Biochemistry

pubs.acs.org/biochemistry Article

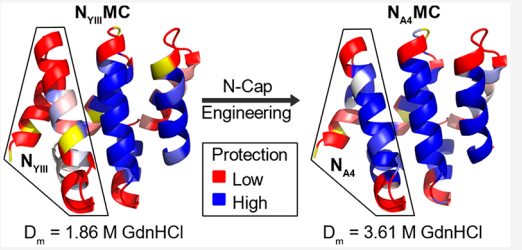
Improved Repeat Protein Stability by Combined Consensus and Computational Protein Design

Erich Michel,* Stefano Cucuzza, Peer R. E. Mittl, Oliver Zerbe, and Andreas Plückthun*

Cite This: https://doi.org/10.1021/acs.biochem.2c00083 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: High protein stability is an important feature for proteins used as therapeutics, as diagnostics, and in basic research. We have previously employed consensus design to engineer optimized Armadillo repeat proteins (ArmRPs) for sequence-specific recognition of linear epitopes with a modular binding mode. These designed ArmRPs (dArmRPs) feature high stability and are composed of M-type internal repeats that are flanked by N- and C-terminal capping repeats that protect the hydrophobic core from solvent exposure. While the overall stability of the designed ArmRPs is remarkably high, subsequent biochemical and biophysical experiments revealed that the N-capping repeat assumes a partially unfolded, solvent-accessible conformation for a small fraction of time that renders it vulnerable to proteolysis and aggregation. To overcome this problem, we have designed new N-caps starting from an M-type internal repeat using the Rosetta software. The superior stability of the computationally refined models was experimentally verified by circular dichroism and nuclear magnetic resonance spectroscopy. A crystal structure of a dArmRP containing the novel N-cap revealed that the enhanced stability correlates with an improved packing of this N-cap onto the hydrophobic core of the dArmRP. Hydrogen exchange experiments further show that the level of local unfolding of the N-cap is reduced by several orders of magnitude, resulting in increased resistance to proteolysis and weakened aggregation. As a first application of the novel N-cap, we determined the solution structure of a dArmRP with four internal repeats, which was previously impeded by the instability of the original N-cap.



The need for binding proteins that recognize linear or structural epitopes with high affinity and specificity is constantly increasing. These binding proteins are used as therapeutics, diagnostics, and research reagents. Today, most commercially available protein binders, in all three categories, are based on the antibody scaffold; however, alternative scaffolds with attractive properties are emerging.[1,2] A particularly interesting scaffold for the recognition of linear epitopes is provided by Armadillo repeat proteins (ArmRPs), an abundant eukaryotic protein family involved in a wide variety of biological functions that include transcription regulation, nuclear transport, cellular adhesion, etc.[3,4]

Naturally occurring ArmRPs (nArmRPs) are typically composed of around 8−12 internal repeats, which are flanked by N- and C-terminal capping repeats.[3,5,6] Each internal repeat contains around 42 amino acids that constitute three helices (H1−H3), which fold into a right-handed triangular staircase. The assembly of multiple repeats thus generates an elongated, right-handed superhelical protein shape that exposes a concave binding surface composed of adjacent helices H3, which interacts with polypeptide segments in an extended conformation.[5,6] This recognition involves specific interactions between the bound peptide side chains and the binding surface of the nArmRPs and is further enhanced by hydrogen bonds between the peptide backbone and conserved asparagine residues in helices H3.[5,6] In nArmRPs, binding is typically concentrated to short patches where, in a first approximation, two or three amino acids of the peptide are recognized per internal repeat. This modular peptide binding mode is less regular in nArmRPs and typically shows an alteration between short bound and unbound peptide stretches. Therefore, in nArmRPs, deviations from the ideal binding stoichiometry of two target amino acids per repeat are frequently observed.[5−7]

Designed ArmRPs (dArmRPs) have been engineered with the aim to create sequence-specific peptide binding scaffolds that feature modular and consecutive recognition of extended peptides with an exact stoichiometry of two amino acids per internal repeat.[7] So-called C-type internal repeats of the dArmRPs were obtained from a consensus design approach based on >240 input sequences from the importin-α and β-catenin/plakoglobin superfamilies. Further computational optimization of three hydrophobic core positions for improved packing in the C-type consensus design[7] and mutation of two lysine residues to glutamines to prevent electrostatic

Special Issue: Protein Engineering

Received: February 11, 2022
Revised: April 7, 2022

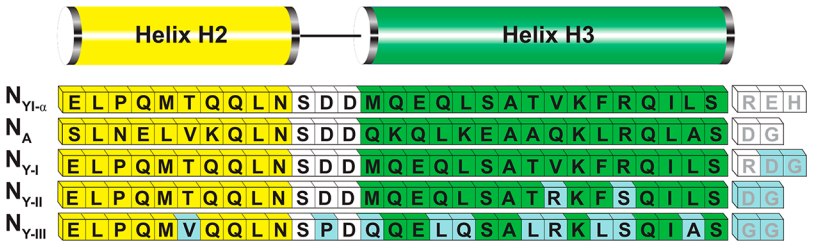


Figure 1. Previous generations of N-caps for dArmRPs. Sequences of previously engineered N-cap variants are shown. Residues in yellow and green boxes indicate helices H2 and H3, respectively. Residues in light blue boxes indicate modified positions. Abbreviations: $N_{YI-\alpha}$, yeast importin-α; N_A, artificial cap derived from consensus design and previous computational optimization; N_{Y-I}, N_{Y-II}, and N_{YIII}, first-, second-, and third-generation caps, respectively, derived from yeast importin-α and computational optimization.

repulsions[8] provided the M-type internal repeat (hereafter termed the internal M repeat).

The significant contribution of capping repeats to the overall protein stability and to prevent aggregation has been demonstrated previously for designed ankyrin repeat proteins (DARPins),[9,10] emphasizing the need for careful cap design. The C-terminal C_{AI}-capping repeat for dArmRPs was designed by replacing hydrophobic surface-exposed residues of the C-type internal repeat with hydrophilic ones, using guidance from available structural and sequence alignment data.[7] The C_{AII}-cap was subsequently generated by introducing two mutations near the C-terminus, which improved packing and solubility.[8] Moreover, replacing the C_{AI}-cap with the C_{AII}-cap in dArmRPs with four internal M repeats significantly increased the melting temperature by ∼7 °C and the transition midpoint in GdnHCl-induced unfolding by more than ∼0.5 M GdnHCl.[8]

Previous data from limited proteolysis experiments and sequence alignments[5−7] did not provide a clear definition for the start of the stable portion of the N-capping repeat in natural ArmRPs (nArmRPs). Moreover, nArmRP crystal structures provided resolved electron density for only helices H2 and H3 in the N-cap,[5,6] probably due to conformational dynamics involving residues corresponding to H1. Invisible N-cap residues in the crystal structure were thus considered unstructured, and the N-cap in nArmRPs was defined to comprise only helices H2 and H3.[7]

The first design of an N-capping repeat (N_A), which was based on optimization of surface-exposed residues in the C-type internal repeat (Figure 1), resulted in very low dArmRP solubility and expression yields.[7] An alternative N-cap design (N_{YI}) used residues E88−H119 of yeast importin-α as a starting scaffold and introduced the R117D and E118G mutations in the linker between helix H3 of the N-cap and helix H1 of the next internal repeat.[7] This N_{YI}-cap displayed enhanced solubility and expression yields; however, molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) experiments suggested significant flexibility in the N_{YI}-cap, which was addressed in the N_{YII}-cap by mutations V24R and R27S and deletion of R32 (Figure 1) to match the linker length between internal M repeats.[8] Exchanging the N_{YI}-cap with the N_{YII}-cap in dArmRPs with four internal M repeats showed rather modest increases of ∼2 °C in the melting temperature and 0.1−0.15 M GdnHCl in the transition midpoint in GdnHCl-induced unfolding.[8]

Despite the improved features, crystal structures of dArmRPs containing the N_{YII}-cap revealed domain swapping of the N_{YII}-cap due to formation of a continuous α-helix comprising H3 of the N_{YII}-cap and H1 of the first internal M repeat.[11] To further stabilize the N_{YII}-cap and to avoid domain swapping, the obtained crystal structures served as templates for a structure-based re-engineering to generate the N_{YIII}-cap: the D41G mutation aimed to minimize the helix propensity of the residues between the N-cap and the first internal M repeat and thus to suppress formation of a continuous helix comprised of helices H3 and H1; mutations T17V, Q28L, T32L, F35L, and L39A intended to improve packing of the hydrophobic core, M25Q and L29Q decreased the hydrophobicity of surface-exposed residues, and D23P enhanced loop formation between helices H1 and H2 (Figure 1). Overall, replacing the N_{YII}-cap with the N_{YIII}-cap increased the melting temperature by another 4.5 °C and the transition midpoint in GdnHCl-induced unfolding by 0.2 M GdnHCl.[11]

The successive engineering of the N-cap from the first N_{YI}-cap to the most recent N_{YIII}-cap provided a combined stabilization that resulted in increases by ∼6.5 °C in thermal unfolding and 0.3−0.35 M GdnHCl in denaturant-induced unfolding experiments.[8,11] Despite these stability improvements, we now provide evidence that the N_{YIII}-cap still shows significant local unfolding, which facilitates proteolytic degradation and aggregation. To provide a more robust N-cap, we report the engineering of significantly stabilized N-cap versions by combining consensus design and computational optimization and provide experimental evidence that highlights the obtained stability improvement.

EXPERIMENTAL SECTION

dArmRPs are composed of N- and C-terminal capping repeats, where in our nomenclature the subscripts refer to the version of the cap, and internal M repeats, where the subscript indicates the number of M repeats. Experimental procedures that include cloning of target genes and protein expression and purification are described in the Supporting Information.

NMR Analysis. NMR experiments were performed at 310.15 K on a Bruker Avance 600 spectrometer equipped with a cryogenic triple-resonance probe head. All NMR samples were supplemented with 5% (v/v) D_2O. Backbone resonances were assigned with two-dimensional (2D) $^{15}N-^{1}H$ HSQC, three-dimensional (3D) HNCA, 3D HNCACB, 3D HNCO, 3D HN(CA)CO, and 3D CBCA(CO)NH experiments.[12] Secondary structure analysis was performed using the C_α and C′ shifts according to the chemical shift index protocol.[13] Backbone amide mobilities were determined from 2D $^{15}N\{^{1}H\}$ heteronuclear NOE data recorded using a relaxation delay of 5 s.[14]

The amide proton exchange experiments were performed at pH 5.5 using 0.1 mM protein in a total volume of 500 μL. Proton exchange was started by redissolving the lyophilized protein sample in 500 μL of D_2O, followed by immediate and

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

continued measurement of 2D $^{15}N-^{1}H$ HSQC experiments after regular time intervals. All measurement and processing parameters were kept identical throughout the data acquisition series, and the sample was kept constantly at 37 °C between NMR measurements. The disappearance of individual amide resonances was followed by cross-peak integration using the software CARA,[15] and the residue-specific observed exchange rates (k_{obs}) were determined from a single-exponential decay fit to the amide cross-peak intensity versus time. Protection factors (P) for individual residues were determined from the ratio of intrinsic and observed exchange rates (k_{in}/k_{obs}).[16,17]

The structure determination of $N_{A4}M_4C_{AII}$ in solution using PCS constraints was performed according to the recently described procedure.[18] Three tag-attachment sites (E15C, Q92C, and S221C) were used for coupling of dia- and paramagnetic tags. The three initial structural models used as templates for the NMR structure calculation were derived from $N_{YIII}M_5C_{AII}$ [Protein Data Bank (PDB) entry 5AEI] by (i) deletion of the C_{AII}-cap and using the PyMOL mutagenesis wizard to first convert the residues of the last internal M repeat into C_{AII}-cap residues and then to convert the N_{YIII}-cap into the N_{A4}-cap residues, (ii) a Rosetta model, obtained by energy minimization of this first structural model using the *relax* protocol, and (iii) the crystal structure of $N_{A4}M_4C_{AII}$ determined in this work.

Computational Protein Design. The computational design in Rosetta was carried out with dArmRP constructs containing only one internal M repeat (generally termed NMC constructs). The used starting structural template $N_{YIII}MC_{AII}$ was created by least-squares superposition of the internal M repeats of the $N_{YIII}M$ and MC_{AII} fragments, derived from the crystal structure of $N_{YIII}M_5C_{AII}$ (PDB entry 5AEI). All Rosetta calculations were performed using the Rosetta 3.9 release and the "beta_nov16" scoring function.[19] Rosetta all-atom refinements of the initial $N_{YIII}MC_{AII}$ structural model were obtained by running the *relax* protocol[20,21] to generate 10 refined structural models, each obtained from a total of 20 cycles of side chain repacking and minimization. The obtained refined structural models served as templates for computational protein design of the N-cap with the *fixbb* protocol,[22] which was run with 500 trajectories for each of the 20 output structures. N-Cap residues chosen for side chain-rotamer optimization by Rosetta were tested for all possible amino acids except cysteine (ALLAAxC). Residues 1, 2, 4, 5, 8, 11–13, 15, 16, 19, 20, 23, 26, and 27 comprised the set of surface-exposed amino acids. The obtained designs were subjected to an all-atom refinement as described above, and the average Rosetta energy was calculated for the 10 output structural models.

Assessment of Protein Stability by Circular Dichroism (CD) Spectroscopy. Equilibrium denaturant-induced and thermal unfolding experiments with the NMC constructs were monitored by CD spectroscopy on a Jasco J-715 instrument using a cylindrical cuvette with a 1 mm path length equipped with temperature control. All measurements were performed using 15 μM protein in NMR buffer with a data pitch of 0.5 nm, a scanning speed of 100 nm/min, a response time of 4 s, a bandwidth of 1 nm, and a sensitivity of 100 mdeg. Denaturant-induced equilibrium unfolding was achieved by overnight incubation at room temperature with various concentrations of GdnHCl (Fluka) and measured via the ellipticity at 222 nm with 25 accumulations at 20 °C. The fraction of unfolded dArmRP at each concentration of GdnHCl was calculated according to eq 1:

$$F_U = \frac{\theta_N - \theta(x)}{\theta_N - \theta_U} \quad (1)$$

where θ_N and θ_U are the mean residue ellipticities for fully native and fully unfolded protein, respectively, and $\theta(x)$ is the observed ellipticity at x M GdnHCl. The obtained sigmoidal unfolding curves were then fit to eq 2:

$$f_U([D]) = \frac{(Y_N + m_N[D] + Y_U + m_U[D])e^{-(\Delta G_0 - m[D])/RT}}{1 + e^{-(\Delta G_0 - m[D])/RT}} \quad (2)$$

where f_U is the fraction of unfolded protein, [D] is the denaturant concentration, Y_N, m_N, Y_U, and m_U are intercepts and slopes for the pre- and post-transition baselines, respectively, ΔG_0 is the free energy in absence of a denaturant, and m describes the dependence of ΔG on the denaturant concentration.[23] The Y_N and Y_U intercepts were fixed to 0 and 1, respectively, while the slopes m_N and m_U were fixed to zero. The denaturation midpoint concentrations (D_m) were derived from eq 3:

$$D_m = \frac{\Delta G_0}{m} \quad (3)$$

Thermal unfolding of the NMC constructs was analyzed with a temperature increase of 1 °C min^{-1} from 25 to 95 °C while the ellipticity at 222 nm was recorded. However, as most constructs with novel N-caps were not completely unfolded at 95 °C, the melting temperature (T_m) was approximated with the inflection point obtained from the derivative of the unfolding curve instead of fitting the data with the usual temperature-dependent Gibbs–Helmholtz equation.[23]

Crystallization and Structure Determination. A 60 mg/mL $N_{A4}M_4C_{AII}$ solution in 10 mM Tris-HCl (pH 7.6) was applied to sparse-matrix screens from Molecular Dimensions and Hampton Research in 96-well plates (Corning) at 20 °C to identify crystallization conditions. Protein solutions were mixed at ratios of 1:1, 1:2, and 1:3 with the reservoir solution to volumes of 300–400 nL and equilibrated against 30 μL of the reservoir solution in sitting-drop vapor diffusion experiments. Crystals obtained in 35% (v/v) dioxane were picked after addition of 30% (v/v) ethylene glycol as a cryoprotectant and flash-frozen in liquid nitrogen. Diffraction data were collected with a Dectris Eiger X 16M detector on the X06SA beamline at the Swiss Light Source (Paul-Scherrer Institute, Villigen, Switzerland) and processed using XDS[24] and Aimless.[25] The crystal structure was determined by molecular replacement using MOLREP[26] with one protein chain from the $N_{YIII}M_5C_{AII}$:$(KR)_5$ complex structure (PDB entry 5AEI)[27] as a search model, followed by structural refinement using REFMAC[28] and model building in COOT.[29] R_{free} was calculated with 5% of separated data, and PROCHECK[30] was used to validate the final structure. All data collection and refinement statistics are listed in Table S3.

Affinity Determination. Affinities of NM_4C_{AII} proteins with various N-caps to the $(KR)_5$ peptide fused to mNeonGreen were determined by fluorescence anisotropy on a Tecan Safire II plate reader equipped with a fluorescence polarization module. A fixed amount of 2 nM $(KR)_5$-sfGFP was titrated with dArmRP in four replicates with 24 dilutions ranging from 160 pM to 20 μM. Excitation and emission wavelengths were

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX–XXX

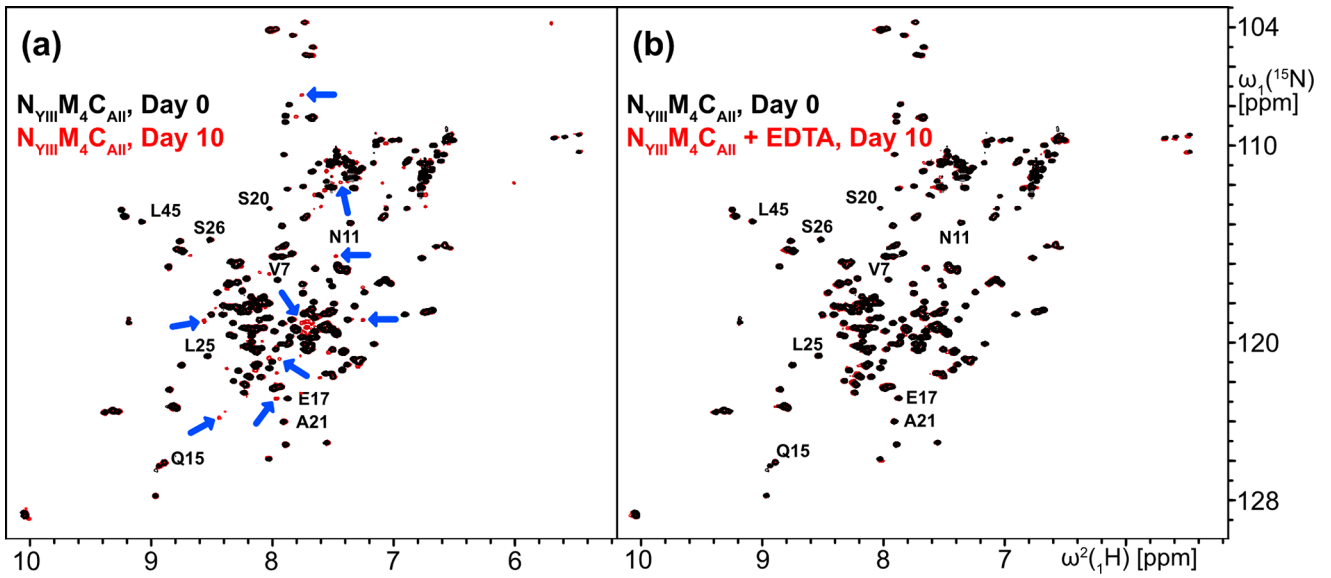


Figure 2. NMR analysis of $N_{YIII}M_4C_{AII}$ reveals sample instability. Superpositions of 2D $^{15}N-^{1}H$ HSQC spectra of 100 μM $N_{YIII}M_4C_{AII}$ in PBS buffer at pH 7 after incubation for 0 and 10 days at 37 °C measured in the (a) absence or (b) presence of 250 μM EDTA. Black and red resonances indicate spectra after 0 and 10 days, respectively, while blue arrows denote additional signals that appear after 10 days. The assignments of some signals have been added for orientation. All spectra were recorded at 37 °C and 600 MHz.

set to 470 and 510 nm, respectively, using a bandwidth of 10 nm. The averages of four replicates were subtracted with the anisotropy obtained with the lowest dArmRP concentration and were fit, as previously described,[27] to eq 4:

$$F_{AP}(c_A) = m\frac{K_d + c_A + c_P - \sqrt{(K_d + c_A + c_P)^2 - 4c_Ac_P}}{2c_P} \quad (4)$$

where F_{AP} is the anisotropy of the peptide at the given concentrations, which is proportional to the fraction of the bound peptide, c_A is the concentration of dArmRP, c_P is the fixed concentration of the peptide, K_d is the dissociation constant, and m is the anisotropy amplitude between the unbound and bound peptide.

RESULTS AND DISCUSSION

Designed Armadillo repeat proteins provide a promising scaffold for the engineering of modular sequence-specific peptide binding proteins. In this context, "peptide" refers to the recognition sequence of a linear epitope. For such applications, dArmRP scaffolds need to provide exceptionally high stability and solubility to compensate for potentially unfavorable structural changes that can be a consequence of introducing and modifying various binding pockets in the internal repeats. To further enhance the overall stability of dArmRPs, we aimed to optimize the N-cap, using a combination of consensus and computational protein design. We were motivated to focus on the N-cap because of a variety of observations summarized below.

NMR Analysis Reveals N_{YIII}-Cap Instability. We intended to use NMR spectroscopy to study the structural and dynamic adaptations of dArmRPs upon binding to their cognate target peptides. The initial isotope-labeled dArmRP prepared for NMR analysis comprised four internal M repeats with the N_{YIII}-cap and C_{AII}-cap as the N- and C-terminal capping repeats,[8,11] respectively. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis of the purified dArmRPs revealed high purity and the absence of undesired protein bands (data not shown). However, 2D $^{15}N-^{1}H$ HSQC spectra of the dArmRP revealed a gradual appearance of a subset of new signals with weak dispersion after several days at 37 °C, suggesting partial sample degradation (Figure 2a).

We had initially speculated that minute amounts of TEV protease, which was used to proteolytically remove the N-terminally $(His)_6$-tagged GB1 fusion domain during purification, might have remained in the NMR sample and exerted off-target cleavage that caused partial degradation of the dArmRP. To further investigate this, we supplied a freshly prepared dArmRP NMR sample with 20 μg of TEV protease and compared the NMR spectra recorded at different time points with those from dArmRP samples without added TEV protease. Unexpectedly, the addition of TEV protease prevented sample degradation and the appearance of new peaks, which we attributed to the protective effect of a storage buffer component such as EDTA rather than to the TEV protease itself. Indeed, supplementing the NMR samples with 0.25 mM EDTA effectively prevented the appearance of additional peaks and protected the protein from degradation (Figure 2b). This protective effect by EDTA suggested the presence of catalytic amounts of a co-purifying metalloprotease from the *Escherichia coli* expression host, which could not be detected by SDS–PAGE. Mass analysis of the partially degraded, ^{15}N-labeled NMR sample revealed a second protein species with a mass difference of 3105 Da from the intact dArmRP, which is in perfect agreement with proteolytic cleavage occurring between residues Q27 and I28, located in helix H3 of the N_{YIII}-cap. A subsequent bioinformatics search for known *E. coli* proteases that recognize this cleavage site provided no obvious results, however.

Protein Dynamics Suggest a Well-Folded and Rigid N_{YIII}-Cap. The available crystal structures of dArmRPs containing the N_{YIII}-cap[11,27] indicate formation of two helices, H2 and H3, in the N_{YIII}-cap. However, proteolytic cleavage requires transient unfolding of helix H3 to provide access of the protease to the backbone of its recognized target site. To

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX–XXX

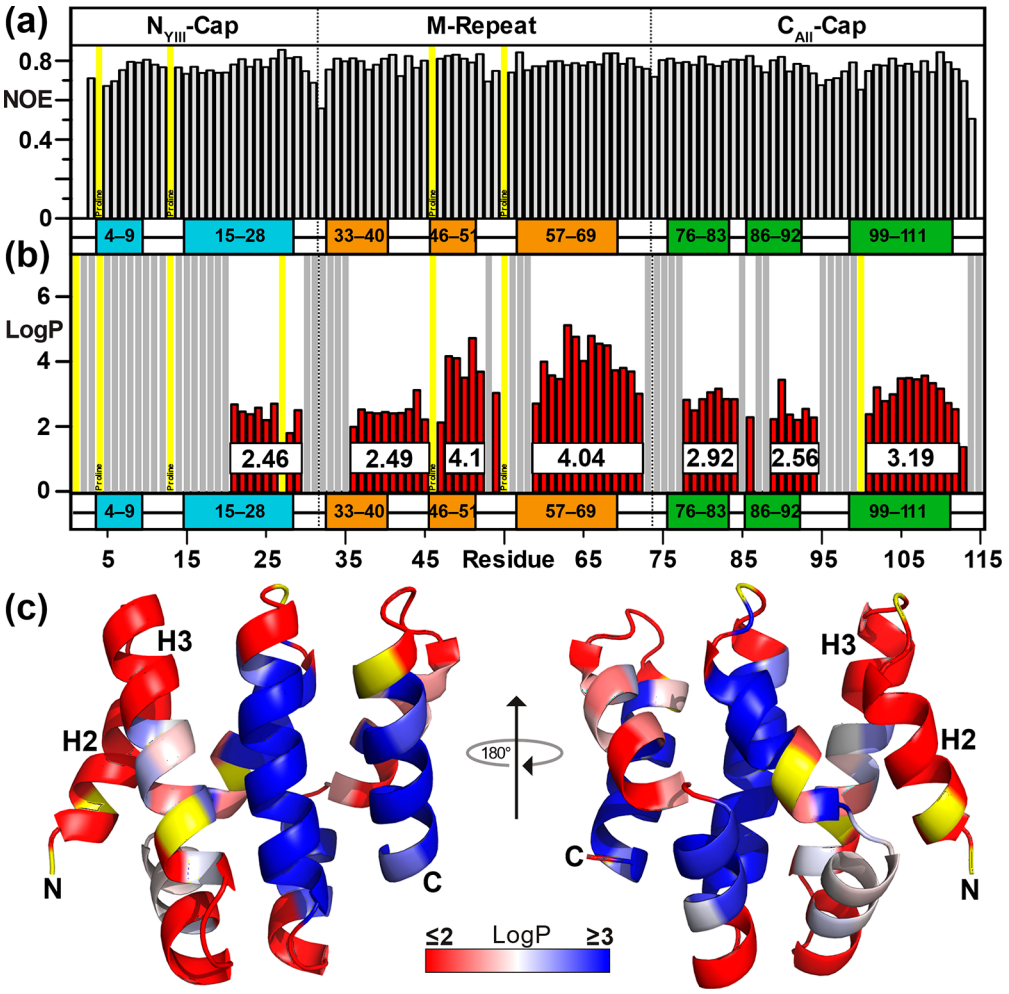


Figure 3. Conformational amide bond mobility and hydrogen exchange analysis for $N_{YIII}MC_{AII}$ at pH 5.5. (a) 2D $^{15}N\{^{1}H\}$ heteronuclear NOE values determined for individual backbone amides in $N_{YIII}MC_{AII}$. Colored boxes indicate helical segments in the N_{YIII}-cap (blue), internal M repeat (orange), and C_{AII}-cap (green) as determined from secondary chemical shift analysis. (b) Logarithm of protection factors (log *P*) obtained from hydrogen exchange analysis of $N_{YIII}MC_{AII}$. Gray bars indicate residues that exchange too fast to determine *P* values, while yellow bars indicate proline residues or residues with overlapping amide resonances for which no *P* values could be obtained. Numbers in white boxes on red bars indicate averaged log *P* values for individual structural units. (c) Mapping of experimental protection factors on a cartoon representation of the Rosetta structural model of $N_{YIII}MC_{AII}$. The color-coded gradient ranges from red, indicating weak protection with a maximum log *P* of 2, to white to blue, representing improved protection with a minimum log *P* of 3. Residues for which no log *P* values were available are colored yellow.

assess the conformational dynamics of the N_{YIII}-cap at atomic resolution by NMR, we prepared a minimalistic $N_{YIII}MC_{AII}$ dArmRP, with only one internal M repeat flanked by the N_{YIII}-cap and C_{AII}-cap. 2D $^{15}N-^{1}H$ HSQC spectra of this construct revealed well-dispersed amide signals without apparent line broadening, suggesting a uniform, well-folded protein population (Figure S1). Peak broadening of the backbone amide resonances was observed for only residues N33 and E34 of the internal M repeat and N75 and E76 of the C_{AII}-cap, indicating conformational dynamics in the intermediate exchange time regime (micro- to milliseconds) for residues that constitute the beginning of helix 1. The assignment of the $N_{YIII}MC_{AII}$ backbone resonances (BMRB entry 51239) provided the basis for a secondary structure analysis using the measured $^{13}C_{\alpha}$ and $^{13}C'$ chemical shift deviations from random coil values[13] (Figure S2). The secondary $^{13}C_{\alpha}$ chemical shifts suggest that helix H2 in the N_{YIII}-cap is comprised of residues P4−Q9 and helix H3 of residues Q15−S30 (Figure S2a). The secondary $^{13}C'$ chemical shifts confirm helical segments for residues P4−Q9 in helix H2 and for residues Q15−Q28 in helix H3 (Figure S2b). A comparison of helices H2 and H3 of the N_{YIII}-cap in solution with those observed in crystal structures[27] reveals identical secondary structure boundaries and thus confirms that the putative proteolytic cleavage site between Q27 and I28 is located within a helix.

To investigate amide bond mobilities on the pico- to nanosecond time scale within the N_{YIII}-cap, we carried out 2D $^{15}N\{^{1}H\}$ heteronuclear NOE (HetNOE) experiments.[14] The data analysis revealed near-maximal positive $^{15}N\{^{1}H\}$ HetNOEs and therefore indicates restricted amide bond motions for most residues within the N_{YIII}-cap, the internal M repeat, and the C_{AII}-cap (Figure 3a). A slight decrease in the HetNOE due to amide bond motions that are slightly faster than the overall tumbling of the protein was observed for residues G31 and G32, which connect the N_{YIII}-cap to the internal M repeat, and for the C-terminus of the protein (Figure 3a). In contrast, no significant increase in the backbone conformational dynamics was observed for the corresponding residues G73 and G74 that connect the internal M repeat with the C_{AII}-cap. These mobility differences may hint at a reduced conforma-

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

Table 1. Designed N-Cap Sequences and Rosetta Energies of the Corresponding NMC Constructs

N-Cap	Helix 2	Helix 3	Rosetta Energy
A4	PDLPKLVKLLKSS	NEEILLKALRALAEIAS	-358.3
A5	PDLPKLVKLLKSS	NEEILLKALKALAEIAS	-357
A7	PDLPKLVKLLKSS	DEETLLKALRTLAEIAS	-356.9
A9	PDLPKLVKLLKSS	DEKTLLEALKTLAEIAS	-356.8
A8	PDLPKLVKLLKSS	DEETLLKALKTLAEIAS	-356.6
A6	GALPALVQLLSSP	DEETLLKALKTLAEIAS	-349.8
H23	GALPALVQLLSSP	NEQILQEALWALSNIAS	-335
Y	GELPQMVQQLNSP	DQQELQSALRKLSQIAS	-332.9

tional stability in the structural vicinity of residues G31 and G32 compared to that near G73 and G74. However, the presented NMR data of $N_{YIII}MC_{AII}$ show a single NMR-observable protein population with an N_{YIII}-cap comprised of two stable helices and do not indicate conformational dynamics from helix unfolding within the N_{YIII}-cap.

Hydrogen Exchange Reveals Otherwise Invisible Transient Unfolded States. The aforementioned NMR analysis did not reveal detectable populations of alternative conformations and instead suggested formation of stable α-helices in the observable population of the N_{YIII}-cap. This implies that a conformation of $N_{YIII}MC_{AII}$ in which helix H3 of the N_{YIII}-cap is unfolded and accessible to proteolytic degradation must be so sparsely populated that it remains invisible to conventional NMR analysis. To illuminate such marginally populated "invisible" states that are in dynamic equilibrium with the native state of $N_{YIII}MC_{AII}$, we decided to measure the amide proton hydrogen exchange (HX) by NMR to reveal the possible existence and relative populations of these states at single-residue resolution. Hydrogen exchange between water and protein amides is directly correlated with the physical access of water molecules to individual amides in the protein, and the observed exchange rates (k_{obs}) can be described by eq 5:

$$k_{obs} = k_{int} \times \frac{k_1}{k_2} \quad (5)$$

where k_{int} is the residue-specific intrinsic exchange rate of a particular solvent-exposed amide proton, k_1 is the rate constant for the conversion from a solvent-protected (closed) into a solvent-exposed (open) state, and k_2 is the rate constant for the reverse process.[16,17,31] The closing equilibrium constant is termed protection factor P and is defined as the k_{int}/k_{obs} ratio. Amide protons engaged in hydrogen bond networks such as in α-helices and those buried in the hydrophobic core of a protein typically reach high P values.[31] An increased level of transient unfolding of helices H2 and H3 in the N_{YIII}-cap should therefore be reflected in significantly decreased P values compared to those of the more compact parts of the protein.[31]

The HX data of $N_{YIII}MC_{AII}$ recorded at pH 5.5 revealed that the first 20 residues of the N_{YIII}-cap exchange too fast to be captured in our experimental setup, indicating that P values for these residues must be smaller than ∼100 and that they spend at least 1% of the time in an open conformation (Figure 3b,c). The only residues of the N_{YIII}-cap for which we could measure protection factors were A21−A29 located within helix H3. The averaged log P value of 2.46 for this segment corresponds to 0.3% of the time spent in an open conformation. Residues S30−Q35, which comprise the linker between H3 of the N_{YIII}-cap and the beginning of H1 of the internal M repeat, also exchanged too fast to be observable. However, residues I36−A47, which constitute the majority of helix H1 of the internal M repeat up to the beginning of helix H2, exchange with an averaged log P value of 2.49, which closely resembles the value of the segment comprising residues A21−A29, suggesting that these segments unfold together as a cooperative unit (Figure 3b). Residues L48−L52 of helix H2 and residues I59−S72 of helix H3 in the internal M repeat show similar log P values of 4.1 and 4.04 that correspond to ∼0.005% and ∼0.003% of the time spent in an open conformation, respectively (Figure 3b). The similar log P values for H2 and H3 could suggest that these helices unfold in a cooperative manner. The helices in the C_{AII}-cap show more similar log P values among themselves, with values of 2.92, 2.56, and 3.19 for residues K78−A84 in helix H1, K89−Q94 in helix H2, and I101−L112 in helix H3, respectively (Figure 3b,c).

The HX data convincingly show that the residues in the N_{YIII}-cap have the lowest protection factors and that they spend at least 0.3% of the time in an open conformation, during which proteases can access the polypeptide chain. Helix 1 of the internal M repeat appears to be weakly protected and unfolds cooperatively with H3 of the N_{YIII}-cap; however, the cooperatively unfolding helices H2 and H3 of the internal M repeat possess protection factors ∼50−75-fold higher than that of helix H1, which can be rationalized by the more protected environment provided by packing against helices H2 and H3 of both N_{YIII}- and C_{AII}-caps. The corresponding P values of the C_{AII}-cap are severalfold increased compared to those of the N_{YIII}-cap, which implies a better overall packing of the C_{AII}-cap and suggests that the stability of the N_{YIII}-cap could possibly be improved by optimization of the repeat packing.

Computational N-Cap Design for Enhanced Stability. The HX experiments mentioned above have revealed that the N_{YIII}-cap spends a small but significant amount of time in an "open" conformation that gives access to the amide protons, while the internal M repeat shows a high protection factor and high stability. Previous experiments have further shown that helices H2 and H3 of the internal M repeat can substitute the N-cap in dArmRPs without significant losses of stability or solubility.[32,33] Due to these favorable properties, we decided to use an N_{H23}-cap composed of helices H2 and H3 of the internal M repeat as a starting template for a new N-cap, in combination with one internal M repeat and a C_{AII}-cap, for an *in silico* design of new N-caps using the Rosetta macromolecular modeling program.[34]

A scanning mutagenesis screen probing each individual position in the N_{H23}-cap showed that the largest energetic gains in Rosetta can be obtained by mutation of surface-exposed residues located in helices H2 and H3 (Table S4), suggesting that the packing of the existing hydrophobic core,

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

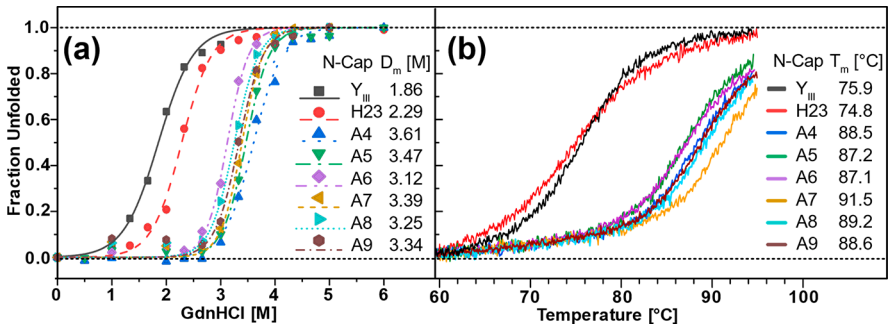


Figure 4. Denaturant-induced and thermal unfolding analysis of NMC constructs with different N-caps. (a) Guanidine hydrochloride (GdnHCl)-induced unfolding and (b) thermal unfolding curves of the different NMC proteins containing either newly designed N-caps or the original N_{YIII}-cap. Protein unfolding was monitored by following the CD signal at 222 nm. The obtained denaturation midpoint concentrations of GdnHCl (D_m) and melting temperatures (T_m) are indicated for each N-cap variant.

transferred from the internal M repeat, is scored favorably by Rosetta. Due to this finding, our design strategies included simultaneous optimization of either all surface-exposed or all residues of the N_{H23}-cap, using a combination of the Rosetta *fixbb* and *relax* protocols. Rosetta-proposed mutations occurred mainly for surface-exposed residues (Table 1), confirming the initial results of the scanning mutagenesis screen (Table S4). The total Rosetta energy units (REUs) of the newly designed NMC variants after energy minimization ranged from ∼350 to ∼358 REUs, which compare favorably to the values of 333 and 335 REUs obtained for the constructs containing the original N_{YIII}-cap and the template N_{H23}-cap, respectively (Table 1).

The N-cap variant A6, a hybrid construct composed of the original helix H2 from the starting template N_{H23} and a newly designed helix H3, scored 17 REUs better than the original $N_{YIII}MC_{AII}$, whereas all variants containing both newly designed helices H2 and H3 scored at least 24 REUs better than $N_{YIII}MC_{AII}$. This indicates that the REU gains were >2-fold larger in helix H3 than in helix H2. All N-cap variants with optimized helices H2 and H3 differ by <1.7 REUs from each other and show only a few conservative sequence variations (Table 1). The sequence composition of the newly designed N-caps shows a large proportion of charged amino acids, which account for approximately one-third of all residues, and an even slightly larger proportion of the helix-forming residues Leu and Ala. Interestingly, all seven Gln residues in the original N_{YIII}-cap sequence have been replaced with either Lys, Glu, or Leu in the new N-cap sequences by Rosetta.

A comparison of Rosetta energies for each residue in the original N_{YIII}-cap with the corresponding residue in the highest-scoring N_{A4}-cap reveals that the five mutations (M6L, Q9L, Q19L, K24A, and S26A), which are located at or in the hydrophobic core, account for a gain of 18.7 REUs. Most surface-exposed residues show smaller individual REU gains but contribute favorably to the overall stability of the new N_{A4}-cap (Table S5). This suggests that transfer of the hydrophobic core from an internal M repeat to the N-cap provided mainly stability, while redesign of surface-exposed residues addressed both protein solubility and stability.

Experimental Stability Assessment of N-Cap Designs. To experimentally assess the stability of the newly designed N-caps, we expressed and purified the corresponding NMC constructs to analyze both denaturant-induced and thermal unfolding of these proteins by CD spectroscopy. Denaturant-induced equilibrium unfolding of the NMC constructs was achieved with increasing concentrations of guanidine hydrochloride (GdnHCl) in PBS buffer (pH 7) and monitored by recording the CD signal at 222 nm. The analysis showed cooperative unfolding for all tested constructs and provided denaturation midpoint concentrations (D_m) of 1.86 and 2.29 M GdnHCl for $N_{YIII}MC_{AII}$ and $N_{H23}MC_{AII}$, respectively, while all NMC constructs containing a newly designed N-cap showed D_m values ranging from 3.12 M GdnHCl for $N_{A6}MC_{AII}$ to 3.61 M GdnHCl for $N_{A4}MC_{AII}$ (Figure 4). The corresponding ΔG values obtained for the $N_{YIII}MC_{AII}$ and $N_{H23}MC_{AII}$ constructs were −14.1 and −20 kJ/mol, respectively, while all new N-cap constructs showed ΔG values of at least −32.9 kJ/mol (Table S6). However, due to the mutual dependency of ΔG and m in the curve fit and the limited accuracy of the m values due to the moderate number of data points in the transition region of the unfolding curve, D_m appears to be a more reliable measure of the stability of the investigated proteins.

The calculated Rosetta energies agree remarkably well with the ranking of experimentally determined stabilities toward denaturant-induced unfolding and show a correlation of 16.5 ± 1.3 REU for a change in D_m of 1 M GdnHCl (Figure S3). The optimization of surface-exposed residues appears to be a very important contributor to the large overall stability improvement because the sole transfer of helices H2 and H3 of an internal M repeat, which provided the stable hydrophobic core, to the N_{H23}-cap increased the D_m value only to 2.29 M GdnHCl. N-Caps obtained after including redesign of surface-exposed residues all showed D_m values of >3 M GdnHCl. The large increase in D_m from 1.86 M for $N_{YIII}MC_{AII}$ to 3.61 M GdnHCl in $N_{A4}MC_{AII}$ underlines the significantly improved stability of the novel N-caps and is ∼5 times larger than all combined D_m gains from previous N-cap engineering efforts.[8,11]

To complement and support the denaturant-induced unfolding data, we followed thermal unfolding of the NMC constructs by recording the CD signal at 222 nm during a slow and steady temperature increase of 1 °C min^{-1} from 25 to 95 °C. In contrast to the denaturant-induced unfolding data, the thermal unfolding stabilities did not follow the exact ranking suggested from the Rosetta energies (Figure 4) but still showed a good correlation of 1.60 ± 0.18 REU for a change in T_m of 1 °C (Figure S3). All NMC constructs containing newly designed N-caps showed significantly increased T_m values between 87.1 and 91.5 °C, compared to T_m values of 75.9 and 74.8 °C for $N_{YIII}MC_{AII}$ and $N_{H23}MC_{AII}$, respectively, which

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

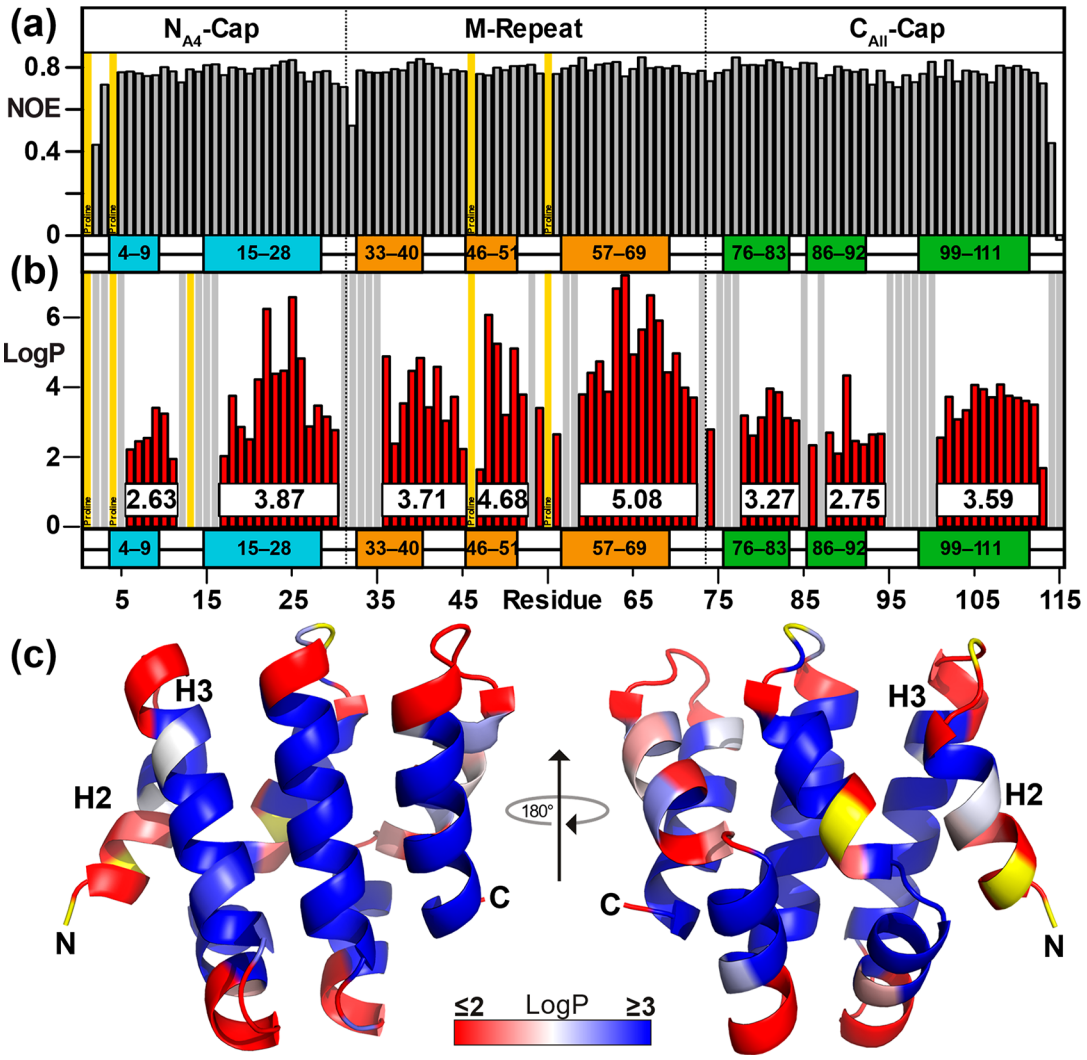


Figure 5. Conformational amide bond mobility and hydrogen exchange analysis for $N_{A4}MC_{AII}$ at pH 5.5. (a) 2D $^{15}N\{^{1}H\}$ heteronuclear NOE values determined for individual backbone amides in $N_{A4}MC_{AII}$. Colored boxes indicate helical segments in the N_{A4}-cap (cyan), the internal M repeat (orange), and C_{AII}-cap (green) as determined from the secondary chemical shift analysis. (b) Logarithm of protection factors (log *P*) obtained from hydrogen exchange analysis of $N_{A4}MC_{AII}$. Gray bars indicate residues that exchange too fast to determine *P* values, while yellow bars indicate proline residues or residues with overlapping amide resonances for which no *P* values could be obtained. Numbers in white boxes on red bars indicate averaged log *P* values for individual structural units. (c) Mapping of experimental protection factors on a cartoon representation of the Rosetta structural model of $N_{YIII}MC_{AII}$. The color-coded gradient ranges from red, indicating weak protection with a maximum log *P* of 2, to white to blue, representing an increased protection factor with a minimum log *P* of 3. Residues for which no log *P* values were available are colored yellow.

thus confirmed the high stability of the new N-caps. Furthermore, all NMC constructs showed completely cooperative and reversible thermal unfolding (data not shown).

NMR Analysis of $N_{A4}MC_{AII}$. The large increase in stability for the $N_{A4}MC_{AII}$ construct prompted us to further characterize the structural and dynamic properties of this protein by NMR spectroscopy. We therefore prepared ^{13}C- and ^{15}N-labeled $N_{A4}MC_{AII}$ to assign the backbone resonances (BMRB entry 51240) and to derive secondary shifts, which indicated no significant differences in the helical properties of the proteins $N_{YIII}MC_{AII}$ and $N_{A4}MC_{AII}$ (Figure S4). Furthermore, HetNOE data showed no increased conformational mobilities of any backbone amides in the $N_{A4}MC_{AII}$ protein, including the newly designed N-cap (Figure 5a), which indicates a rigid conformation of the predominant population. This is comparable to the data of $N_{YIII}MC_{AII}$, for which transient opening of the conformation could be detected only by HX experiments and long-term incubation (see below).

We then analyzed and compared the long-term stabilities of the new $N_{A4}MC_{AII}$ and original $N_{YIII}MC_{AII}$ proteins. In contrast to the previously observed slow degradation of the $N_{YIII}M_4C_{AII}$ protein, the smaller $N_{YIII}MC_{AII}$ completely precipitates upon prolonged incubation at 37 °C (Figure S5), which is likely due to the reduced solubility of the populations with partially unfolded helices and/or repeats in the smaller protein, compared to the proteins containing four internal repeats. $N_{A4}MC_{AII}$, on the contrary, does not show any changes in the spectra after 64 days (Figure S6), indicating that the novel N_{A4}-cap completely prevents adverse sample modifications, such as proteolysis or aggregation, and confirms the increased stability observed in the unfolding experiments.

Hydrogen Exchange of N_{A4}-Cap Indicates Stabilized Folding Units. The previous HX data of $N_{YIII}MC_{AII}$ showed

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

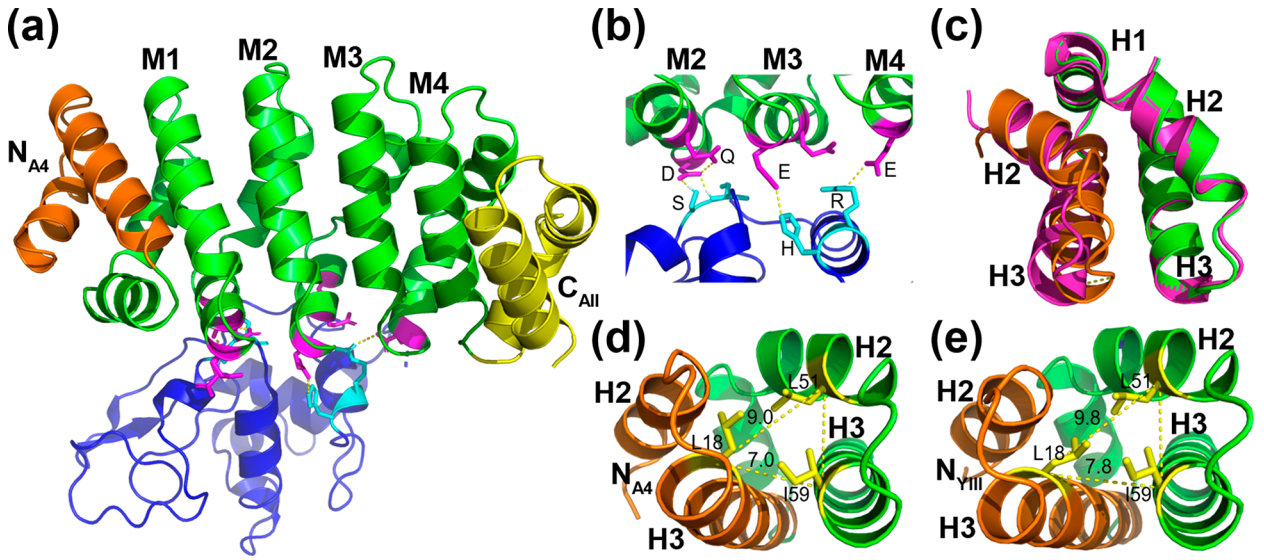


Figure 6. Crystal structure of $N_{A4}M_4C_{AII}$ showing improved helical packing of the N_{A4}-cap against an internal M repeat. (a) Crystal structure of $N_{A4}M_4C_{AII}$ determined in complex with lysozyme (PDB entry 7QNP). The N_{A4}-cap, internal M repeats, and C_{AII}-cap are colored orange, green, and yellow, respectively, while lysozyme is colored blue. (b) Close-up of the contacts observed between $N_{A4}M_4C_{AII}$ and lysozyme. Important residues are indicated by single letters. (c) Superposition of N-caps and first internal M repeats from the crystal structure of $N_{A4}M_4C_{AII}$, colored orange and green, and the crystal structure of $N_{YIII}M_5C_{AII}$ (PDB entry 5AEI), colored magenta. (d and e) Distances (Å) from L18 in helix H3 of the N-cap to L51 in helix H2 and I59 in helix H3 of the first internal M repeat are indicated for $N_{A4}M_4C_{AII}$ and $N_{YIII}M_5C_{AII}$, respectively (PDB entry 5AEI).

that the N_{YIII}-cap is the least stable repeat within the construct, which spends at least 0.3% of the time in an open conformation. To compare its properties with those of the new N-cap in $N_{A4}MC_{AII}$, we analyzed the amide HX in $N_{A4}MC_{AII}$ using the identical setup as for $N_{YIII}MC_{AII}$ (Figure 5b,c). The previously unobservable H2 of the N_{YIII}-cap is now sufficiently stabilized in the N_{A4}-cap to measure its exchange rates, which provides a log *P* of 2.63 for residues L6−K11 and shows that H2 spends 0.23% of the time in an open conformation. The linker segment comprising residues S12−E16 exchanged too fast to be measurable; however, residues I17−S30 showed a significantly increased log *P* of 3.87, which corresponds to only 0.014% of the time in an open conformation.

The only segment in the N_{YIII}-cap that can be observed by HX, which appears to contain the proteolytic target cleavage site in $N_{YIII}M_4C_{AII}$, comprised residues A21−A29 with a log *P* of 2.46 (Figure 3b). In the N_{A4}-cap, the corresponding segment now shows a log *P* of 4.47, increased by >2 orders of magnitude, which explains the increased sample stability (Figure 5b). Moreover, the internal M repeat displays a >15-fold increase in *P* values for helix H1, an ∼4-fold increase for helix H2 and an ∼10-fold increase for helix H3, compared to the *P* values obtained in $N_{YIII}MC_{AII}$. Albeit weakly, this stability increase is further propagated even into the C_{AII}-cap where helices H1−H3 show improvements in *P* values of >2-, >1.5-, and >2.5-fold, respectively. This indicates that cooperative effects arising from the improved stability and tight packing of the N_{A4}-cap are transferred throughout the entire protein.

The Crystal Structure of $N_{A4}M_4C_{AII}$ Highlights Tighter N-Cap Packing. To gain insight into the structural details of the novel N_{A4}-cap, we determined the crystal structure of $N_{A4}M_4C_{AII}$, at 1.59 Å resolution (PDB entry 7QNP). During the refinement, difference electron density emerged, which could not be assigned to a dArmRP molecule. Because the electron density was very clear, a structure was built manually into this difference electron density map. We then used the manually built protein chain to query the PDB database, which unambiguously assigned it to hen egg white lysozyme (PDB entry 132l). Obviously, lysozyme, which was used for lysing the *E. coli* host, was accidentally co-purified and co-crystallized with $N_{A4}M_4C_{AII}$. The binding interface between the dArmRP and lysozyme involves mainly polar interactions between residues of helices H1 in internal repeats M2−M4 of the dArmRP and residues in lysozyme (Figure 6). Affinity measurements between $N_{A4}M_4C_{AII}$ and lysozyme by isothermal titration calorimetry indicate a very weak interaction with a K_d of ∼6.6 μM (data not shown).

The helical boundaries observed in the crystal structure correspond well with the secondary shifts determined by NMR. This confirms that helices H2 and H3 of the N_{A4}-cap are comprised of residues L3−K11 and E15−S28, respectively. A structural comparison between the N_{A4}- and N_{YIII}-caps shows that helix H3 of the N_{A4}-cap packs more closely against helices H2 and H3 of the first internal M repeat (Figure 6), which further supports the increased protection factors for helices located in both the N_{A4}-cap and the neighboring internal M repeat. For example, the $C_\alpha-C_\alpha$ distances from L18, which is a common residue in both N_{A4}- and N_{YIII}-caps, to L51 in helix H2 and I59 in helix H3 of the internal M repeat decrease from 9.8 to 9.0 Å and from 7.8 to 7.0 Å, respectively (Figure 6). Other available crystal structures of dArmRPs containing the N_{YIII}-cap (PDB entries 5MFH, 4V3O, and 5MFD) show values of 10.7−11 and 8.4−9.1 Å for the corresponding L18−L51 and L18−I59 distances, respectively.

Novel N-Caps Do Not Impact Target Peptide Binding. dArmRPs are modular peptide binding molecules that interact with their cognate target peptides via specific interactions mediated by the internal M repeats. The capping repeats provide stability and solubility and are not supposed to contribute to the specific target peptide recognition. To assess the nonbinding properties of the novel N-caps, we determined

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

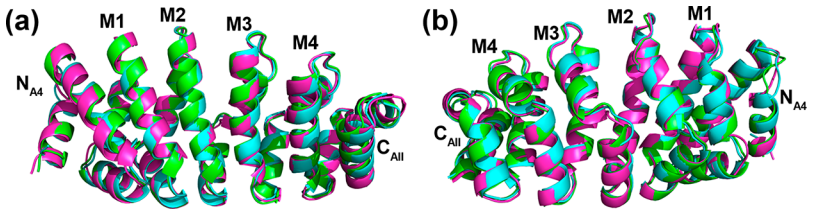


Figure 7. PCS-derived solution structures of $N_{A4}M_4C_{AII}$ (PDB entry 7R0R). (a) Front and (b) back views of a superposition of three PCS-derived NMR solution structures obtained from different starting models. All $N_{A4}M_4C_{AII}$ solution structures reveal N_{A4}-cap conformations that are closely packed against the internal M repeat.

the binding affinity of dArmRPs, containing either the novel N-caps or the original N_{YIII}-cap, four internal M repeats, and the C_{AII}-cap, toward the $(KR)_5$ peptide (Table S7). The obtained results show similar K_d values between 22 and 49 nM for all tested combinations. In particular, the constructs with the well-characterized N_{A4}- and N_{YIII}-caps yield K_d values of 30.5 ± 2.3 and 36.1 ± 2.9 nM, respectively. This suggests that the novel caps do not significantly impact peptide binding, which is one of the desired features of N-caps.

Solution Structure of $N_{A4}M_4C_{AII}$. NMR studies of dArmRPs containing the N_{YIII}-cap proved to be difficult due to the low stability of the N_{YIII}-cap. The recent PCS-derived NMR structure calculation of $N_{YIII}M_4C_{AII}$ revealed multiple conformations of the N_{YIII}-cap, containing contributions from a detachment of fluctuating N_{YIII}-caps from the first internal M repeat,[18] which complicated the structure determination process. As a first application of the new N_{A4}-cap and to assess whether the new N_{A4}-cap facilitates NMR studies, we determined the solution structure of the $N_{A4}M_4C_{AII}$ protein using a combination of NOE- and PCS-derived distance constraints as recently described.[18] The obtained set of three $N_{A4}M_4C_{AII}$ solution structures (PDB entry 7R0R) superimposes with a root-mean-square deviation (RMSD) of 0.39 ± 0.24 Å, indicating good convergence in the structure calculation, and with an RMSD of 1.63 ± 0.17 Å to the $N_{A4}M_4C_{AII}$ crystal structure. In stark contrast to the solution structure of $N_{YIII}M_4C_{AII}$, the PCS-refined structure calculation of the $N_{A4}M_4C_{AII}$ protein provides conformations in which the N_{A4}-cap is firmly packed against the M repeat (Figure 7). Large conformational fluctuations of the N_{A4}-cap are absent, which further highlights the improved stability and overall properties of the novel N_{A4}-cap.

CONCLUSION

We describe here the stabilization of the N-capping repeat of dArmRPs by employing a combination of consensus and computational protein design. The original N_{YIII}-cap was shown to be susceptible to aggregation and degradation, even though NMR analysis of the N_{YIII}-cap did not show any obvious indications for an unstable capping repeat. However, hydrogen exchange experiments revealed a very small but significant transient population of states with unfolded helices in the N_{YIII}-cap, which provide the molecular basis for aggregation and degradation. We decided to employ a previously engineered internal M repeat, obtained from consensus design, as a structural template for a computational optimization using the Rosetta software. Most residues within the hydrophobic core did not require optimization, but most surface-exposed residues were changed during *in silico* design. This optimization resulted in very large stability improvements in GdnHCl-induced equilibrium unfolding, which were ∼5-fold larger than all gains combined from previous engineering efforts. We could furthermore demonstrate that these novel N-caps display a >100-fold reduction in the populations of unfolded states, which provides the basis for the elimination of the previously observed aggregation and degradation propensities. The determined crystal structure of the $N_{A4}M_4C_{AII}$ protein indicated tighter packing of the novel N_{A4}-cap to the first internal M repeat, which provided structural evidence for the improved stability of dArmRPs containing the new N_{A4}-cap. We used the new N_{A4}-cap to determine the solution structure of $N_{A4}M_4C_{AII}$, which, in contrast to the previously determined solution structure of $N_{YIII}M_4C_{AII}$, shows good convergence and a well-packed N_{A4}-cap. This work demonstrates that combining consensus and computational protein design is a very powerful approach for improving protein stability.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.2c00083.

Cloning of target genes, protein expression and purification, NMR spectra, secondary structure analysis from chemical shift indices, correlation between Rosetta energies and experimental protein stabilities, NMR analysis of long-term protein stability, cloning of target genes and expression plasmids, sequences of oligonucleotide primers, X-ray data collection and refinement statistics of the $N_{A4}M_4C_{AII}$:lysozyme structure, computational scanning mutagenesis of N_{H23}-cap, Rosetta energy differences between N_{YIII}- and N_{A4}-caps, and affinities of NM_4C proteins for the $(KR)_5$ peptide (PDF)

Accession Codes

The NMR assignments for the $N_{YIII}MC_{AII}$ and $N_{A4}MC_{AII}$ proteins were deposited at the Biological Magnetic Resonance Bank (BMRB) as entries 51239 and 51240, respectively. The NMR assignments and pseudocontact shifts obtained for the three different cysteine variants of the $N_{A4}M_4C_{AII}$ proteins were deposited as BMRB entries 51290−51292. The X-ray and solution structures of the $N_{A4}M_4C_{AII}$ protein were deposited as PDB entries 7QNP and 7R0R, respectively.

AUTHOR INFORMATION

Corresponding Authors

Erich Michel − *Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland*; Email: e.michel@bioc.uzh.ch

Andreas Plückthun − *Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland*; orcid.org/0000-0003-4191-5306; Email: plueckthun@bioc.uzh.ch

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX

Authors

Stefano Cucuzza − *Department of Chemistry, University of Zurich, CH-8057 Zurich, Switzerland;* Present Address: Technical Research and Development, Global Drug Development, Novartis Pharma AG, Klybeckstrasse 141, CH-4057 Basel, Switzerland

Peer R. E. Mittl − *Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland*

Oliver Zerbe − *Department of Chemistry, University of Zurich, CH-8057 Zurich, Switzerland;* orcid.org/0000-0003-0475-438X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.2c00083

Author Contributions

E.M.: conceptualization, methodology, formal analysis and investigation of biochemical and biophysical experiments, *in silico* protein design and crystal structure determination, writing of the original draft, review and editing, and visualization. S.C.: methodology, formal analysis and investigation of solution structure determination, and review and editing. P.R.E.M.: formal analysis and investigation of crystal structure determination and review and editing. O.Z.: review and editing and funding acquisition. A.P.: review and editing and funding acquisition. All authors have reviewed and given approval to the final version of the manuscript.

Funding

This work was funded by Schweizerische Nationalfonds Grant 40B2-0_176535 and European Union FET-Open Grant PRe-ART (both to A.P.) and Stiftung für wissenschaftliche Forschung der Universität Zürich Grant F-73539-02-01 (to O.Z.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Cornel Fässler and Daniel Burla for their contributions in the initial phase of the project and Beat Blattmann from the Protein Crystallization Center at the University of Zurich for his help with setting up crystallization screens.

■ ABBREVIATIONS

nArmRP, natural Armadillo repeat protein; dArmRP, designed Armadillo repeat protein; NMR, nuclear magnetic resonance; CD, circular dichroism; GdnHCl, guanidine hydrochloride; HX, hydrogen exchange; PCS, pseudocontact shift.

■ REFERENCES

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

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

(3) Peifer, M.; Berg, S.; Reynolds, A. B. A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* **1994**, *76*, 789−791.

(4) Hatzfeld, M. The armadillo family of structural proteins. *Int. Rev. Cytol.* **1998**, *186*, 179−224.

(5) Huber, A. H.; Nelson, W. J.; Weis, W. I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **1997**, *90*, 871−882.

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

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

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

(9) Interlandi, G.; Wetzel, S. K.; Settanni, G.; Plückthun, A.; Caflisch, A. Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. *J. Mol. Biol.* **2008**, *375*, 837−854.

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

(11) Madhurantakam, C.; Varadamsetty, G.; Grütter, M. G.; Plückthun, A.; Mittl, P. R. E. Structure-based optimization of designed Armadillo-repeat proteins. *Protein Sci.* **2012**, *21*, 1015−1028.

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

(13) Wishart, D. S.; Sykes, B. D. 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* **1994**, *4*, 171−180.

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

(15) Keller, R. *The Computer Aided Resonance Assignment Tutorial*; Cantina Verlag: Goldau, Switzerland, 2004.

(16) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **1993**, *17*, 75−86.

(17) Damberger, F. F.; Michel, E.; Ishida, Y.; Leal, W. S.; Wüthrich, K. Pheromone discrimination by a pH-tuned polymorphism of the Bombyx mori pheromone-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18680−18685.

(18) Cucuzza, S.; Güntert, P.; Plückthun, A.; Zerbe, O. An automated iterative approach for protein structure refinement using pseudocontact shifts. *J. Biomol. NMR* **2021**, *75*, 319−334.

(19) Pavlovicz, R. E.; Park, H.; DiMaio, F. Efficient consideration of coordinated water molecules improves computational protein-protein and protein-ligand docking discrimination. *PLoS Comput. Biol.* **2020**, *16*, No. e1008103.

(20) Conway, P.; Tyka, M. D.; DiMaio, F.; Konerding, D. E.; Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **2014**, *23*, 47−55.

(21) Nivon, L. G.; Moretti, R.; Baker, D. A Pareto-optimal refinement method for protein design scaffolds. *PLoS One* **2013**, *8*, No. e59004.

(22) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364−1368.

(23) Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **1990**, *8*, 93−98.

(24) Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125−132.

(25) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204−1214.

(26) Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 22−25.

(27) Hansen, S.; Tremmel, D.; Madhurantakam, C.; Reichen, C.; Mittl, P. R. E.; Plückthun, A. Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. *J. Am. Chem. Soc.* **2016**, *138*, 3526−3532.
(28) Murshudov, G. N.; Vagin, A. A.; Lebedev, A.; Wilson, K. S.; Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 247−255.
(29) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(30) Laskowski, R. A.; Moss, D. S.; Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **1993**, *231*, 1049−1067.
(31) Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **1996**, *6*, 18−23.
(32) Michel, E.; Plückthun, A.; Zerbe, O. Peptide-guided assembly of repeat protein fragments. *Angew. Chem., Int. Ed.* **2018**, *57*, 4576−4579.
(33) Michel, E.; Plückthun, A.; Zerbe, O. Peptide binding affinity redistributes preassembled repeat protein fragments. *Biol. Chem.* **2019**, *400*, 395−404.
(34) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; Basanta, B.; Bender, B. J.; Blacklock, K.; Bonet, J.; Boyken, S. E.; Bradley, P.; Bystroff, C.; Conway, P.; Cooper, S.; Correia, B. E.; Coventry, B.; Das, R.; De Jong, R. M.; DiMaio, F.; Dsilva, L.; Dunbrack, R.; Ford, A. S.; Frenz, B.; Fu, D. Y.; Geniesse, C.; Goldschmidt, L.; Gowthaman, R.; Gray, J. J.; Gront, D.; Guffy, S.; Horowitz, S.; Huang, P. S.; Huber, T.; Jacobs, T. M.; Jeliazkov, J. R.; Johnson, D. K.; Kappel, K.; Karanicolas, J.; Khakzad, H.; Khar, K. R.; Khare, S. D.; Khatib, F.; Khramushin, A.; King, I. C.; Kleffner, R.; Koepnick, B.; Kortemme, T.; Kuenze, G.; Kuhlman, B.; Kuroda, D.; Labonte, J. W.; Lai, J. K.; Lapidoth, G.; Leaver-Fay, A.; Lindert, S.; Linsky, T.; London, N.; Lubin, J. H.; Lyskov, S.; Maguire, J.; Malmstrom, L.; Marcos, E.; Marcu, O.; Marze, N. A.; Meiler, J.; Moretti, R.; Mulligan, V. K.; Nerli, S.; Norn, C.; O'Conchuir, S.; Ollikainen, N.; Ovchinnikov, S.; Pacella, M. S.; Pan, X.; Park, H.; Pavlovicz, R. E.; Pethe, M.; Pierce, B. G.; Pilla, K. B.; Raveh, B.; Renfrew, P. D.; Burman, S. S. R.; Rubenstein, A.; Sauer, M. F.; Scheck, A.; Schief, W.; Schueler-Furman, O.; Sedan, Y.; Sevy, A. M.; Sgourakis, N. G.; Shi, L.; Siegel, J. B.; Silva, D. A.; Smith, S.; Song, Y.; Stein, A.; Szegedy, M.; Teets, F. D.; Thyme, S. B.; Wang, R. Y.; Watkins, A.; Zimmerman, L.; Bonneau, R. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **2020**, *17*, 665−680.

https://doi.org/10.1021/acs.biochem.2c00083
Biochemistry XXXX, XXX, XXX−XXX