microPublication BIOLOGY

11/19/2026 - *Open Access*

Bacterial Growth Inhibition Screen (BGIS): an unconventional phenotype-based workflow for mapping regions in human dark proteins associated with reduced growth in *E. coli.*

Songhe Jin[1], Chutian Chen [1], Yewei Liu [2§*], Ferdinand Kappes[3§*]

[1]Duke Kunshan University, Kunshan, JS, China

[2]Department of Biosciences and Bioinformatics, Xi'an Jiaotong-Liverpool University, Wuzhong District, Suzhou, China

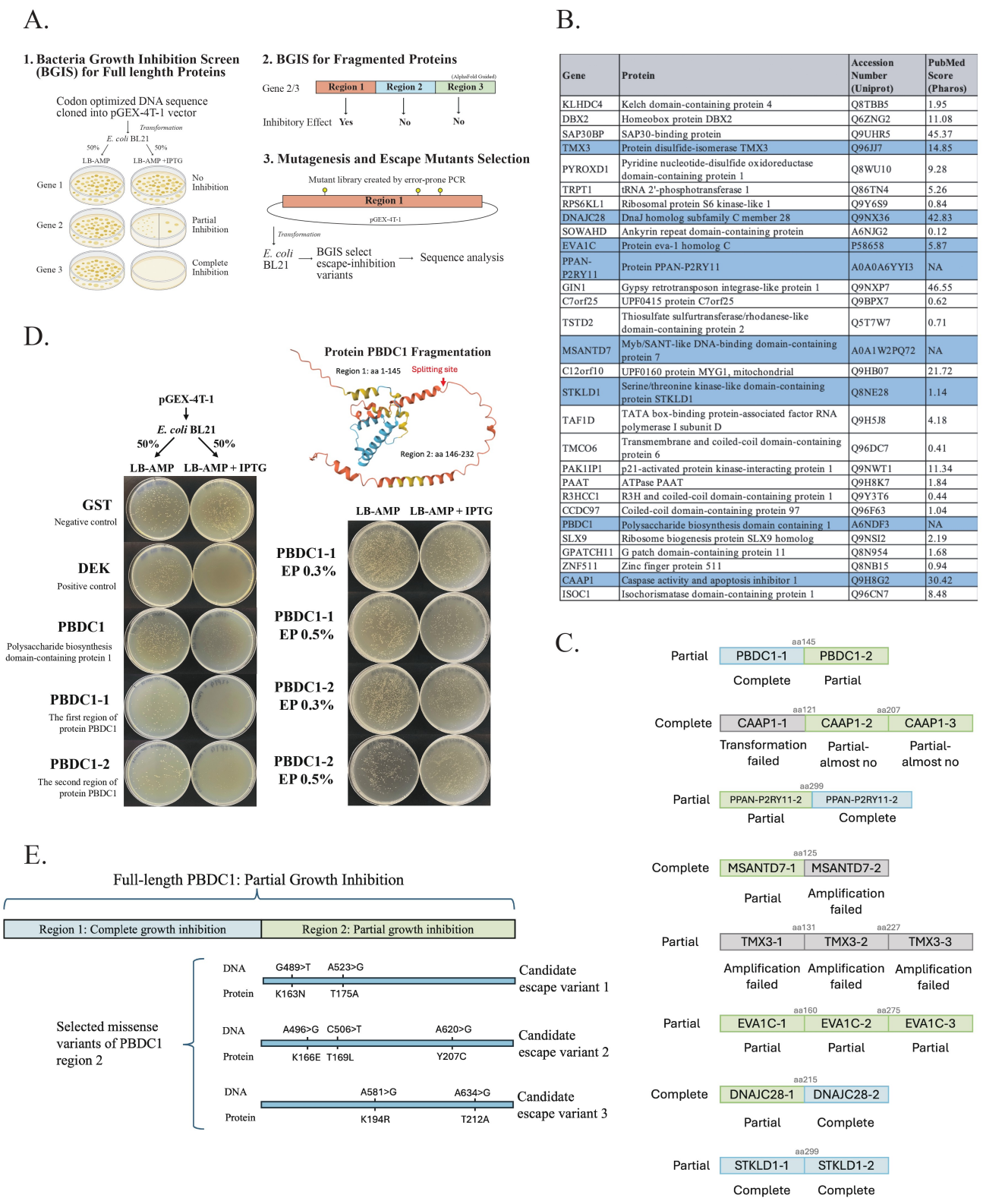
[3]Division of Natural and Applied Sciences, Duke Kunshan University, Kunshan, JS, China

[§]To whom correspondence should be addressed: yewei.liu@dukekunshan.edu.cn; Ferdinand.Kappes@dukekunshan.edu.cn

[*]These authors contributed equally.

Abstract

The human proteome holds immense therapeutic potential, yet many "dark" proteins remain functionally uncharacterized. We apply the Bacterial Growth Inhibition Screen (BGIS), previously validated on the DEK oncoprotein, to 29 dark human proteins. Through a three-phase screening pipeline comprising an initial growth-inhibition screen, AlphaFold-guided fragmentation, and error-prone PCR mutagenesis, BGIS identified regions associated with a growth-inhibitory phenotype in 8 of 29 proteins and, as exemplified by PBDC1, yielded candidate growth-escape mutants. This proof-of-concept study establishes BGIS as a scalable, cost-effective framework for preliminary hypothesis-independent, phenotype-level characterization of dark proteins, providing an additional experimental entry point for dark proteome research.



Gene	Protein	Accession Number (Uniprot)	PubMed Score (Pharos)
KLHDC4	Kelch domain-containing protein 4	Q8TBB5	1.95
DBX2	Homeobox protein DBX2	Q6ZNG2	11.08
SAP30BP	SAP30-binding protein	Q9UHR5	45.37
TMX3	Protein disulfide-isomerase TMX3	Q96JJ7	14.85
PYROXD1	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 1	Q8WU10	9.28
TRPT1	tRNA 2'-phosphotransferase 1	Q86TN4	5.26
RPS6KL1	Ribosomal protein S6 kinase-like 1	Q9Y6S9	0.84
DNAJC28	DnaJ homolog subfamily C member 28	Q9NX36	42.83
SOWAHD	Ankyrin repeat domain-containing protein	A6NJG2	0.12
EVA1C	Protein eva-1 homolog C	P58658	5.87
PPAN-P2RY11	Protein PPAN-P2RY11	A0A0A6YYI3	NA
GIN1	Gypsy retrotransposon integrase-like protein 1	Q9NXP7	46.55
C7orf25	UPF0415 protein C7orf25	Q9BPX7	0.62
TSTD2	Thiosulfate sulfurtransferase/rhodanese-like domain-containing protein 2	Q5T7W7	0.71
MSANTD7	Myb/SANT-like DNA-binding domain-containing protein 7	A0A1W2PQ72	NA
C12orf10	UPF0160 protein MYG1, mitochondrial	Q9HB07	21.72
STKLD1	Serine/threonine kinase-like domain-containing protein STKLD1	Q8NE28	1.14
TAF1D	TATA box-binding protein-associated factor RNA polymerase I subunit D	Q9H5J8	4.18
TMCO6	Transmembrane and coiled-coil domain-containing protein 6	Q96DC7	0.41
PAK1IP1	p21-activated protein kinase-interacting protein 1	Q9NWT1	11.34
PAAT	ATPase PAAT	Q9H8K7	1.84
R3HCC1	R3H and coiled-coil domain-containing protein 1	Q9Y3T6	0.44
CCDC97	Coiled-coil domain-containing protein 97	Q96F63	1.04
PBDC1	Polysaccharide biosynthesis domain containing 1	A6NDF3	NA
SLX9	Ribosome biogenesis protein SLX9 homolog	Q9NSI2	2.19
GPATCH11	G patch domain-containing protein 11	Q8N954	1.68
ZNF511	Zinc finger protein 511	Q8NB15	0.94
CAAP1	Caspase activity and apoptosis inhibitor 1	Q9H8G2	30.42
ISOC1	Isochorismatase domain-containing protein 1	Q96CN7	8.48

Figure 1. Bacterial Growth Inhibition Screen (BGIS) of 29 human dark proteins identifies regions associated with a reduced growth phenotype in *E. coli* and growth-escape mutants in PBDC1:

(A) BGIS workflow. Phase 1, Initial BGIS: full-length dark proteins are expressed as GST fusions in *E. coli* BL21(DE3) and screened for the growth-inhibitory phenotype (GIP) upon IPTG induction. Phase 2, Domain Specific BGIS: inhibitory proteins are fragmented along AlphaFold2-predicted structural boundaries and re-screened to localize the region associated with the GIP. Phase 3, Growth-escape mutant screen: the fragment producing the GIP is mutagenized by error-prone PCR and screened for variants capable of escape. **(B) Summary table of all 29 dark proteins screened in Phase 1.**

PubMed scores are sourced from the Pharos/TCRD database, reflecting the limited literature representation of these dark protein candidates. Detailed selection criteria are described in the Methods section. Proteins whose induced expression produces a complete or partial GIP are highlighted in blue, whereas; proteins that do not produce a GIP are not highlighted. Full plate images are shown in Extended Data Fig. 1A**. (C) Summary of BGIS results for all proteins producing a growth-inhibitory phenotype and their fragments from Phases 1 and 2.** A complete GIP (marked as "complete") indicates no colony formation on +IPTG plates; a partial GIP (marked as "partial") indicates significantly reduced colony size on +IPTG plates relative to −IPTG controls. Full plate images are shown in Extended Data Fig. 1B. **(D-E) Proof-of-concept demonstration using PBDC1. (D) BGIS results of PBDC1, its fragmentations, and error-prone PCR mutagenesis.** GST empty vector and GST-DEK fusion protein served as negative and positive controls respectively. Colony plate images show BGIS results for full-length PBDC1 and its AlphaFold-guided fragments (Region 1 and Region 2), where IPTG induction drives expression of the corresponding GST-fusion protein cloned into pGEX-4T-1. Error-prone PCR (EP) was performed at two $MnCl_2$ concentrations 0.3% (w/v) (EP 0.3%) and 0.5% (w/v) (EP 0.5%), the resulting mutant libraries were transformed into *E. coli* BL21(DE3). Large colonies emerging on +IPTG plates represent candidate growth-escape mutants that no longer produce the growth-inhibitory phenotype and were selected for sequencing. **(E) Candidate growth-escape mutants in PBDC1 Region 2.** After excluding sequences containing indels or nonsense mutations, remaining sequences were further filtered to retain only those harboring missense mutations, discarding synonymous substitutions. The resulting nucleotide changes and corresponding amino acid substitutions are indicated. Three candidate growth-escape mutants were identified in PBDC1 Region 2. These variants have not been re-tested, and their relevance to the native biological function of PBDC1 has not been established.

Description

Despite the Human Genome Project's identification of ~20,000 protein-coding genes (Nurk et al., 2022) and large-scale efforts such as the HUPO Human Proteome Project (Omenn et al., 2024), a substantial portion of the human proteome remains functionally unannotated, the so-called "dark proteins", whose activities and interactions have yet to be experimentally elucidated (Cao et al., 2024).

While recent studies highlighted that many “dark” proteins indeed execute essential functions in organisms (Moitra & Larrouy-Maumus, 2026; Rocha et al., 2023), research on “dark” proteins remains limited, often due to a lack of molecular tools and the labor-intensive nature of the work with uncertain rewards (Sharifi Tabar et al., 2022). In addition, despite structural predictions by AlphaFold, experimentally deciphering molecular function at the domain level in dark proteins presents a methodological challenge. A critical step in such studies is the generation of mutants that reduce or abolish a defined protein activity, which in turn allows that activity to be linked causally to a cellular phenotype, rather than relying solely on gene-level knockdown or knockout approaches that abolish all protein functions simultaneously (Guo, Prell, et al., 2021). Conventional site-directed mutagenesis is the standard strategy for generating such mutants; however, it requires detailed prior knowledge of protein structure and functionally important residues (Brannigan & Wilkinson, 2002). For dark proteins, this information is typically unavailable.

Therefore, for dark proteins lacking any such prior information, an unbiased, hypothesis-independent approach to identifying the regions of a protein that produce a measurable phenotype, and to generating variants that escape that phenotype, is required. To address this challenge, Guo et al. (2021) established the Bacterial Growth Inhibition Screen (BGIS) for the study of proteins. Using this strategy, we previously identified a mutant of the DEK oncoprotein that no longer produces a growth-inhibitory phenotype; for DEK, this phenotype has been linked to direct DNA compaction. These results demonstrate that, in a protein for which an independent activity assay exists, BGIS can point to residues relevant to a defined molecular activity in a fully hypothesis-independent manner (Guo, Prell, et al., 2021; Guo, Xu, et al., 2021).

However, BGIS has only been validated on protein (DEK) with known functions, and its applicability to dark proteins, where no functional or structural reference exists, remains unexplored. One barrier lies in domain fragmentation: localizing the growth-inhibitory phenotype to a specific region of a full-length protein requires rational boundaries for fragmentation, information that is unavailable for dark proteins. The release of AlphaFold2, which provides relatively high-confidence structural predictions for protein sequences, offers a solution to this problem by enabling structure-guided fragmentation without experimental structural data. In this study, we integrate BGIS with AlphaFold2-predicted domain boundaries to systematically screen human dark proteins for the growth-inhibitory phenotype, to localize the regions associated with it, and to generate candidate growth-escape mutants, establishing a proof-of-concept workflow for hypothesis-independent, phenotype-level characterization of the dark proteome.

Throughout this manuscript we use “phenotype” in a strict, operational sense. Heterologous overexpression of a non-native protein in *E. coli* can reduce growth for reasons that are unrelated to the protein's native biological function, including protein misfolding, aggregation, or other effects. We therefore do not directly equate reduced growth with molecular activity, and we do not infer native function from these experiments. Accordingly, we use “growth-inhibitory phenotype (GIP)” for reduced colony formation or reduced colony size upon IPTG-induced expression from the plasmid, “GIP-associated region” for a fragment that still produces the GIP when expressed, and “growth-escape mutant (GEM)”

for a variant that no longer produces such a phenotype. None of these terms implies that the corresponding region, residue, or substitution is required for the native biological function of the human protein; establishing that would require an independent activity assay and validation in a eukaryotic context.

To test if human "dark proteins" produce a GIP in *E. coli*, we screened 29 human dark proteins through a three-phase BGIS pipeline **(Fig. 1A)**. In Phase 1, each protein was expressed as a GST fusion in *E. coli* BL21(DE3) and assessed for growth inhibition upon IPTG induction, using GST alone as a negative control and the DEK oncoprotein as a positive control (Guo, Prell, et al., 2021). Growth inhibition was defined as either complete absence of colony formation or a significant reduction in colony size on inducing plates relative to non-inducing controls (Fig. 1D), identifying candidates that produce a detectable growth-inhibitory phenotype (GIP). Proteins that produced the growth-inhibitory phenotype were carried into Phase 2, where AlphaFold2-predicted structural boundaries guided the design of protein fragments **(Fig. 1D)**, allowing the growth-inhibitory phenotype to be localized to a defined region without relying on experimental structural data. In Phase 3, the inhibitory domain was subjected to error-prone PCR mutagenesis to generate random variant libraries; colonies that escaped growth inhibition were then selected as candidate growth-escape mutants, with confirmed missense substitutions identified by sequencing.

Of the 29 dark proteins screened in Phase 1, eight exhibited complete or partial growth inhibition upon IPTG induction **(Fig. 1B)**, indicating that while BGIS is not universally applicable, a meaningful fraction of dark proteins produces a growth-inhibitory phenotype detectable by this approach. In Phase 2, AlphaFold2-guided fragmentation localized the growth-inhibitory phenotype to a defined region in 6 of these proteins (**Fig. 1C**), supporting the utility of structure-predicted boundaries as a fragmentation strategy. To demonstrate the complete three-phase workflow, PBDC1 was taken through Phase 3 as a proof-of-concept example **(Fig. 1D–1E)**. Phase 2 fragmentation revealed that both region 1 and region 2 of PBDC1 independently retained the growth-inhibitory phenotype, with region 1 producing complete inhibition and Region 2 producing partial inhibition **(Fig. 1D)**. Error-prone PCR mutagenesis was performed on both regions; however, only region 2 yielded candidate growth-escape mutants after sequencing-based filtering to exclude nonsense and frameshift variants. Three variants carrying missense substitutions were identified in region 2, each carrying 2-3 amino acid substitutions, identifying residues that are associated with the growth-inhibitory phenotype in PBDC1 (**Fig. 1E**). These variants are candidates only: they have not been re-tested, and their relevance to the native biological function of PBDC1 has not been established.

Together, the screening of 29 human dark proteins, including phenotype-level mapping of the regions responsible for reduced growth in 6 of them, offers a preliminary reference that may serve as an entry point for future investigations into these uncharacterized proteins. For PBDC1 specifically, Phase 2 fragmentation and Phase 3 mutagenesis further yielded three candidate growth-escape variants in region 2, identifying residues associated with the phenotype and providing starting points for future mechanistic investigation. Moreover, the successful application of BGIS to human dark proteins serves as a proof-of-concept of a fast, low-cost workflow that requires no prior functional assumptions. While not universally applicable, this approach offers a practical methodology for researchers encountering bottlenecks in the phenotypic characterization of uncharacterized proteins and in the generation of variants that escape a given phenotype.

Several caveats and limitations should be noted. First, the mechanism underlying growth inhibition has not been experimentally verified for each dark protein screened; unlike DEK, whose toxicity was linked to direct DNA compaction (Guo, Prell, et al., 2021). Thus, the basis of inhibition in other candidates remains unknown, and growth inhibition alone cannot be taken as evidence of any particular molecular activity, nor of the protein's native biological function, in a eukaryotic context. Second, the AlphaFold2-guided fragmentation strategy relies on empirical identification of structural boundaries based on the visual clustering of secondary structure elements, which is subjective and may not correspond to genuine structural or functional boundaries. The reliability of this approach is further constrained by regional variations in AlphaFold2 prediction confidence (pLDDT scores), particularly in intrinsically disordered regions where boundary assignments are less certain. Third, the screen covers 29 proteins, and conclusions regarding the broader applicability of BGIS to the dark proteome should be interpreted with caution. Fourth, the error-prone PCR mutagenesis condition requires further optimization. The $MnCl_2$ concentration, the primary determinant of mutation load, proved difficult to control in practice: high concentrations introduced multiple substitutions per clone rather than the ideal single missense mutation, making it laborious to determine the key substitution responsible for escape from the phenotype. On the other hand, concentrations that were low failed to yield enough validated missense mutants because of a high proportion of nonsense mutations; a limitation particularly pronounced for Region 1 of PBDC1.

Finally, the candidate growth-escape mutants identified here have not been subjected to reversion mutagenesis, nor have the confirmed substitutions been reintroduced into full-length PBDC1 and re-expressed in *E. coli* to verify their effect in the intact protein context. As all validations were performed solely at the bacterial level, whether these substitutions affect the native biological function of PBDC1 in a eukaryotic context also remains to be determined. Future studies could perform reversion mutagenesis to define the minimal set of substitutions required to escape the phenotype, reintroducing confirmed mutations into full-length PBDC1 for bacterial re-expression, and ultimately characterizing these proteins and their growth-escape variants in human cells to obtain eukaryotic context information.

Methods

Protein Selection and Plasmid Construction

Candidate proteins were selected using the Target Development Level (TDL) classification in Pharos/TCRD (Kelleher et al., 2023), focusing on proteins categorized as Tdark or low-confidence Tbio; tiers representing the least characterized fraction of the human proteome. Candidates were further filtered by two criteria: (i) the presence of at least one structurally discrete domain, as assessed by visual inspection of AlphaFold-predicted models (Jumper et al., 2021), and (ii) a Pharos PubMed Score below 50 — a conservative threshold well below well-studied and early-stage research targets (e.g., EGFR: 19,716; LRRK2: 2,119; DEK: 370) - indicating minimal direct representation in the biomedical literature **(Fig. 1B)**.

Additionally, three proteins absent from the Pharos/TCRD database were included as supplementary candidates to broaden the coverage of understudied targets. A0A1W2PQ72 (MSANTD7) is a reviewed Swiss-Prot entry with a UniProt annotation score of 2/5, reflecting near-absent experimental characterization. A6NDF3 (PBDC1) is an unreviewed TrEMBL entry with an annotation score of 1/5, but notably carries "Evidence at protein level", confirming experimental detection while remaining entirely without functional annotation. A0A0A6YYI3 (PPAN-P2RY11) is unreviewed with an annotation score of 1/5 and protein existence classified as "Predicted", indicating no experimental evidence for its expression or function.

Coding sequences were codon-optimized for *E. coli* BL21 (DE3) expression using the IDT Codon Optimization Tool and synthesized by GenScript. Each fragment was cloned into pGEX-4T-1 (GenScript) between the EcoRI and XhoI restriction sites, enabling IPTG-inducible expression as an N-terminal GST fusion protein. Correct insertion was confirmed by Sanger sequencing prior to use.

Bacterial Growth Inhibition Screen (BGIS)

LB agar was prepared at 40 g/L, autoclaved at 121 °C for 20 min, and supplemented with ampicillin (100 µg/mL) at ~50 °C. For BGIS plates, IPTG was additionally added to 0.1 mM at ~40 °C. Plates were dried in a biosafety cabinet and stored at 4 °C until use.

All plasmids were diluted to 10 ng/µL in nuclease-free water. Chemically competent *E. coli* BL21(DE3) cells (50 µL) were incubated with 1 µL plasmid on ice for 20 min, heat-shocked at 42 °C for 45 s, returned to ice for 5 min, and recovered in 70 µL LB for 5 min. Equal volumes (50 µL) of transformed culture were spread onto paired LB-AMP plates with and without 0.1 mM IPTG and incubated at 37 °C overnight. Colony formation was compared between conditions and proteins were classified as producing "complete inhibition," "partial inhibition," or "no inhibition" of colony formation upon induced expression. This classification is purely phenotypic: it reports growth on solid medium and does not identify a molecular activity of the expressed protein. The empty pGEX-4T-1 vector (GST alone) served as the negative control; a GST-DEK fusion protein served as the positive control (Guo et al., 2021). All BGIS analyses were carried out in triplicate, and the results were reproducible across all replicates.

Protein Expression and Western Blotting (Extended Data Fig. 2)

Single colonies were inoculated into 5 mL LB-AMP and grown overnight at 37 °C. The following day, 1 mL of overnight culture was diluted into 4 mL LB-AMP and incubated at 37 °C for 2 h. The culture was divided into two 1.5 mL aliquots; one was supplemented with 0.1 mM IPTG and both were incubated for a further 1.5 h at 37 °C.

Cells were pelleted at 8,000 rpm (6,080 × g) for 1 min, resuspended in 100 µL 2% SDS, 40% power ultrasonic disrupted (SCIENTZ Ultrasonic Homogenizer JY96-IIN) for 3-5 seconds until the solution lost its viscosity. 12 µL of lysate was combined with 3 µL 5× Laemmli Sample buffer, heated at 95 °C for 5 min, and resolved by SDS-PAGE on 10% polyacrylamide gels at 80 V (stacking) then 120 V (resolving). Proteins were transferred to PVDF membrane by semi-dry transfer at 100 mA for 75 min. After blocking with 5% skim milk in TBST, membranes were probed overnight at 4 °C with an anti-GST rabbit monoclonal antibody (1:1,000; Beyotime), washed three times in TBST, then incubated with HRP-conjugated goat anti-rabbit IgG (1:10,000; Beyotime) for 60 min at room temperature. Signal was detected by chemiluminescence (Cytiva, ImageQuant 800) and images were quantified using ImageJ.

AlphaFold-Guided Protein Fragmentation and Subcloning

Proteins that produced the growth-inhibitory phenotype were fragmented for phenotype-level mapping. Boundaries were determined from AlphaFold2-predicted structures and per-residue confidence (pLDDT) scores (https://alphafold.ebi.ac.uk/): domain interfaces were used where clear tertiary boundaries existed, and transitions from low (yellow) to very low (orange) pLDDT were used otherwise. Fragments were designed to span 80–200 amino acids.

Primers were designed with target T_m of 51–56 °C and GC content of 40–60%, adding restriction enzyme cleavage site,and verified by BLAST against the constructed full length protein's plasmid sequence. PCR was performed in 50 µL reaction system (Reagent.Table 3) with the following cycling conditions: 95 °C for 3 min; 33 cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 1min; final extension at 72 °C for 5 min; hold at 12 °C. Products were resolved on 1% agarose gels

containing Gel Red Nucleic Acid (Adamas Life) at 120 V for 30 min, and bands of correct size were excised and purified using the QIAGEN QIAquick Gel Extraction Kit (eluted in 30 µL nuclease-free water).

Gel-purified fragments and the pGEX-4T-1 vector were digested with EcoRI and XhoI (Reagent. Table 4). For inserts containing an internal EcoRI site (e.g., protein 159), SalI was substituted for EcoRI. Digested products were column-purified prior to ligation. Ligation was carried out in a 20 µL reaction containing insertion sequence and linearized vector at a 3:1 molar ratio, 1× T4 DNA Ligase buffer, and T4 DNA Ligase (Beyotime) (Reagent. Table 5), incubated at 25 °C for 2 h or 16 °C overnight. 10 µL of ligation product was transformed into DH5α competent cells using the same heat-shock protocol described above and plated on LB-AMP agar. Colonies were screened by Sanger sequencing (Sangon Biotech, Suzhou), and verified constructs were propagated in DH5α, extracted by miniprep (QI Aprep Spin Miniprep Kit), and retransformed into BL21(DE3) for secondary BGIS.

Error-Prone PCR Mutagenesis

To generate libraries of candidate growth-escape mutants, error-prone PCR was applied to the coding sequence of each fragment that produced the growth-inhibitory phenotype, introducing random mutations. Two $MnCl_2$ concentrations were used to modulate mutation frequency (Reagent. Table 6); all other conditions were identical to standard PCR. Amplified mutant mixture were gel-purified, digested, ligated into pGEX-4T-1, and transformed into BL21(DE3) as described above. Transformants were plated on LB-AMP-IPTG plates and incubated overnight at 37 °C.

Because all selected proteins either completely or partially inhibit *E. coli* growth upon induction, colonies on +IPTG plates that carried mutated variants were visibly larger than background microcolonies, indicative of a variant that no longer produces the growth-inhibitory phenotype upon induced expression. They were designated candidate growth-escape mutants (GEMs).

Sanger Sequencing and Identification of Growth-Escape Mutants

Large colonies from +IPTG plates were picked and their plasmids extracted for Sanger sequencing (Sangon Biotech, Suzhou). Sequences were aligned to the wild-type reference using SnapGene to identify all nucleotide changes relative to the wild-type coding sequence.

Candidate sequences were then filtered using the following criteria to isolate informative growth-escape variants. First, sequences containing any insertion or deletion were excluded, as frameshift mutations would globally disrupt protein structure and preclude attribution of the phenotype to particular residues. Second, sequences harboring premature stop codons (nonsense mutations) were excluded for the same reason—truncated proteins cannot be used to reliably attribute escape from the phenotype to a specific region or residue. Third, synonymous substitutions—single-nucleotide changes that do not alter the encoded amino acid (silent mutations)—were also excluded, as they are unlikely to contribute to escape from the observed growth-inhibitory phenotype. Only sequences carrying non-synonymous single-nucleotide substitutions (missense mutations) were retained as candidate growth-escape variants.

Reagents

Table 1. Bacterial Strains

Strains	**Supplier**
E. coli BL21(DE3)pLysS	AngyuBio
E. coli DH5α	AngyuBio

Table 2. Plasmids

Plasmids	**Supplier**
pGEX-4T-1	Genescript

Table 3. PCR reaction system (50 µL total volume).

Component	Volume (µL)	Final Concentration
Nuclease-free water	to 50	—
10× BeyoTaq Buffer with Mg^{2+} (Beyotime, #D7020-2)	5	1×
dNTP mixture (25mM) (Beyotime, #D7371)	0.4	0.2 mM
Primer Forward (10 µM)	2	0.4 µM
Primer Reverse (10 µM)	2	0.4 µM
Template plasmid	—	[20 ng]
Beyo Taq DNA polymerase (5U/µL) (Beyotime, #D7219-1)	0.25	[1.25 U]

Table 4. Restriction endonuclease digestion system (50 µL total volume).

Component	Volume (µL)
Nuclease-free water	to 50
10× CutSmart buffer (New England Biolabs, #B6004S)	5
PCR product or vector	30
Restriction endonuclease (XhoI: NEB, #R0146V; EcoRI: NEB, #R3101V)	1 each

Table 5. T4 Ligation System.

Component	Volume (µL)
T4 DNA Ligase Buffer (NEB, #B0202S)	2
T4 DNA Ligase (Beyotime, #D7006-1)	0.5
Digested DNA Fragment (~30ng/µL)	2
Digested linear pGEX-4T-1 (~5ng/µL)	16

Table 6. Error-prone PCR reaction systems (50 µL total volume).

Component	0.3 % (w/v) $MnCl_2$ System (µL)	0.5 % (w/v) $MnCl_2$ System (µL)
Nuclease-free water	Up to 50	Up to 50

10× BeyoTaq Buffer with Mg^{2+} (Beyotime, #D7020-2)	5	5
dNTP mixture (25mM) (Beyotime, #D7371)	4	4
Template	1	1
Forward primer (10 μM)	2	2
Reverse primer (10 μM)	2	2
Beyo Taq DNA polymerase (5U/μL) (Beyotime, #D7219-1)	0.25	0.25
$MnCl_2$ 10% (w/v) stock	1.5	2.5

Table 7: Protein fragmentation primers

Primer_name	**Primer_seq**	**Supplier**
Common_F	CAAAATCGGATCTGGTTCCG	Sangon Biotech
Common_R	TCGTCAGTCAGTCACGATG	Sangon Biotech
CAAP1-1_R	TCACTCGAGCTCCAAGTCGCTGTGTTC	Sangon Biotech
CAAP1-2_F	TCAGAATTCGACTTGGAGGAAGGTGG	Sangon Biotech
CAAP1-2_R	TCACTCGAGCTCCTCCATGTCCGAATCC	Sangon Biotech
CAAP1-3_F	TCAGAATTCGAGGAGGAAGCGGATGATG	Sangon Biotech
DNAJC28-1_R	ACTCTCGAGTCACCTTTAGCCATGCTCTCCT	Sangon Biotech
DNAJC28-2_F	TCAGAATTCGGTGATTTCGATAACCTCTCCG	Sangon Biotech
EVA1C-1_R	TCTCTCGAGTCTTCAGTTCATTCGGCTGG	Sangon Biotech
EVA1C-2_F	TCAGAATTCGTGTCTTTTAAGTGCCAGCC	Sangon Biotech
EVA1C-2_R	TCACTCGAGCGGCTTCAGGTTGGC	Sangon Biotech
EVA1C-3_F	TCAGAATTCGCGATCGCCAACCTG	Sangon Biotech
MSANTD7-1_R	ACACTCGAGGCAACCAAGTTTTGATCGCAG	Sangon Biotech
MSANTD7-2_F	TCAGAATTCCACTGCGATCAAAACTTGGTT	Sangon Biotech
PPAN-P2RY11-1_R	TCACTCGAGCTCTTCCTCAGTTTTGC	Sangon Biotech
PPAN-P2RY11-2_F	TCAGAATTCGGCAAAGTGATGTTTCACTCG	Sangon Biotech

STKLD1-1_R	ACTCTCGAGCTGCCTCTCAGGAAGGTG	Sangon Biotech
STKLD1-2_F	TCAGAATTCGGCAGTTTCAAGTCCAGTTG	Sangon Biotech
TMX3-1_R	ACACTCGAGGAAGTGGCCGAATTAGAG	Sangon Biotech
TMX3-2_F	ACAGAATTCCCACTTCCAAGTCAACAAATGT	Sangon Biotech
TMX3-2_R	ACACTCGAGAAACCTTTCCCTGTTGATCC	Sangon Biotech
TMX3-3_F	TCAGAATTCCAGAATTACCTTGCTATGGATGG	Sangon Biotech
PBDC1-1_R	TCACTCGAGCGTTTCTTCTCAACGGTC	Sangon Biotech
PBDC1-2_F	TCAGAATTCGTTGAGAAGAAACGTACCC	Sangon Biotech

Extended Data

Description: Extended Data 1. Images of BGIS experiments supporting the screening results shown in Fig. 1B–C.

(A) Representative BGIS results of the 29 selected human dark proteins. Each protein was transformed into E. coli BL21(DE3) and plated on LB-AMP plates with (+IPTG) or without (−IPTG) IPTG induction. Growth inhibition was evaluated by comparing colony formation between induced and non-induced conditions. Among the 29 proteins screened, CAAP1, MSANTD7, and DNAJC28 exhibited a complete growth inhibition phenotype, whereas PBDC1, PPAN-P2RY11, TMX3, EVA1C, and STKLD1 exhibited a partial growth inhibition phenotype upon IPTG induction. (B) Representative BGIS results of protein fragments derived from the inhibitory proteins identified in (A). Protein fragments were generated based on AlphaFold2-predicted structural boundaries and screened using BGIS to localize domain-specific growth inhibition activity.

. Resource Type: Dataset. File: Ex 1.jpg. DOI: 10.22002/yrmf4-jre15

Description: Extended Data 2. Expression analysis of all 29 human dark proteins used in this study via western blotting.. Resource Type: Dataset. File: Ex2 (1).jpg. DOI: 10.22002/nmqzn-p5719

References

Brannigan JA, Wilkinson AJ. 2002. Protein engineering 20 years on. Nature Reviews Molecular Cell Biology 3: 964-970. DOI: 10.1038/nrm975

Cao X, Sun S, Xing J. 2024. A Massive Proteogenomic Screen Identifies Thousands of Novel Peptides From the Human “Dark” Proteome. Molecular & Cellular Proteomics 23: 100719. DOI: 10.1016/j.mcpro.2024.100719

Guo H, Prell M, Königs H, Xu N, Waldmann T, Hermans-Sachweh B, et al., Kappes. 2021. Bacterial Growth Inhibition Screen (BGIS) identifies a loss-of-function mutant of the DEK oncogene, indicating DNA modulating activities of DEK in chromatin. FEBS Letters 595: 1438-1453. DOI: 10.1002/1873-3468.14070

Guo H, Xu N, Prell M, Königs H, Hermanns-Sachweh B, Lüscher B, Kappes F. 2021. Bacterial Growth Inhibition Screen (BGIS): harnessing recombinant protein toxicity for rapid and unbiased interrogation of protein function. FEBS Letters 595: 1422-1437. DOI: 10.1002/1873-3468.14072

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al., Hassabis. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596: 583-589. DOI: 10.1038/s41586-021-03819-2

Kelleher KJ, Sheils TK, Mathias SL, Yang JJ, Metzger VT, Siramshetty VB, et al., Oprea. 2022. Pharos 2023: an integrated resource for the understudied human proteome. Nucleic Acids Research 51: D1405-D1416. DOI: 10.1093/nar/gkac1033

Moitra T, Larrouy-Maumus G. 2026. Integrated approaches for discovery and functional annotation of proteins of unknown function. Trends in Biochemical Sciences 51: 80-92. DOI: 10.1016/j.tibs.2025.11.001

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al., Phillippy. 2022. The complete sequence of a human genome. Science 376: 44-53. DOI: 10.1126/science.abj6987

Omenn GS, Orchard S, Lane L, Lindskog C, Pineau C, Overall CM, et al., Deutsch. 2024. The 2024 Report on the Human Proteome from the HUPO Human Proteome Project. Journal of Proteome Research 23: 5296-5311. DOI: 10.1021/acs.jproteome.4c00776

Rocha JoJ, Jayaram SA, Stevens TJ, Muschalik N, Shah RD, Emran S, et al., Munro. 2023. Functional unknomics: Systematic screening of conserved genes of unknown function. PLOS Biology 21: e3002222. DOI: 10.1371/journal.pbio.3002222

Sharifi Tabar M, Parsania C, Chen H, Su XD, Bailey CG, Rasko JEJ. 2022. Illuminating the dark protein-protein interactome. Cell Reports Methods 2: 100275. DOI: 10.1016/j.crmeth.2022.100275

Funding: This work was supported by internal funding from Duke Kunshan University, including the Summer Research Scholarship (SRS), Signature Work Fund (SF), Signature Work Research Grant (SWRG), a grant from the DKU Innovation Incubator (Dii), and a Start-up Fund.

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Songhe Jin: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, validation, visualization, writing - original draft, writing - review editing. Chutian Chen : conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, validation, writing - original draft, writing - review editing. Yewei Liu : conceptualization, supervision, visualization, methodology. Ferdinand Kappes: conceptualization, funding acquisition, supervision, resources, writing - review editing.

Reviewed By: Anonymous

History: Received July 5, 2026 **Revision Received** October 25, 2026 **Accepted** November 18, 2026 **Published Online** November 19, 2026 **Indexed** December 3, 2026

Copyright: © 2026 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Jin S, Chen C, Liu Y, Kappes F. 2026. Bacterial Growth Inhibition Screen (BGIS): an unconventional phenotype-based workflow for mapping regions in human dark proteins associated with reduced growth in *E. coli*.. microPublication Biology. 10.17912/micropub.biology.002268