J. Biochem. **144,** 591–598 (2008)
doi:10.1093/jb/mvn103

Construction of a Fully Synthetic Human scFv Antibody Library with CDR3 Regions Randomized by a Split–Mix–Split Method and Its Application

Chang-Cheng Yin[1,2], Li-Li Ren[2], Lin-Lin Zhu[2], Xiang-Bin Wang[2], Zhong Zhang[2], Hua-Liang Huang[2,*] and Xi-Yun Yan[1,†]

[1]National Laboratory of Biomacromolecules, and CAS-UT joint laboratory of Structural Virology and Immunology, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101; and [2]Beijing ABT Genetic Engineering Technology Co., Ltd, Beijing, 102206, China

Received May 1, 2008; accepted August 13, 2008; published online August 19, 2008

The randomization scheme of hypervariable region takes crucial role in construction of a synthetic antibody library. The codon bias and inevitable 'stop' codon of conventional 'NNK' and 'NNS' codons limit their applications. Here we report a split–mix–split DNA synthesis method that can control over the amino acid composition and distribution of randomized sequences effectually. A fully synthetic human antibody library with a diversity of 1.56×10^9 was successfully generated with complementarity determining region 3 (CDR3) randomized by this strategy. Sequencing analysis indicated that >60% of colonies had completely correct scFv genes and the amino acid composition and distribution were designed well in accordance. The utility was demonstrated by screening of scFv clones against BHL (anti-CD3 × anti-ovarian carcinoma bispecific antibody). These results proved the feasibility of the split–mix–split DNA randomization strategy in library construction and site-directed mutagenesis.

Key words: antibody library, biopanning, phage display, randomization, single-chain antibody.

Abbreviations: BHL, anti-CD3 × anti-ovarian carcinoma bispecific single-chain antibody; CDR, complementarity determining region; HuCAL, fully gene-synthetic human combinatoryial antibody library; scFv, single-chain variable fragment; SMS, split–mix–split DNA synthesis; SOE PCR, splice overlap extension PCR; TRIM, trinucleotide-mutagenesis; V_H, variable region of heavy chain; V_L, variable region of light chain; Vκ, variable regionof κ light chain; V_λ, variable region of λ light chain.

An antibody library is a repertoire, from which various antibodies can be obtained for antibody-based biopharmaceuticals, diagnostics and research reagents. Many innovative display methods, such as ribosome display (*1*, *2*), bacterial display (*3*), yeast and mammalian cell display have been developed, which expands the methodology of display and screening (*4*, *5*). Different strategy may be adopted to construct libraries with specific properties. Among those techniques, phage display is the most popular method for construction and screening of antibody library.

The original intention of generating an antibody library was to cover high diversity with limited number of members. Typically, while a library size of $10^{6–8}$ members produced antibody affinities of 10–100 nM, library of $10^{9–10}$ members led to the isolation of antibodies with dissociation constants (K_d) in the subnanomolar range (*6*, *7*). In practice, it is difficult to construct a phage display library with diversity exceeding for 10^{11} due to physical limits, it will face difficulties in genetic manipulation (*8*). Usually, the framework regions of a synthetic antibody library were designed according to the consensus sequences of human antibody germline families. Representative antibody frameworks can cover most genetic information and have been used to improve the efficiency of library construction. By analysing the human antibody repertoire in terms of structure, amino acid sequence diversity and germline usage, Knappik *et al*. (*9*) found that seven V_H and seven V_L modular consensus frameworks with CDR3 region randomized with trinucleotides could cover > 95% of human antibody diversity used.

For the importance of CDR3 regions in antibody–antigen interaction, the design scheme and randomization strategy of CDR3 play crucial role in synthetic library construction. A 'NNK' or 'NNS' scheme (where, N presents A, G, C or T, K presents T or G and S presents G or C) is an extensively used conventional method in library construction (*10*), both the 'NNK' and 'NNS' triplets can produce 32 codons to present all 20 amino acids and an unexpected 'stop' codon, but they can not actualize the elaborate design of random sequence, for example, introducing amino acids of typical subsets at given frequencies. With the increase of the length of random sequence, the limitations of inevitable codon bias and 'stop' codon will greatly influence the efficiency of randomization. The 'TRIM' (trinucleotide mixtures)

*To whom correspondence should be addressed. Tel/Fax: +86-10-80726906, E-mail: hlhuang@genetics.ac.cn
†Correspondence may also be addressed: Tel: +86-10-64888583, Fax: +86-10-64888584, E-mail: yanxy@sun5.ibp.ac.cn

oligonucleotides synthetic method allows control over the amino acid distributions and compositions in CDRs to mimic the natural pattern of diversity in human immune system (*9*), while this method has not been used widely for technical and material limitations. The split–mix–split method represents another effective and reliable alternative to introduce a complete or partially random sequence to CDR regions (*11*, *12*). For a synthetic antibody library, its functional diversity may be different from the designed size due to sequence redundancy and mutations caused by PCR and gene replication. We have therefore aimed at producing a large functional synthetic antibody library with representative frameworks and CDR3 regions randomized by split–mix–split synthesis strategy.

MATERIALS AND METHODS

Bacterial Strains and Plasmids—Escherichia coli—TG1 and HB2151 (Amersham Pharmacia Biotech, USA) were used for bio-panning and periplasmic expression of scFv. BL21 (Novagen, USA) was used for expression of scFv in pTCM2, which was a T7 promoter-based expression vector constructed in our laboratory. The phagemid pCAK2 was derived from pCANTAB5E (Amersham Pharmacia Biotech, USA) and used for library construction.

DNA Synthesis and Assembly of Framework Region Sequences—Totally seven heavy-chain (V_H1A, V_H1B, V_H2, V_H3, V_H4, V_H5 and V_H6) and seven light-chain (V_κ1, V_κ2, V_κ3, V_κ4, V_λ1, V_λ2 and V_λ3) domains were chosen as master frameworks (*9*). They were back-translated to DNA sequences with *E. coli* preferential codons and then split into 8–12 overlapping fragments of 50–70 bases synthesized by Shanghai Bioasia biotechnology Co. Ltd, China. Splice overlap extension PCR (SOE PCR) was performed on the iQ5 thermal cycler (Bio-Rad, USA) in a 50 μl reaction containing each 5 μl of forward and reverse primers at the concentration of 0.5 μM, 1 U of *Pfu* DNA polymerase (Shanghai Bioasia biotechnology Co. Ltd, China) (pre-denature at 94°C for 3 min, and then 94°C for 1 min, 55°C for 40 s, 72°C for 1 min, 15 cycles). The resulting products were separated on 2% agarose and recovered to be used in next round of assemble until full-length V_H and V_L framework regions have been achieved. Seven V_H frameworks covered from FR1 to FR3 were amplified with forward primers, which will introduce the first *Sfi*I site at the 5′-end and reverse primers. Seven V_L frameworks covered from FR1 to FR3 were amplified with forward primers, which will introduce a $(G_4S)_2$ linker at the 5′-end and reverse primers. The V_H and V_L frameworks were cycled 25 times (94°C for 1 min, 55°C for 1 min and 72°C for 1 min) using primers and templates listed in Table 1, each 50 μl reaction contained 0.5 μM primers, 1 U *Pfu* DNA polymerase and 200 μmol dNTPs. PCR products were separated on 1% agarose and gel purified with spin-column (Beijing Tiangen Biotech, China).

Random CDR3 oligonucleotides were synthesized by Shanghai Bioasia Biotechnology Co. Ltd using split–mix–split method as shown in Fig. 1. To obtain sufficient product and diversity, multiple synthesis reactions should be performed parallelly. The double strands CDR3 segments were produced by PCR with primers and templates listed in Table 1. The CDR3-H1346, CDR3-H2 and CDR3-H5 segments were cycled 20 times (94°C for 40 s, 50°C for 40 s and 72°C for 40 s). Each 50 μl reaction contained 0.5 μM forward primer that will introduce a unique sequence overlapping with FR3 at the 5′-end, 0.5 μM reverse primer that will introduce a $(G_4S)_2$ linker at the 3′-end, 5 μl 10 × PCR buffer, 1 U *Pfu* DNA polymerase and each 200 μmol dNTPs. Pooled (equimolar) CDR3-H-A and CDR3-H-B were used as temples to incorporate different length of random CDR3s together. Four CDR3-K segments and three CDR3-L segments were cycled 20 times (94°C for 40 s, 54°C for 40 s and 72°C for 40 s). Each 50 μl reaction contained 0.5 μM forward primer, which will introduce a sequence overlapping with FR3 at the 5′-end, 0.5 μM reverse primer which will introduce the second *Sfi*I site at the 3′-end, 5 μl 10 × PCR buffer, 1 U *Pfu* DNA polymerase and 200 μmol dNTPs. Pooled (equimolar) CDR3-L-A, CDR3-L-B and CDR3-L-C were used as templates to incorporate different lengths of random CDR3s together. All CDR3 segments were separated on 2% agarose and gel-purified.

Table 1. **Primers and templates used to amplify 14 frameworks and seven CDR3 segments.**

PCR product	Forward primer	Reverse primer	Template
V_H1AFR	VH1F	V_H1346R	Plasmid DNA
V_H1BFR	VH1F	VH1346R	Plasmid DNA
V_H2FR	VH2F	VH2R	Plasmid DNA
V_H3FR	VH3F	VH1346	Plasmid DNA
V_H4FR	VH4F	VH1346	Plasmid DNA
V_H5FR	VH5F	VH5R	Plasmid DNA
V_H6FR	VH6F	VH1346R	Plasmid DNA
V_κ1FR	VK1F	VK1234R	Plasmid DNA
V_κ2FR	VK2F	VK1234R	Plasmid DNA
V_κ3FR	VK3F	VK1234R	Plasmid DNA
V_κ4FR	VK4F	VK1234R	Plasmid DNA
V_λ1FR	VL1F	VL123R	Plasmid DNA
V_λ2FR	VL2F	VL123R	Plasmid DNA
V_λ3FR	VL3F	VL123R	Plasmid DNA
CDR3-H1346	CDR3F	VHR	CDR3-H-A, CDR3-H-B
CDR3-H2	CDR3H2F	VHR	CDR3-H-A, CDR3-H-B
CDR3-H5	CDR3H5F	VHR	CDR3-H-A, CDR3-H-B
CDR3-K1	CDR3-K1F	VKR	CDR3-K-A
CDR3-K24	CDR3-K24F	VKR	CDR3-K-B
CDR3-K3	CDR3 –K3F	VKR	CDR3-K-A
CDR3-L123	CDR3-LF	VLR	CDR3-L-A, CDR3-L-B, CDR3-L-C

Assembly of V_H, V_L and scFv Gene Repertoires—In parallel, seven synthetic heavy chain gene repertoires were individually amplified with forward primer VHXF (where, X is 1A, 1B, 2, 3, 4, 5 or 6) and reverse primer VHR using recovered V_H frameworks and equimolar corresponding CDR3 segments as templates (Table 2). Similarly, four V_κ gene repertoires and three V_λ gene repertoires were amplified with forward primers VKXF (where, X is 1, 2, 3 or 4) or VLXF (where, X is 1, 2 or 3) and reverse primers VKR or VLR, recovered V_κ or V_λ frameworks and corresponding CDR3 segments were

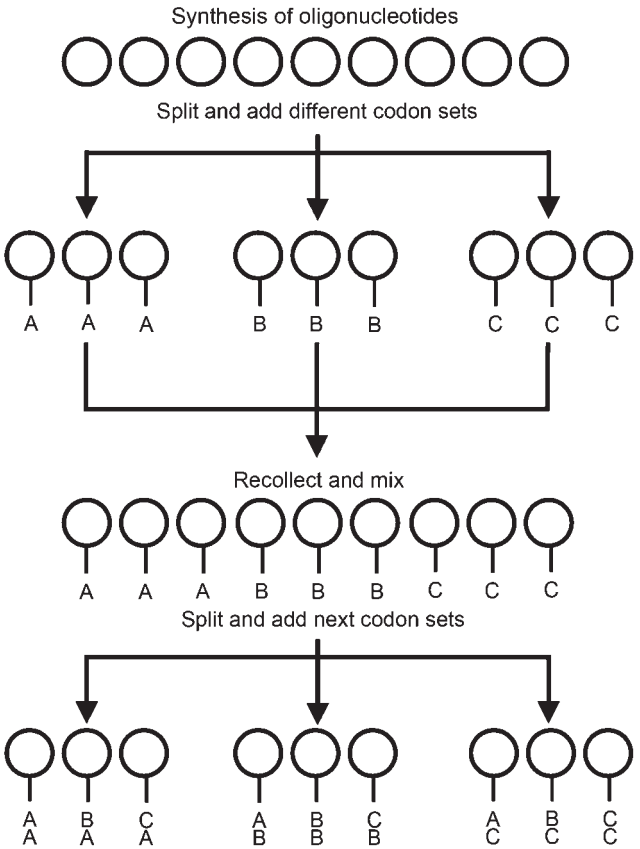

Fig. 1. **Scheme of the split–mix–split DNA synthesis method.** All 19 amino acid (written in one-letter symbol) except for Cysteine are classified into nine degenerate sets, i.e. GA(T,G) = D or E, AA(T,G) = N or K, CA(T,G) = H or Q, A(G,C)C = S or T, (G,C)CG = A or P, (A,C)TG = M or L, (C,G)GT = R or G, (A,G)TT = I or V, T(A,T)C = Y or F and TGG = W. In the scheme, A, B or C denotes one of the codon sets. When a Cys-free random position needs to be introduced, all resins with oligomers are colleted and split into several portions. For each portion, three nucleotides of a codon set are added in turn, and then they are recollected, mixed sufficiently and split to add next triplet codon. This mix-split procedure will be repeated until all random positions have been introduced.

used as templates (Table 2). Each 50 μl assembly reaction contained 0.1 μM forward and reverse primers, 5 μl 10 × PCR buffer, 1 U *Pfu* DNA polymerase and 200 μmol dNTPs. Reactions were cycled 20 times (94°C for 1 min, 55°C for 40 s and 72°C for 1min). V_H-V_κ and V_H-V_λ scFv libraries were created by combining recovered V_H repertoire with pooled V_κ genes and V_λ genes, respectively. Assembly PCR were carried out with forward primer VHXF (where, X is 1A, 1B, 2, 3, 4, 5 or 6) and reverse primer VKR or VLR, respectively. Each 100 μl assembly reaction contained about each 40 ng V_H and V_L template DNA, 0.1 μM forward and reverse primers, 10 μl 10 × PCR buffer, 1 U *Pfu* DNA polymerase and 200 μmol dNTPs. Reactions were cycled 20 times (94°C for 1 min, 53°C for 40 s and 72°C for 2 min).

Cloning of scFv PCR Products into Phagemid Vector—Each 5 μg of pCAK2 DNA or purified scFv fragments were digested with 20 U *Sfi*I (TaKaRa, Japan) for 4 h at 50°C. About 30 μl of mineral oil was added to prevent evaporation. After enzymatic restriction, the linearized pCAK2 vector was separated on 0.8% agarose gel. The scFv samples were directly purified from reaction solution with spin-column. Each ligation was performed in a 100 μl reaction containing 15 U T_4 DNA ligase (Promega, USA), 50 ng linearized vector and 20 ng scFv (molar ration ≈1:3) at 16°C for 20 h. The DNA were precipitated with ethanol and dissolved in 20 μl deionized H_2O.

Table 2. **Primers and DNA templates used for PCR amplification of V_H and V_L repertoires.**

Fv repertoire	Forward primer	Reverse primer	DNA templates	
			Framework fragment	CDR3 segment
V_H1a	VH1F	VHR	V_H1AFR	CDR3-H1346
V_H1b	VH1F	VHR	V_H1BFR	CDR3-H1346
V_H2	VH2F	VHR	V_H2FR	CDR3-H2
V_H3	VH3F	VHR	V_H3FR	CDR3-H1346
V_H4	VH4F	VHR	V_H4FR	CDR3-H1346
V_H5	VH5F	VHR	V_H5FR	CDR3-H5
V_H6	VH6F	VHR	V_H6FR	CDR3-H1346
V_κ1	VK1F	VKR	V_H1FR	CDR3-K1
V_κ2	VK2F	VKR	V_H2FR	CDR3-K24
V_κ3	VK3F	VKR	V_H3FR	CDR3-K3
V_κ4	VK4F	VKR	V_H4FR	CDR3-K24
V_λ1	VL1F	VLR	V_H1FR	CDR3-L123
V_λ2	VL2F	VLR	V_H2FR	CDR3-L123
V_λ3	VL3F	VLR	V_H3FR	CDR3-L123

Transformation and Sequence Analysis of Phage Display Libraries—Approximately 50 ng DNA was transformed into 200 μl electrocompetent TG1 cells by setting the electroporator ECM630 (BTX, USA) to 2.5 KV, 200 Ω and 50 μF. The electroporation sample was then added with 1 ml pre-warmed SOC medium and incubated for 45 min at 37°C. To calculate the diversity of the library, serial dilutions were taken out on 2YT-AG [2YT medium with 100 μg/ml ampicillin (Sigma, USA), 2% w/v glucose and 1.5% agar] plates. For each batch, clones picked up randomly were sequenced. Other culture was added with ampicillin to a final concentration of 100 μg/ml and incubated for about 3 h before harvest. The cell pellets were resuspended in SOC medium (containing 100 μg/ml ampicillin, 2% w/v glucose and 20% v/v glycerol) and stored at −70°C.

Amplification and Rescue of scFv-displaying Phage—Approximately 3×10^8 cells from the −70°C frozen library glycerol stock were inoculated to 100 ml 2YT-AG. The culture was then shaken at 37°C until the $OD_{600} \sim 0.7$, then, 1×10^{12} plaque forming unit (PFU) M13KO7 helper phage prepared according to standard protocols were added. After incubation at 37°C for 45 min, cells were centrifuged and resuspended in 250 ml 2YT containing 100 μg/ml ampicillin and grown for 2 h at 37°C. Kanamycin (Sigma, USA) was then added to a final concentration of 50 μg/ml, and the culture was grown overnight at 37°C followed with spun at 4000*g* for 15 min. Phage were harvested by precipitation with one-fourth of (v/v) 20% polyethylene glycol (PEG) 6000/NaCl

(2.5 M). The titre in number of CFUs was determined by spreading serial dilutions of the freshly infected TG1 onto 2YT-AK agar plates.

Biopanning of Phage Antibody Clone Binding to BHL—A 96-well ELISA plate (Nunc Maxisorp 442404, USA) was coated with 100 µl BHL (~2.5 µg) for 4 h. The wells were washed three times with PBS and then blocked with 200 µl PBSM (PBS containing 2% skim milk) at 37°C for 2 h, 150 µl freshly prepared phage (~10^{12} CFU) in PBS containing 2% BSA were added and incubated for 2 h at 37°C. The plate was washed with 200 µl PBST (PBS containing 0.1% Tween 20) and followed by a wash with PBS. Bound phages were eluted with 50 µl elution buffer (0.1 M HCl, 0.1% BSA, adjust to pH2.2 with glycin). The eluted phages were neutralized with 2 M Tris–HCl (pH 8.0) and used for next round of panning. The titre of the eluted phages was also determined. After the final round of panning, the scFv-phage was used to infect fresh prepared TG1 cells. The PCR-positive clones were sequenced and transformed into HB2151 cells to express soluble scFv fragment.

Expression, Purification and ELISA Assay of scFvs—The selected scFv genes were reamplified with primers containing *Xho*I and *Eco*RI (New England Biolabs, USA) sites, and subcloned into pTCM2. BL21 cells habouring pTCM2-scFv were grown in 50 ml LB broth with 50 µg/ml kanamycin at 37°C. The expression of scFv was induced by addition of IPTG to 0.4 mM final concentration at 30°C, cells were harvested 6 h after IPTG addition by centrifugation and resuspended in PBS. After sonication, the whole cell lysate was spun at 12,000*g* for 40 min. Affinity chromatography on Ni-NTA resin (Qiagen, USA) was performed to purify soluble scFv fragments from supernatant. The pellets containing inclusion bodies were denatured with 8 M urea and renatured by dilution refolding, step-wise dialysis and purified with Ni-NTA resin according to protocols provided by the manufacturer. ELISA of positive clones was performed routinely to determine the binding activities of the selected clones. Briefly, ELISA plate (Nunc Maxisorp 442404, USA) was coated with 100 µl BHL antigen at 25 µg/ml in PBS, blocked with PBSM (PBS containing 2% skim milk), and then washed three times with PBST. Purified scFvs were added in triplicate with serial dilutions. BSA was also added as negative control. Captured scFv was revealed by mouse anti *C-myc* mAb 9E10 (Santa Cruz Biotechnology, USA) as the primary antibody followed by goat anti-mouse IgG-HRP conjugate (R&D Systems, USA) as the secondary antibody. After addition of the chromogenic substrate *o*-phenylenediamine (OPD) (Sigma, USA) solution, the reaction was stopped with 2 M H_2SO_4 and the absorbance was read at 492 nm with a ELx-800 microplate reader (Bio-Tek, USA).

RESULTS

Assembly of V_H and V_L Master Genes and Amplification of Random CDR3 Segments—Seven V_H master genes and seven V_L master genes representing subfamilies frequently used in human antibody repertoire (*9*), were assembled by multi-step SOE-PCR. The PCR products recovered from agarose were used as template in next SOE-PCR until the full-length framework was obtained. Sequence validated clones were used in assembly of V_H and V_L repertoires.

Scheme of the split–mix–split method is shown in Fig. 1 and altogether seven randomized CDR3 fragments were synthesized (Table 3). The random sequences were mainly designed with reference to Knappik's report (*9*). The amplification results of heavy-chain CDR3 (V_H-CDR3), κ light-chain CDR3 ($V_κ$-CDR3) and λ light-chain CDR3 ($V_λ$-CDR3) segments were showed in Fig. 2A.

Sequencing analysis indicated that the clones with randomized V_H-CDR3 of length 6 and 10 amino acids were 37.5% (18/48) and 62.5% (30/48), respectively. For $V_λ$-CDR3, the clones with 3, 4 or 5 randomized residues were 41, 41 and 18%, respectively.

Assembly of scFv Fragments using SOE-PCR—To amplify V_H and V_L genes, assembly PCR was performed with primers and templates listed in Table 2. PCR products were checked by electrophoresis as shown in Fig. 2B. Each V_H gene (~400 bp) covered the sequences from the 5′-*Sfi*I site to an appended 15 bp sequence coding for $(Gly_4Ser)_2$ at the 3′-end. Each $V_κ$ or $V_λ$ gene (~370 bp) covered the sequences from the 5′-$(Gly_4Ser)_2$ linker coding sequence to the 3′-*Sfi*I site. The last 15 bp of V_H genes are identical to the first 15 bp of V_L genes, thus the full-length scFv gene could be assembled by SOE-PCR.

Recovered four $V_κ$ repertoires and three $V_λ$ repertoires were quantitated and pooled at the molar ratio of 1:1:3:1 and 1:1:1 to produce $V_κ$-pool and $V_λ$-pool, respectively. Seven V_H-$V_κ$ and seven V_H-$V_λ$ scFv libraries were created by combining each of seven V_H repertoires with $V_κ$-pool or $V_λ$-pool at molar ration of 1:1. For each combination, more than 20 PCR reactions were performed independently to obtain sufficient products without obvious sequence bias. The PCR products of 14 scFv gene libraries were separated on 1% agarose as shown in Fig. 2C.

Table 3. **Random CDR3 segments used to generate diversity in V_H and V_L domains.**

Random Oligoncleotide	Amino acid sequence deduced from oligonulceotides (standard PROSITE format)	Used in master framework	Planned diversity
CDR3-K-A	D-F-A-[TV]-Y-Y-C-[FHLMQ]-Q-{CP}-[DGNSY]-[DGNS]-{C}-[PS]-{C}-T-F-G-H-S	Vκ1,Vκ3	1.8×10^6
CDR3-K-B	D-V-[AG]-V-Y-Y-C-[FHLMQ]-Q-{CP}-[DGNSY]-[DGNS]-{C}-[PS]-{C}-T-F-G-H-S	Vκ2,Vκ4	4.9×10^5
CDR3-L-A	P-E-[RWY]-D-{CW}(5)-{C}-G-G-G	$V_λ1 \sim V_λ3$	3.3×10^5
CDR3-L-B	P-E-[RWY]-D-{CW}(6)-{C}-G-G-G	$V_λ1 \sim V_λ3$	5.9×10^6
CDR3-L-C	P-E-[RWY]-D-{CW}(7)-{C}-G-G-G	$V_λ1 \sim V_λ3$	4.7×10^7
CDR3-H-A	Y-Y-C-A-R-{C}(6)-D-Y-W-G-Q	$V_H1A \sim V_H6$	4.7×10^7
CDR3-H-B	Y-Y-C-A-R-{C}(10)-[FM]-D-[VY]-W-G-Q	$V_H1A \sim V_H6$	2.5×10^{13}

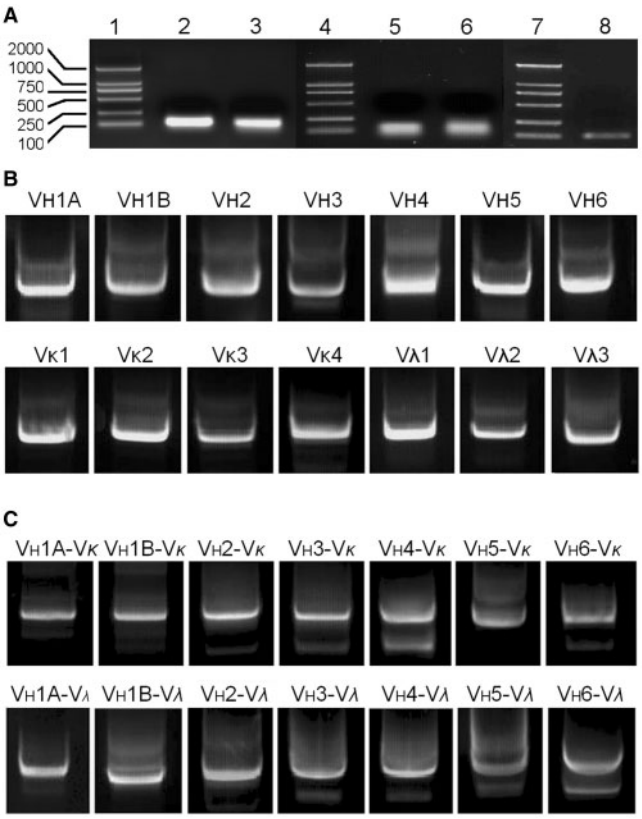

Fig. 2. **Amplification of randomized CDR3s and scFv fragments.** (A) Amplification of randomized CDR3s, 1, 4, 7: DNA molecular mark (TaKaRa); 2: CDR3-H-B, 3: CDR3-H-A, 5: CDR3-K-A, 6: CDR3-K-B, 8: pooled CDR3-L. (B) Assembly of PCR products of seven V_H and seven V_L gene repertoires. (C) Assembly PCR products of 14 V_H–V_L combinations.

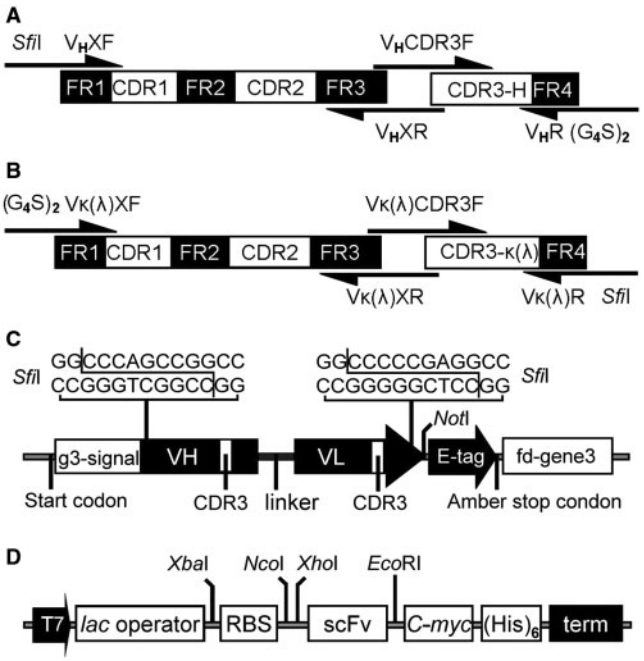

Fig. 3. **Scheme of amplification and cloning procedure.** Both V_H and V_L domains were comprised with two parts: a large fragment covering from FR1 to FR3, and a random segment covering from randomized CDR3 to FR4. (A) PCR amplification of V_H domains. Seven large fragments of V_H domain were amplified with primer pair VHXF and VHXR, which will introduce the first *Sfi*I site at the 5′-end. The random segments of V_H domain were amplified with primer pair VHCDR3F and VHR. (B) PCR amplification of V_L domains. Seven large fragments of V_L domain were amplified with primer pairs VKXF/VKXR and VLXF/VLXR, respectively, the random segment of V_L domain were amplified with primer pair VKCDR3F(VLCDR3F) and VLR, which will incorporate the second *Sfi*I site at the 3′-end. (C) *Sfi*I digestion of amplified scFv fragments and pCAK2 vector. The rare cutting endonuclease *Sfi*I is the only enzyme used for directional cloning. A linearized pCAK2 vector has two different sticky ends that are compatible with *Sfi*I digested scFv fragments, which enables the directional cloning of scFv fragments. (D) Schematic diagram of the expression vector pTCM2. The BHL-scFv product has both (His)6 and a *C-myc* tag at the C-terminus to be used in purification and identification.

As a type II restriction endonuclease, *Sfi*I cleaves the following sequence: GGCCNNNN▼NGGCC (where ▼ is the point of cleavage, N is any base). Two *Sfi*I sites 'GGCCCCAG▼CGGCC' and 'GGCCCCC▼GGGCC' were introduced to the 5′- and 3′-ends of scFv gene, respectively. As shown in Fig. 3, the scFv fragment digested by *Sfi*I would produce two different sticky ends that were compatible with the sticky ends of linearized pCAK2 vector. This design can simplify the genetic manipulation of directional cloning. Totally, about 100 μg of digested scFv DNA fragments and 400 μg of linearized vector were used in ligation, and a total of 657 electroporation transformations were performed to obtain sufficient diversity. The primary library consisting of 14 sublibraries was calculated to contain 1.56×10^9 independent clones (Table 4). The results of PCR revealed that a fraction of ~80% of randomly picked colonies had a ~750 bp insert.

Sequence and Amino Acid Composition Analysis—To check the randomization quality of the V_H-CDR3 and V_L-CDR3, 80 randomly picked colonies were sequenced. The statistical results of amino acid composition and distribution of V_H (48 colonies), V_λ (24 colonies) and V_κ (24 colonies) are given in Tables 5–7, respectively. A total of 48 colonies (60%) contained full-length scFv genes.

Table 4. **The capacity of 14 primary sublibraries.**

Sublibrary	Colonies	Sublibrary	Colonies
H_1A-Vκ	2.06×10^8	H_1A-V_λ	1.34×10^7
H_1B-Vκ	1.93×10^8	H_1B-V_λ	5.80×10^7
H_2-Vκ	1.58×10^8	H_2-V_λ	6.91×10^6
H_3-Vκ	2.55×10^8	H_3-V_λ	1.49×10^8
H_4-Vκ	1.19×10^8	H_4-V_λ	5.28×10^6
H_5-Vκ	1.66×10^8	H_5-Vλ	3.00×10^7
H_6-Vκ	1.90×10^8	H_6-Vλ	8.55×10^6

The amino acid distribution and composition in random regions were in good agreement with designed pattern. Neither 'stop' codon nor Cys codon was observed in all calculated colonies. For V_H-CDR3, all 19 amino acid types except for Cys appeared in the positions 95–100D of V_H-CDR3. The occurrence ratio of each amino acid ranged from 2.9% (Lys) to 7.1% (Gln),

Table 5. **Comparison between designed and experimental amino acid composition in CDR3 regions—V_H.**

Position	93		94		95-100D		100E		101		102		103	
Amino acid	P	F	P	F	P	F	P	F	P	F	P	F	P	F
A	100	96.9			5.3	5.2								
C					0	0								
D					5.3	5.4			100	100				
E					5.3	6.9								
F					5.3	6.6	50	40						
G					5.3	5.4								
H					5.3	4.4								
I					5.3	6.1								
K					5.3	2.9								
L					5.3	3.2								
M					5.3	5.2	50	60						
N					5.3	4.4								
P					5.3	4.2								
Q					5.3	7.1								
R			100	100	5.3	5.2								
S	0	3.1			5.3	6.9								
T					5.3	5.6								
V					5.3	5.4					50	65.6		
W					5.3	5.9							100	100
Y					5.3	4.2					50	34.4		

Table 6. **Comparison between designed and experimental amino acid composition in CDR3 regions—V_λ.**

Position	89		90		91		92		93-95b		96	
Amino acid	P	F	P	F	P	F	P	F	P	F	P	F
A									5.6	11.1	5.3	8.3
C									0	0	0	0
D							100	100	5.6	6.7	5.3	8.3
E									5.6	7.8	5.3	4.2
F									5.6	5.6	5.3	0
G									5.6	3.3	5.3	4.2
H									5.6	5.6	5.3	4.2
I									5.6	5.6	5.3	8.3
K									5.6	3.3	5.3	12.5
L									5.6	1.1	5.3	4.2
M									5.6	3.3	5.3	4.2
N									5.6	2.2	5.3	4.2
P									5.6	3.3	5.3	4.2
Q	100	100							5.6	5.6	5.3	0
R					33	41.7			5.6	6.7	5.3	20.8
S			100	100					5.6	7.8	5.3	0
T									5.6	5.6	5.3	4.2
V									5.6	7.8	5.3	0
W					33	37.5			0	0	5.3	4.2
Y					33	20.8			5.6	7.8	5.3	4.2

which indicated a randomization had been achieved successfully. The amino acid residues at positions with limited residue variability (e.g. 100E and 102) were also in agreement with planned subsets and frequencies. The amino acid distribution of randomized V_κ-CDR3 and V_λ-CDR3 were not as good as that of V_H, which was most likely due to a limitation of the number of samples included. Several clones contained unplanned amino acids (e.g. Ser at position 93 of V_H-CDR3, Arg at position 90 and 95 of V_κ-CDR3) or single nucleotide deletion or insertion, which was mostly due to mutations introduced by PCR during assembling. A slight amino acid distribution bias at some positions (e.g. position 102 of V_H-CDR3 and position 96 of V_λ-CDR3) was probably caused by DNA synthesis manipulation or sampling limitation. The statistical results of 480 amino acids of V_H-CDR3 and more than 110 amino acids of V_L-CDR3 indicated that the split–mix–split method used in DNA synthesis was a reliable approach for creating diversity by sequence and length variation.

Bio-panning of the Library Against BHL and Analysis of scFvs—To evaluate the utility of the library, a 58 kDa anti-CD3 × anti-ovarian carcinoma bispecific antibody, was used as target antigen. Both the anti-CD3 and anti-ovarian carcinoma moieties of BHL were derived from murine monoclonal antibodies, which makes it difficult to raise antibody against BHL from rodent conventionally, while *in vitro* screening of a phage display antibody library derived from human antibody genes will take effect. The V_H3-V_κ sublibrary was subjected to panning against BHL. After three rounds panning, 10 positive BHL-binders (clone B1–B10) were sequenced and used to infect *E. coli* strain HB2151 cells to produce soluble scFv fragments. Among these 10 scFvs, clone B1, B7 and B9 showed higher binding activities than others. The deduced amino acid sequences of V_H-CDR3 and V_L-CDR3 were listed in Table 8. It showed that the occurrence ratios of R96, D96, A97 and S95 at V_H-CDR3, 91L and 92S at L-CDR3 were noticeably different from other amino acids due to selective enrichment of amino acids at these positions.

Specific Binding of scFv to BHL—To identify the binding activities of clone B1, B7 and B9, they were subcloned into expression vector pTCM2 by digesting with *Xho*I and *Eco*RI (Fig. 3D). The products of pTCM2 have both a $(His)_6$ and *C-myc* tag at the C-terminus. The binding activities of clone B7 to BHL were assayed by ELISA. As shown in Fig. 4, refolded clone B7 bound to BHL protein specifically.

DISCUSSION

Amplification of CDR3 regions from immuned or naïve B lymphocytes by PCR is a common method used to generate antigen-induced or natural-biased antibody repertoires. Artificial synthesis of randomized CDR3 region represents an alternative to create diversity in antibody. The amino acid composition and distribution of V_H-CDR3 and V_L-CDR3 domains showed the highest diversity in antibody genes. The length of V_H-CDR3 was a range from 4 to 28 residues (*14*). The lengths of V_κ-CDR3s and V_λ-CDR3 ranged from less than 7 to 10, and from 7 to 12 residues, respectively (*15*, *16*). Due to technical limits, it is difficult to cover such a broad spectrum of sequence diversity in a library; therefore, it is a rational approach to select the representative CDR3 lengths and amino acid compositions in construction of an antibody library. The degeneracy scheme adopted in library construction can greatly influence the quality of library and screening efficiency. Some solutions have been developed to resolve the problem of randomization (*9*, *11*, *17–19*). In this study, the sequences encoding fully

Table 7. **Comparison between designed and experimental amino acid composition in CDR3 regions—V_k.**

Position	89		90		91		92		93		94		95		96	
Amino acid	P	F	P	F	P	F	P	F	P	F	P	F	P	F	P	F
A					2.9	0					5.3	8.3	0	4.2	5.3	8.3
C					0	0					0	0			0	0
D					2.9	8.3	20	16.7	25	25	5.3	6.3			5.3	6.3
E					2.9	4.2					5.3	4.2			5.3	4.3
F	5	6.3			2.9	0					5.3	2.1			5.3	2.1
G					2.9	4.2	20	20.8	25	12.3	5.3	4.2			5.3	4.2
H	5	6.3			2.9	8.3					5.3	6.3			5.3	6.3
I					2.9	4.2					5.3	6.3			5.3	6.3
K					2.9	0					5.3	2.1			5.3	2.1
L	5	6.3			2.9	0					5.3	10.4			5.3	10.4
M	5	6.3			2.9	4.2					5.3	8.3			5.3	8.3
N					2.9	8.3	20	20.8	25	20.8	5.3	4.2			5.3	4.2
P					0	0					5.3	6.3	80	66.7	5.3	6.3
Q	80	75	100	93.8	2.9	4.2					5.3	2.1			5.3	2.1
R			0	6.3	2.9	4.2					5.3	6.3	0	4.2	5.3	6.2
S					2.9	0	20	25	25	37.5	5.3	2.1	20	20.8	5.3	2.1
T					2.9	0			0	4.17	5.3	8.3			5.3	8.3
V					2.9	0					5.3	2.1			5.3	2.1
W					2.9	12.5					5.3	4.2			5.3	4.2
Y					50	37.6	20	16.7			5.3	6.3	0	4.2	5.3	6.3

The numbering of amino acid residues was according to Kabat *et al.* (*13*), the designed amino acid composition (column P) is compared with the experimental composition (column F).

Table 8. **Deduced amino acid sequences in V_H -CDR3 and V_L-CDR3 of selected clones 10 HBL binders containing various CDR3.**

Clone	V_H-CDR3	V_L-CDR3
1	FWHQYS	QQKSDEPQ
2	RDAENI	LQLQKDPW
3	RDLDVA	QQLSSGPD
4	TKHSMA	QQESDRSP
5	NAESWD	QQYSGGPQ
6	PPANEP	MQQSSYPA
7	VYVSGE	QQYSDNPM
8	LDVAEQ	HQYSDKSR
9	VWEDVG	QQYDDKPS
10	RQQESI	QQWSGTPW

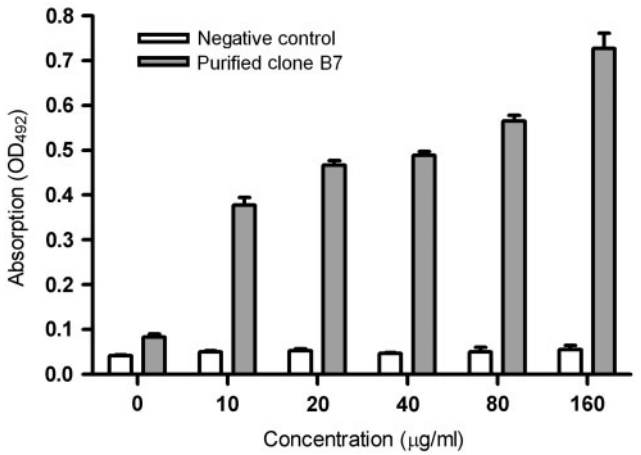

Fig. 4. **ELISA showing the binding of scFv to BHL protein.** The OD_{492} value varied associated with the concentration of clone B7 in a dose-dependent manner, while the negative control (BSA instead of scFv) had no such effect.

randomized or typical subsets of amino acids were synthesized by means of 'split–mix–split' method. The statistical results of 480 amino acids of V_H chain CDR3 and more than 110 amino acids of V_L chain indicated that the 'split–mix–split' method was a reliable approach for creating diversity by sequence and length variation. In 'split–mix–split' method, each codon was synthesized independently; hence, the codon bias was not progressive. The appearance of some unplanned amino acids and slightly amino acid bias were mostly caused by errors of PCR, and besides, the coupling activity difference of four bases, errors in pooling of different portions and insufficient mixing during synthesis would also result in uneven randomization. These defects, however, were not intrinsic and could be improved or even eliminated through elaborate operation.

The redundancy of an antibody library mainly means the disproportion of codons encoding with conventional 'NNK' or 'NNS' triplets. For a PCR-based library construction, redundancy can also be caused from the rapid amplification of gene copies during assembly PCR and cell proliferation. The redundancy of codon encoding for different amino acids must be removed completely to improve the efficiency of randomization, but redundancy of sequences seems to be inevitable. In most cases, a certain degree of colony and sequence redundancy is expected to complete the library construction, preservation and screening. A library with rational redundancy has obvious advantages in library preservation, modification and screening. Although higher primer concentration could improve the yielding of PCR products (*20*), we found that during the assembling of V_H, V_L and scFv fragments, sufficient diversity and appropriate

redundancy could be achieved simultaneously by performing assembly PCR with low-primer concentration and less PCR cycles. After transformation, the copy number of cells increased rapidly in log growth phase, so the cells should be harvested after ~3–4 h of growth. Usually, the library size is estimated according to the number of independent colonies with correct sequences, in consideration of the sequence redundancy, the actual library size may be lower than the estimated library size at least 1–2 orders.

An antibody library constructed based on different frameworks would introduce distinct biophysical and biochemical characteristics to its members, so it is expedient to construct a library based on frameworks with similar expression traits. It was found that a scFv derived from V_H3 (*DP47*) and $V_\kappa3$ (*DPK-22*) combination possessed relative good expression pattern (*12*, *21*, *22*). In this study, soluble portion could be detected in most cases; however, the products were mainly expressed in inclusion body forms in a T7 promoter-based expression vector, which makes it difficult to be purified by simple affinity chromatography method. Besides the sequence of an antibody, some extrinsic factors would also influence the expression and selection of scFv in phage display (*23–26*). More efforts should be made to resolve the problem thoroughly and improve the efficiency of selection.

FUNDING

Chinese Ministry of Sciences 863 Project (2006AA 02A245, 2006AA02Z457); National Protein Project (2006CB910901).

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Hanes, J., Schaffitzel, C., Knappik, A., and Pluckthun, A. (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–1292
2. Rothe, A., Hosse, R.J., and Power, B.E. (2006) Ribosome display for improved biotherapeutic molecules. *Expert. Opin. Biol. Ther.* **6**, 177–187
3. Daugherty, P.S. (2007) Protein engineering with bacterial display. *Curr. Opin. Struct. Biol.* **17**, 474–480
4. Hoogenboom, H.R. (2005) Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23**, 1105–1116
5. Benhar, I. (2007) Design of synthetic antibody libraries. *Expert. Opin. Biol. Ther.* **7**, 763–779
6. Ling, M. (2003) Large antibody display libraries for isolation of high-affinity antibodies. *Comb Chem. High Throughput Screen.* **6**, 421–432
7. Groves, M., Lane, S., Douthwaite, J., Lowne, D., Rees, D.G., Edwards, B., and Jackson, R.H. (2006) Affinity maturation of phage display antibody populations using ribosome display. *J. Immunol. Methods* **313**, 129–139
8. Sblattero, D. and Bradbury, A. (2000) Exploiting recombination in single bacteria to make large phage antibody libraries. *Nat. Biotechnol.* **18**, 75–80
9. Knappik, A., Ge, L., Honegger, A., Pack, P., Fisher, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86
10. Barbas, C. III., Bain, J.D., Hoekstra, D., and Lerner, R. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl Acad. Sci. USA* **89**, 4457–4461
11. Yin, C.C., Liu, J., Lin, Z.H., Wang, S.Z., Wang, X.B., and Huang, H.L. (2003) An efficient approach to the synthesis of a high-quality random peptide repertoire. *Acta Biochem. Biophys. Sin.* **35**, 921–924
12. Wang, S.Z., Yin, C.C., Lin, Z.H., Wang, X.B., Liu, J., Zhang, Y., Yu, W.G., Guan, H.S., and Huang, H.L. (2004) Approach for construction of a completely synthetic human antibody library from well-expressed framework. *High Tech. Lett.* **14**, 32–38
13. Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991) *Sequences of Proteins of immunological interest*, 5th edn, NIH Publication no.91-3242, US Department of Health and Human Services, Washington, DC.
14. Wu, T.T., Johnson, G., and Kabat, E.A. (1983) Length distribution of CDRH3 in antibodies. *Proteins: Struc. Funct. Genet.* **16**, 1–7
15. Tomlinson, I.M., Cox, J., Gherardi, E., Lesk, A., and Chothial, C. (1995) The structural repertoire of the human VK domain. *EMBO J.* **14**, 4628–4638
16. Williams, S.C., Frippiat, J.P., Tomlinson, I.M., Ignatovich, O., Lefranc, M.P., and Winter, G. (1996) Sequence and evolution of the human germline V lambda repertoire. *J. Mol. Biol.* **264**, 220–232
17. Hughes, M., Nagel, D., Santos, A., Sutherland, A., and Hine, A. (2003) Removing the redundancy from randomised gene libraries. *J. Mol. Biol.* **331**, 973–979
18. Tabuchi, I., Soramoto, S., Ueno, S., and Husimi, Y. (2004) Multi-line split DNA synthesis: a novel combinatorial method to make high quality peptide libraries. *BMC Biotechnol.* **4**, 19–26
19. Neylon, C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.* **l32**, 1448–1459
20. Czerny, T. (1996) High primer concentration improves PCR amplification from random pools. *Nucleic Acids Res.* **24**, 985–986
21. Ewert, S., Honegger, A., and Plückthun, A. (2003) Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. *Biochemistry* **42**, 1517–1528
22. Ewert, S., Huber, T., Honegger, A., and Plückthun, A. (2003) biophysical properties of human antibody variable domains. *J. Mol. Biol.* **325**, 531–553
23. Plückthun, A., Krebber, A., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K., and Riesenberg, D. (1996) Producing antibodies in Escherichia coli: from PCR to fermentation in *Antibody Engineering: A Practical Approach* (McCafferty, J., Hoogenboom, H.R., and Chiswell, D.J. eds.) pp. 203–252, Oxford University Press, New York.
24. Pavonia, E., Monteriùa, G., Cianfrigliab, M., and Minenkova, O. (2007) New display vector reduces biological bias for expression of antibodies in *E. coli*. *Gene* **391**, 120–129
25. Soltes, G., Hust, M., Ng, K.K., Bansal, A., Field, J., Stewart, D.I., Dübel, S., Cha, S., and Wiersma, E.J. (2007) On the influence of vector design on antibody phage display. *J. Biotechnol.* **127**, 626–637
26. Kirsch, M., Zaman, M., Meier, D., Dübel, S., and Hust, M. (2005) Parameters affecting the display of antibodies on phage. *J. Immunol. Methods* **301**, 173–185