Supporting Information

In Vitro Construction of Large-scale DNA Libraries from Fragments Containing Random Regions Using Deoxyinosine-containing Oligonucleotides and Endonuclease V

Yasuhide Yamamoto,[†,**] Takuya Terai,[†,**] Shigefumi Kumachi,[‡] and Naoto Nemoto*[,†,‡]

[†]Graduate School of Science and Engineering, Saitama University, 255 Shimo-okubo, Sakura-ku, Saitama City 338-8570, Japan.
[‡]Epsilon Molecular Engineering, Inc., 255 Shimo-okubo, Sakura-ku, Saitama City 338-8570, Japan.

*E-mail: nemoto@fms.saitama-u.ac.jp (N.N.).
** These authors contributed equally to this work.

Supplementary Tables

Table S1 Primers used in this study (from 5′ to 3′).

Name	Sequence (5′ to 3′)
5′ primer	GATCCCGCGAAATTAATACGACTCACTATAGGG
3′ primer	TTTCCACGCCGCCCCCCGTCCT
frag_A_R	TCAG[I]CGCAGGCTTCCTCCAGCCTGCAC
frag_B_F1	GCCT[I]AGCTGCGCTGCTAGCGGA
frag_B_R1	TCACGCTGTC[I]GCGTAGTAGGT
frag_B_F2	GCCT[I]AGCTGCGCCGCTAGCGGAXYZXYZXYZXYZXYZXYZXYZXYZX YZXYZXYZXYZXYZXYZXYZXYZXYZTGGTTCCGCCAGGCTCCTG
frag_B_R2	TCACGCTGTC[I]GCGTAGTAGGTZYXZYXZYXZYXZYXZYXZYXZYXZY XZYXZYXZYXZYXZYXZYXZYXCACTCCCTCGCGCTCCTTTC
frag_C_F	CCGACAGCGT[I]AAGGGACGCTTCACCA
frag_C_R	TCCCCTG[I]CCCCAGTAZYXZYXZYXZYXZYXZYXZYXZYXZYXZYXZY XZYXZYXZYXZYXZYXAGCAGCGCAGTAGTAGATAGCGG
frag_D_F	GCCAGGG[I]ACCCAGGTGACCGTG
M13_F	GTTTTCCCAGTCACGAC
M13_R	CAGGAAACAGCTATGAC

Note: [I] indicates deoxyinosine. XYZ and $\dot{Z}\dot{Y}\dot{X}$ are mixed codons whose composition is described in Figure S2. $\dot{Z}\dot{Y}\dot{X}$ are complementary to XYZ.

Supplementary Figures

DNA sequence (written as 50 nt/line)

GATCCCGCGA AATTAATACG ACTCACTATA GGGGAAGTAT TTTTACAACA
ATTACCAACA ACAACAACAA ACAACAACAA CATTACATTT TACATTCTAC
AACTACAAGC CACCATGGGC GAGGTGCAGC TGGTGGAGAG CGGAGGAGGA
TCCGTGCAGG CTGGAGGAAG CCTGCGCCTG AGCTGCGCTG CTAGCGGAAC
AACATCTATC GATCCTCATT CTTCCTGGTT CCGCCAGGCT CCTGGAAAGG
AGCGCGAGGG AGTGTTATCT GCCAGCATCA ACTTCGACAC CTACTACGCT
GACAGCGTGA AGGGACGCTT CACCATCAGC CAGGACAACG CCAAGAACAC
CGTGTACCTG CAGATGAACA GCCTGAAGCC AGAGGACACC GCTATCTACT
ACTGCGCTGC TGATGCTACT TCTAACCTTG GGACTACCTC GGATTTGCTT
TACTGGGGAC AGGGAACCCA GGTGACCGTG GGAGGAGGCA GCCATCATCA
TCATCATCAC GGCGGAAGCG ACTACAAGGA TGACGATGAC AAGGGTGGTA
GCAGGACGGG GGGCGGCGGG GAAA

Annotation

Region in nt	Explanation
14–33	T7 promotor
37–107	Omega enhancer sequence
110–114	Kozak
115–480	VHH
199–225	CDR1
265–288	CDR2
412–450	CDR3
481–519	His_6 tag + GS linker
520–552	Flag tag + GGS
553–574	Linker hybridization region

Figure S1. Original DNA sequences of a VHH used as a template of PCR. Underlined sequences correspond to the junction sites. Yellow highlighted bases are hybridized to I in the corresponding primers (Table S1).

DNA sequence (written as 50 nt/line)

GATCCCGCGA AATTAATACG ACTCACTATA GGGGAAGTAT TTTTACAACA
ATTACCAACA ACAACAACAA ACAACAACAA CATTACATTT TACATTCTAC
AACTACAAGC CACCATGGGC GAGGTGCAGC TGGTGGAGAG CGGAGGAGGA
TCCGTGCAGG CTGGAGGAAG CCTGCGCCTG AGCTGCGCTG CTAGCGGA**XY**
ZXYZXYZXYZ XYZXYZXYZX YZXYZXYZXY ZXYZXYZXYZ XYZXYZTGGT
TCCGCCAGGC TCCTGGAAAG GAGCGCGAGG GAGTG**XYZXY ZXYZXYZXYZ**
XYZXYZXYZX YZXYZXYZXY ZXYZXYZXYZ XYZACCTACT ACGCCGACAG
CGTGAAGGGA CGCTTCACCA TCAGCCAGGA CAACGCCAAG AACACCGTGT
ACCTGCAGAT GAACAGCCTG AAGCCAGAGG ACACCGCTAT CTACTACTGC
GCTGCT**XYZX YZXYZXYZXY ZXYZXYZXYZ XYZXYZXYZX YZXYZXY**
ZXYZTACTGG GGCCAGGGGA CCCAGGTGAC CGTGGGAGGA GGCAGCCATC
ATCATCATCA TCACGGCGGA AGCGACTACA AGGATGACGA TGACAAGGGT
GGTAGCAGGA CGGGGGGCGG CGTGGAAA

Annotation

Region in nt	Explanation
14–33	T7 promotor
37–107	Omega enhancer sequence
110–114	Kozak
115–480	VHH
199–246	CDR1
286–333	CDR2
457–504	CDR3
535–519	His_6 tag + GS linker
574–606	Flag tag + GGS
607–628	Linker hybridization region

Composition of X, Y, and Z, and their complements, $\dot{X}$, $\dot{Y}$, and $\dot{Z}$

Codon	A	T	G	C
X ($\dot{X}$)	0.35 (0.1)	0.1 (0.35)	0.35 (0.2)	0.2 (0.35)
Y ($\dot{Y}$)	0.4 (0.25)	0.25 (0.4)	0.1 (0.25)	0.25 (0.1)
Z ($\dot{Z}$)	0 (0.4)	0.4 (0)	0.2 (0.4)	0.4 (0.2)

Figure S2. DNA sequences of the produced VHH library. XYZ corresponds to the mixed bases shown in the table above. Bolds indicate CDR regions. Underlined sequences correspond to the junction sites. Yellow highlighted bases are the positions of I in the corresponding primers (Table S1, and compare with Figure S1).

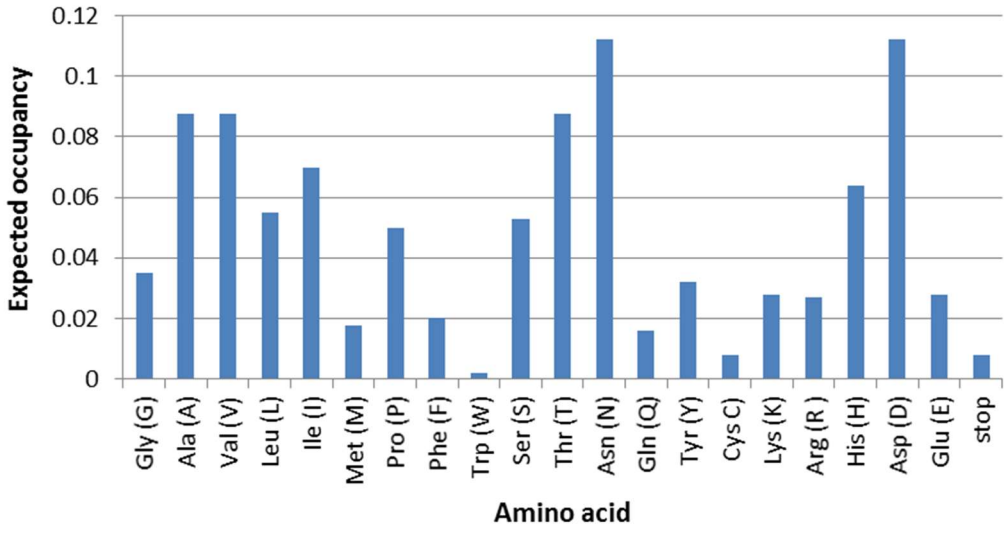

Figure S3. Expected amino acid appearance in the CDR regions coded by the XYZ codon (Figure S2).

Clone 1

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAACTCGCACCAATAAGAATACTAGCCTCACTGCCACTTACGA
GGACAAGTGGTTCCGCCAGGGTCCTGGAAAGGAGCGCGAGGGAGTGACCATCGGTAAT
GGTGGCAAGAATAATCACGATAATGACCCTACTAATACCTACTACGCCGACAGCGTGAAG
GGACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAG
CCTGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTGATAACGACAAGGACACGG
TCGCGGGCCTCAATGACGACGCTATTCGCTACTGGGGCCAGGGGACCCAGGTGACCGTG
GGAGGAGGCAGCCATCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGA
CAAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 2

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATCACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAAACAACAAGAACGCTTATAAGGAGACTCACCATTTCGATC
CTGCTACTTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGCTGTTTACTAGT
ACTGTGATTATTAACATCAACGACCCGAATCACCAGACCTACTACGCCGACAGCGTGAAG
GGACGCTTCACCATCAGCCAGGACAACACCAAGAACACCGTGTACCTGCAGATGAACAG
CCTGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTAAGAACACTCTCACCAACC
ATTATCCTAAGGGGCCGATCGCTGCCCAGTACTGGGGCCAGGGGACCCAGGTGACCGTG
GGAGGAGGCAGCCATCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGA
CAAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 3

−ATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAGCGCATTAGAGTGTGGATAGCTTGCTCGCCCACCATGTCAC
TGATTCGTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGAGCCATGCTGACA
ATAATGACCGTAATTTCAAGGACCTGAAGGCGATCACCTACTACGCCGACAGCGTGAAG
GGACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAG
CCTGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTCCCCACCGGGCGACCACGG
CCGCTTCTGTCATCATTGCCATCGACAAGTACTGGGGCCAGGGGACCCAGGTGACCGTG

GGAGGAGGCAGCCATCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGA
CAAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 4

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGACCCAGCGATCTCCTTACTTATTTCGTTAAGTTGATCGGCGAT
GATCATTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGCACCCGAATATTATC
ACTCATAGCCGTAACTACCATAAGGACGAGAACACCTACTACGCCGACAGCGTGAAGGG
ACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAGCC
TGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTTCCAATGTTGACATTACCGCCG
CTAATGGCAACGATATCTCTGCCGTTTACTGGGGCCAGGGGACCCAGGTGACCGTGGGA
GGAGGCAGCCATCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGACAA
GGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 5

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAAATAACCTTAATCACATTCATAGCGTTCCGGGTAATCTTAAC
AATGTTTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGAACCTCCATACCAA
CATTCGCGCTGATCCTATTGACCGCCATCCCTATACCTACTACGCCGACAGCGTGAAGGGA
CGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAGCCT
GAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTCATAAGACCAACCCCGTTCTCAG
TCCGGATGTCTTCGACATTGCGGTGTACTGGGGCCAGGGGACCCAGGTGACCGTGGGAG
GAGGCAGCCATCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGACAAG
GGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 6

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGATCCGTTGATTAGTGGTTCATTCTCCTTACTATTTGCAACCGT
AGTCTGTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGCTTACTAACGATCC
CAACAACGCGGACAATATGTATAATCACACCAACACCTACTACGCCGACAGCGTGAAGG

GACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAGC
CTGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTGATAATGAGATTAAGAATGAT
GACGACGAGAATGCGGTTACCGCCGCTTACTGGGGCCAGGGGACCCAGGTGACCGTGG
GAGGAGGCAGCCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGAC
AAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 7

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAGCGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAGACATTCATAACGATCATCCCAATATTCAGATGCTCAATAAT
AACAGGTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGATTATTGCTCCCCA
GACCAATTACGTGACTAAGAATCCCCCTCATATGACCTACTACGCCGACAGCGTGAAGGG
ACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAGCC
TGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTGCCGTTCATAATGAGGACCCCG
TCTCAGTCCGAGAAGAGCAATGACCGTTACTGGGGCCAGGGGACCCAGGTGACCGTGG
GAGGAGGCAGCCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGAC
AAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Clone 8

GATCCCGCGAAATTAATACGACTCACTATAGGGGAAGTATTTTTACAACAATTACCAACAA
CAACGACAAACAACAACAACATTACATTTTACATTCTACAACTACAAGCCACCATGGGCG
AGGTGCAGCTGGTGGAGAACGGAGGAGGATCCGTGCAGGCTGGAGGAAGCCTGCGCCT
GAGCTGCGCTGCTAGCGGAAACGCGGACATTGTCCCTCACAACACGAATATGTACCACC
CTTCCATGTGGTTCCGCCAGGCTCCTGGAAAGGAGCGCGAGGGAGTGGACGAGCTCACC
ATTCCGAACGATCACGTCACCGACGCCGATAATGATACCTACTACGCCGACAGCGTGAAG
GGACGCTTCACCATCAGCCAGGACAACGCCAAGAACACCGTGTACCTGCAGATGAACAG
CCTGAAGCCAGAGGACACCGCTATCTACTACTGCGCTGCTCGCATGCATCTCAATCTCAG
GACCGTTACCAACGACCATGTGAGTCACTACTGGGGCCAGGGGACCCAGGTGACCGTGG
GAGGAGGCAGCCATCATCATCATCACGGCGGAAGCGACTACAAGGATGACGATGAC
AAGGGTGGTAGCAGGACGGGGGGCGGCGTGGAAA

Amino acid sequences of CDRs

Clone	CDR1	CDR2	CDR3
1	TRTNKNTSLTATYEDK	TIGNGGKNNHDNDPTN	DNDKDTVAGLNDDAIR
2	NNKNAYKETHHFDPAT	LFTSTVIININDPNHQ	KNTLTNHYPKGPIAAQ
3	AH*SVDSLLAHHVTDS	SHADNNDRNFKDLKAI	PHRATTAASVIIAIDK
4	PSDLLTYFVKLIGDDH	HPNIITHSRNYHKDEN	SNVDITAANGNDISAV
5	NNLNHIHSVPGNLNNV	NLHTNIRADPIDRHPY	HKTNPVLSPDVFDIAV
6	SVD*WFILLTICNRSL	LTNDPNNADNMYNHTN	DNEIKNDDDENAVTAA
7	DIHNDHPNIQMLNNNR	IIAPQTNYVTKNPPHM	AVHNEDPVSVREEQ*P
8	NADIVPHNTNMYHPSM	DELTIPNDHVTDADND	RMHLNLRTVTNDHVSH

Figure S4. Sequences of eight clones of the constructed DNA library. Red indicates unintentional mutation introduced in PCR (for library preparation) and cloning, and underline indicates CDR regions. Yellow highlights indicate the junction sites of I-containing fragments. Amino acid sequences of CDRs are also shown. The stop codon appeared in CDR only at a percentage of 0.8%, which was in good accordance with Figure S3.

Supplementary Methods

Materials and general instruments

General chemicals were of the best grade available supplied by Wako Pure Chemical Industries (Osaka, Japan). Chemicals for molecular biology experiments were obtained from New England Biolabs (NEB, Ipswich, MA, USA), Takara (Tokyo, Japan), and Wako Pure Chemical Industries (Osaka, Japan). UltraPure DNase/RNase-Free Distilled Water (UPDW) was obtained from Thermo Fisher Scientific (Waltham, MA, USA). They were used without further purification. DNA oligos were synthesized by Eurofins Genomics (Luxemburg). PCR were performed with a Biometra TRIO48 thermal cycler. For PCR using deoxyinosine-containing primers, KOD Multi & Epi (Toyobo, Osaka, Japan) was used. For general PCR, PrimeSTAR HS DNA polymerase (Takara) was used for PCR under the conditions recommended by the manufacturer. DNA was purified by FavorPrep PCR Clean-Up Mini Kit (Favorgen Biotech, Taiwan; elution was performed by 40 μL of water), and concentration was measured by Nanodrop 1000 (Thermo Fisher, Waltham, MA, USA). Ethanol precipitation was performed with Quick-Precip Plus Solution (Edge Bio, San Jose, CA, USA). Unless otherwise stated, PAGE analysis of DNA was performed at 60°C using gels containing 8 M urea, with 0.5× TBE as a running buffer (16% acrylamide, 200 V, 150 min). 100 bp DNA ladder (Promega, Madison, WI, USA) was used as a standard. Gel images were taken with a Typhoon FLA9500 imager (GE Healthcare, Chicago, IL, USA) after SYBR Gold staining. DNA sequencing was performed by Eurofins Genomics.

PCR of fragments

The template DNA coding a VHH (Figure S1), which had been prepared in the course of our previous work [S1], was amplified by PCR with primers containing deoxyinosine (see Table S1 for primer sequences). The PCR mixture contained 1× PCR Buffer for KOD Multi & Epi, 400 nM concentration of each primer (see table below; for the sequence, see Table S1), ~0.1 ng template DNA, 0.5 μL enzyme solution, and UPDW in a final volume of 25 μL. The step program for PCR was as follows: 94°C for 2 min, followed by 25 cycles at 98°C for 10 sec, various temperature for 5 sec, and 68°C for 10 sec. The results were analyzed by PAGE. PCR products were then purified by FavorPrep PCR Clean-Up Mini Kit. Fragment B was prepared by two successive PCR.

PCR conditions

Fragment	Forward primer	Reverse primer	Annealing temperature (°C)
A	5′ primer	frag_A_R	None (2-step PCR)
B (1st step)	frag_B_F1	frag_B_R2	62
B (2nd step)	frag_B_F2	frag_B_R1	62

C	frag_C_F	frag_C_R	60
D	frag_D_F	3′ primer	66

Endonuclease V treatment

The PCR-amplified fragments (5 pmol) was diluted in 1× NEBuffer 4 (NEB), and endonuclease V (5 U) was added. Total reaction volume was 25 μL. The mixture was incubated for 1 h at 37°C. Then, 9-fold volume of 8 M urea was added and the mixture was incubated for 30 min at 42°C to strip the cleaved oligo. Note: this step is not necessary if the protruding end is short (<9 nt). Cleaved short fragments containing deoxyinosine were then removed using the FavorPrep PCR Clean-Up Mini Kit. The results were analyzed by PAGE.

Ligation

The endo V-treated fragments (5 pmol each) were mixed in 1X *Taq* DNA ligase reaction buffer (NEB), and *Taq* DNA ligase (40 U) was added. The total reaction volume was 50 μL. The mixture was incubated for 30 min at 45°C. The results were analyzed by PAGE. The products were purified using the FavorPrep PCR Clean-Up Mini Kit.

PCR of full-length library

The ligation mixture of all fragments (1 μL, after clean-up) was amplified by PCR using PrimeSTAR HS DNA polymerase, according to the manufacturer's instruction. 5′ and 3′ primers were used as primers (Table S1). The step program for PCR was as follows: 95°C for 1 min, followed by 8 cycles at 98°C for 10 sec, 55°C for 5 sec, and 72°C for 40 sec. The product was analyzed by PAGE.

Direct sequencing of PCR product

The sequence of the full-length PCR product library was analyzed by Sanger sequencing. 5′ primer (Table S1) was used as a sequence primer. Raw sequence data of CDR regions and junction sites were extracted and are shown in Figure 4 in the main text.

Cloning and sequence analysis of the library

The full-length library was amplified by *Ex Taq* DNA polymerase (Takara) to add A at the 3′ terminal, and the DNA was ligated into pGEM-T Easy Vector (Promega, Madison, WI, USA). The vector was then transformed to DH5α *E. coli* competent cells (Takara), and the cells were spread to LB plates containing ampicillin. All the protocols were according to the manufacturers' instruction. Several colonies were picked up, cultured in liquid LB medium, and the plasmids were purified using a

Plasmid DNA Extraction Mini Kit (Favorgen Biotech). Sequences of eight inserted DNA were analyzed by Sanger sequencing using M13_F and M13_R (Table S1) as sequence primers.

Supporting Reference

S1. Suzuki, T.; Mochizuki, Y.; Kimura, S.; Akazawa-Ogawa, Y.; Hagihara, Y.; Nemoto, N. Anti-survivin single-domain antibodies derived from an artificial library including three synthetic random regions by in vitro selection using cDNA display. *Biophys. Biochem. Res. Commun.* **2018,** 503, 2054–2060.