Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 31

Eukaryotic protein expression systems - II

Eukaryotic protein expression systems-I (lecture 30)

Protein expression in yeast and insect cells

Eukaryotic protein expression systems-II (lecture 31)

Protein expression in mammalian cells

Cell-free protein expression systems

Eukaryotic protein expression systems-III (lecture 32)

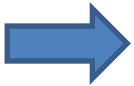
Production of recombinant proteins in plants and farm animals

---

Human gene therapy (lecture 33)

Why should you produce a recombinant protein in mammalian cells?

- rHuman protein must be identical to the natural protein
- Prokaryotic yeast and insect expression systems are generally unable to produce authentic mammalian proteins due to the absence of appropriate mechanisms for carrying out the necessary post-translational modifications to the protein
- **Correct disulphide bond formation.** Reaction mediated by the enzyme disulphide isomerase. An improperly folded protein is unstable and lacks activity
- **Proteolytic cleavage of a precursor form.** Selected segments of amino acid sequences are removed to yield a functional protein



- **Glycosylation.** Gives a protein with stability and, in some instances, its distinctive properties. The most common protein glycosylations occur by the addition of specific sugar residues to serine or threonine (O-linked) or to asparagine (N-linked)
- **Addition of amino acids within proteins**. Modification of this type includes phosphorylation, acetylation, sulfation

ERYTHROPOEITIN (EPO)

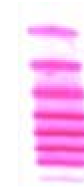
EPO is a glycoprotein that serves as the primary regulator of red blood cells in mammals. It stimulates bone marrow stem cells to differentiate into red blood cells and controls hemoglobin synthesis and red blood cell concentration.

Human EPO is a 30,400-dalton molecule containing 165 amino acids and four carbohydrate chains that incorporate sialic acid residues. There are several forms of EPO, designated by Greek letters, that differ only in the carbohydrate content.

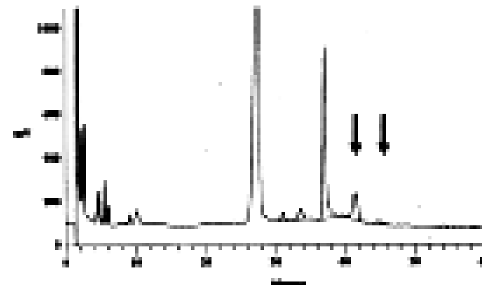
In infants, EPO is produced mostly in the liver, but the kidneys become the primary site of EPO synthesis shortly after birth. EPO production is stimulated by reduced oxygen content in arterial blood in the kidneys. Circulating EPO binds to receptors on the surface of erythroid progenitor cells that in turn mature into red blood cells.

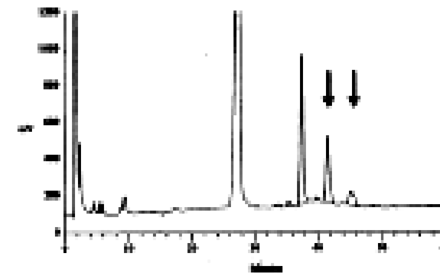
Human EPO was first isolated and later purified from urine in the 1970s. Soon, the gene encoding EPO was cloned and several groups devised recombinant DNA methods to produce EPO by the mid-1980s.

Generic	Trade	Made by	Sold by
Epoetin Alfa	Epogen	Amgen	Amgen
Epoetin Alfa	Procrit	Amgen	Ortho Biotech / J&J
Epoetin Alfa	Eprex	Amgen	Ortho (outside USA)
Darbepoetin Alfa	Aranesp	Amgen	Amgen
Epoetin Beta	NeoRecormon	Roche ?	Roche (outside USA)

	EPO I	EPO II
		
Sialic acid	14.0	14.1
In-vivo activity (U/mg)	224,000	400,000

The carbohydrate structures of the two EPO isoforms are different





Mammalian expression vectors
(non-viral vectors)

Mammalian expression Vectors

- Contain an efficient promoter elements for high level transcriptional initiation
- Contain mRNA processing signals
- Contain selectable markers
- Plasmid sequences for propagation in bacterial hosts
- contains a eukaryotic origin of replication from an animal virus, e.g. Simian virus 40 (SV40)
- Origin of replication from E. coli
- promoter sequences that drive both the cloned gene(s) and the selectable marker gene(s)
- transcription termination sequences - adenylation signals from animal virus e.g. SV40
- Matrix-attachment regions
- Chromatin insulators, locus control regions

Viral genetic elements have been used to construct a number of eukaryotic expression vectors

- ✓ **Viruses are highly efficient replicators & viral gene expression is adapted to eukaryotic systems**
  - very strong promoters (CMV immediate / early promoter)
  - small introns (CMV intron**)**
  - regulatory elements often constitutive – requires only host transcription factor binding

- ✓ **Therefore, the strategy has been to use the regulatory elements of viral genomes**
  - Promoters, enhancers, polyadenylation signals, introns, replication origins, IRES elements.

Eukaryotic promoters commonly used in mammalian expression vectors

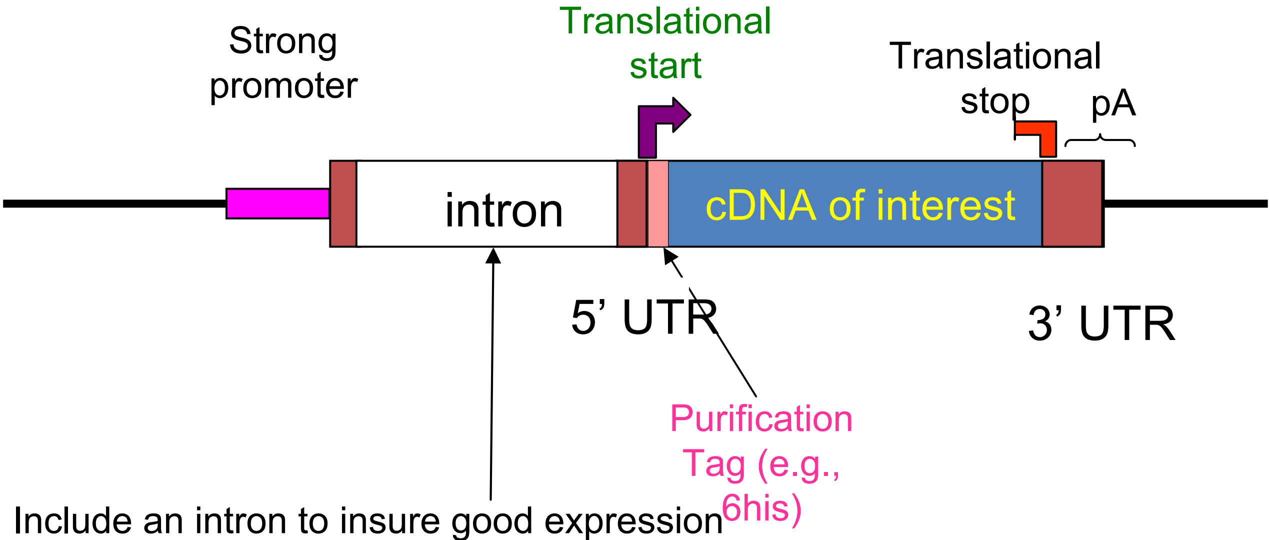
Animal viruses:

Cytomegalovirus (CMV) promoter
SV40 promoter/enhancer
Vaccinia virus promoter
Viral LTRs (MMTV, RSV, HIV)

Mammalian genes:

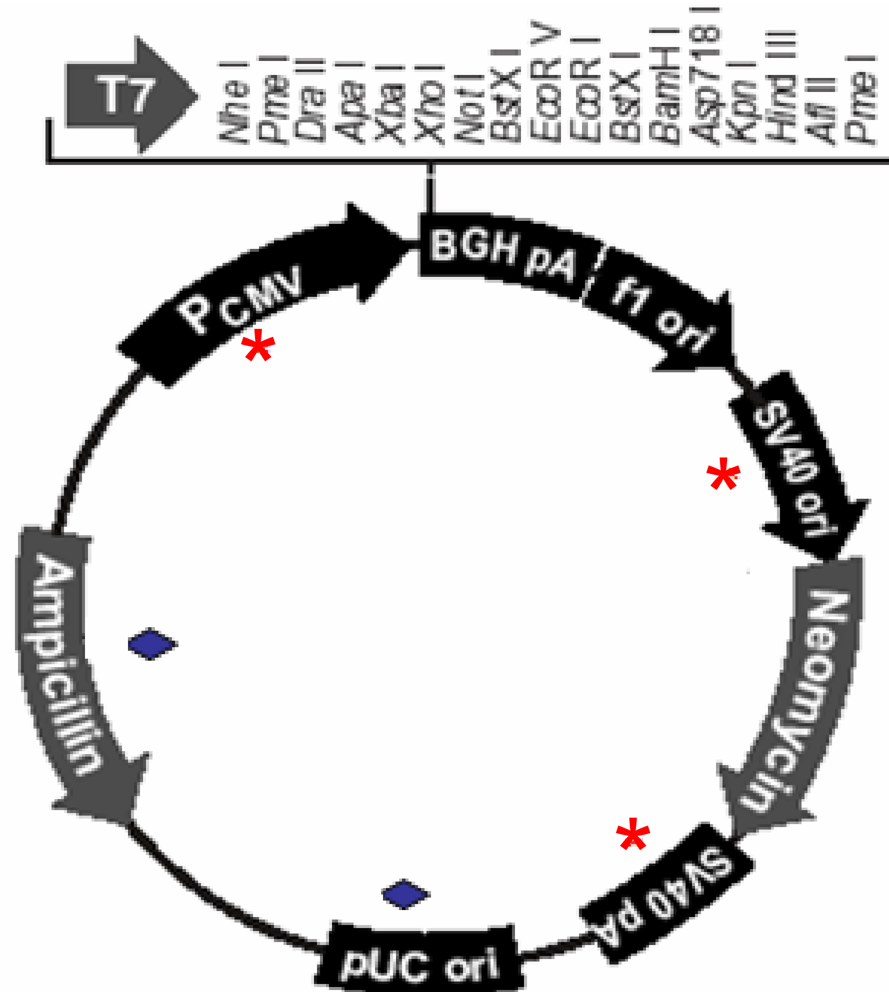
Promoters of constitutively expressed genes (actin, GAPDH),
Promoters of genes expressed in a tissue-specific manner (albumin, NSE)
Promoters of inducible genes (Metallothionin, steroid hormones)

Features of a mammalian expression vector for recombinant protein production:



Also: chromosomal location
amplification

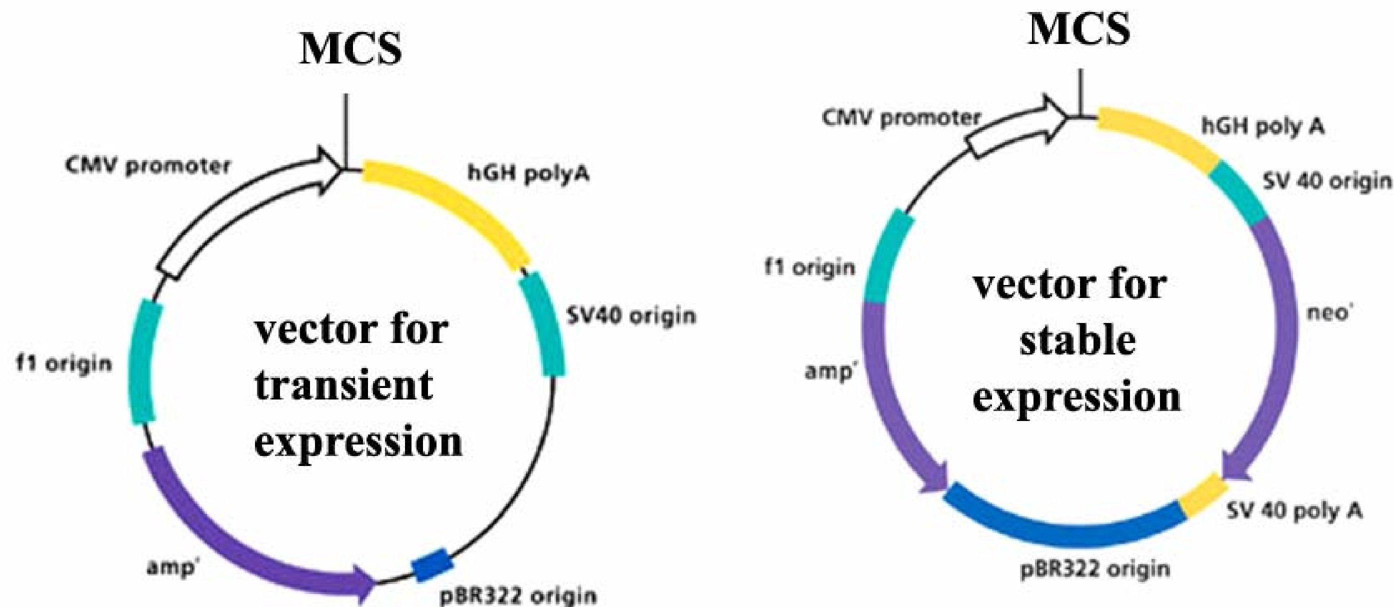
Eukaryotic expression vectors typically contain DNA sequences of viral (*) and bacterial origin (◆)



Transient transfection - standard non-selectable transformations of mammalian cells that replicate the plasmid DNA, but do not faithfully segregate the plasmid to progeny; in time the plasmid will be lost from the cultured cells

Stable transfection - a selectable marker is included in the plasmid vector so that after the cells have been transformed, the rare cell that incorporates the plasmid DNA into its genome (through DNA repair and recombination enzymes) can be isolated and cloned (because progeny of the transformed cell will inherit the plasmid DNA including the selectable marker gene, e.g. antibiotic resistance gene)

Transient *vs* stable expression



The strong human cytomegalovirus (CMV) promoter regulatory region drives constitutive protein expression levels.

The presence of the SV40 replication origin will result in high levels of DNA replication in SV40 replication permissive COS cells.

These vectors contain the pMB1 (derivative of pBR322) origin for replication in bacterial cells, the b-lactamase gene for ampicillin resistance selection in bacteria, hGH polyA, and the f1 origin.

Stable expression vectors carry the aminoglycoside phosphotransferase II gene (neomycin resistance gene or neo^r) that confers resistance to aminoglycosides such as G418 sulfate, allowing selection of stable transfectants.

G418 is used for the selection and maintenance of eukaryotic cells expressing the *neo gene. G418 is an* aminoglycoside antibiotic produced by *Micromonospora rhodorangea.* G418 blocks polypeptide synthesis in eukaryotic cells by binding irreversibly to 80S ribosomes and therefore disrupting their proofreading capability.

RESISTANCE TO G418

Resistance to G418 is conferred by the *neo gene from transposon Tn5 encoding an aminoglycoside 3'-* phosphotransferase, APH 3' II3. This protein inactivates G418 by covalently modifying its amino or hydroxyl functions therefore inhibiting the antibiotic-ribosome interaction.

The working concentration of G418 Sulfate for selection and maintenance of mammalian cell lines transfected with the *neo gene varies with a multitude of factors including cell type.* Optimal concentrations of antibiotic required to kill a specific host cell line is determined by treating the cells with several concentrations ranging from 100 µg/ml to 1 mg/ml. After treatment, cell death occurs rapidly allowing the selection of transfected cells with plasmids carrying the *neo gene in as little as 7 days post-transfection.*

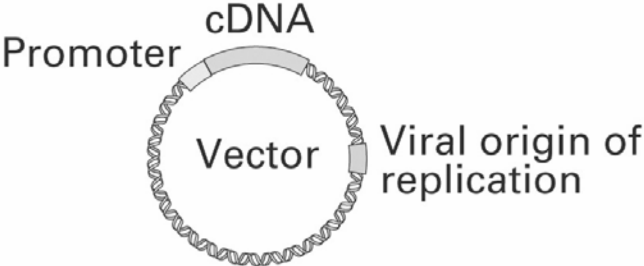
Cell line	**Species**	**Tissue**	**Culture medium**	**G418 (µg/ml)**
HeLa	Human	Uterus	DMEM	200-800
293	Human	Kidney	DMEM	400-1000
B16	Mouse	Melanoma	RPMI	400-1000
CHO	Hamster	Ovary	Ham's	200-400

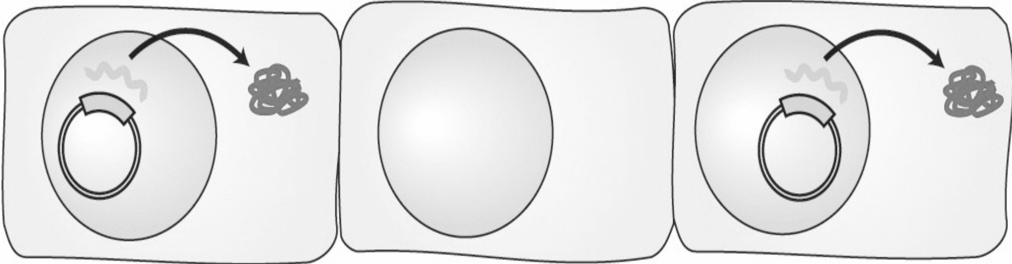
1. Davies J & Jimenez A. 1980. A new selective agent for eukaryotic cloning vectors. Am J Trop Med Hyg 29(5 Suppl):1089-92

2. Bar-Nun S *et al. 1983. G-418, an elongation inhibitor of 80 S ribosomes. Biochim Biophys Acta. 741(1):123-7.*

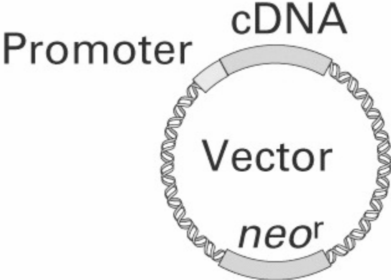
3. Beck E *et al. 1982. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from* transposon Tn5. Gene 19(3):327-36

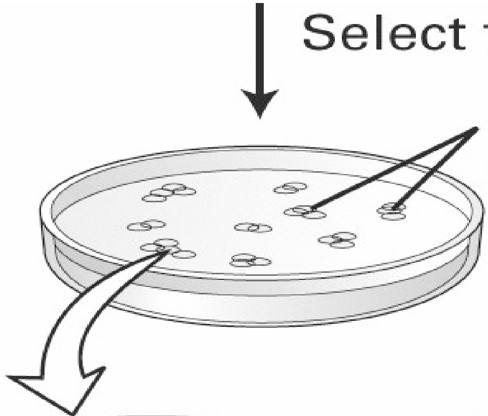
Transient transfection

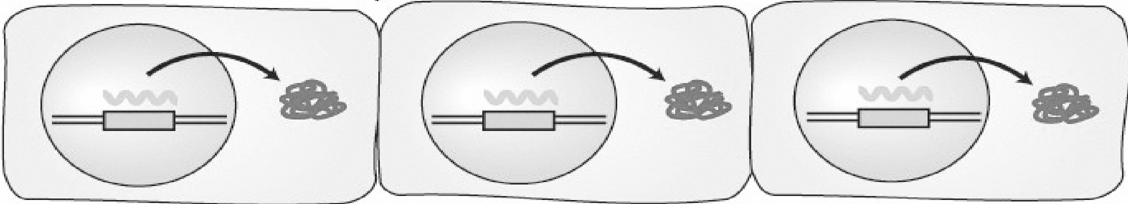




Stable transfection





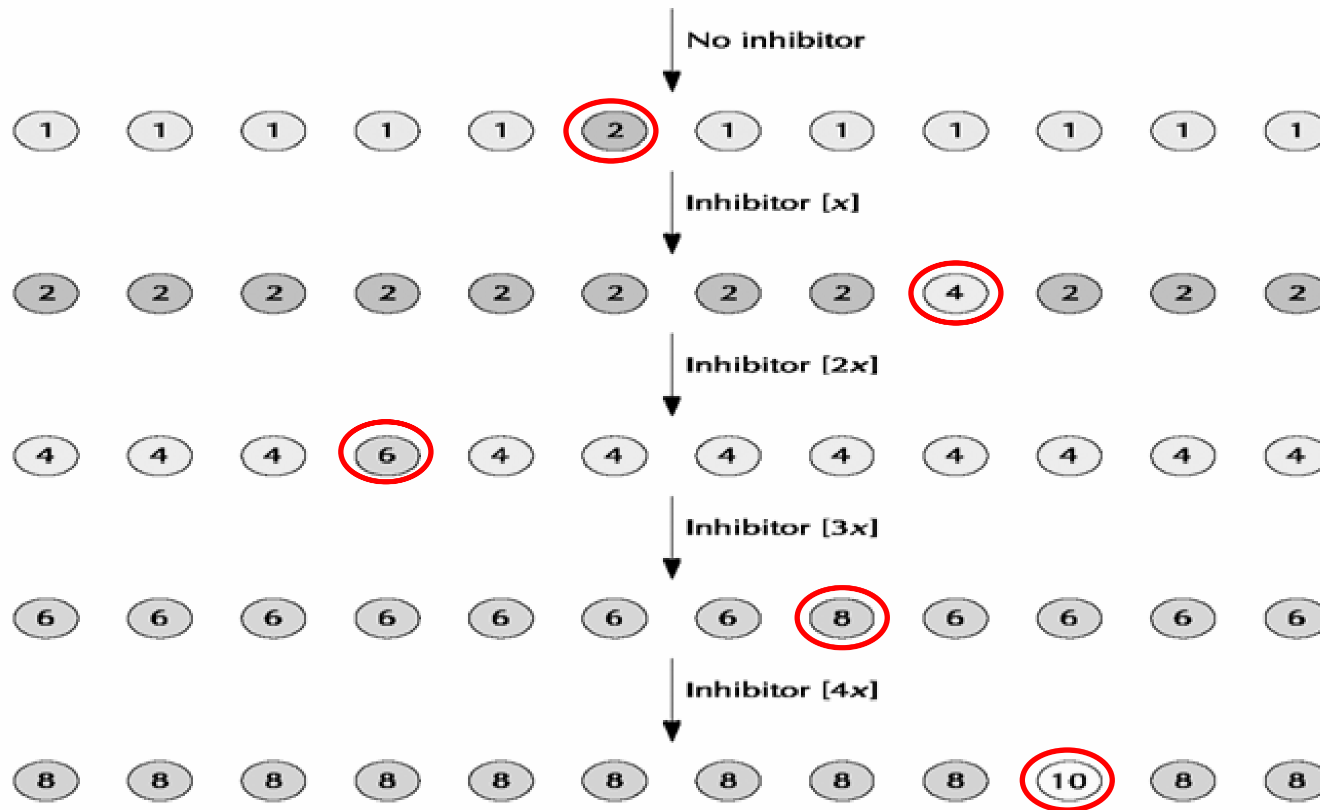


SELECTABLE MARKERS

Enzyme (abbreviation)	Drug for selection
Aminoglycoside phosphotransferase (APH)	G418 (inhibits protein synthesis)
Dihydrofolate reductase (DHFR): Mtx-resistant variant	Methotrexate (Mtx; inhibits DHFR)
Hygromycin-B-phosphotransferase (HPH)	Hygromycin-B (inhibits protein synthesis)
Thymidine kinase (TK)	Aminopterin (inhibits de novo purine and thymidylate synthesis)
Xanthine-guanine phosphoribosyltransferase (XGPRT)	Mycophenolic acid (inhibits de novo GMP synthesis)
Adenosine deaminase (ADA)	9-ß-D-xylofuranosyl adenine (Xyl-A; damages DNA)

- Methotrexate (MTX) inhibits dihydrofolate reductase (DHFR)
- DHFR- host cell with DHFR gene on cloning vector (i.e. linked to target gene)
- Gradually increase MTX concentration in culture
- Gene copy number of DHFR and linked target gene increase to compensate for inhibition of DHFR

Selection Protocol



Gene copy number is shown in circles

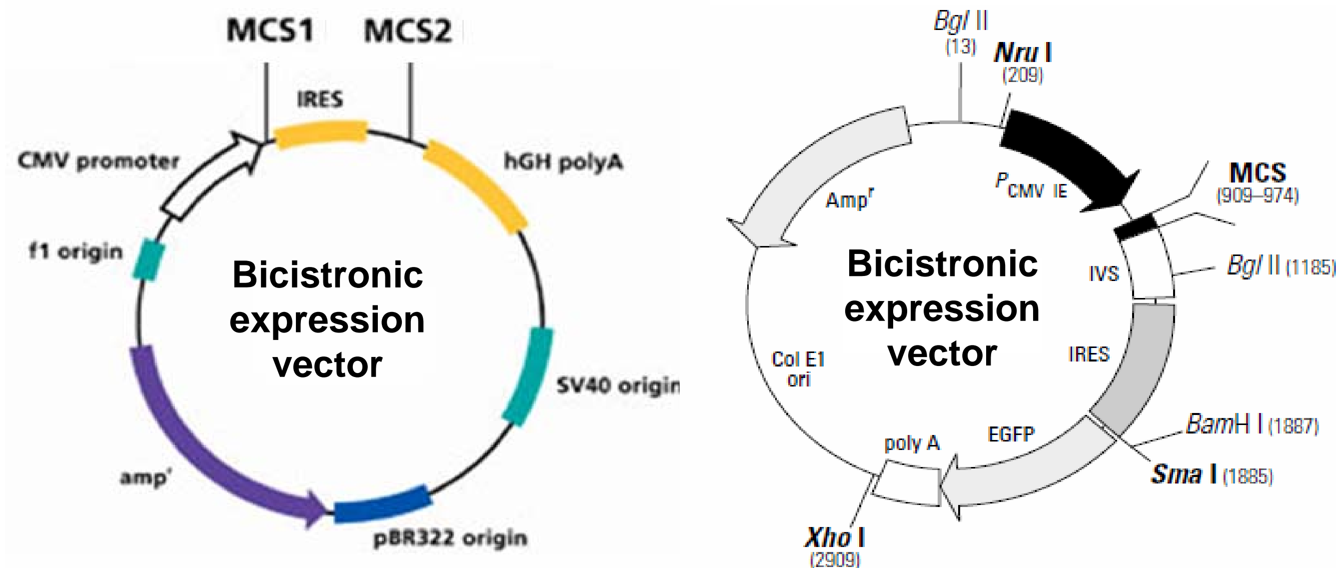
Bistronic mammalian expression vectors

Bistronic expression vectors

Bistronic expression vectors contain an internal ribosomal entry site (IRES) element from the encephalomyocarditis virus (EMCV) for translation of two open reading frames (ORFs) from one bicistronic message.

IRESs are relatively short DNA sequences that can initiate RNA translation in a 5' cap-independent fashion. Placement of the IRES and a second gene of interest (ORF 2) downstream of the first target gene (ORF 1) allows co-expression of ORF 1 in a cap-dependent manner and ORF 2 in a cap-independent fashion, thus facilitating translation of two proteins from one mRNA transcript.

These vectors are designed to drive transcription of the bicistronic message under control of the strong human cytomegalovirus (CMV) promoter regulatory region. BICEP vectors permit co-expression of two genes of interest. Genes cloned into MCS1 are expressed in a cap dependent manner while genes cloned into MCS2 are translated in a cap-independent fashion under control of the EMCV IRES.



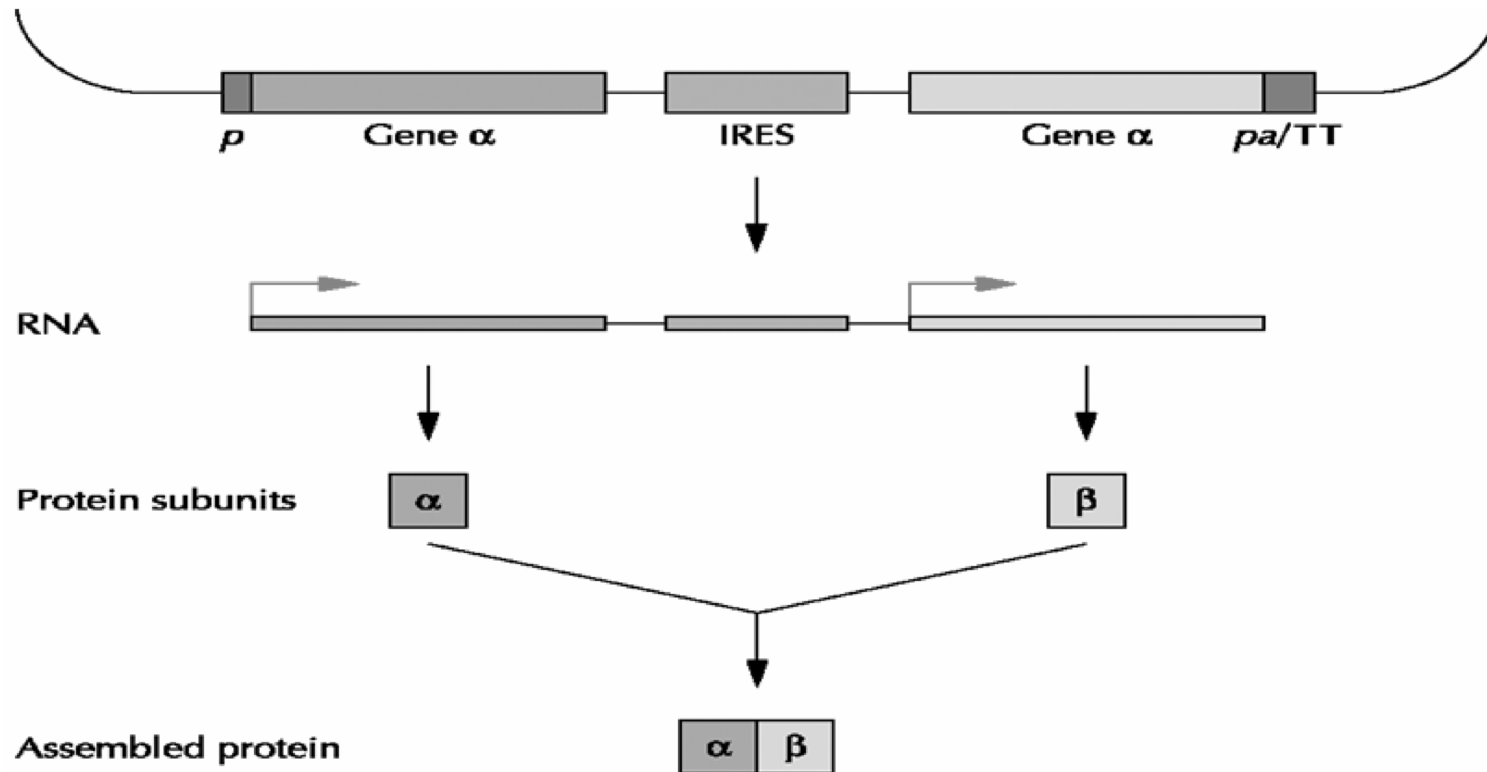
The internal ribosome entry site (IRES) from encephalomyocarditis virus (EMCV) is a noncoding RNA fragment that can initiate high levels of cap-independent protein synthesis in mammalian cells and cell-free extracts.

The IRES is localized to a genome fragment about 430 bases long, immediately 5′ to the AUG, which begins the viral polyprotein open reading frame (ORF). When this region was excised and linked to other portions of the virus ORF, the resulting T7 transcripts were translated even the absence of 5′ capping reactions.

This useful discovery was commercialized in 1990 by Novagen (Madison, WI, USA) in the form of pCITE-1, one of the first vectors that allowed easy linkage of exogenous cistrons onto the cap independent translation enhancer for transcription of hybrid mRNAs and protein expression.

The cap independent translation enhancer (CITE) was later named as internal ribosome entry site, (IRES), and clontech developed a number of bicistronic vectors containing IRES.

Bicistronic expression vectors can be used to express two different sub units of a heterodimeric protein

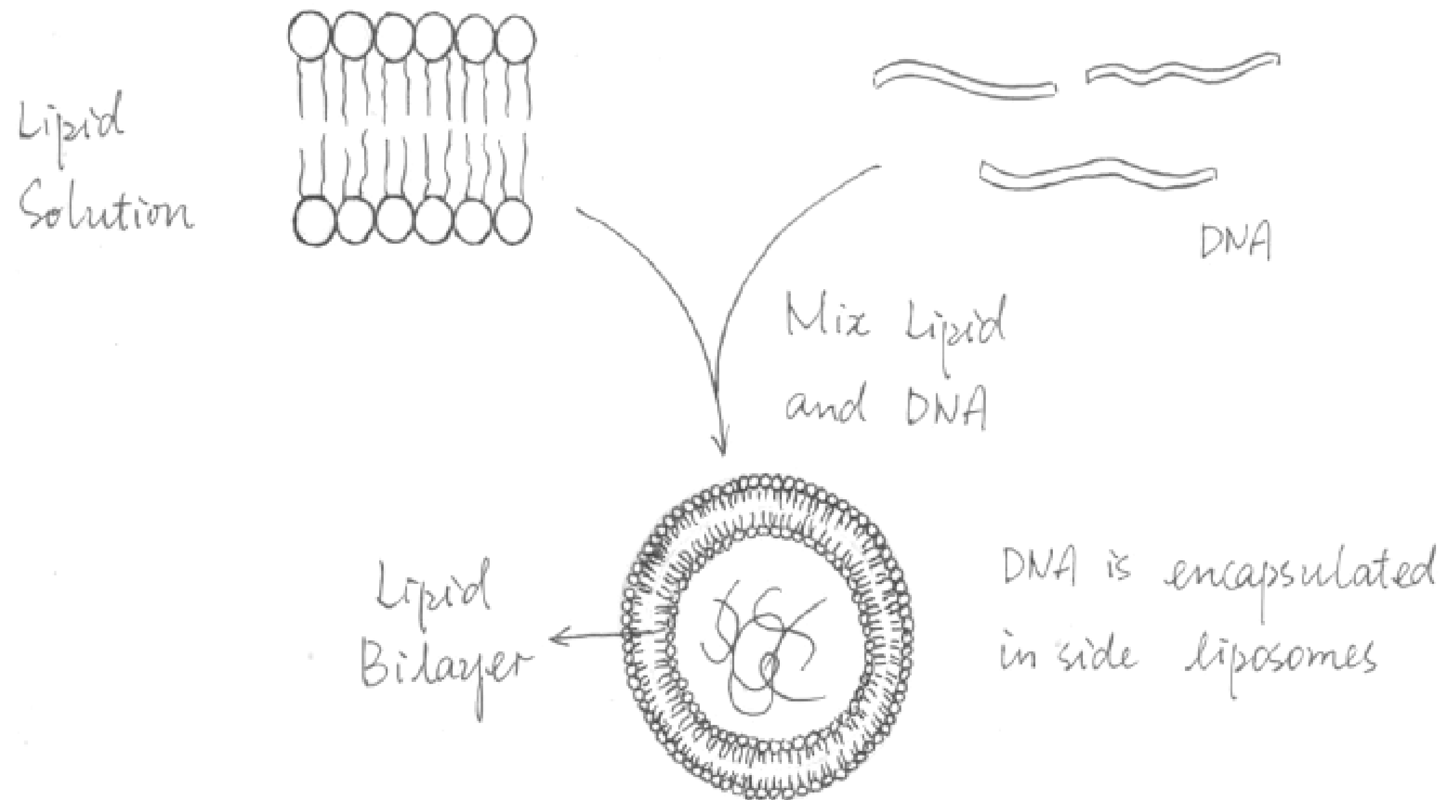


Transfection of DNA into mammalian cells
(non-viral methods)

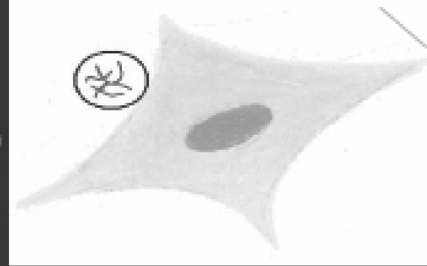
Transfection of DNA into mammalian cells

- Calcium phosphate
  - – Mix cells with DNA in a phosphate buffer
  - – Then solution of calcium salt added to form a precipitate
  - – Cells take up the calcium phosphate crystals which include some DNA
- Liposomes
  - – DNA mixed with lipid to form liposomes, small vesicles with some of the DNA inside
  - – DNA-bearing liposomes fuse with cell membrane carrying DNA inside the cell
- Electroporation

LIPOSOME-MEDIATED GENE TRANSFER INTO MAMMALIAN CELLS

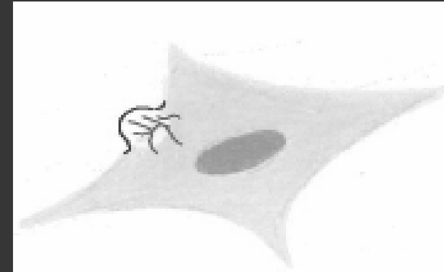


Liposome adheres to cell

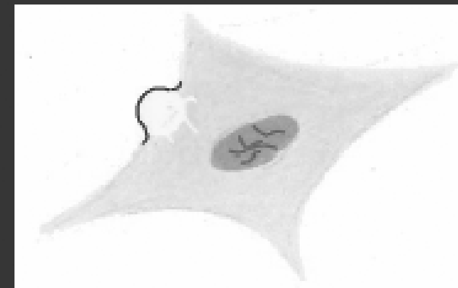


Liposome bilayer fuses with plasma membrane

DNA enters cell

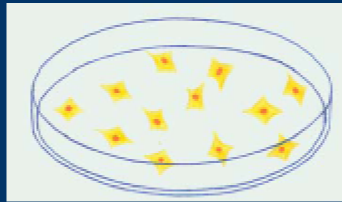


DNA in nucleus



Electroporation

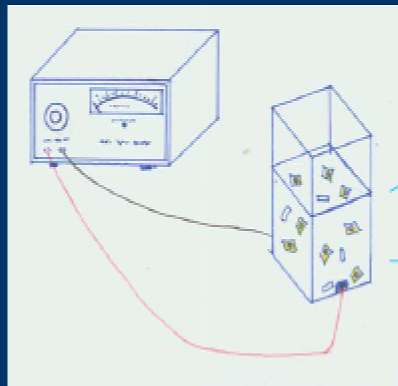
Culture cells



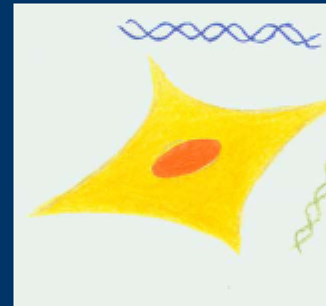
DNA of interest

Selectable gene

Mix DNA and cells in a special cuvette

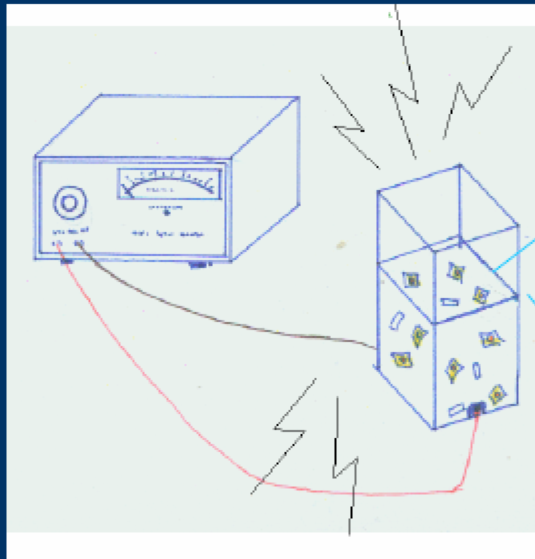


Apply voltage



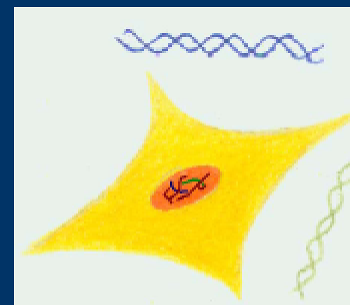
Cells uptake DNA through holes in membrane

Apply voltage



Cells uptake DNA through holes in membrane

DNA enters nucleus



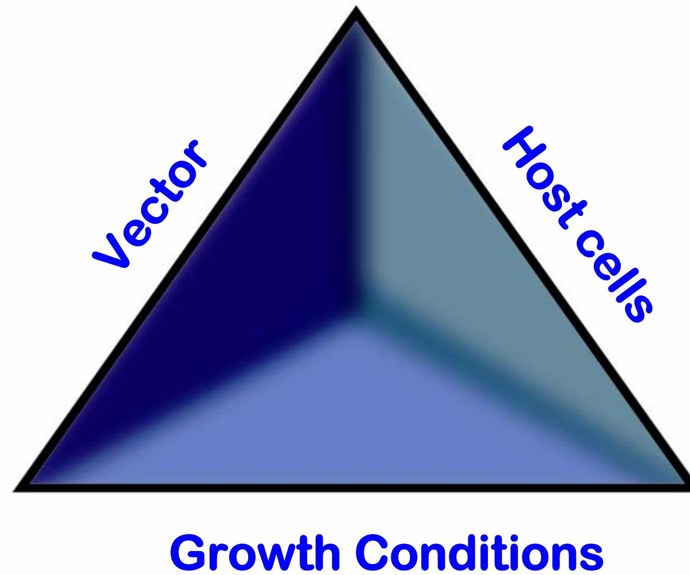
Apply selection

Stable expression

No selection

Transient expression

Three major factors that influence protein expression in mammalian cells



Mammalian cells normally used for recombinant protein production

Transient expression

- African green monkey kidney (COS)
- baby hamster kidney (BHK)
- human embryonic kidney (HEK-293)

Stable expression

- chinese hamster ovary (CHO)

CHINESE HAMSTER OVARY (CHO) CELLS

- Chinese hamster ovary (CHO) cells are the most widely used mammalian cells for transfection, expression, and large-scale recombinant protein production.
- Since CHO cells provide stable and accurate glycosylation, they offer a post-translationally modified product and thus a more accurate *in vitro* rendition of the natural protein.

CHO cells were introduced in the early 1960s as a viable epithelial cell line containing twin female X chromosomes.

Several mammalian expression vectors and cell lines are now commercially available

The FreeStyle™ 293 Expression System is designed to allow large-scale transfection of suspension 293 human embryonic kidney cells in a defined, serum-free medium.

INVITROGEN

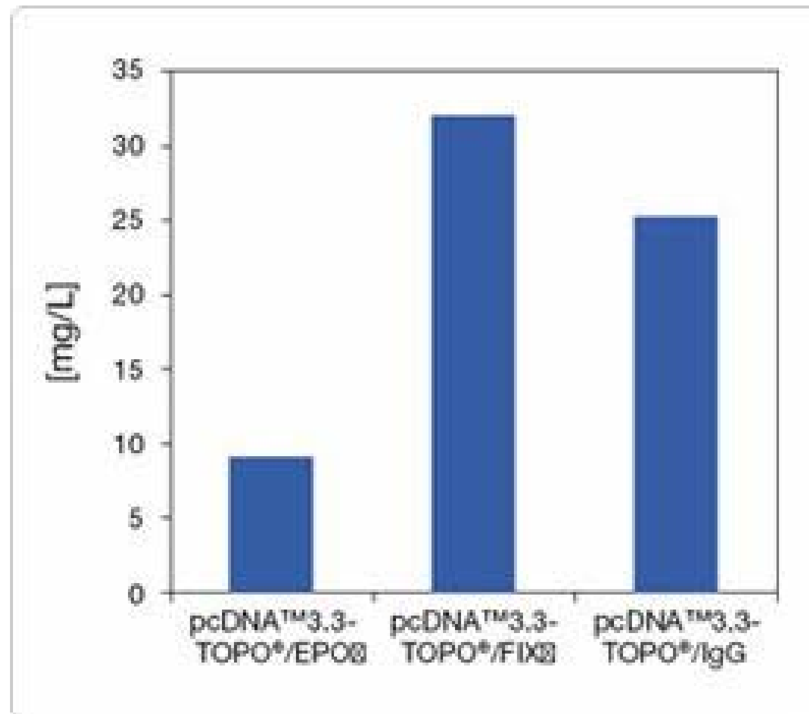
FreeStyle™ 293 System

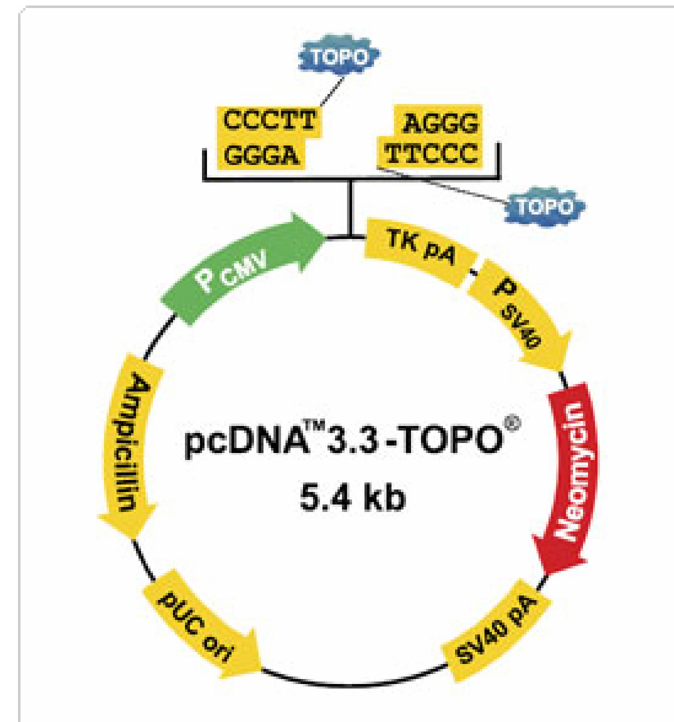
- **Application: Protein production**
- **Components**
  - *FreeStyle™ 293-F suspension cells*
    - Derived from HEK 293 cells for high level protein production in serum-free media
  - *FreeStyle™ 293 expression medium*
    - Chemically defined, serum-free medium optimized for protein production
  - *293fectin™ transfection reagent*
    - Simple, high efficiency transfection of FreeStyle™ 293-F cells

Human growth factor	4.3 g/L
Human growth factor	4.5 g/L
Adhesion protein A	4 mg/L
Adhesion protein B	4 mg/L
Phosphatase enzyme A	5 mg/L

Rapid, scalable protein production in mammalian cells

http://tools.invitrogen.com/content/sfs/manuals/freestyle293_system_man.pdf





Adapted from www.invitrogen.com

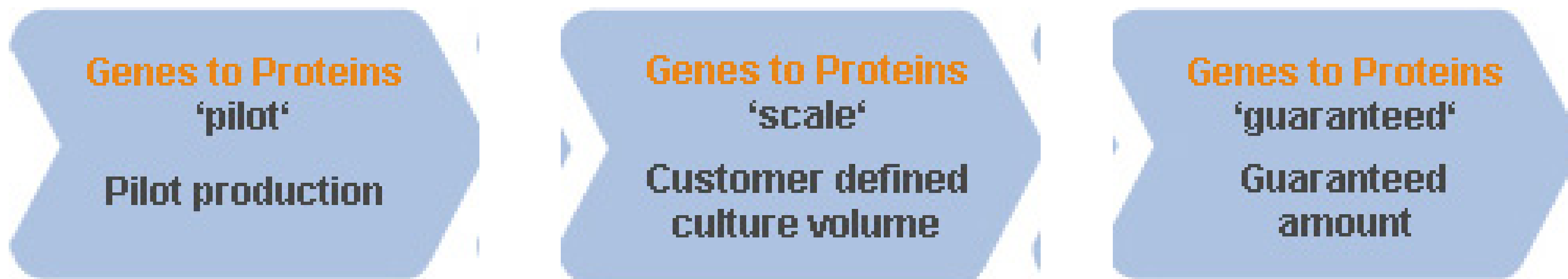
•Improved protein expression reliability through gene optimization

Genes to Expression: Verification of protein expression or determination of the

•best expressing gene (BestExpresser) or gene variant (Best-in-Class) among different optimized versions

•Reliable and fast access to protein

Genes to Proteins: Protein production from stable and transiently transfected cells



http://www.geneart.com

Protein synthesis in cell-free systems

Recombinant protein expression *in vitro:*
* Predates *in vivo expression systems*
1961: Niremberg/Matthaei, polyF synthesis from polyU

* Predominantly used cell free systems:
E.coli 30S extract
Rabbit reticulocyte lysate
Wheat germ extract

* Offers production advantages for:
Proteins toxic to cell division
Proteins produced in inclusion bodies in *E.coli*
Proteins sensitive to intracellular proteases
* For structural biology:
Selective labeling by isotopes for NMR Incorporation of modified amino acids (Se) for crystallography

E.coli cell free system:

* Crude extract (30S):

Endogenous mRNA removed by run-off translation and subsequent degradation

Simple translation apparatus

Comparatively simple control of initiation

CHEAP !!

Rabbit reticulocyte lysate:

Efficient in vitro eukaryotic protein synthesis system

Reticulocytes: Immature red blood cells specialized for hemoglobin synthesis (Hb is 90% of protein content) lacking nuclei but complete translation machinery

Endogenous globin mRNA removed by treatment with Ca2+-dependent micrococcal nuclease, which is then inactivated by EGTA-chelation of Ca2+

Exogenous proteins are synthesized at a rate close to that observed in intact reticulocytes

Both capped (eukaryotic) and uncapped (viral) RNA is translated

Kozak consensus and polyA signal have to be provided on RNA

Synthesizes mainly full-length product

Wheat germ extract:

Convenient alternative to rabbit reticulocyte lysate
Low levels of endogenous mRNA: low background
No micrococcal treatment necessary

Exogenous proteins (mammalian, viral, plant) synthesized at high levels

Advantages of *in vitro* protein expression systems

Express proteins toxic to cells

Express insoluble proteins

Mutational analysis

Protein labeling: Radiolabeling or fluorescent incorporate unusual amino acids

Producing proteins for structural studies, inc. NMR, crystallography, mass spec

Eliminates the time and cost required with cell-based systems:

- Cell line maintenance
- Expression optimization
- Transformation or transfection

Wheat Germ Extract

Product	Size	Cat.#
Wheat Germ Extract	5 × 200µl	L4380
Wheat Germ Extract Plus	40 × 50µl reactions	L3250
	10 × 50µl reactions	L3251

Rabbit Reticulocyte Lysate/Wheat Germ Extract Combination Systems

Product	Size	Cat.#
Rabbit Reticulocyte Lysate/Wheat Germ Extract Combination System	24 reactions	L4330

E. coli S30 Extracts

Product	Size	Cat.#
E. coli S30 Extract System for Circular DNA	30 × 50µl reactions	L1020
E. coli S30 Extract System for Linear Templates	30 × 50µl reactions	L1030
E. coli T7 S30 Extract System for Circular DNA	30 × 50µl reactions	L1130

Scale-up for structural biology:

Major experimental breakthrough:
Coupled transcription-translation systems (*Spirin, 1988)*
Continuous exchange cell-free (CECF) system

Continuous exchange boosts protein yields by:

Replenishing energy sources
Removing inhibitory reaction by-products
Extending translation period
Yields of **several mg/ml have been reported.**

A continuous cell-free protein synthesis system for coupled transcription-translation.
Kigawa T, Yokoyama S. J Biochem. 1991 Aug;110(2):166-8.

Gene expression in a cell-free system on the preparative scale.
Baranov VI, Morozov IYu, Ortlepp SA, Spirin AS. Gene. 1989 Dec 14;84(2):463-6.

The TNT Coupled Reticulocyte Lysate Systems is a single-tube, coupled transcription/translation system.

Standard rabbit reticulocyte lysate translations commonly use RNA synthesized in vitro from SP6, T3 or T7 RNA polymerase promoters and require three separate reactions with several steps between each reaction.

The TNT Systems bypass many of these steps by incorporating transcription directly in the translation mix.

The TNT Systems require plasmid constructs containing a prokaryotic phage RNA polymerase promoter (SP6, T3 or T7) for the initiation of transcription.

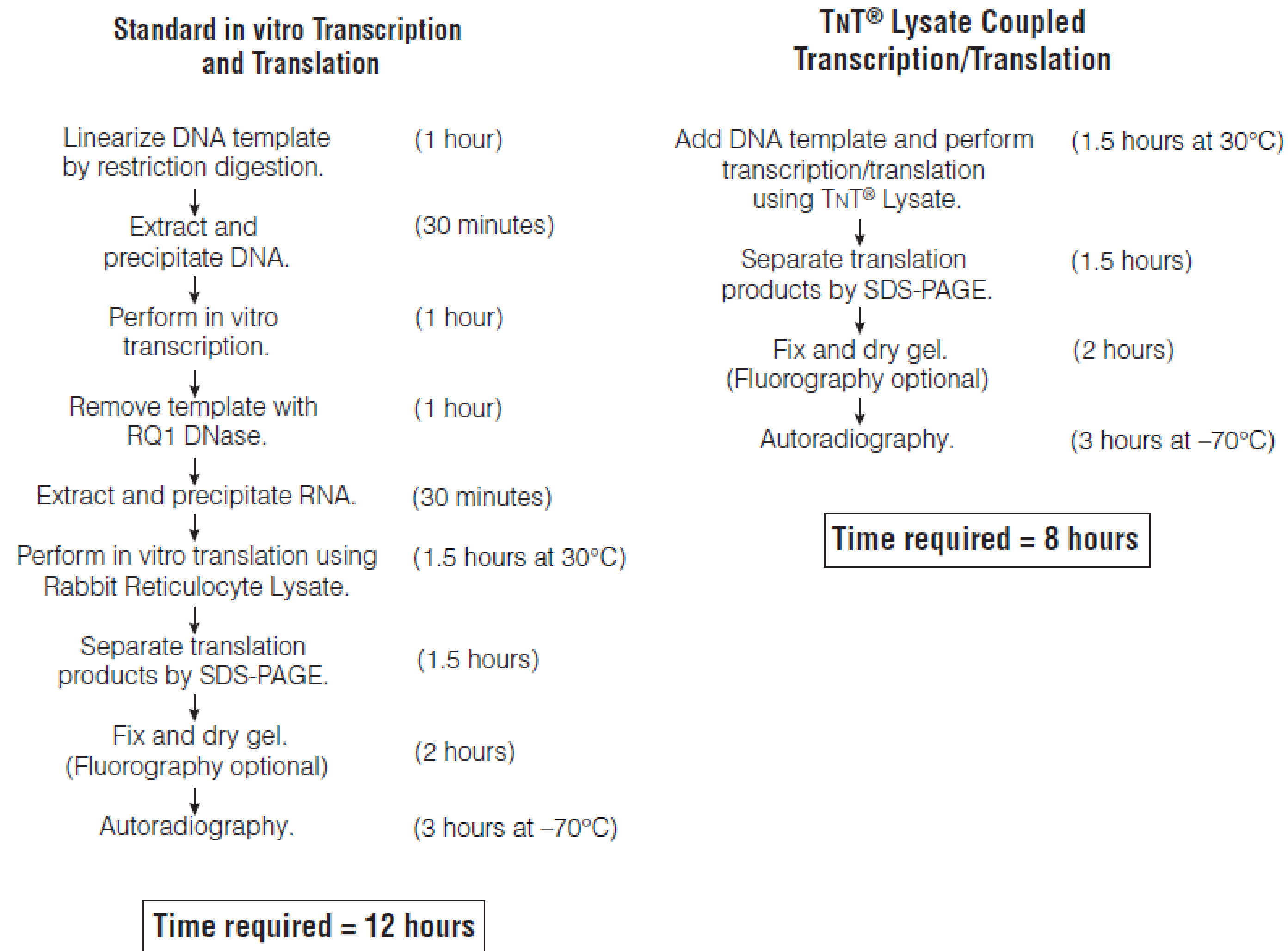
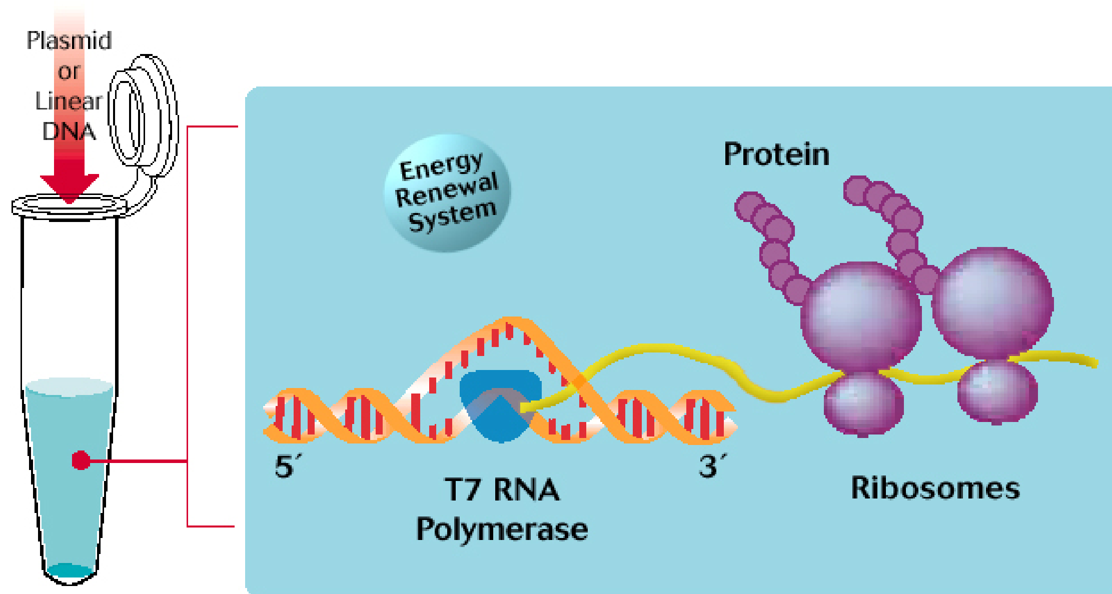


Figure 1. Comparison of standard in vitro transcription and translation procedures to the TNT® Lysate coupled transcription/translation protocol.

Expressway™

- A cell-free expression system optimized for production of full-length, active protein in a simple tube reaction
- Utilizes an *E. coli* lysate
- Coupled transcription/translation system



As the DNA template is transcribed, the 5’ end of the mRNA is bound by ribosome and undergoes translation.

www.invitrogen.com/expressway

http://www.roche-applied-science.com/PROD_INF/BIOCHEMI/no2_05/pdf/p33.pdf

Comparison of protein expression systems

Characteristics	E. coli	Yeast	Insect cells	Mammalian cells
Cell Growth	Rapid (30 Min)	Rapid (90 Min)	Slow (18-24 H)	Slow (24 H)
Complexity of Growth Medium	Minimum	Minimum	Complex	Complex
Cost of Growth Medium	Low	Low	High	High
Protein Folding	Refolding Usually Required	Refolding May Be Required	Proper Folding	Proper Folding
N-linked Glycosylation	None	High Mannose	Simple, No Sialic Acid	Complex
O-linked Glycosylation	No	Yes	Yes	Yes
Phosphorylation	No	Yes	Yes	Yes
Acetylation	No	Yes	Yes	Yes
Acylation	No	Yes	Yes	Yes
γ-Carboxylation	No	No	No	Yes
Yield (mg / L culture)	50-500	10-200	10-200	0.1-100
Cost	Low	Low	Middle	High
Advantage	Simple, robust, lowest cost, highest yield	Simple, low cost, good for certain proteins	Relatively higher yield, better PTM	Natural protein configuration, best PTM
Disadvantage	Least PTM	Longer time, less PTM	Longer time, higher cost	Highest cost, lower yield

PTM = Post-Translational Modification such as glycosylation

Viral-based vectors for gene expression in mammalian cells

Adenovirus
Adeno-associated virus
Baculovirus
Coronavirus
Epstein-Barr virus
Herpes simplex virus
Lentiviruses
Poliovirus
Retroviruses
Semliki Forest virus
Simian virus 40
Sindbis virus
Vaccinia virus