Anticalins: Exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins

A. Richter, E. Eggenstein, A. Skerra

Munich Center for integrated Protein Science (CIPS-M) and Lehrstuhl für Biologische Chemie, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85350 Freising-Weihenstephan, Germany

ABSTRACT

Antibodies, which can recognize a plethora of possible antigens, have been considered as a paradigm of protein engineering performed by nature itself. Lipocalins constitute a distinct family of proteins with functions in ligand binding and transport that occur in many organisms, including man. Like antibodies, lipocalins exhibit a structurally conserved framework – a β-barrel with an attached α-helix – which supports four structurally hypervariable loops forming a cup-shaped binding site. Thus, lipocalins offer an ideal platform for protein engineering to generate novel binding reagents. Using recombinant/synthetic DNA technology and methods of combinatorial library selection, ‘Anticalins’ with prescribed target specificities can be easily generated. Anticalins with picomolar affinities have been developed for three classes of ligands having relevance in basic research and/or medical application: small molecules, peptides, and proteinaceous signalling molecules as well as cell surface receptors. Anticalins derived from human lipocalins have already reached the clinical trial stage. Due to their very small size and simple composition of a single polypeptide chain, which also facilitates the construction of bifunctional fusion proteins, Anticalins promise benefits as a next class of biopharmaceuticals.

© 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Lipocalins as an alternative protein scaffold to immunoglobulins

Antibodies, also known as immunoglobulins (Igs), represent a widely accepted class of biochemical reagents. Due to their specific and tight binding properties against almost all kinds of substances (antigens and haptenes) they have provided valuable tools for biomedical research and therapy for more than a century. Antibodies were initially made amenable to protein engineering by the development of methods for the production of functional Ig fragments in Escherichia coli [1,2], which laid the foundation of current antibody technology. Today, more than 25 Ig-based products are approved for therapeutic use, mainly for the treatment of cancer and autoimmune disorders [3].

While naked monoclonal antibodies (MAbs) – and also their conjugates with radioactive payloads or probes – have already been in use for cancer treatment or in vivo imaging for several years, this toolbox was recently expanded by the development of more specialized formats such as bispecific MAbs or antibody-drug conjugates (ADCs) [4]. To reduce immunogenicity and maintain immunological effector functions, MAbs of murine origin have been replaced by chimeric, humanized or even fully human formats in almost all therapeutic settings. However, owing to their relatively large molecular size of about 150 kDa, their nature as multiple-disulfide-crosslinked and glycosylated proteins and their composition of in total four polypeptide chains (two heavy and two light chains), whole Igs can be efficiently produced only by eukaryotic cells.

From a structural point of view, the antigen-binding site of an antibody is formed by altogether six hypervariable loop regions (also known as complementarity determining regions, CDRs), three in each of the two variable domains (Fig. 1). The CDRs are displayed on a conserved framework comprising a sandwich of disulfide-linked β-sheets which form the typical Ig fold [5]. The high amino acid sequence variability within the six CDRs accounts for the vast number of possible antigen specificities and genetically arises from somatic recombination of an inherited set of gene cassettes in conjunction with hypermutation.

Despite the success of antibodies in many applications, the molecular format of Igs is also associated with some practical disadvantages from a protein engineering perspective and with regard to clinical use: (i) because of the complex architecture the production of full-length Igs is expensive and laborious as complicated mammalian expression systems are usually necessary; (ii) immunological effector functions mediated by the Fc region can...
lead to undesired side reactions during certain therapeutic applications; (iii) poor tissue penetration due to the large molecular size hampers successful treatment of solid tumors; (iv) the long circulation in blood resulting from both the large size and FcRn-mediated endosomal recycling is unfavorable for in vivo imaging and also for therapeutic applications that require prompt adjustment of dosing.

While some of these disadvantages may be overcome by using recombinant Ig fragments or so-called single domain antibodies from cameloids or sharks [6–8], these formats still come with their own peculiarities, such as the necessary chain pairing of Fab, the oligomerization tendency of scFv and the bisected combining site (comprising merely three CDRs) of the single domain antibodies. Given these constraints, the idea of mimicking the highly specific antigen-binding site of Igs by using a different protein scaffold that exhibits structurally variable loops appeared attractive. In this regard, the lipocalin protein family with its simple molecular architecture, small size and beneficial production characteristics offered the potential to engineer novel highly specific protein receptors by using combinatorial mutagenesis and molecular selection methods.

The lipocalins represent a diverse family of robust extracellular proteins, each comprising a single polypeptide chain of 150–190 amino acids. They are abundant in humans, plants, vertebrates, insects, bacteria as well as other genera [9]. So far, twelve different lipocalins have been identified in man [10], where they are secreted into the blood or tissue fluids and serve for the transport and sequestration of typically hydrophobic and/or chemically sensitive compounds. One well-characterized physiological ligand is the human lipocalin 2 (Lcn2), also known as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin [19], and the human lipocalin 1 (Lcn1), also known as tear lipocalin (Tlc)[18], and the human lipocalin 1 (Lcn1), also known as tear lipocalin (Tlc)[18], and the human lipocalin 2 (Lcn2), also known as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin [19].

Despite surprisingly low mutual sequence homology [11], lipocalins share a conserved three-dimensional fold that is dominated by eight antiparallel β-strands which form a circularly closed, cup-shaped β-barrel core with a C-terminal α-helix attached to its site (Fig. 1). At the open end of the cup (greek: calyx) the β-strands are connected in a pair-wise manner by four structurally variable loops that, together with adjoining residues within the β-barrel, form the natural ligand-binding site. Thus, the structural composition of lipocalins with their β-barrel framework which supports an exposed set of loops is remarkably similar to the format of the paired variable domains carrying the antigen-binding site of antibodies [12].

However, there are three notable differences between lipocalins and Igs. First, lipocalins are not diversified by somatic gene recombination/hypermutation mechanisms as provided for antibodies by the immune system. Second, lipocalins exhibit four hypervariable loops instead of six for intact Igs (or three for single domain antibodies); thus, a simple loop grafting from Igs to lipocalins, as performed during antibody humanization, for example [13], is not possible. Third, compared to the combining site of Igs, which usually forms a rather flat interface, lipocalins possess a characteristic deep ligand pocket and, consequently, they are particularly suitable for the tight binding of small, hapten-type compounds or peptides, beside larger macromolecular antigens (Fig. 2).

2. Engineered lipocalins for recognition of prescribed molecular targets

Based on the conceptual similarity with Igs described above, we set out to generate lipocalins with novel ligand-binding functions, dubbed Anticalins [12,14], using the methods of combinatorial protein engineering, i.e. targeted random mutagenesis of the variable loop regions in combination with powerful library selection techniques such as phage display [15]. To this end, four natural lipocalin scaffolds appeared of special interest on the basis of their known three-dimensional structures: the bilin-binding protein (BBP) from the butterfly Pieris brassicae [16], the human apolipoprotein D [17], the human lipocalin 1 (Lcn1), also known as tear lipocalin (Tlc) [18], and the human lipocalin 2 (Lcn2), also known as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin [19].

BBP is a blue pigment protein that protects insects from oxidative stress, which was chosen as a model lipocalin during initial studies to engineer binding sites with novel target specificities. Randomization of 16 defined amino acid positions within the four hypervariable loops – as well as adjoining parts of the β-strands – was well tolerated by this scaffold and resulted in the generation of the first Anticalins with high affinities for the dye fluorescein [14], the plant steroid digoxigenin [20] and a phthalic ester plasticiser [21], all with affinities in the low nanomolar range. These original findings demonstrated the high tolerance of the lipocalin scaffold for artificially introduced side chain substitutions on a wider scale and reconfirmed our notion.

Fig. 1. Comparison of Ig Fv fragments (left) with Anticalins (right). Both comprise a structurally conserved framework (grey) supporting a set of hypervariable loops (colored) that account for target recognition. Fv moieties from six crystal structures of Igs directed against haptens (PDB codes 3QCJ and 3TV3), peptides (PDB codes 3UJJ and 3UJI) or protein antigens (PDB codes 3TCL and 4JDV) were superimposed on the basis of their framework regions. The wild-type lipocalin 2 (Lcn2), which originally binds the metal-chelate FeIII-enterobactin [19], was superimposed with five Lcn2-derived Anticalins (see Fig. 2) recognizing a hapten, two peptides and two protein targets by using a set of 58 Cα positions that are structurally conserved in the β-barrel framework [12].
of the fundamental similarity between the ligand pocket of lipocalsins and the antigen-binding site of antibodies. In case of the digitoxigenin- and fluorescein-specific Anticalins the mechanism of ligand-binding to the engineered lipocalins was investigated in greater detail via site-directed mutagenesis and crystal structure analysis [22–24].

Consequently, the next step was the design of Anticalins based on human lipocalins, in particular Lcn1 and Lcn2, to ensure high tolerability to patients upon medical application [15]. Such Anticalins are generated via combinatorial protein engineering in a four step process that in principle resembles the humoral immune response against an antigen:

First, a genetic library with specifically randomized amino acid positions is prepared. As the number of side chains that form the ligand-binding site in the lipocalin scaffold, which lies in the range between 30 and 50 residues, is far larger than the combinatorial sequence space that can be physically realized [25], the careful choice of the most relevant set of 16–20 positions plays a crucial role (Fig. 3). Several Anticalin libraries have been constructed up to now, for example on the basis of the BBP, Lcn1, Lcn2 and some other human lipocalin scaffolds. In the case of Lcn2, the library design has been subject to iterative optimization [26] involving successive X-ray structural analyses of selected Anticalins as described further below.

Second, lipocalin variants with specific ligand-binding properties are then selected from the naïve random library. So far, the phage display technique has been successfully applied to isolate Anticalins with low nanomolar and sometimes even picomolar dissociation constants from combinatorial lipocalin libraries with complexities of $4 \times 10^8$ [14] to $2 \times 10^{10}$ [26]. While phage display selection via biopanning usually only leads to enrichment of a population of protein variants with some binding function, high throughput screening (HTS) and ELISA techniques subsequently allow the identification of individual Anticalins having specific binding activity, which is further confirmed by including dummy targets as negative controls in parallel experiments [15].

Resulting candidate Anticalins are then expressed at the shake flask scale, purified by affinity chromatography and analyzed (i) for monomeric solution properties via size exclusion chromatography (SEC) and (ii) for target affinity using real-time surface plasmon resonance (SPR) spectroscopy or similar techniques [26]. Recently, Anticalins have also been made amenable to bacterial surface display [27]. In this case, the library of mutated lipocalins is presented on the outer membrane of the Gram-negative bacterium *E. coli* fused to the integral membrane domain of a bacterial autotransporter protein. The lipocalin scaffold appears particularly suitable for this approach, which has been unsuccessful with antibody fragments, probably due to their delicate disulfide bond pattern. This method allows the rapid and convenient selection of cognate Anticalins via fluorescence-activated cell sorting (FACS) after incubation of the bacterial library with the fluorescently labeled target molecule.

Finally, once an Anticalin with specific molecular recognition properties has been selected, its initial properties can be further improved in a process called in vitro affinity maturation, which again has been inspired by the mechanisms of humoral immune response. To this end, the central coding region of the engineered lipocalin, encompassing all four hypervariable loops, is subjected to error-prone mutagenesis, typically involving the polymerase chain reaction (PCR), and the resulting focused library is subjected

---

Fig. 2. Crystal structures of five Anticalins based on the human Lcn2 scaffold with specificities for different types of targets (blue). Anticalins are depicted in complex with the peptide ligands Aβ40 (orange; to be published) and hepcidin (yellow; to be published), the small hapten-like molecule YIII-DTPA (green; PDB code 4IAX) and with the two protein targets ED-B (magenta; PDB code 4GH7) and CTLA-4 (red; PDB code 3BX7).
to another selection against the same target, this time under more stringent conditions [28]. Selection cannot only be driven for improved target affinity or slower dissociation rate but also towards generic properties such as higher protein stability – e.g. via selection at elevated temperature or in the presence of denaturants – or reduced aggregation. These optimization procedures typically involve iterative cycles, which may be complemented by positional saturation mutagenesis once sensitive residues have been identified in the course of an error-prone PCR approach or, possibly, after elucidation of the three-dimensional structure, thus providing a basis for rational protein engineering [29].

3. Selection of Anticalins against targets of biomedical relevance

During the last years, Anticalins directed against various different target formats, that is proteins, peptides and small molecules, have been developed (Fig. 2). Human Lcn2 has emerged as a particularly versatile scaffold to generate Anticalins with exquisite binding activities and superior biochemical properties regarding protein stability and strictly monomeric behavior. Using an Lcn2-based library designed for the binding of small, hapten-like molecules, Anticalins specifically recognizing a metal–chelate complex comprising a lanthanide ion, e.g. Y\(^{3+}\), and a derivative of diethylentriamine pentaacetic acid (DTPA) were selected [28]. A recently published version of this Me-DTPA-specific Anticalin that has been subject to biochemical optimization [29] is now under pre-clinical investigation for in vivo pretargeting radioimmuno therapy (RIT) and diagnostics. Another Lcn2-based random library was constructed to generate Anticalins capable of tightly complexing voluminous, antigen-type protein targets and successfully applied to the extracellular domain of cytotoxic T lymphocyte-associated antigen 4 (CTLA-4, CD152) as a disease-relevant target [30]. Due to its function as a negative regulatory T-cell co-receptor, the Anticalins with subnanomolar affinities, obtained by in vitro affinity maturation, had potent blocking activity and showed an immunostimulatory effect on the T-cell response in an animal model of infectious disease [31].

Based on these experimental results and on a series of X-ray structures of selected Anticalins in complex with their target molecules (cf. Fig. 2), a third generation Lcn2-based random library was recently described [26]. This library combines a presumably optimal distribution of randomized amino acid positions within the binding site of the lipocalin scaffold, amenable to interaction with target molecules of diverse size and shape, with a trinucleotide-based DNA synthesis technique that avoids stop codons and Cys substitutions as well as codon bias due to the degenerated genetic code.

Using this ‘New’ library, Anticalins that specifically recognize a disease-relevant splice variant of the extracellular matrix protein fibronectin (Fn), exhibiting the so-called extra-domain B (ED-B), were developed with affinities in the single-digit nanomolar range [26]. Since ED-B-positive Fn is specifically expressed during tumor angiogenesis, it has emerged as a promising marker for various cancers and constitutes a validated target for in vivo imaging [32]. These Anticalins exhibit specific staining of ED-B positive cells in immunofluorescence microscopy and of human glioma neovascularature in immunohistochemistry. Hence, owing to their high stability and monomeric biochemical behavior Anticalins show promise for applications in oncology.

A series of other Anticalins have also been selected from this Lcn2-based library which potentially address a broad spectrum of human diseases by recognizing well known targets such as VEGFR-3 and Hsp70 in cancer, the Aβ peptide in Alzheimer’s disease and hepcidin, a negative regulator of iron homeostasis, in anemia [33]. The development of Anticalins against the hepcidin peptide has been funded by the European Commission in frame of the EUROCALIN Consortium to promote clinical investigation (http://www.eurocalin-fp7.eu).

Apart from Anticalins derived from the Lcn2 scaffold, other members of the lipocalin family have been successfully applied to generate highly specific and functionally active Anticalins. The human Lcn1/Tic scaffold served to develop a drug candidate for the treatment of solid cancers that effectively blocks the vascular endothelial growth factor (VEGF-A) [34]. This Anticalin, dubbed Angiocaltm (PRS-050), has been investigated as inhibitor of tumor
angiogenesis in a first-in-human Phase I trial, demonstrating safety and high tolerability as well as lack of immunogenicity [35]. Based on the promising data from this repeated dose escalating study, Angiocatal® is recommended to enter clinical Phase II (http://www.pieris-ag.com). Recently, another Anticalin based on the Lcn1/Tlc scaffold that is directed against the hepatocyte growth factor receptor (HGFR; c-Met proto-oncogene) has been described to act as a highly potent and specific MET antagonist with both ligand-dependent and ligand-independent activity [36].

The BBP lipocalin scaffold (discussed above) has been investigated for biomedical uses, too. The fluorescein-specific Anticalin FluA was applied, after genetic fusion to an Ig Fv fragment that recognized the Fn extra-domain A as angiogenesis marker, to pretargeted payload delivery in a mouse tumor xenograft model [37]. Furthermore, the affinity-improved digoxigenin-binding Anticalin (DigiCal) showed promising results in a rat study as antidote for the treatment of digitalis intoxication [38]. Recently, a new BBP-derived Anticalin was selected via ribosome display to recognize the sex hormone estradiol [39].

4. Present and future trends of the Anticalin® technology

Anticalins constitute a promising class of engineered therapeutic proteins that exhibit key features of a potent, safe and economic drug (Fig. 4). Beside fundamental prerequisites for biopharmaceutical development, including high target affinity (nano- to picomolar range) and exquisite specificity, pronounced protein stability and low immunogenicity as well as facile manufacturing and purification, they offer deep tissue penetration due to their small size of just about 20 kDa. To ensure broad therapeutic applicability, several established techniques are available to prolong the intrinsically short plasma half-life of Anticalins, not only by chemical PEGylation but also via fusion to an albumin-binding domain [20] or by applying PASylation® [40].

Due to their monovalent binding function the risk of intermolecular crosslinking of targets or clustering of cell surface receptors is low, which provides an advantage for the design of potent antagonists. Nevertheless, if required, bivalency can be accomplished by genetic fusion of an Anticalin via its flexible N- or C-terminus either to another Anticalin, yielding a so-called Duocalin [41], or to an Ig fragment or a different kind of binding protein. For biomolecular imaging and detection, Anticalins can be genetically fused to a fluorescent protein or chemically coupled to a dye or a radio-metal chelator. Generally, Anticalins show promise for tumor therapy [33], both as antagonists of growth factors (such as VEGF-A), receptors (such as HGFR) or of immunoregulatory molecules (such as CTLA-4) and, if directed against a tumor cell surface marker, as fusion proteins or conjugates with toxin payloads of various kind.

In summary, Anticalins have potential as next-generation biologics for the treatment and diagnosis of human diseases and as research reagents that may complement and even supersede conventional antibodies in many areas.

References


