JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

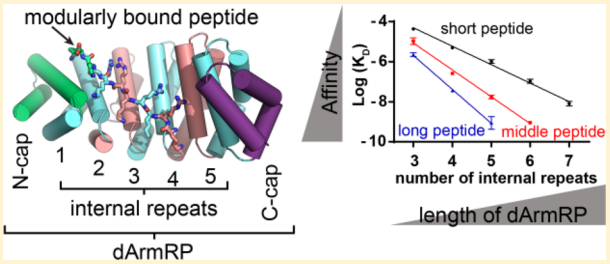
Structure and Energetic Contributions of a Designed Modular Peptide-Binding Protein with Picomolar Affinity

Simon Hansen, Dirk Tremmel, Chaithanya Madhurantakam,[†] Christian Reichen,[‡] Peer R. E. Mittl,* and Andreas Plückthun*

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

Supporting Information

ABSTRACT: Natural armadillo repeat proteins (nArmRP) like importin-α or β-catenin bind their target peptides such that each repeat interacts with a dipeptide unit within the stretched target peptide. However, this modularity is imperfect and also restricted to short peptide stretches of usually four to six consecutive amino acids. Here we report the development and characterization of a regularized and truly modular peptide-specific binding protein, based on designed armadillo repeat proteins (dArmRP), binding to peptides of alternating lysine and arginine residues $(KR)_n$. dArmRP were obtained from nArmRP through cycles of extensive protein engineering, which rendered them more uniform. This regularity is reflected in the consistent binding of dArmRP to (KR)-peptides, where affinities depend on the lengths of target peptides and the number of internal repeats in a very systematic manner, thus confirming the modularity of the interaction. This exponential dependency between affinity and recognition length suggests that each module adds a constant increment of binding energy to sequence-specific recognition. This relationship was confirmed by comprehensive mutagenesis studies that also reveal the importance of individual peptide side chains. The 1.83 Å resolution crystal structure of a dArmRP with five identical internal repeats in complex with the cognate $(KR)_5$ peptide proves a modular binding mode, where each dipeptide is recognized by one internal repeat. The confirmation of this true modularity over longer peptide stretches lays the ground for the design of binders with different specificities and tailored affinities by the assembly of dipeptide-specific modules based on armadillo repeats.



INTRODUCTION

Protein-peptide interactions play an important role in the regulation of many cellular processes such as signaling, replication, metabolism, trafficking, or gene-expression. By the term "protein−peptide interaction" we mean the binding of any unstructured polypeptide stretch to a folded protein domain; thus, the peptide can also be an unstructured region of a folded domain like, e.g., N- or C-termini, a loop or a linker region.

Several natural scaffolds have arisen, e.g., SH2, SH3, PDZ, PTB, WW, GYF, or FHA domains, that are specifically designed to interact with peptides.[1−3] Many of these interaction domains require specific peptide motifs, e.g., SH2 domains require peptides that contain phosphotyrosine residues[4] or SH3 domains bind PxxP motifs.[5] Since these recognition sequences are very short, restricted to particular sequences, and since the binding affinity is in general quite low, such domains are not suited as a tool for general peptide recognition. Nonetheless, engineered variants have been successfully applied for the development of peptide binders such as, e.g., SH2,[6] tetratricopeptide[7,8] or PDZ[9] domains (reviewed in ref 10).

Conversely, the recognition of nucleic acids can be mediated by repeat proteins, where each repeat interacts with one nucleotide, e.g., transcription activator like effectors (TALE) bind dsDNA[11−13] or the pumilio-homology domain[14−16] and artificial pentatricopeptide repeat proteins[17] interact with RNA. This allows to prolong the recognition length by stacking more repeats together, and sequence specificity can be controlled by choosing repeats that interact specifically with certain nucleotides.

*N*atural *Arm*adillo *R*epeat *P*roteins (nArmRP), such as importin-α or β-catenin, exhibit a modular binding mode comparable to the nucleotide-binding repeat proteins: they are also repeat proteins of variable length, and cognate peptides are bound in extended conformation within the concave groove of the solenoid protein. Thus, in a first approximation, each armadillo repeat interacts with two amino acids of the bound peptide.[18−20] However, this modularity is restricted to only short stretches of continuous peptide, the longest stretch of peptide that was structurally confirmed to be bound in a modular fashion has a length of six amino acids[19] (PDB ID: 1bk6), and usually, these stretches are much shorter. Flanking repeats often do not contribute to peptide binding, either because the superhelical curvature does not allow binding of a continuous peptide or because the surface of the nArmRP is not suited to do so.

In this study, we generated a modular system based on designed armadillo repeat proteins (dArmRP) that show a

Received: January 5, 2016
Published: February 15, 2016

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532

modular binding behavior to dipeptide units also in the context of longer peptide stretches. The armadillo repeat comprises three helices (H1, H2, and H3) arranged in a triangular spiral staircase with an extended hydrophobic core (Figure 1a).[18−21]

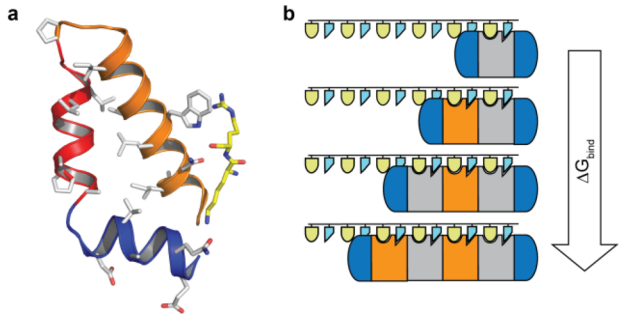


Figure 1. Armadillo repeat as scaffold for peptide binders. (a) Structure of the third armadillo repeat from mouse importin-α (PDB ID: 1ial[20]). Helices 1, 2, and 3 (residues 152 to 193) are shown as a ribbon in blue, red, and orange, respectively, and the bound (KR)-unit with yellow carbon atoms. Residues that are conserved in the M-type internal repeat[22] of $Y_{III}M_xA_{II}$ are shown as sticks with gray carbon atoms (21 out of 42 residues). (b) The free binding energy between the dArmRP and the target peptide depends on the recognition length and can be adjusted by adding peptide binding modules.

Sequences of nArmRPs were used to develop an initial version of a dArmRP using a consensus design approach.[22] The first generation dArmRP was extensively re-engineered by structure-aided design to overcome intrinsic repulsions, instability of caps and to liberate and improve the expected binding interface.[23−25] The final constructs possess the overall composition $Y_{III}M_xA_{II}$, where Y_{III}, M, and A_{II} represent the third generation N-terminal, internal-, and second generation C-terminal repeats, respectively (reviewed in ref 10). The subscript (x) denotes the number of internal repeats, while the roman number subscripts denote the design cycle. All proteins from the $Y_{III}M_xA_{II}$ series are monomeric, express very well in *E. coli* (up to 100 mg of pure protein per liter of culture), and possess improved biophysical stabilities over nArmRPs and previous dArmRP versions.[24] $Y_{III}M_xA_{II}$ is a full-consensus design, where all internal M-repeats are exactly identical.

Because the M-repeat was derived from nArmRPs that recognize positively charged peptide sequences (particularly importin-α which recognizes nuclear localization sequences[18,19,26]) (Figure 1a), $Y_{III}M_xA_{II}$ was designed to recognize, in turn, full-consensus target peptides, comprising identical (KR)-units. This modularity is reflected in additive binding energies when the length of either peptide or dArmRP is extended (Figure 1b). Modularity is an absolute prerequisite and a first step to develop a generic peptide recognition technology with the goal to eventually assemble preselected dipeptide-specific repeats to create binders against novel peptides without performing a selection, and this would thus eventually speed up binder generation to meet the increasing demand for such reagents.[27]

RESULTS AND DISCUSSION

Binding of $Y_{III}M_xA_{II}$ to $(KR)_n$ Peptides. ELISA experiments with biotinylated $(KR)_n$ peptides showed a clear trend: the affinities between $Y_{III}M_xA_{II}$ proteins and $(KR)_n$ peptides increase with the number of M-repeats and the number of (KR)-units. $Y_{III}M_3A_{II}/(KR)_3$ shows the weakest interaction with an ELISA signal slightly above background, whereas the combination $Y_{III}M_5A_{II}/(KR)_5$ shows the highest signal. Binding to the nuclear localization signal (NLS, KKKRKV) is observed but weaker than to $(KR)_3$. This is not surprising since both are hexapeptides that only differ by two point mutations (R2K and R6V) and as shown below, a reduced affinity is expected when arginines are mutated (Figure 2a). Dissociation constants (K_d) were determined for the interaction between $Y_{III}M_xA_{II}$ and fusions of $(KR)_n$ to superfolder green fluorescent protein[28] (sfGFP) by fluorescence anisotropy (Figure 2b). Here, the K_d values cover the range between 43 ± 5 μM and 1.1 ± 0.8 nM for the $Y_{III}M_3A_{II}/(KR)_3$-sfGFP and $Y_{III}M_5A_{II}/(KR)_5$-sfGFP interactions, respectively (Figure 2c, Table 1). The stoichiometry of the interactions were determined by saturation binding of tight interactions between $Y_{III}M_xA_{II}$ and $(KR)_n$-sfGFP and found to be 1:1 (Supporting Information, SI, Figure S1). Neither in the ELISA experiments nor in the fluorescence anisotropy assays noncognate control peptides, such as $(AV)_n$, neurotensin (NT, QLYENKPRRPYIL), or cro peptides (PRTSSF), revealed measurable affinities. Furthermore, the interaction between $Y_{III}M_xA_{II}$ and $(KR)_n$-sfGFP can be inhibited by excess of nonfluorescent $(KR)_n$ peptide (Figure 2d). From the inhibition of the $Y_{III}M_5A_{II}/(KR)_5$-sfGFP interaction by unfused $(KR)_5$ and fusions of $(KR)_5$ to the C-terminus of phage λ protein D (pD), K_d values of 0.94 ± 0.05 nM and 1.2 ± 0.7 nM, respectively, are obtained. These K_d values are almost identical to the K_d value of 1.1 ± 0.8 nM that was obtained from direct titration experiments. Thus, the interaction between $Y_{III}M_xA_{II}$ and $(KR)_n$ peptide is highly specific and independent of any additional domain fused to the $(KR)_n$-construct.

Contributions of Individual Peptide Side Chains to Binding of $Y_{III}M_xA_{II}$. Alanine scanning analysis revealed a significant contribution of each arginine and lysine side chain for binding, as alanine mutations at any position of $(KR)_4$–sfGFP significantly reduce the affinity to $Y_{III}M_5A_{II}$. On average, lysine and arginine contribute binding energies of $\Delta\Delta G$ equal to −4.2 ± 0.3 kJ/mol and −7.2 ± 0.2 kJ/mol, respectively, compared to the unchanged $(KR)_4$ peptide. With the exception of residue R8, which is located very close to the sfGFP-fusion partner, contributions are independent of the position in the peptide, and arginine consistently contributes more binding energy than lysine (p-values <0.0001). Peptides that have two lysine positions mutated to alanine show $\Delta\Delta G$ values of 10.4 ± 0.3 kJ/mol and peptides with two mutated arginines 15.7 ± 0.2 kJ/mol compared to the unchanged peptide. This corresponds approximately to the added effect of two individual lysine- or arginine-mutations, respectively. Therefore, we conclude that the contributions of side chains are independent from each other and two mutations have an additive effect (Figure 2e, Table 2 and Table S1).

Affinity Extrapolation of Very Tight Interactions. The tightest interaction pairs ($Y_{III}M_6A_{II}/(KR)_5$, $Y_{III}M_7A_{II}/(KR)_4$, and $Y_{III}M_7A_{II}/(KR)_5$) could not be measured directly with sufficient accuracy, because a correct affinity determination requires peptide-sfGFP concentrations below the K_d value, which conflicts with the sensitivity requirements of all methods tested. To estimate these affinities, we made use of the additive behavior of side chain contribution, we replaced subsequently all lysine residues in the $(KR)_n$ peptides against alanine and estimated the K_d for the unsubstituted peptide by extrapolation (Figure 2f). In a complementary approach, the dependence on the number of armadillo repeats was used to extrapolate tighter

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532

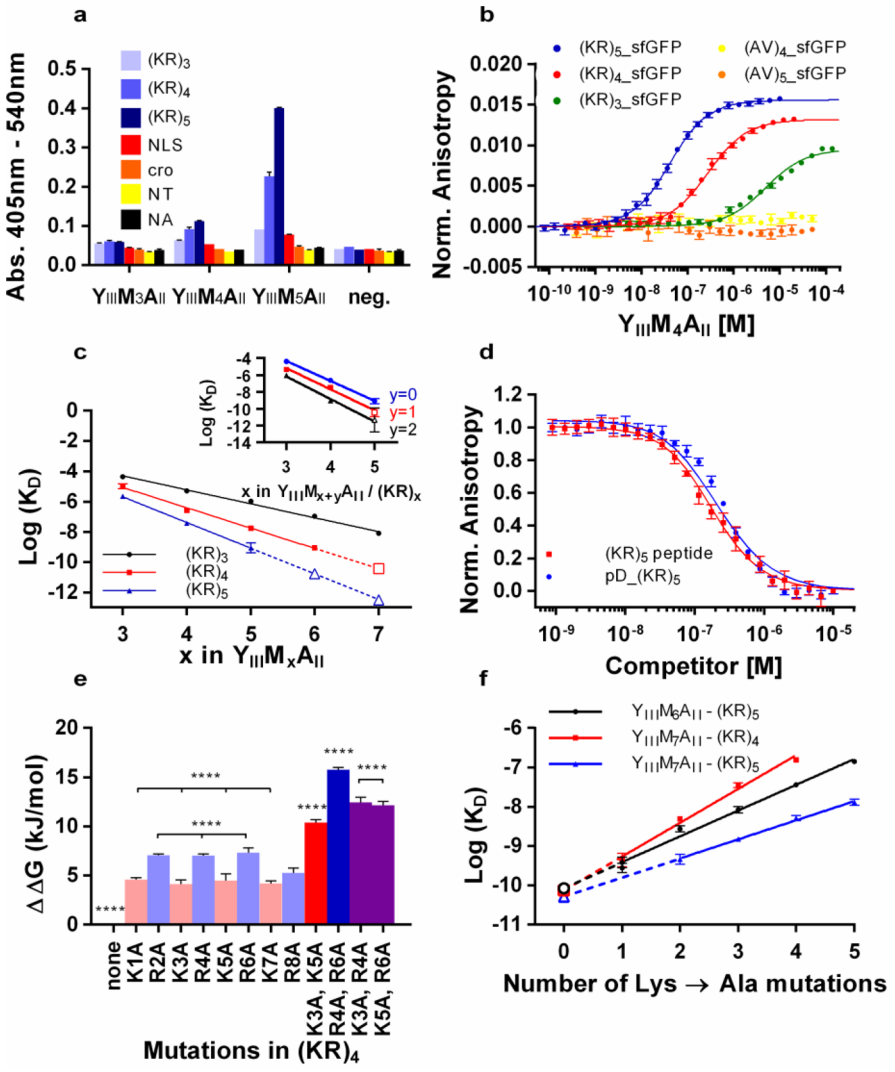


Figure 2. Binding experiments with $Y_{III}M_xA_{II}$ and $(KR)_n$ peptides. (a) ELISA experiments with $Y_{III}M_xA_{II}$ with 3 to 5 internal repeats and different peptides. Names and sequences: nuclear localization sequence (NLS): KKKRKV, cro-peptide (cro): PRTSSF, neurotensin (NT): QLYENKPRRPYIL, only Neutravidin (NA) with no peptide coated. (b) Example curves from direct fluorescence anisotropy assays with fits (solid lines). (c) Dependence of K_ds on the number of armadillo repeats for different peptides with linear regression. Continuous lines show linear regression of measured data. Extrapolation to low K_ds (high affinities) are indicated by dotted lines and open symbols. Inset: Dependence of K_ds when the length of dArmRP and $(KR)_n$ peptide are both increased stepwise. Extrapolated values used are means of the extrapolations from panels 2c and 2f and indicated by open symbols. (d) Example curves of competition assays, fits are indicated by solid lines. (e) $\Delta\Delta G$ values of $(KR)_4$ peptides with specific alanine mutations compared to unmutated $(KR)_4$ binding to $Y_{III}M_5A_{II}$. Significance levels of groups compared to all other groups are indicated (one-way ANOVA; p-value <0.0001 (****)). (f) Dependence of K_ds on the number of lysine to alanine mutations with linear regression for different combinations of dArmRP and $(KR)_n$ peptides. Linear regressions on measured values are indicated by continuous lines and full symbols. Extrapolation to low K_ds (high affinities) are indicated by dotted lines and open symbols. All error bars represent SD.

Table 1. Dissociation Constants of Combinations between $Y_{III}M_xA_{II}$ and $(KR)_n$ Peptides[a,b,c]

	K_d ± std. dev. (nM)				
	$(KR)_3$	$(KR)_4$	$(KR)_5$	$(AV)_4$	$(AV)_5$
$Y_{III}M_3A_{II}$	43550 ± 5300	11200 ± 4000	2290 ± 536	n.i.	n.i.
$Y_{III}M_4A_{II}$	5000 ± 600	265 ± 23	36.4 ± 1.2	n.i.	n.i.
$Y_{III}M_5A_{II}$	1040 ± 220	17.8 ± 3.5	1.1 ± 0.8	n.i.	n.i.
$Y_{III}M_6A_{II}$	108 ± 23	0.91 ± 0.1	0.017, 0.085[d]	n.i.	n.i.
$Y_{III}M_7A_{II}$	8.5 ± 2.3	0.037[d], 0.074[d]	0.0005[d], 0.051[d]	n.i.	n.i.
$Y_{III}M(N^{37}A)_5A_{II}$	6810 ± 730	184 ± 23	36.9 ± 11	n.d.	n.d.

[a]n.d.: not determined. [b]n.i.: no apparent interaction. [c]All values are means of at least three independent assays. [d]These values were not measured directly but extrapolated. Both values have been obtained from different extrapolated data: see text, Figure 2c,f and Table S2.

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532

Table 2. K_d Values between $Y_{III}M_5A_{II}$ and $(KR)_4$ Peptides with Defined Alanine Mutations[a]

peptide sequence	K_d ± std. dev. (nM)[b]	$\Delta\Delta G$ ± std. dev. (kJ/mol)[c]
KRKRKRKR	17.8 ± 3.5	
ARKRKRKR	112 ± 8.9	4.6 ± 0.2
KAKRKRKR	311 ± 13	7.1 ± 0.1
KRARKRKR	94.8 ± 13	4.1 ± 0.4
KRKAKRKR	305 ± 17	7.0 ± 0.2
KRKRARKR	113 ± 23	4.5 ± 0.7
KRKRKAKR	349 ± 50	7.3 ± 0.5
KRKRKRAR	97.4 ± 7.9	4.2 ± 0.2
KRKRKRKA	150 ± 26	5.2 ± 0.5
KRARARKR	1105 ± 28	10.4 ± 0.3
KRKAKAKR	10240 ± 729	15.7 ± 0.2
KRAAKRKR	2740 ± 415	12.5 ± 0.5
KRKRAAKR	2410 ± 289	12.1 ± 0.4

[a]All values are means of at least three independent assays. [b]All K_ds were measured with $Y_{III}M_5A_{II}$ in PBS (pH 7.4) with 0.03% BSA. [c]$\Delta\Delta G = RT\cdot(\ln(K_d \text{ Ala variant}/K_d \text{ unmutated peptide}))$ with $T = 298$ K.

interactions (Figure 2c). These extrapolations resulted in K_ds of 0.5 pM and 51 pM for the tightest interaction between $Y_{III}M_7A_{II}$ and $(KR)_5$-sfGFP, 37 pM and 74 pM between $Y_{III}M_7A_{II}$ and $(KR)_4$-sfGFP, and 17 pM and 85 pM for $Y_{III}M_6A_{II}$ and $(KR)_5$-sfGFP for the extrapolation on repeat number and number of lysine mutations, respectively, bracketing the range of the likely affinity (Figure 2c,f, Table 1 and Table S2).

Affinity Depends Exponentially on Number of Repeats and Peptide Length. It is interesting to note that the directly measured affinities follow a single exponential curve where the $\log(K_d)$ values decrease with the recognition length. The extension of the recognition length by adding (KR)-units to the peptide or M-repeats to the dArmRP increases the likelihood to form additional (KR)/M-repeat interactions in the complex. The increase in binding energy can be deduced from the slopes of the log-plots.

Consistently the energetic contribution of M-repeats increases with the number of (KR)-units in the peptide and vice versa. This is due to two effects: First, each additional (KR)/M-repeat interaction adds a constant amount of binding energy. Second, longer dArmRP and longer (KR) peptides offer more possibilities to bind consecutive peptide units or dArmRP and, therefore, such complexes have higher configurational entropy (Figure 2c, Table S3, Figure S2).

To estimate the binding energy for one (KR)/M-repeat pair, $\log(K_d)$ values were plotted as a function of simultaneous elongation of $Y_{III}M_xA_{II}$ and $(KR)_n$ (Figure 2c inset). The slopes of these curves are almost identical, no matter if the dArmRP exactly fits the length of the $(KR)_n$ peptide (x in $Y_{III}M_xA_{II}$ equal to n in $(KR)_n$; difference length $y = 0$) or if the dArmRP is one or two M-repeats longer than the respective (KR) peptide ($y = 1$ or 2). Thus, each (KR)/M-repeat pair contributes a fixed amount of binding energy of ΔG equal to -14.4 ± 0.7 kJ/mol. For increasing y-values the curves are shifted to lower K_ds, since additional binding sites for the peptide increase the likelihood to form the complex. In summary, the characterization of the $Y_{III}M_xA_{II}/(KR)_n$ interaction showed that affinity can be regulated over a wide range by the length of the binding partners, underlining the modularity of the design.

Crystal Structure of $Y_{III}M_5A_{II}$–$(KR)_5$ Complex Shows Modular Interactions. To confirm the expected modular

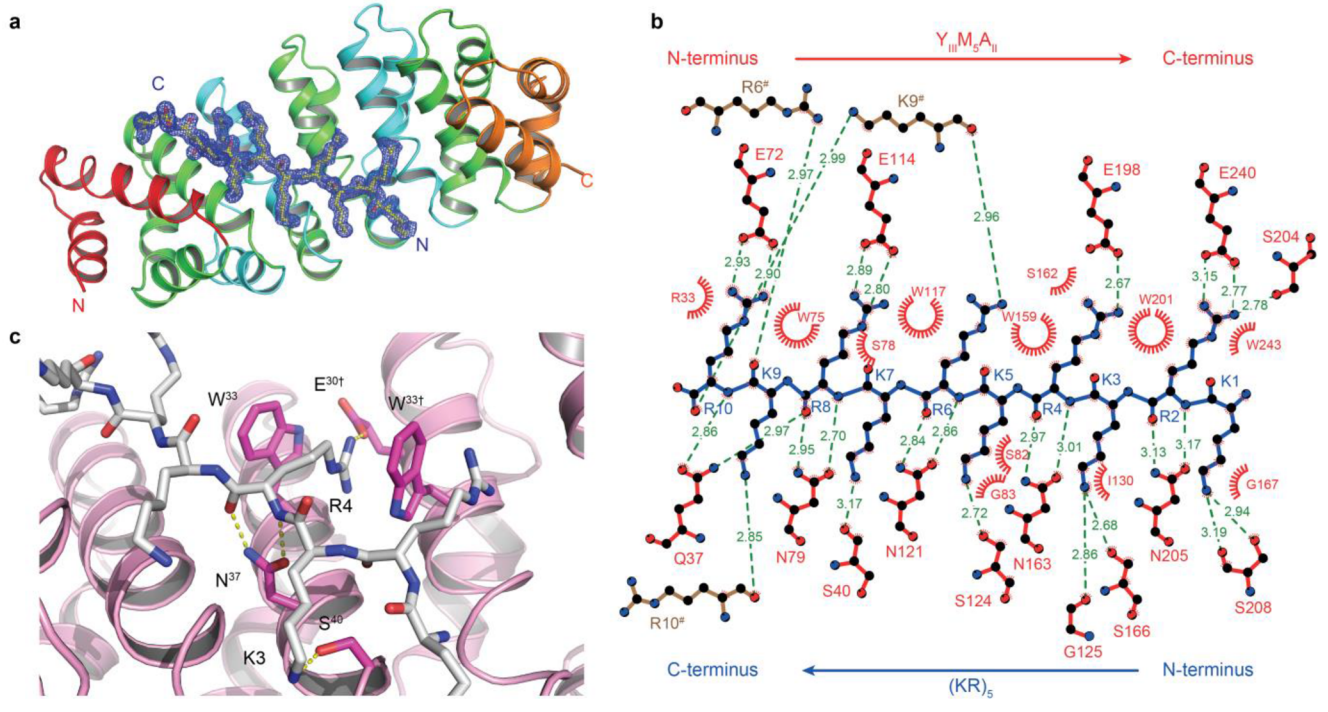


Figure 3. Structure of $Y_{III}M_5A_{II}$–$(KR)_5$ (PDB ID: 5aei) (a) $Y_{III}M_5A_{II}$ (Y_{III}-cap in red, M-repeats alternating green and cyan, A_{II}-cap in orange) in complex with $(KR)_5$ ($2F_o$—F_c map contoured at 1.2 σ), N- and C-termini are indicated for dArmRP and peptide. (b) Schematic interaction map of $Y_{III}M_5A_{II}$ (red) with $(KR)_5$ (blue) residues of a symmetry-related $(KR)_5^{\#}$ in brown, hydrogen bonds in green, numbers indicate distances in Å. Panel created with LigPlot+.[29] (c) Close-up view of typical binding pockets for Arg and Lys residues. Hydrogen bonds in yellow. Parts a and c were made with PyMol.[30]

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526–3532

binding we determined the crystal structure of the $Y_{III}M_5A_{II}$–$(KR)_5$ complex at 1.83 Å resolution. Data collection and refinement statistics are summarized in Table S4. One armadillo repeat contains 42 residues (Figure S3). Positions relative to a repeat are indicated by superscripted numbers (e.g., N^{37}) and positions in context of the whole protein with normal numbers (e.g., N163). All three complexes present in the asymmetric unit are very similar and reveal identical binding topologies (RMSD for Cα atoms ≤0.2 Å). The structure confirms that the peptide binds to the expected binding groove on the concave surface of the dArmRP and that the main chain directions of $Y_{III}M_5A_{II}$ and $(KR)_5$ are antiparallel like in many nArmRPs,[18,19] meaning that the N- to C-terminal directions of $Y_{III}M_5A_{II}$ and $(KR)_5$ run in opposite directions in the complex (Figure 3a, 3b). The main chain of all arginine residues from the peptide (besides R10) are bound by N^{37} side chains from $Y_{III}M_5A_{II}$ via bidentate hydrogen bonds, generating a regular β-sheet-like hydrogen bonding pattern. $(KR)_5$ is shifted toward the N-terminus of $Y_{III}M_5A_{II}$ so that K9 and R10 and repeat M_5 from $Y_{III}M_5A_{II}$ are not involved in the regular binding pattern. $(KR)_5$ side chains bind to well-defined pockets. Most arginine side chains form salt bridges with the side chains of $E^{30\dagger}$ (where † refers to a position in the following repeat) and interact via cation-π stacking with W^{33} and $W^{33\dagger}$. Lysine binding pockets are shallower than arginine pockets and the ε-amino groups of lysine residues form hydrogen bonds with S^{40}—Oγ and in some cases G^{41}—O (Figure 3b,c).

The structural features of the arginine and lysine binding interactions agree well with their contributions to binding energy since arginines show more interactions than lysines and also contribute more to binding energy (see above). All interactions are almost identical for the binding of (KR)-units 1 to 4. However, for the fifth (KR)-unit these interactions are perturbed, because R6, K9, and R10 form a crystal contact with $R6^{\#}$, $K9^{\#}$, and $R10^{\#}$ (where $^{\#}$ symbolizes symmetry-related residues). It can be assumed that the conformation of K9 and R10 and the side-chain of R6 is a crystallization artifact, because the $(KR)_5$ peptide is longer than the $Y_{III}M_5A_{II}$ binding groove and that the complex would adopt a more regular conformation in solution. This is also supported by the alanine scanning measurements (see above), where all lysine and all arginines, respectively, show the same energetic contribution to binding.

The polar interactions seen in the crystal structure predict that the binding of $(KR)_n$ peptides to $Y_{III}M_xA_{II}$ will be pH- and salt-dependent. The analysis of K_d values for the $(KR)_5$/$Y_{III}M_5A_{II}$ interaction as a function of pH and sodium chloride concentration confirms that the peptide binds best at neutral and slightly basic pH values and low salt concentrations (Table S5). The crystal structure also predicts that N^{37} plays an important role for main chain recognition. The $Y_{III}M(N^{37}A)_5A_{II}$ mutant indeed shows significantly decreased affinities for $(KR)_n$ peptides (Table 1). Compared to $Y_{III}M_5A_{II}$ the ΔΔG values for the binding of $(KR)_5$, $(KR)_4$, and $(KR)_3$ are −8.7, −5.8, and −4.7 kJ/mol, respectively. Thus, each N^{37} side chain contributes −1.6 ± 0.2 kJ/mol by binding to the arginine main chain.

CONCLUSIONS

We describe here for the first time a truly modular binding protein for $(KR)_n$ peptides where each part of the targeted peptide is recognized by a well-defined area on the dArmRP.

So far this was only known for DNA–DNA hybridization, protein–DNA interactions (transcription activator like effectors[11−13]) or protein-RNA interactions (pumilio-homology domains[14−16] or artificial pentatricopeptide repeat proteins[17]).

The modular behavior of the $(KR)_n$/$Y_{III}M_xA_{II}$ interaction is demonstrated by the exponential dependence of affinity on peptide and dArmRP lengths and the position-independent contribution of side chains to binding. The modularity is a key feature for generic peptide binders and distinguishes dArmRP from nArmRPs, where "hot-spot" residues are found on target peptides and the additivity of energetic contributions is not straightforward.[31] In the present system of full-consensus dArmRP/peptide complexes, the regularity of binding affinities is remarkable because the sum of the individual contributions by the arginine (−7.2 ± 0.2 kJ/mol) and lysine side chains (−4.2 ± 0.3 kJ/mol) and the asparagine main chain interaction (−1.6 ± 0.2 kJ/mol), resulting in −13.0 kJ/mol per repeat, agrees remarkably well with a ΔΔG value of −14.4 ± 0.7 kJ/mol for a single (KR)/M-repeat interaction in the length series of peptides and proteins. The additivity of the energetic contributions suggests that they are largely independent from each other, which is a crucial prerequisite for a modular detection system. Furthermore, the desired affinity can be designed by varying the recognition length of the binder (Figure 1a). The opportunity to generate very tight binders (with low picomolar and perhaps even better affinities) by adding more internal repeats to the solenoid protein is a unique feature of a modular recognition system.

A modular technology would have the advantage that interaction modules obtained from a selection can be used in multiple contexts to create further binders against different peptide sequences. This is not possible with anti-peptide antibodies, which are still the prevailing detection reagents, nor with other recombinant scaffolds.[27] Since the side chain-binding pockets on neighboring repeats are very close together our initial modularity model might have to be adjusted. One might have to find a compromise for residues shared in two pockets. This could be achieved by rational design or by directed evolution in context of neighboring peptide residues. The demand for affinity reagents is ever increasing, especially with the advent of proteome-wide approaches, and thus novel approaches will be needed to ensure a constant quality of binding reagents[27,32] in the future.

METHODS

Cloning, Protein Expression and Purification. These steps were performed according to slight modifications of methods that were described previously.[25,33] Details can be found in the SI Methods.

Enzyme Labeled Immune Sorbent Assays (ELISA). In-vivo biotinylated pD-peptide fusion proteins or synthetically biotinylated peptides (JPT, Germany) were immobilized on NeutrAvidin-coated (Thermo Scientific, 100 μL, 66 nM) and blocked plates by adding 100 μL of 200 nM fusion-protein or peptide solutions. One hundred microliters of 200 nM ArmRP were added. Binding was detected with an anti-$MRGSH_4$ antibody (Qiagen), a secondary anti-mouse immunoglobulin G alkaline phosphatase conjugate (Sigma), and *p*-nitrophenylphosphate (Fluka, 3 mM in 50 mM $NaHCO_3$ and 50 mM $MgCl_2$). Absorbance at 405 nm was measured using a M1000 plate reader (Tecan). A PBS buffer (pH 7.4) with 0.2% bovine serum albumin and 0.1% Tween 20 was used for all the protein-containing and blocking steps. Each step was incubated for 1 h. Washing after each step was carried out three times with 300 μL PBST (PBS with 0.1% Tween 20). All steps were performed at room temperature.

Fluorescence Anisotropy Assays. The assays were performed in black non-binding 96-well plates (Greiner). A constant concentration of peptide-sfGFP fusion was titrated with increasing concentration of ArmRP. The concentration of peptide-sfGFP was chosen to be

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532

maximally 2.5-fold over the respective K_d, preferentially below K_d. Four replicates for each concentration were pipetted; a dilution series of 24 concentrations of dArmRP was produced. Fluorescence Anisotropy was measured on a M1000 Pro or Safire II plate reader (Tecan) equipped with a fluorescence polarization module. Data were averaged, the anisotropy value with the lowest dArmRP concentration was subtracted from all other values. Fitting to a simple 1:1 binding model (eq 1) was done in SigmaPlot using eq 1,

$$Y(K_d, L_t, R_t) = \frac{m(-K_d - L_t - R_t + \sqrt{(K_d + L_t + R_t)^2 - 4L_tR_t})}{-2R_t} \quad (1)$$

where Y is fraction bound, m is the amplitude of maximal anisotropy increase, K_d is the dissociation constant, L_t is the total ligand concentration (dArmRP), and R_t is the total receptor concentration (peptide-sfGFP). We chose this denomination for L and R, because [peptide-sfGFP] is constant and [dArmRP] is variable.

Competition Assays. Two hundred nanomolar of each $(KR)_5$–sfGFP fusion and $Y_{III}M_5A_{II}$ were titrated with increasing amounts of nonfluorescent peptide competitor (synthetic peptide obtained from LifeTein or peptide-pD-fusion produced in-house). Four replicates for each dilution step were pipetted; in total a dilution series of 24 concentrations of competitor was produced. Fluorescence anisotropy was measured on a M1000 Pro or Safire II plate reader (Tecan) equipped with a fluorescence polarization module. Data were averaged and normalized to values between 0 and 1. Fitting was implemented in an Excel sheet according to ref 34. Two-sample t tests with Welch's correction were performed with the GraphPad Prism 6 software, a p-value <0.05 was regarded as significant.

Alanine Scanning. $(KR)_4$-sfGFP fusions with defined Ala mutations in the peptide were produced as described in the SI methods, each position in the peptide was mutated individually to Ala. Further peptides, where two Lys residues, two Arg residues or one of each were mutated, were produced in the same way (sequences in Figure S3). K_ds of peptides binding to $Y_{III}M_5A_{II}$ were measured as described above, with at least three independent assays per peptide-fusion. $\Delta\Delta G$ values were calculated with the formula: $\Delta\Delta G = RT\cdot(\ln(K_d \text{ Ala variant}/K_d \text{ unchanged peptide}))$. A one-way analysis of variance (ANOVA) was performed with multiple comparisons of all combinations of $\Delta\Delta G$ populations with the GraphPad Prism 6 software, a p-value <0.05 was regarded as significant (Table S1).

Extrapolation of Dissociation Constants of Tight Interactions. The very tight interactions of $Y_{III}M_7A_{II}$ with $(KR)_4$ and $(KR)_5$ and $Y_{III}M_6A_{II}$ with $(KR)_5$ that could not be measured directly with sufficient accuracy were extrapolated by two different methods. One is based on the dependence of K_ds of a given peptide on the number of armadillo repeats (Figure 2c). The decimal logarithm of K_ds of a given peptide ($(KR)_3$, $(KR)_4$, or $(KR)_5$) was plotted against the number of repeats of the respective dArmRP. Values for the K_ds can be found in Table 1. For every given peptide, a linear regression was performed, the equations and R^2 values that were obtained can be found in the SI Methods.

The second extrapolation made use of the dependence of K_ds of a given dArmRP on the number of Lys to Ala mutations in a peptide. The decimal logarithm of K_ds between a given dArmRP was plotted against the number of Lys to Ala mutations in a certain peptide. Values for the K_ds can be found in Table S2. For every peptide/dArmRP combination, a linear regression was performed, the resulting equations and R^2 values can be found in the SI Methods.

The obtained K_d values from both methods are given in Table 1, the first value coming from the extrapolation on the number of armadillo repeats and the second value coming from the extrapolation based on the number of lysine to alanine mutations.

Crystallization, X-ray Data Collection and Refinement. Sparse-matrix screens from Hampton Research and Molecular Dimensions in 96-well Corning plates (Corning Incorporated) at 4 °C were used to identify the preliminary crystallization conditions. Sitting-drop vapor-diffusion experiments were performed using a Phoenix crystallization robot (Art Robbins Instruments, U.S.A.). The protein solution was mixed with reservoir solutions at 1:1, 1:2 or 2:1 ratios (200−300 nL final volume), and the mixtures were equilibrated against 30 μL of reservoir solution. Data collection, and refinement statistics are summarized in Table S4. Twenty percent glycerol was added to the reservoir solution, and the crystals were flash-cooled in liquid nitrogen. Data were collected using a Pilatus detector system (Dectris Ltd., Switzerland) on beamline X06DA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). Data were processed using programs XDS, XSCALE, and XDSCONV[35]

The crystal structure was solved by molecular replacement using the program PHASER.[36] Molecular replacement was performed using a poly-Ala model of $Y_{III}M_5A_{II}$ based on PDB-ID 4plr.[25] After initial phase calculation, structure refinement (rigid-body and NCS restrained) was done using programs REFMAC5[37] and Phenix-Refine,[38,39] followed by model building in COOT.[40] Five percent of data were set aside to calculate the R_{free} value. The final structure was validated using the program PROCHECK.[41]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b00099.

SI Methods; Table S1, significance levels for alanine scan; Table S2, K_d values used in the extrapolation of tight binders; Table S3, $\Delta\Delta G$ values; Table S4, data collection and refinement statistics of $Y_{III}M_5A_{II}$ - $(KR)_5$; Table S5, K_d values at different pH and salt conditions; Figure S1, saturation binding; Figure S2, linear regression of dArmRP binding to $(KR)_n$; and Figure S3, alignment of protein sequences used. Coordinates and X-ray diffraction data were deposited under the PDB ID: 5aei (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*mittl@bioc.uzh.ch

*plueckthun@bioc.uzh.ch

Present Addresses

†Department of Biotechnology, TERI University, 10, Institutional Area, Vasant Kunj, New Delhi 110070, India

‡Molecular Partners AG, Wagistrasse 14, 8952 Schlieren, Switzerland.

Funding

This work was financially supported by a Swiss National Science foundation grant to P.M. and A.P. (Sinergia project CRSI33_122686 and CRSI_141832/1). S.H. is a recipient of the Forschungskredit of the University of Zurich, grant FK-13-028.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful for the help of Fabian Arnold, Jonathan Kiefer, and Ulla Schellhaas with anisotropy measurements. We would like to acknowledge Dr. Erik Sedlak for help with the fitting of competition fluorescence anisotropy data. Thanks also go to Céline Stutz-Ducommun and Beat Blattmann from the UZH Protein Crystallization Center for help with crystallization experiments and the staff from beamlines X06SA and X06DA from the Swiss Light Source (Paul Scherrer Institut, Würenlingen, Switzerland). Thamar Looser is acknowledged for providing the pQE30LIC_3C plasmid.

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532

REFERENCES

(1) Kuriyan, J.; Cowburn, D. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259.
(2) Pawson, T.; Nash, P. *Science* **2003**, *300*, 445.
(3) Pawson, T.; Scott, J. D. *Science* **1997**, *278*, 2075.
(4) Cohen, G. B.; Ren, R. B.; Baltimore, D. *Cell* **1995**, *80*, 237.
(5) Pawson, T. *Nature* **1995**, *373*, 573.
(6) Malabarba, M. G.; Milia, E.; Faretta, M.; Zamponi, R.; Pelicci, P. G.; Di Fiore, P. P. *Oncogene* **2001**, *20*, 5186.
(7) D'Andrea, L. D.; Regan, L. *Trends Biochem. Sci.* **2003**, *28*, 655.
(8) Jackrel, M. E.; Valverde, R.; Regan, L. *Protein Sci.* **2009**, *18*, 762.
(9) Ernst, A.; Gfeller, D.; Kan, Z. Y.; Seshagiri, S.; Kim, P. M.; Bader, G. D.; Sidhu, S. S. *Mol. BioSyst.* **2010**, *6*, 1782.
(10) Reichen, C.; Hansen, S.; Plückthun, A. *J. Struct. Biol.* **2014**, *185*, 147.
(11) Deng, D.; Yan, C. Y.; Pan, X. J.; Mahfouz, M.; Wang, J. W.; Zhu, J. K.; Shi, Y. G.; Yan, N. E. *Science* **2012**, *335*, 720.
(12) Mak, A. N. S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. *Science* **2012**, *335*, 716.
(13) Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. *Science* **2009**, *326*, 1509.
(14) Wang, X.; McLachlan, J.; Zamore, P. D.; Hall, T. M. *Cell* **2002**, *110*, 501.
(15) Campbell, Z. T.; Valley, C. T.; Wickens, M. *Nat. Struct. Mol. Biol.* **2014**, *21*, 732.
(16) Edwards, T. A.; Pyle, S. E.; Wharton, R. P.; Aggarwal, A. K. *Cell* **2001**, *105*, 281.
(17) Coquille, S.; Filipovska, A.; Chia, T.; Rajappa, L.; Lingford, J. P.; Razif, M. F. M.; Thore, S.; Rackham, O. *Nat. Commun.* **2014**, *5*.
(18) Conti, E.; Kuriyan, J. *Structure* **2000**, *8*, 329.
(19) Conti, E.; Uy, M.; Leighton, L.; Blobel, G.; Kuriyan, J. *Cell* **1998**, *94*, 193.
(20) Kobe, B. *Nat. Struct. Biol.* **1999**, *6*, 388.
(21) Huber, A. H.; Nelson, W. J.; Weis, W. I. *Cell* **1997**, *90*, 871.
(22) Parmeggiani, F.; Pellarin, R.; Larsen, A. P.; Varadamsetty, G.; Stumpp, M. T.; Zerbe, O.; Caflisch, A.; Plückthun, A. *J. Mol. Biol.* **2008**, *376*, 1282.
(23) Alfarano, P.; Varadamsetty, G.; Ewald, C.; Parmeggiani, F.; Pellarin, R.; Zerbe, O.; Plückthun, A.; Caflisch, A. *Protein Sci.* **2012**, *21*, 1298.
(24) Madhurantakam, C.; Varadamsetty, G.; Grütter, M. G.; Plückthun, A.; Mittl, P. R. *Protein Sci.* **2012**, *21*, 1015.
(25) Reichen, C.; Madhurantakam, C.; Plückthun, A.; Mittl, P. R. E. *Protein Sci.* **2014**, *23*, 1572.
(26) Kosugi, S.; Hasebe, M.; Matsumura, N.; Takashima, H.; Miyamoto-Sato, E.; Tomita, M.; Yanagawa, H. *J. Biol. Chem.* **2009**, *284*, 478.
(27) Taussig, M. J.; Stoevesandt, O.; Borrebaeck, C. A.; Bradbury, A. R.; Cahill, D.; Cambillau, C.; de Daruvar, A.; Dubel, S.; Eichler, J.; Frank, R.; Gibson, T. J.; Gloriam, D.; Gold, L.; Herberg, F. W.; Hermjakob, H.; Hoheisel, J. D.; Joos, T. O.; Kallioniemi, O.; Koegl, M.; Konthur, Z.; Korn, B.; Kremmer, E.; Krobitsch, S.; Landegren, U.; van der Maarel, S.; McCafferty, J.; Muyldermans, S.; Nygren, P. A.; Palcy, S.; Plückthun, A.; Polic, B.; Przybylski, M.; Saviranta, P.; Sawyer, A.; Sherman, D. J.; Skerra, A.; Templin, M.; Ueffing, M.; Uhlen, M. *Nat. Methods* **2007**, *4*, 13.
(28) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79.
(29) Laskowski, R. A.; Swindells, M. B. *J. Chem. Inf. Model.* **2011**, *51*, 2778.
(30) DeLano, W. L. *The PyMOL Molecular Graphics System*; Delano Scientific: San Carlos, CA, 2002.
(31) Hodel, M. R.; Corbett, A. H.; Hodel, A. E. *J. Biol. Chem.* **2001**, *276*, 1317.
(32) Bradbury, A.; Plückthun, A. *Nature* **2015**, *518*, 27.
(33) Varadamsetty, G.; Tremmel, D.; Hansen, S.; Parmeggiani, F.; Plückthun, A. *J. Mol. Biol.* **2012**, *424*, 68.
(34) Wang, Z. X. *FEBS Lett.* **1995**, *360*, 111.
(35) Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125.
(36) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658.
(37) Murshudov, G. N.; Vagin, A. A.; Lebedev, A.; Wilson, K. S.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 247.
(38) Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Headd, J. J.; Moriarty, N. W.; Richardson, J. S.; Richardson, D. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. *J. Appl. Crystallogr.* **2010**, *43*, 669.
(39) Afonine, P. V.; Mustyakimov, M.; Grosse-Kunstleve, R. W.; Moriarty, N. W.; Langan, P.; Adams, P. D. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 1153.
(40) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126.
(41) Laskowski, R. A.; Moss, D. S.; Thornton, J. M. *J. Mol. Biol.* **1993**, *231*, 1049.

DOI: 10.1021/jacs.6b00099
J. Am. Chem. Soc. 2016, 138, 3526−3532