PROFICS: A bacterial selection system for directed evolution of proteases

Christina Kröß1,2,§, Petra Engele1,2,§, Bernhard Sprenger1,2, Andreas Fischer1,3, Nico Lingg1,3, Magdalena Baier2, Christoph Öhlknecht1,2, Bettina Lier1,4, Chris Oostenbrink1,4, Monika Cserjan-Puschmann1,3, Gerald Striedner1,3, Alois Jungbauer1,3, Rainer Schneider1,2

List of material included

Additional experimental procedures for detection of protein expression in bacterial cells with Npro, cp caspase-2 and Mpro

- Figure S1 Expression control of cells containing cp ATCase and Npro fusion proteins
- Figure S2 Expression control of cells containing cp ATCase fusion variants and cp caspase-2
- Figure S3 Expression control of cells containing cp ATCase fusion variants and Mpro

Table S1 Selected cp caspase-2 mutants

Table S2 FRET assay results for cp caspase-2 S9 variant compared to cp caspase-2 (9) and in silico variant cp caspase-2 D323T H226A (25). Data provided in separate .xlsx format

Supporting materials and methods for P1’ ATCase selections

- Table S3 Activity of cp ATCase with different N-terminal amino acids
- Table S4 Sequences of used oligonucleotides for PCR
- Nucleotide and amino acid sequences of used constructs
- Table S5 Buffer and media compositions

Expression controls and cleavage of proteases in PROFICS selection system

Western blot (Npro-cp ATCase)

The proteins were separated by SDS-PAGE (NuPAGE 4-12 % Bis-Tris gel, Invitrogen, Thermo Fisher Scientific) and blotted onto a methanol activated PVDF membrane applying 30 V constant for 1 h. Blocking was executed in TBS-T (20 mM Tris Base, 150 mM NaCl, 0.05 % Tween, pH 7.6) with 5 % milk powder for one hour. Incubation with primary antibody (1:10,000 dilution of α-Npro antibody, produced in our laboratory, in blocking solution) was at 4 °C overnight. After washing with TBS-T, the membrane was incubated with secondary antibody (α-Goat IgG, Santa Cruz Biotechnology, Inc.; Dallas, TX, USA), diluted 1:4,000 in blocking solution, for 1 h. For detection SuperSignal® West Pico Luminol/Enhancer Solution and SuperSignal® West Pico Stable Peroxide Solution (Thermo Fisher Scientific) and the digital imaging system FUSION FX7 from Peqlab (VWR International; Radnor, PA, USA) were used.
Figure S1: SDS-PAGE and Western blot with antibody against N\textsuperscript{pro} of \textit{E. coli} pyr- cells containing a fusion of active N\textsuperscript{pro} and cp ATCase (Npro-ATCase) or a fusion of inactive N\textsuperscript{pro} and cp ATCase (iNpro-ATCase). Samples were taken before and after induction of expression. Fusion proteins and the ATCase regulatory chain (R-chain) are expressed in both cell types. Only in the cells with active N\textsuperscript{pro} the fusion is cleaved and the ATCase catalytic chain (C-chain) and N\textsuperscript{pro} were detected separately.

Figure S2: SDS-PAGE of \textit{E. coli} pyr- cells containing cp ATCase with different recognition sites (WEHD and VDVAD) in their fusion tags; cp ATCase without fusion and cp caspase-2 without cp ATCase were used as controls. Samples were taken before and after induction of expression. The regulatory chain is expressed in all ATCase containing cells. In addition, the cleaved cp c-chain is visible in the cells containing cp caspase-2 together with 6H-VDVAD-cp ATCase. The uncleaved (tagged) cp c-chain is visible in the cells without cp caspase-2 and in the cells with the WEHD recognition sequence.
Figure S3: SDS-PAGE of *E. coli* pyr- cells containing cp ATCase with a fusion tag comprising the AVLQ-S recognition site. Cells expressing cp ATCase without fusion and only Mpro are shown as controls. Samples were taken before and after induction of expression. The regulatory chain is expressed in all four samples containing a cp ATCase variant. The cleaved and uncleaved catalytic chain can be seen in the cells containing 6H-AVLQ-S-cp ATCase with or without Mpro respectively.

**Selected variants from caspase mutation**

On selective plates colonies with varying sizes were obtained. Colonies larger in size, which were suspected to contain a caspase variant conferring an advantage over others because of the cells’ faster growth were analyzed in more detail.

Table S1: Mutations of variants selected from cp caspase-2 library. The library was transformed in *E. coli* pyr- cells containing tagged cp ATCase. Total volume of the transformation was 500 µl. Aliquots of 250 and 25 µl were pipetted into flasks with selective media (approach Fa and Fb), aliquots of 200 (Pa), and 20 µl (Pb) were plated on selective agar plates. 1 µl was plated on non-selective TY-agar and colonies sequenced to estimate the size of the library and the number of mutations.

| No. | Mutations                          | Silent mutations |
|-----|-----------------------------------|------------------|
| **Colonies from non-selective conditions** |                   |                  |
| I   | none                              | 0                |
| II  | C44R T145A V238F                   | 0                |
| III | K89 F100S H116Q L194Q F217L        | 0                |
| **Agar plate assay Pa** |                   |                  |
| 1   | L45Q K136R                         | 3                |
| 2   | E105V                              | 1                |
| 3   | none                               | 0                |
| 4   | none                               | 0                |
| 5   | T126S                              | 0                |
| 6   | R35S Q144R                         | 2                |
| **Agar plate assay Pb** |                   |                  |
| 13  | K26R                               | 3                |
| 14  | none                               | 2                |
| 15  | none                               | 1                |
| 16  | none                               | 2                |
Table S2: Michaelis-Menten kinetics measured using FRET substrate VDVADXA at 1 µM enzyme concentration. (n.d. not determined). cp caspase-2 S9 variant compared to cp caspase-2 (data from (9)) and in silico variant cp caspase-2 D323T H226A (data from (25))

Data provided in separate .xlsx file

### FRET Results

### Additional experiments for P1’ selections

**Construction of cp ATCase plasmids**

For the cp MA-ATCase construct an alanine was inserted at the N-terminus of cp pyrB by site directed mutagenesis using primers MA-pyrB_forw and MA-pyrB_rev. The original Thr$^{228}$ was mutated to all 19 canonical amino acids with site directed mutagenesis using degenerate primers M-X-ATCase_forw and M-X-ATCase_rev, resulting in cp M-(T→X)-ATCase constructs.

For the deletion variant the Met$^{227}$ in the cp pyrB was removed with site directed mutagenesis. A gene library of Npro-(M→X)-cp ATCase constructs was cloned with degenerate primers Npro-(M-X)-ATCase_forw and Npro-(M-X)-ATCase_rev. Methionine deletion variants (6H-VDVAD-ΔM-X-cp ATCase) were created with site directed mutagenesis.

**Experiments to establish P1’ toolbox**

For the P1’ toolbox experiments untagged cp ATCase constructs (cp MA-ATCase, cp M-(T→X)-ATCase), were transformed into *E. coli* pyr- cells without protease. Cell growth under selective conditions confirmed ATCase activity.

First an alanine was added to the N-terminus of cp ATCase (cp MA-ATCase) to see if the desired P1’ amino acid could simply be inserted upstream of the original terminus. The activity of this construct was tested with the agar plate assay, no colony formation was observed. Next, we generated and selected a gene library in which we mutated the second residue Thr$^{228}$ to all canonical amino acids (cp M-(T→X)-ATCase). Active cp ATCase variants with nine different amino acids in the mutated position were found (A, C, G, P, Q, R, S, T, V). Nearly all these residues are readily accepted by MAP (54).

To express cp ATCase variants without starting methionine we generated a gene library of Npro-cp ATCase. The fusion enabled us to exchange the starting residue of the cp ATCase (Npro-(M→X)-cp ATCase). This approach was unsuccessful, only variants starting with methionine survived the selection.

These experiments indicated that the activity of the cp ATCase depends on removal of the starting methionine rather than on the variation of the Thr$^{228}$ residue (Table S3). Apparently, the spatial
restrictions of the structure and the charges introduced with the new termini do not tolerate much variation. We concluded that removal of the methionine could increase the flexibility of the protein chain. To test the activity of different P1’ cp ATCase variants without their native N-terminal methionine the fusion protein with caspase cleavage tag was used, as it has another start codon upstream of the tag. The disadvantage is, that the process is not only influenced by proper cp ATCase folding but also by the P1’ tolerance of the caspase. The methionine deletion variants were generated with the ten P1’ residues not found in the cp M-(T→X)-ATCase selection. Co-expression with cp caspase-2 under selective conditions confirmed the activity of 5 VDVAD-ΔM-X-cp ATCase variants (X = D, E, I, L, and N). No cell growth was observed on selective plates when the P1’ was F, H, K, W, or Y after 48–72 hours of incubation.

To verify the results, a methionine deletion variant of Npro-cp ATCase as well as the VDVAD-ΔM-cp ATCase together with cp caspase-2 were tested in the shaking flask assay. Cell growth was drastically increased in comparison to cells expressing the respective proteins without the deletion (data not shown). This confirms that the initial methionine is not necessary for enzymatic function but possibly even detrimental and that the subsequent amino acid of cp ATCase (Thr<sup>228</sup>) can be mutated to nearly all canonical amino acids (Table S3), which allows the generation of a toolbox for selection of proteases with a specific P1’ tolerance.

Table S3: Activity of cp ATCase, X is any amino acid despite methionine.

| N-terminal amino acids of cp c-chain | Cell growth                                      |
|-------------------------------------|--------------------------------------------------|
| MTRVQKERL…                         | Yes (original construct)                         |
| N<sup>pro</sup>-MTRVQKERL…         | Yes                                              |
| N<sup>pro</sup>-XTRVQKERL…         | No, except when X=M                               |
| N<sup>pro</sup>-TRVQKERL…          | Yes, improved compared to original construct     |
| MA-MTRVQKERL…                      | No                                               |
| MXRVQKERL…                         | Yes, when X accepted by MAP                       |
| MH<sub>6</sub>GSGVDVAD-MTRVQKERL…   | Yes, when co-expressed with cp caspase-2         |
| MH<sub>6</sub>GSGVDVAD-TRVQKERL…    | Yes, when co-expressed with cp caspase-2         |
|                                     | improved compared to MH<sub>6</sub>GSGVDVAD-MTRVQKERL… |
| MH<sub>6</sub>GSGVDVAD-XRVQKERL…   | Yes, when co-expressed with cp caspase-2         |
|                                     | except X is F, H, K, W, Y                        |

Sequences

Primer Sequences

Table S4: Oligonucleotides used for cloning experiments

| Primer name                  | Sequence 5’→3’                                      |
|------------------------------|-----------------------------------------------------|
| pyrI_genome_forw             | TATACCATGGGCACACACGATAATAAATTGCAG                  |
| pyrI_genome_rev              | TATAGCGGCCGCTTAATTGGCCAGCACCAC                    |
| MA-pyrB_forw                 | atggccATGACCCCGCTGCAAAAA                           |
| MA-pyrB_rev                  | ATGTATATCTCCTTCTTACTTACTTAATATATAC                |
| M-X-ATCase_forw              | TATACATATGmnnCGCGTGAACAAAGGAG                   |
| M-X-ATCase_rev               | TCTCCTTCTTATACTTAATATAC                           |
| Npro-(M-X)-ATCase_forw       | CACTAGTTGCnhnACCCCGCCTGCAAAAA                    |
| Npro-(M-X)-ATCase_rev        | ACCCAGAGTGACGTTAAGTAG                             |
| AVLQ5S_forw                  | gcagageCGCGTGAACAAAGAGCAGGT                      |
AVLQS_rev
ep_caspase_forw (5' phosphorylated)
ep_caspase_rev
vector_forw (5' phosphorylated)
vector_rev

Construct sequences

**Active Npro-cp ATCase**

**Nucleotide sequence: pyrI**

```
ATGGGCACAC ACGATAATAA ATTGCAGGTT GAAGCTATTA AACGCGGCAC GGTAATTGAC
CATATCCCG CCGATACCG TTTTAAGCTG TTGAGTCTGT TCAAGCTGAC CGAAACGGAT
CCGCGATATC CCAACTGATCT CAGATGCGCA TCGAGACGAC ATCTCGGATT AAATCGGAA
CAGCCACGG TTAAACGTAT GCAGAACTAT GCTAGCGGAA CAGAGGATTC AAATCGGAA
TACTGTTTAT CACAGTTTGC CCGTGGGAAT CGGCGCAATT GAAATGATGC
```

**Protein sequence: PyrI**

```
MGTHDNKLQV EAIKRGTVID HIPAQIGFKL LSLFKLTETD QRITIGLNLP
SGEMGRKDLI KIENTFLSED QV
```

**Nucleotide sequence: Npro-pyrB**

```
ATGGAACTCA ATCATTTCGA ACTGCTCTAC AAAACTAGCA AGCAAAAACC TGTTGGCGTT
GAAGAGCCGG TCTACGATAC TGCAGGTCGT CCTCTTTTTG GGAATCCGTC CGAAGTGCAC
CCCCAGTCAA CCCTCAAGCT TCCCCATGAC CGCGGACGCG GTGACATTCG TACAACGCTG
CGCGATCTGC CTCGTAAAGG CGATTGTCGC TCTGGAAACC ACC
```

**Protein Sequence: Npro-pyrB**

```
MELNHFELLY KTSKQKPVGV EEPVYDTAGR PLFGNPSEVH PQSTLKLPHD
```

---

**S-6**
Inactive Npro-cp ATCase
Protein sequence: iNpro-pyrB

MELNHFELLY KTSKQKPVGV EEPVYDATGR PLFGNPSEVH PQSTLKLPHD 50
RGRGDIRTTL RDLPRFKGDCR SGNHLGFVSQ IYIKPGPVYY QDYGTPFYVR 100
APLEPFEDEAQ FCEVTKRGMR TVGSDGKLYH IYVCVDGWQL LKLAKRGTPT 150
TLPWIRFTNM CPLAVTSCTMT RVQRLKDRPS EYANVKQFVP LRSOLDSN 200
ANMKVHLPLP RVDEIATDVQ KTPHAWYFQQ ACGNGFARQA LLLALVLNRA 250
PPLYQKHIISI NDLSRDDLNVL VLATAAKLKA NPQPELLKHK VIASCFFEAS 300
TRTRLSFETS MHRLGASVVG FSDSANTS LG KKGGETLADTI SVISTYVDAI 350
VMRHPQEGAA RLATEFSGNV PVNLAGDSGN QHPTQTLDDL FTTIQETQGRL 400
DNLHVAMVGD LKYGRTVHSL TQALAKFDGN RFYFIAPDAL AMPQYILD 450

cp ATCase c227 PyrI D73E
MCS I Protein Sequence:
MGTHDNKLQV EA1IKRGTIVD HIPAOIQGKLG LSLFOKLTED QRTITGLNL 50
SCEMGKRDLI QNF4LFSLED QVEQLALYAP QAVTNRIDNY EVVKGSKRESL 100
PERIDNVLCV PNSNCISHAE PVSSSFSAVRK RNDIALKCK YCEKEFSNVLV 150

MCS II Protein Sequence:
MTTVQKLERLD FSEYANVKAQ FVLRASDLHN AKANMKVHLHP LPRVDEIATD 50
VDKTPHAWYF QQANGIFISAR QALLALVLNR ANPLYQKHIISI SINDLSRDDL 100
NLVLATAKX KANQEPPELK HKVIASCFFEE ASTRTRLSES TSMHRLGASV 150
VGFSDSANTKS LGKGETLD LIWISTYVDD AIVMRHPQEG AARLATEFSG 200
NPVLNPNGGD SNQHTQPLTL DLFTIQETQGL DRNLNVAVGD GDLKYGRTVH 250
SLTQALAFKD GNRFYFIAPDL ALAMPQYILD MLDEKJAWS LHSSIEEVM 300
EVDILY 306

cp caspase-2
Nucleotide Sequence:
ATGCACCATC ATCACCATCA TGGCAAAAAT CATGCAGGTA GTCCGGGTTG TGAAGAAAGC 50
GCAGCATGTA AAGAAAACGT GCGAAAATGT CGCTCGGGCA CCGTACGCA TATGATTGTG 100
GCTTGATCCTG AAGCAAGGTC AAGAGGAAAT CGACTGATTC CACCTGGCGG GCTGTTTTAT 150
ATGTAAGCTG TGGCAAGGTC AAGCAAGGTC AAGAGGAAAT CGACTGATTC CACCTGGCGG 200
GCTTGATCCTG AAGCAAGGTC AAGAGGAAAT CGACTGATTC CACCTGGCGG GCTGTTTTAT 250
ATGTAAGCTG TGGCAAGGTC AAGCAAGGTC AAGAGGAAAT CGACTGATTC CACCTGGCGG 300

Protein sequence:
MHMHMMHMGKNN HAGSPGCGRSA AAGKEKLKPM RLPLTRSDMIE VACCKCAGTA GAAGAAGACT 50
MRNTKGRSWNY IEALAQFVSE RACDMHVADM LVKVNALIKD REGYAPGTPTP 100
HRCKEMSEYX STLCRLHLFL FGHPTPTSGG VCLQVKFCFP EYQTHDFQLA 150
YRLSQPRGRL ALVLSNVHFT GEKELEFRSG GDVHDSTLVT LFKLGLYDVH 200
VLCQDTAQEM QERQLQNFQAQL PAVRVTDSCL VALSHGVEV GAYVGDGKLL 250
QLQEVFLFGLD NANCPSLONQ PKMFFIQACR GDEETGVDQ QD 292
**S9 (cp caspase-2 E105V)**

**Nucleotide sequence:**

```
ATGCACCATC ATCACCATCA TGGCAAAAAT CATGCAGGTA GTCCGGGTTG TGAAGAAAGC
```

60

```
GCAGCAGGTA AAGAAAAACT GCCGAAAATG CGTCTGCCGA CCCGTAGCGA TATGATTTGT
```

120

```
GGTTATGCAT GTCTGAAAGG CACCGCAGCA ATGCGTAATA CCAAACGTGG TAGCTGGTAT
```

180

```
ATTGAAGCAC TGGCACAGGT TTTTAGCGAA CGTGCATGTG ATATGCATGT TGCAGATATG
```

240

```
CTGGTTAAAG TGAACGCCCT GATTAAAGAT CGTGAAGGTT ATGCACCGGG TACAGAATTT
```

300

```
CATCGTTGTA AAGAATGAG CGAGTATTGT AGCACCCTGT GTCGTCATCT ATACCTGTTT
```

360

```
SARS-CoV-2 Mpro
```

**Nucleotide sequence:**

```
ATGAGCGGTTTTCGTAAAATGGCATTTCCGAGCGGTAAAGTTGAAGGTTGTATGGTTCAGGTTACCTGTGGCACCACCACACTGAATGGTCTGTGGCTGGATGATGTTG
```

192

```
TTTATTGTCCGCGTCATGTTATTTGTA
```

252

```
CCAGCGAAGATATGCTGAACCCGAATTATGAAGATCTGCTGATTCGCAAAAGCAACCATAATTTTCTG
```

312

```
GTTCAGGCAGTTAATGTTCAGCTGCGTTATTGGTCATAGCATGCAGAATTGTGTGCTGAAACTGAA
```

372

```
AGTTGA7ACCGCGAATCCGAAAAAACCGCGAAATATAGTTTGGTGTCTGATCCCGACCGGTGTTCATGCAGGCACCGATCTGGAAGGTAACT
```

432

```
TTTATGGTCCGTTTGTTGATCGTCAGACCGCACAGGCAGCAGGTACAGATACCACCATTACCGTTAAT
```

492

```
GTTCTGGCC
```

552

**Protein sequence:**

```
MSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFL
```

100

```
VQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFT
```

150

```
IKGSFLNGSCGSGFNIDCVSPCYMHMMELPVGHAVIDLEGNYGFVFDRQTQAAGTDITTVN
```

200

```
VLAWLYAAVINGDRWFPLNFRFTLTLNDFNLVMKYNYEPLTQHDVILGPGLSAQTGIAVLDMCASLKEL
```

250

```
LQNGMGRITLGSALEDETPFDVVQRCGSVTFQ
```

300
Table S5 Composition of used buffers and media

| Buffer/Media                        | Composition                                                                 |
|-------------------------------------|-----------------------------------------------------------------------------|
| **PBS** (Phosphate buffered saline, pH 7.4) | 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄                     |
| **TY** (Tryptone yeast) medium      | 1% (w/v) peptone, 0.7% (w/v) yeast extract, 0.25% (w/v) NaCl              |
| **SOC** (Super optimal broth with catabolite repression) | 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 20 mM glucose |
| **Supplemented M9 minimal medium**  | 50 mM Na₂HPO₄, 20 mM KH₂PO₄, 10 mM NaCl, 1 mM MgSO₄, 0.1 mM CaCl₂, 0.4% (w/v) glucose, 0.5% (w/v) casamino acids, 20 mM NH₄Cl, 10 µg/ml FeSO₄, vitamins (0.001 mg/ml of each biotin, thiamine, riboflavin, pyridoxine, niacinamide), 0.025–1 mM IPTG |
| **Used antibiotics**                | Kanamycin (50 µg/ml)                                                       |
|                                     | Ampicillin (100 µg/ml)                                                      |
|                                     | Chloramphenicol (34 µg/ml)                                                  |