In Vitro Nanobody Library Construction by Using Gene Designated-Region Pan-Editing Technology

Zhiyuan Niu,1 Zhixia Luo,1 Pengyang Sun,1 Linwei Ning,1 Xinru Jin,1 Guanxu Chen,1 Changjiang Guo,1 Lingtong Zhi,1 Wei Chang,2 and Wuling Zhu 1

1Synthetic Biology Engineering Lab of Henan Province, School of Life Sciences and Technology, Xinxiang Medical University, Xinxiang, 453003 Henan, China
2Department of Oncology, Xinxiang First People’s Hospital, The Affiliated People’s Hospital of Xinxiang Medical University, Xinxiang 453000, China

Correspondence should be addressed to Zhiyuan Niu; prayernd@126.com and Wuling Zhu; wuling_zhu@163.com

Received 23 January 2022; Accepted 9 June 2022; Published 2 August 2022

Camelid single-domain antibody fragments (nanobodies) are an emerging force in therapeutic biopharmaceuticals and clinical diagnostic reagents in recent years. Nearly all nanobodies available to date have been obtained by animal immunization, a bottleneck restricting the large-scale application of nanobodies. In this study, we developed three kinds of gene designated-region pan-editing (GDP) technologies to introduce multiple mutations in complementarity-determining regions (CDRs) of nanobodies in vitro. Including the integration of G-quadruplex fragments in CDRs, which induces the spontaneous multiple mutations in CDRs; however, these mutant sequences are highly similar, resulting in a lack of sequences diversity in the CDRs. We also used CDR-targeting traditional gRNA-guided base-editors, which effectively diversify the CDRs. And most importantly, we developed the self-assembling gRNAs, which are generated by reprogrammed tracrRNA hijacking of endogenous mRNAs as crRNAs. Using base-editors guided by self-assembling gRNAs, we can realize the iteratively diversify the CDRs. And we believe the last GDP technology is highly promising in immunization-free nanobody library construction, and the full development of this novel nanobody discovery platform can realize the synthetic evolution of nanobodies in vitro.

1. Introduction

Nanobodies provide the remarkable specificity of antibodies within the 15 kDa single-variable domains (VHH) of the heavy-chain-only antibodies found in llamas and other camelids [1]. The advantages of nanobodies include their ability to bind with high affinity to epitopes that are inaccessible to traditional antibodies (≥150 kDa), better stability and the possibility of lower immunogenicity, and Lego-like modularity, and they even increase the efficacy of chimeric antigen receptor (CAR) T cells [2, 3]. There are dozens of active nanobody programs of all stripes in clinical development, covering lots of targets and indications [4–6], even in the diagnosis and treatment of COVID-19 [7, 8]. However, nanobodies discovered to date have been derived from the immunization of camelids [9], which is invariably time-consuming, expensive, and unreliable in quality. Therefore, in this study, we are dedicated to developing a novel platform technology that can be used to expedite the development of nanobodies for therapeutic and diagnostic applications by avoiding the need for animal immunization.

There are mainly two kinds of previously described strategies for in vitro library creation or introducing variation in antibody complementarity-determining region (CDR) loops: the static antibody libraries and the dynamic antibody libraries. The static antibody libraries’ sequence diversity remains constant after it is generated. This library construction scheme mainly relies on the artificial synthesis of antibody sequences and the artificial introduction of sequence diversity in the CDRs of antibodies. For example, libraries that are generated by amplifying CDR sequences from natural antibodies before grafting them into recombinant display systems for selection in vitro [10], using trinucleotide assembly of mutagenic oligonucleotides to create naive libraries on a single antibody framework (i.e., antibody constant region) or a collection of antibody frameworks [11–13], using
synthetic libraries in which the antigen-binding sites were generated by diversifying only four kinds of amino acids (Tyr, Ser, Ala, and Asp) in CDRs [14], and generating synthetic antibody libraries, whose CDRs contain only two kinds of amino acids, including tyrosine and serine [15, 16]. By using these methods, we can accurately limit mutations to specified regions, like CDRs, or design positional amino acid frequencies in CDRs. However, this library building scheme strongly relies on synthetic antibody sequences and artificially introduces sequence diversity in CDRs. Also, binders acquired from these libraries are inferior in quantity and affinity to those obtained from the natural immune system. The dynamic antibody libraries’ antibody diversity is continuously generated. For example, use the most mutagenic RNA viral class, Alphavirus sindbis viruses as vectors for heredity and diversity, which achieved 24-hour selection cycles surpassing $10^3$ mutations/base [17]. Libraries can also be constructed by coexpressing IgG, for example, the heavy chain (HC), light chain (LC), and activation-induced cytidine deaminase (AID) in mammalian cells. The expression of AID alone is sufficient to reproduce the salient features of somatic hypermutation (SHM) in both B cells and other mammalian cells [18, 19]. These currently available dynamic library construction methods allow an iterative fashion to diversify, display, and screen antibodies and more importantly, antibody affinity maturation via iterative cycles of antigen-coated magnetic beads sorting or fluorescence-activated cell sorting (FACS) under increasingly stringent sort conditions. This dynamic library allows for iteratively diversifying the antibody sequences; however, with a relatively low mutation efficiency, the mutations generated by RNA viruses and AID can hardly restrict to a small window of nucleotides within CDRs, and constant region mutations may cause antibody dysfunction.

In summary, it is necessary to further develop a powerful in vitro library construction platform that combines the advantages of both precise mutations of CDRs in currently available static antibody libraries and iterative library building characteristics of dynamic antibody libraries. And this platform mimics the fine natural immune system and continuously generates diverse nanobodies for later displaying, screening, and maturation.

2. Materials and Methods

2.1. Reagents. Puromycin (P8230), Trytome (T8490), Agar Powder (A8190), Yeast Extract Powder (01-012), DNA purification kit (D1300), NaCl (S8210), Trypsin 0.25% (T1360), and PBS (D1040) were from Solarbio, Beijing, China. Stbl3 Chemically Competent Cells (KSTM110L) were from Alpala-Life, Shenzhen, China. Polybrene (40804E576) and Hieff Canace® High-Fidelity DNA Polymerase (10148ES10) were from Yeasen Biotech, Shanghai, China. The Tissue gDNA Isolation Kit (BW-GD2211) was from BIOMiGA, Zhe Jiang, China. EcoRI (1040S), Xba I (1093A), and T4 ligase (2011A) were form TAKARA, Da Lian, China. SanPrep Column DNA Gel Extraction Kit (B518131) and Pfu DNA Polymerase were purchased from Sangon Biotech, Shanghai, China. The E.Z.N.A.® Plasmid Midi Kit I (D6926-03) was purchased from Omega Bio-Tek, Guangzhou, China. Green Taq Mix (P131-01) was from Vazyme Biotech Co., Ltd, Nanjing, China. Goat anti-llama IgG H&L (FITC) (ab112785) was from Abcam, Shanghai, China. The Calcium Phosphate Cell Transfection Kit (C0508) was from Beyotime, Haimen, PR China.

2.2. Cell Culture. Dulbecco’s modified Eagle’s medium (DMEM) and fetal bovine serum (FBS) were from Gibco, Thermo Fisher, Shanghai, China. The HEK293T (GDC0187) cell line was obtained from China Center for Type Culture Collection (CCTCC, Wuhan, China) and cultured according to their instructions. All media were supplemented with 1% penicillin-streptomycin (C0222, Beyotime, Haimen, China) and 10% FBS. All experiments were performed with mycoplasma-free cells.

2.3. Plasmids. pCMV-BE3 (Addgene plasmid # 73021) was a gift from David Liu. Genes expressing LaG-2, LaG-2/G4, LaG-2/G4.1, LaG-2/G4.2, LaG-2/G4.3, AIDmut1, AIDmut2, and membrane-expressed LaG-2 (mLaG-2) were directly synthesized (GenScript, Nanjing, China) and subcloned into pCDH-CMV-MCS-EF-1α-Puro. Genes expressing 3×gRNA, 3×gRNAΔ, 3×gRNAΔ15, and 3×gRNAΔ21 were directly synthesized (GenScript, Nanjing, China) and subcloned into pLL7-U6-MCS-EF-1α-Blast. The full sequences of these genes or plasmids were provided in Table S1.

2.4. Lentivirus Packaging and Infection. Transfection of HEK293T was performed by using Calcium Phosphate Cell Transfection Kit according to the manufacturer’s instructions. We use the vector for expressing LaG-2/G4 (pCDH-CMV-LaG-2/G4-EF-1α-Puro), LaG-2/G4.1, LaG-2/G4.2, LaG-2/G4.3, or LaG-2 (pCDH-CMV-LaG-2-EF-1α-Puro) viral packaging vector (psPAX2):viral envelope vector (pMD2G) at 4:3:1 ratio. Lentivirus containing supernatant was harvested at both 48 and 72 h after transfection and was filtered using a 0.45 μm filter. The acquired lentivirus was mixed with the PEG solution (5% PEG8000 and 0.5 M NaCl) overnight and concentrated by centrifugation at 3000 g for 30 min at 4°C.

Dilution of lentivirus in DMEM complete +10 μg/ml polybrein was used for HEK293T infection. HEK293T expressing LaG-2/G4, LaG-2/G4.1, LaG-2/G4.2, LaG-2/G4.3, or LaG-2 were initially selected on 5 μg/ml puromycin and subsequently grown in the presence of 1 μg/ml puromycin.

2.5. HEK293T Transfection. Transfection of HEK293T was performed by using the Calcium Phosphate Cell Transfection Kit according to the manufacturer’s instructions. 1 × 10^6 HEK293T cells expressing LaG-2 were seeded on 6-well plates. At 1 d, 3 d, and 5 d postseeding, cells were transfected 3 times with 4.5 μg plasmids expressing BE3, AIDmut1, or AIDmut2 base-editors, and 1.5 μg plasmids expressing 3×gRNA, 3×gRNAΔ, 3×gRNAΔ15, or 3×gRNAΔ21.

2.6. Genomic DNA Extraction and Polymerase Chain Reaction. Genomic DNA was isolated from HEK293T cells 48 h after the third time of transfection by using the Tissue gDNA Isolation Kit according to the manufacturer’s instructions. Three nested
pairs of primers were designed for LaG-2/G4, LaG-2/G4.1, LaG-2/G4.2, and LaG-2/G4.3 PCR amplification. The primer sequences used for nested PCR were as follows: Nest 1F: 5′-AGTCGGTTATAGAACGTCG-3′, Nest 1R: 5′-CACA GCAGCCGTTGAACTTG-3′, Nest 2F: 5′-TCCAGGTGG TGTTGACATGC-3′, Nest 2R: 5′-CACCTGGAGGA GTGCTC-3′, Nest 3F: 5′-ATATTCTAGGCGCTTGTC TGCTGAC-3′, and Nest 3R: 5′-ATATGAATTCTCG CGAGTCGAGGT-3′. The primer combination was Nest 3F/Nest 3R for LaG-2 PCR amplification. The reaction conditions consisted of initial denaturation at 98 °C for 3 min, 35 cycles of 98 °C for 10 s, 68 °C for 40 s, and a final extension step for 5 min at 72 °C.

2.7. Flow Cytometry. Flow cytometry samples were run on a Guava easyCyte 8HT. The data were analyzed by using the FlowJo software (v. 7.6.4, TreeStar). The statistics presented are based on at least 10,000 events gated on the population of interest.

2.8. Experimental Repeats and Statistics. All assays had been performed at least thrice to ensure the repeatability of the experiments. The representative pictures and statistical analysis were shown in the final figures. Data for all experiments were analyzed with the Python applets we wrote.

3. Results

3.1. Nanobody Library Construction by Using G-Quadruplex Fragments. Recently, the Liu group developed a series of base-editors, which could benefit point mutation correction by efficiently converting one base pair to a different base pair, such as adenine base-editors (ABEs) [20] and cytosine base-editors (CBEs) [21] or benefit disruption of specific regions and multiplex base-editing applications, such as the ABE8e base-editor [22]. These base-editors will not induce double-stranded DNA breaks or extensive insertions and deletions (indels) to target sites, cause frameshift mutations, and yield a nonsense protein. These advantages make the base-editor an excellent choice for creating mutations in antibody CDRs. However, base-editors only operate on single-stranded DNA but reject double-stranded DNA. Still, this feature is critical to restrict deaminase activity to a small window of nucleotides within the single-stranded DNA structures.

G-quadruplex is a specialized DNA secondary structure consisting of a sequence of guanine-rich nucleic acids. The sequence forming a G-quadruplex requires at least 4 consecutive groups of G bases, joined by 1-7 arbitrary bases and not less than 2 G bases per group (G(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)nG(2)) [23, 24]. The formation of the G-quadruplex allows its complementary strand to exist in a single-stranded state, so gene editing can be performed within the window by finding a protein that directs deaminase to the complementary single strand of the G-quadruplex.

Based on the above facts, the first nanobody library construction strategy we tried was starting from a consensus framework derived from an anti-GFP nanobody LaG-2 [25]. CDR1, CDR2, and CDR3 of nanobody LaG-2 were replaced by the natural G-quadruplex fragments from VEGF, c-myc, and BCL-2 genes, respectively [26], modified LaG-2 was named LaG-2/G4, and the sequence was shown in Figure S1A. And base-editors were directed to the single-strand window by nucleolin, which promotes the formation of G-quadruplexes and stabilizes the structure of G-quadruplexes when it is bound to them. Therefore, it stands to reason that base-editors tethered to nucleolin could be directed to G-quadruplex fragments containing CDRs of LaG-2, as shown in schematic Figure S1B. However, we found that G-quadruplex sequences containing nanobodies could spontaneously generate mutations without the use of nucleolin-tethered base-editors when nanobody sequences were integrated into 293T cells genomes by lentiviral infection. The results showed that the mutations were mainly concentrated in three CDRs, with only a few mutations outside, the overall mutation rate was 161/1000 bp-1 (per bp) (Figure 1), and data of the mutation sites and types, together with the ratio of four nucleotides before and after mutation, were shown in Figure S2. Also, we noticed 3 or 6 amino acids deletions in CDR2; however, these deletions did not cause frameshift mutations. Unfortunately, these mutations lack randomness, and both the mutation sites and the mutated amino acids are preferential. These characteristic are detrimental to the library construction. Moreover, we proved that LaG-2/G4 is stable in the Stbl3 E. coli strain (Figure S3), and high-fidelity PCR is capable of introducing a small number of mutations but far fewer than those introduced by amplification of LaG-2/G4 in 293T cells (Figure S4). To further clarify the pattern of mutations induced by G-quadruplex fragments, we tried to switch CDR2 and CDR3 regions of LaG-2/G4 (Figure S5A) or replaced CDR2 containing G-quadruplex fragments with another G-quadruplex fragment from HIF-1α (Figure S5B) or RET gene (Figure S5C). Then these LaG-2/G4 variants were amplified by Hieff Canace® High-Fidelity DNA Polymerase, and mutations generated by PCR were shown. Moreover, CDR2 containing G-quadruplex fragments from RET genes were amplified by Pfu DNA Polymerase (Figure S5D). However, all these trials failed to improve the diversity of G-quadruplex-induced mutations. More importantly, G-quadruplexes are difficult to amplify in PCR reactions, so it is possible that there is mutant-preferred amplification, and the mutations generated in PCR reactions or HER293T cells are much less than they appeared. Moreover, error-prone PCR has the potential to increase G-quadruplex-induced mutations; however, this does not meet our initial purpose, which involved the G-quadruplex and base-editor. Also, error-prone PCR would accumulate mutations in the constant regions of the nanobody, which could potentially lead to the nanobody's incorrect folding and loss of function. So we quit our plan to use G-quadruplex to induce diversity in the CDRs of nanobodies in this study.

3.2. Nanobody Library Construction by Using Traditional gRNA-Guided and nCas9-Tethered Base-Editors. Besides G-quadruplex, Cas9/gRNA complexes are also perfect choices to create single-stranded bubbles on nanobody CDRs. So in
this section, we try to use gRNAs to direct Cas9-tethered base-editors to nanobody CDRs, as shown in Figure 2(a). The gRNA/Cas9 base-editor complex is recruited to the target sequence by the base-pairing between the gRNA sequence and the complement to the target sequence in the genomic DNA, and then multiple mutations were generated. To create more mutations, we decide to diversify all three CDRs on nanobodies, and this requires at least three gRNAs. Here, we use a polycistronic-tRNA-gRNA (PTG) strategy [27] to quickly assemble three gRNAs in one construct, in which, three CDR-targeting gRNAs are flanked by glycine tRNAs to create polycistronic glycine tRNA-gRNA constructs. Taking advantage of the endogenous tRNA processing system in mammalian cells, we can efficiently transcribe three CDR-targeting gRNAs driven by only one Pol III promoter (Figure 2(b)). Three different nCas9-tethered base-editors were evaluated for nanobody library construction (Figure 2(c)). The BE3 base-editor (APOBEC–XTEN–nCas9–UGI) was engineered in the lab of David R. Liu [21]: AIDmut1 base-editor (AID*Δ–nCas9), in which, AID*Δ is a mutated wild type AID (K10E+T82I+E156G) with partial deletion (196-198AA) of its C terminus nuclear exporting sequence (NES) which ablates its nuclear export signal while increasing somatic hypermutation (SHM). AID*Δ was developed by Hess et al. [28], and we tethered this AID variant to nCas9, thus forming the AIDmut1 base-editor; the last base-editor is AIDmut2, which is generated by removing the full length of NES (183-198AA) from the AIDmut1 base-editor, and deletion of the full-length NES was proved to create more diversification on target site than partial deletion, according to Ma et al. [29].

To test the resulting base-editors, we transfected HEK293T cells which preintegrated the LaG-2 nanobody sequence in the genome, with two-plasmid mixtures in which one plasmid expresses three CDRs-targeting gRNAs and another expresses the base-editor. Transfection was conducted three times, to facilitate the accumulation of mutations. The results demonstrated that the overall mutation rates of BE3, AIDmut1, and AIDmut2 were 2.92/1000 bp$^{-1}$, 10.2/1000 bp$^{-1}$, and 6.23/1000 bp$^{-1}$, respectively. AIDmut1 and AIDmut2 substantially increased the base-editing efficiency compared with BE3. SHM introduces nucleotide alterations at the V regions of heavy and light chain genes at a rate of 1/1000 bp$^{-1}$ [30], which is enough to enable the selection of B cells producing high-affinity antibodies. So, the mutation rates of tested base-editors, especially AIDmut1 and AIDmut2, were far beyond SHM. The base-editing active window spans from −10 to −50 (counting the PAM as positions 1–3) at three CDRs loci (Figures 2(d)–2(f)). Moreover, BE3 and AIDmut1 caused mainly C-to-T conversion (60% and 55.81%, respectively), while AIDmut2 was mainly G-to-A mutation (55.26%). Meanwhile, low levels of unexpected A/T conversion or deletion were observed for BE3 and AIDmut2. Notably, rare long-range fragment deletion was observed for AIDmut1 and AIDmut2 (Figure S6); this is not likely the features of base-editors, so we excluded these minor sequences from analysis in Figures 2(e) and 2(f). And there are no significant differences between the ratio of four nucleotides before and after mutation for all of the base-editors (Figures 2(d)–2(f)). In the collection, we propose that expressing AIDmut1 and AIDmut2 base-editors were feasible strategies for nanobody library construction in vitro.

3.3. Nanobody Library Construction by Using Self-Assembling gRNA-Guided and nCas9-Tethered Base-Editors. Based on the above findings, we know that the expression of three gRNAs can effectively guide the base-editors to the three CDRs and cause mutations in the gRNA recognition region and adjacent sequences (Figure S7). However, gRNA-target recognition relies mainly on the 20 bp seed region at the 5’ end, and a mutation of only 2 bp in the seed region results in a 96% loss of gRNA guidance function [31]. Also, our data demonstrated that mutations generated by AID mutants pile up at a very slow rate. On average, there are only 4 mutation sites on each nanobody sequence, after three times of transfection and mutation (Figure S7). Therefore, when base-editors are guided by the conventional gRNAs, mutations that emerged from the first round of library construction will prevent the generation of more mutations in the next round.

To iteratively diversify the CDRs, firstly, we reviewed the process of generation of conventional gRNA [32], which is a combination of the endogenous bacterial CRISPR RNA
(crRNA) and transactivating crRNA (tracrRNA) into a single chimeric guide RNA (gRNA) transcript. The gRNA combines the targeting specificity of the crRNA with the scaffolding properties of the tracrRNA into a single transcript (Figure 3(a), upper panel). We speculate that the formation of the functional CRISPR/Cas9 complex has no mandatory requirement for the crRNA-tracrRNA-hybridized sequence. This part of the sequence only needs to satisfy the stem-loop structure formed between crRNA and tracrRNA in traditional gRNAs, and we can take advantage of this feature to construct the self-assembling gRNAs. We designed truncated gRNA (gRNAΔ) which does not contain the crRNA part of conventional gRNA. Also, the 14 bp 5′ end tracrRNA sequence that pairs with crRNA was modified, which now pairs with the 3′ constant region adjacent to the CDRs of nanobody mRNA (Figure 3(a), lower panel, and Figure 3(b)). After the pairing of gRNAΔ and the nanobody mRNA, endogenous mammalian RNases will cleave mRNA (generating truncated-mRNA, mRNAΔ) and assist maturation of the gRNAΔ: mRNAΔ duplex (self-assembling gRNA), just as
RNase III cleaves pre-crRNA base-paired with transactivating crRNA (tracrRNA) in the presence of Cas9 [33, 34]. The mature self-assembling gRNAs then guide nCas9-tethered base-editors to create mutations on target DNA sites. Most importantly, the seed regions of our self-assembling gRNAs always come from the latest mutated mRNA, while maintaining self-targeting ability, enabling continuously diversifying the CDRs.

The base-editing efficiency of self-assembling gRNA-guided base-editors was tested under the same condition when using conventional gRNAs. The results demonstrated that the overall mutation rates of AIDmut1 and AIDmut2 were 6.82/1000 bp\(^{-1}\) (Figure 3(c)) and 2.36/1000 bp\(^{-1}\) (Figure 3D), respectively. Both substantially decreased the base-editing efficiency compared with when they were guided by the conventional gRNAs. And more importantly, most of the mutations generated by AIDmut1 and AIDmut2 are outside the CDRs, which is unfavorable for an antibody library building. We speculate that the self-assembling gRNAs may not be processed smoothly and produced three

**Figure 3:** Gene designated-region pan-editing mediated by self-assembling gRNA-guided and nCas9-tethered base-editors. (a) Structures of conventional gRNA and truncated gRNA. (b) Schematic summary of the formation of three self-assembling gRNAs. (c–f) HEK293T cells containing LaG-2 were transfected with indicated combinations of (c) AIDmut1 and \(3 \times \text{gRNA}_\Delta\), (d) AIDmut2 and \(3 \times \text{gRNA}_\Delta\), (e) AIDmut1 and \(3 \times \text{gRNA}_{\Delta 21}\), and (f) AIDmut2 and \(3 \times \text{gRNA}_{\Delta 21}\) for 3 times; then the LaG-2 loci were sequenced. Graphs of the enrichment of the mutations generated by AIDmut1 and AIDmut2 are outside the CDRs, which is unfavorable for an antibody library building. We speculate that the self-assembling gRNAs may not be processed smoothly and produced three
Overall mutation rates of AIDmut1 and AIDmut2 were 9.97/1000 bp.

So we tested the base-editing ability of AIDmut1 and AIDmut2 guided by conventional gRNA directed by truncated gRNAs. Then mLaG-2 expression was evaluated by flow cytometry, and we did not observe any loss in nanobody expression (Figures S8B and S8C). We speculate that this is because the U6 promoter is a weak promoter and only a small amount of gRNAΔ21 will be expressed, then hijack only a few endogenous mRNAs, and thus do not have a significant impact on the expression of nanobodies. We also observed the production of a small number of stop codons (Figures S8D and S8E), which again have not impaired the expression of the nanobodies as shown in Figures S8B and S8C.

To further confirm the iterative evolutionary capacity of our self-assembling gRNAs, we mutated 2 bases on the 20 bp seed region recognition site. Then, we tested the base-editing ability of AIDmut1 and AIDmut2 guided by 5′ end prolonged 15 bp gRNAΔ (gRNAΔ15). As expected, mutation of the seed region recognition site will not affect the binding of truncated gRNAs. Base-editors directed by self-assembling gRNAs substantially increased the base-editing efficiency compared with base-editors directed by conventional gRNAs; moreover, mutations are more concentrated in CDRs for base-editors guided by truncated gRNAs.

![Comparison of mutation tolerance for base-editors guided by conventional gRNAs and self-assembling gRNAs.](image)

**Figure 4:** Comparison of mutation tolerance for base-editors guided by conventional gRNAs and self-assembling gRNAs. (a–d) HEK293T cells containing LaG-2 which contains 2 mutated bases on the 20 bp seed region recognition site (mutation sites marked with two black triangles) were transfected with indicated combinations of (a) AIDmut1 and 3xgRNA, (b) AIDmut2 and 3xgRNA, (c) AIDmut1 and 3xgRNAΔ15, and (d) AIDmut2 and 3xgRNAΔ15 for 3 times; then the LaG-2 loci were sequenced. Graphs of the enrichment of mutation at each base are shown here; we also indicate the type and number of mutations in the lower-left corner, and "Δ" means deletion mutation. The pie chart represents the change in the ratio of the four bases before and after the mutation. In (a–d), mutations from at least 10 LaG-2 sequences were analyzed.
the overall mutation rates of AIDmut1 and AIDmut2 directed by conventional gRNAs were 1.31/1000 bp$^{-1}$ (Figure 4(a)) and 1.67/1000 bp$^{-1}$ (Figure 4(b)), and the CDR rates were 0.00% and 14.29%, respectively. The overall mutation rates of AIDmut1 and AIDmut2 directed by gRNAΔ15 were 9.71/1000 bp$^{-1}$ (Figure 4(c)) and 7.87/1000 bp$^{-1}$ (Figure 4(d)), and the CDR rates were 40.54% and 39.39%, respectively. Thus, by using self-assembling gRNAs, we can realize the continuous diversification of nanobody CDRs.

4. Discussion

The emerging and developing of modern biotechnologies, such as synthetic biology, single base-editing, and photosensitive bioswitch, raise the feasibility of artificial evolution of molecules, cells, and even organisms. To realize the controlled evolution of nanobodies in vitro, we developed several gene designated-region pan-editing (GDP) technologies for library construction. We found that the self-assembling gRNA-guided base-editor is the most promising way to iteratively diversify the CDRs and construct the fully synthetic nanobody library. In each round of mutation, the effective library size was maintained at >10$^7$. In the future, nanobody libraries acquired from a consensus framework derived from a humanized nanobody will greatly improve the safety of clinical applications of nanobodies.

However, more improvements are needed to totally abandon animal immunization and switch to in vitro techniques for producing nanobodies. For example, we still need to improve the efficiency and accuracy of self-assembling gRNA-guided base-editors. Moreover, we only tested CBEs in this study, to build more diverse nanobody libraries, and we also need efficient ABEs, but current ABES are designed to correct for point mutations and thus cannot induce a wide range of mutations in specified regions of the gene. Therefore, we still need to redevelop the ABEs according to the requirements of library construction.

In this study, we observed theoretically and practically the potential of self-assembling gRNAs to consistently introduce random sequences in the gene designated-region. This has important implications not only for the preparation of nanobodies but also for the future introduction of artificial intelligence concepts into synthetic biology. For example, this technology allows the introduction of random function in cells, where we can randomly knock out or mutate any gene on the genome by reutilization of these random sequences with other Cas family members and other self-assembling gRNAs based on these random sequences. Then, by integrating existing photosensitive bioswitches and logic gates for signals input and output, we can design intelligent cells and obtain the desired protein molecules or cellular functions.

Data Availability

DNA sequence data is supplied within the paper and in the supplementary materials.

Conflicts of Interest

The authors report no conflicts of interest regarding the publication of this article.

Authors’ Contributions

ZYN and WLZ conceived experiments and analyzed data. ZYN, ZXL, PYS, LWN, XRJ, GXC, CJG, LTZ, and WC carried out the experiments. ZYN, WLZ, CJG, and LTZ acquired funding. ZYN and WLZ supervised the study. All authors were involved in writing the paper and had final approval of the submitted and published versions.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant 81902916), Natural Science Foundation of Henan Province (222300420513), National Key R&D Program of China (2019YFA0906000), and National Natural Science Foundation of China (Grant 81903187 and 82003261). We thank members of Zhu’s laboratory for helpful discussions, technical assistance, and critical reading of the manuscript.

Supplementary Materials

Figure S1: G-quadruplex and nucleolin-tethered base-editor-mediated GDP technology. Figure S2: enrichment of the LaG-2/G4 spontaneous mutation at each base in HEK293T cells. Figure S3: enrichment of the LaG-2/G4 spontaneous mutation at each base in Stbl3. Figure S4: mutations on LaG-2/G4 generated by Hieff Canace® High-Fidelity DNA Polymerase. Figure S5: mutations on LaG-2/G4 variants generated by different High-Fidelity DNA Polymerases. Figure S6: amino acid mutations on LaG-2 that were generated by conventional gRNA-guided AIDmut1 and AIDmut2. Figure S7: mutations on LaG-2 DNA generated by conventional gRNA-guided AIDmut1. Figure S8: the characteristics of 3×gRNAΔ21-guided base-editors. Table S1: the full sequences of genes, plasmids, and mLaG-2. (Supplementary Materials)

References

[1] C. Hamers-Casterman, T. Atarhouch, S. Muyldermans et al., “Naturally occurring antibodies devoid of light chains,” Nature, vol. 363, no. 6428, pp. 446–448, 1993.
[2] C. Morrison, “Nanobody approval gives domain antibodies a boost,” Nature Reviews Drug Discovery, vol. 18, no. 7, pp. 485–487, 2019.
[3] F. Rahbarizadeh, D. Ahmadvand, and S. M. Moghimi, “CAR T-cell bioengineering: single variable domain of heavy chain antibody targeted CARs,” Advanced Drug Delivery Reviews, vol. 141, pp. 41–46, 2019.
[4] S. Sun, Z. Ding, X. Yang et al., “Nanobody: a small antibody with big implications for tumor therapeutic strategy,” International Journal of Nanomedicine, vol. Volume 16, pp. 2337–2356, 2021.
[5] G. Bao, M. Tang, J. Zhao, and X. Zhu, “Nanobody: a promising toolkit for molecular imaging and disease therapy,” *EJNMMI Research*, vol. 11, no. 1, p. 6, 2021.

[6] I. Jovcevska and S. Muyldermans, “The therapeutic potential of Nanobodies,” *BioDrugs*, vol. 34, no. 1, pp. 11–26, 2020.

[7] T. Güttler, M. Aksu, A. Dickmanns et al., “Neutralization of SARS-CoV-2 by highly potent, hyperthermostable, and mutation-tolerant nanobodies,” *The EMBO Journal*, vol. 40, no. 19, article e107985, 2021.

[8] K. Guo, S. Wustoni, A. Koklu et al., “Rapid single-molecule detection of COVID-19 and MERS antigens via nanobody-functionalized organic electrochemical transistors,” *Nature Biomedical Engineering*, vol. 5, no. 7, pp. 666–677, 2021.

[9] E. Pardon, T. Laeremans, S. Triest et al., “A general protocol for the generation of Nanobodies for structural biology,” *Nature Protocols*, vol. 9, no. 3, pp. 674–693, 2014.

[10] R. M. Hoet, E. H. Cohen, R. B. Kent et al., “Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity,” *Nature Biotechnology*, vol. 23, no. 3, pp. 344–348, 2005.

[11] J. J. Adams and S. S. Sidhu, “Synthetic antibody technologies,” *Current Opinion in Structural Biology*, vol. 24, pp. 1–9, 2014.

[12] S. Moutel, N. Bery, V. Bernard et al., “NaLi-H1: a universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies,” *eLife*, vol. 5, 2016.

[13] C. McMahon, A. S. Baier, R. Pascolutti et al., “Yeast surface display platform for rapid discovery of conformationally selective nanobodies,” *Nature Structural & Molecular Biology*, vol. 25, no. 3, pp. 289–296, 2018.

[14] F. A. Fellouse, P. A. Barthelemy, R. F. Kelley, and S. S. Sidhu, “Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code,” *Journal of Molecular Biology*, vol. 357, no. 1, pp. 100–114, 2006.

[15] F. A. Fellouse, B. Li, D. M. Compaan, A. A. Peden, S. G. Hyrnowitz, and S. S. Sidhu, “Molecular recognition by a binary code,” *Journal of Molecular Biology*, vol. 348, no. 5, pp. 1153–1162, 2005.

[16] F. A. Fellouse, K. Esaki, S. Birtalan et al., “High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries,” *Journal of Molecular Biology*, vol. 373, no. 4, pp. 924–940, 2007.

[17] J. G. English, R. H. J. Olsen, K. Lansu et al., “VEGAS as a platform for facile directed evolution in mammalian cells,” *Cell*, vol. 178, no. 3, pp. 748–761.e17, 2019.

[18] P. M. Bowers, R. A. Horlick, T. Y. Neben et al., “Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 51, pp. 20455–20460, 2011.

[19] P. M. Bowers, R. A. Horlick, M. R. Kehry et al., “Mammalian cell display for the discovery and optimization of antibody therapeutics,” *Methods*, vol. 65, no. 1, pp. 44–56, 2014.

[20] N. M. Gaudelli, A. C. Komor, H. A. Rees et al., “Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage,” *Nature*, vol. 551, no. 7681, pp. 464–471, 2017.

[21] A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage,” *Nature*, vol. 533, no. 7603, pp. 420–424, 2016.

[22] M. F. Richter, K. T. Zhao, E. Eton et al., “Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity,” *Nature Biotechnology*, vol. 38, no. 7, pp. 883–891, 2020.

[23] D. Rhodes and H. J. Lipps, “G-quadruplexes and their regulatory roles in biology,” *Nucleic Acids Research*, vol. 43, no. 18, pp. 8627–8637, 2015.

[24] D. Ji, M. Juhas, C. M. Tsang, C. K. Kwok, Y. Li, and Y. Zhang, “Discovery of G-quadruplex-forming sequences in SARS-CoV-2,” *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 1150–1160, 2021.

[25] P. C. Friddy, Y. Li, S. Keegan et al., “A robust pipeline for rapid production of versatile nanobody repertoires,” *Nature Methods*, vol. 11, no. 12, pp. 1253–1260, 2014.

[26] V. González, K. Guo, L. Hurley, and D. Sun, “Identification and characterization of nucleolin as a c-myc g-quadruplex-binding protein,” *Journal of Biological Chemistry*, vol. 284, pp. 23622–23635, 2009.

[27] K. Xie, B. Minkenberg, and Y. Yang, “Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 11, pp. 3570–3575, 2015.

[28] G. T. Hess, L. Fréard, K. Han et al., “Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells,” *Nature Methods*, vol. 13, no. 12, pp. 1036–1042, 2016.

[29] Y. Ma, J. Zhang, W. Yin, Z. Zhang, Y. Song, and X. Chang, “Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells,” *Nature Methods*, vol. 13, no. 12, pp. 1029–1035, 2016.

[30] K. Rajewsky, I. Förster, and A. Cumano, “Evolutionary and somatic selection of the antibody repertoire in the mouse,” *Science*, vol. 238, no. 4830, pp. 1088–1094, 1987.

[31] E. M. Anderson, A. Haupt, J. A. Schiel et al., “Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity,” *Journal of Biotechnology*, vol. 211, pp. 56–65, 2015.

[32] L. Cong, F. A. Ran, D. Cox et al., “Multiplex genome engineering using CRISPR/Cas systems,” *Science*, vol. 339, no. 6121, pp. 819–823, 2013.

[33] E. Delcheva, K. Chylinski, C. M. Sharma et al., “CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III,” *Nature*, vol. 471, pp. 602–607, 2011.

[34] K. Chylinski, A. Le Rhun, and E. Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems,” *RNA Biology*, vol. 10, no. 5, pp. 726–737, 2013.