SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
COMBINATORIAL LIBRARIES
HIGH-THROUGHPUT SCREENING

Received
1 October 2014

Accepted
2 January 2015

Published
28 January 2015

Correspondence and requests for materials should be addressed to L.G. (Lan.Guan@ttuhsc.edu)

* These authors contributed equally to this work.

† Current address: Computational and Systems Biology Graduate Program, MIT, Cambridge, MA

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.

Combinatorial chemistry is a powerful method for creating biological materials for discovery of novel bioactive reagents[1–3]. Aptamers[4,5], including DNA-, RNA-, and peptide-aptamers, are commonly used materials for building combinatorial libraries[6,7]. Recently, proteins made up of repeating sequences (repeat proteins) have been tested as scaffolds[8–12] for presenting variable surfaces (binding surfaces). Ankyrin repeat proteins (ANK) belong to the adaptor protein family and constitute 6% of eukaryotic proteins with known sequence[13]. They exist in many living forms and modulate numerous critical cellular functions[14–19], such as transcription regulation, cell-cycle control, cell signaling, development and differentiation, and membrane protein targeting and activity. These proteins are also associated with human diseases, such as cancer and neurological disorders[20,21]. Structurally, ANK are composed of tandem repeating motifs, frequently with 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 constructed[8,9,12,17,22–24]. From such a library, several specific ANK proteins with various biological functions were identified by the *in vitro* ribosome-display method[25,26], including crystallography chaperones[27] and therapeutic agents, such as the vascular endothelial growth factor inhibitor[1,26,28]. 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.

Results

Construction of the ANK-N5C combinatorial library. The details in the design and construction are described in Methods. As other repeating proteins[9,10,12], 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 (Fig. 1a), yielding a molecule with a mass of approximate 25 kDa. Based on the available ankyrin sequences and reported library[9], a consensus scaffold for each internal repeat is designed as: DxxGxTPLHxAAxNGHLELVKLLLEKGADINAx, wherein these assigned residues will repeatedly appear at the same framework positions in each internal repeat. The letter "x" denotes codon randomization. The full-length DNA fragment coding for ANK-N5C was divided

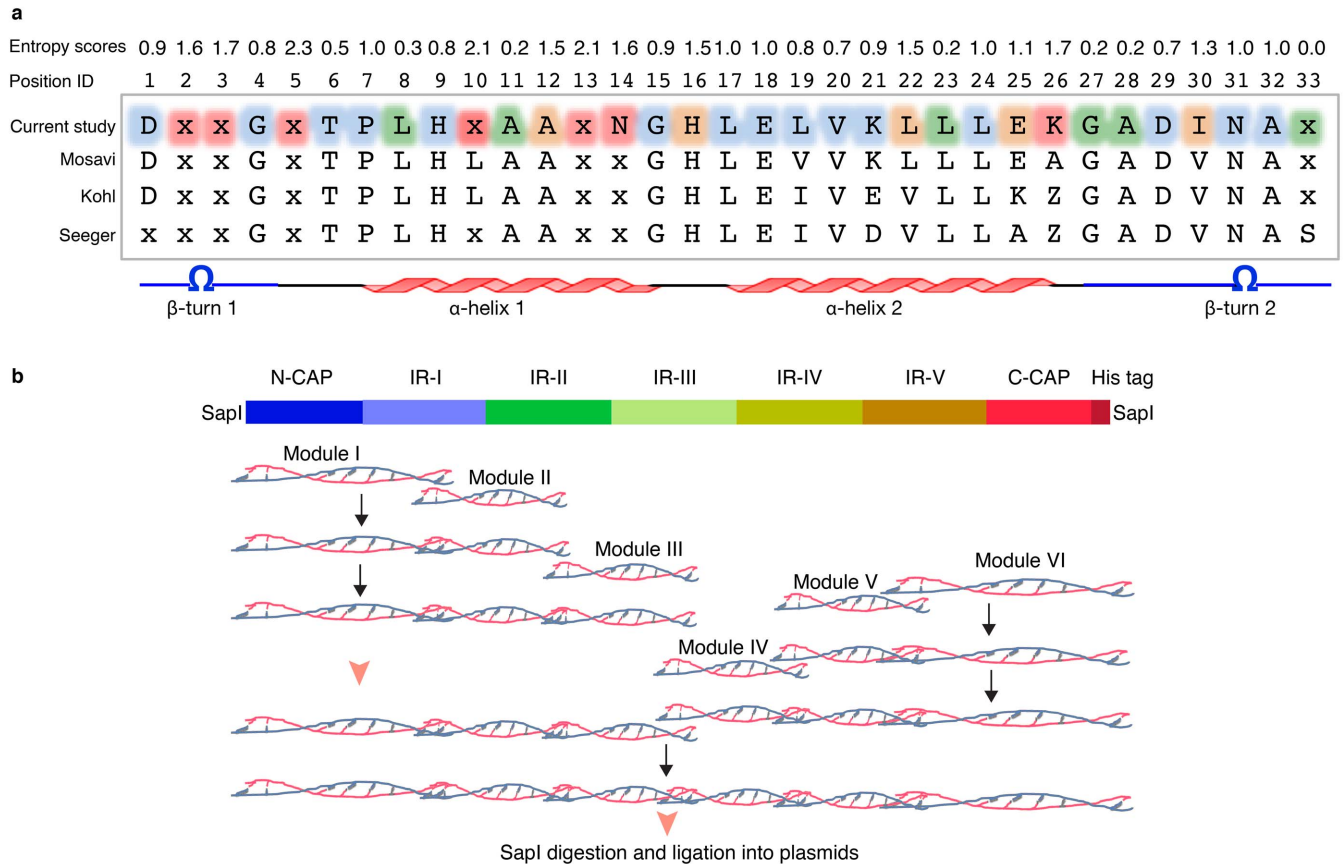

Figure 1 | ANK-N5C combinatorial library. (a), Design of ANK scaffold and randomized positions. Top panel, entropy scores calculated from 28 sequences of internal repeats from available ANK proteins. Inside the box, the designed amino-acid residues are shown at each of the consensus positions; the letter "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.

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 DARPins[9], 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.

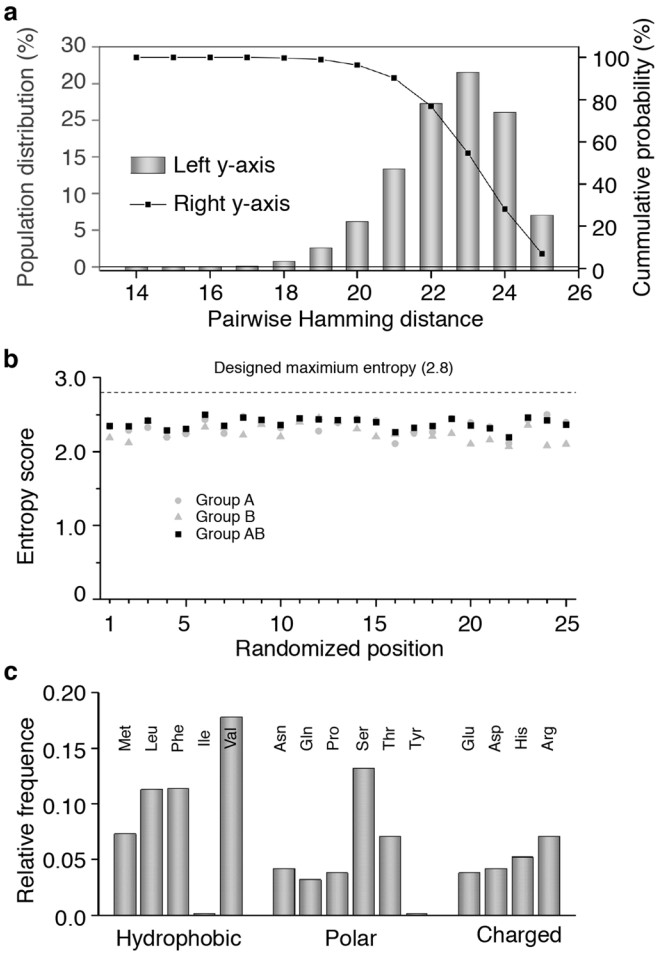

Figure 2 | Diversity evaluation of the ANK-N5C combinatorial library. A total of 105 ANK-N5C clones were used for analysis; among them, 37 clones are N5Cm containing manual corrections at consensus regions; the manual corrections should not affect the levels of randomness introduced at randomized positions. The amino-acid frequency is used to represent the codon frequency because of the single codon usage. (a), Pairwise Hamming distance. From 5525 pairs among the 105 clones, the population distribution (left y-axis) and cumulative probability (right y-axis), respectively, are plotted against the number of variant positions. (b), Shannon entropy at each randomized position. The site-specific diversity, expressed by the Shannon entropy, was calculated from the amino-acid frequency at each randomized position for group A (68 N5C), group B (37 N5Cm), and group AB (105 clones). Dashed line indicates the average entropy score estimated from simulated random sequences with the designed amino acid mixture (2.8), representing the maximum entropy. (c), Overall amino-acid usage. A total of 2625 randomized positions from 105 clones were used for the calculation of amino-acid usage frequency.

***In vivo* functional screen.** In *E. coli*, the *mel* operon (Fig. 3a), which encodes MelA and MelB, is needed for melibiose metabolism[29,30]. For a pilot study, we developed a colony-based functional screening method to identify ANK-N5C proteins inhibiting melibiose fermentation as described in Methods (Fig. 3a). By expressing an ANK-N5C protein encoded by a pCS19/FX-derived plasmid (Table 1) in the Tuner cell (*lacY*$^-$) on melibiose-containing MacConkey agar plates, we identified one yellow colony and 35 other colonies with reduced color from approximately 5×10^5 colonies (Fig. S1a–c). Some clones affect cell growth and some affect glucose fermentation. In this study, we only focus on the one that completely inhibits melibiose fermentation.

Using the cells containing either an empty plasmid or a plasmid encoding ANK-N5C-62 protein that does not affect melibiose fermentation as the controls, we show that the clone ANK-N5C-281 does not inhibit glucose fermentation but inhibits melibiose utilization. This inhibition is concentration dependent, as demonstrated by the level of fermentation, which correlates with expression of the ANK-N5C protein (Fig. 3b). Furthermore, the cells containing ANK-N5C-281 protein fail to grow on melibiose as sole carbon source (Fig. 3c).

With the Tuner cells expressing ANK-N5C-281, the melibiose-induced α-galactosidase activity and melibiose transport are completely abolished with no MelAB proteins detected (Fig. 3d); the melibiose-induced *melA* transcription is also completely prevented as shown by the RT-PCR tests (Fig. 3e).

Transcription inhibition. Activation of the *mel* operon also requires the binding of cAMP-CAP complex. To test if the production, formation, and/or function of the cAMP-CAP complex, are affected by ANK-N5C-281, cAMP was added to the MacConkey media; however, no rescue in melibiose fermentation was detected (Fig. 4a, left panel). Melibiose fermentation is observed by co-expressing MelAB under *lac* promoter of the compatible plasmid pACYC (Fig. 4a, right panel). Consistently, the pACYC-encoded, IPTG-induced melibiose transport catalyzed by MelB or lactose permease (LacY), as well as the expression of MelB, MelA, and LacY are not affected by ANK-N5C-281 (Fig. 4b). These results indicate that ANK-N5C-281 proteins neither inactivate MelA, MelB, or LacY, nor inhibit the production of cAMP or the cAMP-CAP complex activity. It is noteworthy that the cAMP-CAP complex is a global transcription activator.

It is interesting that the IPTG-induced, pACYC-encoded MelR, which is a specific transcription activator for the *melAB* operon, partially rescues the melibiose-dependent MelAB expression and activity (Fig. 4c), as well as melibiose fermentation (Fig. 4d). Purification of MelR for *in vitro* studies was exhaustively attempted and ended on failure, which was also experienced by others[29]. Based on the available functional data, it is possible that ANK-N5C-281 inhibits MelR function and prevents the transcription activation of the *melAB* operon.

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[21,24] (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 β-turn 1 of the internal repeat IV, as pointed by the arrow (Fig. 5b). Each repeat consists of two anti-parallel α-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 ~25 kDa has ~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 β-turn (particularly β-turn 1) and α-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-

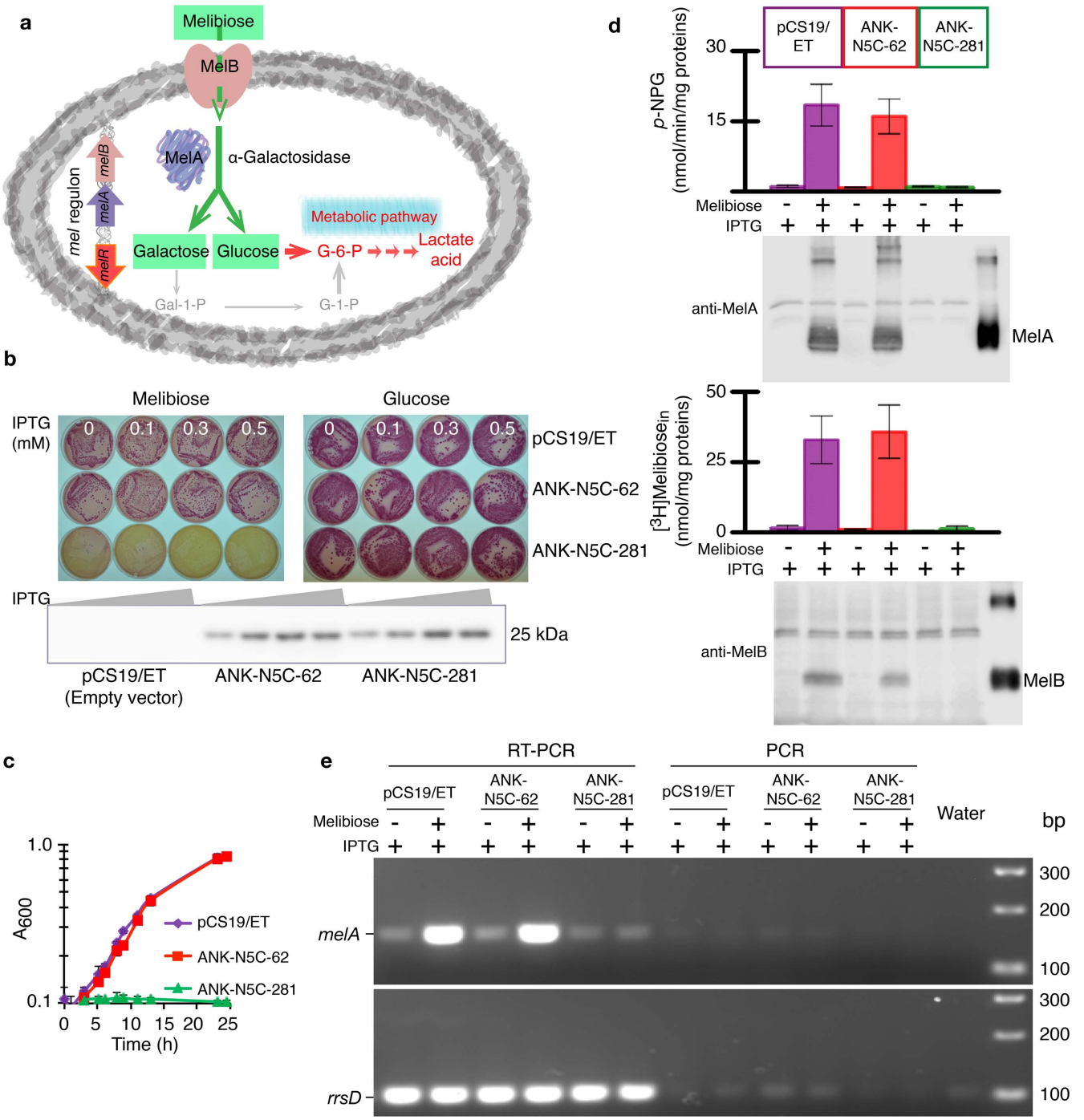

Figure 3 | Isolation of a transcription blocker by an i*n vivo* functional screen. (a), Illustration of the melibiose metabolic pathway in *E. coli*. For all the following panels, *E. coli* Tuner cells harboring pCS19/ET (vector control), pCS19/ANK-N5C-62 (protein control), or pCS19/ANK-N5C-281, respectively, 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-^{3}H]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		
E. coli One Shot® *ccd*B Survival™ 2 T1^R	F′*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lacX74 rec*A1 *ara*Δ139 Δ(*ara-leu*)*7697 gal*U *gal*K *rps*L (Str^R) *end*A1 *nupG fhu*A::*IS2*	Life Technologies
E. coli Stellar™	*F′, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ-*	Clontech
E. coli Tuner™	*lacZY* deletion mutants of BL21	Novagen
E. coli T7 express	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	NEB
E. coli DW2	*melA*$^+$Δ*melB*Δ*lacZY*	[51]
E. coli XL1- Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]*	Agilent Technologies
Plasmids		
pINITIAL	FX cloning vector with SapI and *ccdB* gene; kan^r	[38]
pCS19	pQE60 derivative inserted with gene *lacI*^{q}; amp^r	[52]
pCS19/FX	pCS19 with two SapI sites and *ccdB* gene for FX cloning; amp^r	This study
pCS19/ET	*ccdB* gene was removed from pCS19/FX; amp^r	This study
pCS19/ANK-N5C-62	Expression vector harboring gene encoding for ANK-N5C-62; amp^r	This study
pCS19/ANK-N5C-281	Expression vector harboring gene encoding for ANK-N5C-281; amp^r	This study
pCS19/ANK-N5C-281/P171Q	P171Q mutant of ANK-N5C-281	This study
pCS19/ANK-N5C-281/P171F	P171F mutant of ANK-N5C-281	This study
p7XC3H	Expression vector with SapI and *ccdB* gene for FX cloning; kan^r	[38]
p7XC3H/ANK-N5C-281	Expression vector harboring gene encoding for ANK-N5C-281; kan^r	This study
p7XC3H/ANK-N5C-317	Expression vector harboring gene encoding for ANK-N5C-317; kan^r	This study
p7XNH3	Expression vector with SapI and *ccdB* gene for FX cloning; kan^r	[38]
p7XNH3/MelA	*melA* gene of XL1 Blue subcloned into p7XNH3; kan^r	This study
pT40/MelB-6His	Expression vector containing *melB* gene; amp^r	[51]
pK95 ΔAHB	Expression vector containing *melB* gene; amp^r	[51]
pT7-5/LacY-10His	Expression vector containing *lacY* gene; amp^r	[53]
pACYC/C6 *lacY*	Expression vector containing the C-terminal half of *lacY* gene with *lac* operator/promotor in pACYC184	[54]
pACYC/FX	pACYC with two SapI sites and *ccdB* gene for FX cloning; cam^r	This study
pACYC/ET	*ccdB* gene was removed from pACYC/FX vector; cam^r	This study
pACYC-MelB	*melB* gene of Tuner strain subcloned into pACYC/FX; cam^r	This study
pACYC/MelAB	*melAB* operon of Tuner strain subcloned into pACYC/FX; cam^r	This study
pACYC/LacY	*lacY* gene of Tuner strain subcloned into pACYC/FX; cam^r	This study
pACYC/MelR	*melR* gene of Tuner strain subcloned into pACYC/FX; cam^r	This study

turbed β-turn 1 of the internal repeat IV in ANK-N5C-317 is due to the presence of Pro138, which is conformationally flexible (Fig. 5b, Supplementary Fig. S2a). Pro171 in ANK-N5C-281 may also interfere with formation of the β-turn-1 of the internal repeat V, and the resulting open β-turn constitutes a hinge loop (Figs. 5e, 6a–c, Supplementary Fig. S2b), linking the N-terminal five repeats and the C-terminal two repeats. There are multiple H-bonding interactions between the hinge loops mediated by the consensus Asp167 and Gly170 at positions-1 and -4 of the internal repeat V (Fig. 6c). Within each molecule, a specific salt-bridge interaction (Glu133/Arg199) is established between two positions-33 of internal repeats III and V, which stabilizes the open monomer (Supplementary Fig. S2b) and strengthens the H-bonding interactions between the hinge loops (Fig. 6b). It is noteworthy that Arg199 at position-33 of the internal repeat V is a consensus position; the same position in the internal repeat III is a random position, and in ANK-N5C-281, Glu133 is randomly selected. The two monomers of ANK-N5C-281 form inverted repeats with a significant expansion of the binding area (Figs. 5f) and the creation of additional potential binding areas between the two "hybrid monomers". The surface potential maps calculated from both ANK-N5C proteins reveal that their concave surfaces are negative (Fig. 5d, f). The purified proteins were analyzed by blue native-polyacrylamide gel electrophoresis (BN-PAGE), which shows that ANK-N5C-281 runs much slower than ANK-N5C-62 (Fig. 6d).

The Pro171 of ANK-N5C-281 was replaced with Gln or Phe residue, and the fermentation test shows that both mutants lose the inhibitory effect on melibiose fermentation (Fig. 6e). These data support the notion that Pro171 in ANK-N5C-281 plays a critical role in blocking the transcription activation of *melAB* operon.

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 17^{25} or 5.8×10^{30} 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 $>10^{12}$ 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 development[21], and the vaccinia virus K1 protein that is a host-range protein[31]. Their size and shape are different from the DARPins[24] 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 swapping[10,32]. For ANK-N5C-281, Pro171 at

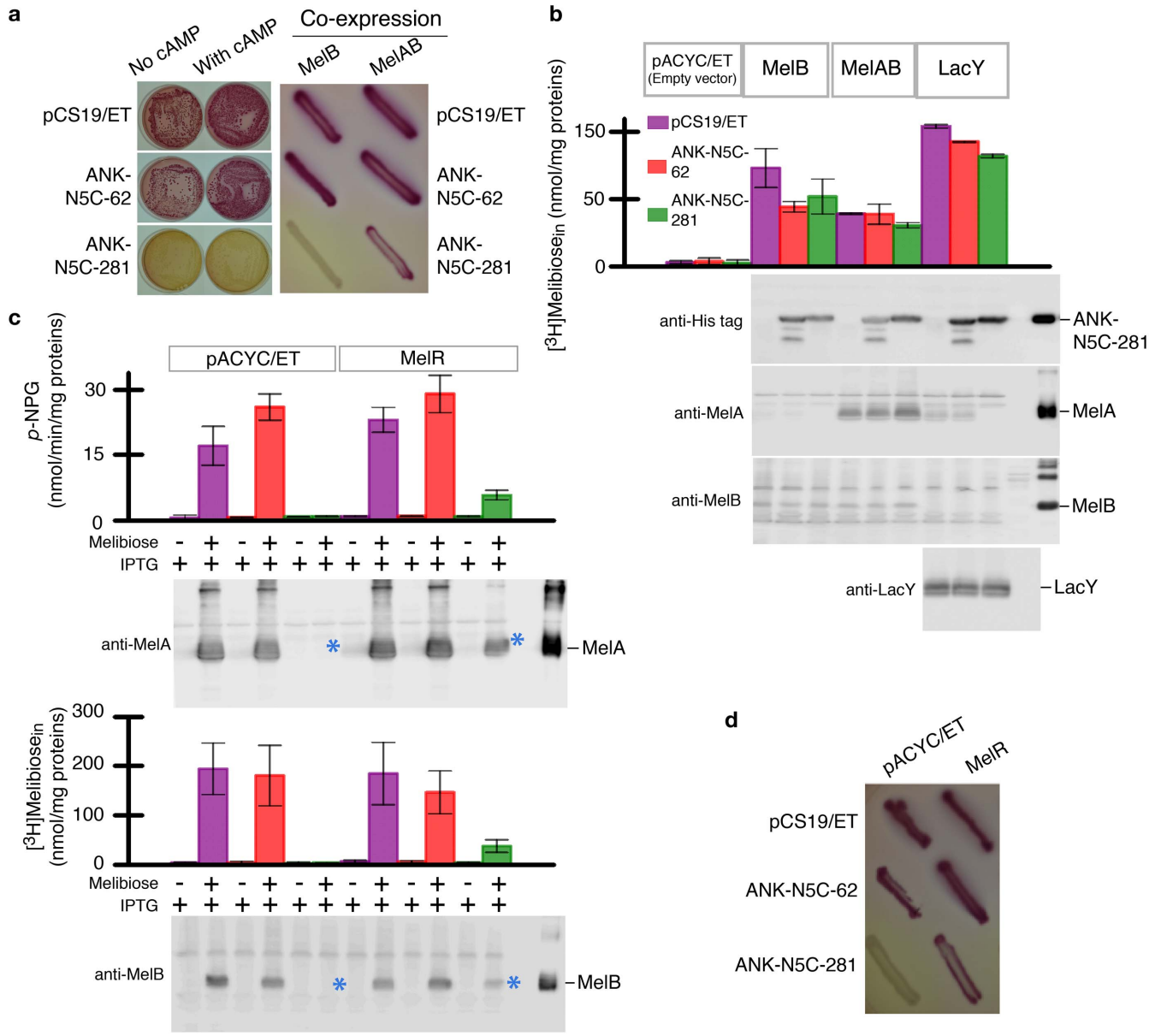

Figure 4 | Specific effect of ANK-N5C-281 on the regulation of *melAB* operon. (a), Melibiose fermentation on the MacConkey agar plates containing melibiose, IPTG, and selection antibiotics. Left panel, Tuner cells with a single plasmid (indicated in the left side) grown with ampicillin and in the absence or presence of 5 mM cAMP. Right panel, Tuner cells containing two compatible plasmids (indicated on the top and on the right side) grown in ampicillin and chloramphenicol. (b), Effect of ANK-N5C-281 on *lac* operon activity carried by a plasmid. Tuner cells harboring pCS19/ANK-N5C-281 (green bars) with compatible plasmid pACYC/ET, pACYC/MelB, pACYC/MelAB, or pACYC/LacY were used to all the studies unless otherwise described. Similar combinations were also applied for pCS19/ET (vector control, purple bars), pCS19/ANK-N5C-62 (protein control, red bars). Ampicillin at 100 μg/mL, chloramphenicol at 25 μg/mL, melibiose at 30 mM, and IPTG at 0.3 mM were used. For Western blot analysis, a total of 40 μg of cell lysates (for the detection of MelA and ANK-N5C proteins) or membranes (for the detection of MelB and LacY) were analyzed by SDS-12% PAGE. Melibiose transport (bars) and protein expression (images) were analyzed with the cells co-expressing ANK-N5C-281 (green bars) with MelB, MelAB, or LacY as indicated by the gray boxes on the top, using ANK-N5C-62 (red bars) as the control. The growth media contained ampicillin, chloramphenicol, and IPTG for inducing the protein expression from the pACYC in the absence of melibiose. Error bars, S.E.M.; n = 2–3. (c), Plasmid-encoded MelR protein partially restores chromosomal *melAB* operon activity. Cell growth conditions and legend are as described in panel b. Detection of melibiose-induced MelA and MelB activity and expression as described in the legend to Fig. 3d. Error bars, S.E.M.; n = 2–4. Blue stars point out the difference in MelA or MelB expression without or with a plasmid-encoding MelR. (d), Co-expression of MelR rescues melibiose fermentation inhibited by ANK-N5C-281.

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

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*	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	119.46, 46.80, 74.24	66.97, 94.34, 69.64
α, β, γ (°)	90.00, 97.80, 90.00	90.0, 91.51, 90.0
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)
$I/\sigma I$	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)	0.006	0.004
Bond angles (°)	1.1	0.89

*Values in parentheses are for highest-resolution shell.

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

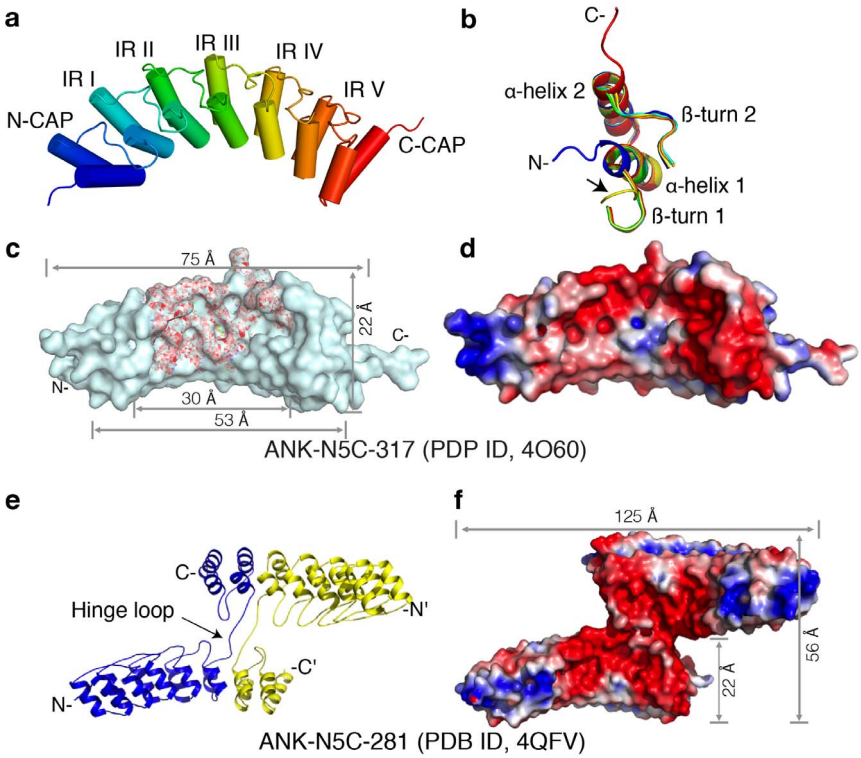

Figure 5 | X-ray crystal structures. (a–d), ANK-N5C-317; (e), (f), ANK-N5C-281. (a), Cartoon representation of the crystal structure of ANK-N5C-317 protein (PDB ID 4O60, 2.5 Å). The N- and C-terminal CAP repeats (N-CAP ad C-CAP) and five internal repeats (IR) are indicated. (b), Superposition of repeats 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-α 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.

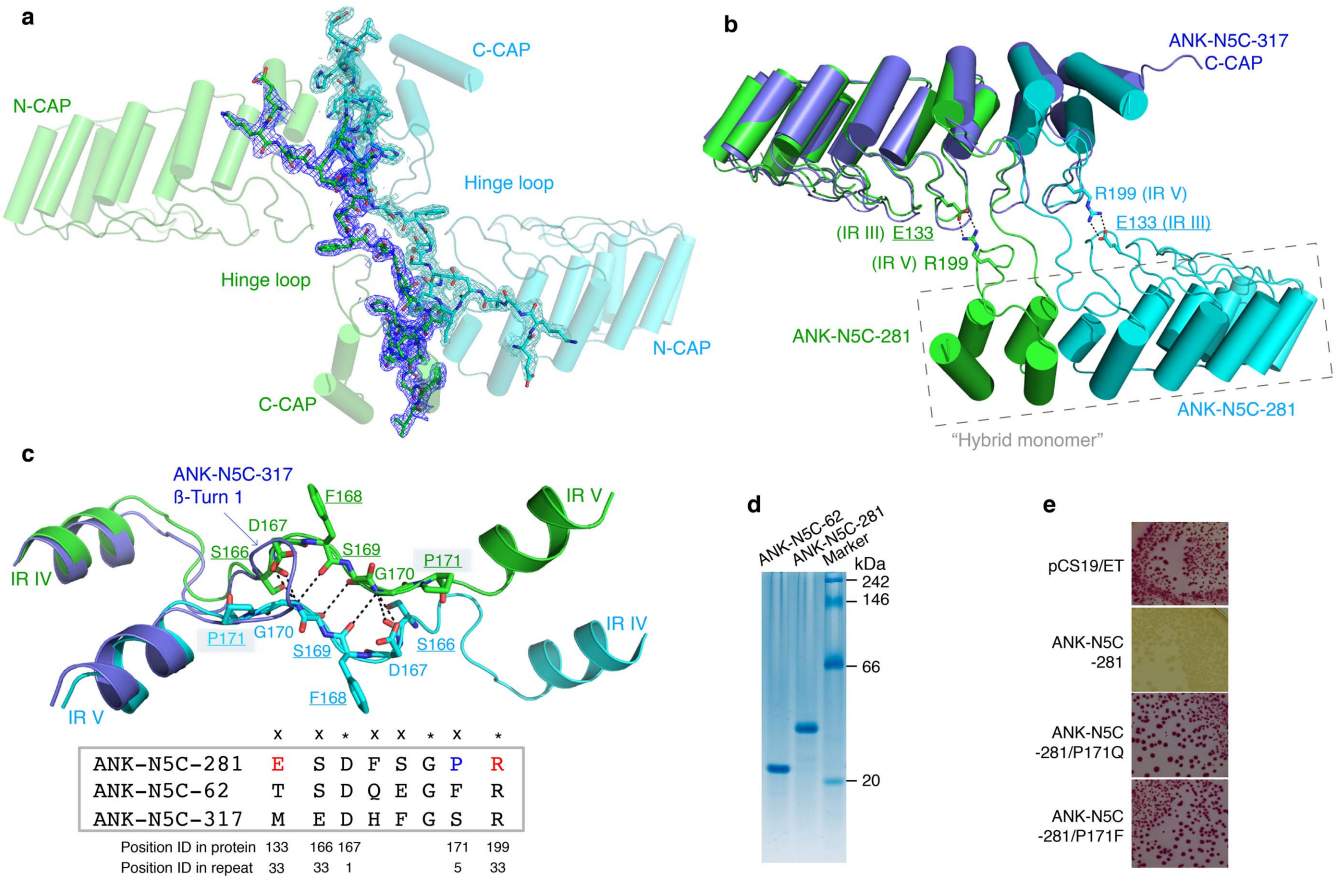

Figure 6 | Dimer interface in ANK-N5C-281. The two swapped monomers in ANK-N5C-281 crystal structure are colored in green and cyan. The structure of ANK-N5C-317 is colored in blue. (a), The 2Fo-Fc electro density map (contour at 1.2 σ) shows the hinge loop linking the N-terminal five repeats with C-terminal 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 β-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 μg each) and the NativeMark™ 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.

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 *in vitro* characterizations, such as the MelR and its homologues of the AraC/XylS family[35,36]. It is likely that such *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, *in vivo* or *in vitro*, may not necessarily produce a binder that is physiologically relevant. The drawback of this *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 library is chemically and topologically diverse. From rather small size of population (5×10^5 clones), a transcription blocker of *melAB* operon was isolated. Among all proteins involved in melibiose fermentation including melibiose uptake, hydrolysis, and glucose metabolism, the transcription of *melAB* operon appears

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 expression[29,30,36]. The current data suggest that MelR's function is inhibited by ANK-N5C-281 because overexpression 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 concave 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[25], or *in vivo* screening method, such as two-hybrid system[37] 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 are 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 information[8,9,23,24]. 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 resides (DIGKKLLEAARAGHDDSVEVLLKKGADINA). The first 18 residues were same as the previously reported N-CAP[9], and the last 13 residues mimic the framework of the internal repeat designed in this study. The C-CAP contains 29 residues (DKFGKTPFDLAIDNGNEDIAEVLQKAARS) that follow the previously optimized sequence[24] 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-center approach (Fig. 1b). The DNA oligonucleotide containing randomized positions was designed as antisense primer, and custom synthesized based on a defined codon usage: most codons constitute 7%, 6 codons encoding hydrophobic residues (Ile, Met, Leu, Val, Trp, and Phe) were 3.8–4%, and no codons were requested for helical breakers Gly and Pro or the potential disulfide former Cys. All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. Construction of plasmid libraries based on a fragment exchange cloning method (FX)[38] 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 fermentation[39–41]. In Tuner cells (*lacZ*$^-$*Y*$^-$), the *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) (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. The clones with reproducible phenotype were selected for plasmid preparation and DNA sequencing analysis. The glucose fermentation was carried out following 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-^{3}H]melibiose as described[40,42,43]. The *E. coli* cells, which were grown with 0.3 mM IPTG for plasmid-encoded protein expression and in the absence or presence of 10 mM melibiose for inducing the *melAB* operon, were washed with 100 mM KP_i (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 descriptions[44] 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*-nitrophenyl moiety as 18380 $M^{-1}cm^{-1}$[45]. 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 rRNA[46]. 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 gels. 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[47] was also used to recognize LacY protein expression. The protein A-conjugated HRP 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[41].

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

Crystallization, data collection and processing. Expression and purification of ANK-N5C-317 and ANK-N5C-281 proteins were described in the Supplementary Note. Crystallization 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 $(NH_3)_2SO_4$, 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 cryoprotectants, 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-281were collected at 100 K with an ADSC QUANTUM 315 and 315R Detector, respectively. Image data were processed with HKL 2000[48] to a resolution of 2.5 Å in *C2* space group with 99.9% completeness for ANK-N5C-317 and 2.0 Å in $P2_1$ 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[49] 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. 56 water molecules were added at the end of the refinement with the final R/R_{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 disallowed regions, 96.18% of residues are in most favored regions, 3.82% in the 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 $S^{166}DFSG^{170}$, 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/R_{free} values of 0.18/0.21 (Table 2). Out of 234 residues, 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 disallowed regions, 96.44% of residues are in most favored regions, 3.56% in the generously allowed regions. Visualization of omit maps and manual model building were performed using Coot 0.7. Surface electro-potential maps were calculated using APBS software[50]. All crystallographic figures were generated with Pymol.

1. Tamaskovic, R., Simon, M., Stefan, N., Schwill, M. & Pluckthun, 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., Stumpp, 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 DARP in library with reduced hydrophobicity. *Protein Sci* **22**, 1239–1257 (2013).
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**, 22703–22709 (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., Stumpp, M. T., Binz, H. K. & Pluckthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett* **539**, 2–6 (2003).
18. Mohler, P. J. *et al*. Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).
19. Li, J., Kline, C. F., Hund, T. J., Anderson, M. E. & Mohler, P. J. Ankyrin-B regulates Kir6.2 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. Krzywda, S. *et al*. The crystal structure of gankyrin, an oncoprotein found in complexes with cyclin-dependent kinase 4, a 19 S proteasomal ATPase regulator, and the tumor suppressors Rb and p53. *J Biol Chem* **279**, 1541–1545 (2004).
22. Binz, H. K. *et al*. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol* **22**, 575–582 (2004).
23. Kohl, A. *et al*. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci U S A* **100**, 1700–1705 (2003).
24. Kramer, M. A., Wetzel, S. K., Pluckthun, A., Mittl, P. R. & Grutter, M. G. Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module. *J Mol Biol* **404**, 381–391 (2010).
25. Zahnd, C., Amstutz, P. & Pluckthun, A. Ribosome display: selecting and evolving proteins *in vitro* that specifically bind to a target. *Nat Methods* **4**, 269–279 (2007).
26. Lipovsek, D. & Pluckthun, A. In-vitro protein evolution by ribosome display and mRNA display. *J Immunol Methods* **290**, 51–67 (2004).
27. Sennhauser, G. & Grutter, M. G. Chaperone-assisted crystallography with DARPins. *Structure* **16**, 1443–1453 (2008).
28. Epa, V. C. *et al*. Structural model for the interaction of a designed Ankyrin Repeat Protein with the human epidermal growth factor receptor 2. *PLoS One* **8**, e59163 (2013).
29. Kahramanoglou, C., Webster, C. L., El-Robh, M. S., Belyaeva, T. A. & Busby, S. J. Mutational analysis of the *Escherichia coli melR* gene suggests a two-state concerted model to explain transcriptional activation and repression in the melibiose operon. *J Bacteriol* **188**, 3199–3207 (2006).
30. Elrobh, M. S., Webster, C. L., Samarasinghe, S., Durose, D. & Busby, S. J. Two DNA sites for MelR in the same orientation are sufficient for optimal MelR-dependent repression at the *Escherichia coli melR* promoter. *FEMS Microbiol Lett* **338**, 62–67 (2013).
31. Li, Y., Meng, X., Xiang, Y. & Deng, J. Structure function studies of vaccinia virus host range protein k1 reveal a novel functional surface for ankyrin repeat proteins. *J Virol* **84**, 3331–3338 (2010).
32. Ferreiro, D. U., Cho, S. S., Komives, E. A. & Wolynes, P. G. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *J Mol Biol* **354**, 679–692 (2005).
33. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* **4**, 2455–2468 (1995).
34. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci* **11**, 1285–1299 (2002).
35. Bourgerie, S. J., Michan, C. M., Thomas, M. S., Busby, S. J. & Hyde, E. I. DNA binding and DNA bending by the MelR transcription activator protein from *Escherichia coli*. *Nucleic Acids Res* **25**, 1685–1693 (1997).
36. Grainger, D. C., Belyaeva, T. A., Lee, D. J., Hyde, E. I. & Busby, S. J. Binding of the *Escherichia coli* MelR protein to the *melAB* promoter: orientation of MelR subunits and investigation of MelR-DNA contacts. *Mol Microbiol* **48**, 335–348 (2003).
37. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* **95**, 5752–5756 (1998).
38. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–3278 (2011).
39. Guan, L., Jakkula, S. V., Hodkoff, A. A. & Su, Y. Role of Gly117 in the cation/melibiose symport of MelB of *Salmonella typhimurium*. *Biochemistry* **51**, 2950–2957 (2012).
40. Amin, A., Ethayathulla, A. S. & Guan, L. Suppression of conformation-compromised mutants of *Salmonella enterica* serovar Typhimurium MelB. *J Bacteriol* **196**, 3134–3139 (2014).
41. Jakkula, S. V. & Guan, L. Reduced Na^+ affinity increases turnover of *Salmonella enterica* serovar Typhimurium MelB. *J Bacteriol* **194**, 5538–5544 (2012).
42. Guan, L., Nurva, S. & Ankeshwarapu, S. P. Mechanism of melibiose/cation symport of the melibiose permease of *Salmonella typhimurium*. *J Biol Chem* **286**, 6367–6374 (2011).
43. Ethayathulla, A. S. *et al*. Structure-based mechanism for Na(+)/melibiose symport by MelB. *Nat Commun* **5**, 3009 (2014).
44. Chakladar, S., Cheng, L., Choi, M., Liu, J. & Bennet, A. J. Mechanistic evaluation of MelA alpha-galactosidase from *Citrobacter freundii*: a family 4 glycosyl hydrolase in which oxidation is rate-limiting. *Biochemistry* **50**, 4298–4308 (2011).
45. Bowers, G. N., Jr., McComb, R. B., Christensen, R. G. & Schaffer, R. High-purity 4-nitrophenol: purification, characterization, and specifications for use as a spectrophotometric reference material. *Clin Chem* **26**, 724–729 (1980).
46. Gao, D. *et al*. Eha, a transcriptional regulator of hemolytic activity of *Edwardsiella tarda*. *FEMS Microbiol Lett* **353**, 132–140 (2014).
47. 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).
48. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Method in Enzymol* **276**, 307–326 (1997).
49. McCoy, A. J. *et al*. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
50. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037–10041 (2001).
51. Pourcher, T., Leclercq, S., Brandolin, G. & Leblanc, G. Melibiose permease of *Escherichia coli*: large scale purification and evidence that H^+, Na^+, and Li^+ sugar symport is catalyzed by a single polypeptide. *Biochemistry* **34**, 4412–4420 (1995).
52. Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339–347 (1999).
53. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. Structural determination of wild-type lactose permease. *Proc Natl Acad Sci U S A* **104**, 15294–15298 (2007).
54. 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* **87**, 4325–4329 (1990).

Acknowledgments

We thank Eric R. Geertsma and Raimund Dutzler for their FX cloning tools. We thank Michael Ehrmann for the gift of plasmid pCS19; Ronald Kaback for the plasmids pACYC/C6 *lacY*, pT7-5/WT LacY 10His, and LacY 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 assembly strategy. L.G. and Y.S. established the protocols for library construction. Y.S. constructed the pCS19/FX vector and all plasmid libraries, and identified the screening conditions. E.B.T. performed all functional studies including primary screening tests, protein expression and activity assays, and RT-PCR. L.G., Y.S., and E.B.T. analyzed the DNA-sequence results. S.X. calculated entropy scores, Hamming distances, and amino-acid usage, and L.G. analyzed the results. H.P. and E.B.T. purified proteins. A.S.E. performed crystallization, X-ray diffraction, data collection and processing, and determined the structure. L.G. and A.S.E. analyzed the structures. All authors contributed to manuscript preparation. L.G. wrote the manuscript.

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 KF981639 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).