ssDNA recombineering boosts *in vivo* evolution of nanobodies displayed on bacterial surfaces

by

Yamal Al-ramahi[1], Akos Nyerges[2]✞, Yago Margolles[3], Lidia Cerdán[3], Gyorgyi Ferenc[2], Csaba Pál[2], Luis Ángel Fernández[3] and Víctor de Lorenzo[1]

[1]*Systems and Synthetic Biology Program,* [3]*Department of Microbial Biotechnology, Centro Nacional de Biotecnología (CNB-CSIC), Campus de Cantoblanco, Madrid 28049, Spain;* [2]*Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Szeged H-6726, Hungary*

Running title: Diversification of surface-displayed nanobodies

Keywords: Nanobody, diversification, recombineering, MAGE, TirM

* Correspondence to: Víctor de Lorenzo and Luis Angel Fernández
Centro Nacional de Biotecnología-CSIC
Campus de Cantoblanco, Madrid 28049, Spain
Tel: 34- 91 585 45 36; Fax: 34-91 585 45 06
E-mails: vdlorenzo@cnb.csic.es, lafdez@cnb.csic.es

✞Current address: Department of Genetics, Harvard Medical School, Boston MA 02115, USA

SUMMARY

In vivo evolution of antibodies facilitates emergence of novel target specificities from pre-existing clones. In this work we show how mutagenic ssDNA recombineering of camel-derived nanobodies encoded in a bacterial genome enables clonal hyper-diversification and the rise of new properties. As a proof-of-principle we used a nanobody recognizing the antigen TirM from enterohaemorrhagic *E. coli* (EHEC) and evolved it towards the otherwise not recognized TirM antigen from enteropathogenic *E. coli* (EPEC). To this end, *E. coli* cells displaying on their surface this nanobody fused to the intimin outer membrane anchor domain were subjected to multiple rounds of mutagenic ssDNA recombineering targeted to the CDR1, CDR2 and CDR3 regions of its genomically encoded VHH sequence. Binders to the new antigen (EPEC TirM) were then selected upon immunomagnetic capture of bacteria bearing the corresponding nanobody variants. As a result, several modified nanobodies were identified which maintained recognition of EHEC TirM but acquired the ability to bind the new antigen with high affinity (Kd ~20 nM). The results highlight the power of combining evolutionary properties of bacteria *in vivo* with oligonucleotide synthesis *in vitro* for the sake of focusing diversification to specific segments of a gene (or protein thereof) of interest. Our experimental workflow empowers the evolution of nanobodies displayed on the surface of bacterial cells for a large number of potential applications in medical and industrial biotechnology.

Antibodies are the best instance of functional biomolecules which owe their performance to the ability of their encoding DNA of combining *in vivo* a limited number of hypervariable sequences with a virtually fixed protein scaffold [1, 2]. While the constant moieties of the different types of antibodies adopt a predetermined fixed structure, their target-binding determinants (the so-called complementarity determinant regions or CDRs) are selected out of a very large diversity pool [3]. Such a process occurs upon exposure of the immune system to given antigens through an intricate molecular mechanism of clonal selection, amplification and affinity/specificity maturation [4, 5]. One remarkable feature of this course is that *in vivo* diversification of the CDR sequences for optimally binding a molecular target is virtually limited to 2-4 residues of the primary protein sequence [5]. The antigen-binding variable domains of the naturally occurring heavy chain camel antibodies (also called VHHs or nanobodies when expressed in a recombinant form) represent an extreme case of this state of affairs [6]. While the structures of the immunoglobulin V-domain scaffold of nanobodies are virtually identical in all cases, their CDRs (Fig. 1) are highly variable, both in length and amino acid composition [7].

Given the structural simplicity of camel-derived nanobodies it comes as little surprise that they have emerged as a favourite platform for recreating in the laboratory some of the mechanisms involved in their natural diversification and thus breed large libraries which can later be panned for selecting best binders to specific targets. In particular, the possibility to express VHHs in multiple surface display platforms (i.e. phages, yeast, bacteria) [8] or ribosome display [9] systems enable the selection of variants after diversification. Variability can be evolved either *in vivo* for the whole of the VHH sequence [10] or *in vitro* for introduction of degenerate codons at CDRs using PCR with oligonucleotide primers [11] or DNA synthesis [12, 13]. Yet, while such methods for creating variability in the CDRs push the number of antibody variants close to those of the natural immune system, they still rely on their generation *in vitro* followed by cloning and selection *in vivo.*

On this background we wondered whether we could boost diversification of the CDRs of VHHs by combining the evolutionary power of bacteria *in vivo* with the ease of synthesizing mutagenic oligonucleotides *in vitro*. In this context, we speculated that the two could be merged if combining surface-display of single-copy genome-encoded nanobodies on *E.* coli cells [14] with the cyclic and DNA-segment directed mutagenesis method called DIvERGE [15]. This *in vivo* ssDNA recombineering strategy progressively hyper-mutagenizes specified segments of any gene of interest (e.g. the CDRs of a given VHH) without off-target mutagenesis in adjacent DNA. This can then be followed by capturing by physical means the bacteria displaying the best nanobody variants on their surface [8, 16].

In this article we show how combination of ssDNA recombineering [17] with surface-displayed nanobodies on *E. coli* cells enables rapid expansion of the antigen binding capacity of a nanobody binding the extracellular domain of the translocated intimin receptor (TirM) of enterohaemorrhagic *E. coli* (EHEC) strains towards the otherwise non-recognized but functionally homologous TirM domain of enteropathogenic *E. coli* (EPEC) strains [18, 19]. The resulting, evolved nanobodies binding the new EPEC TirM antigen were then secreted and easily purified from bacterial culture media by adopting the *E. coli* hemolysin protein secretion system [20] integrated in the experimental pipeline. In this way, the work below results in potentially useful nanobodies for diagnostics and passive immunotherapies against EPEC and EHEC infections causing human diarrhoeal diseases [18, 21]. Furthermore, the data below provides an example of how an all-prokaryotic platform can be designed to artificially embody key features of the immune system for a large number of applications—including diagnostics and other medical and environmental needs.

RESULTS AND DISCUSSION

ssDNA recombineering-based diversification of anti-TirM nanobody TD4. The nanobody called TD4 has high affinity (K_D < 5 nM) and specificity for $TirM^{EHEC}$ [19], a protein domain required for intimate attachment of the pathogen to intestinal epithelial cells with actin pedestal formation during infection by enterohaemorrhagic *E. coli* [22, 23]. However, TD4 nanobody did not show any significant binding for the equivalent TirM antigen of EPEC strains, despite the fact that the 24-amino acid epitope recognized by TD4 is 71% identical in both variants [19]. On this basis, we picked TD4 as a good candidate for sampling whether one given nanobody could be evolved with MAGE/DIvERGE technology [15] towards new binding properties. To this end, we adopted *E. coli* strain EcM1*lux*SAtir [14] (Supplementary Table S1) as the primary experimental platform for V_{HH} diversification. This strain (Fig. 1) has a chromosomal insertion of a genetic construct encoding the TD4 nanobody fused to the C-terminus of the outer membrane anchoring domain of intimin [14]. This fusion is expressed through the constitutive P_{N25} promoter [24] and it is integrated replacing the *flu* (Ag43) locus of *E. coli* K-12 genome [25]. Furthermore, EcM1*lux*SAtir strain was deleted of the *fimA-H* operon, encoding type 1 fimbriae [26], and has an insertion of the *luxCDABE* operon of *Photorhabdus luminescens* [27] into the *matBCDEF* operon, encoding the *E. coli* common pilus [28]. These genetic modifications inactivate the expression of cognate endogenous *E. coli* cell surface structures that could interfere with binding of the displayed nanobody to target antigens by [14]. This makes *E. coli* EcM1luxSAtir an optimal strain for displaying the nanobody moiety of the fusion in an active form for interacting with $TirM^{EHEC}$, whether in solution or bound to a surface [14]. As sketched in Fig. 1, this design links the amino acid sequence of the CDRs of the nanobody to short genomic segments of the *E. coli* carrier. This in turn facilitates application of ssDNA recombineering for targeted diversification of such DNA segments, eventually resulting in variability of the primary sequences of the nanobody's CDRs.

The experimental workflow followed to generate large repertoires of surface-displayed TD4 variants is sketched in Fig. 1 and described in the Methods section. Basically, *E. coli* EcM1luxSAtir was first transformed with plasmid pORTMAGE3 (Supplementary Table S2), which upon thermal induction generates the Redβ ssDNA annealing protein (SSAP) for enabling penetration of synthetic oligos within the replication fork, along with a dominant negative allele of *mutL* (MutLE32K) that transiently supresses mismatch repair—thereby increasing occurrence *in vivo* recombineering events [29, 30]. The pORTMAGE3-bearing strain was then cyclically transformed with a cocktail of target-specific mutagenic 90-mer oligonucleotides. The whole of synthetic ssDNA used in the experiment included 3 distinct pools, each targeting one of the segments of the V_{HH} gene that encoded domains CDR1,

CDR2 and CDR3, respectively. Following the criteria of [30] these oligos were designed to introduce 1.5 mutations on average per 100 nucleotides of the targeted sequence. Following ten consecutive rounds of DIvERGE mutagenesis the DNA sequences corresponding to the V_{HH} genes borne by *E. coli* were amplified from the merged genomic DNA extracted from the culture with primers TirDSF and TirDSR (Supplementary Table S3). The resulting DNA products where then submitted to Pacific Biosciences Circular Consensus (CCS) and Illumina amplicon sequencing for appraisal of the mutagenesis procedure, the outcome being summarized in Fig. 2. The results show that after ten consecutive rounds of DIvERGE mutagenesis, a large number of single-nucleotide replacements accumulated throughout the sequence corresponding to the domains CDR2 and CDR3 and in way less pronounced fashion, in CDRI. While the lower diversification of this last segment could be traced to a suboptimal efficiency of DIvERGE at that target site, it could also happen that structural changes in that protein region are counterselected *in vivo.* Results of Fig. 2 show also that mutagenesis did not spread beyond the boundaries of the targeted CDR sequences and that in any case it was not above non-treated samples. In sum, these experiments benchmarked the procedure and warranted the search of new binding specificities in the mutated TD4 nanobody as explained next.

Selection of high-affinity anti-TirM^{EPEC} nanobodies by magnetic cell sorting. In order to enrich the library of V_{HH} variants derived from the TD4 nanobody for those which have acquired the ability to bind the new antigen, the target TirM^{EPEC} domain was purified (Supplementary Fig. S1) and labelled with biotin as described in the Methods section. The *E. coli* cells bearing the mutagenized V_{HH} pool were next incubated with the biotinylated antigen (Fig. 3), washed to remove unbound TirM^{EPEC} and then mixed with paramagnetic microbeads coated with streptavidin. The sample was subsequently passed through a column held in a magnet. Cells coated with bound biotinylated antigen and anti-biotin microbeads were retained whereas those with no significant amount of antigen on their surface were washed out the column and discarded. Finally, bacteria bound to the column were eluted, regrown and used to start another cycle up to 5 times (Fig. 3). As a negative control, the same workflow was applied to equivalent *E. coli* cells displaying an unrelated anti-GFP nanobody. In order to estimate enrichment of TirM^{EPEC}-binding clones, the colony forming units (CFU) stemming from bacteria in the washout eluate and in the antigen-bound fractions were determined at each cycle. The CFU figures accredited that before sorting, virtually all cells of the input library were in the unbound fraction. In contrast, after five such magnetic-activated cell sorting (MACS) cycles, the number of bound cells was >20% (Supplementary Fig. S2). Expectedly, cells displaying the control anti-GFP nanobody were not retained in the column above background levels using biotinylated TirM^{EPEC} antigen. As an additional control, non-mutated *E. coli* cells displaying the parental TD4 nanobody and treated with

the original biotinylated TirM^{EHEC} antigen were efficiently retained in the column. In order to monitor the emergence and evolution of the new binding activity of the progressively enriched library, aliquots of cultures resulting from each MACS cycle were incubated with biotinylated TirM^{EPEC} or biotinylated BSA (as control antigen), washed and stained with streptavidin-phycoerythrin (PE; see Methods). Samples were then inspected in a fluorescence flow cytometer along with controls of the non-diversified strain (*E. coli* EcM1*lux*SAtir) and the input library (diversified but unselected) treated under identical conditions. The results shown in Fig. 4 indicated that from the second cycle of MACS onwards, the selected populations were mainly composed of cells effectively and specifically binding TirM^{EPEC}. Taken together, these data settled the efficacy of the workflow sketched in Fig. 3 for selecting V_{HH} variants with potentially new specificities.

Analysis of mutations in the evolved V_{HH} clones. In order to examine the changes entered during the DIvERGE procedure in the sequence of the anti-TirM^{EHEC} nanobody enabling binding to TirM^{EPEC}, the genomic regions of the V_{HH} were amplified as above in each of the MACS cycles. The distribution of amino acid replacements was then determined through DNA sequencing of the resulting PCR products as above. Fig. 5 summarizes the emergence and distribution of the most frequent changes found in the clone library at each enrichment cycle. For examining individual clones, samples from the last MACS selection cycle were plated to obtain single colonies. Their V_{HH} gene variants were subsequently amplified by PCR and separately sequenced. By matching the resulting discrete sequences with those of the pooled data we could calculate the occurrence and evolution of each variant relative to the overall mutant frequency after each round of MACS sorting (Fig. 5). Interestingly, some early clones disappeared along the enrichment cycles while others stayed at relatively high frequencies compared to the rest. Eventually, the libraries were gradually enriched with just a few clones, the most abundant being those with V_{HH} encoding nanobodies with single mutations in residues H107Y, T108R and D116G in their primary sequence. Amino acid replacements in these three mutants occurred in their CDR3 segment, which is coherent with earlier observations that largely trace specificity changes to modifications of CDR3 in nanobodies [7]. The 3 mutants were then subsequently picked for further analyses as described below.

Binding of TirM antigens by the evolved nanobodies displayed on the bacterial surface. *E. coli* strains displaying on their surface nanobodies TD4^{H107Y}, TD4^{T108R} and TD4^{D116G} (see above) were tested for their ability to bind the original antigen TirM^{EHEC} in comparison with the one used for selection (TirM^{EPEC}). For this, cells from cultures expressing the intimin-VHH fusions were treated with either biotinylated TirM^{EHEC} or biotinylated TirM^{EPEC} followed by staining with streptavidin-allophycocyanin (APC; see Methods). Bacteria were then subjected to

fluorescence flow cytometry analysis and the results are shown in Fig. 6. Inspection of the histograms revealed that while control cells displaying the non-evolved TD4 nanobody expectedly had no noticeable interaction with $TirM^{EPEC}$, the counterparts expressing $TD4^{H107Y}$, $TD4^{T108R}$ and $TD4^{D116G}$ exhibited a considerable binding to the new antigen (Fig. 6A and 6B). In addition, the evolved nanobodies maintained strong binding signals for the original target $TirM^{EHEC}$, similar to the parental nanobody (Fig. 6C and 6D). These results suggested that the evolved nanobodies had expanded their binding capacity towards the novel antigen—rather than switching specificities completely. This outcome was welcome, as eventually we pursued nanobodies able to bind simultaneously to different types of TirM variants.

To gain some quantitative insight into the expansion in antigen specificity, we next measured the apparent affinities of the TD4 variants toward EHEC and EPEC TirM antigens when displayed on the surface of *E. coli.* For this, bacteria were incubated with different concentrations of the biotinylated antigens ranging from 0 to 250 nM in the case of $TirM^{EPEC}$ and 0-50 nM for $TirM^{EHEC}$ (Fig. 6B and 6D). The median fluorescence intensity values were fitted by a non-linear regression curve based on the model $Y = Bmax \cdot X \cdot (Kd+X)^{-1}$, where Y is the fluorescence intensity, Bmax is the fluorescence signal extrapolated to very high concentrations of antigen, X is the concentration of the antigen and Kd is the equilibrium dissociation constant (i.e., the antigen concentration that binds to half of the binding sites at equilibrium). Based on best fit estimations, the apparent Kd for new antigen $TirM^{EPEC}$ were between ~17 nM and ~33 nM, ranking $TD4^{D116G}$ > $TD4^{T108R}$ > $TD4^{H107Y}$ (Table I). The Kd value for the parental TD4 was ignored as there was no significant binding to $TirM^{EPEC}$ (Table I). Interestingly, parental and evolved nanobodies showed similar apparent Kds for the former antigen $TirM^{EHEC}$, which were in all cases in the subnanomolar range (Table I). Therefore the evolved TD4 variants still maintain an apparent higher affinity for the original antigen $TirM^{EHEC}$ than for the new antigen $TirM^{EPEC}$. These data suggested that the effect of the mutations was not a simple improvement of a low affinity interaction with the new antigen present in the non-evolved TD4, but to create a genuinely new interaction with $TirM^{EPEC}$ while maintaining strong interactions with the former antigen.

Binding affinities of the evolved monovalent nanobodies against TirM antigens. Although the experiments above exposed important features of the nanobody variants $TD4^{T108R}$, $TD4^{H107Y}$ and $TD4^{D116G}$, the data on their apparent affinities might be influenced by their expression in multimeric form on the surface of bacteria [8]. To clarify this, the mutant proteins were purified as monovalent polypeptides following their secretion by *E. coli.* To this end, the corresponding V_{HH} sequences were cloned in fusion with the C-terminal secretion signal of

hemolysin A (HlyA) and the resulting plasmid constructs were then placed in a dedicated *E. coli* strain that expresses HlyB and HlyD components of the hemolysin protein secretion system (Methods section; [20]. The strains bearing these fusions were grown in LB, induced with IPTG, and the secreted nanobodies purified from the culture media (Supplementary Fig. S3). Controls included one strain secreting the non-evolved parental TD4 and another strain producing a non-related nanobody (Vamy [20]. Affinities were determined by coating ELISA plates with purified His-tagged versions of TirM^{EHEC} and TirM^{EPEC} and the values estimated as shown in Fig. 7. In this case, results clearly indicated that the former antigen TirM^{EHEC} is best recognized by the non-evolved TD4 (Kd = 5.5 nM) in a fashion about one order of magnitude better than TD4^{H107Y} (Kd = 47.74 nM) and >20-fold in respect to TD4^{T108R} and TD4^{D116G}. The new antigen TirM^{EPEC} was not recognized by the parental TD4 nanobody at any concentration tested but, the evolved counterparts manifest their newly gained interaction to the new target with Kd values of 20.8 nM (TD4^{H107Y}), 176 nM (TD4^{T108R}) and 397 nM (TD4^{D116G}), respectively. Divergence of the apparent Kd values determined with the purified nanobodies with those estimated with *E. coli* bacteria are likely reflecting avidity effects due to the distinct physics of the nanobody-antigen interaction in monovalent (purified) *versus* multivalent (cell-displayed) formats [31, 32]. Taken together, these results demonstrate the ability of the hereby described workflow to expand the molecular target of a given nanobody towards a different one.

Structural predictions of target expansion by evolved nanobodies. In order to gain some insight into the changes that endowed TD4^{H107Y}, TD4^{T108R} and TD4^{D116G} with new properties, we threaded the primary amino acid sequence of all the TirM variants in the well conserved tridimensional structure of known nanobodies determined by X-ray crystallography using bioinformatic tools (Methods section). The resulting shape of the original TD4 model was then overlapped pair-wise with each of the other 3 evolved variants, with the outcome shown in Fig. 8. The root mean square deviations (RMSD) values between each pair of structures (indicative of divergence from each other) were very low, indicating structures to be virtually identical excepting for the very sites where mutations occurred. Inspection of the amino acid changes revealed that in position 107 a basic histidine is replaced with a bulkier, uncharged and protruding polar tyrosine. In the nearby 108 position an uncharged and polar threonine residue was exchanged by a basic and also bulkier/protruding arginine. This exposed the importance of the site location around 107-108 for recognition of the new antigen, perhaps by direct contacts with specific residues of the TirM^{EPEC} structure that were not amenable for binding by the non-evolved TD4 nanobody [19]. In contrast, an acidic aspartate at position 116 was replaced with a neutral and smaller glycine, suggesting that in this case the change likely eliminated an inhibitory repulsion between a charged amino acid and an incompatible site of TirM^{EPEC}. These replacements are also predicted to change the orientation of the new amino

acid in respect to the original one, although due to the size of the side chains, this was more pronounced in the case of H107Y and T108R. In any case, replacement of a single amino acid in TD4 was sufficient gain of a new binding specificity while their former capacity to bind $TirM^{EHEC}$ was not abolished. This was noticed in the three purified nanobodies, but the tradeoffs were different in each case. $TD4^{H107Y}$ was the one with higher $TirM^{EPEC}$-binding affinity and the mildest loss affinity for $TirM^{EHEC}$. In contrast, acquiring a new binding towards $TirM^{EPEC}$ for $TD4^{T108R}$ and $TD4^{D116G}$ was at a higher cost for their interaction with the former antigen.

Conclusion. The results above showcase the power of merging surface display of nanobodies with genomic-site specific diversification elicited by DIvERGE technology in combination with MACS-based cyclic selection. The CDRs of the encoded nanobody can be easily targeted by ssDNA mutagenic recombineering [15] of the V_{HH} gene integrated as single copy in the bacterial chromosome and fused to intimin outer membrane anchor domain for bacterial display, and the resulting library directly screened and characterized for binding to a new antigen target of choice. However, the apparent affinities for the evolved nanobodies *in vivo* on the bacterial surface do not exactly match with their binding affinities observed as purified monovalent proteins, likely reflecting complex avidity effects of the multimeric format of the nanobody displayed on the bacterial surface. Nonetheless, the hereby proposed experimental pipeline delivers a suite of nanobodies endowed with a range of affinities for the target that cover different needs. In the specific instance addressed in this work we pursued to expand the therapeutic potential of a specific anti-$TirM^{EHEC}$ nanobody that inhibits the attachment of EHEC bacteria to human intestinal tissue towards EPEC strains, which are a major cause of acute and chronic diarrohea in infants [21]. But one can envision a number of scenarios which make the workflow of Fig. 2 and Fig. 3 exceptionally worthy also e.g. for quickly developing existing V_{HH}s against viral proteins towards emerging variants for diagnostic or therapeutic purposes [33, 34, 35]. This can be specially useful when thereby selected nanobodies are subsequently bulk-produced in bacteria with the hemolysin export system—also adopted in our experimental roadmap above. While alternative microbial platforms for *in vivo* generation of nanobody repertoires do exist to the same end [33, 10, 12, 36, 37] we believe that the presented pipeline is simple and contributes to ongoing efforts towards point-of-care technologies for diagnostics and other medical and environmental needs [38, 39, 40].

METHODS

Materials, oligonucleotide synthesis and general procedures. *E. coli* strains used in this work can be found in Supplementary Table S1. Plasmids and other genetic constructs are listed in Supplementary Table S2.

Recombinant DNA manipulations, media preparation and culture conditions followed standard protocols [41]. Purification of $TirM^{EHEC}$ and $TirM^{EPEC}$ was done as described in [19, 20]. Protein concentrations were estimated using Protein assay reagent (BioRad) or BCA protein assay (Thermo Fisher Scientific). Whenever required, proteins were analyzed in denaturing polyacrylamide gels.

DIvERGE oligonucleotides. ssDNAs for recombineering and library diversification were designed to minimize secondary structure, to target the lagging strand of the DNA replication fork and to incorporate 1-2 mutations per CDR1, CDR2 and CDR3 segments of the V_{HH} sequence. For this, 3 randomized regions of the mutagenic oligonucleotides (Supplementary Table S3) were synthesized with 0.5%-0.5%-0.5% soft-randomization with the three other possible nucleobases other than the wild-type. This creates an average chance of having a SNP at a given, randomized position of 1.5% (i.e. 1.5 mutation on average per 100 target bases), which translates into a mutation rate of 1-2 SNP per gene in the resulting library [29, 30]. Each mutagenic oligo pool was synthesized on an ABI3900 DNA synthesizer (Applied Biosystems) following a modified protocol based on phosphoramidite chemistry. The synthesis cycles were applied on a controlled pore glass as solid support and were carried out as follows. First, deprotection was done with trichloroacetic acid 3% (w/v) in dichloromethane. Second, incoming phosphoramidite was dissolved in $55 \cdot 10^{-3}$ M anhydrous acetonitrile, then mixed with the other three amiditesin and subsequently coupled by activation with 5-ethylthio-1H-tetrazole. Third, capping was done with acetic anhydride 10% (v/v) in anhydrous tetrahydrofuran and N-methyl-imidazole 16% (v/v) and pyridine 10% (v/v) in an anhydrous tetrahydrofuran solution. Finally, the oxidation was done with a solution of 5 g of iodine per litre of a mixture of pyridine:water:tetrahydrofuran in the ratios 0.5:2:97.5 respectively. Cycles were repeated until the 90^{th} nucleotide and then, the DNA strands were cleaved from the solid support using concentrated ammonia. Crude oligos were purified by high-performance liquid chromatography (HPLC) on a Shimadzu ADVP-10 HPLC system. HPLC fractions were concentrated and dimethoxytrityl (5´-DMTr) protecting groups were removed using a PolyPak column (Glen Research) following the manufacturer's protocol. Oligos were eluted, lyophilized and suspended in TE buffer (1X) pH 8.0 from Integrated DNA Technologies. Sequences of all all DIvERGE oligos are listed in Supplementary Table S3.

Genomic DNA sequence diversification with MAGE/DIvERGE. A fresh single colony of *E.coli* EcM1 bearing pORTMAGE3 (Addgene Plasmid ID #72678) taken from $LB_{a,Km}$ was inoculated into 2 mL LB_{Km} and incubated overnight at 30 ºC, 170 rpm. This cell culture was used to prepare a 1:100 dilution in 10 mL LB_{Km} that was then incubated at 30 ºC and 170 rpm until reaching 0.4 OD_{600nm}. Next, this was incubated 15 min in a water bath at 42

ºC with shaking at 250 rpm and then the flask containing the cells was kept 10 min in ice. The grown cells were harvested by centrifugation at 4000 rpm for 8 min at 4 ºC and supernatant discarding. Sedimented cells were washed by resuspension in 10 mL ice-cold water, centrifugation at 4000 rpm for 8 min at 4 ºC and supernatant discarding. The sediment of cells from the last wash was suspended in 160 µL of ice cold dH_2O and kept on ice. A mix of 3 µL with primers pools CDR1, CDR2 and CDR3 in 1:1:1 ratio were incorporated to 40 µL of competent cells in a 1.5 mL tube and the resulting mix was kept on ice for 5 min. These cells were electroporated, then suspended in 6 mL of TB and subsequently kept at 30 ºC with shaking at 170 rpm. After 60 min, 10 mL LB supplemented with kanamycin were added and the mix was kept at 30 ºC in a shaker until 0.4 OD_{600nm}. Then, 10 mL of this culture were taken to prepare electrocompetent cells in order to start a new DIvERGE cycle. The other 5 mL were grown overnight, then OD_{600nm} was measured and 1.5 mL were used for genomic DNA extraction and another 1.5 mL to prepare a stock tube in glycerol 20% to store at -80 ºC.

Biotinylation. Fractions of the antigens EHEC TirM or EPEC TirM were diluted in 1.5 mL PBS to 1 $mg \cdot ml^{-1}$. The NHS-Biotin (Sigma-Aldrich B2643) was dissolved in DMSO and immediately added to these protein solutions in a biotin:protein molar ratio of 20:1 and the mix was incubated 30 min at RT with gentle agitation. Then 1 M Tris-HCl, pH7.5 was added for a final concentration of 50 mM and the mixture was incubated 1 h on ice. The sample was concentrated to a final volume of 500 µL in a 3kDa Centricon (Merck) by centrifugation at 5000 g. Finally, the efficiency of biotinylation was checked in a Western blot revealed with POD-anti-biotin (horseradish peroxidase conjugated with Streptavidin, Roche).

Magnetic-activated cell sorting (MACS). An amount of 10^8 bacteria were washed by centrifugation (5 min, 4000g, 4ºC) and supernatant discarding and then were suspended in 2 mL of BSA 0.5% in PBS 1X. The same process was repeated again from the beginning. Then, the cells were washed again and suspended in 200 µL of a solution of BSA 0.5% in PBS 1X. The appropriate amount of the biotinylated antigen $TirM^{EPEC}$ or $TirM^{EHEC}$ was incorporated to reach the final concentration of choice (from 50 nM to 250 nM), mixed gently and incubated 1h at 4 ºC. These cells were washed three times, suspended in 80 µL of a solution of BSA 0.5% in PBS 1X, mixed gently with 20 µL of Anti-Biotin Microbeads and incubated 30 min at 4ºC. Then, the cells were washed three times and suspended in 500 µL of BSA 0.5% in PBS 1X. A column was placed on the magnet and rinsed with 1.5 mL of PBS+BSA 0.5%. Then, the cell suspension was applied onto the column and the flow-through collected. The column was washed three times with 500 µL of PBS+BSA 0.5%, and the three flow through were collected together in the same tube. The column was removed from the separator and placed on a suitable collection tube.

Next, 2 mL of PBS+BSA 0.5% were added onto the column. The magnetically labelled cells were immediately flushed out by firmly pushing the plunger into the column.

***In vivo* determination of nanobody-antigen affinity by flow cytometry.** The strains EcM1T108R, EcM1H107Y and EcM1D116G were inoculated in separate flasks with 25 mL of $LB_{,km}$ and EcM1LuxSATir in a flask with 25 mL of LB. These four flasks were subsequently incubated overnight at 30 ºC and 170 rpm. From each culture, a volume corresponding to 1 OD_{600nm} was centrifuged at 4000 g for 2 min at 4 ºC followed by 20000 g for 2 min at 4 ºC, the supernatant was discarded and the sediment was washed three times with 1 ml of pre-chilled and filter sterilized PBS and centrifugation at 4 ºC and 4000 g for 2 min followed by 20000 g for 2 min at 4 ºC. Then, the cells were harvested by centrifugation (at 4 ºC at 4000 g for 2 min followed by 20000 g for 2 min at 4 ºC) and the sediment suspended in 500 µL of PBS. Next, 9 fractions with 25 µL of this cell suspension were centrifuged at 20000 g and 4 ºC for 2 min, the supernatants discarded and the cells incubated 90 min at 4 ºC with a fixed amount of biotinylated antigen and increasing volumes of PBS from 0.25 to 1.6 mL to attain a final concentration range between 0 nM to 250 nM. After incubation, cells were centrifuged at 4000 g and 4 ºC for 2 min followed by 20000 g at 4 ºC for 2 min, then washed twice with 1 mL of filtered and sterilized PBS at 4 ºC and labeled with Streptavidin-allophycocyanin. After a final washing step with PBS, the median fluorescence intensity of APC was quantified in a cytometer (Gallios, Beckman Coulter). Data of Median Fluorescence Intensity relative to maximum Median Fluorescence Intensity obtained from the cytometer were plotted against the concentration of biotinylated antigen to estimate the dissociation constant (K_D). Curve was fitted according to non-linear least squares regression method and one site-specific binding saturation kinetics model using the data analysis tool in Prism software (GraphPad).

Isolation and sequencing of genomic DNA. Bacterial DNA was isolated from fresh overnight cultures of the libraries rMAGE1, rMAGE2, rMAGE3, rMAGE4, rMAGE5 and rMAGE5control with the kit UltraClean® Microbail DNA Isolation Kit (Mo Bio, QIAGEN). Fractions of the isolated genomes were used as templates for PCR with the primers tirDSF and tirDSR and the obtained amplicons were purified and sent for sequencing with Illumina MiSeq set-up. Isolated genomic DNA from the libraries rMAGE5, rMAGE10, MACS1EHEC, MACS1CR, MACS1EPEC, MACS2EPEC, MACS3EPEC, MACS4EPEC and MACS5EPEC was used as template for a PCR with the primers tirDSF and tirDSR and the resulting amplicons were cleaned NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) and sent for PacBio sequencing. For screening and for sequence confirmation, colony PCR products of around 1060 bp were obtained with the oligonucleotides CheckSATir_F and CheckSATir_Rev using

Q5DNApolymerase (New England Biolabs) and the program as follows: 30 s at 98 ºC, 30 cycles (15 s at 98 ºC, 45 s at 70 ºC and 90 s at 72 ºC) and for final extension 120 s at 72 ºC. PCR products were cleaned with the kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) and sent to Macrogen for sequencing with the oligonucleotide CDRI_Fwd. Results were used to estimate the frequency of mutations in the libraries.

Plasmids for bulk production and purification of nanobodies. Changes found in the DNA of the evolved VHH variants were first recreated in the frame of the wild type TD4 sequence. To this end pCANTAB6_VhhTD4 plasmid (Supplementary Fig. S2) was used as DNA template for PCR using Q5 DNA polymerase (NEB) in three different reactions each containing the oligos 1 and 2 (for H107Y), 3 and 4 (for T108R) or 5 and 6 (for D116G). Conditions were 30s at 98 ºC, (15s at 98 ºC, 30s at 65 ºC, 3 min at 72 ºC) x 18 cycles and 5 min at 72 ºC. Mixes were subsequently digested 1h at 37 ºC with DpnI (NEB) and used for electroporation in *E. coli* CC118. Afterwards, each of the modified plasmids was isolated and confirmed by sequencing using the oligos 7 and 8 (Supplementary Table S3). The resulting plasmids were then named pCANTAB6_H107Y, pCANTAB6_T108R and pCANTAB6_D116G. In parallel, plasmid pEHLYA5_Vamy was isolated, digested with SfiI and NotI and the resulting DNA fragments were separated by electrophoresis in agarose gel. Then, the vector pEHLYA5 in the band corresponding to 3612 bp was cleaned with Nucleospin Gel and PCR clean-up kit (Macherey-Nagel). Plasmids pCANTAB6_H107Y, pCANTAB6_T108R and pCANTAB6_D116G were isolated, digested with SfiI and NotI and the DNA fragments separated by electrophoresis in agarose gel. Each DNA band of 400 bp containing a VHH gene variant was purified and ligated, to the SfiI/NotI-digested pEHLYA5 and the ligation mixed transformed in *E. coli* DH5α. Plasmids from clones growing in LB, ampicillin (100 μg/ml) and glucose (20%) were sequenced with the oligonucleotides 9 and 10 and, once verified were named pEHLYA_H107Y, pEHLYA_T108R and pEHLYA_D116G respectively.

Expression and purification of nanobodies. Competent *E. coli* HB2151 cells were prepared and endowed, by heat shock, with the plasmids for the purification of the Nbs variants from culture supernatants using the hemolysin secretion system [20]. Cells were recovered and then plated on LB agar with glucose and the proper antibiotics. The resulting strains were E. coli HB2151 (pVdL9.3_Vamy), E. coli HB2151 (pVdL9.3_TD4), E. coli HB2151 (pEHLYA_H107Y + pVdL9.3), E. coli HB2151 (pEHLYA_T108R + pVdL9.3) and E. coli HB2151 (pEHLYA_D116G + pVdL9.3) for production of the Nbs Vamy, TD4, H107Y, T108R and D116G respectively fused to the C-terminal end of hemolysin, one isolated colony of each strain was grown in aerobic conditions overnight at 30 ºC in 25 mL of LB with glucose 20% and Ampicillin or Ampicillin and Chloramphenicol accordingly.

The resulting culture was used to prepare 200 mL of at 0.05 OD_{600nm} in LB with antibiotics that was grown until 0.4 OD_{600nm}. Following, IPTG to 10^{-3} M final concentration is added and the culture further incubated for 6 hours at 30ºC and 100 rpm. Next, the culture was centrifuged 4000 g for 15 min, the supernatant was passed to a new tube and centrifuged 4000 g for 15 min, the supernatant was passed again to a new tube and centrifuged 4000 g for 15 min, supernatant was collected in a new tube and supplemented with PMSF to 10^{-3} M final concentration. Each volume was supplemented with PBS 10x to a final PBS 1X concentration. Samples containing the thereby secreted $[His]_6$-tagged VHH-HlyA fusions were next passed through Talon CellThru resin (Clontech). For this, 2 ml of this metalloaffinity matrix, was placed in a 15 ml tube and it was subsequently washed by centrifugation 2 min at 700g, supernatant discarding and suspension in 10 mL of PBS 1X. The resin was washed again, then it was centrifuged 2 min at 700g, the supernatant was discarded and the resin was added to the suspension containing the secreted Nbs in 50 mL tubes. The mix was incubated in a spinning wheel and, after 30 min, it was added gradually to a 0.7 cm x 5 cm chromatography column (Bio-Rad). Once packed, this resin was washed in the column by passing gradually up to 30 mL of $5 \cdot 10^{-3}$ M imidazole in PBS (1X) through the packed resin. Then, the protein was eluted in 1 mL fractions of 0.15 M imidazole in PBS (1X). Fractions were stored at 4 ºC until needed.

Enzyme-linked immuno assay. ELISA conditions were based on a modified version of a previously described protocol [19, 20]. Briefly, Maxisorp 96-well immunoplates (nunc) were coated by adding 50 µL/well of a solution with the antigen EPEC TirM or EHEC TirM at 10 $\mu g \cdot mL^{-1}$ in PBS (1X) and incubating overnight at 4 °C. Then, blocking was done by discarding the liquid from the plate and adding 200 µL/well of PBS (1X) with skimmed milk (3%) and incubation 60 min at 23 ºC. Next, the liquid in the plate was discarded and the plate washed three times. Each wash consisted in adding 200 µL/well of a solution made of Tween 20 (0.05 %) in PBS (1X), incubation for 5 min at 23 ºC and liquid discarding. Then, 50 µL with a Nb solved in a solution of skimmed milk (3%) in PBS (1X) was added per well in the range of 0-100 nM in the case of TD4 and 0 to 400 nM in the case of H107Y, T108R or H116G. Following incubation was 90 min at 23 ºC. Next, the liquid was discarded and the plate washed three times, each consisting in adding 200 µL/well of Tween 20 (0.05 %) in PBS (1X), incubation for 5 min at 23 ºC and liquid discarding. Mouse antibody anti-Etag diluted 1:2000 in a solution of skimmed milk (3%) in PBS (1X) was added and the plate was incubated 60 min at 23 ºC. Then, the plate was washed three times. Following, the anti-mouse-POD antibody was diluted 1:1000 in the solution with skimmed milk (3%) in PBS (1X), the mix was added to the wells and the filled plate was incubated 60 min at 23 ºC. Then, the liquid was discarded and the plate was washed three times. Finally, results were developed as follows. One o-phenylenediamine tablet (OPD, Sigma) and 10 µL hydrogen peroxide (30 % w/v) were solved in 20 mL phosphate-citrate buffer (pH 5) and 80 µL/well

were added from the mix. The plate was incubated 20 min at 23 ºC, protected from light and then, the reaction was stopped with 20 µL HCl (3 M) /well. Finally, OD_{490nm} was determined using a microplate reader (iMark ELISA plate reader, Bio-Rad).

Structural alignment of nanobodies. The amino acid sequences of the Nbs (table IV) were introduced as input in FASTA format in i-TASSER online server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) to obtain the output which included each structural model file. The models of the structural alignments with the parental Nb TD4 were obtained with the function CEalign in PyMOL software. Alignment images were edited in PyMOL and exported as files with the extension png. Next, image cuts to show specific details were done with the software Fiji (https://imagej.net/Fiji).

Data availability. Data supporting the findings of this work are available within the paper and associated Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information. The datasets generated and analyzed during the current study are available from the corresponding author upon request.

REFERENCES

1. Schroeder HW, Jr., Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125, S41-52 (2010).

2. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies* (Basel) 8, 55 (2019).

3. Wong WK, Leem J, Deane CM. Comparative Analysis of the CDR Loops of Antigen Receptors. *Front Immunol* 10, 2454 (2019).

4. Melchers F. Checkpoints that control B cell development. *J Clin Invest* 125, 2203-2210 (2015).

5. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. **Blood** 112, 1570-1580 (2008).

6. Ingram JR, Schmidt FI, Ploegh HL. Exploiting Nanobodies' Singular Traits. *Ann Rev Immunol* 36, 695-715 (2018).

7. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. *Proteins* 86, 697-706 (2018).

8. Salema V, Fernández L. *Escherichia coli* surface display for the selection of nanobodies. *Microb Biotechnol* 10, 1468-1484 (2017).

9. Zimmermann I, et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* 7, e34317 (2018).

10. Wellner A, et al. Rapid generation of potent antibodies by autonomous hypermutation in yeast. *bioRxiv,* 2020.2011.2011.378778 (2020).

11. Tiller KE, et al. Facile Affinity Maturation of Antibody Variable Domains Using Natural Diversity Mutagenesis. *Front Immunol* 8, 986 (2017).

12. Moutel S, et al. NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. *eLife* 5, e16228 (2016)

13. Muyldermans S. A guide to: generation and design of nanobodies. *FEBS J* https://doi.org/10.1111/febs.15515 (2020).

14. Piñero-Lambea C, Bodelón G, Fernández-Periáñez R, Cuesta AM, Álvarez-Vallina L, Fernández L. Programming controlled adhesion of *E. coli* to target surfaces, cells, and tumors with synthetic adhesins. *ACS Synth Biol* 4, 463-473 (2015).

15. Nyerges Á, et al. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc Natl Acad Sci USA* 115, E5726-e5735 (2018).

16. Salema V, et al. Selection of single domain antibodies from immune libraries displayed on the surface of *E. coli* cells with two β-domains of opposite topologies. *PLoS One* 8, e75126 (2013).

17. Wannier TM, et al. Recombineering and MAGE. Nature Reviews Methods Primers 1, 7 (2021).

18. Wong AR, et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli:* even more subversive elements. *Mol Microbiol* 80, 1420-1438 (2011).

19. Ruano-Gallego D, Yara DA, Di Ianni L, Frankel G, Schüller S, Fernández L. A nanobody targeting the translocated intimin receptor inhibits the attachment of enterohemorrhagic *E. coli* to human colonic mucosa. *PLoS Pathog* 15, e1008031 (2019).

20. Ruano-Gallego D, Fraile S, Gutierrez C, Fernández L. Screening and purification of nanobodies from *E. coli* culture supernatants using the hemolysin secretion system. *Microb Cell Fact* 18, 47 (2019).

21. Chen HD, Frankel G. Enteropathogenic *Escherichia coli:* unravelling pathogenesis. *FEMS Microbiol Rev* 29, 83-98 (2005).

22. Frankel G, Phillips AD. Attaching effacing *Escherichia coli* and paradigms of Tir-triggered actin polymerization: getting off the pedestal. *Cell Microbiol* 10, 549-556 (2008).

23. Clements A, Young JC, Constantinou N, Frankel G. Infection strategies of enteric pathogenic *Escherichia coli. Gut Microbes* 3, 71-87 (2012).

24. Deuschle U, Kammerer W, Gentz R, Bujard H. Promoters of *Escherichia coli:* a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J* 5, 2987-2994 (1986).

25. van der Woude MW, Henderson IR. Regulation and function of Ag43. *Annu Rev Microbiol* 62, 153-169 (2008).

26. Dunnick JK, Eustis SL, Huff JE, Haseman JK. Two-year toxicity and carcinogenicity studies of ampicillin trihydrate and penicillin VK in rodents. *Fundam Appl Toxicol* 12, 252-257 (1989).

27. Winson MK, et al. Engineering the *luxCDABE* genes from *Photorhabdus luminescens* to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs. *FEMS Microbiol Lett* 163, 193-202 (1998).

28. Rendón MA, et al. Commensal and pathogenic Escherichia coli use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci USA* 104, 10637-10642 (2007).

29. Nyerges A, et al. Conditional DNA repair mutants enable highly precise genome engineering. *Nucl Acids Res* 42, (2014) 42, e62

30. Nyerges A et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc Natl Acad Sci USA* 113, 2502-2507 (2016).

31. Kitov PI, Bundle DR. On the nature of the multivalency effect: a thermodynamic model. *J Am Chem Soc* 125, 16271-16284 (2003).

32. Vorup-Jensen T. On the roles of polyvalent binding in immune recognition: perspectives in the nanoscience of immunology and the immune response to nanomedicines. *Adv Drug Deliv Rev* 64, 1759-1781 (2012).

33. Wrapp D, et al. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell* 181, 1004-1015.e1015 (2020).

34. Xiang Y, et al. Versatile and multivalent nanobodies efficiently neutralize SARS-CoV-2. Science 370, 1479-1484 (2020).

35. De Vlieger D, Ballegeer M, Rossey I, Schepens B, Saelens X. Single-Domain Antibodies and Their Formatting to Combat Viral Infections. *Antibodies (Basel)* 8, 1.

36. Zimmermann I, et al. Generation of synthetic nanobodies against delicate proteins. *Nat Protoc* 15, 1707-1741 (2020).

37. Uchański T, et al. An improved yeast surface display platform for the screening of nanobody immune libraries. *Sci Rep* 9, 382 (2019).

38. De Meyer T, Muyldermans S, Depicker A. Nanobody-based products as research and diagnostic tools. *Trends Biotechnol* 32, 263-270 (2014).

39. Chakravarty R, Goel S, Cai W. Nanobody: The *magic bullet* for molecular imaging? *Theranostics* 4, 386-398 (2014).

40. Leow CH, Fischer K, Leow CY, Cheng Q, Chuah C, McCarthy J. Single Domain Antibodies as New Biomarker Detectors. *Diagnostics* (Basel) 7, (2017).

41. Sambrook J, Russell DW. *The condensed protocols from molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press (2006).

Acknowledgements. This work was funded by the SETH (RTI2018-095584-B-C42) (MINECO/FEDER) and SyCoLiM (ERA-COBIOTECH 2018 - PCI2019-111859-2) Projects of the Spanish Ministry of Science and Innovation, the MADONNA (H2020-FET-OPEN-RIA-2017-1-766975), BioRoboost (H2020-NMBP-BIO-CSA-2018-820699), SynBio4Flav (H2020-NMBP-TR-IND/H2020-NMBP-BIO-2018-814650) and MIX-UP (MIX-UP H2020-BIO-CN-2019-870294) Contracts of the European Union and the InGEMICS-CM (S2017/BMD-3691) Project of the Comunidad de Madrid - European Structural and Investment Funds (FSE, FECER). LAF is supported by Grant BIO2017-89081-R from *Agencia Española de Investigación* (AEI/MICIU/FEDER, EU).

Author Contributions. VdL, LAF and AN designed the research. YAR and YM performed the research. LC provided technical assistance. CP supported the work throughout. AN and LAF analyzed data. YAR and VdL wrote the original draft of this paper and prepare the Figures. All authors read and approved the final manuscript.

Competing Interests. A.N. and C.P. are inventors on a patent related to directed evolution with random genomic mutations (DIvERGE) (US10669537B2 and WO2018108987: Mutagenizing Intracellular Nucleic Acids).

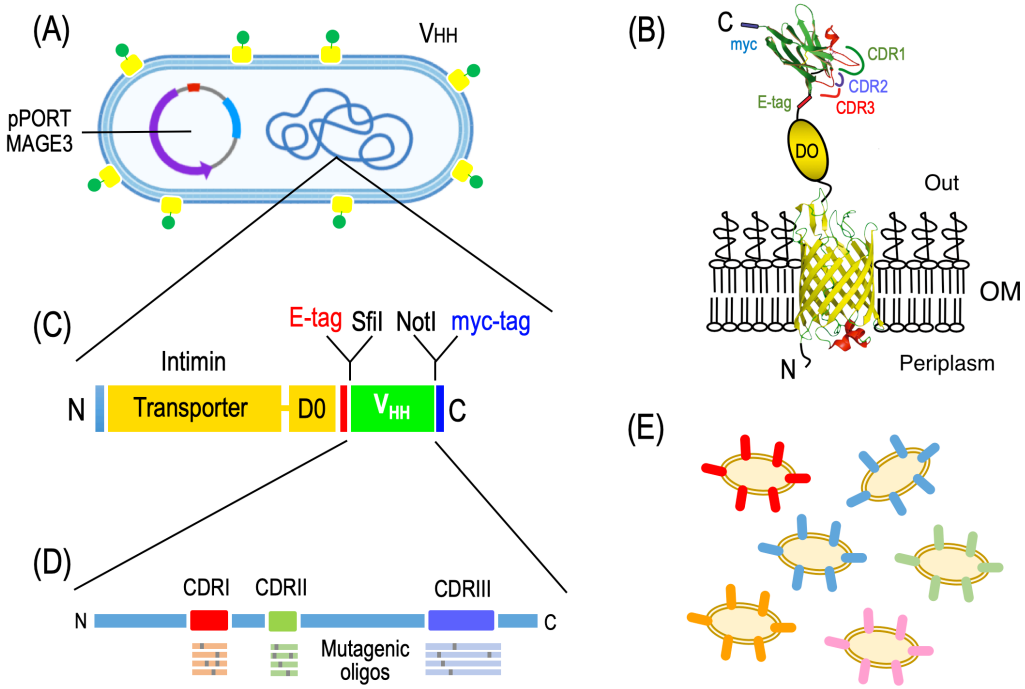
Table 1. Estimated Kd values of evolved nanobodies (x10^{-9} M) by flow cytometry of *E. coli* bacteria.

	New antigen (EPEC TirM)		Former antigen (EHEC TirM)	
Nanobody mutant	Best-fit values	95% Conf Intervals	Best-fit values	95% Conf Intervals
Nb TD4	N.S.	N.S.	0.41	0.26 to 0.55
Nb H107Y	33.86	11.99 to 55.73	0.57	0.36 to 0.78
Nb T108R	27.57	17.64 to 37.50	0.36	0.29 to 0.43
Nb D116G	17.86	11.71 to 24.01	0.58	0.45 to 0.71

Kd values and confidence intervals were estimated using the function “One site - Specific binding” from the software GraphPad (Prism).

FIGURES

Figure 1. A new platform for the generation and screening of VHH variants with new antigen binding specificities.

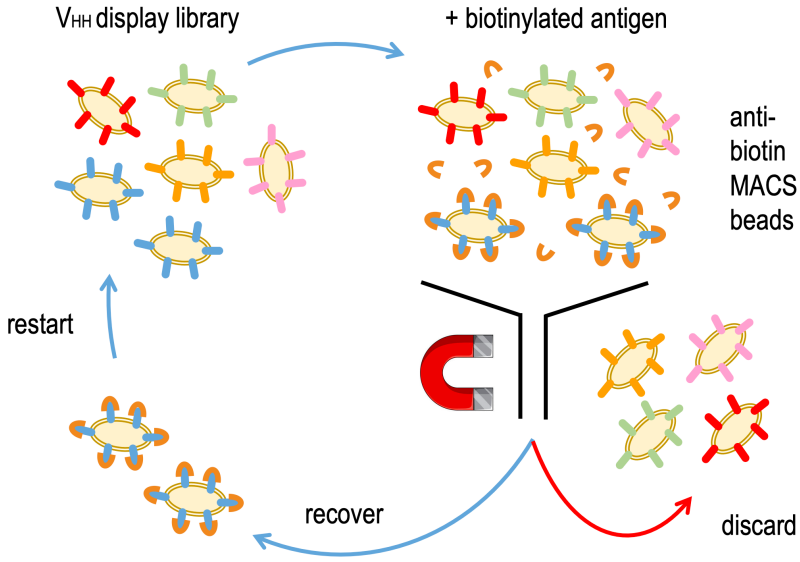


(A) The *E. coli* bacteria used in this work are transformed with plasmid pPORTMAGE (Supplementary Table S2) and display on their surface (B) a fusion between an outer membrane (OM)-bound intimin domain and a nanobody (VHH) which holds the 3 complementarity determinant regions (CDRs) for antigen binding flanked by diagnostic epitopes E-tag and *myc* as indicated. The cognate DNA sequences are encoded in the chromosome (C), where they can be targeted by mutagenic oligonucleotides (D) and generate a population with an increased diversity in the sequences of the CDRs thereby presented on the surface of individual bacteria (C).

Figure 2. Distribution of nucleotide changes through the anti-TirM^{EHEC} VHH sequence following mutagenic recombineering.

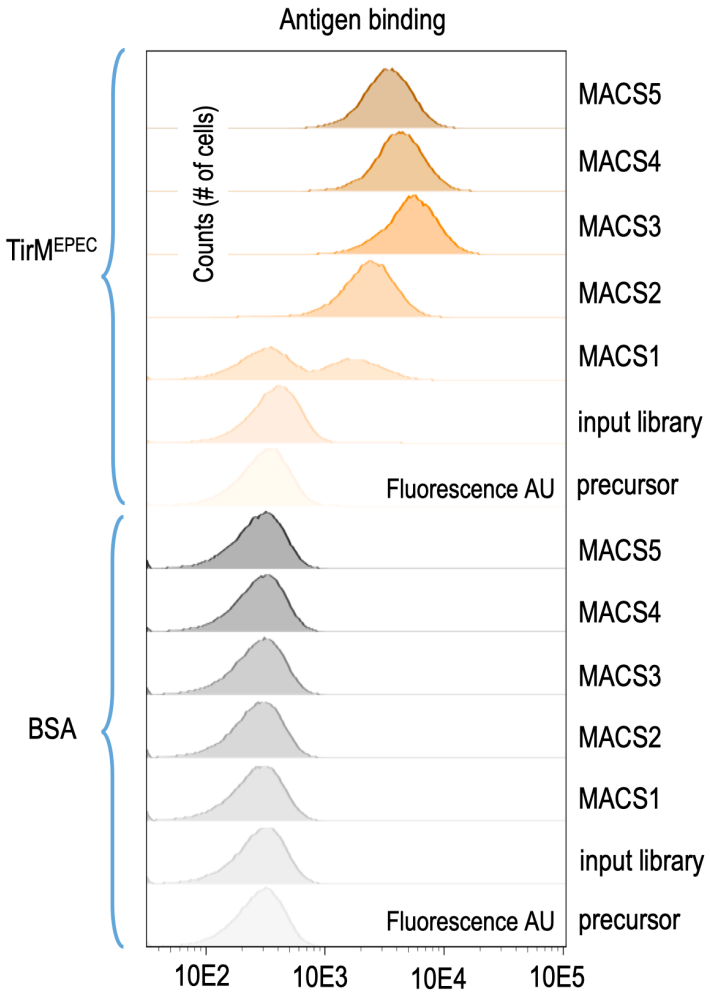
(A) DIvERGE is based on the use of mutagenic oligonucleotide pools that integrate in the three genomic loci encoding the complementarity determining regions (CDRs) of the VHH at stake and thus bring about generation of nanobody libraries. The CDRs are color-coded in the amino acid sequence of the parental nanobody TD4. Two cysteines involved in the formation of disulphide bridges by post-translational modification are highlighted in yellow. (B) Distribution of codon replacements in VHH sequence as exposed by the results of PacBio sequencing. (C) Distribution of single nucleotide replacements after *in vivo* mutagenesis, as determined by sequencing with Illumina technology. (D) Negative control following the same procedure but dispensing the use of mutagenic oligonucleotides. The lower mutagenic efficiency of the oligonucleotide pool targeting CDR1 may have resulted from a higher tendency to form secondary structures which could hinder the transformation process or the incorporation of these oligonucleotides during recombineering. The presence of replacements all over the VHH sequence, at low frequency, are likely to originate in the high error rate of Illumina sequencing [15], as it appears both in the mutagenized libraries and the control experiments.

Figure 3. Pipeline for enrichment of new binding specificities in evolved nanobodies.



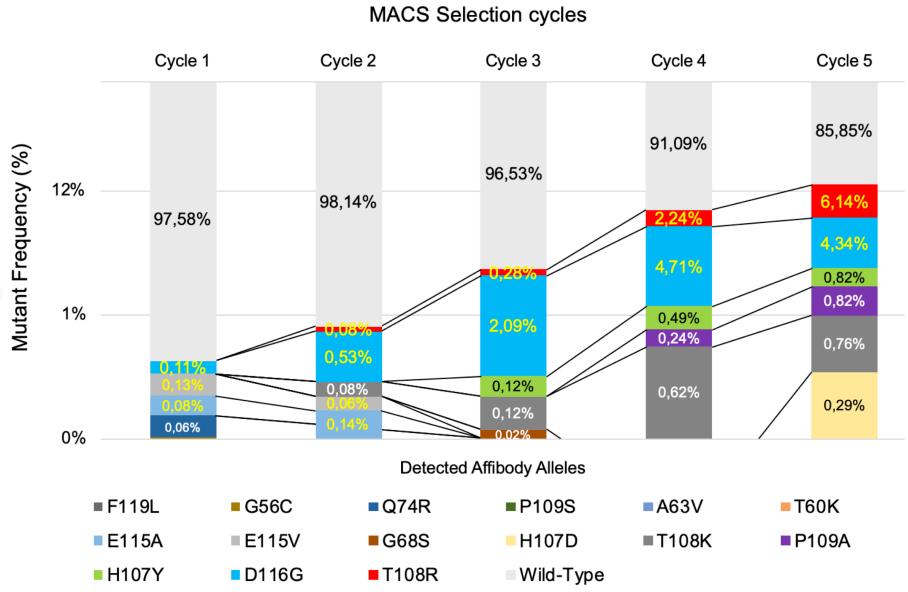
Libraries of *E. coli* bacteria displaying the VHH variants resulting from 10 DIvERGE cycles the are mixed with a biotinylated antigen and then with anti-biotin magnetic-activated cell sorting (MACS) beads. Only the cells displaying a nanobody that specifically binds the antigen can subsequently be bound to the MACS column. Wash and subsequent elution results in the enrichment of the libraries with antigen binding clones.

Figure 4. Evolution of new binding capabilities in *E. coli* populations expressing anti-TirM nanobodies



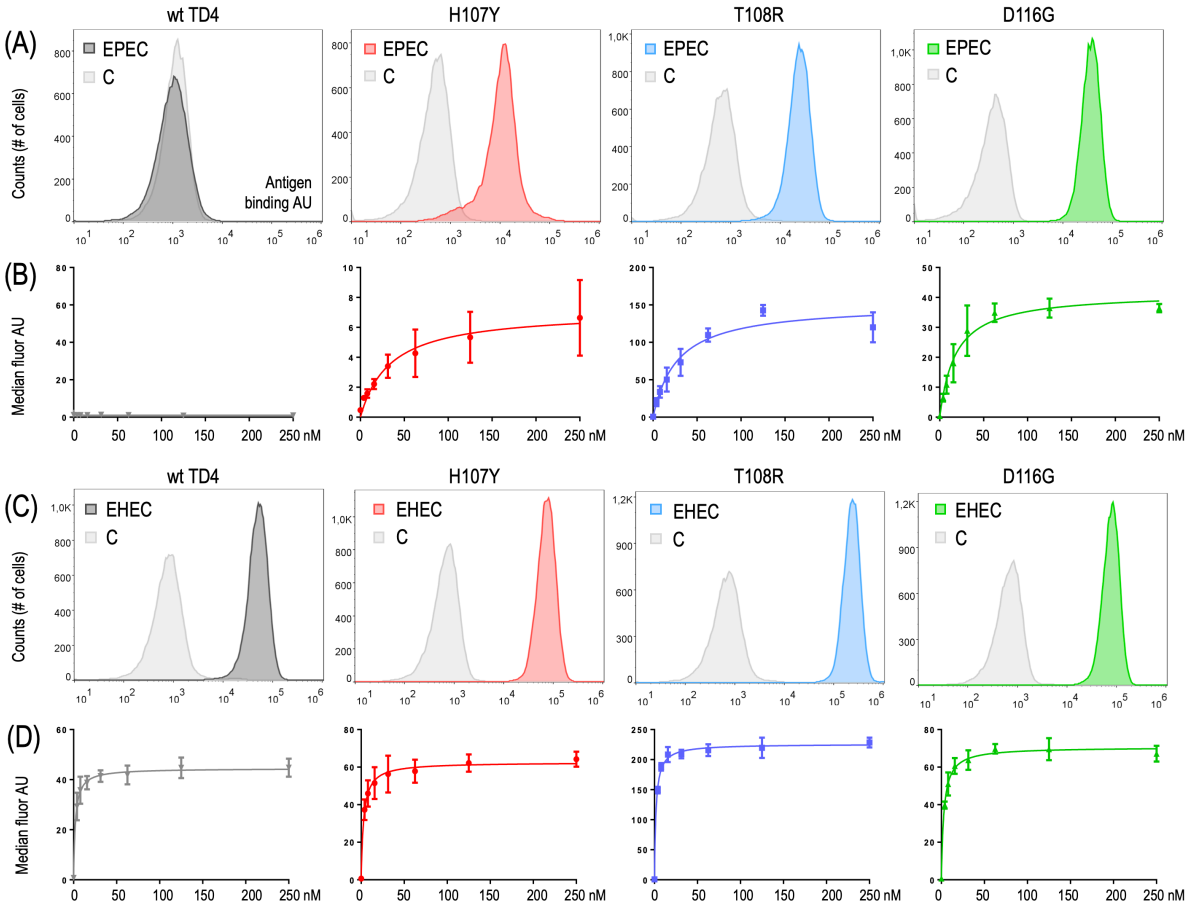
Following expression and display of the VHHs, bacteria retrieved from each of the MACS cycles were exposed to either biotinylated BSA or with biotinylated TirM^{EPEC}, followed by fluorescent staining with streptavidin-phycoerythrin. The results show how the bacterial pool is enriched with clones that bind the new antigen TirM$^{EPEC.}$ right after the first round of immunomagnetic selection. After the second cycle and beyond, the population seems predominantly composed by clones able to bind the new antigen. Note that these analyses just reflect bulk population behaviour, but they do not say anything yet on the nature of the mutations.

Figure 5. Progress of VHH population composition along cycles of immunomagnetic enrichment of TirM^{EPEC}-binding bacteria



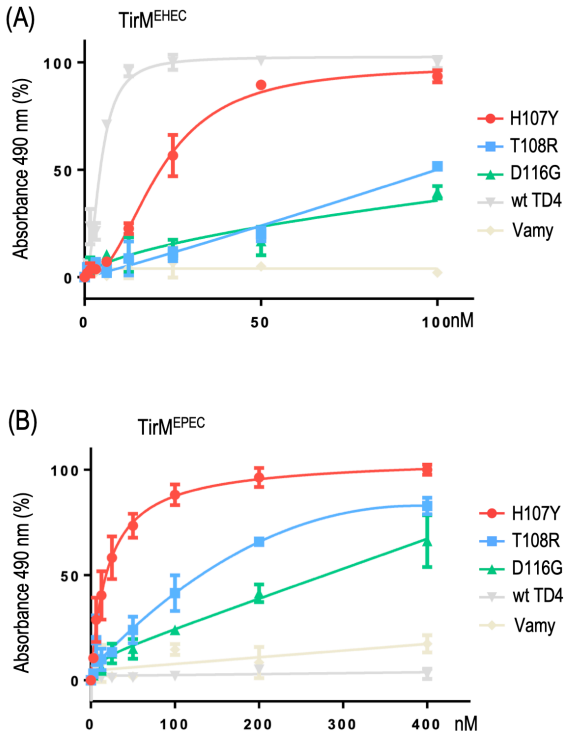
The frequencies and points of emergence of fifteen representative nanobody mutants displayed by bacteria captured in subsequent MACS cycles are shown in logarithmic scale in respect to those bearing the wild type VHH specimen.

Figure 6. Binding properties of the nanobodies displayed on the bacterial surface.



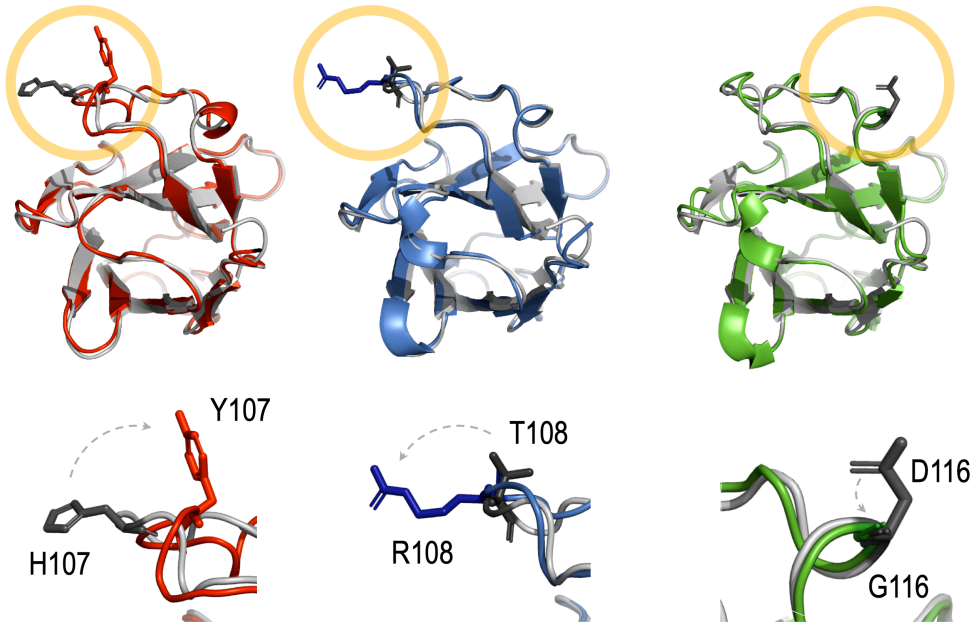
Bacterial clones displaying the nanobody variants indicated in each case on the cell surface were incubated with biotinylated antigens and streptavidin-allophycocyanin and then analysed by fluorescence flow cytometry. TD4 corresponds to the parental nanobody which was the template for the generation of libraries. H107Y, T108R and D116G are the names of three representative mutants and refer to the replacement they carry in the amino acid sequence of the nanobody. Panels (A) and (B) show the interaction of these nanobodies with the new antigen TirM^{EPEC}. Note that in contrast to the parental nanobody TD4, the replacements in the amino acid sequence of the new variants lead to a gain of function that enables them to recognize the new target. Panels (C) and (D) show the interaction with the former antigen TirM^{EHEC}. The four nanobodies tested still bind the former antigen TirM^{EHEC}. Control samples (tagged C in the cytometry panels A and C) were incubated only with buffer and streptavidin-allophycocyanin. Antigen binding was based on fluorescence intensity values measured in arbitrary units (AU). For each concentration of antigen in (B) and (D), median fluorescence values of four replicates were considered. Estimated Kd values are referred to in Table 1.

Figure 7. ELISA of purified nanobodies against TirM antigens.



Interaction of the evolved nanobodies with the former antigen TirM^{EHEC} (A) and new antigen TirM^{EPEC} (B). Vamy is a nanobody specific for alpha amylase used as an unspecific control. Note that TD4 does not bind the new antigen, H107Y binds quite well the former and new antigen (it may be considered as a *generalist*) while T108R and D116G have experienced a minor loss of their ability to bind the former antigen while gaining the ability to bind the new antigen (note that the x axis scale is different in each graph)

Figure 8. Comparison of the structural models of the nanobody variants.



The upper row shows of the different configurations of each new nanobody compared to the structural model of the precursor protein. H107Y (red), Nb T108R (blue) and Nb D116G (green) are shown aligned to TD4 (grey). The bottom row shows a blowup of the replacement positions with former residues is in dark gray and the new ones in dark red, dark blue and dark green respectively. Changes involve replacement of a basic residue by a bulkier, uncharged and polar counterpart in the case of H107Y, the change an uncharged and polar amino acid by a basic and larger one in T108R and the substitution of an acid residue by a neutral and smaller one in D116G. Although more evident in H107Y and T108R, a change in the orientation of the new amino acid with respect to the original alignment occurs in all the three replacements. The RMSD values for the structural alignments with TD4 were all < 1 (i.e. 0.781810 for H107Y, 0.487231 for T108R and 0.637916 for D116G). which is indicative that the overall structures of all these nanobodies can be considered almost identical.