Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Peptide-Guided Assembly of Repeat Protein Fragments

Erich Michel, Andreas Plückthun, and Oliver Zerbe**

anie_201713377_sm_miscellaneous_information.pdf

Author Contributions

O.Z. Conceptualization: Equal; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Project administration: Lead; Supervision: Equal; Writing—original draft: Equal
E.M. Conceptualization: Equal; Data curation: Lead; Investigation: Lead; Methodology: Equal; Writing—original draft: Equal
A.P. Conceptualization: Supporting; Funding acquisition: Lead; Investigation: Supporting; Project administration: Supporting; Writing—original draft: Supporting.

Supplementary Materials and Methods to Michel et al., Peptide-Guided Assembly of Repeat Protein Fragments from Libraries

Molecular cloning

All restriction enzymes, Phusion® High-Fidelity PCR kit, T4 DNA Ligase and T4 Polynucleotide Kinase required for preparation of target expression vectors were purchased from New England Biolabs and were used according to the recommended procedure of the supplier. All oligonucleotides were obtained from Microsynth (Switzerland) and chemicals were purchased from Carl Roth (Switzerland), unless indicated otherwise.

Cloning of the cell-free expression vector pCFX3BT2:

The annealed and phosphorylated oligonucleotides 3Fwd (5'-TAA GAT CTC GAT CCC GCG AAT TC) and 3Rev (5'-AGC GAA TTC GCG GGA TCG AGA TCT) were inserted into the pUC19 vector (1) using the restriction sites NdeI and SapI to yield the intermediate vector pTx1a. Transfer of the 304-nucleotide BglII/EcoRI fragment from pET14b (Novagen) into pTx1a provided the intermediate vector pTx1b. An internal BsaI restriction site was then removed from pTx1b by site-directed mutagenesis with the oligonucleotide primers 5'BSA_Mut (5'-GCA ATG ATA CCG CGG GAA CCA CGC TCA CCG GC) and 3'BSA_Mut (5'-GCC GGT GAG CGT GGT TCC CGC GGT ATC ATT GC), resulting in the intermediate vector pTx1c. In parallel, the multiple cloning site (MCS) of the previously described vector pCFX3 (2) was modified by site-directed mutagenesis using the oligonucleotide primers 3b_Fwd (5'-CCT GTA CTT CCA GAG CGG CAG AAG AGC CAT GGC AAG TCG ACT CGA GC) and 3b_Rev (5'-GCT CGA GTC GAC TTG CCA TGG CTC TTC TGC CGC TCT GGA AGT ACA GG). Further site-directed mutagenesis of this vector with the oligonucleotide primers 3B3_Fwd (5'-CTG TAC TTC CAG AGA AGA GCC ATG G) and 3B3_Rev (5'-CCA TGG CTC TTC TCT GGA AGT ACA G) provided the intermediate vector pCFX3B3 containing a MCS with a SapI site that generates a CAG overhang within the glutamine of the TEV protease cleavage site. The 395-nucleotide fragment obtained by BglII and BamHI cleavage of pCFX3B3 was then transferred into pTx1c to yield the vector pCFX3BT. Site-directed mutagenesis of pCFX3BT with the oligonucleotide primers BsaI_Fwd (5'-GAC TCA CTA TAG GGA GAC GAC AAC CGT TTC CCT CTA G) and BsaI_Rev (5'-CTA GAG GGA AAC GGT TGT CGT CTC CCT ATA GTG AGT C) resulted in the final vector pCFX3BT2 for cell-free expression.

Cloning of the *E. coli* expression vector pEM3BT2:

The 346-nucleotide XbaI/BamHI fragment from pCFX3 (2) was transferred into pET19b (Novagen) to yield the vector pEM1. Site-directed mutagenesis of pEM1 with the oligonucleotide primers 5'BSA_Mut (5'-GCA ATG ATA CCG CGG GAA CCA CGC TCA CCG GC) and 3'BSA_Mut (5'-GCC GGT GAG CGT GGT TCC CGC GGT ATC ATT GC) provided the vector pEM1B devoid of internal BsaI restriction sites. Removal of the internal SapI site of pEM1B by site-directed mutagenesis using the oligonucleotide primers SapI_RM_Fwd (GAG CGA GGA AGC GGA TTT GCG CCT GAT GCG GTA TTT TC) and SapI_RM_Rev (GAA AAT ACC GCA TCA GGC GCA AAT CCG CTT CCT CGC TC) resulted in the vector pEM1C. The final vector pEM3BT2 for recombinant expression in *E. coli* was obtained by transfer of the 337-nucleotide XbaI/BamHI fragment from pCFX3BT2 into pEM1C.

Cloning of target genes:

The target dArmRP genes were PCR-amplified from a custom-synthesized gene encoding YM_3A (GenScript; Table S1) in pCFX3BT2. Unless indicated otherwise, all genes were subcloned into either pCFX3BT2 or pEM3BT2 using the SapI and BamHI sites. The constructs containing individual or combined mutations of Glu, Asn and/or Trp to Ala were prepared by site-directed mutagenesis using the primers indicated in Table S2. The $(KR)_4$-fragment fusion constructs were prepared by insertion of annealed and phosphorylated oligonucleotides encoding $(KR)_4$-$(GS)_4$ into SapI/BamHI-digested pEM3BT2, followed by subsequent insertion of the PCR product encoding either YM_2 or YM_2H1 into the BamHI restriction site.

Protein expression and purification

Unlabeled or uniformly labeled proteins were expressed in *E. coli* BL21 (DE3) cells (Stratagene) growing at 37°C in 500 mL LB or M9 medium (3), respectively. Amino acid-specific isotope labeling was achieved by cell-free protein expression using an *E. coli*-based S30 cell extract (2) according to a previously described protocol (4). All proteins were expressed as fusion constructs with a TEV protease-cleavable N-terminal $(His)_6$-GB1 domain (5). The cell cultures were harvested by centrifugation for 10 min at 5000 × g and 4 °C and the obtained cell pellet was re-suspended in 15 mL buffer A (50 mM sodium phosphate at pH 7.7, 500 mM sodium chloride, 20 mM imidazole, 30 µM sodium azide) on ice. The re-

suspended cells were disrupted in a single passage through a French Press (Thermo Electron Corporation) at 1100 psi pressure and 4°C, and the obtained lysate was mixed with ca. 1 mg of DNaseI (Roche, Switzerland) and cleared by centrifugation for 30 min at 30000 × g and 4°C. The supernatant was filtered through a Filtropur S 0.2 µm sterile filter (Sarstedt, Germany) and was passed over a 5 mL HisTrap HP column (GE Healthcare) in buffer A, which was mounted on an ÄKTA purifier FPLC system (GE Healthcare) equipped with A_{280}, A_{260} and conductivity monitors. After washing with 15 column volumes of buffer A, the target proteins were eluted in a linear gradient of 20–500 mM imidazole in 100 mL buffer A. The fractions with significant absorption at 280 nm were collected and analyzed by SDS-PAGE. The pooled fractions containing the desired target protein were then mixed with 2 mg TEV protease, prepared as previously described (2), and dialyzed overnight at room temperature (RT) in a 3.5 kDa MWCO dialysis membrane against 2 L TEV cleavage buffer (50 mM sodium phosphate at pH 7.7, 100 mM sodium chloride, 0.5 mM DTT, 25 µM EDTA and 30 µM sodium azide). The proteolytically cleaved protein solution was then again passed over the 5 mL HisTrap HP column in buffer A and the eluate containing the desired target protein devoid of the N-terminal $(His)_6$-GB1 domain was collected for two consecutive dialysis steps at RT in a 3.5 kDa MWCO dialysis membrane (Carl Roth, Switzerland), each for 8 h against 2 L of fresh NMR buffer (20 mM sodium phosphate at pH 7.0, 50 mM sodium chloride, 30 µM sodium azide). The protein solution was then concentrated in a 3 kDa MWCO Amicon Ultra-15 centrifugal filter devices (Millipore) at 3500 × g and 16 °C until the desired concentration was obtained.

Size-exclusion chromatography

Analytical size-exclusion chromatography of the dArmRP fragments was conducted at room temperature in NMR buffer on a Superdex 200 Increase 10/300 GL column (GE Healthcare). In each run, a total of 100 µL of 10 µM N-terminal dArmRP fragment with varying concentrations from 2.5–20 µM of the C-terminal interaction partner was applied onto the column with a flow rate of 0.6 mL/min, monitoring the absorbance at 280 nm.

Isothermal Titration Calorimetry

ITC experiments to characterize the association of complementary dArmRP fragments and their interaction with $(KR)_n$-peptides were measured on a VP-ITC instrument (MicroCal Inc., Northampton, MA). The experimental setup comprised 1.4 mL of 5–12.5 µM N-terminal

fragments in the sample cell and 60–220 μM of the C-terminal fragments in the syringe. Both fragments were extensively dialyzed against the same batch of ITC buffer (20 mM sodium phosphate at pH 7.0, 200 mM sodium chloride, 100 μM EDTA, and 30 μM sodium azide). Titrations were conducted at 25 °C with a stirring rate of 300 rpm, and contained 29 injections each with 10 μL volume and 10 seconds duration, and 240 seconds spacing between each injection. The raw data was integrated, corrected for non-specific heats, normalized for concentration and analyzed using the MicroCal Origin software.

NMR spectroscopy and assignments

NMR experiments were recorded with protein solutions of 450 μL volume containing 6% (v/v) D_2O in NMR buffer at 310 K on Bruker Avance 600 and 700 spectrometers equipped with cryogenic triple-resonance probes. The backbone resonances of H23MA were assigned based on the 3D HNCO, 3D HN(CA)CO, 3D HNCA, 3D HNCACB, 3D CBCA(CO)NH and 3D [^{15}N]-resolved [^{1}H,^{1}H]-NOESY experiments (6), which were recorded with a sample of 0.8 mM uniformly [^{13}C,^{15}N]-labeled H23MA. Due to the repetitive nature of the protein sequence many amino acids from different modules exhibited close-to-identical chemical shifts. An important point is that carbonyl chemical shifts were usually sufficiently different to allow unambiguous sequential assignments, while C_α or C_β shifts were often nearly identical. In addition, unique peaks were usually observed in the ^{15}N,^{1}H correlation, which served as anchoring points for the sequence-specific resonance assignment according to previously described approaches (7, 8). We obtained near-complete sequence-specific resonance assignments with the exception of N159 and N201, for which amide resonances were broadened beyond detection. The secondary structure of the H23MA fragment in the free and bound state was determined by analysis of the C_α and C' shifts according to the chemicals shift index protocol (9). The backbone dynamics of the free H23MA fragment were analyzed using the heteronuclear 2D ^{15}N-{^{1}H}-NOE experiment (10, 11).

NMR quantification of fragment populations

The populations of the free and assembled C-terminal fragments were determined by integration of characteristic groups of peaks for each fragment variant in 2D [^{15}N,^{1}H]-HSQC spectra that shift at least two full widths at half height between the free and the assembled state. These peak groups included the amide resonances of L135, L136, L148 and L151 for the [2-EWN]-H23MA fragment, A131, A147, A150, A155, A164 and A168 for the [2-W]-

H23MA fragment and W149 and W191 for the wt-H23MA fragment. All 2D [^{15}N,^{1}H]-HSQC spectra were recorded and processed with identical parameters to provide a reliable and comparable peak integration. A reference 2D [^{15}N,^{1}H]-HSQC spectrum of the mixed three C-terminal variants (wt-H23MA, [2-W]-H23MA, [2-EWN]-H23MA) was measured before the addition of the N-terminal $(KR)_4$-YM_2H1 fragment. NMR titrations were then performed by stepwise addition of 0.5 equivalents of $(KR)_4$-YM_2H1 (9 nmol) to a mixture containing 18 nmol of each of the three H23MA fragments. The characteristic peaks were then integrated at each titration step, corrected for dilution effects, and were divided by the corresponding peak reference value from the spectrum in absence of the N-terminal fragment. The obtained ratios at each titration step were then averaged within the group of characteristic peaks of each fragment variant, e.g. the leucine peak ratios for the [2-EWN]-H23MA variant, and the group standard deviation was calculated. The sum of the three averaged group ratios, which corresponds to the total free population of each set of C-terminal fragments, was then normalized according to the stoichiometry of N- and C-terminal fragments at a particular titration step, e.g. two thirds of the total available H23MA fragments have to be in the free and one third in the assembled state in presence of one equivalent $(KR)_4$-YM_2H1. The assembled population of a particular C-terminal fragment was finally calculated by subtracting its free population from its total population.

Supplementary Tables

Table S1. DNA Sequence of the custom-designed YM_3A gene

```
GCT CTT CAC AGG GCG AAT TGC CGC AGA TGG TTC AGC AGC TGA ACT CCC CGG
ACC AGC AAG AAC TGC AGT CTG CTC TGC GTA AAC TCT CAC AGA TTG CAA GCG
GTG GCA ATG AGC AAA TCC AAG CCG TGA TCG ACG CCG GTG CTC TGC CTG CTT
TGG TGC AGC TCT TAA GCA GTC CAA ATG AGC AGA TCC TTC AGG AAG CCT TGT
GGG CAC TGA GTA ACA TCG CTT CCG GAG GTA ATG AAC AGA TCC AGG CTG TTA
TTG ACG CTG GCG CCC TTC CAG CCC TCG TCC AAT TGC TTT CGT CAC CTA ACG
AAC AGA TTC TCC AAG AGG CTC TCT GGG CGT TAA GCA ATA TCG CGA GTG GCG
GTA ACG AGC AGA TTC AGG CAG TCA TCG ATG CAG GGG CGT TAC CCG CAT TAG
TAC AAC TTC TCA GTT CGC CCA ATG AAC AAA TCT TGC AAG AAG CGC TTT GGG
CTC TTT CTA ACA TTG CCT CTG GTG GGA ACG AAC AAA AAC AAG CGG TAA AAG
AAG CGG GAG CAC TCG AGA AAC TGG AAC AGC TGC AGT CAC ACG AGA ACG AGA
AGA TTC AGA AAG AGG CAC AGG AAG CCT TGG AGA AGC TGC AGT CCC ACT AAT
AAG GAT CC
```

Table S2. Oligonucleotide primer for cloning of the target proteins

Protein	5'-oligonucleotide and 3'-oligonucleotide primer
YM_3A	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-GCT TTG TTA GCA GCC GGA TC
YM_2	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-TTT GGA TCC TTA TTA GCC ACT CGC GAT ATT GCT TAA C
M_2A	5'-AAA GCTCTTC A CAG GGT AAT GAA CAG ATC CAG GCT G 5'-GCT TTG TTA GCA GCC GGA TC
YM_2H1	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-TTT GGA TCC TTA TGC ATC GAT GAC TGC CTG AAT C
H23MA	5'-AAA GCTCTTC A CAG GGC GCC CTT CCA GCC C 5'-GCT TTG TTA GCA GCC GGA TC
YM_2H12	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-TTT GGA TCC TTA ACT GAG AAG TTG TAC TAA TGC GG
H3MA	5'-AAA GCTCTTC A CAG TCA CCT AAC GAA CAG ATT CTC C 5'-GCT TTG TTA GCA GCC GGA TC
[3-EWN]-YM_3A	5'-GCA AGC AGC GCT TGC GGC TCT TTC TGC CAT TG 5'-CAA TGG CAG AAA GAG CCG CAA GCG CTG CTT GC
[3-W]-YM_3A	5'-AAA GCTCTTC A GCG GC TCT TTC TAA CAT TGC CTC 5'-AAA GCT CTT CAC GCA AGC GCT TCT TGC AAG ATT TGT TC
$(KR)_4$-$(GS)_4$G – YM_2H1	5'-AAA GGATCC GGC GAA TTG CCG CAG 5'-TTT GGA TCC TTA TGC ATC GAT GAC TGC CTG AAT C

Supplementary References

1. Norrander J, Kempe T, & Messing J (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26(1):101–106.
2. Michel E & Wüthrich K (2012) High-yield *Escherichia coli*-based cell-free expression of human proteins. *J. Biomol. NMR* 53(1):43–51.
3. Michel E & Allain FH (2015) Selective amino acid segmental labeling of multi-domain proteins. *Methods Enzymol.* 565:389–422.
4. Michel E, Skrisovska L, Wüthrich K, & Allain FH (2013) Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. *ChemBioChem* 14(4):457–466.
5. Michel E & Wüthrich K (2012) Cell-free expression of disulfide-containing eukaryotic proteins for structural biology. *FEBS J.* 279(17):3176–3184.
6. Sattler M, Schleucher J, & Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34:93–158.
7. Wetzel SK*, et al.* (2010) Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. *J. Mol. Biol.* 402(1):241–258.
8. Ewald C*, et al.* (2015) A combined NMR and computational approach to investigate peptide binding to a designed Armadillo repeat protein. *J. Mol. Biol.* 427(10):1916–1933.
9. Wishart DS & Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR* 4(2):171–180.
10. Kay LE, Torchia DA, & Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28(23):8972–8979.
11. Noggle JH & Schirmer RE (1971) *The Nuclear Overhauser Effect: Chemical Applications* (Academic Press, New York).

Supplementary Figures

Figure S1

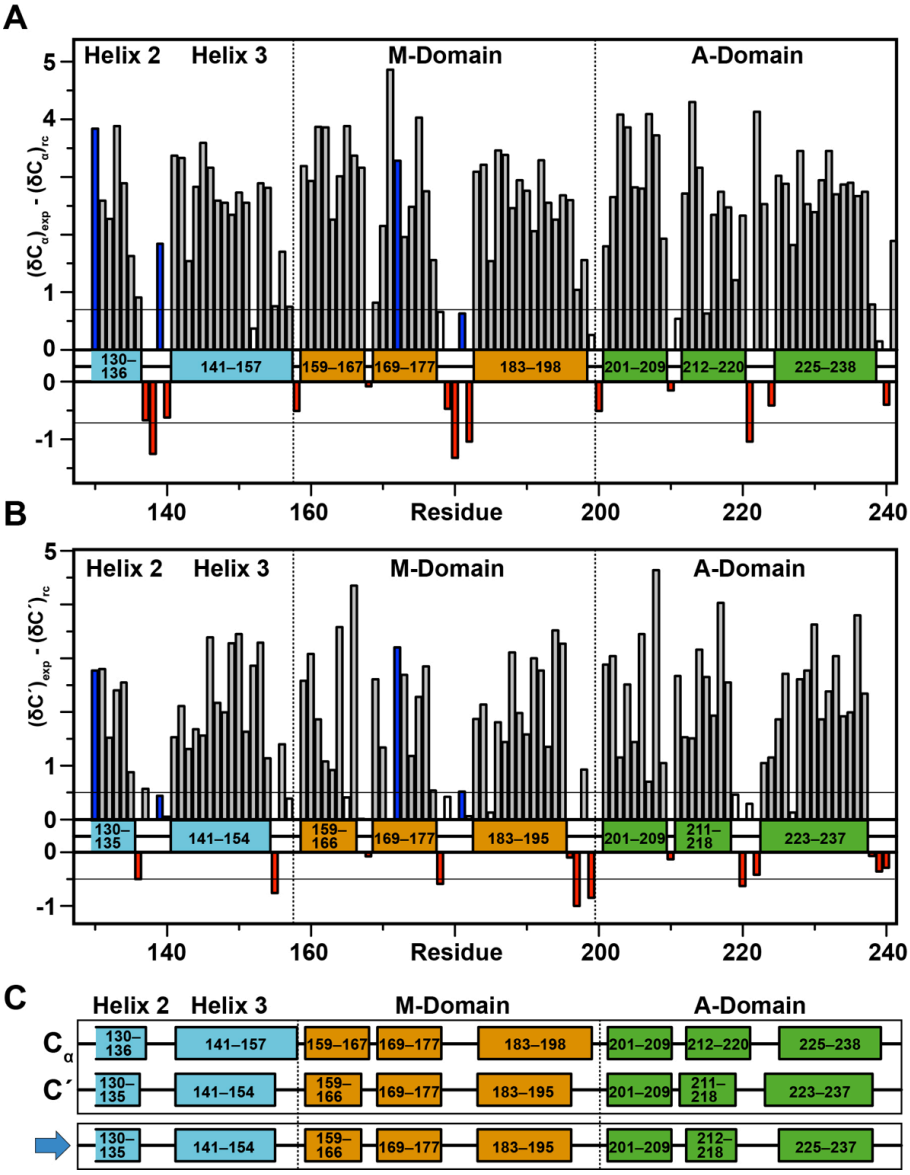

Fig. S1: Secondary structure of the free H23MA dArmRP fragment derived from chemical shift indices. Secondary chemical shifts are indicated for the assigned Cα (A) and C' (B) spins of free H23MA. Red bars correspond to residues with chemical shift index values that are incompatible with the presence of an α-helix while blue bars indicate proline residues. The lines at ordinate values of 0.7 (A) or 0.5 (B) indicate the cut-off to identify helical residues from Cα and C' chemical shifts, respectively. Residues that form a regular α-helix are schematically represented by colored boxes. (C) The consensus secondary structure (blue arrow) is obtained by combination of those helical regions that are supported by both the Cα and C' shifts.

Figure S2

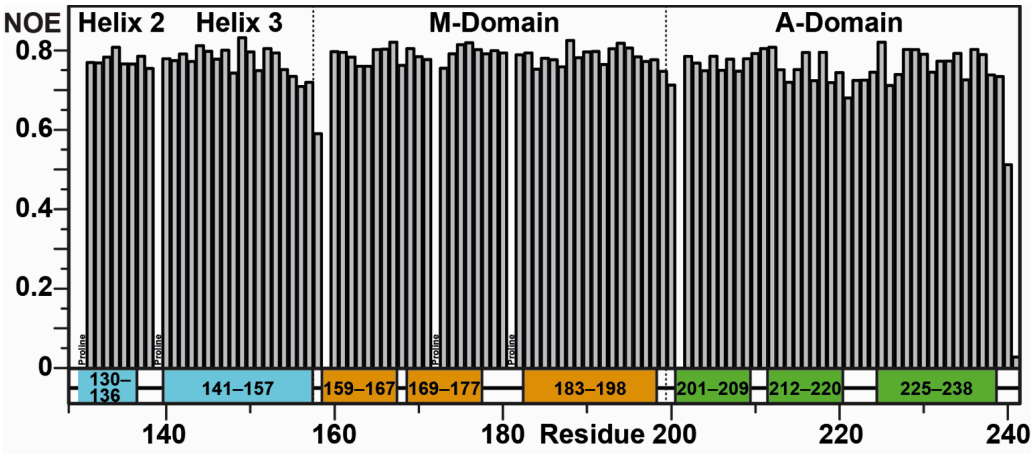

Fig. S2: Analysis of backbone dynamics for free H23MA. The mobility of each assigned amide bond was determined by measuring the heteronuclear 2D $^{15}N\{^{1}H\}$-NOE experiment recorded at 310 K on a 600 MHz spectrometer. The experimentally obtained NOE values for each residue are plotted against the primary and secondary structure of H23MA. The majority of residues display NOE values in the range of 0.7–0.8, which are indicative of a rigid backbone.

Figure S3

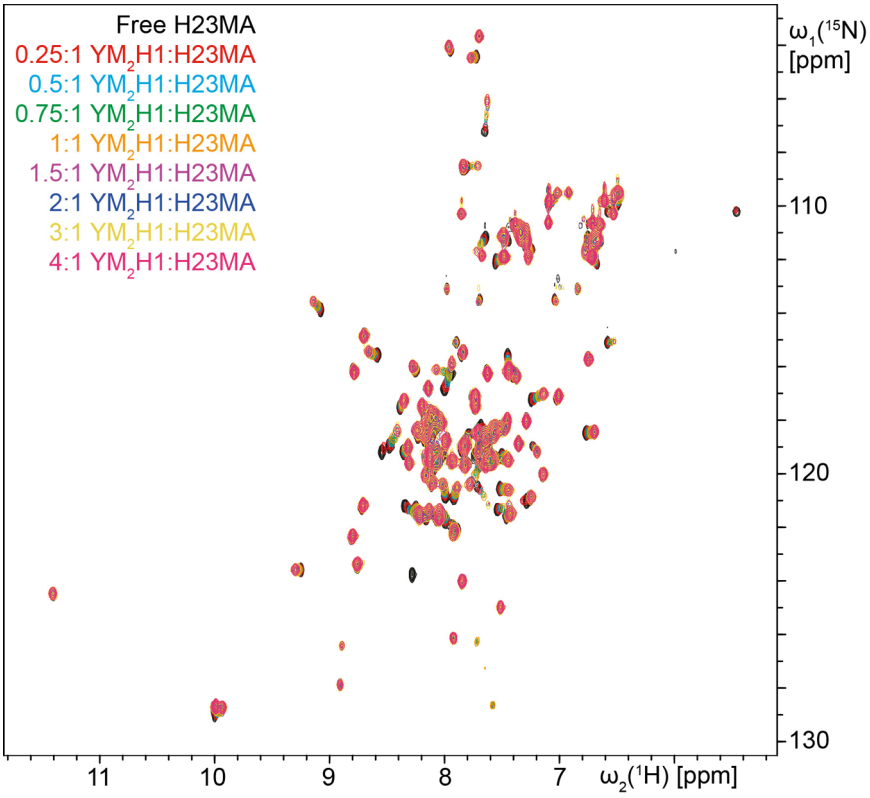

Fig. S3: NMR titration of H23MA with YM_2H1. Superposition of 2D [^{15}N,^{1}H]-HSQC spectra of 450 µM [^{15}N]-labeled H23MA titrated with unlabeled YM_2H1. Peaks in each titration step are color-coded as indicated.

Figure S4

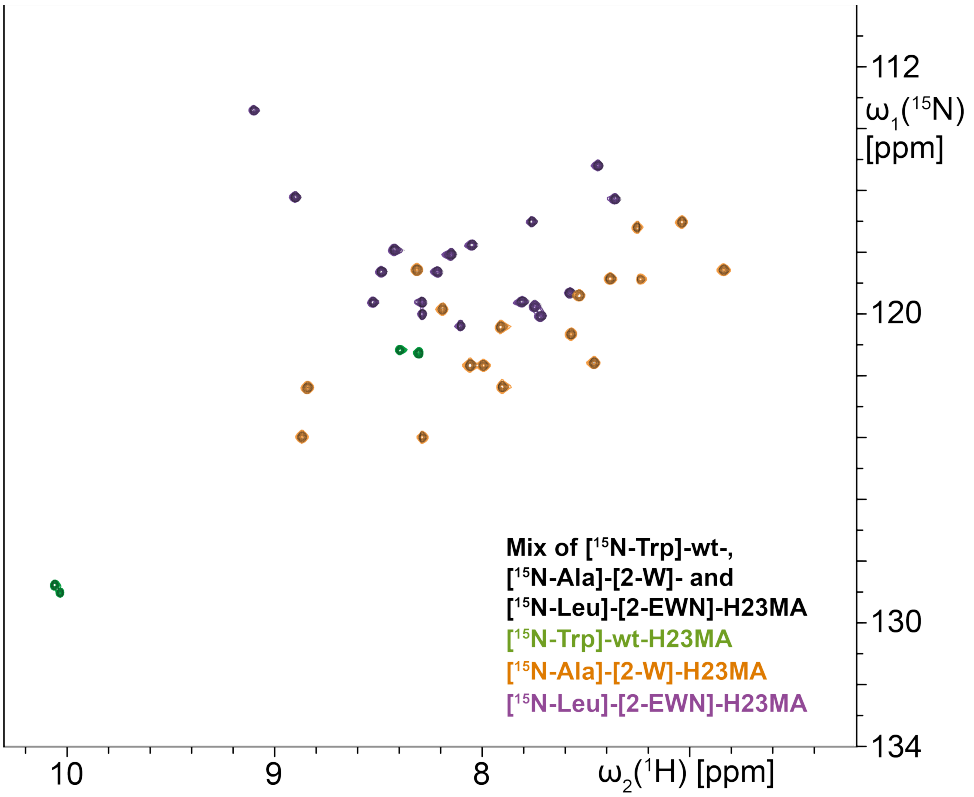

Fig. S4: Superposition of spectra from differently labeled H23MA. Superposition of 2D [^{15}N,^{1}H]-HSQC spectra of [^{15}N-Trp]-wt-H23MA, [^{15}N-Ala]-[2-W]-H23MA and [^{15}N-Leu]-[2-EWN]-H23MA measured either individually or in a mixture containing equimolar amounts of all three proteins. The individual spectra are represented in different colors as indicated.