CLONING, PURIFICATION AND CHARACTERIZATION IN PROKARYOTIC EXPRESSION SYSTEMS

STEPS FOR PROTEIN OBTENTION

OBSTACLES

POTENTIAL SOLUTIONS

STEP 1
Construct design

Consider:
1. Rare codons (when not optimizing)
2. Protein size and domain boundaries
3. Protein sequence (disulfide bonds, contiguous hydrophobic regions...)
4. Protein activity and toxicity*
5. Protein needs signal peptide?
6. Open reading frame (Methionine at the N-terminus and Stop codon at the C-terminus)
7. Lack in data can be supplemented by predictions of Structure (Alphafold, RoseTTAFold, I-TASSER...) & Stability (Aggrescan 3D, ProtParam)

LINKER TYPES · FUSION TAGS

*GeneArt will ask for proof of safety in case that the protein is related to any toxin.

- Protein MW >60 kD → Design construct of single, globular domain; Use solubility-enhancing fusion tag
- Many rare codons or tandem rare codons → Codon optimize target gene; Supplement rare tRNAs or use special strains such as Rosetta or CodonPlus *E.coli*
- Domain boundaries are unknown → Use secondary structure prediction; Thread sequence onto a homologous structure; Use de-novo prediction such as Alphafold
- Post-translational modifications or co-factor needed → Change microorganism or add the co-factor or co-express enzymes responsible for the modification

STEP 2
Vector selection and cloning

Consider:
1. Plasmid copy number
2. Promoter strength
3. Promoter basal expression
4. Promoter mode of induction
5. Fusion tag for placement, solubility enhancement, or affinity properties
6. Removal of fusion tag?

pET VECTORS · PROMOTERS

- Expressing toxic protein → Use tightly controlled promoter system (pLysS)
- Expressing aggregation-prone protein → Coexpress with chaperones; Use solubility-enhancing fusion tag

STEP 3
Choose expression strain

Consider:
1. Protease susceptibility
2. Rare codon usage
3. Disulfide bond requirements

E.COLI STRAINS

- Protein susceptible to proteolytic degradation → Use protease deficient strain
- Protein contains disulfide bonds → Express in trxB/gor mutant strain with Trx tag; Export to periplasm

STEP 4
Expression test

Consider:
1. Induction temperature
2. Inducer concentration
3. Choice of culture media

- Cell death due to high basal expression or toxic protein → Use tightly controlled promotor system; Coexpress with partner protein
- No protein/degraded → Check integrity of plasmid, promoters, sequence, and Met/STOP codons.; Consider harmonization of codon usage[1] and other strains.; Optimize temperature and inducer concentration; Check protease target sites in protein sequence

Soluble expression?

- No (w/ disulfide bonds) → Disulfide bonds not formed → Export to periplasm; Express in trxB/gor mutant strain, with oxidative cytoplasm (e.g. Origami strains), or helper systems (e.g. SOX)
- No (w/o disulfide bonds) → Misfolding or aggregation of protein into inclusion bodies → Coexpress with chaperones (e.g. TAKARA system); Lower induction temperature (20-16 ºC); Lower inducer concentration (1 mM-0.001 mM); Use solubility-enhacing fusion tag or fusion partner: GST (dimerizes, Glutathione-S-transferase), MBP (solubility tag, binds to amylose in resins), SUMO (clean after cleavage, ubiquitin-like modifier), TrxA (cleaved, thioredoxin), NusA..; Harmonization of codon usage[1]; Mild solubilization from inclusion bodies (N-lauroyl); Strong solubilization from inclusion bodies and refolding (Urea 8M or Guanidine hydrochloride 6M); Post-translational modification or co-factor needed?
- Yes

STEP 5
Purification from large-scale expresison

Consider:
1. Effective purification process (IMAC, SEC, IEC)
2. Compatible storage buffer, apt for desired downstream applications

- Degradation → Purify at lower temperatures (4ºC); Increase protease inhibitor concentration; Add stabilizers (glycerol or sugars); Change the cell lysis method
- No bindind to affinity column → Use fresh columns and buffers. Store them at 4ºC; Change purification conditions: Salt concentration, pH (always above 7, ideally 8), Buffer composition, Column type (Ni, Co, GST-based...), Decrease loading flow; Purification tag may be unavailable: Switch position or increase length, Use detergent to relax the protein structure, Use reducing agents in lysis (if S-S present) and/or purification buffers.; Osmotic shock to remove periplasm metallophores
- Impurities (protein, LPS or nucleic acids) → Use additional/alternative purification steps. Mind pH, size, properties or column composition (Co, Ni..); Nucleic acids: Increase salt in purification buffers. Treatment with DNAse/RNAse; High LPS content: Re-purify with specific detergents (e.g DOC)
- Inactive protein → Cleave affinity tag, go back to potential solutions for misfolded proteins or re-design construct

STEP 6
Storage and downstream applications

Consider:
1. Protein pI, hydropathicity, expected interactions
2. Protein size
3. Most appropiate dyalisis method
4. Usage/application of the protein to decide formulation (in vivo, biotech process...)

BUFFERS · ADDITIVES · FDA- APPROVED FORMULATIONS

- Aggregation → Change the storage buffer or consider additives; Avoid freezing; Check that pH is far from pI; Avoid metal leaking from columns, include EDTA in storage buffer
- S-S Dimerization → Use reducing agents during purification[2]; Lower the pH to favor the protonated thiol (SH) form; Remove oxygen from water in dyalisis, degasing with N_2
- Degradation → Store at lower temperature (check if freezing is convenient); Add stabilizers such as glycerol or sugars

EXPANDED BIBLIOGRAPHY

(1) Codon harmonization

(2) Reducing agents during purification

FUSION TAGS FOR PROTEIN SOLUBILITY

Fh8, MBP, NusA, Trx, SUMO, GST, SET, GB1, ZZ, HaloTag, SNUT, Skp, T7PK, EspA, Mocr, Ecotin, CaBP, ArsC, IF2-do-main I, Expressivity tag, RpoA, SlyD, Tsf, RpoS, PotD, Crr, msyB, yjgD, rpoD.

LINKER TYPES

Sequence	Type	Linker function
GGSSRSS	Flexible	Phage display, biological activity
GGGGS	Flexible	Increase stability, folding or biological activity
$(GGGGS)_3$	Flexible	Increase stability, folding or biological activity
$(G)_{6-8}$	Flexible	Increase stability or folding
$(EAAAK)_{1-3}$	Rigid	Increase stability or folding
$A(EAAAK)_4ALEA(EAAAK)_4A$	Rigid	Increase biological activity, expression or alter PK
PAPAP	Rigid	Increase biological activity
AEAAAKEAAAKA	Rigid	Increase biological activity
Disulfide (S-S)	Cleavable	Increase biological activity or alter PK
VSQTSKLTR_AETVFPDV	Cleavable	Enable targeting, cleaved by Factor XIa/FVIIa
PLG_LWA	Cleavable	Enable targeting, cleaved by matrix metalloprotease-1
RVL_AEA	Cleavable	Enable targeting, cleaved by HIV-1 protease HIV PR
EDVVCC_SMSY	Cleavable	Enable targeting, cleaved by HCV protease NS3
GGIEGR_GS	Cleavable	Enable targeting, cleaved by Factor Xa
TRHRQPR_GWE	Cleavable	Enable targeting, cleaved by Furin
AGNRVRR_SVG	Cleavable	Enable targeting, cleaved by Furin
RRRRRRR_R_R	Cleavable	Enable targeting, cleaved by Furin
GFLG_	Cleavable	Enable targeting, cleaved by Cathepsin B
LE	Dipeptide	Alter PK
AAY	Tripeptide	Reduce formation of junctional epitopes

pET VECTORS

(page 21)

Options:

Amp/Kan resistance, oriniter T7, His.tag, S-tag, T7-tag, protease cleavage sites, segnal sequence...

PROMOTERS

Organism	Promoter	Inductor	Plasmid	Resistance / Traits	Observations
E. coli *LB medium*	*T7lac*	IPTG	pET22b pET24a-d pET26b pET28a-d	Amp / Periplasm Kana / T7Tag® Kana / Periplasm Kana / T7Tag®	Strong promoter. This promoter requires DE3 strains that co-express the T7 RNA polymerase. F1 origin for ssDNA plasmid mutagenesis or sequencing.
	P_{BAD}	Arabinose	pBAD33 pBAD/HisA-C	Cm Amp	Medium promoter. No strain restriction.
	lac	IPTG	pBluescript II SK	Amp	Weak promoter. No strain restriction, but ocasionally leaky.
L. lactis *M17 medium + 0.5% glucose after autoclave All processes 30 ºC*	*PnisA*	Nisin A	pNZ8148	Cm	NcoI site for translational fusions at the ATG. Contains a terminator after the MCS. Sequence adaptation for cloning into NcoI site can result in a change in the second amino acid of a protein

E.COLI STRAINS

BL21 (DE3)
Most widely used strain for protein expression. Deficient in lon and OmpT proteases. Lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase under the control of the lac UV5 promoter. It is scalable. Compatible with pET vectors and IPTG induction.

BL21 (DE3) pLysS (R/ Cm)
Derived from BL21 (DE3). Contains the pLysS plasmid, which carries the gene encoding the T7 lysozyme, reducing the basal expression of the genes under the control of the T7 promoter. Promotes stability when the pET plasmid encodes for proteins that affect cell growth and viability.

Lemo21 (DE3)
Derived from BL21 (DE3). Tunable expression of T7 polymerase is achieved by varying the level of lysozyme (lysY) by adding L-rhamnose to the expression culture at levels from zero to 2000 μM. When Lemo21(DE3) is grown without rhamnose, the strain performs the same as a pLysS containing strain. However, optional addition of rhamnose tunes the expression of the protein of interest. Useful for membrane proteins, toxic proteins and proteins prone to insoluble expression.

BL21 (DE3) SOX (R/ Cm)
BL21 with a plasmid encoding for DsbC and Erv1p sulfhydryl oxidase from S.cerevisiae, that participate in disulfide bond formation. Strain from Ario de Marco (UNG). Pre-expression of enzymes induced with Arabinose (30-45' 30ºC).

Origami B (DE3) (R/ Kan Tet)
Derived from BL21. Double mutant in the reductases TrxB and gor. This causes more oxidant conditions in the cytoplasm. Deficient in lon and OmpT proteases. Lysogenic for λ-DE3. KanR and TetR. Recommended only for the expression of proteins that require disulfide bond formation for proper folding.

Origami 2 (DE3) (R/ Tet)
Derived from K-12. Double mutant in the reductases TrxB and gor. This causes more oxidant conditions in the cytoplasm. Lysogenic for λ-DE3. Recommended only for the expression of proteins that require disulfide bond formation for proper folding.

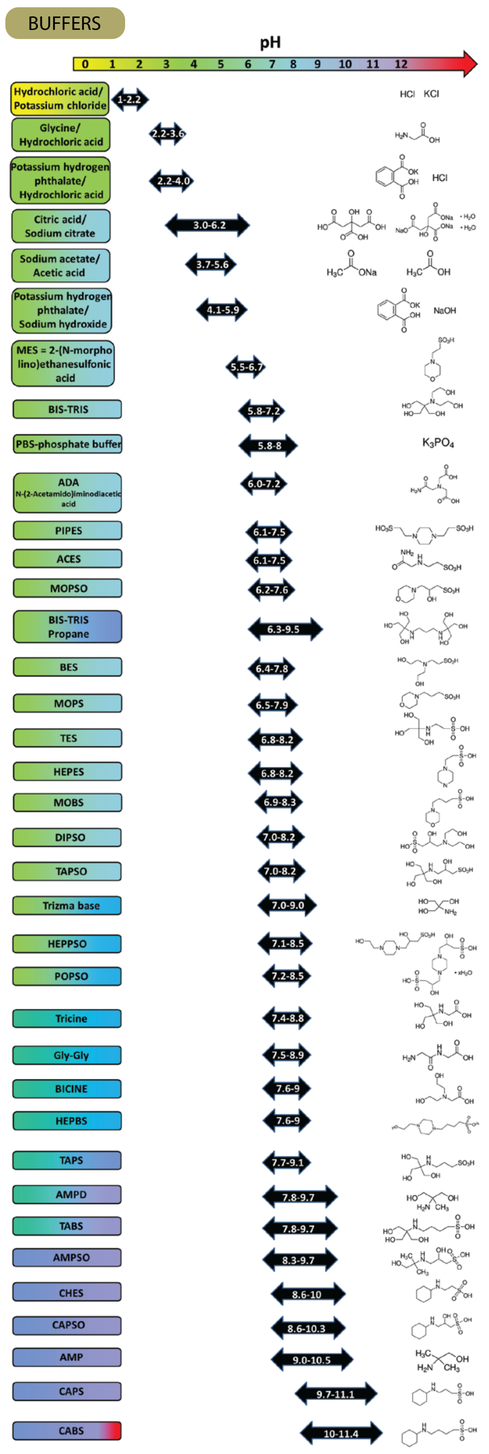
Codon Plus
Derived from BL21. There are different options with genes expressing tRNAs that recognize the rare codons in E.coli. There are also options with or without λ-DE3. Generally CmR.

Rosetta (DE3) (R/ Cm)
Derived from BL21. Contain pRARE (Rosetta) or pRARE2 (Rosetta2) plasmid designed to enhance the expression of eukaryotic proteins that contain codons rarely used in E. coli. Lysogenic for λ-DE3. Both are CmR.

ClearColi® BL21 (DE3)
Genetically modified LPS that does not trigger an endotoxic response in human cells. Considered "endotoxin-free".

DH5α
Derived from K-12. Strain used to amplify the number of plasmids prior to a later transfection. High insert stability due to recA1 mutation. Not commonly used in recombinant protein expression. Cloning strain with multiple mutations that enable high-efficiency transformations.

BUFFERS



ADDITIVES

Detergents & chaotropics
Urea (denaturant)
N-Lauroylsarcosine (anionic surfactant)
Tween 20 (w/ lauric acid)
Tween 80 (w/ oleic acid)
Triton X-100 (non-ionic surfactant)
CHAPS

Reducing agentsts & chaotropics
TCEP (stable longer, less toxic)
DTT (strong reducing power, cheap)
β-mercaptoethanol (denaturant, RNA preservation)

Non-sugar stabilizers
Glycerol (1-20%)
Salts such as NaCl 100-500 mM, KCl or $(NH_4)_2SO_4$
Trifluoroethanol (TFE) or Trimethylamine N-Oxyde (TMAO)
a-Helix stabilizers to promote folding
Specific chaperones (Heat shock proteins or specific ligands)
L-arginine or other amino acids.
PEG 8000 (prevents aggregation)

Sugars
Glucose, Sucrose, Trehalose, Mannitol, Maltose, Xylitol or Sorbitol 5-20% or 50-200 g/L
(useful for general stability and freeze-drying protection)
BSA

Protease inhibitors
Commercial (*e.g.* c0mplete)
PMSF 1mM (Inhibits serine proteases)
EDTA 1-5mM (Inhibits metalloproteases)
Leupeptin (inhibits serine and cysteine proteases.
Others: AEBSF, Benzamidine, Aprotinin, Chymostatin, Antipain

FDA- APPROVED FORMULATIONS

ACKNOWLEDGEMENTS

UAB Universitat Autònoma de Barcelona

ibb · NanoBioTechnology · ciber-bbn Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina · nanbiosis

MADE BY: ELOI PARLADÉ MOLIST