

Lecture 5 - Production of Recombinant Proteins I: Prokaryotic Hosts & Optimization

BENG168

Instructor: Adam M. Feist, Assistant Professor, Shu Chien - Gene Lay Department of Bioengineering

Prokaryotic Host Optimization

- **Production Goals:** The primary objectives of commercial protein production are achieving high quality, high yields, and low cost.
- **System Choice:** Prokaryotic systems are preferred because their genetics are well understood and they grow fast in inexpensive media.
- **Metabolic Efficiency:** Optimization efforts focus on minimizing the metabolic burden on the host to ensure continued cell viability.



Module 1: Genetic Regulation and Translational Optimization (Source Pages: Chapter 2: 91–100)

- Targeting high quality and high protein yields.
- Regulating host transcriptional systems.
- Overcoming host codon bias.

Why Prokaryotic Hosts?

- **Genetic Knowledge:** The genetics, molecular biology, and physiology of common bacteria like *E. coli* are exceptionally well characterized.
- **Industrial Scale-up:** These hosts can be grown to **high cell densities** in large-scale bioreactors using relatively inexpensive growth medium.
- **Optimized Vectors:** Commercial vectors are available that carry strong signals for high levels of both transcription and translation.
- **Protein Limitations:** Bacteria cannot perform post-translational glycosylation, a concept we introduced in Lecture 3 when discussing PTMs.

Table 3.1 Production of recombinant human proteins in various biological hosts

Parameter	Bacteria	Yeast	Mammalian cell culture	Transgenic plants
Glycosylation	None	Incorrect	Correct	Generally correct; small differences
Multimeric proteins assembled	Limited	Limited	Limited	Yes
Production costs	Low-medium	Medium	High	Very low
Protein folding accuracy	Low	Medium	High	High
Protein yield	High	Medium-high	Low-medium	Medium
Scale-up capacity	High	High	Low	Very high
Scale-up costs	High	High	High	Low
Time required	Low	Low-medium	Medium-high	High
Skilled workers required	Medium	Medium	High	Low
Acceptable to regulators	Yes	Yes	Yes	Not yet

Transcriptional Requirements

- **Strong Promoters:** High-level expression requires a promoter with high affinity for RNA polymerase.
- **Consensus Sequences:** Effective promoters closely resemble the –35 and –10 box consensus sequences.
- **Regulatable Control:** Promoters must be "switchable" to prevent energy drain and plasmid loss during growth.

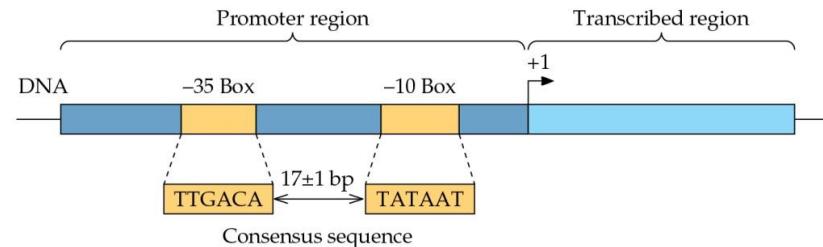


Figure 3.1 A strong *E. coli* promoter resembles the consensus sequences for the –35 and –10 boxes that bind to RNA polymerase. The consensus sequence was determined by aligning many *E. coli* promoters and identifying conserved nucleotides in the sequences centered –35 and –10 bp upstream of the transcription start site (+1). The distance, but not the nucleotide sequence, between the two boxes is also conserved.

The Need for Regulation

- **Resource Drain:** Continuous high-level expression creates an energy drain that impairs essential host cell functions and overall growth.
- **Metabolic Burden:** Overproduction can deplete pools of aminoacyl-tRNAs and decrease cellular ATP or GTP levels significantly.
- **Plasmid Stability:** Without regulation, cells lacking plasmids grow faster and eventually take over the culture, reducing the final yield.

Table 3.3 Factors that may increase the metabolic burden on a prokaryotic host cell that is expressing high levels of a cloned gene

Condition	Effect(s)
High plasmid copy number/size	Requires high amounts of cellular energy and nucleotides for plasmid replication and maintenance; decreases growth rate
Overproduction of target and marker proteins	Depletes the pools of certain aminoacyl-tRNAs and amino acids; decreases cellular ATP or GTP levels
Secretion of high levels of target proteins	Jams export sites; prevents proper localization of essential host proteins
Production of foreign enzymes	May catabolize essential metabolites or produce toxic products
Depletion of dissolved oxygen in medium	Decreases metabolic activity and expression of cloned gene

The lac System – Widely Used

- **Negative Control:** The LacI repressor protein binds to the operator to prevent transcription in the absence of an inducer.
- **Chemical Induction:** Transcription is turned on by adding lactose or the synthetic, non-metabolizable inducer IPTG to the culture medium.
- **Positive Regulation:** The catabolite activator protein (CAP) binds to the DNA to increase promoter affinity for RNA polymerase.
- **Earlier Tools:** We used this system for "Blue/White Screening" in Lecture 2; now we use it for production.

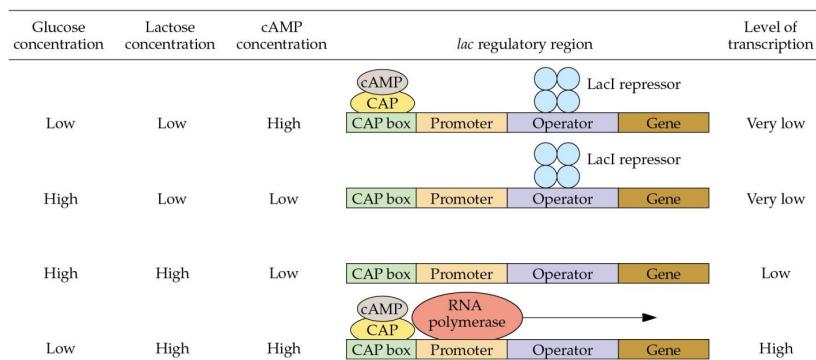


Figure 3.2 The effects of the concentrations of glucose and lactose in the growth medium on the level of transcription from the *E. coli* lac promoter. The cAMP concentration refers to intracellular levels that are regulated by glucose. The LacI repressor is a tetramer that binds to the lac operator in the absence of lactose and prevents transcription. In the presence of cAMP, positive regulator CAP binds to a CAP recognition site (CAP box) on the DNA and activates high levels of transcription. The arrow indicates the direction of transcription.

Glucose concentration	Lactose concentration	cAMP concentration	<i>lac</i> regulatory region	Level of transcription
Low	Low	High	<p>The diagram shows the <i>lac</i> regulatory region consisting of a CAP box (green), Promoter (orange), Operator (purple), and Gene (yellow). A CAP protein dimer (labeled cAMP-CAP) is shown bound to the CAP box. A LacI repressor tetramer (four blue circles) is shown bound to the Operator. An arrow points from the Gene towards the right.</p>	Very low
High	Low	Low	<p>The diagram shows the <i>lac</i> regulatory region consisting of a CAP box (green), Promoter (orange), Operator (purple), and Gene (yellow). A LacI repressor tetramer (four blue circles) is shown bound to the Operator. An arrow points from the Gene towards the right.</p>	Very low
High	High	Low	<p>The diagram shows the <i>lac</i> regulatory region consisting of a CAP box (green), Promoter (orange), Operator (purple), and Gene (yellow). A CAP protein dimer (labeled cAMP-CAP) is shown bound to the CAP box. An arrow points from the Gene towards the right.</p>	Low
Low	High	High	<p>The diagram shows the <i>lac</i> regulatory region consisting of a CAP box (green), Promoter (orange), Operator (purple), and Gene (yellow). A CAP protein dimer (labeled cAMP-CAP) is shown bound to the CAP box. A large red oval labeled "RNA polymerase" is shown bound to the Promoter. An arrow points from the Gene towards the right.</p>	High

Figure 3.2 The effects of the concentrations of glucose and lactose in the growth medium on the level of transcription from the *E. coli lac* promoter. The cAMP concentration refers to intracellular levels that are regulated by glucose. The LacI repressor is a tetramer that binds to the *lac* operator in the absence of lactose and prevents transcription. In the presence of cAMP, positive regulator CAP binds to a CAP recognition site (CAP box) on the DNA and activates high levels of transcription. The arrow indicates the direction of transcription.

Engineered Hybrid: tac Promoter

- **Functional Fusion:** Combines the –35 region of the *trp* promoter and the –10 region of *lac*.
- **Increased Power:** The *tac* promoter is 10x stronger than *lac* while remaining regulatable by IPTG.
- **Commercial Utility:** It is widely used in expression vectors to drive high yields of non-toxic proteins.

A “toolkit” for use in prokaryotic hosts

Table 3.2 Promoters commonly used for expression of cloned genes in prokaryotic hosts

Promoter	Source	Regulator(s)	Features
<i>lac/lacUV5</i>	<i>E. coli</i> lactose operon	Lactose repressor LacI, CAP activator	Induced by lactose (allolactose) or IPTG
Gene 10	Bacteriophage T7	T7 RNA polymerase	Strong promoter
<i>trp</i>	<i>E. coli</i> tryptophan operon	Tryptophan repressor TrpR	Repressed by tryptophan
<i>tet</i>	<i>E. coli</i> Tn10 tetracycline resistance gene	Tetracycline repressor TetR	Induced by tetracycline
<i>tac</i>	Synthetic hybrid of <i>lacUV5</i> and <i>trp</i> promoters	LacI repressor	Induced by IPTG; approximately 10 times stronger than the <i>lac</i> promoter
<i>araBAD</i>	<i>E. coli</i> arabinose operon	Arabinose activator/repressor AraC	Induced by arabinose
<i>pL</i>	Bacteriophage lambda	λ cI repressor/ mutant λ cI ⁸⁵⁷ repressor	Induced by nalidixic acid, which elicits the SOS response; temperature sensitive λ cI ⁸⁵⁷ repressor is unstable above 37°C
<i>cspA</i>	<i>E. coli</i> cold shock gene		Induced at 15°C
<i>nisA</i>	<i>Lactococcus lactis</i> nisin operon	Two component regulators NisK and NisR	Induced by nisin
<i>luxI</i>	<i>Vibrio fischeri</i> luciferase operon	Activator LuxR	Acylhomoserine lactone synthesized by LuxI at high cell densities

T7 Expression System

- **Fast Transcription:** This system uses the T7 RNA polymerase, which transcribes at a rate eight times faster than standard *E. coli* polymerase.
- **Double Switch:** The T7 polymerase gene is typically placed under the control of a lac promoter to regulate its own synthesis.
- **High Specificity:** T7 polymerase only recognizes the T7 gene 10 promoter, ensuring that cellular resources are focused on the target gene.

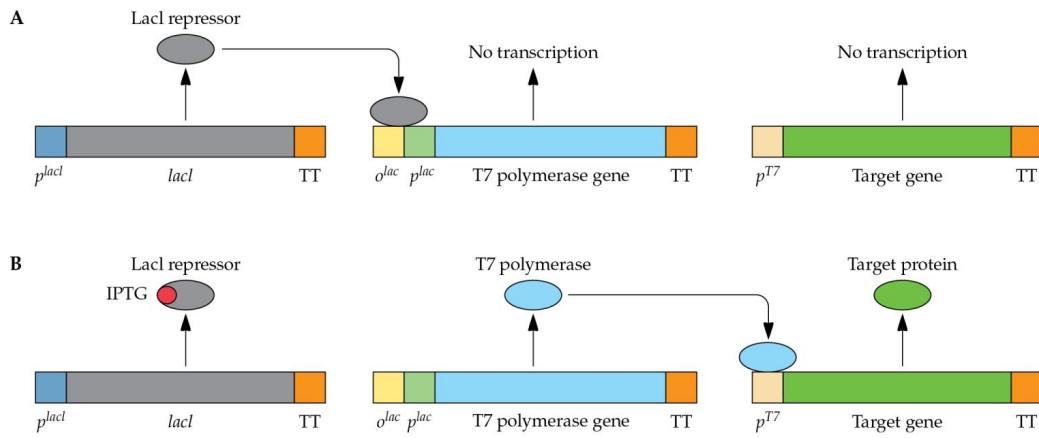


Figure 3.4 Regulation of target gene expression controlled by the promoter for gene 10 from bacteriophage T7 (p^{T7}). **(A)** In the absence of the inducer IPTG, the constitutively produced LacI repressor, the product of the *lacI* gene, which is under the control of the *lacI* promoter (p^{lacI}), represses the synthesis of the T7 RNA polymerase that is transcriptionally controlled by the *lac* operator (o^{lac}) and *lac* promoter (p^{lac}). In the absence of T7 RNA polymerase, the target gene, which is under the transcriptional control of the T7 gene 10 promoter (p^{T7}), is not transcribed. **(B)** When IPTG (or lactose) is added to the medium, it binds to the LacI repressor, thereby preventing it from repressing the transcription of T7 RNA polymerase. In the presence of T7 RNA polymerase, the target gene is transcribed. TT, transcription termination sequence.

Ribosome-Binding Site (RBS) Optimization

- **Sequence Interaction:** The RBS is a 6 to 8 nucleotide sequence that base pairs with the 16S RNA of the small ribosome.
- **Distance Requirement:** The RBS must be located within a specific short distance (2 to 20 nucleotides) from the translational AUG start codon.
- **Binding Strength:** Greater efficiency in translation initiation is generally achieved by increasing the binding strength between the mRNA and the ribosome.

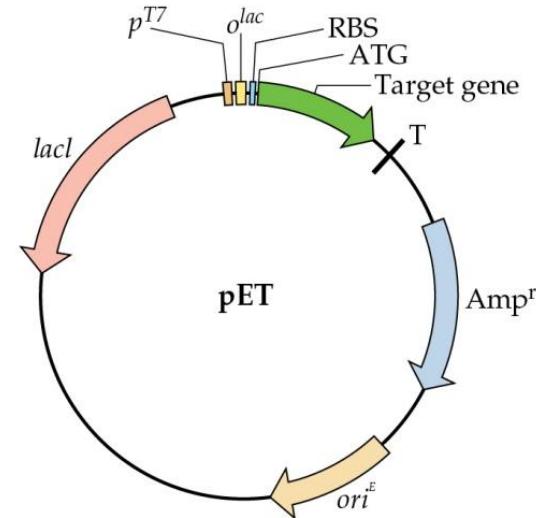


Figure 3.6 The expression vector pET. A target gene is inserted into the vector such that it is in the same reading frame as a start codon (ATG) present on the vector. Transcription of the target gene is controlled by the strong T7 gene 10 promoter (p^{T7}) and terminator (T), and translation is initiated from the high-affinity ribosome-binding site (RBS) from the T7 major capsid protein gene. In the absence of IPTG, the LacI repressor encoded by *lacI* binds to the *lac* operator sequence (o^{lac}) and blocks transcription of the target gene (see Fig. 3.4). The LacI repressor also prevents transcription of the T7 RNA polymerase gene that is integrated into the chromosome of host *E. coli* cells (not shown). The ampicillin resistance gene (Amp^r) is a selectable marker gene that confers resistance to the antibiotic ampicillin, and the origin of replication (ori^E) enables replication of the vector in *E. coli*.

Overcoming mRNA Secondary Structure

- **Inhibitory Folding:** Intrastrand base pairing at the 5' end of the mRNA can create structures that physically shield the ribosome-binding site.
- **Access Blockage:** These secondary structures prevent the small ribosomal subunit from attaching to the mRNA and initiating the translation process.
- **Fine-Tuning:** Engineers often must modify the first few codons of a gene to minimize these structures.

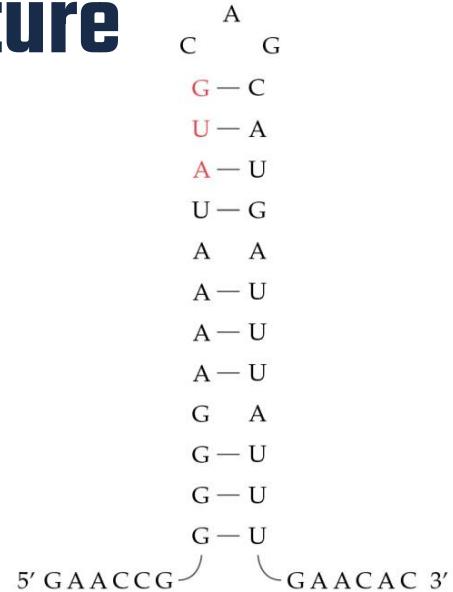


Figure 3.5 Example of the secondary structure of the 5' end of an mRNA that would prevent efficient translation. The ribosome-binding site is GGGGG, the start codon is AUG (shown in red), and the first few codons are CAG-CAU-GAU-UUA-UUU. Note that in addition to the traditional A·U and G·C base pairs in mRNA, G can also base pair to some extent with U.

Table 3.4 Genetic code and codon usage in *E. coli* and humans

Codon Usage Optimization

- **Organism Preferences:** Different organisms prefer different codons for the same amino acid, a phenomenon known as "codon bias".
- **Translation Stalling:** If a gene uses codons rare to the host, translation may stall or incorporate incorrect amino acids due to tRNA depletion.
- **Engineering Solutions:** We can either chemically resynthesize the gene with host-preferred codons or engineer the host to overexpress rare tRNAs.
- **Enabling Tech:** We use the "Functional Prediction" and "Annotation" skills from Lecture 3 to identify these rare codons

Codon	Amino acid	Frequency of use in:	
		<i>E. coli</i>	Humans
GGG	Glycine	0.13	0.23
GGA	Glycine	0.09	0.26
GGU	Glycine	0.38	0.18
GGC	Glycine	0.40	0.33
GAG	Glutamic acid	0.30	0.59
GAA	Glutamic acid	0.70	0.41
GAU	Aspartic acid	0.59	0.44
GAC	Aspartic acid	0.41	0.56
GUG	Valine	0.34	0.48
GUU	Valine	0.17	0.10
GUU	Valine	0.29	0.17
GUC	Valine	0.20	0.25
GCG	Alanine	0.34	0.10
GCA	Alanine	0.22	0.22
GCU	Alanine	0.19	0.28
GCC	Alanine	0.25	0.40
AAG	Lysine	0.24	0.60
AAA	Lysine	0.76	0.40
AAU	Asparagine	0.39	0.44
AAC	Asparagine	0.61	0.56
AUG	Methionine	1.00	1.00
AUA	Isoleucine	0.07	0.14
AUU	Isoleucine	0.47	0.35
AUC	Isoleucine	0.46	0.51
ACG	Threonine	0.23	0.12
ACA	Threonine	0.12	0.27
ACU	Threonine	0.21	0.23
ACC	Threonine	0.43	0.38
UGG	Tryptophan	1.00	1.00
UGU	Cysteine	0.43	0.42
UGC	Cysteine	0.57	0.58
UGA	Stop	0.30	0.61

Codon Usage Optimization

- Table 3.5 shows the impact of chemically resynthesis of the gene with host-preferred codons

Table 3.5 Increases in gene expression that result from altering the codon usage of the wild-type gene (or cDNA) to more closely correspond to the host *E. coli* cell^a

Protein	Improvement (fold)
Human interleukin-2	16
<i>Clostridium tetani</i> tetanus toxin fragment C	4
Human cardiac troponin T	10–40
Mouse c-FOS protein	>200
Spinach plastocyanin	1.2
Human neurofibromin	3
Human glutathione transferase M2-2	140
Human phosphatidylcholine transfer protein	>100
Human interleukin-6	3
Human interleukin-18	5
<i>Plasmodium</i> vaccine candidate antigen	4

^aIn some cases, only a small number of codons were altered, while in others, the entire gene was synthesized with the optimal codon usage for expression in *E. coli*.

Codon Usage Optimization

- An example of an approach to engineer the host to overexpress rare tRNAs

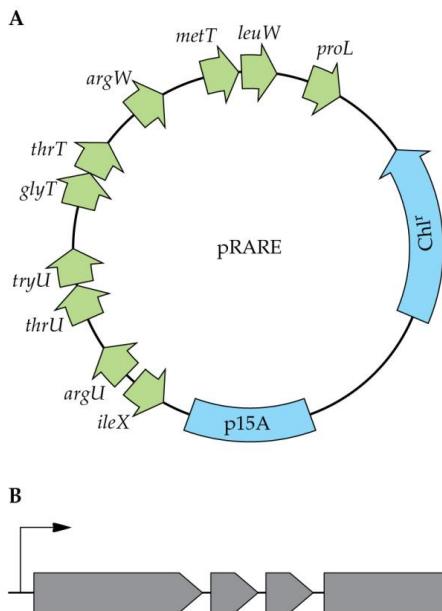


Figure 3.7 Two commercially available systems to increase the pool of certain less abundant tRNAs in *E. coli*. **(A)** Plasmid pRARE carries genes encoding 10 rare *E. coli* tRNAs. The plasmid also has an origin of replication (p15A) for replication in *E. coli* and a gene encoding resistance to the antibiotic chloramphenicol (Chl^r). **(B)** Genes encoding six rare tRNAs (red genes) were integrated into one of the ribosomal RNA operons (gray genes) in the chromosome of an *E. coli* strain (SixPack). The tRNA genes were cloned into a plasmid and integrated into the bacterial chromosome by homologous recombination. Transcription is driven by the strong ribosomal RNA operon promoter (arrow), and abundance of the tRNAs increases during the exponential phase of growth. Data from Lipinszki et al., ACS Synth. Biol. 7:2656–2664, 2018.

Video - Concepts in the module or a demonstration

- Prokaryotic Gene Regulation: Lac Operon (3min53s)

More videos

- How Codon Optimization is Changing Lives (2min37s)

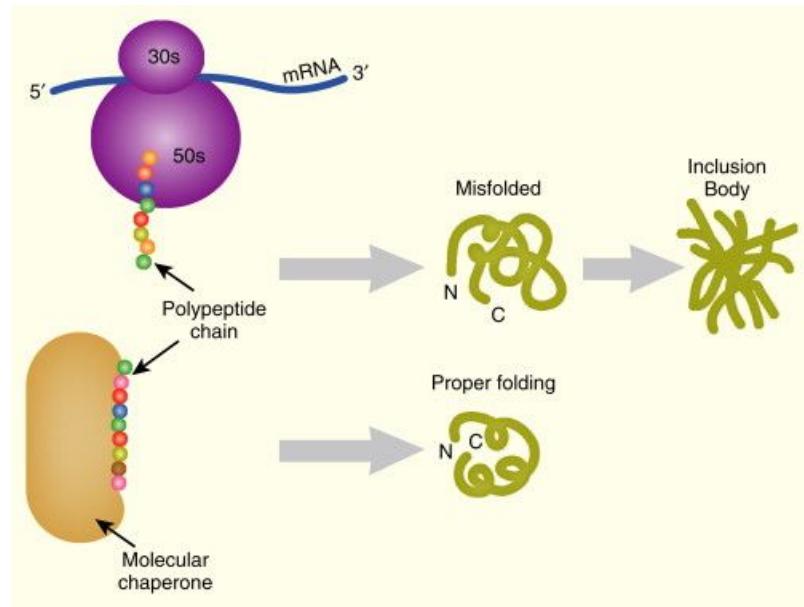
Module 2: Protein Stability, Secretion, and Recovery

(Source Pages: Chapter 2: 100–119)

- Preventing inclusion body formation.**
- Engineering protein secretion pathways.**
- Facilitating rapid affinity purification.**

The Challenge of Inclusion Bodies

- **Protein Aggregation:** Overexpression often leads to the formation of inclusion bodies, which are insoluble and inactive aggregates of misfolded proteins.
- **Host Environment:** Misfolding occurs when the host environment (pH, redox status) differs too much from the protein's native environment.
- **Recovery Hurdles:** Recovering functional protein from these bodies requires difficult and expensive solubilization & refolding steps in the lab.



Molecular Chaperones Help Prevent Inclusion Bodies

If recombinant proteins fold aberrantly, they aggregate and form inclusion bodies. If, during translation, molecular chaperone proteins hold the recombinant protein, the new polypeptide will fold properly and become fully functional.

Facilitating Proper Folding - Molecular Chaperones

- **Folding Assistance:** Chaperones are specialized proteins that assist in the correct three-dimensional folding of nascent polypeptide chains.
- **Diverse Roles:** "Folding" chaperones use ATP to force refolding, while "holding" chaperones prevent aggregation until folding can occur.
- **Coexpression Benefits:** Overexpressing these chaperones alongside our target protein can significantly increase the yield of soluble, active product. e.g., 5-fold

Table 3.6 *E. coli* proteins that facilitate the correct folding of recombinant proteins^a

Localization	Function	Name
Cytoplasm	Holding chaperone	Hsp31 Hsp33 IbpA IbpB Trigger factor
	Folding chaperone	GroEL (Hsp60) DnaK (Hsp70) HscA HscC HtpG (Hsp90)
	Disaggregase	ClpB
	Secretory chaperone	SecB
Periplasm	Generic chaperones	Skp (OmpH) FkpA
	Specialized chaperones	SurA LolA PapD FimC
	Peptidyl-prolyl isomerases	SurA PpiD FkpA PpiA (RotA)
	Disulfide bond formation	DsbA DsbB DsbC DsbD DsbE DsbG CcmH

^aData from Baneyx and Mujacic, *Nat. Biotechnol.* 22:1399–1408, 2004.

Disulfide Bond Engineering

- **Structural Stability:** Many eukaryotic proteins require covalent disulfide bonds between cysteine residues to achieve their stable, active configuration.
- **Redox Environment:** The bacterial cytoplasm is typically a reducing environment that prevents these vital oxidative bonds from forming naturally.
- **Host Engineering:** Specific host strains (e.g., *gor* and *trxR* mutants) can be engineered to create an oxidizing cytoplasm suitable for bond formation.

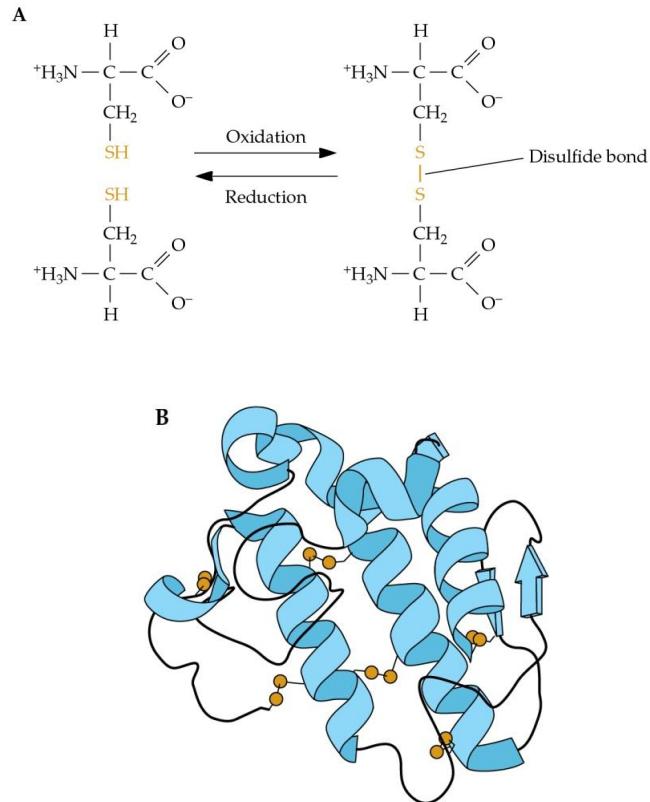


Figure 3.9 Disulfide bond in a protein. **(A)** A covalent, disulfide bond forms by oxidation of sulfhydryl (SH) groups on cysteines. **(B)** Disulfide bonds between cysteines (represented by brown circles) within a polypeptide (ribbon diagram) contribute to the structural stability of the protein.

Disulfide Bond Engineering

- **Structural Stability:** Many eukaryotic proteins require covalent disulfide bonds between cysteine residues to achieve their stable, active configuration.
- **Redox Environment:** The bacterial cytoplasm is typically a reducing environment that prevents these vital oxidative bonds from forming naturally.
- **Host Engineering:** Specific host strains (e.g., *gor* and *trxB* mutants) can be engineered to create an oxidizing cytoplasm suitable for bond formation.

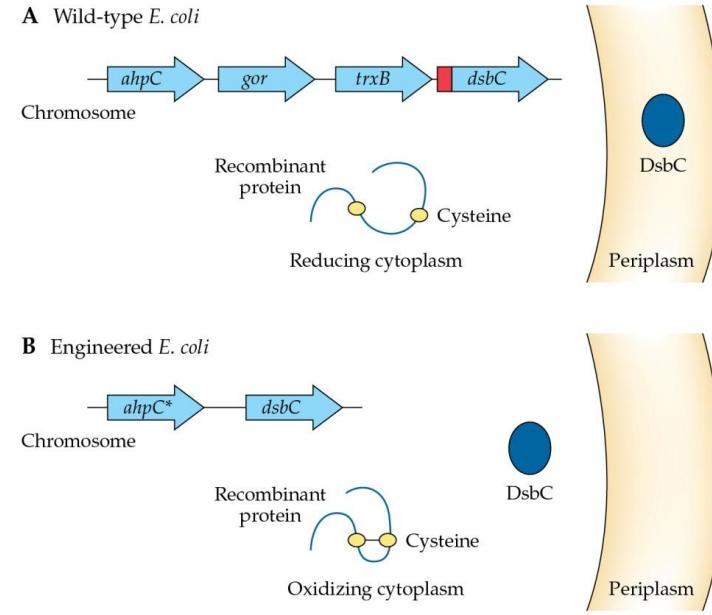


Figure 3.10 An engineered *E. coli* host strain catalyzes the formation of disulfide bonds in cytoplasmic recombinant proteins. **(A)** Wild-type *E. coli* produces reductases (from genes *gor* and *trxB*) that maintain cysteines in cytoplasmic proteins in their reduced state, in which they do not form disulfide bonds (Fig. 3.9). Disulfide bond isomerase DsbC is expressed with a signal peptide (encoded by the DNA sequence represented in red) that directs it to the inner membrane for translocation to the periplasm (Fig. 3.11). **(B)** In the genetically modified *E. coli* strain, the *gor* and *trxB* genes were deleted from the chromosome to create oxidizing conditions in the cytoplasm that are suitable for disulfide bond formation. The lethality of this condition was overcome by altering AhpC (encoded by *ahpC**) to reduce essential proteins. DsbC is overexpressed without a signal peptide and is therefore retained in the cytoplasm to catalyze disulfide bond formation between some cysteines in a protein. Note that the position of the genes in the chromosome is represented schematically.

Decreasing Protein Degradation

- **Protease Attack:** Microbial hosts produce numerous proteases that can rapidly degrade foreign proteins that are perceived as damaged or unneeded.
- **The N-End Rule:** The half-life of a protein is heavily influenced by the nature of its N-terminal amino acid residues.
- **Genetic Defense:** We can increase stability by using **protease-deficient host strains** or by **engineering the protein's N-terminus** to hide it from proteases.

Table 3.8 Stability of β -galactosidase with certain amino acids added to its N terminus^a

Amino acid added	Half-life
Met, Ser, Ala	>20 h
Thr, Val, Gly	>20 h
Ile, Glu	>30 min
Tyr, Gln	~10 min
Pro	~7 min
Phe, Leu, Asp, Lys	~3 min
Arg	~2 min

^aData from Bachmair et al., *Science* 234: 179–186, 1986.

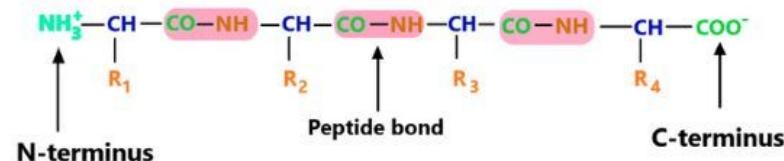


FIG: N-terminus and C-terminus of polypeptide chain

Directing Protein Secretion

- **Secretion Benefits:** Directing proteins to the periplasm or medium avoids cytoplasmic proteases and simplifies purification.
- **Signal Peptides:** Short N-terminal sequences act as 'zip codes' to direct proteins to the Sec translocon.
- **Periplasmic Advantage:** The periplasm provides an oxidative environment conducive to disulfide bond formation.

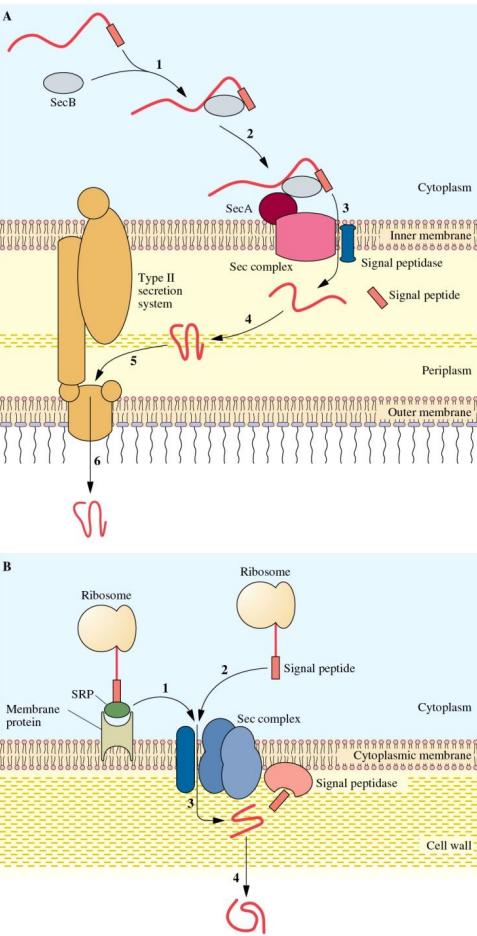


Figure 3.11 Protein secretion in bacteria. (A) General secretion pathway in Gram-negative bacteria. The SecB protein binds to the N-terminal signal peptide (red rectangle) of a secretory protein in the cytoplasm (1). SecB delivers the protein to the SecA protein that is associated with the Sec complex (SecYEG translocon) in the inner membrane (2), and the secretory protein is translocated across the inner membrane (3). A signal peptidase removes the signal peptide, and the secretory protein is properly folded in the periplasm (4). Some secretory proteins are directed to a protein complex (type II secretion system) in the outer membrane (5) and are translocated to the external environment (6). (B) Protein secretion in Gram-positive bacteria. A signal recognition particle (SRP) binds to the signal peptide of a secretory protein, and this complex binds to a membrane protein that directs the secretory protein (1) to the Sec complex. There is also an SRP-independent pathway (2), where a signal peptide alone makes contact with the Sec complex. The secretory protein is translocated through a channel within the Sec complex (3), and the signal peptide is removed by a signal peptidase. Proper folding of the secretory protein occurs as it passes through the cell wall (4).

The Tat Secretion Pathway

- **Folded Export:** The Twin-Arginine Translocase (Tat) system is unique because it exports **proteins that have already completed their folding.**
- **Specific Motifs:** Requires a specific twin-arginine signal sequence from proteins like TorA.
- **High-Level Secretion:** This system has been used to achieve massive secretion levels (30 g/L) for proteins like human growth hormone.

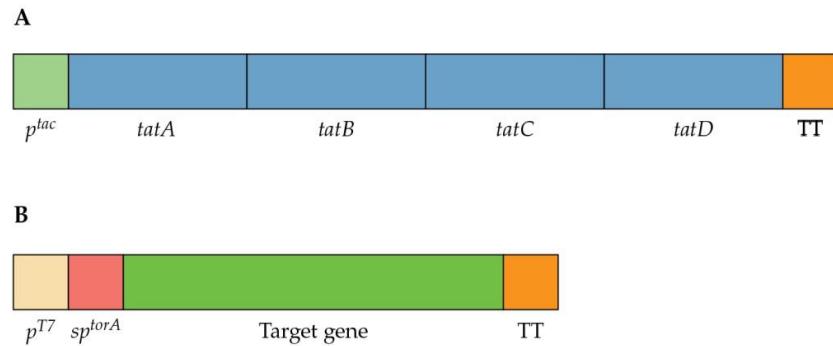


Figure 3.13 Twin-arginine translocation (Tat) pathway for periplasmic secretion of recombinant proteins. **(A)** Production of the Tat translocase was increased in an *E. coli* host strain by inserting the strong, regulatable *lac* promoter (*p^{lac}*) upstream of the *tatABCD* operon in the chromosome. **(B)** To direct a recombinant protein to the Tat translocase, the target gene is fused to the sequence for the N-terminal twin-arginine signal peptide from TorA (*sp^{torA}*). *p^{T7}*, T7 promoter; TT, transcription termination sequence.

Extracellular Secretion (Type III) - T3SS

- **Crossing Both Membranes:** Type III systems span the inner and outer membranes to dump protein directly into the medium.
- **Minimal Systems:** Engineers have created 'minimal' operons of the T3SS to control the timing and level of extracellular export (Natural T3SS are highly complex, ~ 30 structural and regulatory proteins).
- **Continuous Harvesting:** This allows for recovery of protein without lysing the cells, protecting the product from intracellular proteases.

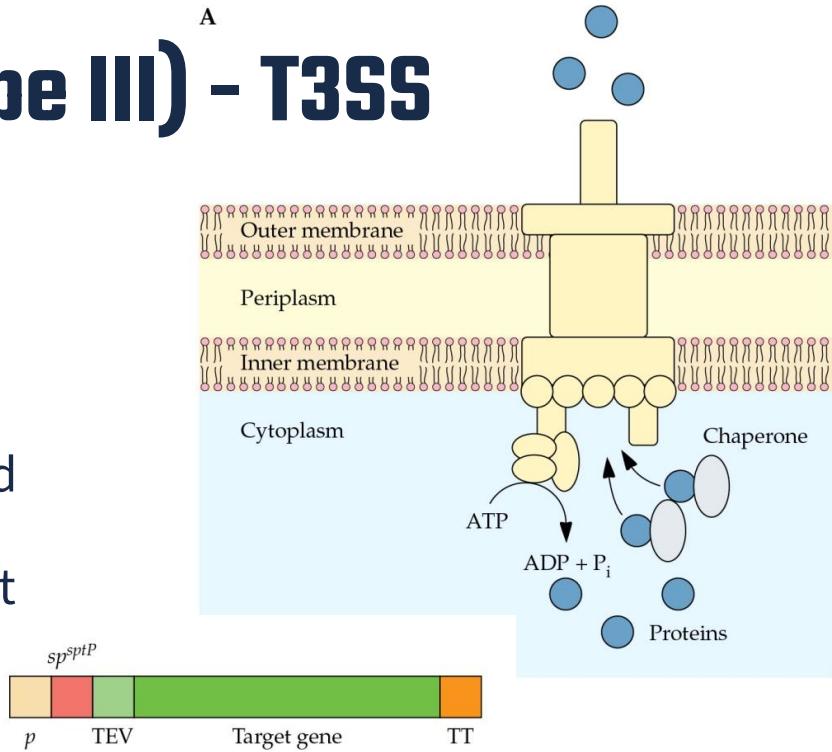


Figure 3.15 Extracellular secretion of recombinant proteins via the type III secretion system. (A) The type III secretion system is a protein complex that spans the inner and outer membranes of a Gram-negative bacterium. Proteins with a specific N-terminal signal peptide are delivered to the secretion apparatus by a specific chaperone and are exported to the culture medium. (B) Recombinant proteins expressed from genes that possess a sequence encoding the N-terminal signal peptide from SptP (sp^{sptP}), a natural type III secretion system substrate, are bound by the chaperone SicP. Following secretion, the signal peptide may be cleaved by the *Tobacco etch virus* (TEV) protease if the DNA sequence for the protease recognition site is placed between the signal peptide coding sequence and the target gene. *p*, promoter; TT, transcription termination sequence.

Affinity Purification Systems

- **Fusion Tags:** Target proteins are expressed as fusions with short peptide tags that have a high affinity for specific ligands.
- **One-Step Recovery:** The His-tag system allows proteins to bind tightly to nickel columns, enabling purification of up to 100-fold in a single step.
- **Enabling Tech:** This connects to the Lecture 4 Tandem Affinity Purification (TAP) for protein teams.
- **Broad Applicability:** A variety of tags (e.g., GST, MBP, Flag) exist.

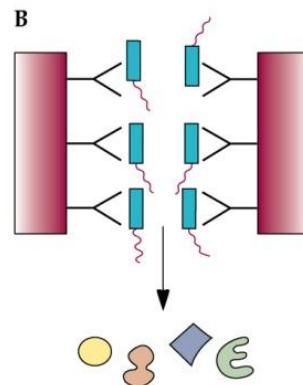
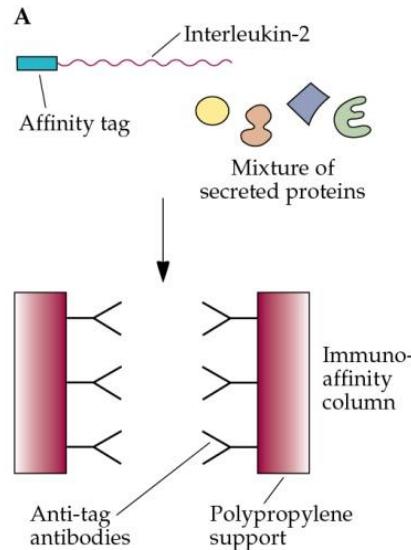


Figure 3.17 Purification of a fusion protein by immunoaffinity chromatography. **(A)** An antibody that binds to a short peptide sequence (affinity tag) of the fusion protein (anti-tag antibody) is attached to a solid polypropylene support. The mixture of secreted proteins is passed through the column containing the bound antibody. **(B)** The affinity tag of the fusion protein binds to the antibody, and the other proteins pass through. The immunopurified fusion protein can then be selectively eluted from the column by the addition of pure affinity tag peptide.

Chromosomal Integration

- **Plasmid Instability:** In large bioreactors, cells often 'lose' their plasmids if they don't provide a growth advantage.
- **Stable Maintenance:** Integrating the target gene into the host chromosome ensures it is passed to all offspring without antibiotics.
- **Enabling Tech:** This is the same Homology-Directed Repair (HDR) mechanism we covered in Lecture 2's CRISPR module - typically 50 nucleotide overhangs

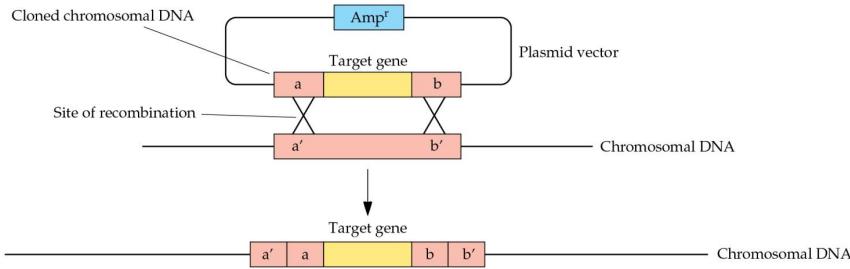


Figure 3.21 Integration of a target gene into a chromosomal site by one-step homologous recombination. The target gene is cloned into a plasmid vector between sequences (a and b) that are homologous to chromosomal DNA sequences (a' and b') at the site targeted for integration. Recombination occurs between homologous plasmid-borne DNA regions a and b and host chromosome DNA regions a' and b', respectively. A double-crossover event (x) results in the stable integration of the target gene into the chromosome.

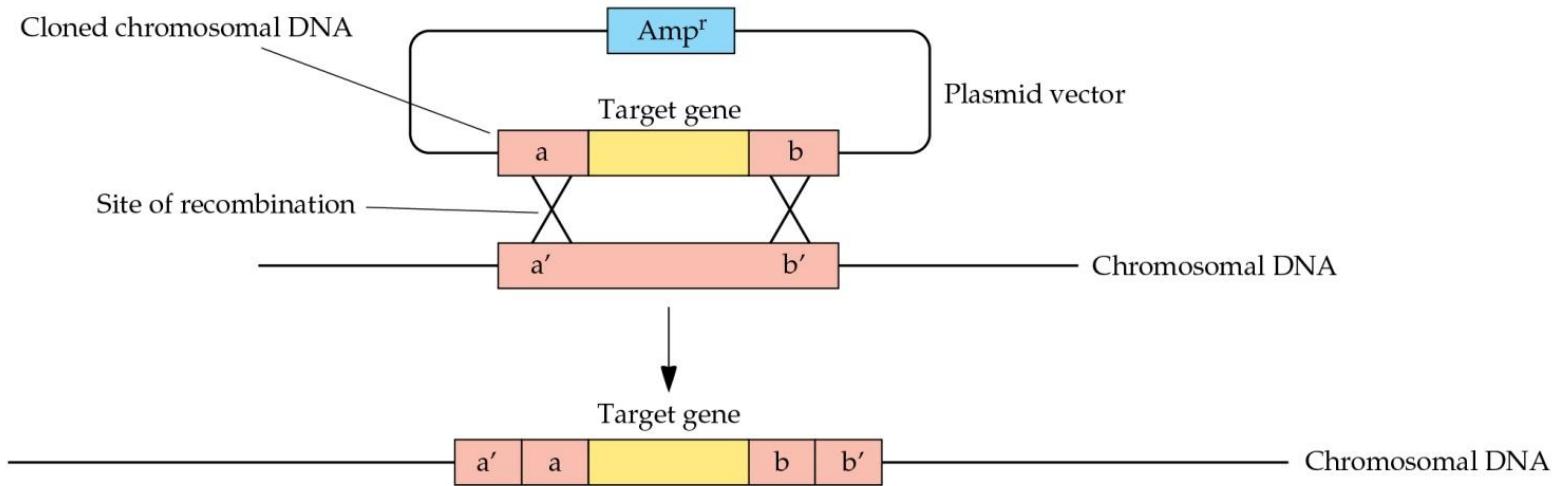


Figure 3.21 Integration of a target gene into a chromosomal site by one-step homologous recombination. The target gene is cloned into a plasmid vector between sequences (a and b) that are homologous to chromosomal DNA sequences (a' and b') at the site targeted for integration. Recombination occurs between homologous plasmid-borne DNA regions a and b and host chromosome DNA regions a' and b', respectively. A double-crossover event (x) results in the stable integration of the target gene into the chromosome.

Removing Selectable Markers

- **Safety Standards:** Antibiotic resistance genes used for selection are often undesirable in the final transgenic strain used for production.
- **Environmental Security:** Removing these markers prevents the potential spread of antibiotic resistance genes into the wider environment.

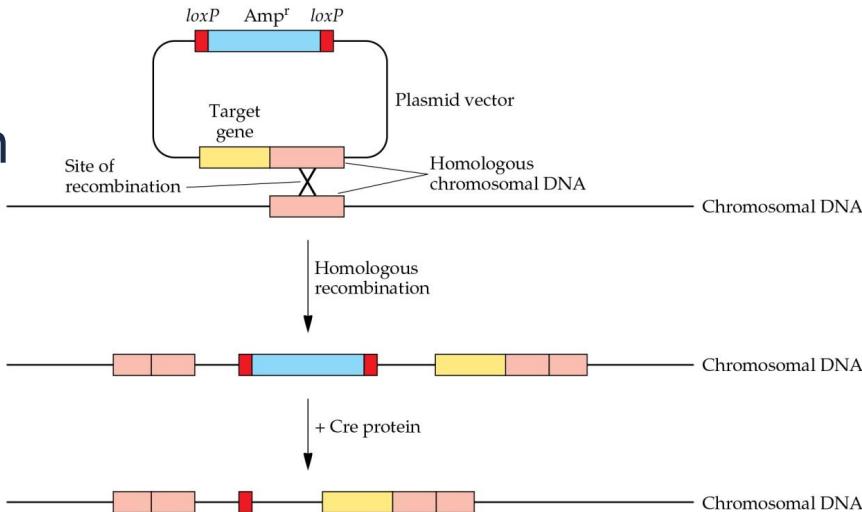


Figure 3.23 Removal of an antibiotic resistance gene following integration of plasmid DNA into a bacterial chromosome. The target gene is cloned into a plasmid adjacent to a sequence that is homologous to a chromosomal DNA sequence at the site targeted for integration. Recombination (x) occurs between homologous sequences in the chromosome and on the plasmid, resulting in the integration of the entire plasmid into the chromosome. The antibiotic resistance gene (*Amp^r*), which is flanked by *loxP* sites, is excised by the action of the Cre recombinase enzyme, leaving one *loxP* site on the integrated plasmid. The gene encoding the Cre enzyme is on a separate plasmid within the same cell under the transcriptional control of the *E. coli lac* promoter so that excision is induced when IPTG is added to the growth medium.

Video - Concepts in the module or a demonstration

- Molecular Chaperones: Heat Shock Proteins (2min13s)

More videos

- TAT Protein Secretion System (1min26s)

The End