Crystal Structure of a Human VH: Requirements for Maintaining a Monomeric Fragment[†,‡]

Tania Dottorini,[§,||] Cara K. Vaughan,[*,§,⊥] Martin A. Walsh,[#] Paola LoSurdo, and Maurizio Sollazzo

Istituto di Ricerche di Biologia Molecolare, Crystallography Unit, Via Pontina km 30600, 00040 Pomezia (Rm), Italy

Received October 5, 2003; Revised Manuscript Received November 14, 2003

ABSTRACT: The variable domain of dromedary immunoglobulins comprises only the heavy chain and is missing the light-chain variable domain. This single domain is sufficient for antigen recognition and binding—half that required by other mammals. Human antibody-VHs have previously been camelized to be soluble stable fragments that retain antigen binding. Such engineered V_HH are of interest in drug development, since they are nonimmunogenic, and in other biotechnology applications. We present the structure of a camelized human antibody fragment (cVH), which is a competitive and reversible inhibitor of the NS3 serine protease of the hepatitis C virus (HCV). In solution, this cVH undergoes a concentration-dependent monomer−dimer equilibrium. The structure confirms the minimum mutational requirements of the VL-binding face. The fragment also suggests a means by which the observed dimerization occurs, highlighting the importance of the composition of the CDR3 in maintaining a truly camelized VH.

Protein engineers can tailor proteins to bind chosen macromolecular ligands using combinatorial methods (making a library of variants), in particular, displaying proteins on the surface of phage. A successful strategy in designing a protein with affinity for a new target starts with choosing a small, stable, single domain protein that is easily expressed in *Escherichia coli*. This is followed by randomization of a patch of exposed (surface) residues, display of the resultant library on the surface of phage, and affinity-selection of this library against the target. This technique has selected variants typically with micromolar dissociation constants. This patch can be from any secondary structural element: examples include randomization of residues on a surface formed by a pair of α-helices in protein A (*1*, *2*); a β-sheet in knottins (*3*); and exposed loops in Tendamistat (*4*), the Minibody (*5*) (β-sheet proteins), and chytochrome b_{562} (a 4-helix bundle) (*6*). In the case of a protein A variant that was selected for binding to *Taq* DNA polymerase, further rounds of affinity maturation improved the binding by 2 orders of magnitude (*1*). Selection of variants from the same scaffold against different targets has shown that ligands cannot always be found (however, experimental considerations such as library size and stringency of selection conditions, as well as the suitability of the starting scaffold, may contribute to this) (*3*). These exercises test the boundaries of protein design and lay the foundations for a future in which we will be able to readily design proteins in vitro that meet the needs of industry and medicine.

Nature uses a parallel strategy when launching an immune response by randomization of the hypervariable loops in the variable domain of both the heavy and the light chain of antibodies. This allows binding with high specificity and low dissociation constants to virtually any antigen presented to them. Each Ig[1] V domain comprises two β-sheets packed at an angle on top of each other to form a sandwich, with the hypervariable loops at one end. Five of these loops, H1/L1, H2/L2, and L3, take one of the canonical structures defined by specific loop and framework interactions (*7*, *8*). No canonical structures have been identified for the sixth loop, H3, which can vary considerably in length and conformation; however, some general rules have been suggested (*9*).

A range of antibody fragments that effectively bind target molecules has been successfully designed (*10*−*12*); these are important clinically due to faster biodistribution and clearance of smaller proteins. The design of these fragments varies, but all have one thing in common—the presence of both VH and VL domains since studies on isolated variable domains of murine antibodies showed reduced binding as compared with the heterodimer and with propensity to aggregation (*13*).

The discovery that the immune response of dromedaries is mediated through only heavy-chain antibody units therefore opened an exciting avenue in immunoglobulin research (*14*). This discovery raised important questions regarding the minimal molecular requirements for raising an effective immune response and for mimicking this response in vitro (*15*).

[†] C.K.V. was supported by a Marie Curie Fellowship.

[‡] The atomic coordinates have been deposited. The PDB ID is 1OL0.

[*] Corresponding author. Tel.: +44-207-7906073. Fax: +44-207-7906055. E-mail: cvaughan@icr.ac.uk.

[§] These authors contributed equally to this work.

[||] Present address: Department of Internal Medicine, Section of Biochemistry, University of Perugia, 06100 Perugia-Italy.

[⊥] Present address: Section of Structural Biology, Institute of Cancer Research, 237 Fulham Rd., London, UK, SW3 6JB.

[#] Present address: MRC France c/o CRG BM14, ESRF, B. P.220, F-38043 Grenoble CEDEX, France.

[1] Abbreviations: Ig, immunoglobulin; VH, heavy-chain antibody variable domain; VL, light-chain antibody variable domain; V_HH, camelid heavy-chain antibody VH; FV, variable fragment; CDR, complementarity determining region; NS3, nonstructural protein 3; HCV, hepatitis C virus; RMSD, root-mean-square deviation; ASA, solvent accessible surface area; PBS, phosphate buffered saline.

10.1021/bi035800b CCC: $27.50
Published on Web 12/25/2003

Sequence comparisons identified key residues in VH on the VL-binding face (*16*). A camelized human VH (cVH), in which these residues were mutated, was found to be a folded and soluble monomer in the absence of its VL counterpart (*17*). Riechmann made a display library in which the hypervariable region CDR3 of the cVH was randomized. Variants with the desired properties could then be selected and amplified in cycles of affinity-selection. After several such rounds, variants with the best properties remain. These may be sufficient for the chosen purpose or may be subjected to further rounds of randomization and selection, thus improving their initial characteristics, a process of affinity maturation in vitro (*18*−*20*).

A few years ago, we used a library of ~10^8 cVH variants in which the CDR3 loop was randomized in length and sequence (*17*) to select variants that could inhibit the hepatitis C virus NS3 serine protease. We identified cVH-E2 to be a competitive and reversible inhibitor (K_i of 150 nM) of NS3 (*21*), making it a suitable target for drug discovery. The HCV NS3 serine protease is a crucial component in replication of the hepatitis C virus, necessary for cleavage of the viral polyprotein downstream of itself. Structural studies of the HCV NS3 serine protease revealed an unusually flat active site; thus, the search for small molecule competitive inhibitors has proved difficult. It was reasoned that a larger inhibitor binding surface may provide clues for designing active compounds (*22*).

This paper presents the crystal structure of cVH-E2. Comparisons of the structure are made with other know VH domain structures, and we propose a means by which protease inhibition and dimerization occurs.

MATERIALS AND METHODS

Construction of the pUC-based vector containing the cVH-E2 antibody fragment has been described previously (*21*).

Expression and Purification of cVH-E2. The cVH-E2 domain was expressed as a seleno-methionine derivative at 23 °C in the *E. coli* strain B834 and secreted into the periplasm to allow disulfide bond formation.

Bacterial cultures were grown at 37 °C in rich 2TY medium, containing ampicillin (50 mg/L) and 1% glucose. When the OD_{600} reached 0.8 cm^{-1}, cells were harvested by centrifugation, and the cell pellet was washed in PBS. The bacterial pellet was resuspended in minimal media containing ampicillin (50 mg/L) enriched with biotin (1 mg/L), thiamine (2.4 mg/L), and $FeSO_4/H_2O$ (4.8 mg/L). The cells were returned to 37 °C, and protein expression was induced with 0.4 mM IPTG in the presence of SIGMA D/L seleno-methionine (50 mg/L). Five hours after induction, the cells were harvested. Protein extraction and purification proceeded as described previously (*21*).

Protein purity was checked by nonreducing SDS-15% PAGE. Protein concentrations were determined as described by Gill and von Hippel (*23*). Typically, 2 mg of pure protein is obtained per liter of *E. coli* culture. Incorporation of seleno-methionine and the exact molecular weight (15 770 Da) were confirmed by mass spectrometry experiments.

Crystallization and Data Collection. Crystallization screens were carried out using the hanging-drop vapor diffusion technique. Small crystals initially grew from 4 μL hanging drops formed by mixing 2 μL of protein at a concentration of 5 mg/mL with 2 μL of reservoir solution containing 15−20% (w/v) PEG 8000, 0.4 M $(NH_4)_2SO_4$, and 0.1 M HEPES, pH 7.5. Seeding of these initial crystals led to long needlelike diffraction quality crystals with typical dimensions of approximately 50 × 50 × 200 μm^3. The crystals show hexagonal symmetry. The unit cell dimensions are $a = b =$ 75.3 Å, c = 101.78 Å, $\alpha = \beta = 90°$, and $\gamma = 120°$. For X-ray diffraction data collection, crystals were mounted in a nylon loop, transferred to a stabilizing buffer containing 40% PEG 8000 as cryoprotectant, and flash frozen in liquid nitrogen. Data were collected using synchrotron radiation from beamline ID14-EH2 at the European Synchrotron Radiation Facility, with an ADSC Quantum-4R CCD detector. All data were reduced using the HKL2000 suite (*24*). A summary of the data collection statistics is shown in Table 1.

Table 1: Crystallographic Data Collection and Refinement Statistics

unit cell parameters (Å)	$a = b = 75.30$, $c = 101.78$, $\alpha = \beta = 90°$, $\gamma = 120°$
resolution range (Å)	25.2−1.8
no. of reflections measured	69437
no. of unique reflections	29910
completeness (%)	94.9 (71.5)[a]
R_{merge} (%)[b]	5.1 (22.7)
$\{\langle I\rangle\}/\{\langle \sigma I\rangle\}$	17.2 (1.5)
B_{wilson} (Å^2)	17.34
R-factor (%)[c]	15.58
free R-factor (%)[d]	18.02
no. of protein atoms	1818
no. of solvent atoms	388
rms deviations from ideal geometry	
bond distance (1−2) (Å)	0.017
angle distance (1−3)	0.038
planar distance (1−4)	0.044
chiral volumes (Å^3)	0.096
planar torsion angles (deg)	6.3
deviations from planarity (peptides) (Å)	0.03
deviations from planarity (aromatics) (Å)	0.01
mean B-factors (Å^2)	
all atoms	23.69
protein atoms	20.34
glycerol (2)	41.77
sulfate ions	49.14
water molecules	37.71
Ramachandran plot statistics	
residues in most favored regions (%)	93.6
residues in additional allowed regions (%)	6.4
residues in disallowed regions (%)	0.01

[a] Values in parentheses are for the highest resolution shell (1.84−1.80 Å). [b] $R_{merge} = \sum_{hkl}\sum_{i=1}^{N} |\langle I^{hkl}\rangle - I_i^{hkl}|/\sum_{hkl}\sum_{i=1}^{N} I_i^{hkl}$. [c] $R = 100\sum_h||F_o| - |F_c||/\sum_h|F_o|$. [d] Free R-factor is calculated from 5% of the data that was omitted during course of the refinement.

Structure Solution and Refinement. The structure of the VH domain was determined by molecular replacement using *AMoRE* (*25*). The search model used was the heavy-chain variable domain from Lama Glama (PDB accession code 1HCV), which has a sequence identity of 69%. Calculation of the translation function in the two enantiomorphic space group settings allowed the unambiguous assignment of the space group as $P3_221$. The initial molecular replacement phases were improved using the MOLREP option in ARP/wARP (*26*). Refinement of the model was carried out with

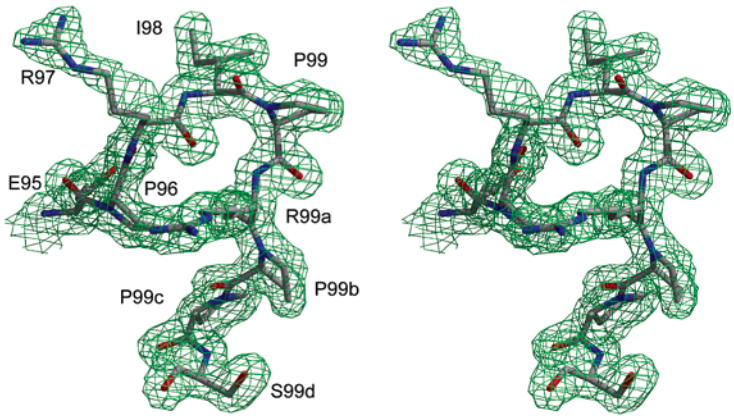

FIGURE 1: Stereo diagram of the $2F_o - F_c$ electron density map of CDR3. Residues 95−99d of the CDR3 are shown. The map is contoured at 1σ. Drawn with BOBSCRIPT and rendered using Raster3D (*54*).

the program REFMAC (*27*) within the CCP4 suite (*28*) using the restrained maximum likelihood method with a bulk solvent correction and overall anisotropic scaling. Water molecules were located automatically with the program ARP (*29*). Difference electron density maps with coefficients $2m|F_o| - D|F_c|$ and $m|F_o| - D|F_c|$ (output weights generated from REFMAC) were used to guide manual fitting of the models in the program O (*30*). During model building and refinement, 5% of the data was flagged for cross-validation to monitor the progress of refinement (*31*). In the final stages of the refinement, all data were included. Details of the refinement statistics are shown in Table 1.

RESULTS

Quality of the cVH-E2 Model and Description of the Structure. The refined model for cVH-E2 consists of 121 residues and 399 water molecules giving an *R*-factor of 15.6% for all data between 30 and 1.8 Å resolution and an R_{free} of 18.0%. The quality of the final model can be assessed from the refinement statistics presented in Table 1. Thr20 is located in a disallowed region of the Ramachandran plot ($\phi = -75°$, $\Psi = 95°$) (*32*) but is located in well-defined density. A representative view of the electron density is shown in Figure 1.

The overall shape of the molecule is displayed as a ribbon diagram in Figure 2. The architecture of the cVH-E2 is the well-known immunoglobulin V domain fold composed of a sandwich of two β-sheets with five inner and four outer strands (*7*). The nomenclature used for describing the structure is the Kabat numbering: residues 3−12 (strand A), 17−25 (strand B), 32−40 (strand C), 44−52 (strand C′), 56−60 (strand C″), 66−73 (strand D), 76−82b (strand E), 87−96 (strand F), and 102−112 (strand G).

cVH-E2 is a dimer in the crystal (Figure 2), and the structure presents a novel subunit arrangement that, we believe, has not been previously observed for Ig V domains. Dimerization occurs in solution at protein concentrations >7 mg mL^{-1} as demonstrated by analytical ultracentrifugation and dynamic light scattering data (not shown).

Comparison of cVH-E2, VH-P8, and FV-POT. To detect features characteristic of the affinity-selected cVH-E2, the crystal structure was compared to the NMR structure of the parent molecule VH-P8 (*33*) (PDB code 1VHP) and to the VH domain of the human FV-POT (*34*) (PDB code 1IGM) since this provides a good model of VH-VL associated

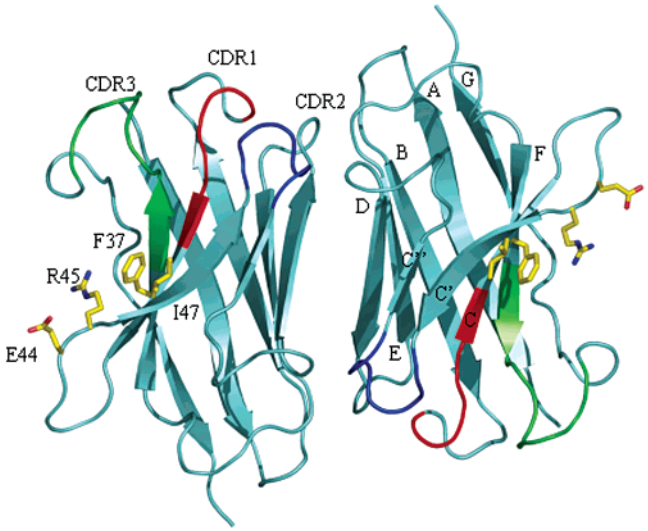

FIGURE 2: Ribbon diagram of the cVH-E2 dimer. Loops that contain all or part of CDR1, 2, and 3 are red, blue, and green, respectively. Camelizing mutations found on the former VL-binding face are shown in ball-and-stick. The β-strands are labeled according to the Kabat nomenclature. Drawn with PyMOL (*55*).

domains with highest sequence identity to cVH-E2. The FV-POT and the camelized VH-P8 originate from the same human VH (subgroup III). VH-P8 and cVH-E2 share a 93% sequence homology as compared with a value of 74% for cVH-E2 and VH-POT (Figure 3a).

The overall folds of cVH-E2, VH-POT, and VH-P8 are essentially the same. The rms deviation for the superposition of cVH-E2 with VH-P8 is 2.11 Å and with VH-POT is 0.62 Å (88 and 114 Cα atoms, respectively, with a deviation of less than 3.5 Å, included in the alignment). The aligned β-sheets are virtually identical, showing that the selected loop, despite its unusual proline-rich content, has not induced any significant change in the framework structure.

The most notable difference in the cVH-E2 crystal structure when compared with other Ig V domains lies in the arrangement of the dimer, in which the subunits of the homodimer are packed in a head to tail fashion (Figure 2). The antigen binding loops of cVH-E2 are displayed in an antiparallel fashion that is contrary to the arrangement of the VL-VH heterodimer in an FV fragment in which both variable domains point in the same direction. Instead, the arrangement corresponds to a quasi-heterodimeric FV structure in which one of the two domains has been rotated 180°

a

```
      1         10        20        30        40        50 a53
cVHE2 QVQLVESGGGLVQPGGSLRLSCAASGFTFS SYAMS WFRQAPGKEREIVS AVSGSGGSTYY
VHP8  EVQLVESGGGLVQPGGSLRLSCAASGFTFS SYAMS WVRQAPGKEREIVS AVSGSGGSTYY
VHPOT EVHLLESGGNLVQPGGSLRLSCAASGFTFN IFVMS WVRQAPGKGLEWVS GVFGSGGNTDY
      :*:*:****.******************.  :.** *.******  * **  .* ****.* *
                                        cdr1                  cdr2

      60        70        80 abc83     90         abcd100
cVHE2 ADSVR GRFTISRDNSKNTLYLQMNSLRAEDTAVYY CAREPRIPRPPSFDYW GQGTLVTVS
VHP8  ADSVK GRFTISRDNSKNTLYLQMNSLRAEDTAVYY CAR----LKKYAFDYW GQGTLVTVS
VHPOT ADAVK GRFTITRDNSKNTLYLQMNSLRAEDTAIYY CAKHRVSYVLTGFDSW GQGTLVTVS
      **:*: *****:**********************:** **:          .** * *********
      cdr2                                      cdr3

      113
cVHE2 S--------
VHP8  S--------
VHPOT SGSASAPTL
      *.::::.:
```

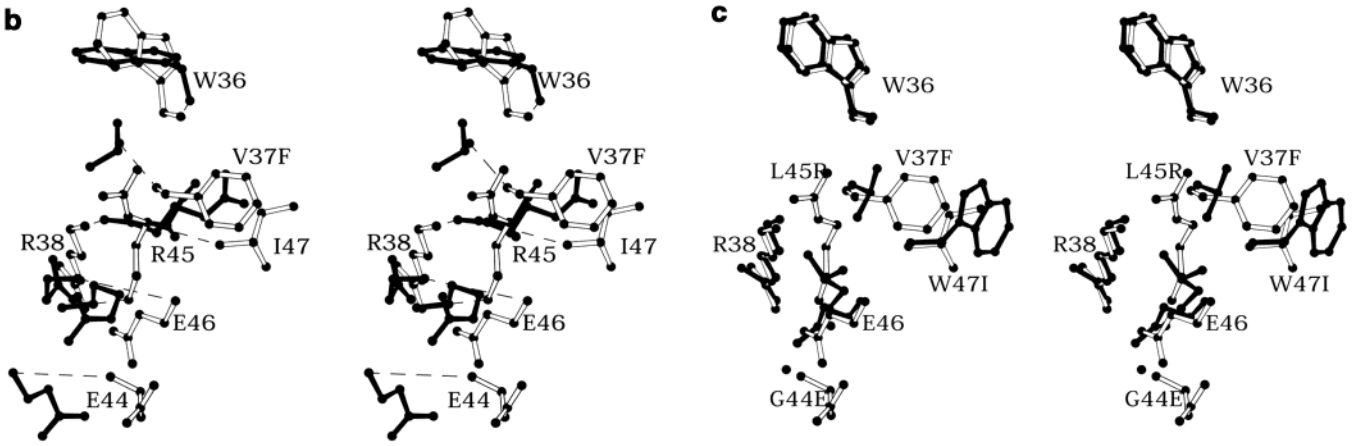

FIGURE 3: (a) Sequence homology for cVH-E2, VH-P8, and VH-POT. Residues are numbered according to Kabat. Points indicate different residues, and double points indicate different residues with the same polarity. The CDRs are boxed. The residues selected from the phage-displayed library are highlighted in bold. Consensus camelizing mutations are highlighted in gray. (b) Stereoview of the side chains in the camelized interface in a least squares superposition of the Cα backbone of cVH-E2 (white) with (i) VH-P8 (black) and (ii) FV-POT (black). Where a large shift in backbone position is observed, the two corresponding Cα atoms are linked by a dashed line. Drawn with BOBSCRIPT (*54*).

with respect to the other so that none of the three CDRs participates in interdomain contacts. The hydrogen bonds stabilizing the new VH-VH interface are discussed next.

Former VL Interface. The VL-binding face is usually formed by the β-strands F and G, including part of the CDR3 loop, and β-strands C and C′. In cVH-E2, this face contains the camelizing amino acid substitutions Val37 → Phe, Gly44 → Glu, Leu45 → Arg, and Trp47 → Ile (*35*, *36*) (Figure 2). The residues Phe37, Glu44, Arg45, and Gly47 are those typically found in camelid V_HH domains. However, Riechmann found that the mutation Gly47 → Ile resulted in a more stable monomer with better expression (*37*). The sum of these camelizing mutations render the single cVH domain less sticky than native VH domains deprived of their VL.

Although all but one (Val37 → Phe) of these camelizing mutations is identical in VH-P8 and cVH-E2, their adopted conformations vary considerably (Figure 3b). In cVH-E2, these residues bear greater similarity to the noncamelized interface of VH-POT (Figure 3c). In cVH-E2, the conformation of these residues is explained by a series of packing interactions. The positively charged guanidinium group of Arg45 is sandwiched between the π-electron clouds of Tyr95 and Trp103. Ile47, Phe37, and the edge of Trp103 pack together in a hydrophobic interaction. The stacking of Trp103 with Phe37 is observed in all camelid V_HH solved to date (*38*−*42*).

The latter interaction is not observed in VH-P8. Instead, a backbone rearrangement to bury the hydrophobic residues Trp103 and Val37 means that these side chains are no longer available for the above packing interactions. This rearrangement causes a bulge in β-strand G and a kink in strand C (*33*). In cVH-E2, however, β-strands G and C are identical to those of VH-POT.

Hypervariable Regions. The CDR1 and CDR2 loops of cVH-E2 are largely homologous with those of FV-POT and VH-P8 (Figure 3a). As expected, the structures are also similar; however, there are some notable differences.

The first hypervariable region contains amino acids 27−35, which form an exposed loop (H1) encompassing the less variable residues 26−30. The exposed loop around CDR1 adopts a type I canonical structure according to Chothia et al. (*8*, *43*). This structure is determined by a sharp turn around Gly26 and by the conformation of residues Ala24, Phe27, Phe29, and Met34. In cVH-E2, this is more closely related to the canonical CDR1 of VH-POT than the parent molecule VH-P8. This is a direct consequence of the buried position of Trp103 in VH-P8, discussed previously. In VH-P8, Trp103 packs against residues that determine the CDR1 structure, namely, Phe27 and Met34, which in turn cause a new orientation for the side chain of Phe29.

The CDR2 (amino acids between Kabat position 50 and 65) forms a type 3 canonical structure for the H2 region between residues 52−56 (*43*). The backbone conformation of Gly52a differs from the structure of the parent VH-P8, in which it is projected toward amino acid 56. Instead, this residue in cVH-E2 points in the opposite direction, as observed in FV-POT.

The identification of canonical structures for the third hypervariable domain of immunoglobulins is difficult considering the high variability in length and sequence. Furthermore, randomisation in vitro, as carried out for cVH-E2, may result in conformations not normally observed in vivo. The CDR3 of cVH-E2 has very well-defined electron density for both the main-chain and the side-chain atoms (Figure 1). The H3 loop (Kabat residues 95−100) of cVH-E2 presents a very different conformation to that of VH-P8 and FV-POT.

The high proline content of the affinity-selected loop and several hydrogen bonds contribute to a rigid CDR3. Within CDR3, two hydrogen bonds maintain the structure (Arg97 O−Arg100 N, Pro100a O−Phe100 N). In addition, at either end of the CDR3, a main-chain−main-chain hydrogen bond is made (Glu95 N−Ala33 O and Pro100b O−Asp101 N). A side-chain−side-chain hydrogen bond is also formed between Tyr32 (OH) and amine group of Arg101. Several crystal contacts also stabilize the position of the affinity selected loop, particularly the packing of residues Pro99 and Pro99b with residues of β-strands C′ and C″ of a neighboring molecule. None of these contacts is direct hydrogen bonds.

VH-VH Interface (Dimerization Interface). The VH-VH interface is formed by the secondary structural units encompassing β-strands D and E and their connecting loop, the hypervariable region H2, and the loop between strands β-A and β-B. The interface is highly charged (Figure 4). Points of direct hydrogen bonding between the monomers occur at the top and bottom of the interface. Stabilizing interactions at the center of the interface occur through water-mediated hydrogen bonds. A total of 29 water molecules are present at the interface with 20 having solvent-accessible surface areas of less than 10 Å^2.

There are 10 direct hydrogen bonds between the monomers, five at either end of the interface involving Ser17, Gly55, Gly66, Arg72, Asp73, Asn74, Ser75, Gln82, and Asn84 (Table 2). Of these, six are main-chain−side-chain interactions, two are side-chain−side-chain, and two are main-chain−main-chain interactions. The remainder of the interface is stabilized by a large number of hydrogen bonds involving water molecules. Twenty-five hydrogen bonds are mediated by one water molecule, and four are mediated by two water molecules.

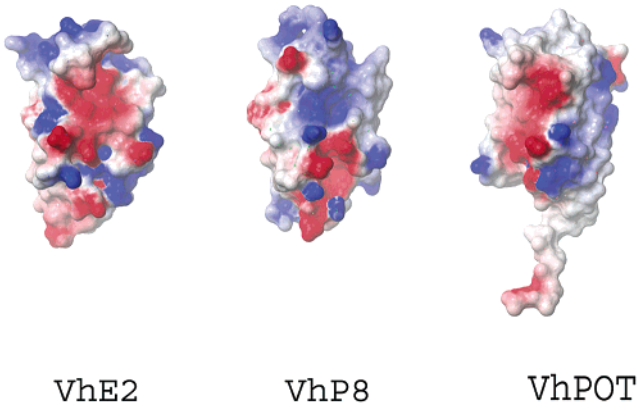

FIGURE 4: Surface of the VL-binding face of (a) cVH-E2, (b) VH-P8, and (c) VH-POT. The surfaces are colored according to electrostatic potential using conventional colors of blue (basic), red (acidic), and white (nonpolar). The potentials were calculated, and the figures were drawn in MolMol (*56*).

Table 2: Hydrogen Bonds Made between the Crystallographic CVH-E2 Dimer[a]

		dist. in chain A (Å)	angle in chain A (deg)	dist. in chain B (Å)	angle in chain B (deg)	inter-action
17 Ser N	75 Ser OG	2.85	130.1	3.05	149.5	mc−sc
17 Ser OG	74 Asn N	3.15	131.9	3.30	105.1	sc−mc
55 Gly O	66 Gly O	3.34	143.7	3.26	93.6	mc−-mc
72 Arg O	84 Asn OD1	2.97	108.6	3.01	143.7	mc−sc
73 Asp OD2	82 Gln NE2	3.25	155.7	3.24	91.0	sc−sc

[a] Distances and angles calculated in Contacts, CPP4 version 4.2.(*28*).

DISCUSSION

We report the X-ray structure of a camelized, human, heavy-chain variable domain that was affinity-selected from a phage-displayed library in which the CDR3 was randomized. The cVH-E2 variant was selected for competitive inhibition of the serine protease activity of HCV NS3 and has an IC_{50} of 300 nM and a dissociation constant of 150 nM (*21*).

The framework structure is, as expected, identical to the native immunoglobulin. The cVH fragment, however, adopts a dimeric conformation not previously observed in human or mouse heavy variable domains in solution. Heavy-chain homodimers are formed in which the three hypervariable regions of each VH domain protrude in opposite directions toward the solvent. Unlike a typical immunoglobulin VH-VL heterodimer, there are therefore no intermolecular interactions made by the complementarity determining regions.

Differences in the structure of cVH-E2 as compared with the parent molecule, VH-P8, are located mainly at the former VL-binding interface. The camelizing mutations here behave differently from those reported in the NMR structure. The Glu44 and Arg45 side chains in cVH-E2 are solvent exposed, but Arg45 shows the same orientation as Leu45 of FV-POT, and Glu44 is pointing toward the VL interface, mimicking a VH-VL interface, as normally present in FV fragments.

The in−out flipping of Val37 and Arg38, reported for VH-P8, is not observed for cVH-E2. In VH-P8, this flipping causes Val37 to be buried in the domain core and exposes Arg38 to solvent, creating a more hydrophilic surface for the VL-free domain. In FV-POT and other VL-associated VH domains, the Val37 side chain is involved in the formation of a hydrophobic pocket that accommodates the side chain of the highly conserved amino acid VL-Phe98. Sequence comparison shows, however, that Val37 is typically mutated to phenylalanine or tyrosine in camelid antibodies (*44*). In cVH-E2, the additional camelizing mutation Val37 → Phe is present. Surprisingly, this bulky hydrophobic side chain is oriented toward the solvent, while its neighbor Arg38 is buried, as observed in FV-POT. This orientation is also observed in all crystal structures of camelid V_HH solved to date (*38*−*42*, *45*, *46*). The mutation, in which Phe37 exposes only the edge of its side chain to solvent, allows Trp103 to reorient itself and pack against Phe37 so that its hydrophilic NE nitrogen is also solvent exposed. Thus, when considered together, the effect on accessible surface area is favorable.

These results confirm the hypothesis by Desmyter et al. (*38*) in their report of the structure of a camelid V_HH in complex with lysozyme. They compared the NMR structure of VH-P8 with their camel antibody and suggested that the mutation of Val37 → Phe in VH-P8 may undo the observed in−out flipping and return these residues to their conventional orientation.

Therefore, in addition to the mutations Gly44 → Glu, Leu45 → Arg, and Tyr47 → Ile/Gly, mutation of residue 37 to Phe (and possibly also Tyr) is necessary to maintain true structural integrity of the camelized domain.

cVH Homodimer. Previous gel filtration analysis has shown that cVH-E2 migrates as an apparent dimer in solution (*47*). In light of the crystallographic results, dynamic light scattering and analytical ultra-centrifugation experiments were performed to ascertain the extent of dimerization in solution. The results of these experiments (not shown) show that at concentrations below 7 mg mL^{-1} (the concentration at which crystallization occurs), the protein is a monomer. At considerably higher protein concentrations of more than 20 mg mL^{-1}, the sample behaves as a monodisperse dimer and reveals that a concentration-dependent dimerization occurs in the low millimolar range.

The total ASA buried in the observed homodimer is 1405 $Å^2$, 707 $Å^2$ from molecule A and 698 $Å^2$ from molecule B. This monomer−monomer interface buries considerably more ASA than other interfaces formed by crystal contacts of the asymmetric unit. (The next largest interface buries 270 $Å^2$ less ASA.) These values are within the range found to be commonly observed at protein−protein interfaces in which association involves only small changes of conformation (*48*). In addition, the number of hydrogen bonds (both direct and water-mediated) lies within the values found for protein complexes (*49*). The percentage of polar surface area buried at the interface (determined on a side-chain basis) is unusually high; polar residues contribute 82% to the total buried ASA. For an average interface, the nonpolar character is generally similar to that of a normal protein surface (*48*); however, a higher percentage of polar character is frequently observed at antibody−antigen complexes (*49*). For interfaces with a high percentage of buried polar residues, water molecules are generally found distributed throughout the interface (*49*), and this is in agreement with the cVH-E2 homodimer.

cVH-E2 was affinity selected from a library of camelized VH molecules in which the CDR3 is randomized. Of those molecules with the desired binding characteristics, cVH-E2 is the only variant for which dimerization in solution is observed. This implies, however, that dimerization is a consequence of the selected CDR3 loop. In the crystallographic dimer, the CDR3 loops do not interact but are at opposite sides of the molecule from the dimer interface. This may suggest that the crystallographic dimer is a result of crystal packing rather than a reflection of the dimer observed in solution. Crystal contacts generated by 2-fold symmetry, as observed for cVH-E2, can bury between 1500 and 2500 $Å^2$ (*50*), which is similar to that observed here (but this range is similar for true recognition sites between proteins). The results raised the possibility that the dimer interaction observed in solution could be mediated by the camelized interface in a pseudo-VH-VL interaction. Using the VH-VL dimer of FV-POT as a template, we modeled a classical VH-VL dimer by least-squares fitting two cVH-E2 monomers on VH- and VL-POT.

As described earlier, the main difference between the backbone of the cVH-E2 and the VH-POT lies in the conformation of the CDR3; therefore, one might anticipate that any structural clashes in our modeled VH-VL dimer would occur here. Surprisingly, the backbones of the affinity selected CDR3s pack perfectly together. Furthermore, the principal residues from VL-POT that contribute to the VH-VL interface (Gln38, Gly42, Ala43, Leu46, and Tyr49) (*34*) lie in the same orientation as the equivalent residues in the pseudo-VL component in our modeled cVH-E2 dimer (Gln39, Lys43, Ile47, and Tyr59).

However, examination of side-chain positions shows that there are structural clashes between the two molecules in our modeled classical dimer (25 distances less that 2.6 Å belonging to three residues from β-strands C and C′ and six residues from β-strands F, G, and CDR3).

Recently, the solution structure of a llama V_HH was determined by NMR (*46*) in which all the loops of the domain were highly mobile. The rigid nature of our crystal structure may not accurately portray the range of possible cVH-E2 CDR3 orientations available, suggesting that any clashes in our modeled dimer could be assuaged in solution. However, since there are steric clashes at other interfacial positions outside of the CDR3 loop, as described previously for β-strands C and C′, CDR3 flexibility may not be sufficient to allow a sterically favorably VH-VL-type dimer to occur. Furthermore, comparison of the electrostatic surface representation of the VL-binding face of cVH-E2 and VH-POT (Figure 4) shows that there are notable differences suggesting that there would be no charge complementarity in our modeled VH-VL-type dimer. It is also notable, however, that there are large differences between cVH-E2 and its parent molecule VH-P8 (Figure 4).

Recombinant immunoglobulin and T-cell antigen receptor V domains have been shown to form dimers in solution. In the crystal structure of the variable domain homodimer of the antiphosphatidylcholine Fv fragment, a dimeric structure is also seen (*51*). In this case, the VL monomers are associated in an Fv-like fashion, facing each other through a hydrophobic association similar to the VH-VL interface.

A similar association is reported for the Vα-Vα dimer of the T-cell antigen receptor (*52*), where one of the two Vα domains functions as V*β*, thereby mimicking a Vα*β* dimer (*53*).

The arguments for and against the crystallographic dimer being a representation of the cVH-E2 dimer in solution are inconclusive. There may be another mode of nonclassical dimerization available to cVH-E2 that involves an interaction between the CDR3 loops, but on which, as yet, we have no further structural information.

The X-ray structure of cVH-E2 confirms that camelization has achieved its scope of avoiding dimerization at the traditional VL-binding interface. We believe that the nature of the CDR3 may be crucial for the success of camelization, particularly at high concentrations, since the affinity-selected, proline-rich CDR3 of cVH-E2 results in a concentration dependent dimerization.

ACKNOWLEDGMENT

We thank the European Synchrotron Radiation Facility (ESRF), Grenoble, for synchrotron beamtime at ID14-2 and Stefania di Marco and Cinzia Volpari for help with data collection.

REFERENCES

1. Gunneriusson, E., Nord, K., Uhlen, M., and Nygren, P. (1999) *Protein Eng. 12*, 873−878.
2. Nord, K., Nilsson, J., Nilsson, B., Uhlen, M., and Nygren, P. (1995) *Protein Eng. 8*, 601−608.
3. Smith, G., Patel, S., Windass, J., Thornton, J., and Winter, G. (1998) *J. Mol. Biol. 277*, 317−332.
4. McConnell, S., and Hoess, R. (1995) *J. Mol. Biol. 250*, 460−470.
5. Martin, F., Toniatti, C., Salvati, A., Ciliberto, G., Cortese, R., and Sollazzo, M. (1996) *J. Mol. Biol. 255*, 86−97.
6. Ku, J., and Schultz, P. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6552−6556.
7. Chothia, C., and Lesk, A. (1987) *J. Mol. Biol. 196*, 901−918.
8. Chothia, C., Lesk, A., Tramontano, A., Levitt, M., Smith-Gill, S., Air, G., Sheriff, S., Padlan, E., Davies, D., Tulips, W., Colman, P., Spinelli, S., Alzari, P., and Poljak, J. (1989) *Nature 342*, 877−883.
9. Morea, V., Tramontano, A., Chothia, C., and Lesk, A. (1998) *J. Mol. Biol. 275*, 269−294.
10. Pluckthun, A., and Pack, P. (1997) *Immunotechnology 3*, 83−105.
11. Poljak, R. J. (1994) *Structure 2*, 1121−1123.
12. Kortt, A. A., Dolezal, O., Power, B. E., and Hudson, P. J. (2001) *Biomolecular Eng. 18*, 95−108.
13. Ward, S., Gussow, D., Griffiths, A., Jones, P., and Winter, G. (1989) *Nature 341*, 544−546.
14. Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Bajyana Songa, E., Bendahman, N., and Hamers, R. (1993) *Nature 363*, 446−448.
15. Sheriff, S., and Constantine, K. (1996) *Nat. Struct. Biol. 3*, 733−736.
16. Nguyen, V., Muyldermans, S., and Hamers, R. (1998) *J. Mol. Biol. 275*, 413−418.
17. Davies, J., and Reichmann, L. (1995) *Bio/Technology 13*, 475−479.
18. Hoogenboom, H. (1997) *Trends Biotech. 15*, 62−70.
19. Griffiths, A., and Duncan, A. (1998) *Curr. Opin. Biotech. 9*, 102−108.
20. Clackson, T., and Wells, J. (1994) *Trends Biotech. 12*, 173−184.
21. Martin, F., Volpari, C., Steinkuhler, C., Dimasi, N., Brunetti, M., Biasiol, G., Altamura, S., Cortese, R., De Francesco, R., and Sollazzo, M. (1997) *Protein Eng. 10*, 607−614.
22. Di Marco, S., Rizzi, M., Volpari, C., Walsh, M., Narjes, F., Colarusso, S., De Francesco, R., Matassa, V., and Sollazzo, M. (2000) *J. Biol. Chem. 275*, 7152−7157.
23. Gill, S., and von Hippel, P. (1989) *Anal. Biochem. 182*, 319−326.
24. Otwinowski, Z., and Minor, W. (1997) *Macromolecular Crystallography, part A*, pp 307−326, Academic Press, San Diego.
25. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
26. Perrakis, A., Morris, R., and Lamzin, V. (1999) *Nat. Struct. Biol. 6*, 458−463.
27. Murshudov, G., Vagin, A., and Dodson, E. (1997) *Acta Crystallogr. D53*, 240−255.
28. Number 4, C. C. P. (1994) *Acta Crystallogr. D50*, 760−763.
29. Lamzin, V., and Wilson, K. (1997) *Methods Enzymol. 277*, 269−305.
30. Jones, T., Zou, J., Cowan, S., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
31. Brunger, A. (1992) *Nature 355*, 472−475.
32. Ramakrishnan, C., and Ramachandran, G. N. (1965) *Biophys. J. 5*, 909−933.
33. Reichmann, L. (1996) *J. Mol. Biol. 259*, 957−969.
34. Fan, Z., Shan, L., Guddat, L., He, X., Gray, W., Raison, R., and Edmundson, A. (1992) *J. Mol. Biol. 288*, 188−207.
35. Chothia, C., Novotny, J., Bruccoleri, R., and Karplus, M. (1985) *J. Mol. Biol. 186*, 651−663.
36. Poljak, R. (1975) *Nature 256*, 373−376.
37. Davies, J., and Reichmann, L. (1996) *Protein Eng. 9*, 531−537.
38. Desmyter, A., Transue, T., Ghahroudi, M., Dao Thi, M., Poortmans, F., Hamers, R., Muyldermans, S., and Wyns, L. (1996) *Nat. Struct. Biol. 3*, 803−811.
39. Spinelli, S., Frenken, L., Bourgeois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C., and Tegoni, M. (1996) *Nat. Struct. Biol. 3*, 752−757.
40. Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M., Muyldermans, S., and Wyns, L. (1999) *Structure 7*, 361−370.
41. Spinelli, S., Frenken, L., Hermans, P., Verrips, T., Brown, K., Tegoni, M., and Cambillau, C. (2000) *Biochemistry 39*, 1217−22.
42. Spinelli, S., Tegoni, M., Frenken, L., van Vliet, C., and Cambillau, C. (2001) *J. Mol. Biol. 311*, 123−129.
43. Chothia, C., Lesk, A., Gherardi, E., Tomlinson, I., Walter, G., Marks, J., Llewelyn, M., and Winter, G. (1992) *J. Mol. Biol. 2227*, 799−817.
44. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J., and Hamers, R. (1994) *Protein Eng. 7*, 1129−1135.
45. Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., and Cambillau, C. (2002) *J. Biol. Chem. 277*, 23645−23650.
46. Vranken, W., Tolkatchev, D., Xu, P., Tanha, J., Chen, Z., Narang, S., and Ni, F. (2002) *Biochemistry 41*, 8570−8579.
47. Martin, F., Dimasi, N., Volpari, C., Perrera, C., Di Marco, S., Brunetti, M., Steinkuhler, C., De Francesco, R., and Sollazzo, M. (1998) *Biochemistry 37*, 11459−11468.
48. Lo Conte, L., Chothia, C., and Janin, J. (1999) *J. Mol. Biol. 285*, 2177−2198.
49. Janin, J. (1999) *Structure Folding Des. 7*, R277−R279.
50. Janin, J., and Rodier, F. (1995) *Proteins 23*, 580−587.
51. Steipe, B., Pluckthun, A., and Huber, R. (1992) *J. Mol. Biol. 225*, 739−753.
52. Fields, B., Ober, B., Malchiodi, E., Lebedeva, M., Braden, B., Ysern, X., Kim, J., Shao, X., Ward, E., and Mariuzza, R. (1995) *Science 270*, 1821−1824.
53. Housset, D., Mazza, G., Gregoire, C., Piras, C., Malissen, B., and Fontecilla-Camps, J. (1997) *EMBO J. 16*, 4205−4216.
54. Esnouf, R. (1999) *Acta Crystallogr. D55 (Pt 4)*, 938−940.
55. DeLano, W. (2002) *PyMol*, Delano Scientific, San Carlos.
56. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics 14*, 51−2 and 29−32.

BI035800B