Supplementary Material

for

Designed Armadillo Repeat Proteins: Library generation, characterization and selection of peptide binders with high specificity

Gautham Varadamsetty[1], Dirk Tremmel[1], Simon Hansen[1], Fabio Parmeggiani[1,2], Andreas Plückthun[1*]

[1]Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

[2]present address: Department of Biochemistry, University of Washington, J Wing, Health Sciences Building, Box 357350 Seattle, WA 98195

*Corresponding author. E-mail address: plueckthun@bioc.uzh.ch

Methods

Synthesis of ArmR modules

A complete list of all oligonucleotides used is given in Supplementary Data Table ST1. The trinucleotides phosphoramidites were obtained from Glen Research (USA) and lib5F was synthesized by Metabion (Germany). Standard oligonucleotides were from Microsynth (Switzerland).

An approach similar to the one described previously[1,2] was adopted for gene assembly (Supplementary Data Fig. S4). All single library modules were assembled by using the combinations of oligonucleotides indicated in Supplementary Data Table ST1 by assembly PCR. As an example, for the M-type consensus, pairs of partially overlapping oligonucleotides (1–2, 3–4, and 5–6) were annealed and the double strand was completed by PCR. Then, 2 µl from these PCR reaction mixtures was combined as template for a second assembly reaction in the presence of oligonucleotides 1 and 6. All the oligonucleotides were used at a final concentration of 1 µM. The annealing temperature was 47°C for the first reaction and 50°C for the second. Thirty PCR cycles were performed with an extension time of 30 s. The same procedure was applied for the other internal and capping repeats. For the C-terminal capping repeat, oligonucleotides Ca1libF, Ca2libR, Ca3F, Ca4R, Ca5F and Ca6R were used. Only four oligonucleotides (Ny1F, Ny2R, Ny3F and Ny4libRbis) were used for the shorter N-terminal capping repeat. The full-length N-cap was amplified with primers N1yF and Nylib4Rbis, while the full length C-cap was amplified with primers Ca1libF and Ca6R, and the internal library modules were amplified with libFor and lib6R. The restriction sites BamHI and KpnI were used for the direct insertion of the capping modules into the plasmid pQE30ss .

Due to randomization at position 41 (Fig. 1, Supplementary Data Fig. S2), the strategy for repeat assembly of the library modules had to be slightly altered compared to the original

module M.[2] The end points of the capping repeats, and the DNA sequence of the internal modules was therefore shifted by 6 bases compared to the original module M.[2]

For assembly of the internal library module (referred to as L type module), the oligonucleotides lib1F1, lib1F2 and lib1F3 together make up the mixture corresponding to the first of the six oligonucleotides used. At position 4, lib1F1 codes for the residues I and T, lib1F2 for E, K and Q, lib1F3 for H and R. These three oligonucleotides were mixed in molar ratio 2:3:2 to obtain the oligonucleotide mixture lib1F, thus containing the 7 different codons in equal amounts. The oligonucleotide lib5F incorporates trinucleotides phosphoramidites[3] for the codons corresponding to the randomized positions. The designed ArmR library modules were generated by assembly PCR using oligonucleotides lib1F1, lib1F2, lib1F3, lib2R, lib3F, lib4R, lib5F and lib6R to result in the sequence shown in Supplementary Data Fig. S2.

The designed ArmR library modules were PCR-assembled and subcloned for sequence analysis and to provide a template for the assembly of the whole library. In total, 2.2×10^7 independent clones for a single library module were obtained, a 20-fold oversampling of the theoretical repeat module diversity.

Synthesis of DNA encoding designed ArmRP library

The cells harboring the module-library plasmids were re-grown in liquid culture to obtain sufficient DNA and the plasmids were purified. This DNA was the starting material for the designed ArmRP library, which was assembled without any additional frameshift selection. The single modules were PCR amplified from the vectors, using external primers pQE_f_1 and pQE_r_1 (Qiagen, Switzerland). Modules were digested with the type IIS restriction enzymes BpiI and BsaI and directly ligated together (Supplementary Data Fig. S2, S4, S5) essentially as described previously,[2] leading to a final library containing three internal randomized modules (N3C).

To construct DNA cassettes encoding the N5C ArmR library domains, a PCR-assembled non-randomized (constant) M-type repeat module was ligated first to an PCR-assembled N-terminal capping repeat and subsequently to a PCR-assembled C-terminal capping repeat by using type IIs restriction enzymes (Supplementary Data Fig. S5).

Subsequently, the ligation products of constant M repeat modules (KK, KQ and QQ type) with N and C-caps were PCR amplified, followed by restriction digestion and ligation to each other: N-capM with MC-cap. The resulting ligation product consists of N-cap, 2 constant modules separated by a DNA linker, and the C-cap. It was then PCR amplified and subcloned into the vector pPANK. In the next step, the PCR assembled randomized library module was ligated stepwise to another randomized library module. The PCR-amplified double library repeat module was then ligated with another randomized repeat module.

Of all the possible combinations of codons introduced in one repeat (9.9×10^6), only one (GAA-GAC, Glu-Asp) at positions 40-41 leads to a recognition site for BpiI that can be cleaved during assembly and would thus be lost during assembly.

After each ligation step the concentration of purified ligation product was measured. A ligation mix of sufficient quantity to represent and oversample the diversity of the designed ArmRP library was tested by transformation and sequence analysis of randomly picked clones. From the amount of DNA used, after each ligation step the practical library size of the N3C library was approximately 10^{11}. BamHI and KpnI restriction sites were used for insertion of the whole genes into the vector pPANK and the plasmids were sequenced to obtain an initial N3C library.

The final libraries are then formed by ligating double-digested PCR product of three library modules (LLL) into the plasmid, where it is inserted between NM and MC to result in proteins with NMLLLMC module orientation. The full-length proteins were termed as N5C library and it was obtained in 4 types, characterized by the residues at position 26 and 29 in each repeat (KK, KQ, QK or QQ).

Since the N-terminal cap from yeast (termed Ny or Y) and the artificial C-terminal cap (termed Ca or A) were to be used for many different constructs, a plasmid containing the two caps was also constructed (pPANK-NyCa). In this construct, the caps are separated by a linker, including the restriction sites for *Bpi*I and *Bsa*I. This construct made it possible to insert any internal repeat module directly in between the two caps (Supplementary Data Fig. S5).

References

1. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489-503.
2. Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A. & Plückthun, A. (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* **376**, 1282-1304.
3. Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600-5607.

Table ST1: Oligonucleotides used for the assembly and cloning of designed ArmR protein genes and target peptides as fusion partners to protein D

name	sequence 5'-3'	description (for=forward, rev=reverse)
KQfor	CCGAACGAGAAGATCCTGCAAGAAGCTCTGTGGGC	for mutation KK->KQ
KQrev	GCCCACAGAGCTTCTTGCAGGATCTTCTCGTTCGG	rev mutation KK->KQ
QKfor	CCTCTCCGAACGAGCAGATCCTGAAAGAAGC	for mutation KK->QK
QKrev	GCTTCTTTCAGGATCTGCTCGTTCGGAGAGG	rev mutation KK->QK
QQfor	CCTCTCCGAACGAGCAGATCCTGCAAGAAGCTCTGTGGGC	for mutation KK->QQ
QQrev	GCCCACAGAGCTTCTTGCAGGATCTGCTCGTTCGGAGAGG	rev mutation KK->QQ
Ny4libRbis	TTCCTGGTACCCTAAGGTCTCATTCGTTACCATCACGAGACAGGATCTG	rev replacement for assembly Ny cap in library format
Ca1libF	CCAGGGATCCTCTAGATAGGAAGACCTCGAACAGAAACAGGC	for replacement for assembly Ca cap in library format
Ca2libR	GTTTCTCCAGAGCACCAGCTTCTTTAACAGCCTGTTTCTGTTCGAGGTC	rev replacement for assembly Ca cap in library format
lib1F1	CCAGGGATCCTAGGAAGACCTCGAACAAAYCCAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes I and T
lib1F2	CCAGGGATCCTAGGAAGACCTCGAACAAVAACAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes E, K and Q
lib1F3	CCAGGGATCCTAGGAAGACCTCGAACAACRCCAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes H and R
lib2R	CCAGAGCCGGCAGAGCACCAGCATCGATAACAGCTTG	rev assembly libraries
lib3F	CTGCCGGCTCTGGTTCAACTGCTGTCCTCTC	for assembly libraries
lib4R	TTTCAGGATCTTCTCGTTCGGAGAGGACAGCAG	rev assembly KK library
lib5F	CGAGAAGATCCTGAAANNNGCTCTGNNNGCTCTGNNNAACATCGCTNNNNNNGGTAACGAATGAG	for assembly KK library, randomized positions 30,33,36,40,41 with trinucleotides, coding for an equimolar mixture of all amino acids, except Cys, Gly and Pro
lib6R	TTCCTGGTACCCTAAGGTCTCATTCGTTACC	rev assembly libraries
libFOR	CCAGGGATCCTAGGAAGACCTCGAAC	for general for library amplification
lib4RKQ	TTGCAGGATCTTCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to KQ library
lib4RQK	TTTCAGGATCTGCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to QK library
lib4RQQ	TTGCAGGATCTGCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to QQ library
lib4RKQ	TTGCAGGATCTTCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to KQ library
lib4RQK	TTTCAGGATCTGCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to QK library
lib4RQQ	TTGCAGGATCTGCTCGTTCGGAGAGGACAGCAG	rev conversion KK library to QQ library
Ca6RhindRD	TTCCTAAGCTTGTGGGAGAACTGCTTCTCCAGAGCTTCC	rev amplification to remove stop codon for ribosome display

Table ST1 (continued): Oligonucleotides used for the assembly and cloning of designed ArmR protein genes and target peptides as fusion partners to protein D (continued)

name	sequence 5'-3'	description (for=forward, rev=reverse)
Ny1F	CCAGGGATCCGAACTGCCGCAGATGACC CAGCAGCTGAACTCTG	for assembly Ny module and amplification
Ny2R	CGGTAGCAGACAGCTGTTCCTGCATGTC GTCAGAGTTCAGCTGCTGGG	rev assembly Ny module
Ny3F	GAACAGCTGTCTGCTACCGTTAAATTCC GTCAGATCCTGTCTCGTGATGG	for assembly Ny module
Ny4R	TTCCTGGTACCCTAAGGTCTCAACCATC ACGAGACAGGATCTG	rev assembly Ny module and amplification
Ny4libRbis	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCATCACGAGACAGGATCTG	rev replacement for assembly Ny cap in library format
Ca1F	CCAGGGATCCTAGGAAGACCTTGGTAAC GAACAGAAACAGGC	for assembly Ca module and amplification
Ca1libF	CCAGGGATCCTCTAGATAGGAAGACCTC GAACAGAAACAGGC	for replacement for assembly Ca cap in library format
Ca2R	GTTTCTCCAGAGCACCAGCTTCTTTAAC AGCCTGTTTCTGTTCGTTACC	rev assembly Ca module
Ca2libR	GTTTCTCCAGAGCACCAGCTTCTTTAAC AGCCTGTTTCTGTTCGAGGTC	rev replacement for assembly Ca cap in library format
Ca3F	GCTGGTGCTCTGGAGAAACTGGAACAGC TGCAGTCCCACGAG	for assembly Ca module
Ca4R	CCTGAGCTTCTTTCTGGATCTTCTCGTT CTCGTGGGACTGCAGC	rev assembly Ca module
Ca5F	GATCCAGAAAGAAGCTCAGGAAGCTCTG GAGAAGCAGTTCTCCC	for assembly Ca module
Ca6R	TTCCTGGTACCTCATTAGTGGGAGAACT GCTTCTCCAG	rev assembly Ca module and amplification
Ca6RhindRD	TTCCTAAGCTTGTGGGAGAACTGCTTCT CCAGAGCTTCC	rev amplification to remove stop codon for ribosome display
lib1F1	CCAGGGATCCTAGGAAGACCTCGAACAA AYCCAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes I and T
lib1F2	CCAGGGATCCTAGGAAGACCTCGAACAA VAACAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes E, K and Q
lib1F3	CCAGGGATCCTAGGAAGACCTCGAACAA CRCCAAGCTGTTATCG	for assembly libraries, randomized position 4, encodes H and R
lib2R	CCAGAGCCGGCAGAGCACCAGCATCGAT AACAGCTTG	rev assembly libraries
lib3F	CTGCCGGCTCTGGTTCAACTGCTGTCCT CTC	for assembly libraries
lib4R	TTTCAGGATCTTCTCGTTCGGAGAGGAC AGCAG	rev assembly KK library
lib5F	CGAGAAGATCCTGAAANNNGCTCTGNNN GCTCTGNNNAACATCGCTNNNNNNGGTA ACGAATGAG	for assembly KK library, randomized positions 30,33,36,40,41 with trinucleotides, coding for an equimolar mixture of all amino acids, except Cys, Gly and Pro
lib6R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACC	rev assembly libraries
libFOR	CCAGGGATCCTAGGAAGACCTCGAAC	for general for library amplification
5KQ	CGAGAAGATCCTGCAA	for conversion KK library to KQ library
5QK	CGAGCAGATCCTGAAA	for conversion KK library to QK library
5QQ	CGAGCAGATCCTGCAA	for conversion KK library to QQ library

Table ST1 (continued): Oligonucleotides used for the assembly and cloning of designed ArmR protein genes and target peptides as fusion partners to protein D (continued)

name	sequence 5'-3'	description (for=forward, rev=reverse)
lib4RKQ	TTGCAGGATCTTCTCGTTCGGAGAGGAC AGCAG	rev conversion KK library to KQ library
lib4RQK	TTTCAGGATCTGCTCGTTCGGAGAGGAC AGCAG	rev conversion KK library to QK library
lib4RQQ	TTGCAGGATCTGCTCGTTCGGAGAGGAC AGCAG	rev conversion KK library to QQ library
lib4RKQ	TTGCAGGATCTTCTCGTTCGGAGAGGAC AGCAG	rev conversion KK library to KQ library
T7b	ATACGAAATTAATACGACTCACTATAGG GAGACCACAACGG	for outer primer ribosome display
tolAk	CCGCACACCAGTAAGGTGTGCGGTTTCA GTTGCCGCTTTCTTTCT	rev outer primer ribosome display
VG_4f	CACCATCACGGATCCGAACTGCCGC	cap specific for inner primer ribosome display
VG_4r	ATAAAGCTTGTGGGAGAACTGTTCTCC	cap specific rev inner primer ribosome display
1_4F	CCAGGGATCCTAGGAAGACCTCGAACAA AAGCAAGCTGT	for unselected library modules amplification
1_4R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCGTGGTGAGCGAT	rev unselected library modules amplification
2_10F	CCAGGGATCCTAGGAAGACCTCGAACAA AAGCAAGCTGT	for unselected library modules amplification
2_10R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCGTGGAAAGCGAT	rev unselected library modules amplification
2_12F	CCAGGGATCCTAGGAAGACCTCGAACAA AAGCAAGCTGT	for selected library modules amplification
2_12R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCGGTTTTAGCGAT	rev unselected library modules amplification
VG_8F	CCAGGGATCCTAGGAAGACCTCGAACAA CGTCAAGCTGT	for selected library modules amplification
VG_8R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCCAGAACAGCGAT	rev unselected library modules amplification
VG_12F	CCAGGGATCCTAGGAAGACCTCGAACAA CGTCAAGCTGT	for selected library modules amplification
VG_12R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCGGTGAAAGCGAT	rev unselected library modules amplification
VG_30F	CCAGGGATCCTAGGAAGACCTCGAACAA CACCAAGCTGT	for selected library modules amplification
VG_30R	TTCCTGGTACCCTAAGGTCTCATTCGTT ACCGGTGGTAGCGAT	rev unselected library modules amplification
NLSF	CTGTGGATCCGGTTCTCCGAAGAAGAAG CGCAAGGTGGAGGGTTAATAAGCTTG	for NLS peptide amplification
NLSR	GGACTACAAGCTTATTAACCCTCCACCT TGCGCTTCTTCTTCGGAGAACCGGA	rev NLS peptide amplification
NTF	TCACGGATCCCAGCTGTACGAAAACAAA CCGCGTCGTCCGTAC	for NT peptide amplification
NTR	CAGGTATATGCGCAAGCTTATTACAGGA TGTACGGACG	rev NT peptide amplification
NT_AV_F	TCACGGATCCCAGCTGTACGAAAACGCT GCTGCTGCTGTGGTG	for NT_AV peptide amplification

Table ST1 (continued): Oligonucleotides used for the assembly and cloning of designed ArmR protein genes and target peptides as fusion partners to protein D (continued)

name	sequence 5'-3'	description (for=forward, rev=reverse)
pD_NT (8-13)_F	TCACGGATCCCGTCGTCCGTACATCCTG	for NT hexa peptide (8-13) amplification
pD_NT (8-13)_R	CAGGTATATGCGCAAGCTTATTACAGGATGTACGGACG	rev NT hexa peptide (8-13) amplification
pD_NT_F15_F	TCACGGATCCCAGCTGTACGAAAACGCTCCGCGTCGTGCTTAC	for pD_NT_F15 peptide amplification
pD_NT_F15_R	CAGGTATATGCGCAAGCTTATTAAGCAGCGTAAGCACG	rev pD_NT_F15 peptide amplification
pD_NT_F16_F	TCACGGATCCCAGCTGTACGAAAACAAAGCTGCTGCTCCGGCT	for pD_NT_F16 peptide amplification
pD_NT_F16_R	CAGGTATATGCGCAAGCTTATTACAGGATAGCCGGAGC	rev pD_NT_F16 peptide amplification

Table ST2: Properties of unselected library members of N5C libraries (KK, KQ, and QQ)

N5C library type[a]	No. of clones analyzed[b] (SEC)	Aggregates[c] (SEC)	Monomers[d] (SEC)
KK	11	9 proteins (minor peaks)	11 proteins (major peaks)
KQ	5	3 proteins (minor peaks)	5 proteins (major peaks)
QQ	5	0 proteins (no aggregates)	5 proteins (major peaks)

[a] The two letters refer to the amino acid at position 26 and 29, respectively
[b] number of clones of KK, KQ and QQ library analyzed in the N5C format
[c] number of clones eluting as mixture of aggregates (minor proportion) and monomers (major proportion)
[d] number of clones eluting as monomers

Fig. S1

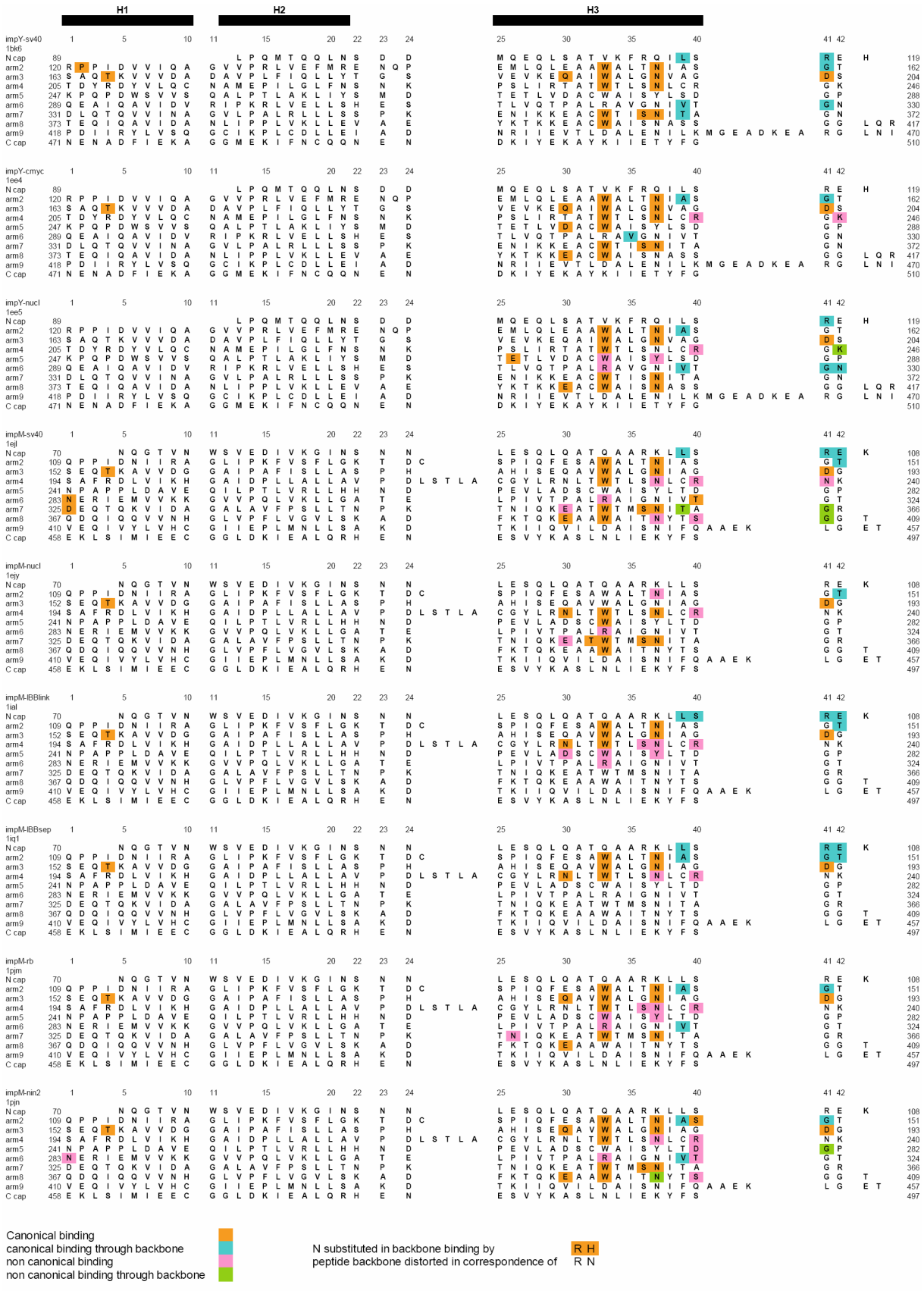

H1
H2
H3
impY-sv40
1bk6
impY-cmyc
1ee4
impY-nucl
1ee5
impM-sv40
1ejl
impM-nucl
1ejy
impM-IBBlink
1ial
impM-IBBsep
1iq1
impM-trb
1pjm
impM-nin2
1pn
N cap
arm2
arm3
arm4
arm5
arm6
arm7
arm8
arm9
C cap
Canonical binding
canonical binding through backbone
non canonical binding
non canonical binding through backbone
N substituted in backbone binding by
peptide backbone distorted in correspondence of

Fig. S1 (continued)

```
bcatM-Pcadherin  1          10            20  22              23 24   25      30         40
1i7w
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatM-cadherin  1          10            20  22              23 24   25      30         40
1i7x
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatM-APC  1               10            20  22              23 24   25      30         40
1jpp
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatM-ICAT  1              10            20  22              23 24   25      30         40
1m1e
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatH-XTcf3  1             10            20  22              23 24   25      30         40
1g3j
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatH-XTcf4  1             10            20  22              23 24   25      30         40
1jdh
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatH-XTcf4bis  1          10            20  22              23 24   25      30         40
1jpw
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatH-ICAT  1              10            20  22              23 24   25      30         40
1luj
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

```
bcatH-APC20P  1            10            20  22              23 24   25      30         40
1th1
N cap   149            AT  RAIPELTKLLND                           ED  QVVVNKAAVMVHQLSK  K      181
arm2    182  EASRHAIMRSP   QMVSAIVRTMQN                          TND  VETARCTAGTLHNLSH  H      224
arm3    225  REGLLAIFKS    GGIPALVKMLGS                           PV  DSVLFYAITTLHNLLL  HQ     266
arm4    267  EGAKMAVRLA    GGLQKMVALLNK                           TN  VKFLAITTDCLQILAY  GN     308
arm5    309  QESKLIILAS    GGPQALVNIMRT                          YTY  EKLLWTTSRVLKVLSV  C      350
arm6    351  SSNKPAIVEA    GGMQALGLHLTD                           PS  QRLVQNCLWTLRNLSD  AA     392
arm7    393     TKQEGME    GLLGTLVQLLGS                           DD  INVVTCAAGILSNLTC  NN     431
arm8    432  YKNKMMVCQV    GGIEALVRTVLRA                         GDR  EDITEPAICALRHLTS  RHQEA  478
arm9    479  EMAQNAVRLH    YGLPVVVKLLHP                          PSH  WPLIKATVGLIRNLAL  C      520
arm10   521  PANHAPLREQ    GAIPRLVQLLVRAHQDTQR  RTSMGGTQQQFVEGVRM  EEIVEGCTGALHILAR  D      583
arm11   584  VHNRIVIRGL    NTIPLFVQLLYS                           PI  ENIQRVAAGVLCELAQ  D      624
C cap   625  KEAAEAIEAE    GATAPLTELLHS                           RN  EGVATYAAAVLFRMSE         664
```

Fig. S1. (Figure on previous 2 pages). Binding residues in natural armadillo repeat proteins. The protein sequences of importins α and β-catenins are depicted with the repeats aligned according to the structural data. The residues involved in binding, as defined by analysis of crystal structures of complexes, are colored. Orange indicates a canonical binding mode; with the target backbone bound every two residues by an asparagine at position 37 in the armadillo repeat. Cyan indicates that the main chain of that armadillo residue is used for the recognition of a target side chain. Magenta is used for residues contacting the target main chain or side chain in alternative ways (e.g. position 36 contacts the backbone instead of position 37) and green when the interaction involves the backbone of a residue of the armadillo domain. Note that this analysis was carried out in 2008.

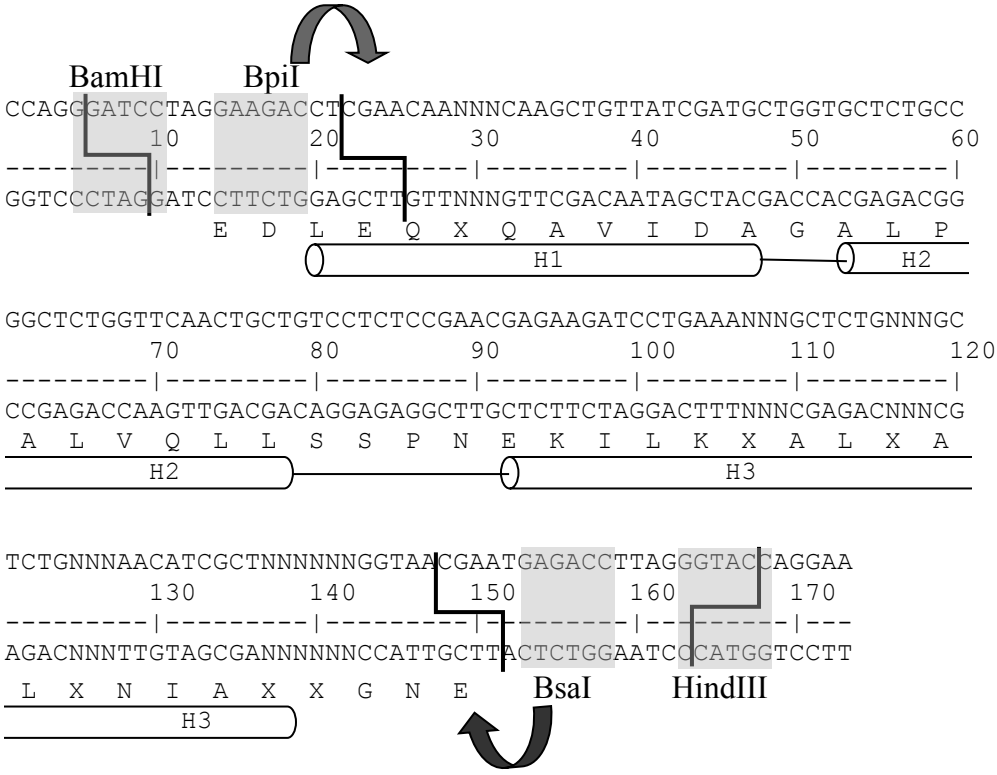


Fig. S2. The assembly PCR product of a single ArmR library module is shown. The restriction enzyme recognition sites are shown as grey boxes and the cutting sites are indicated with continuous lines. Note that the DNA recognition sites of the type IIs restriction enzymes are distant from their cleavage site pointed by curved arrows, and thus these sites are lost upon cleavage and religation.

Position 4: Allowed amino acids: E H K I Q T R

residues	counted (79)	% expected	% observed
E	9	14.3	11.4
H	15	14.3	19.0
K	13	14.3	16.5
I	11	14.3	13.9
Q	9	14.3	11.4
T	11	14.3	13.9
R	11	14.3	13.9

Other randomized positions
(All amino acids except Cys, Gly, Pro)

residues	counted (399)	% expected	% observed
A	27	5.9	6.8
D	17	5.9	4.3
E	27	5.9	6.8
F	21	5.9	5.3
H	34	5.9	8.5
I	31	5.9	7.8
K	21	5.9	5.3
L	14	5.9	3.5
M	24	5.9	6.0
N	15	5.9	3.8
Q	18	5.9	4.5
R	16	5.9	4.0
S	27	5.9	6.8
T	26	5.9	6.5
V	34	5.9	8.5
W	23	5.9	5.8
Y	24	5.9	6.0

Single module

Clones sequenced: 23 individual repeats

Framework point mutation: 1
Framework point del. and ins. (frameshift): 2
Framework del. and ins. (no frameshift): -
Rand. Pos. point mut.: -
Rand. Pos. point del. and ins. (frameshift): 4
Rand. Pos. del. and ins. (no frameshift): 2
Out of frame sequences: 6
In frame sequences: 17

N3C
Clones sequenced: 20 N3C, 60 repeats

Framework point mutation: 13 (4 silent)
Framework point del. and ins. (frameshift): 7
Framework del. and ins. (no frameshift): 3
Rand. Pos. point mut.: 2 (1 silent, 1 stop)
Rand. Pos. point del. and ins. (frameshift): 2
Rand. Pos. del. and ins. (no frameshift): 3
Out of frame sequences: 7
In frame sequences: 13

Rand. Pos., randomized position
Point mut., point mutation
del., deletion
ins., insertion

Fig. S3. Library quality. Expected and observed frequency of residues at randomized positions indicated and mutations observed in single modules and members of the unselected N3C library. Only a restricted number of residues was allowed at position 4 because of steric constraints. Insertions or deletions without frameshift are the result of the insertion or deletion of 3 nucleotides.

Figure S4

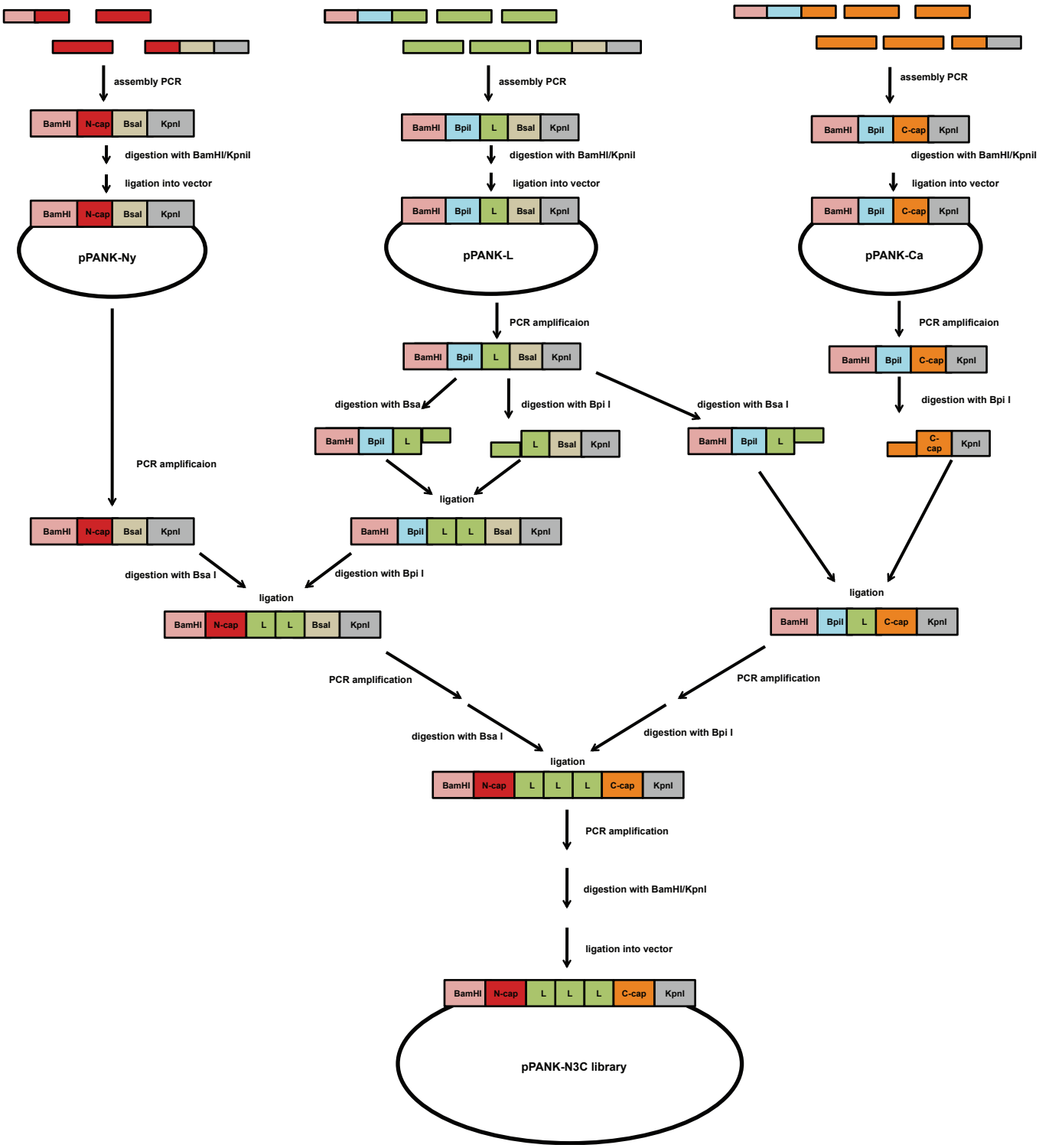

assembly PCR
BamHI
N-cap
BsaI
KpnI
digestion with BamHI/KpnI
ligation into vector
pPANK-Ny
BpiI
L
pPANK-L
C-cap
pPANK-Ca
PCR amplificaion
digestion with Bsa I
digestion with Bpi I
ligation
PCR amplification
pPANK-N3C library

Fig. S4. Assembly of designed ArmRP library at the DNA level starting from single repeat modules. The terminal capping repeat (N-cap, labeled N), C-capping repeat (C-cap, labeled C) and the single library module (L) are generated by assembly PCR. The C-terminal capping repeat is ligated to the first armadillo library module using the type IIs restriction enzymes BpiI and BsaI. A single repeat library module is ligated to another library module followed by ligation to the N-terminal capping repeat using the same set of restriction enzymes as above. The assembled LC modules are combined with assembled NLL modules to finally give a N3C library. By this strategy, designed armadillo libraries of longer lengths can also be generated. The use of the type IIs restriction enzymes ensures the seamless junction of the repeats in a directional manner. Note that the modules are not drawn to scale.

Figure S5

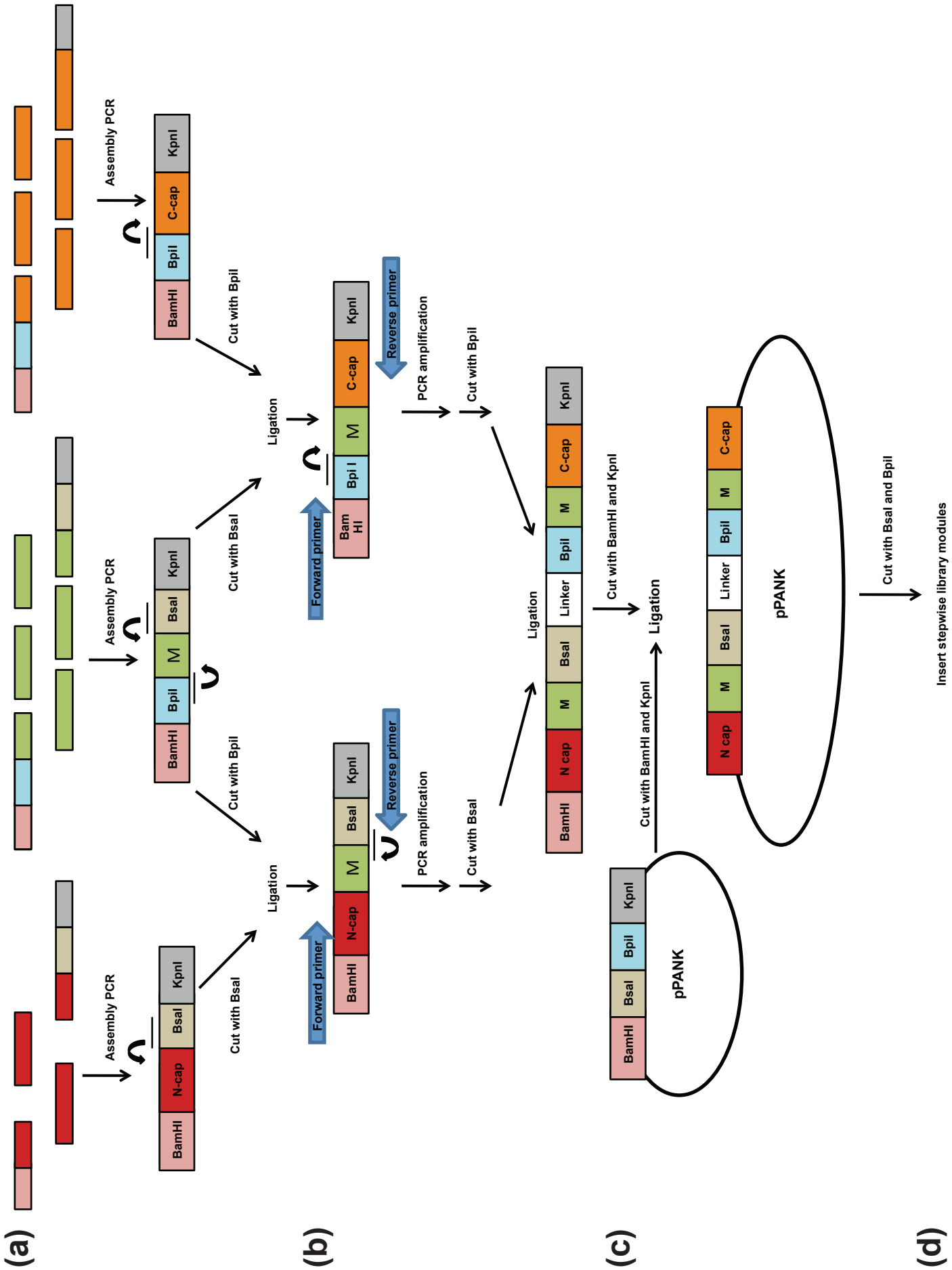

(a)
Assembly PCR
BamHI
N-cap
BsaI
KpnI
Cut with BsaI
Assembly PCR
BamHI
BpiI
M
BsaI
KpnI
Cut with BpiI
Cut with BsaI
Assembly PCR
BamHI
BpiI
C-cap
KpnI
Cut with BpiI
(b)
Ligation
Forward primer
BamHI
N-cap
M
BsaI
KpnI
Reverse primer
PCR amplification
Cut with BsaI
Ligation
Forward primer
Bam HI
Bpi I
M
C-cap
KpnI
Reverse primer
PCR amplification
Cut with BpiI
(c)
Ligation
BamHI
N cap
M
BsaI
Linker
BpiI
M
C-cap
KpnI
Cut with BamHI and KpnI
Ligation
Cut with BamHI and KpnI
BamHI
BsaI
BpiI
KpnI
pPANK
N cap
M
BsaI
Linker
BpiI
M
C-cap
pPANK
(d)
Cut with BsaI and BpiI
Insert stepwise library modules

Fig. S5. Schematic representation of the assembly of the designed ArmRP constructs at the DNA level to generate the N5C library. (a) Oligonucleotides were assembled to a complete single ArmR module, N-cap module and C-cap module by PCR. The N-cap contains external restriction sites for BamHI, KpnI and the type II restriction enzyme site BsaI. The single ArmR modules contain external restriction sites for BamHI and KpnI and the type II restriction enzymes BsaI and BpiI. The C-cap contains external restriction sites for Bam HI, KpnI and the type II restriction enzyme site BpiI. Digestion of one single ArmR module with BsaI and another molecule, digested with with BpiI, followed by ligation to BsaI digested N-cap and BpiI digested C-cap generates a double repeat modules of N-capM and MC-cap **(b)** Using external primers, the double modules are amplified for additional rounds of digestion and ligation of double module N-cap-M with double module M-C-cap joined via DNA linker **(c)** The BamHI and KpnI sites are used to insert the module(s) with terminal capping repeats into an appropriate vector (pPANK). **(d)** Alternatively, the BpiI and BsaI sites can be used to insert the module(s) directly in between the repeats (N-cap-M and M-C-cap) in the pPANK-YM-MA vector, thereby generating the final NxC constructs. After each ligation step the concentration of purified ligation product was measured. A ligation mix of sufficient quantity to represent and oversample the diversity of the designed ArmRP library was tested.

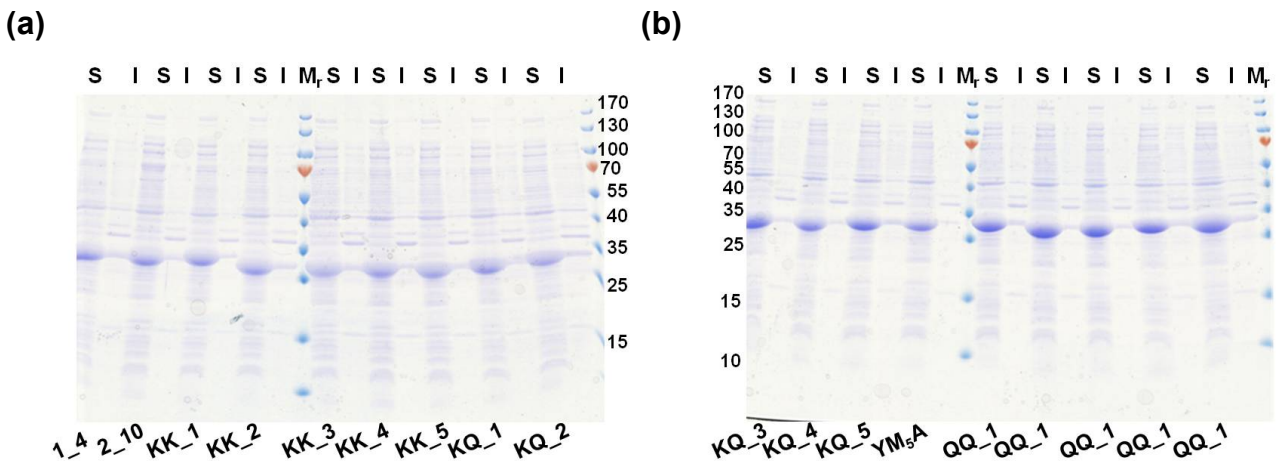
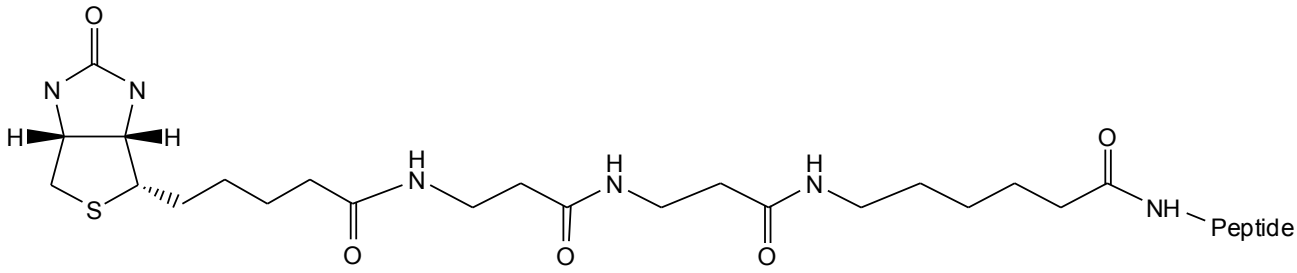


Fig. S6. Expression analysis of unselected designed ArmRP N5C library members. **(a)** and **(b)** Crude extracts of *E. coli* XL1-Blue expressing unselected and consensus-designed ArmRPs. The proteins without letters in front are obtained from the N3C library (in KK format) but were converted to N5C format. Proteins were expressed for four hours (see Materials and Methods) and the cell lysates were analyzed by SDS-PAGE (15%). Soluble fractions (S) and insoluble fractions (I) of expressed proteins were loaded. The expected size is approximately 32 kDa.

(a)



Biotin β-Alanine β-Alanine 6-amino hexanoic acid (LC)

(b)

Fig. S7. Target formats of NT peptide for ribosome display selections. **(a)** Structure of biotin linked via 2 β-alanine units to 6-amino hexanoic acid (LC). The free N-terminus of the peptide (neurotensin (1-13)) was connected to the carboxy group of LC via an amide bond. The C-terminal carboxy group of the peptide was unmodified. **(b)** DNA cassette for in vivo expressed biotinylated neurotensin peptide. The NT peptide target (light green) is fused to phage λ protein D (pink) connected by a His_6 tag (cyan) and a $(Gly\text{-}Ser)_2$ linker (dark grey). An Avitag (blue) is attached to the N terminus of the protein D for in vivo biotinylation. The carboxy-terminus of neurotensin is unmodified.

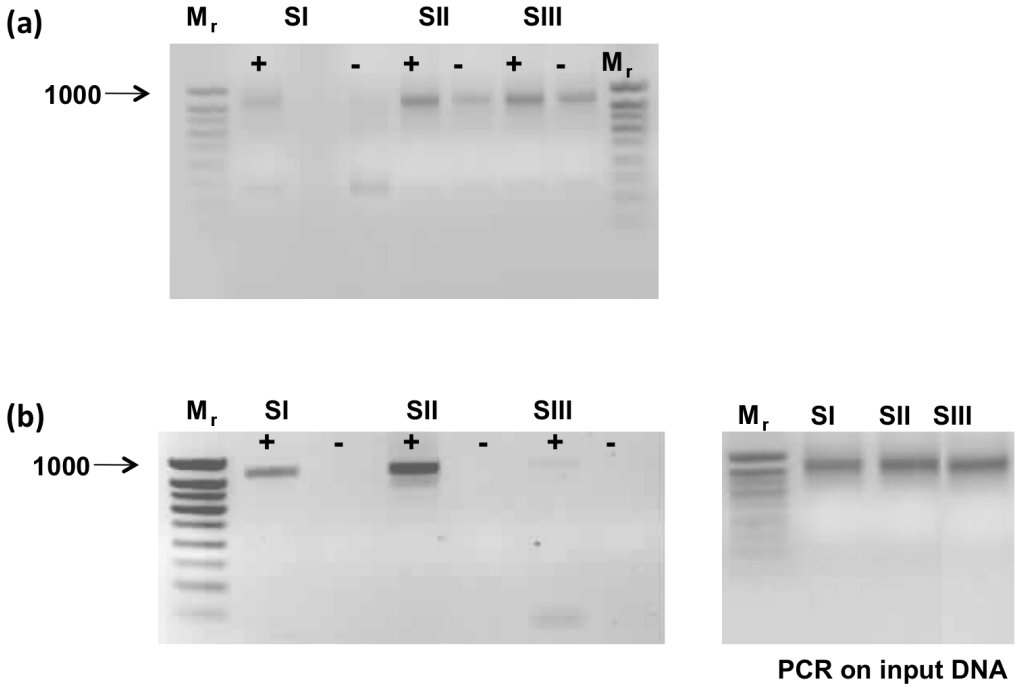


Fig. S8. Analysis of N5C library (QQ) amplified DNA pools after RT-PCR on 1.2% agarose gels. **(a)** Amplification of DNA pools of the third round of RD selection after 30 cycles **(b)** Amplification of DNA pools of the fourth round of RD selection after 25 cycles. All the PCRs were performed using designed ArmRP specific primers VG_4f and VG_4r. The mRNA obtained from selections for binding to peptide target was used for panning either on peptide target (+) or neutravidin only (-). SI, SII, and SIII refer to the three different strategies used for panning (cf. Table 2). To test the quality of input libraries, mRNA obtained from selections was used directly for reverse transcription, skipping the selection cycle (labeled PCR on input mRNA)

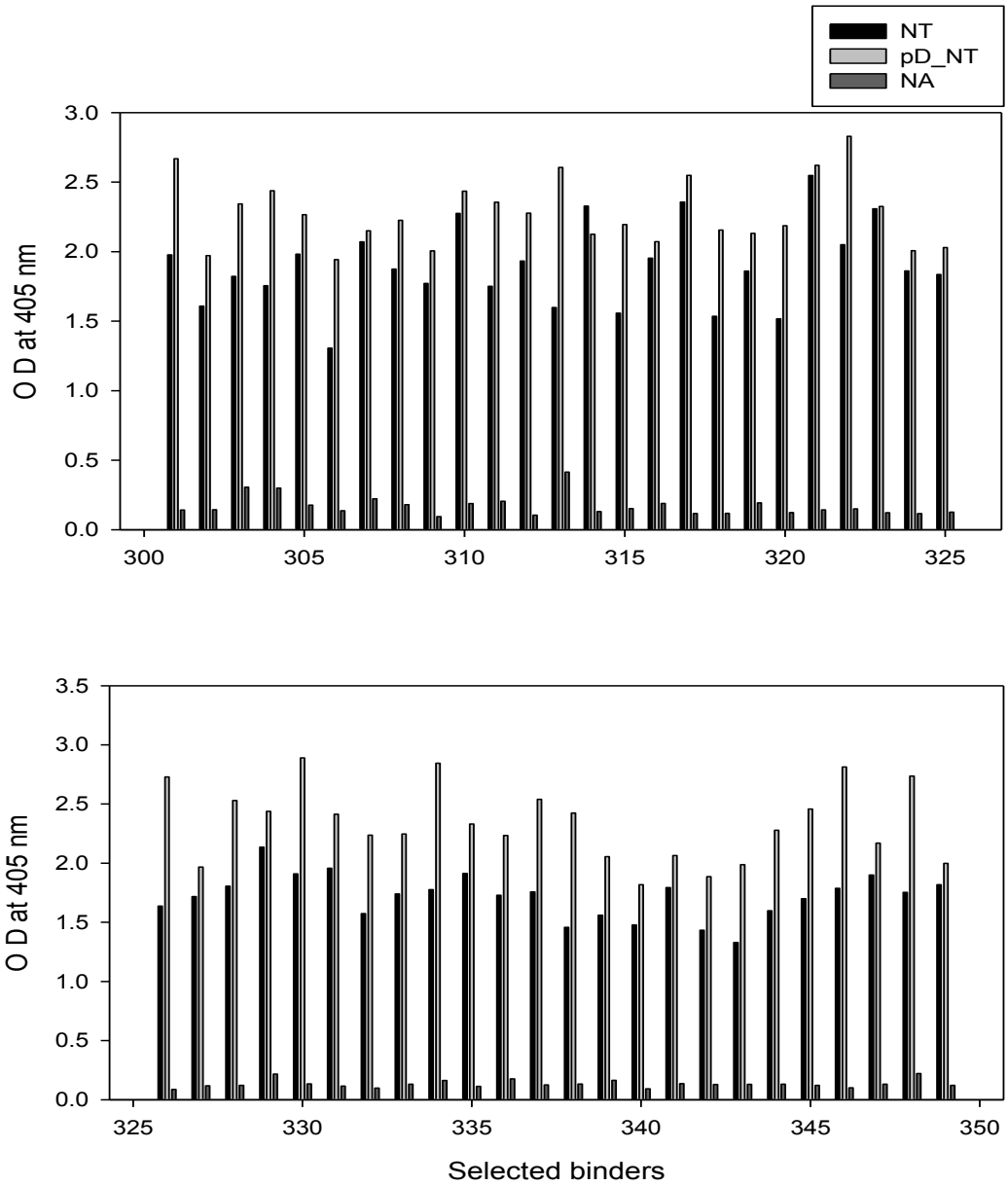


Fig. S9. Target recognition by selected ArmRPs. ELISA analysis (N5C QQ library) of single clones from strategy II after the fourth RD selection cycle. NT, pD_NT and NA indicate the target peptide neurotensin, neurotensin fused to protein D (both in turn bound to neutravidin via biotin), and neutravidin alone, respectively. The binding of the designed ArmRPs was detected using a primary anti-RGSHis-antibody and anti-mouse AP coupled secondary antibody (Qiagen). The signal was developed for 2 h after addition of the substrate.

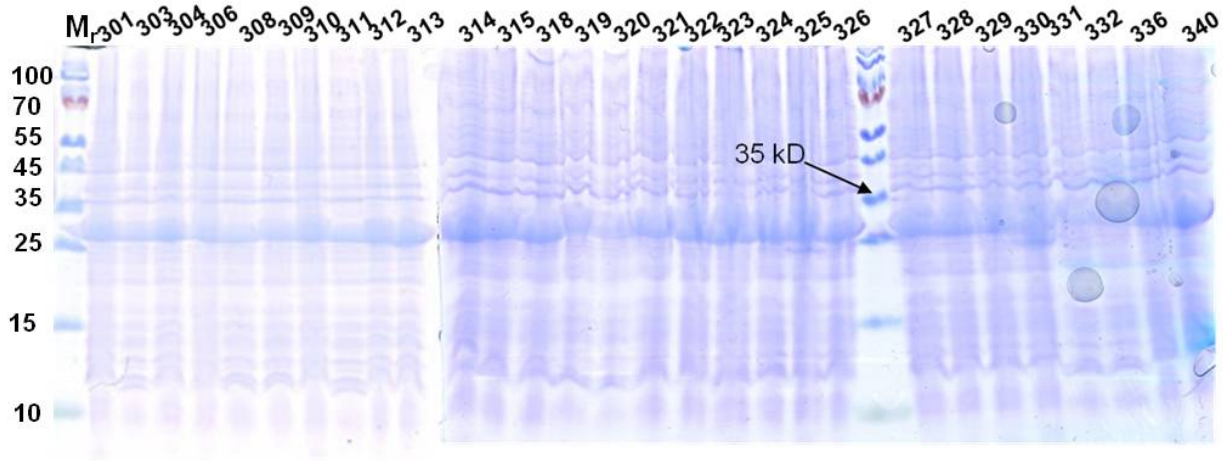


Fig. S10. Secondary screening of designed ArmRPs by 15% SDS-PAGE. Analysis of whole cell extracts of selected binders (Binders from ELISA depicted in Fig. 4.2.16). On each lane, cell extract from ~100 µl of the expression culture was loaded. All the expressed binders run at the expected protein size corresponding to consensus designed protein YM_5A. The expected size is approximately 32 kDa.

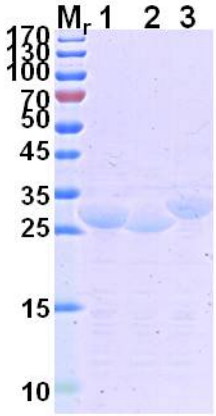


Fig. S11. Single-step IMAC purification of selected binders. (lane 1, YM_5A included as control for N5C proteins; lane 2, lane VG_306; lane 3, VG_328). The expected size is approximately 32 kDa. Proteins were analyzed by SDS-PAGE (15%). The size marker is indicated in kDa.

```
pRDV        117 atggcggactacaaagatgacgatgacaaaggatcccac 155
                 M  A  D  Y  K  D  D  D  D  K  G  S  H

pRDVhis     117 atgagaggatcgcatcaccatcaccatcacggatcccag 155
                 M  R  G  S  H  H  H  H  H  H  G  S  E

pRDVhis_CAG 117 atgcgagggtcgcatcaccatcaccatcacggatcccag 155
                 M  R  G  S  H  H  H  H  H  H  G  S  E
```

Fig. S12. Changes in pRDVhis and pRDVhis_CAG compared to the original pRDV vector (GenBank accession number AY327136) for ribosome display. In yellow the initiator ATG is shown, in grey the original FLAG tag, and in cyan the His_6 tag. In red is shown the region which may form a hairpin with the sequence GTCCTCTC (nucleotide 79 in **Fig. S2**, aa 21-23 in **Fig. 1a**) and in green mutations to disrupt the RNA secondary structure. (Note that in some recent publications, the vector pRDVhis has been called "pRDV" for short)