Module 2 overview

lecture	*lab*
1. Introduction to the module	1. Start-up protein eng.
2. Rational protein design	2. Site-directed mutagenesis
3. Fluorescence and sensors	3. DNA amplification
4. Protein expression	4. Prepare expression system
SPRING BREAK	
5. Review & gene analysis	5. Gene analysis & induction
6. Purification and protein analysis	6. Characterize expression
7. Binding & affinity measurements	7. Assay protein behavior
8. High throughput engineering	8. Data analysis

Lecture 4: Protein expression & purification

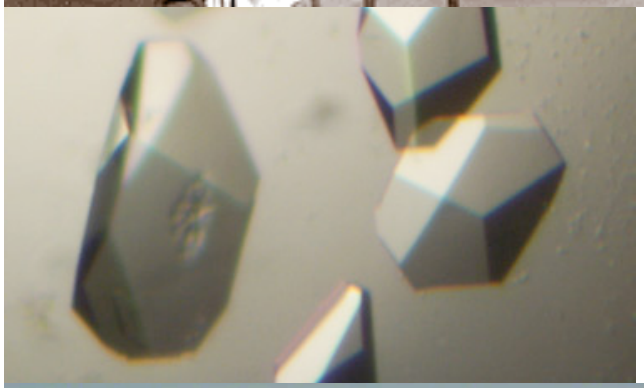
I. Why express & purify proteins?
   A. Scientific applications
   B. Applications in industry, *etc.*

II. Protein expression systems
   A. Alternatives to protein expression
   B. Prokaryotic and eukaryotic systems

Laboratory uses of purified proteins

Biochemistry analysis

Structural biology



Research biochemicals

www.mcgill.ca, images.apple.com,
www.varianinc.com, www.neb.com

Protein therapeutics

Table 1 Top ten recombinant therapeutic proteins and their global sales between 2001 and 2003

Product (generic)/ marketing company	2001 ($million)	2002 ($million)	2003 ($million)	Growth (decline) 2002–2003 (%)
Procrit (epoetin alfa)/ Johnson & Johnson	3,430	4,269	3,986	(6.6)
Epogen (epoetin alfa)/ Amgen	2,108	2,261	2,435	7.7
Neupogen (filgrastim)/ Amgen	1,346	1,380	1,268	(8.1)
Neulasta (pegfilgrastim)/ Amgen	0	464	1,255	170.5
Novolin (insulin systemic)/ Novo Nordisk	2,244	2,255	2,235	(0.9)
Avonex (interferon beta-1a)/ Biogen IDEC	971	1,034	1,170	13.2
PEG-Intron A franchise (pegylated interferon alpha)/ Schering Plough	1,447	2,736	1,851	(32.3)
Enbrel (etanercept)/ Amgen	856	521	1,300	149.5
Aranesp (darbepoetin alfa)/ Amgen	42	416	1,544	271.2
NeoRecormon (epoetin-beta)/ Roche	479	766	1,318	72.1
Top ten product sales	*12,923*	*16,102*	*18,362*	*14.0*
Others	*8,547*	*10,833*	*13,703*	*26.5*
Total market value	*21,470*	*26,935*	*32,065*	*19.0*

Source: Datamonitor and company-reported information.

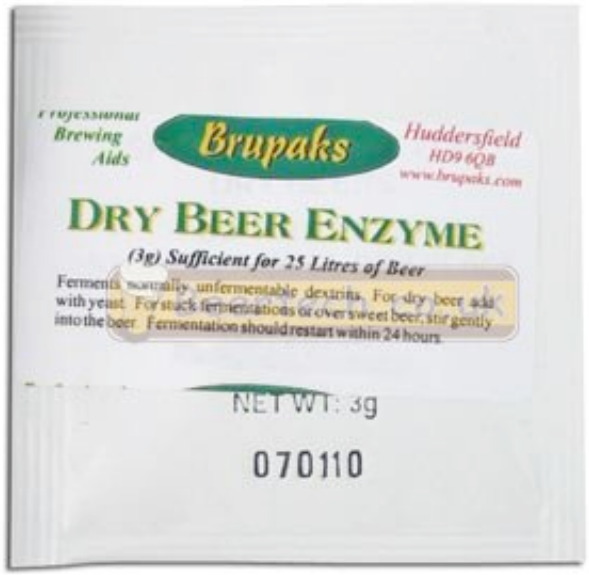
PEGylated

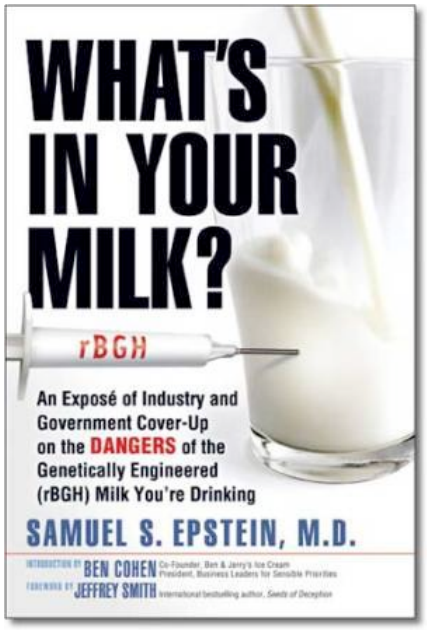
PEGylated

TNF ligand binding domain + Fc antibody domain

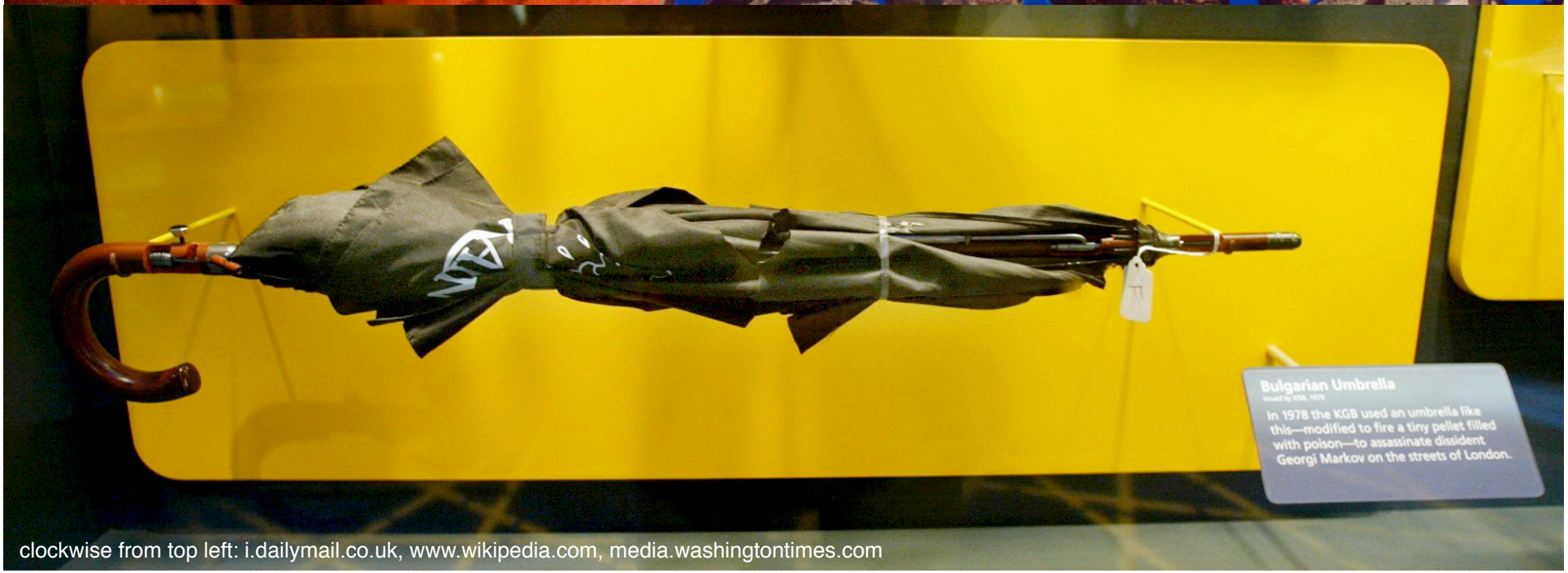
epo engineered to have additional glycoslyation sites

Pavlou & Reichert (2004) *Nat. Biotechnol.*





clockwise from top left: s.sears.com, www.beertech.co.uk, www.treatment-skin.com, www.valleynaturals.com, servekrishna.net



clockwise from top left: i.dailymail.co.uk, www.wikipedia.com, media.washingtontimes.com

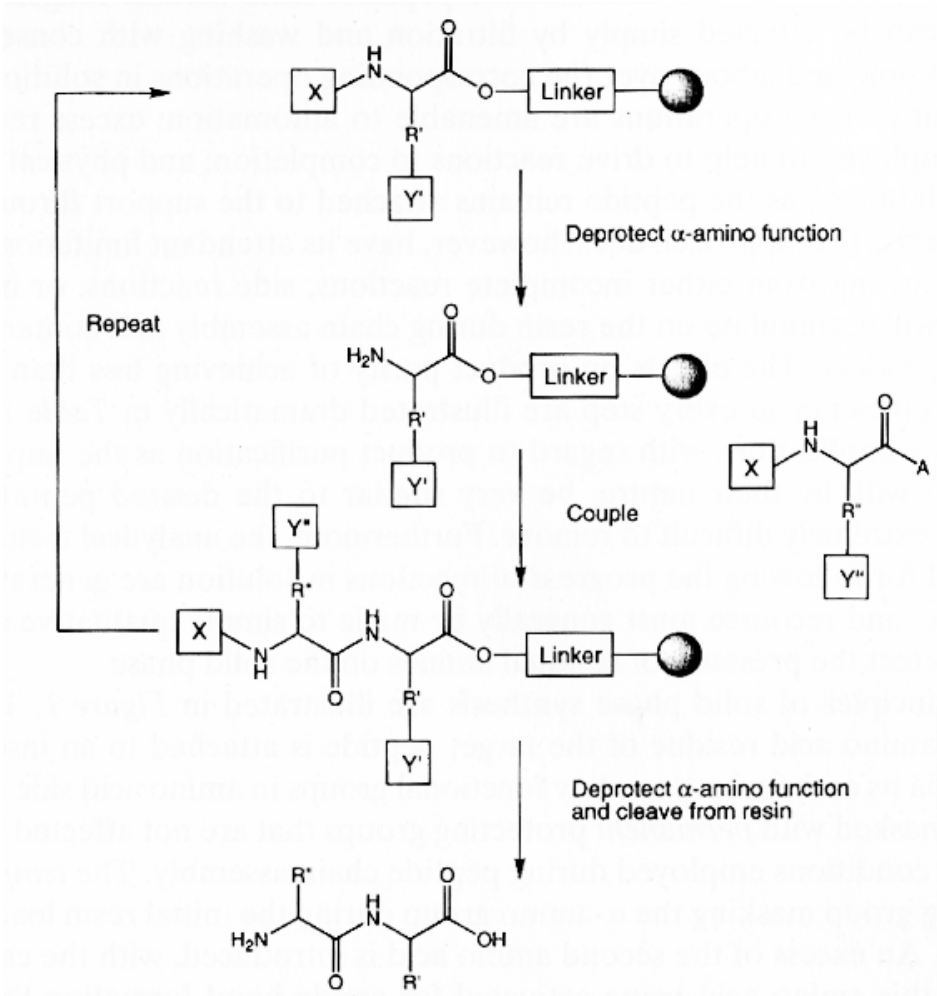
How can proteins be produced?

1. Purify from natural source
   advantages: no chemistry or DNA manipulation required, proteins likely to fold properly, assemble with native cofactors, *etc.*
   disadvantages: usually only practical for high abundance proteins, source-specific purification method required

2. Synthesize *de novo*
   advantages: no DNA manipulation required, synthesis methods well established, proteins produced are relatively pure
   disadvantages: relatively expensive, becomes extremely difficult for polypeptides > 50 amino acids

3. Express and purify from a dedicated expression system
   advantages: cheap and frequently high-yield, versatile expression systems already established
   disadvantages: cloning required, troubleshooting often needed to obtain high expression and proper folding

Solid phase peptide synthesis is a reliable technique for generating short polypeptides



X = Temporary amino protecting group
Y = Permanent side-chain protecting group
A = Carboxy activating group

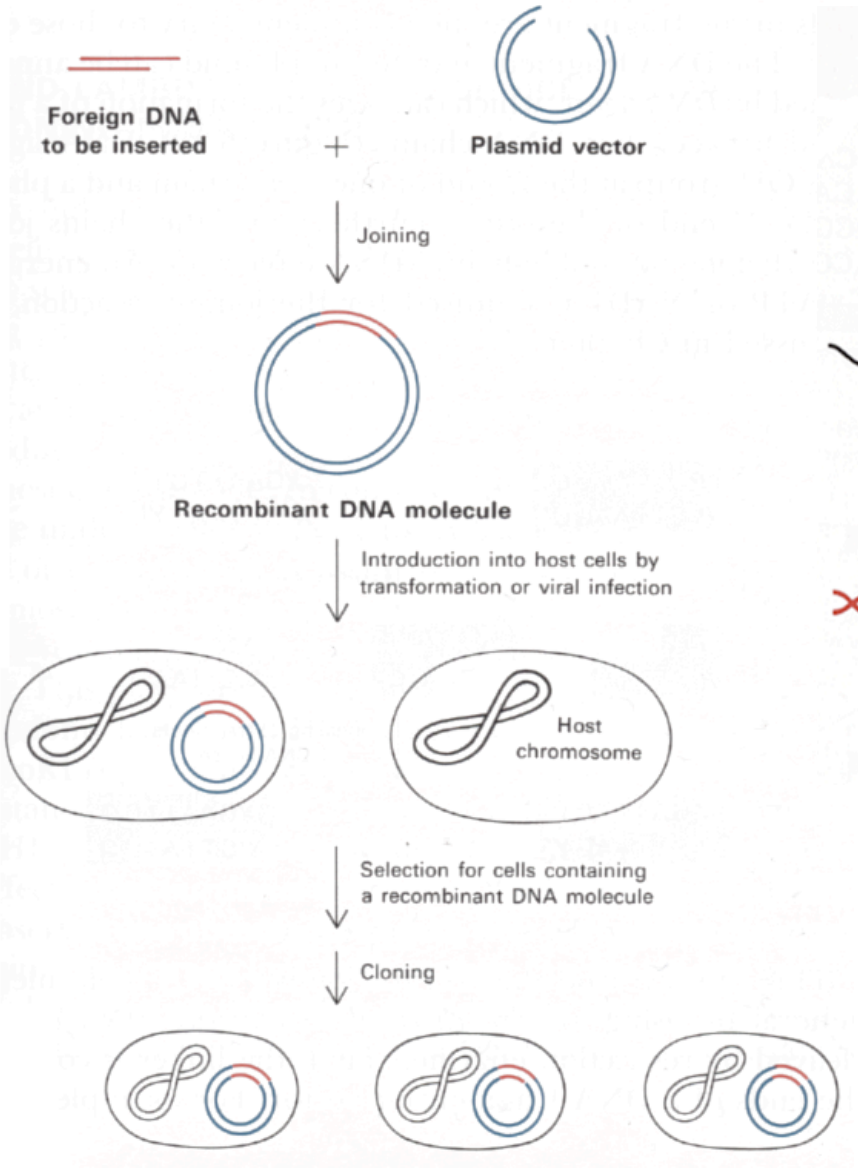
www.pitt.edu

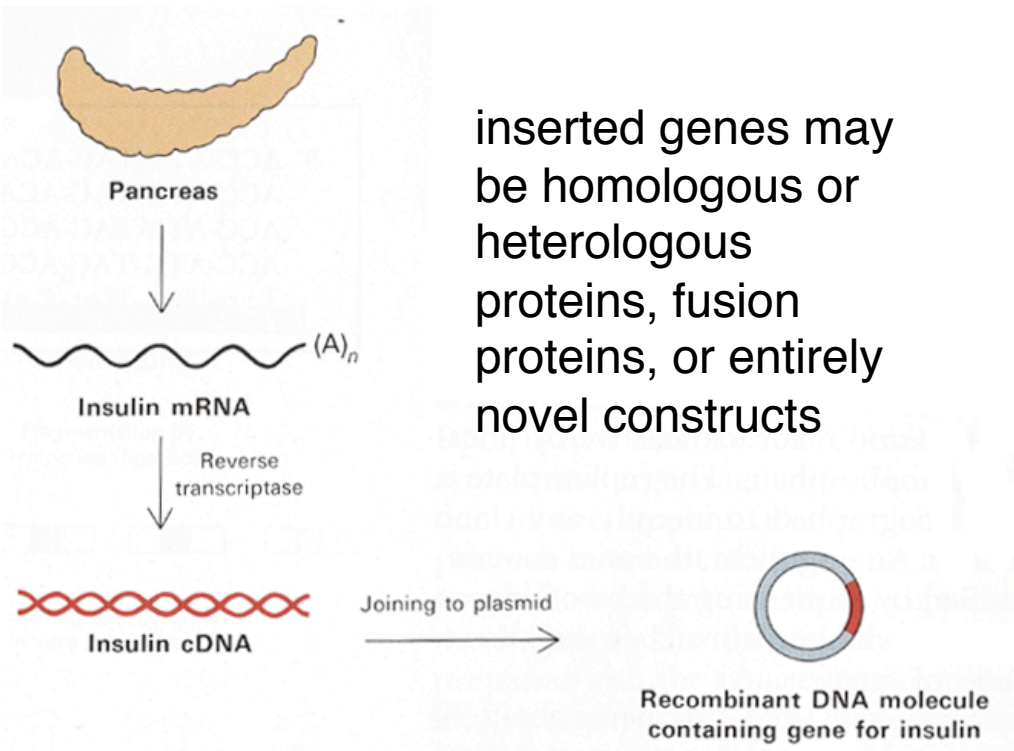
Table 1. Effects of accumulated errors on final product yields

No. of reactions		Yield of each reaction (%)			
		100	**99**	**95**	**90**
10	Overall yields	100	90	60	35
20		100	81	36	12
30		100	74	21	4
40		100	67	13	1
50		100	61	8	< 1

Chan & White (2000) *Fmoc Solid Phase Peptide Synthesis*

E. coli are the most common host for recombinant gene expression



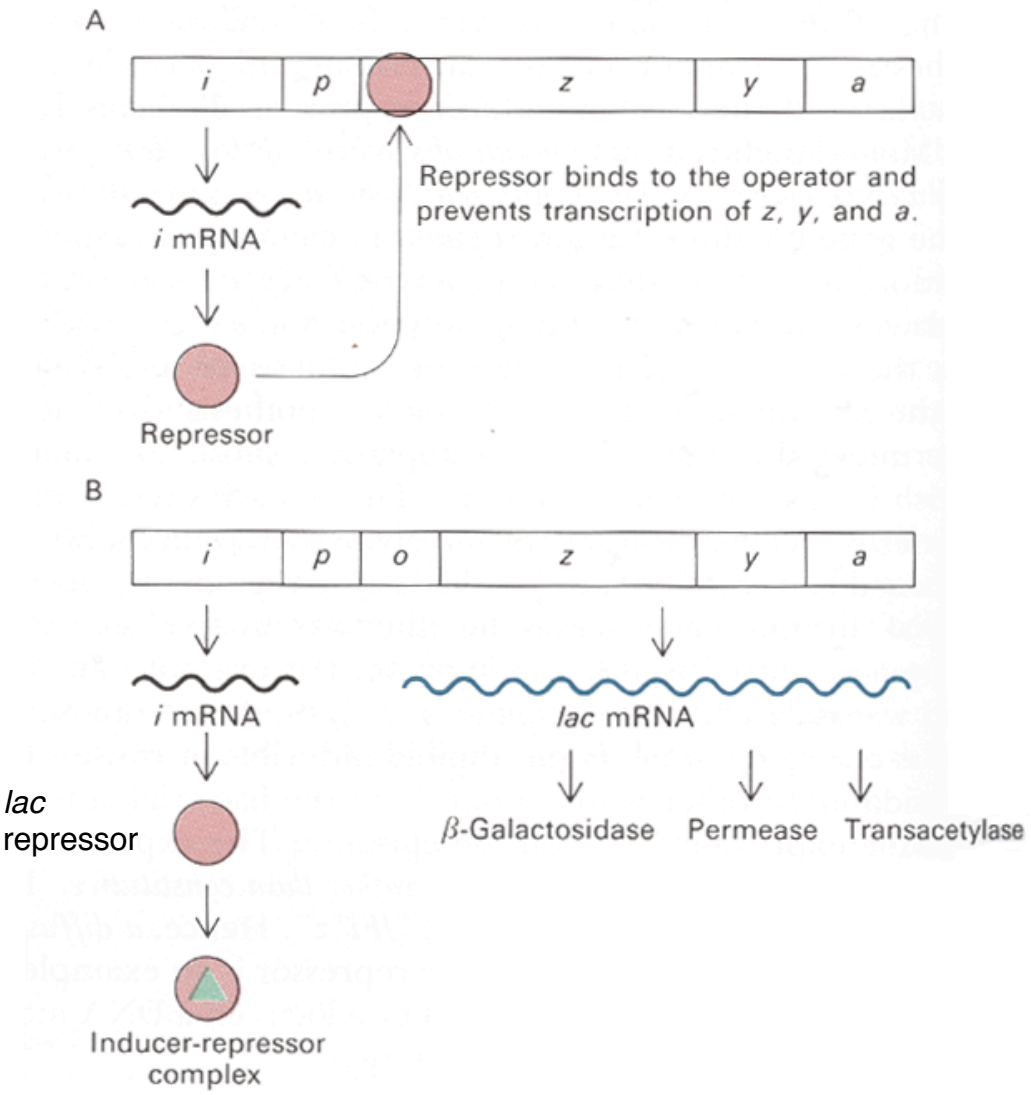


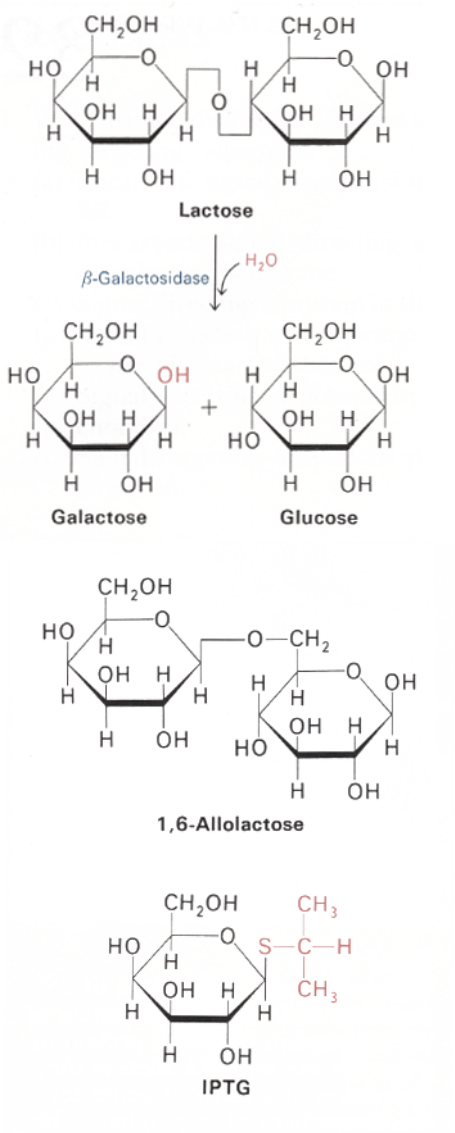
inserted genes may be homologous or heterologous proteins, fusion proteins, or entirely novel constructs

Once a foreign gene has been introduced, how does protein expression take place?

Stryer (1988) *Biochemistry, 3rd ed.*

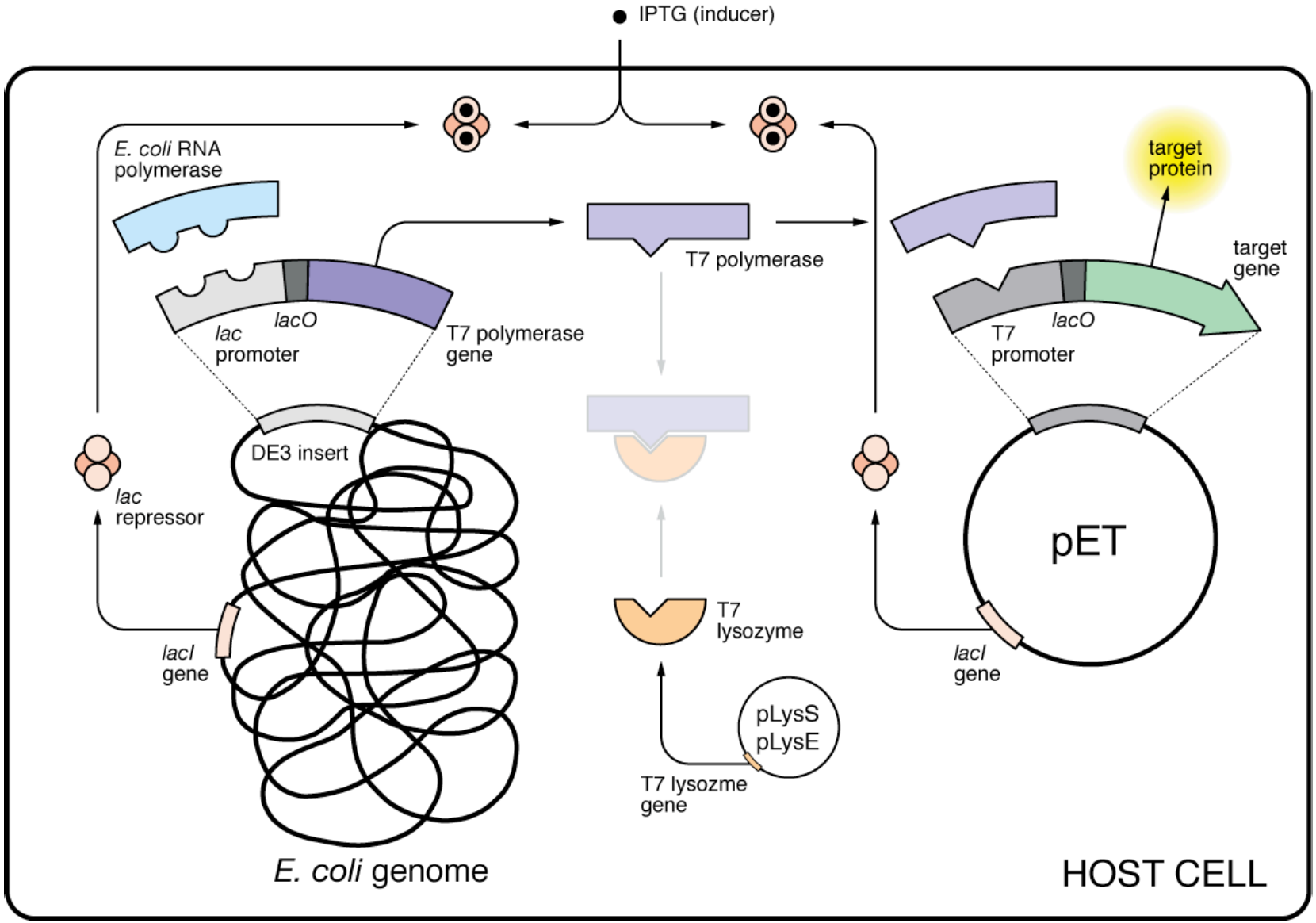
The *lac* operon is the basis for the most common bacterial protein expression systems



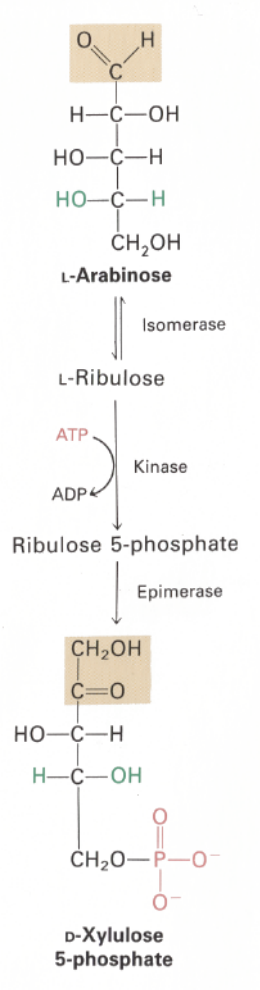


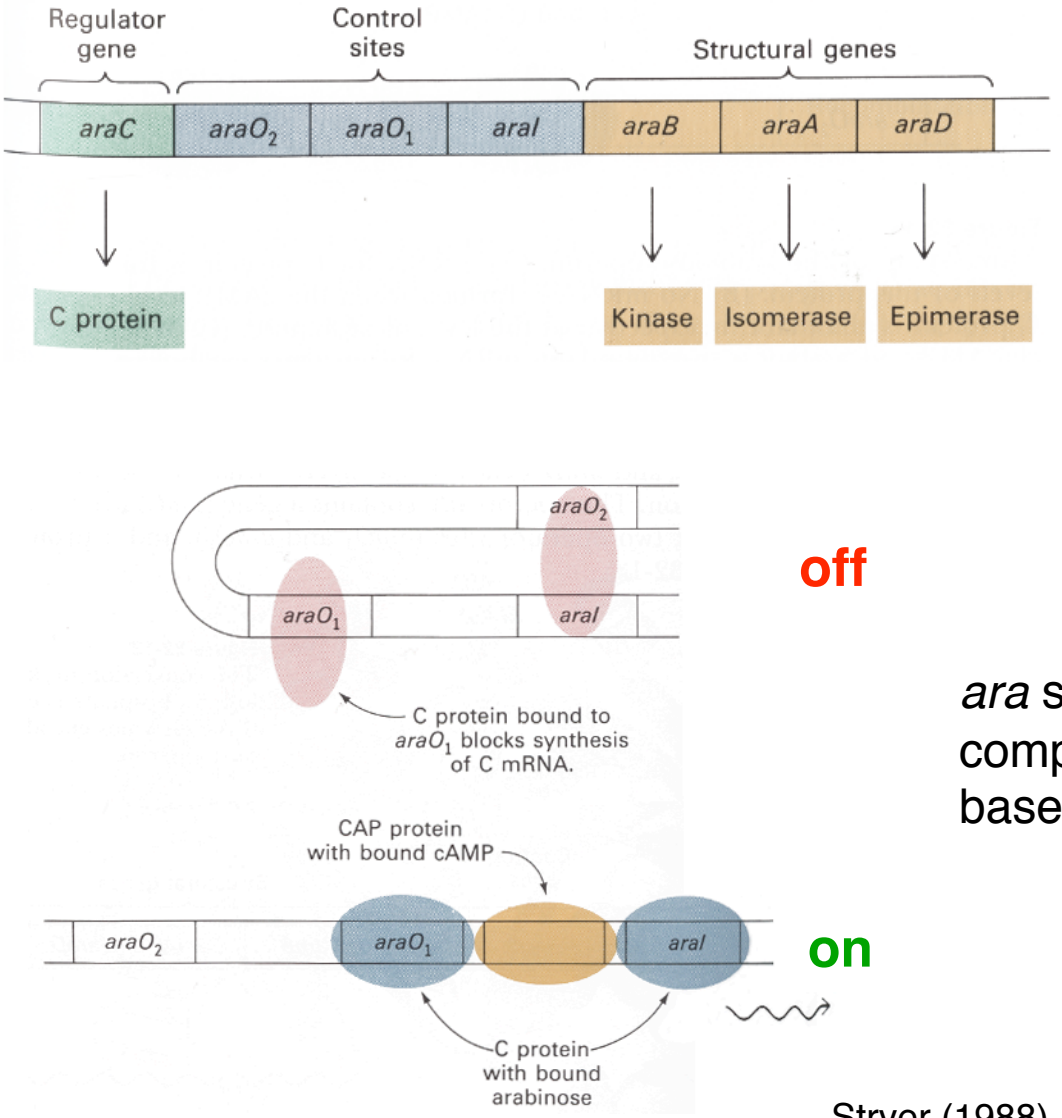
Stryer (1988) *Biochemistry, 3rd ed.*

T7 expression system



Other induction systems can also be used for protein expression in *E. coli:* arabinose system is considered to be more tightly controlled than the *lac* operon



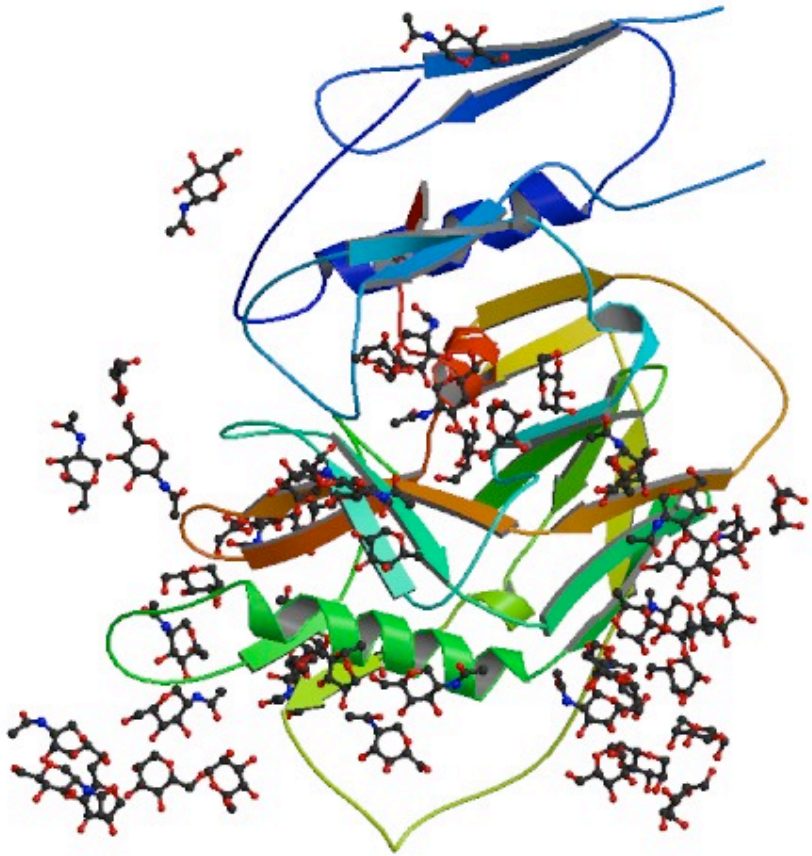


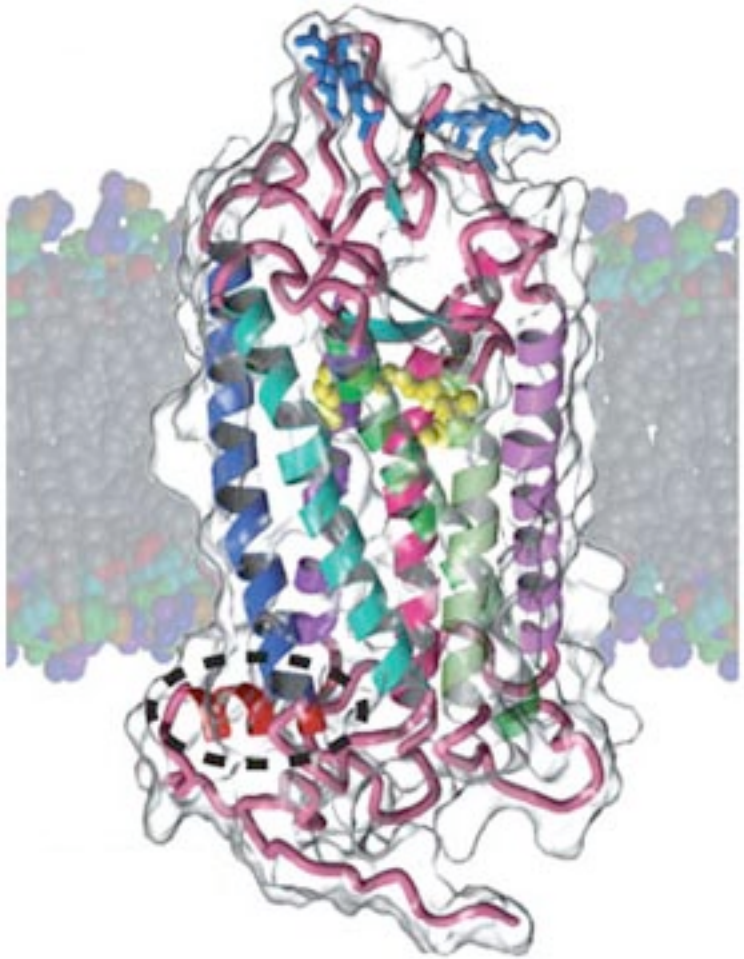
ara system is also compatible with T7-based vectors

Stryer (1988) *Biochemistry, 3rd ed.*

Differences between prokaryotic and eukaryotic proteins sometimes require eukaryotic expression systems.

These two proteins exemplify characteristics that frequently call for eukaryotic expression:

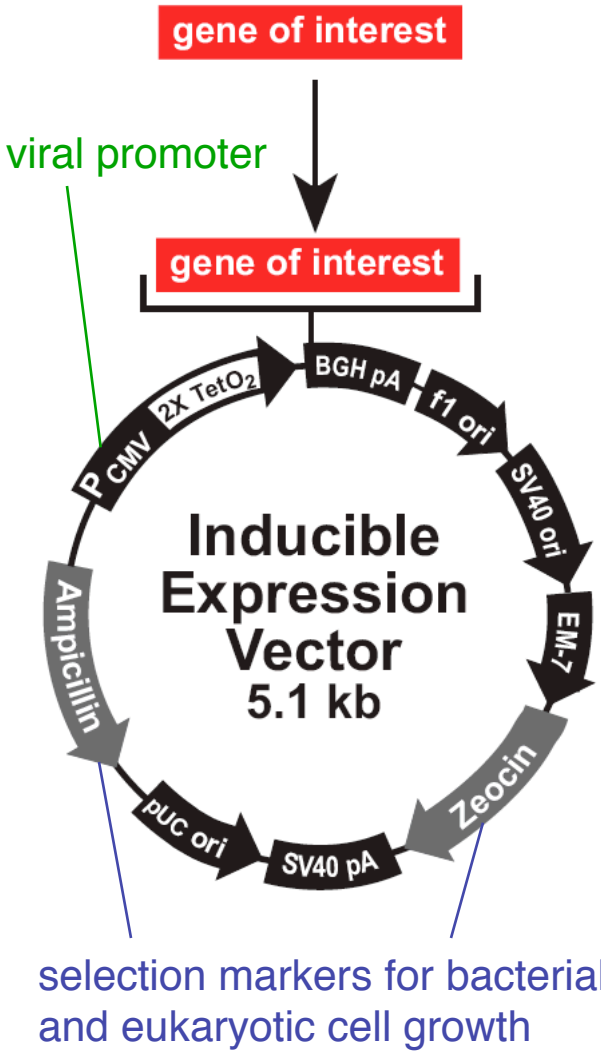




www.rcsb.pdb.org

www.rikenresearch.riken.jp

Eukaryotic expression vectors share features with bacterial systems

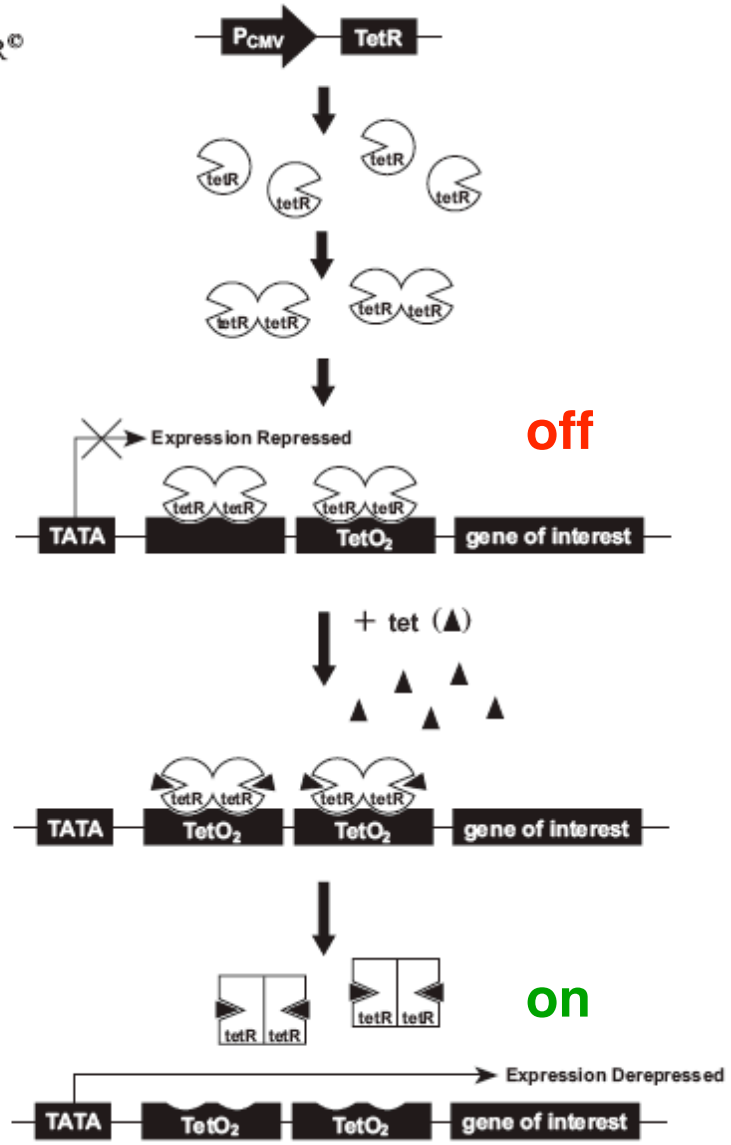


1. Tet repressor (tetR) protein is expressed from pcDNA6/TR© in cultured cells.

2. TetR homodimers bind to Tet operator 2 ($TetO_2$) sequences in the inducible expression vector, repressing transcription of the gene of interest.

3. Upon addition, tetracycline (tet) binds to tetR homodimers.

4. Binding of tet to tetR homodimers causes a conformational change in tetR, release from the Tet operator sequences, and induction of transcription from the gene of interest.



Invitrogen (2006) *T-REx System*

Prokaryotic *vs.* eukaryotic protein expression

property	*prokaryotic*	*higher eukaryotic*
yield/(L culture)	1-100 mg	widely variable
cost/(L medium)	~$5	~$50
introduction of DNA	transformation of competent cells	viral or nonviral transfection
handling	sterile needles, *etc*.	tissue culture hood
cell incubation	shaking incubator	usu. w/CO_2-control
induction	usually IPTG	none, tetracycline
glycosylation, *etc.*	no	yes
notes	best for small, globular proteins	best for complex, eukaryotic proteins