Supporting Online Information

Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids

Moritz Pott, Moritz Johannes Schmidt and Daniel Summerer*

Department of Chemistry and Konstanz Research School Chemical Biology
University of Konstanz
Universitätsstraße 10
78457 Konstanz (Germany)
E-mail: daniel.summerer@uni-konstanz.de

Oligonucleotide sequences 3

Table S 1: Sequences of clones in unselected library 5

Table S 2: Sequences of clones identified in early selection rounds with ncAA 1. 5

Table S 3: Sequences of clones identified in early selection rounds with ncAA 6. 5

Table S 4. Sequences of clones identified in early selection rounds with the tyrosine codon TAT. 6

Table S 5: Sequences of clones identified in early selection rounds with the lysine codon AAA. 6

Figure S 1: Spectra of 1-R5-2(TAT)_TRX-6His expression (from Figure 5A). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum. 7

Figure S 2: Spectra of 1-R5-2_TRX-6His expression in presence of ncAA 1 (from Figure 5A). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum. 8

Figure S 3: Spectra of 8-R11-1(AAA)_TRX-6His expression (Figure 5B). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum. 9

Protein Sequences of TRX constructs 10

ESI-MS mass analyses 10

Oligonucleotide sequences

Name	Descriptor	Sequence
oDaS21	*XbaI*_pEVOL_5´-PCR_primer	TTT TTT CTA GAA TCA ATC ATC CCC ATA ATC CTT GTT
oDaS22	*SphI*_pEVOL-3´-PCR_primer	TTT TTG CAT GCA GGC AGT TAT TGG TGC CCT TAA AC
oDaS86	GFP_*NheI*_fwd	ATG GAT TAC AAA GAT GAT GAT GCT AGC GTG AGC AAG GGC GAG
oDaS87	GFP_*NheI*_rev	AAC AGC TCC TCG CCC TTG CTC ACG CTA GCA TCA TCA T
oDaS338	*NotI*_TRX_rev	TTT TGC GGC CGC TCA ATG ATG ATG ATG ATG ATG CTT GTC
oMoP690	*HindIII*_pBAD_Flag_fwd	CCC GTT TTT TGG GCT AAC AGG AGG AAG CTT CCA TGG ATT ACA AAG ATG ATG
oMoP691	*HindIII*_pBAD_Flag_rev	CAT CAT CTT TGT AAT CCA TGG AAG CTT CCT CCT GTT AGC CCA AAA AAC GGG
oMoP699	*NheI*_CAT_fwd	TTT TTG CTA GCA TGG AGA AAA AAA TCA CTG GAT ATA CCA CCG TTG ATA TA
oMoP700	*NotI*_CAT_rev	A AAA AGC GGC CGC TCA CGC CCC GCCCTG
oMoP817	HindIII_tag_CAT_fwd	T TTT TAA GCT TCC ATG GGT NNK NNK TAG NNK NNK GGT GGT GCT AGC GAG AAA AAA ATC ACT GGA TAT AC
oMoP837	*HindIII*_tag_GFP_fwd	T TTT TAA GCT TCC ATG GGT NNK NNK TAG NNK NNK GGT GGT GCT AGC GTG AGC AAG GGC GAG
oMoP838	*NotI*_GFP_rev	AAA AAG CGG CCG CTC AGT GAT GGT GAT GGT GAT GCT TGT ACA GCT
oMoP1032	*HindIII*_**1**-R5-2(TAT)_CAT_fwd	T TTT TAA GCT TCC ATG GGT AAT AAT TAT ACG AAG GGT GGT GCT AGC GAG AAA AAA ATC ACT GGA TAT A
oMoP1033	*HindIII*_**8**-R11-1(AAA)_CAT_fwd	T TTT TAA GCT TCC ATG GGT CAG AAG AAA AAG CAG GGT GGT GCT AGC GAG AAA AAA ATC ACT GGA T
oMoP1120	*HindIII*_tag(TAT)_CAT_fwd	T TTT TAA GCT TCC ATG GGT NNK NNK TAT NNK NNK GGT GGT GCT AGC GAG AAA AAA ATC ACT GGA TAT AC
oMoP1121	*HindIII*_tag(AAA)_CAT_fwd	T TTT TAA GCT TCC ATG GGT NNK NNK AAA NNK NNK GGT GGT GCT AGC GAG AAA AAA

		ATC ACT GGA TAT AC
oMoP1103	*NcoI*_L1_TRX_fwd	T TTT TCC ATG GGT CCT CTT TAG CTG TTT GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1104	*NcoI*_L2_TRX_fwd	T TTT TCC ATG GGT CAG AGT TAG CCG AAG GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1105	NcoI_L8_TRX_fwd	T TTT TCC ATG GGT CAT CCG TAG ACT CCT GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1119	*NcoI*_**1**-R5-2_TRX_fwd	T TTT TCC ATG GGT AAT AAT TAG ACG AAG GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1120	*NcoI*_**1**-R5-2(TAT)_TRX_fwd	T TTT TCC ATG GGT AAT AAT TAT ACG AAG GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1160	*NcoI*_**8**-R11-1_TRX_fwd	T TTT TCC ATG GGT CAT CAT TAG AAG AAT GGT GGT GCT AGC GGC GAT AAA ATT ATT
oMoP1161	*NcoI*_**8**-R11-1(AAA)_TRX_fwd	T TTT TCC ATG GGT CAT CAT AAA AAG AAT GGT GGT GCT AGC GGC GAT AAA ATT ATT

Table S 1: Sequences of clones in unselected library

Library Clones before Selection		
Name	Sequence[1]	Occurence
0-R0-1	**F F F F** TTT TTT TAG TTT TTT	1x
0-R0-2	**R P W P** CGT CCT TAG TGG CCT	1x
0-R0-3	**P R P P** CCG CGG TAG CCT CCG	1x
0-R0-4	**S A C S** TCT GCT TAG TGT TCT	1x
0-R0-5	**S T Q N** AGT ACT TAG CAG AAT	1x
0-R0-6	**S V F F** TCT GTT TAG TTT TTT	1x
0-R0-7	**K P H A** AAG CCT TAG CAT GCG	1x
0-R0-8	**N Q P T** AAT CAG TAG CCG ACT	1x

Table S 2: Sequences of clones identified in early selection rounds with ncAA 1.

Hits of Selection with **1** - Round 3		
Name	Sequence[1]	Occurence
1-R3-1	**H T T P** CAT ACT TAG ACG CCG	1x
1-R3-2	**H H T T** CAT CAT TAG ACT ACT	1x
1-R3-3	**H Q K P** CAT CAG TAG AAG CCT	1x
1-R3-4	**H T N D** CAT ACT TAG AAT GAT	1x
1-R3-5	**K N K K** AAG AAT TAG AAG AAG	1x

Table S 3: Sequences of clones identified in early selection rounds with ncAA 6.

Hits of Selection with **8** - Round 6		
Name	Sequence	Occurence
8-R6-1	**Q M N N** CAG ATG TAG AAT AAT	1x
8-R6-2	**H H K N** CAT CAT TAG AAG AAT	1x
8-R6-3	**Q K N K** CAG AAG TAG AAT AAG	1x
8-R6-4	**Q K K N** CAG AAG TAG AAG AAT	1x
8-R6-5	**Q H N E** CAG CAT TAG AAT GAG	1x
8-R6-6	**Q Q N E** CAG CAG TAG AAT GAG	1x
8-R6-7	**Q K N N** CAG AAG TAG AAT AAT	1x
8-R6-8	**Q K K Q** CAG AAG TAG AAG CAG	1x

8-R6-9	**Q K E N** CAG AAG TAG GAG AAT	1x

Table S 4. Sequences of clones identified in early selection rounds with the tyrosine codon TAT.

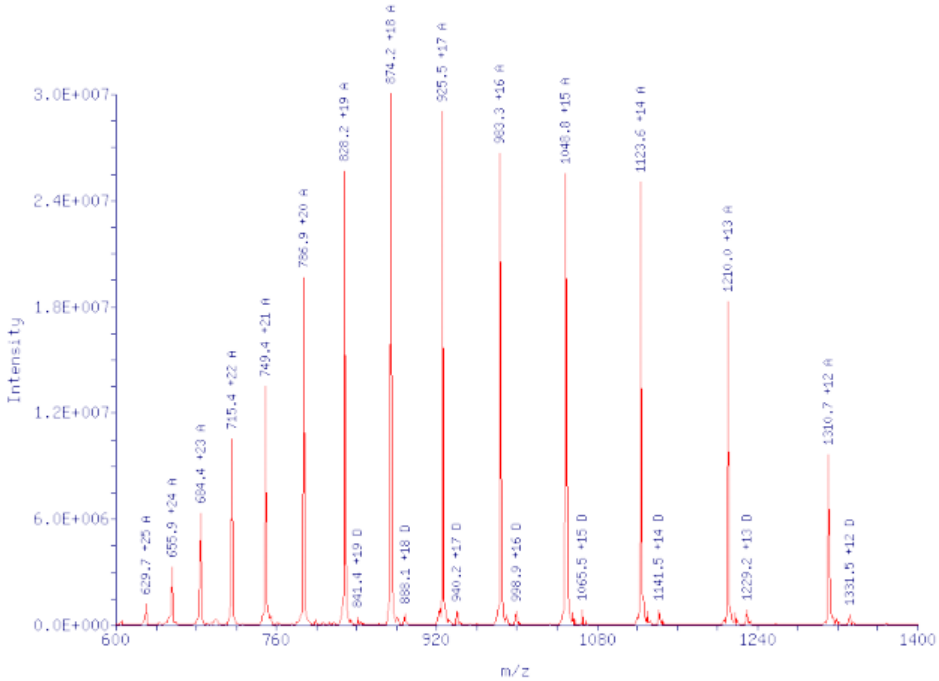
Hits of Selection of pBAD_tag(TAT)_CATwt libraries - Round 9		
Name	Sequence	Occurence
Y-R9-1	**L K Y R N** TTG AAG TAT AGG AAT	1x
Y-R9-2	**D L Y S S** GAT TTG TAT AGT AGT	1x
Y-R9-3	**L N Y V T** CTG AAT TAT GTG ACT	1x
Y-R9-4	**V D Y G W** GTG GAT TAT GGT TGG	1x

Hits of Selection of pBAD_tag(TAT)_CATwt libraries - Round 12		
Name	Sequence	Occurence
Y-R12-1	**N M Y T L** AAT ATG TAT ACT CTG	1x
Y-R12-2	**H R Y A Y** CAT CGT TAT GCG TAT	1x
Y-R12-3	**C F Y N P** TGT TTT TAT AAT CCG	1x
Y-R12-4	**S N Y S C** TCG AAT TAT TCT TGT	1x
Y-R12-5	**S L Y T W** TCT CTT TAT ACT TGG	1x
Y-R12-6	**D L Y S C** GAT CTT TAT AGT TGT	1x
Y-R12-7	**E Y Y F A** GAG TAT TAT TTT GCG	1x
Y-R12-8	**S H Y K I** TCT CAT TAT AAG ATT	1x
Y-R12-9	**P I Y S D** CCT ATT TAT AGT GAT	1x

Table S 5: Sequences of clones identified in early selection rounds with the lysine codon AAA.

Hits of Selection of pBAD_tag(AAA)_CATwt libraries - Round 9		
Name	Sequence	Occurence
K-R9-1	**Q V K F P** CAG GTT AAA TTT CCG	1x
K-R9-2	**L E K Y F** TTG GAG AAA TAT TTG	1x
K-R9-3	**F N K S E** TTT AAT AAA AGT GAG	1x
K-R9-4	**F D K S W** TTT GAT AAA TCT TGG	1x
K-R9-5	**H A K Y G** CAT GCT AAA TAT GGT	1x

Figure S 1: Spectra of 1-R5-2(TAT)_TRX-6His expression (from Figure 5A). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum.



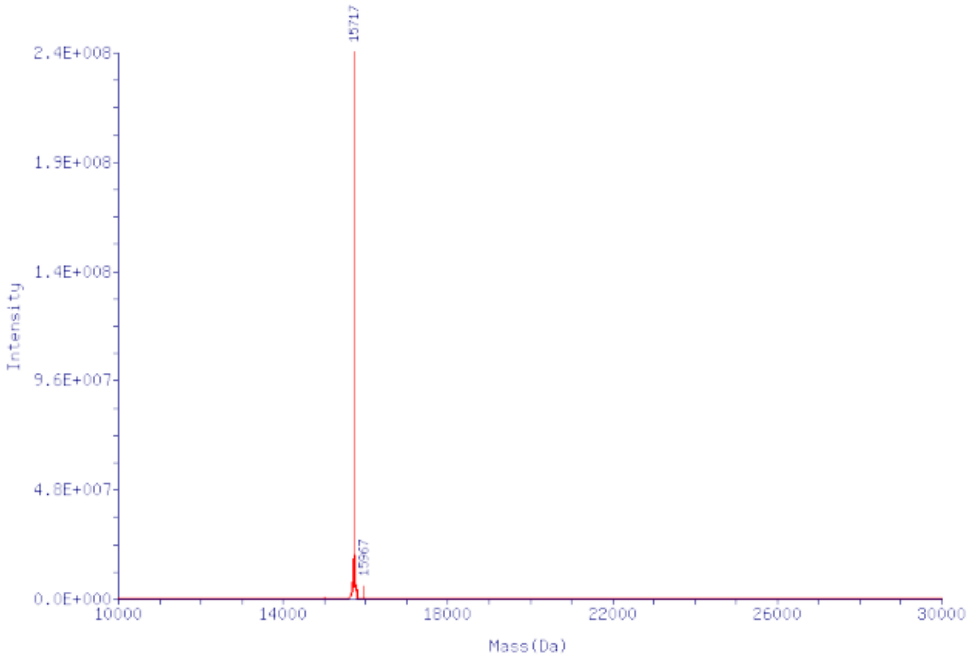
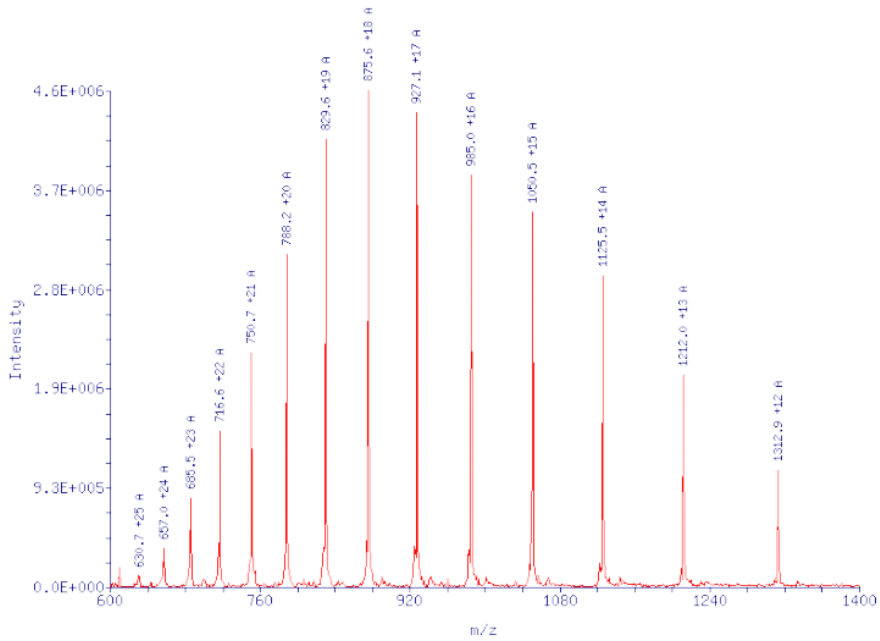


Figure S 2: Spectra of 1-R5-2_TRX-6His expression in presence of ncAA 1 (from Figure 5A). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum.



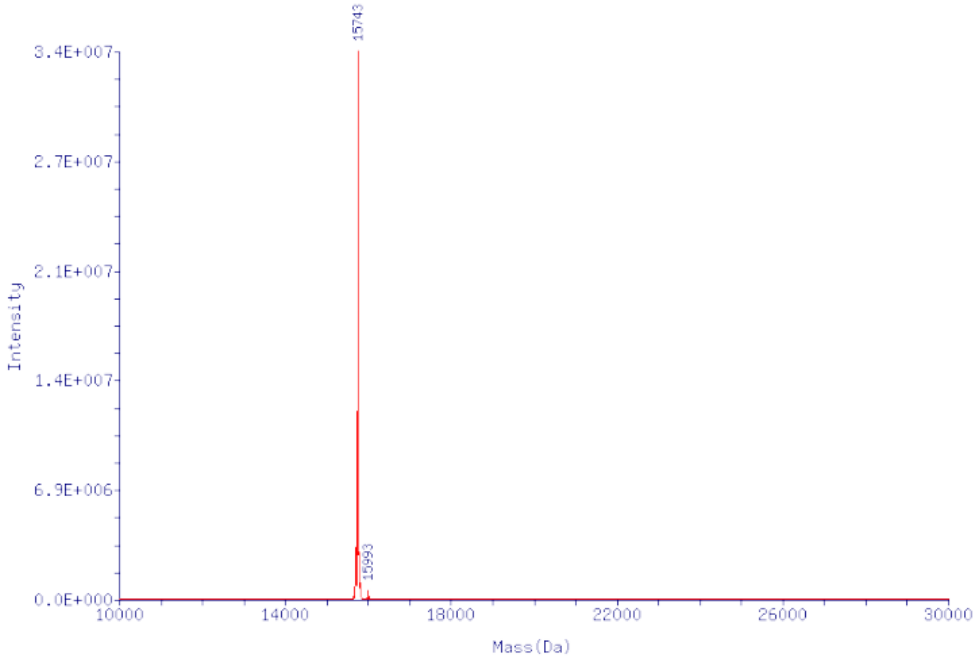
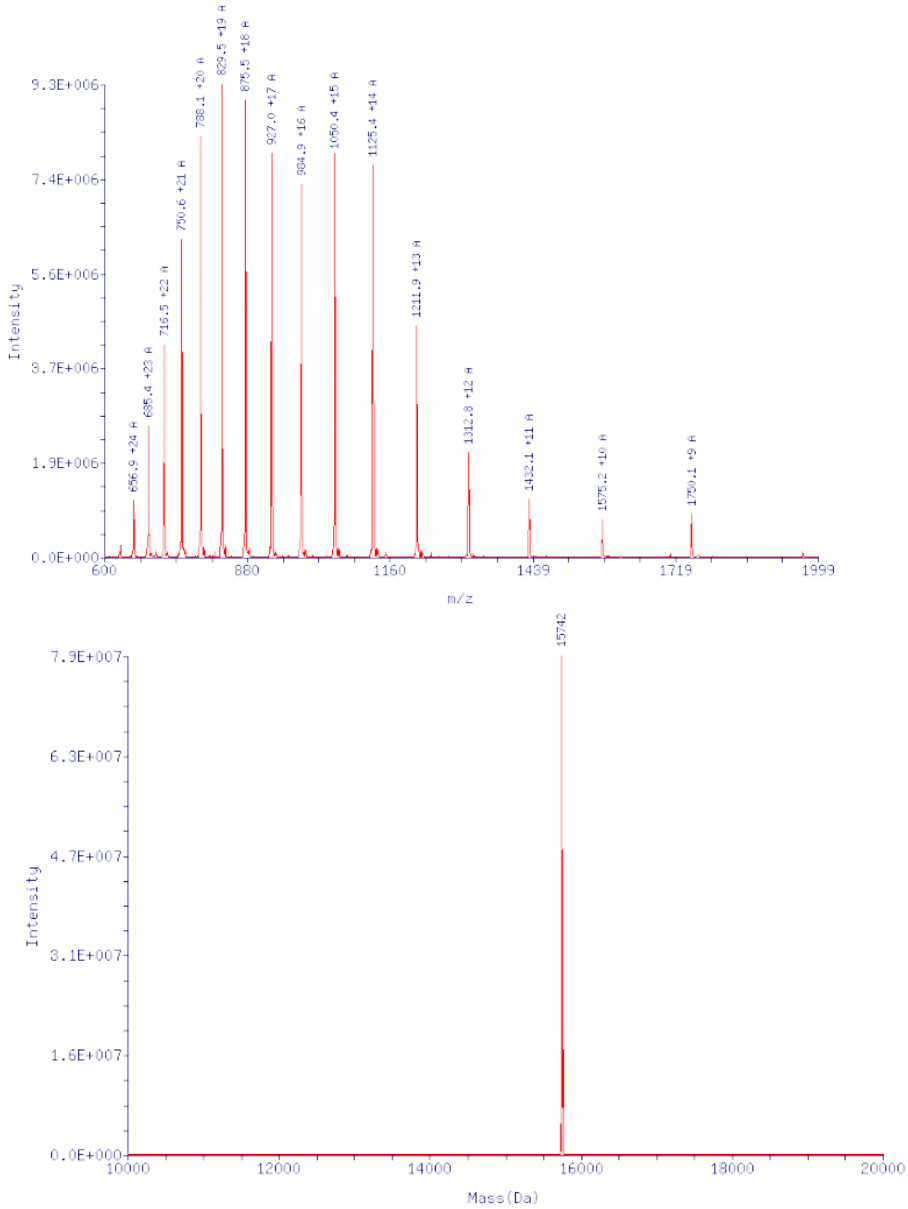


Figure S 3: Spectra of 8-R11-1(AAA)_TRX-6His expression (Figure 5B). Top: Non-deconvoluted spectrum, Bottom: Zoom of deconvoluted spectrum.



Protein Sequences of TRX constructs

>FLAG-TRXwt
MGDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAP
KYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGTDDDDKHHHHH
H*

>TRX_**1**-R5-2(TAT)
MG**NNYTK**GGASGDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKL
NIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGTD
DDDKHHHHHH*

>TRX_**6**-R11-1(AAA)
MG**HHKKN**GGASGDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAK
LNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGT
DDDDKHHHHHH*

Fused sequences are underlined with amino acids encoded by random region and central codon in bold.

ESI-MS mass analyses

FLAG-TRXwt, Mass found[Da]: 14767
TRX_**1**-R5-2(TAT), Mass found[Da]: 15717
TRX_**6**-R11-1(AAA), Mass found[Da]: 15742

Mass difference TRX_1-R5-2(TAT) - FLAG-TRXwt [Da] calc: 949; found: 950.

Mass difference TRX_6-R11-1(AAA) - FLAG-TRXwt [Da] calc:975; found: 976