Al Qaraghuli, Mohammed M. and Ferro, Valerie (2017) Analysis of the binding loops configuration and surface adaptation of different crystallized single-domain antibodies in response to various antigens. Journal of Molecular Recognition, 30 (4). ISSN 0952-3499 , http://dx.doi.org/10.1002/jmr.2592

This version is available at https://strathprints.strath.ac.uk/58537/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

The Strathprints institutional repository (https://strathprints.strath.ac.uk) is a digital archive of University of Strathclyde research outputs. It has been developed to disseminate open access research outputs, expose data about those outputs, and enable the management and persistent access to Strathclyde's intellectual output.

Title:

Analysis of the binding loops configuration and surface adaptation of different crystallised single domain antibodies in response to various antigens

Authors

Mohammed M Al Qaraghuli*, Valerie A Ferro

Affiliation

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE, Scotland, UK.

*corresponding author: mohammed.al-qaraghuli@strath.ac.uk

Short title

Structural analysis of single domain antibodies

Keywords

Single domain antibody/CDR length/hapten/lysozyme/surface area

For Peer Review

Abstract

Monoclonal antibodies have revolutionised the biomedical field through their ubiquitous utilisation in different diagnostics and therapeutic applications. Despite this widespread use, their large size and structural complexity have limited their versatility in specific applications. The antibody variable region that is responsible for binding antigen is embodied within domains that can be rescued individually as single-domain antibody (sdAb) fragments. Due to the unique characteristics of sdAbs, such as low molecular weight, high physico-chemical stability, and the ability to bind antigens inaccessible to conventional antibodies, they represent a viable alternative to full-length antibodies. Consequently, 149 crystal structures of sdAbs, originating from human (VH), camelids (VHH), or sharks (VNAR), were retrieved from the Protein Data Bank, and their structures were compared. The three types of sdAbs displayed complementarity determining regions (CDRs) with different lengths and configurations. CDR3 of the VHH and VNAR domains were dominated by pleated and extended orientations, respectively. While VNAR showed the smallest average molecular weight and Molecular Surface Area (MSA) compared to VHH and VH antibodies. However, the Solvent Accessible Surface Area (SASA) measurements of the three tested sdAbs types were very similar. All the anti-hapten VHH antibodies showed pleated CDR3, which were sufficient to create a binding pocket to accommodate haptens (methotrexate and azo dyes) in terms of shape and electrostatic potential. Whereas the sdAbs that recognised lysozyme, showed more diversity in their CDR3 orientation to enable them to recognise various topographies of lysozyme. Subsequently, the three sdAbs classes were different in size and surface area, and have shown distinguishable ability to optimise their CDRs length and orientation to recognise different antigen classes.

For Peer Review

Introduction

Antibodies are widely used in numerous research and medical applications. Structurally, an antibody consists of two heavy and two light polypeptide chains, based on their sizes [1]. The light chains are either of a lambda (λ) or kappa (k) subtype, which can be linked to any of the nine heavy chain subtypes that creates different antibody classes in humans (IgM, IgD, IgG_{1-4}, IgA_{1-2}, and IgE). However, about 85% of the total immunoglobulins (Igs) in human serum are known to be IgG antibodies [2]. The IgG antibody is composed of three fragments, two identical fragment antigen-binding (Fabs) that each contain the first two domains of the heavy and light chains, and one fragment crystallisable region (Fc) [3,4]. The variable region responsible for antigen binding is formed by amino acids located at the tip of the antibody molecule [5]. Each of the variable heavy (VH) or light (VL) domains consist of three complementarity determining regions (CDRs), which are alternatively distributed across four framework (FW) regions, and are accountable for antigen recognition [6]. These domains are the smallest part of the conventional antibody that preserve the original binding activity. In addition to conventional antibodies, heavy chain only antibodies can be naturally acquired from camelidae (camel, llama, and vicugna), or shark species (smooth dogfish, spotted catfish, wobbegong, banded houndshark, and bamboo shark), and are known as HCAb and IgNAR, respectively (reviewed in [7,8]).

The attraction towards use of antibodies originates from the flexibility and modification-tolerability of their structures to fit any bespoke application. Nevertheless, working with full length antibodies (molecular weight of ~150 kDa) can be associated with some impracticality such as their high cost of production, slow expression, weak tissue penetration, and unsuitable long half-life for imaging applications [9]. Therefore, the adaptation of sdAbs is considered a viable alternative in both industrial and research applications [10,11]. Although

For Peer Review

sdAbs are small and stable, the absence of a Fc region from these domains can counterbalance these benefits due to the subsequent abolishment of cellular and complement activation, and reduction in serum half-life [12]. These effects are normally mediated by the Fc region of the antibody that binds to C1q, Fc receptor (FcR), and neonatal Fc receptor (FcRn) [13]. However, the half-life can still be restored by, for instance, fusing these sdAbs to human serum albumin (HSA) to increase the serum half-life without affecting the binding and activity of the fragments [14]. These VH or VL sdAbs (molecular weight of 12-15 kDa) can be successfully obtained by individual rescuing of the original dimeric VH and VL domains of conventional IgG, and expressing them as monomers [15,16]. Also, the HCAbs or IgNAR are devoid of light chains, and their variable domains (VHH or VNAR) have been rescued as sdAbs utilising various established antibody engineering methodologies [17,18].

The VH, VHH, and VNAR domains represent the major types of sdAbs (Figure 1). The VH domain is composed of two anti-parallel β-sheets, one with six strands (A', G, F, C, C', and C'') and the other with four strands (A, B, E, and D) [19]. A conserved disulphide bond, between two highly conserved cysteines (Cys), links the two sheets [20]. The inter-strand bridges between B-C, C'-C'', and F-G strands normally form CDRs 1, 2, and 3, respectively. The VH and VL interface is mainly constructed through packing of strands C, C', F, and G [21]. A high degree of sequence similarity (~80%) was observed between VH of family III and the variable domain (VHH) of HCAbs [22], and both can be superimposed precisely [23,24]. Despite the high sequence conservation, four positions are constitutively different between VH and VHH antibodies (V37F/Y, G44E/Q, L45R/C, and W47G/S/L/F) [19,25]. These four substitutions represent the hallmark of camelisation/humanisation strategies. The VHH domain displays a Cys residue either in the CDR1 or position 45 (FW region), and to establish a disulphide bond, a second Cys can be introduced in the CDR3 during the variable (V) - diverse (D) - joining (J) genes recombination of VHH domains [22]. The third type of

For Peer Review

sdAb is VNAR protein that represents the smallest (~12 kDa) natural binding vertebrate domains [26]. Only a small sequence similarity (25-30%) to mammalian heavy chains was noticed, and the VNARs were more related to the V regions of T-cell receptor (TCR) or Ig kappa light chains [26,27]. Despite this low sequence similarity, the VNARs can still be folded and superimposed in a similar manner to classical VH or VL domains [28]. This can be attributed to classical canonical Cys residues (positions 35 and 107) that stabilise the standard Ig fold, along with an invariant tryptophan at position 36 [29]. Sequence analysis has permitted the classification of VNAR domains into five types based on the presence or absence of non-canonical Cys at specific positions (reviewed by [7]. The availability of these Cys residues was reflected by the ability of these VNARs to create different paratopes [28,30–32].

Previous research of the sdAbs field has comprehensively analysed individual domains obtained from human, camelidae, or shark species. Some of these studies have exclusively investigated their structures [33–37], while others have focused on their isolation and characterisation processes [14,38–43]. However, a collective structural analysis of the three types of sdAbs in terms of CDR lengths and binding site shapes has still not been fully elucidated, and is therefore the focus of this article. Consequently, this structural analysis uses highly reliable crystal structures, which can be obtained from the Protein Data Bank (PDB) [44]. Although the retrieved structures might not be a full representation of nature, since crystallisation can be dishearteningly limited by technical feasibility and cost, they can still provide high quality structural information, which can always be complemented by literature data to fulfil each criteria of the analysis.

The correlation between the sdAbs' molecular weight and surface area was investigated because different amino acids can fold into various three-dimensional structures of similar surface area. The Molecular Surface Area (MSA) indicates an envelope of solute-solvent

For Peer Review

interface from which the solvent molecules are excluded [45]. MSA can be considered as the proper surface to be used for a quantitative evaluation of the hydrophobic effect [46]. On the other hand, the Solvent Accessible Surface Area (SASA) was originally proposed to represent the area of contact between protein and solvent, and to quantify hydrophobic burial [47]. It also demonstrates the area over which the centre of a solvent molecule can be placed while retaining van der Waals contacts with a specific atom and not penetrating others. Analysis of surface area has been used by researchers to evaluate their individual sdAb [35,41,48]. However, a collective analysis of this not fully explored parameter can provide key information about these three types of sdAbs, in terms of folding or binding conformational changes, as has been shown previously in other protein classes [49].

With respect to antigen binding, sdAbs protruding binding sites can comfortably bind the cleft of many enzymes [50,51], but might not be expected to bind small antigens such as haptens that normally bind in a pocket at the VH–VL interface [52]. Nevertheless, several VHH domains have successfully detected different haptens including herbicides, trinitrotoluene, caffeine, mycotoxins, steroids and therapeutic drugs [53–60]. Consequently, different sdAbs crystal structures raised against two antigen classes, lysozyme and the haptens methotrexate (MTX) and azo dye were used as models to understand this interaction process.

Methods

Antibody selection

The crystal structures of different sdAbs were retrieved from the PDB. The utilised search terms were "single domain antibody", "heavy chain antibody", "heavy chain only antibody",

For Peer Review

"camel antibody", "VHH antibody", "llama antibody", "dromedary antibody", "shark antibody", "shark VNAR", or "shark IgNAR". Only structures with acceptable resolution (3 Å or less) were included in the analysis to allow a confident determination of the molecular interactions and structures [61–63]. Using this search profile, a total of 123 VHH crystal structures were obtained from different species including camel (34), llama (82), and alpaca (7). Also, 16 VNAR structures were examined from nurse shark (6), spiny dogfish (8), and spotted wobbegong (2), whilst only 10 VH crustal structures were available in the PDB. Therefore, the total retrieved crystal structures have summed up to 149 crystal structures.

CDRs length and binding shape analyses

The sdAbs sequences acquired from the PDB, and analysed using BioEdit Sequence Alignment Editor, version 7.2.5 [64]. ClustalW Multiple alignment was used to align sequences of the same formats. The three CDRs of the VH fragments were defined using the standard Kabat numbering system [65]. CDRs of the VHH and VNAR domains were determined following standard definitions [28,39,66,67]. The MSA and SASA were calculated in square angstrom (Å²) using PyMOL (academic version). The surface topography of the sdAbs is majorly affected by the shape and length of CDR3 and, therefore, the paratope shape analysis was mainly based on the orientation of CDR3. Three types of CDR3 shapes were observed and denoted as extended, short/flat, or pleated CDR3. This classification was mainly based on whether any specific CDR3 was extended beyond the other CDRs (extended), or within the same boundaries of other CDRs (short/flat), or flipped to the side of the sdAbs (pleated). The binding shape of one VHH crystal structure (1SJV) was excluded from binding shape (CDR3) analysis because it showed an abnormal extension of CDR3 and

For Peer Review

FW4 (Supplementary Table 1). Two other structures (1VER and 1SHM) were also not included since they did not display CDR3 main chain orientation.

Electrostatic potential and docking analysis

Electrostatic potential of the selected crystal structures were calculated using Python Molecule Viewer (PMV) Version 1.5.6 [68]. The electrostatic potential was measured (Compute>Electrostatics>Compute Potential using APBS), in accordance with Adaptive Poisson-Boltzmann Solver (APBS) Version 0.5.1. The energy was mapped to the surface with medium surface quality and 1 Å distance from the surface (Compute>Electrostatics>Map Potential to Surface). The map colour was coded as white: 0 kT/e, blue: 13.7 kT/e, red: -13.7 kT/e.

The antibody-antigen docking analysis was performed using the molecular docking and visual screening program AutoDock Vina [69]. Both the antibodies and antigen (methotrexate) structures were retrieved from the PDB and saved in pdb format. Polar hydrogen atoms were added to the antibodies' models, and the produced models were saved as pdbqt files. A potential option within AutoDock Tools is the ability to determine the docking site of the antibody by setting the dimensions of the docking grid box (Grid>Grid box). This can be achieved by setting the x, y, and z axes of the grid box to cover the binding sites of the antibody. The docking process was commanded through the utilisation of the command prompt within Windows 8. The commands script has included (>cd "Desktop\(file name)"; >"\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe" –help; >"\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe" --config conf.txt --log log.txt). Upon completion of the docking process, the models were exported to the assigned destination file. The generated models, with a descending order of affinity and root-mean-square deviation (RMSD) values, were subsequently analysed by PyMOL.

For Peer Review

Statistical analysis

The statistical analysis was conducted using GraphPad Prism® version 5. One-way ANOVA statistical test (with Bonferroni's Post-Test analysis) was used to compare the average MSA and SASA of the three sdAb types (Table 1). Statistical no significance (ns) was concluded with $P > 0.05$. While the statistical significance was denoted with one star (*) if $P \leq 0.05$, and three stars (***) if $P \leq 0.001$.

Results

CDRs Length and binding sites shape

The analysis included crystal structures of sdAbs obtained from different species. The CDR length examination of each individual type of sdAb revealed conservation in lengths of CDR1 and CDR2, and expected high diversity in CDR3 lengths (Figure 2). The lengths of CDR1 were identical in each group of VH (6 amino acids) and VNAR (8 amino acids) domains, irrespective of binding specificity and type of antigen target (Figure 2A). The VHH fragments were dominated (92%) by 8 amino acid lengths, and were similar to CDR1 lengths of VNAR fragments. The longest CDR1 (19 amino acids) was displayed by VHH antibody (3K3Q), whereas 5 amino acid CDR1 were observed in four VHH crystal structures (4C58, 4C59, 1OP9, and 3EBA). Examination of CDR2 lengths of VH domains revealed a single length of 16 amino acids (Figure 2A). The VHH domains were mainly represented by CDR2 lengths of 10 (84 sequences) and 9 (26 sequences) amino acids (Figure 2B). The VNAR domains do not display CDR2 and, therefore, were not included in CDR2 length comparisons. CDR3 lengths of VHH domains reflected a normal distribution model with CDR3 lengths ranging from 7-26 amino acids (Figure 2C). The more frequently adopted

For Peer Review

CDR3 lengths were 17, 18, and 8 amino acids for VHH, VNAR, and VH crystal structures, respectively (Figure 2C). The VNAR crystal structure of 3MOQ possessed the longest CDR3 represented by 29 amino acids, while the shortest CDR3 lengths (6 amino acids) were observed in two VH crystal structures, 2UZI and 2VH5.

The examined sdAbs CDR3 have adopted either an extended, flat/short, or pleated configuration, as exemplified in Figure 3, and detailed in Supplementary Tables 1-3. These results were noticed by examining the main chain confirmation of the 146 sdAbs (121 VHH, 15 VNAR, and 10 VH). Figure 4 illustrates the orientation of all these crystal structures, which are divided into 13 short/flat (6 VHH, 5 VH, and 2 VNAR), 28 extended (15 VHH, 10 VNAR, and 3 VH), and 105 pleated (101 VH, 2 VNAR, and 2 VH) CDR3 of these domains.

Surface Area and Molecular Weight

The VNAR domains showed the smallest average molecular weight of ~12 kDa, followed by VH and VHH domains (Table 1). VHH crystal structure (3K3Q) presented the largest molecular weight of 14.47 kDa, in contrast to the smallest crystal structure (11.31 kDa) recorded by a VNAR structure (4HGM). The differences in the molecular weight among the examined sdAbs were reflected in their total MSA. The average MSA of VNAR crystal structures was ~1000 Å² less than for the VHH domains, and this difference was statistically significant (Table 1). Even with this large difference in MSA between VHH and VNAR, the average SASAs were surprisingly very close, with no statistical difference, with VNAR fragments slightly exceeding the VHH antibodies (Table 1). One of the VHH structures (5C2U) had the largest MSA, whilst the VNAR (1VER) domain was the smallest. In contrast, there was no statistical difference between the three sdAb types in terms of SASA (Table 1). A VNAR domain (2Z8W) had the largest exposed surface to solvent, and the VHH antibody

For Peer Review

(4IOS) was the lowest (Table 1). The VH domains were positioned in the middle, between VNAR and VHH, in terms of average MSA and molecular weight. Subsequently, the VHH domains showed the highest molecular weight and surface area, however, this was not reflected in their SASA that was slightly surpassed by VNAR domains.

Binding mode against different antigen classes

Anti-haptens sdAbs

The analysed structures included eight llama VHH crystal structures that were developed against haptens (Supplementary Table 4), and half of these structures were crystallised in complex with the haptens, as summarised in Figure 5 (A-D). All eight VHH structures displayed pleated CDR3 of 17-18 amino acid lengths. Bending of CDR3 successfully created a pocket shape at the side of the antibodies as, for instance, displayed by 1QD0 and 1I3U in complex with azo dyes (Figures 5 A and B). Furthermore, the binding pockets of these two antibodies were positively charged to accommodate the azo dyes (Figures 6 A and B).

Another interesting model of binding was represented by four VHH crystal structures (3QXT, 3QXV, 3QXU, and 3QXW). The former two structures were crystallised in complex with MTX, and the latter two were their free forms (Supplementary Table 4). The positively charged side of MTX would be ideally expected to extend towards a negatively charged cluster within the CDR3 generated by three aspartate (Asp) residues. However, the two complexed VHH antibodies, 3QXT and 3QXV, showed MTX to be immersed into a tunnel-shaped pocket below CDR1 (Fig 4 C and D). These pockets were neutral-slightly positively charged, and did not complement the immersed positively charged part of MTX (Figures 6 C and D). In order to investigate this uncommon binding mode, docking of MTX to 3QXT and 3QXV crystal structures were performed using Autodock vina. MTX displayed a polycyclic

For Peer Review

structure containing five oxygen atoms clustered at one side of the structure, and eight nitrogen atoms, seven of which were located at the pteridine end. The 3QXT-MTX docking model showed a similar orientation at the pteridine end of MTX under CDR1, and the active groups of MTX bound to different amino acids including C24, R28, S30, R32, R74, N79, and T80 (Figure 5 E). The pteridine end of MTX, in the 3QXV-MTX model, was also inserted under CDR1 (Figure 5 F). Two main substitutions (N76K and Y79N) were crucially important in improving the binding affinity of 3QXV over 3QXT.

Anti-lysozyme

Nineteen crystallised sdAbs were reported against lysozyme, including 10 VHH, 3 VH, and 6 VNAR crystal structures (Supplementary Table 5). The binding sites of these sdAbs were variable, and their CDR3 lengths ranged from 17-26, 17-18, and 11-12 amino acids for VHH, VNAR, and VH structures, respectively. The three VH domains displayed short CDR3 that were slightly extended to the side of these antibodies (Figure 7A). In addition, pleated CDR3 configurations were found in all the 10 VHH structures and 2 type I VNARs, as shown in Figures 7B and C. The remaining four type II VNAR crystal structures showed α-helical extended CDR3 conformations (Figure 7D).

These different binding site configurations provided an early indicator that the sdAbs might be recognising different sites of the enzyme and, therefore, it was necessary to investigate this further. Out of the nineteen sdAbs, 3 structures were crystallised as the free form (1OHQ, 2I24, and 2I27), and 16 in complex with lysozyme (Supplementary Table 5). Consequently, it was possible to determine the binding sites of lysozyme-complexed structures by structural alignment, and these sdAbs were found to bind different sites of the enzyme, as shown in Figure 8A. Apart from 3 crystal structures (4IOC, 1OP9, and 3EBA), the majority of these sdAbs recognised different sites of a large groove within the lysozyme structure that

For Peer Review

contained both positive, negative, and neutral patches, and made this location attractive for these antibodies (Figure 8B). Interestingly, two groups of sdAbs (coloured as cyan and magenta in Figures 8C and D) were able to share the same orientation of the middle part of their CDR3 (6 amino acid positions), despite being different sdAbs formats (VNAR and VHH) and configured distantly (Figure 8D).

Discussion

SdAbs are widely used in various biomedical applications [70]. Three widely used formats of sdAbs domains (VH, VHH, and VNAR) broadly share several features to be tagged as single domain binding fragments, such as their small size that is combined with high stability, expression yield, and nanomolar affinity [16,35,39,42,71,72]. However, a closer inspection of their structures can identify specific structural characteristics that are sufficient to explicitly maintain their individual identities.

Determining the CDR lengths of antibodies is highly imperative, as the gross shape of antigen binding sites (pocket, groove, or flat surface) relies fundamentally on the lengths of these loops [73]. Despite this importance, CDR definition was associated with several challenges including different definition approaches, mainly by Kabat, Chothia, and IMGT [6,19,65,74–77]. The correlation process can also be hindered by the fact that different CDR lengths are likely to be developed against countless antigen targets. Since the examined 149 crystal structures were raised against different antigens, it was important to examine whether these sdAb have generally fallen within the expected overall length spectrum and diversity of each species. CDR1 and CDR2 of the analysed VH and CDR1 of VNAR domains have shown relative restriction in length, when compared to the slightly variable VHH antibodies

For Peer Review

(Figure 2). This length restriction was observed previously in CDR H1 and CDR H2 of conventional antibodies [78]. Comparably, CDR1 of VHH domains were found to be more variable than VH antibodies, and this phenomenon is attributed to somatic mutations of VHH germlines [79]. Another article has shown that CDR1 and CDR2 of VHH can display different canonical structures when compared to conventional VH domains [80].

The unassembled VH and VL domains are generally characterised by their instability, and individual VH domains are notoriously known to be highly aggregated [81]. This aggregation tendency was previously attributed to the exposed hydrophobic patches located at the interface between the unassembled VH and VL domains. In the current study, the examined 10 VH crystal structures were characterised by 8 amino acid CDR3 lengths (Figure 2). However, this length, is shorter than the anticipated 9-14 amino acids of human and mouse CDRH3 [82–84]. The four interface positions (37, 44, 45, and 47) within VH sdAb, which are different from VHH domains, were noticed to be hydrophobic, and can enhance the aggregation of VH domains. This aggregation-tendency might be augmented by the more frequent short CDR3 (8aa) of VH domains as observed in Figure 2C. In addition, two VH crystal structures, 2VYR and 3QYC, were noticed with long CDR3 of 15 and 16 aa, respectively. However, only 2VYR showed pleated CDR3 while 3QYC displayed exceptionally extended CDR3 (Figure 4). Therefore, unlike VHH or VNAR domains, the short CDR3 did not provide VH domains with great ability to be extended beyond the other CDRs, or to bend across areas that are supposed to be covered by the VL domain (Figure 3). In comparison to VH antibodies, CDR3 lengths of the VHH and VNAR domains were longer, as represented by an average of 17 and 18 amino acids for VHH and VNAR structures, respectively (Figure 2). Generally, the VHH and VNAR average CDR3 lengths were in agreement with what has been observed by other researchers [22,38,85].

For Peer Review

Despite the small length variance, between VHH and VNAR domains, their CDR3 configurations were strikingly distinguishable. CDR3 represent a large proportion of the sdAbs and has an influence on their surface areas, while long CDR3s are generally required to generate extended or pleated shapes. The more frequent CDR3 lengths were observed to be 17aa and 18aa for VHH domains and VNAR domains, respectively (Figure 2). Despite this similarity in the more frequent lengths, and the general tendency of VNAR and VHH to display long CDR3, pleated CDR3 were observed in a large proportion of VHH domains (~83%), whilst the extended CDR3 represented ~67% of the analysed VNAR domains (Figure 4). Therefore, long CDR3 are a crucial perquisite to generate either pleated or extended shapes, but might not govern the final CDR3 shape or surface area of the sdAbs *per se*. The long and pleated CDR3 of VHH antibodies can reduce their aggregation, when compared to VHH antibodies, since they can potentially cover a large proportion of the VL dimerisation regions. This mechanism can also support the presence of hydrophilic residues at positions 37, 44, 45, and 47 of VHH domains in improving their solubility. The structural bending of CDR3 might not be vital for VNARs, because these domains can display more polar and charged residues at regions corresponding to the VH-VL interface [86]. These well distributed charged residues (Glu46, Lys82, Gln84, Arg101, and Lys104) can provide both a hydrophilic surface to the surrounding environment, and shield the conserved hydrophobic core residues [32]. Although only 2 out of the 16 VNAR structures showed pleated CDR3, the two disulphide bonds within these domains held their CDR3 loops into the direction of HV2 [51]. This tightly packed (type I) VNAR was observed only in nurse shark [29]. Whilst extension of CDR3 can remarkably extend the binding sites of these VNAR domains to penetrate into active sites of different targets, especially enzymes [34].

An exclusive feature of the VNAR domain, which differentiated it from all the other domains, is the absence of CDR2 and the presence of two hypervariable regions (HV2 and

For Peer Review

HV4). The lack of CDR2 originated from the absence of two strands (C' and C'') that are normally available within the conventional VH domains [32,51,87]. This structural property has caused a reduction in VNAR sequence lengths that are shorter than VH and VHH antibodies. Consequently, all these structural differences might result in decreasing the average molecular weight of VNAR fragments when compared to VH and VHH antibodies (Table 1). In addition, these features were reflected in the measured MSA, since the measured average MSA of VNAR domains was ~1000 Å² smaller than VHH antibodies. Despite the large difference in MSA between VNAR and VHH domains, their SASA values were very close (Table 1). The SASA similarity can be attributed to the fact that the number of amino acid residues that become buried when the chain folds increases with monomeric protein size [88]. This folding tendency can reduce the polypeptide chain surface in contact with solvent to replace solvent-solute interaction with solvent-solvent counterparts that are more favourable thermodynamically [89].

Small haptens are not expected to be targeted efficiently by sdAbs since they possess a limited number of conformational epitopes suitable for recognition by protruding sdAbs paratopes. However, hapten-binding VHH domains have been successfully isolated using strong selection systems [53,56,90,91]. All the examined anti-hapten VHH domains showed pleated CDR3 (Supplementary Table 4). Despite the absence of a VL domain, the azo dyes RR1 and RR6 recognition mechanism by VHH domains (1I3U and 1QD0) closely mimics traditional VH/VL interfaces, where the hapten pocket is located at the former VL interface created by their pleated CDR3 (Figure 5). The binding pockets of these two antibodies were positively charged (Figures 6A and B), and have electrostatically complemented the negative charges of 11 (Azo dye RR6) and 7 (Azo dye RR1) oxygen atoms presented exposed to the binding pockets. The generated pockets electrostatic potential within these VHH domains have complemented the charges of the utilised haptens, as was shown in full-length

For Peer Review

antibodies [78]. However, the binding pocket of 1QD0 was not large enough to accommodate the entire azo dye (RR6), and only accommodated parts of this antigen, and CDR1 loop provides a strong interaction for the azo dye Reactive Red 6 [90]. The availability of both RR1 complexed (1L3U) and free (1L3V) VHH crystal structures can provide information on whether CDR3 is involved in an antigen-induced binding by displaying specific conformational changes. Spinelli *et al.*, suggested major involvement of CDR3, followed by CDR2, and a framework residue in the binding process [92]. Also, the authors observed that there were movements of 2.0-3.5 Å of the CDR2 and CDR3 towards the RR1 hapten, which suggests a possible antigen-induced reorientation of CDR3. Another hapten-binding model involved two VHH crystal structures in conjugation with MTX. Fanning and colleagues (2011) have shown, through CDRs grafting experiments, that changing five amino acids at positions 76-80, has resulted in improving the binding affinity by 1000 folds [93]. These results were confirmed by the docking analysis in the present study (Figurse 5 E and F), and positions 76 and 79 have dramatically improved the binding affinity of 3QXV. In addition, the reduced positive charge of 3QXV binding pocket has accommodated the positively charged ptredine end of MTX better than the slightly positive charged pocket of 3QXT (Figures 6 C and D). Haptens are not recognisable by the immune system unless conjugated to carrier molecules, and the design of hapten-carrier protein conjugates is key in the development of anti-hapten antibodies [78]. Fanning *et al.*, used the oxygen rich end of MTX in the conjugation process. Generally, antibodies tend to recognise the outer epitopes of the conjugate, in this case the ptredine end, and if the conjugation process is inverted, the oxygen rich end will be more attractive for antibodies. Also, MTX possess other antigenic groups that can be considered as potential epitopes, and other panels of antibodies might be active against other antigenic groups. Consequently, the proposed model was influenced by the conjugation process and CDRs grafting design and mutations at two key positions (N76K and

For Peer Review

Y79N) that has favoured 3QXV over 3QXT, and dictated the favoured MTX orientation underneath CDR1.

The second binding model was based on the structurally well-established human and hen egg-white lysozyme, with molecular weight of 14.7 and 14.3 kDa, respectively (Supplementary Table 5). Lysozyme is highly immunogenic, and the complete structure of the protein was determined and targeted by various antibodies [94]. In contrast to the anti-hapten antibodies, lysozyme binders have displayed short, pleated, and extended CDR3 (Figure 7). This diversity in binding site configurations might originate from the fact that these sdAbs (VHH, VH, and VNAR domains) have recognised different sites of the active pocket-shaped site of lysozyme (Figure 8A). This pocket was shown previously to attract various VHH antibodies, which was strikingly unfavourable to conventional murine antibodies that preferred planar surfaces located outside the active site of the enzyme [50]. Subsequently, the sdAbs have restrictively developed pleated CDR3 to accommodate haptens, and were more flexible in recognising lysozyme through different CDR3 orientations.

In conclusion, each of the VH, VHH, or VNAR domains has maintained a distinguishable level of surface area and molecular weight to maintain their structural stability. Despite the structural similarity within each class, the analysed sdAbs have shown remarkable ability to orientate their CDR3 in various conformations to recognise diverse range of antigens including proteins, glycoproteins, peptides, enzymes, and even haptens. This remarkable flexibility can extend their expediency beyond their distinct ability to bind enzyme clefts or cryptic epitopes as widely appreciated within this research field. Subsequently, this research suggests that there is potential for these sdAbs to be exploited in various immunodiagnostics, biosensors, photothermal therapies, and nanoparticles conjugation.

For Peer Review

References

[1] Edelman GM. DISSOCIATION OF γ-GLOBULIN. J Am Chem Soc 1959;81:3155–6. doi:10.1021/ja01521a071.

[2] Maynard J, Georgiou G. Antibody engineering. Annu Rev Biomed Eng 2000;2:339–76. doi:10.1146/annurev.bioeng.2.1.339.

[3] Ryle AP, Porter RR. Parapepsins: two proteolytic enzymes associated with porcine pepsin. Biochem J 1959;73:75–86.

[4] Porter RR. The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. Biochem J 1959;73:119–26.

[5] Edelman GM, Gottlieb PD. A Genetic Marker in the Variable Region of Light Chains of Mouse Immunoglobulins*. Proc Natl Acad Sci U S A 1970;67:1192–9.

[6] Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med 1970;132:211–50.

[7] Barelle C, Porter A. VNARs: An Ancient and Unique Repertoire of Molecules That Deliver Small, Soluble, Stable and High Affinity Binders of Proteins. Antibodies 2015;4:240–58. doi:10.3390/antib4030240.

[8] Muyldermans S. Nanobodies: natural single-domain antibodies. Annu Rev Biochem 2013;82:775–97. doi:10.1146/annurev-biochem-063011-092449.

[9] Vincke C, Muyldermans S. Introduction to heavy chain antibodies and derived Nanobodies. Methods Mol Biol Clifton NJ 2012;911:15–26. doi:10.1007/978-1-61779-968-6_2.

[10] Kontermann RE. Alternative antibody formats. Curr Opin Mol Ther 2010;12:176–83.

[11] Wesolowski J, Alzogaray V, Reyelt J, Unger M, Juarez K, Urrutia M, et al. Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. Med Microbiol Immunol (Berl) 2009;198:157–74. doi:10.1007/s00430-009-0116-7.

[12] Chames P, Van Regenmortel M, Weiss E, Baty D. Therapeutic antibodies: successes, limitations and hopes for the future. Br J Pharmacol 2009;157:220–33. doi:10.1111/j.1476-5381.2009.00190.x.

[13] Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. Front Immunol 2014;5:520. doi:10.3389/fimmu.2014.00520.

[14] Müller MR, Saunders K, Grace C, Jin M, Piche-Nicholas N, Steven J, et al. Improving the pharmacokinetic properties of biologics by fusion to an anti-HSA shark VNAR domain. mAbs 2012;4:673–85. doi:10.4161/mabs.22242.

[15] Park S-Y, Lee W-R, Lee S-C, Kwon M-H, Kim Y-S, Kim J-S. Crystal structure of single-domain VL of an anti-DNA binding antibody 3D8 scFv and its active site revealed by complex structures of a small molecule and metals. Proteins 2008;71:2091–6. doi:10.1002/prot.22011.

[16] Kim D-S, Song H-N, Nam HJ, Kim S-G, Park Y-S, Park J-C, et al. Directed Evolution of Human Heavy Chain Variable Domain (V H) Using In Vivo Protein Fitness Filter. PLOS ONE 2014;9:e98178. doi:10.1371/journal.pone.0098178.

[17] Müller MR, O'Dwyer R, Kovaleva M, Rudkin F, Dooley H, Barelle CJ. Generation and isolation of target-specific single-domain antibodies from shark immune repertoires. Methods Mol Biol Clifton NJ 2012;907:177–94. doi:10.1007/978-1-61779-974-7_9.

For Peer Review

[18] Olichon A, de Marco A. Preparation of a naïve library of camelid single domain antibodies. Methods Mol Biol Clifton NJ 2012;911:65–78. doi:10.1007/978-1-61779-968-6_5.
[19] Chothia C, Gelfand I, Kister A. Structural determinants in the sequences of immunoglobulin variable domain. J Mol Biol 1998;278:457–79. doi:10.1006/jmbi.1998.1653.
[20] Lesk AM, Chothia C. Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. J Mol Biol 1982;160:325–42.
[21] Chothia C, Novotný J, Bruccoleri R, Karplus M. Domain association in immunoglobulin molecules. The packing of variable domains. J Mol Biol 1985;186:651–63.
[22] Vu KB, Ghahroudi MA, Wyns L, Muyldermans S. Comparison of llama VH sequences from conventional and heavy chain antibodies. Mol Immunol 1997;34:1121–31. doi:10.1016/S0161-5890(97)00146-6.
[23] Decanniere K, Desmyter A, Lauwereys M, Ghahroudi MA, Muyldermans S, Wyns L. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. Struct Lond Engl 1993 1999;7:361–70.
[24] Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, Hamers R, et al. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol 1996;3:803–11.
[25] Muyldermans S, Cambillau C, Wyns L. Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends Biochem Sci 2001;26:230–5.
[26] Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 1995;374:168–73. doi:10.1038/374168a0.
[27] Richards MH, Nelson JL. The evolution of vertebrate antigen receptors: a phylogenetic approach. Mol Biol Evol 2000;17:146–55.
[28] Kovalenko OV, Olland A, Piché-Nicholas N, Godbole A, King D, Svenson K, et al. Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis. J Biol Chem 2013;288:17408–19. doi:10.1074/jbc.M112.435289.
[29] Barelle C, Gill DS, Charlton K. Shark novel antigen receptors--the next generation of biologic therapeutics? Adv Exp Med Biol 2009;655:49–62. doi:10.1007/978-1-4419-1132-2_6.
[30] Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, et al. Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci U S A 1998;95:11804–9.
[31] Diaz M, Velez J, Singh M, Cerny J, Flajnik MF. Mutational pattern of the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol 1999;11:825–33.
[32] Streltsov VA, Carmichael JA, Nuttall SD. Structure of a shark IgNAR antibody variable domain and modeling of an early-developmental isotype. Protein Sci Publ Protein Soc 2005;14:2901–9. doi:10.1110/ps.051709505.
[33] Fennell BJ, Darmanin-Sheehan A, Hufton SE, Calabro V, Wu L, Müller MR, et al. Dissection of the IgNAR V domain: molecular scanning and orthologue

For Peer Review

database mining define novel IgNAR hallmarks and affinity maturation mechanisms. J Mol Biol 2010;400:155–70. doi:10.1016/j.jmb.2010.04.061.

[34] Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA. Maturation of shark single-domain (IgNAR) antibodies: evidence for induced-fit binding. J Mol Biol 2007;367:358–72. doi:10.1016/j.jmb.2006.12.045.

[35] Desmyter A, Decanniere K, Muyldermans S, Wyns L. Antigen specificity and high affinity binding provided by one single loop of a camel single-domain antibody. J Biol Chem 2001;276:26285–90. doi:10.1074/jbc.M102107200.

[36] De Genst EJ, Guilliams T, Wellens J, O'Day EM, Waudby CA, Meehan S, et al. Structure and properties of a complex of α-synuclein and a single-domain camelid antibody. J Mol Biol 2010;402:326–43. doi:10.1016/j.jmb.2010.07.001.

[37] Jespers L, Schon O, James LC, Veprintsev D, Winter G. Crystal structure of HEL4, a soluble, refoldable human V(H) single domain with a germ-line scaffold. J Mol Biol 2004;337:893–903. doi:10.1016/j.jmb.2004.02.013.

[38] Diaz M, Stanfield RL, Greenberg AS, Flajnik MF. Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development. Immunogenetics 2002;54:501–12. doi:10.1007/s00251-002-0479-z.

[39] Dooley H, Flajnik MF, Porter AJ. Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display. Mol Immunol 2003;40:25–33.

[40] Dooley H, Stanfield RL, Brady RA, Flajnik MF. First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. Proc Natl Acad Sci U S A 2006;103:1846–51. doi:10.1073/pnas.0508341103.

[41] De Genst E, Handelberg F, Van Meirhaeghe A, Vynck S, Loris R, Wyns L, et al. Chemical basis for the affinity maturation of a camel single domain antibody. J Biol Chem 2004;279:53593–601. doi:10.1074/jbc.M407843200.

[42] De Genst E, Silence K, Ghahroudi MA, Decanniere K, Loris R, Kinne J, et al. Strong in vivo maturation compensates for structurally restricted H3 loops in antibody repertoires. J Biol Chem 2005;280:14114–21. doi:10.1074/jbc.M413011200.

[43] Tanaka T, Rabbitts TH. Functional intracellular antibody fragments do not require invariant intra-domain disulfide bonds. J Mol Biol 2008;376:749–57. doi:10.1016/j.jmb.2007.11.085.

[44] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42.

[45] Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 1977;6:151–76. doi:10.1146/annurev.bb.06.060177.001055.

[46] Tuñón I, Silla E, Pascual-Ahuir JL. Molecular surface area and hydrophobic effect. Protein Eng 1992;5:715–6.

[47] Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.

[48] Streltsov VA, Carmichael JA, Nuttall SD. Structure of a shark IgNAR antibody variable domain and modeling of an early-developmental isotype. Protein Sci Publ Protein Soc 2005;14:2901–9. doi:10.1110/ps.051709505.

[49] Marsh JA, Teichmann SA. Relative solvent accessible surface area predicts protein conformational changes upon binding. Struct Lond Engl 1993 2011;19:859–67. doi:10.1016/j.str.2011.03.010.

[50] De Genst E, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain

For Peer Review

574 antibodies. Proc Natl Acad Sci U S A 2006;103:4586–91.
575 doi:10.1073/pnas.0505379103.
576 [51] Stanfield RL, Dooley H, Flajnik MF, Wilson IA. Crystal structure of a shark
577 single-domain antibody V region in complex with lysozyme. Science
578 2004;305:1770–3. doi:10.1126/science.1101148.
579 [52] Sundberg EJ, Mariuzza RA. Molecular recognition in antibody-antigen
580 complexes. Adv Protein Chem 2002;61:119–60.
581 [53] Yau KYF, Groves MAT, Li S, Sheedy C, Lee H, Tanha J, et al. Selection of
582 hapten-specific single-domain antibodies from a non-immunized llama ribosome
583 display library. J Immunol Methods 2003;281:161–75.
584 [54] Sheedy C, Yau KYF, Hirama T, MacKenzie CR, Hall JC. Selection,
585 characterization, and CDR shuffling of naive llama single-domain antibodies
586 selected against auxin and their cross-reactivity with auxinic herbicides from
587 four chemical families. J Agric Food Chem 2006;54:3668–78.
588 [55] Ladenson RC, Crimmins DL, Landt Y, Ladenson JH. Isolation and
589 characterization of a thermally stable recombinant anti-caffeine heavy-chain
590 antibody fragment. Anal Chem 2006;78:4501–8. doi:10.1021/ac058044j.
591 [56] Alvarez-Rueda N, Behar G, Ferré V, Pugnière M, Roquet F, Gastinel L, et al.
592 Generation of llama single-domain antibodies against methotrexate, a
593 prototypical hapten. Mol Immunol 2007;44:1680–90.
594 doi:10.1016/j.molimm.2006.08.007.
595 [57] Doyle PJ, Arbabi-Ghahroudi M, Gaudette N, Furzer G, Savard ME, Gleddie S,
596 et al. Cloning, expression, and characterization of a single-domain antibody
597 fragment with affinity for 15-acetyl-deoxynivalenol. Mol Immunol 2008;45:3703–
598 13. doi:10.1016/j.molimm.2008.06.005.
599 [58] Anderson GP, Liu JL, Hale ML, Bernstein RD, Moore M, Swain MD, et al.
600 Development of antiricin single domain antibodies toward detection and
601 therapeutic reagents. Anal Chem 2008;80:9604–11. doi:10.1021/ac8019398.
602 [59] Doyle PJ, Saeed H, Hermans A, Gleddie SC, Hussack G, Arbabi-Ghahroudi M,
603 et al. Intracellular expression of a single domain antibody reduces cytotoxicity of
604 15-acetyldeoxynivalenol in yeast. J Biol Chem 2009;284:35029–39.
605 doi:10.1074/jbc.M109.045047.
606 [60] Kobayashi N, Oyama H, Nakano M, Kanda T, Banzono E, Kato Y, et al.
607 “Cleavable” hapten-biotin conjugates: preparation and use for the generation of
608 anti-steroid single-domain antibody fragments. Anal Biochem 2009;387:257–66.
609 doi:10.1016/j.ab.2009.01.004.
610 [61] Almagro JC. Identification of differences in the specificity-determining residues
611 of antibodies that recognize antigens of different size: implications for the
612 rational design of antibody repertoires. J Mol Recognit JMR 2004;17:132–43.
613 doi:10.1002/jmr.659.
614 [62] McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J
615 Mol Biol 1994;238:777–93. doi:10.1006/jmbi.1994.1334.
616 [63] Raghunathan G, Smart J, Williams J, Almagro JC. Antigen-binding site anatomy
617 and somatic mutations in antibodies that recognize different types of antigens. J
618 Mol Recognit JMR 2012;25:103–13. doi:10.1002/jmr.2158.
619 [64] Hall T. BioEdit: a user-friendly biological sequence alignment editor and
620 analysis program for Windows 95/98/NT, 1999, p. 95–8.
621 [65] Kabat EA, Wu TT, Foeller C, Perry HM, Gottesman KS. Sequences of Proteins
622 of Immunological Interest. DIANE Publishing; 1992.

For Peer Review

[66] Ditlev SB, Florea R, Nielsen MA, Theander TG, Magez S, Boeuf P, et al. Utilizing Nanobody Technology to Target Non-Immunodominant Domains of VAR2CSA. PLOS ONE 2014;9:e84981. doi:10.1371/journal.pone.0084981.
[67] Nguyen-Duc T, Peeters E, Muyldermans S, Charlier D, Hassanzadeh-Ghassabeh G. Nanobody(R)-based chromatin immunoprecipitation/micro-array analysis for genome-wide identification of transcription factor DNA binding sites. Nucleic Acids Res 2013;41:e59. doi:10.1093/nar/gks1342.
[68] Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model 1999;17:57–61.
[69] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61. doi:10.1002/jcc.21334.
[70] de Marco A. Biotechnological applications of recombinant single-domain antibody fragments. Microb Cell Factories 2011;10:44. doi:10.1186/1475-2859-10-44.
[71] Nuttall SD, Humberstone KS, Krishnan UV, Carmichael JA, Doughty L, Hattarki M, et al. Selection and affinity maturation of IgNAR variable domains targeting Plasmodium falciparum AMA1. Proteins 2004;55:187–97. doi:10.1002/prot.20005.
[72] Tanaka T, Williams RL, Rabbitts TH. Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. EMBO J 2007;26:3250–9. doi:10.1038/sj.emboj.7601744.
[73] Collis AVJ, Brouwer AP, Martin ACR. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 2003;325:337–54.
[74] Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987;196:901–17.
[75] Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 1997;273:927–48. doi:10.1006/jmbi.1997.1354.
[76] Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, et al. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 1999;27:209–12.
[77] Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, et al. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 2000;28:219–21.
[78] Al Qaraghuli MM, Palliyil S, Broadbent G, Cullen DC, Charlton KA, Porter AJ. Defining the complementarities between antibodies and haptens to refine our understanding and aid the prediction of a successful binding interaction. BMC Biotechnol 2015;15:99. doi:10.1186/s12896-015-0217-x.
[79] Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire. EMBO J 2000;19:921–30. doi:10.1093/emboj/19.5.921.
[80] Decanniere K, Muyldermans S, Wyns L. Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? J Mol Biol 2000;300:83–91. doi:10.1006/jmbi.2000.3839.
[81] Ward ES, Güssow D, Griffiths AD, Jones PT, Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. Nature 1989;341:544–6. doi:10.1038/341544a0.

For Peer Review

672 [82] Griffin LM, Snowden JR, Lawson ADG, Wernery U, Kinne J, Baker TS. Analysis
673 of heavy and light chain sequences of conventional camelid antibodies from
674 Camelus dromedarius and Camelus bactrianus species. J Immunol Methods
675 2014;405:35–46. doi:10.1016/j.jim.2014.01.003.
676 [83] Zemlin M, Klinger M, Link J, Zemlin C, Bauer K, Engler JA, et al. Expressed
677 murine and human CDR-H3 intervals of equal length exhibit distinct repertoires
678 that differ in their amino acid composition and predicted range of structures. J
679 Mol Biol 2003;334:733–49.
680 [84] Johnson G, Wu TT. Preferred CDRH3 lengths for antibodies with defined
681 specificities. Int Immunol 1998;10:1801–5.
682 [85] Smith LE, Crouch K, Cao W, Müller MR, Wu L, Steven J, et al. Characterization
683 of the immunoglobulin repertoire of the spiny dogfish (Squalus acanthias). Dev
684 Comp Immunol 2012;36:665–79. doi:10.1016/j.dci.2011.10.007.
685 [86] Flajnik MF. Immunogenetics: alternative strategies in adaptive immunity and the
686 rise of comparative immunogenomics. Curr Opin Immunol 2007;19:522–5.
687 doi:10.1016/j.coi.2007.07.013.
688 [87] Zielonka S, Empting M, Grzeschik J, Könning D, Barelle CJ, Kolmar H.
689 Structural insights and biomedical potential of IgNAR scaffolds from sharks.
690 mAbs 2015;7:15–25. doi:10.4161/19420862.2015.989032.
691 [88] Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric
692 proteins. J Mol Biol 1987;196:641–56.
693 [89] Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of
694 oligomeric proteins. J Mol Biol 1988;204:155–64.
695 [90] Spinelli S, Frenken LG, Hermans P, Verrips T, Brown K, Tegoni M, et al.
696 Camelid heavy-chain variable domains provide efficient combining sites to
697 haptens. Biochemistry (Mosc) 2000;39:1217–22.
698 [91] Harmsen MM, van Solt CB, Fijten HPD, van Keulen L, Rosalia RA,
699 Weerdmeester K, et al. Passive immunization of guinea pigs with llama single-
700 domain antibody fragments against foot-and-mouth disease. Vet Microbiol
701 2007;120:193–206. doi:10.1016/j.vetmic.2006.10.029.
702 [92] Spinelli S, Tegoni M, Frenken L, van Vliet C, Cambillau C. Lateral recognition of
703 a dye hapten by a llama VHH domain. J Mol Biol 2001;311:123–9.
704 doi:10.1006/jmbi.2001.4856.
705 [93] Fanning SW, Horn JR. An anti-hapten camelid antibody reveals a cryptic
706 binding site with significant energetic contributions from a nonhypervariable
707 loop. Protein Sci Publ Protein Soc 2011;20:1196–207. doi:10.1002/pro.648.
708 [94] Smith-Gill SJ. Molecular recognition of lysozyme by monoclonal antibodies.
709 EXS 1996;75:277–300.
710
711
712
713

For Peer Review

Tables

Table 1: Measurements of MSA, SASA, and molecular weight

		VHH	VH	VNAR
Molecular Surface Area (Å²)	Average (±SEM)	12047.5 (49.76)	11563.3 (156.4)	11051.5 (150.33)
	Highest	5C2U: 13633	2VYR: 12435	3MOQ: 12367
	Lowest	1VHP: 10372	4PGJ: 10456	1VER: 10040
	Statistics	VHH vs VH: *	VH vs VNAR: ns	VHH vs VNAR: ***
Solvent Accessible Surface Area (Å²)	Average (±SEM)	6571.3 (25.5)	6454.7 (98.91)	6587.8 (124.28)
	Highest	4B50: 7382	3QYC: 6976	2Z8W: 7434
	Lowest	4IOS: 5818	4PGJ: 6036	1VER: 5893
	Statistics	VHH vs VH: ns	VH vs VNAR: ns	VHH vs VNAR: ns
Molecular Weight (kDa)	Average (±SEM)	13.3 (0.04)	12.9 (0.11)	12.2 (0.12)
	Highest	3K3Q: 14.47	2VYR: 13.74	3MOQ: 13.31
	Lowest	4X7F: 12.32	3ZHD, 2VH5: 12.64	4HGM: 11.31

Peer Review

Figure legends

Figure 1: Crystal structures of sdAbs
Three sdAb types were analysed including A) VH (1OHQ), B) VHH (1BZQ), and VNAR (1VES) as examples from each type. The variable region within these domains are called complementarity determining region (CDR) and hypervariable region (HV). The CDR regions were colour coded for illustration as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR): yellow, and HV4(VNAR): magenta. These crystal structures were selected as examples. Structures were viewed by PyMOL 1.3 (academic version).

Figure 2: CDRs length distribution of sdAbs
Length illustration of **A)** CDR1, **B)** CDR2, and **C)** CDR3. The analysed sequences were 123, 10, and 16 sequences for VHH, VH, and VNAR domains, respectively.

Figure 3: Binding site analysis of different sdAbs
CDRs orientation of VHH, VH, and VNAR domains. These domains characterised by either an extended, flat, or pleated CDR3. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR): yellow, and HV4(VNAR): magenta. These crystal structures were selected as examples, and their PDB entry are depicted at the lower corner of each picture. Structures were viewed by PyMOL 1.3 (academic version).

Figure 4: CDR3 backbone orientation of sdAbs
The CDR3 backbone orientation of sdAbs were grouped into either flat, extended, or pleated CDR3. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue. Structures were grouped and viewed by PyMOL 1.3 (academic version).

Figure 5: VHH domains crystallised or docked with their hapten targets.
The binding surfaces of VHH-hapten complexes are demonstrated in PDB entries **A)** 1QD0 crystal structure (VHH-azo dye Reactive Red (RR6)), **B)** 1l3U crystal structure (VHH-azo dye Reactive Red (RR1)), **C)** 3QXT crystal structure (VHH-Methotrexate), **D)** 3QXV crystal structure (VHH-Methotrexate), **E)** 3QXT-Methotrexate docking model, **F)** 3QXV-Methotrexate docking model. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, CDR4 (in E and F): yellow. Structures were viewed by PyMOL 1.3 (academic version).

Figure 6: Surface-mapped electrostatic potential of VHH domains crystallised with their hapten targets .
The binding surfaces of four VHH-hapten complexes are demonstrated in four PDB entries **A)** 1QD0 (VHH-azo dye Reactive Red (RR6)), **B)** 1l3U (VHH-azo dye Reactive Red (RR1)), **C)** 3QXT (VHH-Methotrexate), and **D)** 3QXV (VHH-Methotrexate). Measurements were calculated utilising Python Molecule Viewer (PMV) Version 1.5.6. The produced energy was mapped to the surface with medium surface quality and 1 Å distance from the surface. The map colour was coded as white: 0 kT/e, Blue: 13.7 kT/e, Red: -13.7 kT/e.

Figure 7: CDR3 backbone configuration of anti-lysozyme sdAbs
The backbone configuration of anti-lysozyme **A)** VH, **B)** VHH, **C)** pleated type I VNAR (1SQ2 and 1T6V), **D)** Extended type II VNAR (2I24, 2I25, 2I26 and 2I27) domains. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR):

For Peer Review

yellow, and HV4 (VNAR): magenta. Structures were viewed by PyMOL 1.3 (academic version).

Figure 8: Binding sites of anti-lysozyme sdAbs
The anti-lysozyme crystal structures were grouped into seven groups. These groups were coloured as orange (4IOC), green (1OP9 and 3EBA), blue (1RI8 and 1RJC), yellow (4PGJ and 4U3X), magentas (1JTO, 1JTT, 1JTP, 1MEL, and1XFP), cyan (2I25 and 2I26), white (1SQ2 and 1T6V), and the lysozyme is red coloured. **A)** Represent the binding sites of anti-lysozyme sdAbs. **B)** electrostatic surface of lysozyme, which was configured in same orientation in image A. **C)** and **D)** illustrate the binding sites of two groups (cyan and magentas) to same binding site as side and top view, respectively. Structures were viewed by PyMOL 1.3 (academic version).

Supplementary information document

Supplementary Table 1: Crystal structures obtained from camelidae

Supplementary Table 2: Crystal structures obtained from human

Supplementary Table 3: Crystal structures obtained from shark

Supplementary Table 4: Anti-hapten crystal structures

Supplementary Table 5: Anti-lysozyme crystal structures

Supplementary references

For Peer Review

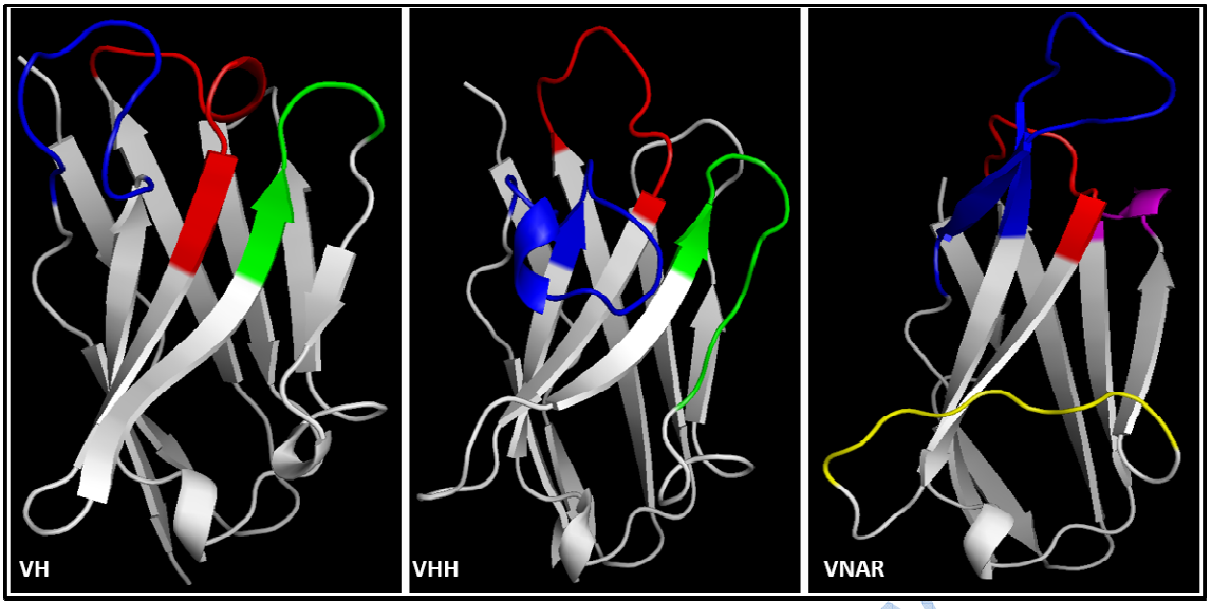
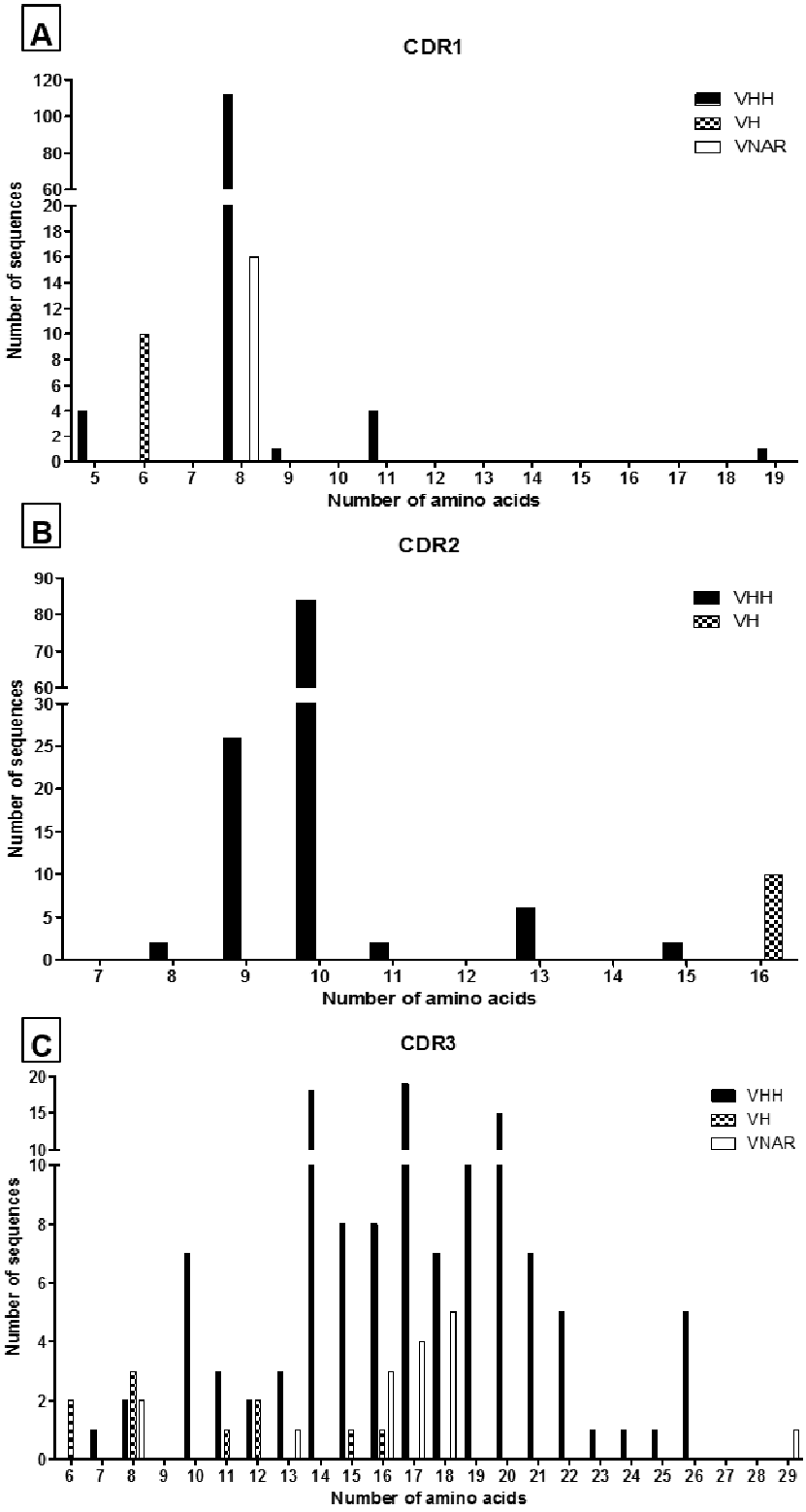


Figure 1: Crystal structures of sdAbs

Three sdAb types were analysed including A) VH (1OHQ), B) VHH (1BZQ), and VNAR (1VES) as examples from each type. The variable region within these domains are called complementarity determining region (CDR) and hypervariable region (HV). The CDR regions were colour coded for illustration as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR): yellow, and HV4(VNAR): magenta. These crystal structures were selected as examples. Structures were viewed by PyMOL 1.3 (academic version).

For Peer Review

811



812
813

814 **Figure 2: CDRs length distribution of sdAbs**

815 Length illustration of **A)** CDR1, **B)** CDR2, and **C)** CDR3. The analysed sequences were 123,
816 10, and 16 sequences for VHH, VH, and VNAR domains, respectively.

817

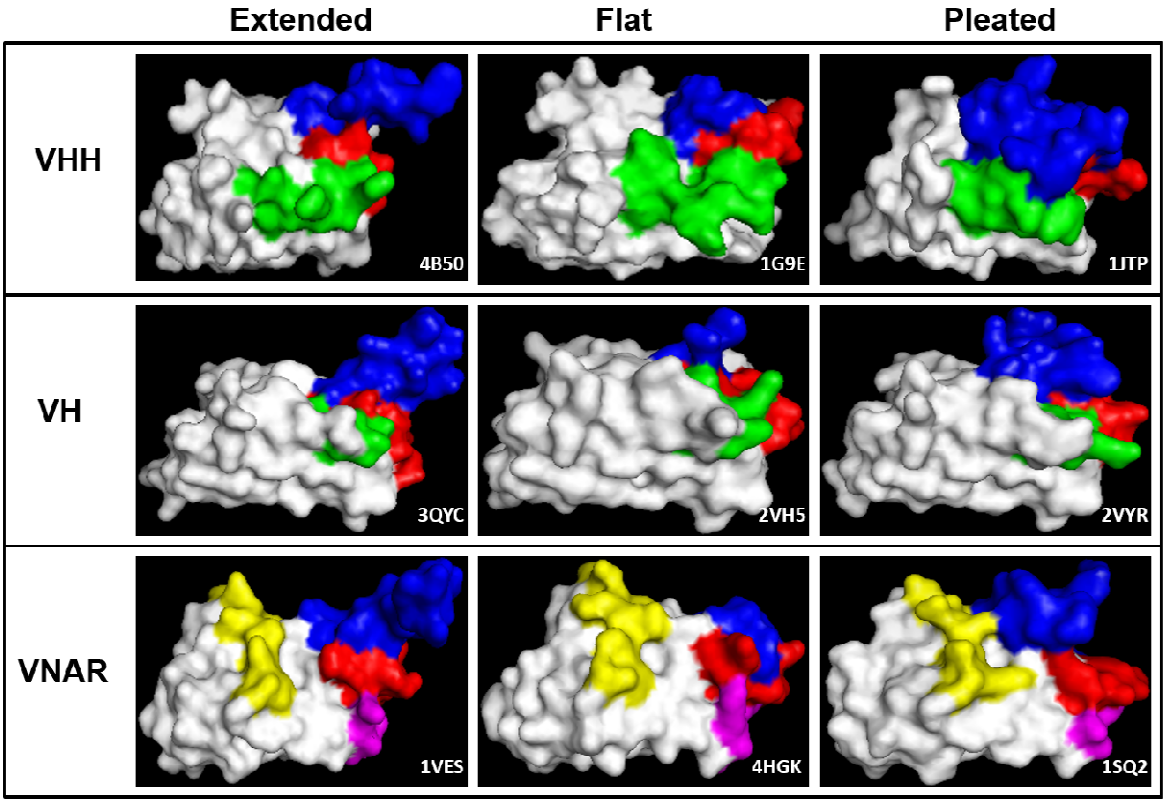
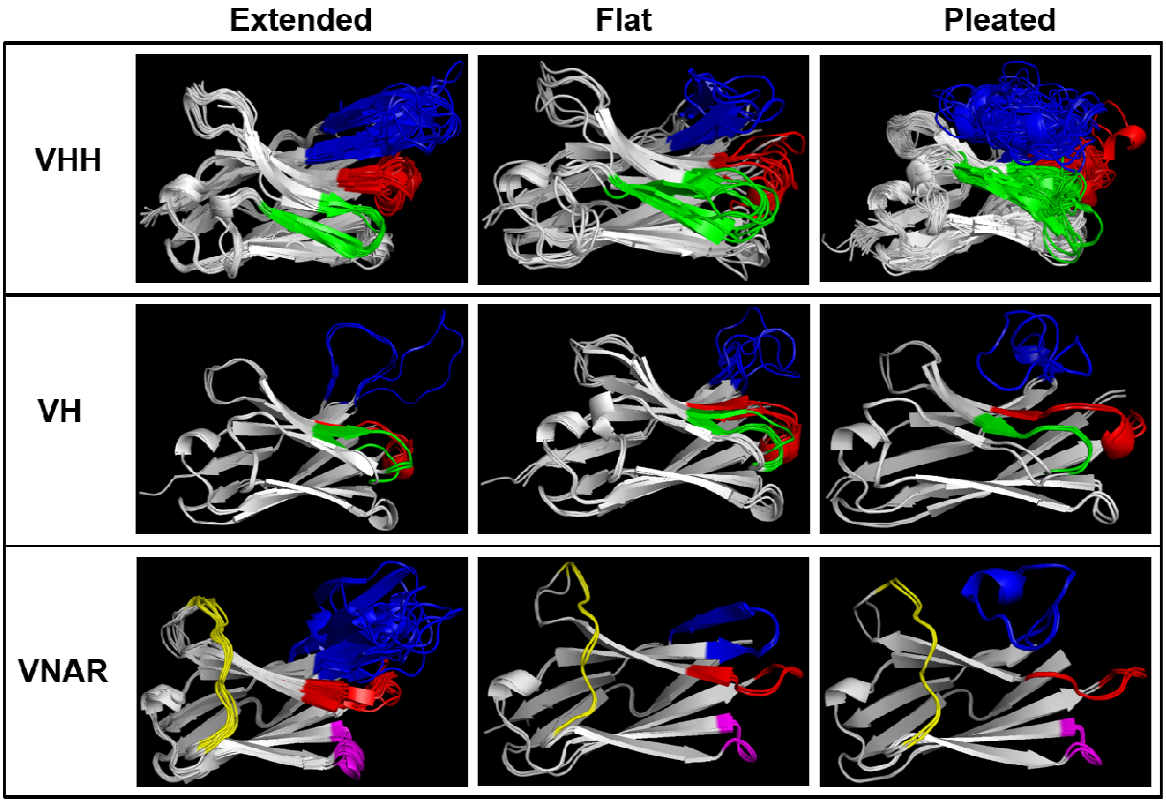


Figure 3: Binding site analysis of different sdAbs

CDRs orientation of VHH, VH, and VNAR domains. These domains characterised by either an extended, flat, or pleated CDR3. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR): yellow, and HV4(VNAR): magenta. These crystal structures were selected as examples, and their PDB entry are depicted at the lower corner of each picture. Structures were viewed by PyMOL 1.3 (academic version).

For Peer Review



828
829
830 **Figure 4: CDR3 backbone orientation of sdAbs**
831 The CDR3 backbone orientation of sdAbs were grouped into either flat, extended, or pleated
832 CDR3. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue.
833 Structures were grouped and viewed by PyMOL 1.3 (academic version).
834

For Peer Review

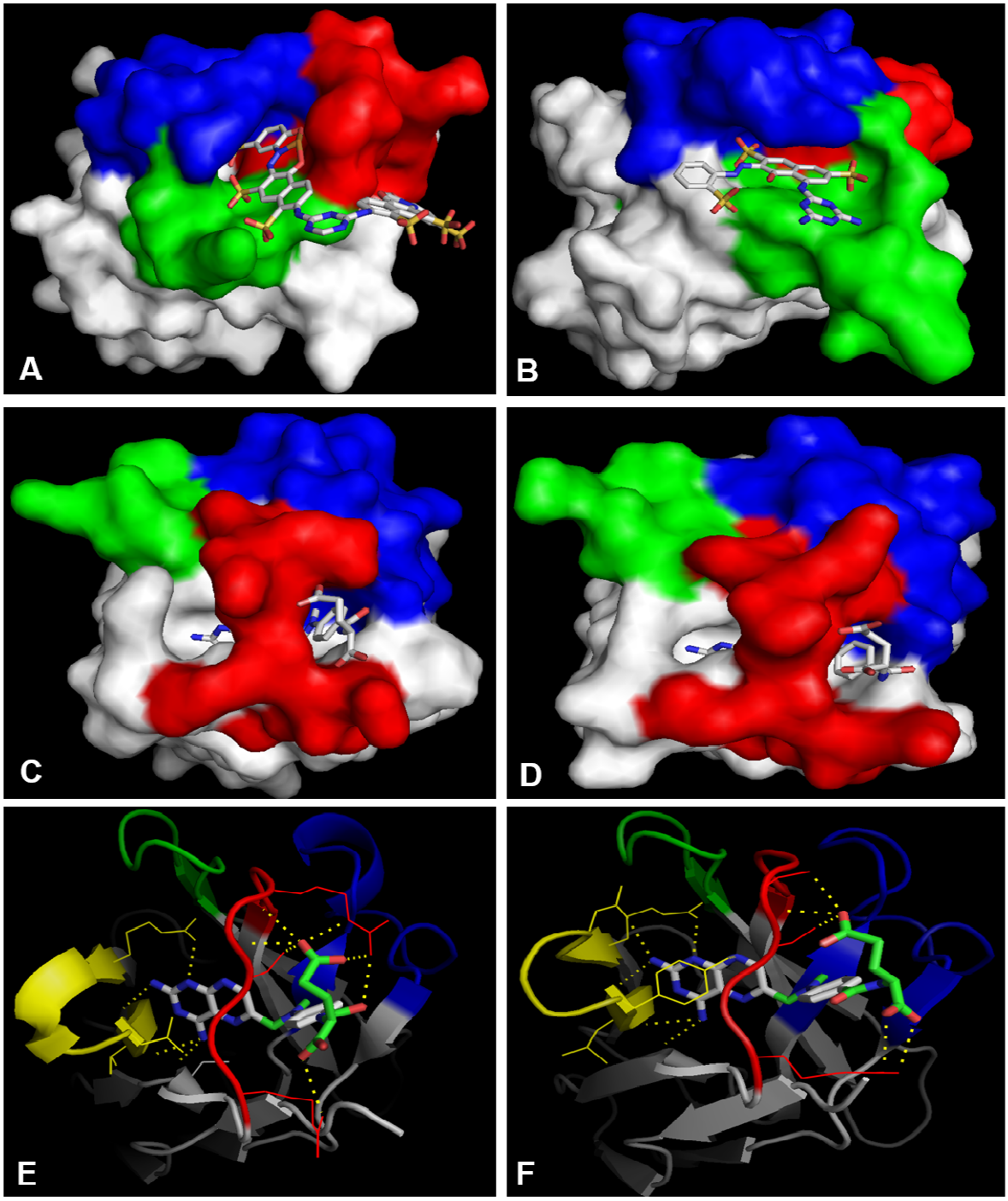


Figure 5: VHH domains crystallised or docked with their hapten targets.
The binding surfaces of VHH-hapten complexes are demonstrated in PDB entries **A)** 1QD0 crystal structure (VHH-azo dye Reactive Red (RR6)), **B)** 1l3U crystal structure (VHH-azo dye Reactive Red (RR1)), **C)** 3QXT crystal structure (VHH-Methotrexate), **D)** 3QXV crystal structure (VHH-Methotrexate), **E)** 3QXT-Methotrexate docking model, **F)** 3QXV-Methotrexate docking model. The CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, CDR4 (in E and F): yellow. Structures were viewed by PyMOL 1.3 (academic version).

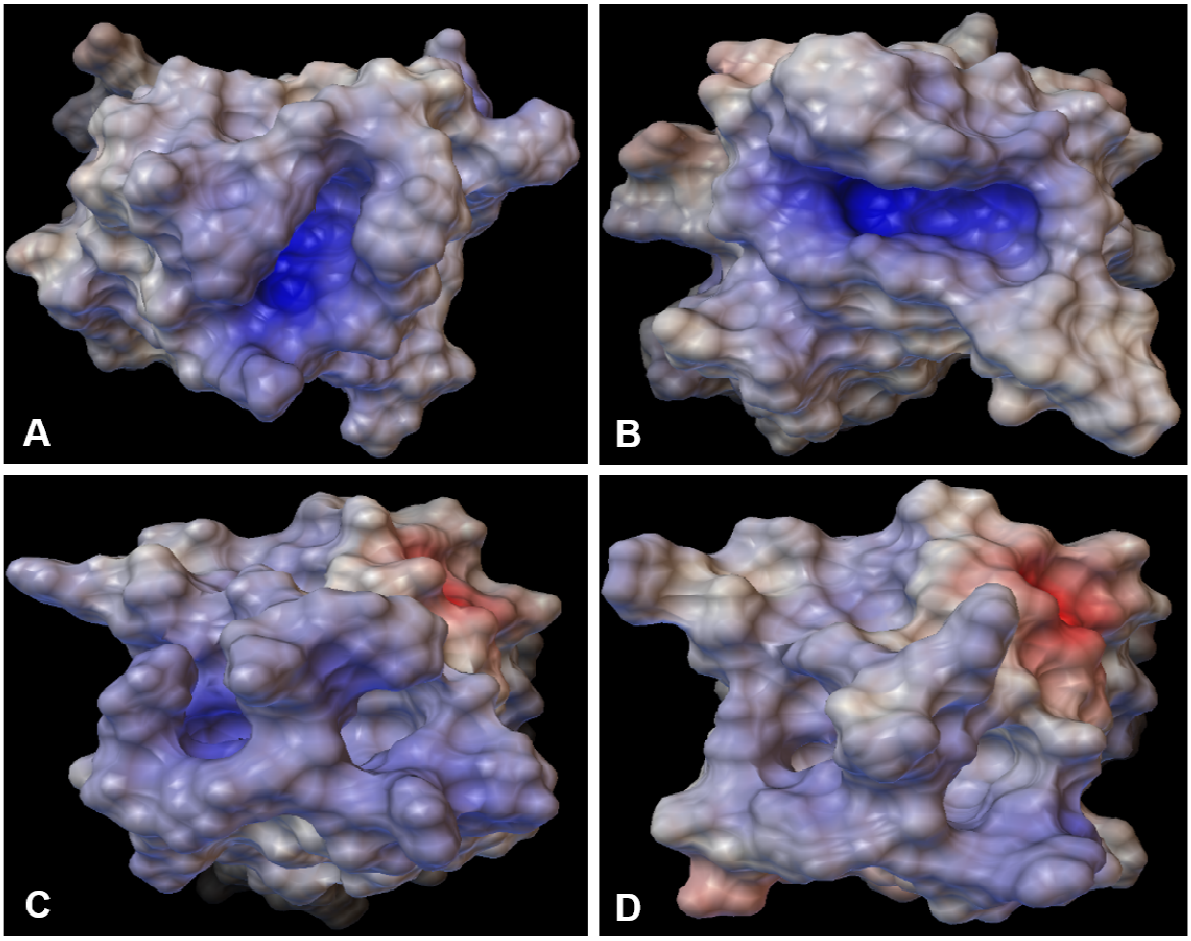
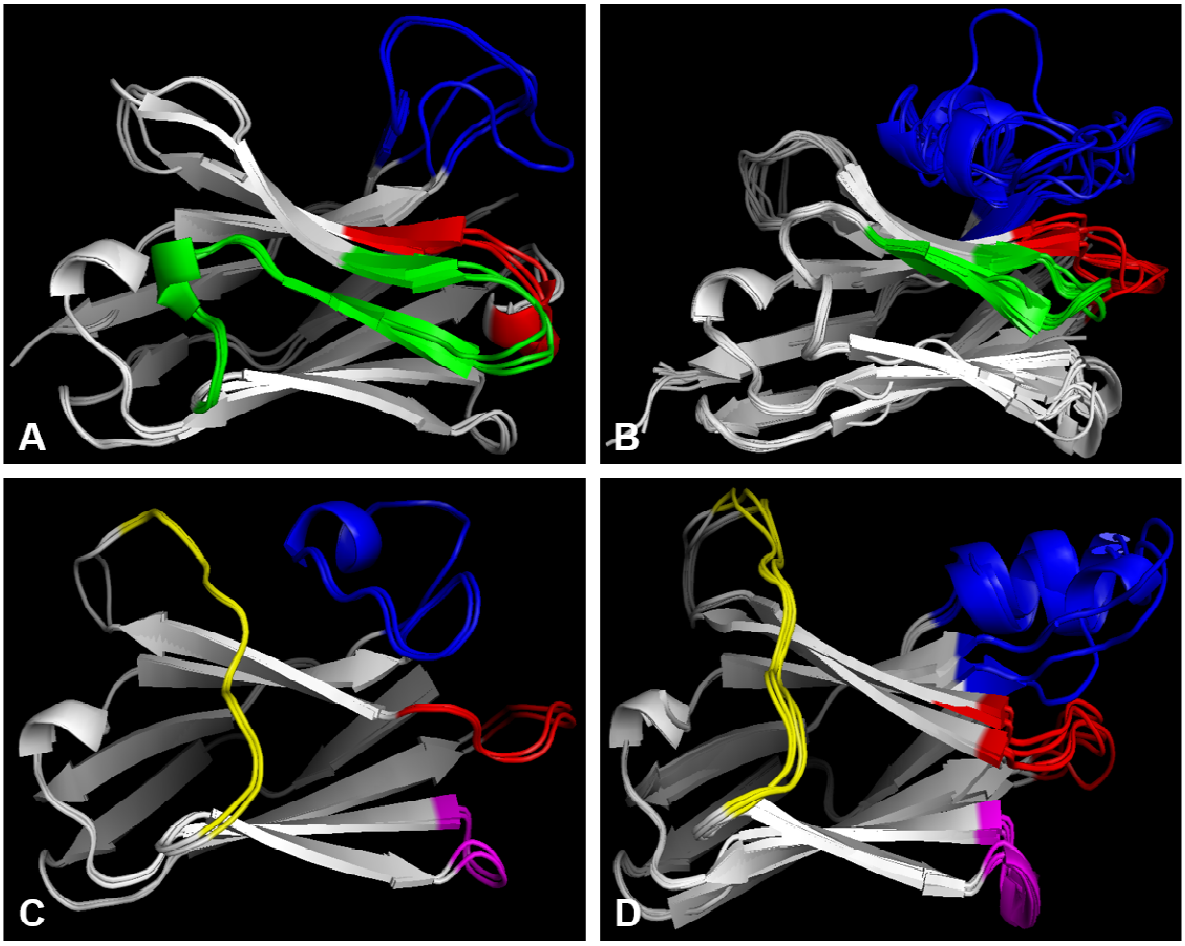


Figure 6: Surface-mapped electrostatic potential of VHH domains crystallised with their hapten targets .

The binding surfaces of four VHH-hapten complexes are demonstrated in four PDB entries **A)** 1QD0 (VHH-azo dye Reactive Red (RR6)), **B)** 1l3U (VHH-azo dye Reactive Red (RR1)), **C)** 3QXT (VHH-Methotrexate), and **D)** 3QXV (VHH-Methotrexate). Measurements were calculated utilising Python Molecule Viewer (PMV) Version 1.5.6. The produced energy was mapped to the surface with medium surface quality and 1 Å distance from the surface. The map colour was coded as white: 0 kT/e, Blue: 13.7 kT/e, Red: -13.7 kT/e.

For Peer

858



859
860

861 **Figure 7: CDR3 backbone configuration of anti-lysozyme sdAbs**
862 The backbone configuration of anti-lysozyme **A)** VH, **B)** VHH, **C)** pleated type I VNAR
863 (1SQ2 and 1T6V), **D)** Extended type II VNAR (2I24, 2I25, 2I26 and 2I27) domains. The
864 CDR regions were colour coded as CDR1: red, CDR2: green, CDR3: blue, HV2 (VNAR):
865 yellow, and HV4 (VNAR): magenta. Structures were viewed by PyMOL 1.3 (academic
866 version).
867

For Peer

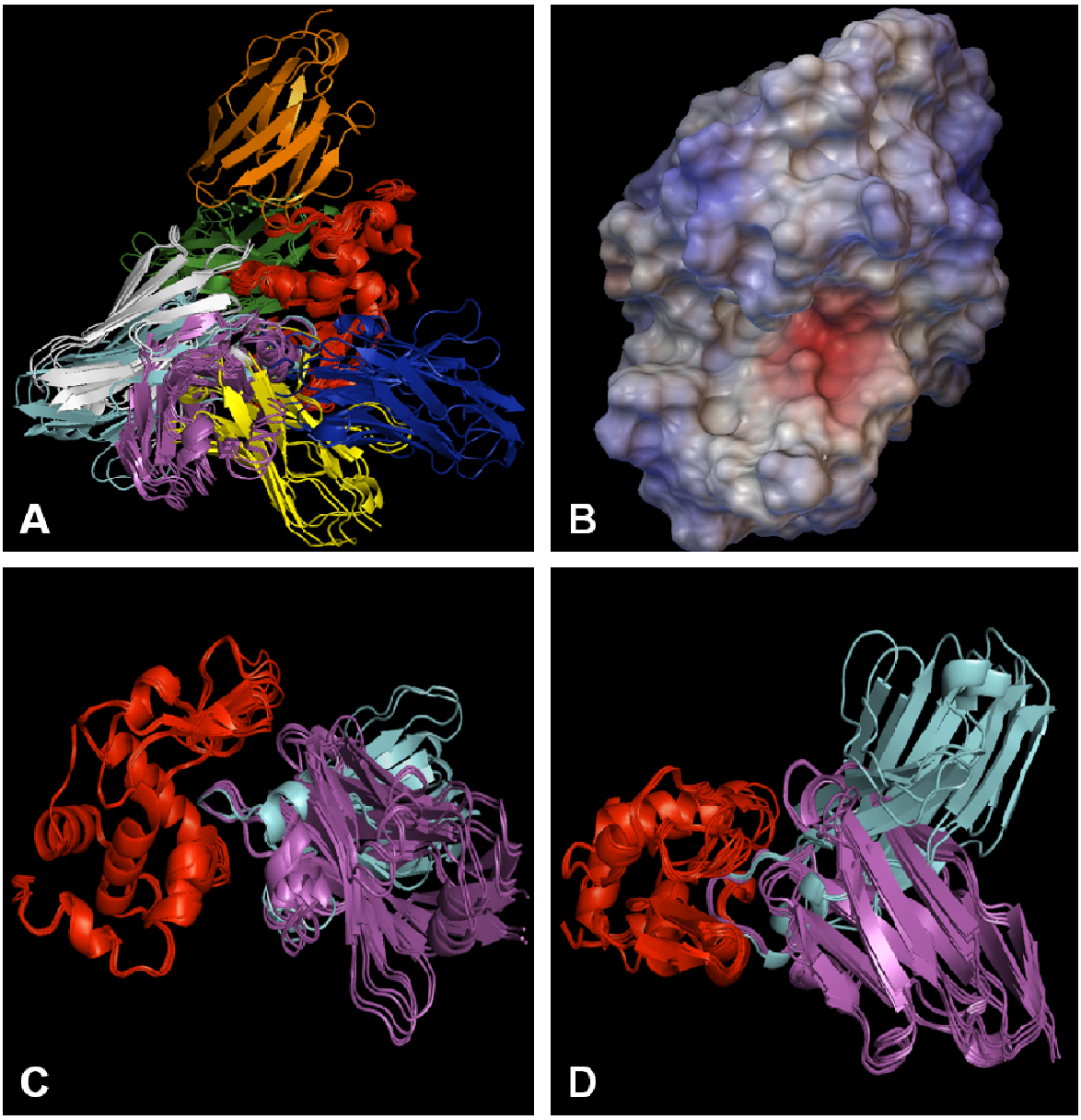


Figure 8: Binding sites of anti-lysozyme sdAbs

The anti-lysozyme crystal structures were grouped into seven groups. These groups were coloured as orange (4IOC), green (1OP9 and 3EBA), blue (1RI8 and 1RJC), yellow (4PGJ and 4U3X), magentas (1JTO, 1JTT, 1JTP, 1MEL, and1XFP), cyan (2I25 and 2I26), white (1SQ2 and 1T6V), and the lysozyme is red coloured. **A)** Represent the binding sites of anti-lysozyme sdAbs. **B)** electrostatic surface of lysozyme, which was configured in same orientation in image A. **C)** and **D)** illustrate the binding sites of two groups (cyan and magentas) to same binding site as side and top view, respectively. Structures were viewed by PyMOL 1.3 (academic version).