Module Overview

Day	Lecture	Lab
1	Introduction	DNA library synthesis (PCR)
2	SELEX I: Building a Library	DNA library purification (agarose gel electrophoresis)
3	SELEX II: Selecting RNA with target functionality	RNA library synthesis (*In vitro* transcription = IVT)
4	SELEX III: Technical advances & problem-solving	RNA purification and heme affinity selection
5	Characterizing aptamers	RNA to DNA by RT-PCR
6	Introduction to porphyrins: chemistry & biology	Post-selection IVT Journal Club 1
7	Aptamer applications in biology & technology	Aptamer binding assay
8	Aptamers as therapeutics	Journal Club 2

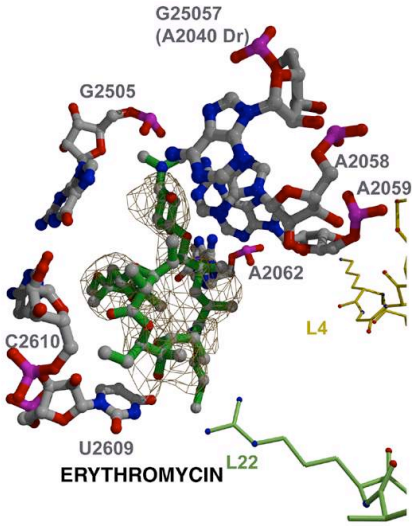
SELEX I

Building a Library

20.109 Lecture 2

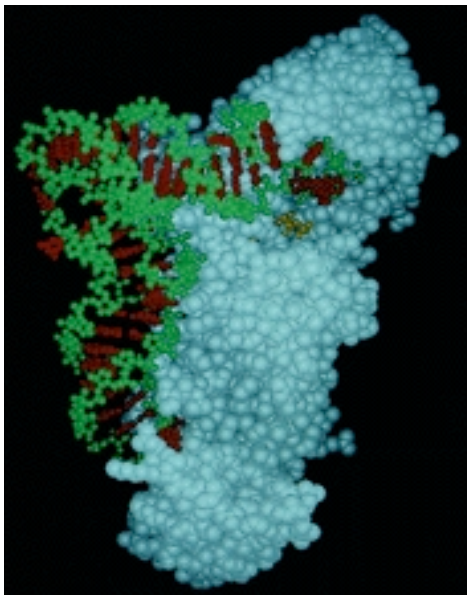
9 February, 2010

Last time …



23S rRNA: erythromycin

- ***Can we discover novel RNA molecules that interact with any target of interest?***



tRNA::aaRS

- In Nature, RNA interacts with both small molecules and proteins
- The 3D structure of the RNA permit stabilizing atomic contacts to be made
- Subtle differences in RNA 3D structure can lead to distinct binding partner interactions

Today's Objectives

- Better conceptualize the SELEX process for selecting RNAs with desired binding affinity (aptamers)
- Understand some basic principles influencing RNA library design
  - Appreciate how practical issues shape library architecture
  - Understand the concept of library diversity
  - Appreciate the limitations in building an ideal library

Discovering your desired RNA

1. Design-oriented approach

2. Selection-based approach

Discovering your desired RNA

"Design-oriented approach"

Decide on target function

Design specific RNA to meet function

Requires

1. *A priori* knowledge of target RNA structure required for function

2. Ability to predict RNA structure based on simple inputs (e.g. sequence)

Challenges

1. Difficult to predetermine the RNA structure required for function

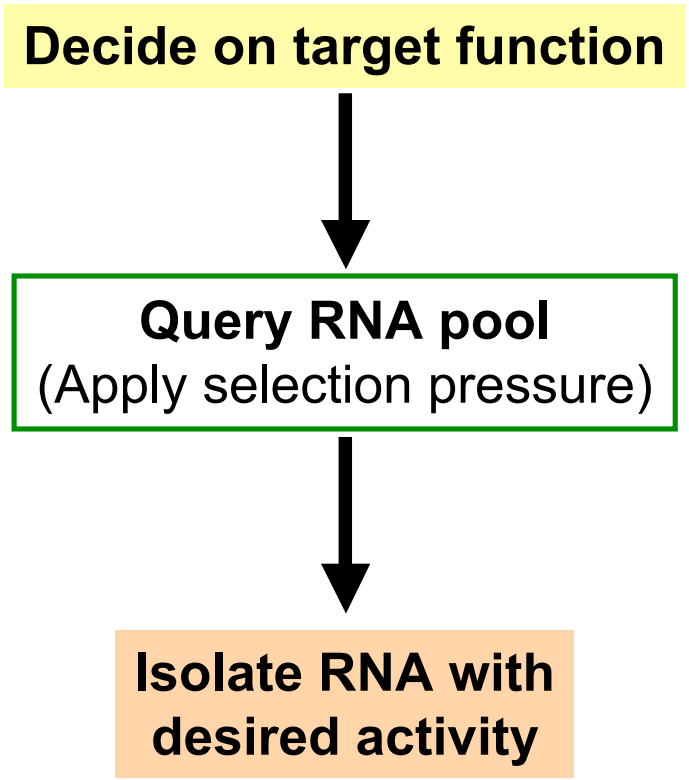
2. Cannot robustly use linear RNA sequence information to completely infer:

   • Structure

   • Function

Discovering your desired RNA

"Selection-oriented approach"



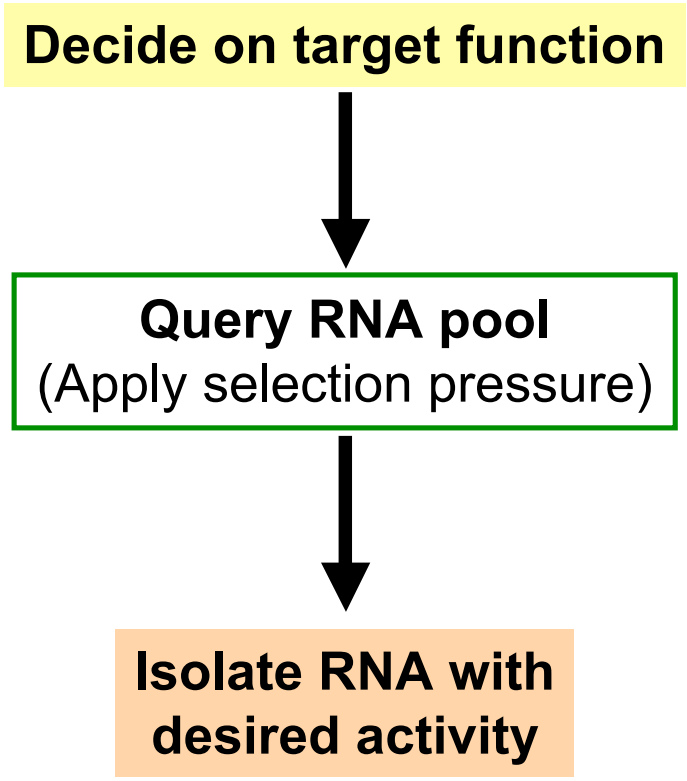
Requires:

1. Access to a sufficiently diverse RNA pool
   - Increased probability that the desired activity is present
2. Effective strategy for eliminating "losers" and enriching for "winners"

Discovering your desired RNA

"Selection-oriented approach"

Presently tenable



Advantages

1. No *a priori* knowledge of structure <=> function relationship required

2. Function drives emergence of a solution
   - By default, "winner" RNA has the requisite structure for function!

Discovering RNA with novel properties

- **SELEX**
  - ***S***ystematic ***E***volution of ***L***igands by ***EX***ponential enrichment
    - A selection-based strategy

Larry Gold
(U. Colorado)

Decide on target

↓

Query RNA pool
(Apply selection pressure)

↓

Isolate RNA with desired binding activity

– Derived from latin word "*aptus*" meaning "*to fit*"

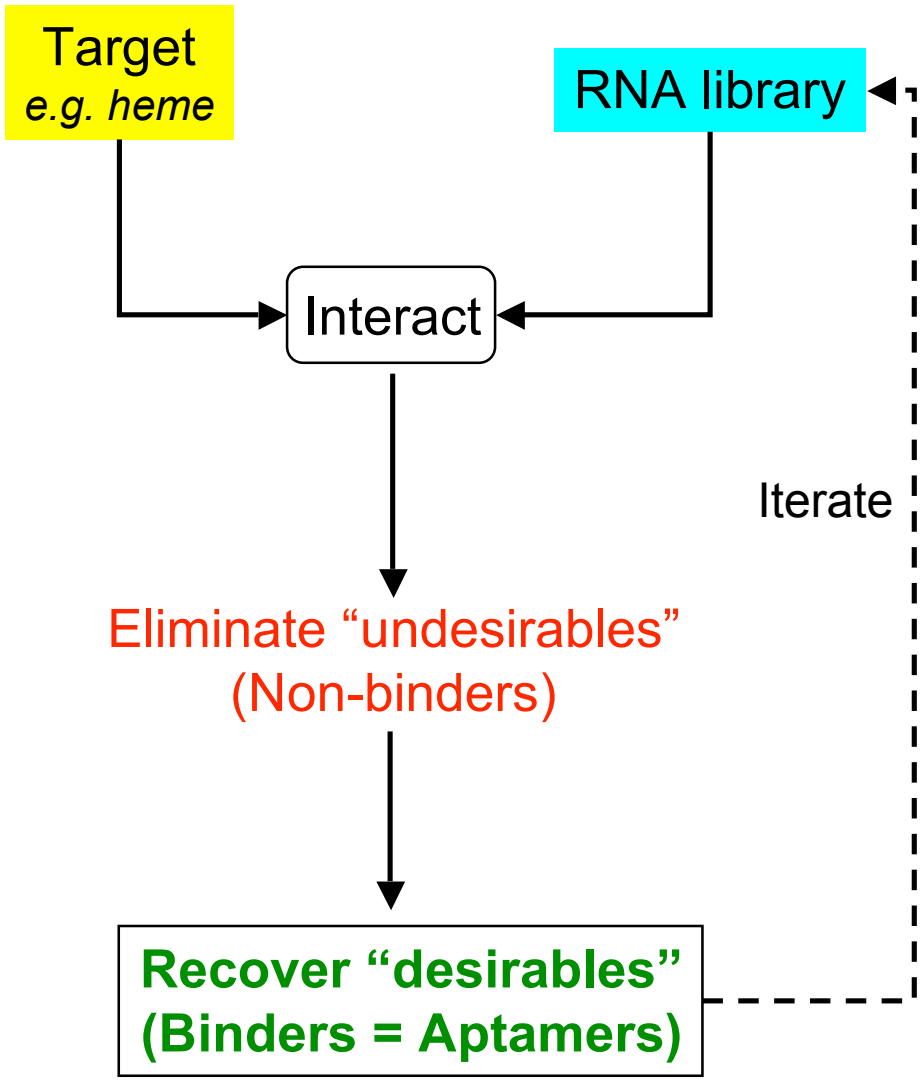
– RNA aptamer = RNA derived from a large pool having specific binding affinity for a target molecule

= *Aptamer*

Jack Szostak
(Harvard U.)

C. Tuerk and L. Gold; ***Science***; 249 (4968), 505-510, 1990.
A.D. Ellington and J.W. Szostak; ***Nature***; 346 (6287), 818-822, 1990.

SELEX: The process (simply)



- ***Materials***:
  - Target of interest
  - RNA library
- ***Need strategies for***:
  - Exposing target to library
  - Eliminating non-binders (partitioning step)
  - Recovering binders
  - Expanding recovered pool after each round

Conceptualizing SELEX

Molecular targets
e.g. heme

Majors

20
1
2
3
4
5
6
.
.
.
24

SELEX
Strategy for efficiently querying your RNA library

Barton
Strategy for efficiently querying the MIT Collections

MITLibraries

History
Science
Engineering
Sub-topic
Book X
Book Y
Book Z
Philosophy

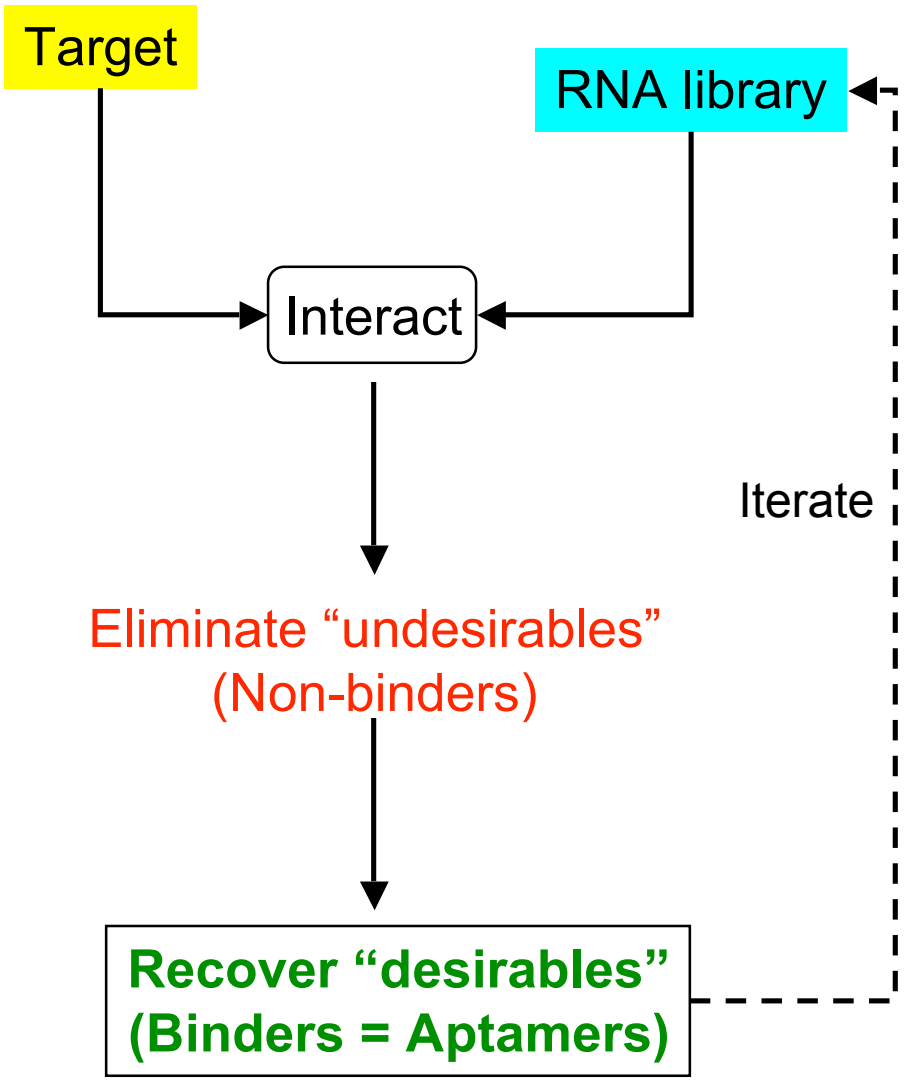
Book collection = RNA sequence collection

Book {x,y,z} = Aptamer

Student major dictates which books will be in demand

Populate library with "books" expected to be in demand by students

Target selection

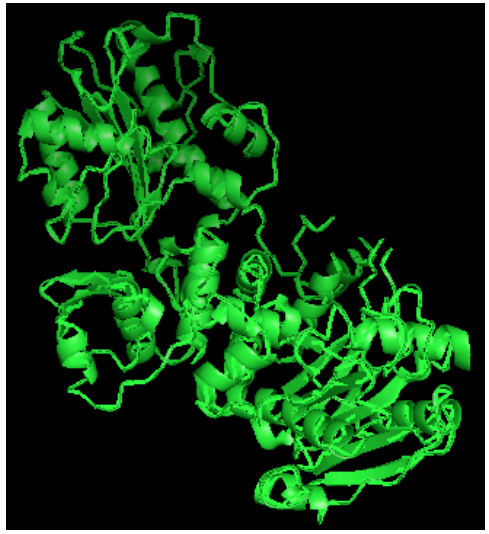


Target selection

Target

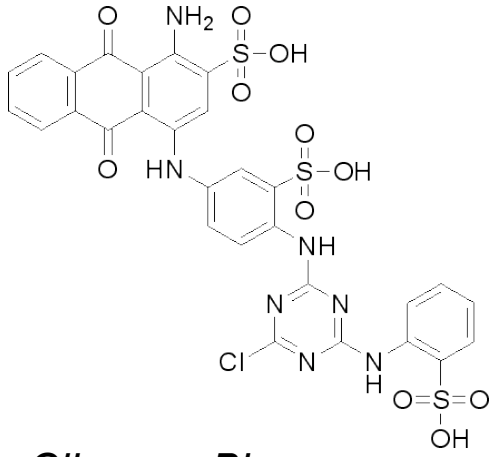
- The (mostly) trivial part
- Driven by investigator's interest(s)

RNA binding to protein



T4 DNA polymerase
Residues 1-388
(www.rcsb.org)

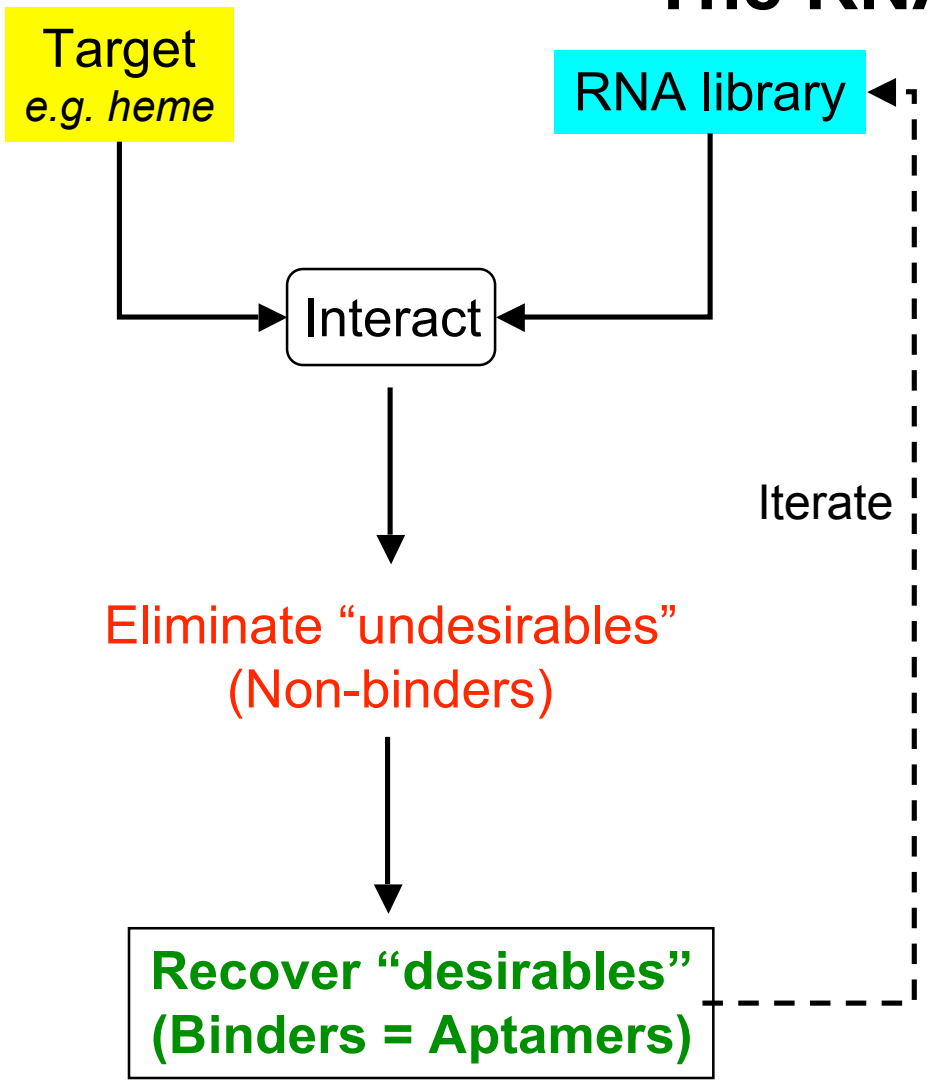
RNA binding to small molecule organic dyes



Cibracon Blue

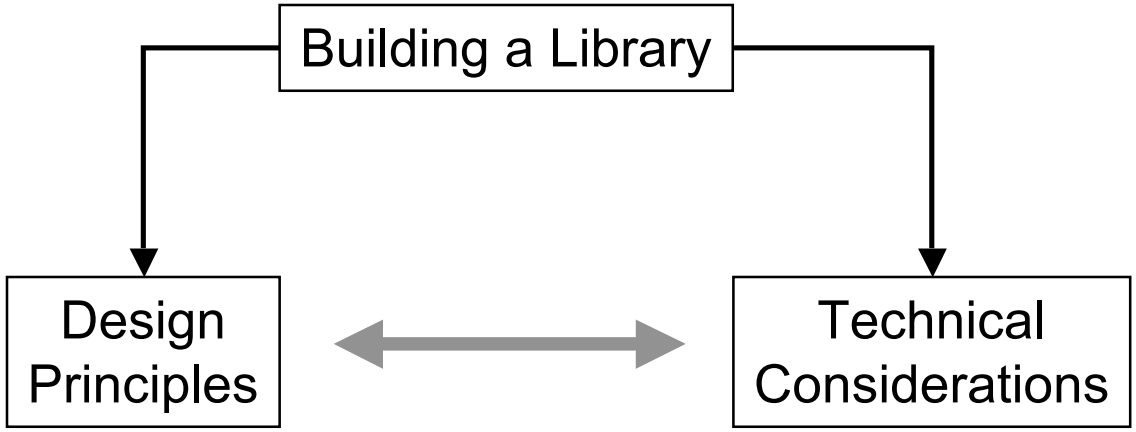
C. Tuerk and L. Gold; ***Science***; 249 (4968), 505-510, 1990.
A.D. Ellington and J.W. Szostak; ***Nature***; 346 (6287), 818-822, 1990.

The RNA Library



The RNA Library (abstracted)

RNA library



- One library per target **or** *one library for all targets*
- Balance between “useful” and “useless” library members
- Maximizing “useful” collection within space constraints

- Stability during storage
- Synthesizing library at reasonable costs
- Availability of efficient methods for manipulating library

- *Overall, library must be in a technical format compatible with all the steps involved in SELEX*

Technical considerations

- **Stability during storage**
  - DNA versus RNA?
  - DNA is more stable than RNA
    - RNA much more susceptible to hydrolysis than DNA;
      - Divalent metal catalyzed
    - RNA *highly* susceptible to ubiquitous RNases
  - *DNA is an excellent long-term form for stably storing library*

Technical considerations

- **Synthesis costs**
  - DNA

www.idtdna.com

Custom Oligonucleotide Synthesis

Desalted custom synthesized DNA oligos are shipped lyophilized or hydrated with Lab Ready Oligo Service. Synthesis scales up to 1 µmole are shipped the next business day. 5 µmole and 10 µmole scales are shipped within 5 business days.

Base Pricing		
Synthesis Scale	**Price**	
25 nmole DNA Oligo	$0.35 USD / Base	Order
100 nmole DNA oligo	$0.55 USD / Base	Order
250 nmole DNA oligo	$0.95 USD / Base	Order
1 µmole DNA oligo	$1.95 USD / Base	Order
5 µmole DNA oligo	$9.50 USD / Base	Order
10 µmole DNA oligo	$17.50 USD / Base	Order

– DNA oligo 100 bases long
– 1 µmol scale

Cost = 100 bases x $1.95/base
= $ 195

Technical considerations

- **Synthesis costs**

Custom RNA Synthesis and Purification

IDT has the expertise to deliver custom-synthesized RNA with the yield and purity that today's researcher demands. RNA is shipped deprotected and desalted in 2-3 business days or deprotected and purified in 4-6 business days. Please inquire for turnaround on 5 µmole and 10 µmole RNA synthesis.

Custom RNA Synthesis Pricing:					
	100 nmole	**250 nmole**	**1 µmole**	**5 µmole**	**10 µmole**
RNA bases	$6.50 USD	$8.50 USD	$18.00 USD	$60.00 USD	$115.00 USD

– RNA oligo 100 bases long
– 1 µmol scale

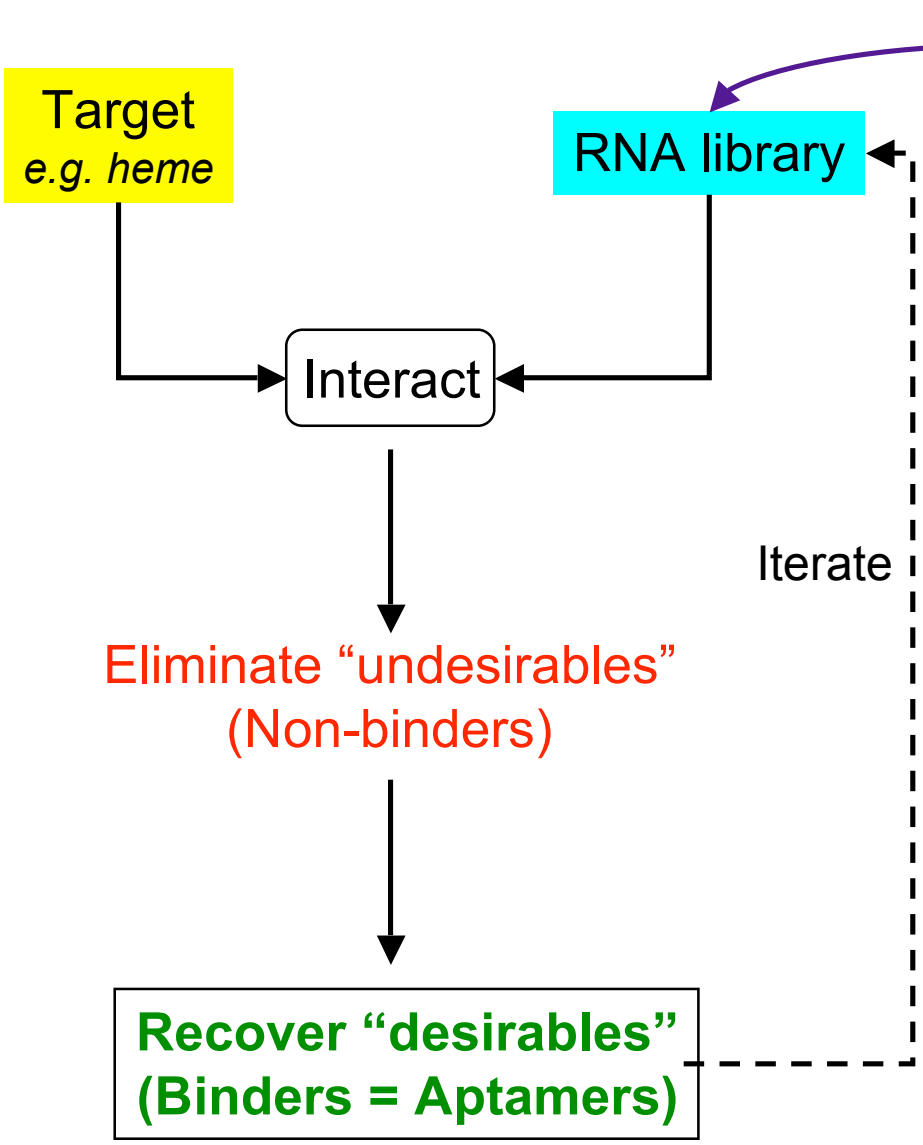
Cost = 100 bases x $18/base = *$ 1800*

www.idtdna.com

Technical considerations

- **Stability during storage**
  - *DNA is an excellent long-term form for stably storing library*
- Cost of synthesis
  - *DNA is more cost-effective and technically simpler to synthesize than RNA*
- **Two very compelling technical reasons for choosing DNA as the storage medium for your library!**

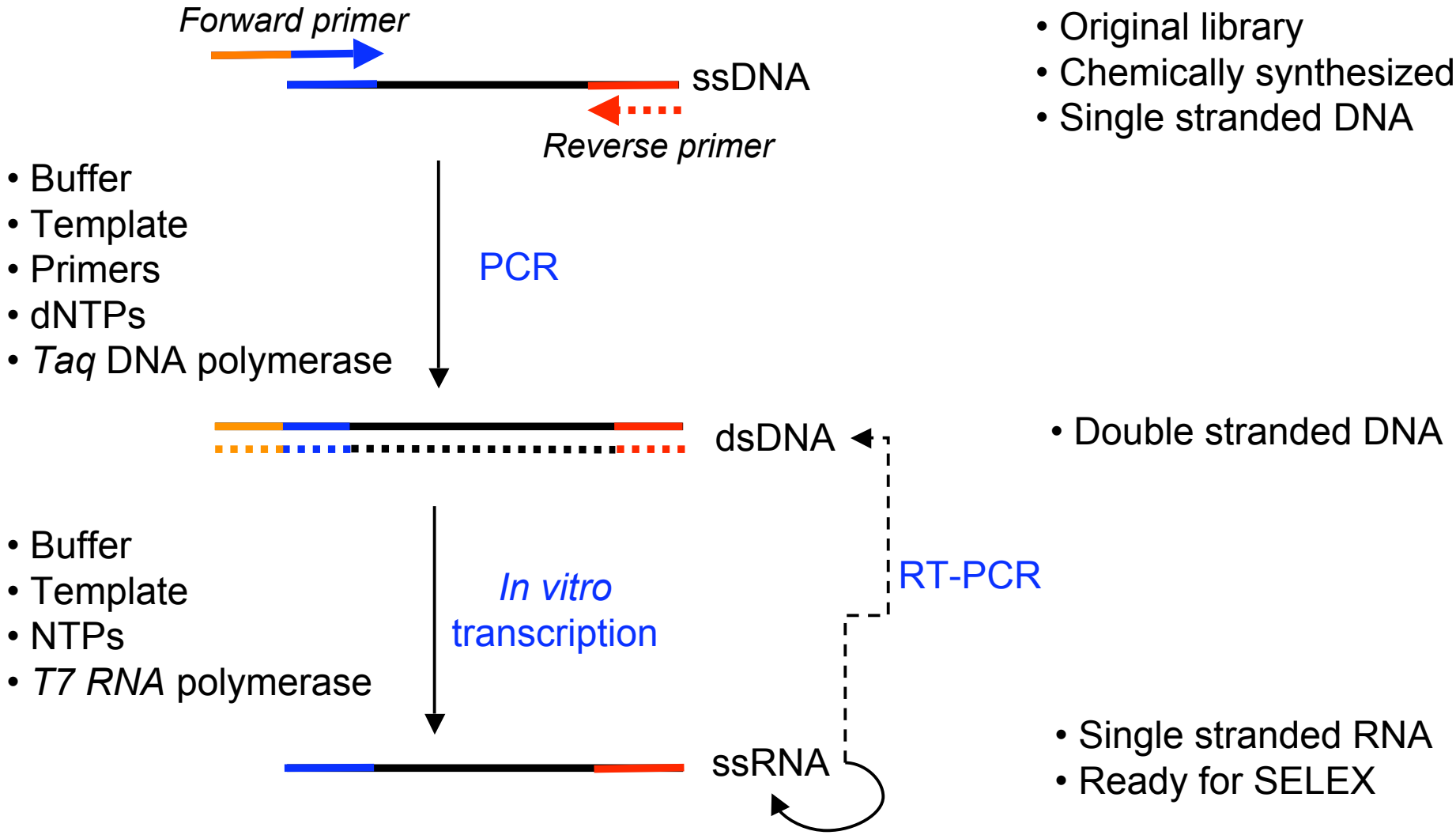
SELEX: The process (simply)



1. How do we derive the RNA library from our original DNA library?

2. How do we amplify the surviving RNA library between selection rounds?

SELEX: DNA Library --> RNA Library & Back



In vitro transcription

T7
promoter

dsDNA

- Buffer
- Template
- NTPs
- *T7 RNA* polymerase

In vitro
transcription

ssRNA

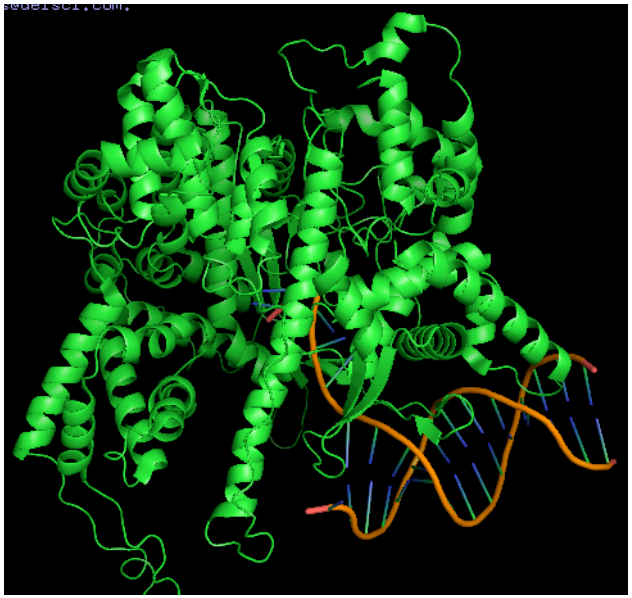
T7 promoter

TAATACGACTCACTATAGGGTACTT...

DNA sense strand

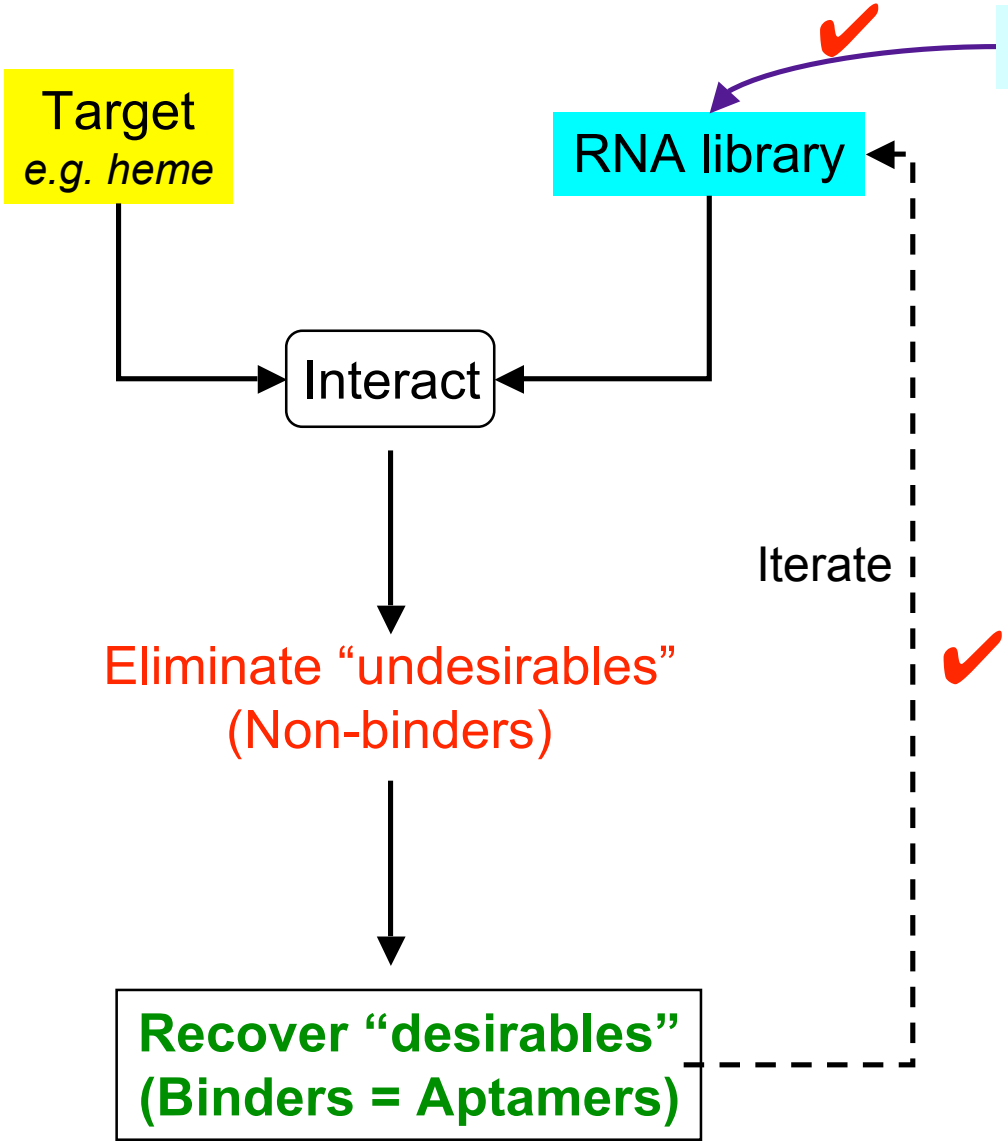
transcribe

GGGUACUU... *ssRNA*



T7 RNAP in complex with its promoter
PDB (www.rcsb.org)

SELEX: The process (simply)



1. How do we derive the RNA library from our original DNA library?

2. How do we amplify the surviving RNA library between selection rounds?

Overall architecture of ds DNA library

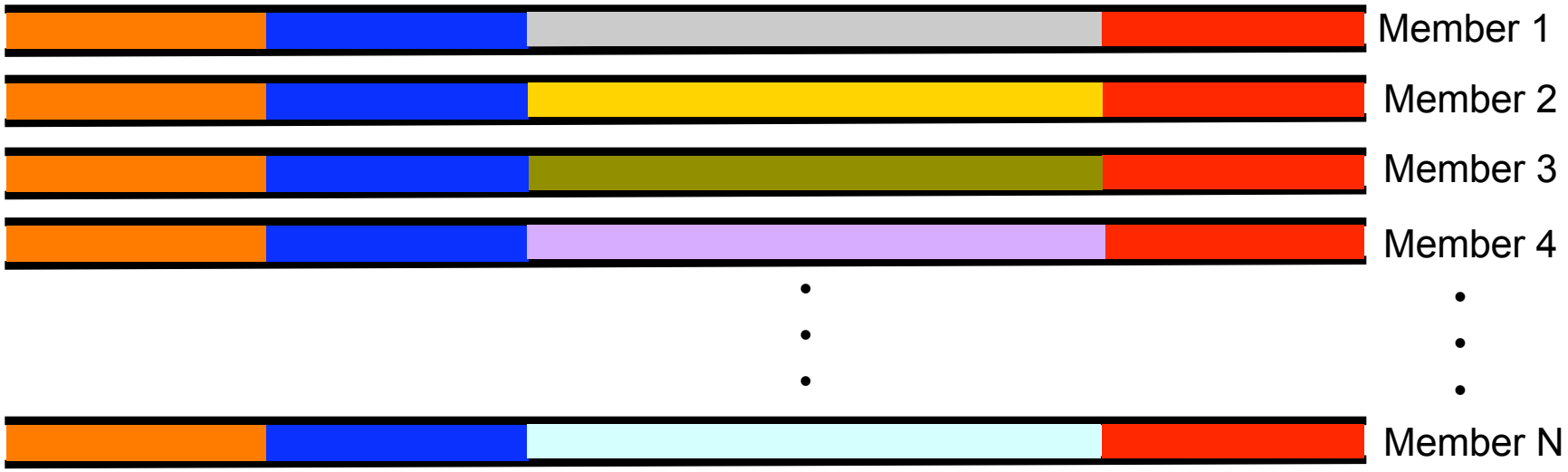
- Technical constraints dictate this architecture

How do we achieve variability between individual library members?

- **Each library member has a unique, defined sequence**
- **Members differ from each other in the variable region**

How do you synthesize such a library?

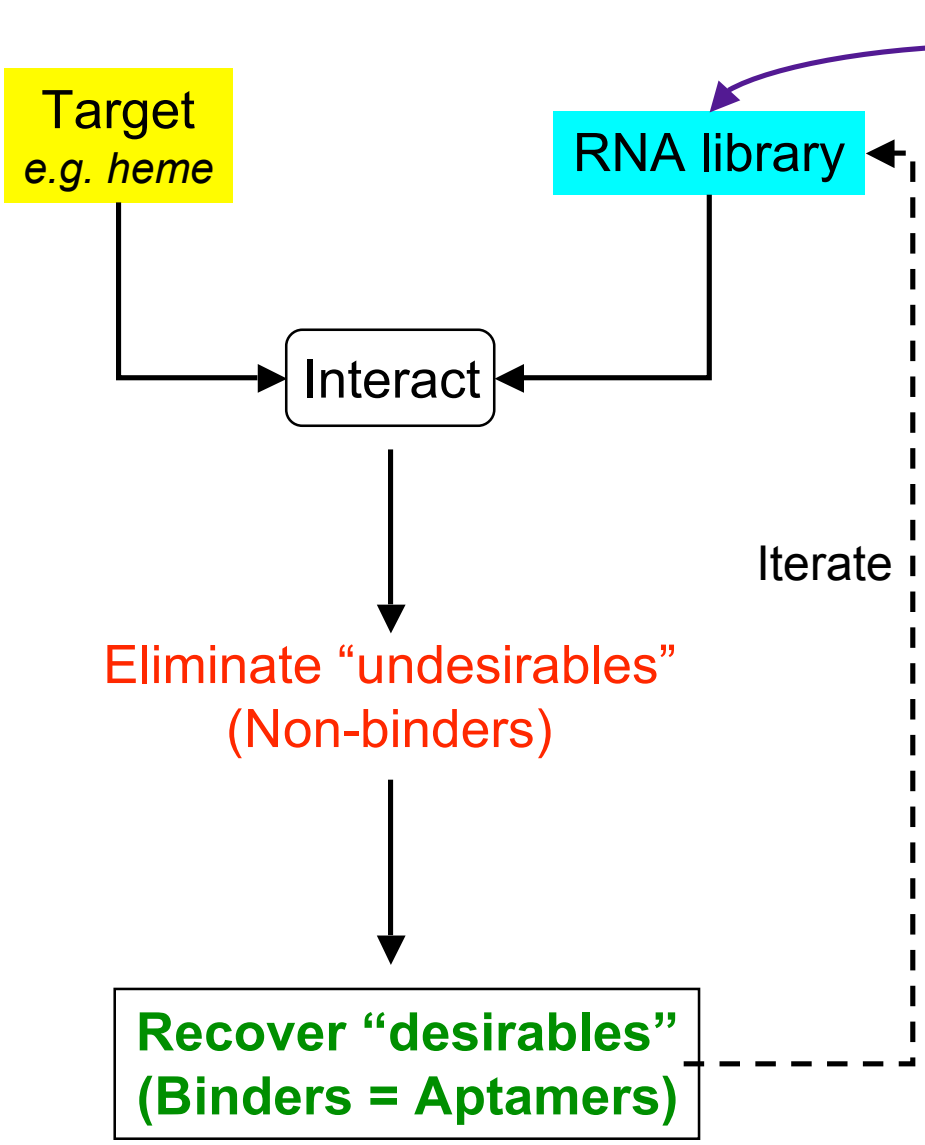
- **DNA synthesis is automated**
  - Program machine to add a specified base at a specified position
  - ***How do you build your target library?***



Exactly as you thought!

- **For fixed regions**:
  - Specify a single nucleotide to be added at that position
- **In the variable region**:
  - Mix the four nucleotides in equal “reactivity” proportions
  - Equal chance of either A, G, T or C being added at that position
  - Many distinct DNA oligonucleotides are being simultaneously synthesized

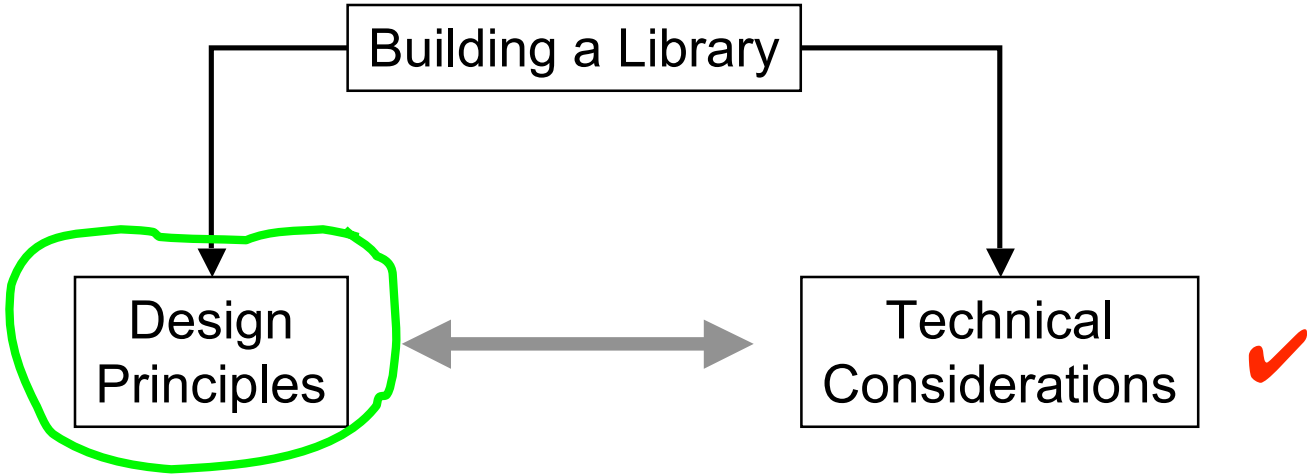
SELEX: The process (simply)



Now, that we understand the technical constraints for manipulating our library:

– *How do we achieve the higher-level desired properties?*

The RNA Library (abstracted)



- One library per target **or** *one library for all targets*
- Balance between “useful” and “useless” library members
- Maximizing “useful” collection within space constraints

- Stability during storage
- Synthesizing library at reasonable costs
- Availability of efficient methods for manipulating library

- *Now, let’s think about what we want in our library!*

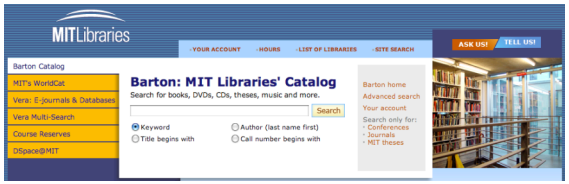
One master library or many libraries?

Molecular targets
e.g. heme

Majors

20
1
2
3
4
5
6
.
.
.
24

SELEX
Strategy for efficiently querying your RNA library



Barton
Strategy for efficiently querying the MIT Collections

MITLibraries

History
Science
Engineering
Philosophy

Book collection = RNA sequence collection

Book {x,y,z} = Aptamer

Sub-topic

Book X
Book Y
Book Z

- ***Known target with a general idea about what its partner RNA should look like --> "custom build" library***
- ***In absence of this data, build "generic" library***

Library design principles

Co-optimize several competing variables:

- **Diversity**
  - Maximize the number of distinct RNA sequences present
- **Space limitations**
  - Maximize the total number of RNA molecules present
  - Practical limitations exist (i.e. How much RNA can you reasonably prepare?)
- **Representation**
  - Each possible RNA sequence is present at least once
- **Adaptability**
  - Have an easy way for increasing the representation of “popular” RNA molecules = **SELEX**!
- **Easily replenished: Chemical synthesis; PCR; *in vitro* transcription**

Diversity

- **How can you increase diversity in your RNA library?**
  - Increase:
    - The length of the variable region;
    - The number of nucleotides from which to choose;
    - The molar quantity of library available (sometimes)
- **How do you calculate your library diversity?**
  - Distinguish theoretical *versus* actual

Calculating theoretical diversity

- Let's fix the nucleotides available = 4 (A, G, T, C)
  - 8 nucleotide variable region:
    - Maximum Diversity = Number of distinct sequences possible
    - = $(4)^8$ ~ 6.6 x 10^4 unique sequences
  - 20 nucleotide variable region:
    - Maximum Diversity = $(4)^{20}$ ~ 1 x 10^{12} unique sequences possible!
  - 50 nucleotide variable region:
    - Maximum Diversity = $(4)^{50}$ ~ 1.3 x 10^{30} unique sequences possible!!
- **Enormous theoretical diversity possible with nucleic acid libraries!**
  - 8 nucleotides (assuming a 5^{th} nucleotide option):
    - Maximum Diversity = $(5)^8$ = 4 x 10^5 unique sequences possible

Alas, there's only so much practical and affordable space for your library

- **How many unique sequences can be represented in this space?**

The Avogadro Constant:
= 6.022 x 10^{23} molecules/mol

(1 nmol = 1 x 10^{-9} mol)

Number of molecules in 1 nmol
~ (1 x 10^{-9} x 6.022 x 10^{23})
~ 6 x 10^{14} molecules!

IDT INTEGRATED DNA TECHNOLOGIES

Base Pricing		
Synthesis Scale	**Price**	
25 nmole DNA Oligo	$0.35 USD / Base	Order
100 nmole DNA oligo	$0.55 USD / Base	Order
250 nmole DNA oligo	$0.95 USD / Base	Order
1 µmole DNA oligo	$1.95 USD / Base	Order
5 µmole DNA oligo	$9.50 USD / Base	Order
10 µmole DNA oligo	$17.50 USD / Base	Order

- **1 µmol scale synthesis**
  - Nice compromise between cost and library mass obtained
  - On larger scale, downstream steps in library prep become limiting

- **From this scale synthesis:**
  - Obtain ~ 1 nmol full-length, useable library

So, what size library (diversity) fits comfortably into the practical space available?

- **Total space = 6 x 10^{14} molecules**
- 8 nucleotide variable region:
  - Number of distinct sequences possible
  - = $(4)^8$ ~ 6.6 x 10^4 unique sequences
- 20 nucleotide variable region:
  - Maximum Diversity = $(4)^{20}$ ~ 1 x 10^{12} unique sequences possible!
- 50 nucleotide variable region:
  - Maximum Diversity = $(4)^{50}$ ~ 1.3 x 10^{30} unique sequences possible!!
- *In which of these libraries can the theoretical diversity be fully represented given our space constraints?*

Representation

- **Total space = 6×10^{14} molecules**
- 8 nucleotide variable region:
  - Maximum Diversity = $(4)^8 \sim 6.6 \times 10^4$ unique sequences
  - Each sequence present @ $(6 \times 10^{14}/6.6 \times 10^4) \sim 1 \times 10^{10}$ copies/library
- 20 nucleotide variable region:
  - Maximum Diversity = $(4)^{20} \sim 1 \times 10^{12}$ unique sequences possible!
  - Each sequence present @ $(6 \times 10^{14}/1 \times 10^{12}) \sim 6 \times 10^2$ copies/library
- 50 nucleotide variable region:
  - Maximum Diversity = $(4)^{50} \sim 1.3 \times 10^{30}$ unique sequences possible!!
  - Each sequence present @ $(6 \times 10^{14}/1.3 \times 10^{30}) \leq 1$ copy/library!

How do you co-optimize across these parameters

Scenario I

- Maximize diversity
- Achieve full representation by ensuring you have the available space.
  - Choose 50-nucleotide variable region (assume 100-base oligo)
  - Require ~ 3 x 10^5 metric tons of oligonucleotide!!!
    - And that's to have each possible sequence represented once!
  - How much diversity is enough?
    - 8, 20 or 50 (or more?)-nucleotide variable region?
    - Can you determine this ahead of time for every possible target?

How do you co-optimize across these parameters

Scenario II

- Set space limit (i.e. reasonable cost)
- Maximize diversity (within this limit)
- Preserve representation at some acceptable (read: arbitrary) limit?
  - You'll saturate your space at ~ 23-nucleotide variable region (~ 10^{14} maximum diversity)
  - (Recall: For 1 µmol synthesis (yield: ~1 nmol) --> ~ 10^{14} molecules present)
  - Is this enough diversity?

How do you co-optimize across these parameters

Scenario III

- Set space limit (i.e. reasonable cost)
- Maximize diversity
- Sacrifice representation
  - A given sequence present only once (if at all) in library
  - Is this problematic?
  - What does this mean for library reuse?
    - Sampling without replacement

What's the best strategy for assembling your library?

Scenario I

- Maximize diversity
- Achieve full representation by ensuring you have the available space

Scenario II

- Set space limit (i.e. reasonable cost)
- Maximize diversity (within this limit)
- Preserve representation at some acceptable (read: arbitrary) limit?

Scenario III

- Set space limit (i.e. reasonable cost)
- Maximize diversity
- Sacrifice representation

The Answer? In the end, it's really up to you!

Summary

- Developed an conceptual framework for SELEX
- Examined some key steps involved in the process:
  - *Target selection*
  - *RNA library construction*
- Library diversity
  - *Calculations*
  - *Maximizing diversity within technical constraints*
  - *Wisely choosing the appropriate library for your needs!*