Curvature of designed armadillo repeat proteins allows modular peptide binding

Simon Hansen1,4, Patrick Ernst1, Sebastian L.B. König, Christian Reichen2, Christina Ewald3, Daniel Nettels, Peer R.E. Mittl, Benjamin Schuler, Andreas Plückthun*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

ARTICLE INFO

Keywords:
Armadillo repeat proteins
Peptide binding
Protein crystallography
Protein engineering
Single-molecule FRET

ABSTRACT

Designed armadillo repeat proteins (dArmRPs) were developed to create a modular peptide binding technology where each of the structural repeats binds two residues of the target peptide. An essential prerequisite for such a technology is a dArmRP geometry that matches the peptide bond length. To this end, we determined a large set (n = 27) of dArmRP X-ray structures, of which 12 were previously unpublished, to calculate curvature parameters that define their geometry. Our analysis shows that consensus dArmRPs exhibit curvatures close to the optimal range for modular peptide recognition. Binding of peptide ligands can induce a curvature within the desired range, as confirmed by single-molecule FRET experiments in solution. On the other hand, computationally designed ArmRPs, where side chains have been chosen with the intention to optimally fit into a geometrically optimized backbone, turned out to be more divergent in reality, and thus not suitable for continuous peptide binding. Furthermore, we show that the formation of a crystal lattice can induce small but significant deviations from the curvature adopted in solution, which can interfere with the evaluation of repeat protein scaffolds when high accuracy is required. This study corroborates the suitability of consensus dArmRPs as a scaffold for the development of modular peptide binders.

1. Introduction

Armadillo repeat proteins consist of a repetitive protein fold found in many proteins, most prominently in importin-α and β-catenin, which are involved in cellular processes including nuclear import, cell adhesion and signaling (Tewari et al., 2010). The name derives from the appearance of a Drosophila mutant with a defect in a segment polarity gene later identified as β-catenin (Nüsslein-Volhard and Wieschaus, 1980). Natural armadillo repeat proteins (nArmRPs) are abundant in the human genome and usually consist of 4-12 consecutive repeats (Gul et al., 2016) where the repeats at either terminus (called N- and C-cap, respectively) are more hydrophilic to protect the hydrophobic core from solvent exposure. Each repeat is usually made from 42 residues that fold into three triangularly arranged helices (H1, H2 and H3), and the angular stacking of several repeats creates a binding groove of adjacent H3 helices (Huber et al., 1997). The best described members of this protein family, importin-α and β-catenin, were shown to bind peptides in a stretched conformation in their binding groove (Conti et al., 1998; Graham et al., 2000). This stretched conformation of the bound peptide is enforced by bidentate hydrogen bonds between the backbone NH and O atoms of every second peptide bond and conserved asparagine residues on H3 (N37, superscripted numbers refer to residue numbering within one repeat), while the specificity is established by interactions with the peptide side chains. nArmRPs bind peptide ligands of up to six amino acid units in length in a modular fashion, i.e., two residues per repeat are recognized (Conti and Kurtjyan, 2000; Conti et al., 1998). Longer modular peptide binding has not been observed in ArmRP, because they do not possess a regular curvature that fits the binding register dictated by the peptide bond lengths (Reichen et al., 2016b). Designed armadillo repeat proteins (dArmRP) were developed to expand this modular peptide recognition by engineering a completely regular scaffold that fits the required

http://dx.doi.org/10.1016/j.jsb.2017.08.009
Received 20 April 2017; Received in revised form 15 July 2017; Accepted 28 August 2017
Available online 31 August 2017
1047-8477/ © 2017 Elsevier Inc. All rights reserved.
Fig. 1. Modular peptide binding. (a) Model of a continuous modular peptide binder made by propagation of internal repeats 1–5 of 5 MFL (chain B, blue) with the bound peptide (red). (b) Schematic illustration of two intertwined helices with parameters to describe the curvature.

register along the whole scaffold length and would therefore allow binding of, in theory, infinitely long peptides (Fig. 1a) (Alfarano et al., 2012; Madhurantakam et al., 2012; Parmeggiani et al., 2008). Based on this scaffold, surface-modified repeats might be developed that specifically bind a certain dipeptide in the context of a longer peptide, and most likely this can be achieved by directed evolution methods. Such preselected repeats could then in turn be assembled to create specific peptide binders against novel primary peptide sequences without additional selections (Reichen et al., 2014a).

The correct curvature of the scaffold is a fundamental prerequisite for the development of such a technology. The curvature of a repeat protein can be described by three parameters that define a superhelix which is formed by solenoid proteins: The rise (h), the angle (Ω), and the radius of the superhelix (r). For the parametrization of experimental structures, these values are usually calculated between the center of mass (CoM) of each repeat and the central axis of the superhelix (DiMaio et al., 2011; Kobe and Kajava, 2000; Park et al., 2015). A peptide in an extended conformation has a distance of 6.7–7.0 Å between the Cα atom of one residue and the Cα two residues further toward the C-terminus (termed Cα(P/P+2) distance). This distance was predicted by modeling of peptides in a relaxed β-strand conformation and was also found in peptide-bound structures by nArmRP and dArmRP (Reichen et al., 2016b). This distance varied only for short stretches. dArmRPs and bound peptides can be described as two intertwined helices with parameters to describe the curvature.

Previously, a computational design approach was described, where predefined curvature restraints were used to design this protein scaffold. These restraints were obtained from a model based on two repeats of yeast importin-α, which exhibited a Cα(P/P+2) distance within the optimal range for continuous peptide binding. Still, the obtained scaffold deviated from the intended curvature, and hence, solely allowed limited modular binding over approximately three consecutive repeats (Reichen et al., 2016b). On the other hand, a consensus design approach resulted in high-affinity binding to peptides with repetitive arginine-lysine sequence ((KR)n) that showed in-solution binding behavior that is consistent with modular binding (Hansen et al., 2016). However, these results still do not prove an optimal curvature of the scaffold since the peptides used were at most ten residues in length. For even longer peptides, small deviations of the scaffold might add up and at some point prevent modular peptide binding.

Here, we analyze 27 experimental X-ray structures of consensus and computational dArmRPs with a focus on their curvature parameters. This analysis allows us to draw conclusions regarding which design approach provides a suitable scaffold for modular peptide binding. The analysis also revealed that, due to the intrinsic flexibility of dArmRP, packing into a crystal lattice can introduce geometric differences between structures of dArmRP that also influence their peptide binding mode. This hampers the characterization of designed scaffolds by crystallography alone, which was therefore supplemented by in-solution FRET measurements.

2. Material and methods

2.1. Cloning

Cloning of the dArmRP genes has been described previously (Madhurantakam et al., 2012; Reichen et al., 2016b). Direct fusions of (KR)-peptides to the dArmRP were introduced by PCR using oligonucleotides encoding the (KR)-peptide, a GS-linker and an overlap with either the N- or C-terminus of dArmRP. All sequences of the different types of dArmRPs used in this study are shown in Fig. 1. Constructs were subcloned with BamHI and HindIII restriction enzymes (FastDigest enzymes, Fermentas) into the vector pQE30LIC_3C, which contains a 3C-protease-cleavable MRGSHis6-tag (Hansen et al., 2016). All constructs were confirmed by DNA sequencing (Microsynth), and glycerol stocks of the respective plasmids were prepared (overnight culture of transformed E. coli XL1 Blue or BL21 (DE3) with 15% glycerol, stored at −80 °C).

2.2. Protein expression and purification

Protein expression was carried out in 1 l of 2xYT medium (containing 100 µg/l ampicillin and 0.5% glucose). Overnight cultures (25 ml) were inoculated from glycerol stocks, grown at 37 °C and used for inoculation. The expression culture was grown to an OD₆₀₀ of 0.7, followed by induction with 750 µM of isopropyl-β-D-thiogalactopyranoside (IPTG) and incubation at 37 °C for 5 h. Cells were centrifuged (5,000 g, 5 min), resuspended in 25 ml TBS_W (50 mM Tris pH 8.0, 400 mM NaCl, 20 mM imidazole) and flash-frozen (liquid N₂, stored at −80 °C). Resuspended cells were thawed and lysed by sonication and passed through a French press system. Insoluble contents were removed by centrifugation (25,000 g, 20 min). The crude extracts were loaded onto Ni-NTA superflow resin columns (3 ml, Qiagen), 30 column volumes (CV) of TBS_W were applied for washing. Elution was achieved by 2.5 CV of TBS_E (TBS_W with 300 mM imidazole). Human rhinovirus 3C-protease (2% w/w) was mixed with eluted protein fractions and the reaction mixtures were dialyzed against 50 mM Tris pH 7.4 and 300 mM NaCl to remove the His₆₅-tags. Protease and uncleaved proteins were trapped by IMAC while cleaved protein was collected in the flow-through. Nonmonomeric protein fractions were isolated by SEC on an Äkta explorer chromatography system using a HiLoad 16/60 Superdex 200 pg column and 10 mM Tris, pH 7.4 with 100 mM NaCl as running buffer. Proteins were concentrated (Amicon Ultra Centrifugal Filters, Merck Millipore) to a final concentration of approximately 20 mg/ml. Peptides (LifeTein) were added with a 1.5–1.8 molar excess over protein. Protein preparations used to determine structures 5MFE, 5MFG, 5MFP and 5MFN were supplemented with 200 mM CaCl₂ to enhance
the chance of crystal hits by enforcing Ca\(^{2+}\)-bridged dArmRP dimers (see below).

### 2.3. Crystallization and structure determination

Screening for initial conditions was done with commercial 96-well sparse-matrix screens (Hampton Research and Molecular Dimensions) at 4 °C in a sitting-drop vapor diffusion set-up. For each condition the reservoir solution was mixed in three ratios with protein solution (1:1, 1:2 and 2:1 or 5:1). Based on the conditions identified in the initial screens, optimization screens (perpendicular pH- and precipitant gradients in 96-well plates) were applied to improve crystal quality. Supplementary Table ST1 summarizes the crystallization conditions as well as data collection and refinement statistics. Crystals were flash-frozen (liquid N\(_2\)) in mother liquor supplemented with 5–20% glycerol or ethylene glycol. Diffraction data were collected on beam lines X06DA and X06SA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a Pilatus detector system (Dectris Ltd). Data were processed using the XDS, XSCALE and XDSCONV software packages (Kabsch, 2010).

PHASER (McCoy et al., 2007) was used for molecular replacement to obtain initial phases, and previously determined structures of dArmRP were used as search models. Refinement was done using the programs REFMAC5 (Murshudov et al., 1999, 2011), BUSTER (Brünger et al., 2016) and Phenix-Refine (Afonine et al., 2010a,b), followed by model building in COOT (Emsley and Cowtan, 2004; Emsley et al., 2010).

### 2.4. Curvature analysis

Coordinates of C\(^{\alpha}\) atoms of neighboring repeat pairs (excluding the flexible loop residue at position 23) were extracted from all experimental structures and used as an input into the make-symmdef_file.pl script from the Rosetta symmetry framework (DiMaio et al., 2011), which calculated the corresponding curvature parameters and generated models with uniform curvature of the input repeats (symm.pdb-files). A model of an armadillo repeat with a bound (KR)-dipeptide was made by averaging the C\(^{\alpha}\)-coordinates of 53 repeats from structures 5AEI, 5MFL and 5MFM that showed bidentate hydrogen bonds shorter than 3.1 Å from the N\(^{37}\) side chain to the peptide bond in the peptide backbone. This model was superimposed onto two consecutive repeats of symm.pdb-files, and the C\(^{\alpha}\)(P/P+2) distance of the bound peptide residues was measured. We analyzed whether copies of dArmRP from the same AU exhibit the same curvature parameters by plotting C\(^{\alpha}\)(P/P+2) distances of repeat pairs of all chains within the same AU followed by manual inspection. To remove bias toward structures with many chains in the AU, all structures were classified and the curvature parameters were averaged accordingly. The following classification was used: (i) single chain in AU (no averaging) (ii) one population in AU, i.e., all molecules are highly similar (averaging among the same repeat pairs of all chains in the AU) and (iii) more than one population in AU (averaging only among the same repeat pairs of chains of the same population). The grouping of all structures and the plots used to decide upon the category are shown in Supplementary Figs. S1, S2, S3 and S4, and are summarized in Supplementary Table ST2. In some cases when the classification was not obvious, additional criteria were considered (e.g., dArmRP chains that form Ca\(^{2+}\)-bridged dimers in the AU often display similar curvatures and were therefore considered as one population; if crystal contacts differ largely between chains (PDB ID: 5MFI) they were treated as different populations, even if they show similar parameters; or if pairs of chains are practically identical, while there is a small difference between the pairs, they were treated as two populations, PDB ID: 4DBA and 4VMF). If the classification was still doubtful (PDB ID: 4V3Q and 4V3O), chains were considered as one population; for

---

**Fig. 2.** Sequence alignment of all dArmRP constructs used in this study. The recognition site of 3C-protease is shown in yellow, the scissile bond is indicated by arrows. Positions of helices H1, H2 and H3 are indicated by green, blue and red rectangles, respectively. Repeats Y\(_{\text{II}}\), Y\(_{\text{III}}\), M\(_{\text{I}}\), M\(_{\text{II}}\), A\(_{\text{II}}\) and A\(_{\text{III}}\) are consensus-based; repeats D\(_{\text{q}}\), D\(_{\text{q}}\).V1, D\(_{\text{q}}\).V2, CPAF and C\(_{\text{qI}}\) originate from the computational design approach. Sequence differences within consensus or computational designs are highlighted in red, differences between consensus and computational designs are shown by a grey background.
2.5. Single-molecule FRET experiments

For fluorescence labeling of YmM₄A₄ with a donor–acceptor dye pair, amino acid exchanges D51C and S273C were introduced by site-directed mutagenesis. The protein was expressed and purified by IMAC as described above, but 1 mM β-mercaptoethanol (Pierce) was added to all buffers to keep the Cys residues reduced. SEC was conducted in PBS containing 1 mM DTT, followed by anion exchange chromatography in 20 mM NaH₂PO₄/Na₂HPO₄, pH 7.2 on a MonoQ 5/50 GL column. For elution, a gradient from 0 to 1000 mM NaCl was applied. For labeling, maleimide derivatives of Alexa 488 and 594 were used. These commercially available fluorophores contain a C5 linker, which ensures rotational averaging of the dye orientations, as well as optimal quantum yields. 1–2 mg of protein were incubated with 0.7 molar equivalents of Alexa Fluor-488 (ThermoFisher Scientific) at 4 °C for 16 h in a total volume of 2 mL. Singly labeled YmM₄A₄ was separated from unlabeled and doubly labeled protein by anion exchange chromatography and incubated with a 1.5-fold excess of Alexa Fluor-594 at 4 °C for 16 h. Finally, donor–acceptor labeled protein was purified by anion exchange chromatography, flash-frozen and stored at −80 °C until further use.

Fluorescently labeled protein was observed on a confocal single-molecule instrument (MicroTime 200, PicoQuant GmbH) at 22 °C as described before (Aznauryan et al., 2013). Dyes were excited using pulsed interleaved excitation (PIE) (Müller et al., 2005) at a pulse repetition rate of 20 MHz. Light of a supercontinuum laser (SC-450, Fianium Ltd., Southampton, UK) selected with a z825/15 band pass filter (Chroma Technology) was used to excite Alexa Fluor-594, and Alexa Fluor-488 was excited using an LDH-D-C-485 diode laser (PicoQuant GmbH). Fluorescence photons were collected by the objective, passed through a chromatic beam splitter (R405/488/594 Chroma Technology), focused onto a 100 μm pinhole, and separated according to polarization and wavelength using a polarizing beam splitter and two 585DCXR dichroic mirrors (Chroma Technology). Finally, photons were filtered (ET525/50 M or HQ650/100 band-pass filters, Chroma-Technology) and detected by four avalanche photodiodes (SPCM-AQR-15, PerkinElmer Inc., Wellesley, MA). All measurements were conducted in phosphate-buffered saline (PBS) supplemented with 0.001% (w/v) Tween-20 (Pierce) to minimize surface adhesion and 143 mM β-mercaptoethanol (Sigma-Aldrich) for photoprotection at a protein concentration of approximately 50 pM. The total acquisition time was 60 min per measurement.

Photon recordings were binned into 1 ms intervals, and fluorescence bursts caused by labeled proteins diffusing through the confocal volume were identified using a threshold of 30 photons per bin. Photon counts were corrected for background, differences in donor/acceptor quantum yields, differences in detection efficiencies, spectral crosstalk, as well as direct acceptor excitation (Schuler, 2007). For each photon burst, the stoichiometry ratio, S, and the transfer efficiency, E, were calculated as $S = \frac{N_{\text{tot},\text{Dex}}}{N_{\text{tot},\text{Dex}} + N_{\text{tot},\text{Aex}}}$ and $E = \frac{N_{\text{tot},\text{Dex}}}{N_{\text{tot},\text{Dex}} + N_{\text{tot},\text{Aex}}}$, respectively. Here, $N_{\text{tot},\text{Dex}}$ and $N_{\text{tot},\text{Aex}}$ correspond to the corrected total numbers of photons emitted following donor or acceptor excitation, respectively, and $N_{\text{tot},\text{Dex}}$ and $N_{\text{tot},\text{Aex}}$ correspond to the corrected donor and acceptor photon counts following donor excitation (Müller et al., 2005). Bursts with S < 0.6 were used to generate transfer efficiency histograms. Probability distribution analysis (PDA) was used to analyze transfer efficiency histograms (Antonik et al., 2006). PDA assumes an intrinsic distribution of experimental transfer efficiencies, which is further broadened by shot noise. Shot-noise broadening is modeled using the photon statistics of the experimental burst-size distribution. Here we applied PDA assuming a very narrow intrinsic transfer efficiency distribution to demonstrate that the width of the histogram peaks is dominated by shot-noise. Geometrically accessible volumes of dyes were calculated for chain B of 5AEI and chain C of 5MFC, which served as model structures for the ligand-bound and ligand-free state, respectively. All calculations were performed using the freely available “FRET Positioning and Screening” software, in which Alexa 488 and 594 were parameterized as described earlier (Kalinin et al., 2012). To take into account the rapid fluctuations in interdye distance that occur on timescales similar to the excited state lifetime of the donor, we modeled the dynamics as diffusive motion in the potential of mean force corresponding to the interdye distance distribution from the accessible volume calculations using the diffusion coefficient of the free dyes (Best et al., 2015).

3. Results

3.1. Crystal structures serving as the basis of the curvature analysis

Here we are analyzing the curvature of all previously published structures of dArmRP (n = 13), as well as another 12 previously unpublished structures. We expect that this large dataset will allow us to estimate the possible curvatures that can be adopted by dArmRP better than single structures that are possibly influenced by crystal packing. Overall, these 27 structures represent two different design approaches (consensus (Alfarano et al., 2012; Madhuranatkum et al., 2012; Parmeggiani et al., 2008) and computational (Reichen et al., 2016b)), cover dArmRPs from 3 to 6 internal repeats, represent several different capping and internal repeat types, and are either apo structures or complexed with different peptide ligands. The structures were obtained in various space groups with 1–8 dArmRP molecules per asymmetric unit (AU) and resolutions ranging from 1.3–2.8 Å. An overview of all structures is given in Table 1, and the primary dArmRP sequences are shown in Fig. 2. Detailed discussions of the previously published structures are provided in the original publications (Table 1). New structures will not be discussed in-depth, although interesting features will be highlighted.

A recurring arrangement in the crystal structures of consensus dArmRP is the formation of “Ca²⁺-bridged dimers”, where Ca²⁺ ions are complexed between two repeats of neighboring chains, thereby forming an antiparallel dimer (Hansen et al., 2017; Reichen et al., 2016a). This motif is found in roughly half of the consensus dArmRP structures but never in computational dArmRPs (Table 1). Different arrangements of Ca²⁺-bridged dimers exist depending on which repeats of two dArmRP are paired by complexing the same Ca²⁺ ion. Most often a shift of 1 (i.e., one repeat) is observed (Fig. 3a, in structures 4V3Q, 4PLS, 5AEI, 5MFE, 5MFG, 5M5H, 5MFL and 5MFO). A shift of 0 is found in structures 5MFD and 5MFN while a shift of 2 is found in structures 4V3O and 5MFN.

5MFF forms tight parallel dArmRP dimers that are not mediated by Ca²⁺ ions; the stoichiometry with the (RR)₅ peptide is one peptide per two dArmRP. This is a consequence of the tight packing of the dArmRP dimers in this lattice, since it is sterically impossible to fit two peptides into the small channel between the two dArmRPs. In size exclusion chromatography experiments with a multi-angle light scattering detector (SEC-MALS), dimer formation by addition of (RR)₅ was not detected (data not shown). Hence we believe the 2:1 stoichiometry is a crystallization artefact and not present in solution. In this conformation the peptide side chains are involved in cation-π stacking interactions with W₉₉ and salt bridges with E₃₀ from both dArmRP chains of the dimer (Fig. 3b).

5MFG and 4PLS form a similar dArmRP tetramer; domain swapping events of the capping repeats connect two dArmRP such that the exposed internal repeats stack together to form one apparent dArmRP.
with ten internal repeats. The tetramer consists of two such dimers connected by Ca\(^{2+}\) ions (Reichen et al., 2014b) (Fig. 3c). This domain swapping event is different from the one observed with the Y\(_{II}\) cap, since most probably no completely homogenous complexes are formed throughout the crystals. In 5MFG a peptide ligand is added to the crystallization mixture, ten residues could be modeled within a full-length dArmRP and the faint additional density was modeled by an additional C-cap. Also for the peptide ligands of 5MFH additional electron density is observed in most chains. Hence peptides were modeled as two chains of alternative conformations that are shifted by one or two (KR) repeats relative to each other with a total occupancy of 1. Here the electron density clearly indicates the occurrence of several binding registers of the peptide along the dArmRP binding groove. Such multiple binding modes were proposed to contribute to the high affinity of dArmRPs (Hansen et al., 2016).

### 3.2. Curvature analysis

Curvature parameters were calculated from neighboring armadillo repeat pairs of all structures listed in Table 1 (see Material and Methods, Fig. 1). A total of 329 repeat pairs was analyzed. All raw data of this analysis is supplied in the Supplementary raw data table.

Analysis of dArmRPs from the same AU showed that in some structures with many chains per AU while retaining the information of Ca\(^{2+}\)-bridged dimers, up to 4 populations with unique curvature parameters are found per dArmRPs per AU (Supplementary Figs. 3 and 4 and Supplementary Table ST2).

We then generated a normalized dataset to remove bias towards structures with many chains per AU while retaining the information of chains that adopt different conformations within the same AU. This was achieved by averaging parameters of dArmRP from the same AU.
exhibiting identical curvatures in a repeat pair-wise fashion (see Material and Methods and Supplementary Material for further details) and thus counting them only once. After averaging, the curated data set consisted of 151 sets of curvature parameters (Supplementary raw data table).

The mean rise in this dataset is 7.34 ± 0.72 Å (mean value ± standard deviation) with minimal and maximal values of 5.95 and 10.45 Å, respectively. The radius rA (to the CoM) has a mean value of 15.26 ± 2.25 Å (7.69 – 20.25 Å), while the angle 2Ω adopts values ranging from 23.0° to 32.0° with a mean of 27.7 ± 1.8°. These parameters are not independent of each other: the angle-radius distribution of superhelical repeat proteins follows a hyperbolic shape (Brunette et al., 2015). In our dataset 2Ω angles plotted as a function of radii show a curve with a negative slope. This is in agreement with the hyperbolic distribution since only a small angle-range is sampled and the negative slope is consistent with the fact that dArmRPs form right-handed super helices. The rise is not directly dependent on the other parameters. For α-helical repeat proteins, values below 16 Å are adopted, since a larger rise does not allow the formation of a continuous hydrophobic core (Brunette et al., 2015). The most informative parameter is the Cα(P/P+2) distance, as it directly assesses whether modular peptide binding over long stretches is possible with the respective curvature.

The curvature parameters of the consensus and computational dArmRPs can be clearly distinguished; the computational design is characterized by a significantly higher rise and angle but a smaller radius (h = 7.95 ± 0.55 Å, 2Ω = 30.0 ± 1.1°, rA = 12.44 ± 1.58 Å). The parameters of the consensus dArmRP show some pronounced outliers that can be explained by domain swapping events in the parental structures. Repeat pairs directly adjacent to domain-swapped YII-caps (in structures 4DBA and 4DB8) entail a pronounced enlargement of rise and Cα(P/P+2) values. On the other hand, the domain swap of the AII-cap in structure 5MFG (chains A and D) leads to a very small rise and Cα(P/P+2) distance in adjacent repeat pairs (Supplementary Figs. S1 and S3). Since we consider that domain-swaps are crystallization artefacts, the respective curvature parameters from those structures were removed from all following analyses (dataset: cons-dds). In this dataset 2Ω angles plotted as a function of radii show a curve with a negative slope. This is in agreement with the hyperbolic distribution since only a small angle-range is sampled and the negative slope is consistent with the fact that dArmRPs form right-handed super helices. The rise is not directly dependent on the other parameters. For α-helical repeat proteins, values below 16 Å are adopted, since a larger rise does not allow the formation of a continuous hydrophobic core (Brunette et al., 2015). The most informative parameter is the Cα(P/P+2) distance, as it directly assesses whether modular peptide binding over long stretches is possible with the respective curvature.

To investigate whether peptide binding can induce a more optimal
curvature, we divided the cons-dds and computational dataset into four sub-datasets based on the ligand binding properties of the respective structure. Henceforth the sub-datasets are referred to as canonical (bidentate hydrogen bonds with N37 residues are observed along the whole dArmRP), partial canonical (bidentate bonds present in some instances), non-canonical (peptide does not interact with N37 residues and no 1:1 stoichiometry is present) and no ligand (no peptide ligand present in the structure). The grouping of the structures is given in Table 1. In consensus dArmRP, canonically bound peptides induce a significantly smaller C\(\alpha\)(P/P+2) value compared to partial canonical and structures without ligand. Structures with non-canonical binding exhibit C\(\alpha\)(P/P+2) values similar to the canonical structures. These data are derived from the structures 5MFF and 5MFG, in which the crystal lattices make it sterically impossible (5MFF) or energetically unfavorable (5MFG) to form canonical complexes, even though the curvature parameters appear to be suitable (see above and Fig. 5a).

The formation of Ca\(^{2+}\)-bridged dimers leads to significantly lower C\(\alpha\)(P/P+2) values (7.27 ± 0.35 Å in Ca\(^{2+}\)-bridged dimers, 7.56 ± 0.51 Å without Ca\(^{2+}\)-bridged dimers). However, the entire canonical sub-dataset falls into the group with Ca\(^{2+}\)-bridged dimers. If the values of the canonical sub-dataset are removed, the difference between the groups is not significant anymore (7.34 ± 0.37 Å with peptide binding seems to reduce protein-to-protein variation in curvature parameters, as the corresponding standard deviations are smaller compared to the ones of the partial canonical and no ligand sub-datasets.

For the computationally designed dArmRPs only structures with partial canonical binding or no ligand could be determined. Their C\(\alpha\)(P/P+2) distances adopt values of 7.96 ± 0.48 Å and 8.15 ± 0.77 Å for the partial canonical and the no ligand sub-dataset, respectively. Not only are these values very similar, they are also considerably greater than the optimal range (Fig. 5b).

To investigate whether capping repeats influence the curvature of neighboring internal repeats, we grouped C\(\alpha\)(P/P+2) distances into N-cap neighboring, central and C-cap neighboring repeat pairs. For both design approaches, repeat pairs adjacent to caps display elevated C\(\alpha\)(P/P+2) values. But only for N-cap neighboring repeats of consensus dArmRPs the C\(\alpha\)(P/P+2) distances are significantly larger than the C\(\alpha\)(P/P+2) values of central repeats (7.59 ± 0.40 Å and 7.21 ± 0.40 Å, respectively; Fig. 5c and d).
Ca^{2+}\text{-bridged dimers, }7.56 \pm 0.51 \text{ Å without Ca}^{2+}\text{-bridged dimers). Thus, Ca}^{2+}\text{-bridged dimer formation might help with the adaptation of consensus dArmRP to the peptide by regularizing the arrangement, but the main contribution stems from the interaction with the peptide.}

### 3.3. Single-molecule FRET experiments

As the analysis of all crystal structures revealed, the curvature of dArmRPs can be influenced by the specific arrangement of the crystal lattice. We were interested in the behavior in solution, which is the relevant environment for the final applications of dArmRP. Hence we wanted to probe the presence of conformational changes upon peptide binding unbiased by the context of the protein crystal. Therefore, we performed single-molecule Förster resonance energy transfer (FRET) spectroscopy experiments of YmM5A5 in solution. Single-molecule FRET relies on the distance-dependent energy transfer between a donor and an acceptor fluorophore (Selvin and Ha, 2008). For this purpose, cysteine residues were introduced at positions 51 and 273 in the consensus dArmRP sequence, followed by sequential coupling of mal- eimide-functionalized Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor). Confocal single-molecule FRET experiments were performed at a protein concentration of approximately 50 pM in the presence and absence of 1 μM unlabeled (KR)5, respectively. Individual FRET-labeled YmM5A5 molecules diffusing through the confocal volume were detected as bursts of fluorescence photons. Transfer efficiencies were calculated for each burst and binned in histograms. Since single molecules are measured, single-molecule FRET enables us to (i) resolve subpopulations that would otherwise be hidden in the ensemble average, and (ii) compare the width of transfer efficiency distributions to the value theoretically expected from photon shot noise.

The observed transfer efficiency histograms are depicted in Fig. 6a. In both cases, a single, well-defined peak is observed. Note that the widths of the peaks are dominated by the stochastic fluctuation of recorded numbers of donor and acceptor photons (shot noise). This is illustrated by calculating the theoretically expected shot noise broadening using probability distribution analysis (PDA) (Fig. 6a, black lines) (Antonik et al., 2006). This observation indicates that the distance fluctuations between donor and acceptor are faster than the burst duration, and only mean transfer efficiencies, \( \langle E \rangle \), can be obtained from the histograms. We found an increase from \( \langle E \rangle_{\text{exp}} = 0.21 \) in the absence to \( \langle E \rangle_{\text{exp}} = 0.31 \) in the presence of (KR)5, indicating a decrease in the average inter-dye distance.

To directly compare single-molecule FRET experiments and crystallographic results, we estimated the expected mean transfer efficiencies based on the atomic-resolution structures. Here, chain B of 5AEI served as a model for bound dArmRP, and chain C of 5MFC as a model of unbound dArmRP (Fig. 6b). Since the linkers used to conjugate the dyes to the proteins are long and flexible, a relatively large range of inter-dye distances can be sampled. To calculate the expected mean transfer efficiencies, we determined the distribution of inter-dye distances from the geometrically accessible volumes of both dyes (Fig. 6b) (Kalinin et al., 2012) and took into account the dynamics of inter-dye distance fluctuations (see Methods). This analysis yields an expected change in transfer efficiency from approximately 0.23 to 0.28 upon peptide binding, very similar to our experimentally determined values. In summary, the single-molecule FRET experiments thus confirm a small but significant distance change that occurs on the subnanometer scale upon canonical peptide-binding to dArmRPs.

### 4. Discussion

In this study, we show that dArmRPs adopt different curvature parameters in different crystal structures, sometimes even within the same AU. This demonstrates an intrinsic flexibility of the scaffold which was also observed for nArmRPs (Huber et al., 1997) and importin-β, belonging to the closely related family of HEAT repeat proteins (Halder et al., 2015; Tauchert et al., 2016; Zachariae and Grubmüller, 2008). However, we expect the flexibility to be smaller in dArmRPs since the scaffold is more regular and stable compared to nArmRP (Madhuranthakam et al., 2012; Parmeggiani et al., 2008; Reichen et al., 2016b).

The packing of the crystal lattice plays a major role for the observed curvature but it is very difficult to predict the specific influence of the lattice on the curvature of the protein, since multiple crystal contacts have to be taken into account that are hard to quantify and to compare between different lattices. Moreover, the differences among all structures are still small. The superposition of Cα’s of chain B of 5AEI (consensus dArmRP with the lowest average Cα(P/P+2) value of 6.93 Å) and chain A of 5MFC (consensus dArmRP with the highest average Cα(P/P+2) value of 7.94 Å) results in a RMSD of 1.49 Å. Thus, for most scaffold design projects such differences would not be considered meaningful. However, for the purpose of designing a modular peptide-binding scaffold, a value close to the optimum is important for long-range peptide binding. In solution, a range of curvatures around an average state will be adopted while in a crystal structure one or few possible states that are locked in the crystal lattice are observed. Hence, it is not possible to characterize the curvature that is adapted in solution by a single crystal structure. However, we deem that the large set of structures used herein samples a sufficiently large conformational space in order to estimate the average curvature in solution.

Our analysis shows that the two design approaches - consensus and computational – can be distinguished by all the parameters used to describe the curvature. It is remarkable that the consensus dArmRP exhibits a curvature that is better suited to bind peptides continuously than the computational dArmRP, since these were explicitly designed by using curvature restraints, while no such restraints were used during the consensus design. It should be noted, however, that several mutations had to be introduced to the computational dArmRPs in order to obtain diffracting crystals. Hence, it remains unclear whether the original computational designs would adopt the desired curvature (Reichen et al., 2016b) as they could not be crystallized.

Consensus dArmRP have a Cα(P/P+2) value of 7.35 ± 0.42 Å which is slightly above the optimal value (6.7–7.0 Å). However, peptide binding can induce curvatures that are in agreement with modular binding as illustrated by structures 5AEI and 5MFC (average Cα(P/P+2) value of 6.94 Å and 6.98 Å, respectively) (Fig. 5). The induction of...
an optimal curvature by peptide binding is corroborated by the fact that none of the ligand-free structures exhibits a similar geometry and by the changes on peptide binding observed in single-molecule FRET measurements (see below). For the crystal lattice and additional parameters influencing the curvature. These appear to be dominant over the binding energy of the peptide, since in a number of structures no ligand is found (even though it was added to the crystallization mix), or only partial canonical binding is observed. We believe that in many cases the lattice enforces a curvature that is not adequate for modular binding or allows only modular binding of short peptide stretches. In addition, peptide binding might be affected by additives used to obtain crystals that lower the affinity of the interaction (e.g. high salt or extreme pH values) (Hansen et al., 2016). In some cases, non-canonical binding is observed even though a suitable curvature is adopted. Here steric or energetic effects are responsible for the deviation from the expected binding mode (e.g., 5MFF and 5MFG, see Results).

The curvature of computational dArmRP was analyzed previously for 4D4E and 4D49 (Reichen et al., 2016b). Here we extend this analysis with additional structures (5MFB, 5MFI, 5MFJ and 5MFK) to obtain a more representative dataset. The findings from the previous study that computational dArmRP possess curvatures which only allow limited modular peptide binding are confirmed by the additional structures analyzed in this study.

Capping repeats are known to be important for the stability of repeat proteins in general; in the case of dArmRP during the consensus design the caps were reengineered several times to improve the stability (Alfarano et al., 2012; Forrer et al., 2004; Interlandi et al., 2008; Madhuranthakam et al., 2012; Wetzel et al., 2008). We found that neighboring repeat pairs of caps, especially of the N-cap from consensus dArmRP, possess slightly higher C°(P/P+2) values (Fig. 5c and d). This might indicate that the interface between these repeats is not optimal. However, this issue can be easily addressed by introducing a “spacer repeat” between caps and binding repeats. Such an approach has been shown to even improve protein stability in library construction (Varamadsetty et al., 2012).

The analysis of the crystal structures suggests that consensus dArmRPs undergo a change in C°(P/P+2) distances upon (KR)5 peptide binding. Single-molecule FRET experiments of donor/acceptor labeled YpMMAAp performed in the presence and absence of (KR)5 show an increase in the mean transfer efficiency from 0.21 to 0.31 upon peptide binding. These values are in good agreement with those calculated from the crystal structures, taking the dynamics of the dye linker into account. Taken together, single-molecule FRET experiments demonstrate that the small conformational change of consensus dArmRPs observed in the crystal structures is also present in dArmRP in free solution.

In summary, our analysis shows that the consensus dArmRP scaffold is suitable as a basis for the development of modular peptide binders. An inherent flexibility allows the scaffold to adopt a suitable curvature for modular peptide binding. Further engineering of the scaffold could be envisaged to optimize the curvature also in the unbound state (i.e., decrease C°(P/P+2) distance) or possibly even by increasing the flexibility of the scaffold. This might increase the affinity towards peptides, even though this does not appear to be a pressing problem, as picomolar affinities can already be reached with the present design (Hansen et al., 2016). Furthermore, we could spotlight that in design approaches, where very high accuracy of the scaffold design is important, protein crystallization can introduce small but significant deviations from possible geometries in solution since it does not reflect the inherent flexibility of scaffolds and lattice forces can be dominant over induced changes by ligand-binding.

Acknowledgements

We would like to thank Céline Stutz-Ducommun and Beat Blattmann from the UZH Protein Crystallization Center for help with crystallization experiments and the staff of beamlines X06DA and X06SA at the Swiss Light Source (Paul Scherrer Institut, Würenlingen, Switzerland) for technical support. This work was supported by the Swiss National Science Foundation (grant number CRSI_141832/1 to A.P. and 31003A_152839 to B.S.), COST Action BM1405 (Non-globular proteins) and the Forschungskredit of the University of Zurich (grant numbers FK-13-028 to S.H. and FK-16-018 to P.E.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.jsb.2017.08.009.

References


