nature biotechnology

High-affinity binders selected from designed ankyrin repeat protein libraries

H Kaspar Binz[1,2], Patrick Amstutz[1,2], Andreas Kohl[1,2], Michael T Stumpp[1], Christophe Briand[1], Patrik Forrer[1], Markus G Grütter[1] & Andreas Plückthun[1]

We report here the evolution of ankyrin repeat (AR) proteins *in vitro* for specific, high-affinity target binding. Using a consensus design strategy, we generated combinatorial libraries of AR proteins of varying repeat numbers with diversified binding surfaces. Libraries of two and three repeats, flanked by 'capping repeats,' were used in ribosome-display selections against maltose binding protein (MBP) and two eukaryotic kinases. We rapidly enriched target-specific binders with affinities in the low nanomolar range and determined the crystal structure of one of the selected AR proteins in complex with MBP at 2.3 Å resolution. The interaction relies on the randomized positions of the designed AR protein and is comparable to natural, heterodimeric protein-protein interactions. Thus, our AR protein libraries are valuable sources for binding molecules and, because of the very favorable biophysical properties of the designed AR proteins, an attractive alternative to antibody libraries.

Repeat proteins are ubiquitous binding molecules fundamental to many biological processes[1–3]. Their modular architecture is presumably the key to their evolutionary success[4]. Repeat proteins are characterized by consecutive homologous structural units (repeats), which stack to form an elongated protein domain with a continuous hydrophobic core[3]. In principle, this architecture allows their binding specificities to evolve not only by point mutations but also by insertion, deletion or shuffling of repeats[5]. This evolutionary strategy might enable repeat proteins to acquire new functions by adjusting their surface without jeopardizing their overall topology. AR proteins are one prominent repeat protein family illustrating the binding versatility of repeat proteins. They occur throughout all phyla and mediate protein-protein interactions in the nucleus or cytoplasm, or while anchored to the membrane or when secreted into the extracellular space[6]. AR proteins are built from stacked, 33 amino acid repeats, each forming a β-turn that is followed by two antiparallel α-helices and a loop reaching the β-turn of the next repeat[7]. In most known complexes, the β-turn and the first α-helix mediate the interactions with the target, and different numbers of adjacent repeats are involved in binding[7]. The reported target binding affinities of natural AR proteins are in the low nanomolar range[8,9].

In biotechnology and biomedical research, antibodies and fragments thereof are the most widely used specific, high-affinity binding molecules. Antibodies can be generated against essentially any target either by immunization or by using natural or rationally designed antibody libraries *in vitro*[10,11]. Yet, many antibodies have relatively low expression yields, a tendency to aggregate and a dependence on disulfide bonds for stability. An ideal alternative protein scaffold would have none of these drawbacks while still exhibiting the same affinity and specificity as antibodies. Previous attempts to generate alternative binding molecules relied on either loop or surface randomization of protein scaffolds, which are typically small and always fixed in dimension[12–14].

We present here the results of a different strategy[15] to design and select alternative protein scaffolds, relying on the modularity of AR proteins (**Fig. 1**). We generated combinatorial libraries of consensus-designed AR proteins of varying sizes (that is, varying repeat numbers) with randomized potential interaction surfaces. Unselected library members are very well expressed, soluble, thermodynamically stable and show the typical AR domain fold[16,17]. Here, we show the successful selection of binding molecules from these libraries, proving that our design strategy works. We selected specific binders with high affinities for the *Escherichia coli* maltose binding protein (MBP) and two eukaryotic mitogen-activated protein kinases (MAPKs). The crystal structure of one of the selected binders in complex with MBP was determined, revealing atomic level insights into the mode of target interaction of this class of designed binding molecules.

RESULTS

Designed AR protein libraries

We designed a consensus AR module consisting of six diversified potential interaction residues (which can be any amino acid except cysteine, glycine and proline) and 27 framework residues (26 are fixed and one is allowed to be asparagine, histidine or tyrosine). This module was designed from sequence alignments and structural analyses (**Fig. 1**)[16]. The randomized potential interaction residues are located in the β-turn and the first α-helix of the AR module. We cloned varying numbers of this repeat module between capping repeats, which are special terminal repeats of AR domains shielding the hydrophobic

[1]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [2]These authors contributed equally to this work. Correspondence should be addressed to A.P. (plueckthun@bioc.unizh.ch) or M.G.G. (gruetter@bioc.unizh.ch).

Published online 18 April 2004; doi:10.1038/nbt962

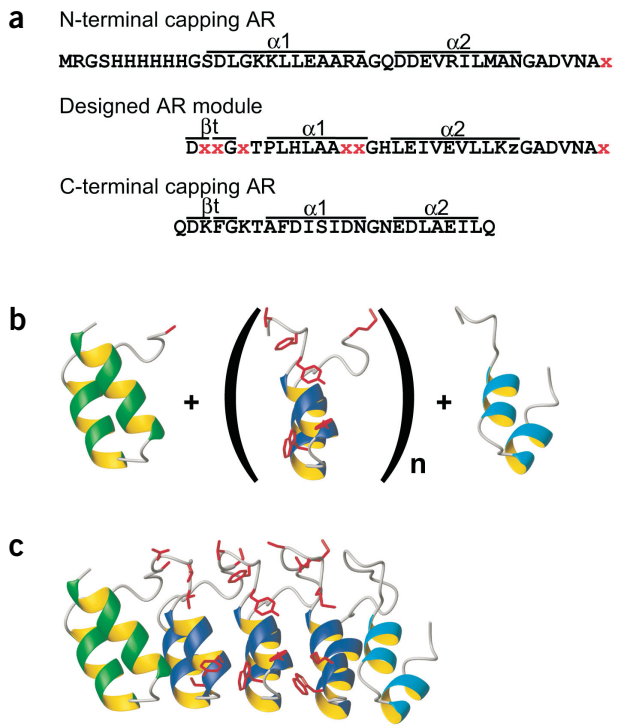

Figure 1 Construction of designed AR protein libraries. (**a**) Sequences of the N-terminal capping AR, the designed AR module and the C-terminal capping AR. The secondary structure elements are indicated above the sequences. The designed AR module consists of 26 defined framework residues, six randomized potential interaction residues (red x, any of the 20 natural amino acids except cysteine, glycine or proline) and one randomized framework residue (z, any of the amino acids asparagine, histidine or tyrosine). The designed AR module was derived via sequence and structure consensus analyses[16]. (**b**) Schematic representation of the library generation of designed AR proteins. Note that this assembly is represented on the protein level, whereas the real library assembly is on the DNA level. By assembling an N-terminal capping AR (green), varying numbers of the designed AR module (blue) and a C-terminal capping AR (cyan), combinatorial libraries of designed AR proteins of different repeat numbers were generated (side chains of the randomized potential interaction residues are shown in stick-mode in red). (**c**) Ribbon representation of the selected MBP binding AR protein off7 (colors as in **b**). This binder is derived from a library consisting of a N-terminal capping AR, three designed AR modules and a C-terminal capping AR. This figure was made with MolMol[49].

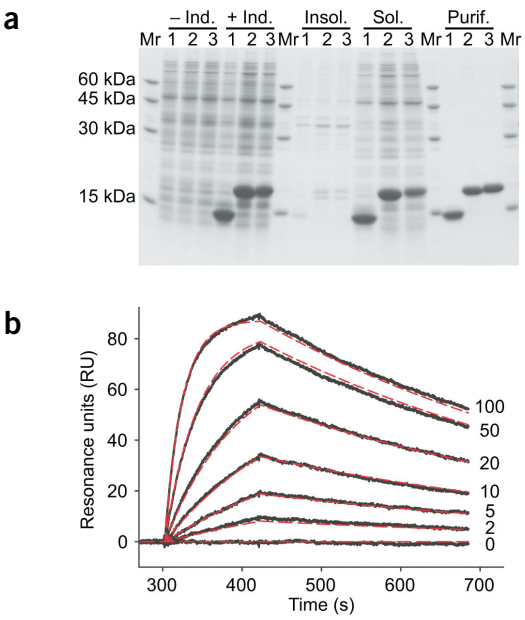

Figure 2 Expression, purification and SPR analysis of selected AR proteins. (**a**) Expression and purification of the selected MBP binders mbp3_16 (1), off7 (2) and mbp3_5 (3). At $OD_{600} = 0.6$, the noninduced (– Ind.) cultures were induced with 0.5 mM IPTG and grown for 4 h at 37 °C (+ Ind.). After cell lysis, the AR proteins are in the soluble fraction (Sol.). Insol., insoluble fraction. The proteins were then purified in a single IMAC purification step (Purif.). (**b**) BIAcore analysis of off7. Different concentrations of off7 (0, 2, 5, 10, 20, 50 and 100 nM) were applied to a flow cell with immobilized MBP for 2 min, followed by washing with buffer flow. The global fit is indicated in the figure by red dashed lines (see **Table 1** for the extracted kinetic data).

core (**Fig. 1**). We thereby increased the size of the potential interaction surface and potentiated its diversity. This strategy provided combinatorial libraries of designed AR proteins with distinct repeat numbers. The libraries were named N2C and N3C, indicating proteins had an N-terminal capping repeat, two and three, respectively, designed (and randomized) AR modules and a C-terminal capping repeat (**Fig. 1**). We used the N2C and N3C libraries for the selections, because AR proteins of this length are very abundant in nature[6]. Unselected members of these libraries were expressed in soluble form at about 200 mg/l in *E. coli* shake flask cultures. These proteins were monomeric, showed circular dichroism (CD) spectra indistinguishable from natural AR proteins[16] and the AR fold was confirmed by a crystal structure of an unselected library member[17]. Similar to designed proteins in other consensus repeat protein studies[18,19], our designed AR proteins showed high thermodynamic stability during unfolding induced by heat[16] and denaturants[17]. Hence, the consensus-designed AR proteins are stable scaffolds with large and modular potential interaction surfaces. The theoretical diversities of these libraries are $5.2 \cdot 10^{15}$ (N2C) and $3.8 \cdot 10^{23}$ (N3C)[16]. The DNA libraries used in the selections contained at least 10^{10} individual members each as estimated from the amount of ligated library DNA. The library diversities were further increased in subsequent PCR cycles.

Ribosome-display selection against MBP

We chose *E. coli* MBP as the first target protein for evaluating our libraries, because it can be obtained in large amounts in pure form and because its structure is known (Protein Data Bank (PDB) entry 1LLS)[20]. We did the ribosome-display selections[21] with biotinylated MBP bound to neutravidin in microtiter plates. An enrichment of binders was observed after the second selection round both for the N2C and the N3C libraries. We performed a total of four to five selection rounds before analyzing single, selected library members.

We screened individual selected AR proteins for MBP binding by an enzyme-linked immunosorbent assay (ELISA) using crude *E. coli* extracts. Of 60 N3C AR proteins screened, 18 gave a specific signal (signal/background ≥ 10), compared to 4 of 56 N2C molecules. Sequencing of the 18 MBP binding N3C AR proteins revealed that they could be divided into at least three sequence groups (see **Supplementary Fig. 1** online). However, identical clones were never found and considerable diversity was left, indicating that an even more stringent selection pressure could be applied. For the N2C clones the sequence analysis was less conclusive because of the limited data set. Nevertheless, some repeats of N2C molecules showed striking sequence similarities to repeats of the selected N3C proteins. In both

Table 1 Kinetic binding data of selected clones determined by surface plasmon resonance

Target	Clone name (length)	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [M]
MBP	off7 (N3C)	$4.2 \cdot 10^5$	$1.9 \cdot 10^{-3}$	$4.4 \cdot 10^{-9}$
	mbp3_5 (N3C)	$2.0 \cdot 10^5$	$4.4 \cdot 10^{-3}$	$22 \cdot 10^{-9}$
	mbp3_16 (N2C)	$6.0 \cdot 10^5$	$1.0 \cdot 10^{-2}$	$17 \cdot 10^{-9}$
JNK2	JNK2_2_3 (N2C)	$9.7 \cdot 10^5$	$2.0 \cdot 10^{-3}$	$2.1 \cdot 10^{-9}$
p38	p38_2_3 (N2C)	$9.5 \cdot 10^5$	$3.5 \cdot 10^{-3}$	$3.7 \cdot 10^{-9}$

N2C and N3C molecules, aromatic residues appeared frequently and thus seem to be important in the selected sequences (see **Supplementary Fig. 1** online).

Framework mutations were randomly scattered and occurred at low frequencies but were present in all molecules (on average, 3.8 amino acid mutations per N3C and 2.4 mutations per N2C molecule; see **Supplementary Fig. 1** online). Each library member went through at least 450 PCR cycles (library generation and selection), which might explain this finding. The mutations are mostly of a conservative nature, and thus the fold of the AR domains is most probably not affected by the alteration of the framework (see description of the crystal structure below).

Selected AR proteins show high affinity and specificity

The selected AR proteins were expressed at high levels in soluble form in the cytoplasm of *E. coli* (up to 200 mg/l) and purified to homogeneity by a single immobilized metal ion affinity chromatography (IMAC) purification step (**Fig. 2a**). We screened 21 clones (4 N2C, 17 N3C) by surface plasmon resonance (SPR). Using the purified AR proteins at 1 μM, we first compared the on- and off-rates of MBP binding. Three AR proteins with slow off-rates were analyzed at multiple concentrations and evaluated with a global kinetic fit (N2C: mbp3_16; N3C: off7 and mbp3_5) (**Fig. 2b** and **Table 1**). off7, a selected N3C molecule, had the highest affinity for MBP (K_D= 4.4 nM). The N2C molecule mbp3_16 had a dissociation constant of K_D= 17 nM. Hence, both N2C and N3C molecules can be selected to bind MBP with high affinity. Clones that went through five selection rounds had higher affinities for MBP than clones that were selected through four rounds. In SPR experiments, the N3C library member off7 was specific and did not cross-react with phage lambda protein D[22], streptavidin or the aminoglycoside-3′-phosphotransferase APH(3′)-IIIa, a bacterial kanamycin resistance protein[23].

To further investigate specificity, ELISA experiments were done with purified MBP-binding AR proteins (**Fig. 3**). mbp3_16, off7 and mbp3_5 are specific for MBP and do not interact with phage lambda protein D[22], APH[23] or neutravidin (**Fig. 3a**). In a competition ELISA experiment, the binding of off7 to immobilized MBP could be inhibited by preincubation with free MBP (**Fig. 3b**). The affinity estimated from these experiments was consistent with the SPR measurements (50% inhibition at 10 nM). The unselected N3C AR library member E3_5[16,17] did not interact with MBP, indicating that the designed AR domain scaffolds per se do not bind MBP (**Fig. 3b**).

At a concentration of 15 μM, off7 is monomeric (as indicated by size exclusion chromatography) and shows a CD spectrum identical to that of E3_5 (ref. 16), which has an AR domain fold[17] (data not shown).

Selection of specific high-affinity MAPK binders

To further evaluate the potential of our AR protein libraries, we chose the eukaryotic protein kinases JNK2 and p38 as our next target proteins (see ref. 24 and references therein). We did a total of four ribosome-display selection rounds with the N2C library before comparing single, selected library members. Screening 15 clones each, we obtained ten ELISA-positive JNK2 binders and ten ELISA-positive p38 binders. The sequence of one representative member for each of these target-specific groups is given in **Supplementary Figure 1** online. These MAPK binders share many features of the selected MBP binders. They have affinities in the low nM range (**Table 1**); they can be expressed at high levels in soluble form in the cytoplasm of *E. coli* and purified to homogeneity by a single IMAC purification step (data not shown); their randomized positions are enriched in aromatic amino acids (see **Supplementary Fig. 1** online). To investigate their target specificity, we did ELISA experiments with purified, selected AR

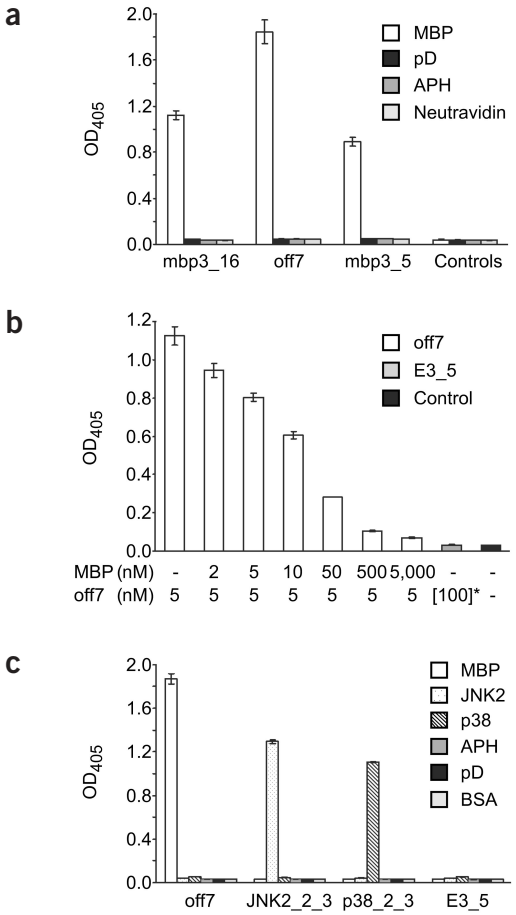

Figure 3 ELISAs with selected AR proteins. (**a**) Specificity of MBP binders. The interaction of the proteins mbp3_16, off7 and mbp3_5 (each 50 nM; control with no AR protein) with immobilized MBP, pD, APH and neutravidin is shown. (**b**) Competition ELISA illustrating the interaction between the selected AR protein off7 and MBP. off7 (5 nM) was incubated with varying concentrations of free MBP before binding on immobilized MBP. The binding to MBP of off7 can be specifically inhibited by increasing concentrations of free MBP in solution. An unselected AR protein of the N3C library (*) showed no interaction with MBP (100 nM of E3_5)[16,17], giving a signal identical to that of the control (no AR protein on immobilized MBP). (**c**) Specificity comparison of an MBP, a JNK2 and a p38 binder. The interaction of 100 nM each of the proteins off7 (binds MBP), JNK2_2_3 (binds JNK2), p38_2_3 (binds p38) and E3_5 (unselected N3C library member) with MBP, JNK2, p38, APH, pD and BSA is shown. Note that in all representations the background binding of the detection antibodies has not been subtracted.

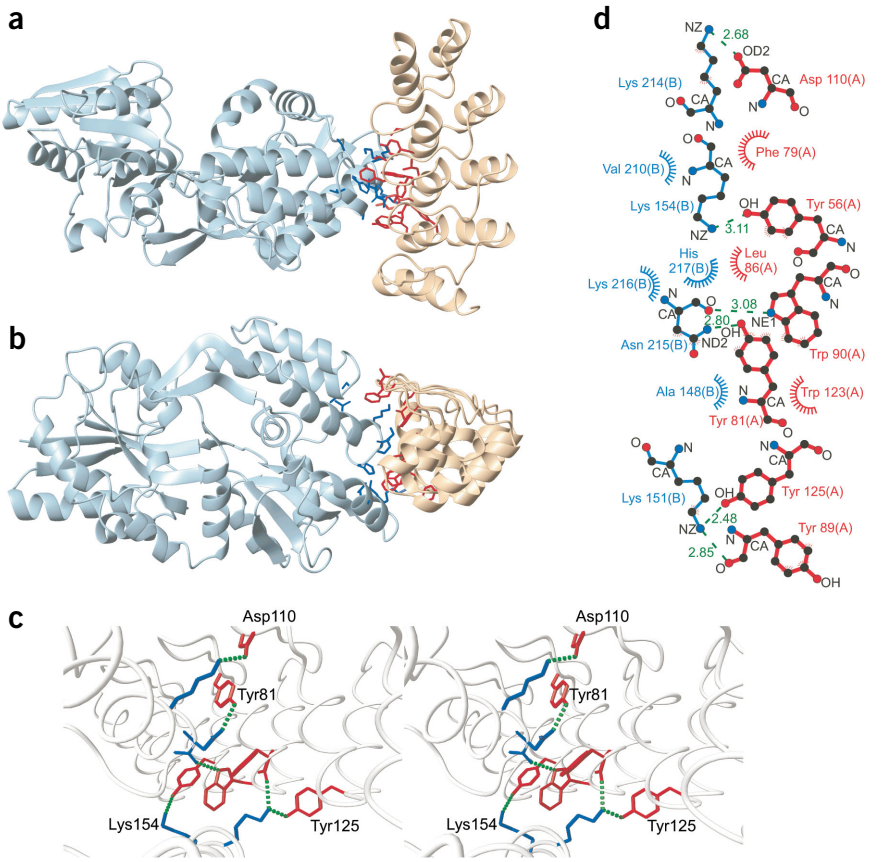

Figure 4 Crystal structure of the designed AR protein off7 in complex with MBP. (**a**,**b**) Two perpendicular views of the complex are shown. MBP is on the left (blue), off7 on the right (ochre). The interaction residues are highlighted in stick-mode in red (off7) and blue (MBP), respectively. (**c**) A close stereo view on the H-bond pattern in similar view as in **b**. Note that in this representation only residues involved in H-bonds (green dashed lines) are shown. For orientation, some residues involved in H-bonding are labeled. **Figure 4a–c** were made with MolMol[49]. (**d**) Ligplot[48] representation of the interaction between MBP (chain B, blue) and off7 (chain A, red). H-bonds (in green) including the H-bond distances as well as residues and atoms involved in hydrophobic contacts (indicated by red or blue rays) are shown.

proteins (**Fig. 3c**). All binders were highly specific for the target proteins that were used to select them. Most importantly, the AR proteins allow perfect discrimination between the homologous MAPKs (51% identity and 59% similarity between JNK2 and p38 on the amino acid level in our format).

Structure determination of off7 in complex with MBP

To validate our AR randomization scheme and to analyze the selected interaction at the atomic level, we determined the crystal structure of one binder (off7) in complex with MBP (see Methods). The phasing problem was solved by molecular replacement using E3_5, a designed N3C AR protein[17] as a search model without using the phases of the larger MBP. Hence, when used in cocrystallization studies, AR proteins may serve as valuable tools to obtain first phases and finally the structure of its binding partner. The results of the data collection and refinement are shown in **Supplementary Table 1** online. As can be seen in the crystal structure (**Figs. 4** and **5**), the AR protein binds the open form of MBP[20] creating an elongated complex. For the AR protein, clear electron density starts at Ser12, but no or only very weak electron density was observed for the N-terminal His_6 tag. The electron density is clear until the second-to-last amino acid (Leu168). In the complex, the AR protein has the typical AR domain fold and is highly similar to the known structure of the designed AR protein E3_5 (root-mean-square deviation of the C_α atoms ($r.m.s.d._{C\alpha}$) < 1 Å)[17]. The main differences between the two structures are found in the β-turn region of repeat module 1 (second repeat), where the C_α chain shows a maximal $r.m.s.d._{C\alpha}$ of 2.1 Å. The H-bond network, which most probably stabilizes the β-turn region[17], is conserved and undisturbed. off7 has three framework mutations (K16R, N74D and H125Y) not present in the library design, but they are conservative and in accordance with the AR domain fold, and only Y125 is involved in MBP binding (see below).

MBP was found in the open conformation with no ligand bound[20]. A superposition on PDB entry 1LLS[20] shows very few differences ($r.m.s.d._{C\alpha}$ < 0.9 Å). The N-terminal His_6 tag and the C terminus of MBP are not defined in the electron density. Clear density extends from Gly19 to Gly387.

Analysis of the interaction of off7 with MBP

The interaction of the AR protein off7 with MBP was analyzed as described in the Methods section. The only direct interaction between off7 and MBP in the crystal lattice is the selected heterodimer interface (**Fig. 4**). Crystal packing contacts are mainly between adjacent off7 molecules and between adjacent MBP molecules. The heterodimer interface is formed by the concave randomized surface of the AR protein off7 (611 $Å^2$ buried surface) and a slightly larger convex surface on the MBP (656 $Å^2$ buried surface), resulting in a total buried surface area of 1,267 $Å^2$ (**Fig. 5**). The off7/MBP complex is further characterized by six H-bonds and a planarity index of 2.1[25]. The details of the interaction are listed in **Table 2** and in **Supplementary Table 2** online.

The interaction of off7 involves residues from all three randomized repeat modules, although the randomized repeat modules 2 and 3 (constituting repeats 3 and 4 in the protein because of the capping repeats) contribute more to the binding than the randomized repeat module 1 does. In total, 9 out of 18 randomized potential interaction residues are involved in the binding to MBP. The interface is characterized by a large number of aromatic residues, which account for 73% of the buried surface area on off7 (**Fig. 4** and **Supplementary Table 2** online). Among these residues, the tyrosines cover 28% of the buried surface area and are involved in four H-bonds (Tyr56, Tyr81, Tyr89 and Tyr125). In the interface (**Fig. 4**), tyrosines have a dual role being both H-bond formers and hydrophobic contact mediators. Besides the four tyrosines, Trp90 and Asp110 also form H-bonds. Three framework residues (Leu86, Asp110 and Tyr125) form part of the interaction surface. Two (Asp110 and Tyr125) form H-bonds to MBP (**Fig. 4**) and the third (Leu86) is engaged in hydrophobic interactions. Interestingly, Tyr125 is a framework mutation (H125Y). With the

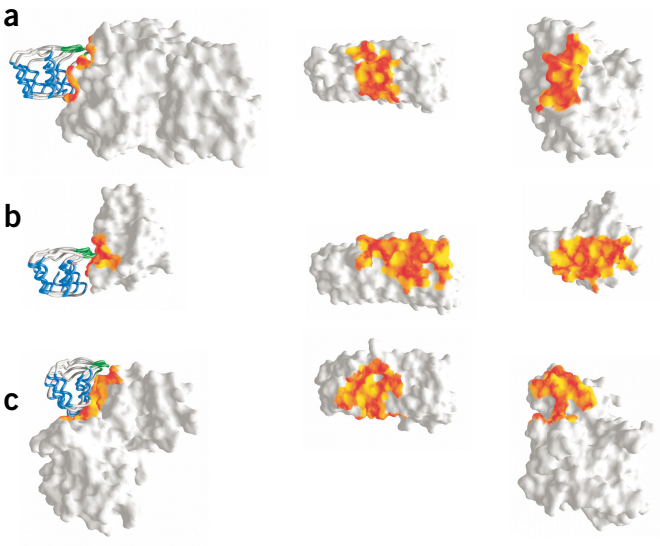

Figure 5 Open sandwich illustrations of the interaction surfaces of AR proteins and their targets. (**a**–**c**) GRASP[50] shape complementarity representations of the interactions between off7 and MBP (shape complementarity, 0.739) (**a**), GABPβ1 and GABPα (PDB entry 1AWC; shape complementarity, 0.665) (**b**), and p18^{INK4c} and CDK6 (PDB entry 1G3N; shape complementarity, 0.688) (**c**), respectively. The complex is shown on the left with the AR proteins in a backbone worm representation (α-helices in blue, β-turns in green) and the target protein in a surface representation. The open sandwich surface representations are shown in the middle (AR proteins) and on the right (targets). The contact areas are stained according to the shape complementarities from orange (low) to red (high).

exception of the high percentage of aromatic amino acids, the residue composition of the off7 interface is comparable to other AR protein complexes. However, the small and functionally biased set of only seven structurally analyzed natural AR protein complexes limits the comparison to their average amino acid composition in the binding interface.

The interaction surface of MBP is located on helices H7, H9 and H17 (**Fig. 4** and **Supplementary Table 2** online). Four lysines (Lys151, Lys154, Lys214 and Lys216) are involved in four H-bonds and together form approximately 60% of the buried surface area on the MBP surface (**Fig. 4**). Every lysine of MBP is in contact with a tyrosine of off7 resulting in three H-bonds and five hydrophobic contacts (see **Supplementary Table 2** online). In general, the surface of the MBP is rather negatively charged except for the spot where off7 binds, where there is a positively charged surface patch formed by the four lysines.

A comparison to natural protein-protein interactions

We compared the off7/MBP complex to all available complex crystal structures of AR proteins, to natural heterodimer complexes, to antibody-antigen complexes and to an affibody complex (see Methods section and **Table 2**). The off7/MBP interaction is comparable to natural protein-protein interactions[25]. The 611 Å^2 buried surface area, one H-bond per 100 Å^2 buried surface area and the planarity value of 2.1 of off7, are all within the normal parameters of heterodimer complexes (**Table 2**)[25]. The buried surface area is at the lower limit of antibody-antigen complexes[26], whereas the affinity is comparable to high affinity monovalent antibody-antigen binding. Otherwise, all values seem to fit the standard parameters for antibody-antigen complexes quite well, but are also similar to those of heterodimer complexes. Natural AR protein interactions can be very diverse in size, composition and in the structural elements involved (**Fig. 5** and **Table 2**). Our randomization scheme of the β-turn and the first α-helix of the designed AR modules was based on crystal structures of natural AR complexes, where it was clear that the AR scaffold is directly used for binding[16]. Because 50% of the randomized target interaction residues that were selected do indeed interact with MBP, the interaction mode is very similar to that of natural AR proteins, such as the GABPβ1 (**Fig. 5**), which partly inspired our library design[16]. In comparison to the natural AR proteins, off7 has a slightly smaller buried surface area but a higher H-bond density (**Table 2**).

Recently the structure of an affibody, another designed binding molecule, based on the staphylococcal three-helix bundle protein A, in complex with its target protein was published[27]. The complex shows a slightly larger buried surface area than the off7/MBP complex but with a comparable number of H-bonds (**Table 2**). The affibody has a thousand-fold lower affinity for its target[28] than the AR protein off7 does for MBP ($K_D = 6\ \mu M$ vs. $K_D = 4.4$ nM, respectively). This is probably because this affibody is in a molten globule state and assumes a defined structure only upon binding[29], leading to a loss of entropy that reduces the overall free energy of binding and thus the observed affinity.

DISCUSSION

We designed AR protein libraries of varying repeat numbers using a consensus design strategy[15–17]. Here we show the successful selection of binding molecules from these libraries. The properties of the designed AR proteins perfectly match the criteria for alternative scaffolds. They are expressed at a high level in soluble form, are monomeric and do not contain any cysteines[16]. Unlike previously presented scaffolds (for reviews see refs. 12,13), which typically use either randomized loops or a randomized surface on a given protein scaffold for binding, AR proteins use both β-turns and a randomized surface. Most importantly, they are not restricted in dimension because of their modular architecture (**Fig. 1**). Thus, the interaction surface can be adapted by adding more repeat modules. The favorable properties of the molecules in the starting libraries seem to positively influence both the speed of selection and its outcome. In only four to five ribosome-display selection rounds, we were able to enrich the pool in specific, high-affinity protein binders from the N2C and N3C libraries, which appears to be faster than ribosome-display selections from antibody single-chain Fv libraries[30].

The selected AR proteins retain the advantageous properties of the designed AR proteins, being expressed in high amounts in soluble form and free of cysteines (**Fig. 2**). The selected clones specifically recognize the target protein against which they were selected and do not cross-react with other proteins as shown by ELISA (**Fig. 3**). The affinities of the selected clones are in the low nanomolar range (**Table 1** and **Fig. 2**), the association rates are in the typical range for protein-protein interactions (that is, 10^5–$10^6\ M^{-1}s^{-1}$)[31] and the dissociation rates are in the range of 10^{-2} to $2{\cdot}10^{-3}\ s^{-1}$ (**Table 1**). We anticipate that these off-rates can probably be improved further by a diversification step followed by an off-rate ribosome-display selection round[32]. Such a diversification step could involve not only classical error-prone PCR, but also, more interestingly, strategies purely amenable to repeat proteins such as repeat shuffling or repeat elongation.

The selected AR protein sequences contain a high number of aromatic residues (see **Supplementary Fig. 1** online) and in the case of off7, seven aromatic amino acids are involved in MBP binding (**Fig. 4**), including four prominent tyrosines (**Supplementary Table 2** online). Interestingly, high tyrosine content has also been noted in antibody binding sites[26]. The dual interaction role of tyrosine as H-bond former and hydrophobic contact mediator (see Results)[33] is probably the

Table 2 Comparison between AR protein complexes and other protein-protein interactions

PDB entry	Resolution	ΔASA[a] in Å^2	No. H-bonds	No. H-bonds/100 Å^2 ΔASA[a]	No. of salt bridges	Planarity	No. of bridging H_2O
1awc	2.2	853.62	5	0.58	0	2.30	3
1bi7	3.4	1205.5	7	0.29	4	2.50	0
1blx	1.9	845.4	11	1.30	1	2.50	12
1g3n	2.9	843.3	12	1.42	1	2.20	0
1ikn	2.3	763.96	2	0.26	n.d.	2.73	0
1oy3	2.1	1593.9	7	0.44	n.d.	4.01	10
1ycs[b]	2.2	713.7	7	1.00	1	2.60	1
1svx (off7/MBP)	2.3	611.2	6	1.00	0	2.10	0
1lp1 (Affibody)	2.3	848.9	6	0.71	0	2.00	2
Antibody-antigen complexes[c]	–	777 ± 135	–	1.1 ± 0.5	–	2.2 ± 0.4	–
Heterodimeric protein-protein complexes [c]	–	983 ± 582	–	1.1 ± 0.5	–	2.8 ± 0.9	–

[a]Surface area per molecule occluded upon complex formation. [b]1ycs uses a different binding surface than the other AR proteins in this table. [c]According to ref. 25.

reason for this accumulation. In addition, aromatic residues are generally enriched in protein-protein interaction interfaces[25].

Apart from the high content of aromatic residues, the crystal structure of the off7/MBP complex reveals an interaction interface that is comparable to that found in natural heterodimer and antibody-antigen complexes (**Table 2**)[25]. It also shows that the AR protein binds its target with the randomized amino acids (**Fig. 4** and **Supplementary Table 2** online), hence validating our randomization scheme. In the crystal structure of the complex, off7 shows the typical AR domain fold with a high similarity to the unselected N3C library member E3_5 (ref. 17). In solution, the uncomplexed, monomeric off7 shows a CD spectrum virtually identical to that of E3_5 (ref. 16). Thus, the backbone of off7 does not seem to be rearranged in a substantial way upon binding to MBP but rather seems to interact in a key-to-lock mechanism. Such a rigid-body interaction may be advantageous both for affinity (low entropic costs upon binding) and specificity (conformational restriction). p18, a natural AR protein interacting with CDK4/6 (**Fig. 5**), also has a low r.m.s.d.$_{C\alpha}$ of < 0.9 Å between the complexed (PDB entry 1G3N) and the uncomplexed (PDB entry 1IHB) AR protein. However, the interaction with CDK4/6 leads to structural alterations in the target[34]. In contrast to those of the rigid AR domain scaffolds, the binding site of antibodies seem to be able to adopt different conformations. The loops of the complementarity-determining region of typical protein binding antibodies can undergo substantial changes upon binding[35] and the loop flexibility might even be used to accomplish multispecificity[36].

Here, we have successfully validated our AR proteins as designed binding molecules using MBP and two MAPKs as model targets, and thus have introduced a binding molecule with very favorable properties. The combination of the high expression level of designed AR proteins (200 mg/l soluble protein in shake flasks), their high thermodynamic stability (9.5 to 21 kcal/mol[17]), the absence of cysteines, the fast enrichment of binders (four selection rounds), their low nanomolar affinities along with high specificities and their modular architecture compares favorably with reports on antibodies and other alternative scaffolds (such as protein A[28], lipocalins[14], fibronectin[37] or green fluorescent protein[38]). Our results open the door for a number of applications. Apart from being useful in capturing molecules on protein chips or in affinity purification—typical applications for designed binding molecules—designed AR proteins are especially suited for intracellular applications. Their high stability, the absence of disulfide bonds and the selectable high affinities are optimal prerequisites for intracellular inhibitors[39], where antibodies are less than ideal. Generally, AR proteins are also conceivable in therapeutic applications, which are currently a domain of recombinant antibodies. As we have shown here, it is possible to both cocrystallize a target protein with a designed binding AR protein and determine its crystal structure with the help of the AR protein. Hence, designed binding AR proteins could be used in cocrystallization and structure determination of proteins difficult to crystallize, similar to what has been shown with antibodies[40].

METHODS

Molecular biology. Unless stated otherwise, all experiments were done according to protocols found in reference 41. Enzymes and buffers were from New England Biolabs (NEB) or Fermentas. All PCR reactions were done using the proofreading Vent-polymerase (NEB).

Vectors used in antigen production. The different vectors that were prepared for the present study are described in detail in the **Supplementary Methods** online. pQEMBP (GenBank accession no. AY327141) was used for the expression of His-tagged, nonbiotinylated MBP. pAT224 (AY327139) was used for the expression of His-tagged, biotinylated MBP. pAT222 (AY327137) was used for the production of His-tagged, biotinylated pD. pAT222_JNK2 and pAT222_p38 were used for the production of His-tagged, biotinylated JNK2 and p38, respectively. All pAT222 and pAT224 constructs carry an Avi tag for biotinylation at the N terminus and a His_6 tag at the C terminus. pBirAcm (Avidity) was used for *in vivo* biotinylation.

Antigen production and purification. The biotinylated proteins pD, MBP, JNK2 and p38 (plasmids pAT222, pAT224, pAT222_JNK2 and pAT222_p38) were produced using *in vivo* biotinylation with plasmid pBirAcm in *E. coli* XL-1 Blue (Stratagene) according to the protocols of Avidity and QIAgen. Efficient biotinylation was confirmed by ELISA and blotting with a streptavidin-alkaline phosphatase conjugate (Roche) and mass spectrometry. Nonbiotinylated MBP for the ELISA analysis and crystallization was produced in the same way as the AR proteins[16] using pQEMBP in *E. coli* XL-1 Blue. The protein purification was carried out as described[16].

Ribosome-display vector (pRDV; AY327136). The cloning of pRDV is described in detail in the **Supplementary Methods** online. pRDV contains all flanking DNA regions necessary for ribosome display: the T7-promoter, the ribosomal binding site and an in-frame *tolA* gene spacer. Hence, by simple ligation of the DNA encoding the combinatorial library into pRDV and by a PCR using this ligation mix as template, all features necessary for ribosome display are added to the library. The use of pRDV has the advantage that it always provides error-free library flanking regions and that it saves a number of working steps compared to the standard PCR approach for library generation[21].

Generation of combinatorial libraries. The AR protein library generation has been described[16]. We changed that protocol in this study in that all ARs, that is,

the N-terminal capping AR, the designed repeat module and the C-terminal capping AR were used as PCR products for the assembly of the libraries. In the present study, both the N- and the C-terminal capping repeat were amplified by PCR from cloned and verified sequences to reduce sequence errors. In this way, AR protein libraries consisting of an N-terminal capping AR, two or three designed AR modules and a C-terminal capping AR (N2C and N3C libraries) were assembled. To convert the libraries to the ribosome-display format, they were amplified by PCR using oligonucleotides EWT4 (5′-TTCCTCATGAGAG GATCGCATCACCATCACCATCACGGATCCGACCTGGG-3′) and WTC4 (5′-TTTGGGAAGCTTTTGCAGGATTTCAGC-3′) and ligated into pRDV using the restriction enzymes *Bsp*HI (or *Nco*I for pRDV) and *Hin*dIII. The ligation product was purified using QIAquick (QIAgen) columns. The purified ligation served as template for a PCR using oligonucleotides RDVf1 (5′-CCTTTTGCT CACATGACCCG-3′) and tolAk (5′-CCGCACACCAGTAAGGTGTGCGGTT TCAGTTGCCGCTTTCTTTCT-3′). Thereby, combinatorial N2C and N3C DNA libraries were generated in the ribosome-display format.

Ribosome display. The PCR-amplified libraries were transcribed and selections were done as described[21]. For the selection, the biotinylated antigen was immobilized as follows: neutravidin (66 nM, 100 µl/well; Pierce) in TBS150 (50 mM Tris HCl, pH 7.4, 150 mM NaCl) was immobilized on a Maxisorp plate (Nunc) by overnight incubation at 4 °C. The wells were then blocked with 300 µl 0.5% BSA (Fluka) in TBS150 for 1 h at 23 °C. Biotinylated antigen (100 µl, 1 µM) in TBS150 with 0.5% BSA was allowed to bind for 1 h at 4 °C. Before the ribosome-display round, the wells were extensively washed with washing buffer WBT (50 mM Tris acetic acid, pH 7.5, 150 mM NaCl, 50 mM $Mg(CH_3COO^-)_2$, 0.05% Tween 20). A ribosome-display round consisted of two 30-min prepanning steps on neutravidin and a 1 h binding step on the target protein. After washing, RNA purification and reverse transcription (with oligonucleotide tolAk), a first PCR was done using oligonucleotides T7B (5′-ATACGAAAT TAATACGACTCACTATAGGGAGACCACAACGG-3′) and tolAk. This RT-PCR product was purified on an agarose gel and reamplified in a second PCR using the same oligonucleotides. The second PCR product served as template for the next round of ribosome display. The number of RT-PCR cycles was reduced from 40 to 30 to 25 in the first three rounds to monitor the enrichment of binders. Binders were analyzed after four or five rounds.

Analysis of selected binders. From the selected DNA pools, the AR open reading frame was amplified by PCR and cloned into pQE30 (QIAgen) via *Bam*HI/*Hin*dIII (oligonucleotides: EWT3: 5′-TTCCGCGGATCCGAC-CTGGG-3′ and WTC4). The DNA sequences were determined using standard techniques. The sequences of the MBP binding proteins off7 (AY326424), MBP3_5 (AY326425) and MBP3_16 (AY326426) have been deposited in GenBank. The amino acid sequences of all sequenced clones are listed in **Supplementary Figure 1** online. For the ELISA screening, the crude extract of 0.6 ml protein expression cultures was used (expression according to QIAgen). The cell pellets were lysed with 50 µl B-Per (Pierce) and the lysates were mixed with 250 µl TBS500 (50 mM Tris HCl, pH 8.0, 500 mM NaCl) each. For quantitative ELISA, BIAcore, CD, analytical gel-filtration and crystallization, single, selected library members were produced on a liter scale and purified as described[16]. CD spectroscopy and analytical gel-filtration were done as described[16] using 15 µM protein in TBS150 (pH 7.4).

ELISA. Biotinylated antigens were immobilized on neutravidin-coated plates as described above. For the screening of the pools, 100 µl of the above crude extracts were applied to wells with or without immobilized antigen for 1 h at 4 °C. After extensive washing with TBS150, binding was detected with an anti-RGS-His antibody (QIAgen; detects only the RGS-His_6-tag of the AR protein, not the His_6-tag of the antigen), an anti-mouse-IgG-alkaline phosphatase conjugate (Pierce) and p-nitrophenylphosphate (Fluka). Quantitative ELISAs were done in the same manner, except purified protein was used (see **Fig. 3**). For competition ELISA, the purified AR protein off7 was incubated with varying amounts of free MBP before (4 °C, 100 min) and during the binding reaction (see **Fig. 3**).

Surface plasmon resonance (SPR). SPR was measured using a BIAcore 3000 instrument (BIAcore). The running buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl and 0.005% Tween 20. A streptavidin SA chip (BIAcore) was used with 480 RU biotinylated MBP immobilized (440 RU JNK2 and 450 RU p38, respectively). The interactions were measured at a flow of 60 µl/min with 5 min buffer flow, 2 min injection of MBP-binding AR protein in varying concentrations (10 pM to 200 nM) and an off-rate measurement of 40 min with buffer flow. The signal of an uncoated reference cell was subtracted from the measurements. Inhibition BIAcore measurements gave results similar to that of the kinetic analyses (data not shown). The p38 and JNK2 binders were measured similarly, but with an injection time of 3 min. The kinetic data of the interaction were evaluated with a global fit using BIAevaluation 3.0 (BIAcore), Scrubber (BioLogic software) and Clamp[42].

Complex purification and crystallization. MBP and the selected MBP-binding AR proteins off7, mbp3_5 and mbp3_16 were produced as described above. The cell pellets of 1-liter bacterial culture of each MBP and off7 (or mbp3_5 or mbp3_16) were pooled and then lysed using an Emulsiflex C5 (Avestin) followed by additional sonication. The proteins were purified using an IMAC column as described[16], followed by a preparative Superdex-75 (Amersham Pharmacia) size exclusion chromatography step in 10 mM Tris HCl, pH 7.6 and 100 mM NaCl. For every protein mixture, the peak fraction with the smallest molecular weight containing both MBP and the AR protein in equimolar amounts, as determined by SDS-PAGE, was collected and used for crystallization. Light scattering of these fractions was measured as described[17]. It showed a monodisperse particle distribution for all three complexes. For the off7/MBP complex (61.4 kDa calculated mass for the 1:1 complex) an average radius of 3.6 nm, equivalent to a hydrated particle of 95 kDa, was estimated, which corresponds to a nonhydrated particle of 65 kDa. The off7/MBP complex crystallized readily and was further analyzed. The protein complex was concentrated to 26 mg/ml for crystallization. Initial crystallization screening was done in 96-well, sitting drop, square well crystallization plates (Greiner Bio-One). The reservoirs were filled with 100 µl reservoir solution using an 8-channel pipette from a 2 ml 96-well, deep-well block into the crystallization plate. Using an 8-channel pipette, 2 µl of reservoir solution were pipetted in the crystallization well and mixed with 2 µl of protein solution. The initial crystals were refined using standard techniques. The crystals used for data collection grew in about 2–3 weeks in 30% PEG 6000, 0.1 M Tris HCl pH 8–9, 100 mM NaCl in a hanging drop experiment with 500 µl reservoir, 2 µl protein solution mixed with 2 µl water and 2 µl reservoir solution. For data collection the crystals were soaked in the mother liquid with 10% ethylene glycol for about 30 s to 1 min and flash frozen in a cryostream at 100 K.

Data collection, reduction, structure solution and refinement. Data were collected at the European Synchrotron Radiation Facility beamline ID14-1. The data were processed using MOSFLM, SCALA and TRUNCATE[43]. The crystal belonged to space group $P2_1$, with a Matthews coefficient of $V_M = 2.1$ Å^3/Da, corresponding to an estimated water content of 39%.

The crystal structure was determined by molecular replacement using the program AmoRe[44], with the structure of the unselected N3C library member E3_5 (PDB entry 1MJ0[17]) as a search model. A conventional AmoRe protocol (rotation, translation, rigid body refinement) was applied yielding a solution. This information was used to obtain a first electron density. At this point only the AR protein was clearly visible in the electron density. A solvent flipping protocol was then applied to modify the map[45]. MBP in its open form (PDB entry 1LLS)[20] was positioned in the resulting electron density using program O[46]. Because about a third of MBP was visible, manual building would have been feasible as well. The rest of the model building was carried out using the program O[46], the structure refinement was done in CNS[45] resulting in a final model with an R-factor of 19.5 % and an R_{free}-factor of 24.9 % (**Supplementary Table 1** online).

Analysis of the complexes. The structural analysis of the complexes was done as suggested[25]. H-bonds and hydrophobic interactions were calculated with HBPLUS[47], LIGPLOT[48] and DIMPLOT[48] using the default settings. Other parameters were calculated using the protein-protein interaction server http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html[25] or CNS[45].

The atomic coordinates of the described complex were deposited in the PDB (PDB-ID: 1SVX).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS
We thank the members of the Plückthun and Grütter laboratories for valuable discussions and David L. Zechel for the critical reading of the manuscript. H.K.B. was supported by a pre-doctoral fellowship of the Roche Research Foundation. M.T.S. was in receipt of a pre-doctoral scholarship from the Fonds der Chemischen Industrie and the Bundesministerium für Bildung und Forschung. This work was supported by the Swiss National Center of Competence in Research in structural biology and the Swiss Krebsliga.

COMPETING INTERESTS STATEMENT
The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 24 December 2003; accepted 14 January 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Groves, M.R. & Barford, D. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9**, 383–389 (1999).
2. Andrade, M.A., Perez-Iratxeta, C. & Ponting, C.P. Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).
3. Kobe, B. & Kajava, A.V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
4. Marcotte, E.M., Pellegrini, M., Yeates, T.O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
5. Ellis, J., Dodds, P. & Pryor, T. Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* **3**, 278–284 (2000).
6. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins: Struct. Funct. Genet.* **17**, 363–374 (1993).
7. Sedgwick, S.G. & Smerdon, S.J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316 (1999).
8. Suzuki, F. *et al.* Functional interactions of transcription factor human GA-binding protein subunits. *J. Biol. Chem.* **273**, 29302–29308 (1998).
9. Malek, S., Huxford, T. & Ghosh, G. IκBα functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-κB. *J. Biol. Chem.* **273**, 25427–25435 (1998).
10. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349**, 293–299 (1991).
11. Plückthun, A. *et al.* in *Antibody Engineering* (eds. McCafferty, J., Hoogenboom, H.J. & Chiswell, D.J.) 203–252 (IRL Press, Oxford, 1996).
12. Nygren, P.-Å. & Uhlén, M. Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463–469 (1997).
13. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187 (2000).
14. Skerra, A. Imitating the humoral immune response. *Curr. Opin. Chem. Biol.* **7**, 683–693 (2003).
15. Forrer, P., Stumpp, M.T., Binz, H.K. & Plückthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2–6 (2003).
16. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
17. Kohl, A. *et al.* Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA* **100**, 1700–1705 (2003).
18. Mosavi, L.K., Minor, D.L. Jr. & Peng, Z.-y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. USA* **99**, 16029–16034 (2002).
19. Main, E.R., Xiong, Y., Cocco, M.J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (Camb)* **11**, 497–508 (2003).
20. Rubin, S.M., Lee, S.Y., Ruiz, E.J., Pines, A. & Wemmer, D.E. Detection and characterization of xenon-binding sites in proteins by ^{129}Xe NMR spectroscopy. *J. Mol. Biol.* **322**, 425–440 (2002).
21. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942 (1997).
22. Yang, F. *et al.* Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat. Struct. Biol.* **7**, 230–237 (2000).
23. Hon, W.C. *et al.* Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895 (1997).
24. Forrer, P., Tamaskovic, R. & Jaussi, R. Enzyme-linked immunosorbent assay for the measurement of JNK, ERK and p38 kinase activities. *Biol. Chem.* **379**, 1101–1111 (1998).
25. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 13–20 (1996).
26. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
27. Högbom, M., Eklund, M., Nygren, P.-Å. & Nordlund, P. Structural basis for recognition by an *in vitro* evolved affibody. *Proc. Natl. Acad. Sci. USA* **100**, 3191–3196 (2003).
28. Eklund, M., Axelsson, L., Uhlén, M. & Nygren, P.-Å. Anti-idiotypic protein domains selected from protein A-based affibody libraries. *Proteins: Struct. Funct. Genet.* **48**, 454–462 (2002).
29. Wahlberg, E. *et al.* An affibody in complex with a target protein: structure and coupled folding. *Proc. Natl. Acad. Sci. USA* **100**, 3185–3190 (2003).
30. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–1292 (2000).
31. Wodak, S.J. & Janin, J. Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61**, 9–73 (2002).
32. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Plückthun, A. Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. USA* **98**, 75–80 (2001).
33. Mian, I.S., Bradwell, A.R. & Olson, A.J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**, 133–151 (1991).
34. Jeffrey, P.D., Tong, L. & Pavletich, N.P. Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev.* **14**, 3115–3125 (2000).
35. Sundberg, E.J. & Mariuzza, R.A. Molecular recognition in antibody-antigen complexes. *Adv. Protein Chem.* **61**, 119–160 (2002).
36. James, L.C., Roversi, P. & Tawfik, D.S. Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 (2003).
37. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9**, 933–942 (2002).
38. Zeytun, A., Jeromin, A., Scalettar, B.A., Waldo, G.S. & Bradbury, A.R. Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies. *Nat. Biotechnol.* **21**, 1473–1479 (2003).
39. Cattaneo, A. & Biocca, S. The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115–121 (1999).
40. Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842–846 (1995).
41. Sambrook, J., Fritsch, E.F. & Maniatis, T. Molecular cloning: A laboratory manual, edn. 2 (Cold Spring Harbor Press, New York, 1989).
42. Myszka, D.G. & Morton, T.A. CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* **23**, 149–150 (1998).
43. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Navaza, J. *Acta Crystallogr. A* **50**, 157–163 (1994).
45. Brünger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. McDonald, I.K. & Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
48. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
49. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
50. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296 (1991).