

Sensitive recovery of recombinant antibody clones after their *in silico* identification within NGS datasets

Spiliotopoulos, Anastasios.^{a,c}, Owen, Jonathan. P.^b, Maddison, Ben. C.^b, Dreveny, Ingrid.^c, Rees, Helen. C.^b, Gough, Kevin. C.^{a*}

^aSchool of Veterinary Medicine and Science, The University of Nottingham, Sutton Bonington, Loughborough, Leicestershire, LE12 5RD, UK

^bADAS UK, School of Veterinary Medicine and Science, The University of Nottingham, Sutton Bonington, Loughborough, Leicestershire, LE12 5RD, UK

^cSchool of Pharmacy, Centre for Biomolecular Sciences, The University of Nottingham, Nottingham NG7 2RD, UK

* Corresponding author

e-mail: Kevin.Gough@nottingham.ac.uk

Telephone: +44 (0)1159 516272

Contact email addresses:

A. Spiliotopoulos: paxas@nottingham.ac.uk

J. P. Owen: Jon.Owen@adas.co.uk

I. Dreveny: Ingrid.Dreveny@nottingham.ac.uk

B. C. Maddison: Ben.Maddison@adas.co.uk

H. C. Rees: Helen.Rees@adas.co.uk

K. C. Gough: Kevin.Gough@nottingham.ac.uk

Abstract

Recently the analytical power of the latest high throughput next generation DNA sequencing platforms has been used to analyse phage that have been selected from the panning of large combinatorial libraries displaying either peptide or antibody ligands. This process, commonly referred to as next generation phage display (NGPD), allows the researcher to determine the identity of specific phage that are being enriched against an antigen target by analysis of the DNA sequence encoding the displayed ligand. This method bypasses several steps in conventional phage panning that include laborious colony picking and functional ligand screening. A downside of this approach is that the only output from such experiments is the DNA sequence information of such enriched phage particles. In the case of peptides, the peptide sequence can be synthesised directly and used for further screening, however this is more difficult with larger antibody fragments such as ScFvs. In the case of ScFvs, their coding sequence would have to be fully elucidated, synthesised and re-cloned before expression. We describe here the application of an inverse PCR-ligation methodology that enables the specific recovery of ScFvs of interest from enriched sub-libraries of phage clones. Phagemid particles are recovered using sequence information derived from their unique heavy chain CDR3/FR4 domains and specific clones can be recovered irrespective of CDR3 size and at levels of abundance that would be refractory to their discovery during conventional phage panning and screening.

Key words: phage display, ScFv, rescue, inverse PCR, CDR3

1.0 Introduction

The isolation of ligands from large combinatorial libraries is a powerful technique allowing the discovery of bioactive molecules with a wide range of applications including diagnostics and therapeutics. This technique is generally carried out within the context of a biological display system, and commonly these include phage, yeast or bacterial cells to display a functional biomolecule (Hoogenboom *et al.* 1998, Löfblom 2011, Gera *et al.* 2013). In doing so, the phenotype of the ligand is directly linked to its genotype, and the isolation of an individual clone displaying a particular binding property allows the isolation of the genetic sequence encoding the phenotype of interest (antibody, peptide etc). This technique has the potential to isolate ligands to a variety of target types from libraries that can contain billions of different specificities.

In all display systems a target molecule (often a protein) is introduced *in vitro* to a library of potential binding partners in a process that is known as biopanning. During the biopanning steps, subsets of the library bind to the immobilised target, and after washing away non-binders, the enriched ligands are eluted, propagated and then reintroduced into another round of biopanning against the same target. Iterative rounds of this biopanning can result in the enrichment of molecules with high specificity to the target. Individual clones of interest are identified by single clone screening (often by ELISA) to identify those specific clones of interest and isolated clones are then generally characterised by Sanger DNA sequencing.

Antibody fragments are commonly displayed within phage-display libraries. Within these libraries an ScFv domain is commonly displayed as a fusion to the gene 3 protein product of the *E. coli* M13 bacteriophage. The ScFv is an engineered fusion protein comprised of antibody heavy and light chain variable regions which are kept in close proximity via a short uncharged linker peptide. ScFv libraries can have antibody diversities as large as 10^{11} , (Lloyd

et al. 2011) and are based on the *in vitro* random re-assortment of the variable heavy and light chains from naïve or primed immune repertoires. The ScFv is a versatile reagent, amenable for further engineering and has found different applications from diagnostics to therapeutics (Hairul Bahara *et al.* 2013). Antibody phage display technology can, in ideal conditions, deliver in just a few weeks immunoreagents to almost any target of interest. However, due to the iterative nature of panning experiments they often enrich for phage particles that exhibit some growth advantage, (for example they possess a less toxic gene product) and as such the sub-libraries can quickly become biased towards such parasitic clones that may not be the best binders or at worst may not exhibit any activity to the target. A more effective approach is to identify potentially enriched phage after just a single round or two of biopanning, before specific immunoreagents can be outgrown and lost in later rounds. Such an approach would be highly labour intensive if carried out by conventional colony screening and success would not be guaranteed even after screening large numbers of individual clones. Recently the application of the latest next generation DNA sequencing (NGS) platforms have revolutionised how we can approach the analysis of such experiments. Instead of laborious colony screening by immunoassay, the complete output of ligand genes from each panning round can now be sequenced in its entirety at low cost giving unparalleled information on the success of a panning experiment. Those domains within the ScFv that have the most diversity, and form the centre of the antigen combining site, the CDR3 domains, have been the sequencing target of choice for *in silico* identification of enriched sequences. Methods for the analysis of the NGS datasets are either based on the comparison of clone frequency within the enriched sub libraries relative to the whole library (Ravn *et al.* 2010) or are based on the comparison of positive and negative panning reactions (panning the same library against a target and an unrelated molecule) and the generation of Z-scores representing individual ScFv enrichment (Zhang *et al.* 2012). Recently, Rebollo *et al.* also described using

the frequency of motif isolation within NGS datasets to identify enriched bicyclic peptide binders to different protein targets (Rebollo *et al.* 2014). The identification of enriched candidate sequences is now a relatively straightforward process using published bioinformatics scripts. In the case of peptide library enrichment, candidate peptides can be identified, synthesised and then screened for binding in an appropriate assay. With ScFvs, which are typically comprised of around 350-400 amino acids, chemical synthesis of the active molecule is not economically viable if indeed technically possible. Gene synthesis of such a coding sequence is possible, but for a range of candidate ScFvs requiring screening it would be very costly. Also, this would rely on the full sequence of the whole ScFv molecule being known, whereas only short read sequencing data around the CDR3 region(s) is usually recovered from NGS data files from most of the current DNA sequencing platforms.

Without efficient methods for the rapid and high throughput rescue of low abundance ScFv sequences of interest (those that may show enrichment during early library panning rounds) the application of NGS analysis to determine enriched ScFv-phage particles is of limited use. Here we demonstrate the sensitive recovery of ScFv clones within a background of sub-library DNA. The recovery method uses an inverse PCR, self-ligation and cloning strategy (Hoskins *et al.* 2005, D'Angelo 2014) to selectively amplify the desired ScFv. Our further development of this methodology demonstrates that this approach can be used to recover ScFv clones with very short CDR3 motifs and at low levels within a sub library, in one instance where a clone was represented at a level of just 0.0025% of the phage population. This higher sensitivity compared to any previously reported method will facilitate the rescue of specific ScFvs identified from NGS datasets of panning experiments for the rapid isolation of the best immunoreagents, and not just those that have selective growth advantages.

2.0 Materials and methods

2.1 ScFv antibodies: All ScFv sequences originated from the MRC Griffin-1 phagemid library. The sub-libraries of phage were generated from the selection of this library against different target proteins, across several experiments. Next generation sequencing datasets were from Ion Torrent PGM sequence analysis using a 318 chip. ScFv sequences that were identified and rescued (table 1) were from multiple panning experiments and different panning rounds. In addition, an ScFv clone (DU15-2) isolated using colony screening was used as a control ScFv for method development, this ScFv was isolated as a potential binder to the N-terminal Dusp-Ubl domains of human USP15 cysteine protease (Ubiquitin specific protease 15).

2.2 Inverse PCR of target ScFv-containing phagemid: Inverse PCR was carried out using primers designed to encompass the upstream region of the ScFv heavy chain CDR3 coding region identified during the sequencing of clones after panning. This was paired with another oligonucleotide annealing to the complementary DNA strand within the ScFv heavy chain framework region 4 (FR4; figure 1). Taking into account the degeneracy of the FR4 within the ScFv library, universal FR4 primers were synthesised based on the alignment of several FR4 domains from individual clones within the V-BASE database. The oligonucleotide complementary to the FR4 domain was TGGGGCCRDGGNACMMYGGTCAC. Where CDR3s were very small, the 5' end of the CDR3 specific primer was designed to be complementary to sequence within the FR4 region (figure 1). In these instances a clone specific FR4 primer was used that was complementary to a region adjacent to the CDR3/FR4 primer (ScFv 9 and ScFv 11, figure 1). These PCR primer sets direct the amplification of the entire ScFv containing phagemid of interest. In order to facilitate the blunt end cloning of these amplified fragments, each oligonucleotide was phosphorylated at the 5' end. Inverse PCR reactions consisted of 200 μ M dNTP mix, 5 pmol of each primer, and 1 unit of Q5 DNA

polymerase (NEB) plus 10µl of GC enhancer buffer, and 10-30 ng of DNA template in a final volume of 50 µl of 1X Q5 reaction buffer. PCR was carried out as follows: samples were 'hot started' by heating the thermocycler to 95 °C, before the addition of samples, whereupon samples were initially denatured for 2 min at 95 °C, and then given a first step of 10 or 15 cycles of 30 s annealing at between 53 - 65 °C (depending on the T_m of the primers) followed by 25 or 20 cycles respectively at a higher annealing temperature, compared to the first step, that varied from 60 to 68 °C; each cycle commenced with a 30 s denaturation step at 95 °C and the extension step was consistently performed at 72 °C for 5 min.

In testing this methodology a previously characterised ScFv (DU15-2) was used. Inverse PCR using the specific DU15-2 primer and the FR4-L degenerate primer (figure 1) was carried out using a dilution series of the monoclonal phagemid spiked into 10 ng of the Griffin-1 library DNA.

2.3 Cloning PCR products: PCR products were visualised by running the whole reaction product on a 1% TAE agarose gel stained with NANC520. Bands of the correct size were excised and purified using a Nucleospin DNA purification protocol. Amplified products contain two 5' phosphate groups making the PCR product amenable for self-ligation.

Ligations were carried out overnight at room temperature using 6 µl of purified PCR product, 400U T4 DNA ligase (NEB), in a final volume of 20 µl of 1X DNA ligase buffer. The ligation products (2 µl) were transformed into TG1 *E. coli* heat shock competent cells and were selected on LB-Amp agar plates overnight at 37 °C. Colonies were picked and used for PCR screening using the CDR3 specific primer and the primer VHFR2Den (a degenerate framework 2 primer: 5'GGVMARGGNCTKGASTGG 3'), binding to a sequence within the Griffin-1 library FR2 region. Positive clones identified by PCR then had the ScFv domain fully sequenced by Sanger sequencing using a commercial service.

3.0 Results

Using phagemid DNA of a defined ScFv that had been isolated by conventional colony screening and previously sequenced (DU15-2), we determined that the inverse PCR methods could generate PCR products of a size that corresponded to that of the linear ScFv phagemid (figure 2A). In order to determine how sensitive this PCR based methodology was, DNA from DU15-2 was serially diluted into DNA extracted from the MRC Griffin-1 ScFv library. PCR products analysed by agarose gel determined that this inverse PCR could successfully amplify specific ScFv clone DNA that was present at an amount equivalent to 0.026% of the total DNA in the PCR; a dilution of 1 in 3,846 (figure 2A). At each DNA dilution PCR products were excised from the agarose, purified and self-ligated. Transformation of these ligation products resulted in a high number of transformants. Transformants were screened by colony PCR using the CDR3 specific oligonucleotide paired with a degenerate primer that annealed to the FR2 region within the ScFv library. From the lowest amount of template spike that was amplified (1 in 3,846) we achieved a colony screen success rate of 100% (10 positive PCRs out of 10 reactions) that gave a PCR product corresponding to the expected amplicon size (data not shown). Of these colonies, three were further analysed by Sanger DNA sequencing, and all three clones were confirmed as containing the correct DU15-2 antibody sequence.

To demonstrate the further applicability of this technique to antibody clone rescue, where clones of interest may exist at very different frequencies and will possess very different heavy chain CDR3 sequences (both sequence composition and length), twelve ScFv clones were targeted from different panning experiments (table 1). These ScFvs represent a diversity of different length heavy chain CDR3 and clone abundance within the library. ScFvs with short heavy chain CDR3 domains (ScFvs 9 and 11) were amplified using primers which extended into the FR4 domain and the FR4-specific primer was shifted by an equal number of bases

into the FR4 region (table 1, figure 1). For each ScFv a PCR product was obtained (examples given in figure 2C). In all cases self-ligation of the PCR product followed by transformation into TG1 cells resulted in successful recovery of clones of correct DNA sequence to the desired ScFv as determined by Sanger sequencing. To demonstrate the sensitive recovery of ScFv clones containing shorter heavy chain CDR3 domains is also achievable using this method we also show the recovery by inverse PCR of DNA from clone 11 DNA at a dilution of 0.026% of the library DNA (Figure 2B).

4.0 Discussion

The recovery of ScFv clones is a major bottleneck to the utilisation of data derived from the next generation DNA sequencing analysis of phage-ScFv library panning. To date, several methodologies have been described for the rescue of specific ScFvs from enriched libraries. The method described by Zhang *et al.* (2012) uses a biotinylated oligonucleotide complimentary to the heavy chain CDR3 ssDNA of the phage DNA. Once ssDNA is enriched with this primer using capture on streptavidin coated beads it is converted to dsDNA and then used to transform *E. coli*. Another method described by Ravn *et al.* (2010) used the heavy chain CDR3 sequence to amplify and recover ScFvs of interest using a two step PCR approach. Two PCRs were used to isolate DNA for the VH and VL domains of the ScFv, these were then assembled using a separate overlap PCR via the specific heavy chain CDR3 region. The cloning of this PCR product back into the library vector allowed the selection of the specific clones of interest. This cloning rescue strategy requires 3 PCR steps, and the necessary restriction digestion of this DNA before ligation. Both of these described methods are more complicated, require additional steps and will be more costly than the inverse PCR methodology described here. These two publications do not determine the sensitivity of the

methods by rescuing serial dilutions of clones within background library DNA, however Zhang *et al.* recovered specific ScFv clones of interest at a level of 1% within background library DNA, and Ravn *et al.* rescued clones to around 0.5% of the library.

Recently, D'Angelo *et al.* (2014) have described a similar inverse PCR methodology to that detailed within the current paper. Both of these methods are a further application of SLIP-self ligation of inverse PCR products, developed by Hoskins *et al.* (2005) for the recovery of specific clones from drosophila cDNA libraries. D'Angelo and co-workers describe the rescue of ScFv clones by inverse PCR using specific PCR primers and demonstrate that this method was able to detect clones that were present at a level of 1.9% of the library, and suggest that they could isolate clones present at a levels down to 0.5% of the library. Their method uses two primers annealing to adjacent sequences within the HCDR3 and extending outwards into the framework regions. These FR-complementary 3' regions of the primers will be common to many ScFv clones and it is assumed that specificity of the primers for the target HCDR3 is due to their higher overall homology to that particular clone. It is likely therefore that the method will be more efficient for larger HCDR3s where the non-specific, FR-complementary 3' regions are proportionally less of the overall primer sequence. Here, we describe further developments of this technique to isolate ScFvs with a far greater sensitivity than any method reported previously. The reported method uses a single primer that is complementary to the HCDR3 along with a primer complementary to FR4. The FR4 primer is likely to be shared by numerous ScFv clones. However, the HCDR3 primer, including its 3' end, does not extend into common FR sequence allowing very high specificity in its annealing to, and extension of the target HCDR3 and therefore affording high specificity and concomitant efficiency in the overall inverse PCR rescue process. Using 2 characterised ScFv sequences with HCDR3 lengths of 11 and 5 amino acids spiked into ScFv library DNA, we could recover the clones when present at a level of 0.026% of the total

DNA. At this level within the sub-library, over 3800 colonies would have to be picked and individually screened to pick up a single clone by conventional colony screening. When we sought to further characterise clones of interest that were identified from sub-libraries from panning we were able to rescue one ScFv clone that was present at a level of 0.0025% of the library, or a frequency of ~ 1 in 40,000. Again, at this level, it would be impractical to use conventional colony picking and screening to isolate such a clone. The application of the presented methodology to the rescue of ScFv clones identified during NGPD experiments could be applied to early panning rounds facilitating the rescue of very low frequency antibodies which may not be efficient at propagation and may be lost on further rounds of selection.

We have demonstrated that this phage rescue methodology can also be applied to the recovery of ScFvs with virtually any length of heavy chain CDR3. In the examples described within this paper the smallest CRD3 domain recovered coded for just 4 amino-acids. Overall, all 12 of the targeted ScFv clones were successfully rescued after minimal optimisation of the standard PCR conditions and they were present over a wide range of frequencies in the enriched library pool (0.0025- 53.9%). The data demonstrates the applicability of the method to rescue a wide range of ScFvs present at even very low frequencies.

5.0 Conclusions: The self-ligation of inverse PCR products directed by primers homologous to the heavy chain CDR3 and framework 4 regions of the ScFv sequence has proven to be a sensitive and robust methodology for the rescue of ScFvs, even when small CDR3 domains are present and when ScFv clones are at very low abundance within enriched libraries. The methodology finally allows a significant bottleneck to be overcome in the application of NGPD approaches for ScFv isolation.

Acknowledgments:

AS was funded by the University of Nottingham (Schools of Pharmacy and Veterinary Medicine and Science).

References:

D'Angelo, S., S. Kumar, L. Naranjo, F. Ferrara, C. Kiss, and A.R. Bradbury. **From deep sequencing to actual clones.** 2014. Protein Eng Des Sel. 27(10): p. 301-7.

Gera N, Hussain M, Rao BM. **Protein selection using yeast surface display.** 2013. Methods. 2013 15;60(1):15-26.

Hairul Bahara NH, Tye GJ, Choong YS, Ong EB, Ismail A, Lim TS. **Phage display antibodies for diagnostic applications.** 2013. Biologicals. 41(4):209-16.

Hoogenboom HR, de Bruijn AP, Hufton SE, Hoet RM, Arends JW, Roovers RC. **Antibody phage display technology and its applications.** 1998 Immunotechnology. 4(1):1-20.

Hoskins, R.A., M. Stapleton, R.A. George, C. Yu, K.H. Wan, J.W. Carlson, and S.E. Celniker. **Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP).** 2005. Nucleic Acids Res. 33(21): p. e185.

Löfblom J. **Bacterial display in combinatorial protein engineering.** 2011. Biotechnol J. 6(9):1115-29.

Lloyd, C., D. Lowe, B. Edwards, F. Welsh, T. Dilks, C. Hardman, and T. Vaughan, **Modelling the human immune response: performance of a 1011 human antibody**

repertoire against a broad panel of therapeutically relevant antigens. 2009 Protein Eng Des Sel. 22(3): p. 159-68.

Ravn U, Gueneau F, Baerlocher L, Osteras M, Desmurs M, Malinge P, Magistrelli G, Farinelli L, Kosco-Vilbois MH, Fischer N. **By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection.** 2010. Nucleic Acids Res. 38(21):e193.

Rebollo RI, Sabisz M, Baeriswyl V, Heinis C. **Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides.** 2014. Nucleic Acids Res. Oct 27.

Zhang H, Torkamani A, Jones TM, Ruiz DI, Pons J, Lerner RA. **Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems.** 2011. Proc Natl Acad Sci U S A. 108(33):13456-61.

Table 1: ScFvs targeted for recovery are listed in order of their frequency within their appropriate sub-library.

ScFv	CDR3 Sequence ^a	Frequency	Total clone no.	Percentage Frequency ^b
1	NLWSPFDY	35315	65438	53.9
2	GKRYFDWFPLYHFDH	46154	97641	47.2
3	CLNSSMTS	41651	97641	42.6
4	MAYHTAPK	20525	61071	33.6
5	SKTLNLSKTL	19299	80972	23.8
6	ATSTEGDRGLFK	10446	65438	15.9
7	LSQGLAEQYAG	7691	51594	14.9
8	STRQSTL	3338	61071	5.4
9	INYD	2166	60141	3.6
10	LGRLARG	754	60141	1.2
11	SSNTF	108	60141	0.1
12	GGPAGYSNA	3	115541	0.0025

^a The amino acid sequence of the heavy chain CDR3

^b Frequency is the number of times this clone was sequenced compared to the total number of clones sequenced within that sequencing experiment.

Figure legends

Figure 1: Inverse PCR schematic. Inverse PCR was used to amplify the whole ScFv clone of interest. The DU15-2 ScFv illustrates the primer design strategy for those ScFvs with relatively long heavy chain CDR3 domains, here the reverse primer is complementary to the whole CDR3 coding sequence, and is paired with a forward FR4 primer that abuts the 5' end of the reverse primer. ScFvs with relatively short heavy chain CDR3 (ScFv 9 and 11) used a reverse primer which encompassed all of the unique ScFv heavy chain CDR3 sequence to direct the specific amplification of the correct ScFv, but this also included sequence from the FR4 domain as illustrated. A forward primer is designed that anneals to the heavy chain FR4 domain to the immediate border of the reverse primer. Both primers were 5' phosphorylated (P).

Figure 2: Inverse PCR products of ScFv that had been spiked into a background of 20 ng MRC Griffin-1 library DNA. (A) The rescue of DU15-2 ScFv (containing an 11 amino acid heavy chain CDR3); 1: positive control amplification containing 10 ng DU15-2 phagemid DNA, 2: 0.53 ng DU15-2 DNA spiked into library, 3: 0.053 ng DU15-2 DNA spiked into library, 4: 0.0053 ng DU15-2 DNA spiked into library. (B) The phagemid DNA of clone 11 (with a 5 amino acid heavy chain CDR3) was also diluted into a background of Griffin 1 library (B) lanes 1-4, as for the DU15-2 phagemid. (C) Examples of PCR products from the inverse PCR amplifications used for each ScFv rescue. Amplified products are identified by their clone number (as detailed in table 1). Samples were analysed on a 1% (w/v) agarose gel, M- 1kb DNA markers (NEB), in A and B arrow indicates full length ScFv products that were excised, re-ligated and then transformed.

Figure 1

ScFv VH-VL

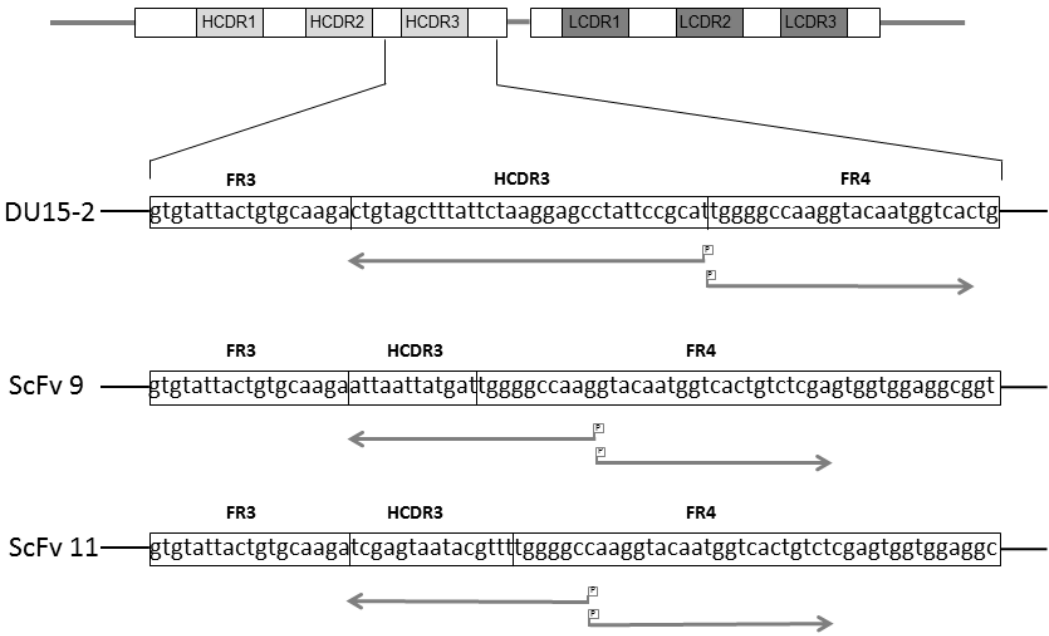
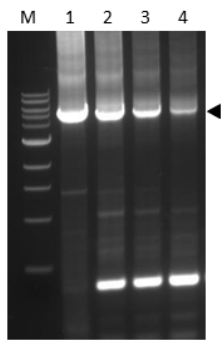
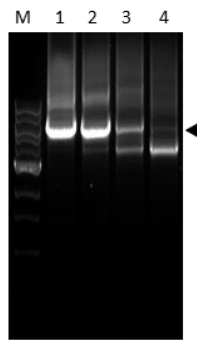


Figure 2

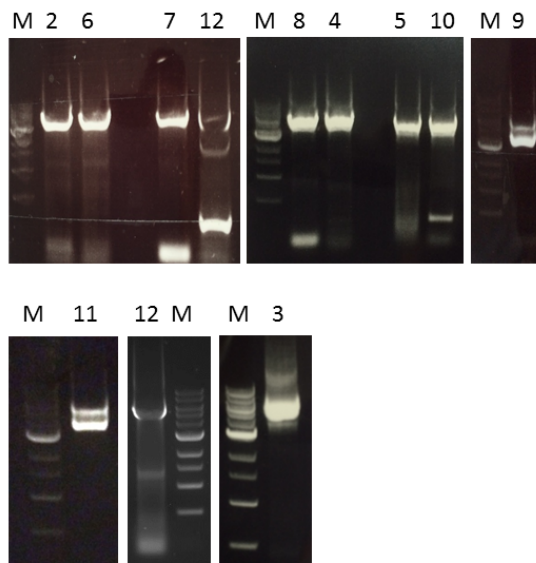
A



B



C



Graphical abstract:

