A transcription blocker isolated from a designed repeat protein combinatorial library by in vivo functional screen

Elena B. Tikhonova, Abdul S. Ethayathulla*, Yue Su*, Parameswaran Hariharan, Shicong Xie† & Lan Guan

Department of Cell Physiology & Molecular Biophysics, Center for Membrane Protein Research, Texas Tech University Health Sciences Center, Lubbock, Texas 79430.

A transcription blocker isolated from a designed repeat protein combinatorial library by in vivo functional screen.

Elena B. Tikhonova, Abdul S. Ethayathulla*, Yue Su*, Parameswaran Hariharan, Shicong Xie† & Lan Guan

Department of Cell Physiology & Molecular Biophysics, Center for Membrane Protein Research, Texas Tech University Health Sciences Center, Lubbock, Texas 79430.

A highly diverse DNA library coding for ankyrin seven-repeat proteins (ANK-N5C) was designed and constructed by a PCR-based combinatorial assembly strategy. A bacterial melibiose fermentation assay was adapted for in vivo functional screen. We isolated a transcription blocker that completely inhibits the melibiose-dependent expression of α-galactosidase (MelA) and melibiose permease (MelB) of Escherichia coli by specifically preventing activation of the melAB operon. High-resolution crystal structural determination reveals that the designed ANK-N5C protein has a typical ankyrin fold, and the specific transcription blocker, ANK-N5C-281, forms a domain-swapped dimer. Functional tests suggest that the activity of MelR, a DNA-binding transcription activator and a member of AraC family of transcription factors, is inhibited by ANK-N5C-281 protein. All ANK-N5C proteins are expected to have a concave binding area with negative surface potential, suggesting that the designed ANK-N5C library proteins may facilitate the discovery of binders recognizing structural motifs with positive surface potential, like in DNA-binding proteins. Overall, our results show that the established library is a useful tool for the discovery of novel bioactive reagents.

Results

Construction of the ANK-N5C combinatorial library. The details in the design and construction are described in Methods. As other repeating proteins9,10,12, each ANK-N5C polypeptide contains N- and C-terminal cap repeats (N-CAP and C-CAP) and five internal repeats containing 33 amino-acid residues. They are mainly involved in protein-protein interactions through their concave surfaces. Combinatorial libraries coding for designed ankyrin proteins (DARPins) with three internal repeats were successfully constructed8,9,12,17,22–24. From such a library, several specific ANK proteins with various biological functions were identified by the in vitro ribosome-display method15,26, including crystallography chaperones27 and therapeutic agents, such as the vascular endothelial growth factor inhibitor1,26,28. To develop bio-reagents or binders for functional and structural studies, we created an ANK-based combinatorial library containing five internal repeats (ANK-N5C) by a ligase-independent, PCR-based combinatorial assembly strategy. By an in vivo functional screening method, we isolated a transcription blocker of the mel operon of Escherichia coli (E. coli). Crystal structure determination reveals that the transcription blocker is a domain-swapped dimer.
into six DNA fragments (Fig. 1b); each one was individually built. A defined codon mixture was supplied at an x-position during oligonucleotide synthesis. Applying an end-to-center sequential assembly approach, we could efficiently assemble the full-length DNA by PCR (Fig. 1b). Each polypeptide contains a total of 25 random residues.

Validation of constructs. DNA sequencing was used for the validation of the created plasmid clones. Consistent results are obtained from 146 clones isolated from three tests (Supplementary Table S1), showing that the average yield for obtaining ANK-N5C clones with an expected DNA length is about 46% (68 of 146). There are six clones containing four randomized internal repeats (ANK-N4C); thus, greater than 50% of clones contain randomized positions. There are 71 clones having open-reading-frame errors due to deletion or insertion of nucleobases. Among them, there are 27 clones (18%) with errors located in or adjacent to a randomized position, and 44 clones (30%) in a framework position. All clones, 68 ANK-N5C (group A) and 37 ANK-N5Cm (with manual correction for those with a mutation at a known framework position, group B), exhibit a unique deduced protein sequence. It is noteworthy that they have identical amino-acid identity at framework regions. The results indicate that each plasmid preparation contains totally different clones.

We further analyzed the degree of diversity by calculating pairwise Hamming distances (the total number of differences) within the clones of groups A and B, or the combined group AB. The distribution of Hamming distances is nearly symmetrical with a mode of 23 (Fig. 2a). More than 96% of the pairs have greater than 20 of the randomized positions occupied with different residues; 7% of the pairs have the maximal Hamming distance of 25; less than 1% of the pairs have distances between 14 and 18. There is no significant difference between groups A and B (data not shown). Furthermore, to quantify the randomness of amino-acid identity at each randomized position, we calculated the site-specific entropy and found that the entropy scores across all randomized positions are consistently high (Fig. 2b), with an empirical average number of 2.38. Compared with the average entropy score 2.80 calculated by simulating random sequences using the designed amino-acid usage, all specific positions in the library are randomized.

The overall amino-acid frequency, which was calculated from a total number of 2,625 randomized positions of the 105 ANK-N5C clones, is slightly biased toward hydrophobic residues (Fig. 2c). We have not observed Ala, Trp, and Lys residues from these samples, although Ala and Lys residues appear in other tests; residues Ile and Tyr are also poorly represented, probably due to the limited sampling number. Similar to the design for DARPin9, codons for Gly, Pro, and Cys residues were excluded from the design; however, Pro appears at a frequency of 3.6%. Examination of 30 x-positions occupied by Pro reveals that all are encoded by a specific codon (CCA) and scattered in all x-positions. While it is unclear, it is less likely due to errors in PCR or oligonucleotide synthesis procedure.

Figure 1 | ANK-N5C combinatorial library. (a), Design of ANK scaffold and randomized positions. Top panel, entropy scores calculated from 28 sequences of internal repeats from available ANK proteins. Inside the box, the designed amino-acid residues are shown at each of the consensus positions; the letter “x” represents any amino-acid residue but Gly, Pro, or Cys. Positions with a calculated entropy score less than 0.3, 0.4–1.0, 1.1–1.5, and greater than 1.6 are filled in green, blue, orange and red, respectively. Other designed ANK repeats are from references 8, 12, 23, where “Z” represents Asn, His, or Tyr. The secondary structure elements derived from the 3-D crystal structure of ANK-N5C-317 are illustrated. (b), Assembly strategy. The N-CAP, C-CAP, and five internal repeats (IR) are represented by rectangles filled with colors consistent with the rainbow presentation of the 3-D cartoon representation of the crystal structure in Fig. 5.
In vivo (encoded by a pCS19/FX-derived plasmid (Table 1) in the Tuner cell described in Methods (Fig. 3a). By expressing an ANK-N5C protein to identify ANK-N5C proteins inhibiting melibiose fermentation as approximately 5 \( \times 10^5 \) clones were used for the calculation of amino-acid usage frequency.

Crystal structure determination. High-resolution crystal structures for two ANK-N5C proteins, ANK-N5C-317 and ANK-N5C-281, were resolved to 2.5 (PDB ID, 4O60) and 2.0 Å (PDB ID, 4QFV), respectively (Table 2). The purified proteins are stable and readily crystallized. The 3-D structure of ANK-N5C-317 protein reveals a typical topology for ANK proteins (Fig. 5a–d); surprisingly, ANK-N5C-281 forms a domain-swapped dimer (Fig. 5e, f).

In ANK-N5C-317, the designed N-CAP, C-CAP, and five internal repeats form a “tiara-like” shape with two-layer helices (Fig. 5a); all the repeats are superimposed well except for the \( \beta \)-turn 1 of the internal repeat IV, as pointed by the arrow (Fig. 5b). Each repeat consists of two anti-parallel \( \alpha \)-helices. The consensus hydrophobic residues of adjacent helices form the continuous hydrophobic core between two-layer helices, which is stabilized by multiple H-bonds within and between repeats. The molecule with a mass of \( \sim 25 \) kDa has \( \sim 75 \) Å in length; its convex surface regularly presents negative and positive charges, and the concave surface is about 53 Å in width (Fig. 5c). The 25 randomized positions distribute in \( \beta \)-turn (particularly \( \beta \)-turn 1) and \( \alpha \)-helix 1, and form a continuous binding surface spanning approximately 25 Å (Fig. 5c).

In ANK-N5C-281, the refined model reveals that two molecules exchange their identical C-terminal two repeats, forming a domain-swapped dimer (Fig. 5e–f, Fig. 6a–c). The helical packing between the internal repeats IV and V of two swapped molecules is similar to that in ANK-N5C-317. The overall fold of the “hybrid monomer” in ANK-N5C-281, which consists of five N-terminal repeats with two C-terminal repeats from another molecule, also superimposes well with ANK-N5C-317 (Fig. 6b).

Pro residues are unexpectedly observed in both proteins. Pro138 of ANK-N5C-317 and Pro171 of ANK-N5C-281 are at position-5 of the internal repeats IV and V, respectively. It is likely that the dis-
Figure 3 | Isolation of a transcription blocker by an in vivo functional screen. (a), Illustration of the melibiose metabolic pathway in *E. coli*. For all the following panels, *E. coli* Tuner cells harboring pCS19/ET (vector control), pCS19/ANK-N5C-62 (protein control), or pCS19/ANK-N5C-281, respectively, were used. The plasmids are described in Table 1. Ampicillin at 100 μg/mL, melibiose at 30 mM, and IPTG at 0.3 mM were used unless otherwise described. (b), Protein-concentration dependent effect. The cells carrying a given plasmid were tested for melibiose and glucose fermentation, respectively, on MacConkey agar plates with ampicillin, sugar, and different concentrations of IPTG. The expression of ANK-N5C proteins was analyzed with His-tag antibody by Western blotting using the cells collected from above melibiose-containing plates. (c), Cell growth on M9 minimal media supplemented with ampicillin, IPTG, and 10 mM melibiose. (d), Detection of MelA and MelB activity and expression. Cells were grown at 30 °C in LB media containing 0.5% glycerol, ampicillin, and IPTG with or without 10 mM melibiose, and used for activity and expression of MelA (top two panels) and MelB (bottom two panels). The activity of MelA and MelB (bars) are expressed as α-NPG hydrolysis and uptake of [1-3H]melibiose (0.4 mM, 10 mCi/mmol), respectively. Expression of MelA and MelB proteins (images) were analyzed by Western blot. A total of 40 μg cell extracts or membrane proteins, as well as 100 ng of purified MelA or MelB were loaded on SDS-12% PAGE. Error bars, S.E.M.; n = 3-5. (e), RT-PCR. Total RNA was purified from cells grown described in panel d. The reverse transcriptase enzyme mix was treated for 10 min at 95 °C before PCR reaction for control. Products were analyzed on 3% agarose gel. Water, instead of cells, was used for control.
Table 1 | E. coli strains and plasmids used in this study

| Strains and vectors | Description | Reference |
|---------------------|-------------|-----------|
| Strains<br>
E. coli One Shot™<br>ccdB Survival™ 2 T1®<br>E. coli Stellar™<br>E. coli Tuner™<br>E. coli T7 express<br>E. coli DW2<br>E. coli XL1-Blue<br>Plasmids<br>pIINTAL<br>pCS19<br>pCS19/FX<br>pCS19/ET<br>pCS19/ANK-N5C-62<br>pCS19/ANK-N5C-281<br>pCS19/ANK-N5C-281/P171Q<br>pCS19/ANK-N5C-281/P171F<br>p7XC3H<br>p7XC3H/ANK-N5C-281<br>p7XC3H/ANK-N5C-317<br>p7XNH3<br>p7XNH3/MelA<br>p140/MelB-6His<br>p95/ΔAHB<br>p175/LacY-10His<br>pACYC/C6 lacY<br>pACYC/FX<br>pACYC/ET<br>pACYC/MelB<br>pACYC/MelAB<br>pACYC/LacY<br>pACYC/MelR | FX cloning vector with Sapl and ccdB gene; kan'<br>pQE60 derivative inserted with gene lacZ; amp'<br>pCS19 with two Sapl sites and ccdB gene for FX cloning; amp'<br>Expression vector harboring gene encoding for ANK-N5C-62; amp'<br>Expression vector harboring gene encoding for ANK-N5C-281; amp'<br>P171Q mutant of ANK-N5C-281<br>Expression vector with Sapl and ccdB gene for FX cloning; kan'<br>Expression vector harboring gene encoding for ANK-N5C-281; kan'<br>Expression vector harboring gene encoding for ANK-N5C-317; kan'<br>Expression vector with Sapl and ccdB gene for FX cloning; kan'<br>melA gene of XL1 Blue subcloned into p7XNH3; kan'<br>Expression vector containing melB gene; amp'<br>Expression vector containing melB gene; amp'<br>Expression vector containing lacY gene; amp'<br>Expression vector containing the C-terminal half of lacY gene with lac operator/promoter in pACYC184<br>pACYC with two Sapl sites and ccdB gene for FX cloning; cam'<br>ccdB gene was removed from pACYC/FX; cam'<br>melB gene of Tuner strain subcloned into pACYC/FX; cam'<br>melB operon of Tuner strain subcloned into pACYC/FX; cam'<br>lacY gene of Tuner strain subcloned into pACYC/FX; cam'<br>melR gene of Tuner strain subcloned into pACYC/FX; cam' | 38<br>52<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study<br>51<br>This study<br>51<br>51<br>54<br>This study<br>This study<br>This study<br>This study<br>38<br>54<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study<br>This study |

Discussion

We optimized efficient PCR-based protocols for constructing a combinatorial DNA library coding seven ankyrin repeats (ANK-N5C). The obtained DNA library has high accuracy (46%) and high diversity. Theoretically, the diversity is calculated to contain 1725 or 5.8 × 1010 unique molecules; certainly, this number is limited by PCR reaction for assembling fragments I–II with III. Practically, each batch of full-length PCR fragments is estimated to have >104 unique molecules; however, completely different ANK-N5C clones can be obtained by re-assembling the available DNA fragments by mix-and-match. It is worthy to mention that a specific selection method determines the diversity of each screen.

Protein ANK-N5C-317, which shows a partial inhibition of melibiose fermentation (Fig. S1), exhibits a typical ankyrin fold. The overall architecture is similar to other natural ankyrins with seven repeats, such as the Gankyrin that is involved in epithelial tumor development4, and the vaccinia virus K1 protein that is a host-range protein11. Their size and shape are different from the DARPinss4 that contains three internal repeats. For the isolated transcription blocker ANK-N5C-281, surprisingly, an unexpected domain-swapped dimer is observed from four crystal structures refined to resolution at 2.0–2.5 Å, and only the structure with highest resolution was reported here.

It is apparent that repeating proteins may have a tendency to form intermolecular domain swapping10,12. For ANK-N5C-281, Pro171 at

Support the notion that Pro171 in ANK-N5C-281 plays a critical role in blocking the transcription activation of melAB operon.
position-5 was randomly selected. It is noteworthy that the main-chain nitrogen atom at position-5 forms a critical H-bond with the negatively charged carboxyl group of Asp at the conserved position-1 (Fig. 1a, Supplementary Fig. S2b) for maintaining the β-turn 1 structure. Pro171 interrupts this critical interaction and imposes conformational flexibility, which substitute the β-turn into a hinge loop. A similar mechanism for generating a domain-swapped dimer was proposed theoretically\(^{33,34}\). The additional contacts between the two hinge loops (Fig. 6c), as well as the randomly selected salt-bridge between internal repeats III and V, make the domain-swapped dimer more favorable thermodynamically (Fig. 6b, c). Both P171Q and P171F mutants of ANK-N5C-281 completely lose the inhibitory effect on melibiose fermentation; however, Pro at position-5 can not be used as a sole evidence for prediction of domain-swapping.
event of ankyrin proteins. In ANK-N5C-317, Pro138 presents also at position-5 but does not induce a domain swapping; instead, it only locally interrupts the β-turn-1 (Supplementary Fig. S2a). The observed additional interaction, particularly the specific salt-bridge interaction between intra-molecular internal repeat III and V seems also critical. Consistently, BN-PAGE analysis indicates that ANK-N5C-281 migrates much slower than that of the control protein ANK-N5C-62 with a similar molecular weight of ~25 kDa (Fig. 6d). More

Table 2 | Data collection and refinement statistics (Molecular replacement)

|                      | ANK-N5C-317 (PDB ID 4O60) | ANK-N5C-281 (PDB ID 4QFV) |
|----------------------|---------------------------|---------------------------|
| **Data collection**  |                           |                           |
| Space group          | C2                        | P2₁                       |
| Cell dimensions      | a, b, c (Å)               | a, b, c (Å)               |
| α, β, γ (°)          | 119.46, 46.80, 74.24      | 66.97, 94.34, 69.64       |
| Resolution (Å)       | 50.0–2.50 [2.59–2.50]*    | 50.0–2.00 [2.03–2.00]     |
| R_sym or R_merge     | 0.09 [0.70]               | 0.07 [0.53]               |
| I/σI                | 30.0 (5.5)                | 18.0 (2.5)                |
| Completeness (%)     | 99.9 (100)                | 98.6 (95.7)               |
| Redundancy           | 6.4 (7.0)                 | 7.4 (6.2)                 |
| **Refinement**       |                           |                           |
| Resolution (Å)       | 50.0–2.50                 | 50.0–2.00                 |
| No. reflections      | 26425                     | 56653                     |
| R_work/R_free        | 0.18/0.24                 | 0.187/0.219               |
| No. atoms            | 3529                      | 7401                      |
| Protein              | 3473                      | 6883                      |
| Water                | 56                        | 518                       |
| B-factors            | 45.2                      | 31.6                      |
| Protein              | 43.0                      | 31.0                      |
| Water                | 38.5                      | 33.8                      |
| R.m.s. deviations    | Bond lengths (Å)          | Bond angles (°)           |
|                      | 0.006                     | 0.004                     |
|                      | 1.1                       | 0.89                      |

*Values in parentheses are for highest-resolution shell.

Figure 5 | X-ray crystal structures. (a–d), ANK-N5C-317; (e), (f), ANK-N5C-281. (a), Cartoon representation of the crystal structure of ANK-N5C-317 protein (PDB ID 4O60, 2.5 Å). The N- and C-terminal CAP repeats (N-CAP ad C-CAP) and five internal repeats (IR) are indicated. (b), Superposition of repeats of ANK-N5C-317 protein. The secondary structure elements are indicated. The arrow points to the disturbed β-turn-1 of the internal repeat IV. (c), Surface representation. All 25 randomized positions are indicated by red-colored C-x positions. (d), Surface electrostatic potential map on ANK-N5C-317 structure was calculated by APBS program. (e), Overall folding of the domain-swapped dimer ANK-N5C-281 (PDB ID 4QFV, 2.0 Å). (f), Surface potential map on ANK-N5C-281 structure was calculated by APBS program.
studies are, however, required to determine if the domain-swapped dimer contributes to its biological function.

The studies presented here may be useful for designing protein-based combinatorial libraries. A common scenario is to exclude helical breakers (Pro and Gly) from codon optimization in order to avoid breaking protein folding \(^9\) and domain swapping \(^{10}\). In contrast to this, we show that the inhibitory activity of the transcription blocker ANK-N5C-281 requires the presence of Pro171 (Fig. 6e). On the other hand, while the diversity of a repeat protein-based library is high in general, a fixed scaffold limits the extent of diversity with regard to topology and architecture. Therefore, inclusion of Pro or Gly for codon optimization may increase the probability of obtaining molecules with unexpected topology/architecture, and some of them may possess novel biological functionality. An inverted dimer as observed in ANK-N5C-281 may favor the capture of dimeric transcription factors.

The colony-based functional screen we developed here is a powerful generic approach, which allows all molecules involved in the same function to undergo selection simultaneously in a physiological condition. The phenotype and genotype of a selected binder are directly coupled. There are many proteins that are not amenable for \textit{in vitro} characterizations, such as the MelR and its homologues of the AraC/XylS family\(^{35,36}\). It is likely that such \textit{in vivo} functional screen may be the simple solution for obtaining a binder that possesses a biological activity. This may be especially relevant when the target protein is a part of complex, as many do. In this case, a single-target protein-based screening method, \textit{in vivo} or \textit{in vitro}, may not necessarily produce a binder that is physiologically relevant. The drawback of this \textit{in vivo} approach is that the dissection of the underlying mechanism is usually time-consuming due to the biological complexity. In any case, it is important that an ANK-N5C protein isolated from such functional screen is active in a physiological condition, as demonstrated here. It is worthy to point out that this method is not a general screening technique but it is for specific targeting of bacterial melibiose uptake and metabolism. The fermentation approach is, however, applicable for targeting other bacterial proteins involving in transport and metabolism of varied sugars, such as, glucose, lactose, or maltose.

The functional and structural studies show that the constructed ANK-N5C-281 library is chemically and topologically diverse. From rather small size of population (\(5 \times 10^5\) clones), a transcription blocker of \textit{melAB} operon was isolated. Among all proteins involved in melibiose fermentation including melibiose uptake, hydrolysis, and glucose metabolism, the transcription of \textit{melAB} operon appears

---

**Figure 6 | Dimer interface in ANK-N5C-281.** The two swapped monomers in ANK-N5C-281 crystal structure are colored in green and cyan. The structure of ANK-N5C-317 is colored in blue. (a), The 2Fo-Fc electro density map (contour at 1.2 \(\sigma\)) shows the hinge loop linking the N-terminal five repeats with C-terminal two repeats. (b), Superposition of ANK-N5C-317 with ANK-N5C-281. Salt-bridge interactions between Glu133 (position-33 of internal repeat III) and Arg199 (position-33 of internal repeat V) within one ANK-N5C-218 monomer are shown by dotted lines. The "hybrid monomer" is indicated. (c), Interactions between the two hinge loops in the domain-swapped dimer. Dotted lines indicate H-bonding interactions. The random residues are underlined. Helices from the internal repeats-IV and -V are indicated, respectively. The \(\beta\)-turn-1 of the internal repeat V of ANK-N5C-317 is colored as blue and indicated by arrow. Partial sequence alignment of ANK-N5C-62, ANK-N5C-317, and ANK-N5C-281 are shown in the box underneath. \(x\), randomized position; \(*\), consensus position. Amino acid positioning in protein and in repeat are indicated. Pro171 and the charge pair Glu133/Arg199 in ANK-N5C-281 are colored in blue and red, respectively. (d), BN-15\%PAGE. ANK-N5C proteins (5 mg each) and the NativeMark\textsuperscript{TM} unstained protein standard were loaded on each well. (e), Site-directed mutagenesis of ANK-N5C-281. Effect of ANK-N5C-281/P171Q or F mutant on melibiose fermentation was carried out as described in the legend to Fig. 3.
to be an easier target for the designed library. The DNA-binding protein MelR, the key protein in transcription activation of the mel operon, is a transcription activator for the melAB operon and also a suppressor for its own expression10,11,12. The current data suggest that MelR’s function is inhibited by ANK-N5C-281 because overexpression of MelR suppresses the effect of ANK-N5C-281; however, the precise inhibitory mechanism is unknown.

It is noteworthy that ANK-N5C proteins possess a negative cave surface in general, implying that a protein with a positively charged surface, such as in DNA-binding, may be easier captured. The ANK-N5C library, like DARpin, is a good resource that can be easily adapted for other in vitro display methods, such as ribosome display, or in vivo screening method, such as two-hybrid system for discovery of blockers, inhibitors, or binders.

**Methods**

**Bacterial strains and plasmids.** The genotype and source of E. coli strains and plasmids used in this study are listed in Table 1. Construction of vectors and expression plasmids is described in Supplementary Note.

**Design of ANK-N5C combinatorial library.** The ANK-N5C combinatorial library (Fig. 1a) was designed based on amino-acid conservation and structural analyses, as well as published information8,10,11. Among the 495 ANK repeats collected from UniProt and RCSB PDB databases, 28 sequences of unique ANK repeat with 33 residues in length were selected. Calculated entropy scores (Shannon entropy) from the 28 protein sequences show that 21 out of the 33 positions are highly conserved with an entropy score lower than 1.0 (Fig. 1a, green and blue shades); accordingly, 20 positions, except for the position-33, were assigned as a framework position. Another 7 positions (positions-12, -14, -16, -22, -25, -26, and -30), with relatively higher entropy scores, were also assigned as framework positions with the most frequent residues because these positions are less likely to contribute to a binding motif. Five positions with higher entropy scores (> 1.6) were assigned as potential randomized positions (position-2, -3, -5, -10, and -13). Position-33, with the lowest entropy was also selected for codon optimization because of its location in close proximity to the cluster of randomized positions. The total number of randomized position per each polypeptide is 25; position-2 in the internal repeat I, position-13 in the internal repeat IV, and positions-10, 13, 33 in the internal repeat V are not randomized for facilitating DNA assembly by PCR. The N-CAP contains 31 residues (D1GKLKDLLASRRGAGHDEYVLLLKGDAD) The first 18 residues were the same as the previously reported N-CAP9, and the last 13 residues mimic the framework of the internal repeat designed in this study. The C-CAP contains 29 residues (D1GKFKTPDLADINGNEDAAELVQQAARS) that follow the previously optimized sequence10 with a 6xHis tag at the C-terminal end to facilitate DNA assembly and protein purification.

**PCR-based assembly strategy.** The entire DNA fragments coding the library ANK-N5C proteins were divided into six overlapping DNA modules (Fig. 1b). Each duplex DNA module was created by conventional annealing and extension reactions. The full-length DNA fragments were obtained by a PCR-based assembly method at a bidirectional end-to-end cloning approach (Fig. 1b). The DNA oligonucleotides were synthesized by Integrated DNA Technologies, Inc. Construction of plasmid libraries based on a fragment exchange cloning method (EX)46 was described in the Supplementary Note.

**Melibiose fermentation.** The rich media MacConkey agar plates containing melibiose as the sole carbohydrate source was used for melibiose fermentation. Red colonies grown on MacConkey agar indicate melibiose utilization; yellow colonies denote no melibiose fermentation47,48. In Tuner cells (laccZ), mel operon is solely responsible for melibiose transport and hydrolysis, and transcription activation of melAB is induced by melibiose, not by isopropyl β-D-1-thiogalactopyranoside (IPTG) (IPTG) (Fig. 3d).

Tuner competent cells were transformed with pCS19/ANK-N5C library plasmids and plated onto the lactose-free MacConkey agar plate containing 30 mM melibiose (inducer for mel operon) as the sole carbohydrate source, 100 mg/L ampicillin, and 0.1 mM IPTG (inducer for expression of pCS19/ANK-N5C), and incubated in 37°C overnight. Colonies from the covalently reproducible phage plaques were selected for random plasmid preparation and DNA sequencing analysis. The fluorescence culturing was followed after the protocols using 30 mM glucose instead of melibiose.

**Melibiose transport assay.** Melibiose transport assays with intact cells were carried out by fast filtration assay with [1-3H]melibiose as described10,11,12. The E. coli cells, which are grown with 0.3 mM IPTG for plasmid expression and in the absence or presence of 10 mM melibiose for inducing the melAB operon, were washed with 100 mM KP, (pH 7.5) for transport assay at 0.4 mM melibiose and 20 mM NaCl.

**MelA activity assay.** The Tuner cells were grown in the absence or presence of 10 mM melibiose, and broken by sonication. The cell extracts were used to detect the α-galactosidase activity using p-nitrophenyl-α-galactoside (α-NPG) as the substrate, following published descriptions10 with minor modifications. Absorption at 405 nm was measured after 15 min incubation at 37°C. Total amount of hydrolysis product was estimated using the extinction coefficient value for p-nitrophenol moiety as 18300 M⁻¹ cm⁻¹. MelA activity was expressed as nmol α-NPG/min/mg total cell proteins.

**RT-PCR.** Total RNA samples from the E. coli Tuner cells were isolated by RNAeasy Mini Kit (Qiagen). An equal amount of RNA (200 ng) was used for each 50-μL RT-PCR reaction. The melA-specific primers were designed for amplifying a 146-bp fragment, and the control is a 101-bp fragment for the rrsD gene that encodes for the 16S RNA8. The reaction was performed using Transcriptor One-Step RT-PCR kit (Roche) with 20, 25 and 30 cycles for monitoring the dynamics of amplification. Amplicons were analyzed by DNA electrophoresis on 3% agarose gel. The reverse transcriptase was heat-inactivated for verification of potential chromosomal DNA contamination.

**Cell growth on M9 media.** The overnight cultures were prepared in LB media containing 100 μg/ml ampicillin, and cells were washed with M9 media and re-inoculated into M9 media supplemented with 10 mM melibiose, 100 μg/ml ampicillin, and 0.3 mM IPTG, and shaken at 37°C. Absorption at 600 nm was monitored.

**Antibody preparation and Western blot analysis.** MelA and MelB proteins purified as described in the Supplementary Note were used to raise rabbit polyclonal antibody samples by the Covance Research Products Inc. Polyclonal anti-C-terminal LacY antibody was also used to recognize LacY protein expression. The protein A-agarose affinity chromatography was used for the detection of the specific antibody-bound MelA, MelB, and LacY. Penta-His HPR conjugate antibody (Qiagen) was applied to detect ANK-N5C expression encoded by pCS19/FX-derived vector. A total of 40 μg cell extracts or membrane proteins were separated on SDS-12% PAGE, and Western blot analysis were carried out as described.

**Blue native-polyacrylamide gel electrophoresis.** Proteins were analyzed by BN-15%PAGE at 4°C following the protocol provided by Life Technologies.

**Crystalization, data collection and processing.** Expression and purification of ANK-N5C-317 and ANK-N5C-281 proteins were described in the Supplementary Note. Crystalization trials were carried out by the hanging-drop vapor-diffusion method at 23°C by mixing 2 μl of protein sample at a protein concentration of about 10 mg/ml with 2 μl of reservoir containing 100 mM sodium acetate trihydrate (pH 4.2), 200 mM (NH₄)₂SO₄, 18–20% PEG 3350, and 10% glycerol. Crystals were frozen in liquid nitrogen after soaking in the mother liquid supplemented with 25% PEG 3350 and 10% glycerol as cryo-protectants, and tested for X-ray diffraction at the Lawrence Berkeley National Laboratory, Advanced Light Source BL 8.2.2 or 5.0.1 via remote data collection. The complete diffraction datasets for ANK-N5C-317 and ANK-N5C-281 were collected at 100 K with an ADSC QUANTUM 315 and 315R Detector, respectively. Image data were processed with HKL200048 to a resolution of 2.5 A. A 2.2 space group with 99.9% completeness for ANK-N5C-317 and 2.0 A in P2₁ space group with 98% completeness for ANK-N5C-281 (Table 2).

**Structure solution and refinement.** The structure of ANK-N5C-317 was solved by molecular replacement using the DARPin protein containing three internal repeats (PDB ID, 2XEE) as the search probe and the Phaser 2.52 program in Phenix suite. The asymmetric unit contains two closely packed molecules with 40% solvent content. An initial model was built using the Phenix SP AutoBuild program. OMIT maps and simulated annealing and density modification yielded an interpretable density map at 2.5 Å resolution. With iterative rounds of manual model building and refinement, the complete model for the ANK-N5C-317 was built. Water molecules were added at the end of the refinement with the final R/AR free values of 0.18/0.24 (Table 2). Out of 234 residues including the C-terminal six-His tag, the side-chain positioning for residues 2–231 in Mol-A and 4–231 in Mol-B were well resolved. No residues are in the disordered regions, 96.18% of residues are in most favored regions, 3.16% generously allowed regions. For the structure determination of ANK-N5C-281, the structure of ANK-N5C-317 was used as a searching model for molecular replacement. During model refinement, we observed strong positive difference Fourier between neighboring molecules and main-chain clashes in the regions of S166DFSG170, suggesting domain-swapping event. The model was re-built according to the density, yielding two domain-swapped dimers in the asymmetric unit. A total of 518 water molecules were added at the end of the refinement with the final R/AR free values of 0.18/0.21 (Table 2). Out of 234 residues including the C-terminal six-His tag, the side-chain positioning for residues 1–229 in Mol-A, 3–229 in Mol-B, and 4–230 in Mol-C and Mol-D were well resolved. No residues are in the disordered regions, 96.18% of residues are in most favored regions, 3.16% generously allowed regions. Visualization of omit maps and manual model building were performed using Coot0.7. Surface electro-potential maps were calculated using APBS software49. All crystallographic figures were generated with Pymol. 1. Tamaskovic, R., Simon, M., Stefan, N., Schwill, M. & Pucilovski, A. Designed ankyrin repeat proteins (DARPins) from research to therapy. Methods Enzymol 503, 101–134 (2012).
2. Colas, P. Combinatorial protein reagents to manipulate protein function. Curr Opin Chem Biol 4, 54–59 (2000).
3. Kossiakoff, A. A. & Koide, S. Understanding mechanisms governing protein-protein interactions from synthetic binding interfaces. Curr Opin Struct Biol 18, 499–506 (2008).
4. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505–510 (1990).
5. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818–822 (1990).
6. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. Nature 344, 467–468 (1990).
7. Colas, P. et al. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature 380, 548–550 (1996).
8. Mosavi, L. K., Minor, D. L., Jr. & Peng, Z. Y. Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci U S A 99, 16029–16034 (2002).
9. Binz, H. K., Stumpf, M. T., Forrer, P., Amstutz, P. & Pluckthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 332, 489–503 (2003).
10. Madhurantakam, C., Varadamsetty, G., Grutter, M. G., Pluckthun, A. & Mittl, P. R. Structure-based optimization of designed Armadillo-repeat proteins. Protein Sci 21, 1015–1028 (2012).
11. Gilbreth, R. N. & Koide, S. Structural insights for engineering binding proteins based on non-antibody scaffolds. Curr Opin Struct Biol 22, 413–420 (2012).
12. Seeger, M. A. et al. Design, construction, and characterization of a second-generation DARPin in library with reduced hydrophilicity. Protein Sci 12, 1239–1257 (2003).
13. Ferreiro, D. U. & Komives, E. A. The plastic landscape of repeat proteins. Proc Natl Acad Sci U S A 104, 7735–7736 (2007).
14. Michaely, P. & Bennett, V. The membrane-binding domain of ankyrin contains four independently folded subdomains, each comprised of six ankyrin repeats. J Biol Chem 268, 27703–27709 (1993).
15. Lux, S. E., John, K. M. & Bennett, V. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. Nature 344, 36–42 (1990).
16. Li, J., Mahajan, A. & Tsai, M. D. Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry 45, 15168–15178 (2006).
17. Forrer, P., Stumpf, M. T., Binz, H. K. & Pluckthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett 559, 2–6 (2003).
18. Mohler, P. J. et al. Ankyrin-B mutation causes type 4 long QT cardiac arrhythmia and sudden cardiac death. Nat Med 421, 634–639 (2003).
19. Li, J., Kline, C. F., Hund, T. I., Anderson, M. E. & Mohler, P. J. Ankyrin-B regulates Kir2.6 membrane expression and function in heart. J Biol Chem 285, 28723–28730 (2010).
20. Lambert, S. & Bennett, V. From anemia to cerebellar dysfunction. A review of the ankyrin gene family. Eur J Biochem 211, 1–6 (1993).
21. Krzywy, S. et al. The crystal structure of gankyrin, an oncoprotein found in the human epidermal growth factor receptor 2. J Biol Chem 285, 269–279 (2010).
22. Bowers, G. N., Jr., McComb, R. B., Christensen, R. G. & Schaffer, R. High-purity 4-melibiose symport by MelB. Nature 295, 809 (1980).
23. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning and the tumor suppressors Rb and p53. Nature Cell 8, 22703–22709 (1993).
24. Kahramanoglou, C., Webster, C. L., El-Robh, M. S., Belyaeva, T. A. & Busby, S. J. In vitro selection of mAb alpha-galactosidase from Citrobacter freundii: a family 4 glycosyl hydroxide in which oxidation is rate-limiting. Biochemistry 50, 4298–4308 (2011).
25. Bowers, G. N., Jr., McComb, R. B., Christensen, R. G. & Schaffer, R. High-purity 4-melibiose symport by MelB. Nat Commun 5, 3009 (2014).
26. Chakladar, S., Cheng, L., Chiu, M., Liu, J. & Bennett, A. J. Mechanistic evaluation of MelA alpha-galactosidase from Citrobacter freundii: a family 4 glycosyl hydroxide in which oxidation is rate-limiting. Biochemistry 50, 4298–4308 (2011).
27. Sennhauser, G. & Grutter, M. G. Chaperone-assisted crystallography with the human epidermal growth factor receptor 2. J Biol Chem 285, 269–279 (2010).
28. Eha, a transcriptional regulator of hemolytic activity of Escherichia coli: localization of helix VI. J Mol Biol 312, 69–77 (2001).
29. Otwinski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol 276, 307–326 (1997).
30. McCoy, A. J. et al. Phaser crystallographic software. J Appl Crystallogr 40, 658–674 (2007).
31. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of large scale purification and evidence that H+, Na+, and Li+ sugar symport is catalyzed by a single polypeptide. Biochemistry 34, 4412–4420 (1995).
32. Spiess, C., Bell, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell 97, 339–347 (1999).
33. Guo, D. et al. Eha, a transcriptional regulator of hemolytic activity of Edwardsiella tarda. FEMS Microbiol Lett 353, 132–140 (2014).
34. Guan, L., Weinglass, A. B. & Kaback, H. R. Helix packing in the lactose permease of Escherichia coli: localization of helix VI. J Mol Biol 312, 69–77 (2001).
35. Bibi, E. & Kaback, H. R. In vivo expression of the lacY gene in two segments leads to functional lac permease. Proc Natl Acad Sci USA 94, 13425–13429 (1997).

Acknowledgments

We thank Eric R. Geertsma and Raimund Dutler for their FX cloning tools. We thank Michael Ehrmann for the gift of plasmid pCS19; Ronald Kaback for the plasmids pACYC/Cfr lac, pET-55WT LacY 10His, and Lac-Y antibody; and Gerard Leblanc for a MelB expressing vector and DW2 strain. We thank Ronald Kaback for encouragement of this project, and Luis Reuss for reading the manuscript. This work was supported by the National Institutes of Health Grant R01 GM095538 to L.G.

Author contributions

L.G. conceived and directed this research. L.G. designed the ANK-N5C library and chaperone to protease in a widely conserved heat shock protein. J Biol Chem 285, 28723–28730 (2010).

Additional information

Author Information

The constructed genes and protein structures (ANK-N5C-317 and ANK-N5C-281) have been deposited in the Gene Bank (accession numbers KP816369 and
KJ633117), and in the Protein Data Bank with the accession numbers (PDB IDs, 4O60 and 4QFV), respectively.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Tikhonova, E.B. et al. A transcription blocker isolated from a designed repeat protein combinatorial library by in vivo functional screen. Sci. Rep. 5, 8070; DOI:10.1038/srep08070 (2015).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/