Eliminating OFF-Frame Clones in Randomized Gene Libraries

An Improved Split β-Lactamase Enrichment System

Mejias-Gomez, Oscar; Madsen, Andreas V.; Pedersen, Lasse E.; Kristensen, Peter; Goletz, Steffen

Published in:
New Biotechnology

DOI (link to publication from Publisher):
10.1016/j.nbt.2023.03.002

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

New BIOTECHNOLOGY 75 (2023) 13–20

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

journal homepage: www.elsevier.com/locate/nbt

Eliminating OFF-frame clones in randomized gene libraries: An improved split β-lactamase enrichment system

Oscar Mejias-Gomez [a], Andreas V. Madsen [a], Lasse E. Pedersen [a], Peter Kristensen [b], Steffen Goletz [a,*]

[a] Department of Biotechnology and Biomedicine, Section for Protein Science and Biotherapeutics, Technical University of Denmark, Kongens Lyngby, Denmark
[b] Department of Chemistry and Bioscience, Section for Bioscience and Engineering, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Keywords:
Randomized gene libraries
Split beta-lactamase
Functional diversity
Curation system
Frameshifts
Antibody libraries

ABSTRACT

Large, randomized libraries are a key technology for many biotechnological applications. While genetic diversity is the main parameter most libraries direct their resources on, less focus is devoted to ensuring functional IN-frame expression. This study describes a faster and more efficient system based on a split β-lactamase complementation for removal of OFF-frame clones and increase of functional diversity, suitable for construction of randomized libraries. The gene of interest is inserted between two fragments of the β-lactamase gene, conferring resistance to β-lactam drugs only upon expression of an inserted IN-frame gene without stop codons or frameshifts. The preinduction-free system was capable of eliminating OFF-frame clones in starting mixtures of as little as 1% IN-frame clones and enriching to about 70% IN-frame clones, even when their starting rate was as low as 0.001%. The curation system was verified by constructing a single-domain antibody phage display library using trinucleotide phosphoramidites for randomizing a complementary determining region, while eliminating OFF-frame clones and maximizing functional diversity.

1. Introduction

Generation of libraries containing randomized genes is a process utilized in several molecular biology and biotechnology applications, such as directed evolution, enzyme, protein and antibody engineering, functional genomics and antibody phage display [1–3]. In most cases, the main goal is to create large, randomized libraries with maximum diversity; however, they are prone to result in significant numbers of genes where the reading frame is unintentionally altered, thus creating clones where translation is out of frame (OFF-frame) [4]. This is especially true when using randomization techniques such as error-prone PCR, site-directed mutagenesis (SDM), splice-overlapping PCR (SOE-PCR) and trimer phosphoramidites (TRIM) [5,6]. Libraries of highest quality are based on a combination of high functional diversity, originating from a high and meaningful genetic diversity, as well as a high abundance of IN-frame open reading frames (ORFs). An application where maximum removal of OFF-frame clones is especially important is where construction of large libraries is restricted by large workload as e.g. for antibody phage display libraries with high functional diversity. In phage display, a key aspect to determine whether a library is successful is its quality, which is commonly reflected by its ability to produce binders against a diverse set of antigens, which in turn depends on the functional diversity and relative abundance of IN-frame clones. In phage-display libraries, only IN-frame clones will result in an ORF comprising the fusion protein and the minor capsid protein III (pIII) and therefore be successfully displayed at the surface of filamentous phages, since pIII is located downstream of the antibody fragments. However, if no selection for ORFs is performed prior to phage production, the

List of abbreviations: AlBla-7B6-OmBla, gene block with alpha beta-lactamase, 7B6 clone and omega beta-lactamase; CDR, complementary determining region; CFU/mL, colony forming units per milliliter; ET, extension time; IGV, Integrated Genomics Viewer; IN-frame, clones where translation is in frame; LB agar, Luria broth agar; LB broth, Luria Broth broth; OFF-frame, clones where translation is out of frame; ONT, Oxford Nanopore Technologies; POI, Protein of interest; pTP_AlBla-7B6-OmBla, A plasmid with the gene-block AlBla-7B6-OmBla; pTP_AlBla-7B6-OmBla_OFF, A plasmid with the gene-block AlBla-7B6-OmBla, with a deletion; pTP_Lib, TP1122 plasmid with ssDNA library oligos cloned; pTP1122, plasmid used as template for curation system; QV, Quality value; scFv, Single chain antibody fragment; sdAb, Single domain antibody; SDM, Site-directed mutagenesis; SOE-PCR, Splice-overlapping PCR; Ta, Annealing temperature; TRIM, Trimer phosphoramidites.

* Correspondence to: Søltofts Plads, 224, 122, 2800 Kongens Lyngby, Denmark.
E-mail address: sgoletz@dtu.dk (S. Goletz).

https://doi.org/10.1016/j.nbt.2023.03.002
Received 30 January 2023; Received in revised form 20 February 2023; Accepted 4 March 2023
Available online 6 March 2023

functional diversity of the library (filamentous phages successfully displaying an antibody fragment) will be significantly reduced.

Several attempts aimed at increasing the ratio of IN-frame clones have been developed. These can be divided into whether the curation happens before or after phage production. The methods after phage production rely on the ability of some antibody fragments to bind specific affinity ligands like protein A [7] or protein L [8] when correctly folded. While this can filter libraries for correctly folded antibody fragments, they are limited to certain human frameworks and reduce the overall diversity of the library to include antibody frameworks capable of binding protein A or protein L, thereby introducing a bias in curation. The methods used before phage production are based on cloning the antibody fragment genes upstream of reporter genes. In 1992, Seehaus et al. developed a plasmid in which single chain antibody fragments (scFv) were cloned upstream of a β-lactamase resistance gene, which successfully selected for ORFs [9]. This concept was further explored by Zachi et al. who introduced homologous recombination sites flanking the resistance gene, which enabled a rapid downstream phage-display application [10]. Since then, the β-lactamase approach has been used for enrichment for IN-frame clones in phage-display libraries [9–13], and in other applications such as gene annotation [14]. In recent years, an improved version based on a split β-lactamase protein complementation assay has been developed, which reduced background resistance typically seen when using the full β-lactamase. In brief, β-lactamase, which confers resistance to β-lactam antibiotics, can be separated into two fragments (Alpha and Omega). For β-lactamase to be active, both fragments must be present and interact with each other. Moreover, these two fragments can regain enzymatic activity when brought into proximity by the IN-frame expression of a protein of interest (POI). If the POI is OFF-frame or the POI contains a stop codon, only the Alpha fragment is expressed, which is not sufficient to confer resistance against β-lactam antibiotics and will ultimately render the cells carrying the plasmid unable to survive under selective antibiotic pressure (Fig. 1). This version has also been utilized outside the phage display field for monitoring protein-protein interactions in vivo [15], construction of gene-fragment libraries [16], directed evolution of aggregation-prone scFvs [17], identification of aggregation inhibitors [18], and to assess the impact of aminoacid substitutions in proteins susceptible to aggregation [19].

Here we describe a further improved, faster and more robust version of the split β-lactamase IN-frame enrichment system that requires no pre-induction. The method is more efficient than traditional methods, providing a greater yield of IN-frame clones in a shorter time. The system shows high IN-frame enrichment rates and is able to fully eliminate OFF-frame clones. The system has been used for the curation of a single-domain antibody (sdAb) library randomized via Trimer Phosphoramidites (TRIM) in one complementary determining region (CDR) prior to phage production, which eliminated OFF-frame clones and maximized functional diversity.

2. Materials and methods

2.1. Materials

Materials were obtained from the following providers: E.coli strains NEB 5-alpha (C2987) (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) and NEB10beta electrocompetent (*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*), high fidelity hot-start DNA polymerase Q5 (M0493S), Q5 Site-directed mutagenesis kit (E0554S), and NEBuilder HiFi DNA Assembly master mixes (E2621S) from New England Biolabs (Ipswich, MA, USA). GeneJet Miniprep Kit (K0502) for plasmid DNA extraction, 1 mm electropotation cuvettes (551011), Qubit Fluorometer and Qubit HS DNA kit (Q32851) for DNA concentration measurement, and Nanodrop One for DNA quality assessment were purchased from Thermo Fisher Scientific (Waltham, MA, USA). The plasmid pTP1122 was a gift from Dirk Görlich (Addgene plasmid # 104159; http://n2t.net/addgene:104159; RRID:Addgene_104159) (Watertown, MA, USA). The gene-block AlBla-7B6-OmBla was purchased from Twist Bioscience (South San Francisco, CA, USA) (Table S1). The primers for cloning (Primer 1–4, and 7–8) and site-directed mutagenesis (Primers 5 and 6) were from Integrated DNA Technologies (Coralville, IO, USA) (Table S1). The ssDNA TRIM library oligo was from ELLA Biotech GmbH (Fürstenfeldbruck, Germany). Mag-Bind Total Pure NGS (M1378–01) for magnetic-bead cleanup from Omega Biotek (Norcross, GA, USA). The

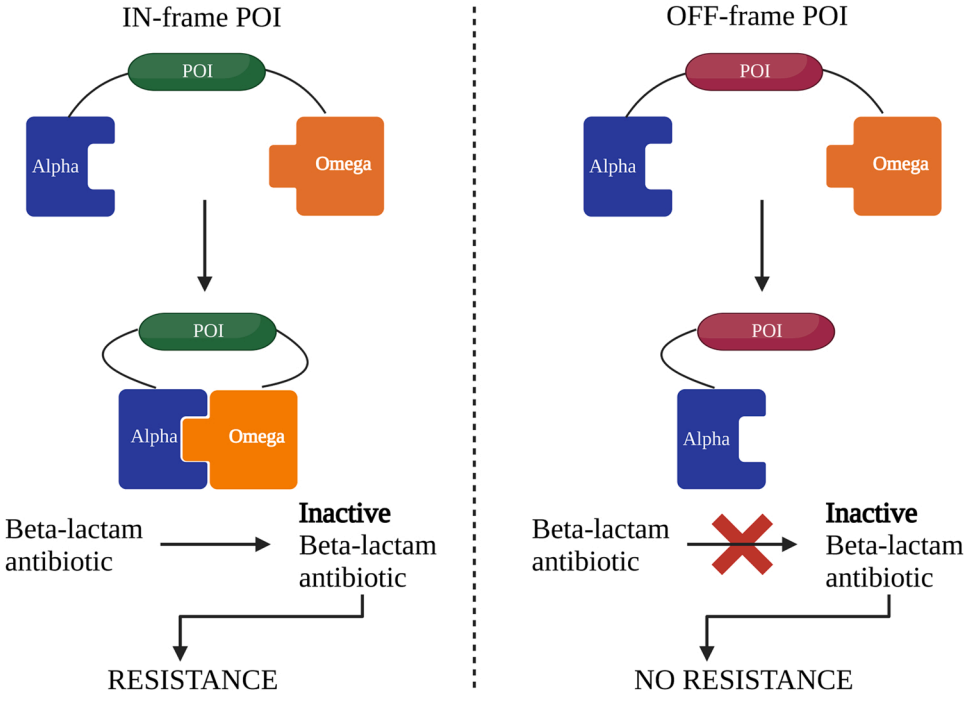

Fig. 1. Split-β-lactamase complementation assay concept. If the protein of interest (POI) has been cloned IN-frame of Alpha β-lactamase and does not contain a stop codon, both fragments, Alpha and Omega of β-lactamase, will be expressed. The fragments will then be able to regain β-lactamase functionality and confer resistance to β-lactam antibiotics by inactivating them. If the POI has been cloned OFF-frame or other frameshifts or a stop codon are present, only the Alpha fragment will be expressed, which is not enough to confer resistance against β-lactams antibiotics. Image generated with Biorender under publication agreement number "NN24XOW6HG".

rapid barcoding kit (RBK004), R9.4.1 flow cell and MinION were from Oxford Nanopore Technologies (ONT) (Oxford, UK). CLC Main Workbench (Qiagen, Hilden, Germany) software for analysis of sequencing results. Integrated Genomics Viewer (IGV) (University of California, CA, USA) was used for analyzing ONT sequencing results.

2.2. Methods

For generation of pTP_AlBla-7B6-OmBla, 10 ng of pTP1122 were used as template for high-fidelity PCR linearization using Primers 1 and 2 and Q5 Hot-start Master Mix (Q5 HS MM) as polymerase (30 cycles, Ta=66 °C, ET=120 s). In addition, Q5-HS-MM was used to amplify 1 ng of the gene-block AlBla-7B6-OmBla using primers 3 and 4, (30 cycles, Ta=66 °C, ET=120 s). The PCRs were analyzed via 1% agarose gel electrophoresis. PCR-linearized pTP1122 and PCR-amplified AlBla-7B6-OmBla were mixed in a 10:1 insert/vector molar ratio together with NEBuilder HiFi DNA Assembly Master Mix. The reaction was incubated for 30 min at 50 °C. The assembled product was transformed into NEB 5-alpha using heat-shock: 1 µL of DNA was mixed with 25 µL of cells and incubated for 30 min on ice, the mixture was then introduced into a water bath at 42°C for exactly 30 s and placed back on ice for 5 min, 975 µL of pre-warmed SOC recovery media was added, and the cells were recovered for 1 h shaking at 220 rpm and 37 °C. The cells were plated on Luria Broth Agar plates supplemented with 50 µg/mL of kanamycin. Assembly was verified after picking isolated colonies and miniprep plasmid extraction via Sanger sequencing.

For generation of pTP_AlBla-7B6-OmBla_OFF, the Sanger-verified assembled product pTP_AlBla-7B6-OmBla was PCR-linearized with Q5-HS-MM and primers 5 and 6 (30 cycles, Ta=60 °C, ET=160 s). The PCR was verified on a 1% agarose gel. The verified PCR-linearized pTP_AlBla-7B6-OmBla_OFF was added to the Kinase-Ligase-*Dpn*I (KLD) mix and incubated for 5 min at room temperature. The product was transformed into NEB 5-alpha chemical competent cells as described above. The deletion was verified via Sanger sequencing as described above.

For the construction of pTP_Lib, 10 ng of pTP_AlBla-7B6-OmBla were used as template for high-fidelity PCR linearization using Q5-HS-MM and primers 7 and 8 (30 cycles, Ta=60 °C, ET=160 s). After PCR linearization, the product was treated with *Dpn*I. The PCR-linearized and *Dpn*I treated product was cleaned using magnetic beads at a 0.55x ratio, following manufacturer's instructions. The PCR, *Dpn*I treatment and cleanup were verified on a 1% agarose gel and using Nanodrop. The ssDNA TRIM oligos were cloned into pTP_AlBla-7B6-OmBla in a 20:1 insert/vector molar ratio using NEBuilder HiFi Assembly MM supplemented with 5% DMSO for 15 min at 50 °C. The assembled product was transformed via electroporation into NEB10beta: 1 µL of DNA was mixed with 25 µL of cells, the mixture was transferred to 1 mm electroporation cuvettes and subjected to electroporation at 1700 V, 975 µL of pre-warmed recovery media were added immediately after and the cells were incubated for 90 min at 37 °C and 220 rpm. Serial dilutions of the transformed cells were made and directly plated into LB agar plates supplemented with 50 µg/mL kanamycin, 100 µg/mL streptomycin, and 100 µg/mL of carbenicillin (Selective) or 2% glucose (Unselective) and incubated ON at 30 °C. For the mixes, accurate DNA measurements of pTP_AlBla-7B6-OmBla and pTP_AlBla-7B6-OmBla_OFF were performed using Qubit and diluted to 100 ng/µL. Mix 1 corresponded to 100% of pTP_AlBla-7B6-OmBla; Mix 2–15% of pTP_AlBla-7B6-OmBla and 85% of pTP_AlBla-7B6-OmBla_OFF; Mix 3 to, 1% of pTP_AlBla-7B6-OmBla and 99% of pTP_AlBla-7B6-OmBla_OFF, and Mix 4–0001% pTP_AlBla-7B6-OmBla and 99,999% of pTP_AlBla-7B6-OmBla_OFF. 100 µL of mixes were transformed separately into NEB10beta as described above. For verification, ten randomly picked colonies from each mix plate and 30 from the library were inoculated into 5 mL of LB-broth supplemented with 50 µg/mL of kanamycin, 100 µg/mL streptomycin and 2% glucose. DNA plasmid extraction was performed as described above and the DNA was Sanger-sequenced. Undiluted transformed bacteria were plated, scraped after growth with ice-cold LB broth and immediately pelleted and DNA extracted. Rapid Barcoding Kit for ONT library preparation was used to attach unique barcodes to each mix-plate and each library-plate pair following manufacturer's instructions. The barcoded library was sequenced in MinION using a R9.4.1 (FLO-MIN106) flow cell for 1.89 h with super-accurate basecalling (Guppy 6.0.7). The raw reads were mapped to the reference file using minimap2 v2.24 and variant called using haploid variant Medaka v1.6.0 with r941_min_sup_variant_g507 as model. The alignments were visualized using Integrated Genomics Viewer (IGV) [20]. ONT sequencing raw data is available in Supplementary materials.

3. Results

3.1. Design and construction of a rapid pre-induction free curation system

The curation system is based on the split β-lactamase complementation assay. To establish and validate the curation system, a preexisting single domain antibody (sdAb) in the plasmid pTP1122 was exchanged for a gene block comprising the Alpha and Omega β-lactamase fragments separated by a sdAb clone (7B6) previously selected from the pREDATOR library [21]. (Fig. 2). The 1352 bp gene block AlBla-7B6-OmBla (Table S1) was designed to contain the following components: a β-lactamase signal sequence (BSS), Alpha β-lactamase fragment (AlBla), NGR tripeptide, which was identified as increasing enzyme activity 4-fold [22], $(GGGGS)_3$ linker (GSlink), *Nco*I site, 7B6 clone, *Not*I site, $(GGGGS)_3$ linker (GSlink), Omega β-lactamase fragment (OmBla). The gene-block was cloned IN-frame between RBS and the t0 terminator of pTP1122 in exchange of the preexisting nanobody, forming the curation system called pTP_AlBla-7B6-OmBla. The correct insertion was verified by agarose gel electrophoresis, PCR and Sanger sequencing (data not shown).

3.2. Evaluation of the rapid curation system

To test the system's performance, an OFF-frame version of the above plasmid was created by deleting a single nucleotide in position 930 of

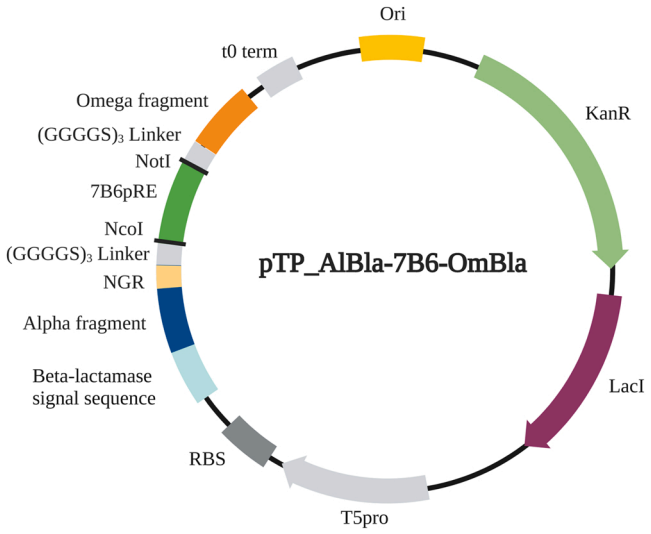

Fig. 2. Map of the curation system pTP_AlBla-7B6-OmBla. The plasmid contained a kanamycin resistance gene (KanR), lac repressor (lacI), ColE1 origin of replication (ori), T5 promoter with embedded lac operator (T5pro), ribosome binding site (RBS), and the gene-block AlBla-7B6-OmBla, which entails: a β-lactamase signal sequence, Alpha fragment, NGR tripeptide, (GGGGS)3 linker, *Nco*I site, 7B6 His-tagged sdAb clone, *Not*I site, (GGGGS)3 linker, and Omega fragment. Image generated with Biorender under publication agreement number "DC24XOX22B".

the construct using site directed mutagenesis. Four different mixes with different ratios of IN/OFF-frame plasmids were transformed in NEB10 electrocompetent cells and recovered (Fig. 3, step 1 and 2). The transformed cells were incubated in recovery media for 1.5 h at 37 °C and plated on LB agar supplemented with 50 μg/mL kanamycin, 100 μg/mL streptomycin, and 100 μg/mL carbenicillin (selective) or 2% glucose (unselective) (Fig. 3, step 3).

Initially the optimal concentration of carbenicillin for antibiotic pressure optimization was established by testing increasing concentrations of carbenicillin including 10, 50, 100, 200, and 500 μg/mL in the presence or absence of 2% glucose (Fig. 4 A and B) in Mix 2 with an excess of 85% of OFF-frame clones representing a worst-case scenario for library constructions. After transformation of NEB-10 cells, 10 colonies from each plate were sequenced via Sanger sequencing. The system could reach full curation with 100% IN-frame clones in the absence of glucose with carbenicillin concentrations from 50 μg/mL to 500 μg/mL (Fig. 4 A). Moreover, the CFU/mL corresponded to a reduction of approximately 85% of the unselective population for carbenicillin concentrations above 50 μg/mL when the system was applied (Fig. 4B). From this, 100 μg/mL, which corresponds to the standard concentration of carbenicillin in literature, was chosen for evaluating further the curation potential of the system involving Mixes 1–4, with even lower percentage of IN-frame clones, as well as for a TRIM-randomized library model.

Colony forming units per milliliter (CFU/mL) were calculated as a proxy for evaluating the efficiency of IN-frame clonal selection where a reduction of CFU/mL equivalent to the percentage of OFF-frame clones was expected (Fig. 3, expected results). The reduction in CFU/mL was 28%, 86%, 98.4% and 99.99% for Mixes 1, 2, 3 and 4, respectively (Fig. 4C, Table S2).

Ten colonies were picked for each mix in both selective and unselective plates and inoculated into 5 mL of LB supplemented with kanamycin, streptomycin, and glucose. DNA plasmid extraction was performed for each culture and the plasmid was sequenced using Sanger sequencing. As a quality control, sequences with less than 300 bp of QV≥ 20 were discarded from the analysis (Table S3) and the rest were used for assessing the performance of the system in terms of relative abundance of IN-frame clones before and after selective pressure. Mixes 2 and 3 showed no OFF-frame clones after selective pressure resulting in a statistically significant increase in relative abundance of IN-frame clones from 10% (Mix 2) and 0% (Mix 3) to 100% (Fisher's exact test p-values of 0.0001 and 0.02 for mixes 2 and 3, respectively) after curation (Fig. 4D and Table S3). Furthermore, Mix 4 had a statistically significant increase from 0.001% to 50% (Fisher's exact test p-value of

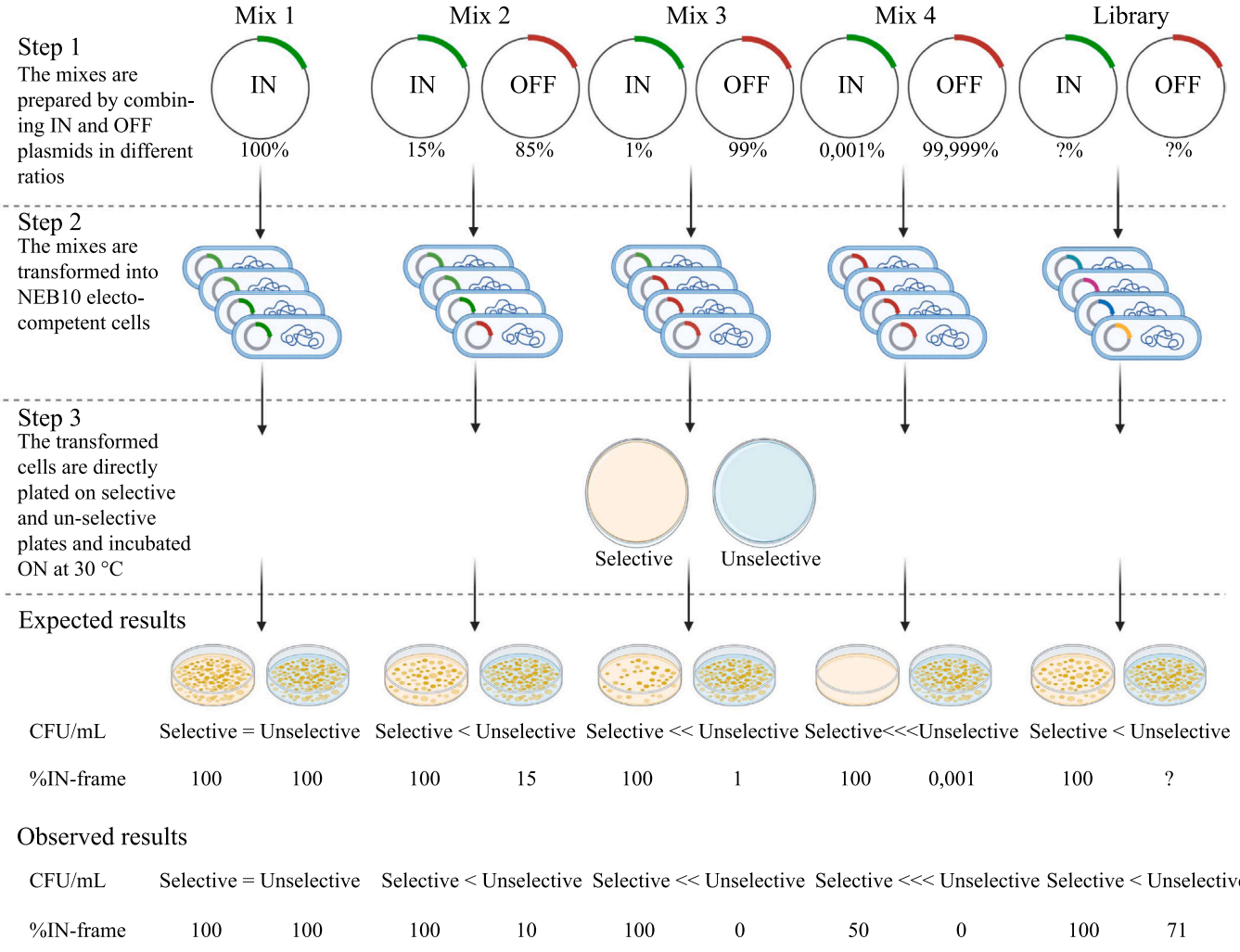

Fig. 3. Design and evaluation of the curation system. The curation system was designed to be as fast as possible by directly plating the transformed bacteria without pre-induction. Step 1. The system performance was evaluated by creating mixes with different ratios of IN- and OFF-frame plasmids. Step 2. The mixes were transformed into NEB10 electrocompetent cells. Step 3. After direct plating on LB-agar supplemented with kanamycin, streptomycin, and carbenicillin (selective) or glucose (unselective), CFU/mL and relative abundance of IN-frame clones were calculated and compared to the expected results. The system was further evaluated by curation of a TRIM-randomized library with unknown percentage of IN-frame clones. Image partially created in Biorender under publication license agreement "TE24XOXE3A".

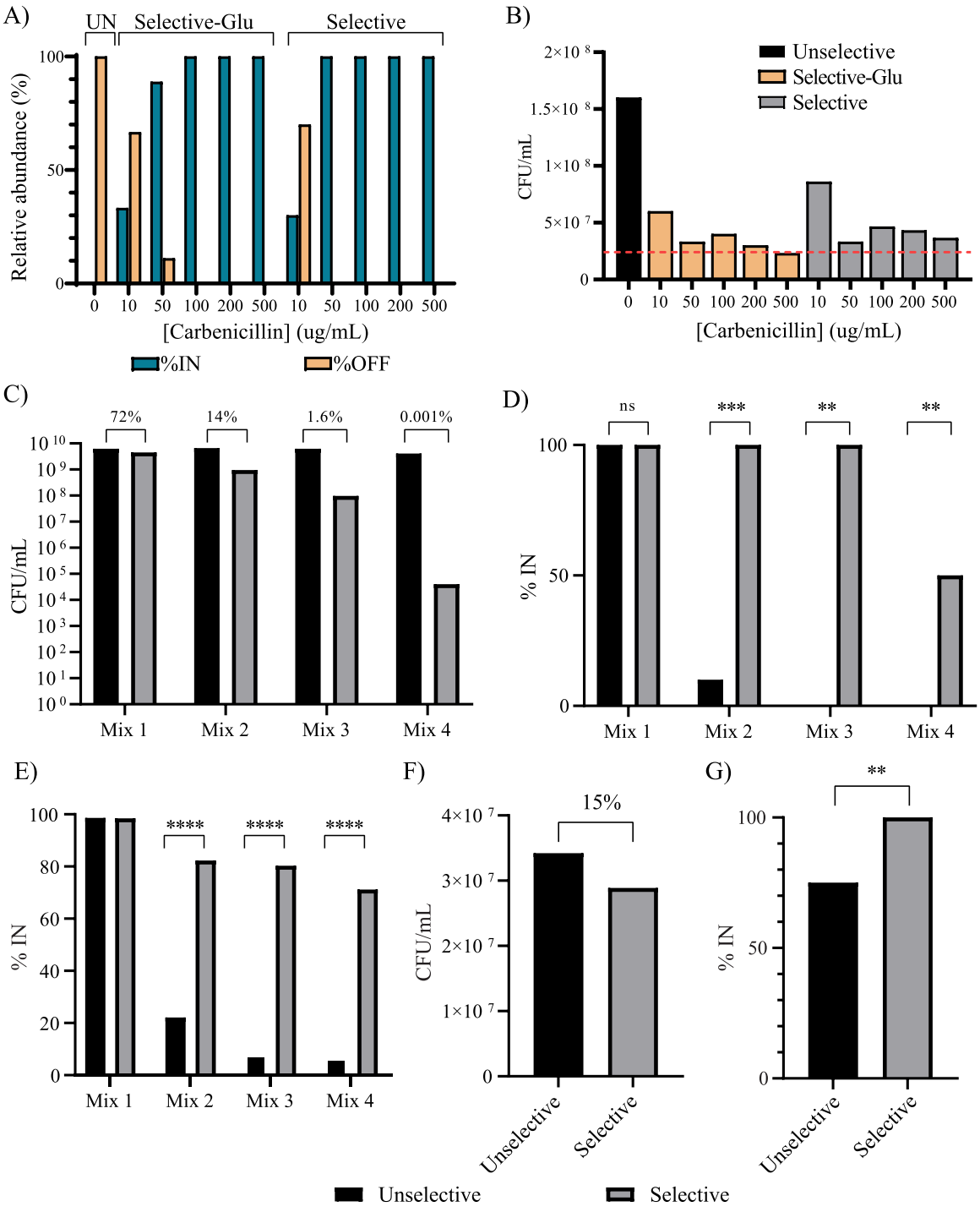

Fig. 4. A) Carbenicillin concentration optimization. The relative abundance of IN (blue) and OFF-frame (orange) clones is displayed for all carbenicillin concentrations tested (10, 50, 100, 200, and 500 μg/mL) in presence or absence of 2% glucose. B) CFU/mL in each sample of carbenicillin concentration optimization experiment. The CFU/mL is displayed for all carbenicillin concentrations tested in presence or absence of 2% glucose (orange and gray, respectively). Red dotted line indicates the expected CFU/mL if all OFF-frame clones are eliminated from uncurated sample with 15% IN-frame clones. C) CFU/mL of mixes 1–4 in unselective (black) and selective (gray) plates. The y-axis shows colony forming units per milliliter of bacteria calculated by considering number of colonies, dilution and volume of transformed bacteria plated. Percentages show CFU/mL of selective plate in comparison with unselective plate. D) Relative abundance of IN-frame clone in mixes before and after selection. Data from Sanger sequencing was used to determine which colonies contained an IN or OFF-frame clone. The enrichment for IN -frame clones was statistically significant for mixes 2, 3, and 4. E) Oxford Nanopore Technologies sequencing results for mixes before and after selection. IGV was used to visualize and analyze the aligned reads to the reference. The reads and deletion counts were used to estimate the relative abundance of IN- and OFF-frame clones for each mix before and after curation. There was a statistically significant selection for IN-frame clones after curation system was applied F) CFU/mL of library before and after selection. A reduction of 15% was observed for the library after the curation system was applied. G) Relative abundance of IN-frame clones in single-domain TRIM-randomized phage display library before and after selection. Sanger sequencing results were used to determine which clones had both a different CDR2 to each other and to the reference sequence as well as presence of IN-frame clones. 70% of the clones were OFF-frame before the curation system was applied (unselective) while none of them was OFF-frame after the selection system was applied (selective).

0.03) (Table S3 and Fig. 4D). The IN-frame selection was furthermore investigated using Oxford Nanopore Technologies (ONT) sequencing with a rapid barcoding library preparation method where a unique barcode is attached to each sample. 194,674 reads were analyzed to yield 880.2 Mbases with an average quality score of 14.62. The average read length was estimated to be 4521 kb with 100% basecalled reads. The average number of reads per barcode was 26534. The maximum number of reads were obtained for Mix1-unselective and minimum for Mix 4-selective with 30679 and 4391 reads, respectively. The Integrated Genomics Viewer (IGV) was used to visualize and analyze the aligned reads to the reference. The reads and deletion counts were used to estimate the relative abundance of IN- and OFF-frame clones for each mix before and after curation (Fig. 4E, Supplementary Table S4). Medaka variant calling indicated that Mixes 2, 3, and 4 were populated by a majority of variants with a deletion of 1 nucleotide in position 930. Moreover, it confirmed that relative abundance of OFF-frame clones was as expected for Mixes 2 and 3 (82.1% and 97.7%, respectively). Additionally, Mix 4 was confirmed to have at least 98.3% of OFF-frame clones. Medaka did not detect any variants in Mixes 1 (selective and unselective) and Mixes 2, 3, and 4 (selective), indicating efficient removal of OFF-frame clones in the mixes after the curation system was applied.

3.3. Applicability of the system using a single-domain TRIM-randomized phage

To evaluate the efficiency of IN-frame selection on an antibody phage display library, a sdAb TRIM randomized phage display library was subjected to selection pressure essentially as described above. A 15% reduction in CFU/mL was observed for selective in comparison with unselective plate (Fig. 4 F). Thirty colonies were picked from each plate and DNA plasmid extraction and Sanger sequencing was conducted as described above. In this case, quality filtering discarded 7 sequences from unselective plate and only 2 for selective plate. For the unselective plate, out of the 23 remaining sequences, 17 showed variability in CDR2 with 12 of them encoding IN-frame clones. For the selective plate, all sequences encoded IN-frame clones with 93% of the clones showing successful randomization of the CDR residues. The relative abundances of IN-frame clones were then 70.6% and 100% for unselective and selective plates, respectively, indicating a full curation (Fisher's exact test p-value of 0.0064) (Fig. 4G, Table S5). ONT sequencing analysis confirmed variation in the randomized area, while the Medaka base caller was not able to identify OFF-frame variants after curation.

4. Discussion

Previously a number of libraries have been enriched for IN-frame fusions using full length β-lactamase fusions to optimize the functional diversity [12,13,23–25]. This paper describes a further improved rapid system for enrichment of IN-frame clones based on the split β-lactamase complementation suitable for large, randomized libraries for applications where having maximum abundance of ORF is desired or advantageous, as for example in antibody phage display. The main advantage of this system is its flexibility and speed, since there is no need for pre-induction with IPTG prior to plating under selective pressure, unlike previously published methods [16].

This method demonstrates an exceptionally high curation rate and the introduction of two unique restriction sites flanking the gene-of-interest (in this paper referred to as 7B6pRE) makes the system applicable to any method where a high abundance of IN-frame clones is desirable. It was demonstrated that the transformed cells could be directly plated after recovery in selective plates, obtaining efficient selection for IN-frame fusion proteins in all samples tested. OFF-frame clones could be fully eliminated in starting mixtures containing as little as 1% IN-frame clones (Mixes 2, 3, and library, p value < 0.05). The highest observed enrichment was a factor of a 70,000-fold for IN-frame clones (Mix 4 from 0.001% to 70%, Table S4 and Fig. 4E), demonstrating that the system could function even in instances where the relative abundance of IN-frame clones was exceptionally limited. When the output was analyzed using Sanger sequencing and ONT sequencing, similar results were obtained (Fig. 4D-E). Nevertheless, ONT sequencing revealed an enrichment for IN-frame clones to more than 70% for Mix 4, which could not be observed with Sanger sequencing (Fig. 4E). The deviations in the numbers obtained with Sanger and ONT sequencing, could be attributed to the error rate of long-read sequencing technologies for ONT [26–29]. Additionally, Medaka variant caller was able to identify variants for only the Mixes 2, 3, and 4 before the curation system was applied (unselective plates). Afterwards (selective plates), Medaka was not able to identify variants, which indicates that the curation system was able to select for IN-frame clones to a full or close to a full curation. In the case of the single-domain antibody library, Sanger sequencing results indicated that the curation system was also able to effectively eliminate OFF-frame clones. However, ONT analysis of the library did not show any significant difference before and after application of the curation selection pressure, which might be due to the rather small enrichment that occurred for the library (75–100%) and the inability of Medaka variant caller to detect low frequency variants below the error rate threshold of long-read sequencing [30].

In phage display, some libraries use a post-phage-production approach for selecting IN-frame clones. These are based on the capacity of some antibody fragments to bind specific affinity ligands such as protein A [7,31–37], protein L [38] and hemagglutinin[39]. These methods are advantageous especially for single chain antibody fragments because they filter out those which do not fold correctly, thus facilitating soluble expression, affinity determination, crystallization, and large-scale biomanufacturing. However, these methods could potentially introduce biases towards paratopes recognizing the affinity ligands, thus reducing library diversity and versatility, and are not suitable for those libraries where e.g., protein A binding is not necessary or wanted. Herein a pre-phage-production method is described that can be applicable for any format and type of randomized library and that effectively eliminates OFF-frame clones.

The rapid curation mechanism proposed in this study was designed not to need IPTG and therefore avoid extensive pre-induction steps by the combination of several factors. The system's plasmid design contained a bacteriophage T5 promoter, which is a strong promoter recognized by any *E. coli* strain RNA polymerase [40,41]. Moreover, the *E. coli* strain used in this study (NEB10) has the key characteristic of having a deletion of the full lac-operon, thereby being deficient in the lactose permease *LacY*, which prevents induction of gene expression with IPTG or lactose. Adding IPTG to the selective plates resulted in a dramatic reduction in the CFU/mL (data not shown), which could be attributed to a toxic effect of IPTG on non-lactose utilizing *E. coli* strains (demonstrated in other *E. coli* strains by [42]). Additionally, this strain is a derivative of the commonly used DH10B, containing several mutations for enhanced DNA preparation such as *recA1*, *endA1*, *galE*⁻, and *deoR* [43–46]. The absence of glucose in the selective plates eliminates the tight regulation from the lacI repressor of the embedded lac operon in T5 promoter, thus enabling expression of the split β-lactamase governed by T5 promoter.

5. Conclusion

In this study, the design, creation, and validation of a rapid IN-frame selection system was demonstrated. The system was effective in eliminating OFF-frame clones in starting mixtures with as much as 99% OFF-frame clones and only 1% IN-frame clones and to enrich IN-frame clones by up to about 70,000-fold when present as low as 0.001% before selection. The preinduction-free split β-lactamase complementation-based IN-frame selection system is suitable for fast and efficient increase of functional diversity of libraries where OFF-frame clones are abundant or should be eliminated. Due to the introduction of unique restriction

enzymes flanking the gene-of-interest, the system is applicable to any type of large-scale and high-quality cloning method where high abundance of IN-frame clones is necessary.

Funding sources

This work was supported by The Novo Nordisk Foundation Grant NNF19SA0056783, NNF19SA0057794, and NNF20SA0066621.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available at https://figshare.com/s/db050fa0b81592a51755 and has an official doi reserved which will be active after publication.
ONT sequencing data (Original data) (Figshare)

Acknowledgements

The authors would like to acknowledge the valuable contributions of Kerstin Skovgaard in providing feedback to the design, execution, and analysis of this study. The authors also extend their gratitude to the reviewers for their valuable criticism and suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.nbt.2023.03.002.

References

[1] Abou-Nader M, Benedik MJ. Rapid generation of random mutant libraries. Bioeng Bugs 2010;1:337–40. https://doi.org/10.4161/bbug.1.5.12942.
[2] Neylon C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res 2004;32:1448–59. https://doi.org/10.1093/nar/gkh315.
[3] Hoskins RA. Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP. e185–e185 Nucleic Acids Res 2005;33. https://doi.org/10.1093/nar/gni184.
[4] D'Angelo S, Velappan N, Mignone F, Santoro C, Sblattero D, Kiss C, et al. Filtering "genic" open reading frames from genomic DNA samples for advanced annotation. BMC Genom 2011;12:S5. https://doi.org/10.1186/1471-2164-12-S1-S5.
[5] Knappik A, Ge L, Honegger A, Pack P, Fischer M, Wellnhofer G, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol 2000; 296:57–86. https://doi.org/10.1006/JMBI.1999.3444.
[6] Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. Genome Res 1992;2:28–33. https://doi.org/10.1101/gr.2.1.28.
[7] Valadon P, Pérez-Tapia SM, Nelson RS, Guzmán-Bringas OU, Arrieta-Oliva HI, Gómez-Castellano KM, et al. ALTHEA Gold Libraries™: antibody libraries for therapeutic antibody discovery. MAbs 2019;11:516–31. https://doi.org/10.1080/19420862.2019.1571879.
[8] Løset GÅ, Løbersli I, Kavlie A, Stacy JE, Borgen T, Kausmally L, et al. Construction, evaluation and refinement of a large human antibody phage library based on the IgD and IgM variable gene repertoire. J Immunol Methods 2005;299:47–62. https://doi.org/10.1016/J.JIM.2005.01.014.
[9] Seehaus T, Breitling F, Dübel S, Klewinghaus I, Little M. A vector for the removal of deletion mutants from antibody libraries. Gene 1992;114:235–7. https://doi.org/10.1016/0378-1119(92)90580-I.
[10] Zacchi P, Sblattero D, Florian F, Marzari R, Bradbury ARM. Selecting open reading frames from DNA. Genome Res 2003;13:980–90. https://doi.org/10.1101/gr.861503.
[11] Ge X, Mazor Y, Hunicke-Smith SP, Ellington AD, Georgiou G. Rapid construction and characterization of synthetic antibody libraries without DNA amplification. Biotechnol Bioeng 2010;106:347–57. https://doi.org/10.1002/BIT.22712.
[12] Rothe C, Urlinger S, Löhning C, Prassler J, Stark Y, Jäger U, et al. The Human Combinatorial Antibody Library HuCAL GOLD Combines Diversification of All Six CDRs According to the Natural Immune System with a Novel Display Method for Efficient Selection of High-Affinity Antibodies. J Mol Biol 2008;376:1182–200. https://doi.org/10.1016/j.jmb.2007.12.018.
[13] Yang HY, Kang KJ, Chung JE, Shim H. Construction of a large synthetic human scFv library with six diversified CDRs and high functional diversity. Mol Cells 2009;27:225–35. https://doi.org/10.1007/s10059-009-0028-9.
[14] D'Angelo S, Velappan N, Mignone F, Santoro C, Sblattero D, Kiss C, et al. Filtering "genic" open reading frames from genomic DNA samples for advanced annotation. BMC Genom 2011;12:S5. https://doi.org/10.1186/1471-2164-12-S1-S5.
[15] Galarneau A, Primeau M, Trudeau LE, Michnick SW. β-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein–protein interactions. Nat Biotechnol 2002 20:6 2002;20:619–22. https://doi.org/10.1038/nbt0602-619.
[16] Verma V, Joshi G, Gupta A, Chaudhary VK. An efficient ORF selection system for DNA fragment libraries based on split betalactamase complementation. PLoS One 2020:15. https://doi.org/10.1371/journal.pone.0235853.
[17] Ebo JS, Saunders JC, Devine PWA, Gordon AM, Warwick AS, Schiffrin B, et al. An in vivo platform to select and evolve aggregation-resistant proteins. Nat Commun 2020;11:1816. https://doi.org/10.1038/s41467-020-15667-1.
[18] Saunders JC, Young LM, Mahood RA, Jackson MP, Revill CH, Foster RJ, et al. An in vivo platform for identifying inhibitors of protein aggregation. Nat Chem Biol 2016;12:94–101. https://doi.org/10.1038/nchembio.1988.
[19] Guthertz N, van der Kant R, Martinez RM, Xu Y, Trinh CH, Iorga BI, et al. The effect of mutation on an aggregation-prone protein: An in vivo, in vitro, and in silico analysis. Proc Natl Acad Sci 2022:119. https://doi.org/10.1073/pnas.2200468119.
[20] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011 2011;29(1):24–6. https://doi.org/10.1038/nbt.1754.
[21] Mandrup OA, Friis NA, Lykkemark S, Just J, Kristensen P. A novel heavy domain antibody library with functionally optimized complementarity determining regions. PLoS One 2013;8:1–15. https://doi.org/10.1371/journal.pone.0076834.
[22] Wehrman T, Kleaveland B, Her JH, Balint RF, Blau HM. Protein-protein interactions monitored in mammalian cells via complementation of β-lactamase enzyme fragments. Proc Natl Acad Sci USA 2002;99:3469–74. https://doi.org/10.1073/PNAS.062043699/ASSET/DC3EB6DF-6B89-4172-B2D3-6F239A093DB2/ASSETS/GRAPHIC/PQ0620436004.JPEG.
[23] Prassler J, Thiel S, Pracht C, Polzer A, Peters S, Bauer M, et al. HuCAL PLATINUM, a synthetic fab library optimized for sequence diversity and superior performance in mammalian expression systems. J Mol Biol 2011;413:261–78. https://doi.org/10.1016/J.JMB.2011.08.012.
[24] Huovinen T, Syrjänpää M, Sanmark H, Brockmann EC, Azhayev A, Wang Q, et al. Two ScFv antibody libraries derived from identical VL–VH framework with different binding site designs display distinct binding profiles. Protein Eng, Des Sel 2013;26:683–93. https://doi.org/10.1093/PROTEIN/GZT037.
[25] Goldman ER, Andersen GP, Liu JL, Delehanty JB, Sherwood LJ, Osborn LE, et al. Facile generation of heat-stable antiviral and antitoxin single domain antibodies from a semisynthetic llama library. Anal Chem 2006;78:8245–55. https://doi.org/10.1021/AC0610053.
[26] Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PLoS One 2021;16:e0257521. https://doi.org/10.1371/JOURNAL.PONE.0257521.
[27] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020 21 2020;1 (21):1–16. https://doi.org/10.1186/S13059-020-1935-5.
[28] Sahlin K, Sipos B, James PL, Medvedev P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. Nat Commun 2021 12:1 2021;12:1–13. https://doi.org/10.1038/s41467-020-20340-8.
[29] Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. BMC Genom 2020;21:1–15. https://doi.org/10.1186/S12864-020-07227-0/TABLES/9.
[30] Feng Z, Clemente JC, Wong B, Schadt EE. Detecting and phasing minor single-nucleotide variants from long-read sequencing data. Nat Commun 2021 12:1 2021; 12:1–13. https://doi.org/10.1038/s41467-021-23289-4.
[31] Jespers L, Schon O, Famm K, Winter G. Aggregation-resistant domain antibodies selected on phage by heat denaturation. Nat Biotechnol 2004:22. https://doi.org/10.1038/nbt1000.
[32] Christ D, Famm K, Winter G. Repertoires of aggregation-resistant human antibody domains. Protein Eng, Des Sel 2007;20:413–6. https://doi.org/10.1093/PROTEIN/GZM037.
[33] Barthelemy PA, Raab H, Appleton BA, Bond CJ, Wu P, Wiesmann C, et al. Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. J Biol Chem 2008;283:3639–54. https://doi.org/10.1074/JBC.M708536200.
[34] Arbabi-Ghahroudi M, To R, Gaudette N, Hirama T, Ding W, MacKenzie R, et al. Aggregation-resistant VHs selected by in vitro evolution tend to have disulfide-bonded loops and acidic isoelectric points. Protein Eng, Des Sel 2009;22:59–66. https://doi.org/10.1093/PROTEIN/GZN071.
[35] Ignatovich O, Jespers L, Tomlinson IM, de Wildt RMT. Creation of the large and highly functional synthetic repertoire of human VH and Vκ domain antibodies. Methods Mol Biol 2012;911:39–63. https://doi.org/10.1007/978-1-61779-968-6_4/COVER/.
[36] Ma X, Barthelemy PA, Rouge L, Wiesmann C, Sidhu SS. Design of synthetic autonomous VH domain libraries and structural analysis of a VH domain bound to vascular endothelial growth factor. J Mol Biol 2013;425:2247–59. https://doi.org/10.1016/J.JMB.2013.03.020.
[37] Henry KA, Kim DY, Kandalaft H, Lowden MJ, Yang Q, Schrag JD, et al. Stability-Diversity Tradeoffs Impose Fundamental Constraints on Selection of Synthetic Human V H/V L Single-Domain Antibodies from In Vitro Display Libraries. Front Immunol 2017:8. https://doi.org/10.3389/FIMMU.2017.01759.

[38] Hussack G, Keklikian A, Alsughayyir J, Hanifi-Moghaddam P, Arbabi-Ghahroudi M, van Faassen H, et al. A VL single-domain antibody library shows a high-propensity to yield non-aggregating binders. Protein Eng, Des Sel 2012;25:313–8. https://doi.org/10.1093/PROTEIN/GZS014.
[39] Bai X, Kim J, Kang S, Kim W, Shim H. A Novel Human scFv Library with Non-Combinatorial Synthetic CDR Diversity. PLoS One 2015;10:e0141045. https://doi.org/10.1371/JOURNAL.PONE.0141045.
[40] Bujard H, Gentz R, Lanzer M, Stueber D, Mueller M, Ibrahimi I, et al. [26] A T5 promoter-based transcription-translation system for the analysis of proteins in Vitro and in Vivo. Methods Enzym 1987;155:416–33. https://doi.org/10.1016/0076-6879(87)55028-5.
[41] Gentz R, Bujard H. Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J Bacteriol 1985;164:70–7. https://doi.org/10.1128/JB.164.1.70-77.1985.
[42] Dvorak P, Chrast L, Nikel PI, Fedr R, Soucek K, Sedlackova M, et al. Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. Micro Cell Fact 2015;14:1–15. https://doi.org/10.1186/S12934-015-0393-3/FIGURES/5.
[43] Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, et al. The complete genome sequence of Escherichia coli DH10B: Insights into the biology of a laboratory workhorse. J Bacteriol 2008;190:2597–606. https://doi.org/10.1128/JB.01695-07/SUPPL_FILE/SUPPLEMENTAL_TABLE_3.ZIP.
[44] Casadaban MJ, Cohen SN. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. J Mol Biol 1980;138:179–207. https://doi.org/10.1016/0022-2836(80)90283-1.
[45] Grant SGN, Jessee J, Bloom FR, Hanahan D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. Proc Natl Acad Sci USA 1990;87:4645–9. https://doi.org/10.1073/PNAS.87.12.4645.
[46] Hanahan D, Jessee J, Bloom FR. [4] Plasmid transformation of Escherichia coli and other bacteria. Methods Enzym 1991;204:63–113. https://doi.org/10.1016/0076-6879(91)04006-A.