DNA and RNA aptamers and their discovery using In-vivo Evolution & SELECT Protocols

From Lectures of: Daniel Kalderon & Larry Chasin, Columbia University and Peter Quin, USC

Topics:

- Aptamers
- SELECT
- DNA enzymes (DNAzymes)
- RNA enzymes (Ribozymes)

Nucleic acid aptamers

Aptamers: molecules that bind other molecules with good affinity and specificity.

- Usually these are proteins But they can also be RNA or DNA.
- That is, single stranded RNA or DNA molecules can and will fold up into secondary and tertiary structures depending on their sequence.
- DNA can be synthesized as very large numbers of different random sequences, and some may be aptamers for a specific target.

Some examples of aptamer targets

Small molecule Targets:

- Zn₂
- ATP
- adenosine
- cyclic AMP
- GDP
- FMN (RNA aptamer in E.coli)
- cocaine
- dopamine
- amino acids (arginine)
- porphyrin
- biotin
- organic dyes (cibacron blue, malachite green)
- neutral disaccharides (cellobiose & cellulose)
- oligopeptides
- aminoglycoside antibiotics (tobramycin)

Protein Targets

- thrombin
- HIVtat
- HIV rev
- Factor IX
- VEGF
- PDGF
- ricin
- large glycoproteins
- anthrax spores

Binding Affinity by Aptamers to their Targets

Table 1. Nucleic acid aptamers for which three-dimensional structures have been determined. ND, not determined.

Ligand	Nucleic acid*	Affinity K_d [μ M]	3D structure†
Theophylline	RNA (4)	~0.3	NMR, 1EHT (5)
FMN	RNA (6)	~0.5	NMR, 1FMN (7)
AMP	DNA (9)	~6	NMR, 1AW4 (12)
	RNA (8)	\sim 10	NMR, 1AM0, 1RAW (10, 11)
Arginine	2 DNA (15)	~125	NMR, 10LD, 2ARG (18, 20)
	RNA (16)	~60	NMR, 1KOC (19)
Citrulline	RNA (16)	~65	NMR, 1KOD (19)
Tobramycin	2 RNA (25)	~ 0.009	NMR, 1TOB (32)
		\sim 0.012	NMR, 2TOB (33)
Neomycin B	RNA (26)	~ 0.115	NMR, 1NEM (34)
HIV-1 Rev peptide	2 RNA (40)	~ 0.004	NMR, 1ULL, 484D (41, 42)
HTLV-1 Rex peptide	RNA (43)	~ 0.025	NMR, 1C4J (44)
MS2 coat protein	3 RNA (45)	ND	X-ray, 5-7MSF (45, 46)
Thrombin	DNA (47)	~ 0.025	NMR, 148D (38); x-ray, 1HAO (39)

*The number of different sequences that have been studied is indicated. The structure determination method (e.g., NMR, nuclear magnetic resonance) and the Protein Data Bank entry for the atomic coordinates are given.

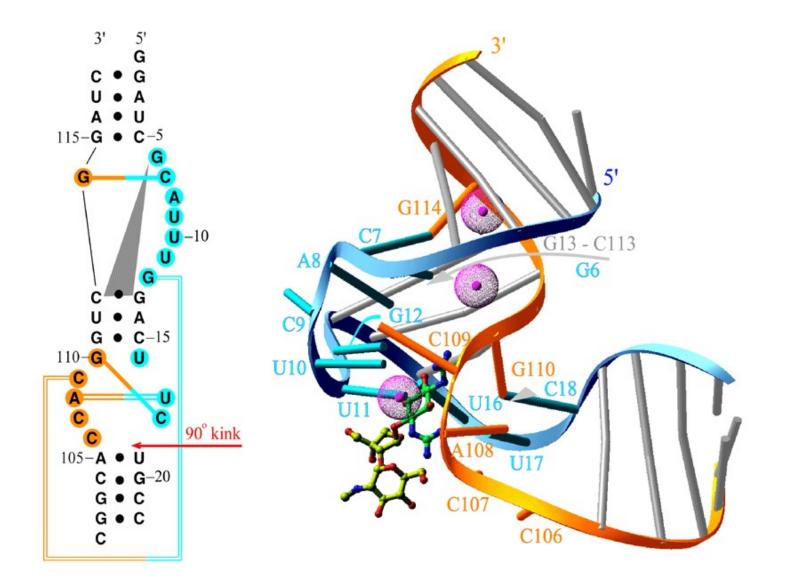
Hermann, T. and Patel, D.J.\2000. Adaptive recognition by nucleic acid aptamers. Science **287**: 820-825.

Example: AMP-binding aptamer

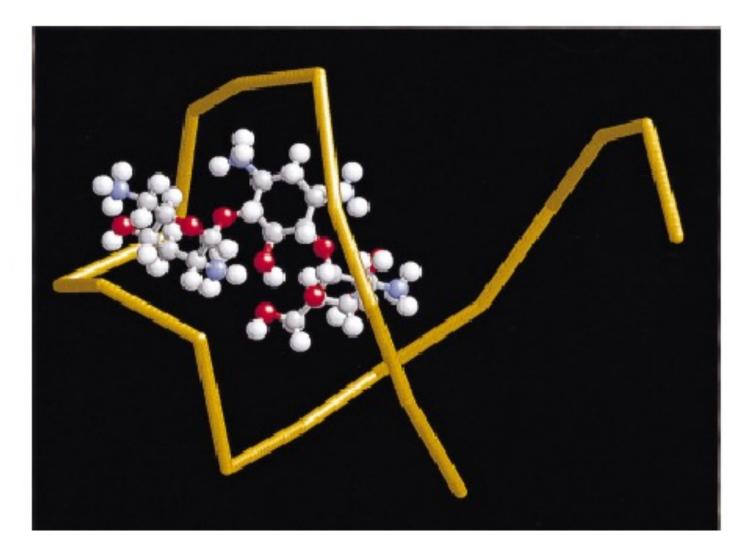
$$\begin{array}{c} & G & A \\ & 10 & A \\ & 10 & A \\ & A & AMP & C15 \\ & G & U \\ & 5 & G & U \\ & 5 & G & 20 \\ & 0 & 0 & G & CAC^U U^{25} \\ & G & G & U & G & G & CAC^U U^{25} \\ & C & C & A & C \\ & 40 & 35 & G & 30 & G \\ & & 34 & & 30 & G \end{array}$$

Daniel Kalderon & Larry Chasin, Columbia University

Example: Streptomycin-binding aptamer



Example: Tobramycin (antibiotic) bound to its aptamer (backbone)



Example:

An aptamer-gated nanorobot for targeted transport of molecular payloads

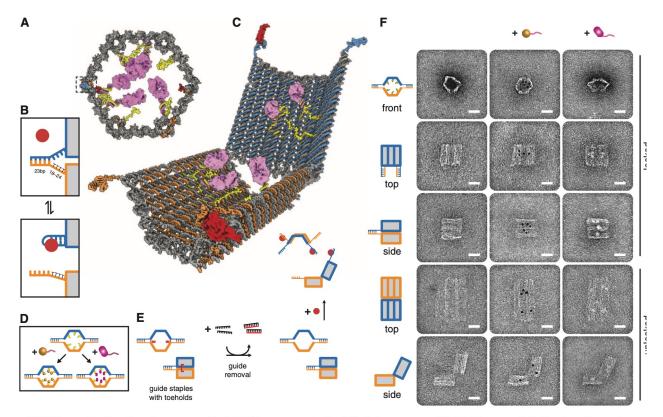
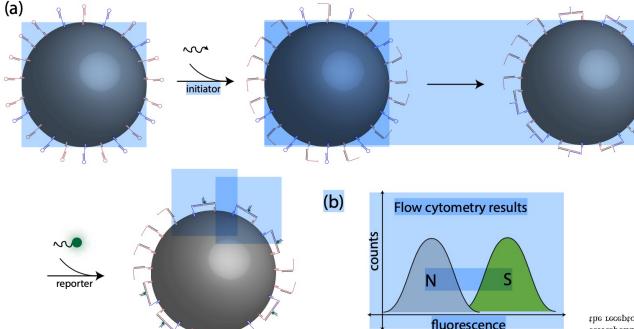


Fig. 1. Design and TEM analysis of aptamer-gated DNA nanorobot. **(A)** Schematic front orthographic view of closed nanorobot loaded with a protein payload. Two DNA-aptamer locks fasten the front of the device on the left (boxed) and right. **(B)** Aptamer lock mechanism, consisting of a DNA aptamer (blue) and a partially complementary strand (orange). The lock can be stabilized in a dissociated state by its antigen key (red). Unless otherwise noted, the lock duplex length is 24 bp, with an 18- to 24-base thymine spacer in the nonaptamer strand. **(C)** Perspective view of nanorobot opened by protein displacement of aptamer locks. The two domains (blue and orange) are constrained in the rear by

scaffold hinges. (**D**) Payloads such as gold nanoparticles (gold) and antibody Fab´ fragments (magenta) can be loaded inside the nanorobot. (**E**) Front and side views show guide staples (red) bearing 8-base toeholds aid assembly of nanorobot to 97.5% yield in closed state as assessed by manual counting. After folding, guide staples are removed by addition of fully complementary oligos (black). Nanorobots can be subsequently activated by interaction with antigen keys (red). (**F**) TEM images of robots in closed and open conformations. Left column, unloaded; center column, robots loaded with 5-nm gold nanoparticles; right column, robots loaded with Fab´ fragments. Scale bars, 20 nm.

Tiangi Song, Shalin Shah, Hieu Bui, Sudhanshu Garg, Abeer Eshra, Ming Yang, and John Reif, Programming DNA-Based Biomolecular Reaction Networks on Cancer Cell Membranes, Journal of the American Chemical Society (JACS), Vol. 141, No. 42, pp. 16539-16543. (Oct 2019).

the receptors. corresponding receptor, and the nodes can move on the cell membrane because of the mobility of Note that the density of each node on the cell membrane is determined by the density of the reaction network on the membrane. Group "N": cancer cells of the same type without fluorophore. and can be recognized by flow cytometry. (b) Group "S": cancer cells labeled by fluorophore via a strand (conjugated with a fluorophore), such that the cancer cells are labeled by the fluorophore up the blue hairpin. The output strand from the blue hairpin can hybridize with a reporter DNA the red hairpin by DNA strand displacement. Then, the output strand from the red hairpin opens the initiator strands, the 2-layer linear cascade reaction is started. First, the initiator opens up filter out the free nodes in the buffer to exclude potential non-localized reactions. By introducing membrane, both nodes will be localized on the membrane by aptamer-receptor binding. We then first mix the nodes with the cancer cells in a reaction buffer. If both targeted receptors exist on the using a DNA aptamer via aptamer-receptor binding. When operating the reaction network, we to proteins.^{30–38} Each node targets a designated cancer cell (the gray sphere) membrane receptor designed³⁴ and recognize a particular cell membrane receptor which can range from small molecules connected with a DNA aptamer, where a DNA aptamer is a DNA strand that can be rationally nodes in the network that are indicated by two different colors, where each node is a DNA hairpin by our architecture. This is a 2-layer linear cascade reaction network, and there are two types of Figure 1: A high-level description of our architecture. (a) An example reaction network



Cancer Cell Membranes Uses Aptamers to bind DNA hairpin devices to Cell Membrane Proteins

Example: Programming DNA-Based Biomolecular Reaction Networks on ⁹

But how can we discover aptamers?

=> Use Evolution !!

The RNA World Hypothesis

RNA has the essential properties needed for life: it can serve both as a repository of information (in its sequence of nucleotides) and as a catalyst.

RNA may have supported cellular or pre-cellular life – the "RNA World"

The RNA world evolved into the protein/DNA world of today.

Dinosaur 65 extinction Major 250 extinction Cambrian 550 radiation $\mathbf{0}_2$ 2.5 MILLION YEARS Abundant life 3.5 Oldest 3.8 rocks RMEnd of bombardment 4.5 BILLION BIG 15.8 Time lines

The RNA World, 2nd Ed.

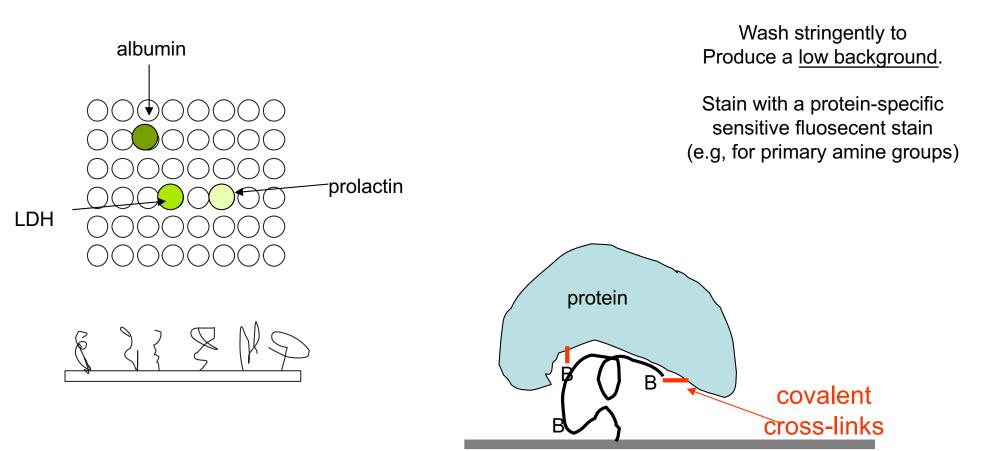
Evolution of Nucleic acid aptamers

- Aptamers can be selected from among these molecules based on their ability to bind an immobilized ligand. The tiny fraction found by chance to be able to bind to your favorite ligand can by amplified by PCR (along with background molecules).
- Re-iteration of the procedure will enrich for the aptamer until they dominate the population. At this point they can be cloned and sequenced.
- RNA molecules can be selected by synthesizing them from a randomized DNA population using the T7 promoter appended to each DNA molecule.
- This enrichment procedure is just the SELEX method described earlier for finding the RNA substrate for RNA binding proteins. In this case it's the same procedure, looked from the opposite point of view: not what RNA will the protein bind best, but what RNA binds the protein best.

SELEX Procedure for Evolving Aptamers

C. Tuerk and L. Gold. "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," Science, 249:505-10, 1990

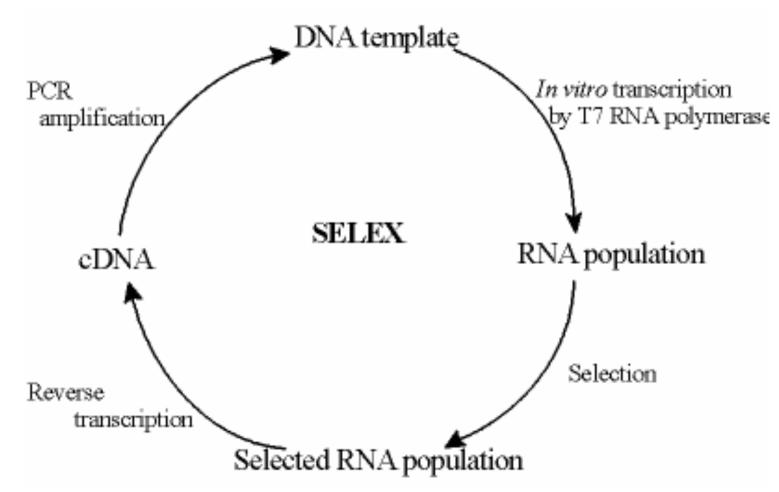
More recently: Somalogic, Inc.: Photoaptamers



In Vitro Selection & Evolution

- SELEX: Systematic Evolution of Ligands by EXponential enrichment. Also known as "In Vitro Selection" or "In Vitro Evolution".
- Allows simultaneous screening of > 10¹⁵ individual nucleic acid molecules for different functionalities.
- Developed in 1990 in the laboratories of Gerald F. Joyce (La Jolla), Jack W. Szostak (Boston), and Larry Gold (Boulder).
- > An important and wildly used tool in molecular biology.

for a review, see Silverman, 2003, RNA, <u>9</u>: 377-83



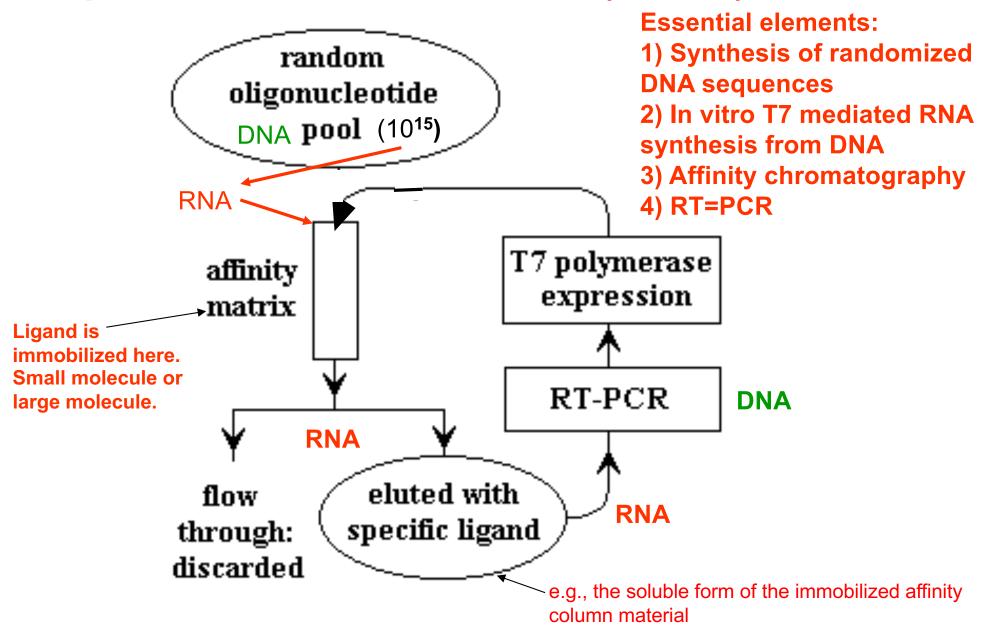
The general strategy:

- 1) Introduction of genetic variation or mutations.
- 2) Selection of variant molecules best suited for a target function.
- 3) Amplification of the selected molecules.

Application of SELEX:

- Engineering an enzyme with a novel function.
- Investigating the RNA world hypothesis.
- Designing molecules for clinical applications.

SELEX: Systematic Evolution of Ligands by Exponential Enrichment for RNA (or DNA)

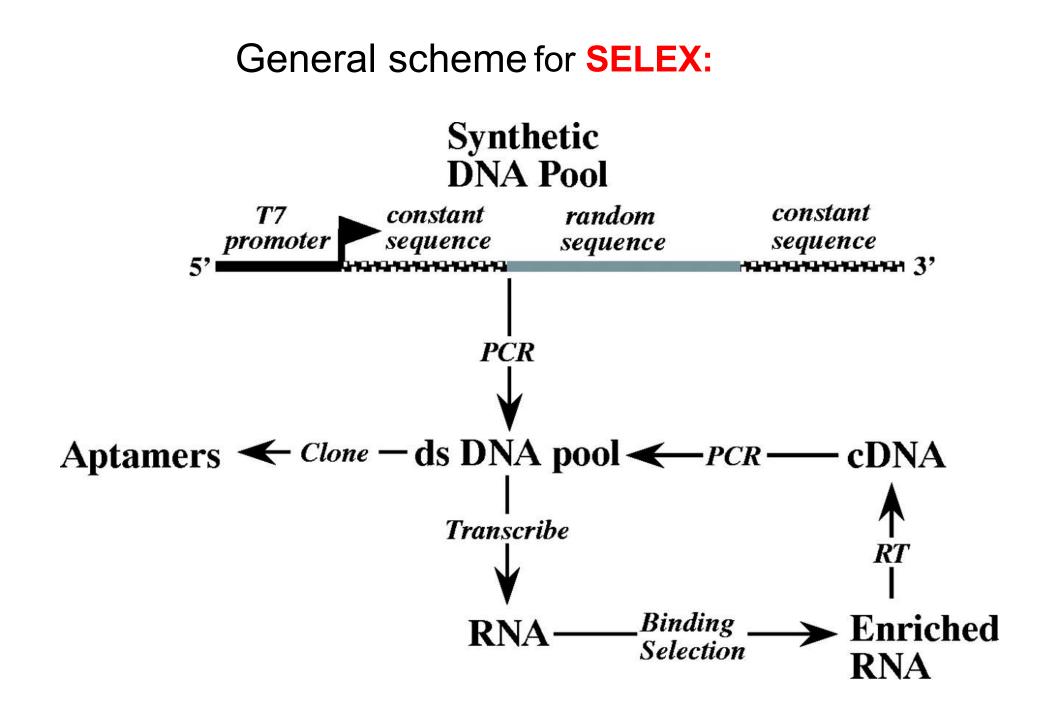


Building Random DNA Library for SELEX:

A library of random 40-mers (DNA sequences with 40 bases) synthesized, between 2 arbitrary 20-mers (PCR sites)

20-mer Random 40 20-mer

- Total number of possible 40-mers: $4^{40} = 10^{24}$
- Practical limit = 10^{15} = ~ 2 nmoles = ~ 50 ug DNA
- 10¹⁵ is a large number.
 Very large (e.g., 500,000 times as many as all the unique 40-mers in the human genome.)
- These 10¹⁵ sequences are known as "sequence space"
- Each DNA molecule of these 10¹⁵ (or RNA molecule copied from them) can fold into a particular 3-D structure, but we know little as yet about these structures.
- But can select the molecules that bind to our target by: AFFINITY CHROMATOGRAPHY

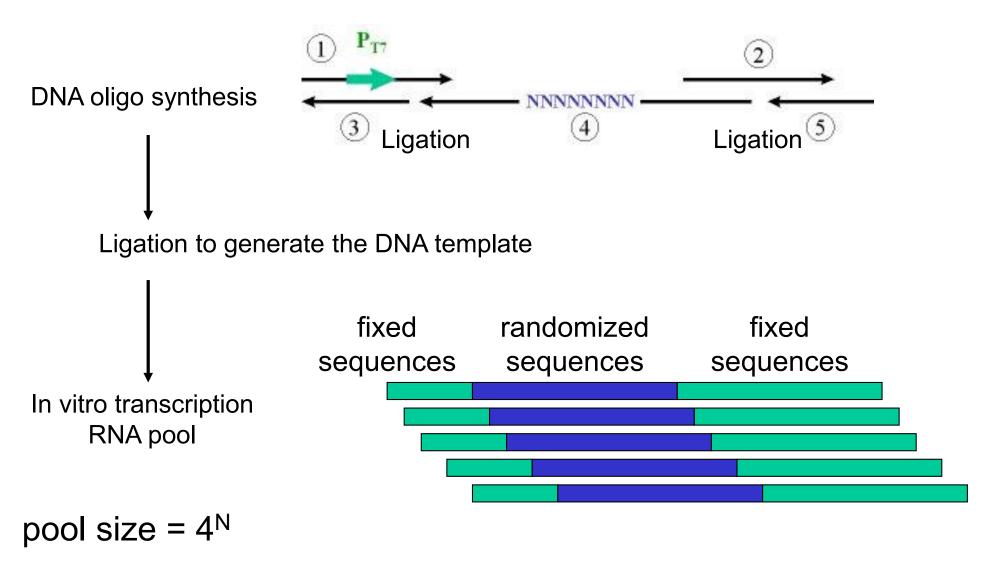


SELEX Example:

Example: Aptamer selected for binding to a protein, T4 DNA polymerase (Tuerk C, Gold L. 1990. *Science* 249:505-10)

 Selection via binding to T4 DNA polymerase wash immobilized on nitrocellulose filters • The filters are washed to remove non-specifically bound RNA molecules elute Tight (specific) binding **RNA** molecules are eluted specific from the T4 DNA pol binding sequences protein and are collected.

Construction of the randomized RNA pool for **SELEX**:

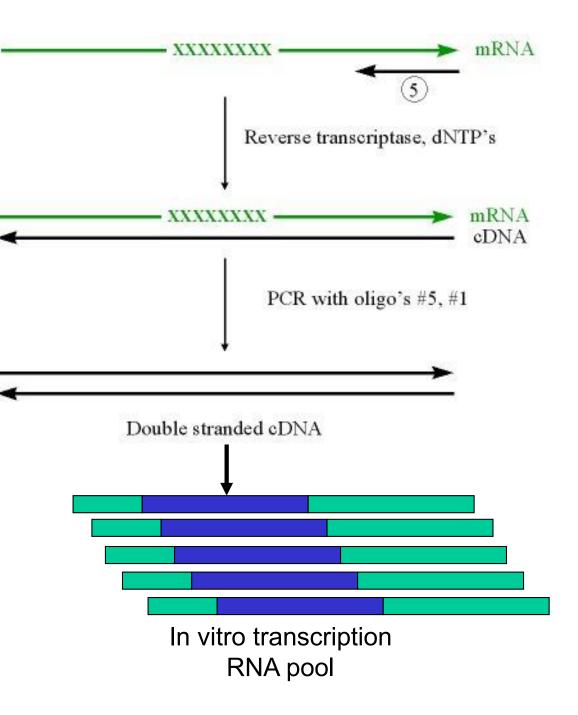


N = 8, pool = 65,536; N = 40, pool ~ 10²⁴; N = 100, pool ~ 10⁶⁰

Amplification: for **SELEX**:

 The eluted RNA molecules are converted to single strand cDNA using oligonucleotide #5 as a primer, and adding Reverse transcriptase and dNTP's. Duplex DNA is produced from the single strand cDNA by PCR using oligonucleotides #1 and #5. The duplex DNA is amplified via PCR using primer #1 and #5

•In vitro transcription generated a new pool of RNA with increased activity.



Results and Analyses for **SELEX**:

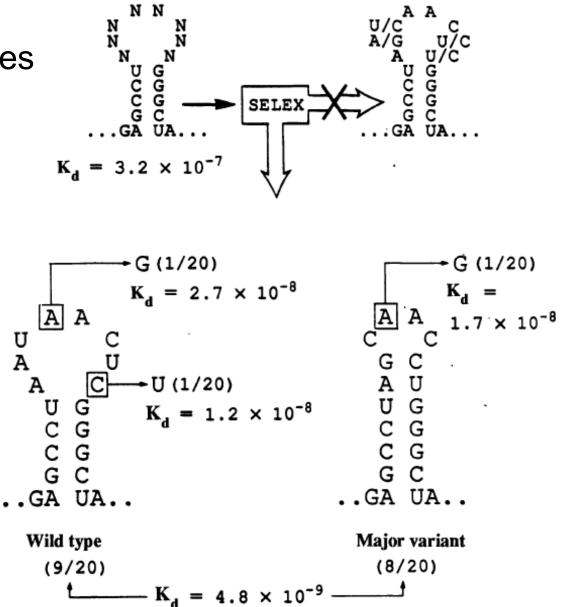


Fig. 8. Summary of results. The use of SELEX did not yield the apparent consensus one would expect from the batch sequences shown in Fig. 5, but yielded the wild-type and major variant species with three single mutants. The frequencies of each species in the 20 isolates tested are shown with the approximate K_d 's derived from the filter binding assays shown in Fig. 7.

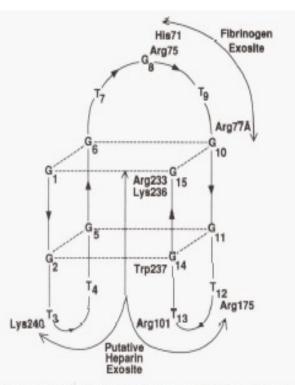


Fig. 1. Schematic of the folded 15-mer DNA and its interaction with α -thrombin.

G-quartets dominate the structure of antithrombin DNA aptamers

60-18(38)

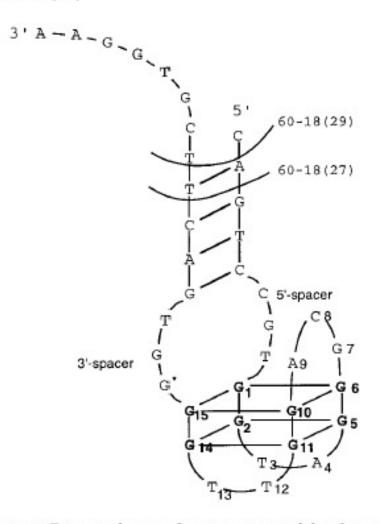


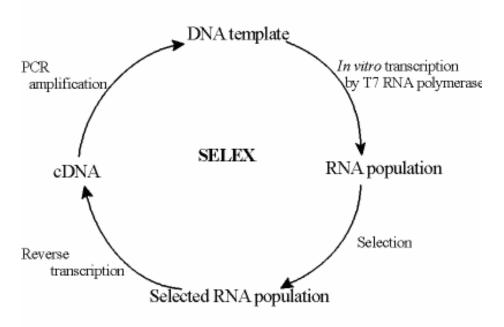
Figure 4. Proposed secondary structure of the thrombin DNA ligand 60-18[29] is shown. The 15 nucleotides of the G-quadruplex core sequence are numbered. Conserved G nucleotides required for G quadruplex formation are indicated in bold. The three nucleotides of the 5'- and 3'-spacer regions are indicated. An * indicates the first G residue of the 3'-spacer involved in high-affinity binding.

In vitro evolution of peptide/protein using **SELEX**:



✤ A peptide sequence <u>can</u> <u>NOT</u> "reverse-translate" to DNA/RNA, a different scheme of amplification is needed to link function (being selected) and sequence.

Current Methods: Phage display; mRNA display; …



SELECT for Ribozymes (RNA enzymes)

Ribozyme: a ribonucleic acid (RNA) enzyme that catalyzes a chemical reaction

Ribozymes = RNA enzymes

1982 Tom Cech: Tetrahymena rRNA intron is self-spliced out (Guanosine [GR] + Mg++)

Altman and Pace: Ribonuclease P is an RNP: RNA component alone can process the 5' ends of tRNAs

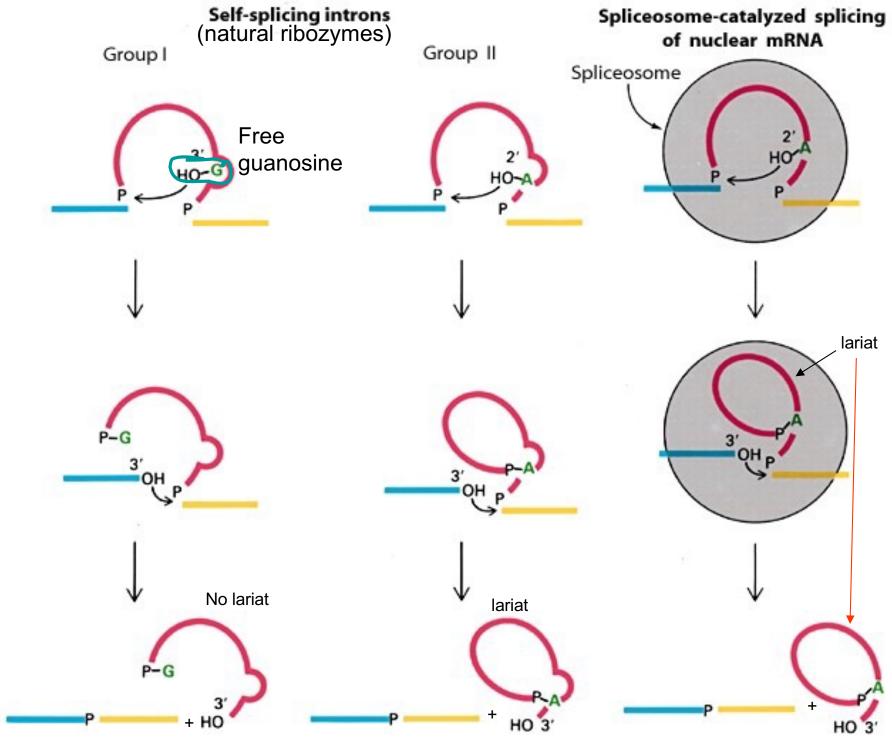
Mitochondrial group I introns (GR –catalyzed) also can self-splice

Then group II introns in mitochondria (lariat-formers)

Mutations (100's) revealed required attributes: Internal guide sequence GR-binding site secondary structure

Conserved base analysis (100's) \rightarrow confirms structure

X-ray diffraction: a few 3-D structures



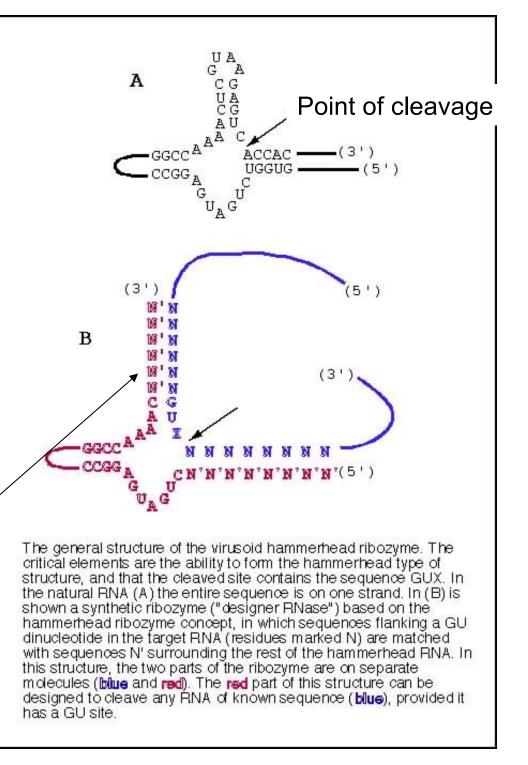
Daniel Kalderon & Larry Chasin, Columbia University

Hammerhead Ribozymes:

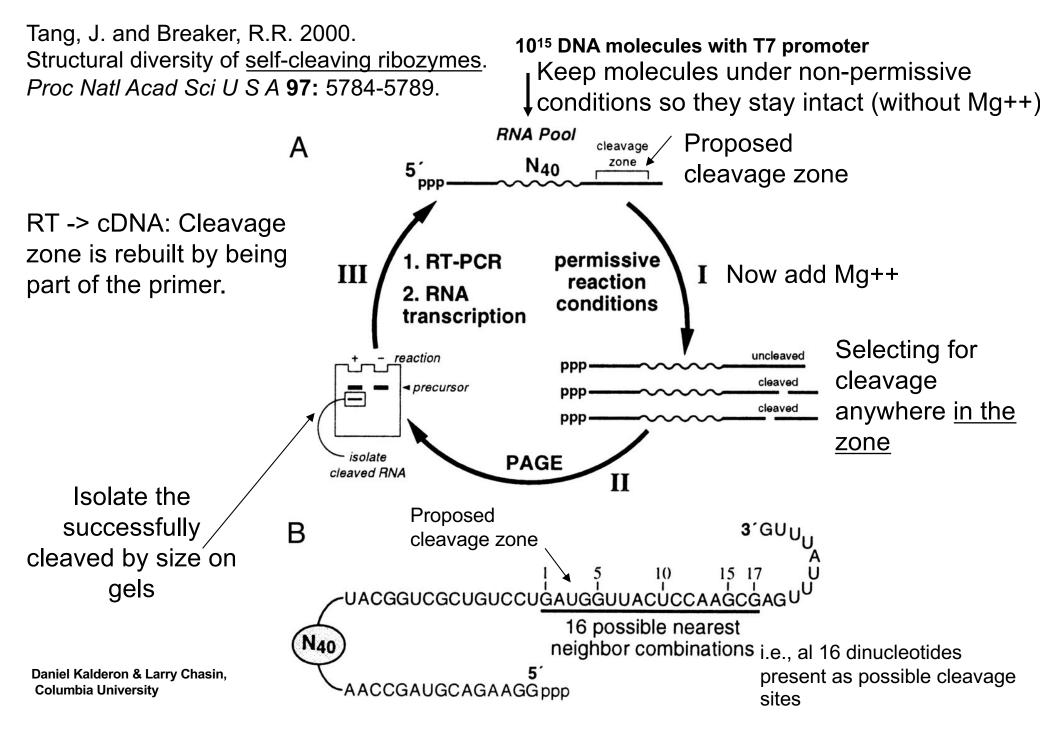
Hammerhead ribozyme (RNase) can cleave in cis ("hammer head" is upside down)

Synthetic variation: cleaves in *trans*

You are in charge of what it will cleave (you fill in the N's)



You can use SELEX to isolate new artificial ribozymes



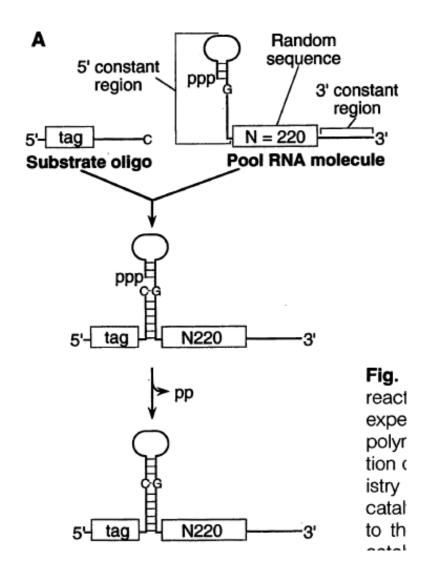
Example: Creating new ribozymes: "Isolation of New Ribozymes from a Large Pool of Random Sequences." Bartel & Szostak, SCIENCE, 261 (1993) 1411

Iterative in vitro selection; isoate a new class of catalytic RNAs from a large pool of random-sequence RNA molecules.

Ribozymes ligate two RNA molecules that are aligned on a template by catalyzing the attack of a 3'-hydroxyl on an adjacent 5'triphosphate-a reaction;

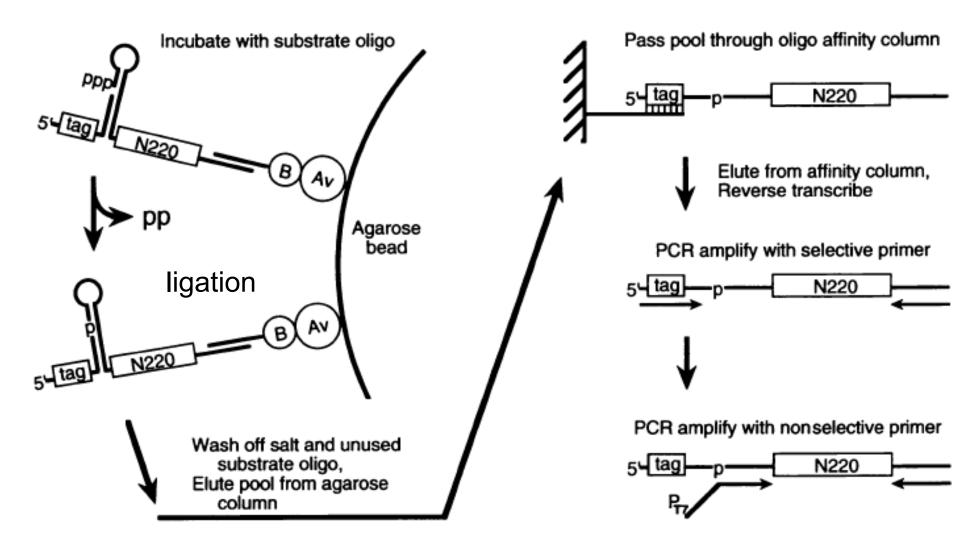
Rxn similar to that employed by the familiar protein enzymes that synthesize RNA.

In vitro evolution gave ribozymes with reaction rates 7 million times faster than the uncat-alyzed reaction rate.



Selection Scheme

selection



Ribozyme Catalyzed

Un-Catalyzed

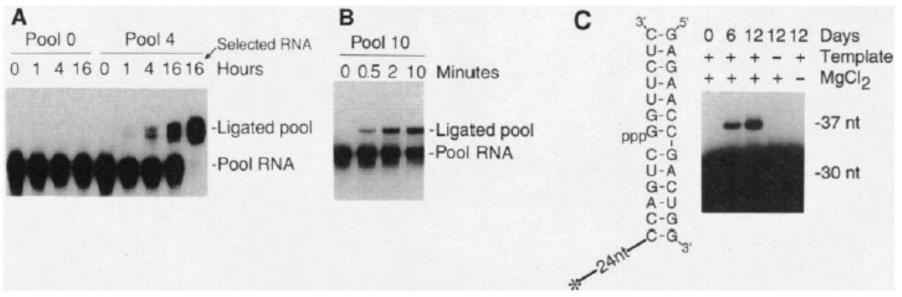


Fig. 4. (**A**) Time course of the catalyzed reaction. Phosphorus-32–labeled RNA of pools 0 and 4 was immobilized on agarose beads and incubated with substrate oligonucleotide. At the indicated time samples of the reaction were removed and stopped by addition of EDTA. RNA was eluted from the beads and separated by electrophoresis on a 4 percent acrylamide–7 M urea gel. A portion of the 16-hour pool 4 reaction was enriched for ligated molecules on an oligonucleotide affinity column ("selected RNA"). Immobilization, ligation, elution, and affinity selection conditions were as described for the first rounds of selection (Fig. 3). PhosphorImager (Molecular Dynamics) or Betagen (IntelliGenetics) scans were used for ³²P quantitation throughout. (**B**) Time course of pool 10 RNA ligation. Uniformly labeled pool RNA (0.4 μ M) was annealed (*33*), and then incubated at 25°C with substrate oligonucleotide (2.5 μ M) in the ligation buffer of the initial rounds (Fig. 3). Samples of the reaction were stopped with EDTA, and the RNA was separated on a 6 percent acrylamide–7 M urea gel. (**C**) Time course of the uncatalyzed reaction and template and magnesium dependence of this reaction. The 30-nt molecule that had been used as a substrate

Example of Ribozyme Evolution:

RNA Polymerase

Evolution of RNA Polymerase:

From ligase to RNA polymerase: "RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension"

Science 292, 1319 (2001) (A) A limited polymerase ĊĊŬĊ=ĊĠ_ſŪĠĠ(based on the ligase GGAG GCAACCGⁿⁿ CCUC=CGGUGGC core (black line). Template-ribozyme pairing required С (B) Randomized pool ^U UG u ["]G c G G (*C*) Round 10, polymerase; no pairing required; (D) Further improved AGG CCC polymerase ĠÁŪĠ ŪŪĊŪĊÁÁ GGAG GCAACCG GGAG GCAACCGcg

Evolution of RNA Polymerase:

Crystal Structure of the Catalytic Core of an RNApolymerase Ribozyme (SCIENCE, 326, (2009), 1271)

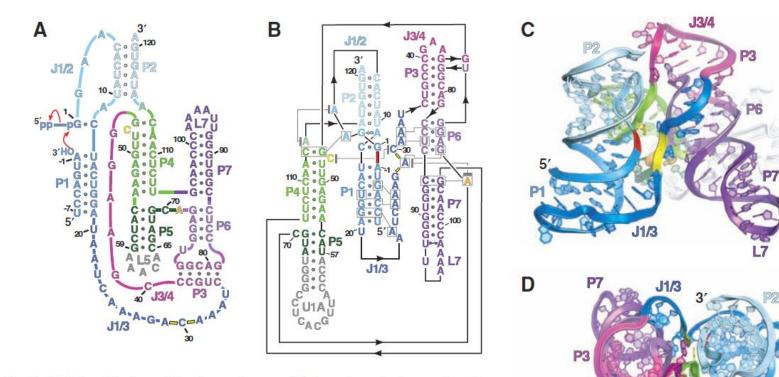
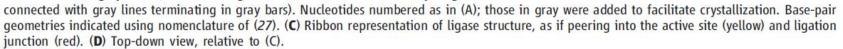


Fig. 1. Global architecture of the ligase ribozyme. (**A**) Secondary structure and reaction scheme of a ligase variant with decreased Mg^{2+} dependence (*10*). It is depicted undergoing ligation, by using the classical secondary-structure representation (*15*). Red arrows indicate attack by the substrate 3'-hydroxyl on the ribozyme α -phosphate with concomitant loss of pyrophosphate. (**B**) Revised secondary structure of the crystallization construct, reflecting the coaxial stacking and relative domain orientation. Indicated is the ligation junction (thick red dash), backbone phosphates at the active site (yellow dashes), base triples (boxed residues connected with gray lines), and stacking interactions (residues vertically aligned or



U1

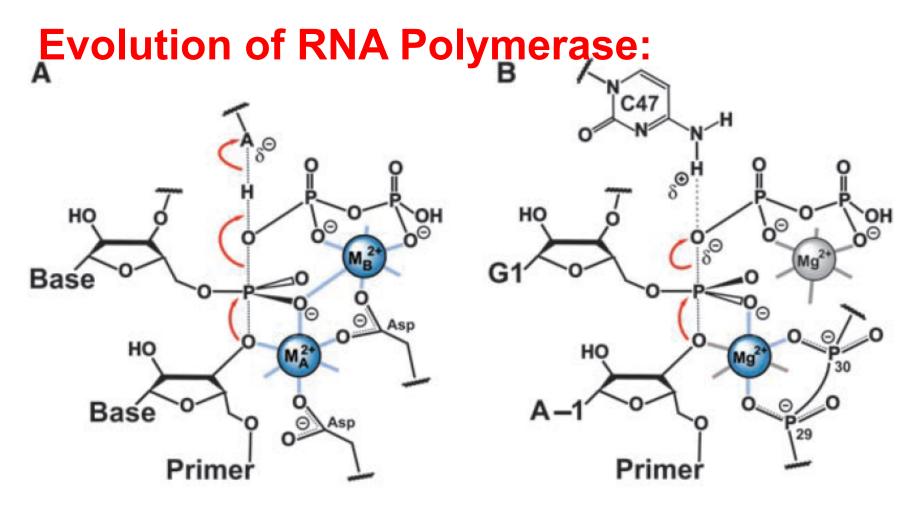
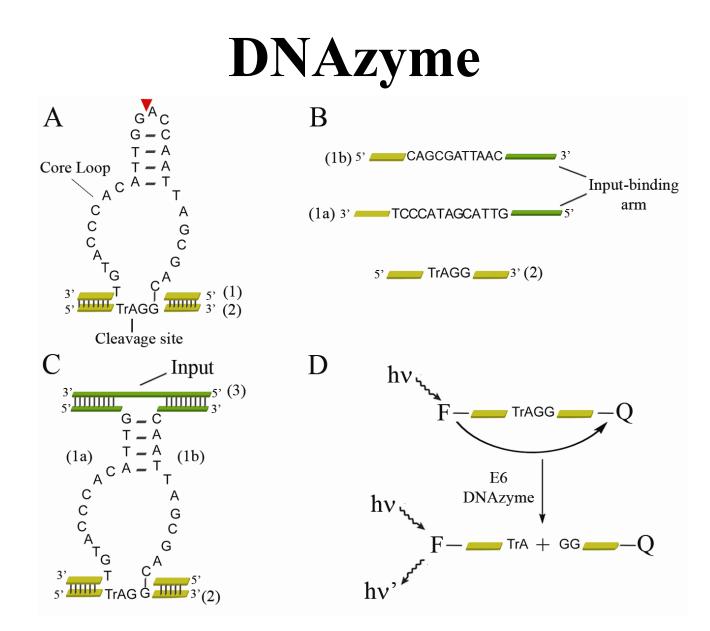


Fig. 4. Transition-state stabilization by polymerases built from either protein or RNA. (**A**) Catalysis by proteinaceous polymerases (23, 24). Indicated are bonds formed or broken during the transition state (red arrows), coordination of catalytic metal ions, M_A and M_B (blue solid lines), and an active-site acid (A•••H). (**B**) Model for catalysis by the ligase ribozyme. Notation as in (A), with the addition of a hydrogen bond between C47 N4 and the leaving group (dashed gray line). Some magnesium ligands are not specified; for those that are, relative orientations are unknown. A proposed contact to the reactive phosphate pro- R_P oxygen (28) and two speculative contacts implied by NAIM are in blue. Metal ion and coordinations not supported (or refuted) by structural or biochemical evidence are in gray.

SELECT for DNAzymes (DNA enzymes)

DNAzymes: are DNA oligonucleotides that can catalyze specific chemical reactions, such as restriction cuts.

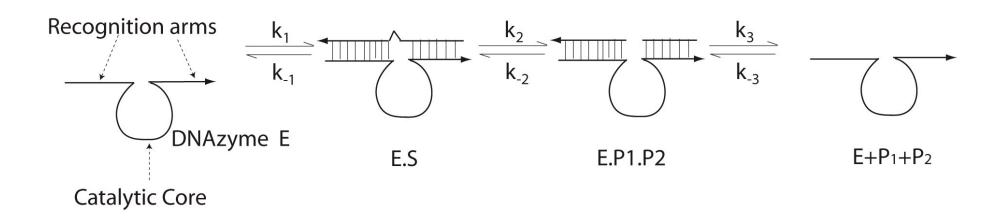
- DNAzymes can be discovered by In vitro selection or In vitro evolution
- DNAzymes are also named: Deoxyribozymes, DNA enzymes or catalytic DNA.



(from Wilner)

DNAzyme kinetics

RNA Substrate S



2nd step is rate determining
Requires metal ion as cofactor
k₂ >> k₋₂, k₁ >> k₋₁, k₃ >> k₂

[Santoro]

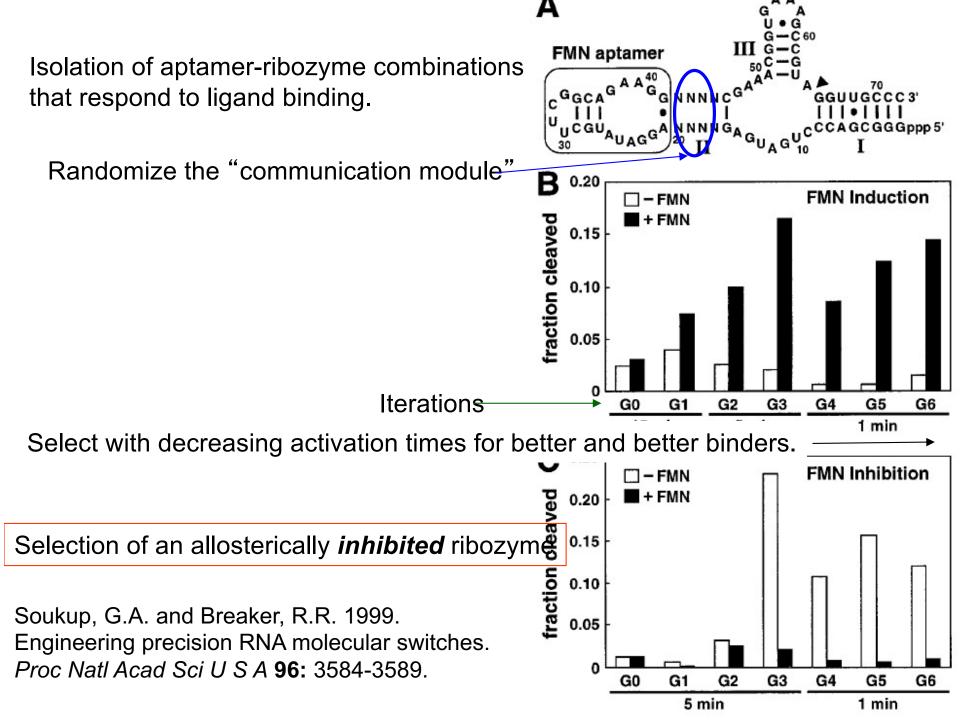
SELECT for DNAzymes (DNA enzymes)

- Start with 10¹⁵ DNA molecules again
- Select for enzyme activity:
 - E.g., cleaves itself off a solid support in the presence of Mg++
- Many different activities have been selected.
 Most have to do with nucleic acid transformations; RNase, ligase, kinase, etc.
 But not all (C-C bond formation possible).
- Generally much slower than protein enzymes.
- Most work has been on RNases (usually associated with the word "ribozymes")

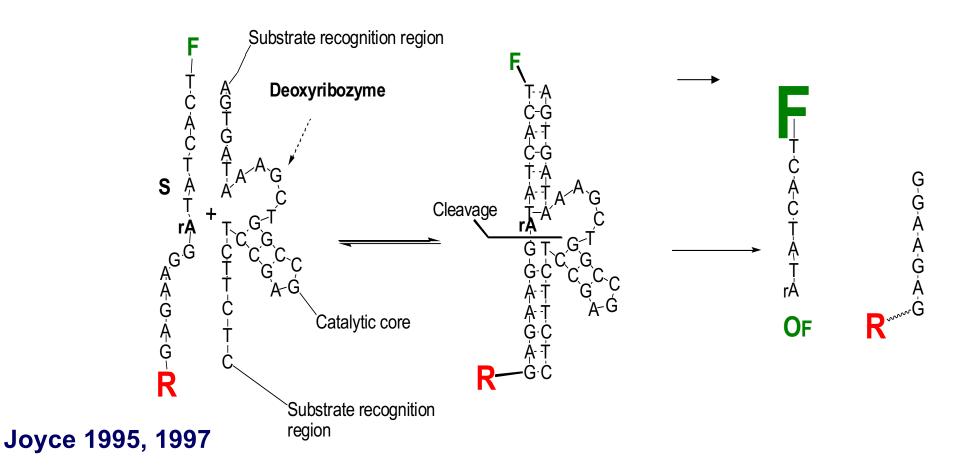
Allosteric ribozymes

- Allosteric: the alteration of the activity of a protein through the binding of an effector molecule at a specific site.
- Ribozyme: a ribonucleic acid (RNA) enzyme that catalyzes a chemical reaction
- Combine an aptamer and a ribozyme → Allosteric ribozyme
- Catalytic activity can be controlled by ligand binding !
- Positive or negative.
- Modular
- Molecular switches, biosensors

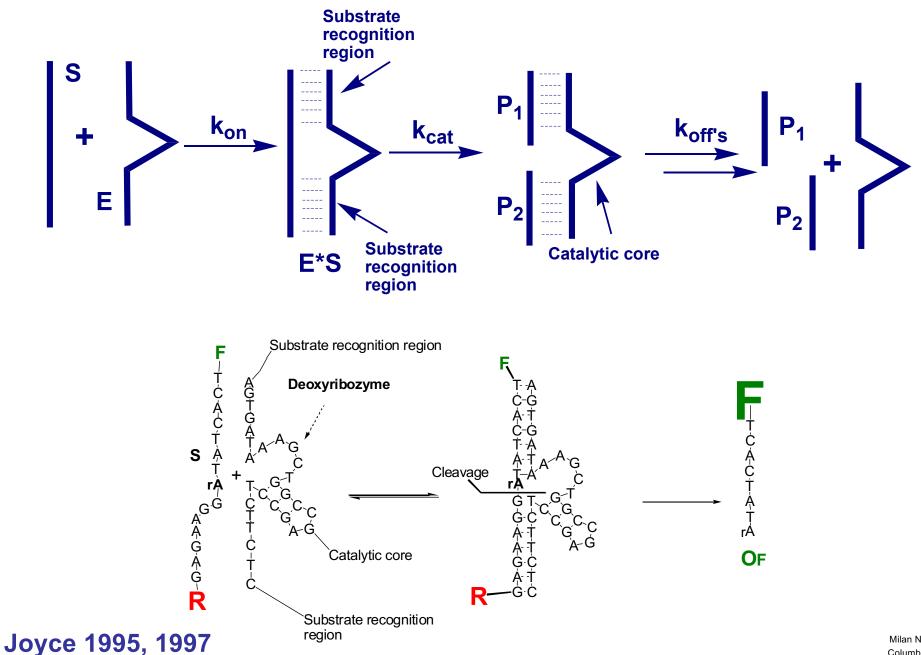
Selection of an allosterically activated ribozyme



Ribozymes used to cut double stranded DNA at Recognition Sites:



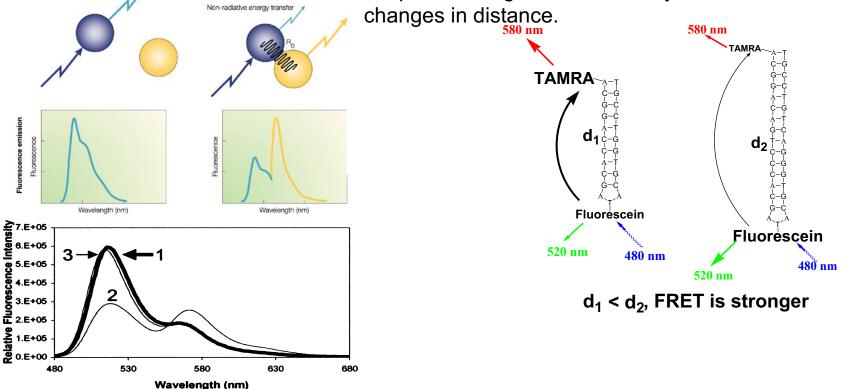
Ribozymes used to cut double stranded DNA at Recognition Sites:



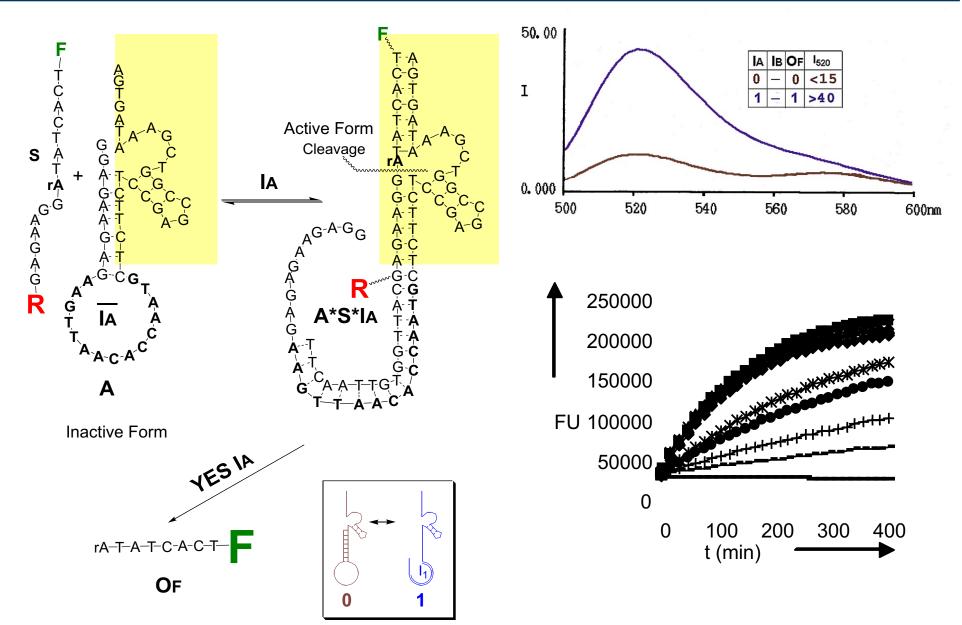
Milan N. Stojanovic Columbia University

Förster or fluorescence resonance energy transfer (FRET)

- FRET is mechanism describing energy transfer between two light-sensitive molecules (chromophores).
- A donor chromophore, initially in its electronic excited state, may transfer energy to an acceptor chromophore through nonradiative dipole–dipole coupling.
- The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor, making FRET extremely sensitive to small



Catalytic Molecular Beacons as Sensor Gates Detector



Joyce 1995, Breaker 1999, Tyagi, Kramer 1996

Stojanovic et al., *ChemBioChem* **2001** Stojanovic et al., *J. Am. Chem. Soc.* **2002**

Using an allosteric ribozyme to create a chemical sensor

Frauendorf, C. and Jaschke, A. 2001. Detection of small organic analytes by fluorescing molecular switches. Bioorg Med Chem **9:** 2521-2524.

> Start with a theophyllinedependent ribozyme:

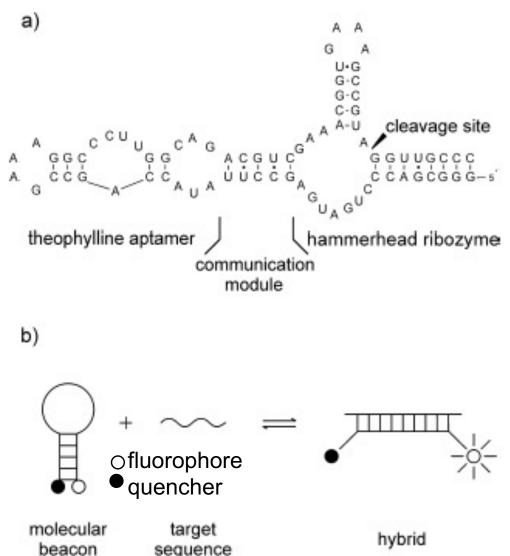
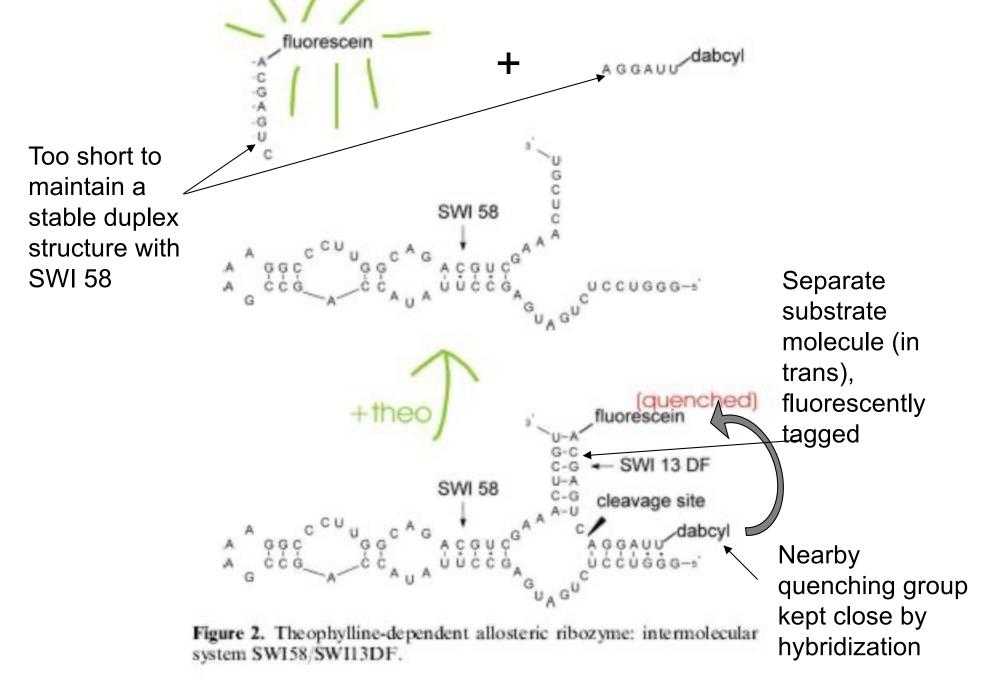


Figure 1. (a) Theophylline-dependent allosteric ribozyme⁶ and (b) operating principle of molecular beacons.

Analogy: A molecular "beacon" that respond to nucleic acid hybridization

Using an allosteric ribozyme to create a chemical sensor



н H₃C₅ CH3 0.6 theophylline a) 5X over background 0.5 сн_з H₃C₅ 0.4 (rFU/s 0.3 theophylline 0 caffeine rate CH3 0.2 caffeine 0.1 good specificity 0 2 6 0 c (analyte) (mM) Not so sensitive Daniel Kalderon & Larry Chasin, Columbia University (0.3 mM)

Using an allosteric ribozyme to create a chemical sensor

Some DNAzyme activities

Emilsson, G. M. and R. R. Breaker (2002). Deoxyribozymes: new activities and new applications. <u>Cell Mol Life Sci</u> **59**(4): 596-607.

Chemical reaction: Enzyme name	Cofactor/covalent modifications	Maximum k (min ⁻¹)	Rate enhancement	Ref.
RNA transesterification:	Compare protein enzymes, over spontaneous reaction			
G5 dominant clone	Pb ²⁺ Typically 6000 on this scale	1	10 ⁵	29
E6	Mg ²⁺ (100/sec)	0.04	10 ⁵	38
Mg5/8-17/17E ^a	Zn ²⁺	1.35	10 ⁸ calc	73-76
10-23	Mg ²⁺	5	10 ⁵ calc.	31, 32
Na8	_	0.0067	10 ⁸	27
HD2	L-histidine	0.2	106	28
1611	imidazole-modified DNA, Zn2+	4.3	10 ^s calc.	42
9 ₂₅ -11	imidazole-, amine-modified DNA	0.044	106 calc.	44
DNA cleavage:				
Class II	Cu^{2+}	0.2	107	49-51
Rd10 dominant clone	phosphoramidate substrate			
	Mg ²⁺ , oligonucleotide	n/a	10 ³	54
N-glycosylase activity:				
10-28	Ca ²⁺	0.2	106	53
Phosphorylation:				
NTP-A2.1	Ca ²⁺	0.012	109	59
	Ca	0.012	10	
Adenylation:	C-2+ 14-2+	0.0048	1010	12
Class I	Cu ²⁺ , Mg ²⁺	0.0048	10.5	62
DNA ligation:			N-30122	
E47	imidazole substrate, Zn ²⁺	0.07	10 ⁵	65
L78 ^b	AMP-capped substrate, Mn ²⁺	0.0001	10 ⁵	64
C14	phosphorothioester linkage formation	0.023	10 ²	67
Porphyrin metalation:				
PS5.M		1.3	103	55, 77

Table 1. Reactions and rate constants of deoxyribozymes.