Erich Michel*, Andreas Plückthun and Oliver Zerbe*

Peptide binding affinity redistributes preassembled repeat protein fragments

https://doi.org/10.1515/hsz-2018-0355
Received August 24, 2018; accepted November 21, 2018; previously published online December 4, 2018

Abstract: Designed armadillo repeat proteins (dArmRPs) are modular peptide binders composed of N- and C-terminal capping repeats Y and A and a variable number of internal modules M that each specifically recognize two amino acids of the target peptide. Complementary fragments of dArmRPs obtained by splitting the protein between helices H1 and H2 of an internal module show conditional and specific assembly only in the presence of a target peptide (Michel, E., Plückthun, A., and Zerbe, O. (2018). Peptide-guided assembly of repeat protein fragments. Angew. Chem. Int. Ed. 57, 4576–4579). Here, we investigate dArmRP fragments that already spontaneously assemble with high affinity, e.g. those obtained from splits between entire modules or between helices H2 and H3. We find that the interaction of the peptide with the assembled fragments induces distal conformational rearrangements that suggest an induced fit on a global protein level. A population analysis of an equimolar mixture of an N-terminal and three C-terminal fragments with various affinities for the target peptide revealed predominant assembly of the weakest peptide binder. However, adding a target peptide to this mixture altered the population of the protein complexes such that the combination with the highest affinity for the peptide increased and becomes predominant when adding excess of peptide, highlighting the feasibility of peptide-induced enrichment of best binders from intermodular fragment mixtures.

Keywords: NMR spectroscopy; peptide binding; protein design; protein fragment complementation; repeat protein.

***Corresponding authors: Erich Michel and Oliver Zerbe,** Department of Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, e-mail: erichxmichel@gmail.com (E. Michel); oliver.zerbe@chem.uzh.ch (O. Zerbe)
Andreas Plückthun: Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Introduction

Binding proteins that recognize proteins, peptides and small molecules with high affinity and high selectivity have been constructed on either antibody or non-antibody scaffolds. They provide complementary surfaces to the target for sufficient intermolecular contacts (Kuriyan and Cowburn, 1997; Binz et al., 2005; Gebauer and Skerra, 2009; McCafferty and Schofield, 2015; Plückthun, 2015). This interaction surface could theoretically also be reconstituted from two complementary protein fragments that, either in the presence of the ligand or on their own, form a protein complex capable of binding the target. Such an approach has been introduced with the open-sandwich immunoassay, where isolated V_H and V_L domains of an antibody recognize non-overlapping epitopes of a common antigen and assemble only in its presence (Ueda et al., 1996). However, split scaffold proteins that retain their fragment dimerization properties, independent of residues that need to be chosen for particular target recognition, are difficult to obtain. Designed armadillo repeat proteins (dArmRPs) provide an exceptional case of a repeat protein scaffold where complementary fragments form high-affinity complexes that are capable of binding their target peptides (Watson et al., 2014). dArmRPs are constituted of a variable number of tightly packed internal modules M that are flanked by N- and C-terminal capping modules Y_{iii} and A_{ii}, respectively (Figure 1A) (Conti et al., 1998; Parmeggiani et al., 2008; Alfarano et al., 2012; Reichen et al., 2016). Each internal repeat module M is composed of 42 residues that form three helices H1, H2 and H3, which fold into a right-handed triangular spiral that results in an elongated, super-helical protein molecule (Figure 1A). The intermolecular contacts to the target peptide are provided on the concave surface generated by the tight arrangement of adjacent helices H3 (Figure 1A, B). One particular feature of dArmRPs is its modular peptide recognition, i.e. each dipeptide of the extended target peptide is recognized by one internal module M (Figure 1A, B). This property offers a promising potential for the design of protein binders against virtually any extended target peptide from scratch, once internal modules against all possible 400 dipeptide combinations are available.

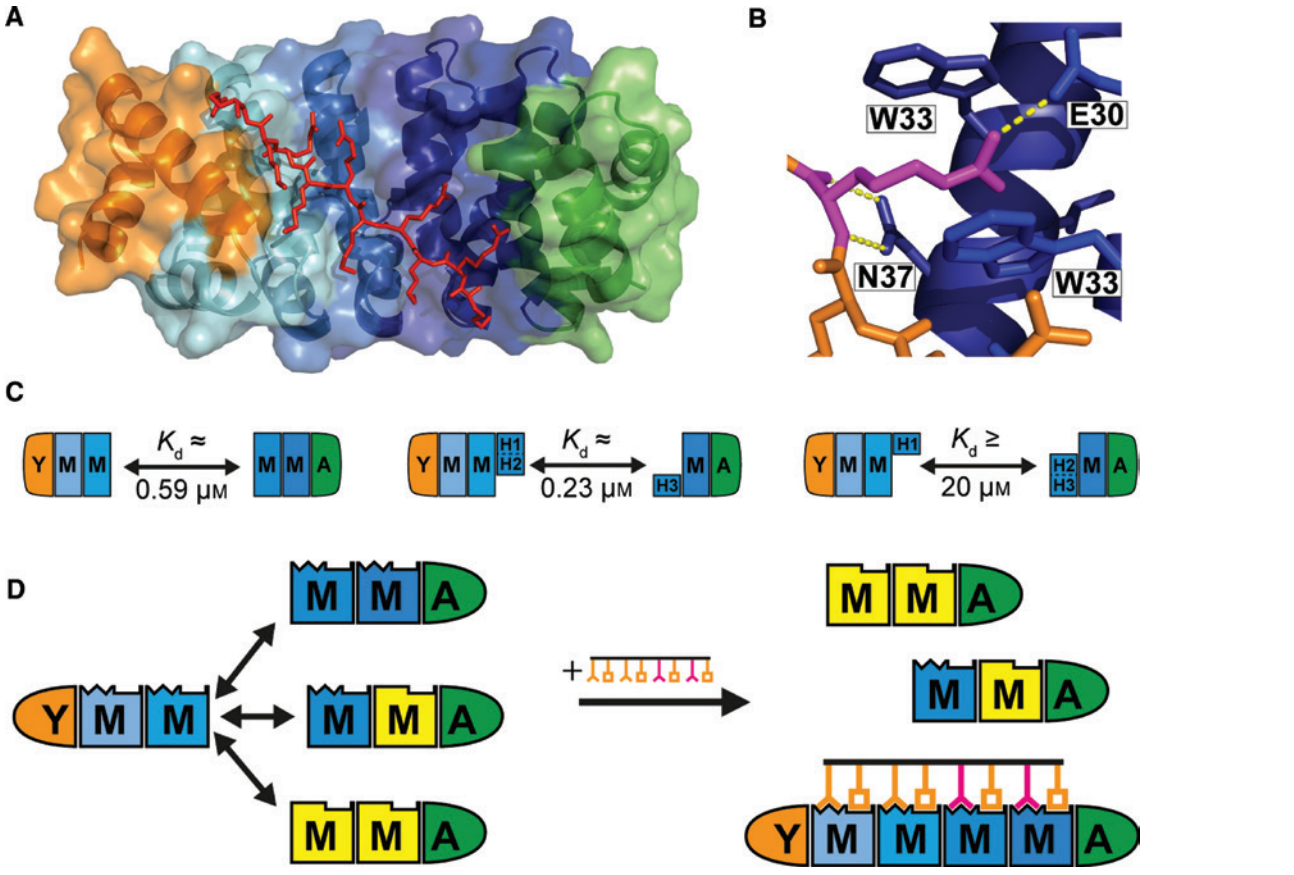

Figure 1: Structural and biochemical features of designed armadillo repeat proteins.
(A) Crystal structure of YM_3A in complex with the $(KR)_5$-peptide (Hansen et al., 2016). dArmRPs bind target peptides with high affinity in a modular fashion, where each M module specifically recognizes two amino acids of the target peptide. (B) The Arg residues of the target peptide are specifically recognized in a pocket formed by the dArmRP, contributing ionic interactions by Glu residues and π-cation interactions with Trp residues, and the peptide is further stabilized by a bifurcated hydrogen bond formed between the peptide backbone amide and the sidechain amide of an Asn residue in each dArmR. (C) Complementary fragments of dArmRPs with various affinities for each other can be prepared by splitting the protein between helices of an M module, leading to fragments pairs that differ up to 100-fold in affinity. (D) Peptide-induced re-shuffling of pre-assembled fragment combinations: An equimolar mixture of one N-terminal and three C-terminal dArmRPs fragments with different affinities for the target peptide (mutated modules shown in yellow) spontaneously assembles upon mixing. Addition of the target peptide enriches the fragment combination that constitutes the best binder. The two discriminated target peptide residues are depicted in magenta.

We recently demonstrated that complementary dArmRP fragments assemble into non-covalent protein complexes with high affinity, provided that the split site is between internal repeat modules: the complex YM_2:MA, derived from the protein YM_3A, structurally resembles the full-length protein YM_3A and retains the ability to bind its peptide ligand neurotensin (Watson et al., 2014). During an investigation of alternative dArmRP split sites (Figure 1C) we discovered an intra-modular site between helices H1 and H2 of an internal module M that yields complementary fragments that assemble only in the presence of a target peptide (Michel et al., 2018). The high discriminatory power of this system enabled enrichment of target peptide-binding fragment combinations that differed in only one out of eight interacting residues (Michel et al., 2018).

In this study, we set out to characterize the high-affinity assembly of dArmRP fragments obtained from an inter-modular split site and to analyze their target binding properties by nuclear magnetic resonance (NMR) spectroscopy. We further used NMR spectroscopy to quantify the complemented populations in an equimolar mixture of an N-terminal YMM fragment with three complementary C-terminal MMA fragments that display incremental affinity towards the target peptide $(KR)_5$ (Figure 1D). Despite the high affinity of the complementary fragments and a clear bias for assembly, in the absence of the target peptide, of the fragment which happens to be the weakest target

binder, we highlight the feasibility of peptide-induced reshuffling of the assembled fragment combinations to a clear enrichment of the best binder to the target peptide.

Results

We here report on the complementation of protein fragments derived from dArmRPs, and how the assembly of fragment mixtures is influenced by the addition of peptide ligands. In our previous report we focused on a detailed analysis of the behavior of two complementary dArmRP fragments that do not significantly interact with each other in absence of a ligand peptide (Michel et al., 2018). Here, we study the assembly of complementary dArmRP fragment pairs derived from an intermodular split (between helices 3 and 1 of the subsequent repeat) or from a split between helix 2 and 3 within an dArmRP module, respectively. We term these YMM:MMA or YMMH12:H3MA, respectively.

In our analysis, we use secondary chemical shifts (Wishart and Sykes, 1994) and heteronuclear $^{15}N\{^{1}H\}$-NOE data (Palmer, 2015) to analyze the structure and dynamics of free and assembled fragments. Chemical shift mapping with NMR titrations are then used to follow the interaction of the assembled YMM:MMA complex with the $(KR)_5$ peptide. We further apply selective amino acid labeling of different C-terminal fragments to readily identify the various mutants and their association in the mixture. Integration of peaks that stem from the free fragments in 2D $[^{15}N,^{1}H]$-HSQC spectra allows us to quantify the assembly of protein complexes in equimolar mixtures of one YMM and three C-terminal MMA fragments with various affinities for a target peptide, and to analyze how the composition is altered in the presence of the peptide ligand.

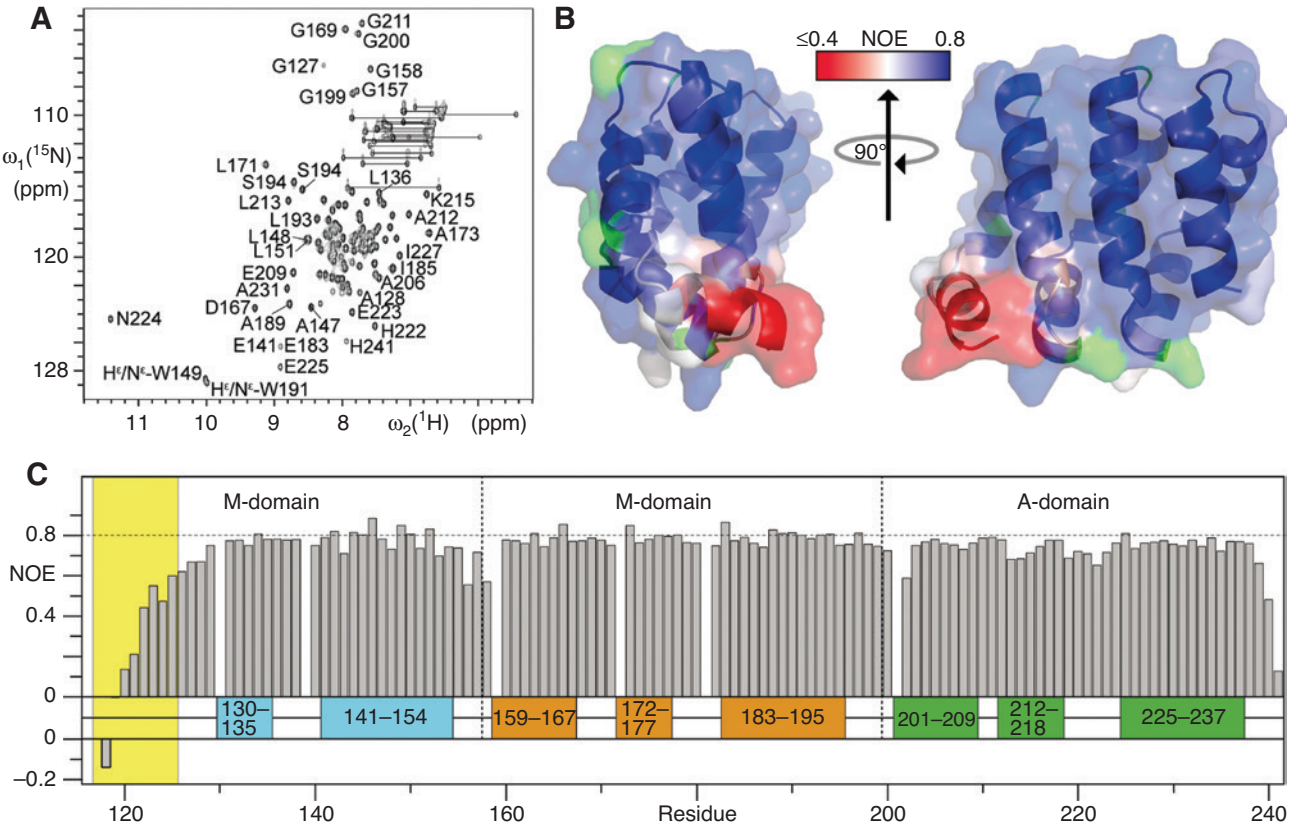

Figure 2: NMR assignments and backbone dynamics of MMA.
(A) 2D $[^{15}N,^{1}H]$-HSQC spectrum showing exemplary amide resonance assignments of free MMA. (B) The mobility of each assigned amide bond was determined by measuring the $^{15}N\{^{1}H\}$-NOE and is color-coded on a structural model of MMA, derived from the YM_5A crystal structure (Hansen et al., 2016). Red and blue patches on the surface representation indicate flexible and rigid amide moieties, respectively, while green represents unassigned and Pro residues. (C) $^{15}N\{^{1}H\}$-NOE values for each residue are plotted against the sequence of MMA. The vast majority of residues display NOE values in the range of 0.7–0.8, which are indicative of a rigid backbone. In contrast to this, residues in the first M module that correspond to helix H1 in the full-length protein (highlighted with yellow background) display significantly increased backbone dynamics.

YMM:MMA complementation induces formation of helix 1 in MMA

Uniformly [^{13}C,^{15}N]-isotope labeled MMA in combination with three-dimensional (3D) triple-resonance NMR provided unambiguous sequence-specific resonance assignments (Figure 2A) with the exception of N117, N159 and N201, for which amide resonances were broadened beyond detection. Secondary chemical shifts revealed that helix H1 in the first M module is not formed in the free MMA fragment (Supplementary Figures 1A, 2A), which is further supported by reduced values of the ^{15}N{^{1}H}-NOE (Figure 2B, C). The remainder of the MMA fragment represents a rigidly folded polypeptide chain, in agreement with the tight packing and high thermodynamic stability of C-terminal dArmRP fragments (Figure 2B, C) (Watson et al., 2014).

To investigate the molecular details of structural changes associated with complex formation between MMA and YMM, we titrated unlabeled YMM in steps of 0.25 equivalents into labeled MMA (Figure 3A). Up to one equivalent of YMM a significant number of MMA amide resonances gradually disappeared and reappeared at new positions, indicating slow chemical exchange on the NMR timescale that is usually observed for association with a sub-micromolar K_d (Figure 3A). Thereafter, no further spectral changes were observed, in agreement with the K_d of ca. 600 nM that we determined by isothermal titration calorimetry (ITC) for the association between the two protein fragments (Supplementary Figure 6A).

All resonances of MMA in the complex with YMM were assigned, again with the exception of N117, N159 and N201 because their resonances were broadened beyond detection. The largest chemical shift changes upon complex formation were observed for the N-terminal residues of the first M module that usually form helix H1 in the full-length dArmRP (Figure 4A). Indeed, an analysis of secondary chemical shifts indicates the induced formation of helix H1 in MMA upon complementation with YMM (Supplementary Figures 1B, 2B), consistent with the formation of a structure similar to the unsplit YM$_4$A protein (Watson et al., 2014).

NMR characterization of free H3MA and its complementation to YMMH12

An alternative fragmentation of a YM$_4$A dArmRP between helices H2 and H3 within the third M module results in YMMH12 and H3MA fragments that assemble with a K_d of 234 ± 62 nM, which is of a similar magnitude as the complementation of YMM with MMA (Figure 1C) (Michel et al., 2018). To further characterize these fragments, we assigned the backbone resonances of free H3MA and

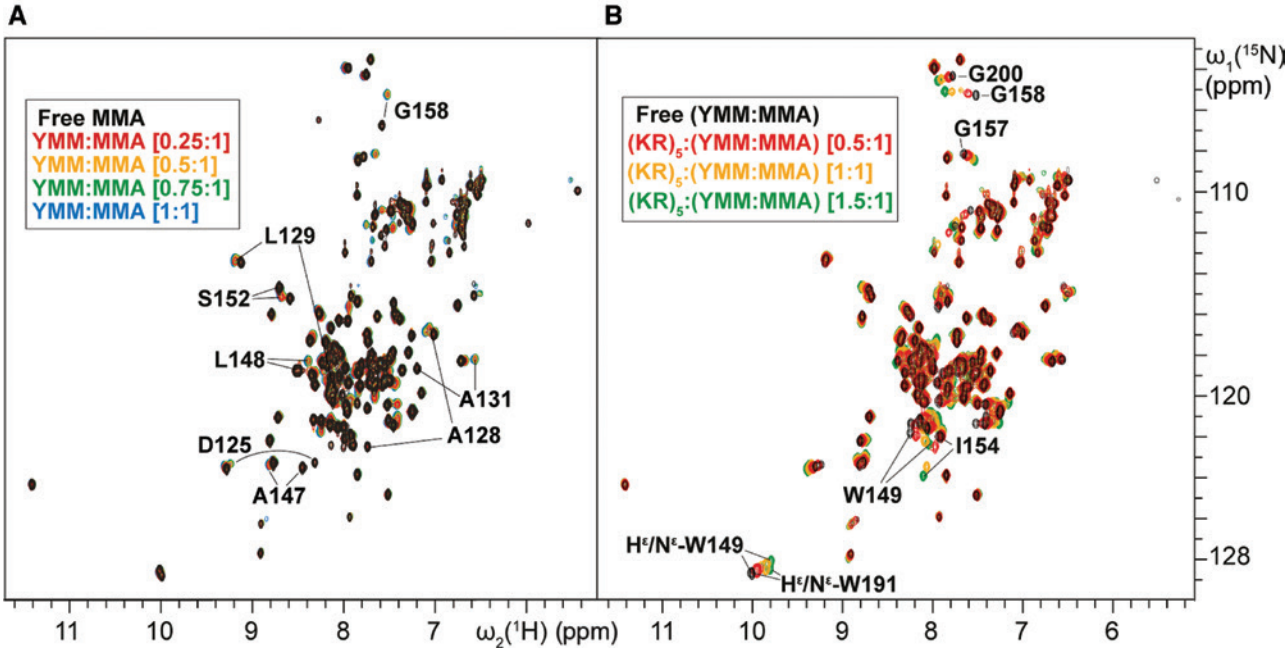

Figure 3: Chemical shift mapping of the YMM:MMA complementation and formation of the ternary complex with (KR)$_5$.
(A) Superposition of [^{15}N,^{1}H]-HSQC spectra of the titration of C-terminal [*u*-^{15}N]-MMA with unlabeled [*u*-^{14}N]-N-terminal YMM (left) and (B) of the [*u*-^{14}N]-YMM:[*u*-^{15}N]-MMA complex titrated with unlabeled (KR)$_5$ peptide. Peaks corresponding to individual titration steps are color-coded as indicated.

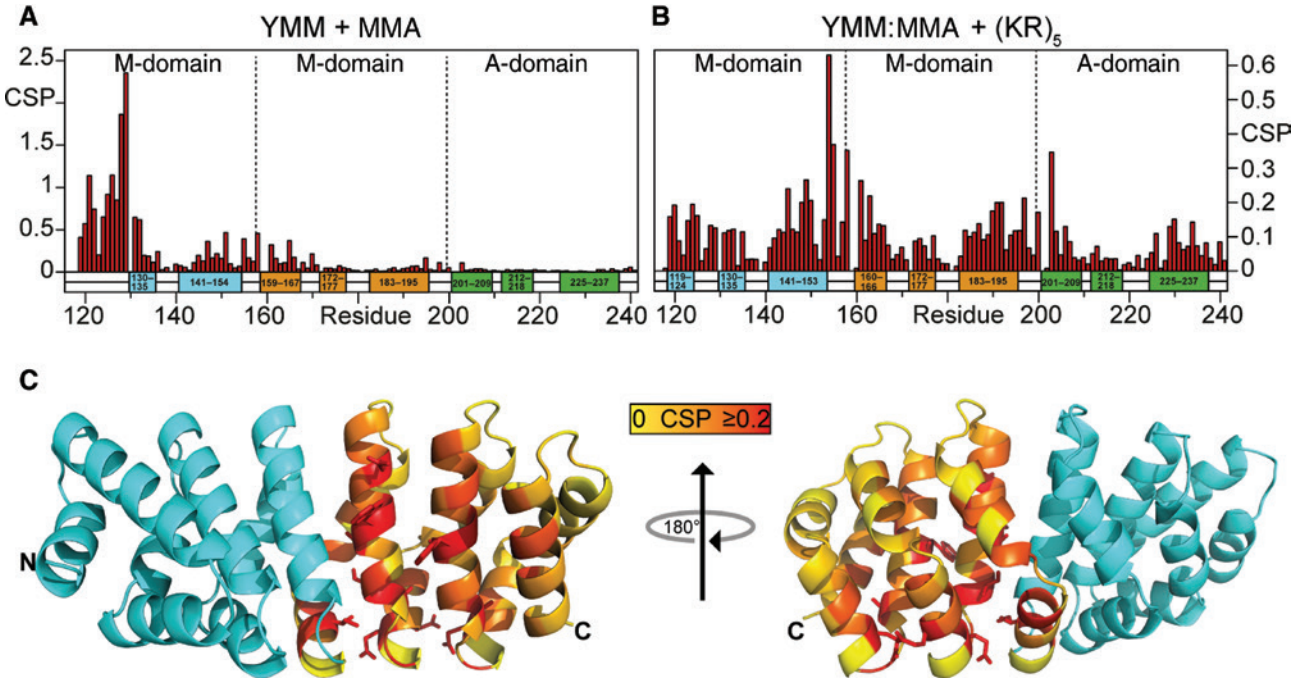

Figure 4: YMM:MMA complex formation and structural adaptations.
The residue-specific chemical shift perturbations (CSPs) of (A) the [YMM]:[MMA] and (B) the [YMM:MMA]:[$(KR)_5$] titrations are plotted against the sequence and secondary structure of MMA. (C) The C-terminal MMA modules of the YM_5A crystal structure (Hansen et al., 2016) are linearly color-coded from yellow (no CSP) to red (CSP ≥ 0.2) according to the CSPs of MMA experienced upon $(KR)_5$ peptide binding to YMM:MMA. The unlabeled YMM fragment is shown in blue.

performed a secondary chemical shift analysis to assess the type of secondary structure of free H3MA (Supplementary Figure 3A). In contrast to MMA, all helices observed in the full-length dArmRP also appear to be present in H3MA. Investigation of the backbone amide mobilities in H3MA using the heteronuclear NOE experiment further revealed a rigid backbone throughout the protein (Supplementary Figure 3B). We then followed the complementation of H3MA to YMMH12 with NMR by titrating unlabeled YMMH12 to ^{15}N-labeled H3MA (Supplementary Figure 3C). The complementation results in the formation of new distinct peaks with increasing intensity, reflecting slow exchange, and is completed at a molar ratio of 1:1, which is in agreement with a K_d in the nM-range (Michel et al., 2018). To conclude, the YMMH12:H3MA fragment pair behaves very similar to the fragments derived from the inter-modular split YMM:MMA.

$(KR)_5$ binding induces long-range conformational rearrangements in YMM:MMA

Binding of the ligand $(KR)_5$ to YMM:MMA occurs in fast exchange on the NMR timescale (Figure 3B), indicating that the interaction of the peptide with the protein complex is much weaker than the interaction of the two complementary protein fragments with each other. Furthermore, spectral changes are observed up to a ratio of ligand to YMM:MMA of 1.5:1, suggesting a K_d of YMM:MMA for $(KR)_5$ in the micromolar range. Interestingly, chemical shift perturbations (CSPs) of the MMA amide resonances after adding $(KR)_5$ are observed throughout the entire MMA fragment, including the capping module A (Figure 4). This indicates concerted structural adaptations of the entire protein upon peptide binding, which we currently further investigate.

A particularly useful feature of dArmRPs, due to the modular nature of the proteins, is that the affinity of the protein-peptide interaction can be adjusted by altering the number of (KR)-dipeptide repeats in the peptide and/or the number of internal modules M of the protein, as was shown for the un-split dArmRPs $Y(M)_xA$ and $(KR)_n$ peptides for the cases of x = 3–7 and n = 3–5 (Hansen et al., 2016). Binding of $(KR)_4$ and $(KR)_5$ to the YMMM:MMA complex, carrying one internal repeat more than the system described, revealed a shift from fast to the fast/intermediate exchange regime, saturation with 1.33 equivalents of $(KR)_5$ and slightly reduced CSPs when compared to the titration of YMM:MMA with $(KR)_5$ (Supplementary Figure 4). This shows that the tighter binding of $(KR)_5$ to

YM_5A than to YM_4A, previously described for the intact proteins (Hansen et al., 2016), is seen for the complexes assembled from fragments as well. When adding $(KR)_4$ to YMMM:MMA, the system is in fast exchange, saturated at ca. 1.66–2 equivalents of $(KR)_4$ and shows the smallest peak perturbations (Supplementary Figure S4). Due to the more favorable properties, we decided to continue with the YMM:MMA:$(KR)_5$ model system.

Design of dArmRP variants with different affinities towards (KR)-peptides

Our ultimate goal was to have a set of C-terminal fragments that all would form complexes with a single N-terminal fragment, but which would show different affinities to the peptide. We thus started with mutants which in the context of an uncleaved YM_5A protein would display different affinities for the peptide ligand $(KR)_5$. The design of these mutants was based on a crystal structure of YM_5A complexed to $(KR)_5$ that revealed that each arginine of the $(KR)_n$ peptide interacts with internal M modules through electrostatic interactions with Glu30 and π-cation interactions with Trp33, as well as by the peptide bond forming H-bonds to the conserved Asn37 (Figure 1B) (Hansen et al., 2016).

To obtain MMA variants with reduced affinity towards the target peptide, we first tested the effect of E30A and N37A mutations in full-length YM_5A. For this purpose, we introduced mutations into the fourth and fifth M module of YM_5A and determined the K_d for binding of a GFP-$(KR)_4$ fusion by fluorescence anisotropy. Mutation of both E30 and N37 to alanine in either M5 alone (referred to as [5-EN]-YM_5A) or in both M4 and M5 ([4,5-EN]-YM_5A) reduced the affinity towards $(KR)_4$ from a K_d of 45 nM for the wild-type interaction to 186 nM for [5-EN]-YM_5A and 4090 nM for [4,5-EN]-YM_5A. Based on these mutants, we prepared the corresponding C-terminal fragments [2-EN]-MMA and [1,2-EN]-MMA for the competition experiments.

Quantitative dArmRP fragment distribution analysis by NMR

To quantify the populations of each C-terminal MMA fragment in the free and assembled states, we prepared variants with a single [^{15}N]-labeled amino acid type. We chose to label Ala, Trp and Leu residues, because many of these residues form different contacts in the free and assembled states (Figure 3A). To introduce the labels, we used our home-made *Escherichia coli*-based cell-free expression system (Michel and Wüthrich, 2012a,b). Integration of characteristic signals provides a convenient means to simultaneously quantify the relative populations of free and bound species of the three C-terminal fragments in the mixture (Table 1). This quantification relies on integration of characteristic peaks of the free state of the MMA fragments, and derives the population of the bound state from the signal disappearance of the free population. This approach is feasible as the interaction of YMM with MMA is in slow exchange on the NMR timescale, which provides distinct peaks for the free and bound states, and as the characteristic peaks are well-separated in the NMR spectrum. Superposition of [^{15}N,^{1}H]-HSQC spectra, measured for each individual fragment and in a mixture containing equimolar amounts of all three C-terminal MMA variants (Supplementary Figure S5) confirmed that these fragments do not interact with each other.

Table 1: Population analysis for formation of YMM:MMA complexes upon addition of $(KR)_5$ in mixtures of YMM with three MMA variants as derived from integration of corresponding peaks in [^{15}N,^{1}H]-HSQC spectra.

Equivalents of $(KR)_5$ added	Population bound to one equiv. YMM (%)[a]		
	wt-MMA	[2-EN]-MMA	[1,2-EN]-MMA
0	5.1 ± 0.8	23.8 ± 6.1	71.1 ± 4.1
0.5	33.6 ± 5.7	12.6 ± 3.5	53.8 ± 8.3
1	55.0 ± 4.6	12.9 ± 2.9	32.1 ± 5.5
1.5	62.1 ± 4.4	14.5 ± 1.6	23.4 ± 2.9
2	67.9 ± 4.9	16.4 ± 1.7	15.7 ± 3.0
3	67.5 ± 1.6	17.1 ± 1.5	15.4 ± 7.8
4	70.7 ± 7.0	16.9 ± 0.8	12.4 ± 2.8
5	70.7 ± 3.3	16.9 ± 1.2	12.4 ± 1.7
10	69.6 ± 5.8	18.1 ± 4.1	12.3 ± 2.6

[a]Based on integration of [^{15}N,^{1}H]-HSQC signals of W149, W191, A128, A131, A147, A150, A155, A164, A168, L129, L132, L135, L136, L148 and L151.

We first investigated the distribution in the absence of peptide. The addition of one equivalent YMM to the equimolar mixture of [^{15}N-Trp]-labeled MMA, [^{15}N-Ala]-labeled [2-EN]-MMA and [^{15}N-Leu]-labeled [1,2-EN]-MMA induced mostly chemical shift perturbations for Leu residues (e.g. Leu136), significantly smaller changes for Ala residues (e.g. Ala147) and no detectable CSPs for Trp residues (e.g. Trp149) (Figure 5). This strongly suggested that YMM preferentially binds to the [^{15}N-Leu]-labeled [1,2-EN]-MMA fragment in the absence of a target peptide. Integration of NMR signals reveals that, at equilibrium, YMM binds approx. 71% [1,2-EN]-MMA, 24% [2-EN]-MMA and only 5% of the wt-MMA in the absence of peptide ligand (Table 1). As these differences were at first surprising, we confirmed them with ITC measurements with YMM complementing

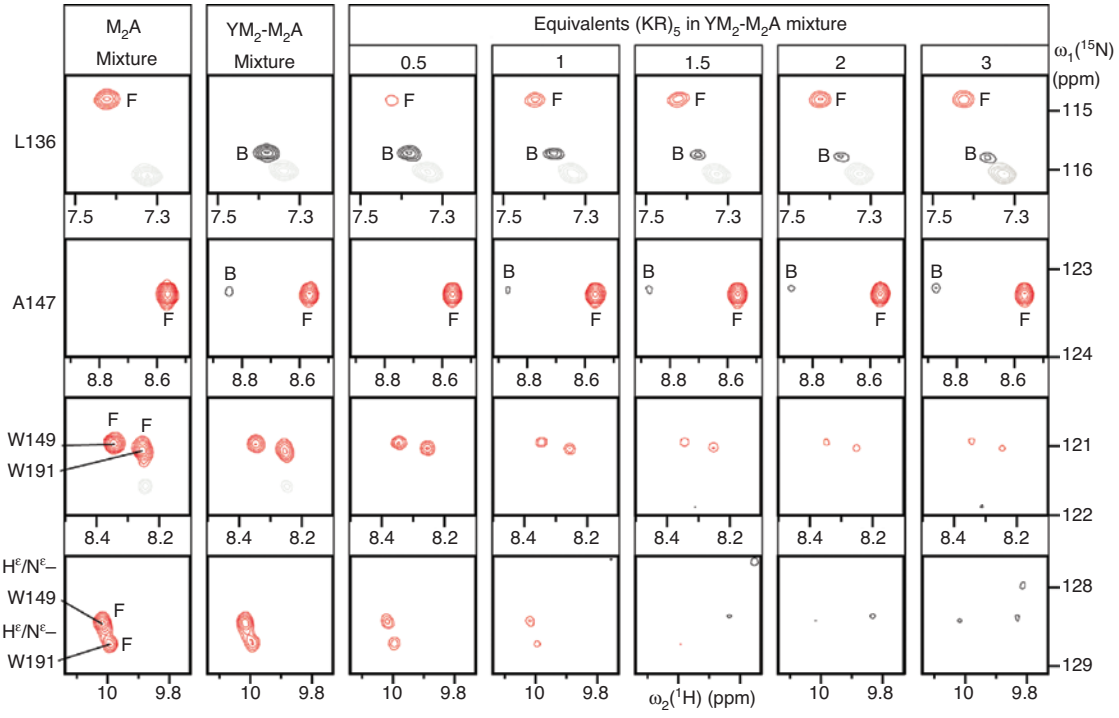

Figure 5: Peptide-induced re-distribution of pre-assembled YMM:MMA complexes.
NMR population analysis of an equimolar mixture of unlabeled YMM and three uniquely amino acid-type labeled MMA variants [^{15}N-Leu]-[1,2-EN]-MMA, [^{15}N-Ala]-[2-EN]-MMA and [^{15}N-Trp]-wt-MMA in the absence and presence of increasing amounts of the peptide $(KR)_5$. Shown are selected resonances of the [^{15}N,^{1}H]-HSQC spectrum that display a significant shift from the free (red contours) to the assembled state (black contours). Gray contours indicate resonances that are not part of the quantification. Characteristic peaks corresponding to the free and assembled states of MMA are labeled with an 'F' (free) or 'B' (bound), respectively.

either of the three C-terminal fragments. We obtained a K_d of 24 nM for the [1,2-EN]-MMA fragment interacting with YMM, 295 nM for the [2-EN]-MMA fragment and 593 nM for the wt-MMA fragment (Supplementary Figure 6), all in the absence of peptide.

We then investigated the effect of adding the target peptide. The stepwise addition of $(KR)_5$ gradually increased the population of free [^{15}N-Leu]-labeled [1,2-EN]-MMA fragment while decreasing the population of free [^{15}N-Trp]-labeled MMA, indicating that the target peptide enriches the combination of YMM with the wt-MMA, which represents the combination with the highest resulting affinity of the unsplit dArmRP for the target peptide (Figure 5). This population inversion is most clearly seen when adding between 0 and 2 equivalents of $(KR)_5$. At four equivalents of $(KR)_5$, 71% of YMM binds to wt-MMA, 17% to [2-EN]-MMA and 12% to [1,2-EN]-MMA (Table 1). These data demonstrate that the peptide-guided fragment re-distribution even overcomes the ca. 20-fold difference in K_d favoring the [1,2-EN]-MMA fragment in the absence of peptide. The experiments indicate that mutations at the fragment interface, such as the E-to-A and N-to-A mutations in the first module M, can significantly bias fragment assembly in the absence of target peptides, because they simultaneously influence the self-assembly of the protein complex and peptide binding.

Computational modeling rationalizes preferential binding of YMM to [1,2-EN]-MMA

To understand the preferred interaction of YMM with [1,2-EN]-MMA over [2-EN]-MMA or wt-MMA in the absence of a target peptide, we used the Rosetta software suite (Alford et al., 2017) to compute all-atom refinements of the structural models of YMM bound to either of the three C-terminal MMA variants. The resulting total Rosetta energy units (REU) of the three complexes reveal a clear distinction that correlates well with our experimental observations (Table 2). Dissection of the total REU into residue-specific contributions reveals particularly large energetic differences for Trp residues located within the two modules at the fragment interface, which gradually decrease in energy with each mutated module (Supplementary Table S1). This effect can also be seen within the full-length YM_4A variants but is completely absent in the

Table 2: Rosetta energy units (REU) difference between YMM:MMA assemblies and isolated YMM and MMA fragment variants.

Assembled fragments (REU)		Free fragments (REU)				Δ (Bound-ΣFree) (REU)
YMM:wt-MMA	−736.3	YMM	−326.1	wt-MMA	−380.8	−29.4
YMM:[2-EN]-MMA	−739.5	YMM	−326.1	[2-EN]-MMA	−374.3	−39.1
YMM:[1,2-EN]-MMA	−757.5	YMM	−326.1	[1,2-EN]-MMA	−382.4	−49.1

isolated MMA fragment variants (Supplementary Tables S2, S3), suggesting that packing of more than two internal modules enforces non-ideal side chain conformations for these residues. Structural analysis of the YMM:MMA complexes uncovers a rotation of the Trp side chain whenever the succeeding module contains the Glu-to-Ala mutation, which results in a hydrophobic contact between the Trp side chain and the newly introduced Ala (Supplementary Figure S7). This Trp side chain conformation is disfavored with succeeding wild-type modules due to a steric clash of the Trp and Glu side chains (Supplementary Figure S7). This re-orientation of the Trp side chain, which acts across the fragment interface, appears to be the main driving force for the preferential binding of YMM to [1,2-EN]-MMA. In the peptide complex, however, the favorable interactions between $(KR)_5$ and the dArmRP easily compensate for this effect and provide an explanation for the observed enrichment of the best binder upon addition of the target peptide (Supplementary Tables S4, S5).

Discussion

Many novel binding proteins have been generated from diverse protein libraries of very different architecture, and they mostly have in common that a randomized surface or a series of adjacent randomized loops is used for target binding in the context of a single contiguous protein domain. Different selected amino acids provide complementary contacts to the desired targets, while the overall fold of the scaffold protein is ideally preserved. While this overall strategy has been modeled on the immune system, the humoral immune system of most animals uses antibodies with two chains that both contain complementary regions to the target molecule. This two-chain strategy immensely increases diversification at the low cost of only doubling the number of genes, which was successfully exploited for antibody redesign with random combinatorial immunoglobulin libraries that achieve vast binder diversification by the shuffling of variable light and heavy chains (Kang et al., 1991). Inspired from nature's two-chain strategy, we placed the paratopes into two complementary protein fragments derived from an intermodular split of dArmRPs (Figure 1C) that form a high-affinity complex in the absence of a target peptide.

Here, we evaluate the dynamic behavior of the pre-assembled dArmRP fragments after addition of the peptide ligand. In particular, we analyze whether peptide addition alters populations of pre-assembled fragments such that the best binder for the target peptide is enriched. Furthermore, we evaluated the potential of the peptide to discriminate between pre-assembled fragment complexes that show relatively small differences in binding affinity to the peptide. To this end, we created a small model library represented by equimolar mixtures of one N-terminal fragment and three different C-terminal fragments with different affinities to the target peptide (Figure 1D). The selective labeling of each of the three different C-terminal fragments with one unique ^{15}N-labeled amino acid type allowed the simultaneous analysis of the assembly state of each variant in the mixture by NMR (Figures 3 and 4). In the initial YMM:MMA mixture of fragments in the absence of the target peptide, we observed a clear preference of the N-terminal YMM fragment to complement with the C-terminal [1,2-EN]-MMA fragment – a complex which shows the lowest affinity for the peptide (Figure 5, Table 1). This complementation bias was verified by an ITC interaction analysis yielding a K_d of 24 nM for the complex of YMM and [1,2-EN]-MMA (Supplementary Figure S6). We have used Rosetta modeling to show that this effect arises from an energetically favored re-orientation of the Trp sidechain when the Glu sidechain of the succeeding module is mutated to Ala. However, addition of the peptide ligand counteracted these effects and shifted the YMM-bound population from ca. 5% wt-MMA and ca. 71% [1,2-EN]-MMA in the absence of the peptide to ca. 70% wt-MMA and ca. 12% [1,2-EN]-MMA in its presence. This clearly demonstrates the discriminatory power of the target peptide but also suggests that a potentially increased enrichment is possible with complementary fragments that display no preference for each other in the absence of the peptide.

The employed dArmRP scaffold offers features that makes it also particularly suitable for the development of fragment-based selection libraries: (i) They interact with peptides in an extended conformation and provide a binding pocket for each amino acid of the bound peptide

(Hansen et al., 2016). (ii) They can be split in several ways into well-behaving complementary fragments, which assemble into a complex that structurally resembles the full-length protein and which retains the ability to bind target peptides (Watson et al., 2014; Michel et al., 2018). (iii) Mutations which affect the assembly of the complementary fragments can be overcompensated by the discriminatory power of the peptide, which ensures selection of fragments based on the affinity to the target peptide.

We conclude that splits of dArmRP between entire modules or between helices 2 and 3 result in fragments that form high-affinity interactions with one another, and that the resulting protein complexes all can bind peptide ligands developed against the unsplit dArmRP proteins, again underlining that they come together to rebuild the native armadillo protein (Watson et al., 2014). In the case of the intermodular split the putative N-terminal helix in the C-terminal fragment is absent but forms in the complex. Moreover, pre-assembled states of complementary dArmRP fragments can be efficiently re-shuffled by target peptides to enrich the best binder. The robustness and discriminatory power of the presented fragment-based recognition system could find promising applications in biochemistry, biotechnology and synthetic biology.

Materials and methods

The target genes were PCR-amplified from a custom-synthesized gene encoding YM3A (Michel et al., 2018) using the oligonucleotide primers indicated in Supplementary Table S6. The obtained PCR products were sub-cloned either into the cell-free expression vector pCFX3BT2 (Michel et al., 2018) or the *E. coli* expression vectors pEM3BT2 (Michel et al., 2018) and pEM3BT2-1D (prepared as described in the Supplementary Material). Unlabeled or uniformly labeled proteins were expressed in *E. coli* BL21 (DE3) cells (Stratagene, La Jolla, CA, USA) growing at 37°C in 500 ml LB or M9 medium (Michel and Allain, 2015), respectively. Protein expression was induced with 1 mM IPTG at an OD_{600} of ca. 0.6 for 16 h at 30°C. Amino acid-specific isotope labeling was achieved by cell-free protein expression using an *E. coli*-based S30 cell extract (Michel and Wüthrich, 2012a,b) according to a previously described protocol (Michel et al., 2013). All proteins were expressed as fusion constructs with a TEV protease-cleavable N-terminal $(His)_6$-GB1 domain (Michel and Wüthrich, 2012a,b) and were purified as described in the Supplementary Information. The purified protein samples devoid of the N-terminal $(His)_6$-GB1 domain were dialyzed against NMR buffer (20 mM sodium phosphate at pH 7.0, 50 mM sodium chloride, and 30 μM sodium azide). The N-terminally acetylated $(KR)_4$ and $(KR)_5$ peptides were prepared by solid phase peptide synthesis and were purified by C18-reverse phase high-performance liquid chromatography (RP-HPLC).

ITC measurements of the association between complementary dArmRP fragments and the interaction of assembled fragments with the GB1-$(KR)_5$-peptide fusion construct were conducted on a VP-ITC instrument (MicroCal Inc., Northampton, MA, USA) using 1.4 ml of 12.5 μM YMM in the sample cell and 220 μM MMA fragment in the syringe. Interaction of the peptide with the complemented YMM:MMA fragments were measured with 300 μM GB1-$(KR)_4$ in the syringe and either 12.5 or 25 μM of the complementary fragments in the sample cell. All protein solutions were extensively dialyzed against NMR buffer before analysis. The ITC experiments were performed at 25°C with stirring (300 rpm) and comprised 29 injections of 10 μl applied within 10 s, with a pause of 240 s between the individual injections. The obtained raw data was analyzed using the MicroCal Origin software (MicroCal Inc., Northampton, MA, USA). The interaction of wild-type and mutant full-length YM_5A dArmRP proteins with the GFP-$(KR)_4$ peptide was determined by fluorescence anisotropy as described previously (Hansen et al., 2016).

All NMR experiments were measured at 310 K with protein solutions in NMR buffer containing 6% (v/v) D_2O on Bruker Avance 600 and 700 spectrometers (Bruker, Fällanden, Switzerland) equipped with cryogenic triple-resonance probes. The backbone resonances of [^{13}C,^{15}N]-MMA and [^{13}C,^{15}N]-H3MA were assigned using the 3D HNCO, 3D HN(CA)CO, 3D HNCA, 3D HNCACB, 3D CBCA(CO)NH and 3D [^{15}N]-resolved [^{1}H,^{1}H]-NOESY experiments (Sattler et al., 1999). The secondary structure of the free and assembled MMA and H3MA fragments was determined by analysis of the C_α and C′ shifts following the chemical shift index protocol (Wishart and Sykes, 1994). The backbone amide mobilities of the free MMA and H3MA fragments were analyzed using two-dimensional (2D) ^{15}N{^{1}H}-NOE data recorded at 600 MHz (Noggle and Schirmer, 1971; Kay et al., 1989). Quantification of the free and assembled fragment populations by NMR spectroscopy was performed by signal integration of well-separated peaks in the 2D [^{15}N,^{1}H]-HSQC spectrum that are characteristic for the individual free C-terminal fragments as previously described (Michel et al., 2018). These characteristic peak groups comprised the amide resonances of L129, L132, L135, L136, L148 and L151 for the [1,2-EN]-MMA fragment, A128, A131, A147, A150, A155, A164 and A168 for the [2-EN]-MMA fragment and W149 and W191 for the wt-MMA fragment.

The Rosetta all-atom refinements of the structural models were performed using the Relax protocol in the Rosetta 3.9 release in combination with the beta_nov2016 scoring function. Each refinement comprised the calculation of 200 structures, and the 10 structures with the lowest Rosetta energy units were considered for the subsequent analysis.

Acknowledgments: We gratefully acknowledge financial support by a SINERGIA grant from the Swiss National Science Foundation to A.P and O.Z. (funder id: 10.13039/501100001711, grant no. 122686) and from the Research Council of the University of Zurich (grant no. K-73539-01-01) to E.M.

References

Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., Zerbe, O., Plückthun, A., and Caflisch, A. (2012). Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. Protein Sci. *21*, 1298–1314.

Alford, R.F., Leaver-Fay, A., Jeliazko, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.M., Kappel, K., et al. (2017). The Rosetta all-atom energy function for macromolecular modeling and design. J. Chem. Theory Comput. *13*, 3031–3038.

Binz, H.K., Amstutz, P., and Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. *23*, 1257–1268.

Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell. *94*, 193–204.

Gebauer, M. and Skerra, A. (2009). Engineered protein scaffolds as next-generation antibody therapeutics. Curr. Opin. Chem. Biol. *13*, 245–255.

Hansen, S., Tremmel, D., Madhurantakam, C., Reichen, C., Mittl, P.R., and Plückthun, A. (2016). Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. J. Am. Chem. Soc. *138*, 3526–3532.

Kang, A.S., Jones, T.M., and Burton, D.R. (1991). Antibody redesign by chain shuffling from random combinatorial immunoglobulin libraries. Proc. Natl. Acad. Sci. USA *88*, 11120–11123.

Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry. *28*, 8972–8979.

Kuriyan, J. and Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. Annu. Rev. Biophys. Biomol. Struct. *26*, 259–288.

McCafferty, J. and Schofield, D. (2015). Identification of optimal protein binders through the use of large genetically encoded display libraries. Curr. Opin. Chem. Biol. *26*, 16–24.

Michel, E. and Wüthrich, K. (2012a). High-yield *Escherichia coli*-based cell-free expression of human proteins. J. Biomol. NMR. *53*, 43–51.

Michel, E. and Wüthrich, K. (2012b). Cell-free expression of disulfide-containing eukaryotic proteins for structural biology. FEBS J. *279*, 3176–3184.

Michel, E. and Allain, F.H. (2015). Selective amino acid segmental labeling of multi-domain proteins. Methods Enzymol. *565*, 389–422.

Michel, E., Skrisovska, L., Wüthrich, K., and Allain, F.H. (2013). Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. ChemBioChem. *14*, 457–466.

Michel, E., Plückthun, A., and Zerbe, O. (2018). Peptide-guided assembly of repeat protein fragments. Angew. Chem. Int. Ed. *57*, 4576–4579.

Noggle, J.H. and Schirmer, R.E. (1971). The Nuclear Overhauser Effect: Chemical Applications (New York: Academic Press).

Palmer, A.G. (2015). Enzyme dynamics from NMR spectroscopy. Acc. Chem. Res. *48*, 457–465.

Parmeggiani, F., Pellarin, R., Larsen, A.P., Varadamsetty, G., Stumpp, M.T., Zerbe, O., Caflisch, A., and Plückthun, A. (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J. Mol. Biol. *376*, 1282–1304.

Plückthun, A. (2015). Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. Annu. Rev. Pharmacol. Toxicol. *55*, 489–511.

Reichen, C., Hansen, S., Forzani, C., Honegger, A., Fleishman, S.J., Zhou, T., Parmeggiani, F., Ernst, P., Madhurantakam, C., Ewald, C., et al. (2016). Computationally designed Armadillo repeat proteins for modular peptide recognition. J. Mol. Biol. *428*, 4467–4489.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. *34*, 93–158.

Ueda, H., Tsumoto, K., Kubota, K., Suzuki, E., Nagamune, T., Nishimura, H., Schueler, P.A., Winter, G., Kumagai, I., and Mohoney, W.C. (1996). Open sandwich ELISA: a novel immunoassay based on the interchain interaction of antibody variable region. Nat. Biotechnol. *14*, 1714-1718.

Watson, R.P., Christen, M.T., Ewald, C., Bumbak, F., Reichen, C., Mihajlovic, M., Schmidt, E., Güntert, P., Caflisch, A., Plückthun, A., et al. (2014). Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure. Structure *22*, 985–995.

Wishart, D.S. and Sykes, B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J. Biomol. NMR *4*, 171–180.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/hsz-2018-0355).