PNAS

Supporting Information for

Modular Binder Technology by NGS-Aided, High-Resolution Selection in Yeast of Designed Armadillo Modules

Yvonne Stark[a,1], Faye Menard[a,2], Jeliazko R. Jeliazkov[a,3], Patrick Ernst[a,4], Anupama Chembath[b], Mohammed Ashraf[b], Anna V. Hine[b] and Andreas Plückthun[a]

[a] *Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.*
[b] *College of Health and Life Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK*

[1] *Present address: Mabylon, 8952 Schlieren, Switzerland*
[2] *Present address: Neurimmune, 8952 Schlieren, Switzerland*
[3] *Present address: Protein Design and Informatics, Data and Predictive Science, Glaxo Smith Kline R&D, Collegeville, PA 19426, USA.*
[4] *Present address: University of Zurich, Dean's office of the Medical Faculty*

Correspondence should be addressed to A.P. (plueckthun@bioc.uzh.ch)
Email: plueckthun@bioc.uzh.ch

This PDF file includes:

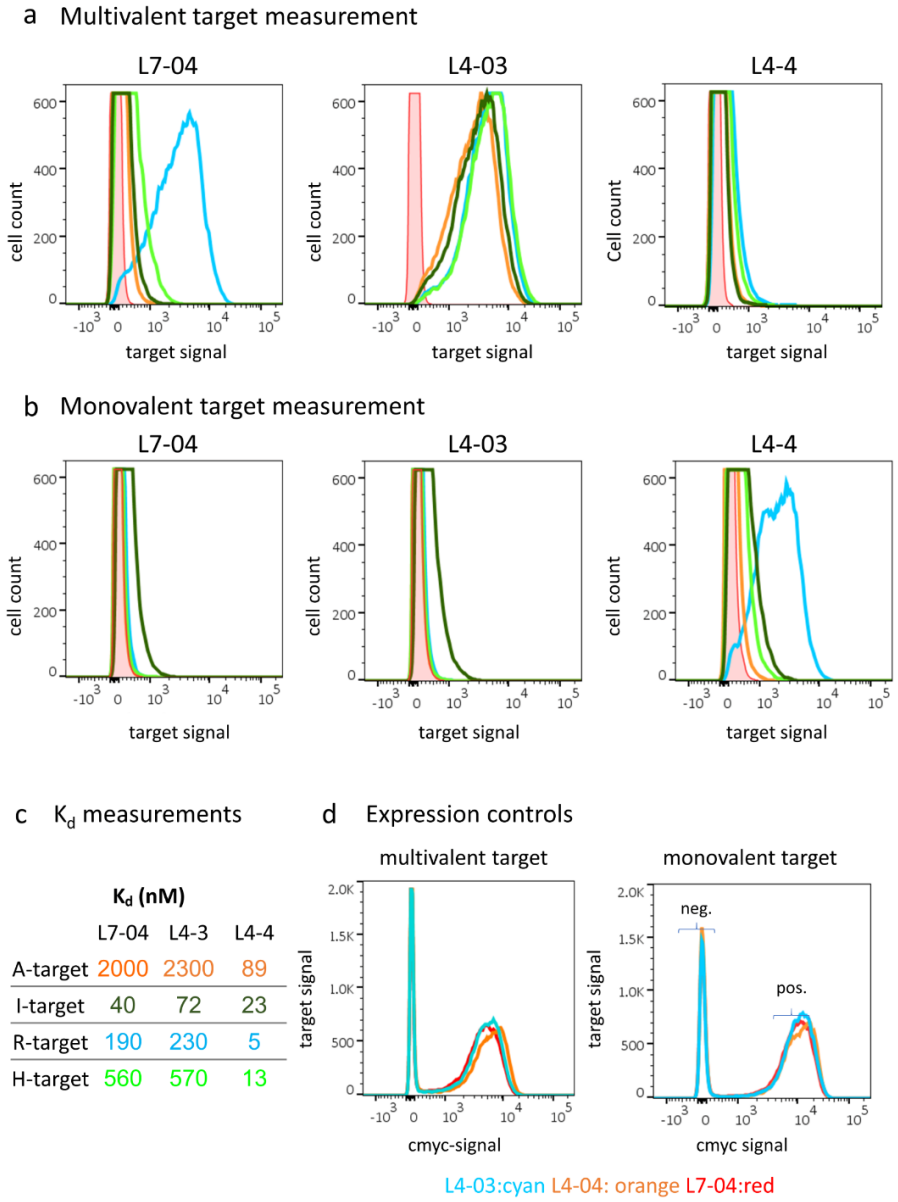
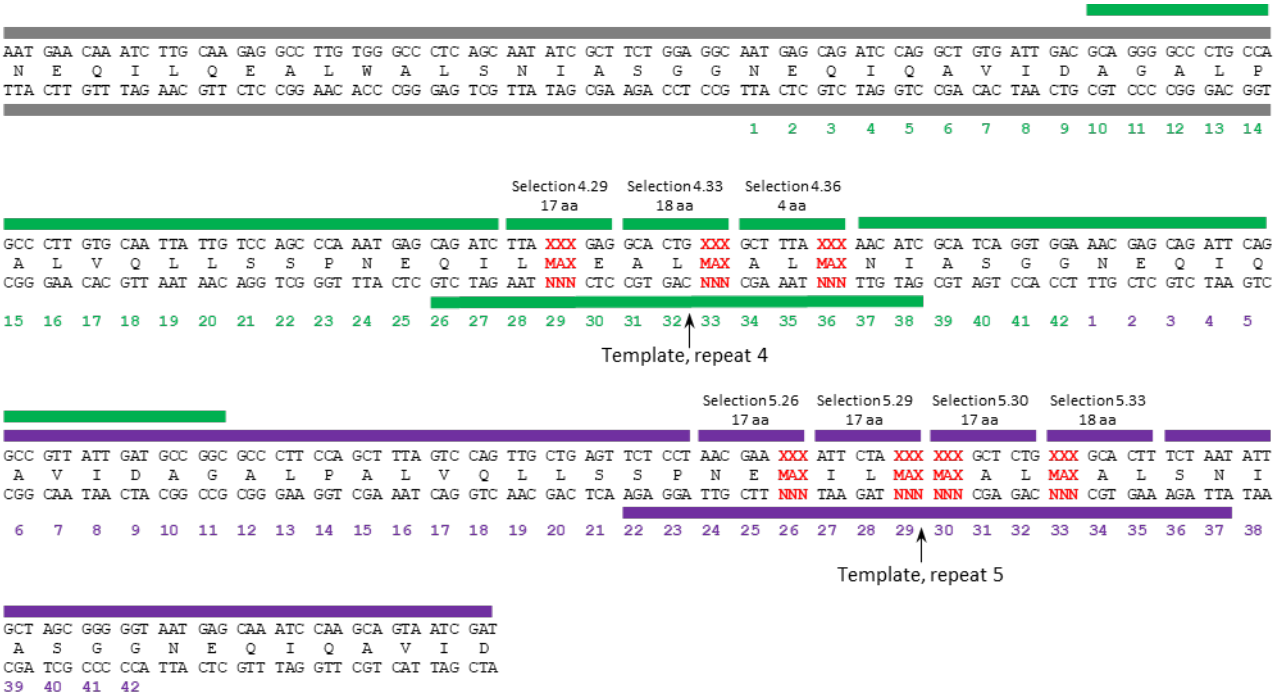


Fig. S1. Comparison of single-clone yeast analysis for multi- and monovalent targets. a) Measurement with multivalent target (SA-based). L7-04, L4-3 and L4-4 denote different ArmRPs, the colors denote different targets, as shown in the table with matching colors. When comparing the flow cytometry results with the K_d data obtained from the purified protein expressed in *E. coli* and measured by fluorescence anisotropy (c), it can be seen that the signal intensity does not correlate to the K_d data and seems randomly distributed. The highest affinity binders do not show the strongest shifts in flow cytometry. Shown with a red filled curve is the signal generated by SA only. b) Experiments using monovalent target. The signal intensities correlate better with the Kd data. c) Measured K_d data for the targets and dArmRP variants are shown. d) Comparison of the expression signal for the individual single clones used for the comparison between multivalent (SA based) and monovalent targets (GFP fusions). For both plots, a negative (non-expressing) and a positive (expression) population is observed. The expression control is performed by detecting the C-terminal c-myc-tag on the displayed ArmRP. The negative population results from freshly divided cells that show no display yet. As shown here, the expression for the investigated clones L4-03 (cyan), L4-04 (orange) and L7-04 (red) can be considered the same and does not explain the inconsistencies observed for the target signal.

a MAX randomisation strategy and design



b NGS analysis of codon representation

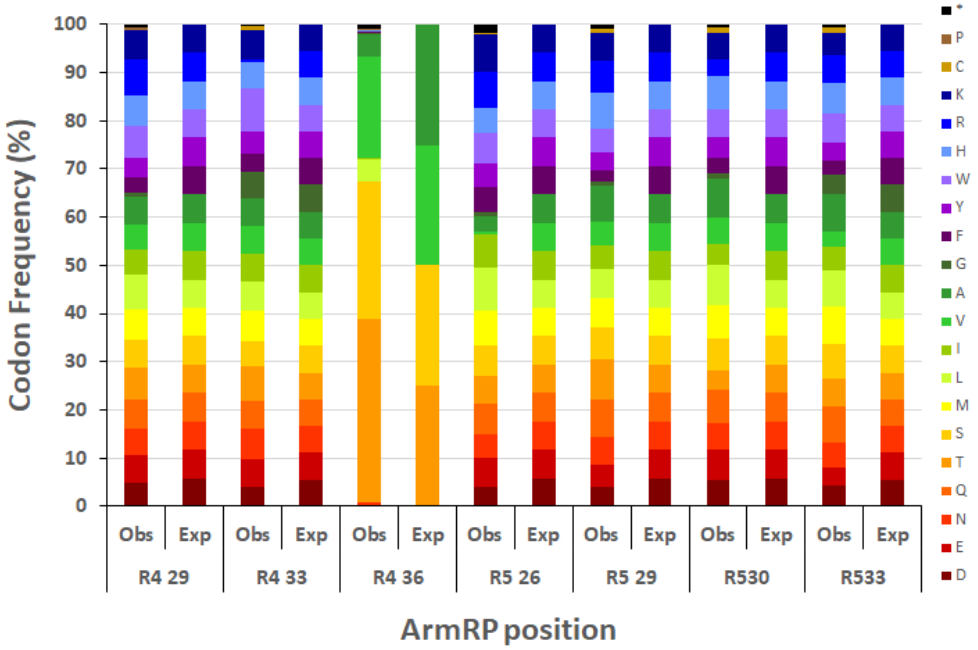


Figure S2. MAX randomization of insert for the second yeast display library (Tyr pocket). a) A region of the dArmRP gene encompassing repeats 3-6 was divided into three sections, comprising a conserved region (grey), repeat 4 (green) and repeat 5-6 (purple). The conserved region was synthesized from two overlapping oligonucleotides with the sequence 5'-AATGAACAAATCTTGCAAGAGGCCTTGTGGGCCCTCAGCAATATCGCTTCTGGAGGC-3' and 5'-TGGCAGGGCCCCTGCGTCAATCACAGCCTGGATCTGCTCATTGCCTCCAGAAGCGAT-3', which were hybridized, extended and cloned. Having confirmed the correct sequence, the resulting fragment was amplified with primers 5'-AATGAACAAATCTTGCAA-3' and 5'-GGCTGGCAGGGCCCCTGC-3'. Repeats 4 and 5-6 were created using MAX

randomization (21) with selection, template and end oligonucleotides as indicated by green and purple lines, respectively, where “XXX” indicates MAX codons of sequence Ala=GCT, Asp=GAT, Glu=GAA, Phe=TTT, Gly=GGT, His=CAT, Ile=ATT, Lys=AAG, Leu=TTG, Met=ATG, Asn=AAT, Gln=CAA, Arg=AGA, Ser=TCT, Thr=ACT, Val=GTT, Trp=TGG and Tyr=TAT as required. All selection oligonucleotides and both “End 2” oligonucleotides were obtained pre-phosphorylated. Repeat 4 was then amplified with primers 5’-GCAGGGGCCCTGCCAGCC-3’ and 5’-GCCGGCATCAATAACGGC-3’, and repeat 5-6 was amplified with primers 5’-GTTATTGATGCCGGCGCC-3’ and 5’-ATCGATTACTGCTTGGAT-3’. Thereafter, the three fragments were joined by overlap PCR and the completed randomized section was amplified by primers 5’-AATGAACAAATCTTGCAA-3’ and 5’-ATCGATTACTGCTTGGAT-3’. The resulting cassette, randomized at positions R429 (X1), R433 (X2), R436 (X3), R526 (X4), R529 (X5), R530 (X6) and R533 (X7) was assessed by NGS sequencing. b) NGS data of the randomized cassette were analyzed as described (21) and show good agreement between expected (design) and observed (experimental) codon frequencies at each randomized position.

Table S1. Amino acids sequence of all binders. Indicated in bold are the residues that deviate from the consensus designed sequence to bind the target amino acid.

Binder	*Amino acid sequence*
TyroM4	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**K**EAL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**Q**ALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
TyroM6	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**K**EAL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**ALVNIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**Q**ALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
Tyr pocket	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**K**EAL**E**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**L**IL**IR**AL**Q**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His Top1	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**D**EAL**Y**AL**T**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**WQ**AL**A**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top2	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**D**EAL**T**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**S**IL**WH**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top3	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**D**EAL**V**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**WH**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN

His-top4	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**Q**EAL**T**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**Q**IL**WY**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top5	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**F**EAL**T**AL**M**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**N**IL**LR**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top6	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**D**EAL**Y**AL**A**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**WQ**AL**D**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top7	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**E**EAL**W**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**FW**AL**G**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top8	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**W**EAL**F**AL**Y**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**ED**AL**G**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
His-top9	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**T**EAL**V**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**Q**IL**EY**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN

His-top10	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**Q**EAL**V**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**LY**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
enrichF4-1	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**Q**EAL**T**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNESILFFALEALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
enrichF4-2	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**D**EAL**L**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**A**IL**WH**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
enrichF4-3	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**H**EAL**I**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**Q**IL**FD**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN
enrichF4-4	GPGSELPQMVQQLNSPDQQELQSALWKLRNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILSSALGALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQLALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQIL**V**EAL**V**AL**V**NIASGG NEQIQAVIDAGALPALVQLLSSPNE**D**IL**QY**AL**E**ALSNIASGG NEQIQAVIDAGALPALVQLLSSPNEQILQEALWALSNIASGG NEQKQAVKEAGALEKLEQLQSHENEKIQKEAQEALEKLQSHKLN

Table S2. Data collection and refinement statistics.

PDB ID: 8QZN	
Resolution range (Å)	44.68 - 1.4 (1.45 - 1.4)
Space group	P 2_1 2_1 2_1
Unit cell	
a,b,c (Å)	50.85 93.61 76.84
α,β,γ (°)	90 90 90
Total reflections	957,100 (96,242)
Unique reflections	72,884 (7188)
Multiplicity	13.1 (13.4)
Completeness (%)	99.93 (99.94)
Mean I/sigma(I)	15.32 (1.08)
Wilson B-factor	22.89
R-merge	0.07441 (2.353)
R-meas	0.07755 (2.445)
R-pim	0.02154 (0.6606)
CC1/2	0.998 (0.52)
CC*	0.999 (0.827)
Reflections used in refinement	72879 (7188)
Reflections used for R-free	3645 (360)
R-work	0.1691 (0.3269)
R-free	0.1804 (0.3426)
CC(work)	0.969 (0.752)
CC(free)	0.971 (0.690)
Number of non-hydrogen atoms	3200
macromolecules	2762
ligands	200
solvent	328
Protein residues	346
RMS(bonds)	0.003
RMS(angles)	0.64
Ramachandran favored (%)	99.42
Ramachandran allowed (%)	0.58
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.05
Clashscore	2.61
Average B-factor	29.60
macromolecules	27.43
ligands	56.66
solvent	38.81
Number of TLS groups	1