4.2: M13 Phage

The bacteriophage known as "M13" forms the basis of cloning systems designed to easily introduce mutations into genes inserted into the phage genome. It also has been used in various "phage display" methodologies and "combinatorial" DNA and peptide libraries.

M13 infection and replication

M13 is a filamentous bacteriophage which infects *E. coli* host. The M13 genome has the following characteristics:

- Circular *single-stranded* DNA
- 6400 base pairs long
- The genome codes for a total of 10 genes (named using Roman numerals I through X)
Figure 4.2.1: M13 genome

- **Gene VIII** codes for the major structural protein of the bacteriophage particles
- **Gene III** codes for the minor coat protein

~2,700 copies of gene VIII protein

5-8 copies of gene III protein
**Figure 4.2.2: Gene III and gene VIII**

- The gene VIII protein forms a tubular array of approx. 2,700 identical subunits surrounding the viral genome
- Approximately five to eight copies of the gene III protein are located at the ends of the filamentous phage (i.e. genome plus gene VIII assembly)
- Allows binding to bacterial "sex" pilus
  - Pilus is a bacterial surface structure of *E. coli* which harbor the "F factor" extrachromosomal element

### Infection

- Single strand genome (designated '+') strand) attached to pilus enters host cell
  - Major coat protein (gene VIII) stripped off
  - Minor coat protein (gene III) remains attached
- Host components convert single strand (+) genome to double stranded circular DNA (called the replicative or "RF" form)
- Transcription begins
  - Series of promoters
    - Provides a gradient of transcription such that gene nearest the two transcription terminators are transcribed the most
  - Two terminators
    - One at the end of gene VIII
    - One at the end of gene IV
  - Transcription of all 10 genes proceeds in same direction

### Amplification of viral genome

- Gene II protein introduces 'nick' in (+) strand
- Pol I extends the (+) strand using *strand displacement* (and the '-' strand as template)
- After one trip around the genome the gene II protein nicks again to release a completed (linear) '+'' genome
  - Linear (+) genome is circularized
- During first 15-20 minutes of DNA replication the progeny (+) strands are converted to double stranded (RF) form
  - These serve as additional templates for further transcription
- Gene V protein builds up
  - This is a *single stranded DNA binding protein*
  - Prevents conversion of single (+) strand to the RF form
- Now get a buildup of circular single stranded (+) DNA (M13 genome)
Phage packaging

- Major coat protein (Gene VIII) present in *E. coli* membrane
- M13 (+) genome, covered in ss binding protein - Gene V protein, move to cell membrane
- Gene V protein stripped off and the major coat protein (Gene VIII) covers phage DNA as it is extruded out
  - Packaging process is therefore *not linked to any size constraint* of the M13 genome
  - Length of the filamentous phage is determined by size of the DNA in the genome
  - Inserts of up 42 Kb have been introduced into M13 genome and packaged (7x genome size)
- ~8 copies of the Gene III protein are attached at the end of the extruded genome

Development of M13 into a cloning vector

M13 was developed into a useful cloning vector by inserting the following elements into the genome:

- a gene for the *lac* repressor (*lac I*) protein to allow regulation of the *lac* promoter
- the operator-proximal region of the *lac Z* gene (to allow for a-complementation in a host with operator-proximal deletion of the *lac Z* gene).
- a *lac* promoter upstream of the *lac Z* gene
- a polylinker (multiple cloning site) region inserted several codons into the *lac Z* gene
**Figure 4.2.4: Insertions into genome**

- The vectors were named according to the specific polylinker region they contained
- The vectors were typically constructed in pairs, with the polylinker regions in opposite orientations

**Figure 4.2.5: Polylinker regions**

### Cloning into M13mp vectors

- The RF (double stranded) form of the M13 phage can be isolated and treated just like any other plasmid
- The polylinker region can be "opened" using restriction endonucleases appropriate for accepting the fragment of interest
- The fragment is ligated into the plink region
- The availability of inverse oriented plink's (e.g. mp18, mp19) means that inserted fragments with non-complementary ends can be inserted in either orientation

### Single stranded forms of the phage

The ability to isolated a single stranded form of the phage has advantages in both sequencing and mutagenesis.

- Single stranded DNA template can be read further than double stranded template

An efficient mutagenesis method (the "Kunkel" method) was developed using the single stranded form of the phage.

- The M13mp vector with insert is first grown in a mutant *E. coli* host (e.g. CJ236) which would occasionally incorporate uracil into the DNA instead of thymidine.
- *E. coli* normally synthesizes an enzyme (uracil-N-glycosidase) that removes uracil residues in DNA. However, in *ung*- strains, the uracil is not removed.

- The level of uracil mis-incorporation into DNA is enhanced in strains which have a deficiency in dUTPase. This enzyme converts dUTP to dUDP, and therefore, in *dut*- strains the levels of dUTP are elevated and enhance misincorporation of dUTP into host DNA.

- The *E. coli* strain CJ236 has a genotype which includes dut-/ung- features.
  - A mutagenic primer would be annealed to this single strand template (e.g. to produce a point mutation)
  - The primer is extended using the four dNTP's, and is ligated to produce duplex DNA
  - The duplex DNA is inserted into a different host *E. coli* (e.g. *JM101*) which recognizes and excises (degrades) uracil containing DNA (i.e. a strain which is *ung+*)
  - The parent (wild type) strand is preferentially degraded and the mutagenic strand is replicated.
  - The phage progeny typically have a high incidence (80-90%) of the desired mutation.

*Figure 4.2.6: Kunkel method*