JBC Papers in Press. Published on March 3, 2015 as Manuscript M114.614842
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M114.614842

Fully human V_H single domains that rival the stability and cleft recognition of camelid antibodies*

Romain Rouet[1a], Kip Dudgeon[1a], Mary Christie[1], David Langley[1] and Daniel Christ[12]

[1] From the Department of Immunology, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, NSW 2010, Australia

[2] The University of New South Wales, Faculty of Medicine, St Vincent's Clinical School, Darlinghurst, Sydney, NSW 2010, Australia

[a] Both authors contributed equally to this work

*Running title: *Human V_H single domains that rival camelid antibodies*

To whom the correspondence should be addressed: Daniel Christ, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, NSW 2010, Australia, Tel (61) 9295 8458; Fax (61) 9295 8404; E-mail: d.christ@garvan.org.au

Keywords: Antibody engineering, single domains, camelid antibodies, stability, cleft binding

Background: Camelid antibody domains are naturally stable and capable of cleft binding.
Results: Protein engineering can endow such properties onto human antibody domains.
Conclusion: Our strategy does not require undesirable antibody framework changes.
Significance: Robust building blocks for human therapeutic applications.

ABSTRACT

Human V_H single domains represent a promising class of antibody fragments with applications as therapeutic modalities. Unfortunately, isolated human V_H domains also generally display poor biophysical properties and a propensity to aggregate. This has encouraged the development of non-human antibody domains as alternative means of antigen recognition and, in particular, camelid (V_HH) domains. Naturally devoid of light chain partners, these domains are characterized by favourable biophysical properties and propensity for cleft binding, a highly desirable characteristic, allowing the targeting of cryptic epitopes. In contrast, previously reported structures of human V_H single domains had failed to recapitulate this property. Here we report the engineering and characterization of phage display libraries of stable human V_H domains, and the selection of binders against a diverse set of antigens. Unlike 'camelized' human domains, the domains do not rely on potentially immunogenic framework mutations and maintain the structure of the V_H/V_L interface. Structure determination in complex with hen egg-white lysozyme revealed an extended V_H binding interface, with complementarity determining region 3 (CDR3) deeply penetrating into the active site cleft, highly reminiscent to what has been observed for camelid domains. Taken together, our results demonstrate that fully human V_H domains can be constructed that are not only stable and well-expressed, but also rival the cleft binding properties of camelid antibodies.

Human antibodies are exclusively expressed as paired species, containing both heavy and light chains. In marked contrast to humans, llamas and camels have been shown to also produce non-paired species (1,2). Antigen binding of these heavy chain-only antibodies resides exclusively within a single variable domain (V_HH domain) (3). Since their discovery in the mid-1990s (1), camelid V_HH domains have attracted considerable interest and investment, due to their potential as building blocks for therapeutic and biotechnology

applications (4). The domains are biophysically and structurally well characterized, and have been shown to possess high thermodynamic and colloidal stabilities, together with a preference for binding within pockets of protein structure (high cleft binding propensity) (5,6). These features are achieved by means of an extended CDR3 loop, which has been shown to be capable of protruding deeply into enzyme active sites and otherwise cryptic epitopes of viruses and G-protein coupled receptors (7-9). Residues in the light chain interface differ between in human domains and V_HH domains, resulting in a hydrophilic nature of the camelid surface (2). In addition, the extended CDR3 loop of camelid domains is capable of looping backwards, further shielding the V_H/V_L interface area through interactions with framework residues. (5). However, this feature has also been shown to restrict camelid CDR3 diversity, which is highlighted by randomization studies revealing a strong consensus towards naturally occurring hydrophobic sequence motifs (10).

The favourable properties of camelid domains have inspired attempts to generate binders based on isolated human V_H domains, although initially with limited success. The main problem holding back the field has not been thermodynamic stability, but rather *colloidal* stability (aggregation propensity) (11,12). While thermodynamic stability can be achieved by the use of stable human V_H families (such as VH3) (13), colloidal stability remains difficult to control (14). It results in overall poor biophysical properties, as indicated by low solubility, expression and purification yields, and a lack of heat-refoldability (11,15).

'Camelization' strategies, as pioneered by Riechmann and co-workers, have been developed to improve the biophysical properties of human domains (16). These strategies have focused on a set of framework residues, which are hydrophobic in humans, but are largely hydrophilic in camelid domains (V_HH tetrad; positions 37, 44, 45 and 47 (2); all numbering according to Kabat (17)). However, initial attempts at directly transplanting such residues into a human framework were met with mixed success, with the resulting domains suffering from limited stabilities and structural deformations in NMR experiments (18). Structural changes have also been observed upon engineering of human domains through non-native disulfide links (19). Recent camelization attempts have therefore focused on other interface positions (such as 39) and on the introduction of novel substitutions at tetrad positions. A study of human single domains derived from the antibody therapeutic Trastuzumab (Herceptin 4D5) (20) revealed several such novel stabilizing framework mutations (10). The crystal structure of a soluble 4D5 quadruple mutant was reported in the same study (H35G, Q39E, L45E and R50S - clone B1a). Using an elegant display approach on phage (retention of superantigen binding) the authors also demonstrated compatibility of these mutations with CDR3 diversity, suggesting that domains stabilized in this manner may be capable of antigen binding. Indeed, the same group recently reported the structure of a V_H 4D5 triple mutant in complex with vascular epithelial growth factor (VEGF - V_H V1a complex) (21).

While such prior camelization approaches have undoubtedly resulted in improvements of biophysical properties, these engineered human domains nevertheless fall short of what would be considered ideal for many applications, and human therapy in particular. For instance, changes of multiple conserved framework residues may well result in the generation of novel B- and T-cell epitopes and an increase of immunogenicity in humans (22). In addition, the introduction of mutations into the V_H/V_L interface prevents pairing with light chain, which restricts the developability of the domains and limits their use in antibody bispecifics (23). Previously reported structures of human V_H single domain in complex with antigen have also largely failed to recapitulate the cleft binding properties of camelid domains (21,24). To overcome these limitations we decided to bypass the mutation of human framework residues altogether, instead exclusively focussing on the engineering of CDR regions. Here we investigate strategies for library design, and report biophysical and structural properties of fully human antibody V_H domains.

EXPERIMENTAL PROCEDURES

Construction of synthetic human V_H antibody repertoires - Synthetic libraries were constructed essentially as previously as described (25). For the generation of the Garvan I repertoire, single-

stranded DNA encoding the model V_H domain HEL4 in the phage vector FdMyc was isolated using a QIAprep spin M13 kit (Qiagen). Randomization of CDR regions was carried out by combinatorial mutagenesis using partially degenerate oligonucleotides (Table I). For the generation of the Garvan II repertoire, single-stranded DNA encoding DP47 germline (V3-23/DP-47 V-segment and JH4b J-segment derived from the immunoglobulin heavy chain locus of chromosome 14) in the phagemid vector pHEN1 was isolated using a QIAprep spin M13 kit (Qiagen). Randomization of CDR1 and CDR2 regions was carried out by Kunkel mutagenesis using reverse-complemented degenerated oligonucleotides. For CDR3 regions of the Garvan II library TRIM codon mutagenesis was utilized and diversity was introduced by splice-overlap extension PCR. After the mutagenesis steps, repertoires were transformed into *E. coli* TG1 bacteria and validated by DNA Sanger sequencing of random clones (>100 clones analyzed). Phages were purified by two precipitations with PEG/NaCl either directly from the culture supernatant in the case of FdMyc, or after a rescue step using KM13 helper phage in the case of pHEN1.

Phage display selection - Phage display selections were carried essentially out as described (26). In brief, 100 nM of biotinylated antigen was utilized in rounds 1 and 2, 10 nM in round 3 and 1 nM in round 4, which were immobilized on a neutravidin coated Maxisorp Immunoplate (Nunc) (rounds 1 and 3) or on streptavidin-coated magnetic beads (Dynal – rounds 2 and 4). Plates and beads were blocked with 5% milk powder (w/v) in PBS supplemented with 1% Tween-20 (MPBST) for 1 h. After washes with PBST (PBS buffer supplemented with 0.05% Tween-20), 1 x 10^{12} (round 1) or 1 x 10^{11} (rounds 2, 3 and 4) phages were pre-blocked in MPBST for 30 min, and then incubated with the antigen for 1 h. After washes with PBST, bound phages were eluted with 100 µg/ml trypsin protease (Sigma-Aldrich) in 10 mM Tris pH 7.4, supplemented with 137 mM NaCl and 1 mM $CaCl_2$ for 1 h. Eluted phages were used to infect mid-log phase *E. coli* TG1 bacteria and plated onto TYE-agar plates supplemented with 4% glucose and 100 µg/ml of ampicillin (pHEN1) or 15 µg/ml of tetracycline (FdMyc).

Enzyme-linked immunosorbent assay (ELISA) - For phage ELISA, a Maxisorp Immunoplate was coated with antigens and blocked with 5% milk powder (w/v) in PBS buffer. Phage supernatant was blocked with MPBST for 30 min and added to wells of the ELISA plate for 1 h. After washes with PBST, bound phage were detected using HRP conjugated anti-M13 antibody (GE Healthcare) and 3,3',5,5'-tetramethylbenzidine (TMB) substrate (Becton Dickinson). For soluble ELISA, a Maxisorp Immunoplate was coated, blocked and washed as above. Bound domains were detected using HRP conjugated anti-c-myc antibody (ICL Lab) and TMB substrate.

Heat refolding on phage - Aggregation resistance was analyzed by measuring retention of signal after heating in a phage ELISA format utilizing vector FdMyc as described (the pHEN1-based Garvan II library was re-cloned for this purpose) (27). A Maxisorp Immunoplate was coated with *Staphylococcus aureus* protein A (Sigma), and blocked with 5% milk powder (w/v) in PBS buffer. Single colonies were picked from agar plates and grown overnight at 30°C in 2xTY medium supplemented with 15 µg/ml tetracycline. Supernatant was cleared by centrifugation and phages were biotinylated by addition of NHS-PEG_4-biotin (Pierce) to a final concentration of 50 µM for 2 h. The reaction was then quenched by addition of Tris pH 7.5 to a final concentration of 100 mM and incubated for 1 h. For heat-induced refolding, biotinylated phage supernatant was incubated at 80°C for 10 minutes, followed by incubation at 4°C for 10 minutes. Supernatant was added to wells of the blocked ELISA plate and incubated for 1 h. After washes with PBST, bound phage were detected using HRP conjugated extravidin (Sigma) and TMB substrate. Finally, the level of retained protein A binding after heat treatment was calculated as a percentage of the untreated phage sample.

Heat refolding in solution - Refolding after heat denaturation in solution was determined using size-exclusion chromatography (SEC). For this purpose, purified protein at 10 µM in PBS was heated at 80°C for 10 minutes followed by cooling at 4°C for 10 minutes. Samples were analyzed on a Superdex-S75 gel-filtration column (GE

Healthcare) using an AKTA Purifier (GE Healthcare) chromatography system and PBS running buffer. Recovery of each variant was determined by measuring the area under the curve after heating, expressed as a percentage of the unheated protein sample.

Protein expression and purification - Expression and purification of human V_H domains was performed essentially as described (28). In brief, genes encoding V_H domains were recloned from phage or phagemid vectors into the expression vector pET12a (Novagen) and transformed into *E. coli* BL21-Gold (Stratagene). Transformants were used to inoculate 2xTY medium supplemented with 4% glucose, 100 µg/ml ampicillin, 15 µg/ml tetracycline at an OD_{600nm} of 0.05 in baffled flasks, and grown in a shaking incubator at 37°C. At an $OD_{600\ nm}$ of 0.5, bacteria were pelleted by centrifugation at 3200 x *g*, resuspended in 2xTY medium supplemented with 1 mM IPTG, 100 µg/ml ampicillin, 15 µg/ml tetracycline and incubated for 36-48 h at 30°C (with a second induction with 1 mM IPTG after 24 h). Supernatant was cleared by centrifugation at 10 000 x *g*, passed through a 0.45 µm filtration unit and pH adjusted to 7.0 – 8.0 by addition of 1 M HCl. Protein A resin (GE Healthcare) was applied to a gravity flow column (BioRad), and washed successively with 5 column volumes of mQ-H_2O and PBS. Supernatant was applied to the column by gravity flow and washed with 1 column volume of PBS. Protein was eluted with 100 mM glycine pH 2.7 and the eluted fractions neutralized with 100 mM Tris pH 8.8. Finally, the fractions were dialysed extensively against PBS and the purified domains concentrated using Amicon-Ultra 10kDa concentration devices (Millipore).

Surface plasmon resonance - Binding kinetics were determined by surface plasmon resonance (SPR) using a Biacore 2000 instrument (GE Healthcare). Biotinylated antigen was immobilized on a streptavidin chip (GE Healthcare) and serial dilutions of purified V_H domains in PBS were injected. Association and dissociation curves were fitted using a 1:1 Langmuir binding model and analyzed using BIAevaluation software (GE Healthcare).

Circular dichroism - Thermal unfolding was measured by circular dichroism (CD) using a J-815 spectrometer (Jasco) in a quartz cuvette (2 mm path length). Protein samples at a final concentration of 20 µM in PBS and melting curves were obtained by recording CD signal at 235 nm with a 1 nm bandwidth and 1 s integration time while heating the solutions from 20°C to 80°C at 1°C/min. To measure the reversibility of unfolding, signal was measured during cooling from 80°C to 20°C at 1°C/min.

Crystal growth, structure solution, refinement and analysis – The V_H H04-HEL complex was purified by size exclusion chromatography using 50 mM Tris-HCl pH 7.5 supplemented with 150 mM NaCl as running buffer (as described above). Peak fractions were collected and concentrated using Amicon-Ultra 10kDa concentration devices (Millipore). Initial crystallization hits were obtained with the JCSG-plus crystal screen (Molecular Dimensions) using a Mosquito crystallization robot (TTP Labtech) and 96-well MRC2 sitting drop crystallization plates (Swissci). Crystals were grown at room temperature in a hanging-drop format where 2 µL of V_H H04-HEL complex (4.37 mg/mL, in 50 mM Tris-HCl pH 7.5 supplemented with 150 mM NaCl) was combined with an equivalent volume of well solution (100 mM sodium citrate pH 5.4 supplemented with 28% (w/v) $PEG_{1,500}$ for PDB ID: 4PGJ; 300 mM sodium citrate pH 5.5 supplemented with 16% (w/v) $PEG_{3,350}$ for PBD ID: 4U3X). Crystals were snap frozen in liquid nitrogen (100 K) and diffraction data were recorded on the beamline MX2 at the Australian Synchrotron. Reflections were indexed and integrated with iMOSFLM (29), scrutinized for symmetry with POINTLESS (30), scaled with SCALA (30), and imported into the CCP4i software package (31). Reflection data statistics are shown in Table S2. Structures were solved by molecular replacement using PHASER (32). The search models for the V_H and HEL components were PDB entry 1OHQ (V_H) and PDB entry 1ZVY (HEL), both stripped of side-chain moieties. Two essentially identical V_H-HEL complexes were found in the asymmetric unit, where H04 domains are chained A and C, and corresponding HEL molecules are chained B and D. During refinement the two complexes were not averaged using non-crystallographic symmetry but

were treated independently of each other. Initial rounds of rigid body refinement were followed by B-factor restrained refinement using REFMAC5 (33) and employed translation-libration-screw (TLS) parameterization (34). During later stages of refinement water molecules were modeled into appropriate features of difference density. Models were compared with maps and manipulated in real space using COOT (35). Models were scrutinized using the MOLPROBITY validation server (36). Buried surfaces were calculated with PDBePISA (37). Coordinates for V_H H04-HEL structure have been deposited in the PDB as entry 4PGJ and 4U3X respectively.

RESULTS

Generation of the Garvan I human V_H phage repertoire - As a starting point for the design of a first generation repertoire of stable V_H domains we utilized the HEL4 model protein. This human V_H domain had been identified by selection on phage and its structure had been determined by crystallography (27,38). It is characterized by low aggregation propensity (high colloidal stability), heat-refoldability and excellent expression and purification yields (27,39). HEL4 differs from the DP47 germline domain it had been derived from at over twenty CDR and framework positions (Figure 1A), and the specific role of individual changes had largely remained unclear (27,38). We initially focused our attention onto the complementarity determining region 3 (CDR3) (Figure 1B – in blue), which provides the majority of binding energy in antibody/antigen interactions. Using site directed mutagenesis we introduced diversity at residues 95 to 100a/c, predominantly utilizing DVK codons to introduce hydrophilic amino acids into CDR3 (Table 1). The use of such codons has previously been successfully utilized in the generation of paired antibody fragment libraries (40-42). Transformation into *E. coli* yielded a library of 3 x 10^9 individual clones.

Characterisation of the Garvan I phage repertoire - For initial characterisation, resistance to heat-denaturation was determined using a phage method originally developed by Jespers *et al* (27). The method is based on the multivalent display of human antibody domains on phage, coupled with heating to 80ºC and cooling of the phage particle. Resistant domains can be captured by the conformation dependent superantigen protein A, which binds to folded but not to unfolded or aggregated V_H domains (27,43). The phage method is an excellent predictor of solution properties of human antibody V_H domains, including heat-refoldability, expression and purification yields (44). Human antibody variable domains, such as DP47 germline, aggregate under the conditions of the phage method, however this is not observed for the HEL4 domain, which resists heat-induced aggregation (Figure 1C).

Initial CDR grafting studies carried out by our group (Figure 1C) (45) and more recently by others (46) had indicated the aggregation determinants of HEL4 largely reside within the CDR1 region. As this region had not been targeted in our randomization strategy (which focused on CDR3 instead), we speculated that the Garvan I library may retain the aggregation resistant properties of the HEL4 domain. Further analyses of randomly selected clones from the naïve Garvan-1 repertoire revealed that this was indeed the case, and that the majority of clones displayed considerable resistance against heat-induced aggregation on phage (Figure 1D), with a median level of resistance of 64% compared to 88% for HEL4 and 2% for DP47 germline.

Phage-display selection using the Garvan I phage repertoire - We next decided to assess the potential of the library for the selection of antigen specific V_H domains. For this purpose a set of representative antigens was chosen, covering a wide range of molecular weights and structural classes (human tumour necrosis factor alpha, murine interleukin 21, human prolactin receptor, beta-galactosidase) (Figure 1E). The antigens were immobilized on solid supports and single domains selected by phage-display. After three rounds of selection, phage ELISA confirmed that binders had been obtained for all antigens, with 22-57% of all clones displaying antigen binding. After re-cloning into the pET12a expression vector, monoclonal ELISA was performed on soluble protein, and non-redundant binders identified by DNA sequencing. This revealed that multiple unique V_H domains had been selected for each target antigen.

Characterisation of a stable human anti-TNFα single domain – For further analyses, clone G07 was expressed in large scale and its biophysical properties and affinity were characterized. Thermal denaturation as monitored by circular dichroism revealed that the domain was fully heat-refoldable (Figure 2A, T_m = 63°C), similar to what has been observed for HEL4 and camelid domains (but not for other human V_H domains). Moreover, the domain was also well-expressed in bacteria (at around 5 mg/litre in shaking flasks, compared to 3.3 mg/litre for HEL4 and 0.9 mg/litre for DP47). Affinity measurements by SPR Biacore revealed moderate affinity interactions with the human tumour necrosis factor alpha (TNFα) antigen against which the domain had been selected (k_a = 4.8 x 10^3 $s^{-1}M^{-1}$, k_d = 8.9 x 10^{-3} s^{-1}, K_D = 1.9 μM) (Figure 2B).

Generation of the Garvan II human V_H phagemid repertoire – We next investigated strategies to further increase the affinity of the human domains. The first generation library had been constructed in a multivalent phage vector, thereby streamlining analysis of aggregation resistance using the heat/cool method described by Jespers *et al.* However, it is also evident that multivalent display can promote the enrichment of low affinity clones (through increased avidity, resulting in multiple interactions with antigen immobilized on a solid support) (47). We therefore decided to utilize a phagemid system for the construction of a second-generation library, allowing for monovalent display on the phage surface. In addition to changes to the display system, we also considerably increased the number of CDR positions diversified in the library (Figure 3A), while avoiding positions important for biophysical properties.

Recent work in our laboratory had demonstrated that the introduction of negatively charged amino acids at CDR1 positions significantly improved the colloidal stability of human V_H domains (aspartate or glutamate at 28, 30, 31, 32, 33 and 35) (48). This is in agreement with the observations for the HEL4 model domain, which carries a negatively charged triad within this region (31-33DED) (Figure 3A). Importantly, introduction of multiple negatively charged amino acids at CDR1 positions is required (two or more) (48). This can be readily observed upon step-wise introduction of mutations into DP47, resulting in considerable increases of heat-refoldability (Figure 3B). In addition to reversible unfolding the introduction of mutations at the above CDR1 positions also significantly increases expression and purification yields (44,48), mimicking important and highly desirable characteristic of camelid domains. We also retained two other CDR mutations (28R, 35G) within the libraries that are present in the HEL4 model domain (Figure 3A). These have been reported to result in modest improvement of V_H properties (10,27,38,49). In contrast to HEL4 and previously reported camelized domains all of the mutations utilized here are located at CDR positions, fully retaining the human DP47 framework sequence (Figure 3A).

More specifically, negatively charged amino acids were introduced at CDR positions 32 and 33, while extensively randomizing CDR1, CDR2 and CDR3 positions to obtain a maximum degree of repertoire diversity (Figure 3A/C – Table I). Based on our experience with the Garvan I library, we predominantly utilized codons encoding hydrophilic amino acids at CDR1 and CDR2 positions (Table I). For the diversification of CDR3 we utilized tri-phosphoramidite codon mutagenesis (TRIM) (50). This enabled us to encode a high proportion of amino acids commonly observed in the human repertoire (in particular Tyr, Gly, Ala and Ser) with other amino acids at lower defined proportions (Table I) (51). A library of 3 x 10^9 individual clones was obtained after transformation into *E. coli* bacteria (Garvan II library). From the naïve library clones were randomly selected and analysed using the phage heat/cool method (27). This revealed that a large proportion of the library displayed resistance against heat-induced aggregation (with a median resistance of 40%), despite the randomization of up to sixteen positions within three CDR regions (Figure 3D).

Phage-display selections using the Garvan II phagemid repertoire – Antigen specific binders were next selected from the second generation repertoire. As described for the Garvan I library, antigens were immobilized on solid supports, and after three to four rounds of sections binders were

identified by phagemid ELISA, and verified by soluble ELISA. After selection, a large proportion of clones (14% - 58% depending on the target antigen) displayed antigen binding, with DNA sequencing revealing non-redundant sets of 1-6 (unique) clones per target (Figure 3E). Two of the antigens (VEGF and CD25) represent validated therapeutic targets and had been included in the selections for this reason. Biosensor measurements of representative V_H domains selected against these targets revealed high affinity interactions with equilibrium binding constants in the mid nanomolar range (VEGF: $k_a = 5.7 \times 10^4$ $s^{-1}M^{-1}$, $k_d = 2.3 \times 10^{-2}$ s^{-1}, $K_D = 407$ nM) (CD25*:* $k_a = 6.6 \times 10^4$ $s^{-1}M^{-1}$, $k_d = 1.4 \times 10^{-2}$ s^{-1}, $K_D = 208$ nM).

Characterisation of Garvan II V_H single domains and epitope binning – In addition to therapeutic targets we had also included the common model antigen hen egg white lysozyme (HEL) in the selections, to streamline comparisons with previously reported structures of camelid V_HH complexes (52). The selected anti-HEL V_H single domains were cloned into the pET12a expression vector, expressed in large scale and further characterized (clones H04, G08, D05, C01 and F05). Surface plasmon resonance was used to determine binding kinetics, revealing equilibrium binding down to the low nanomolar range (Figure 4A/B - 26 to 257 nM). In addition to high affinity antigen binding, all of the selected clones also displayed high levels of resistance against heat-induced aggregation (Figure 4B).

To gain further insight into the epitope distribution of the selected clones, we performed competition assays against three previously reported anti-HEL camelid domains (clones D3L11, D2L19 and D2L24) (52). D3L11 and D2L19 bind to the active site cleft of the HEL antigen (Figure 4C - asterisk). This represents the dominant mode of binding of camelid domains to this antigen, which is well characterized through previously reported co-crystal structures (5,52). In contrast, D2L24 binds to a planar epitope on the back of the molecule, distant from the active site cleft (Figure 4C) (52). Epitope binning experiments were carried out using surface plasmon resonance and one of the human V_H single domains selected here (clone H04 characterized by a high affinity and stable SPR baselines). These experiments revealed that H04 competed for binding with D3L11 and D2L19, but not with D2L24 (Figure 4D). Moreover, H04 also completed for binding to antigen with G08, D05, C01 and F05 (Figure 4E). This strongly indicated that all of the human domains selected in this study bind to epitopes within or proximal to the lysozyme active site cleft.

Crystal structure of a human V_H domain in complex with hen-egg white lysozyme - To gain further insights into the interaction of V_H H04 with HEL antigen, molecular details were investigated by X-ray crystallography. The V_H-antigen complex was isolated by size exclusion chromatography and crystals were obtained by hanging drop vapour diffusion under two distinct conditions (see experimental procedures for details). Diffraction data was collected at the Australian Synchrotron, structures solved by molecular replacement, and refined to 2.60 Å and 2.26 Å resolution (Table II). The structures obtained from both datasets displayed two complexes within the asymmetric unit and overall high levels of structural similarity (with root-mean-square-deviations (RMSD) of 0.38-0.52 Å for 231-246 Cα atoms). Further investigations therefore focused on the 2.26 Å resolution dataset.

Nature of the single domain-antigen interaction and comparative analyses - Initial structural analyses revealed a 1:1 stoichiometry and binding of a single V_H H04 domain to a single HEL antigen (Figure 5A). The interaction is centred on the CDR3 region of V_H H04, which protrudes deeply into the lysozyme active site cleft (Figure 5B), burying extensive surface area (Table III: 535 $Å^2$). Notably, this interaction surface is achieved within the context of a CDR3 loop of nine residues, considerably shorter than what had been observed for camelid V_HH and shark V_{NAR} domains raised against HEL antigen (12-18 residues; Table III) (52-54). More specifically, an extensive network of hydrogen bonds is observed between CDR3 residues (Y95, S97, P99, Q100, N100a, H100b) and the HEL cleft surface (Figure 5B). Intriguingly, two prolines (at positions 98 and 99) are observed in the centre of the loop. The presence of sequential prolines in CDR3 is unusual, and results in a distinct bend of backbone orientation, projecting key contact residues

towards the HEL active site. Complementing CDR3 interactions, minor contacts from CDR1 and CDR2 are observed in the H04-HEL complex (buried surface areas of 20 Å^2 and 60 Å^2, respectively). Notably, an asparate residue at CDR1 position 33, which is a determinant of aggregation resistance (48), participates in a salt bridge with Arg61 of the antigen (Figure 5B). Outside the CDR regions of V_H H04, an additional 168 Å^2 of single domain surface is buried by the HEL surface. These interactions involve a number of mainly hydrophobic residues (Val37, Leu45 and Trp47) within the light chain interface of the domain, packing against the surface of the HEL antigen (Figure 5C).

In total, a surface area of 783 Å^2 is buried by V_H H04 in the lysozyme complex, which is comparable to that of a shark V_{NAR} single domain (755 Å^2), but higher than what is observed for a set of camelid domains raised against this antigen (561 – 645 Å^2) (Table III). All of the analysed structures display a large degree of shape complementarity between single domains and the HEL antigen (H04: 0.81, V_{NAR}: 0.77, Camelids: 0.76-0.80). In the majority (5/6) of complexes this high level of complementarity is achieved by means of an extended CDR3 interacting with the active site cleft of the lysozyme antigen. In contrast to H04, shark and camelid domains, previously reported structures of human single domains (V_H9 (24) and V1a (21)) display lower levels of shape complementarity with antigen (0.69 and 0.70). This is reflected by the essentially planar epitope surfaces of their respective antigens (MDM4 and VEGF) (Figure 6). Although both domains bind to antigen predominantly through CDR3 contacts, the loop is not inserted into structural clefts, but rather packs against a hydrophobic helix surface (in the case of V_H9-MDM4) (24), or interfaces with an outer beta strand (in the case of V1a-VEGF) (21).

Structural features of CDR3 and the V_H/V_L interface - In addition to contacts with antigen, the structure of the human H04 single domain also revealed important structural roles of CDR3 residues. As outlined above, a hallmark of camelid domains involves the presence of a hydrophobic pocket formed by CDR3 packing against hydrophobic chains of the V_HH framework (5,52,55-57) (as observed in the structure of the D2L24 anti-lysozyme V_HH domain - Figure 7A). Although these framework residues are not strictly conserved between human and camelid V_HH domains, an analogous core is nonetheless observed in the V_H H04 structure with a tyrosine residue at CDR position 95 packing against a hydrophobic framework core formed by residues at positions 35, 37, 93 and 100d (Figure 7B). Tyrosines are common at Garvan II CDR3 positions and within the human repertoire rendering it likely that similar structural features could be observed in other Garvan II domains. (Table I)(51). This finding is in contrast to the previously reported structure the camelized B1a domain in which the interaction is mediated through a rare CDR tryptophan residue (Figure 7C) (10). Taken together, the formation of a hydrophobic core at the base of H04 CDR3, and the restriction of backbone dihedrals through proline residues, result in the stabilization of CDR conformations, allowing antigen binding through extensive cleft interactions.

The observed hydrophobic core is partially formed by a residue at CDR1 position 35, which is occupied by glycine in H04 and conserved in the Garvan II library. This amino acid is also observed in D2L24 (Figure 7A), B1a (Figure 7C), as well as in HEL4, where results in moderate improvements of gel filtration profiles (but not heat-refoldabilty) (27,38). This is in agreement with the findings Barthelmy *et al.* who observed an enrichment of glycine at position 35 when selecting for protein A superantigen binding as a (gentle) proxy of biophysical properties (no heat selection was utilized) (10). In the HEL4 structure, the presence of Gly35 results in a marked conformational shift of Trp47 within the V_H/V_L interface (Figure 5D; in red) resulting in an increase of surface hydrophilicity (38). However, such changes are not observed in the V_H H04 structure reported here with the conformation of Trp47 closely aligned with that observed in representative human Fab antibody structures (Figure 5D). This suggests that the main role of G35 may not be the rearrangement of Trp47 conformations (38), but rather the formation of a hydrophobic pocket centred on positions 35, 37, 93 and 100x which can act as an acceptor for hydrophobic CDR3

residues (as seen in the H04, B1a and D2L24 structures – Figures 7A, 7B and 7C).

In addition to Trp47 the conformation of the remainder of the H04 V_H/V_L interface (Figure 8A), formed by residues at positions 37, 44 and 45, is also highly similar to that of native (light chain paired) human antibodies (Figure 8B – PDB identifier 3QOS). In contrast, the H04 interface is noticeably different to that of camelized human domains (Figure 8C) and camelid V_HH domains (Figure 8D).

DISCUSSION

Our study of human V_H single domains demonstrates that, unlike what has been reported for camelized human domains, improvements of biophysical properties do not necessarily rely on mutational changes of framework residues. Rather, we demonstrate that CDR composition alone can endow favourable properties onto human single domains. Importantly, we demonstrate that these properties are largely independent of CDR3 diversity, allowing the randomization of this important antigen-binding region. This enabled us to generate a stable first-generation library (capable of refolding after heating to 80ºC) and to select antigen binders against a diverse set of antigens.

In a second step, we extensively randomized additional CDR1 and CDR2 positions, while maintaining key determinants of biophysical properties (two or more negatively charged amino acids at Kabat positions 28, 30, 31, 32, 33, 35). This second-generation V_H library retained a high proportion of stable V_H domains. While we had observed the selection of predominantly low-medium affinity clones for the Garvan I library, affinities in the nanomolar range were observed for the second-generation Garvan II repertoire, comparable to what has been observed for camelid and shark domains (Table III). Moreover, the selected human single domains also displayed structural modes (cleft-binding) and shape complementarities that were highly reminiscent of shark and camelid domains. This is in marked contrast to previously reported structures of human and 'camelized' domains, which are characterized the absence of cleft interactions and lower shape complementarities.

In addition to the 'camelization' of human V_H, strategies for the humanization of camelid single domains have also become available in recent years. For instance, in a detailed study Vinke *et al.* have reported the generation of a partially humanised domain (h-NbBCII10$_{FGLA}$) suitable as an acceptor framework for CDR grafting approaches (58). However, the study also highlighted limitations of this strategy, namely the loss of heat-refoldability upon humanization, and the requirement to maintain multiple camelid residues within the framework 2 region (58).

Taken together, our results demonstrate that libraries of human V_H single domains can be constructed that rival the stability and cleft binding properties of camelid domains. The availability of fully human domains will minimize potential immunogenicity risks, by removing the necessity to change large conserved antibody framework residues. While the effects of framework changes on the immunogenicity of human antibody therapeutics are unknown, changes to CDR regions are well studied, with the majority of antibodies in clinical practice having been generated through CDR-grafting or CDR affinity maturation approaches (59,60). It is becoming increasingly evident that such changes have little or no detectable effect on immunogenicity in patients (22), further strengthening the rationale for the CDR-only approach outlined here. Moreover, by maintaining the structure of the human V_H/V_L interface the domains developed in this study retain the potential of light chain pairing; this may allow the modular assembly of bispecifics through native chain interactions. We conclude that the availability of fully human V_H domains has the potential to open up new and exciting opportunities for the development of human therapeutics.

REFERENCES

1. Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N., and Hamers, R. (1993) Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448
2. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A., and Hamers, R. (1994) Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng* **7**, 1129-1135
3. Muyldermans, S. (2001) Single domain camel antibodies: current status. *J Biotechnol* **74**, 277-302
4. Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* **82**, 775-797
5. Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., Muyldermans, S., and Wyns, L. (1996) Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature structural biology* **3**, 803-811
6. Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., and Cambillau, C. (2002) Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. *J Biol Chem* **277**, 23645-23650
7. Wesolowski, J., Alzogaray, V., Reyelt, J., Unger, M., Juarez, K., Urrutia, M., Cauerhff, A., Danquah, W., Rissiek, B., Scheuplein, F., Schwarz, N., Adriouch, S., Boyer, O., Seman, M., Licea, A., Serreze, D. V., Goldbaum, F. A., Haag, F., and Koch-Nolte, F. (2009) Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. *Med Microbiol Immunol* **198**, 157-174
8. Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., Devree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzki, I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. I., and Kobilka, B. K. (2011) Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **469**, 175-180
9. Lulf, S., Matz, J., Rouyez, M. C., Jarviluoma, A., Saksela, K., Benichou, S., and Geyer, M. (2014) Structural basis for the inhibition of HIV-1 Nef by a high-affinity binding single-domain antibody. *Retrovirology* **11**, 24
10. Barthelemy, P. A., Raab, H., Appleton, B. A., Bond, C. J., Wu, P., Wiesmann, C., and Sidhu, S. S. (2008) Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. *J Biol Chem* **283**, 3639-3654
11. Ewert, S., Cambillau, C., Conrath, K., and Pluckthun, A. (2002) Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains. *Biochemistry* **41**, 3628-3636
12. Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T., and Winter, G. (1989) Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature* **341**, 544-546
13. Ewert, S., Huber, T., Honegger, A., and Pluckthun, A. (2003) Biophysical properties of human antibody variable domains. *J Mol Biol* **325**, 531-553
14. Lowe, D., Dudgeon, K., Rouet, R., Schofield, P., Jermutus, L., and Christ, D. (2011) Aggregation, stability, and formulation of human antibody therapeutics. *Adv Protein Chem Struct Biol* **84**, 41-61
15. Rouet, R., Lowe, D., and Christ, D. (2014) Stability engineering of the human antibody repertoire. *FEBS Lett* **588**, 269-277
16. Riechmann, L., and Muyldermans, S. (1999) Single domain antibodies: comparison of camel VH and camelised human VH domains. *J Immunol Methods* **231**, 25-38

17. Kabat, E., Wu, T. T., Perry, H. M., Kay, S., and Gottesman, C. F. (1992) *Sequences of Proteins of Immunological Interest*, 5 ed., DIANE Publishing
18. Riechmann, L. (1996) Rearrangement of the former VL interface in the solution structure of a camelised, single antibody VH domain. *J Mol Biol* **259**, 957-969
19. Kim, D. Y., Kandalaft, H., Ding, W., Ryan, S., van Faassen, H., Hirama, T., Foote, S. J., MacKenzie, R., and Tanha, J. (2012) Disulfide linkage engineering for improving biophysical properties of human VH domains. *Protein Eng Des Sel* **25**, 581-589
20. Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B., Henner, D., Wong, W. L., Rowland, A. M., Kotts, C., Carver, M. E., and Shepard, H. M. (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci U S A* **89**, 4285-4289
21. Ma, X., Barthelemy, P. A., Rouge, L., Wiesmann, C., and Sidhu, S. S. (2013) Design of synthetic autonomous VH domain libraries and structural analysis of a VH domain bound to vascular endothelial growth factor. *J Mol Biol* **425**, 2247-2259
22. Hwang, W. Y., and Foote, J. (2005) Immunogenicity of engineered antibodies. *Methods (San Diego, Calif* **36**, 3-10
23. Rouet, R., and Christ, D. (2014) Bispecific antibodies with native chain structure. *Nat Biotechnol* **32**, 136-137
24. Yu, G. W., Vaysburd, M., Allen, M. D., Settanni, G., and Fersht, A. R. (2008) Structure of Human MDM4 N-Terminal Domain Bound to a Single-Domain Antibody. *J Mol Biol*
25. Rouet, R., Dudgeon, K., and Christ, D. (2012) Generation of human single domain antibody repertoires by Kunkel mutagenesis. *Methods Mol Biol* **907**, 195-209
26. Lee, C. M., Iorno, N., Sierro, F., and Christ, D. (2007) Selection of human antibody fragments by phage display. *Nat Protoc* **2**, 3001-3008
27. Jespers, L., Schon, O., Famm, K., and Winter, G. (2004) Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat Biotechnol* **22**, 1161-1165
28. Rouet, R., Lowe, D., Dudgeon, K., Roome, B., Schofield, P., Langley, D., Andrews, J., Whitfeld, P., Jermutus, L., and Christ, D. (2012) Expression of high-affinity human antibody fragments in bacteria. *Nat Protoc* **7**, 364-373
29. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281
30. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82
31. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
34. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **57**, 122-133
35. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
36. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-

atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21
37. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797
38. Jespers, L., Schon, O., James, L. C., Veprintsev, D., and Winter, G. (2004) Crystal structure of HEL4, a soluble, refoldable human V(H) single domain with a germ-line scaffold. *J Mol Biol* **337**, 893-903
39. Olichon, A., Schweizer, D., Muyldermans, S., and de Marco, A. (2007) Heating as a rapid purification method for recovering correctly-folded thermotolerant VH and VHH domains. *BMC Biotechnol* **7**, 7
40. Sidhu, S. S., Li, B., Chen, Y., Fellouse, F. A., Eigenbrot, C., and Fuh, G. (2004) Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol* **338**, 299-310
41. Lee, C. V., Liang, W. C., Dennis, M. S., Eigenbrot, C., Sidhu, S. S., and Fuh, G. (2004) High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J Mol Biol* **340**, 1073-1093
42. de Wildt, R. M., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat Biotechnol* **18**, 989-994
43. Jansson, B., Uhlen, M., and Nygren, P. A. (1998) All individual domains of staphylococcal protein A show Fab binding. *FEMS Immunol Med Microbiol* **20**, 69-78
44. Dudgeon, K., Rouet, R., and Christ, D. (2013) Rapid prediction of expression and refolding yields using phage display. *Protein Eng Des Sel* **26**, 671-674
45. Christ, D., and Dudgeon, K. (2011) Modified variable domain molecules and methods for producing and using same. *WO/2011/047442* **PCT/AU2010/001416**
46. Perchiacca, J. M., Bhattacharya, M., and Tessier, P. M. (2011) Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions. *Proteins* **79**, 2637-2647
47. O'Connell, D., Becerril, B., Roy-Burman, A., Daws, M., and Marks, J. D. (2002) Phage versus phagemid libraries for generation of human monoclonal antibodies. *J Mol Biol* **321**, 49-56
48. Dudgeon, K., Rouet, R., Kokmeijer, I., Schofield, P., Stolp, J., Langley, D., Stock, D., and Christ, D. (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc Natl Acad Sci U S A* **109**, 10879-10884
49. Famm, K., Hansen, L., Christ, D., and Winter, G. (2008) Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J Mol Biol* **376**, 926-931
50. Virnekas, B., Ge, L., Pluckthun, A., Schneider, K. C., Wellnhofer, G., and Moroney, S. E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res* **22**, 5600-5607
51. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A., and Virnekas, B. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol* **296**, 57-86
52. De Genst, E., Silence, K., Decanniere, K., Conrath, K., Loris, R., Kinne, J., Muyldermans, S., and Wyns, L. (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci U S A* **103**, 4586-4591

53. De Genst, E., Silence, K., Ghahroudi, M. A., Decanniere, K., Loris, R., Kinne, J., Wyns, L., and Muyldermans, S. (2005) Strong in vivo maturation compensates for structurally restricted H3 loops in antibody repertoires. *J Biol Chem* **280**, 14114-14121
54. Stanfield, R. L., Dooley, H., Flajnik, M. F., and Wilson, I. A. (2004) Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* **305**, 1770-1773
55. De Genst, E., Chan, P.-H., Pardon, E., Hsu, S.-T. D., Kumita, J. R., Christodoulou, J., Menzer, L., Chirgadze, D. Y., Robinson, C. V., Muyldermans, S., Matagne, A., Wyns, L., Dobson, C. M., and Dumoulin, M. (2013) A Nanobody Binding to Non-Amyloidogenic Regions of the Protein Human Lysozyme Enhances Partial Unfolding but Inhibits Amyloid Fibril Formation. *The Journal of Physical Chemistry B* **117**, 13245-13258
56. De Genst, E., Handelberg, F., Van Meirhaeghe, A., Vynck, S., Loris, R., Wyns, L., and Muyldermans, S. (2004) Chemical basis for the affinity maturation of a camel single domain antibody. *J Biol Chem* **279**, 53593-53601
57. Decanniere, K., Transue, T. R., Desmyter, A., Maes, D., Muyldermans, S., and Wyns, L. (2001) Degenerate interfaces in antigen-antibody complexes. *Journal of Molecular Biology* **313**, 473-478
58. Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S., and Conrath, K. (2009) General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J Biol Chem* **284**, 3273-3284
59. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S., and Winter, G. (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522-525
60. Nelson, A. L., Dhimolea, E., and Reichert, J. M. (2010) Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* **9**, 767-774
61. Yu, G. W., Vaysburd, M., Allen, M. D., Settanni, G., and Fersht, A. R. (2009) Structure of Human MDM4 N-Terminal Domain Bound to a Single-Domain Antibody. *Journal of Molecular Biology* **385**, 1578-1589
62. Lawrence, M. C., and Colman, P. M. (1993) Shape Complementarity at Protein/Protein Interfaces. *Journal of Molecular Biology* **234**, 946-950

FOOTNOTES

Atomic coordinates (codes 4PGJ and 4U3X) have been deposited in the Protein Data Bank

TABLE I. Oligonucleotides used for library construction[$]

Garvan I	
CDR3	5′-GTATATTATTGCGCGAGTGCT**DVKDVKDVKDVKDVKDVKDVK**CTGGGCTTTTGGGGTCAGGGAAC-3′
	5′-GTATATTATTGCGCGAGT**NNKDVKDVKDVKDVKDVKDVKDVKNNK**GGCTTTTGGGGTCAGGGAAC-3′
Garvan II	
CDR1	5′-CTCTCCTGTGCAGCCTCCGGATTTAGGTTT**KMTKMT**GAGGATATGGGCTGGGTCC-3′
CDR2	5′-GAAGGGTCTAGAGTGGGTATCA**KMT**ATT**KMTRRTSMT**GGT**KMT**AGCACATACTACGCAGAC-3′
CDR3	5′-CACCGCGGTATATTATTGCGCGAAA**XxxXxxXxxXxxXxxXxxXxx**TTTGACTACTGGGGGTCAGGGAACCCTG-3
	5′-CACCGCGGTATATTATTGCGCGAAA**XxxXxxXxxXxxXxxXxxXxxXxxXxx**TTTGACTACTGGGGGTCAGGGAACCCTG-3′

[$] Diversified positions in bold. Encoded amino acids (codons):

DVK 16.7% Ser, 11.1% Ala, Gly, Thr, 5.6% Arg, Asn, Asp, Cys, Glu, Lys, Ser, Trp and Tyr
KMT 25% A, D, S and Y
RRT 25% D, G, N and S
SMT 25% A, D, H and P
Xxx 19.7% Y(TAC), 16.7% G(GGT), 15.2% S(TCT), 6.6% A(GCT), 6.6% D(GAC), 3.9% I(ATC), L(CTG), P(CCG), R(CGT), T(ACT), V(GTT), 1.6%E(GAA), F(TTC), H(CAT), K(AAA), M(ATG), N(AAC), Q(CAG) and W(TGG); all TRIM codons

TABLE II. V_H H04 diffraction data, structure refinement and structural features

Diffraction Data		
PDB entry	**4PGJ**	**4U3X**
Spacegroup	I1 2 1	P1 2_1 1
Unit cell dimensions: a,b,c (Å); β (°)	100.1, 40.3, 137.0; *101.2*	68.1, 39.2, 100.2; *104.2*
Wavelength (Å)	0.9537	0.9537
Resolution range (Å)	49.08 - 2.60	97.14 - 2.26
Observed reflections[$]	117868	162966
Unique reflections[$]	16831	24456
Completeness[$]* (%)	99.3 (99.0)	98.2 (95.5)
Multiplicity[$]*	7.0 (7.1)	6.0 (5.8)
R_{merge} [$]*	0.152 (0.951)	0.188 (0.910)
Mean (I/sd) [$]*	9.0 (2.0)	7.8 (2.4)
Wilson B ($Å^2$)	45.1	25.3
Refinement		
Protein molecules/asu	2*VH (chains A, C) 2*HEL (chains B, D)	2*VH (chains A, C) 2*HEL (chains B, D)
Atoms modeled/asu	3380	3685
Ramachandran[#] - favored (%)	95.07	97.47
- outliers (%)	0	0
R	0.225	0.192
R_{free} (5% data)	0.287	0.248
RMSD bond lengths (Å)	0.0153	0.0156
RMSD bond angles (°)	1.5828	1.7087
Interface area[@]		
VH-HEL (chains A-B) ($Å^2$)	785 each chain	826 each chain
VH-HEL (chains C-D) ($Å^2$)	807 each chain	815 each chain

[$] As output by SCALA
* Values in parentheses are of the highest resolution shell
[#] As calculated by the MOLPROBITY validation server
[@] As calculated by PDBePISA

TABLE III. Structural features of single domains in complex with antigen

	Human V_H			**Camelid V_HH**	**Camelid V_HH**	**Camelid V_HH**	**Camelid V_HH**	**Shark VIg_{NAR}**
Framework	Native		Mutant	Native				
PDB ID[a] Clone	4U3X V_H H04	2VYR V_H9	3P9W V_H V1a	1ZVY D3L11	1RI8 D2L19	1ZV5 D2L29	1ZVH D2L24	1SQ2 5A7
CDR3 length (AA)	9	12	12	18	17	12	16	18
Additional disulfide bonds	0	0	0	0	1	1	0	2
Affinity (nM)	26	44	16	0.35	3	10	60	22
Hydrogen bonds[b]	21	7	13	17	19	15	9	15
Salt bridges[b]	3	0	2	2	1	1	0	2
Shape complementarity[c]	0.808	0.686	0.699	0.801	0.761	0.766	0.783	0.769
Buried surface area (Å^2)[b]								
Total	783.3	805.3	905.6	644.8	634.0	635.3	561.0	755.3
CDR1	20.2	74.6	9.8	70.8	0	72.7	69.9	161.9
CDR2	60.1	0	30.3	114.6	6.4	2.0	0	0
CDR3	535.4	613.0	498.3	395.7	469.0	268.6	449.9	584.9
Framework	167.6	117.7	367.2	63.7	158.6	292.0	41.2	28.5
Epitope surface	Cleft (HEL)	Planar (p53)	Planar (VEGF)	Cleft (HEL)	Cleft (HEL)	Cleft (HEL)	Planar (HEL)	Cleft (HEL)

[a]References:1ZVY, 1ZV5, 1ZVH (52) ; 1RI8 (53); 1SQ2 (54); 2VYR (61); 3P9W (21).
[b]Calculated by the PISA server (37)
[c]Calculated by the program SC (62)

FIGURE 1. A library of stable human V_H single domains derived from the HEL4 model domain (Garvan I) (A) Amino acid sequence of HEL4 model domain, DP47 germline and Garvan I library with numbering according to Kabat, with CDR1, CDR2 and CDR3 regions highlighted in green, pink and blue respectively. (B) Cartoon representation of human V_H single domain with CDR3 region highlighted. (C) Colloidal stability (aggregation resistance) of HEL4-DP47 chimeras, as measured by retained binding to superantigen after heating to 80ºC (mean, n=3). (D) Colloidal stability of Garvan I library, as measured by retained binding to superantigen after heating to 80ºC. (E) Selection of binders.

FIGURE 2. A stable human anti-TNF V_H single domain selected from the Garvan I library. A) Thermal unfolding (red) and refolding (blue) of clone G07 as monitored by circular dichroism. (B) Determination of G07 binding kinetics by surface plasmon resonance.

FIGURE 3. A second-generation library of stable human V_H domains (Garvan II). (A) Amino acid sequence of HEL4 model domain, DP47 germline and Garvan II library with numbering according to Kabat, with CDR1, CDR2 and CDR3 regions highlighted in green, pink and blue respectively. (B) Thermal unfolding (red) and refolding (blue) of HEL4, DP47, DP47 CDR1 double and triple mutants as monitored by circular dichroism. (C) Cartoon representation of human V_H single domain with CDR1, CDR2 and CDR3 regions highlighted D) Colloidal stability of Garvan II library, as measured by retained binding to superantigen after heating to 80ºC. E) Selection of binders.

FIGURE 4. Binding kinetics and epitope distribution of human V_H single domains selected from the Garvan II library against hen egg white lysozyme antigen (HEL). (A) Determination of clone H04 binding kinetics by surface plasmon resonance. (B) Kinetic association (k_a) and dissociation constants (k_d), equilibrium binding constants (K_D) and heat refoldability in solution of anti-lysozyme human single domains. (C) Cartoon representation of lysozyme antigen with cleft region (asterisk) and epitopes of D3L11, D2L19 and D2L24 camelid domains highlighted. (D) Epitope binning of camelid domain by surface plasmon resonance. (E) Epitope binning of human domain by surface plasmon resonance.

FIGURE 5. Structural basis of cleft recognition by the fully human V_H single domain H04. (A) Structure of the human V_H H04 in complex with hen egg white lysozyme (HEL) antigen (PDB identifier: 4U3X). Cartoon representations with antigen in tan, H04 in grey and CDR1,

CDR2 and CDR3 regions highlighted in green, pink and blue respectively. (B) Hydrogen bonding network with antigen residues in italics. (C) Framework contacts with lysozyme antigen. (D) Conformation of Trp47 residue in V_H/V_L interface, with H04 rendered in grey, HEL4 in orange and representative human Fabs in blue (PDB identifiers: 3QOS, 2VXS, 3KDM, 3BN9).

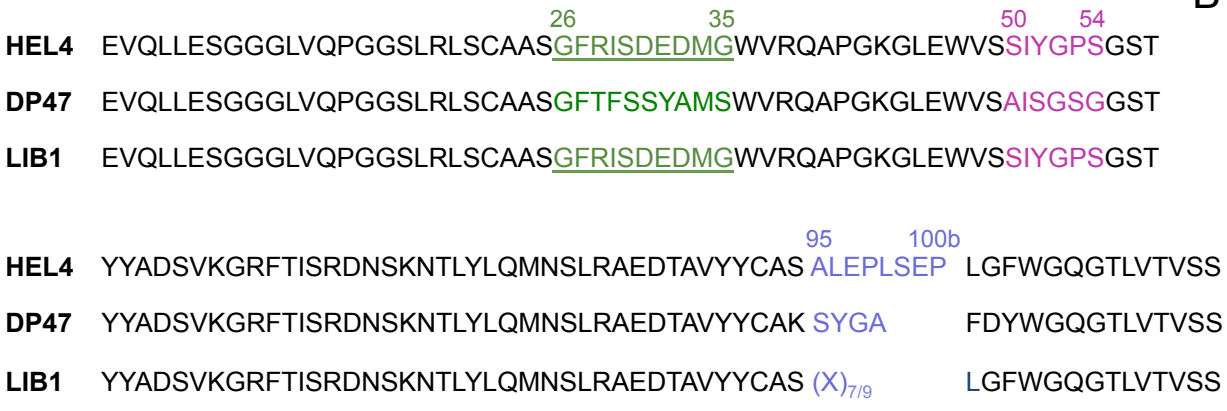
FIGURE 6. Antigen recognition by V_H H04 and other human V_H domains. (A) V_H H04 in complex with hen egg white lysozyme (HEL) antigen, with structural cleft indicated (asterisk) (PDB identifier: 4U3X). (B) V_H 9 in complex with MDM2 (PDB identifier: 2VYR). (C) V_H V1a in complex with vascular endothelial growth factor (VEGF) (PDB identifier: 3P9W).

FIGURE 7. Interaction of CDR3 with framework residues at positions 35, 37, 93 and 100x. (A) Camelid V_HH D2L24 (PDB identifier: 1ZVH). (B) Human V_H H04 (PDB identifier: 1ZVH). (C) Camelized V_H B1a (PDB identifier: 3P9W).

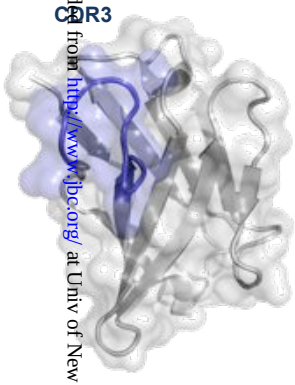
FIGURE 8. Structural features of the V_H/V_L interface. (A) Human V_H H04 (PDB identifier: 4U3X). B) Human DP47 Fab (PDB identifier: 3QOS). (C) Camelized V_H B1a (PDB identifier: 3P9W). (D) Camelid V_HH D2L24 (PDB identifier: 1ZVH).

Figure 1

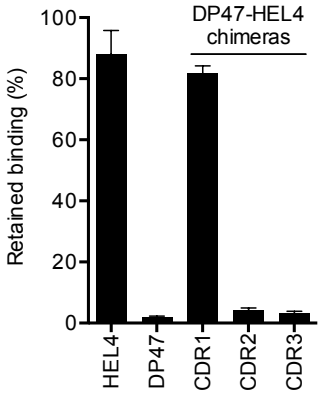
A



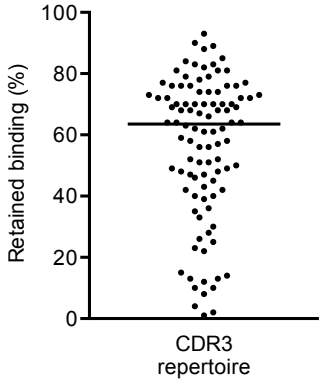
B



C



D



E

Antigen	Binders (phage)	Binders (soluble nr)
Human tumour necrosis factor (TNF-alpha)	121/ 564 (22%)	13/57 (23%)
Murine interleukin 21 (mIL21)	27/47 (57%)	3/5 (60%)
Human prolactin receptor (hPRLR)	23/47 (49%)	9/10 (90%)
Beta-galactosidase	14/47% (30%)	2/13 (15%)

Figure 2

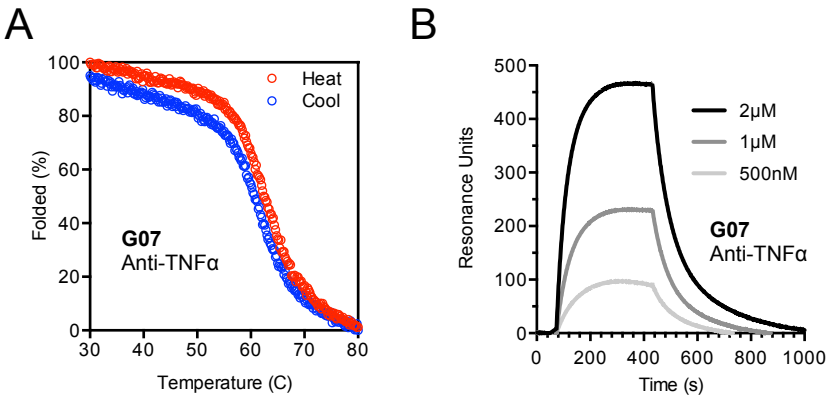
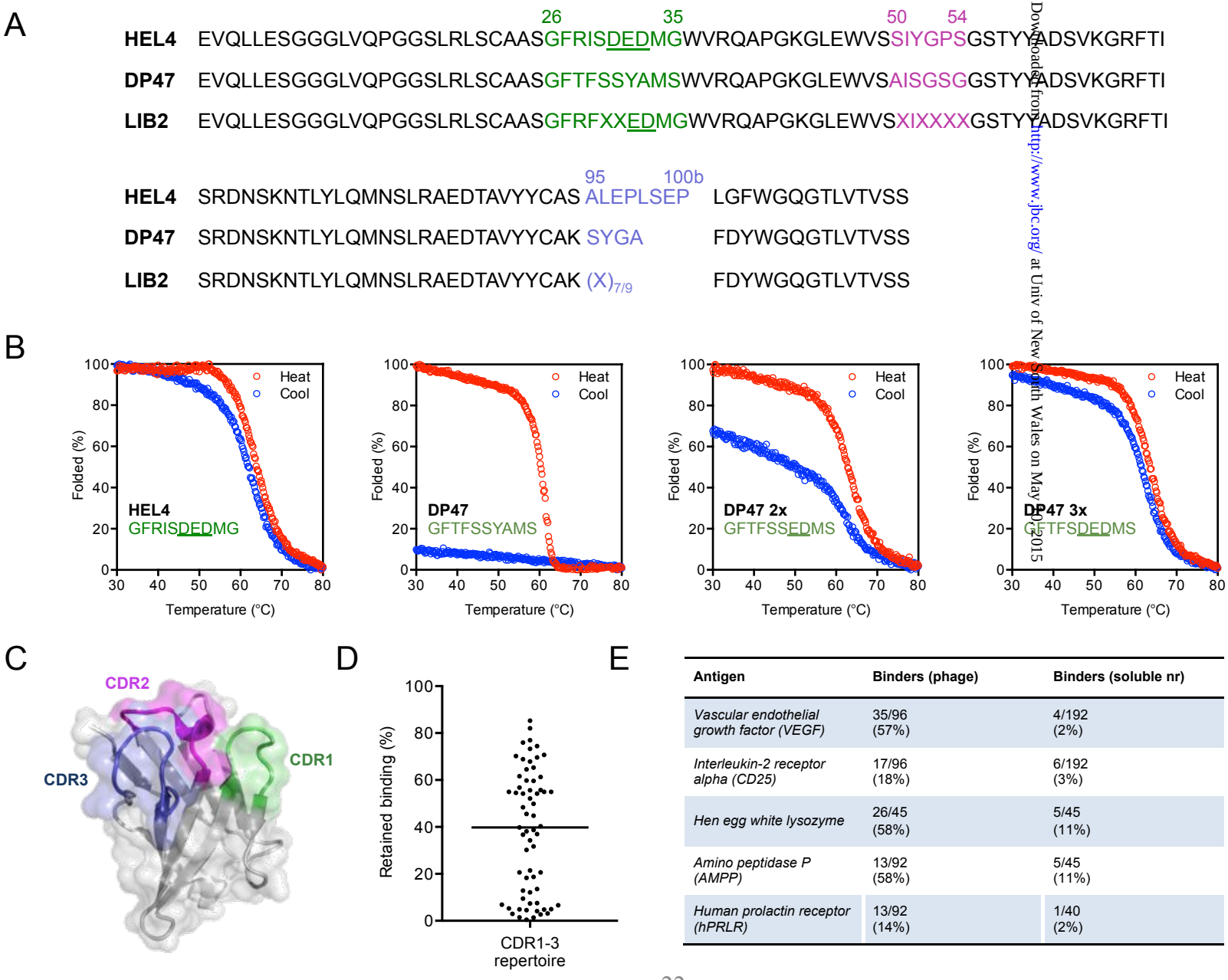


Figure 3

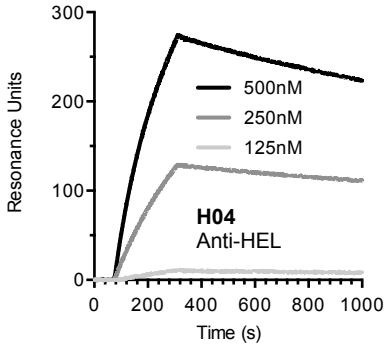


E

Antigen	Binders (phage)	Binders (soluble nr)
Vascular endothelial growth factor (VEGF)	35/96 (57%)	4/192 (2%)
Interleukin-2 receptor alpha (CD25)	17/96 (18%)	6/192 (3%)
Hen egg white lysozyme	26/45 (58%)	5/45 (11%)
Amino peptidase P (AMPP)	13/92 (58%)	5/45 (11%)
Human prolactin receptor (hPRLR)	13/92 (14%)	1/40 (2%)

Figure 4

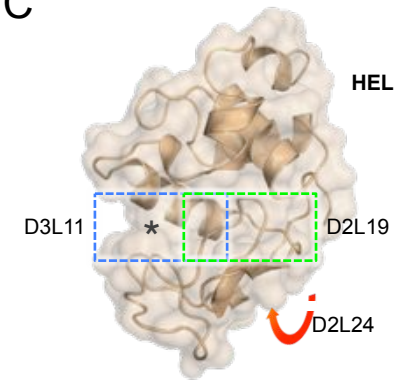
A



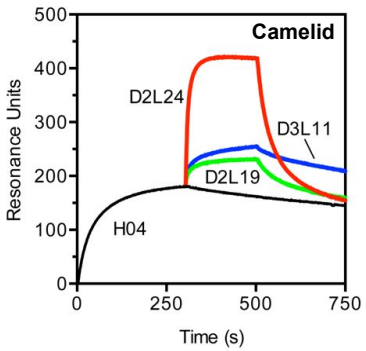
B

Clone	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_D (nM)	Heat refolding
H04	1.0×10^4	2.6×10^{-3}	26	111%
G08	1.9×10^4	1.1×10^{-3}	59	79%
F05	3.8×10^4	8.0×10^{-3}	210	85%
D05	2.1×10^4	5.4×10^{-3}	257	88%
C01	1.4×10^4	2.6×10^{-3}	191	121%

C



D



E

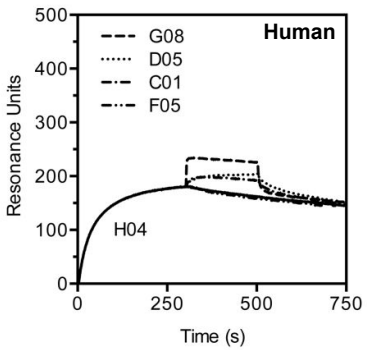


Figure 5

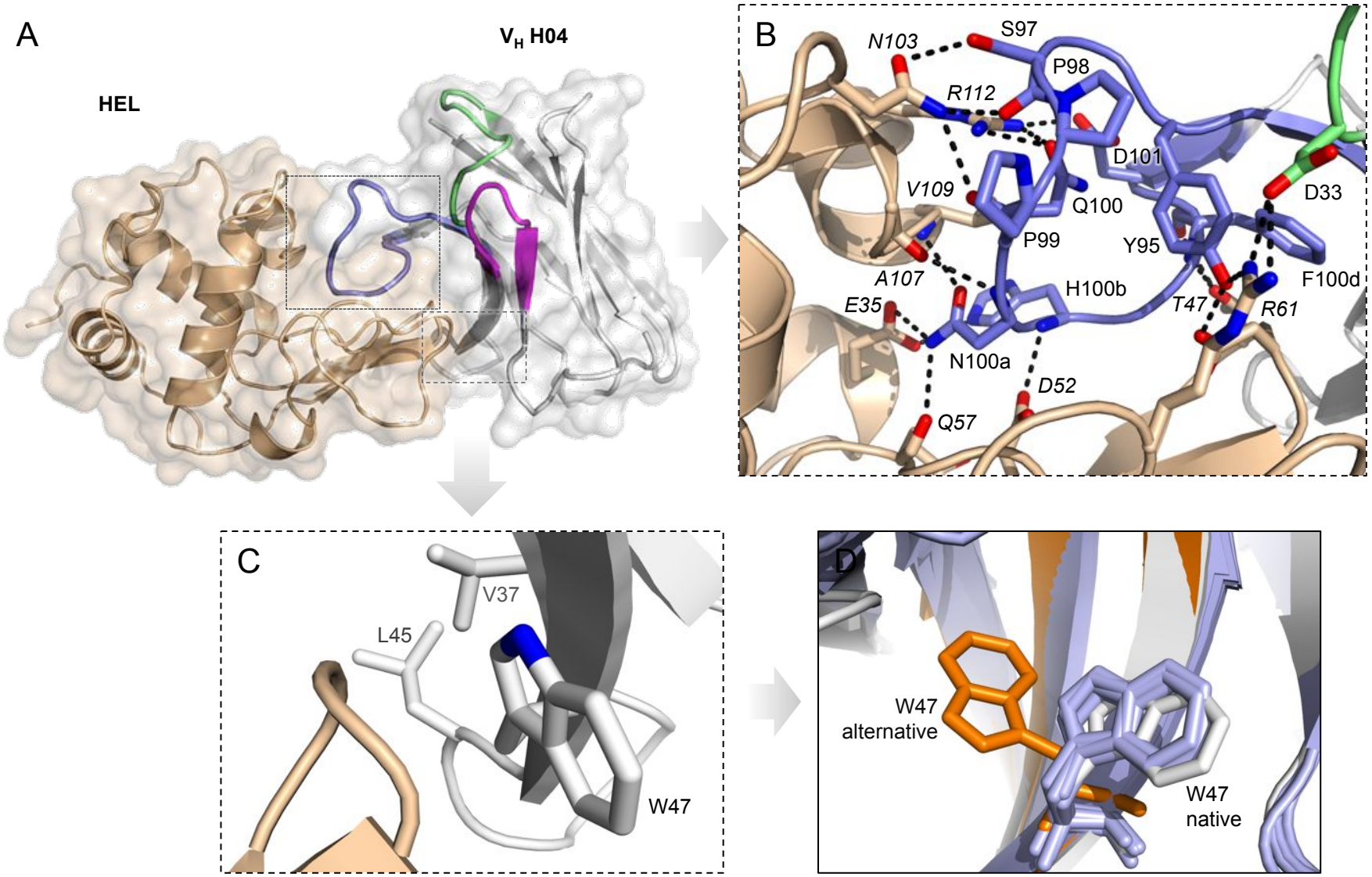

A
HEL
V_H H04
B
N103
S97
P98
R112
D101
V109
Q100
D33
P99
Y95
A107
H100b
F100d
E35
T47
R61
N100a
D52
Q57
C
V37
L45
W47
D
W47
alternative
W47
native

Figure 6

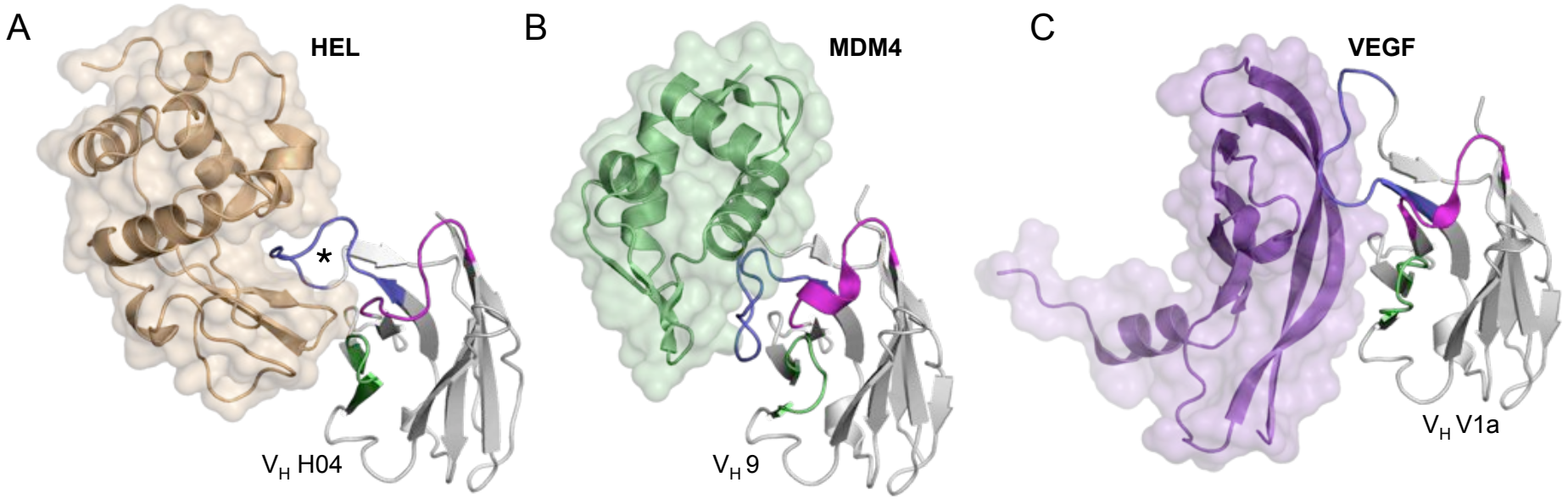


Figure 7

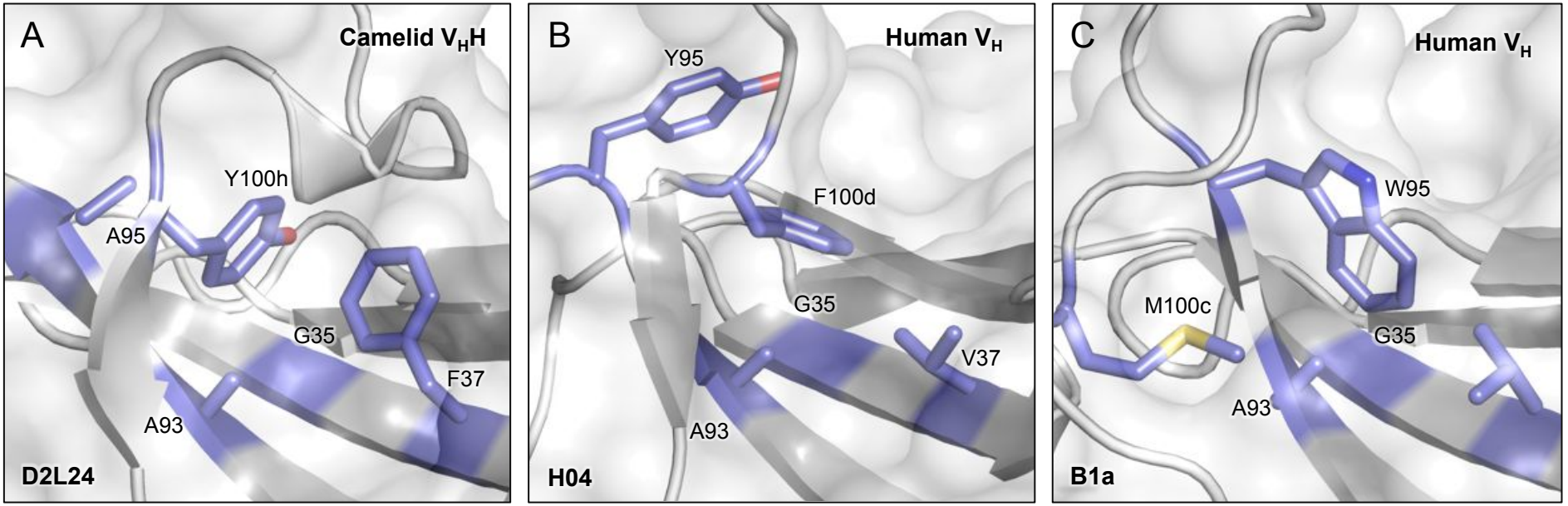
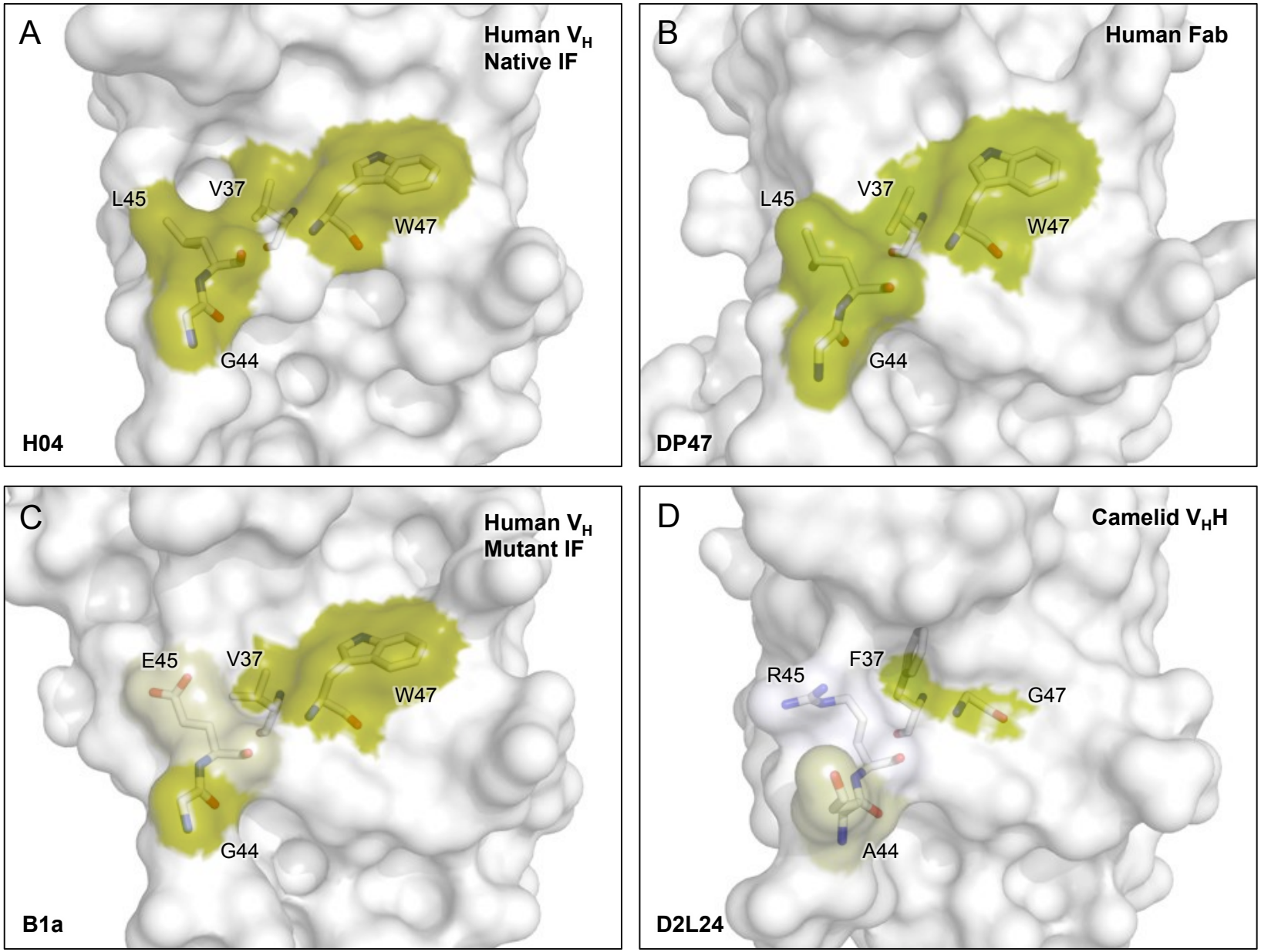
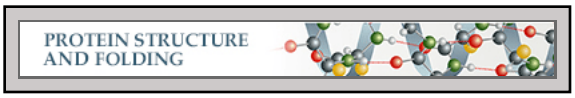

A
Camelid V_HH
Y100h
A95
G35
F37
A93
D2L24
B
Human V_H
Y95
F100d
G35
V37
A93
H04
C
Human V_H
W95
M100c
G35
A93
B1a

Figure 8


A
Human V_H
Native IF
L45
V37
W47
G44
H04
B
Human Fab
L45
V37
W47
G44
DP47
C
Human V_H
Mutant IF
E45
V37
W47
G44
B1a
D
Camelid V_HH
R45
F37
G47
A44
D2L24

Protein Structure and Folding:

Fully human VH single domains that rival the stability and cleft recognition of camelid antibodies



Romain Rouet, Kip Dudgeon, Mary Christie, David Langley and Daniel Christ

J. Biol. Chem. published online March 3, 2015

Access the most updated version of this article at doi: 10.1074/jbc.M114.614842

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2015/03/03/jbc.M114.614842.full.html#ref-list-1