

## Structural plasticity in the loop region of engineered lipocalins with novel ligand specificities, so-called Anticalins

S. Achatz, A. Jarasch, A. Skerra\*

Lehrstuhl für Biologische Chemie, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany

### ARTICLE INFO

#### Keywords:

Beta-barrel  
CDR  
Hypervariable loop  
Lipocalin  
Protein engineering  
Protein scaffold

### ABSTRACT

Anticalins are generated via combinatorial protein design on the basis of the lipocalin protein scaffold and constitute a novel class of small and robust engineered binding proteins that offer prospects for applications in medical therapy as well as *in vivo* diagnostics as an alternative to antibodies. The lipocalins are natural binding proteins with diverse ligand specificities which share a simple architecture with a central eight-stranded anti-parallel  $\beta$ -barrel and an  $\alpha$ -helix attached to its side. At the open end of the  $\beta$ -barrel, four structurally variable loops connect the  $\beta$ -strands in a pair-wise manner and, together, shape the ligand pocket. Using targeted random mutagenesis in combination with molecular selection techniques, this loop region can be reshaped to generate pockets for the tight binding of various ligands ranging from small molecules over peptides to proteins. While such Anticalin proteins can be derived from different natural lipocalins, the human lipocalin 2 (Lcn2) scaffold proved particularly successful for the design of binding proteins with novel specificities and, over the years, more than 20 crystal structures of Lcn2-based Anticalins have been elucidated. In this graphical structural biology review we illustrate the conformational variability that emerged in the loop region of these functionally diverse artificial binding proteins in comparison with the natural scaffold. Our present analysis provides picturesque evidence of the high structural plasticity around the binding site of the lipocalins which explains the proven tolerance toward excessive mutagenesis, thus demonstrating remarkable resemblance to the complementarity-determining region of antibodies (immunoglobulins).

### Introduction

The lipocalins are a family of evolutionarily related proteins that are found ubiquitously in many phyla of life where they are involved in the transport, storage or scavenging of vitamins, hormones and metabolites (Åkerström et al., 2006; Diez-Hermano et al., 2021; Flower, 1996). Despite high sequence diversity – with only a few conserved residues throughout the family – the lipocalins share a highly conserved common fold which is dominated by the central  $\beta$ -barrel backed by an  $\alpha$ -helix and an extended strand. The  $\beta$ -barrel is formed by eight antiparallel  $\beta$ -strands which are arranged in a circular manner around a central axis. Closed by short loops and a hydrophobic core of densely packed aromatic side chains at one end, the  $\beta$ -barrel is open to the solvent at the other end, where four loop segments connect each pair of  $\beta$ -strands and, thus, create a pocket to accommodate a ligand (Skerra, 2000). While the  $\beta$ -barrel with the attached  $\alpha$ -helix is strictly conserved in the lipocalin fold, the set of four loops is structurally highly variable in terms of length, amino acid sequence and backbone conformation, which

explains the broad spectrum of natural ligand specificities that range from vitamin A to Fe<sup>III</sup>-siderophore complexes (Schiefner and Skerra, 2015).

This bipartite protein architecture prompted efforts to reshape the ligand-binding site of natural lipocalins via combinatorial protein design to generate proteins with novel binding functions, so-called Anticalins (Beste et al., 1999; Richter et al., 2014; Skerra, 2001). This was accomplished by preparing genetic libraries encoding lipocalin variants with random mutations targeted at specific positions within the loop regions and applying powerful selection techniques such as phage display and, more recently, bacterial surface display (Gebauer and Skerra, 2012; Richter et al., 2014). X-ray crystallographic analyses of the first Anticalin examples – with specificities towards fluorescein and digoxigenin, respectively, compared with biliverdin as a natural ligand – revealed considerable changes in the loop conformations of the bilin-binding protein (BBP), a structurally well characterized lipocalin from a butterfly that was initially employed as a scaffold.

Hence, a picture emerged revealing features of the lipocalins similar

\* Corresponding author.

E-mail address: [skerra@tum.de](mailto:skerra@tum.de) (A. Skerra).

<https://doi.org/10.1016/j.yjsbx.2021.100054>

Received 28 September 2021; Received in revised form 25 November 2021; Accepted 26 November 2021

Available online 30 November 2021

2590-1524/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Table 1**  
Engineered lipocalins with novel ligand specificities used for the graphical analysis.

Ligand specificity	Anticalin	Ligand type	PDB ID
Y-DTPA	Tb7.N9	Hapten	3DSZ
Y-DTPA	C26	Hapten	4IAW
Y-DTPA	CL31	Hapten	4IAX
Catacalin-TSA	C3A5	Hapten	6Z2C
Colchicine	$\Delta 4$ -D6.2(M69Q)	Hapten	5NKN
Colchicine (apo)	D6.2	Hapten	6Z6Z
Petrobactin	$\Delta 4$ M2	Hapten	6GR0
Petrobactin (apo)	$\Delta 4$ M2	Hapten	6QZ
A $\beta$ peptide	H1GA	Peptide	4MVL
A $\beta$ peptide	US7	Peptide	4MVI
Hepcidin	Ac3 (I24)	Peptide	4QAE
CTLA-4 ectodomain	PRS-010#3 (O10)	Protein	3BX7
Fn ED-B	N7A	Protein	4GH7
Fn ED-B	N7E	Protein	5 N47
Fn ED-B	N9B	Protein	5 N48
huCD98hcED	P3D11	Protein	6S8V
muCD98hcED	C1B12	Protein	6SUA
Wild-type Lcn2	UniProt ID P80188	Siderophore	1L6M

to immunoglobulins (Igs). Both protein classes comprise a highly conserved framework that supports a structurally variable loop region (known as hypervariable loops or complementarity-determining region (CDR) in the case of Igs) which confers the specific antigen or ligand binding activity (Skerra, 2003). However, there is one crucial biological difference: whereas the mammalian immune system is capable of constantly generating antibodies with new antigen specificities via somatic gene recombination and hypermutation, the lipocalins are genetically fixed in a species, thus comprising an inherited spectrum of ligand-binding activities. In humans, for example, there are not more than a dozen distinct lipocalins – plus some isotypes – all of which have been structurally characterized (Schiefner and Skerra, 2015).

With a maturing Anticalin technology, the focus was directed at medical applications to provide a viable alternative to antibodies, a well-established and most successful class of biopharmaceuticals today (Strohl and Strohl, 2012). Compared with Igs, with their large size (~1500 residues), a complicated quaternary structure and complex disulfide bridge and glycosylation patterns, the small and robust lipocalin proteins simply comprise a single polypeptide chain of approximately 180 amino acid residues. This offers several benefits, such as much easier biochemical manipulation and recombinant production as well as the facile construction of fusion proteins incorporating additional functions (Deuschle et al., 2021; Richter et al., 2014). To minimize undesired immunogenicity upon administration to patients, human lipocalins were chosen as appropriate scaffolds for protein engineering in this context, in particular the human lipocalin 1 (Lcn1), also known as tear lipocalin, and the human lipocalin 2 (Lcn2), also known as neutrophil gelatinase-associated lipocalin or siderocalin (Schiefner and Skerra, 2015). Several promising Anticalin drug candidates exhibiting specificities towards different medically relevant target proteins, mainly in the areas of oncology and inflammatory diseases, were generated in this manner and are currently subject to clinical studies at various stages (Deuschle et al., 2021; Rothe and Skerra, 2018).

The human Lcn2 has emerged as a particularly fruitful scaffold to yield Anticalin proteins directed against many different targets (Table 1). In this case, the design of the Lcn2-based random library underwent iterative improvement (Gebauer et al., 2013), taking into

account X-ray structural data of early Anticalin representatives (Kim et al., 2009; Schönfeld et al., 2009) in the light of theoretical considerations on the efficient physical sampling of an astronomically large sequence space (Richter et al., 2014; Skerra, 2003). Over the years, more than 20 crystal structures of Lcn2-based Anticalins with specificities from small molecules, over peptides with varying lengths, up to macromolecular protein targets have been elucidated (Table 1). Many of these Anticalins were raised against disease-related molecular targets. Notably, all of these Lcn2-based Anticalin proteins share exactly the same lengths for all four loops, since the library design always was limited to amino acid exchanges, without insertions or deletions (Gebauer et al., 2013). Consequently, the three-dimensional structures of these Anticalins can be compared in a straightforward manner, both mutually and *versus* the natural Lcn2 scaffold, after superposition via the conserved  $\beta$ -barrel (Skerra, 2000). Here, we present a graphical review of these Lcn2-based Anticalins, which offers interesting insights into their structural biology.

## Graphical review

Basis of our structural comparison was a set of 17 Anticalins with crystal structures deposited at the Protein Data Bank (PDB), Research Collaboratory for Structural Bioinformatics (Rutgers University, New Brunswick, NJ), showing resolutions of 1.40–3.00 Å, complemented with the one of human wild-type Lcn2 (PDB ID: 1L6M) (Goetz et al., 2002) (see Table 1). Only those Anticalin structures whose coordinate sets showed a continuously modelled polypeptide chain were considered for this analysis. An amino acid sequence alignment prepared with ANTICALIGN software (Jarasch et al., 2016) is shown in Fig. 1. Side chain exchanges resulting from different random library designs (Gebauer et al., 2013), followed by selection against various molecular targets, are clustered within the four structurally variable loop regions of the lipocalin scaffold (Skerra, 2000).

For the purpose of comparison, a comprehensive set of 18 published crystal structures of the unmutated human Lcn2 protein, co-crystallized with different ligands and/or in different crystal forms (Clifton et al., 2019), was prepared. To this end, in total 22 coordinate sets deposited at the PDB with 100 % sequence identity to PDB ID: 1L6M (Goetz et al., 2002), which was also included as reference in the group of the Anticalin proteins described above, were considered; of those, the four entries with the lowest resolutions were omitted, resulting in a collection of 18 X-ray coordinate sets showing 2.04–2.55 Å resolution.

Manipulation of both the Anticalin and the wtLcn2 coordinate sets as well as three-dimensional graphics were prepared with UCSF Chimera 1.15 (<http://www.cgl.ucsf.edu/chimera>) (Pettersen et al., 2004). Generally, the polypeptide chain with identifier A was chosen for the analysis. Alternate side chain conformations, if deposited, were neglected and some incomplete side chains in the wild-type Lcn2 crystal structures (mostly surface-exposed Lys) were reconstituted with the most plausible rotamer using PyMOL 2.3.3 (Schrödinger, New York, NY). First, the 18 polypeptide chains for each group of proteins were structurally superimposed via the C $\alpha$  positions of those 58 residues (positions 28–37, 52–58, 63–69, 77–84, 91–94, 106–113, 118–124, 133–139 in the mature wtLcn2 polypeptide) which are structurally conserved throughout the lipocalins (Skerra, 2000) using the wtLcn2 (PDB ID: 1L6M) as template. Then, for each of the residues defined in all crystal structures (sequence positions 7–176) the arithmetic mean position of the 18 C $\alpha$  atoms was calculated, together with the sample standard deviation of the distances of these 18 C $\alpha$  positions to the

```

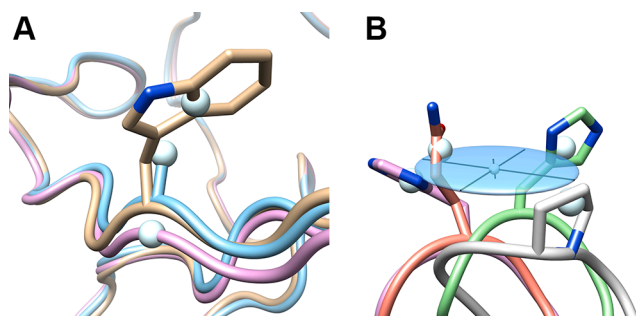
1      10      20      30      40      50      60
1L6M QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKQWVVGLAGNAILREDKDPQKMYATIYELKE
3BX7 .....H.....R...D.QH.MN.....
3DSZ .....H...Q.R...A.....T.Q.....
4GH7 .....H...K...HD.....R.Q.....
4IAW .....H...Q.R...AP.....LL.T.QT...
4IAX .....H...Q.R...APPM.PELLLL.T.QT...
4MVI .....H...V...KS.....W.....
4MVL .....H...A...VL.....L.....
4QAE .....H.....EV.....M.W.....E.
5N47 .....H.....E...SL.....R.....
5N48 .....H...R...MR.....A.V.....
5NKN ----H...V...GF.....I.A.....
6GQZ ----H...F.E.IQQ.....P.I.....
6GR0 ----H...F.E.IQQ.....P.I.....
6S8V .....H...R...TG.....G.F.....
6SUA .....H...GN.....L.....
6Z2C --H...M...GD.....Y.R.....
6Z6Z ----H...V...GF.....I.A.....
=====A=====---#1---B=====

      70      80      90      100      110      120
1L6M DKSYNVTSVLFRRKKKCDYWIRTFVPGSQPGEFTLGNIKSYPGLTSLVLRVWVSTNYNQHAM
3BX7 .....ISSH...E.T.A.....GDK.....D...Y.V
3DSZ .....A.R.....LTM.....
4GH7 .....N.R.VH...N.R.W.....W...W...
4IAW .....A.R...M.E.LTM.....
4IAX ...D.A.R...M.E.LTM...SP.....N...F...
4MVI .....G.GT...H.K.....R...R.R.A...
4MVL .....G.DD...L.K.....R...E.G.W...
4QAE .....I.MPLAE...E.LFQ.....G...G.R.G...
5N47 .....R.S...H.L.....L...K.H.F...
5N48 .....K.M.QR...K.M.N.....A...P.P.T...
5NKN .....FQK.PM...Q.MTD.L.....E.Y.W...
6GQZ .....N.ASNWE...T.R.K.....E.R.M...
6GR0 .....N.ASNWE...T.R.K.....E.R.M...
6S8V .....Y.W.GQ...M.S.G.....A.R.W...
6SUA .....D...ID...H.L.N.....S...TL.S.R...
6Z2C .....W.W.LP...K.F.N.....L...Q.L...
6Z6Z .....FMK.PM...Q.MTD.L.....E.Y.W...
===C===---#2---D===E===---#3---F===

      130      140      150      160      170      178
1L6M VFFKQVSONREYFKITLYGRTEKLTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDG
3BX7 ...LAED.A.F.A.I...A...
3DSZ .....S...L...A...
4GH7 ...R.Y...L.E...N...
4IAW .....Q...S.S.L...A...
4IAX .....Q...S.S.L...A...
4MVI ...V.Q...S.N...
4MVL ...E.A...T.N...
4QAE ...V.W...V.WV...
5N47 ...T.A...F...
5N48 ...H.F...H...
5NKN ...A.Q...D.F...
6GQZ ...T.V...K.W...
6GR0 ...T.V...K.W...
6S8V ...S.T...G.A...
6SUA ...W.H...W.Y...
6Z2C ...T.I...R.Y...
6Z6Z ...A.Q...D.F...
G===---#4---H===

```

Fig. 1. Amino acid sequence alignment of 17 engineered lipocalins (Anticalins; see Table 1) with the natural Lcn2 scaffold (PDB ID: 1L6M; UniProt ID: P80188; mature sequence) shown at the top. The 4 structurally variable loop regions (#1–#4; residues 38–51, 70–76, 95–105 and 125–132) are highlighted. The eight  $\beta$ -strands of the barrel are labeled underneath (A–H). Note that the amino acid exchange Q28H is fixed in all Anticalins (for cloning purposes) while the free Cys residue present at position 87 of the natural protein (with its reactive thiol group) was substituted by Ser throughout.



**Fig. 2.** Exemplary illustration of how the side chain mass centers were calculated and visualized. A centroid is defined as the unweighted, arithmetic mean position of all non-hydrogen positions of a residue excluding its backbone atoms ( $C\alpha$ , N, O and C) – except for glycine, where the  $C\alpha$  position was used to represent the (missing) side chain. Anisotropic displacement ellipsoids were generated for illustration of the spatial side chain variance on the basis of these centroids, thus representing both conformational variations and differences in size as a result of mutagenesis. (A) Centroids calculated for tryptophan, alanine and glycine (residue position 81 in PDB IDs: 3BX7, 4GH7 and 6S8V). The centroid for alanine corresponds to  $C\beta$  (i.e., the methyl group), for glycine it is  $C\alpha$ . (B) Side chain variance visualized as an ellipsoid (light blue) together with its principal axes, here calculated from the covariance matrix for four side chain centroids (residue position 73 in PDB IDs: 3BX7, 4GH7, 6S8V and 6Z2C).

corresponding mean coordinate. These values reflect the local structural deviations of the main chain for the engineered as well as the natural lipocalins.

In the next step, for each residue side chain (i.e. all non-hydrogen atom positions beyond  $C\alpha$ ) the unweighted center of mass was calculated. In the case of the amino acid glycine, which lacks a side chain, the  $C\alpha$  position was used. Again, the arithmetic mean position of all 18 side chain centers of mass was calculated, together with the sample standard deviation of the distances of these individual positions to the mean coordinate. Furthermore, for each amino acid position in the polypeptide chain the three-dimensional covariance matrix was calculated for these 18 side chain centers of mass using the function 'numpy.cov()' as implemented in Chimera (see <https://numpy.org>). These data reflect both the local structural side chain deviations for the engineered as well as the natural lipocalins and their chemical diversity (with side chains differing in size) in the case of the Anticalin group (Fig. 2).

For both groups of proteins a coordinate set was created comprising the artificial coordinates of the mean  $C\alpha$  atom positions as  $C\alpha$  trace of the polypeptide chain A, together with the sample standard deviations stored as instance attribute 'bfactor' for the class 'Atom', thus replacing the data type normally representing the crystallographic temperature B-factors. In addition, a chain X was appended with the side chain center coordinates as well as the corresponding covariance matrices, this time replacing the anisotropic temperature B-factors (PDB, 2008). These data sets were visualized with Chimera by displaying the mean  $C\alpha$  positions as spheres and drawing lines to the corresponding side chain center positions (Fig. 3). On top of that, the  $C\alpha$  standard deviations were visualized using tubes with varying proportional radii for the polypeptide backbone, and the three-dimensional covariances of the side chains were displayed as ellipsoids using the built-in function 'aniso' of Chimera to visualize anisotropic temperature B-factors.

In another representation, the distances in space from the individual  $C\alpha$  atoms of each of the 18 engineered or the natural lipocalins to the corresponding mean  $C\alpha$  position and, similarly, the corresponding

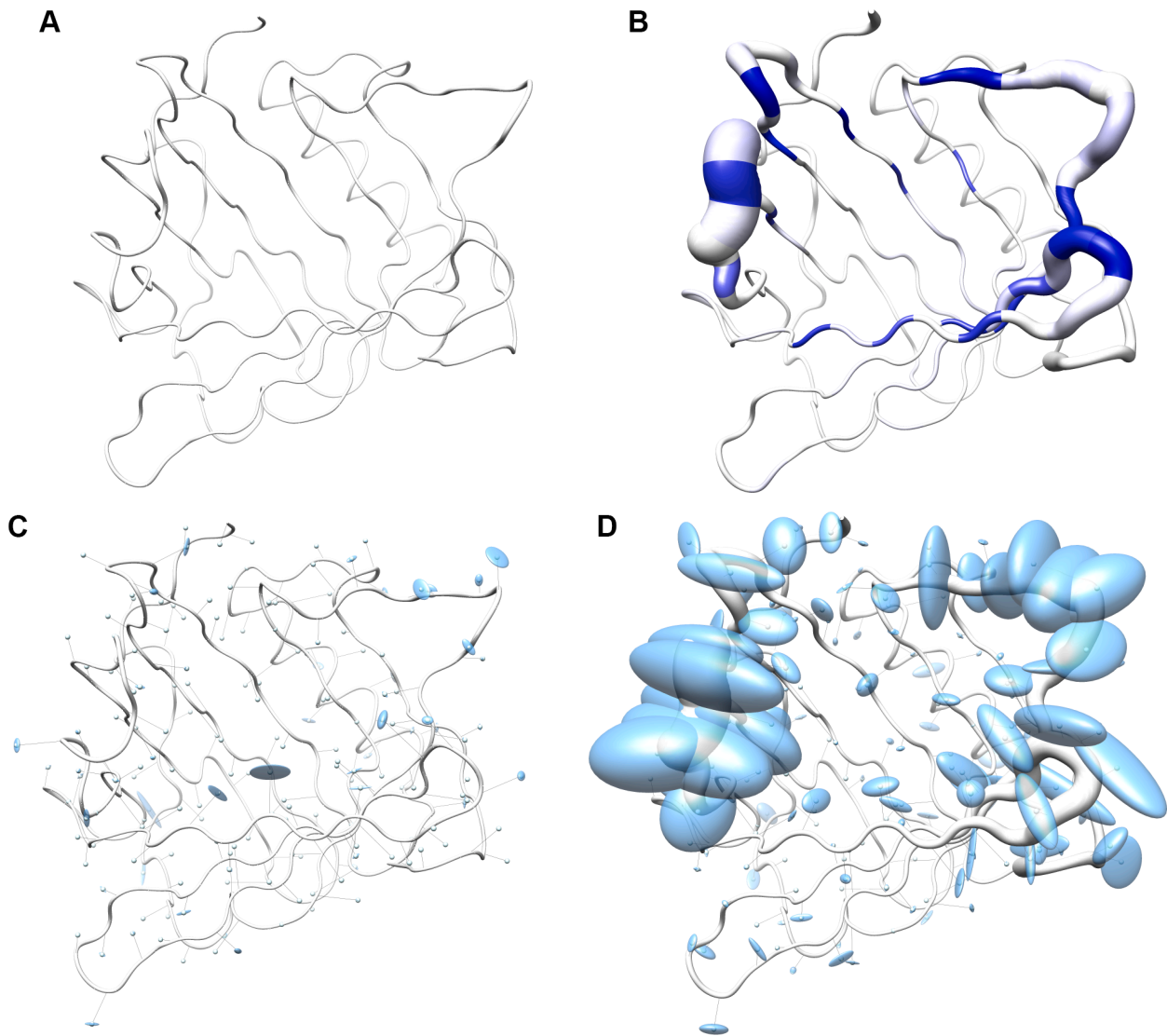
deviations of the side chain center positions, were plotted against the lipocalin amino acid sequence positions using Gnuplot 5.0 (<http://www.gnuplot.info>) (Fig. 4). In these graphs, the standard deviations calculated for each group of 18 proteins were included for reference. These plots reveal remarkable structural variance both for the polypeptide main chain and the side chains within the group of the engineered Anticalins. Interestingly, at both levels, i.e. main chain and side chain, this results in four pronounced peak areas which coincide with those structurally variable loops that were previously defined based on a structural comparison of different natural members of the lipocalin protein family with known three-dimensional structures (Skerra, 2000) – notably, lipocalins originating from different species and recognizing diverse ligands. Thus, the engineered Anticalins with their spectrum of novel ligand specificities, here exclusively based on human Lcn2, appear like a new class of natural lipocalins.

Interestingly, the precise location of the mutated side chains in the individual Anticalins, which are scattered across these four loops, does not seem to locally modulate the structural variance; rather, each loop region conformationally responds as a whole to the varying set of amino acid exchanges that were introduced *versus* wtLcn2. The deviation in individual Anticalins from the mean residue position can be as large as 16.8 Å for the main chain ( $C\alpha$ ) and 19.2 Å for the side chains centers. This contrasts with a similar plot for the natural Lcn2 which was crystallized with various ligands and in different crystal forms and space groups, where the corresponding deviations are much smaller, with  $\leq 1.0$  and  $\leq 3.6$  Å, respectively. This supports the notion of conformational flexibility and structural plasticity as two different phenomena in the case of engineered as well as natural lipocalin proteins (Skerra, 2000; Skerra, 2003). In fact, for the natural Lcn2 structural flexibility is very low – both in the conserved  $\beta$ -barrel part and in the loop regions that shape its ligand pocket – regardless of the changing environment of the protein in different crystal forms or if differing ligands are bound.

On the other hand, if a certain number of amino acid side chains in this protein scaffold are replaced, such as in the Anticalins investigated here, the entire loop region gets reshaped, thus creating binding pockets for ligands diverse in size and shape (Deuschle et al., 2021). Interestingly, however, the (hypothetical) mean  $C\alpha$  trace of the group of Anticalins still closely resembles the one of wtLcn2 (as seen for the low deviation of the thick green line representing wtLcn2 in Fig. 4B and 4D), which means that the natural lipocalin still constitutes a consensus structure for the ensemble of its mutated versions, i.e. the Anticalins. This feature provides a biophysical link to the recognition of consensus sequences representing the energetic minimum of the Ig fold (Steipe et al., 1994) as well as of other protein classes whose members share high homology (Sternke et al., 2020).

The pronounced structural variability among the Anticalins is even better illustrated when looking at a superposition of the engineered *versus* the natural lipocalin structures in three-dimensional space (Fig. 3). Here, the backbone deviations, represented by a tube with varying radius, substantially coincide with the structural deviations of the side chains as depicted by ellipsoids. Notably, while for randomized positions these ellipsoids both depict changes in side chain size and in orientation, conformational changes alone give rise to similarly extended ellipsoids in neighboring positions that harbour conserved residues (cf. Fig. 1). Thus, while merely about half of all possible residues within the loop region of the wtLcn2 scaffold were actually randomized in the different Anticalin libraries described so far (Fig. 5) – in order to allow efficient physical sampling of the resulting sequence space as explained elsewhere (Richter et al., 2014) – many more residues within the loop region structurally respond to the mutagenesis. This





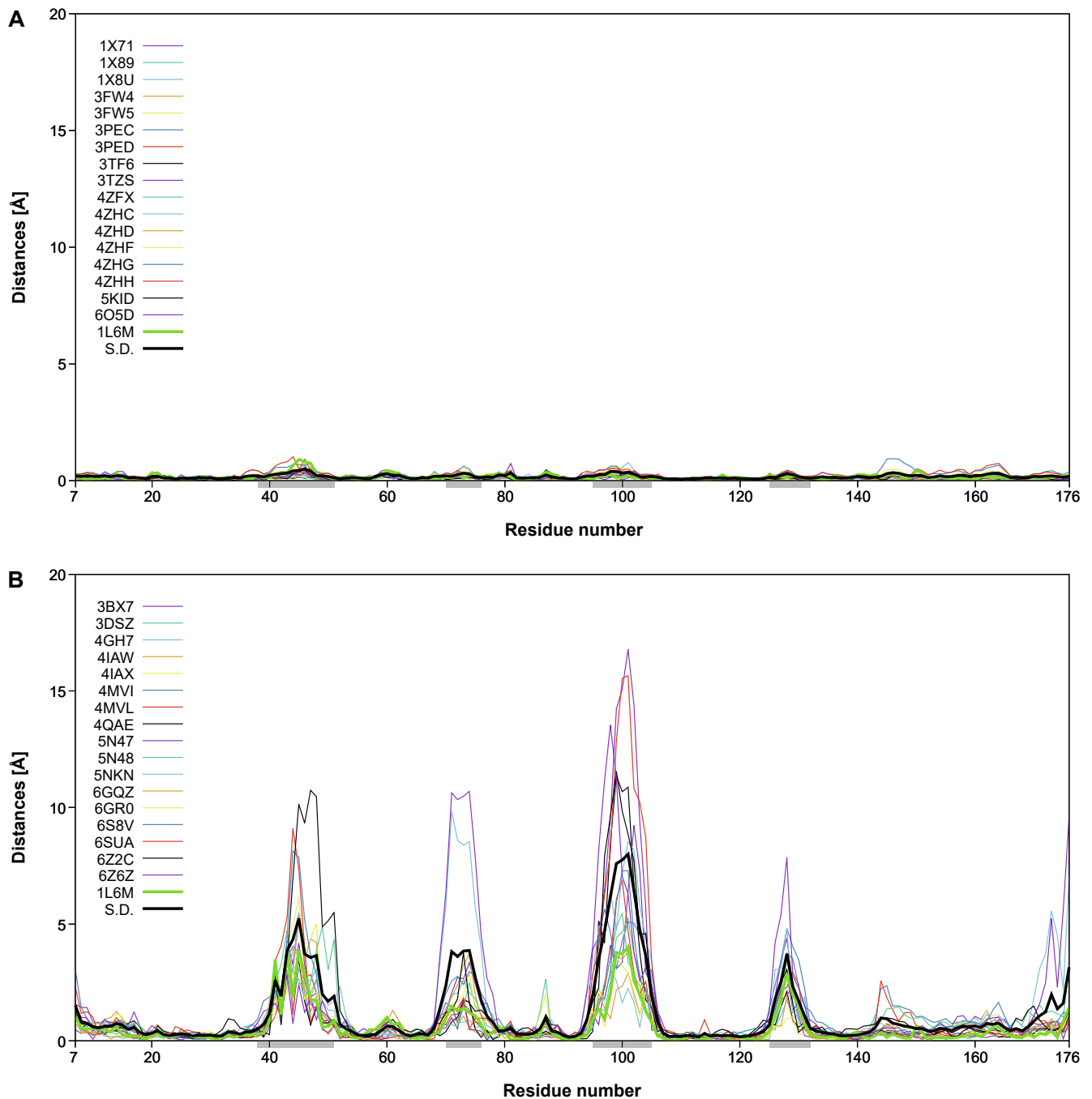
**Fig. 3.** Visualization of backbone and side chain variance for engineered *versus* natural lipocalins in three-dimensional space. The backbone variance is defined as the standard deviation of the  $C\alpha$  distances to their mean positions calculated for a set of 18 representative protein structures in both groups. The corresponding tube has a proportional radius scaled from 0.1 Å to 2 Å. Note that the tube varies in radius only while the calculated standard deviations include both radial as well as axial displacements. In contrast, the spatial variance of the side chains is depicted as translucent ellipsoids (after deriving the side chain mass centers as explained in Fig. 2). (A) Tube representation for the wtLcn2 scaffold, which shows low variance for the backbone despite of its crystallization with different ligands and in different crystal forms. (B) Tube representation for Anticalins, which exhibit high variance for the backbone, apparently due to the mutagenesis in the loop region (see Fig. 1). The intensity of the blue color illustrates the number of different amino acids at a position observed within the data set (from 0, for unmutated residues, up to 10, for fully randomized positions; see Fig. 1). (C) Tube representation with side chain ellipsoids for the natural lipocalin scaffold from (A), again indicating only minor positional variations for a few side chains. (D) Tube representation with side chain ellipsoids for the Anticalins from (B). In (C) and (D), vectors are drawn from the mean  $C\alpha$  position to the center of each corresponding side chain ellipsoid. To this end, cardinal path calculation without smoothing was applied in Chimera to follow the guide atoms (mean  $C\alpha$  positions) exactly. For the ellipsoids a scale factor of 1.0 was applied.

results in fully reshaped ligand pockets, as noted before (Schönfeld et al., 2009), and explains why the engineered lipocalins can efficiently bind diverse ligands representing different classes, such as proteins, peptides and small molecules (Deuschle et al., 2021; Richter et al., 2014).

Of course, it would be interesting to see to which extent this structural deviation is attributable to the effect of ligand binding. In fact, a case of induced fit had been described for an Anticalin selected against the immunological checkpoint receptor CTLA-4 (Schönfeld et al., 2009) where loops #3 and #4 changed their conformations considerably when comparing the crystal structure of the unbound Anticalin with the one of the ligand complex, and loop #3 even appeared structurally disordered in the absence of CTLA-4. On the other hand, in the case of Anticalins

selected against the hapten-type ligand Y-DTPA, the crystal structures of two related mutants solved in the presence or absence of the ligand revealed more subtle effects, with relevant conformational changes only visible for loop #1 (Kim et al., 2009). Crystallographic analysis of an Anticalin engineered to tightly complex and scavenge the siderophore petrobactin revealed even fewer differences upon ligand binding (Dauener et al., 2018). However, so far the number of Anticalins that have been structurally characterized both in the ligand-bound and the free state is still too small to allow a dedicated comparison of the kind presented here.

In conclusion, the general feature of a loop region with pronounced structural plasticity, including high tolerance towards side chain



**Fig. 4.** Structural variability of main chain ( $C\alpha$ ) and side chain center positions in a group of 18 engineered lipocalins (Anticalins; see Table 1) versus the wtLcn2 scaffold (as structurally elucidated in 18 different crystallographic environments; see text). The distances between the individual  $C\alpha$  or side chain center positions and their arithmetic mean positions (for the corresponding group of 18 structures) are plotted for each Anticalin/lipocalin against the residue number in the polypeptide chain. The sample standard deviation for the 18 plotted traces (black) as well as the trace for wtLcn2 (green) are highlighted as thick lines. The four structurally variable loop regions (38–51, 70–76, 95–105, 125–132) are marked as light grey horizontal bars underneath the x-axis. (A)  $C\alpha$  distances and (C) side chain center distances for wtLcn2. (B)  $C\alpha$  distances and (D) side chain center distances for the Anticalins.

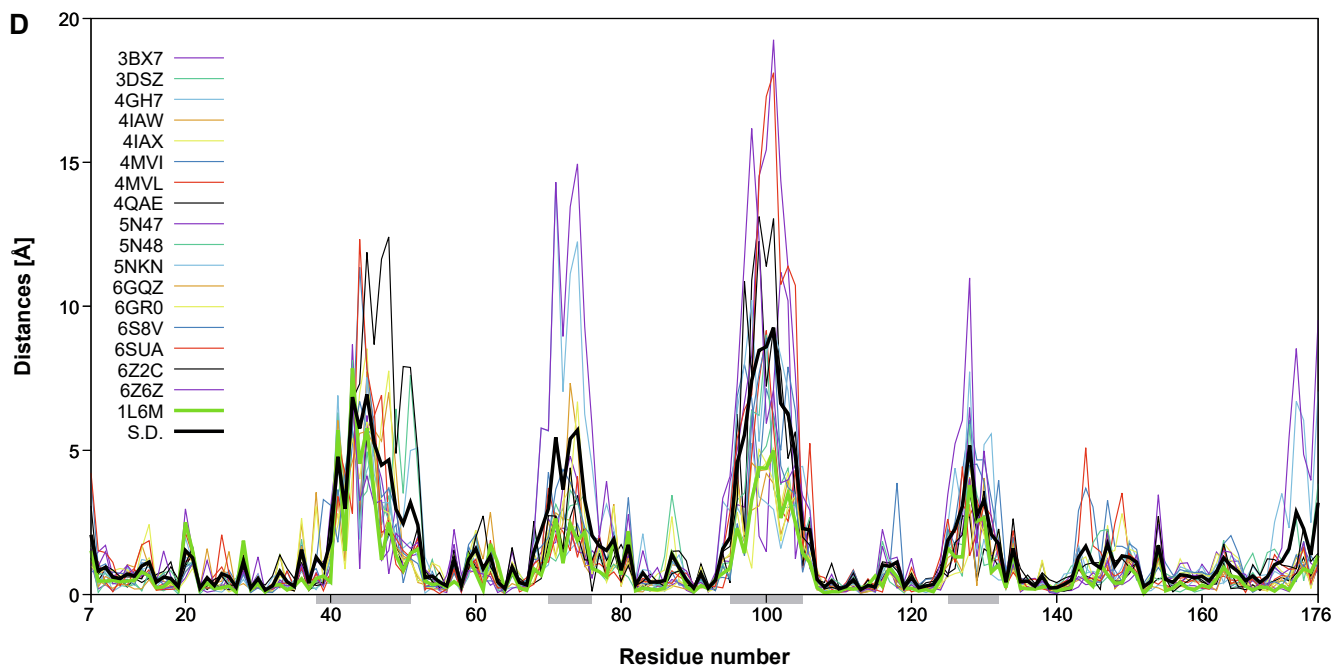
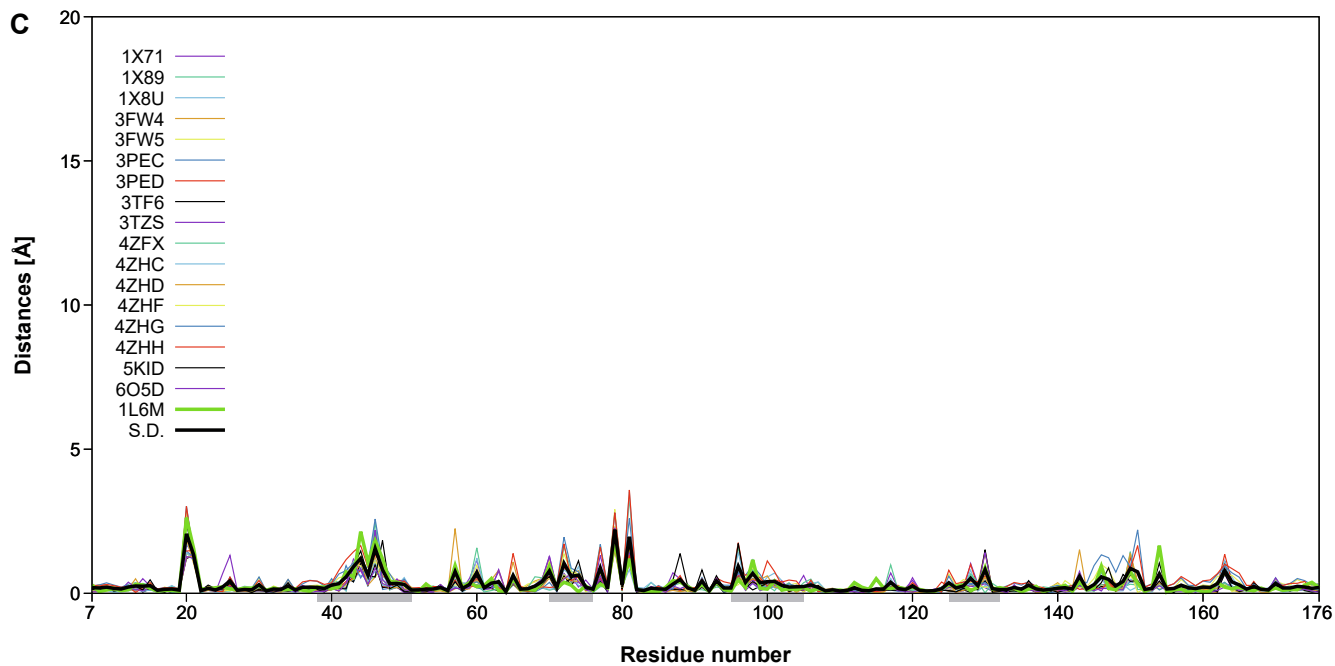
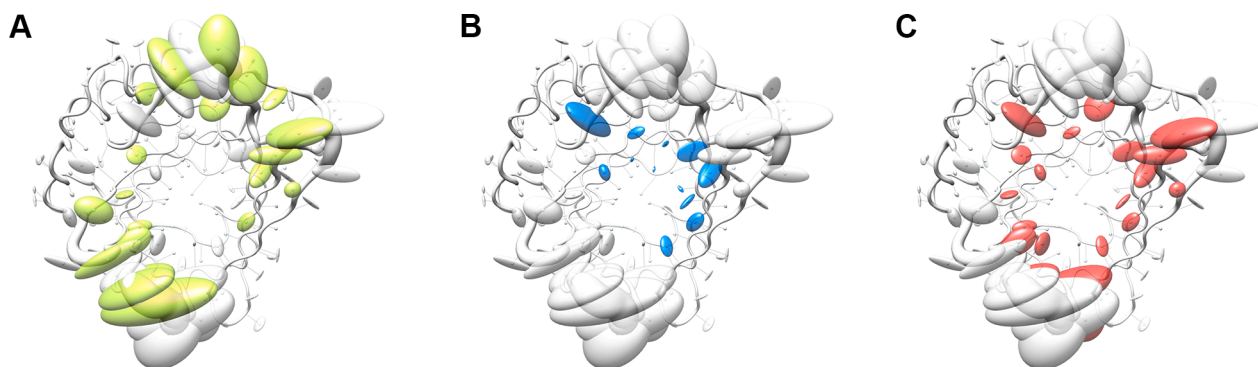


Fig. 4. (continued).



**Fig. 5.** Illustration of mutated amino acid positions in different Lcn2-based random libraries using similar graphical visualization as in Fig. 3D. (A) First-generation library designed to select Anticalins specific for protein targets (Schönfeld et al., 2009). (B) Second-generation library designed to select Anticalins specific for small-molecule targets (Kim et al., 2009). (C) Third-generation library designed to select Anticalins against protein and peptide antigens as well as hapten targets building upon structural analyses of Anticalins successfully selected from the first two libraries (Gebauer et al., 2013).

replacements, as noted early on (Skerra, 2000), distinguishes the lipocalin scaffold from other alternative binding proteins that are derived from proteins with more rigid secondary structural elements, for example Affibodies and DARPinS (Gebauer and Skerra, 2019). Taken together, the structural plasticity of their binding sites, as illustrated in this graphical review, renders engineered lipocalins functionally more similar to Igs, or antibodies, regarding the ability to specifically recognize and tightly bind vastly diverse antigens.

#### Declaration of Competing Interest

A.S. is cofounder and shareholder of Pieris Pharmaceuticals, Inc. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Anticalin® is a registered trademark of Pieris Pharmaceuticals GmbH.

#### References

- Åkerström, B., Borregaard, N., Flower, D.A., Salier, J.-S., 2006. Lipocalins. Landes Bioscience, Georgetown, Texas.
- Beste, G., Schmidt, F.S., Stibora, T., Skerra, A., 1999. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *PNAS* 96 (5), 1898–1903.
- Clifton, M.C., Rupert, P.B., Hoette, T.M., Raymond, K.N., Abergel, R.J., Strong, R.K., 2019. Parsing the functional specificity of Siderocalin/Lipocalin 2/NGAL for siderophores and related small-molecule ligands. *J. Struct. Biol.* X 2, 100008.
- Dauner, M., Eichinger, A., Lücking, G., Scherer, S., Skerra, A., 2018. Reprogramming human siderocalin to neutralize petrobactin, the essential iron scavenger of *Anthrax bacillus*. *Angew. Chem. Int. Ed. Engl.* 57 (44), 14619–14623.
- Deuschle, F.-C., Ilyukhina, E., Skerra, A., 2021. Anticalin® proteins: from bench to bedside. *Expert Opin. Biol. Ther.* 21 (4), 509–518.
- Diez-Hermano, S., Ganfornina, M.D., Skerra, A., Gutiérrez, G., Sanchez, D., 2021. An evolutionary perspective of the lipocalin protein family. *Front. Physiol.* 12, 718983.
- Flower, D.R., 1996. The lipocalin protein family: structure and function. *Biochem. J.* 318 (1), 1–14.
- Gebauer, M., Skerra, A., 2012. Anticalins: small engineered binding proteins based on the lipocalin scaffold. *Methods Enzymol.* 503, 157–188.
- Gebauer, M., Skerra, A., 2019. Engineering of binding functions into proteins. *Curr. Opin. Biotechnol.* 60, 230–241.
- Gebauer, M., Schiefner, A., Matschiner, G., Skerra, A., 2013. Combinatorial design of an Anticalin directed against the extra-domain B for the specific targeting of oncofetal fibronectin. *J. Mol. Biol.* 425 (4), 780–802.
- Goetz, D.H., Holmes, M.A., Borregaard, N., Bluhm, M.E., Raymond, K.N., Strong, R.K., 2002. The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol. Cell* 10 (5), 1033–1043.
- Jarasch, A., Kopp, M., Eggenstein, E., Richter, A., Gebauer, M., Skerra, A., 2016. ANTICALIGN: visualizing, editing and analyzing combined nucleotide and amino acid sequence alignments for combinatorial protein engineering. *Protein Eng. Des. Sel.* 29 (7), 263–270.
- Kim, H.J., Eichinger, A., Skerra, A., 2009. High-affinity recognition of lanthanide(III) chelate complexes by a reprogrammed human lipocalin 2. *J. Am. Chem. Soc.* 131 (10), 3565–3576.
- PDB, 2008. Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description Version 3.20 – Document Published by the wwPDB.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25 (13), 1605–1612.
- Richter, A., Eggenstein, E., Skerra, A., 2014. Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* 588 (2), 213–218.
- Rothe, C., Skerra, A., 2018. Anticalin® proteins as therapeutic agents in human diseases. *BioDrugs* 32 (3), 233–243.
- Schiefner, A., Skerra, A., 2015. The menagerie of human lipocalins: a natural protein scaffold for molecular recognition of physiological compounds. *Accounts Chem Res* 48 (4), 976–985.
- Schönfeld, D., Matschiner, G., Chatwell, L., Trentmann, S., Gille, H., Hülsmeier, M., Brown, N., Kaye, P.M., Schlehüser, S., Hohlbaum, A.M., Skerra, A., 2009. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. *Proc. Natl. Acad. Sci. USA* 106 (20), 8198–8203.
- Skerra, A., 2000. Lipocalins as a scaffold. *Biochim. Biophys. Acta* 1482 (1–2), 337–350.
- Skerra, A., 2001. 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. *J. Biotechnol.* 74 (4), 257–275.
- Skerra, A., 2003. Imitating the humoral immune response. *Curr. Opin. Chem. Biol.* 7 (6), 683–693.
- Steipe, B., Schiller, B., Plückthun, A., Steinbacher, S., 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* 240 (3), 188–192.
- Sternke, M., Tripp, K.W., Barrick, D., 2020. The use of consensus sequence information to engineer stability and activity in proteins. *Methods Enzymol.* 643, 149–179.
- Strohl, W.R., Strohl, L.M., 2012. *Therapeutic Antibody Engineering: Current and Future Advances Driving the Strongest Growth Area in the Pharmaceutical Industry*. Woodhead, Cambridge, UK.