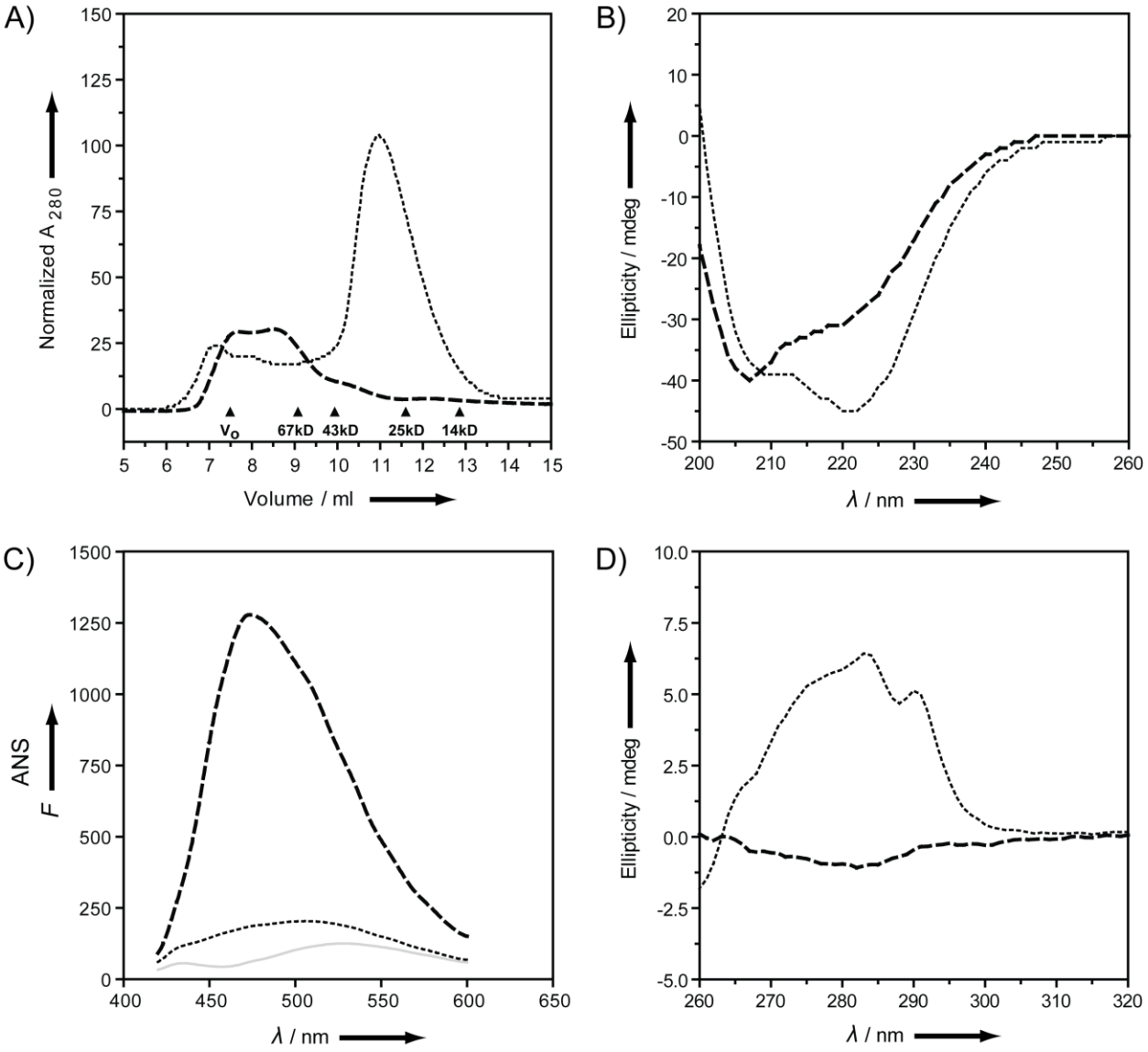

Figure S1. Biophysical characterization of GDPDwt (black dotted line) and GDPDmut (black dashed line) proteins. A) Investigation of quaternary structure by size exclusion chromatography. Monomeric GDPDwt elutes at 11 ml while oligomeric GDPDmut elutes at 8 ml. B) Far-UV circular dichroism spectroscopy to study the secondary structure. GDPDwt exhibits characteristic alpha-helical peaks at 208 nm and 222 nm while GDPDmut has lost some secondary structure, as indicated by signal decrease in 222 nm. C) 1-Anilinonaphthalene-8-sulfonic acid (ANS) fluorescence measurements (tertiary structure). ANS fluorescence is known to increase upon interaction with exposed hydrophobic protein patches, suggesting increased presence of exposed hydrophobic patches in GDPDmut, associated with the loss of the native GDPDwt tertiary structure. Buffer is shown as solid light gray line. D) Near-UV circular dichroism spectroscopy reveals that GDPDmut has significantly less tertiary structure relative to GDPDwt.

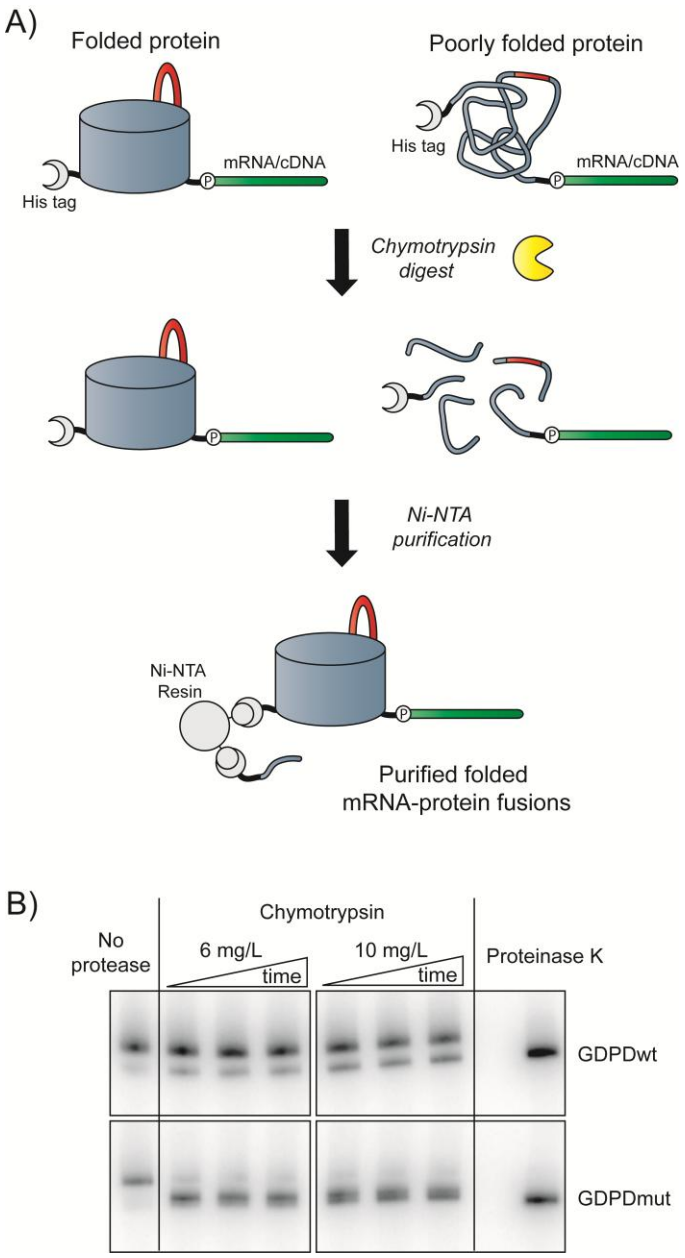

Figure S2. Folding selection by *in vitro* protease digestion. A) Schematic of the protease-digestion based selection by mRNA display. A mixture of folded and unfolded proteins that are covalently linked to their encoding mRNA/cDNA hybrid via puromycin (P) is subjected to chymotrypsin digest and then purified under denaturing conditions via Ni-NTA affinity chromatography. Only the cDNA of the well folded proteins is immobilized on the Ni-NTA resin, and amplified by PCR for downstream applications. In contrast, cleavage of unfolded proteins severs the link between His_6-tag and cDNA, thereby preventing immobilization of the cDNA. B) Analysis of chymotrypsin digestion during initial optimization steps. mRNA-displayed GDPDwt and GDPDmut proteins were incubated with chymotrypsin for 10, 15 and 20 min prior to analysis of crude digest reactions by SDS-PAGE. Undigested mRNA-protein fusions and fusions treated with proteinase K (to degrade proteins non-specifically) were used as controls for 0% and 100% digestion, respectively. Under these conditions GDPDmut is preferentially digested by chymotrypsin. Final optimization steps included the additional use of detergents and an analysis of digestion based on comparison of His_6-tag purified digested and undigested samples via scintillation counting.

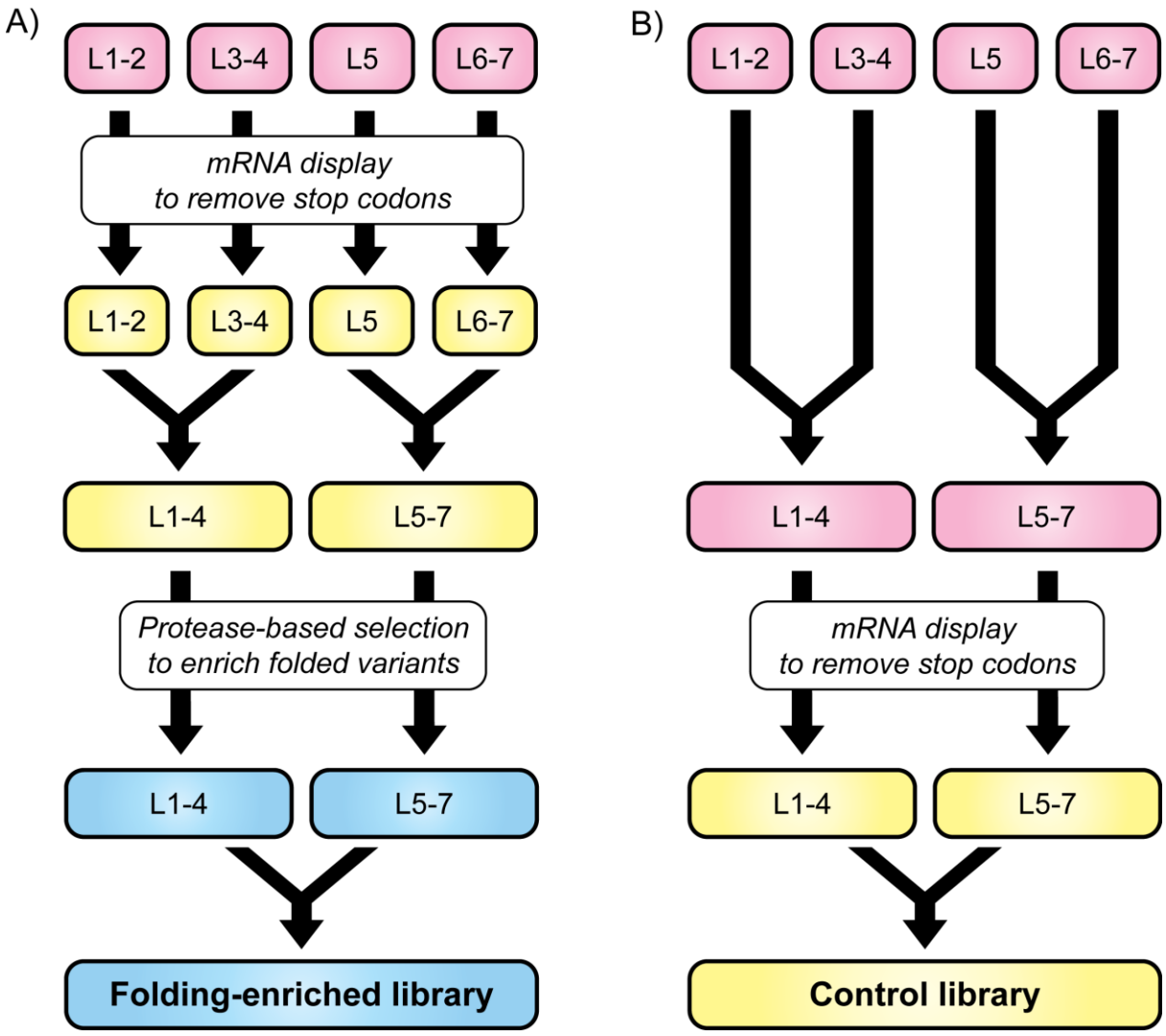

Figure S3. Step-wise strategy for the construction of libraries based on the $(\beta/\alpha)_8$ fold. Each box represents an individual library with randomized loop(s) (e.g. library L1-2). Libraries containing only a single randomized loop are not shown for clarity. The libraries were mRNA-displayed to remove stop codons or, in addition, were subjected to a folding selection by *in vitro* protease digestion (shown in yellow and blue, respectively). A) Construction of the folding-enriched library. Full length $(\beta/\alpha)_8$ barrel libraries were used in mRNA display and folding selection. B) Construction of the control library. $(\beta/\alpha)_4$ half-gene fragments of the barrel libraries were used in mRNA display to prevent the native $(\beta/\alpha)_8$ fold from biasing the randomized regions in the control library.

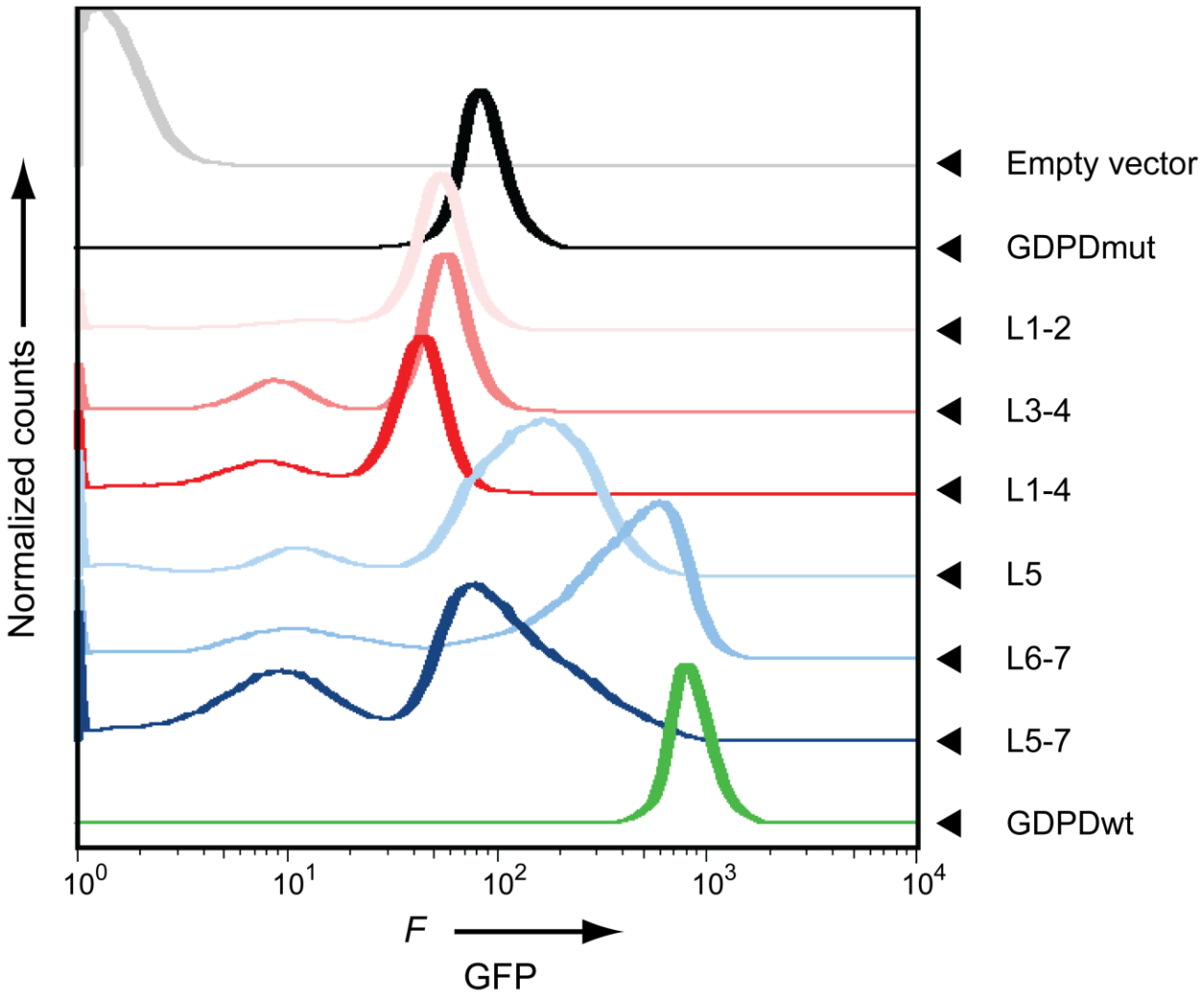

Figure S4. Assessment of folding by GFP-fusion assay. Fluorescence histograms of *E. coli* BL21(DE3) cells expressing the GFP-fused control constructs and intermediate libraries. The empty vector population was gated out on the histograms of cells transformed with the GDPD constructs. None of the libraries shown here had been selected yet for folding.

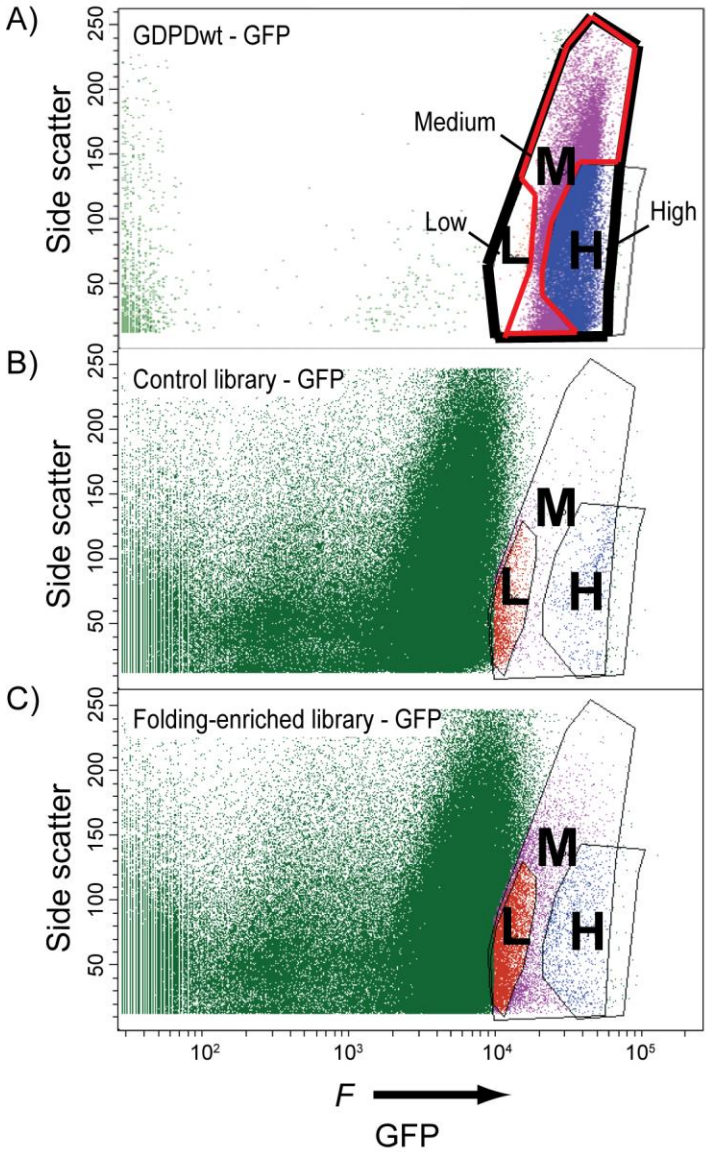

Figure S5. Analysis of library populations by fluorescence-activated cell sorting experiments (FACS). Side scatter versus GFP fluorescence dot plots are shown for A) GDPDwt-GFP, B) control library and C) folding-enriched library. In the top panel, the GDPDwt-like population of cells is framed with a thick black line. The low (L) and high (H) populations of library-GFP cells shown in the lower two panels were collected during the sorting experiment for further analysis. The number of cells in the medium (M) population was estimated (see Table S2).

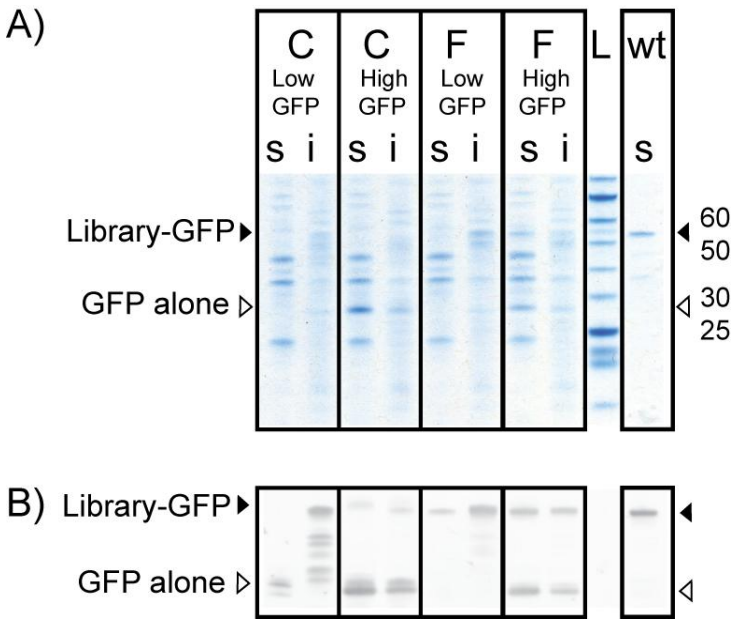

Figure S6. Comparison of the soluble and insoluble fractions of the FACS sorted library populations. A) SDS-PAGE gel Coomassie-stained. B) Western blot of SDS-PAGE gel using polyclonal anti-GFP antibody. Approximate molecular weights for library-GFP fusions (57 kD, black triangle) and for GFP alone (27 kD, white triangle). Equal amounts of soluble (s) and insoluble (i) fractions were loaded for each library. The soluble fraction of over-expressed GDPDwt-GFP fusion is shown as positive control.

Table S1. GFP-fused *in vivo* folding assessment of intermediate libraries. [a]

Species	Mode of GFP fluorescence [b]	% of cells with GDPDwt-GFP fluorescence [c]
GDPDmut	10.4	0.02
L1-2	6.7	0.90
L3-4	7.2	0.05
L1-4	5.4	0.09
L5	21.3	9.3
L6-7	72.3	51.6
L5-7	10	6.9
GDPDwt	100	98.5

[a] Constructs were transformed into *E.coli* BL21(DE3) strain. Stop-codon containing libraries that have not been subjected to mRNA display. Prior to analysis all data were gated to exclude cell populations that matched fluorescence and scatter profiles of cells transformed with empty plasmid.
[b] Values normalized to the mode GFP fluorescence of GDPDwt-GFP.
[c] Wild type cells were gated on the forward scatter versus GFP contour plot to include ~98% of all wild type cells.

Table S2. Fraction of soluble library-GFP fusions in the FACS-sorted populations.

Library population [a], [b]	% Non-empty cells [c]	% soluble GFP fusions of sorted cells		% soluble GFP fusions of all cells [e]
		By Western blot [d]	By SDS-PAGE	
Control, low GFP	0.86%	0.24%	<5% (0/18 clones)	0.002%
Control, medium GFP	0.23% [f]	0.24% - 5.76% [g]	not available	0.001% - 0.013%
Control, high GFP	0.31%	5.76%	<5% (0/19 clones)	0.018%
Estimated % soluble proteins in control library				0.020% - 0.033% [h]
Folding-enriched, low GFP	3.18%	18.20%	29% (4/14 clones)	0.579%
Folding enriched, mid GFP	1.68% [f]	18.20% - 26.70%[g]	not available	0.306% - 0.449%
Folding-enriched, high GFP	0.54%	26.70%	24.6% (20/81 clones)	0.144%
Estimated % soluble proteins in folding-enriched library				1.029% - 1.172% [h]
Enrichment of soluble proteins after folding selection [i]				35 to 50-fold

[a] Library populations fell into the GDPDwt-like profile window defined by side scatter vs. GFP fluorescence dot plot of GDPDwt-GFP construct (Figure S5).
[b] Only the low and high GFP populations were sorted during the FACS experiment.
[c] % Non-empty cells defined as the ratio (# of cells analyzed - # of cells with empty-vector) / (# of cells analyzed). A total of 1.4% of the control library and 5.4% of the folding-enriched library fell into the GDPDwt-like window used for sorting and analysis.
[d] % Soluble GFP fusions of sorted cells calculated from Western blot analysis (Figure S6) using Image J to quantitate the intensities of all anti-GFP stained bands. Defined as the ratio (intensity of soluble GFP-fusions) / Σ (intensity of all anti-GFP stained bands).
[e] % Soluble GFP fusions of all cells calculated as (% non-empty cells x % soluble GFP fusions of sorted cells) for the population of interest.
[f] % Non-empty cells for the unsorted medium GFP population was calculated as the difference in % non-empty empty cells populations (GDPDwt-like – low GFP – high GFP).
[g] Values are lower and upper estimates based on % soluble GFP fusions of sorted cells in the low and high GFP populations.
[h] Estimated % soluble proteins in a library calculated as Σ (% soluble GFP fusions of all cells) for low, medium and high GFP populations.
[i] Fold improvement defined as the ratio (estimated % soluble proteins in the folding-enriched library)/ (estimated % soluble proteins in the control library).

Table S3. List of primers used during library construction.

Primer	Sequence
041B [a]	GCCTTCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGGGCAGCGATAAGATCCACC
042 [a]	TTAATAGCCGGTGCCAGATCCAGACATTCC
018	TATGACCAAGCTTCCAGGGTGTTTTCCAGATACTTGGCGGAGTAACCSNNSNNGCCCAGCACAATCAC
019	CTGGAAAACACCCTGGAAG
047	AACAACAAGGTCTCAGGCGCTTTAAATCSNNSNNSNNSNNSNNGCTCACGACCACCTTGCC
048	AACAACAAGGTCTCGCGCCTGTTCGGTCTGGACG
028	AACAACAAGGTCTCCCTCACGTTCSNNSNNSNNSNNGATTTCGATGTTGATGATCTTG
029	AACAACAAGGTCTCGTGAGGCCGCGGACGCAGTGCTGGAGATCAGCAAAAAGCGTAAG
022	TGGAGGTCCTTGGTACCCTTGAATTTTTCATCCAGCAGGTCCAGATCSNNSNNSNNSNNGGAGCTGAAAATCAGGTTCTTAC
023	GATCTGGACCTGCTGGATG
012	AGTAGAGGTACCAAATACGGTTATNNSATCNNSNNSNNSNNSNNSTACGGTTCCATTGAAAATTTCG
015	CTTCGTCGATCAGATAACCG
003	AGTAAAGAGCTCAAAGGCCTGSNNSNNSNNSNNSNNSNNCACGTGCAGAGAGTACGGAC
017	AGGCCCTATCAGGCCTTTGAG
013	AGTAGATACGTATTTTGCGGTAGATTTCCGGATCSNNSNNSNNSNNSNNCACAAAAATCACGATGCC
016	CTGAAAGAGCTGACCGATGG
039	TTAATAGCCGGTGCCAGATCCAGACATTCCCATTTTGTCATCGTCATCCTTATAGTCGGAGCCACCGGTCTCACCCTTGAATTTTTCATCCAGC
040B	TTCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGCACCATCACCATCACCATGGTCTCAAGGGTACCAAATACGGTTAT
035B	GACTGAACTGATGCCTATATGCTTGTCCGTCCAGATTGGTCTCGTGTTTGAGAACGTGTCCGATG
036B	AACTCAAGTATCGCTATGCCGGTCAATAACCGAGGTCTCAAACACTTCCTTCAGGGTGG
049	CTGCGTGGTAGCTCGTAGCACAGACTCAGCGGATACACACAGAGGTCTCAAAAATTCAAGGGTACCAAATACGG
050	GCACTCCGCTTAGATAGATAGCCAGAAGACAGACAAGGTCTCATTTTTCATCCAGCAGGTCCAG
037B	AAGTAGCATAGAGTGTCGCTCTGGATGTCAAGGTCTCAAAAGGAGCGTCCGTACTCTCTG
038B	ATGATAGCAGATGGACTTAGATTTCCGGTCAGGTGCCGAGGTCTCCCTTTTCCACGCGCTCCACG
GDPDx_001 [b]	TCTGTAAACCATGGATGGGCAGCGATAAGATCCAC
GDPDx_002 [b]	CTGTGCGCTCGAGTTAATAGCCGGTGCCAGATCC
GDPDx_003 [c]	GAAGGAGATATACATATGGGCAGCGATAAGATC
GDPDx_004 [c]	GGAGCCAGCGCGGCCGCCATAGCCGGTGCCAGATCCAG
GDPDmut_Fw [d]	GAAGCCGGCGCGAATCGTGAGGAGCTGGATGTG
GDPDmut_Rev [d]	CACATCCAGCTCCTCACGATTCGCGCCGGCTTC
030 [e]	TTCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGGGCAGCGATAAGATCGTGATTGTGCTGGGCCATCGCGG

[a] Primers used as standard primers to amplify any full length GDPD-based template during the library construction.
[b] Primers used to amplify template DNA for insertion into the pET28a plasmid for protein expression.
[c] Primers used to amplify template DNA for insertion into the pER13 plasmid for the GFP-fusion assay.
[d] Primers used to generate pET28/GDPDmut from the pET28/GDPDwt template.
[e] Primers 030 and 042 were used to generate GDPDmut(-His_6) from the pET28/GDPDmut template.

Table S4. Fragments used in library assembly. [a], [b]

	5'-Fragment				3'-Fragment				
Library [d]	Name	Template	FW	REV	Name	Template	FW	REV	Restriction enzyme used [c]
L1	L1_A	pET28/GDPDwt	041B	018	L1_B	pET28/GDPDwt	019	042	*HindIII*
L2	L2_A	pET28/GDPDwt	041B	047	L2_B	pET28/GDPDwt	048	042	*BsaI*
L3	L3_A	pET28/GDPDwt	041B	028	L3_B	pET28/GDPDwt	029	042	*BsaI*
L3(-His6)	L3_A(-His6)	pET28/GDPDwt	030	028	L3_B	pET28/GDPDwt	029	042	*BsaI*
L4	L4_A	pET28/GDPDwt	041B	022	L4_B	pET28/GDPDwt	023	042	*KpnI*
L5	L5_A	pET28/GDPDwt	041B	015	L5_B	pET28/GDPDwt	012	042	*KpnI*
L6	L6_A	pET28/GDPDwt	041B	003	L6_B	pET28/GDPDwt	017	042	*SacI*
L7	L7_A	pET28/GDPDwt	041B	013	L7_B	pET28/GDPDwt	016	042	*SnaBI*
L1-2 [e]	L1-2_A	L1	041B	047	L2_B	pET28/GDPDwt	048	042	*BsaI*
L3-4 [e]	L3-4_A	pET28/GDPDwt	041B	028	L3-4_B	L4	029	042	*BsaI*
L6-7 [e]	L6-7_A	L6	041B	034	L6-7_B	L7	033	042	*BsaI*
L1-4 [f]	C-L1-4_A	L1-2	041B	036B	C-L1-4_B	L3-4	035B	042	*BsaI*
L5-7 [g]	C-L5-7_A	L5	041B	038B	C-L5-7_B	L6-7	037B	042	*BsaI*
L1-4 (m)	F-L1-4_A	L1-2 (m)	041B	036B	F-L1-4_B	L3-4 (m)	035B	042	*BsaI*
L5-7 (m)	F-L5-7_A	L5 (m)	041B	038B	F-L5-7_B	L6-7 (m)	037B	042	*BsaI*
Control library	C-frag_A	L1-4 frag (m)	041B	050	C-frag_B	L5-7 frag (m)	049	042	*BsaI*
Folding-enriched library	F-frag_A	L1-4 (m&p)	041B	050	F-frag_B	L5-7 (m&p)	049	042	*BsaI*

[a] Individual libraries were generated by amplifying the 5'- and 3'-fragments of the library using the specified template, and FW /REV primer pair.
[b] Libraries denoted with (m) have been subjected to mRNA display; libraries denoted with (p) have been subjected to *in vitro* folding selection by protease digestion.
[c] 5'- and 3'-fragments were digested with the indicated restriction enzyme (RE), gel-purified and ligated to produce the individual libraries. The *BsaI* restriction site, which is absent in the parent GDPDwt scaffold, was introduced into the 5-‘ and 3’ fragments by PCR amplification and was removed again during the gel purification of digested fragments.
[d] Individual libraries were gel purified after the ligation reaction. A fraction of the purified library was used as a template for the successive steps in library assembly.
[e] Libraries were subjected to mRNA display during folding-enriched library construction and then PCR amplified with 041B/042 primer pair to restore T7 promoter sequence lost during the transcription process.
[f] Library was used as template with 041B/039 primer pair to generate the control library fragment (L1-4 frag) subjected to mRNA display.
[g] Library was used as template with 040B/042 primer pair to generate the control library fragment (L5-7 frag) subjected to mRNA display.

.

Supplementary Data

DNA sequence of the GDPDwt scaffold used as template for library assembly

GCCTTCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGGGCAGCGAT
AAGATCCACCATCACCATCACCATGTGATTGTGCTGGGCCATCGCGGTTACTCCGCCAAG
TATCTGGAAAACACCCTGGAAGCTTTCATGAAAGCGATCGAAGCCGGCGCGAATGGTGTG
GAGCTGGATGTGCGCCTGTCTAAAGACGGCAAGGTGGTCGTGAGCCATGATGAAGATTTA
AAGCGCCTGTTCGGTCTGGACGTCAAAATCCGTGACGCCACCGTGTCTGAACTGAAAGAG
CTGACCGATGGCAAAATTACCACCCTGAAGGAAGTGTTTGAGAACGTGTCCGATGACAAG
ATCATCAACATCGAAATCAAGGAACGTGAGGCCGCGGACGCAGTGCTGGAGATCAGCAAA
AAGCGTAAGAACCTGATTTTCAGCTCCTTTGATCTGGACCTGCTGGATGAAAAATTCAAG
GGTACCAAATACGGTTATCTGATCGACGAAGAGAACTACGGTTCCATTGAAAATTTCGTG
GAGCGCGTGGAAAAGGAGCGTCCGTACTCTCTGCACGTGCCCTATCAGGCCTTTGAGCTC
GAATATGCGGTGGAGGTGCTGCGCTCCTTCCGTAAAAAGGGCATCGTGATTTTTGTGTGG
ACCCTGAATGATCCGGAAATCTACCGCAAAATACGTAGAGAGATCGATGGTGTGATTACC
GACGAAGTGGAGCTGTTTGTGAAACTGCGTGGCGGCAGCGGTGGCTCCGACTATAAGGAT
GACGATGACAAAATGGGAATGTCTGGATCTGGCACCGGCTATTAA

Color code:
T7 transcription promoter / TMV translation enhancer
Thio6/His_6
GDPDwt
$(GGS)_2$ spacer
FLAG tag
Puromycin-crosslinking region

Primers 041B and 042 were used as standard primers to amplify any full length GDPD-based template during the library construction. Primers GDPDx_001 and GDPDx_002 were used to amplify template DNA for insertion into the pET28a plasmid for protein expression. Primers GDPDx_003 and GDPDx_004 were used to amplify template DNA for insertion into the pER13 plasmid for the GFP-fusion assay. Primers GDPDmut_Fw and GDPDmut_Rev were used to generate pET28/GDPDmut from the pET28/GDPDwt template. Primers 030 and 042 were used to generate GDPDmut(-His_6) from the pET28/GDPDmut template.

Sequence alignment of the six soluble F(s) clones characterized in this manuscript (Figure 4).

Alignment was performed using Clustal Omega server, Clustal O(1.1.0)

Loop residues randomized during library construction are highlighted in green.

* = column contains identical amino acid

: = column contains different but highly conserved amino acids

```
GDPDwt      MGSDKIHHHHHHVIVLGHRGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)1       MGSDKIHHHHHHVIVLGSRGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)2       MGSDKIHHHHHHVIVLGHRGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)3       MGSDKIHHHHHHVIVLGRLGYSAKYLENTLEAFMRAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)4       MGSDKIHHHHHHVIVLGNGGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)5       MGSDKIHHHHHHVIVLGRVGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
F(s)6       MGSDKIHHHHHHVIVLGRLGYSAKYLENTLEAFMKAIEAGANGVELDVRLSKDGKVVVSHDEDLKRLFG
            ****************  ***************:**********************************

GDPDwt      LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)1       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKTINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)2       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)3       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)4       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)5       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
F(s)6       LDVKIRDATVSELKELTDGKITTLKEVFENVSDDKIINIEIKEREAADAVLEISKKRKNLIFSSFDLDL
            ********************************** ********************************

GDPDwt      LDEKFKGTKYGYLIDE-ENYGSIENFVERVEKERPYSLHVP----YQAFELEYAVEVLRSFRKKGIVIF
F(s)1       LDEKFKGTKYGYLIDE-ENYGSIENFVERVEKERPYSLHVTPTLLSQAFELEYAVEVLRSFRKKGIVIF
F(s)2       LDEKFKGTKYGYKISLWASYGSIENFVERVEKERPYSLHVSSTKDAQAFELEYAVEVLRSFRKKGIVIF
F(s)3       LDEKFKGTKYGYKIGRGGGYGSIENFVERVEKERPYSLHVYSGSPLQAFELEYAVEVLRSFRKKGIVIF
F(s)4       LDEKFKGTKYGYIISLKDTYGSIENFVERVEKERPYSLHVQRASFKQAFELEYAVEVLRSLRKKGIVIF
F(s)5       LDEKFKGTKYGYGIAEGLVYGSIENFVERVEKERPYSLHVELEFMIQAFELEYAVEVLRSFRKKGIVIF
F(s)6       LDEKFKGTKYGYLIDE-ENYGSIENFVERVEKERPYSLHVAVGRVLQAFELEYAVEVLRSFRKKGIVIF
            ************ *     ********************      **************:********

GDPDwt      VWT-LNDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)1       VKNNVCDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)2       VPCLRCDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)3       VASSTHDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)4       VAPDLPDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)5       VRADMSDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
F(s)6       VTSVTRDPEIYRKIRREIDGVITDEVELFVKLRGGSGGSDYKDDDDKMGMSGSGTGY
            *     ************************************************
```