Bio Sci 203 Lecture 18 - cDNA libraries, etc

• Bruce Blumberg (blumberg@uci.edu)
  – office - 4203 Bio Sci II
  – 824-8573
  – lab 5427 (x46873), 5305 (x43116), 5345 BSII
  – open office hours. If my door is open you are welcome to drop in.
  – or if you can find me in the lab or elsewhere
• Link is on main class web site

• Today
  – wrap up genomic libraries
  – factors affecting sequence clonability in E. coli
  – cDNA library theory and construction
### Comparison of cloning systems

<table>
<thead>
<tr>
<th></th>
<th>YAC</th>
<th>BAC</th>
<th>PAC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Host cells</strong></td>
<td>S. cerevisiae AB1380, J57D</td>
<td>E. coli DH10B</td>
<td>E. coli DH10B</td>
</tr>
<tr>
<td><strong>Transformation method</strong></td>
<td>Spheroplast transformation</td>
<td>Electroporation</td>
<td>Electroporation</td>
</tr>
<tr>
<td><strong>DNA topology of recombinants</strong></td>
<td>Linear</td>
<td>Circular supercoiled</td>
<td>Circular supercoiled</td>
</tr>
<tr>
<td><strong>Maximum insert size</strong></td>
<td>&gt;&gt;1 Mb</td>
<td>~300 kb</td>
<td>~300 kb</td>
</tr>
<tr>
<td><strong>Selection for recombinants</strong></td>
<td>Ade2 supF red-white color selection</td>
<td>LacZ blue-white</td>
<td>SacIiIb selective growth</td>
</tr>
<tr>
<td><strong>Selection for vector</strong></td>
<td>Dropout medium (lacking trp and ura)</td>
<td>Chloramphenicol</td>
<td>Kanamycin</td>
</tr>
<tr>
<td><strong>Enzyme for partial digests</strong></td>
<td>EcoRI</td>
<td>HindIII</td>
<td><em>MboI</em> or <em>Sau3Al</em></td>
</tr>
<tr>
<td><strong>Stability</strong></td>
<td>Variable but can be very unstable</td>
<td>Very stable</td>
<td>Very stable</td>
</tr>
<tr>
<td><strong>Degree of chimerism</strong></td>
<td>Varies but can be &gt;50%</td>
<td>Very low</td>
<td>Very low</td>
</tr>
<tr>
<td><strong>Degree of co-cloning</strong></td>
<td>Occasional</td>
<td>Undetectable</td>
<td>Undetectable</td>
</tr>
<tr>
<td><strong>Purification of intact inserts</strong></td>
<td>Difficult</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td><strong>Copies/cell</strong></td>
<td>One</td>
<td>One</td>
<td>One/inducible</td>
</tr>
<tr>
<td><strong>Direct sequencing of insert</strong></td>
<td>Difficult</td>
<td>Relatively easy</td>
<td>Relatively easy</td>
</tr>
<tr>
<td><strong>Clone mating</strong></td>
<td>Yes</td>
<td>No</td>
<td>No'</td>
</tr>
</tbody>
</table>
Which type of library to make

• Do I need to make a new library at all?
  – Is the library I need available?
    • PAC libraries are suitable for most purposes and may be the method of choice
    • If your organism only has YAC libraries available you may wish to make PACs
    • Much easier to buy pools or gridded libraries for screening
      – doesn’t always work
  – What is the intended use?
    • Will this library be used many times?
      – E.g. for isolation of clones for knockouts
      – if so, it pays to do it right and make a top quality library
  – who should make the library?
    • Going rate for custom BAC or PAC library is 50K. Most labs do not have these resources
    • if care is taken, construction is not so difficult
      – we will soon have an automated colony picker that can make high density gridded libraries
Screening of genomic libraries

- What types of probes are suitable for screening genomic libraries?
Screening of genomic libraries

• What types of probes are suitable for screening genomic libraries?
  – Not suitable
    • antibodies (no protein expression)
    • degenerate (mixed) oligonucleotides (genome complexity)
    • DNA binding proteins (genome complexity)
  – suitable
    • cDNAs (or mRNAs)
    • genomic fragments
    • longer oligonucleotides (> 30 mers)
Sequence stability in E. coli

- What are the sorts of factors that might modulate whether a sequence can be stably propagated in E. coli?

  - 1
  
  - 2
  
  - 3
Sequence stability in E. coli

- What are the sorts of factors that might modulate whether a sequence can be stably propagated in E. coli?
  - toxicity
  - restriction
  - Recombination

- toxicity
  - sequence may lead to the production of a toxic product or toxic levels of an otherwise innocuous product
  - more problematic with cDNA than genomic clones

- restriction - Raleigh 1987 Meth. Enzymol. 152, 130-141
  - virtually all microorganisms have systems to destroy non-endogenous DNA host range restriction
    - four classes of restriction endonucleases
  - very important for cloning purposes are recently discovered systems that degrade DNA containing 5-methyl cytosine or 6-methyl adenine.
  - If you are cloning genomic DNA, or hemimethylated cDNA these are very important!
    - Because virtually all eukaryotic DNA contains 5-methyl cytosine and/or 6-methyl adenine
      - mcrA,B,C - methylcytosine
      - mrr - methyl adenine
Sequence stability in E. coli (contd)

• Restriction (contd)
  – frequency at which foreign DNA escapes restriction varies from 1/10^5 for EcoK and EcoB to 1/10 for mcrA.
  – Bottom line is that if you are going to clone genomic DNA or make cDNA libraries, one needs to be conscious of the mcr and mrr restriction status of strains and packaging extracts to be used.

• Recombination - Wyman and Wertman (1987) Meth Enzymol 152, 173-180
  – genomic DNA contains lots of repeated sequences
    • direct repeats
    • inverted repeats
    • interspersed repeats (e.g. Alu)
  – repeated sequences are not stable in recombination proficient E. coli if present in multiple copies per cell
    • lambda
    • plasmid
    • cosmid
  – seems not to apply to single copy vectors such as BAC and PAC
  – on the order of 30% of the human genome is unstable in plasmid or phage clones
    • observation is that phages with such sequences either don’t grow at all or get shorter with time
Sequence stability in E. coli (contd)

• Recombination (contd)
  – E. coli has a variety of recombination pathways. These are the major players in causing sequence underrepresentation
    • recA required for all pathways
    • rec BCD - major recombination pathway
    • sbc B,C - suppressor of B,C
    • minor pathways
      – rec E
      – rec F
      – rec J
    • rule of thumb - the more recombination pathways mutated, the sicker the cells and the slower they grow
  – major players for inverted repeats are recBCD and sbc
  – recA is most important for stabilizing direct repeats and preventing plasmid concatamerization
Sequence stability in E. coli (contd)

• Plating a genomic library
  – whenever possible, select a cell type that is recA, recD, sbcB and deficient in all restriction systems.
    • Conveniently, EcoK, mcrB,C and mrr are all linked and often deleted together in strains
    • can get more than 100 fold difference in numbers of phage between wild type and recombination deficient
  – recD is preferred over recB,C because recD promotes rolling circle replication in lambda which improves yields
What do I need to know about E. coli genetics?

• You look in a supplier’s catalog and see lots of E. coli with different genotypes of the following general form:
  – F’{lacIq Tn10 (TetR)} mcrA, Δ(mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(StrR), endA1, nupG

• Does this make any difference for your experiments?
  – Or should you simply follow the supplier’s instructions?
  – Or just use whatever people in the next lab are using without thinking about it?
What do I need to know about E. coli genetics?

- $F'\{\text{lac}^{\text{Ig}} \text{Tn}10 (\text{Tet}^R)\}$ mcrA, Δ(mrr-\text{hsdRMS-mcrBC}), \Phi80\text{lacZΔM15}, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(StxlR), endA1, nupG

- restriction systems
  - mcrA - cuts $\text{Cm}^5\text{CGG}$
  - mcrB,C - complex cuts at $\text{Gm}^5\text{C}$
  - mrr - restricts 6-methyl adenine containing DNA
  - hsdRMS - EcoK restriction system
    - R cuts $5'-\text{AAC(N)}_6\text{GTGC-3'}$
    - M/S methylates A residues in this sequence

- for stability of long repeated sequences
  - recA1 - deficient in general recombination
  - recD - deficiency in Exonuclease V
  - sbcB,C - Exonuclease I
  - deoR - allows uptake of large DNA

- for lac color selection
  - lacZ ΔM15 either on F’ or on Φ80 prophage
  - lacIg - high level expression of lac repressor. Prevents leaky expression of promoters containing lac operator
What do I need to know about E. coli genetics? (contd)

- for high quality DNA preps
  - recA1 - deficient in general recombination
  - endA1 - deficient in endonuclease I

- if you buy ESTs from Research Genetics or Genome systems
  - tonA - resistant to bacteriophage T1

- for recombinant protein expression
  - lon - protease deficiency
  - OmpT - protease found in periplasmic space
  - ** most important protease inhibitor for E. coli protein preps is pepstatin A

- suppressors
  - supE - inserts glutamine at UAG (amber) codons
  - supF - inserts tyrosine at UAG (amber) codons
    - many older phages have S100\textsuperscript{am} which can only be suppressed by supF
Construction of cDNA libraries

- What is a cDNA library?

- What are they good for?
Construction of cDNA libraries

• What is a cDNA library?
  – Collection of DNA copies representing the expressed mRNA population of a cell, tissue, organ or embryo

• what are they good for?
  – Identifying and isolating expressed mRNAs
  – functional identification of gene products
  – cataloging expression patterns for a particular tissue
    • EST sequencing and microarray analysis
Determinants of library quality

• What constitutes a full-length cDNA?
  – Strictly it is an exact copy of the mRNA
  – full-length protein coding sequence considered acceptable for most purposes

• mRNA
  – full-length, capped mRNAs are critical to making full-length libraries
  – cytoplasmic mRNAs are best

• 1st strand synthesis
  – complete first strand needs to be synthesized
  – issues about enzymes

• 2nd strand synthesis
  – thought to be less important than 1st strand (probably not)

• choice of vector
  – plasmids are best for EST sequencing
  – phages are best for manual screening

• how will library quality be evaluated
  – test with 2,4,6,8 kb probes to ensure that these are well represented
cDNA synthesis

• Scheme
  – mRNA is isolated from source of interest
  – 1-2 ug is denatured and annealed to primer containing d(T)ₙ
  – reverse transcriptase copies mRNA into cDNA
  – DNA polymerase I and Rnase H convert remaining mRNA into DNA
  – cDNA is rendered blunt ended
  – linkers or adapters are added for cloning
  – cDNA is ligated into a suitable vector
  – vector is introduced into bacteria

• Caveats
  – there is lots of bad information out there
    • much is derived from vendors who want to increase sales of their enzymes or kits
  – all manufacturers do not make equal quality enzymes
  – most kits are optimized for speed at the expense of quality
  – small points can make a big difference in the final outcome
cDNA synthesis (contd)

- Preparation of mRNA
  - want minimum of non poly A+ mRNAs
  - affinity chromatography oligo d(T) or (U)
  - Oligo d(T)$_{30}$ latex (Nippon Roche) works best overall (a.k.a. OligoTex Quiagen)
  - 2 successive runs gives ~90% pure A+ mRNA

- denaturation of mRNA
  - critical step
  - most protocols use heat denaturation
  - CH$_3$HgOH is method of choice for best libraries

- First strand synthesis
  - lots of misinformation about enzymes
  - reverse transcriptase contains 2 subunits
    - polymerase
    - Rnase H
  - Rnase H subunit is critical for processivity of the enzyme!
  - Manufacturers want you to buy MMLV Rnase H- RT because it is cloned and almost free to make
  - best enzyme is AMV RT from Seikagaku America
  - thought that 1st strand is main failure point in cDNA synthesis - NOT
  - great improvement in 1st strand synthesis is addition of 0.6M trehalose to reaction
    - allows rxns to run at ~60 C
cDNA synthesis (contd)

- 2nd strand
  - must remove mRNA
  - best way is with RNAse H so that fragments serve as primers for DNA pol I
  - inclusion of RNAse H in 2nd strand reaction makes it silly to exclude RNAse H from RT
  - in my experience, 2nd strand synthesis is the point of failure in cDNA
    - virtually all kits shortcut this step (1-2 hrs)
    - should be overnight
    - recent improvement is to use thermostable RNAse H, DNA ligase and DNA polymerase to maximize production of 2nd strand.
cDNA synthesis (contd)
cDNA synthesis (contd)

- Cloning
  - after 2nd strand is made, the ends must be blunted and linkers or adapters added
    - usually T4 DNA polymerase
  - perfect cDNAs will retain 2-20 bp of RNA at the 5’ end.
    - Linkers can not be added to this by any DNA ligase!
    - But T4 RNA ligase can ligate DNA-RNA and stimulates blunt end ligation 10x
    - no commercial products use T4 RNA ligase so it is no wonder that full-length cDNAs are lost
  - if internal restriction sites have not been protected, they need to be methylated now before linkers are added.
    - Most methylase preps are not clean

\[ \text{mRNA} \]
\[ \text{cDNA} \]

\[ \text{mRNA!} \quad 2\text{nd strand cDNA} \]
\[ 1\text{st strand cDNA} \]
Full-length mRNA isolation and cDNA synthesis

- Ways to capture cap structures and presumably full-length mRNAs
  - affinity chromatography with eIF-4E (cap binding protein a.k.a. Capture)
  - selection with antibody to cap structure
  - oligo capping
  - biotinylated cap trapper

  - Principle is that uncapped mRNAs are dephosphorylated so that they cannot be ligated
  - cap structure is removed leaving only previously capped mRNAs with 5’ PO₄
  - RNA ligase can ligate a 5’-OH oligo to the 5’ end of the mRNA
  - This can be used to prime 2nd strand synthesis

<table>
<thead>
<tr>
<th>Classes of starting RNA</th>
<th>After BAP treatment</th>
<th>After TAP, only previously capped mRNAs carry phosphate</th>
<th>Only previously capped mRNAs will accept a linker ligation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>BAP treatment</td>
<td>TAP treatment</td>
<td>RNA ligase + (r-oligo) OH</td>
</tr>
<tr>
<td>Gppp</td>
<td>Gppp</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>ppp</td>
<td>pp</td>
<td>HO</td>
<td></td>
</tr>
<tr>
<td>pp</td>
<td>p</td>
<td>HO</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>HO</td>
<td>HO</td>
<td></td>
</tr>
<tr>
<td>HO</td>
<td></td>
<td>HO</td>
<td></td>
</tr>
</tbody>
</table>
Full-length mRNA isolation and cDNA synthesis (contd)

- 5’ oligo capping (contd)
  - advantages
    - very simple
    - no homopolymeric regions to worry about
    - can put arbitrary sequence at 5’ end.
      - Enables custom vector construction
      - also enables PCR to make driver for normalization
  - disadvantages
    - cap trapper paper claims this method only gives 70% full-length cDNAs
    - high quality TAP is not easy to find
    - original paper used PCR between 5’ and 3’ primer to make cDNAs
      - PCR => bias!
Full-length mRNA isolation and cDNA synthesis (contd)

  - Principle is that a biotin residue is chemically added to the cap structure
  - approach
    - 1st strand cDNA is synthesized
    - treatment with RNase I cuts any cDNA:mRNA duplexes which are not absolutely complete
    - complete cDNAs are isolated by streptavidin chromatography
    - RNA is hydrolyzed
    - cDNA is tailed with dG
    - 2nd strand synthesis is primed with dC
    - adapter added
    - cloned
  - advantages
    - claimed to give 90% recovery of full-length cDNAs
    - lots of history at RIKEN
  - disadvantages
    - homopolymeric region may make functional analysis of cDNAs difficult or impossible
    - many steps -> points of failure
Full-length mRNA isolation and cDNA synthesis (contd)
Full-length mRNA isolation and cDNA synthesis (contd)

- Cloning of cDNAs
  - most methods require linker or adapter addition followed by restriction digestion
  - relies on methylation to protect internal sites or use of rare cutters
  - A new alternative is ExoIII-mediated subcloning
    - no methylation
    - no restriction digestion
    - no ligation
    - no multimerization of vector or inserts
    - 100% oriented

```
<table>
<thead>
<tr>
<th>full-length mRNA iso</th>
<th>cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTGCAGGATCCCTGCAG</td>
<td>A\textsubscript{10}CTCGAGGCTCTAGAACTATCTGAGCTCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
</tr>
<tr>
<td>AAGCTCTAGGAGATGCTCCCTTGAGT</td>
<td>cDNA</td>
</tr>
</tbody>
</table>
```

after ExoIII digestion

```
<table>
<thead>
<tr>
<th>full-length mRNA iso</th>
<th>cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTTGCAGGATCCCTGCAG</td>
<td>A\textsubscript{10}CTCGAGGCTCTAGAACTATCTGAGCTCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
</tr>
<tr>
<td>TCTAGAAGCTCTAGGAGATGCTCCCTTGAGT</td>
<td>cDNA</td>
</tr>
</tbody>
</table>
```

vector polylinker

```
<table>
<thead>
<tr>
<th>vector polylinker</th>
<th>vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAATGGTCTTTTTTGGAGATCCCTGAGGAAATTCCTCGAGGATGTATATATATCATCTGTGGAGCAGCTCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
<tr>
<td>GTGAAACCAAGAAAAACGTGGGCCGCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
</tbody>
</table>
```

vector after BamHI/ Xhol and ExoIII digestion

```
<table>
<thead>
<tr>
<th>vector before digestion</th>
<th>vector after digestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAATGGTCTTTTTTGGAGATCCCTGAGGAAATTCCTCGAGGATGTATATATATCATCTGTGGAGCAGCTCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
<tr>
<td>GTGAAACCAAGAAAAACGTGGGCCGCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
</tbody>
</table>
```

vector and cDNA after annealing

```
<table>
<thead>
<tr>
<th>vector before annealing</th>
<th>cDNA after annealing</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTAATGGTCTTTTTTGGAGATCCCTGAGGAAATTCCTCGAGGATGTATATATATCATCTGTGGAGCAGCTCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
<tr>
<td>GTGAAACCAAGAAAAACGTGGGCCGCCCTTAGCTGGAAGCAGTCGAGGATCTTTGAT</td>
<td></td>
</tr>
</tbody>
</table>
```

BioSci 203 lecture 18   page 26 ©copyright Bruce Blumberg 2001. All rights reserved
Vectors for cDNA cloning

- Plasmids vs phage
  - phage are preferred for high density manual screening with all types of probes
  - plasmids are preferred for functional screening
    - microinjection
    - transfection
    - panning
  - phage packaging and infection is more routinely efficient than electroporation
    - 10-100x better than best transformation frequency

- what will the library be used for?
  - Consider the intended use as well as other contemplated uses
    - will the library go to an EST project?
      - Plasmid
    - will it be screened manually
      - phage
    - or arrayed and screened on high density filters
      - plasmid
    - will we normalize it?
      - Probably plasmid
Vectors for cDNA cloning (contd)

- Analysis of cDNAs obtained
  - rate limiting step in clone analysis is getting them into a usable form
    - usually a plasmid
  - cloning is tedious, particularly if one has many positives
    - some tricks can be used but this is still the bottleneck

- in about 1985 or so, Stratagene introduced lambda ZAP
  - phage with an embedded plasmid and M13 packaging signals
  - plasmid can be automatically excised by adding a helper phage
    - gene II protein replicates plasmid into ss phagemid which is secreted
  - this was a major advance and many phage libraries today are made in ZAP or its derivatives
  - early protocols had problems with helper phage but this has been overcome

- later, others developed a Cre-lox based system
  - instead of M13 used loxP sites.
  - When Cre recombinase is added, recombination between the loxP sites excises a plasmid

- both methods work very well and make analysis of many clones very straightforward
Vectors for cDNA cloning (contd)

1. Construct DNA library
2. Isolate positive clone

3. Excise the pBluescript plasmid containing the cloned DNA insert by co-infection with helper phage
Vectors for cDNA cloning (contd)

- f1 origin allows single stranded DNA production from plasmid subclones
- SP6 promoter proximal to cloning sites allows unfused RNA synthesis
- Downstream Pme I site among rarest in cDNA, allowing full-length probe synthesis (8)