984.
- Gu,H., Yi,Q., Bray,S.T., Riddle,D.S., Shiau,A.K. and Baker,D. (1995) *Protein Sci.*, **4**, 1108–1117.
- Guerois,R., Nielsen,J.E. and Serrano,L. (2002) *J. Mol. Biol.*, **320**, 369–387.
- Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
- Lahijani,R., Hulley,G., Soriano,G., Horn,N.A. and Marquet,M. (1996) *Hum. Gene Ther.*, **7**, 1971–1980.
- Lee,A.J. and Crothers,D.M. (1998) *Structure (Camb.)*, **6**, 993–1005.
- Lim,W.A. and Sauer,R.T. (1989) *Nature*, **339**, 31–36.
- Lim,W.A. and Sauer,R.T. (1991) *J. Mol. Biol.*, **219**, 359–376.
- Lin-Chao,S., Chen,W.T. and Wong,T.T. (1992) *Mol. Microbiol.*, **6**, 3385–3393.
- Martinez,E., Bartolome,B. and de la Cruz,F. (1988) *Gene*, **68**, 159–162.
- Mathews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.
- Munson,M., O'Brien,R., Sturtevant,J.M. and Regan,L. (1994a) *Protein Sci.*, **3**, 2015–2022.
- Munson,M., Predki,P.F. and Regan,L. (1994b) *Gene*, **144**, 59–62.
- Munson,M., Balasubramanian,S., Fleming,K.G., Nagi,A.D., O'Brien,R., Sturtevant,J.M. and Regan,L. (1996) *Protein Sci.*, **5**, 1584–1593.
- Nagi,A.D. and Regan,L. (1997) *Fold Des.*, **2**, 67–75.
- Nagi,A.D., Anderson,K.S. and Regan,L. (1999) *J. Mol. Biol.*, **286**, 257–265.
- Predki,P.F., Nayak,L.M., Gottlieb,M.B. and Regan,L. (1995) *Cell*, **80**, 41–50.
- Prusiner,S.B. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 13363–13383.
- Richards,F.M. (1997) *Cell. Mol. Life Sci.*, **53**, 790–802.
- Sauer,R.T., Jordan,S.R. and Pabo,C.O. (1990) *Adv. Protein Chem.*, **40**, 1–61.
- Schleif,R. (2000) *Trends Genet.*, **16**, 559–565.
- Siegele,D.A. and Hu,J.C. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 8168–8172.
- Taylor,S.V., Kast,P. and Hilvert,D. (2001) *Angew. Chem. Int. Ed. Engl.*, **40**, 3310–3335.
- Tomizawa,J. (1990) *J. Mol. Biol.*, **212**, 695–708.
- Tomizawa,J. and Som,T. (1984) *Cell*, **38**, 871–878.

Received October 1, 2003; accepted October 27, 2003

Edited by Dr Andrew Griffiths