Supplementary Materials and Methods

Michel et al., **Peptide binding affinity redistributes preassembled repeat protein fragments**

Cloning of target genes:

All enzymes required for generation of the target protein expression vectors were purchased from New England Biolabs (Ipswich, MA, USA) and were used following the suppliers' recommendations. DNA oligonucleotides were purchased from Microsynth (Arlesheim, Switzerland) and chemicals from Carl Roth (Balgach, Switzerland), unless indicated otherwise. The required dArmRP gene constructs were amplified from a custom-synthesized gene encoding YM_3A (GenScript, Piscataway, NJ, USA) (1) by PCR using the oligonucleotide primers listed in Table S6 and were subcloned into the SapI and BamHI sites of either the vector pCFX3BT2 for cell-free expression or the vector pEM3BT2 for recombinant protein expression in *E. coli* (1). The genes encoding wt-YM_5A, [5-EN]-YM_5A and [4,5-EN]-YM_5A were prepared by ligation of a SapI-digested PCR product encoding YM_2 with the corresponding SapI/BamHI-digested M_3A variant into pEM3BT2. The mutations of Glu and Asn to Ala were subsequently prepared by site-directed mutagenesis using the primers indicated in Table S6. The GB1-$(KR)_4$ construct was prepared by ligation of annealed and phosphorylated oligonucleotides (Table S6) into the SapI/BamHI-digested pEM3BT2-1D vector (see below). The fusion construct of the $(KR)_5$ peptide with GFP for the fluorescence anisotropy measurements was prepared as previously described (2).

Cloning of pEM3BT2-1D:

The vector pEM3BT2-1D for expression of the GB1-$(KR)_4$ fusion construct was prepared as follows. The cell-free expression vector pCFX3BT2 (1) served as a template for a site-directed mutagenesis with the oligonucleotide primers pCFX15C_Fwd (CAT CAT CAT CAT AGC AGC GGC GAA AAC CTG TAC) and pCFX15C_Rev (GTA CAG GTT TTC GCC GCT GCT ATG ATG ATG ATG) to yield the vector pCFX15C, which contains a TEV protease cleavage site directly after the N-terminal His-tag. The GB1 domain was then PCR-amplified from pCFX3BT2 using the oligonucleotide primers pCFX1C_Fwd (AAA GCT CTT CAC AGA GCG GAC AGT ACA AAC TGA TCC TGA ACG) and pCFX1C_Rev (TTT GGA TCC GCT GCT TTC GGT AAC) and inserted into pCFX15C using the SapI and BamHI restriction sites to yield the vector pCFX1C, which was further PCR-amplified with the oligonucleotide primers pCFX1D_Fwd (AAA GCT CTT CAA AGA GCT TTC GGA

TCC GGC TGC TAA C) and pCFX1D_Rev (AAA GCT CTT CAC TTC TGC TGC TTT CGG TAA CGG TG), digested with SapI and ligated to yield the vector pCFX1D. The XbaI/BamHI insert from pCFX1D was then subcloned into pEM3BT2 to yield the final vector pEM3BT2-1D.

Protein purification

The *E. coli* cell cultures were centrifuged after expression (10 min at 5000 *g*, 4°C), and the cell pellet was directly re-suspended on ice with 15 ml buffer A (50 mM sodium phosphate at pH 7.7, 500 mM sodium chloride, 20 mM imidazole, 30 µM sodium azide) containing 1 mg DNaseI (Roche, Basel, Switzerland) and 1 mM $MgCl_2$. The cell suspension was then lysed in a French Press (Thermo Electron Corporation, Waltham, MA, USA) with a single passage at 1100 psi pressure and 4°C. The cell lysate was cleared by centrifugation (30 min at 30000 *g*, 4°C) and the supernatant was passed through a Filtropur S 0.2 µm filter device (Sarstedt, Nümbrecht, Germany) and was applied on a 5 ml HisTrap HP column (GE Healthcare, Chicago, IL, USA) in buffer A. The immobilized target proteins were washed with 15–20 column volumes buffer A and were eluted with a linear gradient of 20–500 mM imidazole in 100 ml buffer A. The fractions containing the desired target protein were identified based on the absorption at 280 nm and SDS-PAGE analysis and were subsequently dialyzed overnight with 2 mg TEV protease 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) at room temperature in a 3.5 kDa MWCO dialysis membrane. The dialyzed protein solution was subsequently re-applied onto the 5 ml HisTrap HP column in buffer A to separate the proteolytically processed target protein without the $(His)_6$-GB1 tag from the $(His)_6$-tagged protein species. The target protein collected from the flow-through was dialyzed two times for 8 h against 2 l of fresh NMR buffer in a 3.5 kDa MWCO dialysis membrane and was finally concentrated in a 3 kDa MWCO Amicon Ultra-15 centrifugal filter devices (Millipore, Burlington, MA, USA) at 3500 *g* and 16°C.

Supplementary Figures

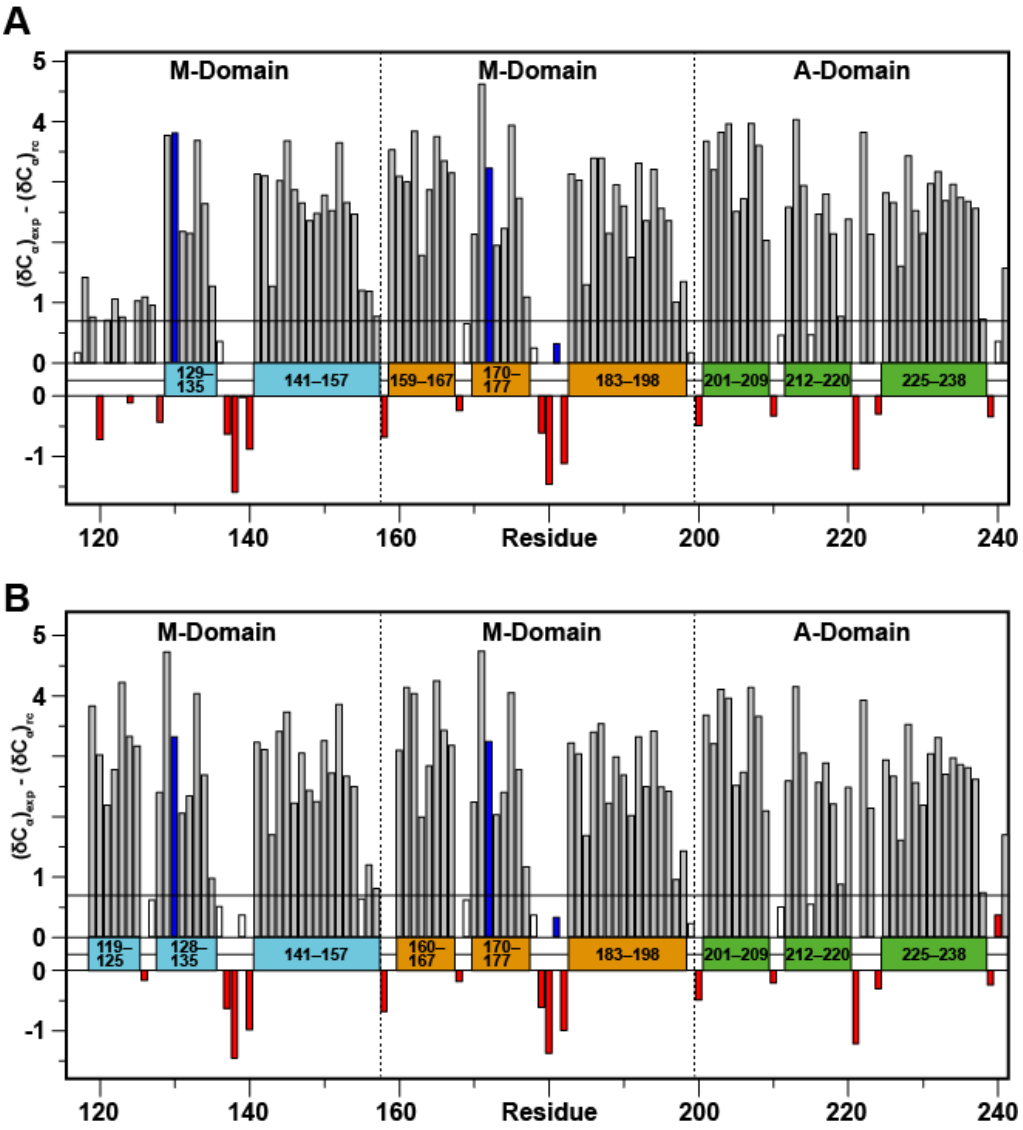


Figure S1: Secondary C_α chemical shift analysis of the free and YMM-bound MMA. The secondary chemical shifts of the assigned C_α spins of the (A) free MMA and (B) of MMA in complex with YMM are plotted against the primary sequence. Grey and red bars indicate residues with chemical shift index values that support or disagree with the presence of an α-helix, respectively, while blue bars indicate proline residues. The line at an ordinate value of 0.7 indicates the cut-off to identify helical residues from C_α chemical shifts. Residues that form a regular α-helix are schematically represented by colored boxes. Residues 118–124 of MMA show no regular secondary structure in the free form and form a helix upon complexation with YMM.

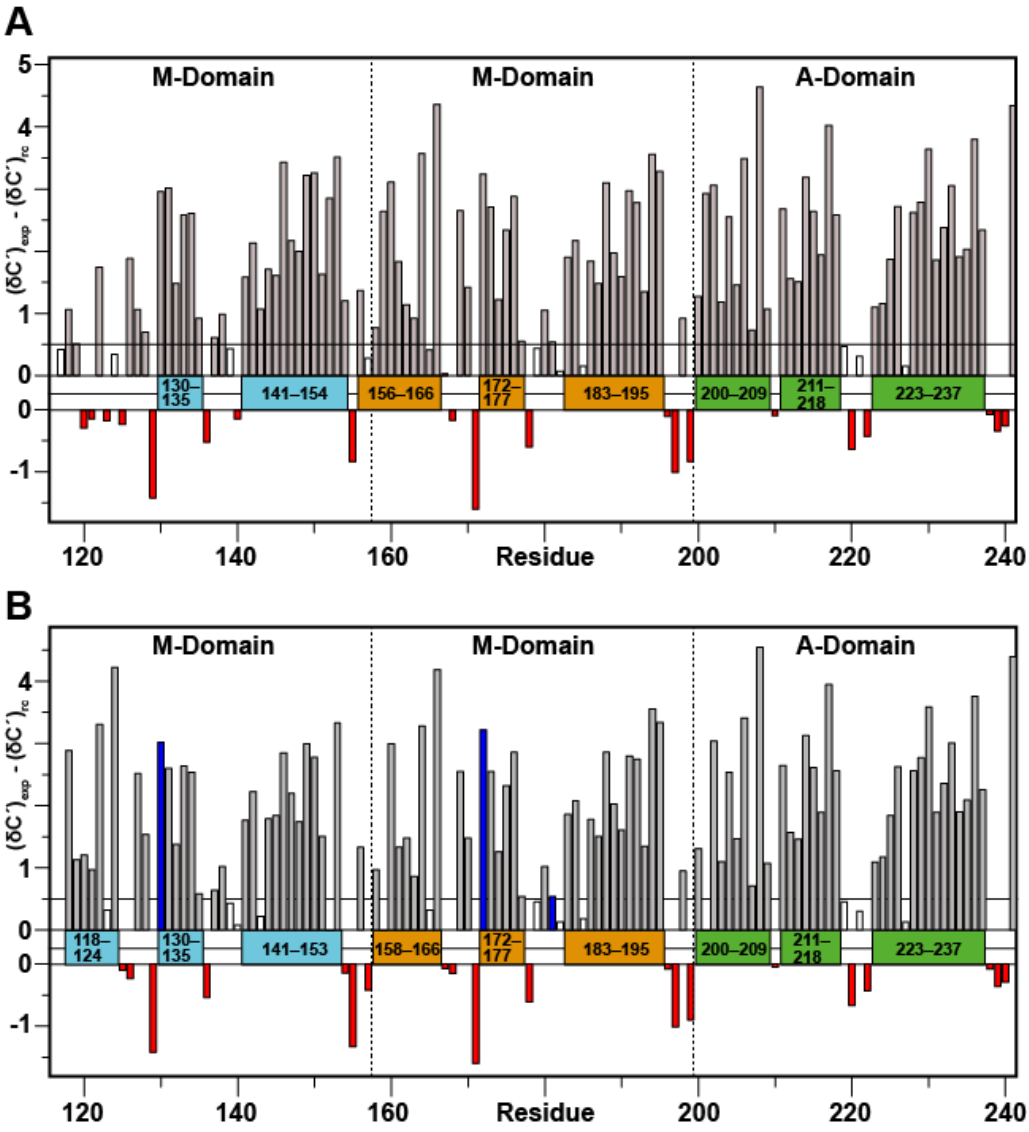
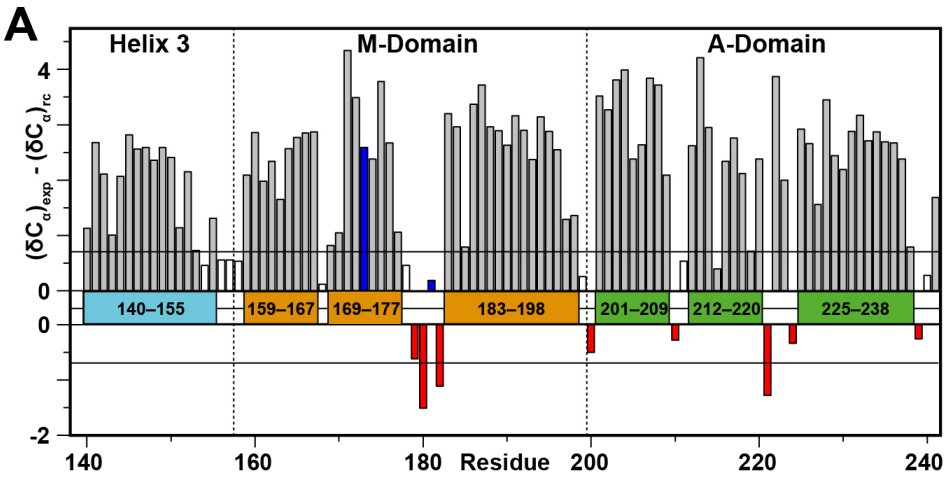
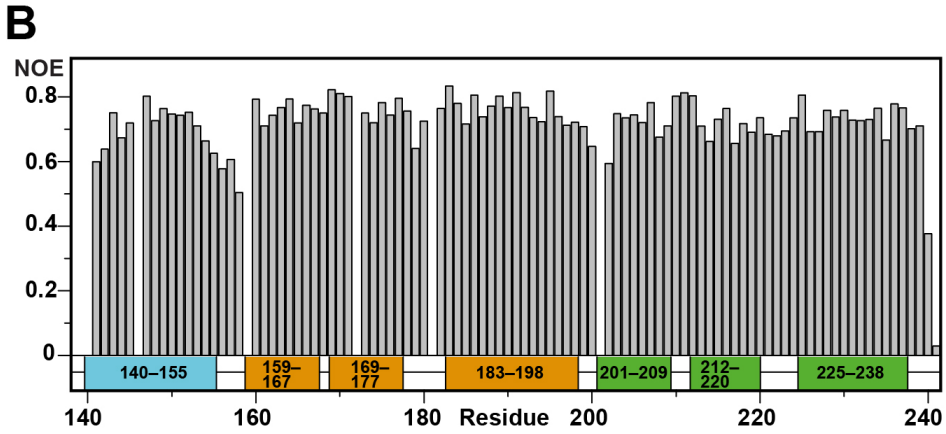


Figure S2: Secondary C' chemical shift analysis of the free and YMM-bound MMA dArmRP fragment. The secondary chemical shifts of the assigned C' spins of the (A) free MMA and (B) of MMA in complex with YMM are plotted against the primary sequence. Grey and red bars indicate residues with chemical shift index values that support or disagree with the presence of an α-helix, respectively, and blue bars indicate proline residues. The line at an ordinate value of 0.5 indicates the cut-off to identify helical residues from C' chemical shifts. Residues that form a regular α-helix according to the C' chemical shift analysis are schematically represented by colored boxes. The C' chemical shift analysis confirms the observation that helix formation in residues 118–124 is induced upon assembly with YMM.


A
Helix 3
M-Domain
A-Domain
$(\delta C_\alpha)_{exp}$ - $(\delta C_\alpha)_{rc}$
4
0
0
-2
140–155
159–167
169–177
183–198
201–209
212–220
225–238
140
160
180
Residue
200
220
240


B
NOE
0.8
0.6
0.4
0.2
0
140–155
159–167
169–177
183–198
201–209
212–220
225–238
140
160
180
Residue
200
220
240

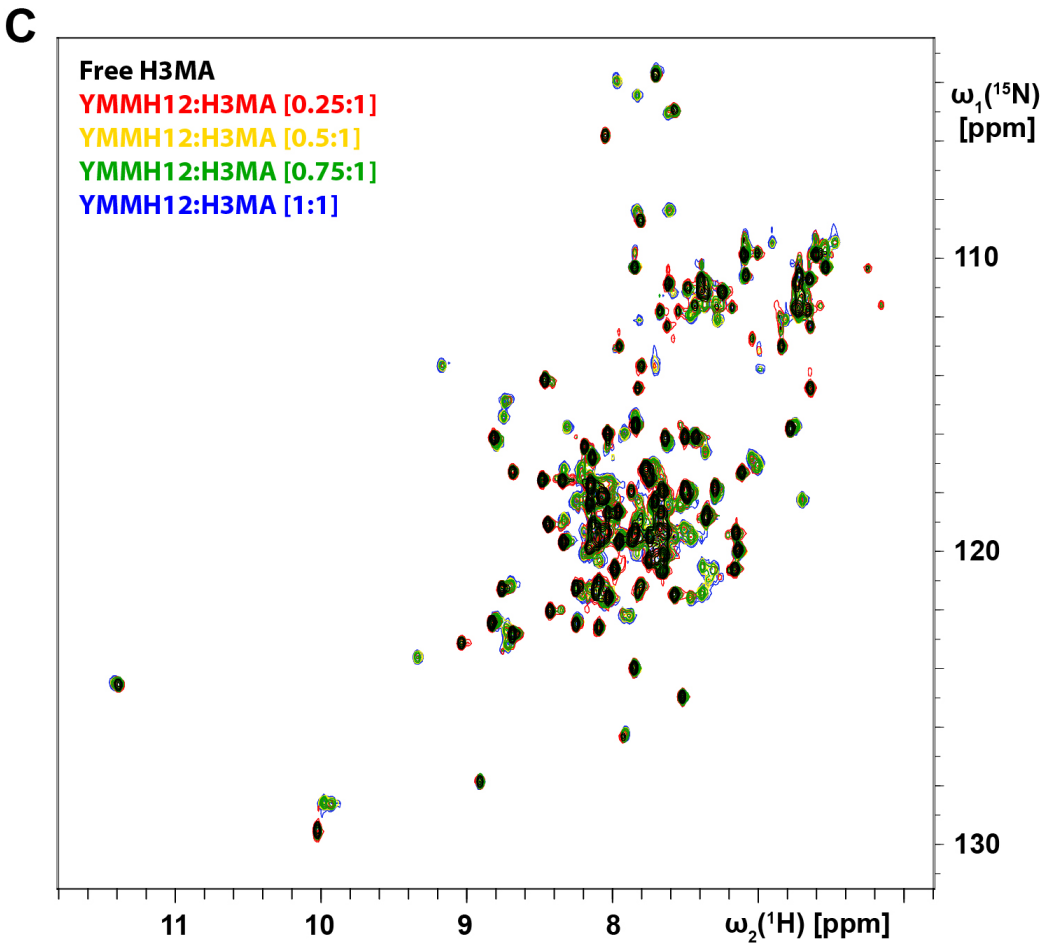

C
Free H3MA
YMMH12:H3MA [0.25:1]
YMMH12:H3MA [0.5:1]
YMMH12:H3MA [0.75:1]
YMMH12:H3MA [1:1]
ω_1(^{15}N) [ppm]
110
120
130
11
10
9
8
ω_2(^{1}H) [ppm]

Figure S3: NMR analysis of free H3MA and in complex with YMMH12. (A) Secondary C_α chemical shift analysis of free H3MA. Grey and red bars indicate residues with chemical shift index values that support or disagree with the presence of an α-helix, respectively, while blue bars indicate proline residues. The secondary chemical shifts indicate helix formation for those residues that also form a helix in full-length dArmRPs. (B) Backbone amide mobility assessment with the heteronuclear 2D $^{15}N\{^1H\}$-NOE experiment. The obtained NOE values for each residue are plotted against the primary and secondary structure of MMA. The majority of residues display NOE values in the range of 0.7–0.8, which indicates a rigid protein backbone. (C) NMR titration of [^{15}N]-labeled H3MA with unlabeled YMMH12. 2D [^{15}N,^1H]-HSQC spectra of the individual titration steps are color-coded as indicated. The titration indicates slow exchange on the NMR timescale and shows no further spectral changes after the addition of one equivalent YMMH12, which agrees with a K_d of 234 nM for the YMMH12:H3MA complementation determined by ITC.

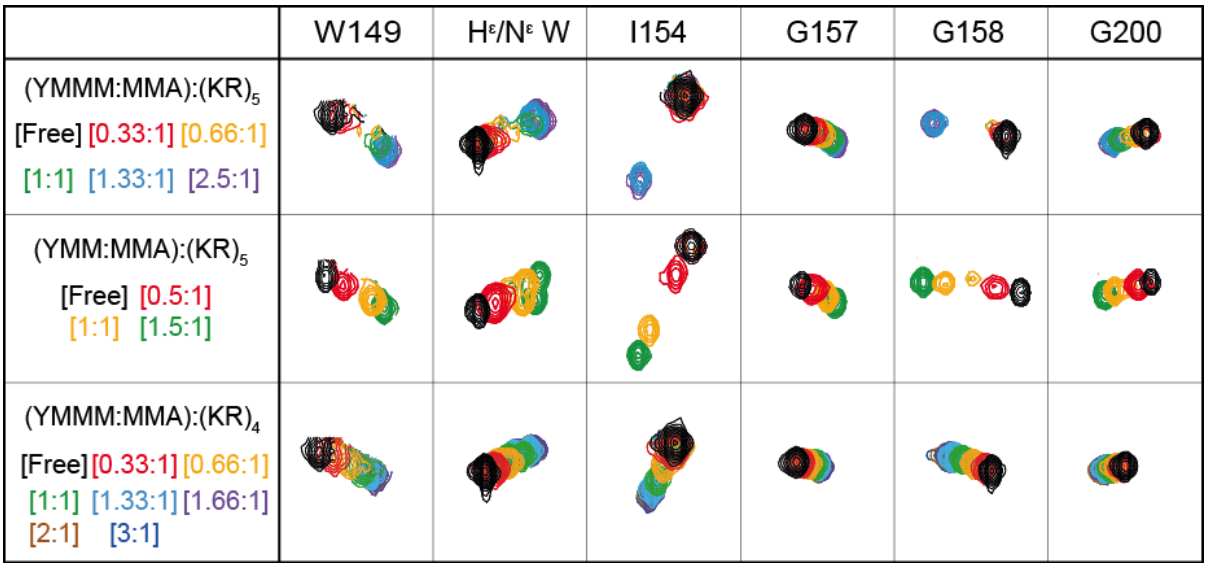


Figure S4: NMR titration analysis of YMMM:MMA and YMM:MMA complexes with either $(KR)_4$ or $(KR)_5$ peptides. Defined regions of the 2D [^{15}N,^1H]-HSQC spectrum with characteristic amide resonances are superimposed for all titration steps, each of which is represented with a different color. "$H^\varepsilon/N^\varepsilon$ W" denotes the indole amide resonances of residues W149 and W191.

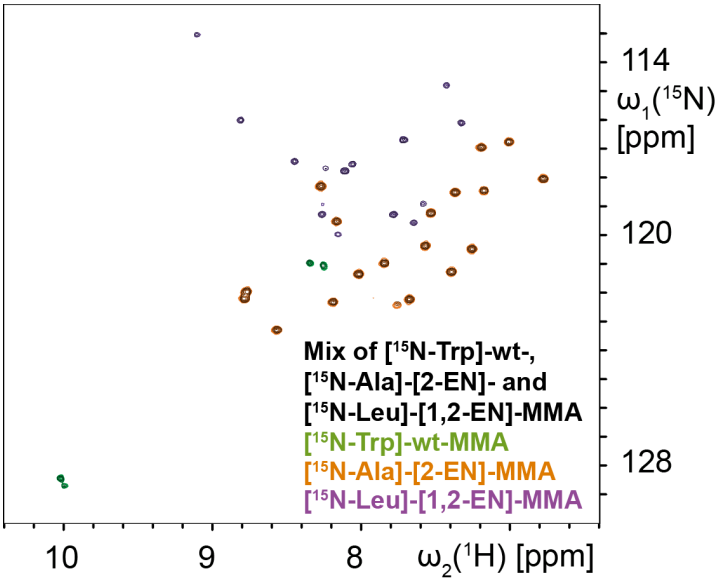


Figure S5: Superposition of 2D [^{15}N,^{1}H]-HSQC spectra from different MMA variants with unique amino acid-type labeling. Superposition of 2D [^{15}N,^{1}H]-HSQC spectra of [^{15}N-Trp]-wt-MMA, [^{15}N-Ala]-[2-EN]-MMA and [^{15}N-Leu]-[1,2-EN]-MMA measured either individually or in a mixture containing equimolar amounts of all three proteins. The individual spectra are color-coded as indicated.

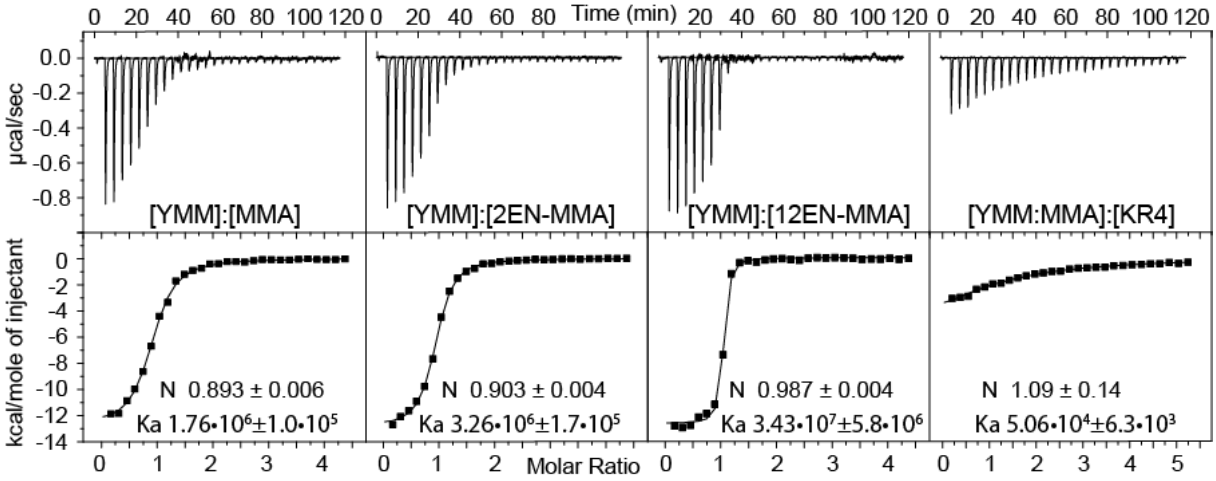


Figure S6: ITC analysis of YMM:MMA complementation using three different MMA variants, and interaction of the YMM:MMA complex with the $(KR)_4$ peptide. Titration of YMM with either (A) wt-MMA, (B) [2-EN]-MMA, or (C) [1,2-EN]-MMA. (D) Titration of the assembled YMM:MMA complex with a GB1-$(KR)_4$ fusion protein. The stoichiometry of binding sites, N, and the association constant, K_a, are indicated for each binding curve.

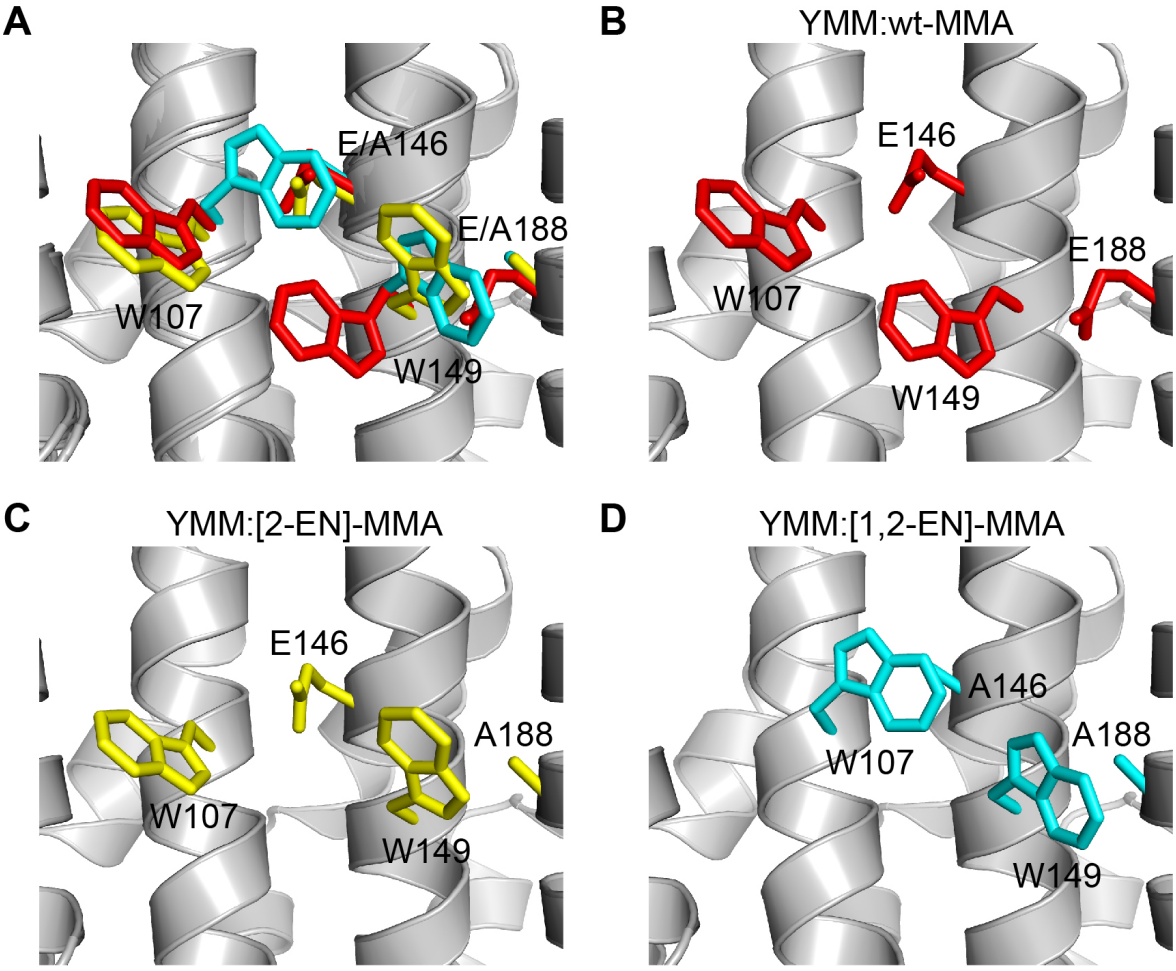


Figure S7: Rosetta modeling of YMM:MMA complexes reveals repacking of the Trp side chains as the main driving force for the preferential binding of YMM to [1,2-EN]-MMA. (A) Superposition and separate cartoon representations of the (B) YMM:wt-MMA, (C) YMM:[2-EN]-MMA, and (D) YMM:[1,2-EN]-MMA complexes. The side chains of the important Trp, Glu, and Ala residues are shown as sticks and are color-coded in red, yellow, and cyan for the wt, 2-EN, and 1,2-EN variants of the complexes, respectively. The Trp side chain adopts an energetically favorable conformation when the succeeding module contains a Glu-to-Ala mutation.

Supplementary Tables

Table S1: Individual contribution of selected residues to the total Rosetta energy units (REU) of YMM:MMA assemblies in absence and in presence of the $(KR)_5$ peptide ligand

Fragment Assembly	Repeat M2 [REU]			Repeat M3 [REU]			Repeat M4 [REU]		
YMM:wt-MMA	E104	W107	N111	E146	W149	N153	E188	W191	N195
	-1.8	1.2	-1.8	-1.5	1.6	-1.8	-3.4	0.6	-1.5
YMM:[2-EN]-MMA	E104	W107	N111	E146	W149	N153	A188	W191	A195
	-1.6	1.5	-2.4	-2.3	-1.6	-1.0	-2.6	0.1	-1.5
YMM:[1,2-EN]-MMA	E104	W107	N111	A146	W149	A153	A188	W191	A195
	-2.0	-1.5	-1.6	-3.6	-3.2	-2.4	-2.4	0.3	-1.9
YMM:wt-MMA	E104	W107	N111	E146	W149	N153	E188	W191	N195
$+(KR)_5$	-2.4	-1.4	-4.1	-1.6	-2.2	-4.8	-2.3	-1.3	-3.1
YMM:[2-EN]-MMA	E104	W107	N111	E146	W149	N153	A188	W191	A195
$+(KR)_5$	-2.8	-1.0	-4.3	-2.5	-2.4	-4.7	-2.8	-1.0	-1.2
YMM:[1,2-EN]-MMA	E104	W107	N111	A146	W149	A153	A188	W191	A195
$+(KR)_5$	-2.9	-3.2	-3.6	-3.5	-3.6	-2.3	-3.4	-1.0	-1.5

Table S2: Individual contribution of selected residues to the total Rosetta energy units (REU) of full-length YM_4A in absence of the $(KR)_5$ peptide ligand

Full-length dArmRP	Repeat M2 [REU]			Repeat M3 [REU]			Repeat M4 [REU]		
wt-YM_4A	E104	W107	N111	E146	W149	N153	E188	W191	N195
	-1.7	0.1	-2.2	-2.2	1.1	-2.6	-2.3	0.8	-2.1
[4-EN]-YM_4A	E104	W107	N111	E146	W149	N153	A188	W191	A195
	-1.8	0.8	-2.2	-2.5	-1.6	-1.8	-3.4	0.5	-1.3
[3,4-EN]-YM_4A	E104	W107	N111	A146	W149	A153	A188	W191	A195
	-1.9	-1.1	-1.9	-3.3	-2.1	-1.7	-3.4	0.0	-1.4

Table S3: Individual contribution of selected residues to the total Rosetta energy units (REU) of free MMA fragment variants

C-terminal Fragment	Repeat M1 [REU]			Repeat M2 [REU]		
wt-MMA	E146	W149	N153	E188	W191	N195
	-2.2	-0.6	-0.1	-2.0	0.8	-1.7
[2-EN]-MMA	E146	W149	N153	A188	W191	A195
	-2.5	-0.5	0.1	-2.9	0.1	-1.3
[1,2-EN]-MMA	A146	W149	A153	A188	W191	A195
	-0.5	-0.9	-0.8	-2.9	0.4	-1.2

Table S4: Rosetta energy unit (REU) differences between $(KR)_5$-bound and free YMM:MMA assemblies and between $(KR)_5$-bound and free full-length YM_4A variants

Bound State	Free State(s)	Δ(Bound-Free) [REU]
YMM:wt-MMA:$(KR)_5$	YMM:wt-MMA	-45.4 ± 3.2
YMM:[2-EN]-MMA:$(KR)_5$	YMM:[2-EN]-MMA	-39.6 ± 2.4
YMM:[1,2-EN]-MMA:$(KR)_5$	YMM:[1,2-EN]-MMA	-31.4 ± 2.4
wt-YM_4A:$(KR)_5$	wt-YM_4A	-48.9 ± 0.2
[4-EN]-YM_4A:$(KR)_5$	[4-EN]-YM_4A	-40.1 ± 0.6
[3,4-EN]-YM_4A:$(KR)_5$	[3,4-EN]-YM_4A	-34.8 ± 0.8

Table S5: Rotetta energy unit (REU) contribution of individual $(KR)_5$ peptide residues to the total REU of YMM:MMA:$(KR)_5$ complexes

MMA variant	Lys 1	Arg 2	Lys 3	Arg 4	Lys 5	Arg 6	Lys 7	Arg 8	Lys 9	Arg 10
wt	0.3	-2.2	-1.7	-4.1	-2.4	-3.4	-1.6	-4.3	-0.8	-3.5
2-EN	1.1	-0.1	-2.3	-2.2	-2.5	-2.4	-1.2	-3.5	-1.1	-3.5
1,2-EN	0.8	-0.9	-1.0	-0.3	-1.8	-2.4	-2.1	-4.3	-1.3	-4.3

Table S6: Oligonucleotide primer for cloning of the target proteins

Protein	5'-oligonucleotide and 3'-oligonucleotide primer
YMM	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-TTT GGA TCC TTA TTA GCC ACT CGC GAT ATT GCT TAA C
MMA	5'-AAA GCTCTTC A CAG GGT AAT GAA CAG ATC CAG GCT G 5'-GCT TTG TTA GCA GCC GGA TC
YMMH12	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
YM3A	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-GCT TTG TTA GCA GCC GGA TC
YM_5A (YM_2-M_3A)	5'-AAA GCTCTTC A CAG GGC GAA TTG CCG CAG 5'-TTT GCT CTT CTC CCG CCA CTC GCG ATA TTG 5'-TTT GCT CTT CTG GGA ATG AGC AAA TCC AAG CCG TG 5'-GCT TTG TTA GCA GCC GGA TC
[3-E]-YM_3A	5'-GAA CAA ATC TTG CAA GCA GCG CTT TGG GCT C 5'-GAG CCC AAA GCG CTG CTT GCA AGA TTT GTT C
[3-N]-YM_3A	5'-GCT TTG GGC TCT TTC TGC CAT TGC CTC TGG TGG 5'-CCA CCA GAG GCA ATG GCA GAA AGA GCC CAA AGC
[2-E]-YM_3A	5'-CAG ATT CTC CAA GCG GCT CTC TGG GCG 5'-CGC CCA GAG AGC CGC TTG GAG AAT CTG
[2-N]-YM_3A	5'-CTG GGC GTT AAG CGC TAT CGC GAG TGG CG 5'-CGC CAC TCG CGA TAG CGC TTA ACG CCC AG
$(KR)_4$-GB1	5'-AGC GGT AGC GGC TCT AAA CGC AAG CGT AAG CGC AAG CGC TAA 5'-GAT CTT AGC GCT TGC GCT TAC GCT TGC GTT TAG AGC CGC TAC C

Supplementary References

Hansen S*, et al.* (2016) Structure and Energetic Contributions of a Designed Modular Peptide-Binding Protein with Picomolar Affinity. *J. Am. Chem. Soc.* 138(10):3526-3532.

Michel E, Plückthun A, & Zerbe O (2018) Peptide-Guided Assembly of Repeat Protein Fragments. *Angew. Chem. Int. Ed.* 57(17):4576-4579.