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

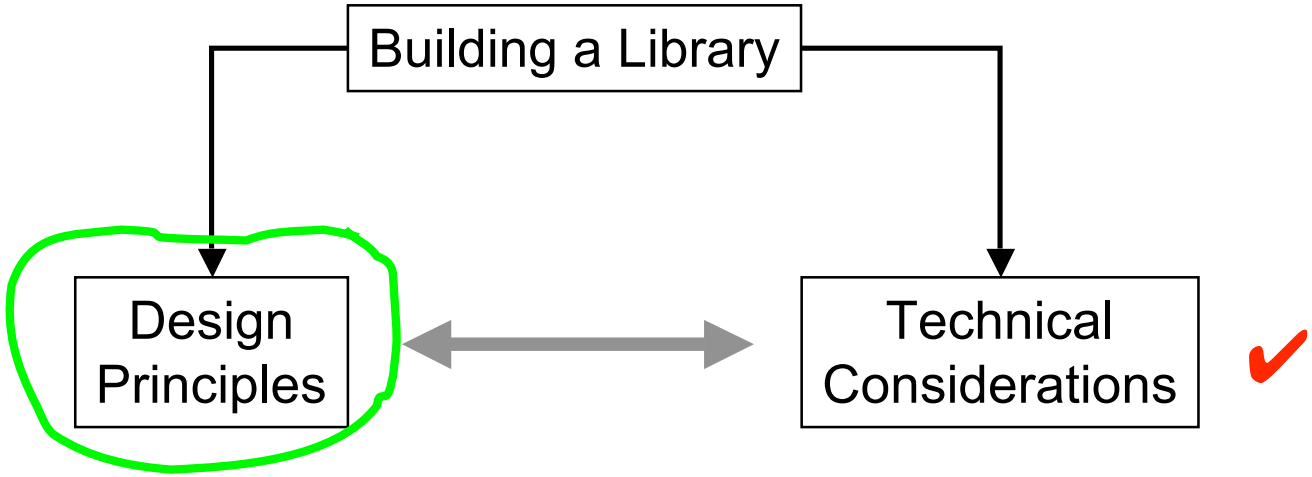
SELEX II

Selecting RNA with target functionality

20.109 Lecture 3

11 February, 2010

The 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

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

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

One master library or many libraries?

Molecular targets

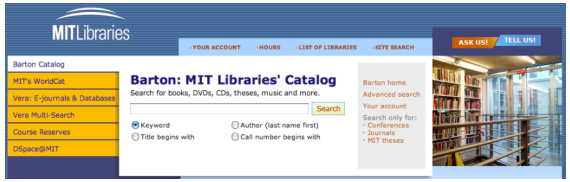
e.g. heme

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 @ $(6 \times 10^{14}/1.3 \times 10^{30})$: 0 or 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

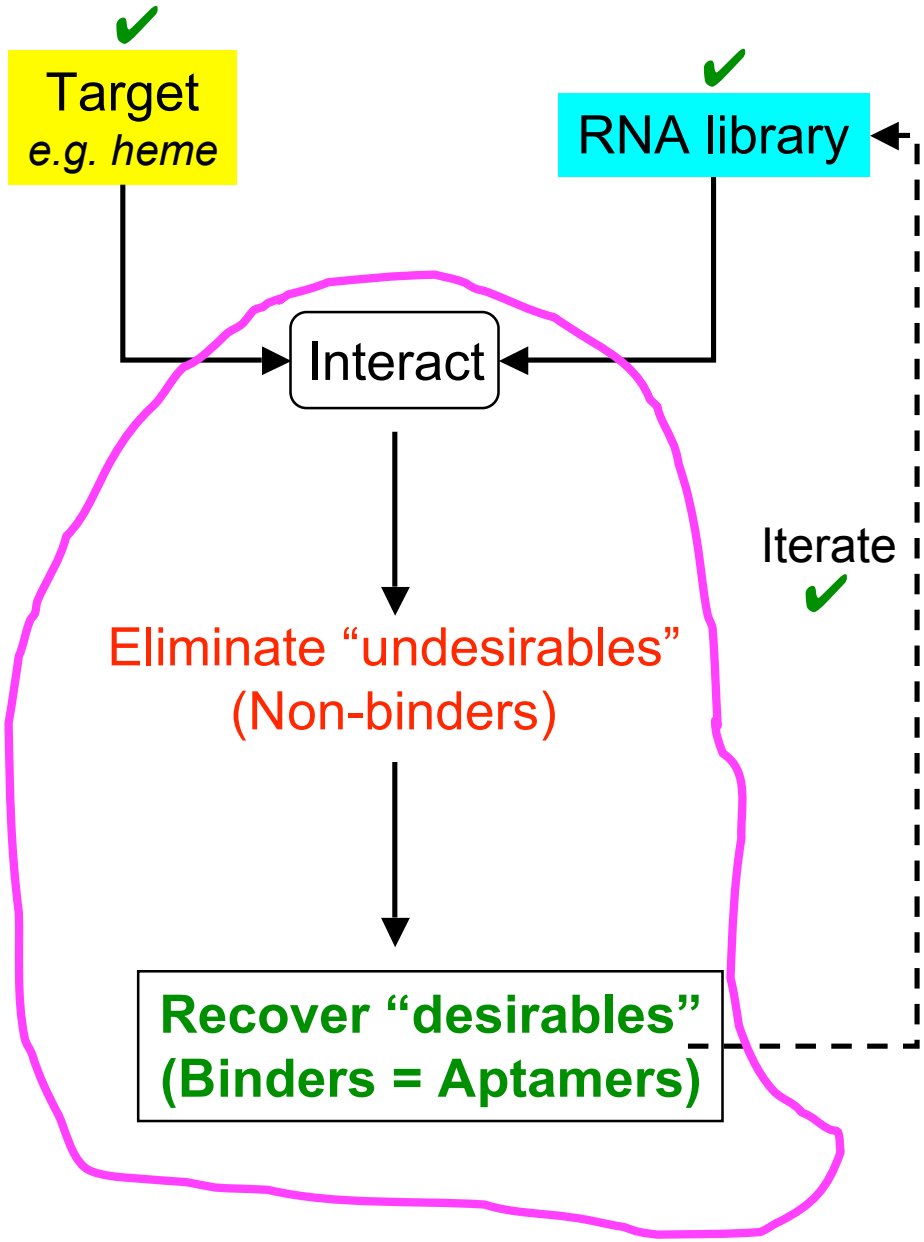
- Set space limit (i.e. reasonable cost)
- Maximize diversity (within this limit)
- Preserve representation at some acceptable (read: arbitrary) limit?

Scenario III

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

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

SELEX: The process (simply)



Now, that we understand:

- Target selection
- Library construction & manipulation

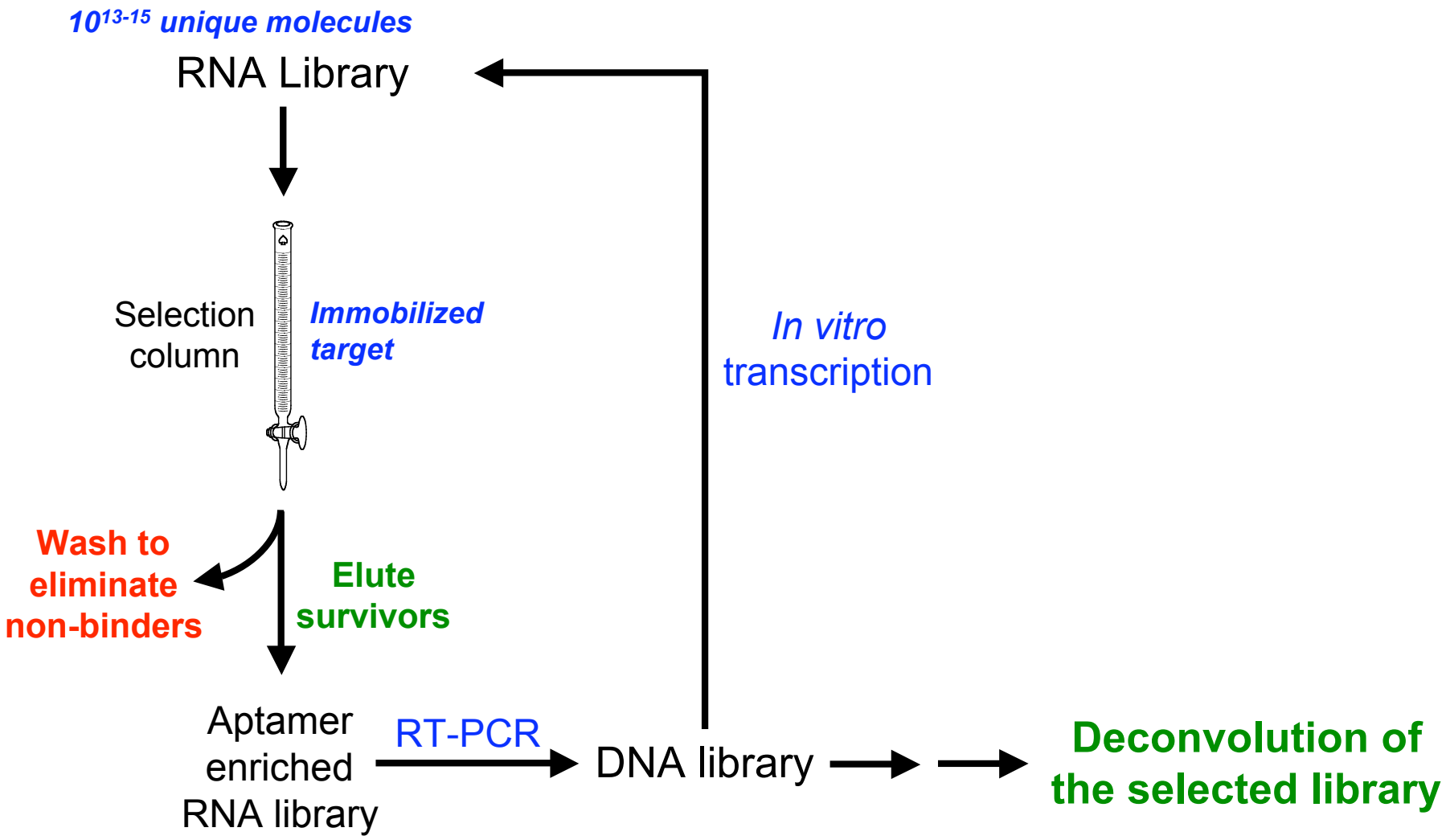
– How do we enrich for binders?

– *How do we put this all together into a workflow?*

Enriching your library for binders

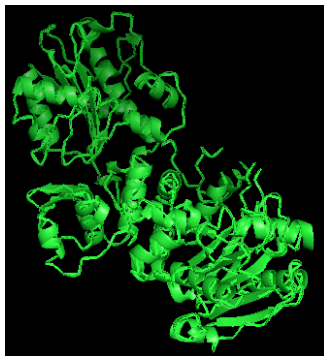
- Need a partitioning strategy:
  - Separate target bound RNA from unbound fraction
  - Selectively release target bound RNA
- Most commonly involves immobilizing target on:
  - A membrane (e.g. nitrocellulose)
  - Solid support (usually some kind of bead)
    - Column format
    - Magnetic separation
- Other approaches, for e.g.:
  - Electrophoretic methods to separate {RNA:target complex} from free RNA [J. Club paper on this].

Putting it all together: A typical SELEX workflow



SELEX *à la* Tuerk & Gold

T4 DNA polymerase

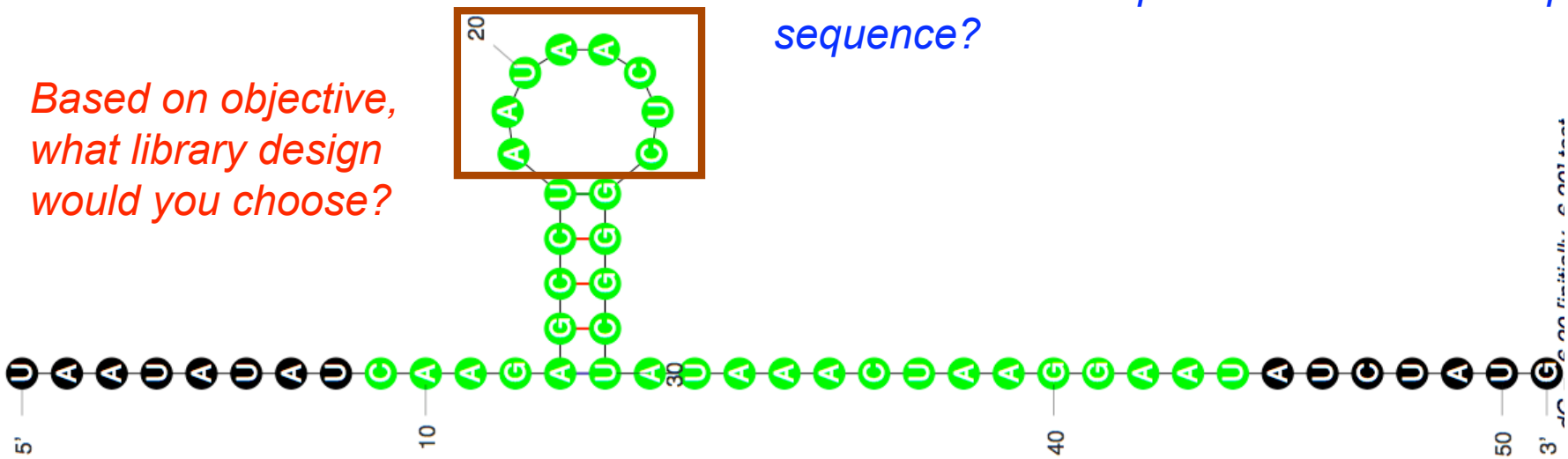


Target

Structure for residues 1-388 from the PDB (www.rcsb.org)

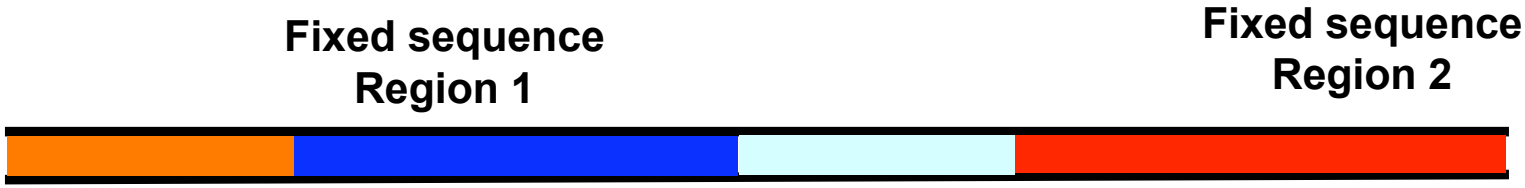
- Target known to interact with RNA from prior work
  - Sequence below found in the mRNA encoding the T4 DNA polymerase
  - Regulatory mechanism:
    - T4 DNA polymerase binds its own mRNA decreases its own synthesis
- 8 nucleotides [AAUAACUC] are critical for the interaction
  - *What underlies the preference for this loop sequence?*

Based on objective, what library design would you choose?



C. Tuerk and L. Gold; ***Science***; 249 (4968), 505-510, 1990

SELEX *à la* Tuerk & Gold



RNA library

T7 promoter

Variable Region:
= 8 nucleotides

- **Total space = 6 x 10^{14} molecules**
- 8 nucleotide variable region:
  - Maximum Diversity = $(4)^8$ ~ 65, 556 unique sequences
  - Each sequence present @ (6 x 10^{14}/~6.6 x 10^4) ~ 1 x 10^{10} copies/library
  - The known RNA target present @ 2 in 10^5 molecules!

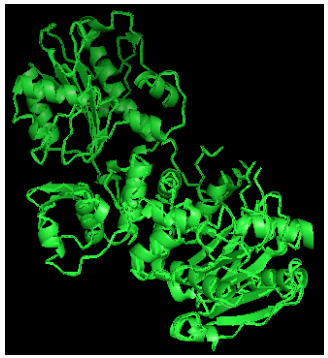
C. Tuerk and L. Gold; ***Science***; 249 (4968), 505-510, 1990

SELEX *à la* Tuerk & Gold

Immobilize on nitrocellulose

- Works well for many protein targets

T4 DNA polymerase



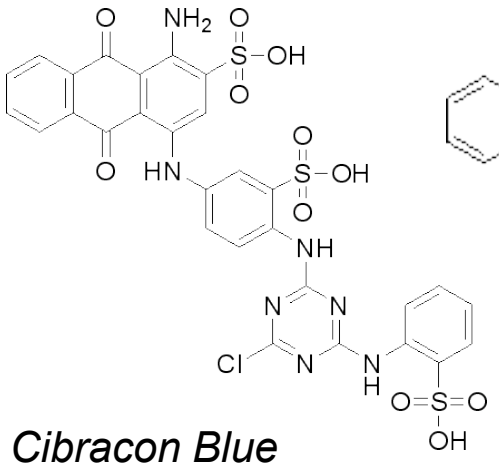
Structure for residues 1-388 from the PDB (www.rcsb.org)

- *Advantages*
  - Very easy and inexpensive!
  - Well-developed and straightforward protocols available
- *Disadvantages*
  - Protein can denature during immobilization step
    - Selected aptamers cannot recognize native protein
  - Not all proteins stick strongly enough to survive washing steps to remove unbound library

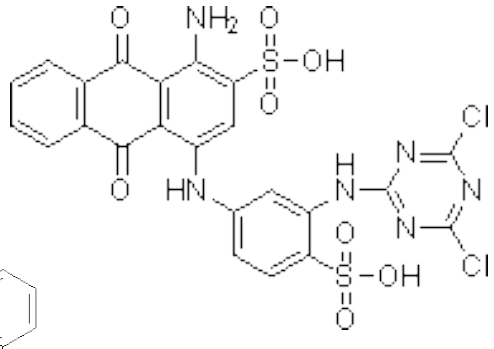
C. Tuerk and L. Gold; ***Science***; 249 (4968), 505-510, 1990

SELEX *à la* Ellington & Szostak

Target



Cibracon Blue



Reactive Blue 4

www.sigmaaldrich.com

- Discover RNA binding to small molecule organic dyes
  - No prior knowledge of their RNA binding capacity
- *Can RNA specifically interacting with these molecules be discovered?*

Based on objective, what library design would you choose?

A.D. Ellington and J.W. Szostak; ***Nature***; 346 (6287), 818-822, 1990

SELEX *à la* Tuerk & Gold

RNA library

T7 promoter

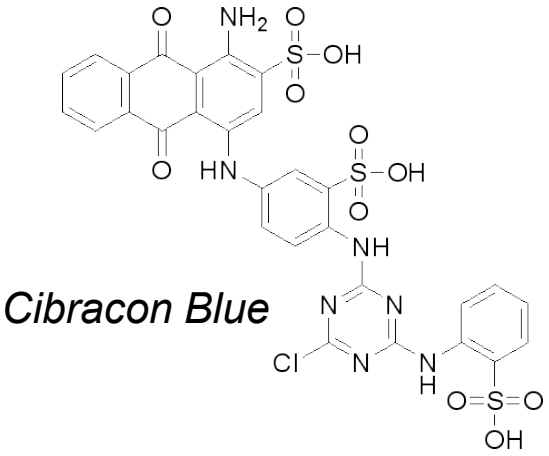
Variable Region:
= 100 nucleotides!

- **Total space ~ 6 x 10^{14} molecules**
- 100 nucleotide variable region:
  - Maximum Diversity = $(4)^{100}$ ~ **2 x 10^{60}** unique sequences possible!
  - Each sequence present @ (6 x 10^{14}/~2 x 10^{60}): Absent or 1copy/library
  - The known RNA target present @ *??? frequency*

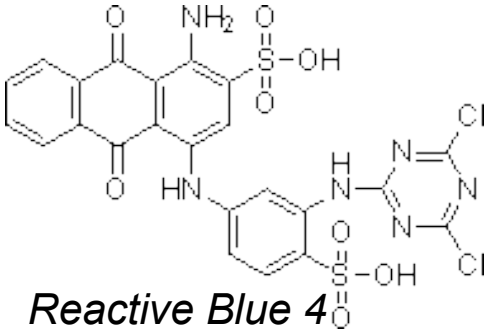
A.D. Ellington and J.W. Szostak; ***Nature***; 346 (6287), 818-822, 1990

SELEX *à la* Ellington & Szostak

Immobilize on agarose beads



Cibracon Blue



Reactive Blue 4

www.sigmaaldrich.com

- Very common strategy
  - Low molecular weight compounds
  - Macromolecules (e.g. proteins)

- *Advantages*
  - Extremely convenient and adaptable to many formats (e.g. column)
  - Better define how your target is displayed for binding (though not completely)

- *Disadvantages*
  - Not all immobilized molecules will be able to interact (even with its cognate RNA)
  - Immobilized form recognized is distinct from the free form of the target

A.D. Ellington and J.W. Szostak; ***Nature***; 346 (6287), 818-822, 1990

Summary

- Developed a conceptual framework for SELEX

- Library diversity
  - *Calculations*
  - *Maximizing diversity within technical constraints*
  - *Choosing the appropriate library for your needs!*

- Examined some key steps involved in the process:
  - *Target selection*
  - *RNA library construction*
  - *Partitioning strategies*

- SELEX can be successfully executed on:
  - *Very distinct targets*
  - *Using distinct library design (diversity, representation, etc)*
  - *Using distinct partitioning strategies*
  - *Fairly robust and generally applicable strategy*