

Supplementary material

The structural components of the *Azotobacter vinelandii* iron-only nitrogenase, AnfDKG, form a protein complex within the plant mitochondrial matrix

Johnston E.^{1,5}, Okada S.¹, Gregg C. M.², Warden A. C.¹, Rolland V.², Gillespie V.², Byrne K.³, Colgrave M. L.^{3,4}, Eamens A. L.⁶, Allen R. S.^{2*}, Wood C. C.²

¹CSIRO Environment, GPO Box 1700, Acton, ACT 2601, Australia

²CSIRO Agriculture and Food, GPO Box 1700, Acton, ACT 2601, Australia

³CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia

⁴Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, 306 Carmody Rd, St. Lucia, QLD 4067, Australia

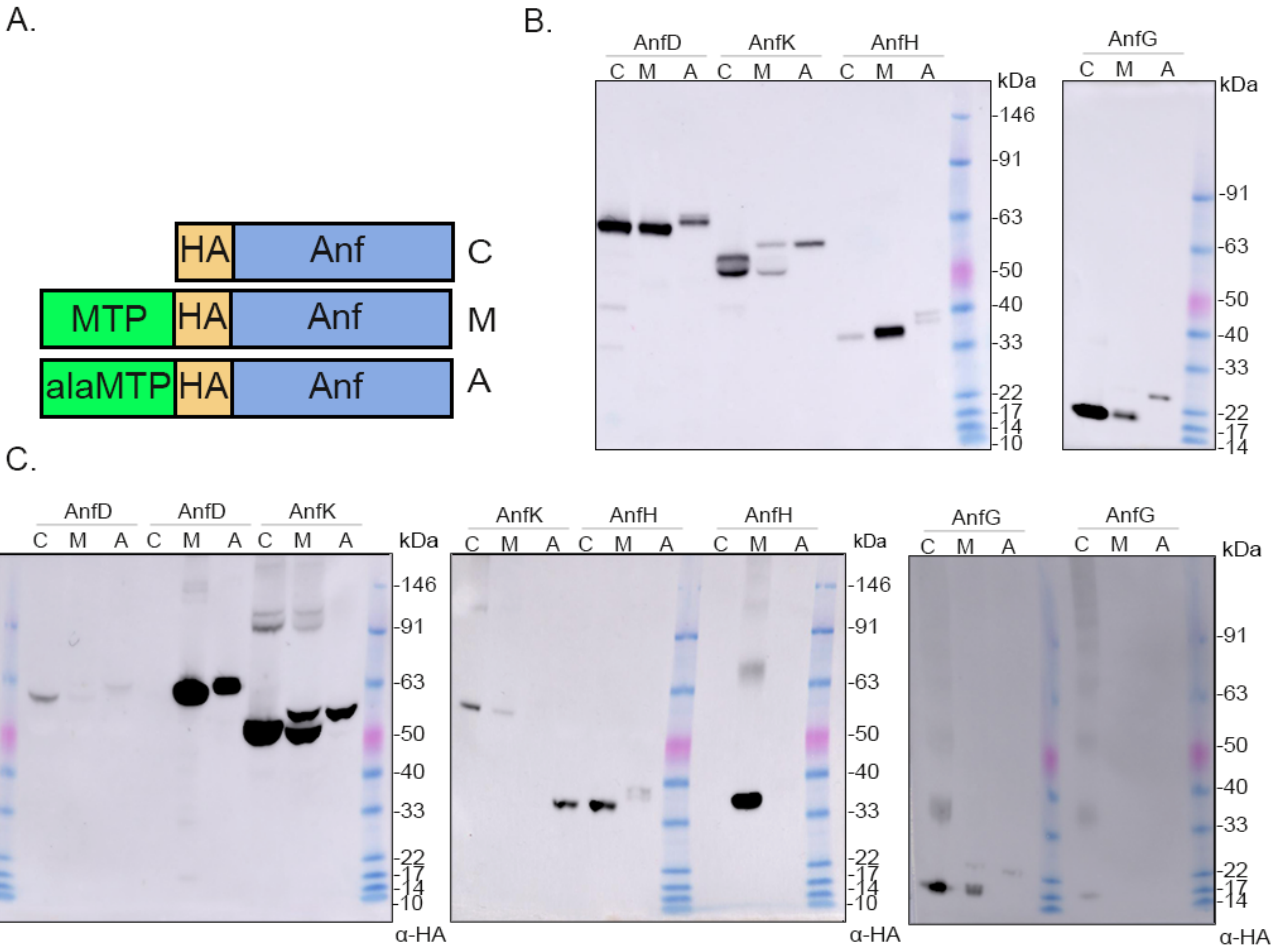
⁵School of Environmental and Life Sciences, University of Newcastle, University Dr, Callaghan NSW 2308, Australia

⁶School of Health, University of the Sunshine Coast, Maroochydore, QLD 4558, Australia

*corresponding author rob.allen@csiro.au

Table of Contents

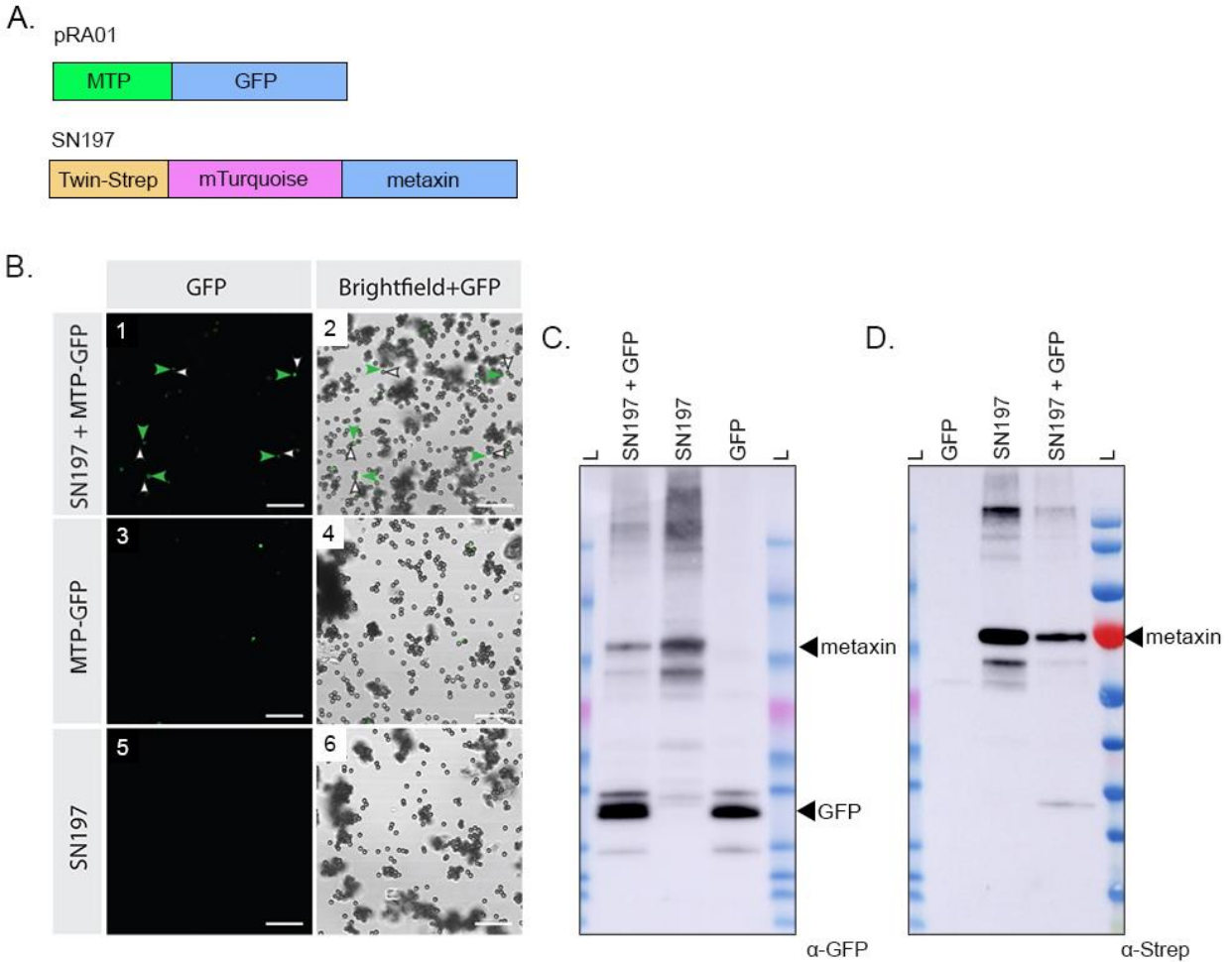
Supplementary Figure 1	Analysis of total, soluble and insoluble protein fractions of individual Anf proteins when targeted to the plant cytosol or mitochondrial matrix.....	3
Supplementary Figure 2	Mitochondrial enrichment by affinity purification	4
Supplementary Figure 3	Extracted ion chromatogram (XIC) for AnfD tryptic peptides from the eluates after streptactin purification from the multigene construct SL37.....	5
Supplementary Figure 4	Extracted ion chromatogram (XIC) for AnfK tryptic peptides from the eluates after streptactin purification from the multigene construct SL37.....	6
Supplementary Figure 5	Extracted ion chromatogram (XIC) for AnfH tryptic peptides from the eluates after streptactin purification from the multigene constructs SL37.	7
Supplementary Figure 6	Extracted ion chromatograms (XIC) for four selected peptides (I) ITNDDVIK, (II) EFLEEFATAIGTK, (III) EYGELAR and (IV) DNIVQK using MRM for AnfH.....	8
Supplementary Figure 7	Extracted ion chromatogram (XIC) for AnfG tryptic peptides from the eluates after streptactin purification from the multigene construct SL37.....	9
Supplementary Figure 8	Extracted ion chromatograms (XIC) for three selected peptides; (I) IDELTDYIMK, (II) ELLCGEPVDLSTSHDR and (III) EEIGSLMQGLK using MRM for AnfG.....	10
Supplementary Table 1	Multiple reaction monitoring transitions of AnfG, H, D and K peptides for targeted liquid chromatography – mass spectrometry (LC-MS).	11
Supplementary Data Set 1	Schematic outline of the peptides detected by LC-MS/MS for AnfDKGH isolated from a range of plant-based expression experiments	17
Supplementary Data Set 2	Schematic outline of the peptides detected by LC-MS/MS for MTP-AnfDKH from enriched plant mitochondria.....	18
Supplementary Data Set 3	Codon-optimized DNA sequences and translated protein sequences of <i>anfD</i> , <i>anfK</i> , <i>anfG</i> and <i>anfH</i> genes used in this study.	20



50 **Supplementary Figure 1** | Analysis of total, soluble and insoluble protein fractions of individual Anf
51 proteins when targeted to the plant cytosol or mitochondrial matrix.

52 **A)** Schematic representation of the MTP-HA-Anf, alaMTP-HA-Anf and the HA-Anf constructs. **B)** Western
53 blot analysis of AnfD, AnfK, AnfH and AnfG in total protein extracts from *N. benthamiana* leaf. **C)**

54 Western blot analysis of AnfD, AnfK, AnfH and AnfG in soluble and insoluble protein extracts from *N.*
55 *benthamiana* leaf. C, cytosolic expression; M, mitochondrial targeted; A, alanine scanned MTP. Parts of
56 these western blot images were used to construct Fig. 2 in the main text.



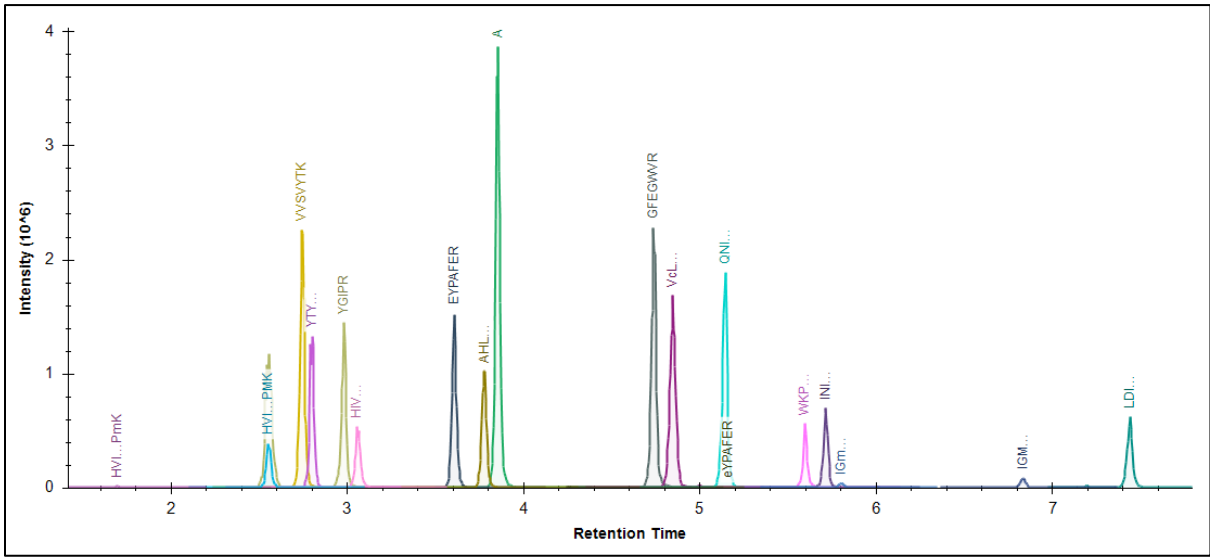
Supplementary Figure 2 | Mitochondrial enrichment by affinity purification

A) Schematic representation of constructs MTP-GFP, pRA01, and Twin-Strep-mTurquoise-metaxin, SN197.

B) Images of mitochondrial enrichment by affinity purification using Twin-Strep-mTurquoise-metaxin.

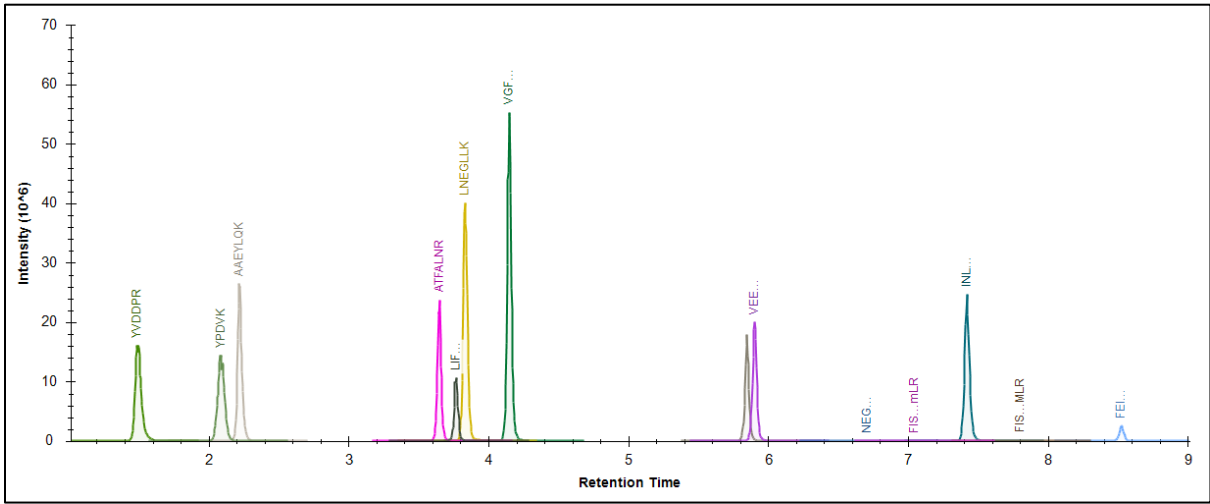
Panel 1 and 2 show the enrichment of fluorescent mitochondria from leaves expressing both, MTP-GFP and Twin-Strep-mTurquoise-metaxin. Panel 3 and 4 show the eluate of the negative control sample, from leaves expressing MTP-GFP only. Panel 5 and 6 show the eluate from leaves expressing Twin-Strep-mTurquoise-metaxin only. The imaging revealed an increase in GFP fluorescence for the sample expressing both MTP-GFP and Twin-Strep-mTurquoise-metaxin compared to either of the controls. Panels 1, 3, 5 show the fluorescent channel, dark field image. Panels 2, 4 and 6 show the fluorescent channel, bright field image. The images shown are representative of three biological replicates. **C)** Western blot analysis of eluate samples from the mitochondrial enrichment. Western blots were probed with the α -GFP antibody for of GFP and α -strep antibody for visualization of metaxin. Twin-Strep-mTurquoise-metaxin was also detected as indicated by α -GFP due to the sequence similarity of mTurquoise and GFP.

AnfD



Supplementary Figure 3 | Extracted ion chromatogram (XIC) for AnfD tryptic peptides from the eluate of the StrepTactin purification from leaves infiltrated with SL37. Fourteen MRM peptides were assessed for AnfD. Each peptide displayed peak profiles with high response and the product ions (Q3 m/z, n = 4) profiled with the same retention times.

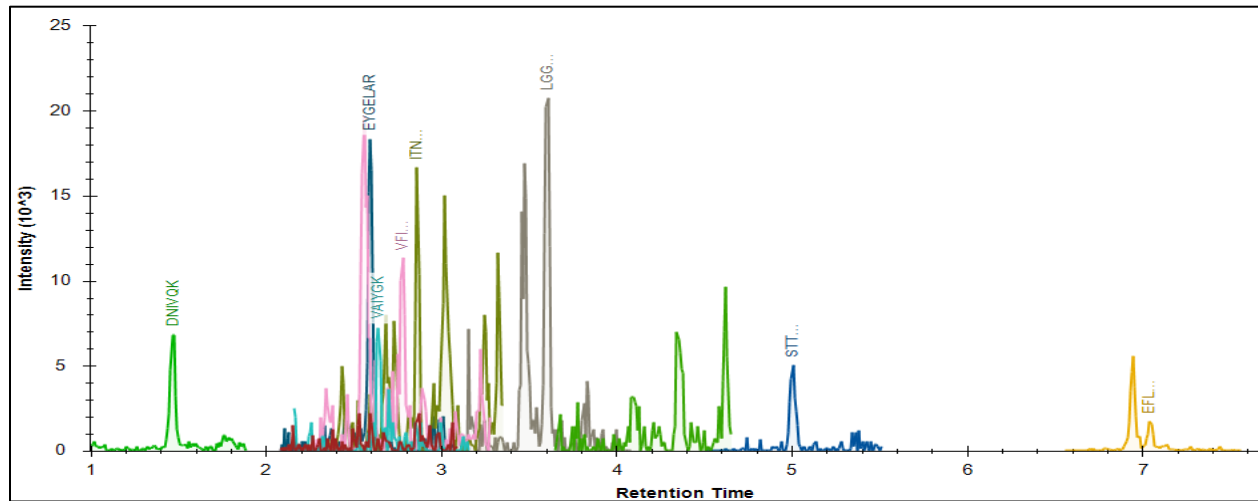
AnfK



Supplementary Figure 4 | Extracted ion chromatogram (XIC) for AnfK tryptic peptides from the eluate of the streptactin purification from leaves infiltrated with SL37.

Thirteen MRM peptides were assessed for AnfK. Each peptide displayed peak profiles with high response and the product ions (Q3 m/z, n = 4) profiled with the same retention times.

87 AnfH

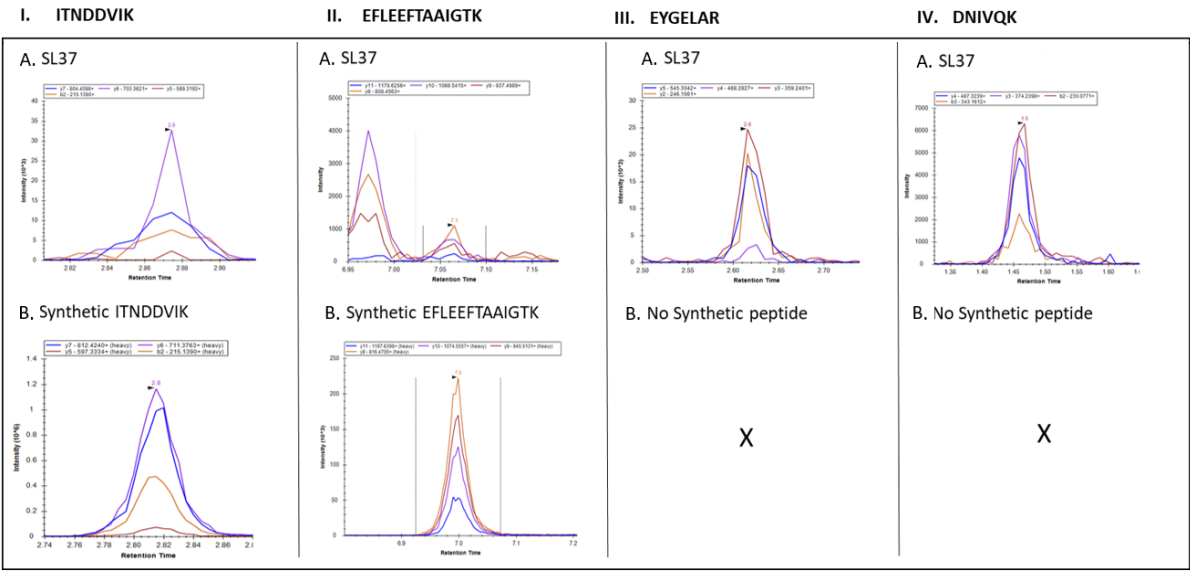


88

89 **Supplementary Figure 5** | Extracted ion chromatogram (XIC) for AnfH tryptic peptides from the
 90 eluate of the streptactin purification from leaves infiltrated with SL37.

91 Eight MRM peptides were assessed for AnfH. Each peptide displayed peak profiles with low but
 92 detectable response and the product ions (Q3 m/z, n = 4) profiled with the same retention times.

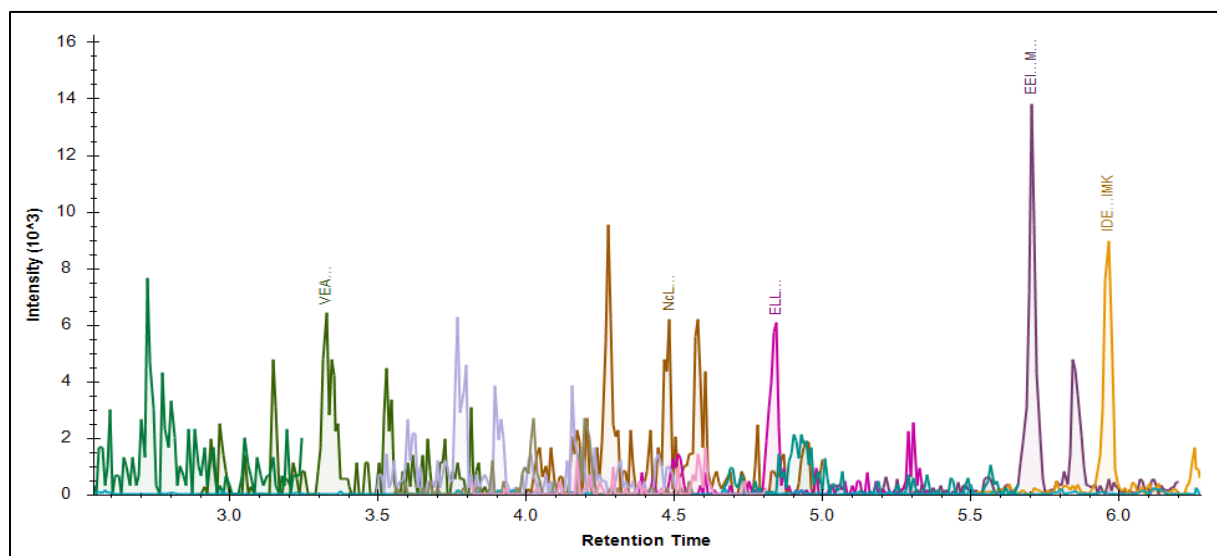
93



Supplementary Figure 6| Extracted ion chromatograms (XIC) for four selected peptides (I) ITNDDVIK, (II) EFLEEFATAIGTK, (III) EYGELAR and (IV) DNIVQK using MRM for AnfH.

The eluate after StrepTactin purification from leaves infiltrated with SL37 shows the presence of AnfH at a low level. (B) Two peptides (ITNDDVIK and EFLEEFATAIGTK) were synthesised as heavy labelled peptides (JPT Peptide Technologies GmbH, Berlin, Germany) to distinguish interference peaks (e.g. EFLEEFATAIGTK interference at RT= 6.95) and confirm that peptides detected displayed peak profiles with the expected product ion profile (Q3 m/z, n = 4), with the transition ions showing similar retention times (RT), 2.82 and 7.0, respectively. This supports the minimum threshold of protein identification with the identification of two unique AnfH peptides and is supported with the identification of ITNDDVIK using 6600TF LC-MS/MS in eluate SL37.

107 **AnfG**



108

109 **Supplementary Figure 7** | Extracted ion chromatogram (XIC) for AnfG tryptic peptides from the

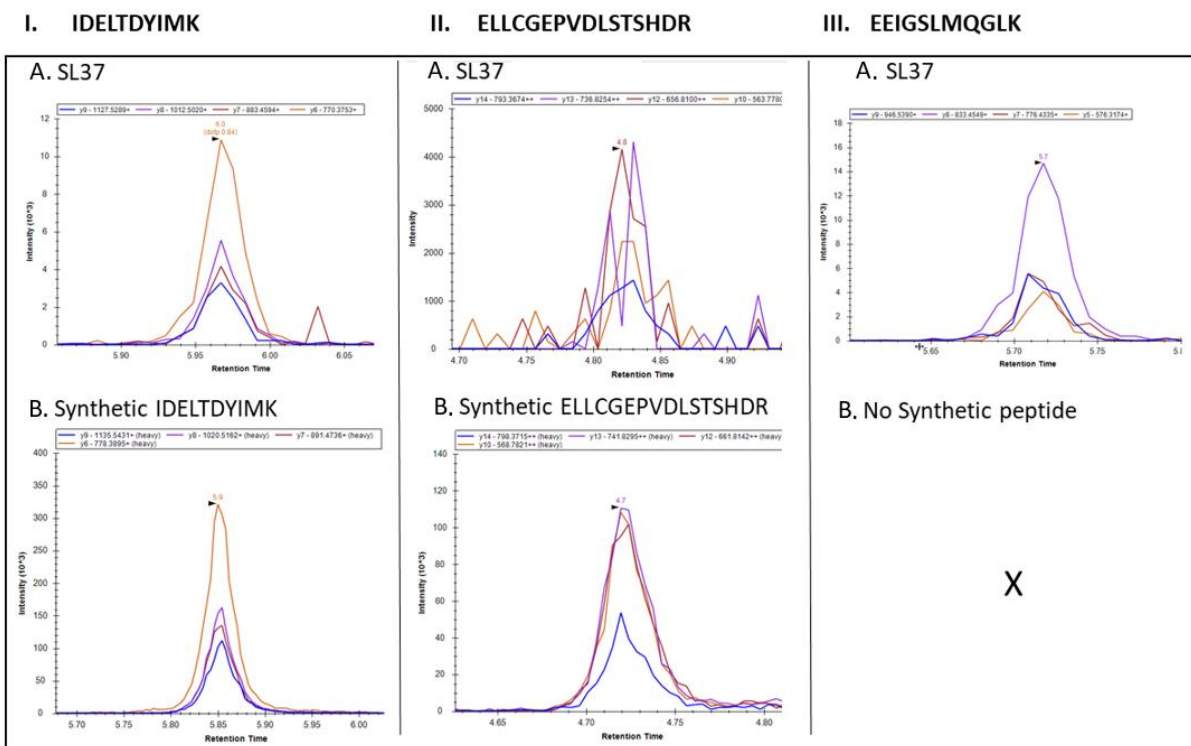
110 eluate of the StrepTactin purification from leaves infiltrated with SL37.

111 Seven MRM peptides were assessed for AnfG. Five out of seven peptides displayed peak profiles

112 with low but detectable response and the product ions (Q3 m/z, n = 4) profiled with the same

113 retention times.

114



116

117 **Supplementary Figure 8** | Extracted ion chromatograms (XIC) for three selected peptides;

118 (I) IDELTDYIMK, (II) ELLCGEPVDLSTSHDR and (III) EEIGSLMQGLK using MRM for AnfG.

119 The eluate after streptactin purification from leaves infiltrated with SL37 shows very low level of

120 AnfG. Two peptides, IDELTDYIMK and ELLCGEPVDLSTSHDR (B) were synthesised as heavy labelled

121 peptides (JPT Peptide Technologies GmbH, Berlin, Germany) to distinguish interference peaks

122 and confirm that both the peptide peak profiles display the expected product ion profile (Q3 m/z,

123 n = 4), and the transition ions show similar retention times (RT), 5.9 and 4.7, respectively. There

124 is evidence of matrix effects on both of these peptides with slight RT shifts between the dilute

125 eluate (IDELTDYIMK, RT=6.0 and ELLCGEPVDLSTSHDR, RT=4.8) compared to the synthetic

126 peptides spiked into total protein extract from *N. benthamiana* prior to the trypsin digest (B). The

127 evidence for the presence of AnfG is not conclusive with the confirmation of the unique

128 IDELTDYIMK AnfG peptide, but very low detection for ELLCGEPVDLSTSHDR (DNQ: detected but

129 not quantifiable).

131 **Supplementary Table 1|** Multiple reaction monitoring transitions of AnfG, H, D and K peptides
132 for targeted liquid chromatography – mass spectrometry (LC-MS).

Protein	Peptide ^a	RT (min) ^b	Q1 <i>m/z</i> ^b	<i>z</i> ^b	Q3 <i>m/z</i> ^b	Fragment	CE ^c
AnfG	VEAPVHPMDAR	3.39	407.87	3+	229.12 561.77 589.28 497.25	b2+ y10++ y5+ y9++	17.58 17.58 17.58 17.58
AnfG	VEAPVHPM[Oxi]DAR	2.66	413.20	3+	229.12 569.77 469.73 505.25	b2+ y10++ y5+ y9++	17.83 17.83 17.83 17.83
AnfG	IDELTDYIMK	6.02	620.81	2+	770.38 883.46 1012.50 1127.53	y6+ y7+ y8+ y9+	29.42 29.42 29.42 29.42
AnfG	IDELTDYIM[Oxi]K	5.16	628.81	2+	786.37 899.45 1028.50 1143.52	y6+ y7+ y8+ y9+	29.81 29.81 29.81 29.81
AnfG	NC[CAM]LWQFHSR	4.50	416.53	3+	399.21 674.34 430.71 487.25	y3+ y5+ y6++ y7++	17.99 17.99 17.99 17.99
AnfG	QNAEILK	2.74	408.23	2+	147.11 260.20 373.28 573.36	y1+ y2+ y3+ y5+	19.00 19.00 19.00 19.00
AnfG	Q[PGQ]NAEILK	4.00	399.72	2+	147.11 260.20 373.28 573.36	y1+ y2+ y3+ y5+	18.59 18.59 18.59 18.59
AnfG	ELLC[CAM]GEPVDLSTSHDR	4.87	609.96	3+	563.78 656.81 736.83 793.37	y10++ y12++ y13++ y14++	27.28 27.28 27.28 27.28
AnfG	EHYPWINSM[Oxi]SK	4.10	704.32	2+	267.11 430.17 978.47 1141.53	b2+ b3+ y8+ y9+	33.51 33.51 33.51 33.51
					576.32	y5+	28.54

AnfG	EEIGSLMQGLK	5.70	602.82	2+	776.43 833.45 946.54	y7+ y8+ y9+	28.54 28.54 28.54
AnfG	EEIGSLM[Oxi]QGLK	4.31	610.81	2+	792.43 849.45 962.53 481.77	y7+ y8+ y9+ y9++	28.93 28.93 28.93 28.93
AnfH	VAIYGK	2.66	325.70	2+	171.11 367.20 480.28 551.32	b2+ y3+ y4+ y5+	14.96 14.96 14.96 14.96
AnfH	STTTQNTAAALAYFHDK	5.01	613.97	3+	826.41 709.33 780.37 1035.53	y15++ y5+ y6+ y9+	27.47 27.47 27.47 27.47
AnfH	VFIHGC[CAM]DPK	2.78	358.18	3+	415.19 244.17 413.70 487.23	b7++ y2+ y7++ y8++	15.19 15.19 15.19 15.19
AnfH	ITNDDVIK	2.85	459.25	2+	215.14 589.32 703.36 804.41	b2+ y5+ y6+ y7+	21.50 21.50 21.50 21.50
AnfH	LGGIIC[CAM]NSR	3.59	495.26	2+	536.22 649.31 819.41 876.44	y4+ y5+ y7+ y8+	23.27 23.27 23.27 23.27
AnfH	EFLEEFTAAIGTK	7.06	728.37	2+	1066.54 1179.63 808.46 937.50	y10+ y11+ y8+ y9+	34.69 34.69 34.69 34.69
AnfH	DNIVQK	1.39	358.70	2+	230.08 343.16 374.24 487.32	b2+ b3+ y3+ y4+	16.58 16.58 16.58 16.58
AnfH	EYGELAR	2.59	419.21	2+	246.16 359.24 488.28 545.30	y2+ y3+ y4+ y5+	19.54 19.54 19.54 19.54
AnfH	E[PGE]YGELAR	2.59	410.20	2+	461.17 246.16 359.24	b4+ y2+ y3+	19.10 19.10 19.10

					545.30	y5+	19.10
AnfD	HVIGTPMK	2.60	441.75	2+	237.13	b2+	20.65
					375.21	y3+	20.65
					646.36	y6+	20.65
					745.43	y7+	20.65
AnfD	HVIGTPM[Oxi]K	1.70	449.74	2+	237.13	b2+	21.04
					508.29	b5+	21.04
					662.35	y6+	21.04
					761.42	y7+	21.04
AnfD	YTYATDVK	2.81	480.74	2+	265.12	b2+	22.56
					533.29	y5+	22.56
					696.36	y6+	22.56
					797.40	y7+	22.56
AnfD	HIVFGAEK	3.07	450.75	2+	251.15	b2+	21.09
					551.28	y5+	21.09
					650.35	y6+	21.09
					763.43	y7+	21.09
AnfD	QNIIEAFK	5.16	481.77	2+	365.22	y3+	22.61
					494.26	y4+	22.61
					607.34	y5+	22.61
					720.43	y6+	22.61
AnfD	INIAWINQK	5.75	550.32	2+	228.13	b2+	25.97
					688.38	y5+	25.97
					759.41	y6+	25.97
					872.50	y7+	25.97
AnfD	AHLNVLEC[CAM]AR	3.81	394.87	3+	436.23	b4+	16.95
					535.30	b5+	16.95
					406.19	y3+	16.95
					535.23	y4+	16.95
AnfD	LDIDGFGFKPLADSLR	7.48	588.65	3+	711.87	y13++	26.26
					768.41	y14++	26.26
					825.93	y15++	26.26
					771.44	y7+	26.26
AnfD	IGMFFGIEDR	6.87	592.79	2+	589.29	y5+	28.05
					736.36	y6+	28.05
					883.43	y7+	28.05
					1071.49	y9+	28.05
AnfD	IGM[Oxi]FFGIEDR	5.90	600.79	2+	589.29	y5+	28.44
					736.36	y6+	28.44
					883.43	y7+	28.44
					1087.49	y9+	28.44
					185.13	b2+	23.91

AnfD	AIIDEEVAR	3.89	508.27	2+	603.31 718.34 831.42	y5+ y6+ y7+	23.91 23.91 23.91
AnfD	WKPELDWYK	5.64	422.22	3+	496.26 611.28 950.46 539.78	y3+ y4+ y7+ y8++	18.27 18.27 18.27 18.27
AnfD	VC[CAM]LWPGGSK	4.86	502.26	2+	260.11 445.24 631.32 744.40	b2+ y5+ y6+ y7+	23.61 23.61 23.61 23.61
AnfD	VVSVYTK	2.74	398.23	2+	199.14 411.22 597.32 696.39	b2+ y3+ y5+ y6+	18.51 18.51 18.51 18.51
AnfD	GFEGWVR	4.75	425.71	2+	205.10 460.27 517.29 646.33	b2+ y3+ y4+ y5+	19.86 19.86 19.86 19.86
AnfD	EYPAFER	3.63	456.22	2+	451.23 522.27 619.32 310.16	y3+ y4+ y5+ y5++	21.35 21.35 21.35 21.35
AnfK	LIFSQHYK	3.79	345.86	3+	227.18 662.33 405.20 461.74	b2+ y5+ y6++ y7++	14.60 14.60 14.60 14.60
AnfK	VEEAVDVLLSR	5.94	615.34	2+	702.41 801.48 872.52 1001.56	y6+ y7+ y8+ y9+	29.15 29.15 29.15 29.15
AnfK	YPDVK	2.07	311.17	2+	376.15 246.18 361.21 458.26	b3+ y2+ y3+ y4+	14.25 14.25 14.25 14.25
AnfK	LNEGLLK	3.85	393.74	2+	228.13 430.30 559.34 673.39	b2+ y4+ y5+ y6+	18.29 18.29 18.29 18.29
AnfK	INLLTGWVNPGDVK	7.47	763.42	2+	228.13 341.22 1072.54	b2+ b3+ y10+	36.41 36.41 36.41

					971.49	y9+	36.41
AnfK	ATFALNR	3.67	396.72	2+	402.25	y3+	18.44
					473.28	y4+	18.44
					620.35	y5+	18.44
					310.68	y5++	18.44
AnfK	AAEYLQK	2.21	411.72	2+	143.08	b2+	19.17
					551.32	y4+	19.17
					680.36	y5+	19.17
					751.40	y6+	19.17
AnfK	FEIPAIIGPTPIGIR	8.55	797.47	2+	1204.74	y12+	38.08
					753.46	y7+	38.08
					810.48	y8+	38.08
					923.57	y9+	38.08
AnfK	NTDIFLQNLK	5.88	603.33	2+	615.38	y5+	28.56
					762.45	y6+	28.56
					875.53	y7+	28.56
					990.56	y8+	28.56
AnfK	YVDDPR	1.43	382.68	2+	263.14	b2+	17.75
					493.19	b4+	17.75
					272.17	y2+	17.75
					502.23	y4+	17.75
AnfK	NEGLELDLILGHSK	6.72	513.28	3+	647.87	y12++	22.64
					428.23	y4+	22.64
					541.31	y5+	22.64
					654.39	y6+	22.64
AnfK	FISIDYNIPMLR	7.81	741.39	2+	1221.63	y10+	35.33
					516.30	y4+	35.33
					906.49	y7+	35.33
					1021.51	y8+	35.33
AnfK	FISIDYNIPM[Oxi]LR	7.12	749.39	2+	1237.62	y10+	35.72
					532.29	y4+	35.72
					922.48	y7+	35.72
					1037.51	y8+	35.72
AnfK	VGFPYDR	4.18	477.74	2+	651.31	y5+	22.41
					326.16	y5++	22.41
					399.69	y6++	22.41
					855.40	y7+	22.41

- a. The peptide sequence is represented by single amino acid code. C[CAM] refers to carbamidomethylation of cysteine; M[Oxi] refers to oxidation of methionine; E[PGE] refers to pyroglutamic acid formed from glutamic acid.
- b. RT, retention time (min); Q1 m/z , precursor ion mass-to-charge ratio; z, charge state; Q3 m/z , fragment ion m/z ; CE, collision energy in V.

138 c. Collision energy settings were calculated for 2+ ions ($CE = \text{slope } (0.049) \times (\text{precursor } m/z) + \text{intercept}$
139 (-1.0)) and 3+ ions ($CE = \text{slope } (0.048) \times (\text{precursor } m/z) + \text{intercept } (-2.0)$) for a 6500 QTRAP mass
140 spectrometer (SCIEX, Redwood City, USA).
141

142

Supplementary Data Set 1 | Schematic outline of the peptides detected by LC-MS/MS for AnfDKGH isolated from a range of plant-based expression experiments

Using a variety of protein sources expressed from multigene and single gene constructs (Table 1, Table 2), the Anf proteins were characterized allowing selection of peptides for MRM development. Green, peptides identified with >95% confidence; yellow, peptides identified with 50-95% confidence; red, peptides identified with <50% confidence; grey, not detected. Peptides used for MRM are underlined. Solid and dashed underline are used to distinguish adjacent tryptic peptides (5-20 amino acids in length) in the MRM method.

AnfD

MPHHEFECSKVIPERKKHAVIKGKGETLADALPQGYLNTIPGSISERGCAYCGAKHVI~~GT~~PMKDVIHISHGPGV~~GCTY~~
DTWQTKRYISDNDNFQLKYTYATDVKEKHIVFGAEKLLKQNIIEAFKAF~~PQIKRMTIYQTCATALIGDDINAI~~AEEV
MEEMPEVDIFVCNSPGFAGPSQSGGHHKINIAWINQKVGTVPEPEITGDHVINYVGEYNIQGDQEV~~MVDYFKRMGIQV~~
LSTFTGNGSYDGLRAMHRAHLNVLECARSAEYICNELRVRYGIPRLDIDGFGFKPLADSLRKIGMFFGIEDRAKAI
DEEVARWKP~~ELDWYKERLMGKKVCLWPGGSKLWHWAHVIEEEMGLKVSVYTKFGHQDMEKGIARC~~GEGTLAIDDP
NELEGLEALEMLKPDIIILTGKRPGEVAKKVRVPYLNHAYHNGPYKGFEGWVRFARDIYNAIYSPIHQLSGIDITKD
NAPEWGN~~GFRTRQMLSDGNLSDAVRNSETLRQY~~TGGYDSVSKLRE~~REYPA~~FERKVG

AnfK

MTCEVKEKGRVGTINPIFTCQPAGAQFVSI~~G~~KDCIGIVHGGQGCVMFVRLIFSQHYKESFELASSSLHEDGAVFGA
CGRVEEA~~VDVLLSRYPDV~~KVVPPIITTCSTEIIGDDVDGVIKKLNEGLLKEKFPDREVHLIAMHTPSFVGSMISGYDV
AVRDVVRHFAKREAPNDKINLLTGWVNP~~GDVKELKHL~~LGEMDIEANVLFIEISFDSPILPDGSAVSHGNTTIEDLID
TGNARATFALNRYEGTKAAEYLQKKFEIPAIIGPTPIGIRNTDIFLONLKKATGKPIPQSLAHERGVAIDALADLTH
MFLAEKRVAIYGAPDLVIGLAEFCLDLEMKPVLLLLGDDNSKYVDDPRIKALQENV~~DYGMEIVTNADFWEL~~ENRIKN
EGLELDLILGH~~SKGRFISIDYNIPMLRVGFPTYDRAGLFR~~YPTVGYGGAIWLAEQMANTLFADMEHKKNKEWVLNVW

AnfH

MTRKVAIYGKGIGKSTTTQNTAAALAYFHD~~KKVF~~IHGCDPKADSTRLLLGK~~PKQET~~LM~~DMLR~~DKGAEKITND~~DVIK~~
KGFLDIQC~~VESGGPEPGVC~~AGRGVITAILMEENGAYTDDLDFVFFDVLGDVVC~~GGFAMP~~IRDGKAQEVYIVASGE
MMAIYAANNICKGLVKYAKQSGVR~~LG~~GIICNSRKVDGEREFLEEFTAAIGTKMIHFVPRDNIVQKAEFNKKT~~VT~~EFA
PEENQAKEYGELARKIIENDEFVIPKPLTMDQLED~~MDVVKY~~GIAD

AnfG

MSTASAAAVVKQKVEAPVHPMDARIDELTDYIMKNCLWQFHSRSWDRERQNAEILKKT~~ELL~~CGEPVDLSTSHDR~~CY~~
WVDAVCLADDYREHYPWINSMSKEEIGSLMOGLKDRMDYLTITGSLNEELSDKH~~Y~~

179 **Supplementary Data Set 2 | Schematic outline of the peptides detected by LC-MS/MS for MTP-**
180 **AnfDKH from enriched plant mitochondria.**

181 Duplicate gel slices of protein bands corresponding to AnfD, K, H and G were excised from
182 Coomassie stained gels and subjected to in-gel digestion (IGD) with trypsin. Peptide
183 identifications and the sequence coverage is shown for AnfH (IGD_9-1 and -2), AnfK (IGD_10-1
184 and -2) and AnfD (IGD_11-1 and -2). The duplicate gel digests (-1 or -2) show similar peptide
185 identifications. No peptides were identified for AnfG in either digest. Green, peptides identified
186 with >95% confidence; yellow, peptides identified with 50-95% confidence; red, peptides
187 identified with <50% confidence; grey, not detected.

188
189 IGD_9-1. pFAy51-HA-AnfH (3 peptides were identified at 95% confidence, 10% sequence coverage).
190 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRSLISTQVVRNRGGYPYDVPDYAGGMTRKVAIYKGKGI
191 GKSTTTQNTAAALAYFHDKKVFIHGCDPKADSTRILIGGKPQETLMDMLRDKGAEKITNDDVIKGFLDIQCVESSG
192 PEPGVCAGRGVITAI DLMEENGAYTDDLDFVFFDVLGDVVC GGFAMP I RDGKAQEVYIVASGEMMAIYAANNICKG
193 LVKYAKQSGVRLGGI ICNSRKVDGEREFLEEF TAAIGTKMIHFVPRDNIVQKAEFNKKTVTEFAPEENQAKKEYGELA
194 RKIIENDEFVIPKPLTMDQLED MVVKYGIAD

195 IGD_9-2. pFAy51-HA-AnfH (3 peptides were identified at 95% confidence, 9.7% sequence coverage).
196 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRSLISTQVVRNRGGYPYDVPDYAGGMTRKVAIYKGKGI
197 GKSTTTQNTAAALAYFHDKKVFIHGCDPKADSTRILIGGKPQETLMDMLRDKGAEKITNDDVIKGFLDIQCVESSG
198 PEPGVCAGRGVITAI DLMEENGAYTDDLDFVFFDVLGDVVC GGFAMP I RDGKAQEVYIVASGEMMAIYAANNICKG
199 LVKYAKQSGVRLGGI ICNSRKVDGEREFLEEF TAAIGTKMIHFVPRDNIVQKAEFNKKTVTEFAPEENQAKKEYGELA
200 RKIIENDEFVIPKPLTMDQLED MVVKYGIAD

201 IGD_10-1. pFAy51-HA-AnfK (5 peptides were identified at 95% confidence, 10.8% sequence coverage).
202 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRSLISTQVVRNRGGYPYDVPDYAGGMTCEVKEKGRVGT
203 INPIFTCQPAQAQFVSIKDCIGIVHGGQGCVMFVRLIFSQHYKESFELASSSLHEDGAVFGACGRVEEAVDVLLS
204 RYPDVKVVPPIITTCSTEIIGDDVDGVIKKLEGLLKEKFPDREVHLIAMHTPSFVGSMISGYDVAVRDVVRHFAKRE
205 APNDKINLLTGWVNP GDVKE LKHL LGEMDIEANVLF EIESFDSPILPDGSAVSHGNTTIEDLIDTGNARATFALNRY
206 EGTKAAEYLQKKFEIPAIIGPTPIGIRNTDIFLQNLK KATGKPI PQSLAHERGVAIDALADLTHMFLAEKRVAIYGA
207 PDLVIGLAEFCLDLEMKPVLLLLGDDNSKYVDDPRIKALQENV DYGMEIVTNADFWEL ENRIKNEGLELDLILGHSK
208 GRFISIDYNIPMLRVGFPTYDRAGLFRYPTVGYGGAIWLAEQMAN TLFADMEHKKNKKEWVLNVW

209 IGD_10-2. pFAy51-HA-AnfK (4 peptides were identified at 95% confidence, 10.1% sequence coverage).
210 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRSLISTQVVRNRGGYPYDVPDYAGGMTCEVKEKGRVGT
211 INPIFTCQPAQAQFVSIKDCIGIVHGGQGCVMFVRLIFSQHYKESFELASSSLHEDGAVFGACGRVEEAVDVLLS
212 RYPDVKVVPPIITTCSTEIIGDDVDGVIKKLEGLLKEKFPDREVHLIAMHTPSFVGSMISGYDVAVRDVVRHFAKRE
213 APNDKINLLTGWVNP GDVKE LKHL LGEMDIEANVLF EIESFDSPILPDGSAVSHGNTTIEDLIDTGNARATFALNRY
214 EGTKAAEYLQKKFEIPAIIGPTPIGIRNTDIFLQNLK KATGKPI PQSLAHERGVAIDALADLTHMFLAEKRVAIYGA
215 PDLVIGLAEFCLDLEMKPVLLLLGDDNSKYVDDPRIKALQENV DYGMEIVTNADFWEL ENRIKNEGLELDLILGHSK
216 GRFISIDYNIPMLRVGFPTYDRAGLFRYPTVGYGGAIWLAEQMAN TLFADMEHKKNKKEWVLNVW

217 IGD_11-1. pFAy51-HA-AnfD (9 peptides were identified at 95% confidence, 17.0% sequence coverage).
218 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVRSLISTQVVRNRGGYPYDVPDYAGGMPHHEFECSKVI P
219 ERKKHAVIKGKETLADALPQGYLNTIPGSI SERGCAYCGAKHVIGTPMKDVIHISHGPVGCTYDTWTQTKRYISDND
220 NFQLKYYATDVKEKHIVFGAEKLLKQNIIEAFKAF PQIKRMTIYQTCATALIGDDINAIAEEVMEEMPEVDIFVCN

221 SPGFAGPSQSGGHHK**INIAWINQK**VGTVPEITGDHVINYVGEYNIQGDQEVMDYFKRMGIQVLSTFTGNGSYDGL
 222 RAMHRAHLNVLECAR**SAEYICNELRVRYGIPRL**DIDGFGFKPLADSLRKIGMFFGIEDRAKAIIDEEVAR**WKPELDW**
 223 **YKERLMGKKVCLWP**GGSKLWHWAH**VIEEEMGLKVSVYTK**FGHQDMEKGIARCGEGTLAIDDPNELEGLEALEMLK
 224 PDIILTGKRPGEVAKKVRVPYLNAYHNGPYK**GFEGWVRFARDIYNAIYSPIHQ**LSGIDIT**KDNAPEWNGFR**TRQ
 225 **MLSDGNLSDAVR**NSETLR**QY**TGGYDSVSKLRER**EYPAFER**KVG

226 IGD_11-2. pFAy51-HA-AnfD (10 peptides were identified at 95% confidence, 18.9% sequence coverage).

227 MAMAVFRREGRRLLPSIAARPIAAIRSPSSDQEEGLLGVR**ISTQVVR**NRGGYPYDVPDYAGGMPHHEFECSKVIP
 228 ERKKHAVIK**GKGETLADALPQGYLNTIPGSISER**GCAYCGAK**HVIGTPMK**DVIHISHGPVGCTYDTWQTKRY**ISDND**
 229 **NFQLKYTYATDVKEKHIVFGAEKLLKQNIIEAFK**AFPQIKRMTIYQTCATALIGDDINAIAEEVMEEMPEVDIFVCN
 230 SPGFAGPSQSGGHHKINIAWINQKVGTVPEITGDHVINYVGEYNIQGDQEVMDYFKRMGIQVLSTFTGNGSYDGL
 231 RAMHRAHLNVLECAR**SAEYICNELRVRYGIPRL**DIDGFGFKPLADSLRKIGMFFGIEDRAK**AIIDEEVAR**WKPELDW
 232 YKERLMGKKVCLWPGGSKLWHWAH**VIEEEMGLKVSVYTK**FGHQDMEKGIARCGEGTLAIDDPNELEGLEALEMLK
 233 PDIILTGKRPGEVAKKVRVPYLNAYHNGPYK**GFEGWVRFARDIYNAIYSPIHQ**LSGIDIT**KDNAPEWNGFR**TRQ
 234 **MLSDGNLSDAVR**NSETLR**QY**TGGYDSVSKLRER**EYPAFER**KVG

Supplementary Data Set 3 | Codon-optimized DNA sequences and translated protein sequences
of *anfD*, *anfK*, *anfG* and *anfH* genes used in this study.

Amino acids underlined represent the MTP sequence, amino acids in **bold** represent epitope tags, HA and Twin-Strep.

SN81, pFAy51-AnfD-HA:

```
atgatggcaatggctgttttccgctcgcggaagggaggcgtctcctcccttcaatcgccgct
M M A M A V F R R E G R R L L P S I A A
cgcccaatcgctgctatccgatctcctctctcttctgaccaggaggaaggacttcttgga
R P I A A I R S P L S S D Q E E G L L G
gttcgatctatctcaactcaagtgggtgcgtaaccgcggaggtatgcctcatcatgagttc
V R S I S T Q V V R N R G G M P H H E F
gagtgttccaaagtaatcccggagagaaagaagcatgccgttataaaaggaaagggcgag
E C S K V I P E R K K H A V I K G K G E
acactagcggacgcgcttcctcaaggatatctaaacacaatacctgggtcaatttccgaa
T L A D A L P Q G Y L N T I P G S I S E
agaggttggtgcttatttgccgtgcaaagcacgttatttggaaccccaatgaaagatgttata
R G C A Y C G A K H V I G T P M K D V I
cacatttcccacggaccagttggctgtacttacgatacatggcagacgaaacggttacata
H I S H G P V G C T Y D T W Q T K R Y I
tccgacaatgacaatttccagctaaagtatacttatgcgacggacgtaaaagaaaaaacac
S D N D N F Q L K Y T Y A T D V K E K H
atagtctttggagctgaaaaattactgaaacaaaacatcattgaggcgtttaaggcattc
I V F G A E K L L K Q N I I E A F K A F
ccgcaaatacaagcgaatgactatctaccagacttggtgcgactgctttgatcggagatgat
P Q I K R M T I Y Q T C A T A L I G D D
attaatgccatcgcggaagaggtaatggaggagatgcctgaagttgacatattttgtttgc
I N A I A E E V M E E M P E V D I F V C
aactctcccggattcgctggaccatctcagagtggcgggcatcacaagatcaatattgcc
N S P G F A G P S Q S G G H H K I N I A
tggattaatcagaaggtcggcacagtagagcctgaaattaccggtgaccatgtcataaac
W I N Q K V G T V E P E I T G D H V I N
tatgtgggggaatataacatacaaggcgaccaggaggttatggttgactactttaagaga
Y V G E Y N I Q G D Q E V M V D Y F K R
atgggggatccaggtcttgtccactttcacggcaacggaagttatgatgggttgagagct
M G I Q V L S T F T G N G S Y D G L R A
atgcaccgtgcgcaacttgaatgtattagagtgtgctcgtagcgcagagtacatttgcaat
M H R A H L N V L E C A R S A E Y I C N
gaactaagagtttagatacggcatccctcgacttgatatagacggttttgattcaaacca
E L R V R Y G I P R L D I D G F G F K P
cttgcggactcactccgtaagattggaatgttctttgggatagaggaccgtgcaaaagcc
L A D S L R K I G M F F G I E D R A K A
atcatcgacgaggaagtagctcggttggaacccggaattggattggtataaggagcgactc
```

278 I I D E E V A R W K P E L D W Y K E R L
 279 atgggcaagaaggtgtgcctatggcctgggtggctccaaattgtggcattgggcgacgctc
 280 M G K K V C L W P G G S K L W H W A H V
 281 atcgaggaagaaatgggattgaaggtcgtgtccgtctacaccaagttcggacatcaaggc
 282 I E E E M G L K V V S V Y T K F G H Q G
 283 gacatggagaaggggaatagccccgatgtggcgaaggtactcttgcgatcgatgatccaaat
 284 D M E K G I A R C G E G T L A I D D P N
 285 gaactcgaaggttttagaagctcttgagatgcttaaaccagatatataattttgacgggaaag
 286 E L E G L E A L E M L K P D I I L T G K
 287 cgtccaggtgaggtggcaaaaaaagtcggttccttacctcaatgccacgcttaccac
 288 R P G E V A K K V R V P Y L N A H A Y H
 289 aacgggccttacaaagggttttgaaggttgggtgcgttttgctagggatatttataatgct
 290 N G P Y K G F E G W V R F A R D I Y N A
 291 atctacagtccaatccatcagctgtccggaatcgatataactaaggataatgccccctgaa
 292 I Y S P I H Q L S G I D I T K D N A P E
 293 tgggggaatggcttttcgtacgcgacagatgctcagtgacgggaatttgagcgacgctgtc
 294 W G N G F R T R Q M L S D G N L S D A V
 295 aggaattcagaaacactgcgtcagtacacaggcggttatgactctgtatcaaaactgcga
 296 R N S E T L R Q Y T G G Y D S V S K L R
 297 gaaagagagtatcccgcggttgaacgtaagggttggtgggtggtacccttacgacgttcct
 298 E R E Y P A F E R K V G G G **Y P Y D V P**
 299 gattacgct
 300 **D Y A**

301

302 **SN82, HA-AnfD:**

303 atgtacccttacgacgttcctgattacgctggaggtatgcctcatcatgagttcgagtgt
 304 M **Y P Y D V P D Y** A G G M P H H E F E C
 305 tccaaagtaatcccgagagaaagaagcatgccgttataaaaggaaagggcgagacacta
 306 S K V I P E R K K H A V I K G K G E T L
 307 gcgacgcgcttcctcaaggatatctaaacacaatacctgggtcaatttcgaaagaggt
 308 A D A L P Q G Y L N T I P G S I S E R G
 309 tgtgcttattgcggtgcaaagcacgttattggaacccaatgaaagatgttatacacatt
 310 C A Y C G A K H V I G T P M K D V I H I
 311 tcccacggaccagttggctgtacttacgatacatggcagacgaaacgttacatatccgac
 312 S H G P V G C T Y D T W Q T K R Y I S D
 313 aatgacaatttccagctaaagtatacttatgacgacggacgtaaaagaaaaacacatagtc
 314 N D N F Q L K Y T Y A T D V K E K H I V
 315 tttggagctgaaaaattactgaaacaaaacatcattgaggcggtttaaggcattccccgcaa
 316 F G A E K L L K Q N I I E A F K A F P Q
 317 atcaagcgaatgactatctaccagacttgtgacgactgctttgatcgagatgatattaat
 318 I K R M T I Y Q T C A T A L I G D D I N
 319 gccatcgcggaagaggtaatggaggagatgcctgaagttgacatatattgtttgcaactct
 320 A I A E E V M E E M P E V D I F V C N S
 321 cccggattcgctggaccatctcagagtggcgggcatcacaagatcaatattgcctggatt
 322 P G F A G P S Q S G G H K I N I A W I
 323 aatcagaaggtcggcacagtagagcctgaaattaccggtgaccatgtcataaactatgtg

324 N Q K V G T V E P E I T G D H V I N Y V
 325 ggggaatataacatacaaggcgaccaggaggttatggttgactactttaagagaatgggg
 326 G E Y N I Q G D Q E V M V D Y F K R M G
 327 atccaggtcttgtccactttcaccggcaacggaagttatgatgggttgagagctatgcac
 328 I Q V L S T F T G N G S Y D G L R A M H
 329 cgtgcgcacttgaatgtattagagtgtgctcgtagcgcagagtacatttgcaatgaacta
 330 R A H L N V L E C A R S A E Y I C N E L
 331 agagtttagatacggcatccctcgacttgatatagacgggttttggttcaaaccacttgcg
 332 R V R Y G I P R L D I D G F G F K P L A
 333 gactcactccgtaagattggaatgttctttgggatagaggaccgtgcaaaagccatcatc
 334 D S L R K I G M F F G I E D R A K A I I
 335 gacgaggaagtagctcgttgaaaccggaattggattggtataaggagcgactcatgggc
 336 D E E V A R W K P E L D W Y K E R L M G
 337 aagaaggtgtgcctatggcctgggtggctccaaattgtggcattgggcgcacgtcatcgag
 338 K K V C L W P G G S K L W H W A H V I E
 339 gaagaaatgggattgaaggctcgtgtccgtctacaccaagttcggacatcaaggcgacatg
 340 E E M G L K V V S V Y T K F G H Q G D M
 341 gagaaggggaatagcccgatgtggcgaaggtactcttgcgatcgatgatccaaatgaactc
 342 E K G I A R C G E G T L A I D D P N E L
 343 gaaggtttagaagctcttgagatgcttaaaccagatatataattttgacgggaaagcgtcca
 344 E G L E A L E M L K P D I I L T G K R P
 345 ggtgaggtggcaaaaaaagtcctgttcccttacctcaatgccacgcttaccacaacggg
 346 G E V A K K V R V P Y L N A H A Y H N G
 347 ccttaciaaaggttttgaagggttggtgctgttttgctagggatatttataatgctatctac
 348 P Y K G F E G W V R F A R D I Y N A I Y
 349 agtccaatccatcagctgtccggaatcgatataactaaggataatgccctgaatggggg
 350 S P I H Q L S G I D I T K D N A P E W G
 351 aatggcttttcgtacgcgacagatgctcagtgacgggaatttgagcgacgctgtcaggaat
 352 N G F R T R Q M L S D G N L S D A V R N
 353 tcagaaacactgcgtcagtacacaggcggttatgactctgtatcaaaactgcgagaaaga
 354 S E T L R Q Y T G G Y D S V S K L R E R
 355 gagtatcccgcggtttgaacgtaaggttggg
 356 E Y P A F E R K V G

357

358 **SN129, pFAy51-HA-AnfK:**

359 atggcaatggctgttttccgtcgcgaaggaggcgctctcctcccttcaatcgccgctcgc
 360 M A M A V F R R E G R R L L P S I A A R
 361 ccaatcgctgctatccgatctcctctctcttctgaccaggaggaaggacttcttggagtt
 362 P I A A I R S P L S S D Q E E G L L G V
 363 cgatctatctcaactcaagtgggtgcgtaaccgcggaggatacccttacgacgttcctgat
 364 R S I S T Q V V R N R G G **Y P Y D V P D**
 365 tacgctggaggtatgacgtgtgaggtgaaggagaagggccgagttgggacgataaatcct
 366 **Y A** G G M T C E V K E K G R V G T I N P
 367 atttttacctgccagcctgcgggcgctcagtttgtatctatcgggatcaaagactgtatc
 368 I F T C Q P A G A Q F V S I G I K D C I
 369 ggtatcgtgcatggaggacaagggtgcgtgatgttcgtgcgactcatcttctcacagcat

370 G I V H G G Q G C V M F V R L I F S Q H
 371 tacaaggagagctttgagctggcatcatcctctctccacgaggacggggccgtgttcggg
 372 Y K E S F E L A S S S L H E D G A V F G
 373 gcctgtggccgtgttgaagaggcagttgatgtcttgctttcccgttatcccgacgtaaag
 374 A C G R V E E A V D V L L S R Y P D V K
 375 gttgtaccaattattactacatgctccacggaaatcataggcgacgatgtggatgggggtt
 376 V V P I I T T C S T E I I G D D V D G V
 377 atcaagaagttgaatgagggctctgctgaaagagaagttccctgaccgagaagtccacctt
 378 I K K L N E G L L K E K F P D R E V H L
 379 atagcgatgcatactcctagttttgtaggatccatgataagtggctacgatgttgctgtg
 380 I A M H T P S F V G S M I S G Y D V A V
 381 cgtgatgtagtacgacacttcgccaaagagagaagccccgaacgataaaataaatcttctt
 382 R D V V R H F A K R E A P N D K I N L L
 383 acaggatgggtcaatcctgggtgatgtgaaagaacttaaacattttactcggcgagatggat
 384 T G W V N P G D V K E L K H L L G E M D
 385 attgaggcgaacgtactttttgaaattgaatcttttgactctcctatcttacctgacggc
 386 I E A N V L F E I E S F D S P I L P D G
 387 tccgcggttagtcacggaaatacaactatagaggacctaattgacacaggtaacgcccga
 388 S A V S H G N T T I E D L I D T G N A R
 389 gcgacgttcgctctaaacagatatgaaggaacgaaagcggcagagtatctacaaaagaaa
 390 A T F A L N R Y E G T K A A E Y L Q K K
 391 tttgagattcctgccattataggcccaacgcccatagggattcgaacacggacatatctc
 392 F E I P A I I G P T P I G I R N T D I F
 393 cttcagaatctgaagaaagctacagggaacccaattccgcagagcttggcccacgagcgt
 394 L Q N L K K A T G K P I P Q S L A H E R
 395 ggagtggctattgatgcttttggcggacctgacacacatgtttctagcagaaaagagagtc
 396 G V A I D A L A D L T H M F L A E K R V
 397 gctatctatggggcaccggatttagtgataggccttgacagagttctgcctagatctagag
 398 A I Y G A P D L V I G L A E F C L D L E
 399 atgaagccggctccttttattacttggtgacgacaattccaaatacgtcgacgatcccagg
 400 M K P V L L L L G D D N S K Y V D D P R
 401 attaaagccctacaagaaaacgtcgactacgggatggagattgttacgaacgcggacttc
 402 I K A L Q E N V D Y G M E I V T N A D F
 403 tgggagctcgagaataggatcaagaatgagggcctcgaactcgacctgatactggggcac
 404 W E L E N R I K N E G L E L D L I L G H
 405 tctaagggttaggtttatcagcatcgactacaatattccgatgttgagagtgggcttcccg
 406 S K G R F I S I D Y N I P M L R V G F P
 407 acgtatgatcgagctggacttttttaggtatcccaccgttggttacggaggggctatatgg
 408 T Y D R A G L F R Y P T V G Y G G A I W
 409 ttagcggagcaaatggcgaacactttgttcgcagacatggaacacaagaaaaacaaggag
 410 L A E Q M A N T L F A D M E H K K N K E
 411 tgggtattgaatgtatgg
 412 W V L N V W

413

414 **SN130, pFAy51-HA-AnfH:**

415 atggcaatggctgttttccgtcgcgaagggaggcgctctcctcccttcaatcgccgctcgc
416 M A M A V F R R E G R R L L P S I A A R
417 ccaatcgctgctatccgatctcctctctcttctgaccaggaggaaggacttcttgagtt
418 P I A A I R S P L S S D Q E E G L L G V
419 cgatctatctcaactcaagtgggtgcgtaaccgcggaggatacccttacgacgttcctgat
420 R S I S T Q V V R N R G G Y P Y D V P D
421 tacgctggaggtatgacgcgaaaagtggcgatttatggcaaaggcggtataggtaaataca
422 **Y A** G G M T R K V A I Y G K G G I G K S
423 acaacgacccaaaataaccgctgctgctttggcctacttccatgataaaaagggttttcac
424 T T T Q N T A A A L A Y F H D K K V F I
425 catggctgcgaccctaaagcggacagtacacgattaataactaggcggaaagccccaggaa
426 H G C D P K A D S T R L I L G G K P Q E
427 accctaattggatatgctacgtgataagggagctgagaaaattactaatgatgacgtaatc
428 T L M D M L R D K G A E K I T N D D V I
429 aaaaaggggttcctcgacatacagtgcgttgagagtgggtggacccgaaccaggcgctcgg
430 K K G F L D I Q C V E S G G P E P G V G
431 tgtgcaggtaggggtgttataacggccatcgacctcatggaagagaatggagcgtaaca
432 C A G R G V I T A I D L M E E N G A Y T
433 gacgacctagattttgtctttttcgatgtccttggtgatgtggtgtgcggcggttttgct
434 D D L D F V F F D V L G D V V C G G F A
435 atgcccattagagatggaaaggcacaggaggtgtatattgttgccctccggggaaatgatg
436 M P I R D G K A Q E V Y I V A S G E M M
437 gcgatttacgcggcaacaatatctgcaagggctctagtgaatatgcaagcagtctgga
438 A I Y A A N N I C K G L V K Y A K Q S G
439 gttagactggggggtataatttgaacagtcgaaaagtggacggtgagaggggaattcctc
440 V R L G G I I C N S R K V D G E R E F L
441 gaggagttttaccgctgcaatagggacgaaaatgatacatttctgtgccagagacaatatc
442 E E F T A A I G T K M I H F V P R D N I
443 gtgcaaaaagctgaattcaataaaaagaccgtgacggagttcgcctccagaggaaaaccaa
444 V Q K A E F N K K T V T E F A P E E N Q
445 gctaaagaatatggagaactcgcccgtaaaatcatcgagaatgatgagttcgttatccca
446 A K E Y G E L A R K I I E N D E F V I P
447 aagcctctcactatggaccagctagaggacatggtagtgaagtatgggatagccgat
448 K P L T M D Q L E D M V V K Y G I A D

449

450 **SN131, pFAy51-HA-AnfG:**

451 atggcaatggctgttttccgtcgcgaagggaggcgctctcctcccttcaatcgccgctcgc
452 M A M A V F R R E G R R L L P S I A A R
453 ccaatcgctgctatccgatctcctctctcttctgaccaggaggaaggacttcttgagtt
454 P I A A I R S P L S S D Q E E G L L G V
455 cgatctatctcaactcaagtgggtgcgtaaccgcggaggatacccttacgacgttcctgat
456 R S I S T Q V V R N R G G Y P Y D V P D
457 tacgctggaggtatgtccacggcctccgccgctgccgtagtcaaacagaaagtagaggcg
458 **Y A** G G M S T A S A A A V V K Q K V E A

459 cccgtacacccgatggatgcgagaatcgacgaacttacagattacataatgaagaactgc
 460 P V H P M D A R I D E L T D Y I M K N C
 461 ctctggcagttccacagcaggagttgggacagggagaggcagaatgccgaaattctcaag
 462 L W Q F H S R S W D R E R Q N A E I L K
 463 aaaacgaaagaattactttgtggcgagcctgtagatctatcaacatcacacgacccgatgc
 464 K T K E L L C G E P V D L S T S H D R C
 465 tattgggttgacgctgtctgcctagcggatgactaccgtgagcactaccctgggataaac
 466 Y W V D A V C L A D D Y R E H Y P W I N
 467 agtatgtccaaggaggagataggaagtctaaggaaggattaaaggaccgtatggactac
 468 S M S K E E I G S L M Q G L K D R M D Y
 469 ttgaccataacggggagcctaaatgaggagttatctgacaagcattac
 470 L T I T G S L N E E L S D K H Y

471

472 **SN152, HA-AnfK:**

473 atgtacccttacgacgttcctgattacgctggaggtatgacgtgtgaggtgaaggagaag
 474 M **Y P Y D V P D Y A** G G M T C E V K E K
 475 ggccgagttgggacgataaaatcctatttttacctgccagcctgcgggcgctcagtttgta
 476 G R V G T I N P I F T C Q P A G A Q F V
 477 tctatcgggatcaaagactgtatcggtatcgtgcatggaggacaaggggtgctgatgttc
 478 S I G I K D C I G I V H G G Q G C V M F
 479 gtgcgactcatcttctcacagcattacaaggagagctttgagctggcatcatcctctctc
 480 V R L I F S Q H Y K E S F E L A S S S L
 481 cacgaggacggggccgtgttcggggcctgtggccgtgttgaagaggcagttgatgtcttg
 482 H E D G A V F G A C G R V E E A V D V L
 483 ctttcccgttatcccgacgtaaagggttgtagcaattattactacatgctccacggaaatc
 484 L S R Y P D V K V V P I I T T C S T E I
 485 ataggcgacgatgtggatggggttatcaagaagttgaatgagggctctgctgaaagagaag
 486 I G D D V D G V I K K L N E G L L K E K
 487 ttccctgaccgagaagtccaccttatagcgatgcatactcctagttttgtaggatccatg
 488 F P D R E V H L I A M H T P S F V G S M
 489 ataagtggctacgatgttgctgtgctgatgtagtacgacacttcgccaagagagaagcc
 490 I S G Y D V A V R D V V R H F A K R E A
 491 ccgaacgataaaataaatcttcttacaggatgggtcaatcctggtgatgtgaaagaactt
 492 P N D K I N L L T G W V N P G D V K E L
 493 aaacattttactcggcgagatggatattgaggcgaacgtactttttgaaattgaatctttt
 494 K H L L G E M D I E A N V L F E I E S F
 495 gactctcctatcttacctgacggctccgcggttagtcacggaaatacaactatagaggac
 496 D S P I L P D G S A V S H G N T T I E D
 497 ctaattgacacaggtaacgcccgcgacgttcgctctaaacagatatgaaggaacgaaa
 498 L I D T G N A R A T F A L N R Y E G T K
 499 gcggcagagtatctacaaaagaaatttgagattcctgccattataggcccaacgcccata
 500 A A E Y L Q K K F E I P A I I G P T P I
 501 gggattcgaaacacggacatattccttcagaatctgaagaaagctacagggaaaccaatt
 502 G I R N T D I F L Q N L K K A T G K P I
 503 ccgcagagcttggcccacgagcgtggagtggctattgatgctttggcggacctgacacac
 504 P Q S L A H E R G V A I D A L A D L T H

505 atgttttctagcagaaaaagagagtcgctatctatggggcaccgcgatttagtgataggcctt
 506 M F L A E K R V A I Y G A P D L V I G L
 507 gcagagtttctgcctagatctagagatgaagccggctcttttattacttggtgacgacaat
 508 A E F C L D L E M K P V L L L L G D D N
 509 tccaaatacgtcgacgatcccaggattaaagccctacaagaaaacgtcgactacgggatg
 510 S K Y V D D P R I K A L Q E N V D Y G M
 511 gagattgttacgaacgcggacttctgggagctcgagaataggatcaagaatgagggcctc
 512 E I V T N A D F W E L E N R I K N E G L
 513 gaactcgacctgatactggggcactctaagggtaggtttatcagcatcgactacaatatt
 514 E L D L I L G H S K G R F I S I D Y N I
 515 ccgatgttgagagtgggcttcccgacgtatgatcgagctggacttttttaggtatcccacc
 516 P M L R V G F P T Y D R A G L F R Y P T
 517 gttggctacggaggggctatatgtgtagcggagcaaatggcgaacactttgttcgcagac
 518 V G Y G G A I W L A E Q M A N T L F A D
 519 atggaacacacagaaaaacaaggagtgggtattgaatgtatgg
 520 M E H K K N K E W V L N V W

521

522 **SN153, HA-AnfH:**

523 atgtacccttacgacgttcctgattacgctggaggtatgacgcgaaaagtggcgatttat
 524 M **Y P Y D V P D Y A** G G M T R K V A I Y
 525 ggcaaaggcgggtataggtaaatacaaacgacccaaaataccgctgctgctttggcctac
 526 G K G G I G K S T T T Q N T A A A L A Y
 527 ttccatgataaaaaaggttttcatccatggctgacacctaagcggacagtacacgatta
 528 F H D K K V F I H G C D P K A D S T R L
 529 atactaggcggaaagccccaggaaaccctaattggatatgctacgtgataagggagctgag
 530 I L G G K P Q E T L M D M L R D K G A E
 531 aaaattactaatgatgacgtaatacaaaaaggggttcctcgacatacagtgcgttgagagt
 532 K I T N D D V I K K G F L D I Q C V E S
 533 ggtggacccgaaccaggcgtcggttgtgcaggtaggggtgttataacggccatcgacctc
 534 G G P E P G V G C A G R G V I T A I D L
 535 atggaagagaatggagcgtacacagacgacctagattttgtctttttcgatgtccttggt
 536 M E E N G A Y T D D L D F V F F D V L G
 537 gatgtggtgtgcggcggttttgcctatgccattagagatggaaaggcacaggaggtgtat
 538 D V V C G G F A M P I R D G K A Q E V Y
 539 attgttgcctccgggggaaatgatggcgattttacgcggcaaacaatatctgcaagggtcta
 540 I V A S G E M M A I Y A A N N I C K G L
 541 gtgaaatatgcaagcagtcctggagtttagactgggggtataatttgtaacagtcgaaaa
 542 V K Y A K Q S G V R L G G I I C N S R K
 543 gtggacggtgagaggggaattcctcgaggagttttaccgctgcaatagggacgaaaaatgata
 544 V D G E R E F L E E F T A A I G T K M I
 545 catttcgtgcccagagacaatatcgtgcaaaaagctgaattcaataaaaagaccgtgacg
 546 H F V P R D N I V Q K A E F N K K T V T
 547 gagttcgtctccagagggaaaaccaagctaagaatatggagaactcgcccgtaaaatcatc
 548 E F A P E E N Q A K E Y G E L A R K I I
 549 gagaatgatgagttcgttatcccaaagcctctcactatggaccagctagaggacatggta
 550 E N D E F V I P K P L T M D Q L E D M V

551 gtgaagtatgggatagccgat
552 V K Y G I A D

553

554 **SN154, HA-AnfG:**

555 atgtacccttaacgacgttcctgattacgctggaggtatgtccacggcctccgccgctgcc
556 M Y P Y D V P D Y A G G M S T A S A A A
557 gtagtcaaacagaaagtagagggcgcccgtacacccgatggatgcgagaatcgacgaactt
558 V V K Q K V E A P V H P M D A R I D E L
559 acagattacataatgaagaactgcctctggcagttccacagcaggagttgggacagggag
560 T D Y I M K N C L W Q F H S R S W D R E
561 aggcagaatgccgaaattctcaagaaaacgaaagaattactttgtggcgagcctgtagat
562 R Q N A E I L K K T K E L L C G E P V D
563 ctatcaacatcacacgaccgatgctattgggttgacgctgtctgcctagcggatgactac
564 L S T S H D R C Y W V D A V C L A D D Y
565 cgtgagcactacccgtggataaacagtatgtccaaggaggagataggaagtctaattgcaa
566 R E H Y P W I N S M S K E E I G S L M Q
567 ggattaaaggaccgtatggactacttgaccataacggggagcctaataatgaggagttatct
568 G L K D R M D Y L T I T G S L N E E L S
569 gacaagcattac
570 D K H Y

571

572 **SN155, alaFAy51-HA-AnfK:**

573 atggcaatggctgtttttccgtcgcgaagcgggcggtctctcctcccttcaatcgccgctcgc
574 M A M A V F R R E A A A L L P S I A A R
575 ccaatcgctgctgctgcagctgccgcttcttctgaccaggaggaaggacttcttctgctgcc
576 P I A A A A A A S S D Q E E G L L A A
577 gcagcggctgcggccgctgtggtgcgtaaccgcggaggttacccttacgacgttcctgat
578 A A A A A A V V R N R G G Y P Y D V P D
579 tacgctggaggtatgacgtgtgaggtgaaggagaagggccgagttgggacgataaatcct
580 Y A G G M T C E V K E K G R V G T I N P
581 atttttacctgccagcctgcgggcgctcagtttgtatctatcgggatcaaagactgtatc
582 I F T C Q P A G A Q F V S I G I K D C I
583 ggtatcgtgcatggaggacaaggggtgctgatgttcgtgcgactcatcttctcacagcat
584 G I V H G G Q G C V M F V R L I F S Q H
585 tacaaggagagctttgagctggcatcatcctctctccacgaggacggggccgtgttcggg
586 Y K E S F E L A S S L H E D G A V F G
587 gcctgtggccgtgttgaagaggcagttgatgtcttgccttcccggttatcccgcgtaaaag
588 A C G R V E E A V D V L L S R Y P D V K
589 gttgtaccaattattactacatgctccacggaaatcataggcgacgatgtggatgggggtt
590 V V P I I T T C S T E I I G D D V D G V
591 atcaagaagttgaatgaggggtctgctgaaagagaagttccctgaccgagaagtccacctt
592 I K K L N E G L L K E K F P D R E V H L
593 atagcgatgcatactcctagttttgtaggatccatgataagtggctacgatgttgctgtg
594 I A M H T P S F V G S M I S G Y D V A V

595 cgtgatgtagtacgacacttcgccaaagagagaagccccgaacgataaaaataaatcttctt
 596 R D V V R H F A K R E A P N D K I N L L
 597 acaggatgggtcaatcctggtgatgtgaaagaacttaaacatttactcggcgagatggat
 598 T G W V N P G D V K E L K H L L G E M D
 599 attgaggcgaacgtactttttgaaattgaatcttttgactctcctatcttacctgacggc
 600 I E A N V L F E I E S F D S P I L P D G
 601 tccgcggttagtcacggaaatacaactatagaggacctaattgacacaggtaacgcccga
 602 S A V S H G N T T I E D L I D T G N A R
 603 gcgacgttcgctctaaacagatatgaaggaacgaaagcggcgagagtatctacaaaagaaa
 604 A T F A L N R Y E G T K A A E Y L Q K K
 605 tttgagattcctgccattataggcccaacgcccatagggattcgaacacggacatatctc
 606 F E I P A I I G P T P I G I R N T D I F
 607 cttcagaatctgaagaaagctacagggaaccaattccgcagagcttggcccacgagcgt
 608 L Q N L K K A T G K P I P Q S L A H E R
 609 ggagtggctattgatgctttggcgacctgacacacatgtttctagcagaaaagagagtc
 610 G V A I D A L A D L T H M F L A E K R V
 611 gctatctatggggcaccgatttagtgataggccttgacagagttctgcctagatctagag
 612 A I Y G A P D L V I G L A E F C L D L E
 613 atgaagccggtcctttttattacttgggtgacgacaattccaaatacgtcgacgatcccagg
 614 M K P V L L L L G D D N S K Y V D D P R
 615 attaaagccctacaagaaaacgtcgactacgggatggagattgttacgaacgcggacttc
 616 I K A L Q E N V D Y G M E I V T N A D F
 617 tgggagctcgagaataggatcaagaatgagggcctcgaactcgacctgatactggggcac
 618 W E L E N R I K N E G L E L D L I L G H
 619 tctaagggttaggtttatcagcatcgactacaatatccgatgttgagagtgggcttcccg
 620 S K G R F I S I D Y N I P M L R V G F P
 621 acgtatgatcgagctggacttttttaggtatcccaccgttggctacggaggggctatatgg
 622 T Y D R A G L F R Y P T V G Y G G A I W
 623 ttagcggagcaaatggcgaacactttgttcgcagacatggaacacaagaaaaacaaggag
 624 L A E Q M A N T L F A D M E H K K N K E
 625 tgggtattgaatgtatgg
 626 W V L N V W

627

628 **SN156, alaFAy51-HA-AnfH:**

629 atggcaatggctgttttccgtcgcgaaagcggcggtctctcctcccttcaatcgccgctcgc
 630 M A M A V F R R E A A A L L P S I A A R
 631 ccaatcgctgctgctgcagctgccgcttcttctgaccaggaggaaggacttcttgctgcc
 632 P I A A A A A A S S D Q E E G L L A A
 633 gcagcggctgcggccgctgtggtgcgtaaccgcggaggttacccttacgacgttcctgat
 634 A A A A A A V V R N R G G **Y P Y D V P D**
 635 tacgctggaggtatgacgcgaaaagtggcgatttatggcaaaggcggataggtaaatca
 636 **Y A** G G M T R K V A I Y G K G G I G K S
 637 acaacgacccaaaataaccgctgctgctttggcctacttccatgataaaaagggttttcac
 638 T T T Q N T A A A L A Y F H D K K V F I
 639 catggctgcgaccctaaagcggacagtacacgattaatactaggcggaaagccccaggaa
 640 H G C D P K A D S T R L I L G G K P Q E

641 accctaattggatatgctacgtgataagggagctgagaaaattactaatgatgacgtaatc
 642 T L M D M L R D K G A E K I T N D D V I
 643 aaaaaggggttcctcgacatacagtgcggttgagagtggtggacccgaaccaggcgctcggg
 644 K K G F L D I Q C V E S G G P E P G V G
 645 tgtgcaggtaggggtgttataacggccatcgacctcatggaagagaatggagcggtacaca
 646 C A G R G V I T A I D L M E E N G A Y T
 647 gacgacctagatTTTTgtctTTTTcgatgtccttggtgatgtggtgtgcggcggttttgct
 648 D D L D F V F F D V L G D V V C G G F A
 649 atgcccattagagatggaaaggcacaggaggtgtatatattgttgcctccggggaaatgatg
 650 M P I R D G K A Q E V Y I V A S G E M M
 651 gcgatttacgcggcaacaatatctgcaaggggtctagtgaatatgcaagcagtcctgga
 652 A I Y A A N N I C K G L V K Y A K Q S G
 653 gttagactggggggtataaatttgtaacagtcgaaaagtggacggtgagaggggaattcctc
 654 V R L G G I I C N S R K V D G E R E F L
 655 gaggagtttaccgctgcaatagggacgaaaatgatacatttcgtgccagagacaatatc
 656 E E F T A A I G T K M I H F V P R D N I
 657 gtgcaaaaagctgaattcaataaaaagaccgtgacggagttcgtccagaggaaaaccaa
 658 V Q K A E F N K K T V T E F A P E E N Q
 659 gctaaagaatatggagaactcgcccgtaaaatcatcgagaatgatgagttcgttatccca
 660 A K E Y G E L A R K I I E N D E F V I P
 661 aagcctctcactatggaccagctagaggacatggtagtgaagtatgggatagccgat
 662 K P L T M D Q L E D M V V K Y G I A D

663

664 **SN157, alaFAy51-HA-AnfG:**

665 atggcaatggctgttttccgtcgcaagcggcggtctctcctcccttcaatcgccgctcgc
 666 M A M A V F R R E A A A L L P S I A A R
 667 ccaatcgctgctgctgcagctgccgcttcttctgaccaggaggaaggacttcttgctgcc
 668 P I A A A A A A A S S D Q E E G L L A A
 669 gcagcggctgcggccgctgtggtgcgtaaccgcggaggttacccttacgacgttcctgat
 670 A A A A A A V V R N R G G **Y P Y D V P D**
 671 tacgctggaggtatgtccacggcctccgccgctgccgtagtcaaacagaaagtagaggcg
 672 **Y A** G G M S T A S A A A V V K Q K V E A
 673 cccgtacaccgatggatgcgagaatcgacgaacttacagattacataatgaagaactgc
 674 P V H P M D A R I D E L T D Y I M K N C
 675 ctctggcagttccacagcaggagttgggacagggagaggcagaatgccgaaattctcaag
 676 L W Q F H S R S W D R E R Q N A E I L K
 677 aaaacgaaagaattactttgtggcgagcctgtagatctatcaacatcacacgaccgatgc
 678 K T K E L L C G E P V D L S T S H D R C
 679 tattgggttgacgctgtctgcctagcggatgactaccgtgagcactaccggtggataaac
 680 Y W V D A V C L A D D Y R E H Y P W I N
 681 agtatgtccaaggaggagataggaagtctaattgcaaggattaaaggaccgtatggactac
 682 S M S K E E I G S L M Q G L K D R M D Y
 683 ttgaccataacggggagcctaaatgaggagttatctgacaagcattac
 684 L T I T G S L N E E L S D K H Y

685

686 **SN158, alaFAy51-HA-AnfD:**

687 atggcaatggctgttttccgtcgcgaagcggcggtctctcctcccttcaatcgccgctcgc
688 M A M A V F R R E A A A L L P S I A A R
689 ccaatcgctgctgctgcagctgccgcttcttctgaccaggaggaaggacttcttgctgcc
690 P I A A A A A A S S D Q E E G L L A A
691 gcagcggctgcggccgctgtggtgcgtaaccgcggaggttacccttacgacgttcctgat
692 A A A A A A V V R N R G G Y P Y D V P D
693 tacgctggaggtatgcctcatcatgagttcgagtggtccaaagtaatcccgagagaaaag
694 **Y A** G G M P H H E F E C S K V I P E R K
695 aagcatgccgttataaaaggaaagggcgagacactagcggacgcgcttcctcaaggatat
696 K H A V I K G K G E T L A D A L P Q G Y
697 ctaaacacaatacctgggtcaatttccgaaagaggttggtgcttattgcggtgcaaagcac
698 L N T I P G S I S E R G C A Y C G A K H
699 gttattggaaccccaatgaaagatgttatacacatttcccacggaccagttggctgtact
700 V I G T P M K D V I H I S H G P V G C T
701 tacgatacatggcagacgaaacgttacatatccgacaatgacaatttccagctaaagtat
702 Y D T W Q T K R Y I S D N D N F Q L K Y
703 acttatgcgacggacgtaaaagaaaaacacatagtccttggagctgaaaaattactgaaa
704 T Y A T D V K E K H I V F G A E K L L K
705 caaaacatcattgaggcgtttaaggcattcccgcaaatcaagcgaatgactatctaccag
706 Q N I I E A F K A F P Q I K R M T I Y Q
707 acttggtgcgactgctttgatcggagatgatattaatgccatcgcggaagaggtaatggag
708 T C A T A L I G D D I N A I A E E V M E
709 gagatgcctgaagttgacatatgttggcaactctcccgattcgctggaccatctcag
710 E M P E V D I F V C N S P G F A G P S Q
711 agtggcgggcatcacaagatcaatattgcctggattaatcagaaggcggcacagtagag
712 S G G H H K I N I A W I N Q K V G T V E
713 cctgaaattaccggtgaccatgtcataaactatgtgggggaatataacatacaaggcgac
714 P E I T G D H V I N Y V G E Y N I Q G D
715 caggaggttatggttgactactttaagagaatggggatccaggtcttgtccactttcacc
716 Q E V M V D Y F K R M G I Q V L S T F T
717 ggcaacggaagttatgatgggttgagagctatgcaccgtgcgcacttgaatgtattagag
718 G N G S Y D G L R A M H R A H L N V L E
719 tgtgctcgtagecgcagagtacatttgcaatgaactaagagttagatacggcatccctcga
720 C A R S A E Y I C N E L R V R Y G I P R
721 cttgatatagacgggttttgattcaaaccacttgcggaactcactccgtaagattggaatg
722 L D I D G F G F K P L A D S L R K I G M
723 ttctttgggatagaggaccgtgcaaaagccatcatcgacgaggaagtagctcgttgaaa
724 F F G I E D R A K A I I D E E V A R W K
725 ccggaattggattggtataaggagcgactcatgggcaagaaggtgtgcctatggcctgg
726 P E L D W Y K E R L M G K K V C L W P G
727 ggctccaaattgtggcattgggcgcacgtcatcgaggaagaaatgggattgaaggtcgtg
728 G S K L W H W A H V I E E E M G L K V V
729 tccgtctacaccaagttcggacatcaaggcgacatggagaagggaatagcccgatgtggc
730 S V Y T K F G H Q G D M E K G I A R C G
731 gaaggtactcttgcgatcgatgatccaaatgaactcgaaggtttagaagctcttgagatg
732 E G T L A I D D P N E L E G L E A L E M

733 cttaaaccagatatataatTTTTGACGGGAAAGCGTCCAGGTGAGGTGGCAAAAAAGTCCGT
 734 L K P D I I L T G K R P G E V A K K V R
 735 gttccttacctcaatgccacgcttaccacaacgggccttacaaaggTTTTGAAGGTTGG
 736 V P Y L N A H A Y H N G P Y K G F E G W
 737 gtgCGTTTTGCTAGGGATATTTATAATGCTATCTACAGTCCAATCCATCAGCTGTCCGGA
 738 V R F A R D I Y N A I Y S P I H Q L S G
 739 atcgatatataactaaggataatgcccctgaatgggggaatggctttcgtacgcgcacagatg
 740 I D I T K D N A P E W G N G F R T R Q M
 741 ctCAGTGACGGGAATTTGAGCGACGCTGTcaggaattcagaaacactgcgtcagtacaca
 742 L S D G N L S D A V R N S E T L R Q Y T
 743 ggCGGTTATGACTCTGTATCAAAAactgcgagaaagagagtatcccgCGTTTGAACGTAAG
 744 G G Y D S V S K L R E R E Y P A F E R K
 745 gttggt V G

746

747 **SN161, pFAy51-HA-AnfD:**

748 atggcaatggctgttttccgtcgCGAAGGGAGGCGTCTCCTCCCTTCAATCGCCGCTCGC
 749 M A M A V F R R E G R R L L P S I A A R
 750 ccaatCGCTGCTATCCGATCTCCTCTCTCTTCTGACCAGGAGGAAGGACTTCTTGAGTT
 751 P I A A I R S P L S S D Q E E G L L G V
 752 CGATCTATCTCAACTCAAGTGGTGCgtaaccgcggaggatacccttacgacgttcctgat
 753 R S I S T Q V V R N R G G Y P Y D V P D
 754 tacgctggaggtatgcctcatcatgagttcgagtggtccaaagtaatcccgGAGAGAAAG
 755 **Y A** G G M P H H E F E C S K V I P E R K
 756 aagcatgccgtttataaaaggaaagggcgagacactagcggacgcgcttcctcaaggatat
 757 K H A V I K G K G E T L A D A L P Q G Y
 758 ctaaacacaatacctgggtcaattttccgaaagaggttggtgcttattgcggtgcaaagcac
 759 L N T I P G S I S E R G C A Y C G A K H
 760 gttattggaaccccaatgaaagatgttatacacattttccacggaccagttggctgtact
 761 V I G T P M K D V I H I S H G P V G C T
 762 tacgatacatggcagacgaaacgttacatatccgacaatgacaatttccagctaaagtat
 763 Y D T W Q T K R Y I S D N D N F Q L K Y
 764 acttatgcgacggacgtaaaagaaaaacacatagTCTTTGGAGCTGAAAAATTACTGAAA
 765 T Y A T D V K E K H I V F G A E K L L K
 766 caaaacatcattgaggcgTTTAAGGCATTCCCGCAATCAAGCGAATGACTATCTACCAG
 767 Q N I I E A F K A F P Q I K R M T I Y Q
 768 acttgTgcgactgctttgatcggagatgatattaatgccatcgCGGAAGAGGTAATGGAG
 769 T C A T A L I G D D I N A I A E E V M E
 770 gagatgcctgaagttgacatatTTTgtttgcaactctcccgattcgctggaccatctcag
 771 E M P E V D I F V C N S P G F A G P S Q
 772 agtggcgggcatcacaagatcaatattgcctggattaatcagaaggTcggcacagtagag
 773 S G G H H K I N I A W I N Q K V G T V E
 774 cctgaaattaccggtgaccatgtcataaactatgtgggggaatataacatacaaggcgac
 775 P E I T G D H V I N Y V G E Y N I Q G D
 776 caggaggttatggttgactactTTAAGAGAATGGGGATCCAGGTCTTGTCCACTTTCACC
 777 Q E V M V D Y F K R M G I Q V L S T F T
 778 ggcaacggaagttatgatgggttgagagctatgcaccgtgcgcacttgaatgtattagag

779 G N G S Y D G L R A M H R A H L N V L E
 780 tgtgctcgtagcgcagagtacatttgcaatgaactaagagttagatacggcatccctcga
 781 C A R S A E Y I C N E L R V R Y G I P R
 782 cttgatatagacggttttggattcaaaccacttgcgactcactccgtaagattggaatg
 783 L D I D G F G F K P L A D S L R K I G M
 784 ttctttgggatagaggaccgtgcaaaagccatcatcgacgaggaagtagctcgttgga
 785 F F G I E D R A K A I I D E E V A R W K
 786 ccggaattggattggtataaggagcgactcatgggcaagaagggtgtgcctatggcctggt
 787 P E L D W Y K E R L M G K K V C L W P G
 788 ggctccaaattgtggcattgggcgacgtcatcgaggaagaaatgggattgaaggctcgtg
 789 G S K L W H W A H V I E E E M G L K V V
 790 tccgtctacaccaagttcggacatcaaggcgacatggagaagggaatagcccgatgtggc
 791 S V Y T K F G H Q G D M E K G I A R C G
 792 gaaggtagctccttgcgatcgatgatccaaatgaactcgaaggtttagaagctccttgagatg
 793 E G T L A I D D P N E L E G L E A L E M
 794 cttaaaccagatataattttgacgggaaagcggtccaggtgaggtggcaaaaaaagtcctg
 795 L K P D I I L T G K R P G E V A K K V R
 796 gttccttacctcaatgccacgcttaccacaacgggccttacaaggttttgaaggttgg
 797 V P Y L N A H A Y H N G P Y K G F E G W
 798 gtgcgttttgcctagggatatttataatgctatctacagtccaatccatcagctgtccgga
 799 V R F A R D I Y N A I Y S P I H Q L S G
 800 atcgatataactaaggataatgccctgaatgggggaatggctttcgtacgcgacagatg
 801 I D I T K D N A P E W G N G F R T R Q M
 802 cttagtgacgggaatttgagcgacgctgtcaggaattcagaaacactgcgtcagtacaca
 803 L S D G N L S D A V R N S E T L R Q Y T
 804 ggcggttatgactctgtatcaaaactgcgagaaagagagtatcccgcggtttgaacgtaag
 805 G G Y D S V S K L R E R E Y P A F E R K
 806 gttggt
 807 V G

808

809 SN177, pFAy51-AnfD-Twin Strep:

810 atgatggcaatggctgttttccgctcgcaagggaggcgtctcctcccttcaatcgccgct
 811 M M A M A V F R R E G R R L L P S I A A
 812 cgcccaatcgtgctatccgatctcctctctcttctgaccaggaggaaggacttcttgga
 813 R P I A A I R S P L S S D Q E E G L L G
 814 gttcgatctatctcaactcaagtgggtgcgtaaccgcggaggtatgcctcatcatgagttc
 815 V R S I S T Q V V R N R G G M P H H E F
 816 gagtgttccaaagtaatcccgagagaaagaagcatgccgttataaaaggaaagggcgag
 817 E C S K V I P E R K K H A V I K G K G E
 818 aactagcggacgcgcttcctcaaggatatctaaacacaatacctgggtcaatttccgaa
 819 T L A D A L P Q G Y L N T I P G S I S E
 820 agaggttggtgcttattgcggtgcaaaagcaggttattggaaccccaatgaaagatgttata
 821 R G C A Y C G A K H V I G T P M K D V I
 822 cacatttcccacggaccagttggctgtacttacgatacatggcagacgaaacggttacata
 823 H I S H G P V G C T Y D T W Q T K R Y I
 824 tccgacaatgacaatttccagctaaagtatacttatgcgacggacgtaaaagaaaaaac

825 S D N D N F Q L K Y T Y A T D V K E K H
826 atagtcctttggagctgaaaaattactgaaacaaacatcattgaggcgctttaaggcattc
827 I V F G A E K L L K Q N I I E A F K A F
828 ccgcaaatacaagcgaatgactatctaccagacttgtgcgactgctttgatcggagatgat
829 P Q I K R M T I Y Q T C A T A L I G D D
830 attaatgccatcgcggaagaggtaatggaggagatgcctgaagttgacatatttgtttgc
831 I N A I A E E V M E E M P E V D I F V C
832 aactctcccgattcgctggaccatctcagagtggcgggcatcacaaagatcaatatattgcc
833 N S P G F A G P S Q S G G H H K I N I A
834 tggattaatcagaaggtcggcacagtagagcctgaaattaccggtgaccatgtcataaac
835 W I N Q K V G T V E P E I T G D H V I N
836 tatgtgggggaatataacatacaaggcgaccaggaggttatgggttgactactttaagaga
837 Y V G E Y N I Q G D Q E V M V D Y F K R
838 atgggggatccagggtcttgtccactttcaccggcaacggaagttatgatgggttgagagct
839 M G I Q V L S T F T G N G S Y D G L R A
840 atgcaccgtgcgacttgaatgtattagagtgtgctcgtagcgcagagtacatttgcaat
841 M H R A H L N V L E C A R S A E Y I C N
842 gaactaagagttagatacggcatccctcgacttgatatagacggttttggattcaaacca
843 E L R V R Y G I P R L D I D G F G F K P
844 cttgcggaactcactccgtaagattggaatgttctttgggatagaggaccgtgcaaaagcc
845 L A D S L R K I G M F F G I E D R A K A
846 atcatcgacgaggaagtagctcgttggaaaccggaattggattggtataaggagcgactc
847 I I D E E V A R W K P E L D W Y K E R L
848 atgggcaagaaggtgtgcctatggcctgggtggctccaaattgtggcattggggcgacgctc
849 M G K K V C L W P G G S K L W H W A H V
850 atcgaggaagaaatgggattgaaggtcgtgtccgtctacaccaagttcggacatcaaggc
851 I E E E M G L K V V S V Y T K F G H Q G
852 gacatggagaagggaatagcccgatgtggcgaagggtactccttgcgatcgatgatccaat
853 D M E K G I A R C G E G T L A I D D P N
854 gaactcgaaggtttagaagctcttgagatgcttaaaccagatataattttgacgggaaag
855 E L E G L E A L E M L K P D I I L T G K
856 cgtccagggtgaggtggcaaaaaaagtccgtgttccttacctcaatgcccacgcttaccac
857 R P G E V A K K V R V P Y L N A H A Y H
858 aacgggccttacaaaggttttgaaggttgggtgctgttttgctagggatattttataatgct
859 N G P Y K G F E G W V R F A R D I Y N A
860 atctacagtccaatccatcagctgtccggaatcgatataactaaggataatgcccctgaa
861 I Y S P I H Q L S G I D I T K D N A P E
862 tgggggaatggcttttcgtacgcgacagatgctcagtgacgggaatttgagcgacgctgtc
863 W G N G F R T R Q M L S D G N L S D A V
864 aggaattcagaaacactgcgtcagtacacaggcggttatgactctgtatcaaaaactgcga
865 R N S E T L R Q Y T G G Y D S V S K L R
866 gaaagagagtatcccgctttgaacgtaaggttgggtgggtcagcatggagtcacatcct
867 E R E Y P A F E R K V G G G **S A W S H P**
868 cagtttgagaaaggtggaggttcaggtgggtggaagcggtggatctgcttgggtcacatcca
869 **Q F E K G G G S G G G S G G S A W S H P**
870 caatttgaaaaa
871 **Q F E K**

872 **SN195, CoxIV-Twin Strep-AnfK:**

873 atgcttttcacttagacaatctattagatTTTTcaagccagctacaagaactttgtgttct
874 M L S L R Q S I R F F K P A T R T L C S
875 tctagatatcttcttcagcaaaaaccttcagcatggagtcctcctcagtttgagaaaggt
876 S R Y L L Q Q K P S A W S H P Q F E K G
877 ggaggttcaggtggtggaagcgggtggatctgcttggtcacatccacaatttgaaaaagga
878 **G G S G G G S G G S A W S H P Q F E K G**
879 ggtatgacgtgtgaggtgaaggagaagggccgagttgggacgataaatcctatttttacc
880 G M T C E V K E K G R V G T I N P I F T
881 tgccagcctgcgggcgctcagtttgtatctatcgggatcaaagactgtatcggtatcgtg
882 C Q P A G A Q F V S I G I K D C I G I V
883 catggaggacaaggggtgctgatgttcgtgcgactcatcttctcacagcattacaaggag
884 H G G Q G C V M F V R L I F S Q H Y K E
885 agctttgagctggcatcatcctctctccacgaggacggggccgtgttcggggcctgtggc
886 S F E L A S S S L H E D G A V F G A C G
887 cgtgttgaagaggcagttgatgtcttgccttcccggttatcccgacgtaaaggttgacca
888 R V E E A V D V L L S R Y P D V K V V P
889 attattactacatgctccacggaaatcataggcgacgatgtggatgggggttatcaagaag
890 I I T T C S T E I I G D D V D G V I K K
891 ttgaatgagggctctgctgaaagagaagttccctgaccgagaagtccaccttatagcgatg
892 L N E G L L K E K F P D R E V H L I A M
893 catactcctagttttgtaggatccatgataagtggctacgatgttgctgtgcgtgatgta
894 H T P S F V G S M I S G Y D V A V R D V
895 gtacgacacttcgccaaagagagaagccccgaacgataaaataaatcttcttacaggatgg
896 V R H F A K R E A P N D K I N L L T G W
897 gtcaatcctggtgatgtgaaagaacttaaacattttactcggcgagatggatattgaggcg
898 V N P G D V K E L K H L L G E M D I E A
899 aacgtactttttgaaattgaatcttttgactctcctatcttacctgacggctccgcggtt
900 N V L F E I E S F D S P I L P D G S A V
901 agtcacggaaatacaactatagaggacctaattgacacaggtaacgcccagcgacgttc
902 S H G N T T I E D L I D T G N A R A T F
903 gctctaaacagatatgaaggaacgaaagcggcagagtatctacaaaagaaatttgagatt
904 A L N R Y E G T K A A E Y L Q K K F E I
905 cctgccattataggcccaacgcccatagggattcgaaacacggacatattccttcagaat
906 P A I I G P T P I G I R N T D I F L Q N
907 ctgaagaaagctacagggaaaccaattccgcagagcttggcccacgagcgtggagtggt
908 L K K A T G K P I P Q S L A H E R G V A
909 attgatgctttggcggacctgacacacatgtttctagcagaaaagagagtcgctatctat
910 I D A L A D L T H M F L A E K R V A I Y
911 ggggcacccgatttagtgataggccttgacagagttctgcctagatctagagatgaagccg
912 G A P D L V I G L A E F C L D L E M K P
913 gtcctttttattacttggtgacgacaattccaaatacgtcgacgatcccaggattaaagcc
914 V L L L L G D D N S K Y V D D P R I K A
915 ctacaagaaaacgtcgactacgggatggagattgttacgaacgcggacttctgggagctc
916 L Q E N V D Y G M E I V T N A D F W E L
917 gagaataggatcaagaatgagggcctcgaactcgacctgatactggggcactctaaggggt
918 E N R I K N E G L E L D L I L G H S K G

919 aggtttatcagcatcgactacaatatccgatgttgagagtgggcttcccgacgtatgat
 920 R F I S I D Y N I P M L R V G F P T Y D
 921 cgagctggacttttttaggtatcccaccgttggtctacggaggggctatatgggttagcggag
 922 R A G L F R Y P T V G Y G G A I W L A E
 923 caaatggcgaacactttgttcgcagacatggaacacaagaaaaacaaggagtgggtattg
 924 Q M A N T L F A D M E H K K N K E W V L
 925 aatgtatgg
 926 N V W

927

928 **SN227, pFAy51-AnfD-HA:**

929 atgatggcaatggctgttttccgtcgcgaagggaggcgtctcctcccttcaatcgccgct
 930 M M A M A V F R R E G R R L L P S I A A
 931 cgcccaatcgctgctatccgatctcctctctcttctgaccaggaggaaggacttcttggg
 932 R P I A A I R S P L S S D Q E E G L L G
 933 gttcgatctatctcaactcaagtgggtgcgtaaccgcggaggtatgcctcatcatgagttc
 934 V R S I S T Q V V R N R G G M P H H E F
 935 gagtgttccaaagtaatcccggagagaaagaagcatgccgttataaaaggaaagggcgag
 936 E C S K V I P E R K K H A V I K G K G E
 937 acactagcggacgcgcttcctcaaggatatctaaacacaatacctgggtcaatttccgaa
 938 T L A D A L P Q G Y L N T I P G S I S E
 939 agaggttggtgcttattgcggtgcaaagcacgttattggaaccccaatgaaagatgttata
 940 R G C A Y C G A K H V I G T P M K D V I
 941 cacatttcccacggaccagttgggtgtacttacgatacatggcagacgaaacggttacata
 942 H I S H G P V G C T Y D T W Q T K R Y I
 943 tccgacaatgacaatttccagctaaagtatacttatgcgacggacgtaaaagaaaaacac
 944 S D N D N F Q L K Y T Y A T D V K E K H
 945 atagtctttggagctgaaaaattactgaaacaaaacatcattgaggcgtttaaggcattc
 946 I V F G A E K L L K Q N I I E A F K A F
 947 ccgcaaataagcgaatgactatctaccagacttggtgcgactgctttgatcggagatgat
 948 P Q I K R M T I Y Q T C A T A L I G D D
 949 attaatgccatcgcggaagaggtaatggaggagatgcctgaagttgacatatttgtttgc
 950 I N A I A E E V M E E M P E V D I F V C
 951 aactctcccgattcgctggaccatctcagagtggcgggcatcacaaagatcaatatggcc
 952 N S P G F A G P S Q S G G H H K I N I A
 953 tggattaatcagaaggtcggcacagtagagcctgaaattaccggtgaccatgtcataaac
 954 W I N Q K V G T V E P E I T G D H V I N
 955 tatgtgggggaatataacatacaaggcgaccaggaggttatggttgactactttaagaga
 956 Y V G E Y N I Q G D Q E V M V D Y F K R
 957 atgggggatccaggtcttgtccactttcaccggcaacggaagttatgatgggttgagagct
 958 M G I Q V L S T F T G N G S Y D G L R A
 959 atgcaccgtgcgacacttgaatgtattagagtgtgctcgtagcgcagagtacatttgcaat
 960 M H R A H L N V L E C A R S A E Y I C N
 961 gaactaagagtttagatacggcatccctcgacttgatatagacggttttggattcaaacca
 962 E L R V R Y G I P R L D I D G F G F K P
 963 cttgcggactcactccgtaagattggaatgttctttgggatagaggaccgtgcaaaagcc
 964 L A D S L R K I G M F F G I E D R A K A

965 atcatcgacgaggaagtagctcggttgaaaccggaattggattggtataaggagcgactc
 966 I I D E E V A R W K P E L D W Y K E R L
 967 atgggcaagaaggtgtgcctatggcctgggtggctccaaattgtggcattgggcgacgctc
 968 M G K K V C L W P G G S K L W H W A H V
 969 atcgaggaagaaatgggattgaaggtcgtgtccgtctacaccaagttcggacatcaaggc
 970 I E E E M G L K V V S V Y T K F G H Q G
 971 gacatggagaagggaatagccccgatgtggcggaaggtactcttgcgatcgatgatccaaat
 972 D M E K G I A R C G E G T L A I D D P N
 973 gaactcgaaggtttagaagctcttgagatgcttaaaccagatataattttgacgggaaag
 974 E L E G L E A L E M L K P D I I L T G K
 975 cgtccaggtgaggtggcaaaaaaagtcctgttccttacctcaatgccacgcttaccac
 976 R P G E V A K K V R V P Y L N A H A Y H
 977 aacgggccttacaaagggtttgaagggtgggtgcgttttgctagggatatttataatgct
 978 N G P Y K G F E G W V R F A R D I Y N A
 979 atctacagtccaatccatcagctgtccggaatcgatataactaaggataatgcccctgaa
 980 I Y S P I H Q L S G I D I T K D N A P E
 981 tgggggaatggctttcgtacgcgacagatgctcagtgacgggaatttgagcgacgctgtc
 982 W G N G F R T R Q M L S D G N L S D A V
 983 aggaattcagaaacactgcgtcagtacacaggcggttatgactctgtatcaaaaactgcca
 984 R N S E T L R Q Y T G G Y D S V S K L R
 985 gaaagagagtatcccgcgtttgaacgtaagggttggtgggtggtacccttacgacgttcct
 986 E R E Y P A F E R K V G G G **Y P Y D V P**
 987 gattacgct
 988 **D Y A**

989

990 **SN371, CoxIV-Twin Strep-AnfG:**

991 atgcttttcacttagacaatctattagatttttcaagccagctacaagaactttgtgttct
 992 M L S L R Q S I R F F K P A T R T L C S
 993 tctagatatcttcttcagcaaaaaccttcagcatggagtcacctcagtttgagaaaggt
 994 S R Y L L Q Q K P **S A W S H P Q F E K G**
 995 ggaggttcaggtggtggaagcgggtggatctgcttggtcacatccacaatttgaaaaagga
 996 **G G S G G G S G G S A W S H P Q F E K G**
 997 ggtatgtccacggcctccgcgctgccgtagtcaaacagaaagtagaggcgcccgtagac
 998 G M S T A S A A A V V K Q K V E A P V H
 999 ccgatggatgcgagaatcgacgaacttacagattacataatgaagaactgcctctggcag
 1000 P M D A R I D E L T D Y I M K N C L W Q
 1001 ttccacagcaggagttgggacaggagaggcagaatgccgaaattctcaagaaaacgaaa
 1002 F H S R S W D R E R Q N A E I L K K T K
 1003 gaattactttgtggcgagcctgtagatctatcaacatcacacgaccgatgctattgggtt
 1004 E L L C G E P V D L S T S H D R C Y W V
 1005 gacgctgtctgcctagcggatgactaccgtgagcactaccctgggataaacagtatgtcc
 1006 D A V C L A D D Y R E H Y P W I N S M S
 1007 aaggaggagataggaagtctaatagcaaggattaaaggaccgtatggactacttgaccata
 1008 K E E I G S L M Q G L K D R M D Y L T I
 1009 acggggagcctaaatgaggagttatctgacaagcattac
 1010 T G S L N E E L S D K H Y