11.3.4 Analysis of RNA-Binding Proteins

1. Bandshift assay

2. Toeprint assay (Primer extension inhibition)
The Principle of the Toeprint Assay

Control: mRNA without protein

mRNA plus protein
Toeprint Assay

DD Hartz (1988)
Methods in Enzymology
164: 419
11.3.5 Analysis of Protein-Protein Interaction

How to study protein-protein interactions?

- Biochemical techniques
- Molecular biological *in vivo* techniques
- Genetic techniques
Biochemical Techniques

- Protein affinity chromatography
- Chemical crosslinking
- Photoaffinity crosslinking
- Protein probing
Protein Affinity Chromatography

Specific Labeling of an Interacting Protein With a Labeled Photoactivatable Cross-Linking Reagent

Two-Dimensional Gels to Identify Cross-Linked Proteins in a Complex

Use of a Labeled Protein to Probe an Expression Library

Molecular Biological Techniques

- The classical two-hybrid system
- The LexA repressor system
- *Bordetella pertussis* adenylate cyclase
- The three-hybrid system
The Two-Hybrid System

Objectives:

1. To identify an unknown protein interacting with a known protein
2. To demonstrate interaction between two known proteins
The Gal4 Protein Contains Two Domains
Gal4 Acts as a Transcriptional Activator for the Galactose Metabolic Enzymes in Yeast
These Two Domains Are Functionally Separable in the Polypeptide Domain
The Two Proteins X and Y Form a Stable Complex
Construction of Two Fusion Proteins

The Fusion Protein Complex Can Act as a Functional Transcriptional Activator

The gene coding for the galactose metabolic enzyme is replaced by reporter genes such as \(\text{lacZ}\).
The Yeast-Two-Hybrid System (Y2H)

T Berggard (2007) Proteomics 7: 2833
A Two-Hybrid System Based on the *Bordetella pertussis* Adenylate Cyclase

Schematic Representation of the Two-Hybrid Assay

The Three-Hybrid System

Objective:
To detect protein-RNA interactions

Principle:
Binding of a bifunctional RNA to each of two hybrid proteins activates transcription of a yeast reporter gene \textit{in vivo}
The Three Hybrid System

Genetic Technique

Isolation of allele-specific mutants

**Example:**
- GrpE interacts with DnaK
- GrpE is essential, DnaK is non-essential
- *grpEts42*: non-viable at 42°C
- Isolation of a *dnaK* mutation allowing growth at 42°C
11.3.6 Surface Display

**Goal:**
To identify ligands interacting with a protein or peptide

1. *in vivo* systems
   - prokaryotic
   - eukaryotic
2. *in vitro* systems
Display Systems

What are protein/peptide display systems?

The protein of interest (passenger protein) is fused to the C or N terminus or within a polypeptide sequence (carrier protein) that targets the resulting chimera onto the surface of biological particles where it will be anchored.

Protein display systems represent one of the most powerful tools for protein engineering.
C-Terminal Fusion

carrier protein           passenger protein

N-terminal Fusion

passenger protein           carrier protein
Internal Fusion
Prokaryotic Display Systems

1. Phage display
2. Cell surface display
3. Spore surface display
11.3.6.1 Phage Display

A protein of interest is incorporated into the coat of a filamentous phage of *E. coli* (M13, fd, f1) by genetically fusing it with one of several phage coat proteins.

George P. Smith  
University of Missouri  
Tucker Hall  
Columbia, USA
The M13 Particle

- **Gene VIII**: major structural protein
- **Gene III**: minor coat protein; necessary for adsorption

M13 phage s.s. DNA genome

- ~2,700 copies of gene VIII protein
- 5-8 copies of gene III protein
Phage M13 Display

Bio-panning

GP Smith (1985) Science 228: 1315
Only Phages with Recombinant pVIII Will Bind to the Matrix
Inverse PCR Introduction of Amino-Terminal Extension to Major Coat Protein VIII
Where is the phage coat assembled?

In the periplasm

That means:

Each protein has to be translocated through the inner membrane into the periplasm
Background:

Based on fusions of proteins or peptides to the capsid proteins pIII and pVIII.

Translocated through the inner membrane via the Sec pathway, where they become incorporated into the phage particle.

- Phage display depends on its effective translocation.
Four basic limitations are associated with the translocation pathway:

1. Protein has to be present in an unfolded state
2. No ATP-dependent chaperones present in the periplasm
3. Reducing environment in the cytoplasm can effect folding
4. Cryptic stop-transfer sequences or the local accumulation of positively charged residues can result in incomplete translocation
By Which Mechanisms Proteins Are Translocated Through the Inner Membrane?

1. The Sec pathway
2. The Tat pathway
3. The SRP pathway
Components of the Sec Pathway

1. Proteins contain a cleavable N-terminal signal sequence
2. SecB binds to hydrophobic regions and prevents premature folding
3. SecA acts as a motor protein
4. SecYEG heterotrimer forms the membrane-embedded translocon

**Important:**
Sec-dependent proteins are translocated in a translocation-competent form
Components of the Tat Pathway

1. Proteins contain a cleavable N-terminal signal sequence with a twin-arginine motif (RR)
2. The Tat proteins

**Important:**
Tat-dependent proteins are translocated post-translationally in their **folded** state
Components of the SRP Pathway

1. Proteins contain a cleavable N-terminal signal sequence with high hydrophobicity

2. The signal recognition particle (SRP):
   - 4.5S RNA
   - 49.6-kDa protein homologue Ffh
   - FtsY protein
A protein of interest is attached to cell surface protein

**Gram-negatives:**
- outer membrane protein
- periplasmic protein
- inner membrane protein

**Gram-positives:**
- cytoplasmic membrane protein
- cell wall
Cell Surface Display Systems in Gram-negative Bacteria
Diagram Showing the Principle of APEX (Anchored Periplasmic Expression)

BR Harvey (2004) PNAS 101: 9193
Cell Surface Display Systems in Gram-positive Bacteria

choline-subst.

P Cossart (2000) PNAS 97: 5013
The sortase mechanism needs the following components:

1. The **sortase** (anchored in the cytoplasmic membrane facing the outside)
2. The **substrate protein** (signal peptide and sorting sequence)
3. The **peptide crossbridge** of the cell wall
Surface Display Using Spore Coat Proteins

Reasons why the protein of interest might fail to be displayed:

- Can aggregate in the periplasm or cytoplasm
- Disulfide-containing proteins: form non-native disulfide bonds
- Local sequence stretches are incompatible with translocation
- Degradation by proteases
Attempts to circumvent poor display:

- Reducing toxicity by reducing expression
- Improving periplasmic folding by overexpressing periplasmic chaperones Skp or FkpA

Principle:

Production of stabilized antibody-ribosome and mRNA complexes \textit{in vitro}

Andreas Plückthun
University of Zürich
Principle of *in vitro* Ribosome Display for Screening Native Protein Libraries for Ligand Binding

mRNA without stop codon

Limitations:

- Efficient *in vitro* translation and stalling can be technically challenging.
- Concentrations of cellular factors required for efficient scFv folding may differ from concentrations found *in vivo*. 
Principle of Intracellular Ribosome Display: Affinity of SCA Libraries

Objective:
To isolate a SCA (scFv) recognizing a protein of interest

1. A scFv library is PCR-amplified, detection tags are added to both ends and a flexible linker followed by a SecM17-mediated arrest site is added to the 3' end

Schematic Drawing of the SecM17-Fused scFv

F = FLAG tag
scFv13 = SCA library
M = cMyc tag
H = His-tag
T = Thrombin cleavage site
GS = Flexible Gly-Ser linker (AGSA AGSG)
SecM17 = SecM stall sequence
* = Stop codon
2. The modified scFv library is cloned in a plasmid vector
3. Induction of \textit{in vivo} transcription
4. SecM causes translation arrest
5. Desired complexes are affinity selected
6. Ribosome complexes are dissociated by addition of EDTA
7. Isolation of mRNA followed by reverse transcription
11.3.7 Surface Plasmon Resonance (SPR)

- Method to monitor the kinetics of a reaction in real time
- Measures the affinity and speed of binding between a capture agent and a probe
- Yields apparent rate constants for the association and dissociation phases of the reaction
Surface Plasmon Resonance Detection Unit

L = light source
P = prisma
D = photodiode array
S = sensor surface
F = flow cell
Application of Surface Plasmon Resonance
11.3.8 SELEX

Systematic Evolution of Ligands by Exponential Enrichment

T Tuerk and L Gold (1990) Science 249: 505
T4 DNA Polymerase Translational Operator Region

Which nucleotides in the loop region are critical for T4 DNA polymerase binding?
Generation of a Library of mRNA Molecules With Random Nucleotide Sequence in the Loop Region

- The library will have a size of $4^8$ or 65,536 uniquely different sequences
- PCR makes use of five different oligos
- A synthetic gene of 110 nucleotides in length is synthesized
In vitro Transcription Produces 92 Nucleotide Long mRNA Molecules

There will be a library of 65,536 unique RNA molecules
Selection by Binding to T4 DNA Polymerase

T4 DNA polymerase has been immobilized on nitrocellulose filters
cDNA Construction

Reverse transcriptase, dNTP’s

PCR with oligo’s #5, #1

Double stranded cDNA
The SELEX Method Identified the Wild Type Sequence and an Alternative Sequence With High Affinity T4 DNA Polymerase Binding Properties
RNA Aptamers

1. Single stranded RNA molecules
2. Average length ~ 100 nucleotides
3. Binds with high specificity to
   - organic dyes
   - amino acids
   - nucleotides
   - antibiotics
   - peptides
   - etc.
4. Fold into an intricate structure

RNA Aptamer Can Distinguish Between Theophylline and Koffeine by 10,000-fold in Affinity
The SELEX Procedure Allows Selection of RNA Molecules with New Functions

Objective:
Identification of an RNA molecule binding to ATP

1. $10^{15}$ random RNA sequences
2. ATP coupled to resin
3. RNA sequences that do not bind ATP (discard)
4. RNA sequences that bind ATP
5. RNA sequences enriched for ATP-binding function

Amplify (reverse transcriptase, RNA polymerase)
Structure of the RNA Aptamer Binding ATP

K_d < 50 µM
Riboswitches

Riboswitches are natural aptamers present in the 5'-untranslated region of operons involved, e.g., in the synthesis of vitamins and amino acids.
11.3.9 Gene Targeting by Homologous Recombination

Objective:

Integration of DNA sequences at predetermined (mostly homologous) sites in eukaryotic cells
Two Configurations of Constructs Used for Homologous Recombination

- **Insertion Construct**
  - Homologous sequences: construct and target gene
  - Nonhomologous sequences: positive selectable marker, target gene, vector

- **Replacement Construct**
  - Homologous sequences: construct and target gene
  - Nonhomologous sequences: positive selectable marker, target gene, vector
Enrichment for Homologous Recombinants by Positive-Negative Selection Using the TK Gene

- 2'-Deoxyguanosine analog
- Phosphorylated
- Chain termination
Enrichment for Homologous Recombinants Using a Positive Selectable Marker (*neo*) Lacking a Promoter
Using the Cre-*lox* System to Integrate an Expression Construct at a Specific Reproducible Site

![Diagram of Cre-lox system]
Using the Cre-\textit{lox} System to Remove a Selectable Marker After Integration at a Unique Site

Transgenic plants
11.3.10 Incorporation of non-native amino acids into proteins in *E. coli* cells
Incorporation of non-native amino acids into proteins in *E. coli* cells

The genetic code codes for only 20 different amino acids though there are 61 codons

**Exceptions:** Selenocysteine (Sel) and pyrrolysine (Pyr)

These amino acids contain a limited number of functional groups: carboxylic acids, amides, thiol, thiol ether, alcohols, basic amines, alkyl and aryl groups
Additional groups are introduced through posttranslational modifications: phosphorylation, glycosylation and hydroxylation

Desirable: fluorescent, photoreactive, redox active amino acids, glycosyl

Objective: Expansion of the genetic code

P. G. Schultz
Which codon can be used to expand the genetic code?

TAG: amber stop codon

Why?

- The least used of the three stop codons
- Some *E. coli* strains contain naturally occurring amber-suppressor tRNAs
- Amber suppression is well tolerated
Two new translational components have to be designed:

1. A tRNA recognising the amber stop codon and inserting the new amino acid
2. Its cognate tRNA synthetase
Important requirements for both components:

**tRNA:**
1. Must be expressed and processed correctly
2. Not aminoacylated by the host synthetases with endogenous amino acids

**Synthetase:**
1. Not allowed to transfer the new amino acid to host tRNAs

The new tRNA – synthetase pair is orthogonal to the host
Requirements for the new amino acid:

1. Non-toxic to the cells
2. Must be efficiently transported from the growth medium into the cytoplasm
3. Or synthesized within the cell
First orthogonal pair:

$tRNA^{Tyr}$ and its cognate synthetase derived from *Methanococcus jannaschii*

Inserts tyrosine instead of an amber codon

**Problem:**
tRNA recognized to some degree by *E. coli* synthetases

**Objective:**
tRNA should not be recognized by *E. coli* synthetases
Solution:
1. Coding sequence was randomized partially using mutagenic PCR
2. Library passed through a negative selection in *E. coli*
3. Positive selection in the presence of the archeal synthetase
Additional orthogonal pairs:

1. Leucyl tRNA – synthetase
2. Lysyl tRNA – synthetase
3. Glutamyl tRNA – synthetase

Can be used in the future alone or in combination with the existing one to

- Decode the TGA opal codon
- The four-base AGGA codon
How to isolate synthetases inserting new amino acids?

Randomization of all the amino acids interacting with tyrosine
Two-step selection procedure:

1. Cells containing the Cm gene with amber mutations at sites permissive to mutations in the presence of the new amino acid and chloramphenicol
2. Cells containing the barnase gene with one amber mutation in the absence of the new amino acid
New Amino Acids
Characteristics of Some New Amino Acids

p-azido-L-phenylalanine and p-benzoyl-L-phenyl-alanine: crosslinking amino acids allow capture of covalent protein-protein or protein-DNA interactions in vitro and in vivo

p-acetylphenylalanine: allows selective modification with non-peptide molecules such as drugs, dyes and polymers

β-GlcNAc-serine and α-GalNAC-threonine: glycosylated amino acids; can be elaborated into more complex carbohydrates using glycosyl transferases
Biosynthesis of p-Aminophenylalalanine

Biosynthetic gene cluster derived from *Streptomyces venezuelae*
Literature:


