

Discovery and Characterization of a New Cell-Penetrating Protein

Supporting Information

Rudo L. Simeon^a, Ana Maria Chamoun^a, Thomas McMillin^a, Zhilei Chen^{a, b*}

^a Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station,
TX 77843

^b Department of Microbial and Molecular Pathogenesis, Texas A&M Health Science Center,
College Station, TX 77843

Abbreviations: CPP, cell-penetrating polypeptide

Keywords: Intracellular, biopharmaceuticals, biotherapeutics, biomolecules

Running Title: New cell-penetrating protein for effective protein/nucleic acids delivery

*Corresponding author. Mailing address:
3122 TAMU, Artie McFerrin Department of Chemical Engineering, Texas A&M University,
College Station, TX 77843
Phone: 979-862-1610. Fax: 979-845-6446. E-mail: zchen4@tamu.edu

Supplementary materials and methods.....Page 3

Figures

Figure S1. Comparison of B1 and EGFP nucleotide and amino acid sequences.....Page 7

Figure S2. Purification of recombinant B1.....Page 8

Figure S3. GFP-L-B1 is not toxic to mammalian cells.....Page 8

Figure S4. Cell membrane permeability of B1-treated cellsPage 9

Figure S5. Secondary structure prediction of B1 amino acid sequencePage 10

Sequences of Proteins Used in this Study.....Page 11

References.....Page 14

Supplementary materials and methods

Chemicals, bacterial strains and cell lines

BSR.LNL.tdTomato reporter cells were a gift from D. Liu (Harvard University). 293T cells were purchased from Invitrogen. TZM-bl cells were obtained from J. Kappes, X. Wu and Tranzyme Inc. through the NIH AIDS Reagent Program, Division of AIDS, NIAID. Jurkat cells were obtained from P. Lindahl (Texas A&M University). Huh-7.5 cells were obtained from C. Rice (Rockefeller University).

All adherent cell lines were cultured in complete growth medium (Dulbecco's Modified Eagle Medium (DMEM) containing 4500 mg/L glucose, 4.0 mM L-Glutamine, and 110 mg/L sodium pyruvate (Thermo Scientific HyClone) supplemented with 10 % fetal bovine serum (Atlanta Biologicals) and 1X non-essential amino acids (Thermo Scientific HyClone)). Jurkat cells were cultured in Roswell Park Memorial Institute Medium (RPMI) 1640 (Lonza) supplemented with 10% fetal bovine serum.

Escherichia coli strains DH5 α (Invitrogen) and BL21(DE3) (New England Biolabs) were used for recombinant DNA cloning and recombinant protein expression, respectively.

Amiloride and dynasore were purchased from Enzo Life Sciences. Chlorpromazine, mannan and cytochalasin B were obtained from Sigma-Aldrich. Nystatin was purchased from Amresco. CellTiter-Glo Luminescent Cell Viability Assay Kit and Luciferase Assay System were purchased from Promega. Bioluminescence Assay Kit was purchased from New England Biolabs. Dulbecco's Phosphate-Buffered Saline (DPBS) was purchased from Thermo Scientific HyClone. OptiMEM and Lipofectamine 2000 were purchased from Invitrogen. 0.4% Trypan Blue solution was purchased from Lonza. Luria-Bertani (LB) broth, Mirus TransIT LT1 transfection reagent and ampicillin were obtained from Thermo Fisher Scientific. Isopropyl- β -D-thiogalactopyranoside (IPTG), 0.25% trypsin-EDTA and imidazole were from VWR International. TransMessenger transfection reagent was purchased from Qiagen.

Protein purification

BL21 (DE3) cells were transformed with the expression plasmid and plated on a Luria–Bertani (LB) agar plate containing 100 µg/mL ampicillin. The next day, a single colony was picked and cultured at 37°C with shaking at 250 rpm. At OD₆₀₀ ~0.6, the cells were cooled to 18°C and isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM) was added to the culture to induce protein expression. After overnight expression (~18 hours) at 18°C, cells were harvested by centrifugation at 6000 x g and 4°C for 20 minutes and pellets were stored at –80 °C until use.

For recombinant protein purification, cell pellets were resuspended in lysis buffer (500 mM NaCl, 50 mM Tris, pH 8, 10 mL per 1 g wet cell pellet), and disrupted by sonication on ice (Qsonica Misonix 200 (Qsonica), 1 second pulse 6 second pause for 3 minutes 10 amp). Whole cell lysate was clarified by centrifugation at 20,000 x g for 20 minutes at 4°C. Soluble cell lysates were loaded onto 5 mL Ni–NTA agarose beads (Qiagen). For constructs expressing B1 or +36GFP, the loaded resin was washed with lysis buffer supplemented with 150 mM imidazole and the product was eluted in lysis buffer supplemented 500 mM imidazole. For all other constructs the loaded resin was washed with lysis buffer supplemented with 50 mM imidazole and the product was eluted in lysis buffer supplemented 150 mM imidazole. Purified protein samples were dialyzed overnight at 4°C into Buffer A (2 M NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄·2H₂O, 2 mM KH₂PO₄, pH 7.4). The proteins +36GFP and Cre-L-B1 were further purified by cation exchange chromatography using a GE HiTrap SP HP column (GE Biosciences, PA). Dialyzed proteins were diluted 1:5 into PBS to reduce the salt concentration and then loaded onto the SP column. The column was washed with Buffer B (0.5 M NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄·2H₂O, 2 mM KH₂PO₄, pH 7.4) and the target protein was eluted in Buffer A. Purified proteins were concentrated via ultracentrifugation through an Amicon 10 kDa MWCO centrifugal column (EMD Millipore). The final protein concentrations were determined using a Pierce Coomassie Plus Protein Assay Kit (Thermo Fisher Scientific). Typically the purification yield of GFP-L-B1 using this approach is ~15 mg per liter of *E. coli* culture. It is worth noting that although a high concentration of NaCl (0.5M) is needed to maintain B1-

fusion protein soluble at the stock concentration (i.e. when B1 is concentrated at $\sim 100 \mu\text{M}$), such high concentrations of NaCl were not needed during cell transduction. The highest final salt concentration used in these experiments was $\sim 230 \text{ mM}$ (Figure 2, $5 \mu\text{M}$ GFP-L-B1). Minimal cellular toxicity was observed under this condition (Figure S3).

Gel shift of GFP-L-B1-conjugated DNA/RNA

pCMV5–Gluc DNA was linearized by digestion with XbaI. pIRF mRNA was transcribed and capped from the XhoI digested plasmid DNA template using the Ampliscribe T7 High Yield Transcription Kit (Epicentre) with the addition of the monomethylated RNA cap analog (Cell Script) following the manufacturer’s protocol. Linearized pCMV5–Gluc plasmid DNA (14.5 femtomoles or 50 ng) or pIRF mRNA (86 femtomoles or 80 ng) were mixed with GFP–L–B1 at the following protein:DNA/RNA molar ratios: 0:1, 10:1, 50:1, 100:1, 500:1, 1000:1 in EMSA buffer (4% glycerol, 1 mM MgCl_2 , 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, 10 mM Tris–HCl, pH 7.4). The mixtures were incubated at room temperature for 10 minutes to allow protein–DNA/RNA association. Formation of protein/nucleic acids complexes was confirmed by analysis on a 0.8% agarose gels (100 V for 40 minutes for DNA and 120 V for 15 minutes for RNA). The ethidium bromide–stained gel was then visualized under UV exposure.

Interrogation of role of cellular endocytic pathways

293T and TZM–bl cells were seeded in 24–well plates at 1.5×10^5 cells per well. The following day, cells were washed with 500 μl OptiMEM to remove residual serum and incubated in 200 μL OptiMEM containing endocytic inhibitors for 1 hour at $37^\circ\text{C}/5\% \text{ CO}_2$. 20 μM GFP–L–B1 (in 20 μL) or mixtures of GFP–L–B1 (8 μM) and pIRF mRNA (5 ng (5 fmol)) were added to the cells in the continued presence of the inhibitors and incubation at $37^\circ\text{C}/5\% \text{ CO}_2$ for one hour. For analysis of cellular uptake/functional cargo delivery, cells were washed once with cold DPBS, incubated with DPBS containing 0.04% Trypan

Blue at room temperature for 2 minutes to quench the fluorescence of extracellular GFP⁽¹⁾, and analyzed via fluorescence microscopy (Zeiss Axiovert 200M) or flow cytometry (excitation/emission 488/530). At least 1×10^4 cells were analyzed for each sample. Firefly luciferase (Fluc) activity deriving from pIRF mRNA delivery was quantified 6 hours post exposure to GFP-L-B1/mRNA complex using the Luciferase Assay System (Promega).

Interrogation of role of cell-surface glycans

For GFP-L-B1 delivery, 293T and TZM-bl cells were seeded in 24-well plates at 7.5×10^4 and 1.5×10^5 cells/well, respectively. For the assessment of functional mRNA delivery, 293T cells were seeded in 96-well plates at 1.5×10^4 cells per well. The following day, cells were washed with OptiMEM to remove residual serum and incubated with either 2 μ M GFP-L-B1 for one hour or mixtures of 1.6 μ M GFP-L-B1 and 5 ng (5 fmol) pIRF mRNA for 6 hours. For assessment of GFP-L-B1 uptake, cells were washed once with cold DPBS, incubated in DPBS containing 0.04% Trypan Blue at room temperature for 2 minutes, and analyzed via fluorescence microscopy (Zeiss Axiovert 200M) or flow cytometry (excitation/emission 488/530). At least 1×10^4 cells were analyzed for each sample. Functional delivery of pIRF mRNA was determined by measuring intracellular Fluc activity 6 hours post transfection using the Luciferase Assay System.

Cloning

The codon-optimized B1 gene was synthesized by Genscript (Piscataway). All constructs were expressed using the pET15b vector, which contains an N-terminal 6xHistidine tag. mCherry was amplified from pET-36GFP-L-mCherry (a gift from D. Liu, Harvard University). Cre was amplified from pET-36GFP-L-Cre (gift from D. Liu). DNA fragments encoding linkers were synthesized by Integrated DNA Technologies and inserted into the appropriate vectors via overlap extension PCR.

Supplementary Figures

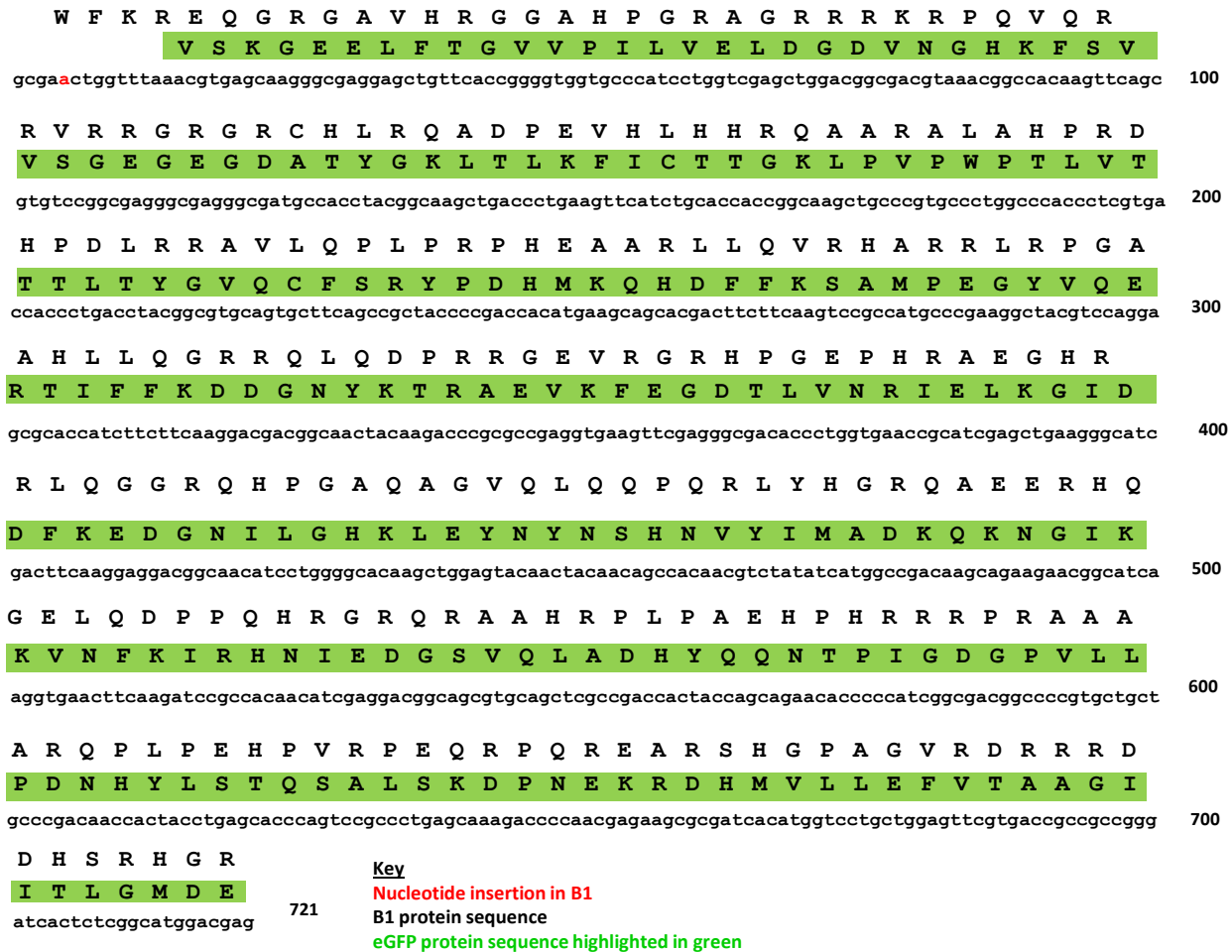


Figure S1. Comparison of B1 and EGFP nucleotide and amino acid sequences. A single base insertion resulted in a frameshift leading to lead to the frame shift that generated B1 is shown in red.

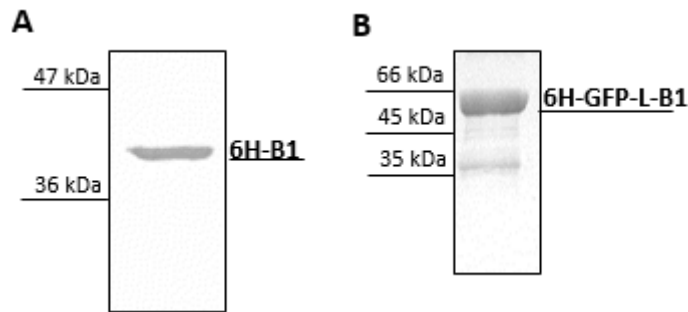


Figure S2. Purification of recombinant B1. **A.** *E. coli* pellet containing recombinant 6H-B1 was suspended in lysis buffer (500 mM NaCl, 50 mM Tris, pH 8) and lysed using a French press (Microfluidics M-110P). The soluble lysate containing 6H-B1 was loaded onto a Ni-NTA agarose column. Impurities were removed by extensive washing with lysis buffer supplemented with 150 mM imidazole. 6H-B1 was eluted from the column with lysis buffer supplemented with 0.5 M imidazole. The purified protein was analyzed via SDS-PAGE (12 % acrylamide). **B.** 6H-GFP-L-B1 was purified in the same way as 6H-B1 except that the *E. coli* cells were lysed using sonication (Qsonica Misonix 200 (Qsonica), 1 second pulse 6 second pause for 3 minutes 10 amp).

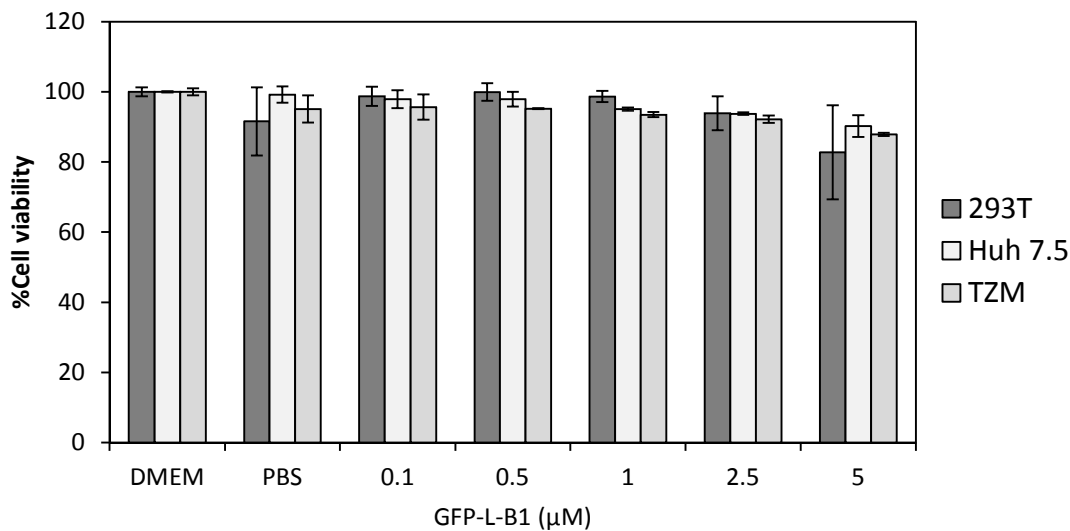


Figure S3. GFP-L-B1 is not toxic to mammalian cells. 293T, TZM and Huh-7.5 cells were seeded in 96-well plates at 2×10^4 , 1.8×10^4 and 2×10^4 cells per well, respectively. The next day, cells were washed with 100 μl OptiMEM before being exposed to 0-5 μM GFP-L-B1 prepared in OptiMEM. 24 hours later, the relative cell viability was quantified using the CellTiter-Glo Assay Kit (Promega) following the manufacturer's protocol. Error bars represent the standard deviation of 2 independent experiments.

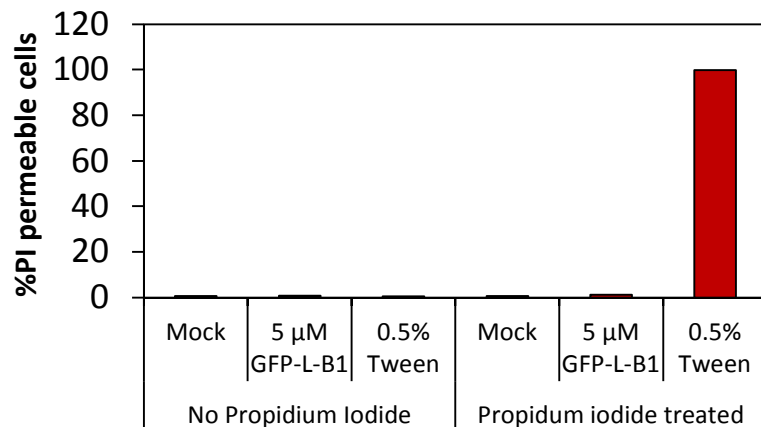


Figure S4. Cell membrane permeability of B1-treated cells. TZM-bl cells were seeded in 24-well plates at 1.5×10^5 cells/well. The following day, cells were washed once with OptiMEM before exposure to 5 μ M GFP-L-B1 or 0.5% Tween for one hour at 37°C/5% CO₂ in the presence of 3 ng/ml propidium iodide. After treatment cells were carefully washed once with cold DPBS followed by three washes with cold heparin solution (diluted to 50 μ g/ml in OptiMEM). After treatment cells were trypsinized and resuspended, and cell permeability was quantified by flow cytometric analysis of propidium iodide fluorescence.



Figure S5. Secondary structure prediction of B1 amino acid sequence. Secondary structure predictions of the B1 protein were made using the GOR4 ⁽²⁾ **(A)** and Chou-Fasman ⁽³⁾ **(B)** algorithms. Several helical motifs were predicted by both algorithms **(C)**.

Sequences of Proteins Used in this Study

B1 (1)

MGSSHHHHHHSSGLVPRGSHMWFKREQRGAHVHRGGAHPGRAGRRRKRQVQRVRRGRGRCHLRQADPEVHLH
HRQAARALAHPRDHPDLRRAVLQPLRPHEAARLLQVRHARRLRPGAHHLLQGRRQLQDPRRGEVRGRHPGEPHRA
EGHRLQGGRQHPGAQAGVQLQQPQRLYHGRQAEERHQGELQDPPQHRGRQRAAHRPLPAEHPHRRRPRAAARQP
LPEHPVRPEQRPQREARSHGPAGVRDRRRDHSRHGRGLN

GFP-L-B1 (2)

MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICT
TGKLPVPWPTLVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIG
DGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELEACGGGGSGGGGSASWFKREQGR
GAVHRGGAHPGRAGRRRKRQVQRVRRGRGRCHLRQADPEVHLHRRQAARALAHPRDHPDLRRAVLQPLP
RPHEAARLLQVRHARRLRPGAHHLLQGRRQLQDPRRGEVRGRHPGEPHRAEGHRLQGGRQHPGAQAGVQL
QQPQRLYHGRQAEERHQGELQDPPQHRGRQRAAHRPLPAEHPHRRRPRAAARQPLPEHPVRPEQRPQREA
RSHGPAGVRDRRRDHSRHGRGLN

GFP-L-B1-F (3)

MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICT
TGKLPVPWPTLVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIG
DGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELEACGGGGSGGGGSASWFKREQGR
GAVHRGGAHPGRAGRRRKRQVQRVRRGRGRCHLRQADPEVHLHRRQAARALAHPRDHPDLRRAVLQPLP
RPLEDPAANKARKEAELAAATAEQ

GFP-L-B1-M (4)

MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICT
TGKLPVPWPTLVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIG
DGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELEACGGGGSGGGGSASQPLRPHE
AARLLQVRHARRLRPGAHHLLQGRRQLQDPRRGEVRGRHPGEPHRAEGHRLQGGRQHPGAQAGVQLQQPQ
RLYHGRQAEERHQGELQDPP

GFP-L-B1-R (5)

MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICT

TGKLPVPWPTLVTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGDIFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIG
DGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELEACGGGGSGGGGSASELQDPPQH
RGRQRAAHRPLPAEHPHRRRPRAAARQPLPEHPVRPEQRPQREARSHGPAGVRDRRRDHSRHGRGLN

mCherry-L-B1 (6)

MGSSHHHHHSSGLVPRGSHMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAK
LKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD
GEFIYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKK
PVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKTGGSGSGSGSGSGSGSGSGG
SSASWFKREQGRGAVHRGGAHPGRAGRRRKRQVQRVRRGRGRCHLRQADPEVHLHHRQAARALAHPRDHPDLR
RAVLQPLRPHEAARLLQVRHARRLRPGAHHLLQGRRQLQDPRRGEVRGRHPGEPHRAEGHRLQGGRQHPGAQAG
VQLQQPQRLYHGRQAERHQGELQDPPQHRGRQRAAHRPLPAEHPHRRRPRAAARQPLPEHPVRPEQRPQREA
RSHGPAGVRDRRRDHSRHGRGLN

mCherry-L-Arg10 (7)

MGSSHHHHHSSGLVPRGSHMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAK
LKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD
GEFIYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKK
PVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKTGGSGSGSGSGSGSGSGSGG
SSASRRRRRRRRRR

mCherry-L-Tat (8)

MGSSHHHHHSSGLVPRGSHMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAK
LKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD
GEFIYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKK
PVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKTGGSGSGSGSGSGSGSGSGG
SSASGRKKRRQRRR

Cre-L-B1 (9)

MGSSHHHHHSSGLVPRGSHMASNLLTVHQNLPALPVDATSDEVKRNLMDFRDRQAFSEHTWKMLLSVC
RSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRR
IRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIA EIARIVKDISRTD
GGRMLIHIGRTKTLVSTAGVEKALS LGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATS QLSTR
ALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWNTNVNIVMNYIRNLD
SETGAMVRLLEDGDGAPSGGGGGSGGGGSASWFKREQGRGAVHRGGAHPGRAGRRRKRQVQRVRRGRGR
HLRQADPEVHLHHRQAARALAHPRDHPDLRRAVLQPLRPHEAARLLQVRHARRLRPGAHHLLQGRRQLQ
DPRRGEVRGRHPGEPHRAEGHRLQGGRQHPGAQAGVQLQQPQRLYHGRQAERHQGELQDPPQHRGRQRA

AHRPLPAEHPHRRRPRAAARQLPEHPVRPEQRPQREARSHGPAGVRDRRRDHSRHGRGLN

Cre-L-Arg10 (10)

MGSSHHHHHSSGLVPRGSHMASNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVC
RSWAAWCKLNNRKWFPAEPEDVRDYLQYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRR
IRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAIEIARIRVKDISRTD
GGRMLIHIGRTKTLVSTAGVEKALS LGVT KLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQ LSTR
ALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWNTNVNIVMNYIRNLD
SETGAMVRLLEDGDGAPSGGGGGSGGGSASRRRRRRRRRR

+36GFP (11)

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGKGDATR GKLT LKFICTTGKLPVPWPTL
VTTLYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL VNRIKLG RDF
KEKGNILGHKLRYNFN SHKVIYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNH
LSTRSKLSDPKERD H MVLLEFVTAAGIKHGRDERYK

+36GFP-L-mCherry (12)

MASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGKGDATR GKLT LKFICTTGKLPVPWPTLVTTLYGVQ
CFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL VNRIKLG RDFKEKGNILGH
KLRYNFN SHKVIYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSK
DPKEKRDH MVLLEFVTAAGIKHGRDERYKGGSGGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG
FMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIP
DYLKLSFPEGFKWERVMN FEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFP SDGPVMQKKTMGWEASSERM
YPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGR
HSTGGMDELYKLEHHHHHH

+36GFP-L-Cre (13)

MASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGKGDATR GKLT LKFICTTGKLPVPWPTLVTTLYGVQCFSRYPKH
KRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL VNRIKLG RDFKEKGNILGHKLRYNFN SHKVIYITADKR
NGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSDPKERD H MVLLEFVTAAGIKHGRDERY
KGGSGGGSGG
VCRSWAAWCKLNNRKWFPAEPEDVRDYLQYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKE
NVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAIEIARIRVKDISRTD GGRMLIHIGRTKTLV
STAGVEKALS LGVT KLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQ LSTRALEGIFEATHRLIYGAKDDSGQRYL
AWSGHSARVGAARDMARAGVSIPEIMQAGGWNTNVNIVMNYIRNLDSETGAMVRLLEDGGGSHHHHHH

References:

1. Miller, V. M., Nelson, R. F., Gouvion, C. M., Williams, A., Rodriguez-Lebron, E., Harper, S. Q., Davidson, B. L., Rebagliati, M. R., and Paulson, H. L. (2005) CHIP Suppresses Polyglutamine Aggregation and Toxicity In Vitro and In Vivo, *The Journal of Neuroscience* 25, 9152-9161.
2. Garnier, J., Gibrat, J. F., and Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence, *Methods Enzymol* 266, 540-553.
3. Prevelige Jr, P., and Fasman, G. D. (1989) Chou-Fasman prediction of the secondary structure of proteins, In *Prediction of protein structure and the principles of protein conformation*, pp 391-416, Springer.