JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2000;**13**:167–187

Review

Engineered protein scaffolds for molecular recognition

Arne Skerra*

Lehrstuhl für Biologische Chemie, Technische Universität München, D-85350 Freising-Weihenstephan, Germany

The use of so-called protein scaffolds has recently attracted considerable attention in biochemistry in the context of generating novel types of ligand receptors for various applications in research and medicine. This development started with the notion that immunoglobulins owe their function to the composition of a conserved framework region and a spatially well-defined antigen-binding site made of peptide segments that are hypervariable both in sequence and in conformation. After the application of antibody engineering methods along with library techniques had resulted in first successes in the selection of functional antibody fragments, several laboratories began to exploit other types of protein architectures for the construction of practically useful binding proteins. Properties like small size of the receptor protein, stability and ease of production were the focus of this work. Hence, among others, single domains of antibodies or of the immunoglobulin superfamily, protease inhibitors, helix-bundle proteins, disulphide-knotted peptides and lipocalins were investigated. Recently, the scaffold concept has even been adopted for the construction of enzymes. However, it appears that not all kinds of polypeptide fold which may appear attractive for the engineering of loop regions at a first glance will indeed permit the construction of independent ligand-binding sites with high affinities and specificities. This review will therefore concentrate on the critical description of the structural properties of experimentally tested protein scaffolds and of the novel functions that have been achieved on their basis, rather than on the methodology of how to best select a particular mutant with a certain activity. An overview will be provided about the current approaches, and some emerging trends will be identified.

Keywords: **ligand-binding; molecular libraries; polypeptide fold; protein design; structural framework**

INTRODUCTION: WHAT IS A MOLECULAR 'SCAFFOLD'?

Originally the term scaffold describes a mechanical construction which supports machines and workers during various activities concerning the building, repair or cleaning of an apparatus or structure. In modern biochemistry, especially in the area of protein engineering, 'scaffold' has adopted another meaning on the molecular level. The term protein scaffold is frequently used in a more or less vague way for a type of polypeptide fold that is observed in differing contexts and with distinct biochemical functions, so that it is thought to be amenable to protein engineering purposes because of a certain intrinsic conformational stability. This review will attempt to provide a better definition of this term, which will help in the identification of future scaffolds that might be appropriate for practical use. It will also give an overview about current experimental concepts for their application in the engineering or design of biomolecules with novel functions.

Even though nature has provided us with an immense collection of proteins for use in medical diagnosis, therapy and biotechnology, and the traditional limitations of their supply have been abolished by the development of recombinant DNA techniques, there is no doubt that still more useful proteins will be accessible via the creation of artificial biomolecules. The generation of receptor proteins with specific ligand-binding properties has become a general aim, essentially because of initial successes in the field of antibody engineering, but the design of novel enzymes will certainly be the next step.

The many attempts at designing proteins *de novo* which have been made during the past 15 years have provided

* Correspondence to: A. Skerra, Lehrstuhl für Biologische Chemie, Technische Universität München, D-85350 Freising-Weihenstephan, Germany.
E-mail: Skerra@Weihenstephan.de

Abbreviations used: ABD, albumin-binding domain of protein G; APPI, Alzheimer's amyloid *β*-protein precursor inhibitor; BBP, bilin-binding protein; BPTI, bovine (or basic) pancreatic trypsin inhibitor; BSA, bovine serum albumin; CBD, cellulose-binding domain of cellobiohydrolase I; CD, circular dichroism; Cdk2, human cyclin-dependent kinase 2; CDR, complementarity-determining region; CTLA-4, human cytotoxic T-lymphocyte associated protein-4; FN3, fibronectin type III domain; GSH, glutathione; GST, glutathione *S*-transferase; hIL-6, human interleukin-6; HSA, human serum albumin; IC_{50}, half-maximal inhibitory concentration; Ig, immunoglobulin; IMAC, immobilized metal affinity chromatography; K_D, equilibrium constant of dissociation; K_i, equilibrium dissociation constant of enzyme inhibitor; LACI-D1, human lipoprotein-associated coagulation inhibitor; pIII, gene III minor coat protein from filamentous bacteriophage f1; PCR, polymerase-chain reaction; PDB, Protein Data Bank; PSTI, human pancreatic secretory trypsin inhibitor; RBP, retinol-binding protein; SPR, surface plasmon resonance; TrxA, *E. coli* thioredoxin.

some insights into the mechanism of protein folding; however, they have not resulted in practically applicable biomolecules. The major reason is the well-known protein folding problem and there is no hope that this dilemma will be principally solved in the near future. Consequently, the idea arose to adapt structurally well-defined polypeptide frameworks for the introduction of novel functions by locally reshaping a part of the protein surface that is thought to be less important for the protein folding process or its stability.

Generally, candidates for artificial receptor proteins were considered to be able to present surface loops of varying sequence and length, including exposed hydrophobic residues, without significant changes in the structural framework (Ku and Schultz, 1995). Naturally, the ligand-binding properties of such artificial proteins will depend on the number, spatial distribution, and diversity of the loop regions. According to practical demands they should be based on monomeric and small polypeptides which are robust, easily engineered, and efficiently produced, if necessary at the fermenter scale, in inexpensive prokaryotic expression systems (Nord *et al.*, 1997).

The question appears whether this more or less intuitive approach can be rationalized and some rules can be found or features identified which make it likely that a certain protein fold will indeed be successfully used as a scaffold. For thermodynamic reasons, an optimal binding site for a given ligand should not only exhibit perfect shape complementarity but it should also be rigid because the loss of conformational entropy upon complexation is then minimal. This aspect has been well recognized in the field of peptide ligand selection where various strategies are being used in order to constrain backbone flexibility (reviewed by Marshall, 1992; Hoess, 1993; Cortese *et al.*, 1995).

Hence an ideal protein scaffold should provide a rigid folding unit which spatially brings together several exposed loops, forming an extended interface that ensures tight binding of the target. However, protein folding is a highly cooperative event, especially in the case of small globular proteins, which constitute the practically most desirable candidates. Therefore, how can it be assessed whether such a polypeptide tolerates a certain number of amino acid substitutions in the chosen loop region without losing its ability to adobt a defined tertiary structure or without becoming much less stably folded at least? Classical studies have indicated that the defined sequence and conformation of an exposed loop may be rather important for the folding of a protein, and even though the three-dimensional topology of a polypeptide chain is primarily determined by the packing of its secondary structural elements, its surface turns contribute significantly to the stability of the folded state (Nagi and Regan, 1996; Predki *et al.*, 1996; see also references in Brunet *et al.*, 1993). Consequently, the occurrence of some nicely arranged loops *per se* does not suffice to indicate the presence of a useful scaffold.

The answer is that a generic protein scaffold should exhibit the peculiar feature of having structurally separated the information and stability of its polypeptide conformation on one hand and the local shape and molecular recognition function of its active site on the other (Fig. 1). But how is a protein with such a modular architecture best identified? In the era of genome sequencing we experience the appearance of an increasing number of structural protein families with a common fold. However, high sequence diversity and differing biochemical function in conjunction with a conserved tertiary fold is probably just a necessary but not a sufficient criterion for a protein that may be suitable as a scaffold in the sense of protein engineering.

As will become apparent from the examples discussed herein, two additional criteria seem to characterize an applicable scaffold: (i) the protein family should possess a well-defined hydrophobic core, which is structurally superimposable among its individual members and can provide a major contribution to the free energy of folding; (ii) it should possess a solvent-accessible active site or binding pocket, which is spatially well separated from the core, and which is ideally involved in the recognition of clearly different targets. This contrasts with a kind of fold that participates in interactions with other biomolecules at separate surface regions.

Once such a promising architecture has been identified, a whole series of currently available biochemical methods may be applied in order to modify it for the desired ligand-binding function. Although rational design experiments have also been carried out, the methods of combinatorial molecular biology are mostly applied. Usually a random library is first generated on the genetic level, whereby substitutions are targeted at amino acid positions within the presumed active site. Second, mutants with the prescribed binding specificities are selected from such a library by means of filamentous phage display and panning in the presence of the ligand (reviewed by Smith, 1991; Wells and Lowman, 1992; Hoess, 1993). More specialized selection techniques may be utilized as well, e.g. colony screening assays (Schlehuber *et al.*, 2000), use of plasmid-binding repressor fusion proteins (Cull *et al.*, 1992) or other classes of phages for display (Huse *et al.*, 1989; Houshmand *et al.*, 1999), the yeast two-hybrid system (Young, 1998) as well as various types of 'ribosome display' (Roberts, 1999).

In this context it is important to be aware of the combinatorial complexity that may be theoretically supported by a certain scaffold in comparison with the practical size of the library, which is usually significantly limited by the experimental setup. Even though it might be considered more satisfactory to work with a redundant library, it should be noted that nature itself employs the concept of sampling, for example in the development of an immune response starting from a huge but clearly restricted set of functionally recombined Ig genes. However, rather than dealing with experimental techniques, which have been extensively reviewed elsewhere, this account will focus on the structural and functional properties of those protein scaffolds which have been investigated so far or are in practical use already.

ANTIBODIES: NATURE'S PARADIGM FOR RECEPTOR PROTEIN ENGINEERING

Immunoglobulins comprise a natural type of biomolecular scaffold which is utilized by the immune system of higher organisms in order to mount an effective humoral response

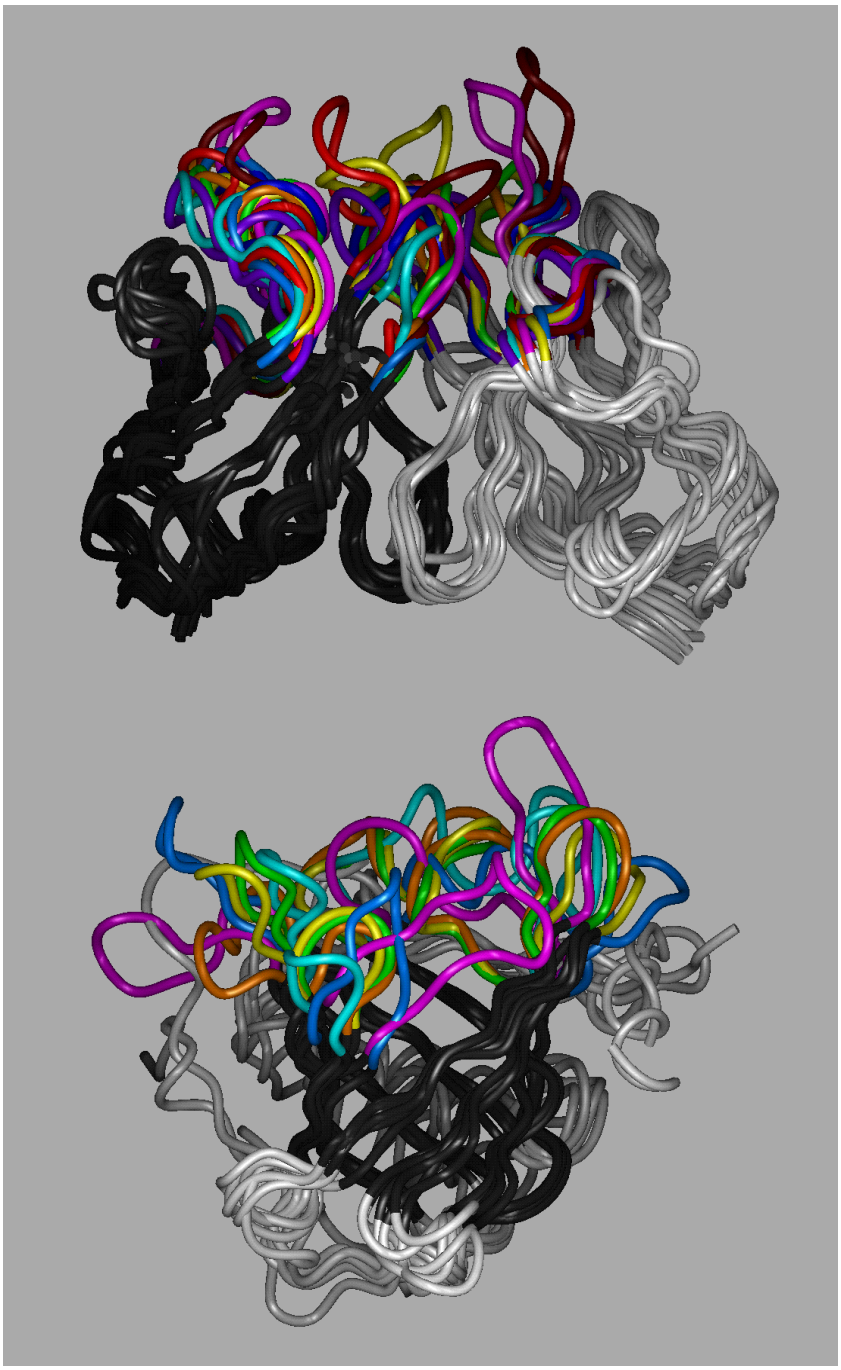

Figure 1. Natural protein scaffolds with a partitioned architecture, i.e. comprising a structurally conserved framework and hypervariable loops: immunoglobulins (top) and lipocalins (bottom). Top: the coordinates (PDB entries 1MCP, 1F19, 2FBJ, 2FB4, 3FAB, 2HFL, 3HFM, 1FDL, 1NCA, 1FLR) of the F_v moieties from ten crystal structures were superimposed (black: V_H frameworks; grey: V_L frameworks). Bottom: the coordinates (PDB entries 1RBP, 1BBP, 1MUP, 1EPA, 1BEB, 1BJ7) of six different lipocalins were superimposed (black: -barrel core; grey: extraneous conserved structural elements).

against pathogenic invaders or their toxins. Approximately 10^8 different 'antibodies' are assumed to circulate in the human body, which is obviously sufficient in order to recognize almost any given foreign substance. Following initial complex formation with the antigen via membrane-attached IgMs on the surface of B-cells, the affinity and specificity of the antibody is continuously raised by cycles of somatic hypermutation and selection, often described as a process of molecular evolution. These events result in the synthesis of proteins with highly evolved molecular recognition properties, which serve as taylored adapter molecules between the foreign substance or organism and receptors on several effector lymphocytes, triggering them for cellular action. Apart from their pivotal role *in vivo* antibodies have become valuable tools in biochemistry and medicine because of their specific and tight binding properties for a huge variety of biomolecular compounds.

Antibodies owe their unique recognition capability to a modular type of structure, which has been well identified (Padlan, 1994; Bork *et al.*, 1994). On the first level they consist of two types of domains which are either of constant or of variable character, although both are based on the structurally well-conserved so-called Ig fold. Several constant domains assemble into a structure denominated F_c, at the stem of the Y-shaped protein, which is responsible for most effector functions, i.e. binding to several cellular receptors or to complement. On each tip of the molecule a pair of variable domains, V_H from the heavy and V_L from the light polypeptide chain, forms the antigen-binding region. Within these variable domains, which are symmetrical about a two-fold axis, there is another level of separation between structurally conserved and variable polypeptide segments.

Altogether six hypervariable loops, also called complementarity-determining regions (CDRs), three within each variable domain, form the combining site. This is the very part of the antibody surface that intimately contacts the antigen upon complex formation. Each set of three loops is supported by a sandwich of two antiparallel β-sheets, with five and four strands, respectively, which forms—together with several conserved loops—the framework region of the variable domain. This discontinuous region exhibits an extremely high conservation of three-dimensional backbone architecture. By means of proper association between V_H and V_L via their five-membered inner β-sheets the hypervariable loops are brought together and form a contiguous surface.

This structural principle explains the success of this protein class as an almost universal molecular tool that is extensively used in the biological sciences for the complexation of a vast number of structures. Obviously the β-sheet fold of the Ig framework is conformationally stable enough in order to structurally support a more or less unlimited number of CDR peptide sequences and to impose a defined three-dimensional structure on them by providing the proper geometrical constraints. As a consequence, the number of possible backbone conformations, called canonical structures, seems in fact to be limited (Al-Lazikani *et al.*, 1997). Nevertheless, the conformational stability of the Ig fold is not arbitrary. Recent studies have shown that the folding energy of the Ig framework statistically varies about some consensus sequence, which appears to constitute the thermodynamic minimum (Steipe *et al.*, 1994). The stability of the whole variable domain is clearly influenced by the sequences that are present within the hypervariable loops (Jung *et al.*, 1999).

The development of methods for the facile production and manipulation of functional Ig fragments in bacterial host systems (Skerra, 1993) has stimulated the field of antibody engineering. Consequently, the scaffold-like properties of the Ig fold have been recognized from the structural point of view. Three types of antigen-binding antibody fragments are commonly in use. The F_{ab} fragment consists of an entire light chain and the variable and first constant domain of the heavy chain (also called the F_d fragment). Due to the tight packing between the pair of constant domains from both chains and an interchain disulphide bond, which is normally present, the F_{ab} fragment is a robust protein with comparable stability as an intact immunoglobulin.

The F_v fragment is the smallest antibody fragment that still carries a functional combining site. It is made of the paired variable regions from both Ig chains, which are held together by non-covalent forces so that its stability is somewhat reduced as compared with an F_{ab} fragment. Chain dissociation may be prevented by introducing Cys residues at appropriate locations into the framework of V_H and V_L in order to form a disulphide crosslink (Glockshuber *et al.*, 1990; Reiter *et al.*, 1996). In another approach the two variable domains are artificially linked via a flexible peptide segment, yielding the so-called single chain, scF_v fragment (Bird and Walker, 1991). Hence, V_H and V_L are covalently held together and the difficulty of coexpressing two different polypeptide chains on the genetic level is partially avoided. Nevertheless, several disadvantages have been associated with this strategy, including hampered folding or stability under practical circumstances, pronounced oligomerization behaviour, and reduced antigen-binding activity as a result of connecting one N-terminus—being close to the combining site in the variable domain—with the peptide linker.

The application of antibody fragments as scaffolds has been promoted by the development of molecular library techniques, especially with the help of bacterial phage display. Initially, libraries of natural Ig sequences, which were prepared from non-immunized or antigen-primed donor organisms, were employed for the *in vitro* selection of artificial antibody fragments, either scF_v or F_{ab}, with binding specificities for prescribed antigens and haptens. Antibody fragments obtained in this way were still largely natural, although new pairings between heavy and light chains were allowed by this technique. The potential of Ig fragments as scaffolds was furthermore exploited by the construction of partially or even entirely synthetic libraries. In this case either a collection of Ig frameworks was combined with one or more synthetic and randomized CDR-encoding regions (mostly CDR-H3) via PCR or one stable framework was equipped with synthetic sets of hypervariable sequences. This kind of work has been reviewed in detail elsewhere (Clackson and Wells, 1994; Hoogenboom, 1997) and will not be further considered here.

The first time that antibodies were purposely used in the sense of a scaffold was in the so-called humanization or CDR-grafting. The set of altogether six CDRs was transplanted from a murine monoclonal antibody onto

human framework regions, resulting in a hybrid F_v moiety retaining the original antigen specificity (Riechmann *et al.*, 1988). Meanwhile this strategy is in widespread use for the construction of antibodies with minimal foreign sequence elements to be applied in human therapy (Hurle and Gross, 1994). The notion of a structural separation into a conserved framework, which provides the proper geometrical constraints, and a functionally diverse active site, made of hypervariable loops, was hence demonstrated in practice, even though later work revealed the importance of peculiar contacts between amino acids in the CDRs and in the framework for full antigen-binding activity.

The approach of CDR-transplantation was taken further in a theoretical study that predicted the grafting of an independent secondary antigen-binding site, the so-called χ-site, onto that part of the F_v surface which diametrically opposes the natural combining site (Keck and Huston, 1996). Although conceptionally attractive, an experimental proof of utilizing the bottom loops of the Ig fold for the development of a 'chimeric bispecific antibody binding site', i.e. χBABS protein, has not yet been described. Nevertheless, an attempt has been made to reshape the original antigen-binding site of an F_v fragment via rational protein design (Essen and Skerra, 1994; Schiweck and Skerra, 1995). The crystal structure of an anti-lysozyme antibody served as a template for the reconstruction of its combining site in order to recognize the thiol-protease inhibitor cystatin as prescribed antigen. Altogether 19 side chains were replaced within the CDRs for this purpose, whereby a five-residue insertion was introduced into CDR-L1. The X-ray crystallographic analysis of the resulting artifical F_{ab} fragment 'M41' revealed indeed close similarity with the modelled structure (Schiweck and Skerra, 1997). However, some local deviations were also seen, explaining the lack of affinity for the new antigen at this stage.

Finally, it was suggested to extend the functional capabilities of the antibody scaffold by altering certain structural features of its combining site. For example, the heavy chain complementarity-determining region 2 (CDR-H2) was elongated beyond the length found in murine germline genes in order to improve the hapten-binding properties of an anti-estradiol antibody (Lamminmäki *et al.*, 1999). Moreover, in the case of the anti-NP [(4-hydroxy-3-nitro-phenyl)acetyl] antibody B1-8 it was found that a loop within framework 3, which neighbours the CDRs adjacent to the combining site, tolerates the insertion of four residues without interfering with hapten-binding. Therefore, it was proposed to use this loop as an additional CDR in order to increase antigen contact beyond that normally possible (Simon and Rajewsky, 1992).

In essence, the importance of the Ig framework for the function of the antigen-binding site in the highly specific and tight complexation of diverse molecular structures by bringing together several peptide segments, being separated in the primary structure, in a geometrically constrained fashion is generally recognized. In contrast, unconstrained single peptides—sometimes called minimal recognition units, MRUs—are usually not able to form tight non-covalent complexes because, first, they provide a limited interaction surface and, second, the loss of conformational entropy upon association with the target significantly reduces the free energy of binding. Consequently, attempts have been made to search for other types of scaffolds, which are still easier to manipulate than antibodies.

SINGLE IG AND IG-LIKE DOMAINS

The use of single Ig domains as a scaffold for the generation of small binding proteins was taken into consideration early. An obvious advantage is that the cloning of two separate coding regions is avoided on the genetic level and, because chain association is not required, the yield of a correctly folded Ig domain can be expected to be higher than that of an F_v or F_{ab} heterodimer. However, although single Ig domains with their three CDRs appear indeed to be useful for the recognition of macromolecular antigens, the disunited Ig scaffold lacks the ability to form pockets, as they are normally necessary for the binding of haptens.

The idea of preparing a library of single V_H domains and searching for members with specific binding properties was first put forward by Ward *et al.* (1989). Several thousand genes were cloned in parallel from spleen genomic DNA of mice immunized with lysozyme or keyhole-limpet haemocyanin and the V_H domains were secreted from *E. coli*. A number of V_H molecules with binding activity against either one antigen or the other was identified and their affinities were found to be in the nanomolar range. However, although these results were encouraging, they seemed to be difficult to reproduce, and in fact the intrinsic stickiness of the isolated Ig domains, their poor solubilities as well as low expression yields abolished further exploitation of this approach.

Generally, an unpaired Ig domain exposes a significant area of hydrophobic surface to solvent, which is otherwise shielded upon association with the partnering domain. This effect causes a pronounced tendency to aggregation (Glockshuber *et al.*, 1990; Davies and Riechmann, 1994) and non-specific adsorption. This situation changed after a peculiar class of antibody molecules was discovered (Hamers-Casterman *et al.*, 1993) which are devoid of light chains so that their V_H domains have obviously evolved to remain soluble without dimerization. Such 'heavy-chain' antibodies, which merely consist of a pair of heavy Ig chains, have been found as a natural subclass of the IgG pool in camels (*Camelus dromedarius*) and other camelids, like llamas (Muyldermans and Lauwereys, 1999).

Sequence analysis of the natural heavy-chain antibodies from camels (Muyldermans *et al.*, 1994) and several X-ray structural analyses of corresponding V_H antibody fragments (later denominated VHH) from camels, dromedaries (Desmyter *et al.*, 1996; Decanniere *et al.*, 1999), and even from llamas (Spinelli *et al.*, 1996) revealed the reasons for the observed high solubility of these Ig domains. The camelid V_H domains have increased surface hydrophilicity, especially in that area which faces the V_L domain in ordinary Igs, and they have a much longer CDR-H3, often participating in an additional disulphide bridge, which partially shields the interface region from solvent.

These findings were adopted by Davies and Riechmann (1995) in order to generate libraries of soluble human V_H domains. A 'camelized' V_H domain was constructed as a template for the preparation of a molecular repertoire starting from a domain belonging to the human VH3 family,

which had previously been characterized as part of a human scF_v fragment and which was known to be well produced in bacteria. Three mutations were introduced within its exposed hydrophobic region so that non-specific binding of the isolated Ig domain through its interface for the V_L domain was prevented. The structural influences of these mutations were later investigated in an NMR study (Riechmann, 1996). CDR-H3 of the modified V_H was then randomized (with 12 degenerate codons) yielding a repertoire of 5×10^5 variants, which were displayed on phage and selected for the binding of two traditional immunological haptens, phenyloxazolone (Ox) and nitrophenacetyl (NIP), conjugated with BSA. With the additional exchange of a Gly by an Ile residue at one of the three deliberately mutated positions, the V_H domains could be secreted as soluble and stable proteins in *E. coli*. Surprisingly, the presence of the Gly residue resulted in a more specific phage enrichment, whereas it hampered the soluble production of the Ig domain. Fluorescence quenching measurements revealed K_D values ranging between 100 and 400 nM for the complexation of the prescribed ligands.

Subsequently, a bigger library of 2×10^8 camelized V_H variants was prepared (as a mixture of libraries with 5–12 randomized residues in CDR-H3) and panned on the immobilized HIV regulatory protein rev (Davies and Riechmann, 1995). Seven clones were selected, and two of them were produced in a soluble form and shown in an ELISA to recognize rev in a specific and concentration-dependent manner, although their affinities were not determined. The same library was later on used in another laboratory for the selection of camelized V_H domains against the NS3 serine protease encoded on the hepatitis C virus genome (Martin *et al.*, 1997). One variant with inhibitory activity, cV_HE2, was isolated and its IC_{50} value was estimated to be 300 nM. Surprisingly, the protein appeared as a dimeric species when characterized by gel filtration, in spite of its 'camelizing' mutations.

Further investigation of the V_H domains that had been selected against the phenyloxazolone group revealed that the folding stabilities were considerably lower than those of a corresponding human VH-Ox13 V_H domain (melting temperature of 56.6 °C compared with 71.2 °C) and an attempt was made to improve one of them by altering four additional amino acids, which were then also thought to be characteristic for the camelid antibodies (Davies and Riechmann, 1996). Determination of their stabilities by reversible thermodenaturation and measurement of ellipticity at 235 nm revealed only marginal effects for the individual mutants. Nevertheless, a new library of 2×10^6 variants was prepared by CDR-H3 randomization. In this case the additional disulphide bridge found in many camel V_H domains had been introduced and one further side chain replaced. Clones with specificities for the phenyloxazolone group (K_D of 246 nM) and for lysozyme (K_D between 1 and 10 μM) were successfully selected and, as a result, their melting points were found to be significantly enhanced (72.6 and 75.3 °C, respectively).

In summary, the 'camelized' human V_H domains might indeed provide a promising scaffold for the generation of small recognition units, especially for the selection of novel enzyme inhibitors (Riechmann and Muyldermans, 1999). However, a useful application has yet to be reported. In addition, their presumed favourable properties in medical therapy, i.e. faster biodistribution and clearance rate (Davies and Riechmann, 1995), must still be proven and it has to be seen whether these may be compromised by an increased immunogenicity due to the alteration of a larger part of the protein's surface, including the unnaturally long CDR-H3 with its extraneous disulphide bridge.

It is worth to be noted in this respect, that human or murine variable domains have been described, which do obviously not depend on the association with another domain—either V_L or V_H—and which could be directly considered as human scaffolds for the engineering of single Ig domains. For example, a melanoma-specific 'V_H antibody' was isolated from an scF_v phage library derived from a patient that had been immunized with genetically modified autologous tumor cells (Cai and Garen, 1996). Furthermore, a natural V_H domain from a mouse hybridoma clone which carried an unusual mutation at one of the interface residues with its V_L domain was successfully used for the preparation of a phage display library with 4×10^8 independent clones after randomization of nine residues within CDR-H3 (Reiter *et al.*, 1999). Panning against either tumour necrosis factor (TNF) or immunoglobulins yielded V_H domains with strong and specific binding activities, which could be efficiently produced and refolded as soluble monomeric proteins from inclusion bodies. In the case of the Ig target the V_H domain was found to be stable against aggregation and exerted a K_D value of 19 nM as determined by SPR measurement.

On the other hand, the monoclonal antibody NEMO, which was obtained by conventional immunization of a mouse, consists solely of κ light chain (Masat *et al.*, 1994). Only the monomeric form of this Ig fragment was shown to be active and to recognize an antigen expressed by human cells of the melanocytic lineage. Contrasting with this, Bence-Jones proteins are a well-characterized species of aberrant Ig molecules that are found in certain disorders of the immune system, consisting of a paired light chain or its variable domain (Stevens *et al.*, 1991). In an approach to controlling and elucidating bioactive conformations of peptides, the dimeric Ig V_L domain REI was successfully utilized as a presentation scaffold (Zhao *et al.*, 1995). The integrin-binding sequence RIPRGDMP was inserted into the CDR-L1 region, thus replacing the SQD wild-type sequence. The resulting molecule, REI-RGD34, which was produced by secretion in *E. coli*, exhibited an IC_{50} value of 130 nM for inhibition of fibrinogen-binding to the integrin $\alpha_{IIb}\beta_3$. From the crystal structure, which was determined at 2.4 Å resolution, it appeared that the inserted RGD sequences showed remarkably similar conformation in both molecules of the dimer occupying the asymmetric unit, which were independently refined. The structural information was then successfully used in order to search small molecule peptidomimetics from a compound database.

TRIMMED IG DOMAINS AND SUBUNITS FROM THE IG SUPERFAMILY

In an attempt to reduce the structural framework of an Ig

domain even further, the so-called minibody was constructed as a 61-residue protein possessing two hypervariable loops (Pessi *et al.*, 1993). The design was based on the heavy chain variable domain of the antibody McPC603 (Skerra and Plückthun, 1988). Just three strands from each of its two β-sheets were included and the central disulphide bond was omitted, but the regions that structurally correspond to CDR-H1 and CDR-H2 were retained. Hydrophobic residues buried in the native molecule, but exposed as a consequence of the trimming, were suitably replaced and one non-CDR loop making extensive contacts with the light chain in the parent antibody was substituted.

Finally, a metal-binding site was engineered by introducing one His side chain into CDR-H1 and two into CDR-H2, using the C_α–C_α distances of the three His residues in the Zn(II)-binding site of carbonic anhydrase B as a guidance. The minibody was produced by solid-phase peptide synthesis. Although it had a rather low solubility of 10 μM, its CD spectrum was consistent with an all-β protein. Some qualitative data on the binding of ^{65}Zn(II) were presented, but no affinity value was determined. Therefore, evidence on the tertiary structure of the minibody remained circumstantial, especially in the light of the crucial role of the missing central disulphide bond in the V_H domain for antibody stability (Glockshuber *et al.*, 1992).

The solubility properties of the original minibody, termed MB1, were improved in a subsequent study by mutagenesis of the β-sheet framework residues and adding some artificial 'solubilizing' motives, made of three Lys residues, at the N- or C-termini (Bianchi *et al.*, 1994). Some of the resulting mutants, which were produced in *E. coli* and solubilized from inclusion bodies, appeared to be soluble up to millimolar levels in aqueous buffer and had an unchanged CD-spectroscopic behaviour.

The minibody MB1 was employed for the construction of a phage display library, followed by selection experiments with human interleukin-6 (Martin *et al.*, 1994). Both hypervariable regions in MB1 were subjected to randomization (six residues each in loop H1 and H2) and a library with approximately 5×10^7 members was prepared. After multiple rounds of selection against immobilized hIL-6, one variant (termed MB02) was identified which still carried the wild-type H1 sequence but whose variegated H2 region revealed remarkable similarity to a short amino acid sequence in the mouse and human IL-6 receptors (IL-6Rα), which is known to play a role in the binding to the cytokine. MB02 was produced by refolding from inclusion bodies and its K_D value for complex formation with hIL-6 was estimated to be 0.23 μM. MB02 did not bind to two structurally related cytokines (human oncostatin M, hOM, and human ciliary neurotrophic factor, hCNTF). However, it inhibited complex formation between hIL-6 and a genetically engineered soluble form of hIL-6Rα *in vitro*, with an IC_{50} of 2 μM. Furthermore, it antagonized receptor binding of hIL-6 and signal transduction in a cell culture assay.

MB02 was optimized in a subsequent study in terms of affinity for hIL-6 and specificity (Martin *et al.*, 1996). For this purpose, the six amino acids in the H1 region were subjected to randomization. Selection was performed via monovalent phagemid display using hIL-6 immobilized in an oriented fashion by means of a cognate monoclonal antibody. In addition, the structurally related cytokine hOM was applied in excess in order to suppress cross-reactivity. After three rounds of affinity selection 24 clones were isolated, sequenced, and the one with the highest level of detectable minibody displayed by the phagemid (MBk), whose sequence was present in four of the clones, was chosen for characterization as a soluble protein after refolding from inclusion bodies produced in *E. coli*.

Its CD spectrum was compatible with a general β-sheet structure. The affinity towards the immobilized cytokine was determined by SPR analysis, yielding a K_D value of 220 nM compared with 730 nM—under the same conditions—for the non-optimized molecule. In addition, MBk inhibited binding of hIL-6 to a recombinant hIL-6Rα fragment with an IC_{50} of 625 nM, a value ca. four-fold lower than that of MB02. In competition assays with the related cytokines hOM and hCNTF, no cross-reactivity of MBk was detected. By investigating the activity towards a panel of hIL-6 mutants it appeared that the region recognized by the minibody was the C-terminal segment of the presumed D-helix, which is also part of the combining site with hIL-6Rα. Finally, the resulting minibody appeared to be a more potent hIL-6 antagonist than its predessessor in tissue culture assays.

Recently, the same group reported another approach (Martin *et al.*, 1999) in order to identify a minibody mutant which competitively inhibits the NS3 serine protease of hepatitis C virus. This study started with the evaluation of ca. 1000 randomly sequenced minibody mutants out of the original repertoire (Martin *et al.*, 1994) and made use of the knowledge about amino acid preferences in the natural protease substrates as well as previously delevoped specific inhibitors. A panel of 14 minibodies with corresponding characteristics was thus identified and subsequently produced as soluble proteins for experimental investigation.

One clone (Mbic) almost completely abolished enzymatic activity in a screening assay with the recombinant NS3 protease using a peptidic substrate. Mbic turned out to be a specific inhibitor, with an IC_{50} value of ca. 1 μM (comparable to the activity of the product inhibitor derived from the natural cleavage site), and no inhibitory activity on human leukocyte elastase or porcine kallikrein. Similar as in the case of the hIL-6 target, only the H2 loop of Mbic was recombinant with respect to the wild-type minibody, bearing the sequence GIEELD. Mutagenesis experiments indicated that the H1 loop did not contribute to protease recognition. Using alanine scanning mutagenesis the pattern of neighbouring Glu residues in the H2 sequence was shown to be mainly responsible for the protease recognition.

The H2 loop still exerted protease-inhibitory activity when grafted into the H3 hypervariable loop of a human 'camelized' V_H domain and even when synthesized as an end-to-tail cyclic peptide. Although the IC_{50} value was raised to ca. 500 μM in the latter case the authors suggested that the cyclic peptide could be suitable as a lead compound for drug development. The minibody did thus serve as a transient peptide scaffold. However, one might ask whether the same result could have been obtained more quickly by applying combinatorial peptide chemistry. Notably, a preceding attempt to select NS3 protease inhibitors directly from the minibody library via phage display had failed (Dimasi *et al.*, 1997).

In principle, the minibody concept seems to be of interest due to the small size of the molecule, which even enables its synthesis by solid phase peptide chemistry. Nevertheless, this scaffold suffers from notorious solubility problems, which are apparently seen for all variants investigated so far. In addition, sound data on the three-dimensional structure of this biomolecule are still missing and the stability of its folded conformation is questionable. Finally, the interaction interface that is provided by the two remaining hypervariable loops is rather limited compared with the scaffolds based on intact Ig domains as described above.

In contrast, the fibronectin type III domain (FN3) constitutes a small, monomeric natural β-sandwich protein with resemblance to a trimmed Ig V_H domain. It consists of 94 amino acids and possesses seven β-strands instead of nine, with three loops connecting the strands in a pairwise fashion at one end of the β-sheet. The loop that carries the integrin-binding RGD sequence is topologically equivalent to the CDR-H3 of a V_H domain. Notably, FN3 is one of the rare members of the Ig superfamily devoid of disulphide bonds. FN3 domains are ubiquitous and occur in cell adhesion molecules, cell surface hormone and cytokine receptors, chaperonins, and carbohydrate-binding domains, all of which are involved in a molecular recognition function. Koide *et al.* (1998) chose the 10th out of 15 repeating units of FN3 in human fibronectin as a scaffold for the generation of novel receptor proteins. Ubiquitin, a small protein involved in intracellular protein degradation pathways, was employed as an examplary target.

The so-called BC and FG loops form a contiguous molecular surface. Based on their large sequence variations among the numerous FN3 units that are present in human fibronectin, these segments were considered not to be crucial to stability and hence subjected to randomization starting from a synthetic gene. A phagemid display library with 10^8 mutant FN3 domains was prepared using 10 randomized residues, five in the BC loop (26–30) and five in the FG loop (77–81). Due to the exceptional length of the FG loop in the chosen FN3 unit, three amino acids (82–84) were removed at the same time. After five rounds of panning against ubiquitin one dominant sequence was identified and dubbed Ubi4. All except one of the randomized positions differed from the wild-type FN3 sequence.

Using phagemid ELISA the Ubi4 variant was shown to bind specifically to its target with respect to unrelated proteins, exhibiting an IC_{50} value of approximately 5 μM. An alanine-scanning mutagenesis revealed that both of the variegated loops contributed equally to the binding. Ubi4 could be readily produced as a soluble protein in *E. coli*, however its solubility was significantly lowered compared with wild-type FN3. This effect was reduced upon addition of a positively charged C-terminal GKKGK tail. Nevertheless, the protein retained considerable background binding, precluding its analysis by size exclusion chromatography or on an SPR biosensor. Gdn–HCl-induced unfolding experiments indicated reversible folding characteristics of the engineered FN3 domain, even though with significantly diminished overall stability. Finally, heteronuclear NMR spectroscopy of the ^{15}N-labelled protein at low pH indicated that it was monomeric and retained the global fold of the FN3 domain.

The V-like extracellular domain of the human cytotoxic T-lymphocyte associated protein-4 (CTLA-4) may provide another potential scaffold based on the Ig architecture, yet possessing two intramolecular disulphide bonds. Normally CTLA-4 constitutes a plasma membrane protein expressed by activated T-cells, which homodimerizes via a disulphide bridge within a stalk region. The V-like domain binds via CDR-equivalent loops to the B7.1 (CD80) or B7.2 (CD86) co-receptors on antigen-presenting cells, playing a role in the negative regulation of T-cell response. Although monomeric CTLA-4 V-domains were successfully synthesized in eukaryotic CHO or *Pichia pastoris* expression systems, attempts to produce active soluble protein in *E. coli* remained unsuccessful. In order to exploit CTLA-4 as an Ig-like single domain scaffold Nuttall *et al.* (1999) investigated the bacterial expression and phage display of a loop-grafted N-terminal domain.

The loops either eqivalent to CDR1 or to CDR3 were substituted with somatostatin, a 14-residue intra-disulphide-linked peptide hormone. As a result, proteins with superior solubility compared with the wild-type CTLA-4 V-domain were obtained following secretion into the periplasm of *E. coli*. Both variants, CTLA-4R-Som1 and -Som3, were predominantly recovered in the monomeric state as revealed by gel filtration. In contrast, the recombinant wild-type CTLA-4R protein was largely present as a dimer or higher oligomer, even though the naturally cross-linking Cys residue had been deleted from the construct.

Both the wild-type and the mutant CTLA-4 domains were used for phage display. In the case of the wild-type protein, recombinant bacteriophages were shown to bind to CD80-Ig as well as to CD86-Ig protein. In the case of the CTLA-4R-Som3 variant binding to an immobilized anti-somatostatin antibody and to the somatostatin receptor subtype 4, which had been expressed in a non-glycosylated form in transfected CHO cells, was detected. Based on their results the authors proposed that the CTLA-4 domain could be useful for the construction of libraries following randomization of one or more of the CDR-like loops. Hence, this scaffold might be employed instead of camelid V_H domains, offering the advantage of a human framework. However, a still pronounced tendency of this biomolecule towards oligomerization should constitute a serious caveat to broader application.

PROTEASE INHIBITORS AS STABLE FOLDING UNITS

Protease inhibitors are widely known as small and remarkably stable proteins. In most cases their protease-binding site comprises a short, more or less extended peptide stretch with varying sequence being presented as an exposed loop by a structural framework that is specific for the inhibitor family. Consequently, the idea appeared attractive to employ a protease inhibitor as a scaffold for a structurally constrained peptide loop. Naturally, these studies aimed first at the generation of protease inhibitors with altered enzyme specificities.

Bovine (or basic) pancreatic trypsin inhibitor, BPTI, the prototypic example of the Kunitz-type protease inhibitors, was used as a scaffold in a study by Roberts *et al.* (1992).

Generally, Kunitz domains are stable proteins with ca. 60 residues, and possessing three disulphide bonds, which act as slow tight-binding, reversible inhibitors of serine proteases. A phage display library of ca. 10^3 BPTI variants (carrying four purposely introduced mutations) was constructed by targeted randomization of the amino acid positions 15–19, i.e. the specificity-determining residues P1 to P4′. Using human neutrophil elastase as a target protein, several variants were selected, which could be produced in *E. coli* and purified by affinity chromatography on immobilized elastase. One of them exhibited a $K_D = 1.0$ pM. This value was 3.6×10^6-fold better than that of the wild-type inhibitor and 50-fold better than the highest affinity of a reversible elastase inhibitor known at the time. In addition, the BPTI variant appeared to be specific for this enzyme, without inhibitory activity for trypsin, porcine pancreatic elastase or human cathepsin G.

Alzheimer's amyloid β-protein precursor inhibitor (APPI) served as a scaffold for the selection of novel Kunitz domain inhibitors for the human TF–FVIIa (tissue factor/factor VIIa) complex, which plays a crucial role at the beginning of the blood coagulation cascade (Dennis and Lazarus, 1994). Three different libraries with complexities ranging from 9×10^6 to 5×10^8 were constructed by randomizing four or five positions in the APPI sequence and presenting the protein as a pIII fusion protein on M13 phagemids. The variegated residues were distributed both among the extended primary binding loop and two separate positions in a closely neighbouring loop, thus comprising a discontinuous kind of paratope.

Four selection cycles against the immobilized TF–FVIIa complex, followed by sequencing of several clones, resulted in identifyable amino acid preferences at each of the randomized positions. From these data an overall consensus sequence, designated TF71-C, was generated. This protein was produced by secretion from *E. coli* and purified via trypsin affinity chromatography as well as reverse-phase HPLC. TF7I-C had an apparent dissociation constant, K_i*, for the complex with TF–FVIIa of 1.9 nM, compared with 300 nM in the control experiment with APPI. Investigation of the specificity revealed that TF7I-C was also a potent inhibitor of FXIa, similarly as APPI, and furthermore of plasma kallikrein. TF7I-C significantly prolonged the clotting time in a tissue factor-initiated prothrombin time assay.

In a conceptionally similar study the first Kunitz domain of human lipoprotein-associated coagulation inhibitor (LACI-D1) was used for the selection of a highly potent inhibitor of plasmin (Markland *et al.*, 1996a). Using a synthetic gene, positions in the P1 region were first mutagenized, followed by enrichment via phagemid display, and then additional diversity was introduced at spatially neighbouring residues in the 'second loop' as mentioned above. Phagemid selection resulted in the isolation of the variant EPI-P211, which inhibited human plasmin with $K_i = 2$ nM. Based on the identified sequences of the selected LACI-D1 variants a biased library for the P1 region was constructed, from which finally the variant EPI-P302 was selected. This inhibitor had a K_i of 87 pM, i.e. 12 500-fold better than that of LACI-D1. EPI-P302 exhibited very high specificity for plasmin in comparison with other human proteases and appeared to be resistant to inactivation by oxidants and extremes of temperature or pH. Employing essentially the same approach of iterative selection, LACI-D1 variants were also selected with specificity for human plasma kallikrein ($K_i = 40$ pM) and human thrombin (Markland *et al.*, 1996b). Although two of the variants, which were investigated as soluble proteins, clearly bound to thrombin in the latter case, they did not measurably inhibit its enzymatic activity. Therefore, they probably formed a complex with a surface region other than the catalytic site.

Röttgen and Collins (1995) used the human pancreatic secretory trypsin inhibitor (PSTI), a small 56-residue protein, for the presentation of peptides with lengths of seven or eight amino acids. This hypervariable segment was held in a highly exposed position between two disulphide bridges at the exposed tip of the inhibitor, replacing the seven amino acid trypsin-inhibitory loop. PSTI was functionally displayed as a pIII fusion protein, as was demonstrated by binding of the phagemids to trypsin or to an anti-pIII antibody. Novel inhibitors of bovine α-chymotrypsin were selected from the phagemid display libraries with a combined complexity of 3.1×10^7 mutants. Taken together with previous work on the rational design of PSTI variants with altered protease specificity (Szardenings *et al.*, 1995) the authors concluded that structural changes of the inhibitory loop do not affect protein folding of PSTI so that this small protein might serve as a scaffold in the development of lead substances for therapy.

A bacterial serine protease inhibitor with a broad range of substrate specificity, ecotin, was employed for selection experiments by Wang *et al.* (1995). The 142-amino acid inhibitor, which normally resides in the periplasm of *E. coli* as a dimer, inhibits human urokinase-type plasminogen activator (uPA) 10 000-fold less efficiently than bovine trypsin, although their catalytic domains are highly homologous and their protease substrate specificities are virtually identical. Ecotin was displayed as a pIII fusion protein on M13 phagemids and just two of its reactive-site amino acids (residues P1 and P1′) were randomized. After screening with an immobilized low molecular weight form of uPA a few variants were selected. Analysis of the isolated ecotin variants revealed an up to 2800-fold increase in binding affinity, corresponding to an apparent dissociation constant of 1 nM for the interaction with the low molecular weight form of uPA.

Tendamistat, a 74-amino-acid inhibitor of α-amylase was used as a molecular scaffold in a study by McConnell and Hoess (1995). The authors sought to exploit a certain similarity of its β-sheet topology, whose fold is stabilized by two rather than one disulphide bond, with that of an Ig domain in order to present constrained random peptides. In this case two loops, comprising residues 38–40 and 60–65, were mutagenized via PCR using a synthetic gene, and a phage display library of ca. 10^8 variants was prepared. The loop around residues 18–20, which contains the conserved residues necessary for the recognition of α-amylase, was thus left unaffected.

This library was used for selection against the monoclonal antibody A8, which is directed against the 21-residue peptide hormone endothelin. As a result, specific tendamistat mutants were isolated with a variety of sequences in both of the loops, none of which matched a linear sequence

from endothelin. The most frequently occurring isolate was further tested and produced in a soluble form. It was still able to inhibit starch hydrolysis by α-amylase, indicating a correct fold. The relative affinity of the tendamistat variant for the monoclonal antibody was only 2.5 times lower than that of endothelin. Site-directed mutagenesis experiments indicated that residues 61–64 in tendamistat were the primary determinants for the binding to the monoclonal antibody, even though the original epitope on endothelin was of discontinuous nature, comprising four side chains on one face of an α-helix.

Because of their usually small size and robust nature inhibitors are certainly attractive as protein scaffolds, even though the number of loops that can be modified is just one or two. However, it has still to be shown whether engineered inhibitor proteins will also be effective in the complexation of macromolecular targets other than proteases so that they could be amenable to wider application, for example in the targeting of cell surface receptors.

NATURAL AND ARTIFICIAL HELIX BUNDLES

Staphylococcal protein A is in widespread use as an immunochemical reagent for the purification or detection of antibodies and as a corresponding affinity tag in fusion proteins (Uhlún *et al.*, 1992). It contains a repeat of five IgG-binding, α-helical bacterial receptor domains, each of which can form a specific complex with the F_c region of an antibody. In particular, an engineered single domain, called 'Z', serves as a monovalent recombinant version of protein A consisting of 58 amino acids. This domain has a simple fold as a bundle of three α-helices. It is highly soluble and devoid of disulphide bridges, can be expressed and secreted at high yield in *E. coli*, and is extremely stable against proteolysis and heat-induced unfolding.

Obviously, it is the tight pairwise packing of the amphipathic helices which provides the remarkable conformational stability. Nevertheless, it was recently possible to remove the third α-helix after, altogether, 12 subsidizing mutations had been introduced into the two remaining helices in a step-wise combinatorial approach (Braisted and Wells, 1996). As a result, a slim version of the protein A domain with just 33 residues in length was obtained. This two-helix minimal scaffold still bound IgG_1 with a K_D of 43 nM, compared with 10 nM for the 59-residue three-helix bundle.

The conventional Z domain was employed as a scaffold for the generation of novel binding activities by Nord *et al.* (1995). Thirteen exposed side chains located within the F_c-binding surface and distributed across the bodies of helices 1 and 2 were chosen for the generation of a random library. The central residue Ile^{31} was unfortunately not included in the randomization because it was assumed to be important for the helix/helix-packing. The library was established by insertion of the synthesized gene into a phagemid vector, as a fusion with an albumin-binding affinity tag (ABD), i.e. a structurally similar three-helix bundle derived from *streptococcal* protein G. The random substitution of nucleotides within the protein A domain was confirmed by sequencing of several clones. Four arbitrarily chosen mutants were produced as soluble ABD fusions in *E. coli* and purified via affinity chromatography on HSA. All four variants were recovered at high yield (1.5–2.5 mg/l culture), but only three of them had their α-helical secondary structure retained, as revealed by CD spectroscopy.

Following to this preliminary study the repertoire of mutagenized Z domains (4×10^7 representatives) was applied to selection experiments against the macromolecular targets *Taq* DNA polymerase, human insulin, and human apolipoprotein A-1 (Nord *et al.*, 1997). Significant enrichments were achieved in all three cases by monovalent phagemid display. In the case of apolipoprotein A-1 another selection experiment was carried out with a second library of similar size that had been prepared by use of (C/A/G)NN degeneracy instead of NN(G/T)-degenerated codons. Whereas in the original library the NN(G/T)-triplets coded for all 20 amino acids together with the amber termination codon, all stop as well as Cys codons were avoided in the second library, albeit at the expense of the aromatic amino acids Trp, Tyr and Phe, whose codons were also excluded. Sequence analysis of the clones from all four selection experiments revealed fully substituted sets of the 13 mutagenized amino acid positions.

All variants could be produced in *E. coli* and purified by means of the ABD tag (which was in some cases present together with a His_6 tag plus an *in vivo* biotinylated domain). Two variants selected against the apolipoprotein carried a single Cys residue and were almost quantitatively recovered as disulphide-bonded dimers. Target recognition was investigated via SPR and found to be specific for each of the three proteins. In the case of the *Taq* DNA polymerase K_D values around 2 μM were found, whereas a K_D of approximately 30 μM was detected in the case of insulin. One of the monomeric Z domain variants directed against the apolipoprotein exhibited a K_D value of 3 μM. CD-spectroscopic analysis finally revealed that there was no apparent difference in the secondary structure content compared with the wild-type Z domain. The authors termed their artificial α-helical receptor proteins as 'affibodies' and described them as interesting candidates for use as ligands in affinity chromatography applications.

Nevertheless, it was apparent that the affinities were not high enough for tight complex formation with the target proteins and corresponding detection purposes. Therefore, an attempt was made to subject one of the affibodies, directed against *Taq* DNA polymerase, to an affinity maturation (Gunneriusson *et al.*, 1999). Using a primary affibody derived from the preceding study ($Z_{Taq4:8}$) six amino acid positions in helix 2 were chosen for selective re-randomization, followed by phagemid display selection at high stringency against the same target protein. Second generation variants were thus identified with improved K_D values in the order of 30–50 nM. The affinity between the resulting affibodies and *Taq* DNA polymerase was hence comparable to that between the wild-type Z domain and an IgG F_c fragment (see above). Taking advantage of an avidity effect, similarly to the oligomeric natural protein A, the construction of a tandem dimer of the best of the selected affibodies ($Z_{TaqS1\text{-}1)}$ led to a three-fold increase of its apparent affinity when the target protein had been immobilized to a sensor chip surface.

A different type of α-helical protein, the four helix bundle protein cytochrome b_{562}, was chosen as an alternate framework for the selection of an artificial hapten-binding protein in a study by Ku and Schultz (1995). Four-helix bundle proteins are relatively small, provide a rigid framework due to the well-packed hydrophobic core, are thermally stable, and can be expressed in *E. coli* at high levels. Two loops, each connecting one pair of the α-helices and coming together at one end of the helix bundle—opposite to the heme-binding site—appeared as a potential motif for the formation of a ligand pocket in between. When the structurally related myohemerythrin was used for generating a first loop library (with five randomized positions in each of the two loops) most of the mutants were unstable and aggregated. Consequently, the authors turned to cytochrome b_{562}, a periplasmic heme protein of *E. coli* that appeared from previous work to be more tolerant of mutations in the loop sequence connecting its third and fourth α-helices (Brunet *et al.*, 1993).

Nine amino acids, five in the first loop and four in the second, were subjected to combined randomization using cassette mutagenesis with degenerate oligodeoxynucleotides. Pilot expression experiments with solubly produced mutants revealed that almost half of the variants folded correctly because they were still capable of complexing heme as their prosthetic group. The library with a complexity of 2×10^8 variants was then used for selection by multivalent phage display against an *N*-methyl-*p*-nitrobenzylamine derivative (a transition state analogue, against which a monoclonal antibody had previously been raised) coupled to BSA as a carrier protein. Twenty clones with hapten-binding activity were sequenced and four of the randomized residues, two in each loop, were found to be conserved.

Based on these results a second, 'doped-loop' library was prepared, where these four of the altogether nine randomized positions were mutated at a decreased frequency, and further selection cycles were performed. Sequence analysis of the enriched variants revealed a fifth consensus position. Four of the mutants were finally produced in *E. coli* using the T7 promoter with yields of 30 mg per 500 ml culture. UV–visible and CD spectra indicated that the mutants folded into a similar structure as the wild-type protein and still bound the heme group. SPR analyses for the binding of the hapten-BSA conjugate revealed K_D values ranging from 5 to 22 μM. Binding to BSA alone was not detectable but, notably, the hapten itself was no longer recognized when coupled to ovalbumin as an alternative carrier protein. Hence, the cytochrome library probably does not provide a cavity of sufficient size and structural diversity in order to complex the hapten tightly alone, but additional interactions with the carrier protein seem to be required. Attempts to examine the ability of this framework to bind larger protein antigens have not yet been reported.

Besides, the α-helix bundle motif is one of the few scaffold structures that was also approached by rational protein design in order to generate new binding functions. *De novo* designed four-helix bundles were used for the grafting of several types of Zn(II)-binding sites. For example, the His_3 motif from the active centre of carbonic anhydrase was transplanted onto the designed four-helix bundle 'α_4' (Handel and DeGrado, 1990). In this case the metal complexation contributed to the conformational specificity in the α-helix bundle, decreasing some of its molten globule-like characteristics and forcing it to adopt a unique topology (for review see DeGrado *et al.*, 1999). The α_4-fold was also used as a template for engineering a fully coordinating Cys_2His_2 Zn(II)-binding site, similar to the one in His_2Cys_2 zinc fingers (Regan and Clarke, 1990). The general observation of a relaxed conformational definition in the designed four-helix bundle proteins indicates that this type of fold probably does not provide the kind of rigid framework as is intended for practical purposes.

LIPOCALINS AS POCKET-FORMING β-BARREL PROTEINS

The lipocalins constitute a structural family of small, functionally diverse proteins, comprising 160–180 residues, with rather weak homology on the amino acid sequence level. The first member with known three-dimensional structure was the retinol-binding protein (RBP) from human serum (Newcomer *et al.*, 1984; Cowan *et al.*, 1990). Since then this family has rapidly grown (Flower, 1996) and comprises nowadays more than 150 members, which are found in a variety of higher organisms but as well in bacteria (Bishop *et al.*, 1995). In general, their primary physiological role seems to lie in the storage or transport of mostly hydrophobic and/or chemically sensitive organic compounds. Thus, each lipocalin is highly optimized for the recognition of its individual ligand, e.g. vitamin A in the case of RBP, and the degree of sequence conservation for a particular lipocalin from different species is rather high.

From the structural point of view the lipocalins share a conserved β-barrel as their central folding motif, which is made of eight antiparallel β-strands winding around a central axis. At the open end of the resulting conical structure these strands are connected in a pair-wise fashion by four loops, which form the entrance to the ligand-binding pocket. At the opposite end the β-barrel is closed by short loops, and densely packed side chains fill its lower part, forming the hydrophobic core. Contrasting with the highly conserved β-barrel topology, the loop regions around the binding pocket differ considerably among individual lipocalins, both in conformation and length of their polypeptide segments.

RBP was first used as a scaffold in a rational protein design study which aimed at the grafting of a metal-binding site from carbonic anhydrase onto the solvent-exposed outer surface of the β-barrel (Müller and Skerra, 1994). In human carbonic anhydrase II a Zn(II)-binding site is formed by the three His residues 94, 96, and 119, whose imidazole side chains protrude into the solvent-filled active site cavity of the enzyme and complex the transition metal ion in a tridentate manner. The three liganding His residues are located on two hydrogen-bonded antiparallel β-strands, which belong to the same type of secondary structure that provides the cylindrical β-sheet in RBP. Using computer modelling, altogether 44 superpositions between this His_3 structural motif and paired segments of β-strands from RBP were investigated in order to identify sites where the three

His side chains from the enzyme could be introduced, while retaining the backbone fine structure and the relative spatial positioning of their C_α and C_β atoms.

Several mutants of RBP were generated and produced in *E. coli* and their Zn(II)-binding properties were investigated in equilibrium dialysis experiments. One mutant, RBP/H_3(A), with His residues introduced at the positions 46, 54 and 56 in the polypeptide sequence, was shown to bind Zn(II) specifically with a stoichiometry of 1 and a remarkably low dissociation constant of 36 nM. Cu(II) and Ni(II) were complexed with lower affinity. The complexation of Zn(II) had no influence on the binding of retinoic acid, one of the natural ligands of RBP, so that a truly bifunctional protein was obtained. In Gdn–HCl-induced unfolding experiments the mutant was found to be significantly stabilized in the presence of small concentrations of the metal ion. Furthermore, it was demonstrated that the protein-bound Zn(II) is accessible to iminodiacetic acid as an additional chelating ligand without competition for the metal ion. Thus it became possible to use the grafted metal-binding site for the efficient purification of the engineered, bacterially produced RBP via IMAC (Müller and Skerra, 1994; Schmidt *et al.*, 1996).

With their set of four variable loops on top of a rigid β-barrel the lipocalins exhibit some similarity with the antigen-binding region of immunoglobulins, whose six hypervariable loops confer specificity for a huge variety of ligands. Consequently, the question arose whether lipocalins could be used as a scaffold for the complexation of prescribed target molecules at their natural ligand pocket, permitting the generation of small and stable receptor proteins. The 174-residue bilin-binding protein (BBP) from *Pieris brassicae* with its rather wide and shallow ligand pocket—which normally complexes biliverdin IX_γ—served as a model protein to be engineered for the recognition of fluorescein, an established immunological hapten (Beste *et al.*, 1999). Sixteen residues at the centre of the binding site, distributed across all four loops, were identified by molecular modelling and subjected to concerted random mutagenesis using a convenient PCR assembly strategy. From the mutant library with a complexity of 3.7×10^8 independent transformants, fluorescein-binding BBP variants were then selected via bacterial phagemid display.

Three variants were identified that were able to complex fluorescein with high affinity, exhibiting dissociation constants as low as 35.2 nM. Interestingly, one of these variants effected almost complete quenching of the ligand's fluorescence, similarly to an antifluorescein antibody known before. Detailed ligand-binding studies and site-directed mutagenesis experiments indicated not only that the molecular recognition of fluorescein is specific but also that positively charged residues at the centre of the pocket are responsible for the tight complex formation. This novel class of engineered stable receptor proteins based on a one-domain scaffold was denominated 'anticalins'.

In a subsequent study an anticalin with specificity for the hydrophilic cardiac glycoside digoxigenin was isolated from the same library of BBP mutants (Schlehuber *et al.*, 2000). This time an enrichment of digoxigenin-binding variants via phagemid display was followed by a specialized colony screening assay in order to quickly spot individual mutants with the desired properties. In this way a BBP variant with a K_D value of 295 nM was obtained. Its affinity for digoxigenin was further raised by selective random mutagenesis of one of the four loops, again followed by phagemid display and colony screening, resulting in a K_D value of 30.2 nM. Notably, the BBP variant recognized the digoxigenin group as a true hapten, without detectable context-dependence. The binding signals were independent of whether this group carried three sugar molecules attached to C-3 of the steroid system, as in the natural compound digoxin, or it was conjugated with differing carrier proteins, i.e. BSA, ovalbumin, and RNase A.

Detailed ligand-binding studies revealed that a chemically similar cardiac glycoside, digitoxin, which just differs by a single missing hydroxyl group from digoxin, was even bound stronger, with a K_D value equal to 3.2 nM. Complexation of the related steroid ouabain, however, which often shows cross-reaction with antibodies directed against digoxin, could not be detected, and the same was the case for testosterone or 4-aminofluorescein. Hence, a specific biomolecular reagent was obtained, which should be useful for the detection of the digoxigenin group, a hapten that is generally applied for the non-radioactive labelling of proteins, nucleic acids, and other macromolecules. In addition, it was demonstrated that anticalins can be produced at high amounts in *E. coli* (Beste *et al.*, 1999) and purified to homogeneity in one step using the *Strep*-tag method (Schmidt and Skerra, 1994). The successful construction of functional fusion proteins with alkaline phosphatase, a wide-spread reporter enzyme, should further contribute to their rapid application in bioanalytics (Schlehuber *et al.*, 2000).

Taken together, these findings demonstrate that the ligand pocket of a lipocalin can be effectively reshaped in order to complex unrelated ligands with high affinity and specificity, similar to antibodies. Consequently, the β-barrel of the lipocalins possesses the potential of supporting ligand pockets with a variety of shapes and surface properties. A compilation of the amino acid replacements observed in engineered BBP variants with several independent specificities demonstrated that all of the 16 randomized positions tolerate each type of amino acid side chain. The ability of the β-barrel to support loop segments with high structural plasticity should not be restricted to the binding of haptens, an effect which rather served as a rigorous test in the previous studies. The binding of macromolecular antigens should be permitted by this scaffold as well, given that appropriate amino acids in the more exposed positions of the four loops will be used for randomization. RBP, for example, is known to form a stable complex with prealbumin when travelling through the blood, involving the loops at the open end of its β-barrel as part of the interface (Monaco *et al.*, 1995).

Lipocalins with engineered ligand-binding sites should therefore provide an attractive alternative to recombinant antibody fragments, combining the advantage of a much smaller size with their composition of a single polypeptide chain. Anticalins could thus be well suitable for *in vivo* diagnostics or even therapy, especially when engineered on the basis of a human framework. Apart from the human RBP described above, lipocalins like apolipoprotein D (Milne *et al.*, 1993), tear lipocalin (Gachon and Lacazette,

1998), and β-Trace, also known as prostaglandin D synthase (Tanaka *et al.*, 1997), which all occur in human body fluids, might be suitable in this respect.

KNOTTINS: AT THE BORDERLINE BETWEEN PROTEINS AND PEPTIDES

The so-called 'knottins' (Le Nguyen *et al.*, 1990) comprise a structural family of rather small proteins, typically 25–35 amino acids, that bind to a range of molecular targets like proteins, sugars and lipids. Their three-dimensional structure is essentially defined by a peculiar arrangement of mostly three disulphide bonds. A characteristic knotted topology with one disulphide bridge crossing the macrocycle limited by the two other intrachain disulphide bonds, which was found in several different microproteins with the same cysteine network, lent its name to this class of biomolecules. Although their secondary structure content is generally low, the knottins share a small triple-stranded antiparallel β-sheet, which is stabilized by the disulphide bond framework. Biochemically well defined members of the knottin family include the trypsin inhibitor EETI-II from *Ecballium elaterium* seeds, the neuronal N-type Ca^{2+} channel blocker ω-conotoxin from the venom of the predatory cone snail *Conus geographus*, and the C-terminal cellulose-binding domain (CBD) of cellobiohydrolase I from the fungus *T. reesei*. Naturally occurring knottins have little sequence homology except the set of Cys residues, which gives rise to the conserved pattern of disulphide bridges. The interspersed peptide loops are highly variable both in length and sequence so that the core of the knottin structure could be a suitable scaffold in order to create interfaces for novel binding activities.

The serin protease inhibitor EETI-II, a squash-type inhibitor isolated from jumping cucumber seeds, was the prototype of the 'knottin' family (Le Nguyen *et al.*, 1990). It consists of 28 amino acids and three disulphide bridges, which are essential for the bioactive conformation of this microprotein. Remarkably, this biomolecule possesses the unfailing ability to refold and correctly close its three disulphide bonds, which facilitates synthesis by peptide chemistry. Close similarity between the protease-inhibiting loop of EETI-II and the corresponding segment of BPTI was noted during structural analysis, leading to the hypothesis of a loop that is variable in sequence but stable in conformation. Substitution of Arg^4 in the inhibitory loop of EETI-II by hydrophobic side chains resulted in altered protease specificities, especially directed against pancreatic and leucocyte elastase as well as chymotrypsin (Le Nguyen *et al.*, 1990). Furthermore, the structural similarity between EETI-II and a carboxypeptidase inhibitor from potato leaves (CPI) permitted the synthesis of a chimeric peptide from the sequence of EETI-II plus a C-terminal tetrapeptide extension with inhibitory function from CPI. Thus, a peptide was obtained with the properties of a double-headed inhibitor complexing both trypsin (K_D = 1.8 nM) and carboxypeptidase (K_D = 3 nM).

Christmann *et al.* (1999) established the bacterial production of EETI-II and investigated to what extent this folding motif tolerates modifications to its loop structures. Using gene synthesis two derivatives were constructed where the six-residue protease-inhibiting loop was either replaced by a 13-residue epitope of Sendai virus L-protein or by a 17-residue epitope from the human bone Gla-protein (also called E-tag). Both wild-type EETI-II and its variants were produced in a correctly folded state—despite being contaminated with disulphide-crosslinked oligomers—at high yields (10–20 mg/l culture after IMAC purification) via secretion into the periplasm of *E. coli* as fusions with the maltose-binding protein (MalE). The cysteine residues were shown to be fully oxidized and the engineered knottins were reactive with monoclonal antibodies directed against the epitopes. Following fusion with fragments of the Lpp and OmpA proteins it was possible to functionally present the EETI-II variants on the bacterial outer membrane, permitting efficient enrichment by fluorescence-activated cell sorting after labelling with cognate monoclonal antibodies.

Smith *et al.* (1998) utilized the carbohydrate-binding CBD, which possesses just two disulphide bonds and thus does not represent a true topological knot, for the generation of a molecular library. They used the flat hydrophilic face of the wedge-shaped protein, which was known to mediate affinity for cellulose, in order to introduce random mutations at seven side chains. In the primary structure these residues cluster in two distinct regions of sequence, either close to the N-terminus or close to the C-terminus, so that the PCR-based library construction was facilitated, resulting in a complexity of 5.5×10^8. Phagemid display selection experiments were performed with four different macromolecular targets: cellulose, α-amylase (TAKA), bovine alkaline phosphatase, and *E. coli* β-glucuronidase. In the search for CBD variants with cellulose-binding activity, a few alternative sequences were indeed identified, although the amino acid exchanges seemed to be more or less conservative in six of the seven positions.

Whereas the selection experiments with α-amylase and β-glucuronidase remained unsuccessful, CBD variants with affinity towards alkaline phosphatase could be isolated. The selected phagemid population did not give rise to ELISA signals for the binding of either cellulose or homologous enzymes from *E. coli* or shrimp, and recognition of alkaline phosphatase was not dependent on its N-linked glycosylation. Sequencing of four clones revealed that just one of the mutagenized residues was unchanged from the wild-type, but the altered residues had certain features in common. The selected CBD variants were finally produced in *E. coli* as MalE fusion proteins and a competition ELISA with alkaline phosphatase was performed, revealing a K_D of ca. 10 μM in the best case. Control experiments with the phage-displayed mutant knottins demonstrated that binding of the enzyme was abolished when the disulphide bonds were reduced or when alanine mutations were introduced in either half of the engineered CBD surface.

Scorpion toxins with their specificities for differing ion channels are structurally related to the knottins as they consist of merely 37 amino acids and their fold, which comprises a short α-helix on one face and an antiparallel triple-stranded β-sheet on the other, is also stabilized by three disulphide bonds. The charybdotoxin structure was employed as a scaffold in order to graft the His_3 metal binding site—along with two further side chains—from carbonic anhydrase B onto the two larger neighbouring strands of its β-

sheet (Vita *et al.*, 1995). Four other mutations were introduced into the charybdotoxin to prevent unfavourable contacts with the metal-binding site and to remove an endogenous His residue. The engineered microprotein was prepared by chemical synthesis, whereby the disulphide bonds were readily formed by controlled oxidation. Affinity for Cu(II) ions was measured by fluorescence titration of the unique Trp residue in the peptide, yielding a K_D of 42 nM. From competition experiments the affinity for Zn(II) ions appeared to be significantly lower, with $K_D = 5.3\ \mu M$. CD spectroscopic and preliminary NMR measurements indicated that the tertiary structure of the mutated peptide was similar to that of charybdotoxin, albeit the K^+-channel blockage activity was lost.

The charybdotoxin structural motif was also used in order to transfer other functional sites onto its β-sheet, in particular the curaremimetic loop of a snake neurotoxin and the CDR2-like loop of human CD4 (Vita *et al.*, 1998). In the first case eight functionally important residues that form a β-hairpin loop in the toxin α from *Naja nigricollis* replaced a structurally related β-hairpin in the scorpion toxin, and six dispensable residues at the N-terminus of the charybdotoxin were concomitantly deleted in order to enhance its accessibility. The chimera was prepared by peptide synthesis and shown to bind specifically to the nicotinic acetylcholine receptor from *Torpedo marmorata* electric organ with an IC_{50} of 50 μM (compared with 0.3 nM for the snake neurotoxin). In the second example the CDR2-like region of human CD4, which conformationally resembles the same β-hairpin in the charybdotoxin, was used for grafting, again accompanied by truncation of the N-terminus of the scorpion toxin (this time by four residues). The chemically synthesized chimera was able to inhibit the binding of soluble recombinant gp120 to a HSA-CD4 hybrid protein with an apparent IC_{50} of 20 μM (compared with 0.8 nM for HSA-CD4).

The knottins were proposed to be principally suitable not only for the generation of modified surfaces, giving rise to interaction with a prescribed macromolecular target, but also for the display of conformationally constrained peptides in one of the loop regions. It remains to be seen, however, whether this scaffold permits the creation of receptor peptides with high affinity and specificity. An important question relates to the proper folding of engineered knottins. Although it seems that EETI-II, and also charybdotoxin, efficiently forms the proper set of disulphide bonds under *in vitro* conditions and there is considerable variability in the interspersed sequences of its natural relatives, there appear to be other knottins which require additional measures in order to adopt a functional conformation. Conotoxins for example, whose remarkable sequence variability was recognized early, are released by proteolytic cleavage from propeptides. Their N-terminal part is obviously essential for the folding and formation of the specific set of disulphide bonds within the conotoxins (Woodward *et al.*, 1990; Olivera *et al.*, 1990), hence comprising the real scaffold under these circumstances.

Data on the purification and yield of the CBD variants were not yet reported but at least in the case of the EETI-II derivatives the amounts obtained by bacterial production were remarkable. This may however be the result of a fortunate choice of the site for inserting the epitope sequences (Christmann *et al.*, 1999) because the N-terminal inhibitory loop of EETI-II is known to be fixed to the core via formation of the third disulphide bond in the final step of the folding process. It is therefore not clear whether the knottin motif provides a true scaffold that will indeed fold independently of side chain exchanges as soon as the inhibitory loop is altered or loop insertions are performed at other sites.

In contrast with the knottins, which are secretory peptides, the Cys_2His_2 zinc-finger is part of cytosolic proteins, comprising one of the most common eukaryotic DNA-binding motifs. Its conformation is stabilized by covalent metal chelate bonds involving two pairs of Cys and His side chains each. This globular structure does not contain disulphide-bonds, but similarly to the cystine knot peptides, specialized side chains, i.e. two Cys and two His residues, occur in fixed positions relative to the secondary structure elements, which are small in size and in themselves not sufficient to confer folding stability. In a study with the ultimate goal of designing peptidomimetics, five residues within the α-helical subdomain of a 26 amino acid Cys_2His_2 consensus zinc-finger motif, well exposed and closely situated in space, were subjected to randomization. The resulting molecules were screened for binding activity towards a monoclonal IgA raised against the *Shigella flexneri* lipopolysaccharide (Bianchi *et al.*, 1995).

The corresponding molecular library was produced both as fusion with the pVIII coat protein of M13 phage (9.7×10^6 transformants) and as soluble peptides by chemical synthesis (6.4×10^6 different sequences), in the latter case also incorporating non-proteinaceous amino acids. The two libraries were independently screened against the monoclonal antibody, yielding very similar consensus sequences. In an ELISA with the synthesized single peptides one of the most common sequences exhibited strong and Zn(II)-dependent binding, suggesting an important role of the proper three-dimensional structure for the molecular recognition. All the selected peptides had spectroscopic properties (CD as well as Co(II)-complex absorption) indistinguishable from the parent zinc finger peptide, CP1 (Krizek *et al.*, 1991). The presence of the α-helix and the two β-strands as well as their mutual packing was also confirmed by NMR. The authors concluded that their α-helical library based on the zinc finger scaffold was conformationally homogeneous. However, data on the affinities and specificities of the selected variants for the target Ig were not reported.

ENZYME ACTIVE SITES AS LIGAND POCKETS: TOWARDS NEW BIOCATALYSTS

Glutathione *S*-transferase (GST) has been used as an enzymatic framework rather for the engineering of proteins with novel binding specificities than for altered catalytic properties. GSTs are a class of cytosolic enzymes that catalyze the first step in the detoxification of electrophilic and usually lipophilic xenobiotics by conjugation with the hydrophilic cellular nucleophile glutathione (GSH). The active site of GSTs is composed of two subsites. The G-site,

which specifically binds the tripeptide GSH, is largely conserved among the whole class. GSH is properly oriented by electrostatic interactions and hydrogen bonds so that its sulfhydryl group becomes activated for nucleophilic attack. The H-site is responsible for the binding of the hydrophobic co-substrate and structurally more diverse. In addition, its specificity for the electrophile is broad and the corresponding K_M values are usually rather high. All structurally known GSTs are homo-dimeric enzymes, whereby each subunit has two domains. The smaller N-terminal domain provides most of the residues for the G-site, whereas the larger C-terminal domain is essentially responsible for the formation of the H-site.

Widersten and Mannervik (1995) employed the human GST A1-1 for the engineering of the H-site cavity in order to create novel specificities for active-site ligands. Ten of the hydrophobic residues aligning the pocket for the electrophilic co-substrate, which are distributed over two regions in the primary structure, were targeted for random mutagenesis. Degenerate primers were used to introduce mutations at the corresponding positions by amplification of the 3′ half of the cDNA. A phagemid display library with a complexity of 5×10^6 transformants was prepared by fusing the mutant GSTs with pIII. Affinity selection experiments were performed using immobilized *S*-(4-carboxybenzyl)-glutathione and BSA-coupled *p*-carboxybenzyl and *p*-nitrobenzyl ligands. In the case of the two smaller benzyl ligands, sets of mutually similar sequences were selected. None of the 10 mutated wild-type residues appeared to be conserved. The larger carboxybenzyl-glutathione ligand did not give rise to a clear selection of sequences, but some of them were identical to those found for the smaller targets. In this case competitive elution of bound phagemid particles in the course of the selection was probably driven by the complexation of the GSH moiety at the unmutated G-site.

Three mutants were produced as soluble proteins, purified, and shown to bind their ligands in an ELISA, whereas the wild-type enzyme did not recognize the same BSA conjugates. When comparing the kinetic properties of the mutants with GST A1-1, their activities were found to be diminished by several orders of magnitude. Even though the novel ligand specificities did not markedly differ from the natural spectrum of accepted electrophilic substrates, this study at least demonstrated that the H-site of this enzyme tolerates a series of small hydrophilic side chains that had been substituted for exclusively apolar residues.

Another enzyme from the same family, human GST P1-1, served as a model system by another laboratory (Napolitano *et al.*, 1996). Being inspired by structural considerations, two ligand-recognition segments (called LRSs) close to the enzyme's active site (residues 36–43 and C-terminal residues 204–210) were identified, which participate in substrate recognition but whose sequences seemed not to be crucial for the structural integrity of the protein. In the authors' view these sites may be functionally compared with the CDR3 hypervariable loops in antibodies. Three libraries (containing $1–5 \times 10^6$ mutants) were generated by PCR mutagenesis, one with the first LRS randomized, one with the second LRS randomized, and one with a partially randomized six amino acid insertion between residues 206 and 207. Forty clones from all three libraries were subjected to a chromogenic activity test with 1-chloro-2,4-dinitrobenzene as electrophilic standard co-substrate. Twenty-one mutants retained enzymatic activity above background. Since all of these clones carried random mutations as expected, the authors concluded that modifications to the LRSs did not abolish binding of either of the two substrates.

Seventeen of these clones were produced in *E. coli* at acceptable amounts as revealed by immunochemical detection via the *myc* tag. Notably, this C-terminal affinity tag was found to have an effect on the compound pattern of inhibition of the enzyme. The 17 mutants were screened for enzymatic inhibition against a panel of 20 organic substances, and five of the mutant GSTs were then examined in greater detail. It appeared that they exhibited markedly different inhibitor profiles with mutual binding affinities typically in the higher micromolar range. For example, two of the so-called 'glubodies' recognized more inhibitory compounds than the recombinant parent enzyme, whereas one of the mutants had a more restricted range of inhibitor susceptibility and another one, which had an extreme C-terminal alteration due to a reading frame shift, seemed to be generally more resistant to inhibition. Interestingly, no correlation was detected between inhibitor binding specificity and specific activity against the chromogenic substrate, as revealed by statistical analysis.

Taken together, these two studies demonstrate that the H-site substrate pocket in members of the GST family of enzymes may be engineered as a recognition site for the binding of small ligands. However, the affinites that resulted from the described indirect measurements were remarkably weak and a useful application of this approach has not yet been described.

Another enzyme that uses GSH as a co-substrate, *E. coli* thioredoxin (TrxA), was employed as a scaffold for the display of single conformationally constrained peptides replacing its active site loop in two studies. TrxA is a small cytosolic enzyme normally involved in maintaining the thiol/disulphide equilibrium inside the cell. It is highly soluble, rigid, can be overexpressed in large amounts (LaVallie *et al.*, 1993), and its three-dimensional structure is known.

TrxA was functionally presented on the *E. coli* flagellum by Lu *et al.* (1995), who successfully inserted its coding sequence into a dispensable region of the gene for flagellin (*fli*C). The active site sequence -CGPC-, which forms a tight, disulphide-constrained and solvent-accessible loop in the oxidized state of the enzyme, was known from a previous study to be highly permissive to the insertion of various peptide sequences (LaVallie *et al.*, 1993). Based on the flagellin fusion system, a library with a diversity of 1.77×10^8 was prepared by replacing the loop with random dodecapeptide sequences. Using a whole bacterial cell panning procedure this library was selected against three monoclonal antibodies that had been adsorbed to plastic surfaces. As a result, significant enrichments were obtained, revealing a distinct consensus sequence of three to five residues in each case. The position of the consensus sequence was variable within the randomized dodecamer, but for each of the three antibodies it could be readily aligned with a contiguous stretch in the antigen sequence. The interaction between the TrxA variants and their cognate antibodies was also probed on a Western blot, revealing varying binding strength. However, affinity values were not determined.

Colas *et al.* (1996) replaced the same TrxA active site loop by a 20-residue random sequence and selected for binding to human cyclin-dependent kinase 2 (Cdk2) using a yeast two-hybrid system. After screening 6×10^6 yeast transformants, 14 TrxA variants, called 'aptamers', were identified carrying mutually different, mostly charged inserts. The amino acid sequences did not have notable similarity to other known protein sequences. Further analysis of six of the aptamers revealed that interaction with Cdk2 was specific in the yeast two-hybrid system as compared with unrelated proteins. SPR measurements with the solubly produced TrxA variants and immobilized Cdk2 yielded K_D values between 30 and 120 nM. The aptamers were moreover seen to inhibit the biochemical activity of Cdk2 in a phosphorylation assay of histone H1 in the presence of cyclin E. Therefore, the selected TrxA variants appeared to be useful reagents for the investigation of intracellular regulatory pathways mediated by protein–protein recognition.

In a further application of this system thioredoxin-20mer aptamers were isolated, which specifically interact with the DNA-binding and dimerization domains of E2F (Fabbrizio *et al.*, 1999). This protein plays an important role in the p16-cyclin D-pRB-E2F regulatory pathway of cell proliferation, by heterodimerizing with a DP protein. Six positive interactors were selected from 5×10^5 bait/prey co-transformants. One of them strongly inhibited E2F activity both *in vitro* and in mammalian fibroblasts. Interestingly, the synthesized variegated peptide segment itself proved to be a potent inhibitor of cell proliferation by blocking cells in G1, suggesting that this peptide aptamer may be structured even in the absence of the thioredoxin scaffold.

A catalytically inactive version of *staphylococcal* nuclease was employed in a related approach as a scaffold in order to display a peptamer library consisting of 16 random amino acids within budding yeast cells, again followed by selection for inhibitors of biological pathways (Norman *et al.*, 1999). *Staphylococcal* nuclease was chosen as a carrier protein because it is small, folds spontaneously in the absence of chaperones, can be produced at high levels in eukaryotes as well as prokaryotes, and has a prominently exposed loop on its surface. This loop, comprising amino acids 19–27 in the mature enzyme, was replaced by the 16 residue random sequence using a synthetic gene, which was equipped with the hemagglutinin epitope tag at the N-terminus and a His_6 tag at the C-terminus of the inactive nuclease.

A library of 6.5×10^6 yeast transformants was selected for inhibitors of the spindle checkpoint, allowing colony formation of an Mps1-overexpressing strain. Three peptamers with divergent sequences were identified, two of which appeared to reduce the amount of the Mps1 protein by affecting its folding or stability, whereas the third one turned out to be a general inhibitor of the spindle checkpoint. A putative interaction target of this inhibitor (Ydr517w) was subsequently identified using the yeast two-hybrid technique. In a similar manner, inhibitors of the mating pheromone response pathway were isolated. Of 29 peptamers that allowed **a** mating-type cell proliferation in the presence of α factor, 20 were silencing inhibitors and nine interfered directly with pheromone signalling.

In all the studies quoted so far enzymes were merely utilized as passive frameworks for the generation of receptor proteins. However, the scaffold concept was recently adopted to the engineering of biocatalysts, as well. Quéméneur *et al.* (1998) used *E. coli* cyclophilin, i.e. a peptidyl-prolyl *cis*/*trans*-isomerase, in order to convert its active site into that of a proline-specific endopeptidase. This remarkable result was achieved by site-directed mutagenesis of three amino acids, close to the peptide-binding cleft, to form a catalytic triad similar to that found in serine proteases. In a study by Nixon *et al.* (1999), the nuclear transport factor 2 was used as a scaffold for the rational re-engineering of the active site derived from the structurally homologous scylatone dehydratase. Although still weak, the achieved enzymatic activity was above background by a factor of 150, more or less demonstrating the feasibility of this approach.

Finally, an investigation by Altamirano *et al.* (2000) focused on two members of the triose phosphate isomerase (TIM) family of enzymes, whose well conserved $(\alpha/\beta)_8$ barrel obviously represents Nature's own preferred scaffold for the creation of biocatalysts during evolution. An attractive property of these enzymes is the bipartite character of the active centre, with the arrangement of substrate-binding residues primarily within the barrel itself and the catalytic residues mostly in the connecting loop regions. Using a combination of rational engineering and random mutagenesis/recombination along with metabolic selection of *E. coli*, indole-3-glycerol-phosphate synthase was converted into a phosphoribosyl-anthranilate isomerase, i.e. that kind of biocatalyst which normally provides the substrate for the former enzyme as part of the tryptophan biosynthetic pathway. The resulting enzyme was even more active than the natural one of *E. coli*, mostly owing to the better binding of its substrate.

Even though these three examples served to mimick biocatalytic functions which were already known, it is conceivable that fully novel enzymatic functions may be realized following similar strategies. In this respect it appears that the TIM barrel is likely to provide the more promising basis as compared with the Ig fold, for example. Catalytic antibodies are generated by immunizing animals with a transition state analogue of a chemical reaction as a hapten (for a review see Lerner *et al.*, 1991). This concept is of interest because it has proven that some classes of chemical transformation that were previously unknown in biochemistry can in principle be tackled (Schultz and Lerner, 1995). However, catalytic antibodies suffer from notoriously weak activities, probably because the Ig topology is hardly capable of encapsulating one or even two substrates from solvent, as it is often characteristic for enzymatic reaction mechanisms. The fact that the $(\alpha/\beta)_8$ barrel has been adopted by nature for a huge variety of different metabolic reactions clearly indicates that this should be the better scaffold for the engineering of artificial biocatalysts.

CONCLUSIONS: WHICH SCAFFOLD TO CHOOSE?

The concept of scaffolds emerged from the field of antibody engineering, after it had been recognized that the modular

Table 1. Prominent protein scaffolds to be engineered for molecular recognition

Name	Template	Size	No. of loops	Crosslinks
scFv	Immunoglobulin	ca. 260	3 + 3	2 S-S
Camelized V_H	VHH/human V_H	ca. 130	3	2 S-S
Minibody	Trimmed Ig V_H	61	2	—
FN3	Fibronectin III	94	2	—
Anticalin	Lipocalin (BBP)	160–180 (174)	4	0–3 (2) S-S
Affibody	Protein A	58	(α)[a]	—
(—)	Cytochrome b_{562}	106	2	Bound heme
Knottin	CBD/EETI II	36/28	(β)[a]	2/3 S-S
(—)	Zinc-finger	26	(α)[a]	Bound Zn^{2+}
Kunitz domain	BPTI/APPI	58	1–2	3 S-S
(—)	Tendamistat	74	2–3	2 S-S
Aptamer	Thioredoxin	108	1	1 S-S
(—)	Staph. nuclease	149	1	—
Glubody	GST P1-1/A1-1	2 × 209/222	2 × 2	—

Most of the listed proteins have been tested for the generation of molecular random libraries. For references see text.
[a] Mutated residues were displayed on rigid secondary structure elements instead of loop segments.

character of their biomolecular architecture might similarly be found in other proteins and thus exploited for novel purposes. The technology of generating and screening libraries of antibodies, mainly via phage display, has proven to be practically useful for the preparation of artificial binding proteins on the basis of Ig fragments, and without doubt this technology is currently still dominant. However, with increasing maturation in the antibody engineering field the intrinsic disadvantages of the immunoglobulin architecture become more and more apparent.

First, antibodies possess considerable size and even their smallest functional fragment, i.e. the F_v fragment composed of ca. 250 amino acids, is quite large. In terms of economy of production, tissue penetration in medical treatment, and several technical applications, smaller entities are desired. Second, antibodies are composed of two different polypeptide chains, which can lead to unstable association for their minor fragments and necessitates complicated cloning steps for the handling of a pair of coding regions in parallel. The introduction of a linker segment in the so-called scF_v fragments has only partially solved the problem because their stabilities are often still low and the antigen-binding capability may be affected by the extraneous polypeptide segment. In addition, not much is known about its immunogenicity in human patients. Third, there are altogether six hypervariable loops in the combining site, which are difficult to manipulate simultaneously and all of which are probably not needed. Therefore, tiny proteins with a potential active site composed of a smaller number of loops is certainly favourable for biotechnological purposes.

However, although the use of protein scaffolds has been generally accepted in modern biochemistry, not all alternative protein scaffolds that have been considered up to now are likely to be useful. Apart from those examples which have already been discussed here, some further polypeptide topologies have been proposed as scaffolds, however without experimental proof to date. For example, the observation that the superfold of the pleckstrin homology (PH) domain occurs in several protein families with distinct functions, despite absent sequence similarity, prompted its suggestion as a promising candidate (Steipe *et al.*, 1997; Blomberg *et al.*, 1999). The so-called β-prism motif (Shimizu and Morikawa, 1996), which was discovered in the two carbohydrate-binding proteins vitelline membrane outer layer protein I and δ-endotoxin, again lacking any discernible sequence similarity, was proposed as a framework structure as well.

In an evaluation of the current approaches (Table 1) not only the potential structural plasticity of the binding site that may be created on the basis of a given protein scaffold (Fig. 2), but also the affinities and specificities that have been experimentally achieved for prescribed ligands must serve as criteria. In the first respect, scaffolds which only provide a single loop for variegation, i.e. most protease inhibitors and the engineered enzymes with loop insertions in their active sites, are probably of a less broad use. In the second respect, one has to distinguish between macromolecular 'antigens' and low molecular weight 'haptens' as target ligands.

Many of the scaffolds tested so far are probably capable of complexing other proteins, especially because these macromolecular targets usually provide grooves into which exposed loop segments or convex surface regions can intimately bind. Single Ig and Ig-like domains, affibodies and knottins are likely to be suitable in this respect. However, it should be kept in mind that knottins have not yet proven to yield tight-binding variants, and even in the case of affibodies considerable effort was necessary—i.e. repeated rounds of engineering and affinity maturation—in order to obtain one variant with a sub-micromolar K_D value for the complexation of a huge DNA-polymerase.

In the case of small haptens the situation is even more critical. Albeit several scaffolds have been investigated in this regard, including single Ig domains, helix bundles and GST variants, the affinites were usually low, and a context dependence in terms of the macromolecular hapten carrier could be observed. Just one type of scaffold, the anticalins,

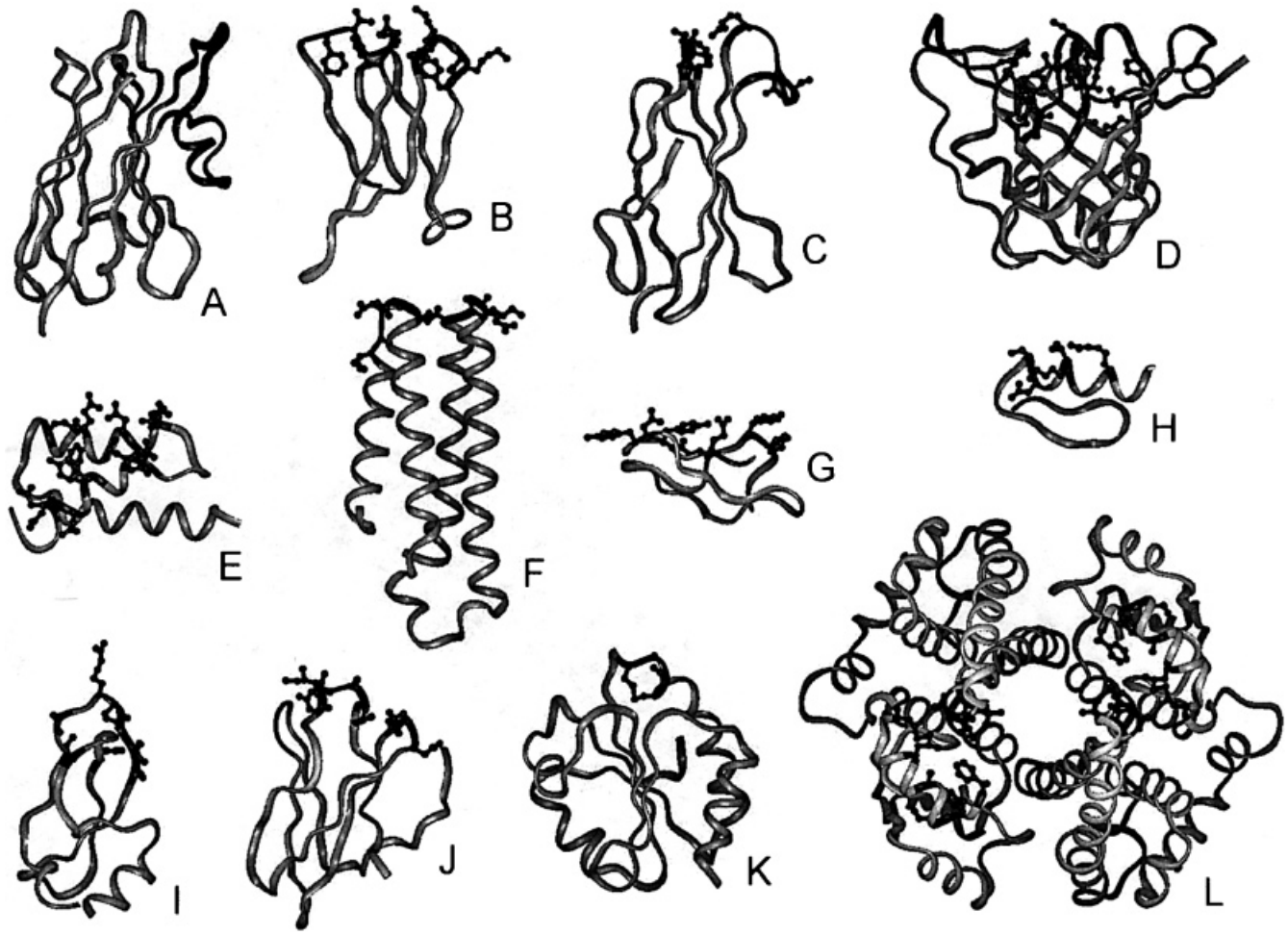

Figure 2. Various protein scaffolds used for library construction and selection of variants with novel ligand-binding function (cf. Table 1). Loops that are thought to be variable in their primary structure are coloured black and the side chains of those amino acids which have actually been variegated for the generation of a molecular library are displayed. (A) Camel VHH (chain A from PDB entry 1MEL); (B) minibody (appropriate segment of chain H from 1MCP); (C) FN3 (1FNA); (D) anticalin (chain A from 1BBP); (E) affibody (1BDD); (F) cytochrome b_{562} (256B); (G) knottin (1CBH); (H) zinc-finger (middle domain of chain C from 1MEY); (I) Kunitz domain (chain A from 1AAP); (J) tendamistat (1HOE); (K) thioredoxin (chain A from 2TRX); (L) GST A1-1 (1GUH).

whose design rests on the natural ligand-binding proteins from the lipocalin family, has clearly been shown to provide receptor proteins with high hapten affinity and extraordinary specificity. They might even have a greater potential than antibodies in corresponding applications because their extended pocket ensures more complete complexation of smaller compounds than is possible in the Ig combining site, which merely provides a pocket with limited size at the interface between the V_H and V_L domains. Finally, the lipocalin family possesses a partitioned architecture, i.e. a structurally conserved β-barrel supporting four hypervariable loops, with a clear resemblance to that of Igs, which was probably optimized in this modular fashion in the course of divergent evolution.

According to the definition given further above a versatile protein scaffold should generally constitute a conformationally stable folding entity that is able to displaying a multitude of loop structures or amino-acid sequences in a localized surface region. This is a prerequisite for the engineering of novel biochemically active sites. Isolated examples describing the successful swapping of a peptide loop from one protein to another (Vita, 1997) do not necessarily demonstrate that the utilized protein core tolerates extended structural variability. However, the question whether a given tertiary structure could provide a promising framework might be better answered using theoretical analyses, taking advantage of the rapidly growing database of experimentally known protein structures. Li *et al.* (1999) have recently presented an interesting compilation of loop conformational diversity. Once extended to more than single loop segments, such a study would certainly aid in the search for architectures that should be useful for functional protein engineering.

In summary, the choice of a suitable polypeptide scaffold opens the way to many applications of protein design starting from a pragmatic approach. Therefore, in contrast with the earlier view, this rapidly developing area of biochemistry has no longer to wait for the protein folding problem to be solved before new and useful biomolecules will be developed and employed in biotechnology and medicine.

REFERENCES

Al-Lazikani B, Lesk AM, Chothia C. 1997. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**: 927–948.

Altamirano MM, Blackburn JM, Aguayo C, Fersht AR. 2000. Directed evolution of new catalytic activity using the α/β-barrel scaffold. *Nature* **403**: 617–622.

Beste G, Schmidt FS, Stibora T, Skerra A. 1999. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc. Natl Acad. Sci. USA* **96**: 1898–1903.

Bianchi E, Venturini S, Pessi A, Tramontano A, Sollazzo M. 1994. High level expression and rational mutagenesis of a designed protein, the minibody. *J. Mol. Biol.* **236**: 649–659.

Bianchi E, Folgori A, Wallace A, Nicotra M, Acali S, Phalipon A, Barbato G, Bazzo R, Cortese R, Felici F, Pessi A. 1995. A conformationally homogeneous combinatorial peptide library. *J. Mol. Biol.* **247**: 154–160.

Bird RE, Walker BW. 1991. Single chain antibody variable regions. *Trends Biotechnol.* **9**: 132–137.

Bishop RE, Penfold SS, Frost LS, Höltje J-V, Weiner JH. 1995. Stationary phase expression of a novel *Escherichia coli* outer membrane lipoprotein and its relationship with mammalian apolipoprotein D. *J. Biol. Chem.* **270**: 23097–23103.

Blomberg N, Baraldi E, Nilges M, Saraste M. 1999. The PH superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.* **24**: 441–445.

Bork P, Holm L, Sander C. 1994. The immunoglobulin fold—structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**: 309–320.

Braisted AC, Wells JA. 1996. Minimizing a binding domain from protein A. *Proc. Natl Acad. Sci. USA* **93**: 5688–5692.

Brunet AP, Huang ES, Huffine ME, Loeb JE, Weltman RJ, Hecht MH. 1993. The role of turns in the structure of an α-helical protein. *Nature* **364**: 355–358.

Cai X, Garen A. 1996. A melanoma-specific V_H antibody cloned from a fusion phage library of a vaccinated melanoma patient. *Proc. Natl Acad. Sci. USA* **93**: 6280–6285.

Christmann A, Walter K, Wentzel A, Krätzner R, Kolmar H. 1999. The cystine knot of a squash-type protease inhibitor as a structural scaffold for *Escherichia coli* cell surface display of conformationally constrained peptides. *Protein Engng* **12**: 797–806.

Clackson T, Wells JA. 1994. *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* **12**: 173–184.

Colas P, Cohen B, Jessen T, Grishina I, McCoy J, Brent R. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**: 548–550.

Cortese R, Monaci P, Nicosia A, Luzzage A, Felici F, Galfré G, Pessi A, Tramontano A, Sollazzo M. 1995. Indentification of biologically active peptides using random libraries displayed on phage. *Curr. Opin. Biotechnol.* **6**: 73–80.

Cowan SW, Newcomer ME, Jones TA. 1990. Crystallographic refinement of human serum retinol-binding protein at 2 Å resolution. *Proteins Struct. Funct. Genet.* **8**: 44–61.

Cull MG, Miller JF, Schatz PJ. 1992. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. *Proc. Natl Acad. Sci. USA* **89**: 1865–1869.

Davies J, Riechmann L. 1995. Antibody VH domains as small recognition units. *Bio/Technology* **13**: 475–479.

Davies J, Riechmann L. 1996. Single antibody domains as small recognition units: design and *in vitro* antigen selection of camelized, human VH domains with improved protein stability. *Protein Engng* **9**: 531–537.

Decanniere K, Desmyter A, Lauwereys M, Ghahroudi MA, Muyldermans S, Wyns L. 1999. A single-domain antibody fragment in complex with RNaseA: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **7**: 361–370.

DeGrado WF, Summa CM, Pavone V, Nastri F, Lombardi A. 1999. De novo design and structural characterization of proteins and metalloproteins. *A. Rev. Biochem.* **68**: 779–819.

Dennis MS, Lazarus RA. 1994. Kunitz domain inhibitors of tissue factor-factor VIIa. *J. Biol. Chem.* **269**: 22129–22136.

Desmyter A, Transue TR, Ghahroudi MA, Dao Thi M-H, Poortmans F, Hamers R, Muyldermans S, Wyns L. 1996. Crystal structure of a camel single-domain V_H antibody fragment in complex with lysozyme. *Nature Struct. Biol.* **3**: 803–811.

Dimasi N, Matin F, Volpari C, Brunetti M, Biasiol G, Altamura S, Cortese R, De Francesco R, Steinkühler C, Sollazzo M. 1997. Characterization of engineered hepatitis C virus NS3 protease inhibitors affinity selected from human pancreatic secretory trypsin inhibitor and minibody repertoires. *J. Virol.* **71**: 7461–7469.

Essen L-O, Skerra A. 1994. The *de novo* design of an antibody combining site: crystallographic analysis of the V_L domain confirms the structural model. *J. Mol. Biol.* **238**: 226–244.

Fabbrizio E, Le Cham L, Polanowska J, Kaczorek M, Lamb N, Brent R, Sardet C. 1999. Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* **18**: 4357–4363.

Flower DR. 1996. The lipocalin protein family: structure and function. *Biochem. J.* **318**: 1–14.

Gachon A-MF, Lacazette E. 1998. Tear lipocalin and the eye's frontline of defense. *Br. J. Ophthalmol.* **82**: 453–455.

Glockshuber R, Malia M, Pfitzinger I, Plückthun A. 1990. A comparison of strategies to stabilize immunoglobin Fv-fragments. *Biochemistry* **29**: 1362–1367.

Glockshuber R, Schmidt T, Plückthun A. 1992. The disulfide bonds in antibody variable domains: effects on stability, folding *in vitro* and functional expression in *Escherichia coli*. *Biochemistry* **31**: 1270–1279.

Gunneriusson E, Nord K, Uhlén M, Nygren P-Å. 1999. Affinity maturation of a *Taq* DNA polymerase specific affibody by helix shuffling. *Protein Engng* **12**: 873–878.

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Bajyana Songa E, Bendahman N, Hamers R. 1993. Naturally occurring antibodies devoid of light chains. *Nature* **363**: 446–448.

Handel T, DeGrado WF. 1990. De novo design of a Zn^{2+}-binding protein. *J. Am. Chem. Soc.* **112**: 6710–6711.

Hoess RH. 1993. Phage display of peptides and protein domains. *Curr. Opin. Struct. Biol.* **3**: 572–579.

Hoogenboom HR. 1997. Designing and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol.* **15**: 62–70.

Houshmand H, Froman G, Magnusson G. 1999. Use of bacteriophage T7 displayed peptides for determination of monoclonal antibody specificity and biosensor analysis of the binding reaction. *Anal. Biochem.* **268**: 363–370.

Hurle MR, Gross M. 1994. Protein engineering techniques for antibody humanization. *Curr. Opin. Biotechnol.* **5**: 428–433.

Huse WD, Sastry L, Iverson SA, Kang AS, Alting-Mees M, Burton DR, Benkovic SJ, Lerner RA. 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246**: 1275–1281.

Jung S, Honegger A, Plückthun A. 1999. Selection for improved protein stability by phage display. *J. Mol. Biol.* **294**: 163–180.

Keck PC, Huston JS. 1996. Symmetry of Fv architecture is conducive to grafting a second antibody binding site in the Fv region. *Biophys. J.* **71**: 2002–2011.

Koide A, Bailey CW, Huang X, Koide S. 1998. The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* **284**: 1141–1151.

Krizek BA, Amann BT, Kilfoil VJ, Merkle DL, Berg JM. 1991. A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc.* **113**: 4518–4523.

Ku J, Schultz PG. 1995. Alternate protein frameworks for molecular recognition. *Proc. Natl Acad. Sci. USA* **92**: 6552–6556.

Lamminmäki U, Paupério S, Westerlund-Karlsson A, Karvinen J,

Virtanen PL, Lövgren T, Saviranta P. 1999. Expanding the conformational diversity by random insertions to CDRH2 results in improved anti-estradiol antibodies. *J. Mol. Biol.* **291**: 589–602.

LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/Technology* **11**: 187–194.

Le Nguyen DL, Heitz A, Chiche L, Castro B, Boigegrain RA, Favel A, Coletti-Previero MA. 1990. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie* **72**: 431–435.

Lerner RA, Benkovic SJ, Schultz PG. 1991. At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **252**: 659–667.

Li W, Liang S, Wang R, Lai L, Han Y. 1999. Exploring the conformational diversity of loops on conserved frameworks. *Protein Engng* **12**: 1075–1086.

Lu Z, Murray KS, Van Cleave V, LaVallie ER, Stahl ML, McCoy JM. 1995. Expression of thioredoxin random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: a system designed for exploring protein–protein interactions. *Bio/Technology* **13**: 366–372.

Markland W, Ley AC, Lee SW, Ladner RC. 1996a. Iterative optimization of high-affinity protease inhibitors using phage display. 1. Plasmin. *Biochemistry* **35**: 8045–8057.

Markland W, Ley AC, Ladner RC. 1996b. Iterative optimization of high-affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin. *Biochemistry* **35**: 8058–8067.

Martin F, Toniatti C, Salvati AL, Venturini S, Ciliberto G, Cortese R, Sollazzo M. 1994. The affinity-selection of a minibody polypeptide inhibitor of human interleukin-6. *EMBO J.* **13**: 5303–5309.

Martin F, Toniatti C, Salvati AL, Ciliberto G, Cortese R, Sollazzo M. 1996. Coupling protein design and *in vitro* selection strategies: improving specificity and affinity of a designed β-protein IL-6 antagonist. *J. Mol. Biol.* **255**: 86–97.

Martin F, Volpari C, Steinkühler C, Dimasi N, Brunetti M, Biasiol G, Altamura S, Cortese R, De Francesco R, Sollazzo M. 1997. Affinity selection of a camelized V_H domain antibody inhibitor of hepatitis C virus NS3 protease. *Protein Engng* **10**: 607–614.

Martin F, Steinkühler C, Brunetti M, Pessi A, Cortese R, De Francesco R, Sollazzo M. 1999. A loop-mimetic inhibitor of the HCV-NS3 protease derived from a minibody. *Protein Engng* **12**: 1005–1011.

Marshall GR. 1992. Three-dimensional structure of peptide-protein complexes: implications for recognition. *Curr. Opin. Struct. Biol.* **2**: 904–919.

Masat L, Wabl M, Johnson JP. 1994. A simpler sort of antibody. *Proc. Natl Acad. Sci. USA* **91**: 893–896.

McConnell SJ, Hoess RH. 1995. Tendamistat as a scaffold for conformationally constrained phage peptide libraries. *J. Mol. Biol.* **250**: 460–470.

Milne RW, Rassart E, Marcel YL. 1993. Molecular biology of apolipoprotein D. *Curr. Opin. Lipidol.* **4**: 100–106.

Monaco HL, Rizzi M, Coda A. 1995. Structure of a complex of two plasma proteins: transthyretin and retinol-binding protein. *Science* **268**: 1039–1041.

Müller HN, Skerra A. 1994. Grafting of a high-affinity Zn(II)-binding site on the β-barrel of retinol-binding protein results in enhanced folding stability and enables simplified purification. *Biochemistry* **33**: 14126–14135.

Muyldermans S, Lauwereys M. 1999. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. *J. Mol. Recogn.* **12**: 131–140.

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JARG, Hamers R. 1994. Sequence and structure of V_H domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Engng* **7**: 1129–1135.

Nagi AD, Regan L. 1996. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold. Des.* **2**: 67–75.

Napolitano EW, Villar HO, Kauvar LM, Higgins DL, Roberts D, Mandac J, Lee SK, Engqvist-Goldstein A, Bukar R, Calio BL, Jack HM, Tainer JA. 1996. Glubodies: randomized libraries of glutathione transferase enzymes. *Chem. Biol.* **3**: 359–367.

Newcomer ME, Jones TA, Åqvist J, Sundelin J, Eriksson U, Rask L, Peterson PA. 1984. The three-dimensional structure of retinol-binding protein. *EMBO J.* **3**: 1451–1454.

Nixon AE, Firestine SM, Salinas FG, Benkovic SJ. 1999. Rational design of a scytalone dehydratase-like enzyme using a structurally homologous protein scaffold. *Proc. Natl Acad. Sci. USA* **96**: 3568–3571.

Nord K, Nilsson J, Nilsson B, Uhlén M, Nygren P-Å. 1995. A combinatorial library of an α-helical bacterial receptor domain. *Protein Engng* **8**: 601–608.

Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlén M, Nygren P-Å. 1997. Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nature Biotechnol.* **15**: 772–777.

Norman TC, Smith DL, Sorger PK, Drees BL, O'Rourke SM, Hughes TR, Roberts CJ, Friend SH, Fields S, Murray AW. 1999. Genetic selection of peptide inhibitors of biological pathways. *Science* **285**: 591–595.

Nuttall SD, Rousch MJM, Irving RA, Hufton SE, Hoogenboom HR, Hudson PJ. 1999. Design and expression of soluble CTLA-4 variable domain as a scaffold for the display of functional polypeptides. *Proteins Struct. Funct. Genet.* **36**: 217–227.

Olivera BM, Rivier J, Clark C, Ramilo CA, Corpuz GP, Abogadie FC, Mena EE, Woodward SR, Hillyard DR, Cruz LJ. 1990. Diversity of *Conus* neuropeptides. *Science* **249**: 257–263.

Padlan EA. 1994. Anatomy of the antibody molecule. *Mol. Immunol.* **31**: 169–217.

Pessi A, Bianchi E, Crameri A, Venturini S, Tramontano A, Sollazzo M. 1993. A designed metal-binding protein with a novel fold. *Nature* **362**: 367–369.

Predki PF, Agrawal V, Brünger AT, Regan L. 1996. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct. Biol.* **3**: 54–58.

Quéméneur E, Moutiez M, Charbonnier J-B, Ménez A. 1998. Engineering cyclophilin into a proline-specific endopeptidase. *Nature* **391**: 301–304.

Regan L, Clarke ND. 1990. A tetrahedral zinc(II)-binding site introduced into a designed protein. *Biochemistry* **29**: 10878–10883.

Reiter Y, Brinkmann U, Lee B, Pastan I. 1996. Engineering antibody Fv fragments for cancer detection and therapy: disulfide-stabilized Fv fragments. *Nature Biotechnol.* **14**: 1239–1245.

Reiter Y, Schuck P, Boyd LF, Plaksin D. 1999. An antibody single-domain phage display library of a native heavy chain variable region: isolation of functional single-domain VH molecules with a unique interface. *J. Mol. Biol.* **290**: 685–698.

Riechmann L. 1996. Rearrangement of the former VL interface in the solution structure of a camelised, single antibody VH domain. *J. Mol. Biol.* **259**: 957–969.

Riechmann L, Muyldermans S. 1999. Single domain antibodies: comparison of camel VH and camelised human VH domains. *J. Immunol. Meth.* **231**: 25–38.

Riechmann L, Clark M, Waldmann H, Winter G. 1988. Reshaping human antibodies for therapy. *Nature* **332**: 323–327.

Roberts BL, Markland W, Ley AC, Kent RB, White DW, Guterman SK, Ladner RC. 1992. Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl Acad. Sci. USA* **89**: 2429–2433.

Roberts RW. 1999. Totally *in vitro* protein selection suing mRNA-protein fusions and ribosome display. *Curr. Opin. Chem. Biol.* **3**: 268–273.

Röttgen P, Collins J. 1995. A human pancreatic secretory trypsin inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. *Gene* **164**: 243–250.

Schiweck W, Skerra A. 1995. Fermenter production of an artificial Fab fragment, rationally designed for the antigen cystatin, and its optimized crystallization through constant domain shuffling. *Proteins Struct. Funct. Genet.* **23**: 561–565.

Schiweck W, Skerra A. 1997. The rational construction of an antibody against cystatin: lessons from the crystal structure of an artificial F_{ab} fragment. *J. Mol. Biol.* **268**: 934–951.

Schlehuber S, Beste G, Skerra A. 2000. A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J. Mol. Biol.* **297**: 1105–1120.

Schmidt TGM, Skerra A. 1994. One-step affinity purification of bacterially produced proteins by means of the "Strep tag" and immobilized recombinant core streptavidin. *J. Chromatogr. A* **676**: 337–345.

Schmidt AM, Müller HN, Skerra A. 1996. A Zn(II)-binding site engineered into retinol-binding protein exhibits metal-ion specificity and allows highly efficient affinity purification with a newly designed metal ligand. *Chem. Biol.* **3**: 645–653.

Schultz PG, Lerner RA. 1995. From molecular diversity to catalysis: lessons from the immune system. *Science* **269**: 1835–1842.

Shimizu T, Morikawa K. 1996. The β-prism: a new folding motif. *Trends Biochem. Sci.* **21**: 3–6.

Simon T, Rajewsky K. 1992. A functional antibody mutant with an insertion in the framework region 3 loop of the V_H domain: implications for antibody engineering. *Protein Engng* **5**: 229–234.

Skerra A. 1993. Bacterial expression of immunoglobulin fragments. *Curr. Opin. Immunol.* **5**: 256–262.

Skerra A, Plückthun A. 1988. Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* **240**: 1038–1041.

Smith GP. 1991. Surface presentation of protein epitopes using bacteriophage expression systems. *Curr. Opin. Biotechnol.* **2**: 668–673.

Smith GP, Patel SU, Windass JD, Thornton JM, Winter G, Griffiths AD. 1998. Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J. Mol. Biol.* **277**: 317–332.

Spinelli S, Frenken L, Bourgeois D, de Ron L, Bos W, Verrips T, Anguille C, Cambillau C, Tegoni M. 1996. The crystal structure of a llama heavy chain variable domain. *Nature Struct. Biol.* **3**: 752–757.

Steipe B, Plücktun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**: 188–192.

Steipe B, Bruhn H, Funk M, Henkel T. 1997. Novel synthetic protein structural templates for the generation, screening and evolution of functional molecular surfaces. International Patent Publication WO 97/45538.

Stevens FJ, Solomon A, Schiffer M. 1991. Bence Jones proteins: a powerful tool for the fundamental study of protein chemistry and pathophysiology. *Biochemistry* **30**: 6803–6805.

Szardenings M, Vasel B, Hecht HJ, Collins J, Schomburg D. 1995. Highly effective protease inhibitors from variants of human pancreatic secretory trypsin inhibitor (hPSTI): an assessment of 3-D structure-based protein design. *Protein Engng* **8**: 45–52.

Tanaka T, Urade Y, Kimura H, Eguchi N, Nishikawa A, Hayaishi O. 1997. Lipocalin-type prostaglandin D synthase (β-trace) is a newly recognized type of retinoid transporter. *J. Biol. Chem.* **272**: 15789–15795.

Uhlén M, Forsberg G, Moks T, Hartmanis M, Nilsson B. 1992. Fusion proteins in biotechnology. *Curr. Opin. Biotechnol.* **3**: 363–369.

Vita C. 1997. Engineering novel proteins by transfer of active sites to natural scaffolds. *Curr. Opin. Biotechnol.* **8**: 429–434.

Vita C, Roumestand C, Toma F, Ménez A. 1995. Scorpion toxins as natural scaffolds for protein engineering. *Proc. Natl Acad. Sci. USA* **92**: 6404–6408.

Vita C, Vizzavona J, Drakopoulou E, Zinn-Justin S, Gilquin B, Ménez A. 1998. Novel miniproteins engineered by the transfer of active sites to small natural scaffolds. *Biopolymers* **47**: 93–100.

Wang CI, Yang Q, Craik CS. 1995. Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin. *J. Biol. Chem.* **270**: 12250–12256.

Ward ES, Güssow D, Griffiths AD, Jones PT, Winter G. 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* **341**: 544–546.

Wells JA, Lowman HB. 1992. Rapid evolution of peptide and protein binding properties *in vitro*. *Curr. Opin. Struct. Biol.* **2**: 597–604.

Widersten M, Mannervik B. 1995. Glutathione transferases with novel active sites isolated by phage display from a library of random mutants. *J. Mol. Biol.* **250**: 115–122.

Woodward SR, Lourdes JC, Olivera BM, Hillyard DR. 1990. Constant and hypervariable regions in conotoxin propeptides. *EMBO J.* **9**: 1015–1020.

Young KH. 1998. Yeast two-hybrid: so many interactions, in so little time. … *Biol. Reprod.* **58**: 302–311.

Zhao B, Helms LR, DesJarlais RL, Abdel-Meguid SS, Wetzel R. 1995. A paradigm for drug discovery using a conformation from the crystal structure of a presentation scaffold. *Nature Struct. Biol.* **2**: 1131–1137.

