ORIGINAL PAPER

Degenerate primer design to clone the human repertoire of immunoglobulin heavy chain variable regions

Ying Sun · Hong-Yan Liu · Ling Mu ·
En-Jie Luo

Abstract Amplifying the variable (Fv or V) regions of immunoglobulins (Ig) has become a challenge in cloning antibody genes for phage display, a technique used to study protein–protein, protein–peptide, and protein–DNA interactions using bacteriophages to connect proteins with the genetic information that encodes them. Key parameters affecting the amplification of full antibody repertoires includes the availability of primers that can amplify as many V genes as possible; however the strategy used to design these primers and programs used to make the necessary alignments have not been well studied and clearly detailed in the literature. Here, we present a set of primers computationally designed by iCODEHOP based on a database of human germline Ig sequences. We used reverse transcription polymerase chain reaction (RT-PCR) protocols that would recognize the V_H genes from human peripheral blood mononuclear cells. We identified the most highly conserved region in framework 1 and framework 4 of the Ig cDNA, and designed a set of degenerated 5′ primers. The V_H genes were successfully amplified by RT-PCR. This new primer has facilitated the creation of more diverse V_H libraries than has been previously possible. Moreover, iCODEHOP improved the primer design efficiency and was found useful both for cloning unknown genes in gene families and for building V_H gene libraries.

Introduction

Antibodies are indispensable tools for basic molecular biology and biotechnology research as well as for diagnosing and treating many diseases. Some neutralizing monoclonal antibodies (Mabs) have already been developed to treat disorders such as anti-severe acute respiratory syndrome (SARS) (Traggiai et al. 2004) and anti-human immunodeficiency virus (HIV) infection (Mehandru et al. 2004; Ferrantelli et al. 2004). A recent development in immunodiagnosis of human disease has been the introduction of molecular recombination techniques leading to the emergence of new analytic approaches based on the essential properties of the single-chain variable fragment (ScFv) (Hafner et al. 2000), or the selection of a relevant recombinant antibody with the use of ScFv phage-display libraries (McCafferty et al. 1990; Clackson et al. 1991; Tanaka and Rabbitts 2009; Pansri et al. 2009) and the production of the recombinant antibodies using different expression systems (Pollock et al. 1999; Yang et al. 2005). The long-term goal of our laboratory is to characterize the Ab repertoire generated against the Hantaan virus following construction of an immune phage ScFv library.

Approaches to define and characterize the Ab repertoire of an endless library require amplifying and/or cloning the Ig V_H and V_L genes. Investigators have tried to devise an universal primer or set of primers for amplifying all possible human V genes (Welschof et al. 1995; McCafferty and Johnson 1996; Sblattero and Bradbury 1998). The primers reported in these studies were designed based on the human Ig sequence data available at the time these

Y. Sun · E.-J. Luo (✉)
Department of Microbiology and Parasitology and Institute of Pathology and Pathophysiology, China Medical University, No. 92, Beier Road, Heping District, Shen Yang 110001, China
e-mail: enjie359@yahoo.cn

H.-Y. Liu · L. Mu
Shenyang Hospital for Infectious Diseases, No. 85, Heping South Avenue, Heping District, Shen Yang 110006, China

Table 1 Oligonucleotide primers for (RT-)PCR amplification of Ig V_H genes

Family or gene[a]	Total no. of genes (%)[b]	Primer	Primer sequence[c]	Fold deg[d]
VH1	30 (14)	VH1/5for	CAGTCCGGCGCCgargtbaaraa	12
VH2	21 (10)	VH2for	GGCCCCGCCCTGgtbaarccnac	24
VH3	84 (40)	VH3for	GGCGGCGGCstggtscagcc	4
VH4	62 (30)	VH4for1	GGCCCCGGCytvgtbaarcc	36
		VH4for2	CAGGTGCAGCTGcarcartgggg	4
VH5	7 (3)	VH1/5for		12
VH6	2 (1)	VH6for	CAGCTGCAGCAGTCCggnccngg	16
VH7	4 (2)	VH7for	GGCCACGAGCTGaaraarccngg	4
JH	13	Jrev[e]	tgaggagacggtgaccrkkgtbcc	24

[a] The 37 pseudogenes in the human germline V_H IgBLAST database were diverse in sequence and were not analysed for designing primers. 19 sequences were incomplete and were excluded from the analysis

[b] The number of germline sequences present in the family analysed. The proportion of genes contributed by a given family to the total number of genes present in our V_H database is given in parenthesis

[c] All oligonucleotide primers are listed in the 5′–3′ direction. A CODEHOP primer is degenerate at the 3′ core region, which is shown in lower case, and is non-degenerate at the 5′ consensus clamp region, which is shown in upper case. IUPAC code for mix bases: M, a/c; R, a/g; W, a/t; Y, c/t; S, c/g; K, g/t; H, a/c/t; V, a/c/g; D, a/g/t; B, c/g/t; N, a/c/g/t

[d] Fold degeneracy

[e] The reverse primer is commonly used with all forward primers

primers were devised. This raises the question whether these primers still cover all known genes adequately. One research group has attempted to overcome the shortcomings of this earlier research by using a set of different V-gene specific primers in combination with Ig-class specific reverse primers (Lim et al. 2010). However, Lim's study did not design primers for building ScFv directly.

In our study, we initially used commercially available human primer sets from Novagen Corporation designed for amplifying Ig variable region cDNAs. The V_L genes were amplified successfully. However, the V_H genes were also inappropriate to build ScFv directly. This prompted us to design our own primers. To accomplish this, we used a new interactive web application that simplified the process of designing and selecting COnsensus DEgenerate Hybrid Oligonucleotide Primers (CODEHOP) from multiply-aligned protein sequences. The iCODEHOP program is described in more detail at https://icodehop.cphi.washington.edu/i-codehop-context/Welcome (Boyce et al. 2009), We then designed a set of specific degenerate primers from the iCODEHOP application, which are indispensable for applications requiring broad V gene family coverage.

Materials and methods

Bioinformatic analysis

Sequences corresponding to the functional V and J genes for Ig V_H genes were downloaded from IMGT®, the international ImMunoGeneTics® information system http://www.imgt.org (Giudicelli et al. 2006) and IgBLAST databases described at http://www.ncbi.nlm.nih.gov/igblast/. These sequences were grouped into various Ig gene families according to IMGT nomenclature. The 229 VH and 13 JH germline gene sequences were retrieved to constitute our working data set.

Primer design

A CODEHOP is a hybrid primer comprising a degenerate 'core' and nondegenerate 'clamp' region (Table 1). The core region locates on the 3′-end of the primer and contains the nucleotide sequences providing all possible codons for a highly conserved amino acid motif of 3–4 residues identified in a protein multiple alignment. The nondegenerate clamp region consists of the most common nucleotides in each position of the codons for 5–7 amino acid positions immediately successive to the conserved motif. This region is usually between 15 and 20 bases, a length that can be adjusted by the user.

Design of the clamp was based on an alignment of available Ig germline genes and codon usage of the target organism. The primers are completely degenerate at their 3′-end and we ensured a high probability of annealing to the target sequences by using the CODEHOP strategy based on the multiply-aligned sequences. The multiple alignment and blocks format were generated by the BlockMaker server (http://blocks.fhcrc.org/blocks/). Primers were designed using the default parameters of the iCODEHOP server, then 7 forward primers were selected.

The reverse primer, which was designed manually according to the alignment of 13 JH germline nucleotide sequences, was commonly used with all forward primers (Table 1). The primers could amplify the V_H genes from framework region 1 (FR 1) to the joining region.

RT-PCR

Peripheral-blood mononuclear cells (PBMCs) were isolated from convalescent patients with hemorrhagic fever with renal syndrome (HFRS) diagnosed by density gradient centrifugation (Lymphocyte Separation Medium, Applygen Technologies Inc). Total RNA was extracted from 1×10^6 PBMCs using the RNAiso Plus (TaKaRa Bio Inc). A total of 400 ng of RNA was reversely transcribed in a 20 μl reaction volume using the TaKaRa RNA PCR Kit (AMV) Version 3.0 (TaKaRa Bio Inc), which was used as template for PCR (5 μl of cDNA for each reaction in 50 μl reaction volume). For further cloning purposes, primers were designed to add a HindIII restriction site on the forward primer and an EcoR1 site on the reverse primer. Amplification conditions were 30 s at 94°C, 30 s at 53°C, and 30 s at 72°C for 35 cycles and final extension at 72°C for 10 min. All reactions were carried out using a TaKaRa PCR Thermal Cycler Dice™ instrument (TaKaRa Bio Inc). Primers used in this study are listed in Table 1. PCR products were purified from 1.5% agarose gel with the TaKaRa Agarose Gel DNA Purification Kit Ver.2.0 (TaKaRa Bio Inc) and blunt-cloned in the PMD18T- vector (TaKaRa Bio Inc). Ligations were used to transform *E. coli* TG1 cells and plated on LB/Amp/IPTG/X-gal plates for blue-white screening. For each V_H family, up to 4 random clones were sequenced using a standard M13-47primer (5'-CGCCAGGGTTTCCCAGTCACGAC-3').

Results

Data analysis and primer design

To develop highly specific and sensitive PCR that can potentially amplify human rearranged/expressed V_H gene belonging to any V gene family, we adopted a RT-PCR strategy (Fig. 1). We took the human Ig locus biology into account while designing the primer set. The application of PCR technology for analysing V genes is a challenging task, primarily because of the diverse character of the V genes. The primers were designed to cover all functional germline Ig genes. The pseudogenes could not be included in designing the primers either, because the sequences were too divergent or the available sequences were truncated and did not cover the primer binding sites (Table 1).
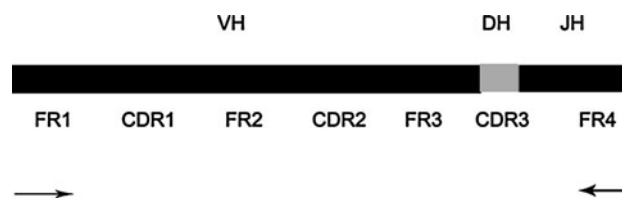

Fig. 1 Schematic diagram outlining the (RT-)PCR based strategy used. The (RT-)PCR strategy for amplifying the human rearranged/expressed V_H genes using peripheral blood mononuclear cells as starting template is shown. The variable (V_H), diversity (D_H) and joining (J_H) regions of the Ig H chain corresponding to the rearranged framework1 (FR1), complementary determining region1 (CDR1), FR2, CDR2, FR3, CDR3 and FR4 are indicated. The *arrows* pointing to the *right* and *left* indicate the orientation of the forward and reverse primers. The binding site for the 5' and 3' primer is located in FR1 and FR4, respectively

In the first step of designing the primers, we downloaded human germline Ig sequences from IMGT and IgBLAST databases. The sequences were grouped in 7 VH (229 protein sequences) and 1 JH (13 nucleotide sequences) families according to the IMGT nomenclature. The germline V gene sequences belonging to the same Ig gene family were aligned and the most conserved region was chosen through the BlockMaker server.

In the second step of designing the primers, we used the iCODEHOP program of conserved blocks to design forward degenerate primers for the V region. The reverse primers were designed according to the JH nucleotide sequences manually. With these tools, we generated a novel set of primers (Table 1) that make it theoretically feasible to amplify and clone the entire Ig V_H repertoire. In most cases, a single primer was sufficient to cover all the members of an Ig gene family—for example, a single primer was sufficient to cover V_H1 and V_H5 genes. However, there were some exceptions: two primers were required to cover the 62 genes belonging to the V_H4 family.

RT-PCR with the primer designed by iCODEHOP

To check whether the primers designed by computer simulation were suitable to clone V_H specificities, we performed RT-PCR with all the primer pairs. As shown in Fig. 2, all the reactions of the VH forward primers produced PCR fragments of the expected size when the new primer set was adopted. The VH fragment size was approximately 350 bp. The PCR products showed minor variations in size across families, depending on the position of the 5' primer binding site (Fig. 2).

To confirm the specificity and diversity of the amplification products, each PCR fragment for V_H amplifications was purified, blunt-cloned, and independently used to transform *E. coli* cells. Several random clones from each transformation were sequenced (Table 2). Among all
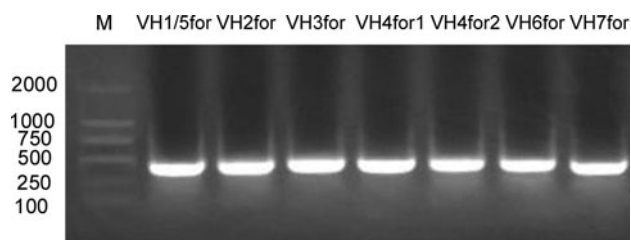
Fig. 2 PCR amplification of V gene families from human peripheral blood mononuclear cells. PCR products obtained for individual VH gene families using primers described in Table 1, were resolved on agarose gels. The name of the forward primer is indicated. The expected size (~ 350 bp) of the PCR products were obtained for 7 V_H gene families. *M* molecular weight marker (2,000, 1,000, 750, 500, and 250 bp)

random sequenced clones, we did not find any no-V_H gene sequences, a finding that confirms the selectivity of our primers. In all instances, sequence analysis revealed that the rearranged V genes recovered were functional and showed a broad V gene usage pattern.

Discussion

The availability of databases comprising gene sequences encoding all Ig genes (IMGT/GENE-DB) has allowed PCR-mediated cloning of antibody repertoires and has shed light on the immune responses in human and mouse. Germline V, D, and J gene sequences encoding V_H chains, were retrieved from the IMGT information system.

Furthermore, the engineering of synthetic antibodies has become an important methodology for generating

diagnostic and therapeutic molecules, such as those used by phage display or ribosome display. A good coverage of V genes is desirable when generating antibody libraries for phage display to ensure the retrieval of diverse set of binding antibodies during selection. Because the 400 bp length of an antibody V gene has approximately 10^8 variations, amplifying a Fv is more complicated than an unknown gene in other gene families. To amplify the V_H genes of Ig cDNAs from PBMCs, we first had to design primers of relatively low degeneracy to realize the natural benefits of a degenerate primer to cover every family sequence and minimize the number of primers. To accomplish this, we focused on selecting conserved regions of the V genes and the degeneracy of primers.

Although there are several other publicly available tools for designing degenerate primers, iCODEHOP is the only tool that contributes to designing primers with the consensus-degenerate hybrid format. It is also one of only a few programs that design primers from amino acid multiple alignments. Several existing systems, including SCPrimer (Jabado et al. 2006), Amplicon (Jarman 2004), and HY-DEN (Linhart and Shamir 2005) design degenerate primers entirely from nucleic acid multiple alignments. This approach is profitable when nucleotide information is available and avoids the problem of back-translating amino acid sequences. However, it is sometimes impossible to recognize conserved regions among the many nucleotide sequences that can encode a protein family of interest, even when conservation is visible at the amino acid level. To the best of our knowledge, GeneFisher (Giegerich et al. 1996; Lamprecht et al. 2008) is the only other publicly available web application besides iCODEHOP that both designs

Table 2 Ig V_H primers validation

Primer specificity validation by DNA sequencing. PCR products (see Fig. 2) were cloned in pmd18/T vector and up to four clones randomly selected and sequenced

Forward primer	Recombinants analysed	Independent rearrangements		
		V	D	J
VH1/5for	3	IGHV5-51*01	IGHD1-7*01	IGHJ4*02
		IGHV1-8*01	IGHD6-6*01	IGHJ4*01
		IGHV5-51*01	IGHD4-17*01	IGHJ3*02
VH2for	4	IGHV2-5*09	IGHD3-16*02	IGHJ4*02
		IGHV5-51*01	IGHD4-17*01	IGHJ4*01
		IGHV4-31*03	IGHD2-15*01	IGHJ5*02
		IGHV2-5*01	IGHD-19*01	IGHJ4*02
VH3for	2	IGHV3-66*02	IGHD2-8*02	IGHJ3*02
		IGHV3-30*04	IGHD1-26*01	IGHJ4*02
VH4for1	1	IGHV4-59*01	IGHD3-16*01	IGHJ6*02
VH4for2	1	IGHV4-59*01	IGHD7-27*01	IGHJ4*02
VH6for	2	IGHV6-1*02	IGHD3-22*01	IGHJ4*01
		IGHV4-31*10	IGHD1-26*01	IGHJ4*02
VH7for	1	IGHV7-4-1*02	IGHD4-23*01	IGHJ1*01

degenerate primers from amino acid sequences and generates degenerate primers that have been verified in laboratory experiments. GeneFisher2 appears to allow degenerate positions at any position along the primers it produces. In contrast, the CODEHOPs created by iCODEHOP have degenerate positions only within 11–12 bases at their 3′ end. The remaining positions of a CODEHOP are derived from the input amino acid multiple alignment by determining the most common codons of the consensus amino acids created from the alignment. As discussed above, we selected iCODEHOP to design degenerate primers for Ig V_H genes.

The iCODEHOP program, which can be employed to design degenerate primers, also has shortcomings. Under certain conditions, iCODEHOP is incapable of designing primers for a particular protein multiple alignment because conservation is too poor among the selected sequences for the program to detect primers that satisfy degeneracy and clamp length constraints. We used two methods of alignment to find the more conserved region. The first method was to use all human Ig gene sequences in all families defined by IMGT aligned as one group. However, the conserved region was not found, so we abandoned this alignment approach. The second method we tried was to take all human Ig gene sequences listed and align them within each family. This way, we selected the optimal region, which was in the begin or in the middle of the V_H FR1. Based on this alignment, the number of primers designed by our program at the 5′ end of the V_H region was 7, less than the number of primers designed by other authors. For 3′ primer design, known FR4 sequences are normally chosen as the target sequences. The JH genes retrieved from IMGT databases were 13, so we appreciated designing degenerate primers manually.

To validate whether this set of primers designed computationally was suitable to clone V_H specificities, several random clones were sequenced and analysed. We showed that although we sequenced a relative small quantity of clones, we found a high percentage of all potential genes. However, the primers were not family-specific because the human V genes showed intra- and inter-family sequence variability, and they were somatically rearranged in a way that generated nearly limitless Ab diversity. Additionally, somatic hypermutation events that coincide with the primer binding sites can potentially affect the amplification efficiency. Therefore, some primers matched more than one V gene family and some V genes were matched by more than one primer.

Although this might limit the utility of the primer set described for clonotypic analyses, this considerably increased the chances to clone most, if not all, V_H gene transcripts, and turned out useful for creating libraries representative of V_H gene repertoires.

Conclusions

Our purpose was to create a primer set able to optimally amplify all Ig V_H genes, an objective we accomplished with the set of primers designed by iCODEHOP. This set will allow us and others to profile the V_H repertoire as well as create libraries such as those based on ScFv. Furthermore, this approach could be expanded to immunoglobulins from other species and to other members of the immunoglobulin family such as the T receptor.

References

Boyce R, Chilana P, Rose TM (2009) iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. Nucleic Acids Res 37(Web Server issue):W222–W228

Clackson T, Hoogenboom HR, Griffiths AD, Winter G (1991) Making antibody fragments using phage display libraries. Nature 352:624–628

Ferrantelli F, Kitabwalla M, Rasmussen AR, Cao C, Chou CT, Katinger H, Stiegler G, Cavacini AL (2004) Potent cross-group neutralization of primary human immunodeficiency virus isolates with monoclonal antibodies—implications for acquired immunodeficiency syndrome vaccine. J Infect Dis 189(1):71–74

Giegerich R, Meyer F, Schleiermacher C (1996) GeneFisher-Software Support for the Detection of postulated genes. In: Proceedings of the fourth international conference on intelligent systems for molecular biology. AAAI Press, Menlo Park, CA 94025, pp 68–77

Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, Lefranc MP (2006) IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. Nucleic Acids Res 34:D781–D784

Hafner FT, Kautz RA, Iverson BL, Tim RC, Karger BL (2000) Noncompetitive immunoassay of small analytes at femtomolar level by affinity probe capillary electrophoresis: direct analysis of digoxin using a uniform-labeled ScFv immunoreagent. Anal Chem 72:5779–5786

Jabado OJ, Palacios G, Kapoor V, Hui J, Renwick N, Zhai J, Briese T, Lipkin WI (2006) Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments. Nucleic Acids Res 34:6605–6611

Jarman SN (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. Bioinformatics 20:1644–1645

Lamprecht AL, Margaria T, Steffen B, Sczyrba A, Hartmeier S, Giegerich R (2008) GeneFisher-P: variations of GeneFisher as processes in Bio-jETI. BMC Bioinformatics 9(Suppl. 4):S13

Lim TS, Mollova S, Rubelt F, Sievert V, Dübel S, Lehrach H, Konthur Z (2010) V-gene amplification revisited—An optimised procedure for amplification of rearranged human antibody genes of different isotypes. New Biotechnol. doi:10.1016/j.nbt.2010.01.001

Linhart C, Shamir R (2005) The degenerate primer design problem: theory and applications. J Comput Biol 12:431–456

McCafferty J, Johnson KS (1996) Construction and screening of antibody display libraries. In: Kay BK, Winter J, McCafferty J (eds) Phage display of peptides and proteins. Academic Press, San Diego, pp 79–111

McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: filamentous phage displaying antibody variable domains. Nature 348:552–554

Mehandru S, Wrin T, Galovich J, Stiegler G, Vcelar B, Hurley A, Hogan C, Vasan S (2004) Neutralization profiles of newly transmitted human immunodeficiency virus type 1 by monoclonal antibodies 2G12, 2F5, and 4E10. J Virol 78(24):14039–14042

Pansri P, Jaruseranee N, Rangnoi K, Kristensen P, Yamabhai M (2009) A compact phage display human scFv library for selection of antibodies to a wide variety of antigens. BMC Biotechnology 9:6. doi:10.1186/1472-6750-9-6

Pollock DP, Kutzko JP, Birck-Wilson E, Williams JL, Echelard Y, Meade HM (1999) Transgenic milk as a method for the production of recombinant antibodies. J Immunol Methods 231:147–157

Sblattero D, Bradbury A (1998) A definitive set of oligonucleotide primers for amplifying human V regions. Immunotechnology 3:271–278

Tanaka T, Rabbitts TH (2009) Selection of complementary single-variable domains for building monoclonal antibodies to native proteins. Nucleic Acids Res 37(5):e41

Traggiai E, Becker S, Subbarao K, Kolesnikova L, Uematsu Y, Gismondo RM, Murphy RB, Rappuoli R (2004) An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med 10:871–875

Welschof M et al (1995) Amino acid sequence based PCR primers for amplification of rearranged human heavy and light chain immunoglobulin variable region genes. J Immunol Methods 179:203–214

Yang XX, Hu WG, Li F, Xia HC, Zhang ZC (2005) Gene cloning, bacterial expression, in vitro refolding, and characterization of a single-chain Fv antibody against PreS1(21–47) fragment of HbsAg. Protein Expr Purif 41:341–348