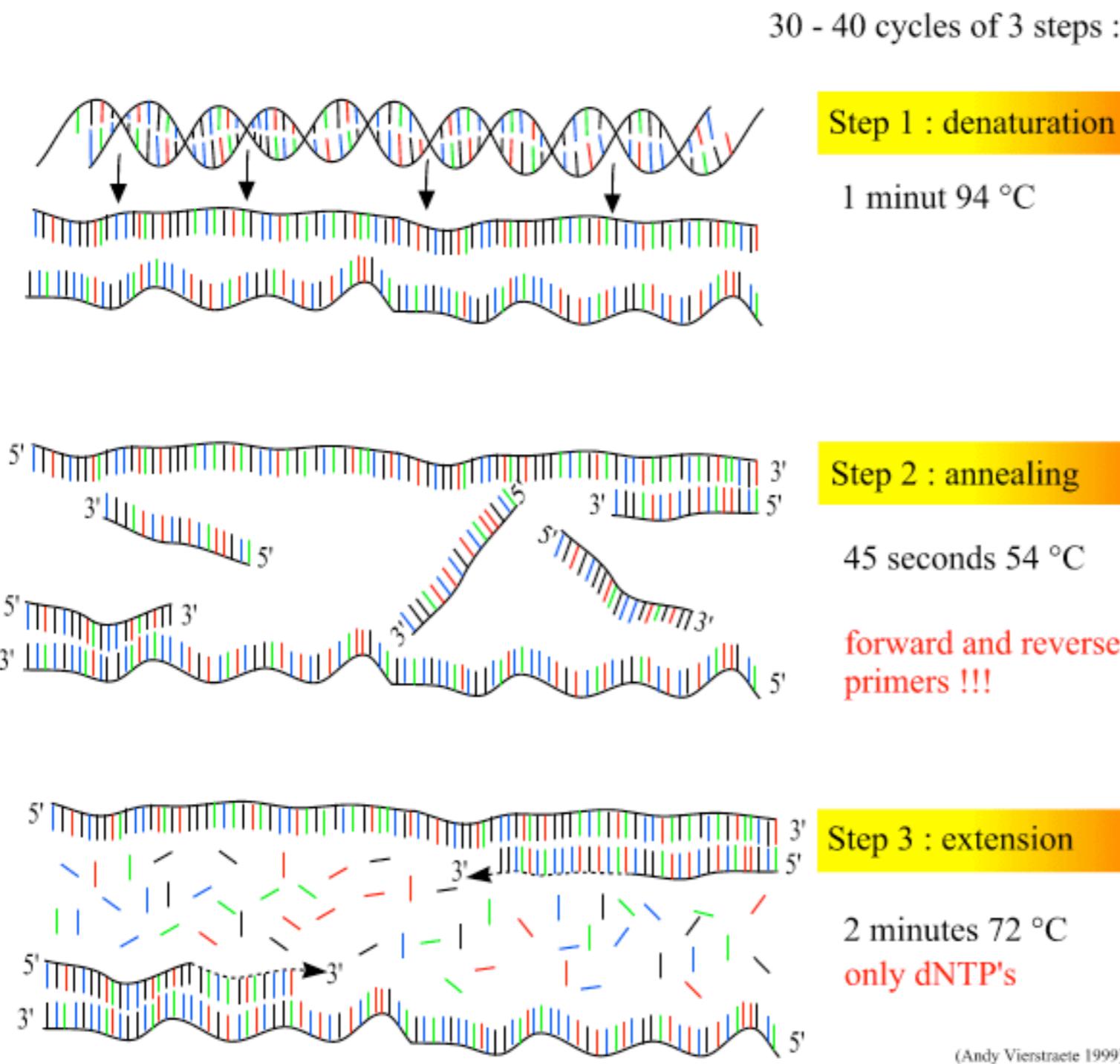


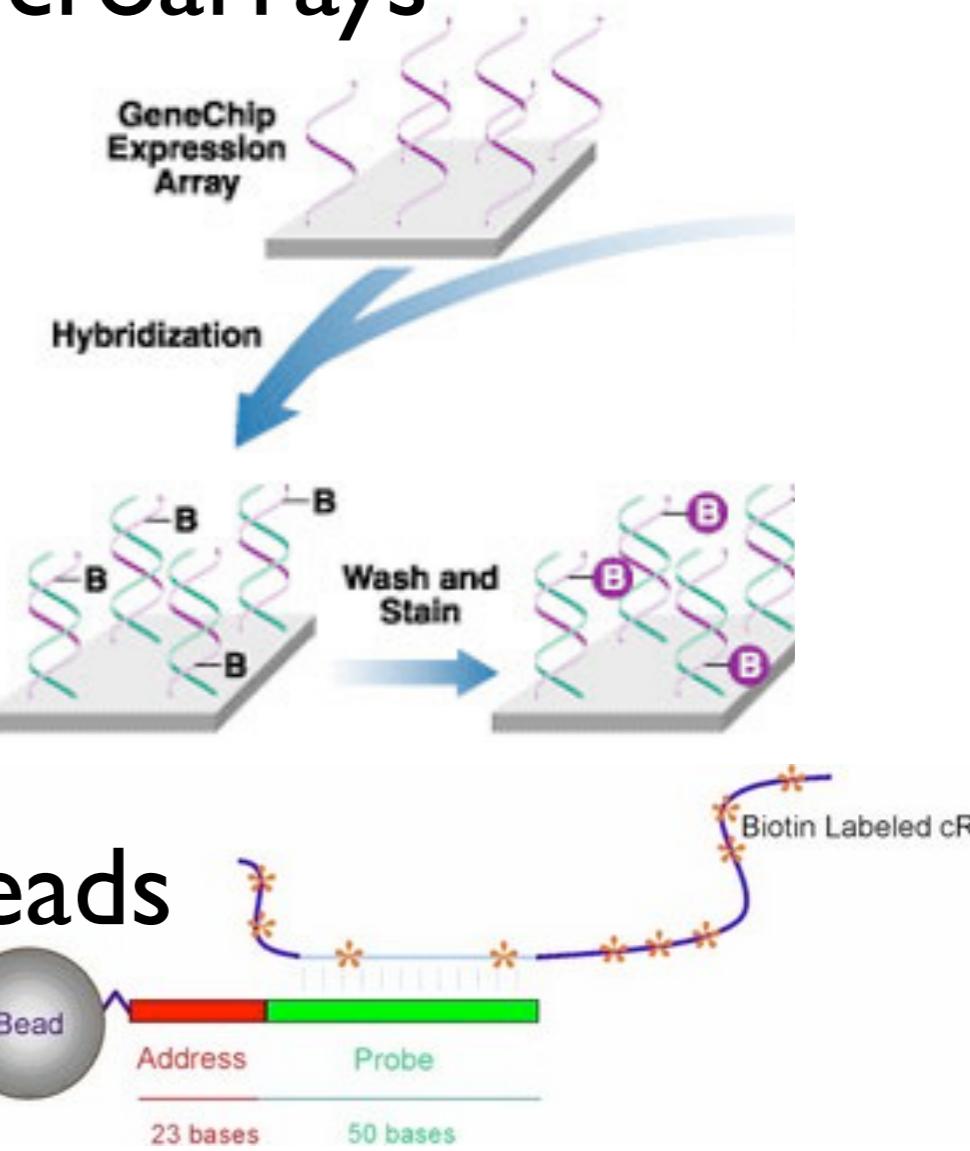
# RNA Engineering

BioE131

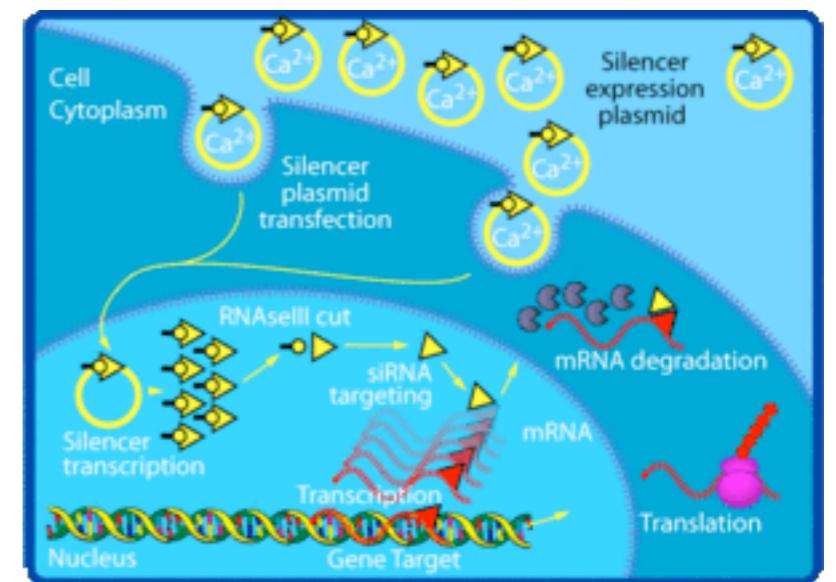
## PCR : Polymerase Chain Reaction



## Microarrays



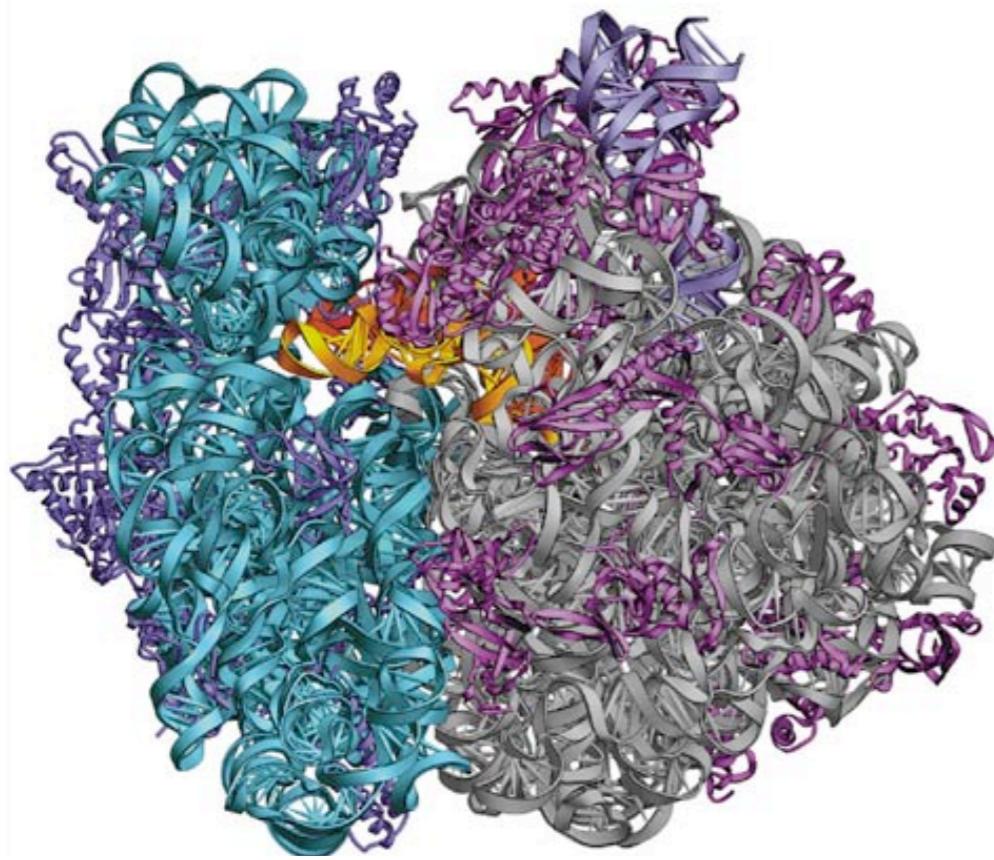
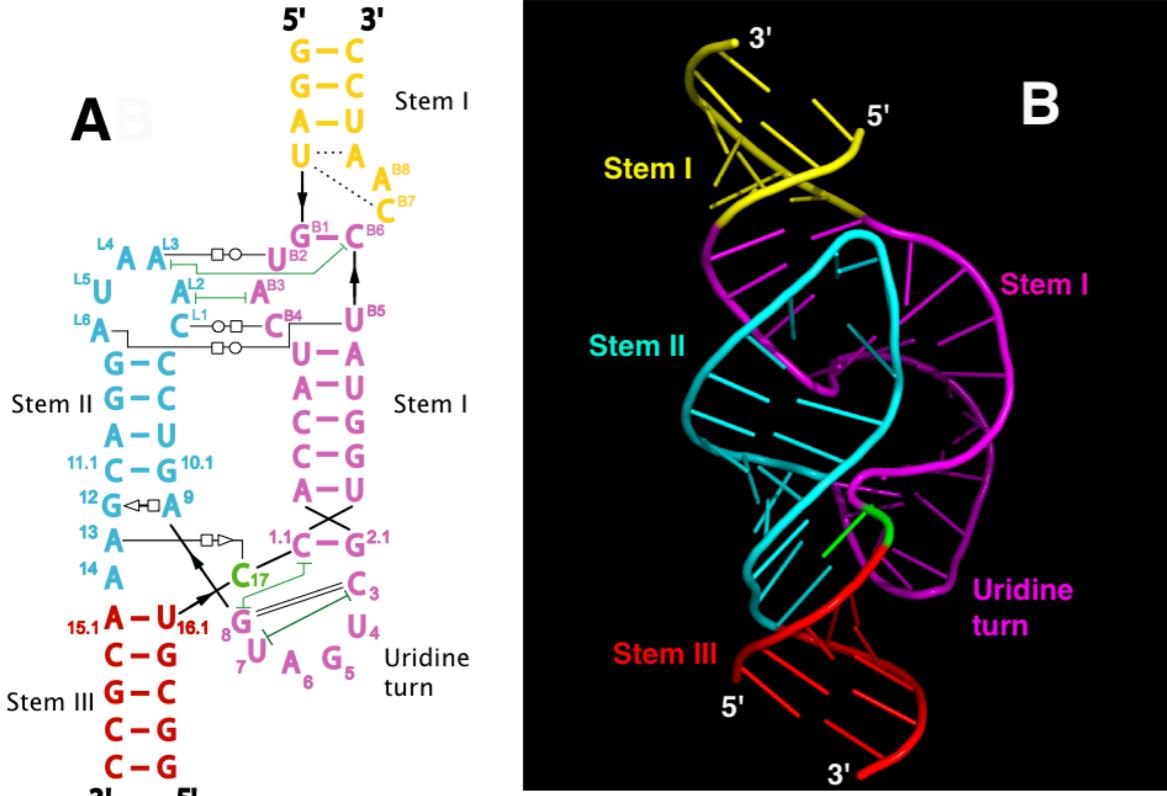
## siRNA



**Widely-used  
engineered RNAs**

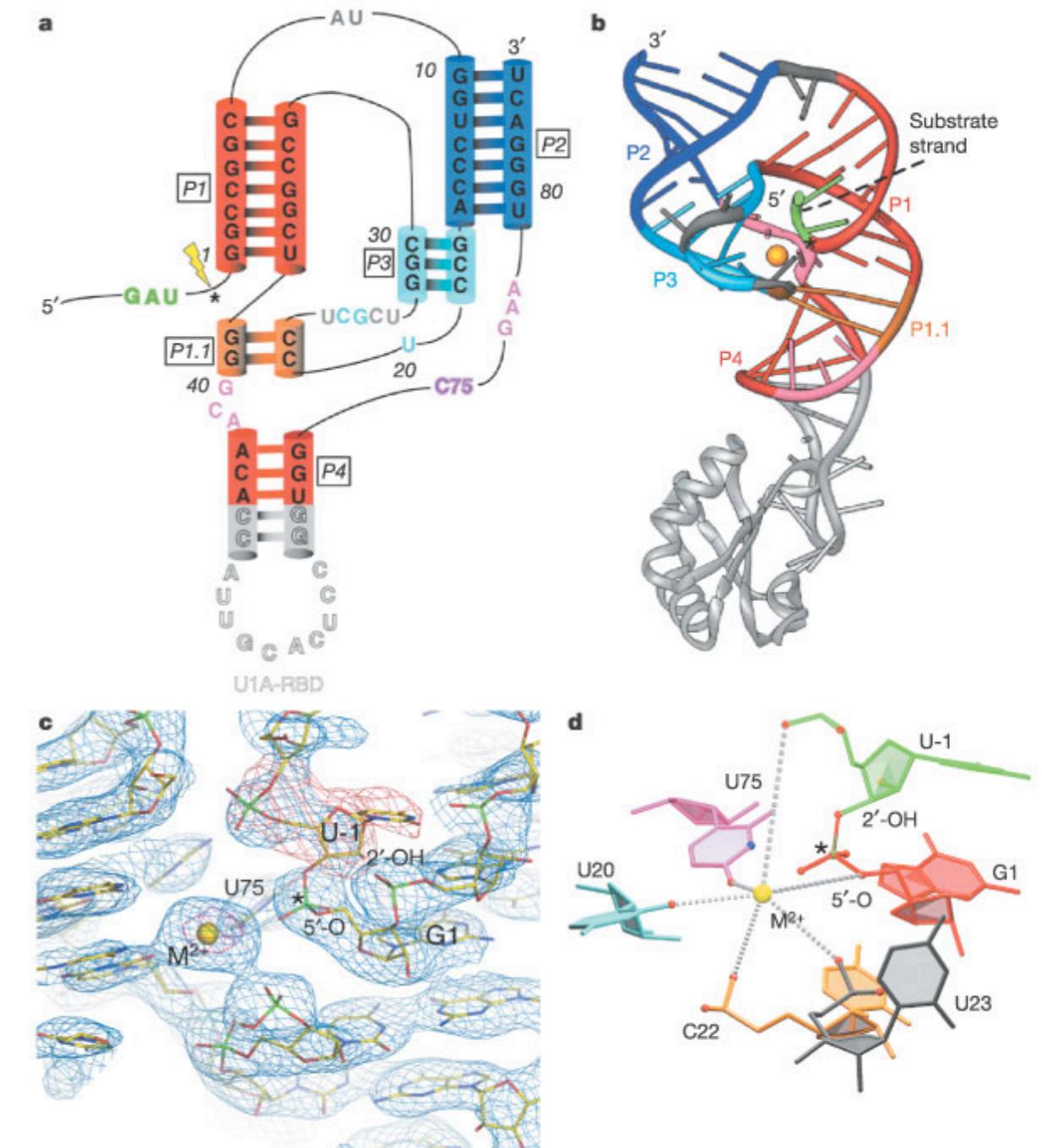
**PCR**

# Hammerhead ribozyme



Ribosomal RNA

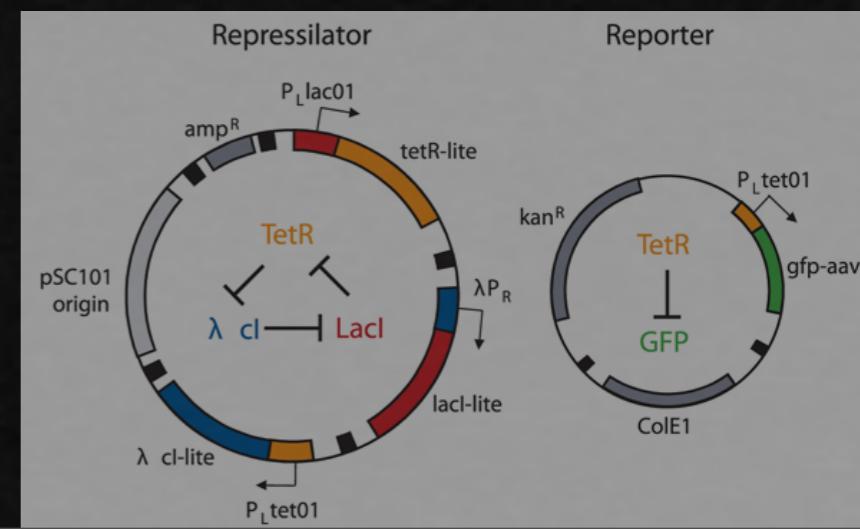
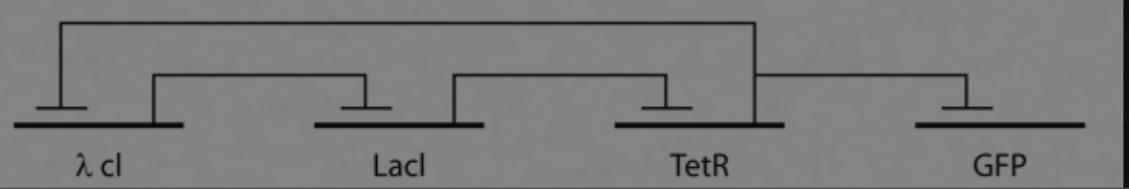
# Possibilities...



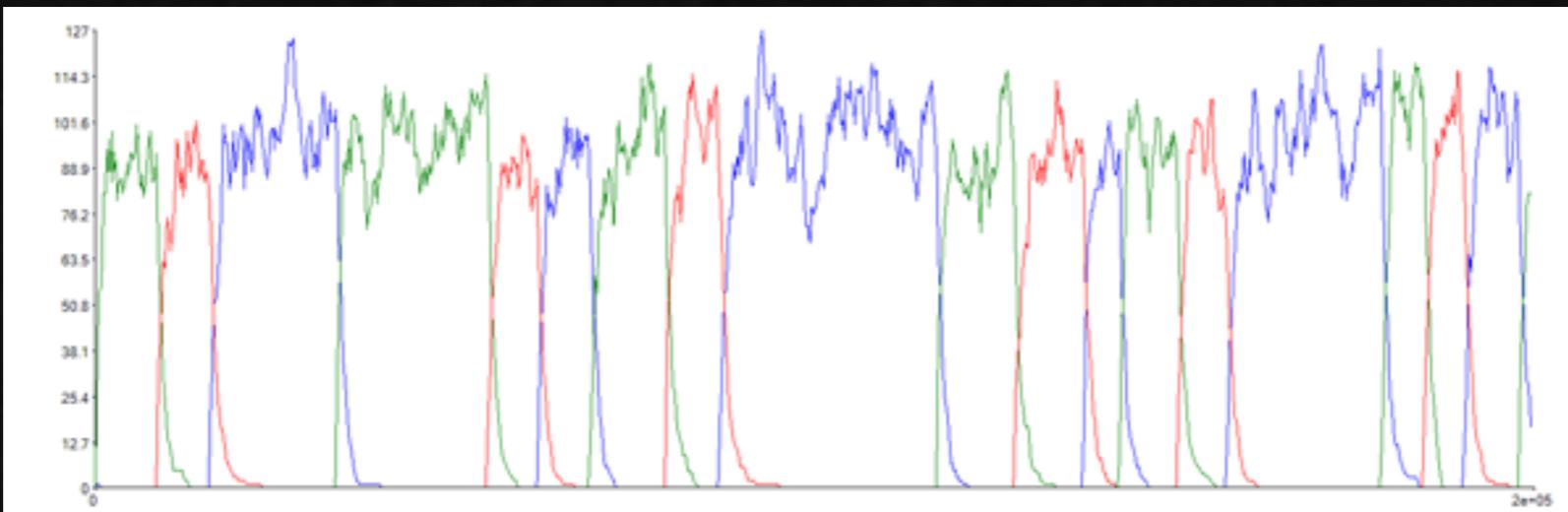
Hepatitis Delta Virus  
ribozyme

# Early synthetic gene circuits used DNA-binding protein transcription factors

The “Repressilator”  
Elowitz & Leibler, Nature, 2000

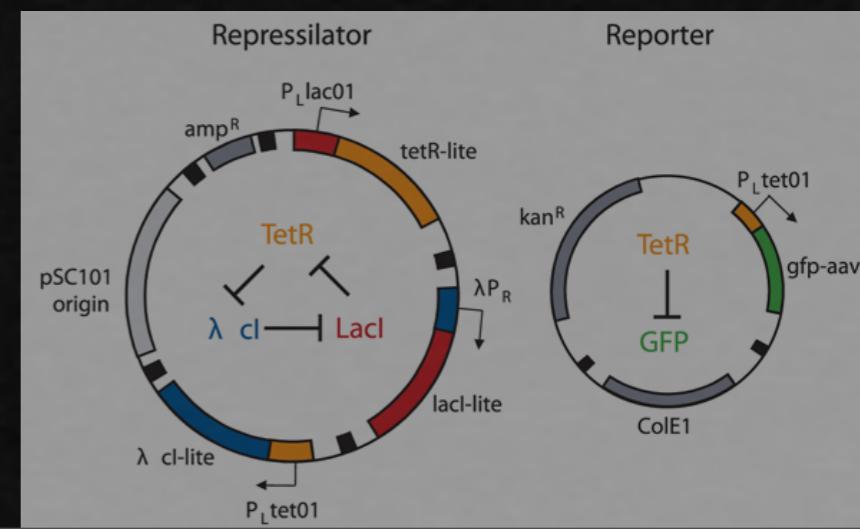
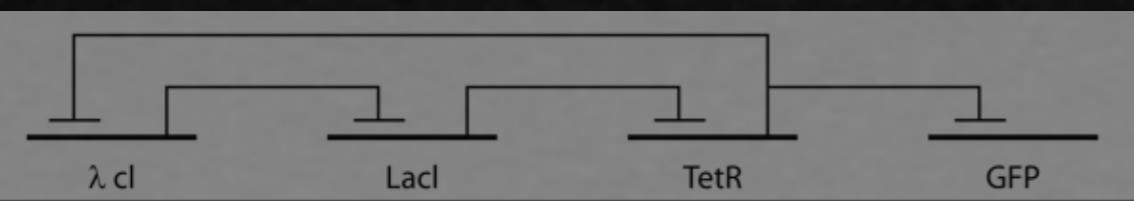


# Early synthetic gene circuits used DNA-binding protein transcription factors

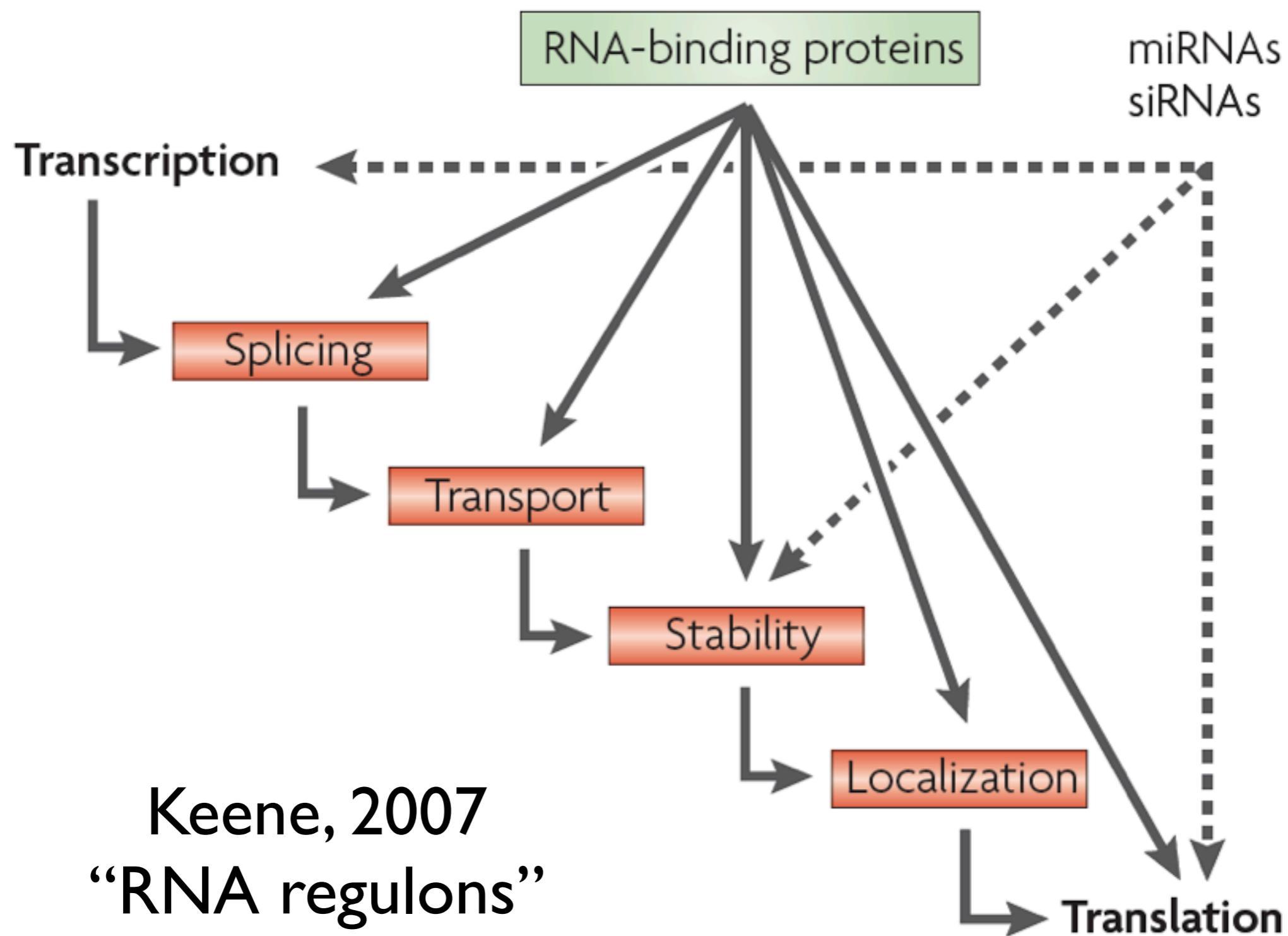


Stochastic simulations (above) suggested oscillations were favored by (1) tight promoters, to prevent “leaky” transcription; (2) carboxyl-tagging of TFs by cis-acting peptide signals triggering prompt degradation by proteases, so that protein & mRNA decay rates are matched. These features were incorporated into the design

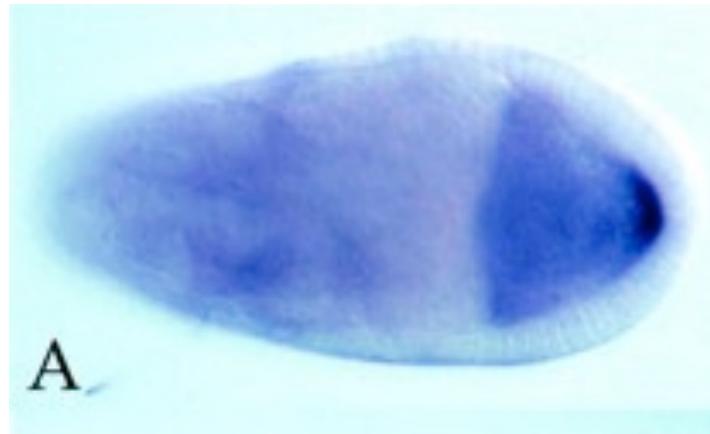
The “Repressilator”  
Elowitz & Leibler, Nature, 2000



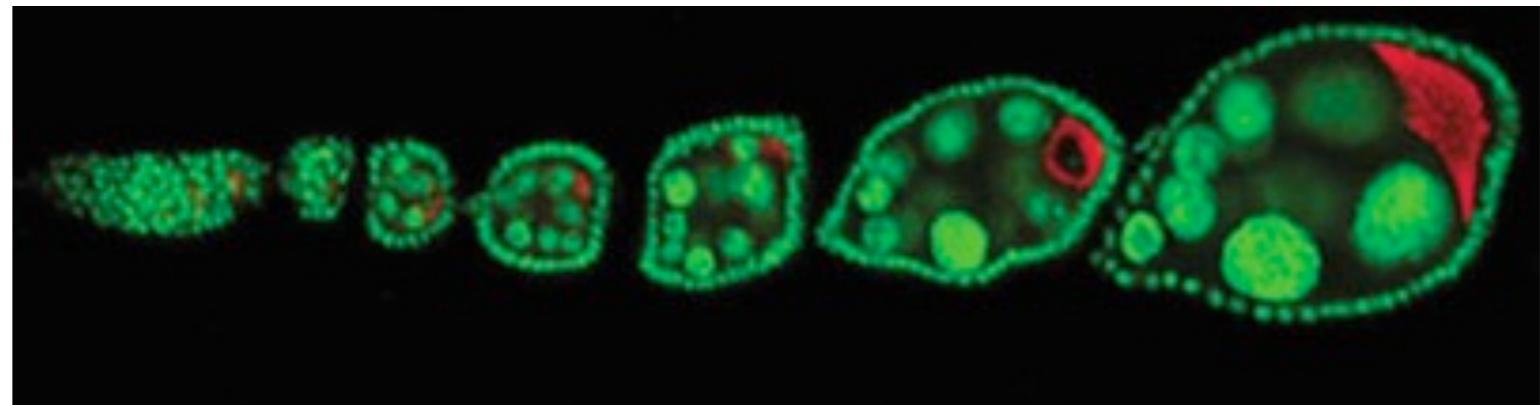
# Post-transcriptional regulation



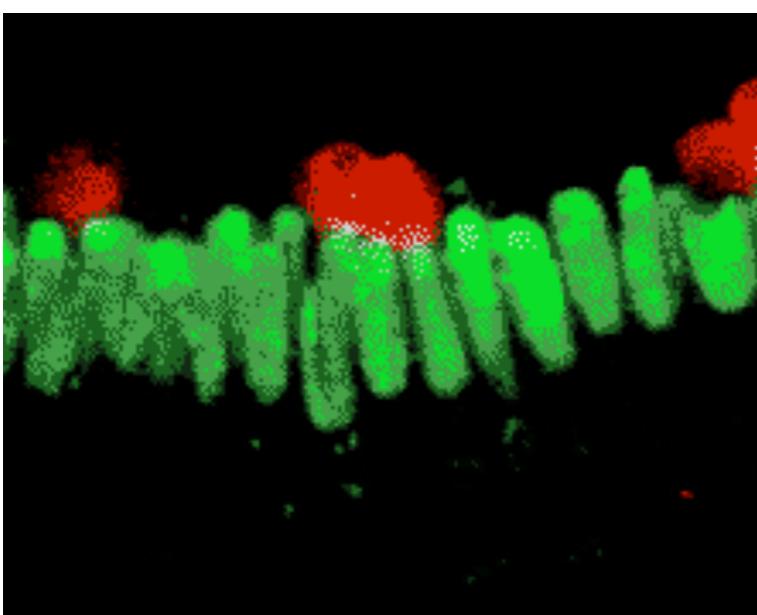
# RNA localization in *Drosophila*



*osk* (Micklem et al, 2000)



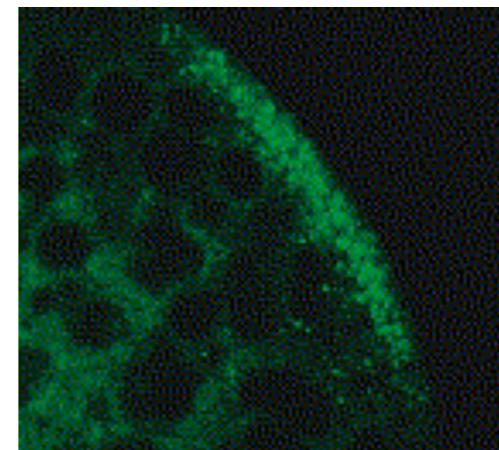
*orb* (Navarro et al, 2004)



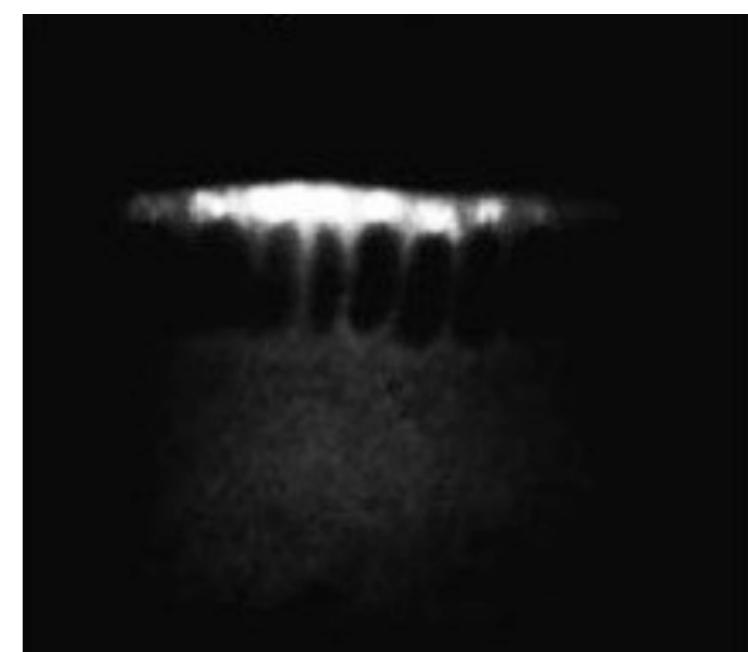
*ftz* (Lall et al, 1999)



*bicoid* (Irion & St Johnston, 2007)



*nanos* (Forrest & Gavis, 2003)



*K10* (Bullock & Ish-Horowicz, 2001)

# Zipcodes

a

AA  
U U  
U A C  
AU  
AU  
UA  
UA  
UA  
UA  
UA  
C  
AU  
UA  
GC  
UA  
A  
UA  
AU  
GU  
UA  
UA  
CG

stg-K10TLS

A fluorescence micrograph showing a single neuron from a C. elegans worm. The neuron exhibits a distinct, elongated pattern of bright, punctate fluorescence along its length, indicating the presence of the transgene product. The background is dark, making the bright spots stand out.

b

AA  
U U  
U U  
A C  
AU  
AU  
a A  
a A  
a A  
a A  
a A C  
AU  
UA  
GC  
UA A  
UA  
AU  
GU  
UA  
UA  
CG

Kstem5'

KIC

(Bullock & Ish-Horowicz, 2001)  
(Serano & Cohen, 1995)

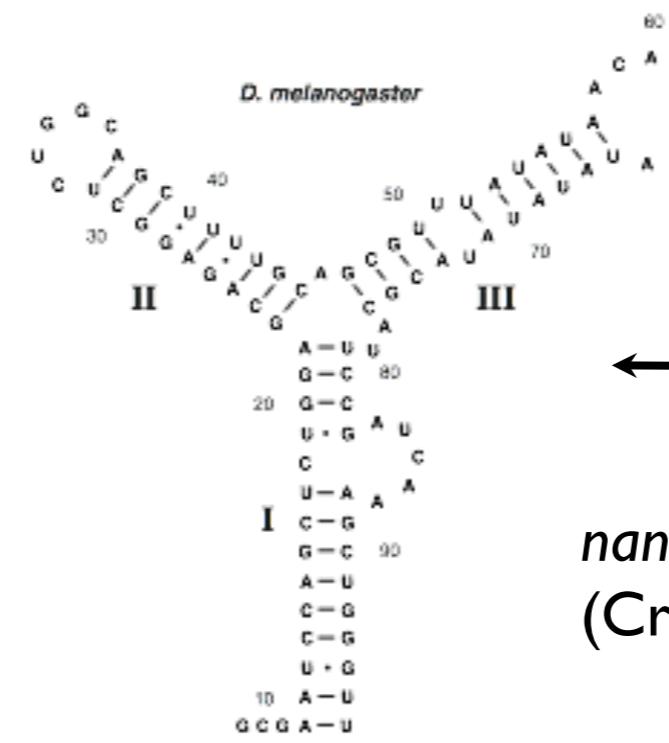
C

AA		
U	U	U
U	A	C
	AU	
		C
	AU	
	UA	
	GC	
	UA	A
	UA	
	AU	
	GU	
	UA	
	UA	
	CG	

Kstem5'3'

*Kstem*5'3'

A



nanos

(Crucks et al, 2000)

# Zipcodes

a

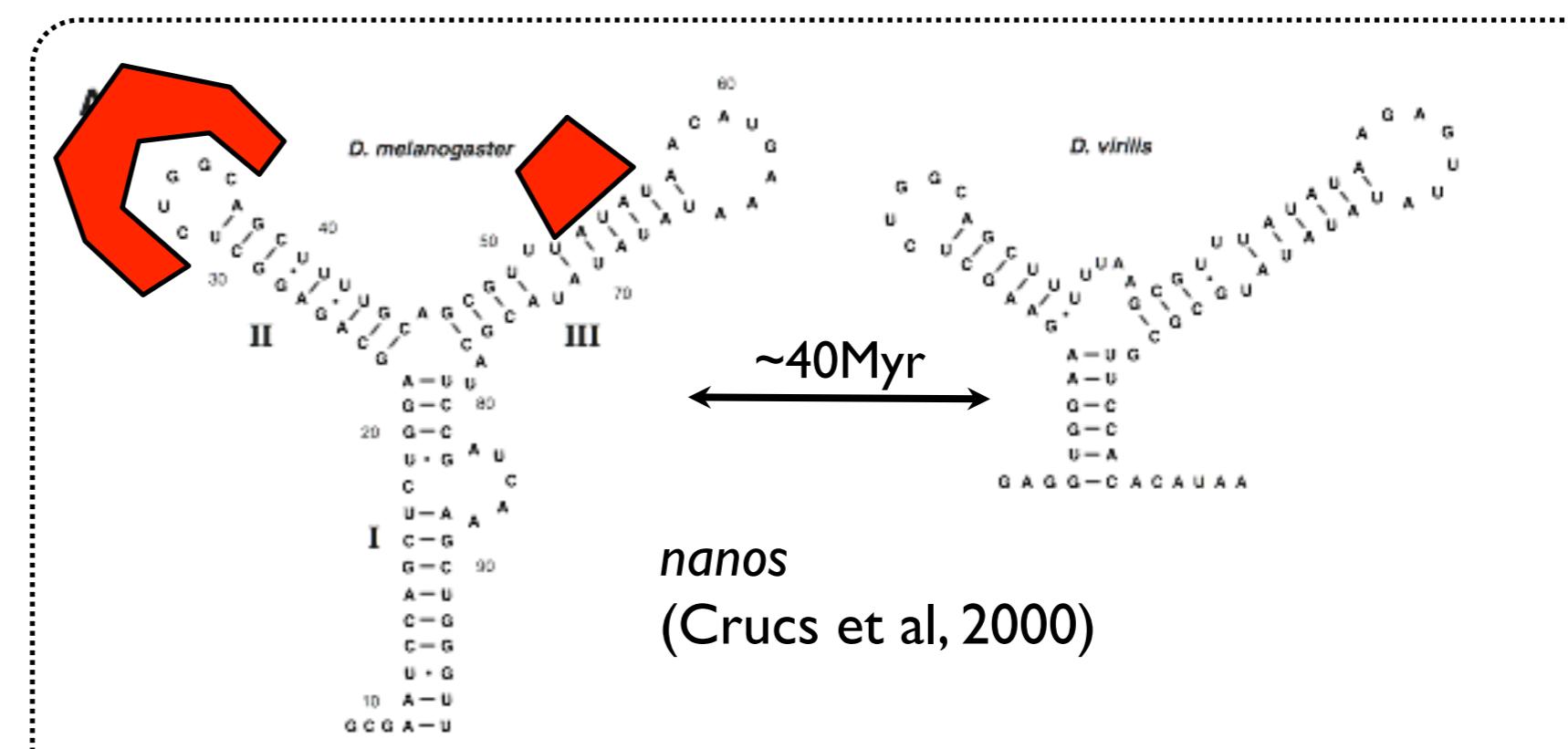
*stg-K10TLS*

b

A fluorescence micrograph showing a dense, irregular cluster of bright, granular spots against a dark background. The spots are concentrated in the center and taper off towards the edges, suggesting a cell cluster or a specific cellular localization.

C

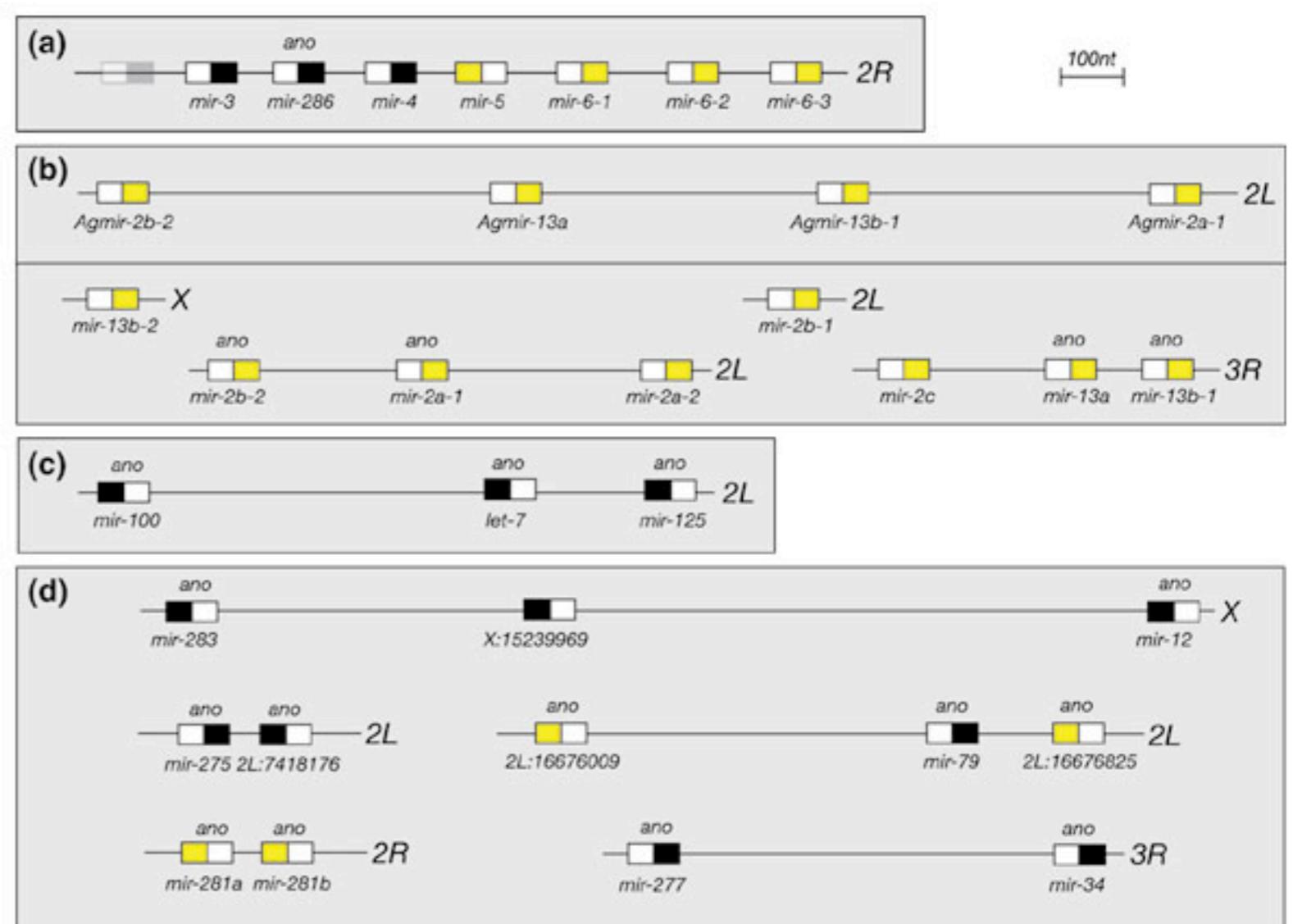
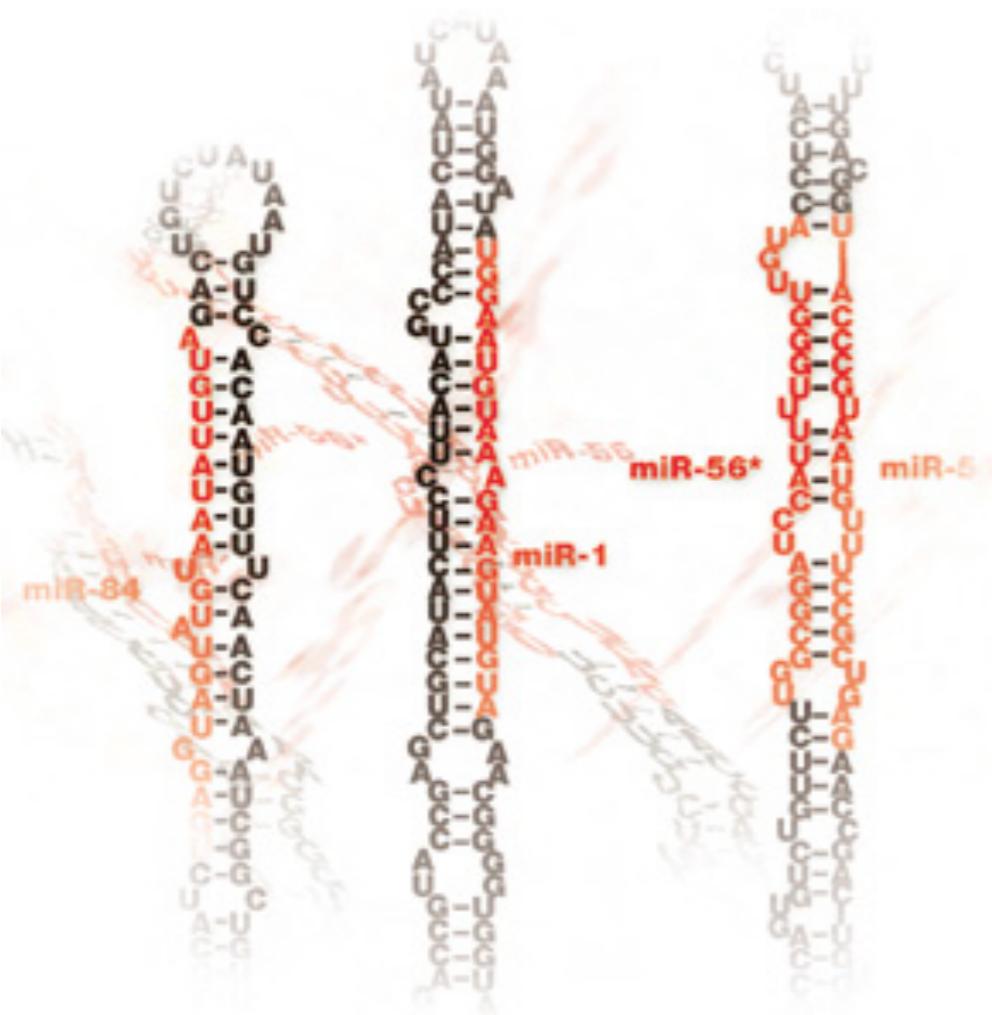
The figure displays a gel electrophoresis pattern with multiple RNA bands. The lanes are labeled on the left with sequence motifs: AA, UU, AU, AU, au, au, au, au, au, C, AU, UA, GC, UA, A, UA, AU, GU, UA, UA, and CC. The bands show varying intensities across the lanes, indicating differential RNA processing or expression levels for each motif.



# Regulatory RNA genes

## miRNAs...

# miRNAs...

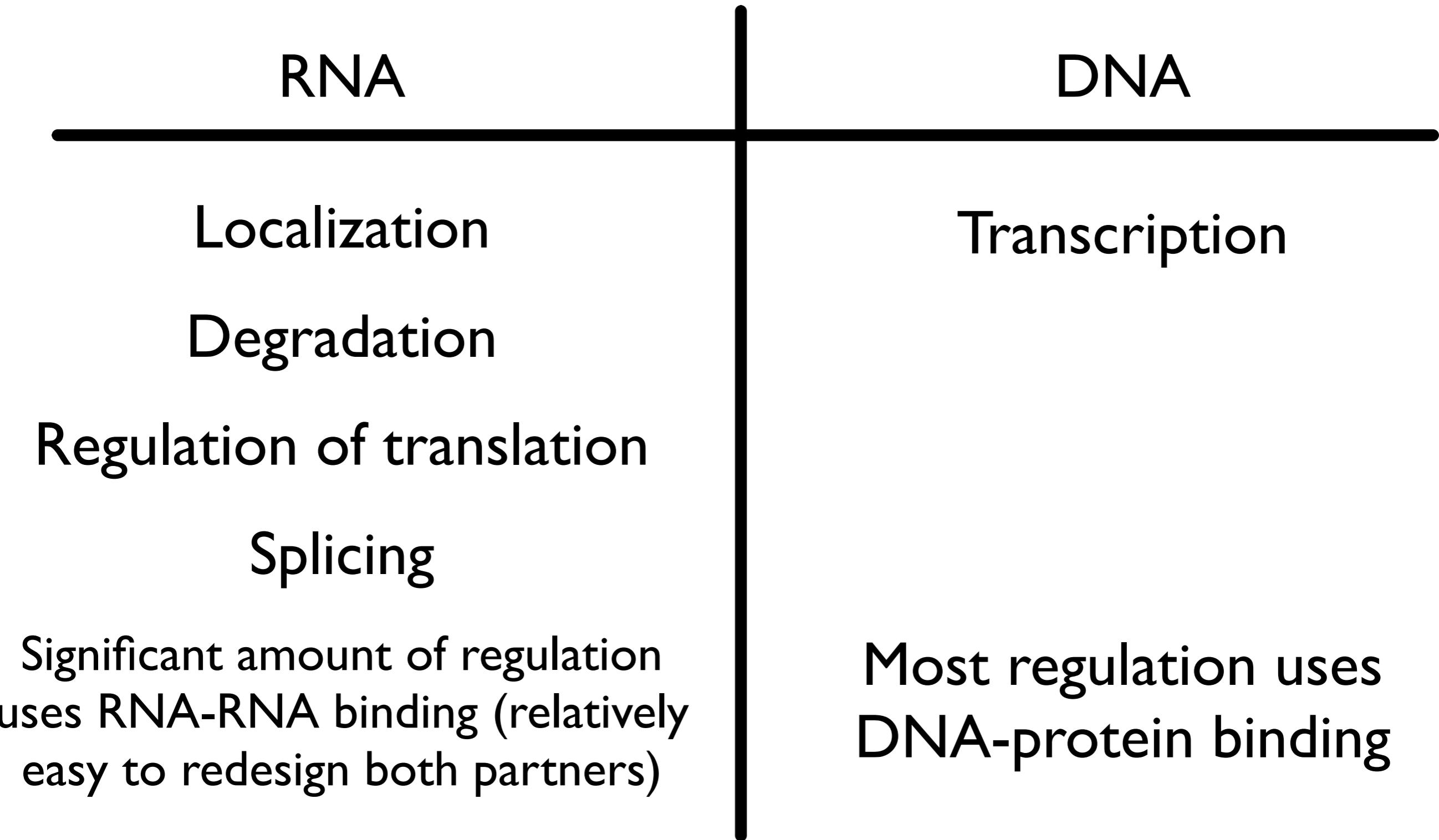


From [www.ionchannels.org](http://www.ionchannels.org)

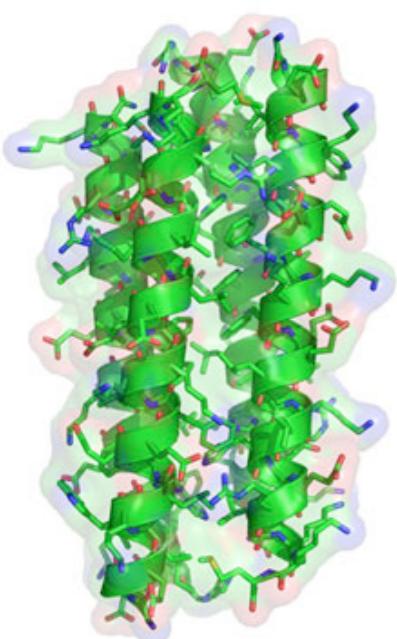
Lai et al, 2000

Also (in flies) *rox*, *hsr-omega*, *bithoraxoid*(?)...

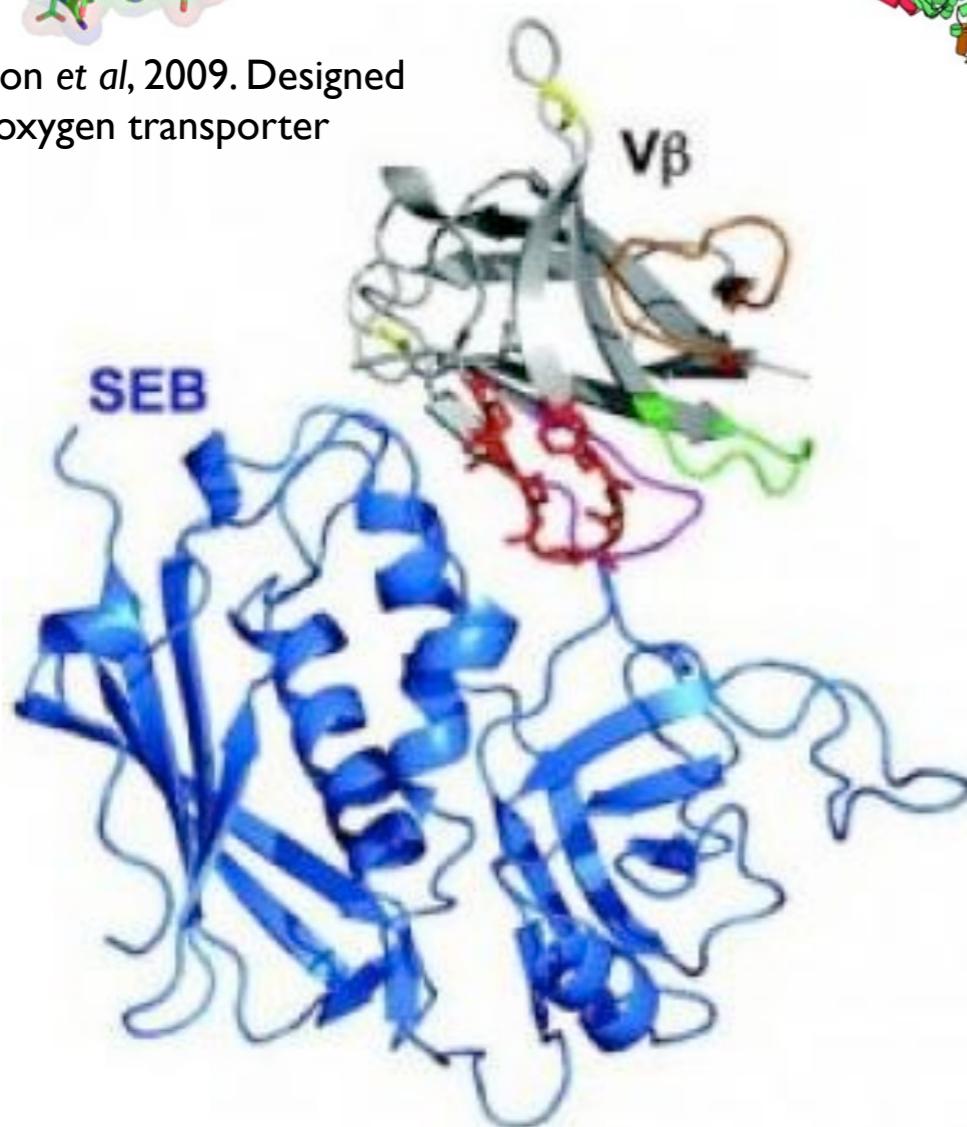
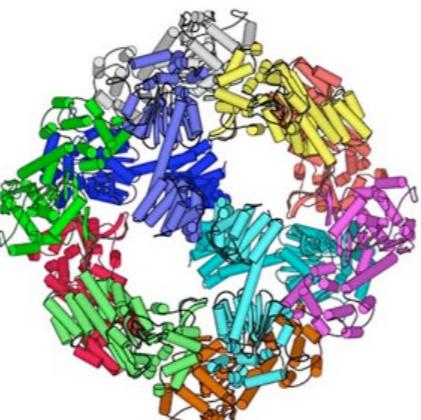
# Designed regulation of gene expression: RNA vs DNA



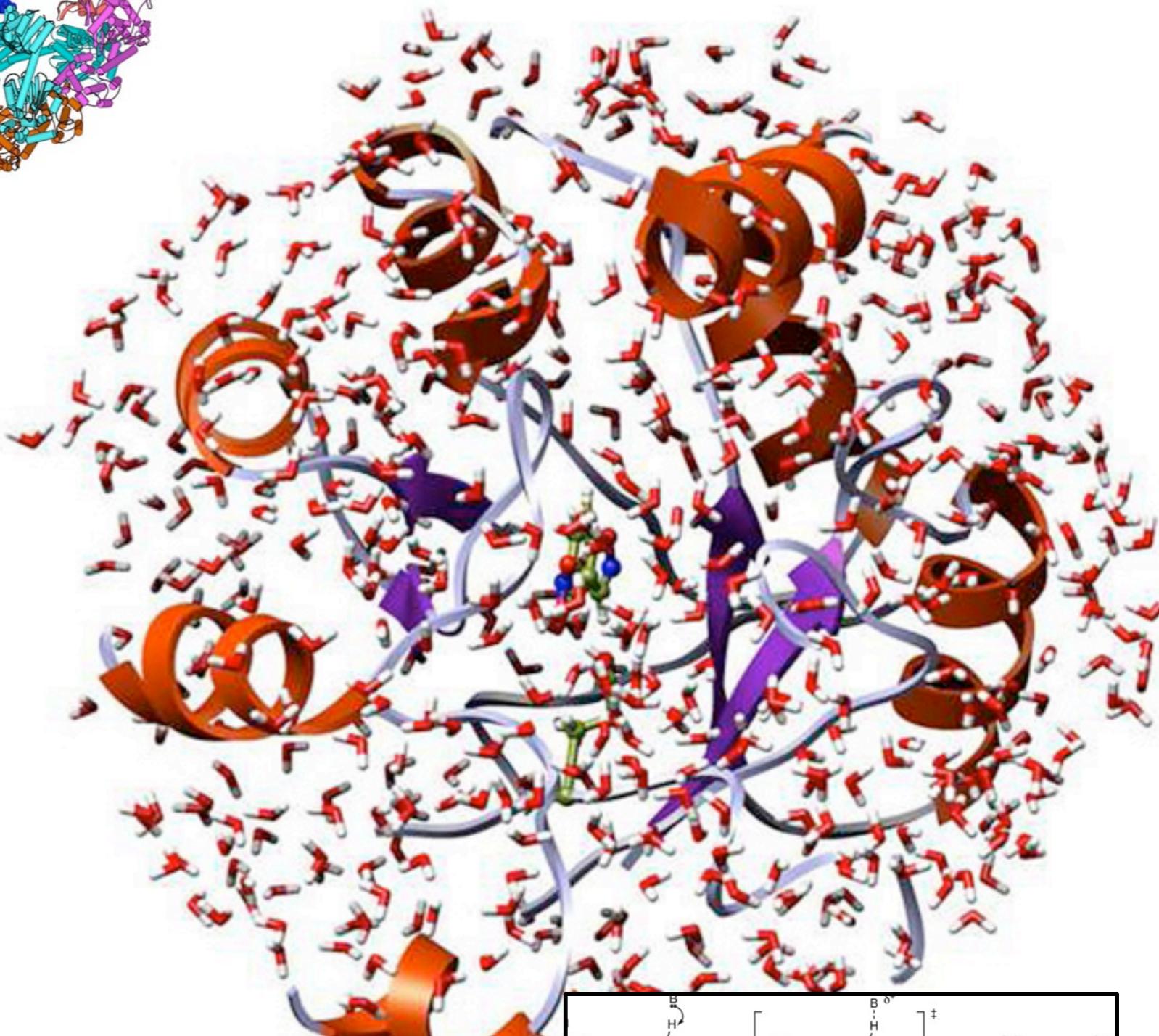
# Protein design



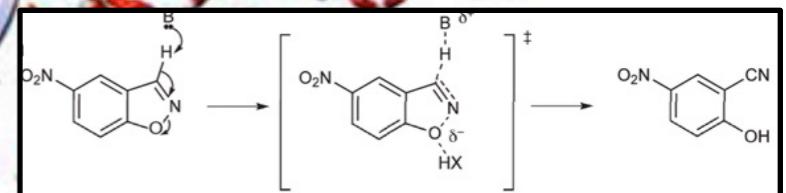
Dutton et al, 2009. Designed oxygen transporter



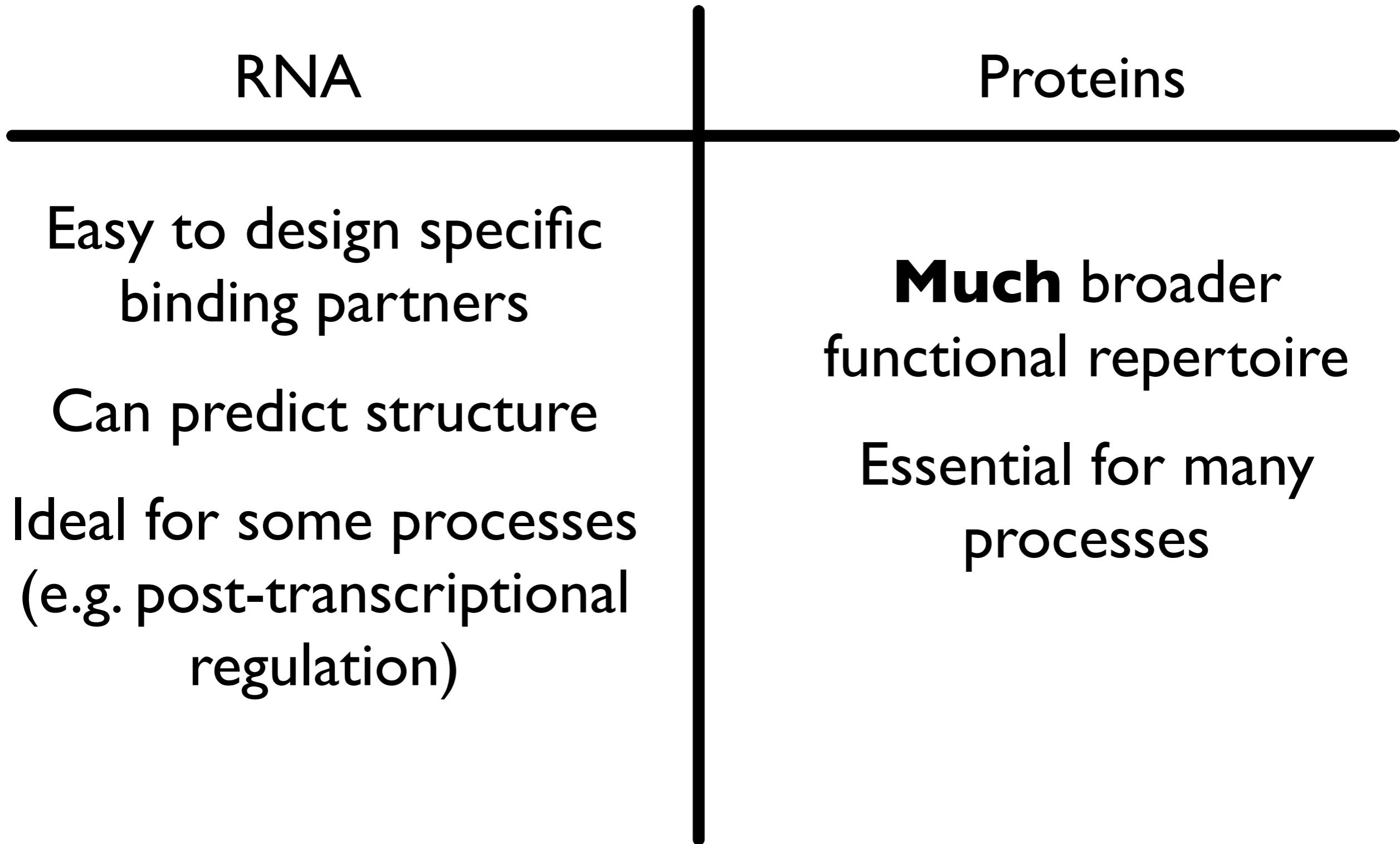
Kranz et al, 2007. Engineered high-affinity binding partner for *S.aureus* enterotoxin B (SEB).



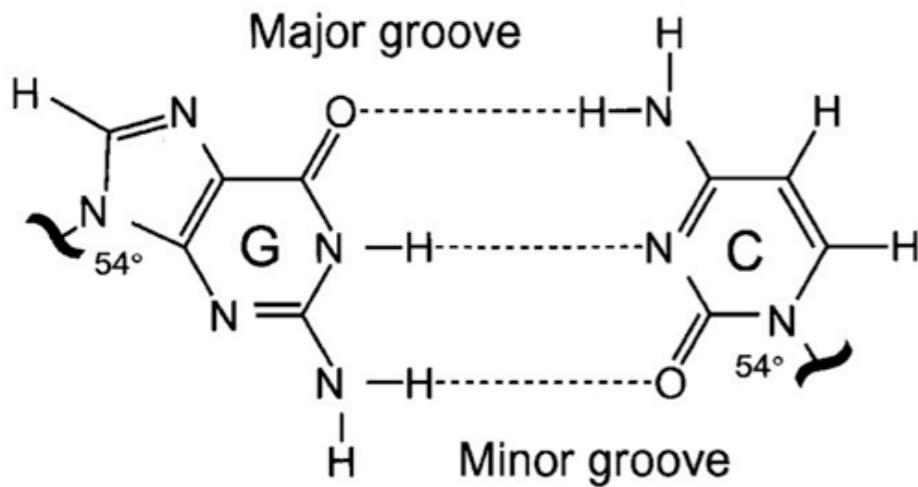
Baker et al, 2008. Computationally designed catalytic enzyme. The catalyzed reaction is shown inset



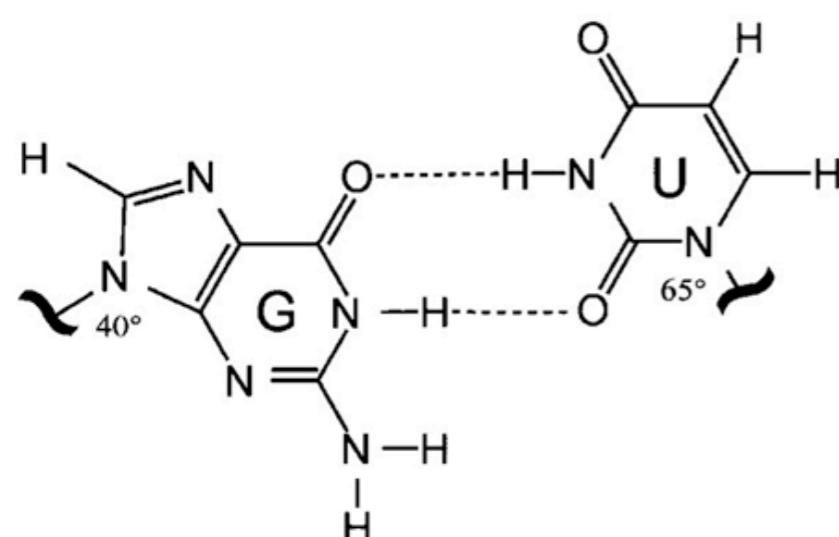
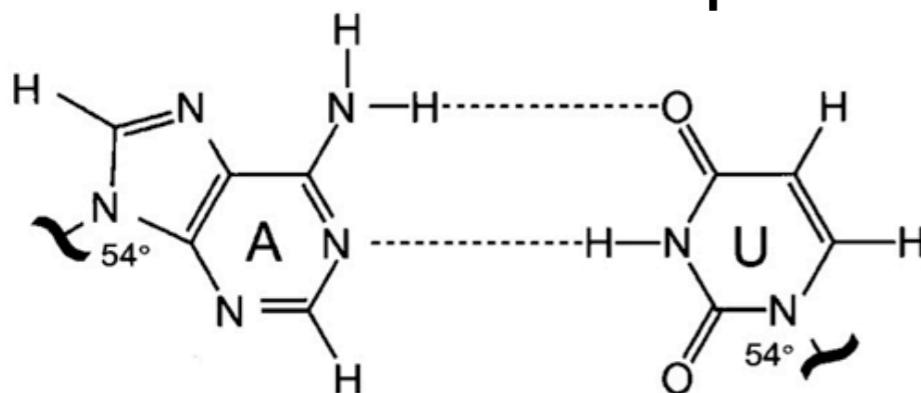
# RNA vs protein design



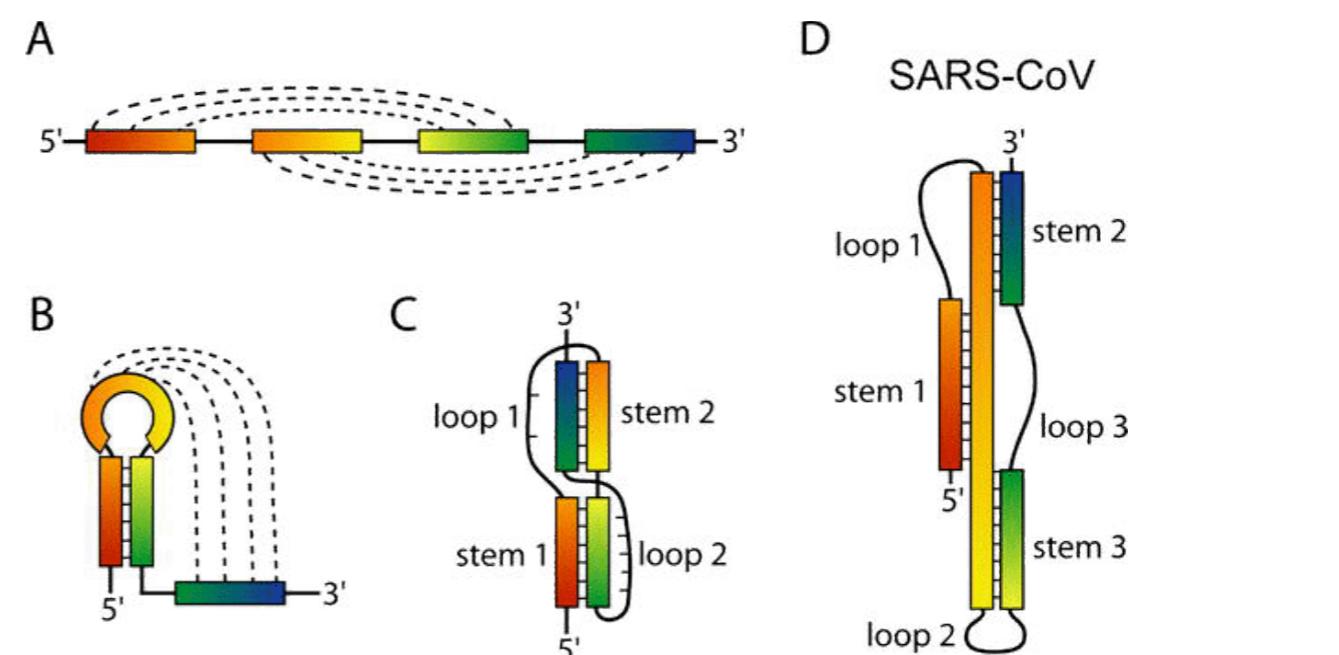
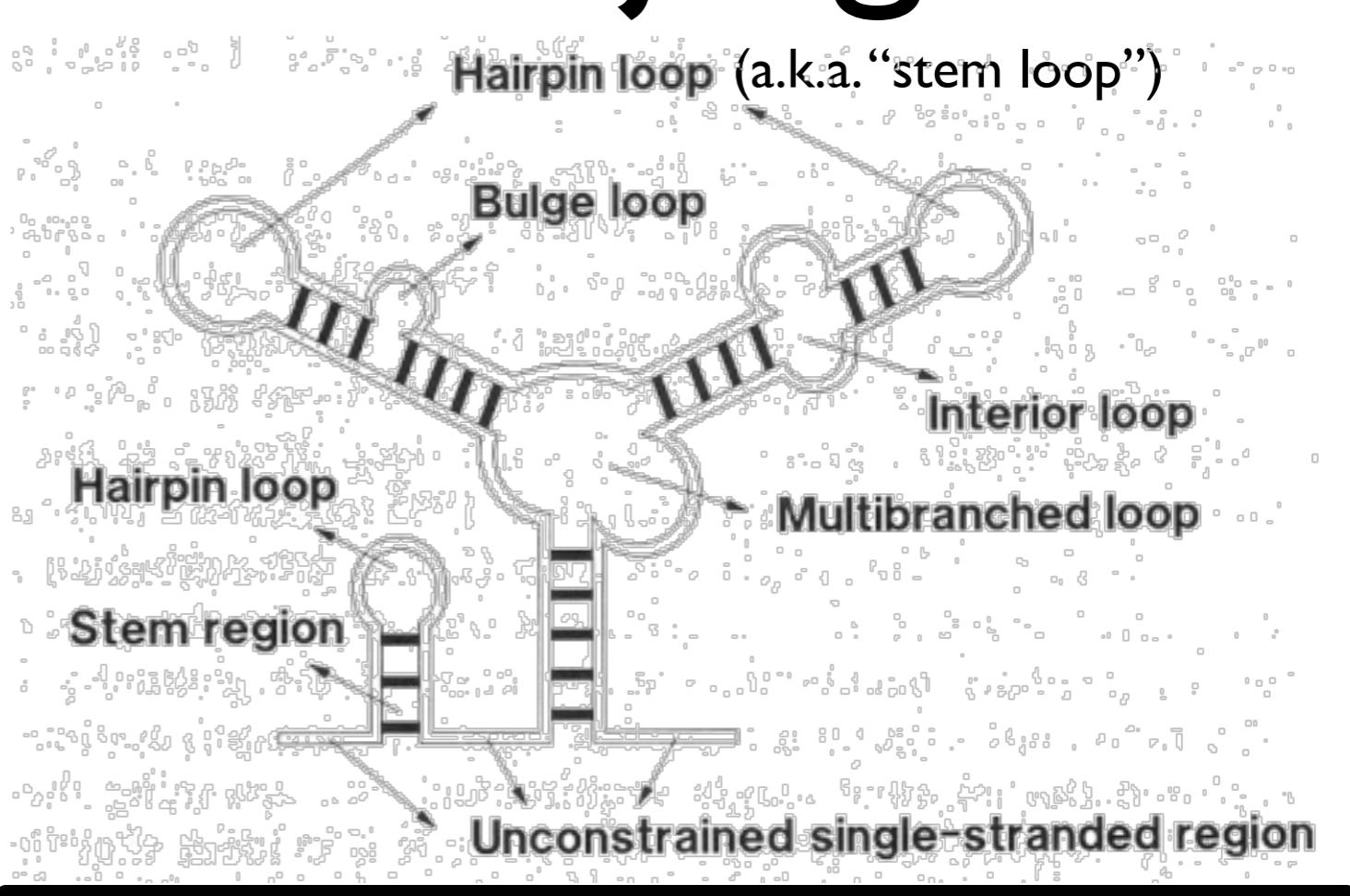
# RNA structure jargon



“Watson-Crick” base pairs

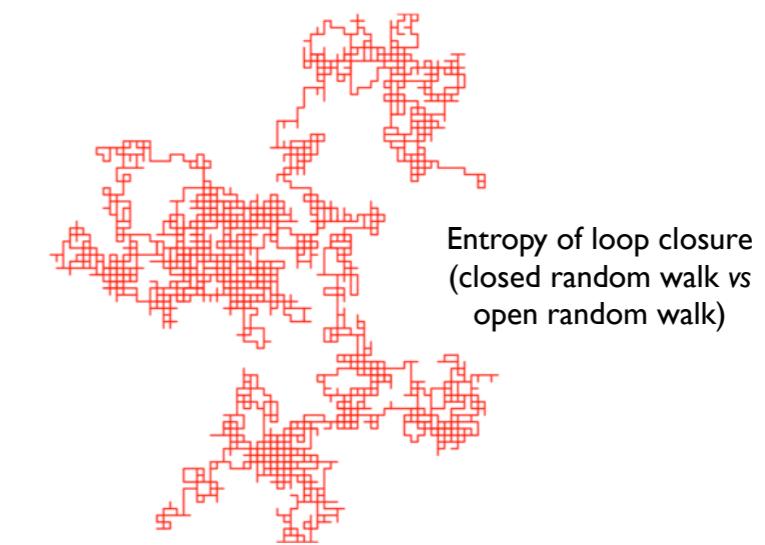


“Wobble” base pair



Pseudoknot

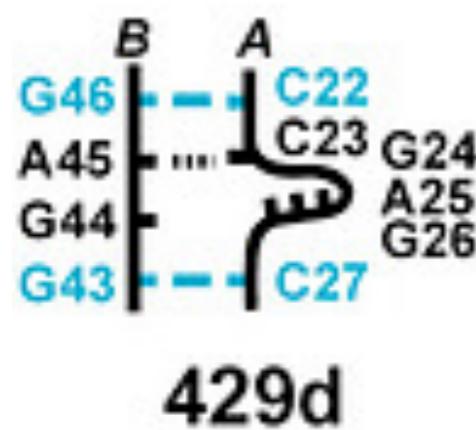
# Subtleties of RNA folding energetics



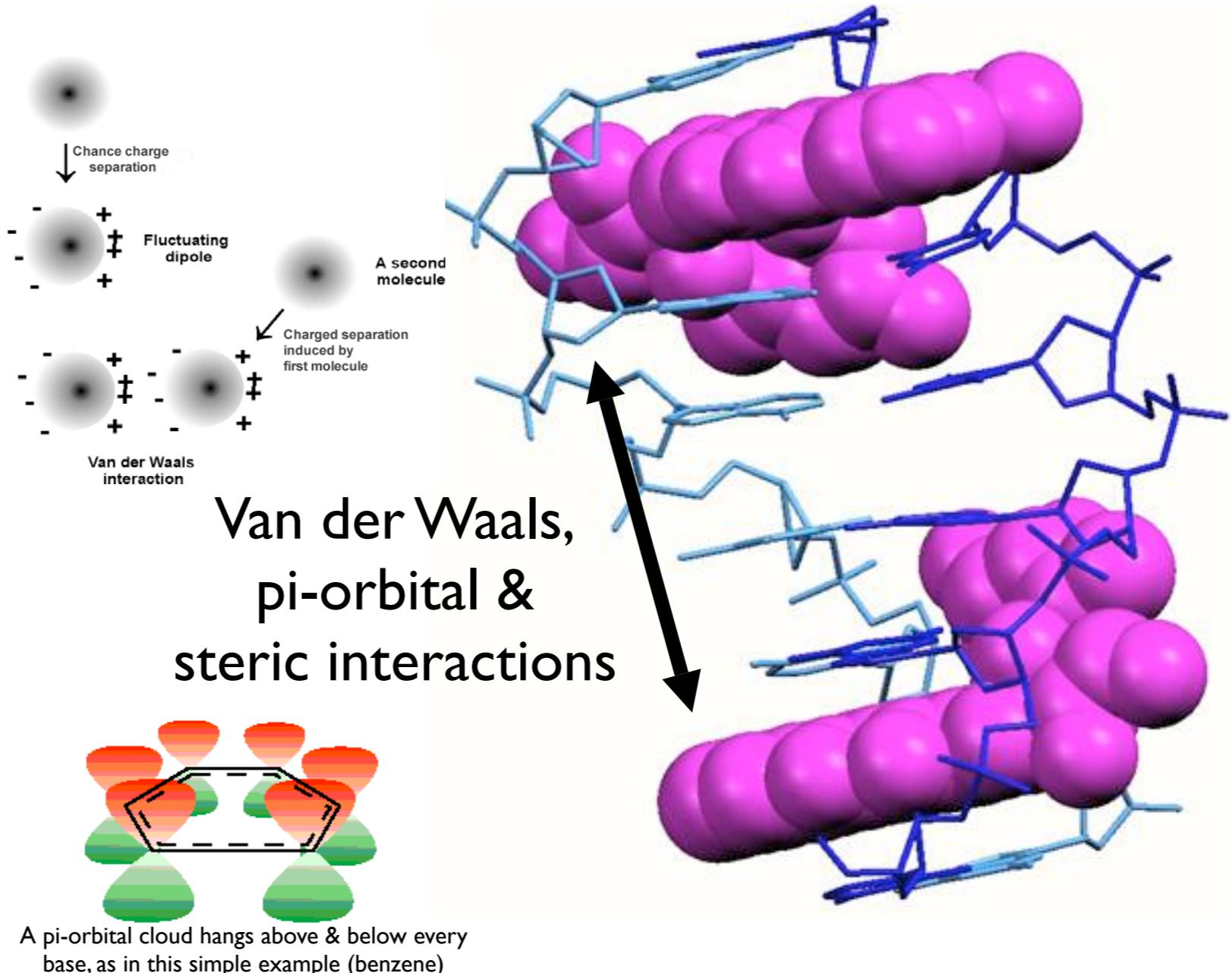
Favored loops,  
e.g. tetraloop:



Favored bulges:

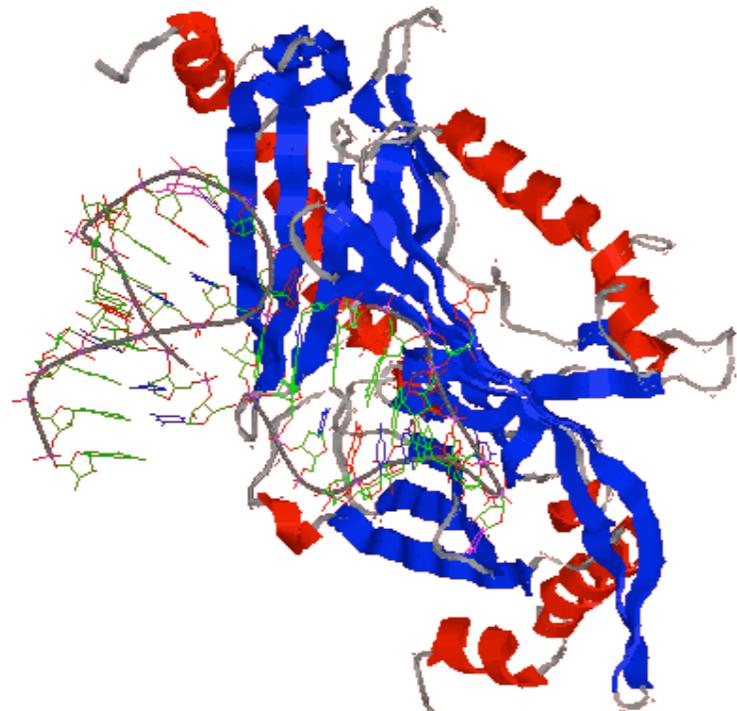


Base pair “stacking”:

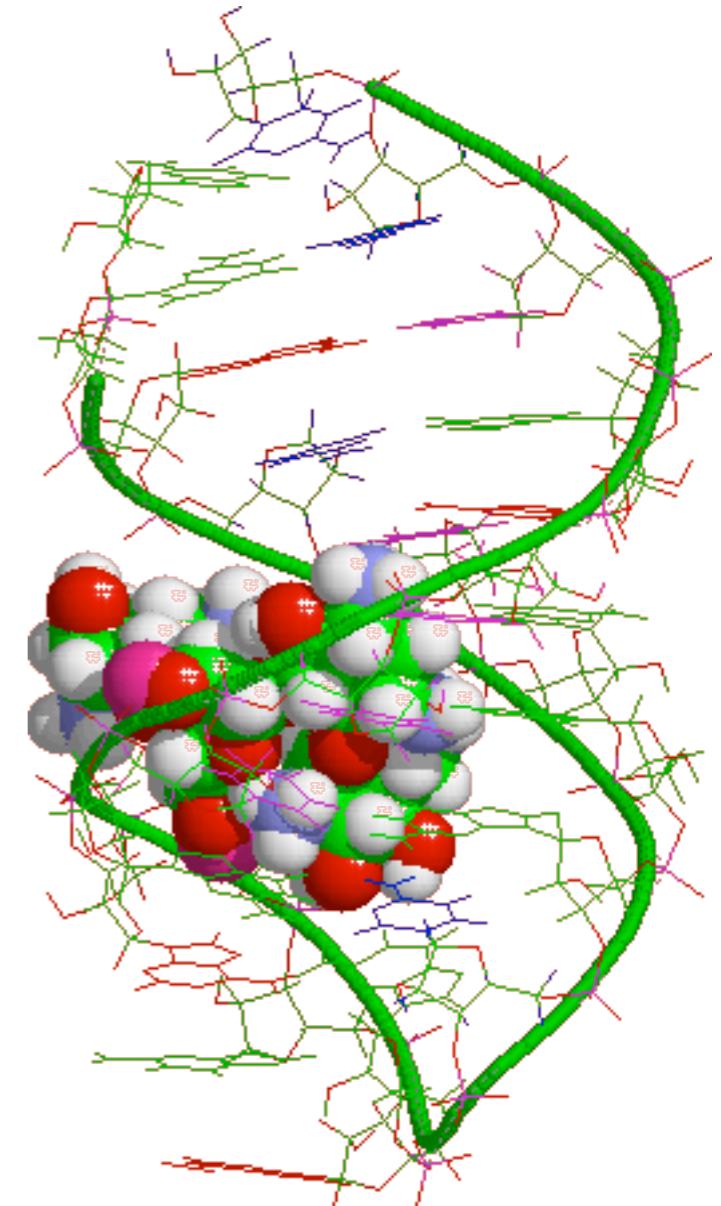


# Aptamers

- Structured RNA or DNA
- Forms binding pocket for specific ligand
- Generated by *in vitro* selection
- c.f. **antibodies**

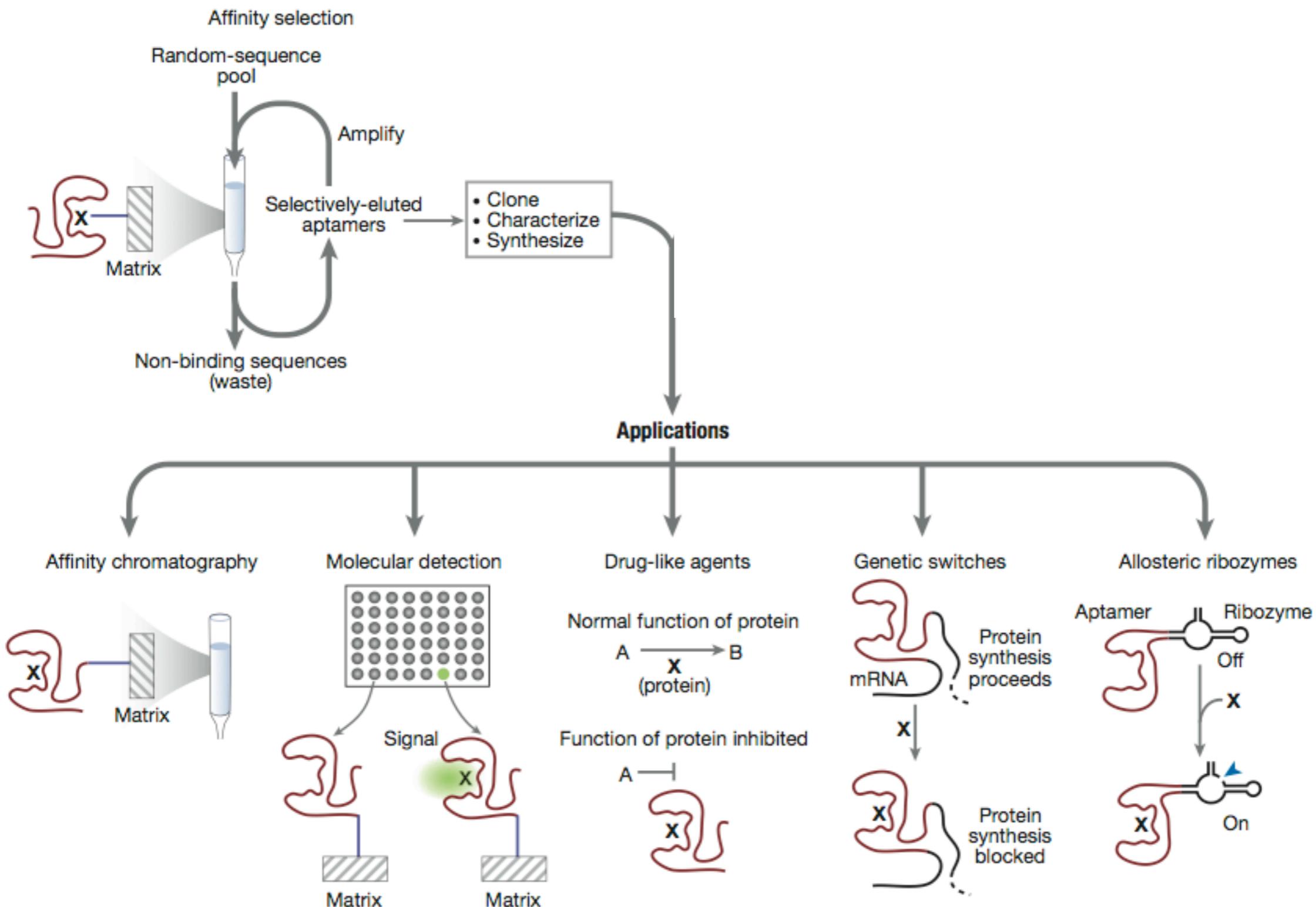


Bacteriophage MS2 coat protein aptamer



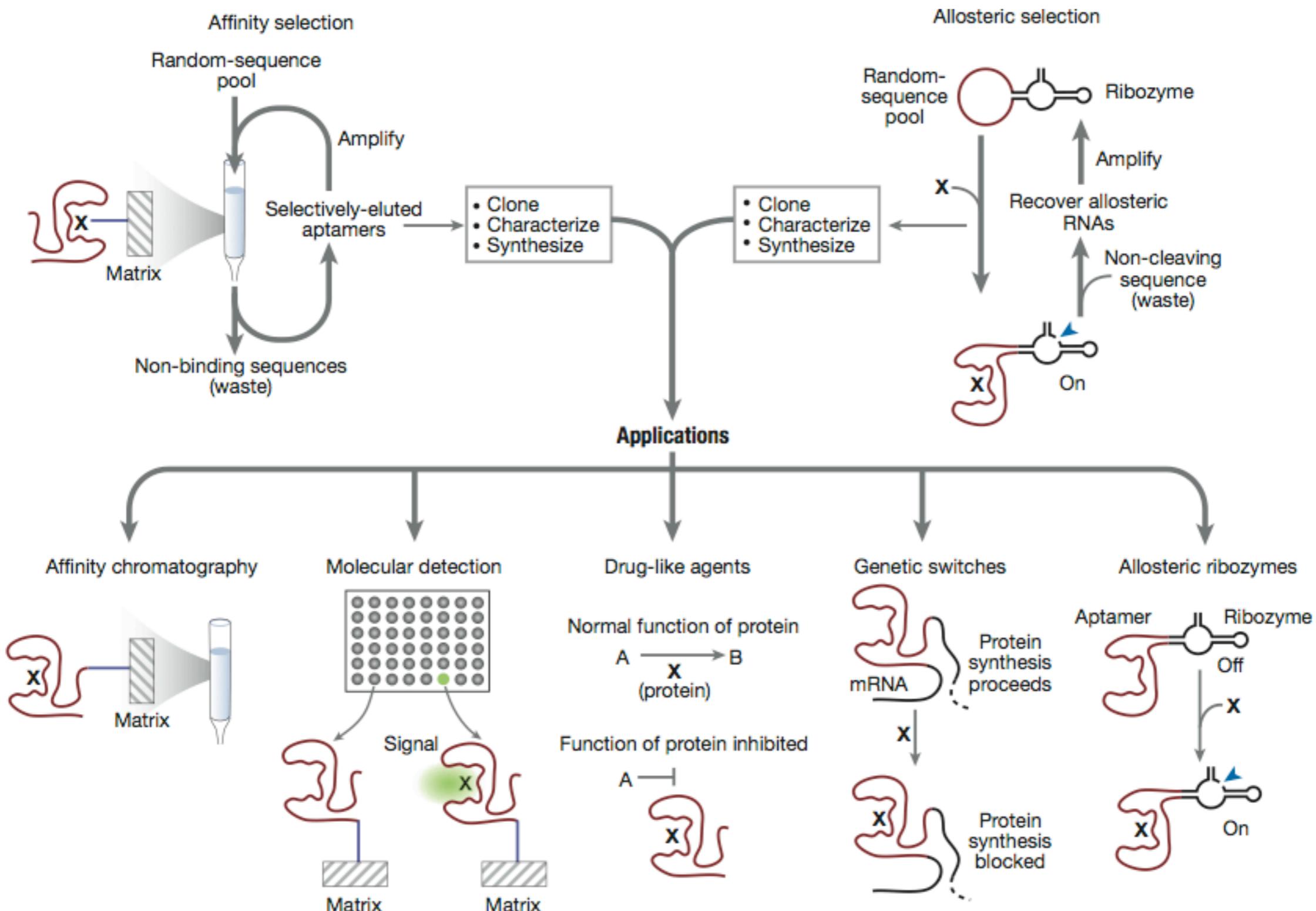
Saccharide aptamer

# *In vitro* selection



Increasingly automated process

# In vitro selection

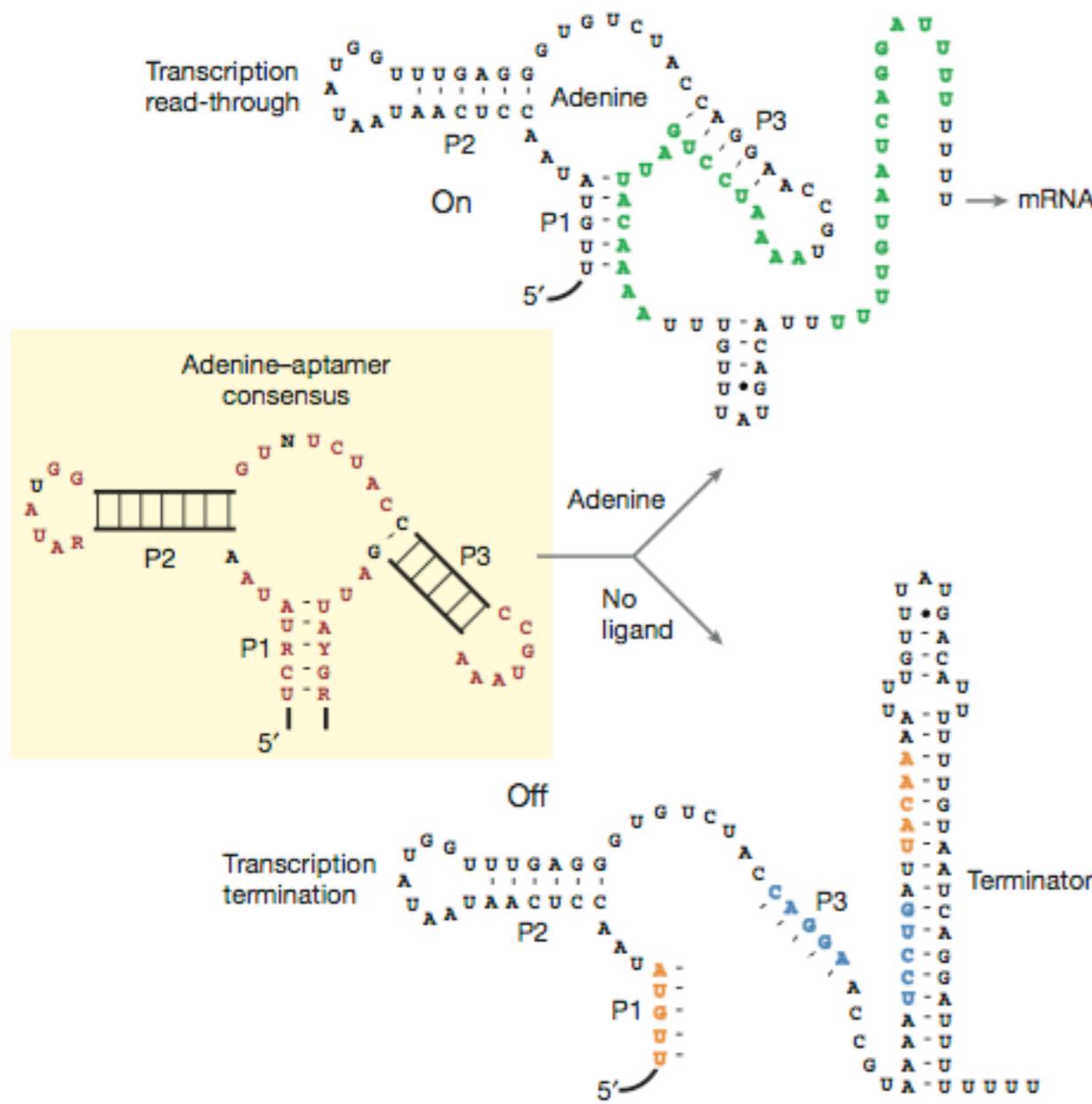


Increasingly automated process

# Riboswitches

- Naturally occurring aptamers
- Coupled to genetic switches
- Regulators of gene expression
  - Widespread in prokaryotes
    - At least 2% of genes in *Bacillus subtilis*
    - Known in eukaryotes (*Arabidopsis*, *Neurospora*...)

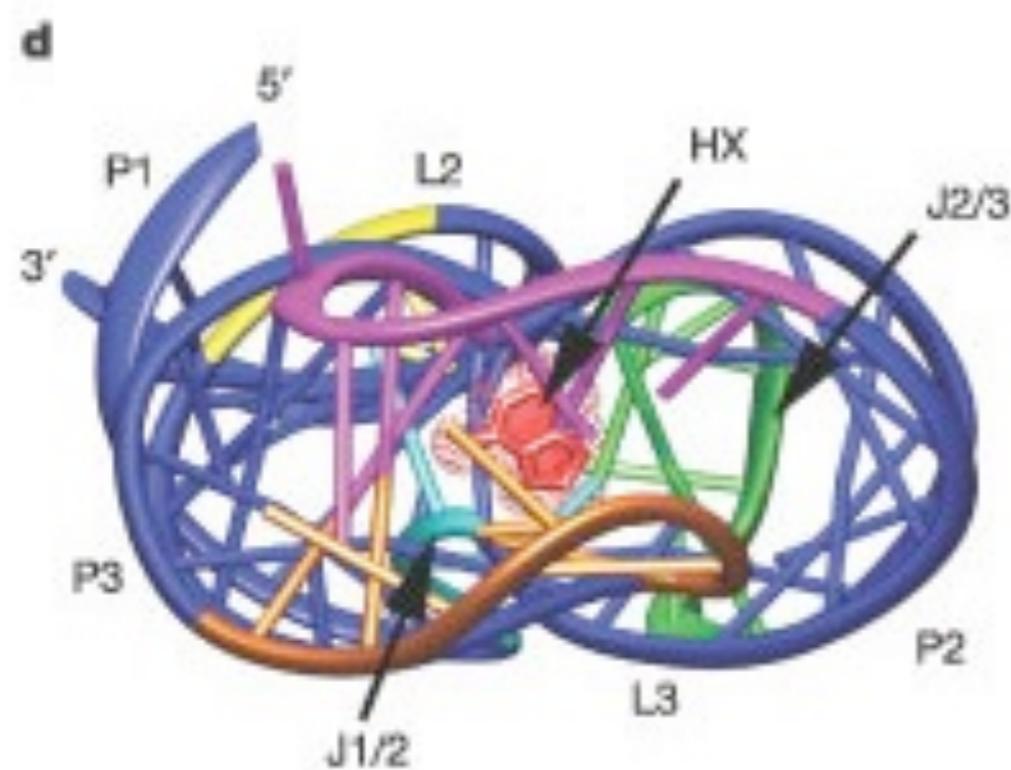
# pbuE Adenine riboswitch (*B.subtilis*)



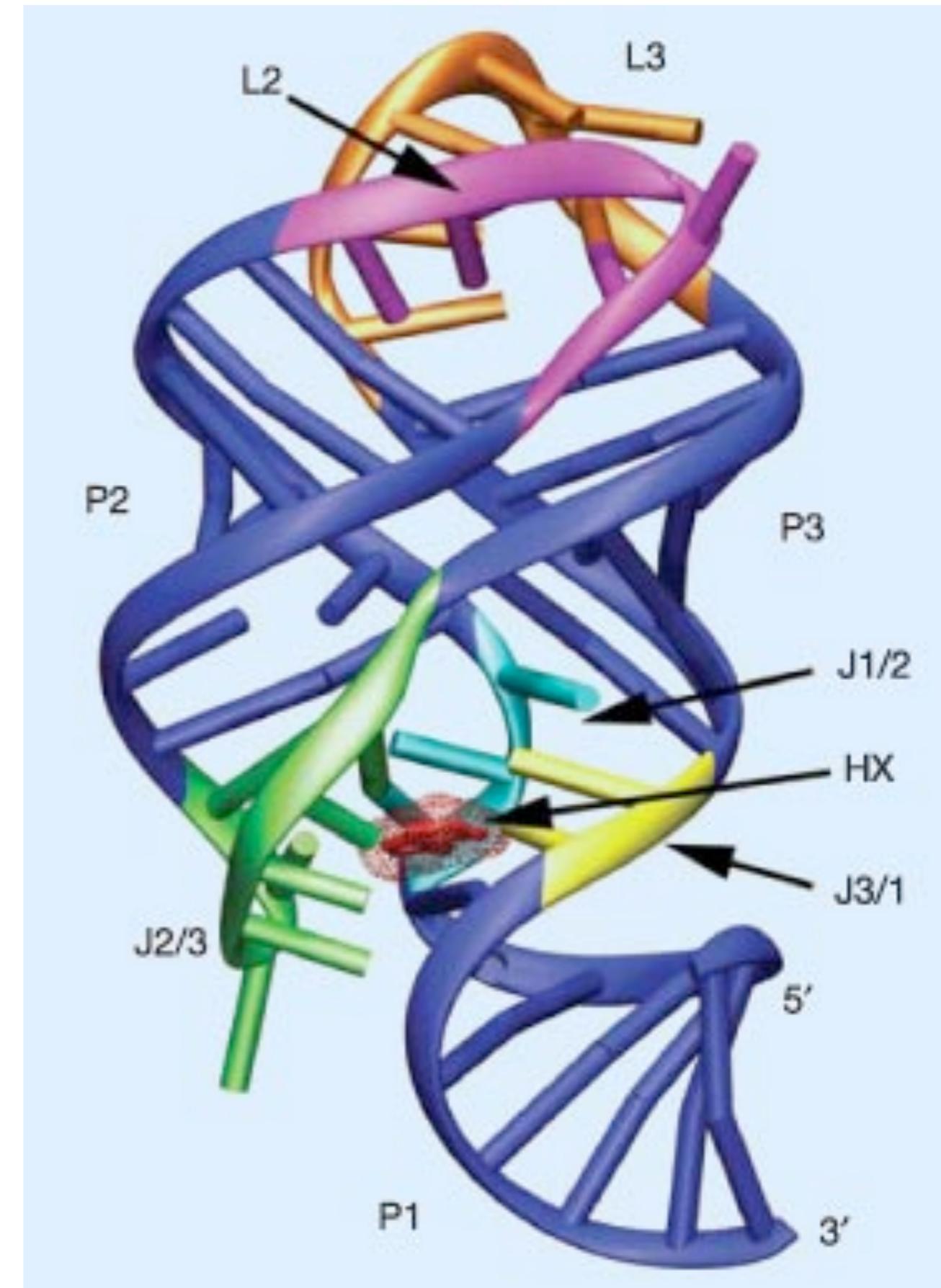
In the presence of sufficient adenine, the terminator stem-loop does not form and the mRNA is transcribed

When adenine is not present, the stem-loop forms and transcription is prematurely terminated

# Riboswitch-ligand complex is 3-dimensional

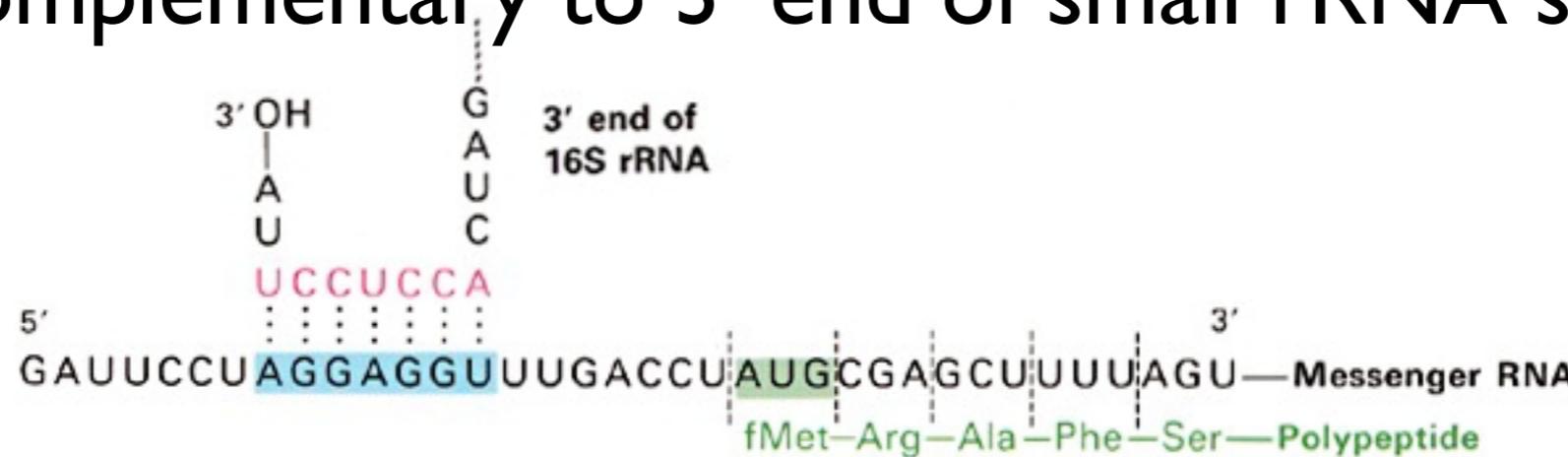


Ligand is hypoxanthine, a metabolite of guanine



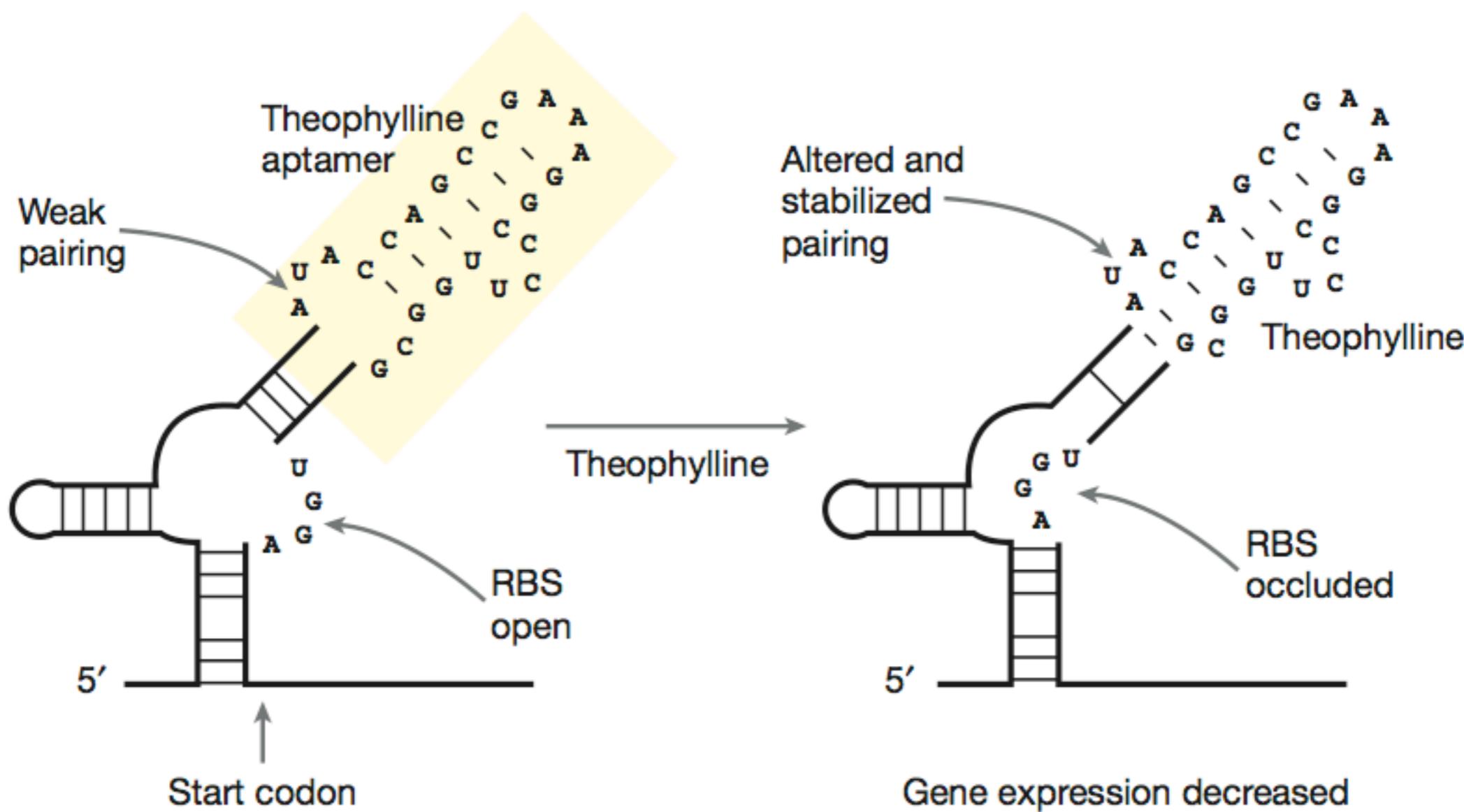
# A synthetic riboswitch

- Theophylline (dimethylxanthine)
  - Anti-asthmatic drug; naturally found in tea
- Riboswitch: presence of theophylline prevents translation
- Works by blocking the Ribosome Binding Site (**RBS**), a.k.a.
  - the **Shine-Dalgarno** sequence (prokaryotes)
  - complementary to 3' end of small rRNA subunit



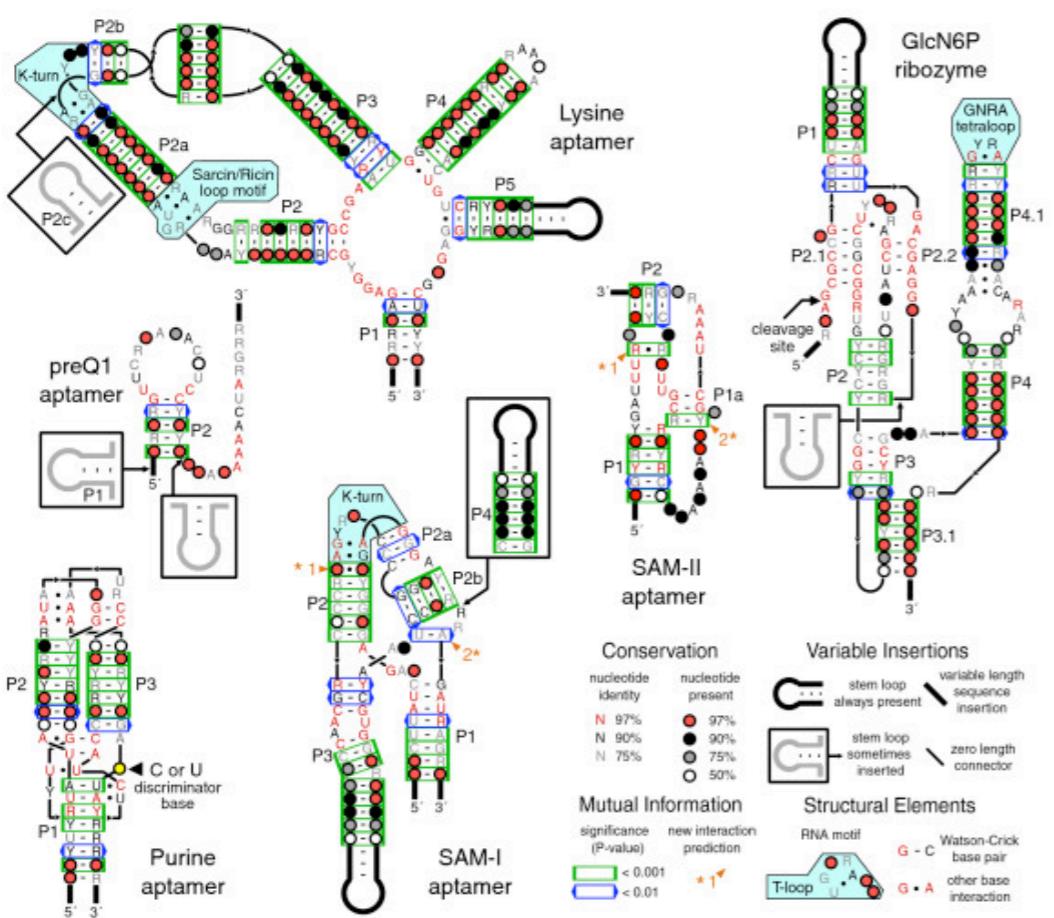
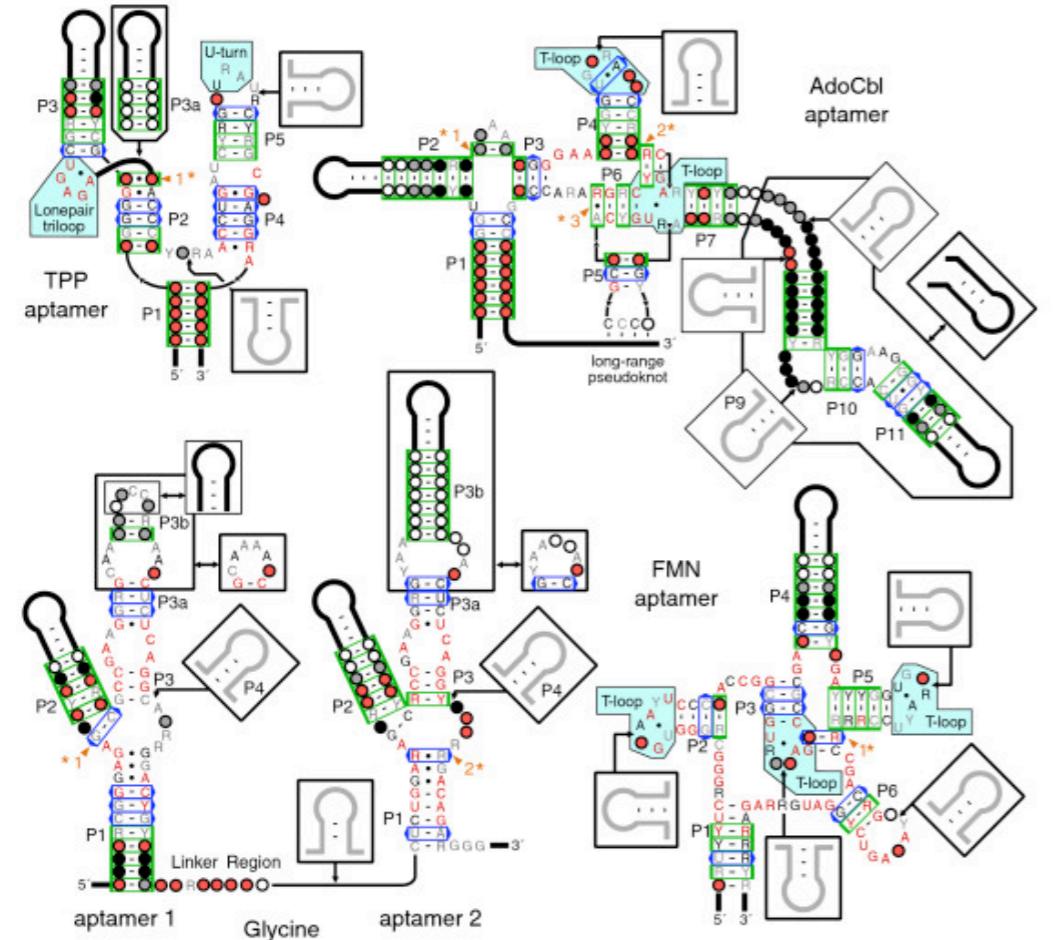
- Eukaryotic equivalent is the **Kozak** sequence

# Theophylline riboswitch



Two parts, each independently selected *in vitro*.  
Mechanism of RBS occlusion unclear (helix slippage?)

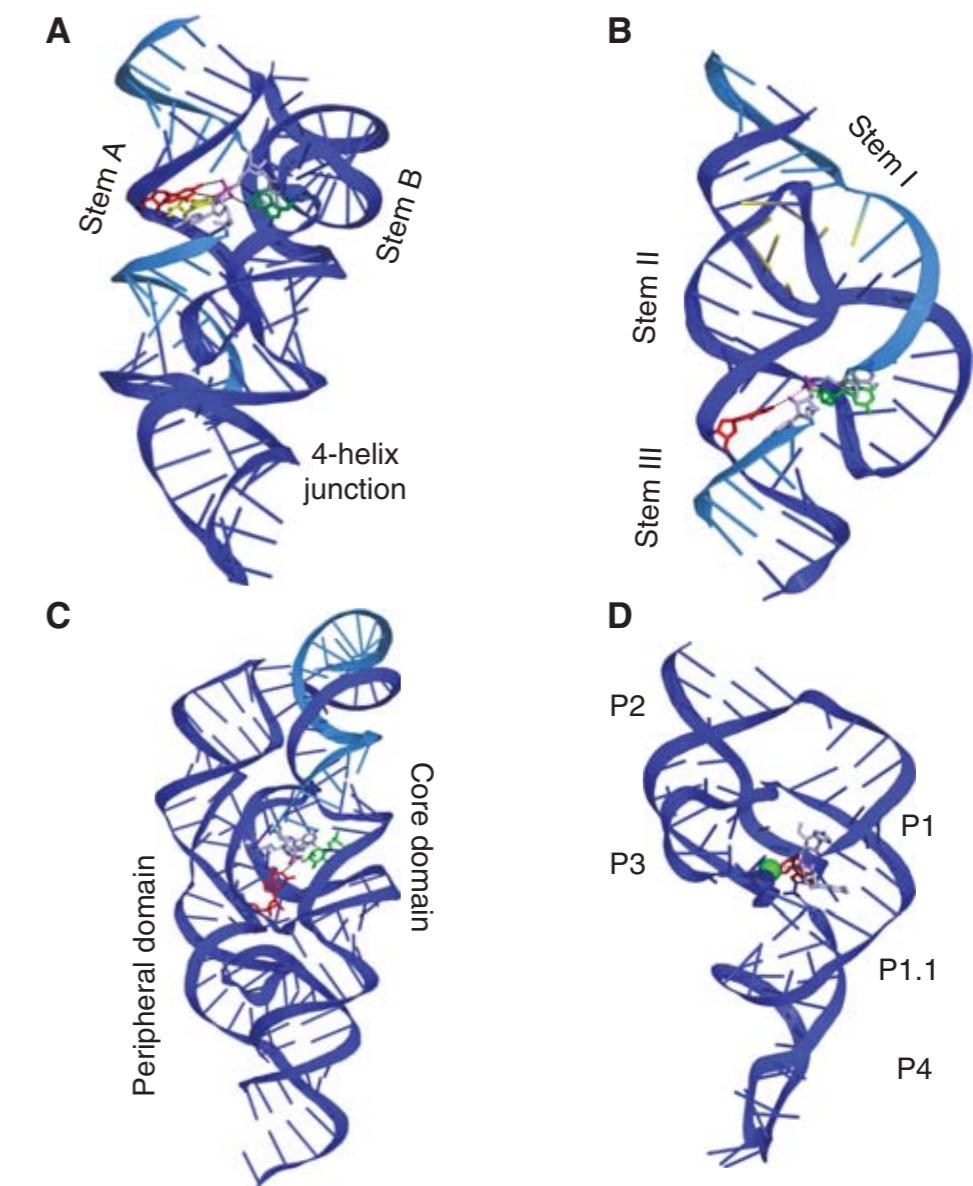
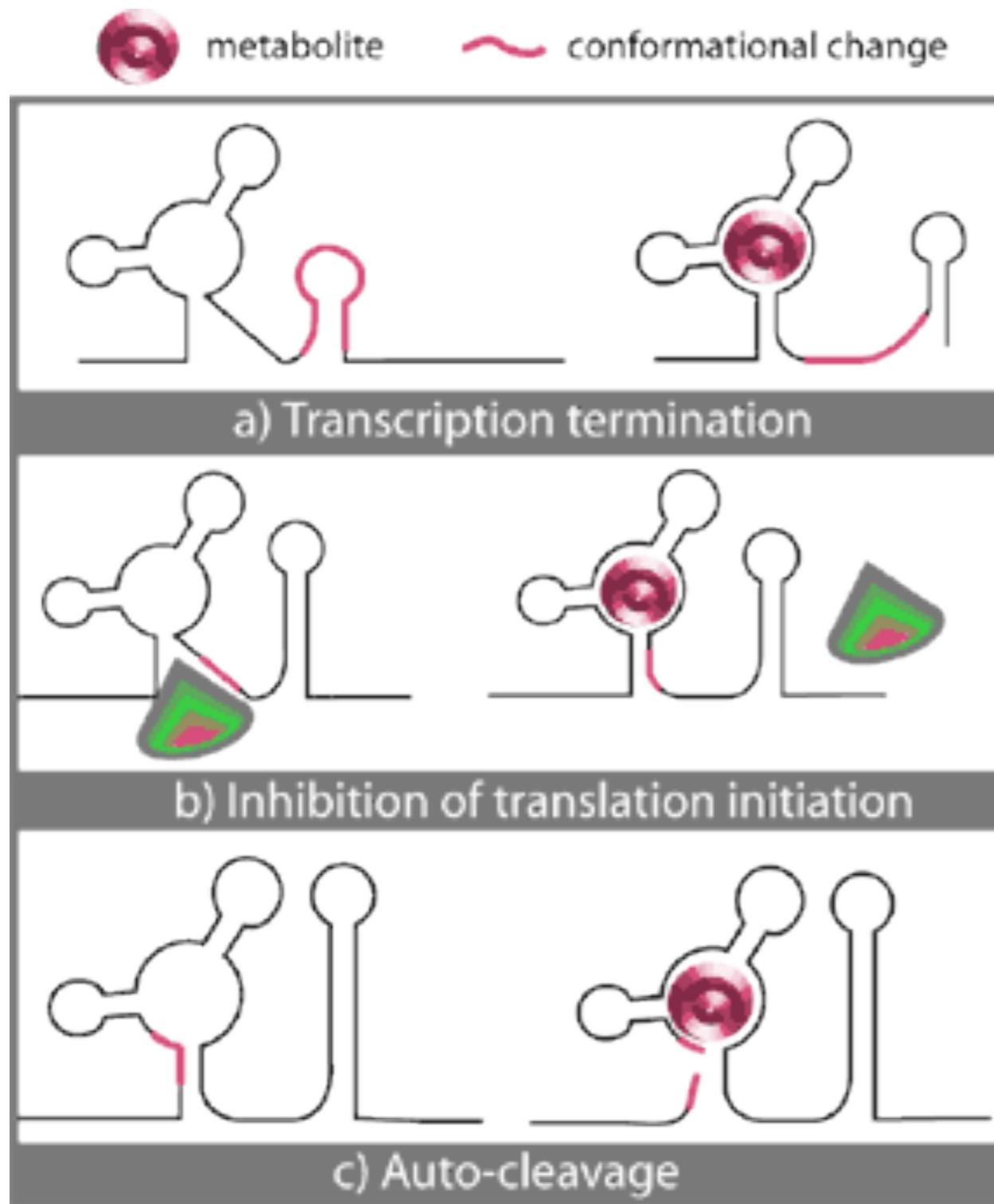
# Riboswitch structures



Barrick and Breaker  
Genome Biology 2007  
8:R239

<http://biowiki.org/RfamRiboswitches>

# Riboswitch mechanisms



**Figure 3.** Cartoon representations of the overall structures of four self-cleaving ribozymes. (A) The hairpin ribozyme, (B) the hammerhead ribozyme, (C) the *glmS* ribozyme-riboswitch, and (D) the hepatitis delta virus (HDV) ribozyme. Residues implicated in general acid and base catalysis in the cleavage reaction are green and red, respectively. The scissile phosphate and nucleophilic 2'-O atoms are magenta, flanked by substrate residues shown in light blue.

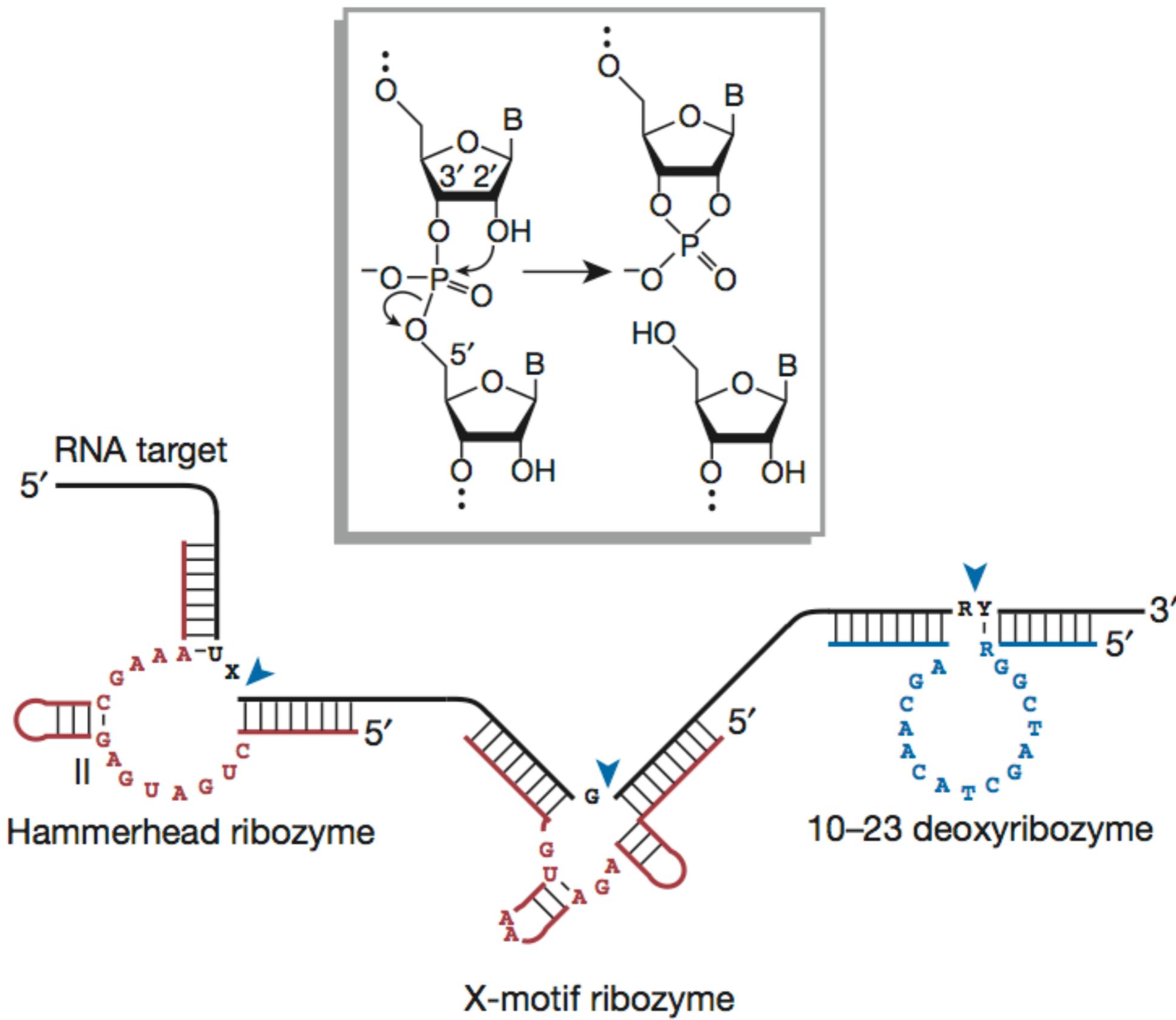
# Aptamers as inputs

- Aptamers/riboswitches are like inputs
- Example applications:
  - Immobilize on gel/support for affinity **chromatography**
  - Use as **biosensors**
  - Combine with **fluorescent reporters**
  - Trigger downstream genetic **circuits**

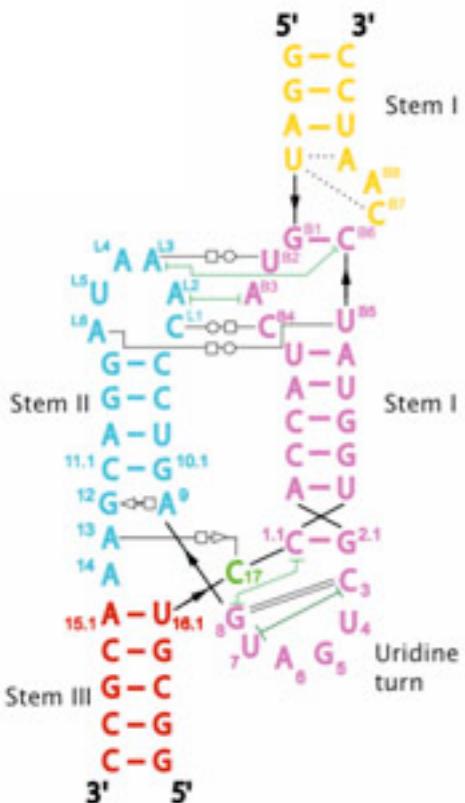
# Ribozymes

- Ribozyme = RNA enzyme
- Limited chemical repertoire in nature:
  - Peptide bond formation (ribosome)
  - Phosphodiester bond cleavage/formation
  - Hammerhead, HDV, RNase P, self-splicing introns
- *In vitro* evolution has found more reactions, e.g.:
  - Covalent attachment of RNA to proteins

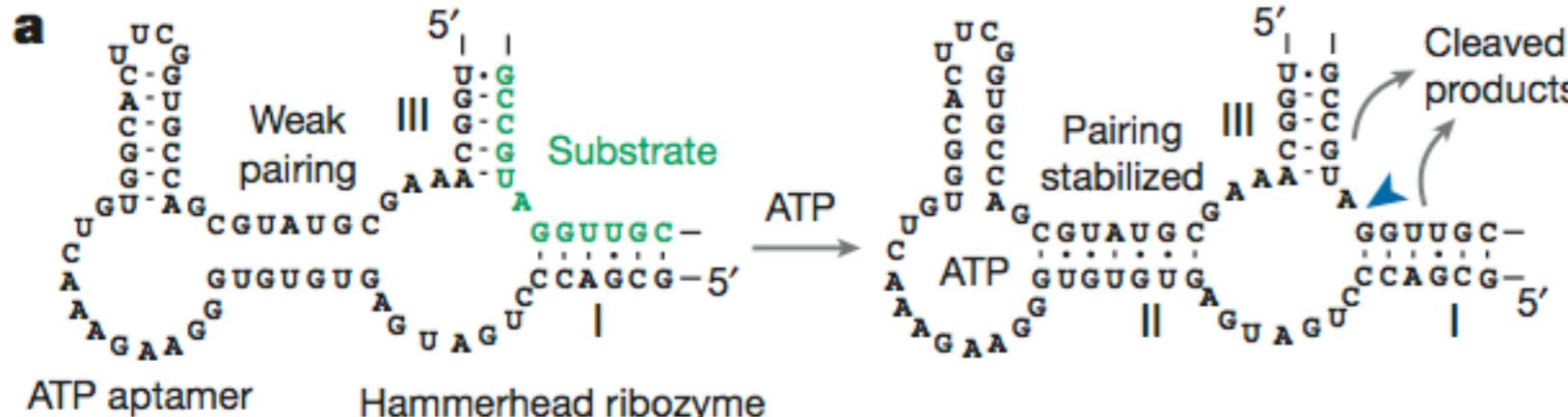
# Phosphodiester-cleaving ribozymes



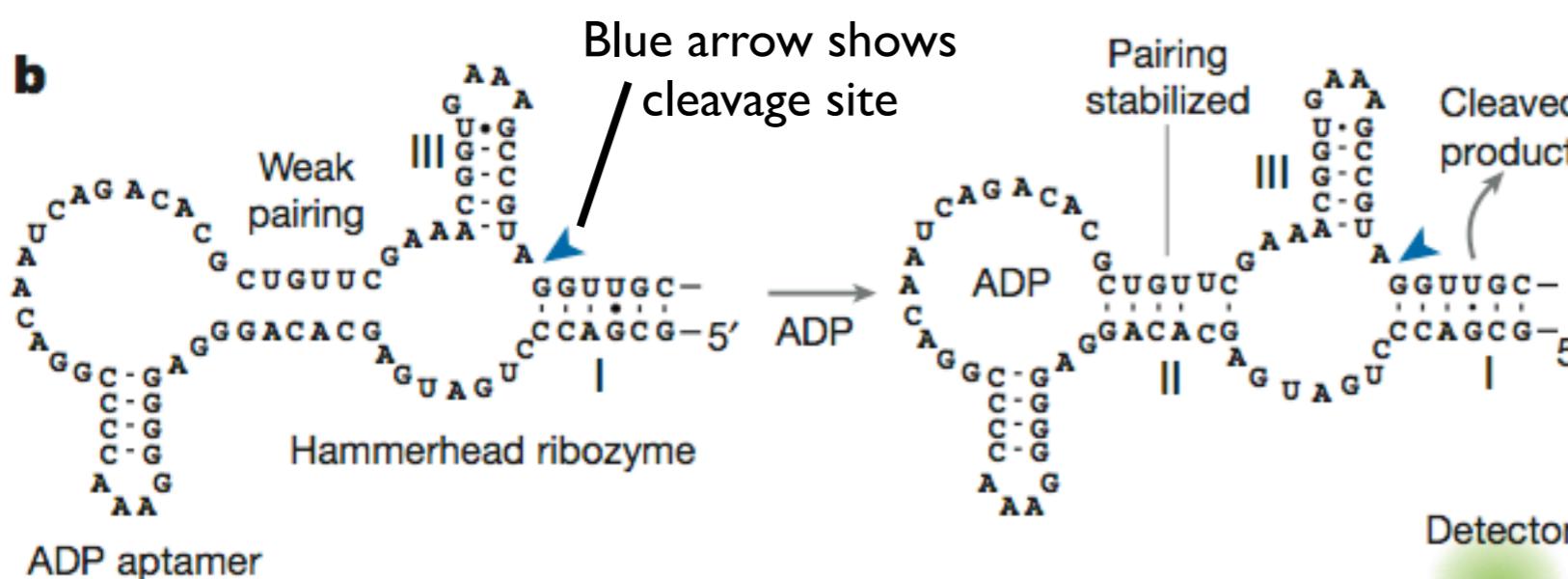
# Hammerhead ribozyme



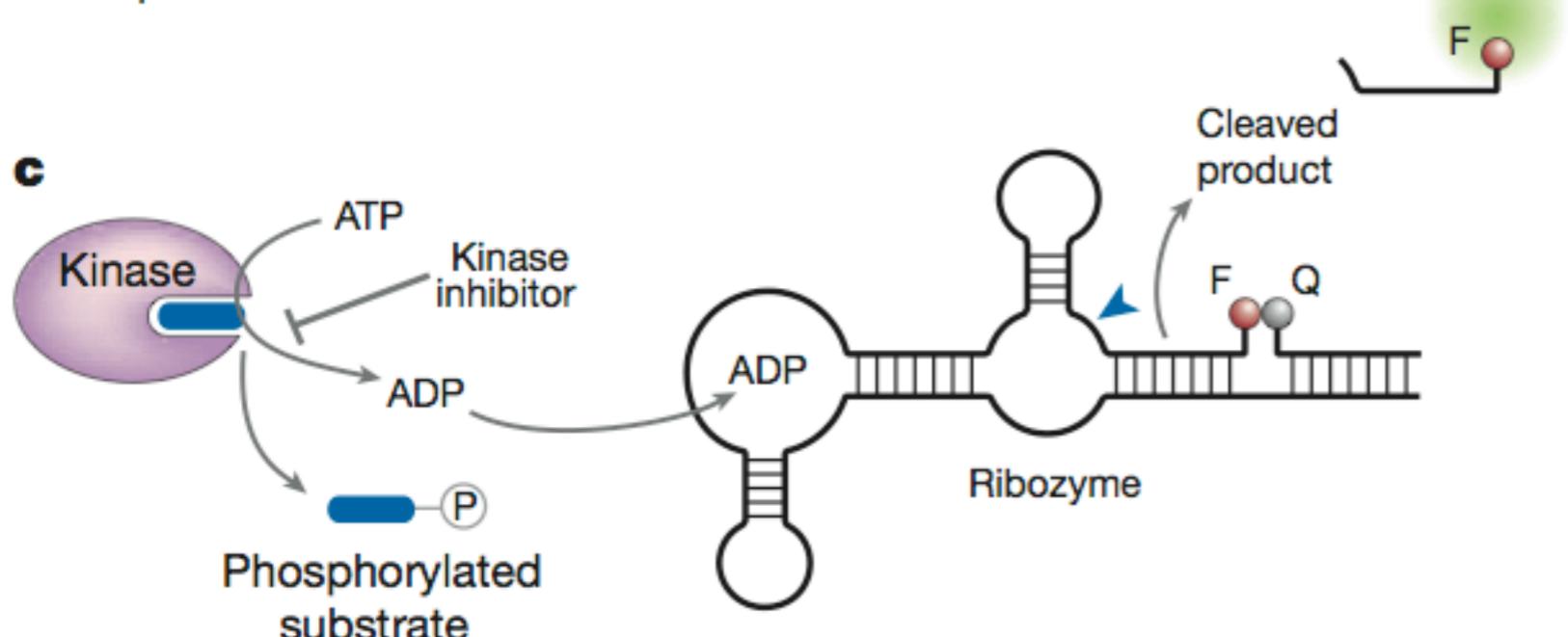
# Engineered allosteric ribozymes as biosensors



ATP-sensing aptamer +  
self-cleaving ribozyme



This variant favors ADP  
binding 100-fold over ATP



Can be used as an *in vitro*  
sensor for enzyme activity  
(or for enzyme inhibition)

# Outputs



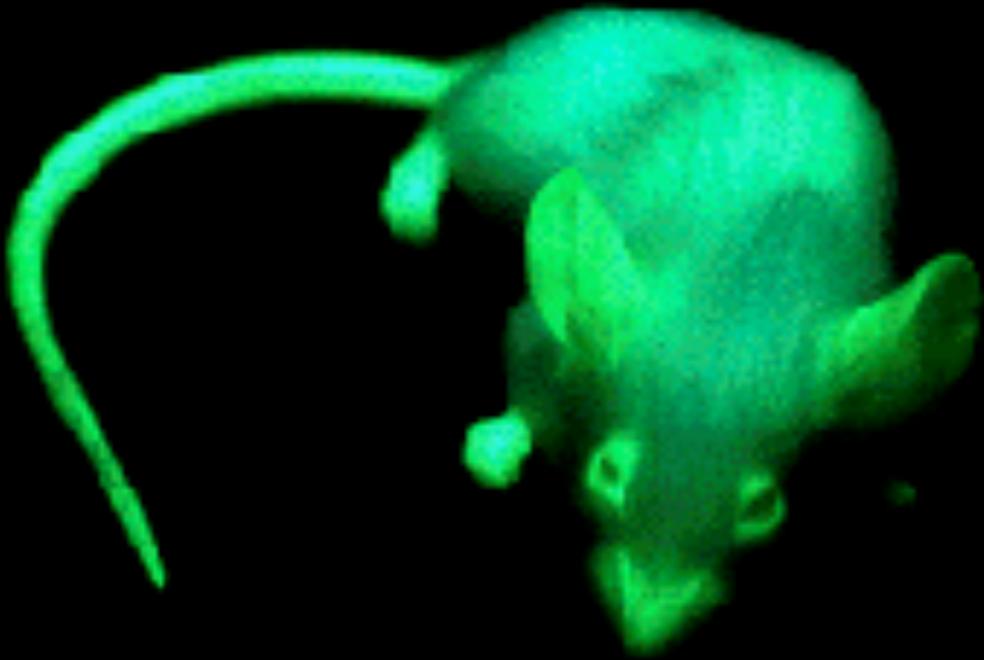
Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



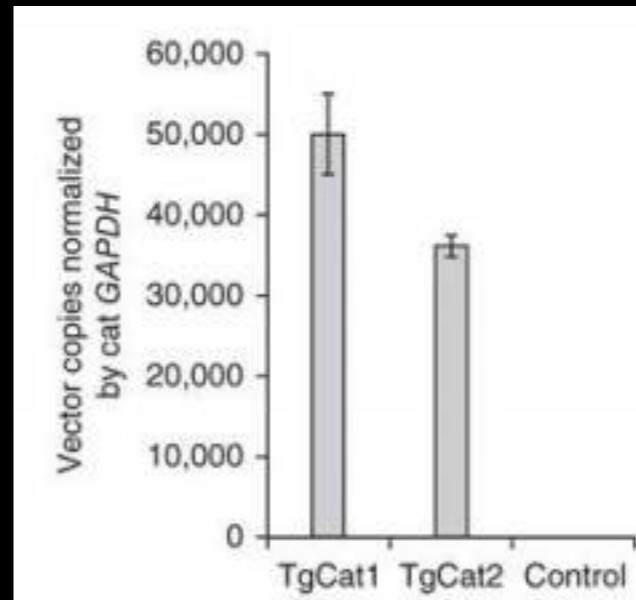
Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



Most common output is **Green Fluorescent Protein (GFP)**



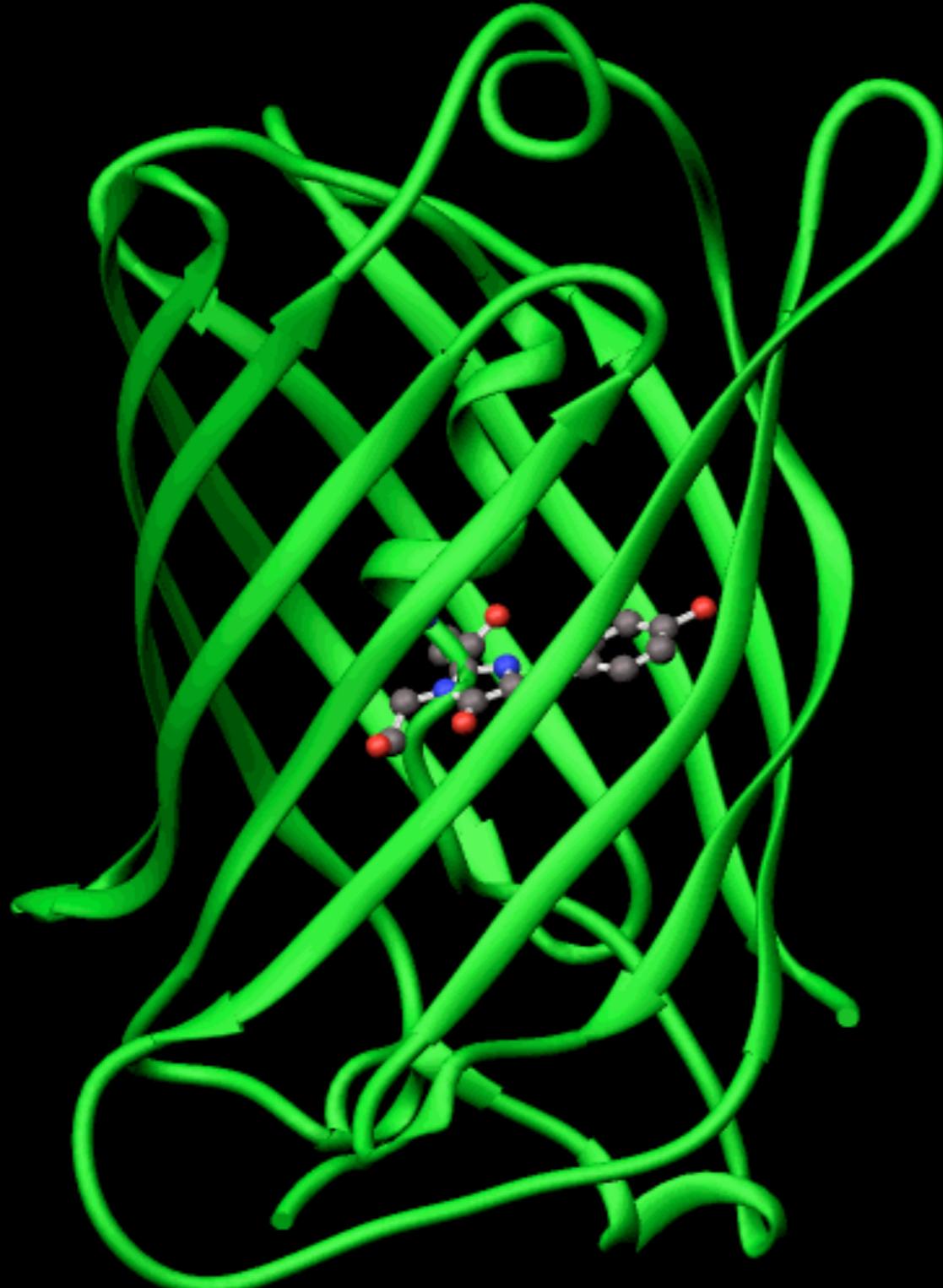
# Outputs

# Your glowstix

Wrote a dem

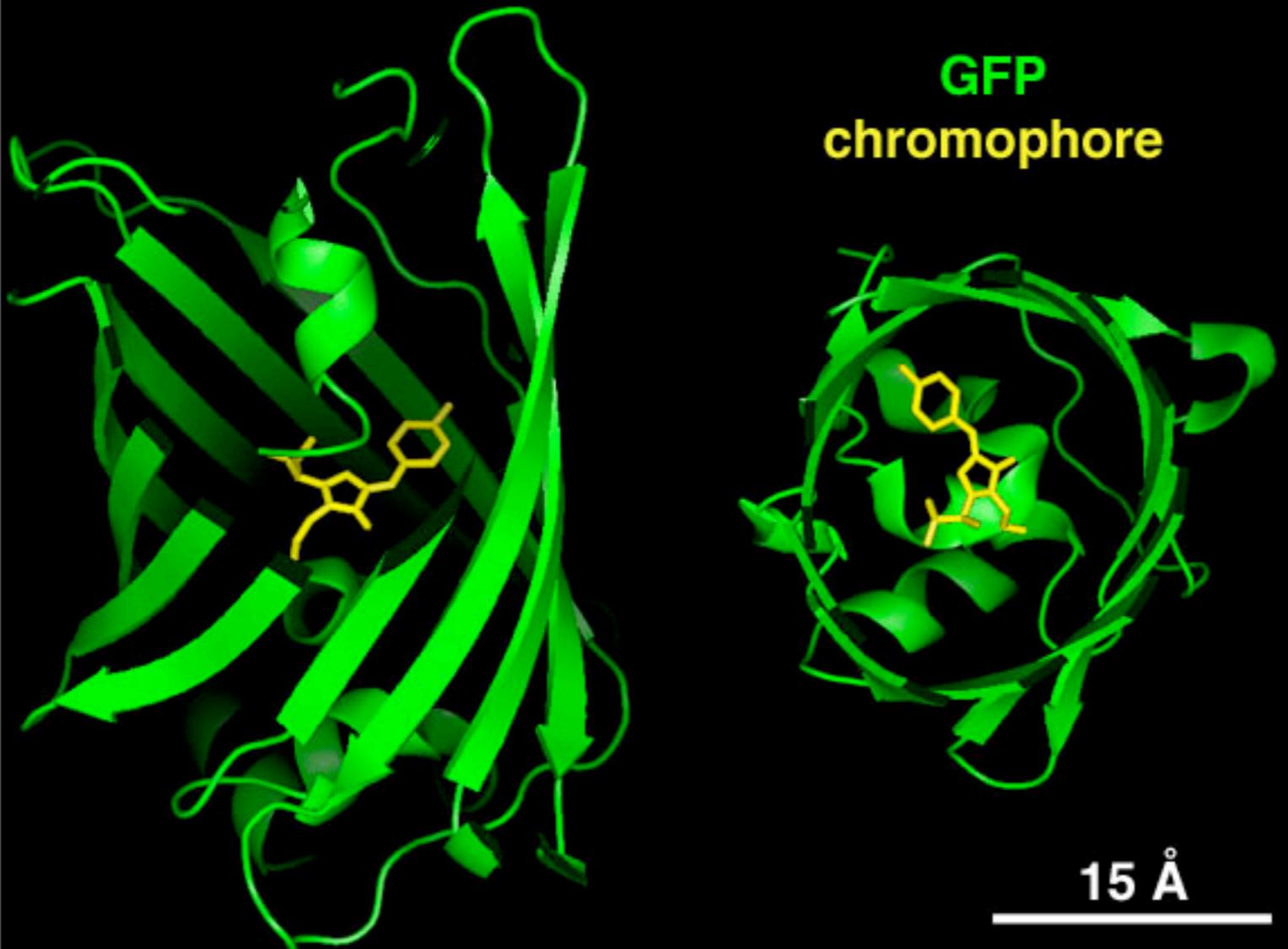
Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



Most common output is **Green Fluorescent Protein (GFP)**

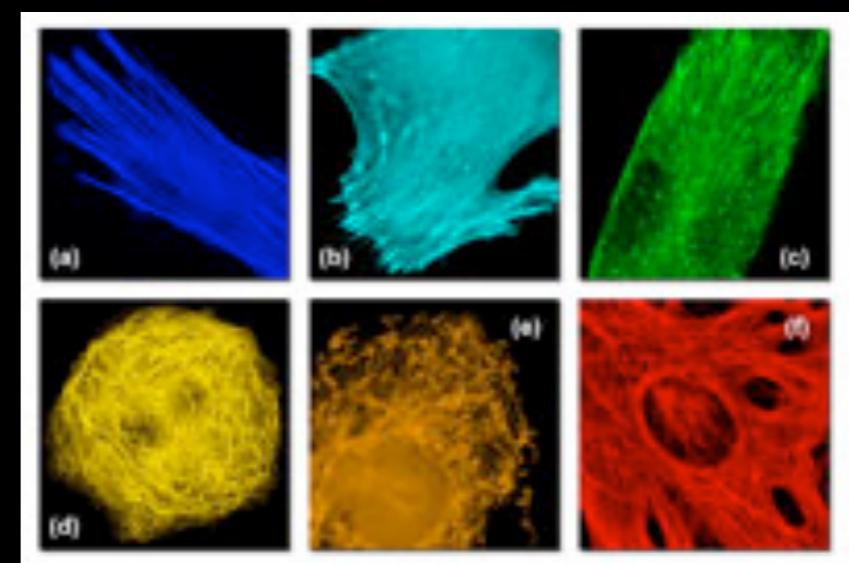
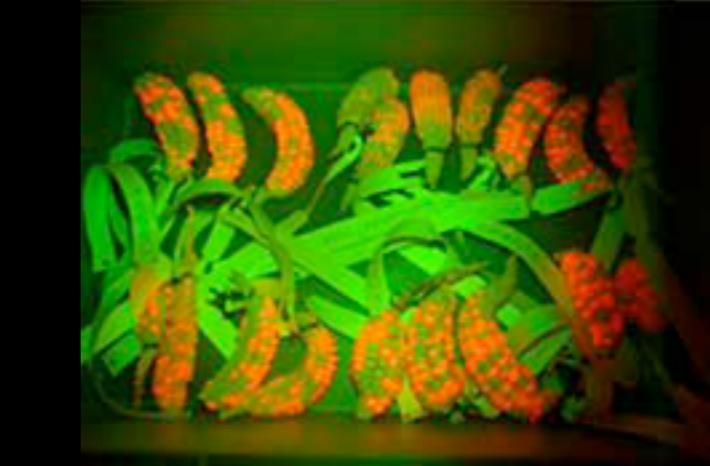
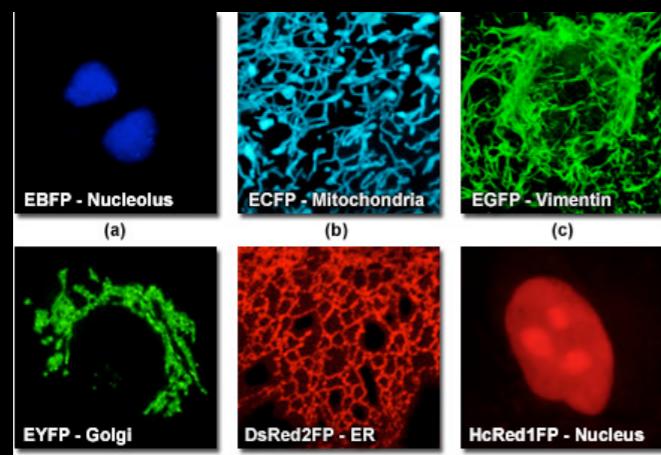
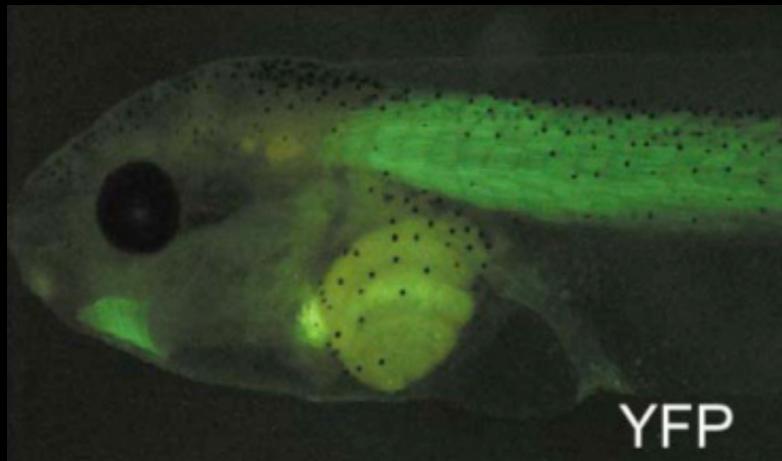
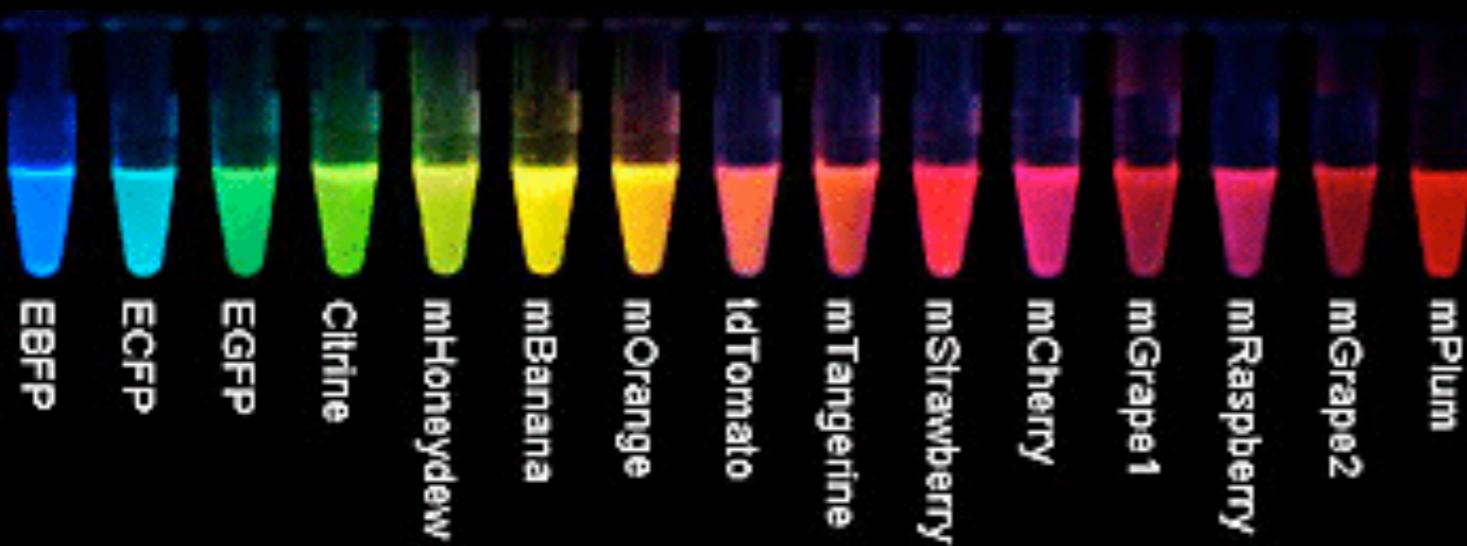
# Outputs



(Chromophore is cyclized tripeptide Ser-Tyr-Gly)

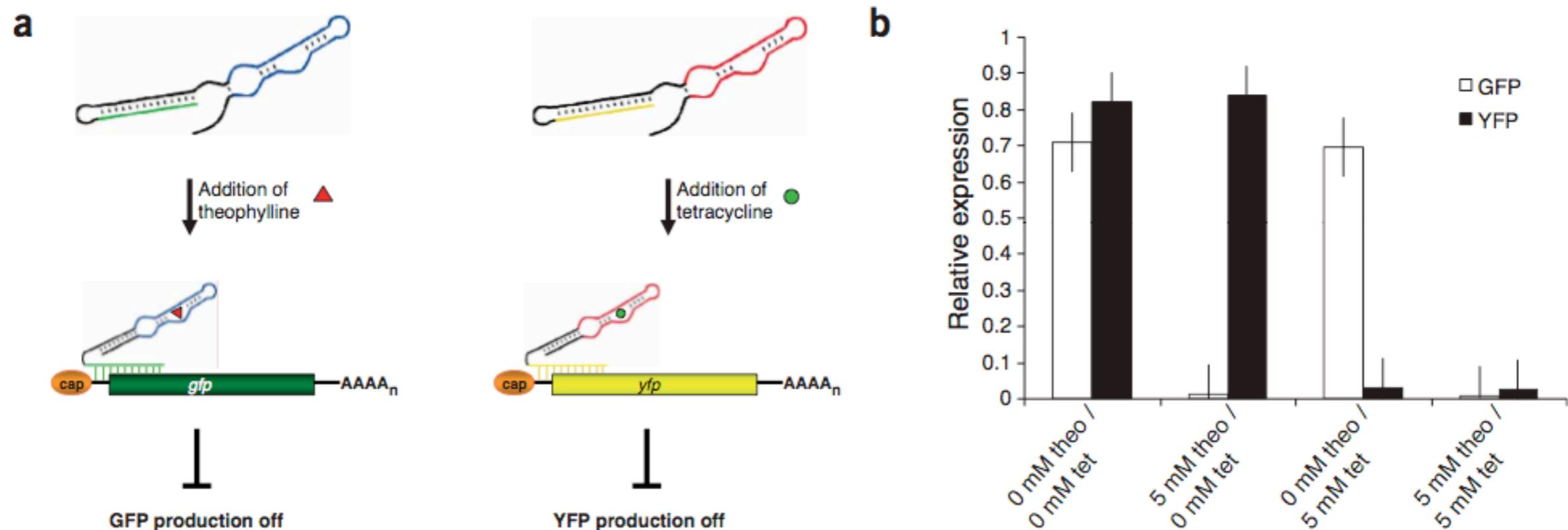
Most common output is **Green Fluorescent Protein (GFP)**

# Outputs



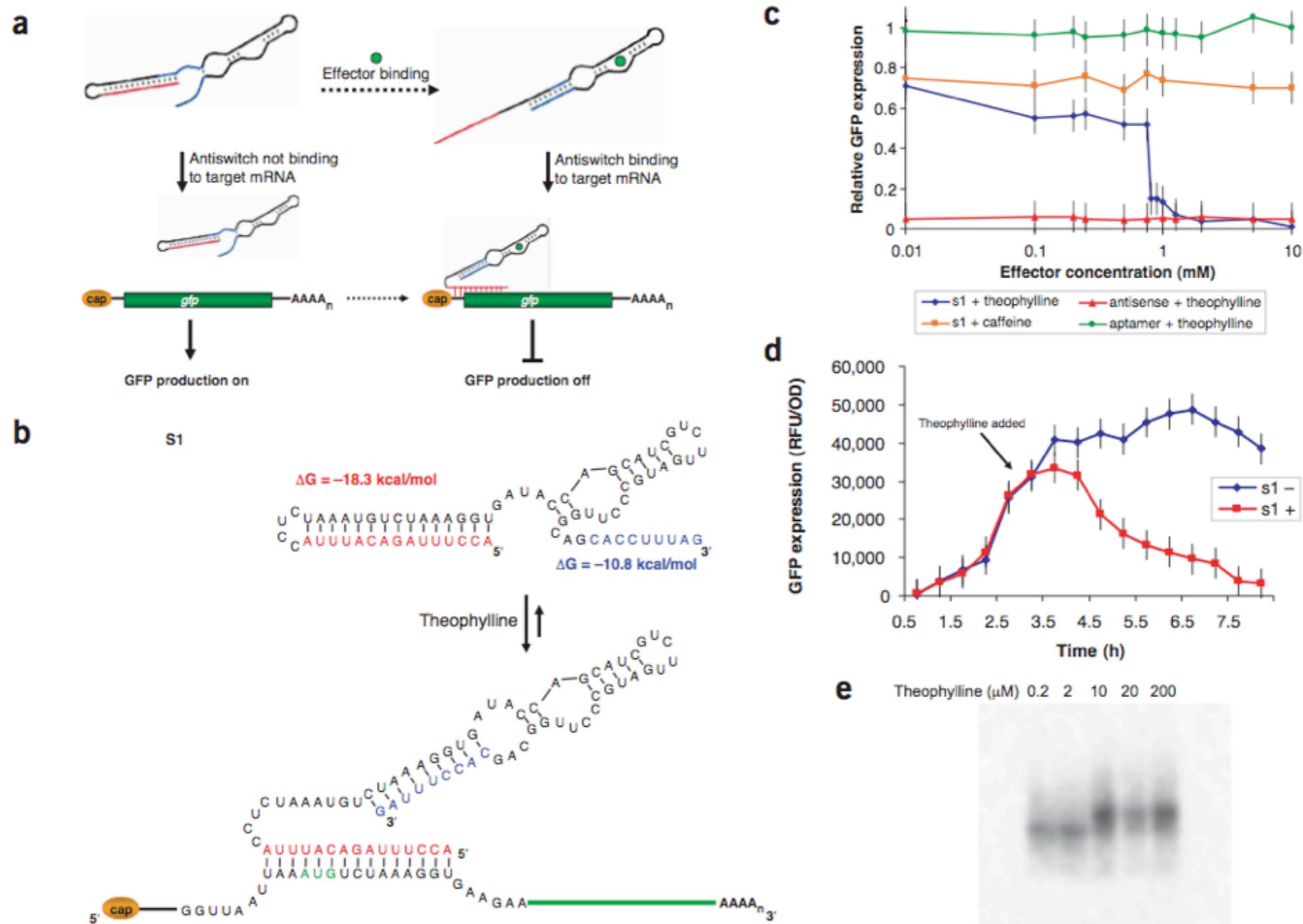
...of course, there is now a full spectrum to choose from...

# Bayer & Smolke, 2005



**Figure 4** Simultaneous regulation of multiple genes through multiple antiswitch regulators. **(a)** Illustration of the mechanism by which two independent antiswitch molecules act to regulate the expression of multiple target genes *in vivo*. In the absence of their respective effectors, the antiswitches are in the 'off' state and are unable to bind to their target transcripts. In this state, both GFP and YFP production is on. In the presence of theophylline, one antiswitch switches its conformation to the 'on' state and turns off GFP production. In the presence of tetracycline, the second antiswitch switches its conformation to the 'on' state and turns off YFP production. These antiswitches act independently of each other to provide combinatorial control over genetic circuits. **(b)** *In vivo* regulation activity of two antiswitch constructs (s1, s9) against their respective targets (GFP, YFP) in the presence or absence of their respective effector molecules (theo, theophylline; tet, tetracycline). Relative YFP expression (black); relative GFP expression (white).

(antiswitch itself is expressed within a self-cleaving ribozyme construct)



**Figure 1** Design and functional activity of an antiswitch regulator. **(a)** General illustration of the mechanism by which an antiswitch molecule acts to regulate gene expression *in vivo*. The antisense sequence is indicated in red; switching ‘aptamer stem’ is shown in blue. In the absence of effector, the antisense domain is bound in a double-stranded region of the RNA referred to as the ‘antisense stem’ and the antiswitch is in the ‘off’ state. In this state the antiswitch is unable to bind to its target transcript, which has a *gfp* coding region, and as a result, GFP production is on. In the presence of effector, the antiswitch binds the molecule, forcing the aptamer stem to form, switching its confirmation to the ‘on’ state. In this state the antisense domain of the antiswitch will bind to its target transcript and through an antisense mechanism turn the production of GFP off. **(b)** Sequence and predicted structural switching of a theophylline-responsive antiswitch, s1, and its target mRNA. On s1, the antisense sequence is indicated in red; switching aptamer stem sequence is indicated in blue, the stability of each switching stem is indicated. On the target mRNA, the start codon is indicated in green. **(c)** *In vivo* GFP regulation activity of s1 and controls across different effector concentrations: aptamer construct (negative control) in the presence of theophylline (green); antisense construct (positive control) in the presence of theophylline (red); s1 in the presence of caffeine (negative control, orange); s1 in the presence of theophylline (blue). Data are presented as relative, normalized GFP expression in cells harboring these constructs against expression levels from induced and uninduced cells harboring only the GFP expression construct. **(d)** *In vivo* temporal response of s1 inhibiting GFP expression upon addition of effector to cells that have accumulated steady-state levels of GFP and antiswitch s1. No theophylline, blue; 2 mM theophylline, red. **(e)** *In vitro* affinity assays of s1 to target and effector molecules. The mobility of radiolabeled s1 was monitored in the presence of equimolar concentrations of target transcript and varying concentrations of theophylline as indicated.

# RNA logic

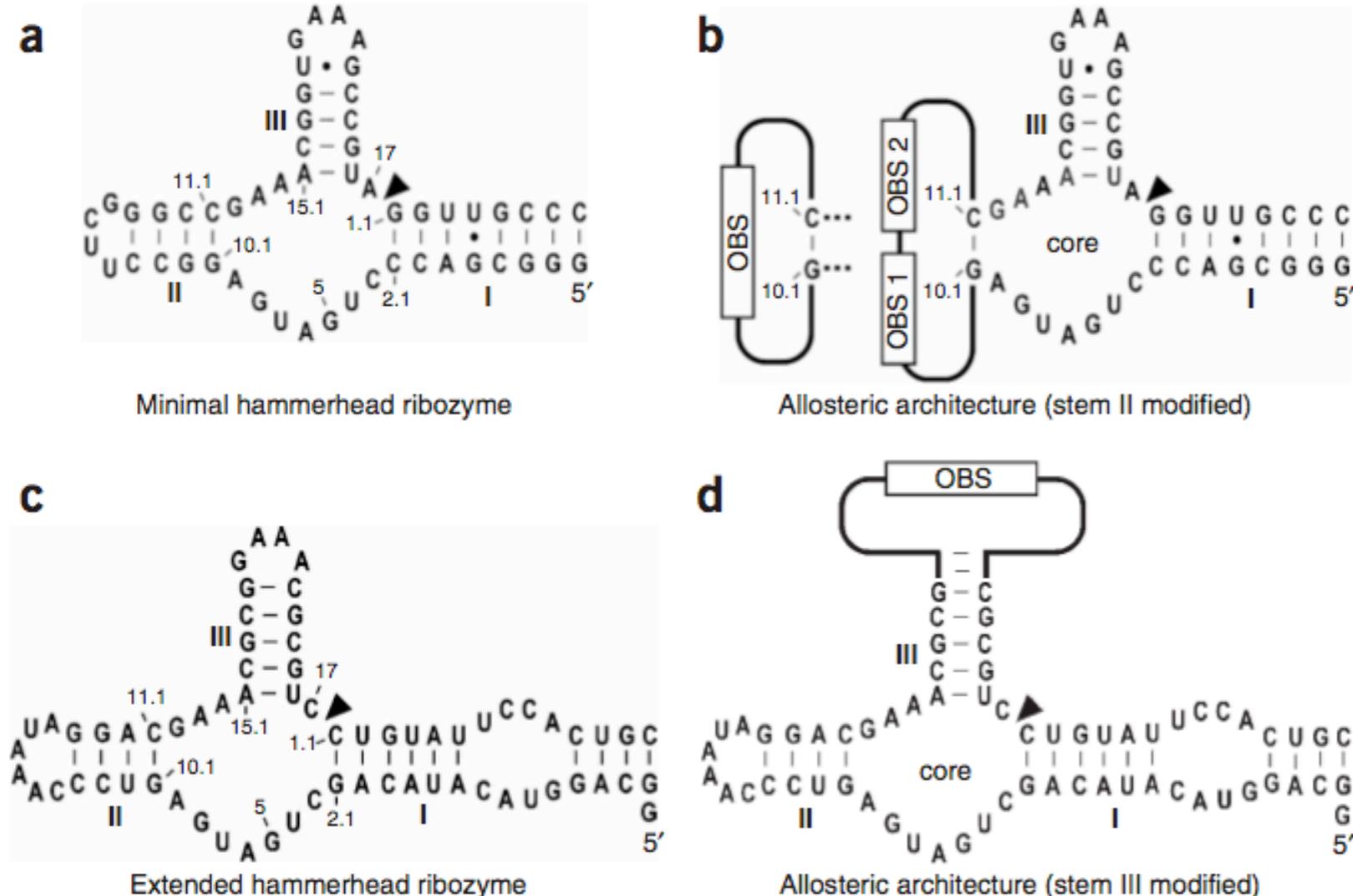
- One vision of synthetic biology is to be able to design genetic circuits as rationally as we can design electronic circuits
- Fundamental to digital electronics is the concept of the **logic gate**
- Reproducing such components in genetic systems has been a recurrent theme in synthetic biology research

Name	Graphic Symbol	Algebraic Function	Truth Table
AND		$F = A + B$ or $F = AB$	A B   F 0 0   0 0 1   0 1 0   0 1 1   1
OR		$F = A + B$	A B   F 0 0   0 0 1   1 1 0   1 1 1   1
NOT		$F = \bar{A}$ or $F = A'$	A   F 0   1 1   0
NAND		$F = (\overline{AB})$	A B   F 0 0   1 0 1   1 1 0   1 1 1   0
NOR		$F = (\overline{A + B})$	A B   F 0 0   1 0 1   0 1 0   0 1 1   0

# Penchovsky & Breaker, 2005

- One attempt at the “RNA logic gate” problem
- Uses hammerhead ribozyme with allosteric shifts triggered by complementary RNA
- “Firing” of the gate involves RNA cleavage, so it must be continuously “primed” with fresh RNA (bug or feature?)
  - Also requires continuous energy input
  - Only quite simple circuits tested (so far)
  - Other solutions do exist
    - e.g. Seelig, Winfree et al, Science, 2006

# Penchovsky & Breaker, 2005

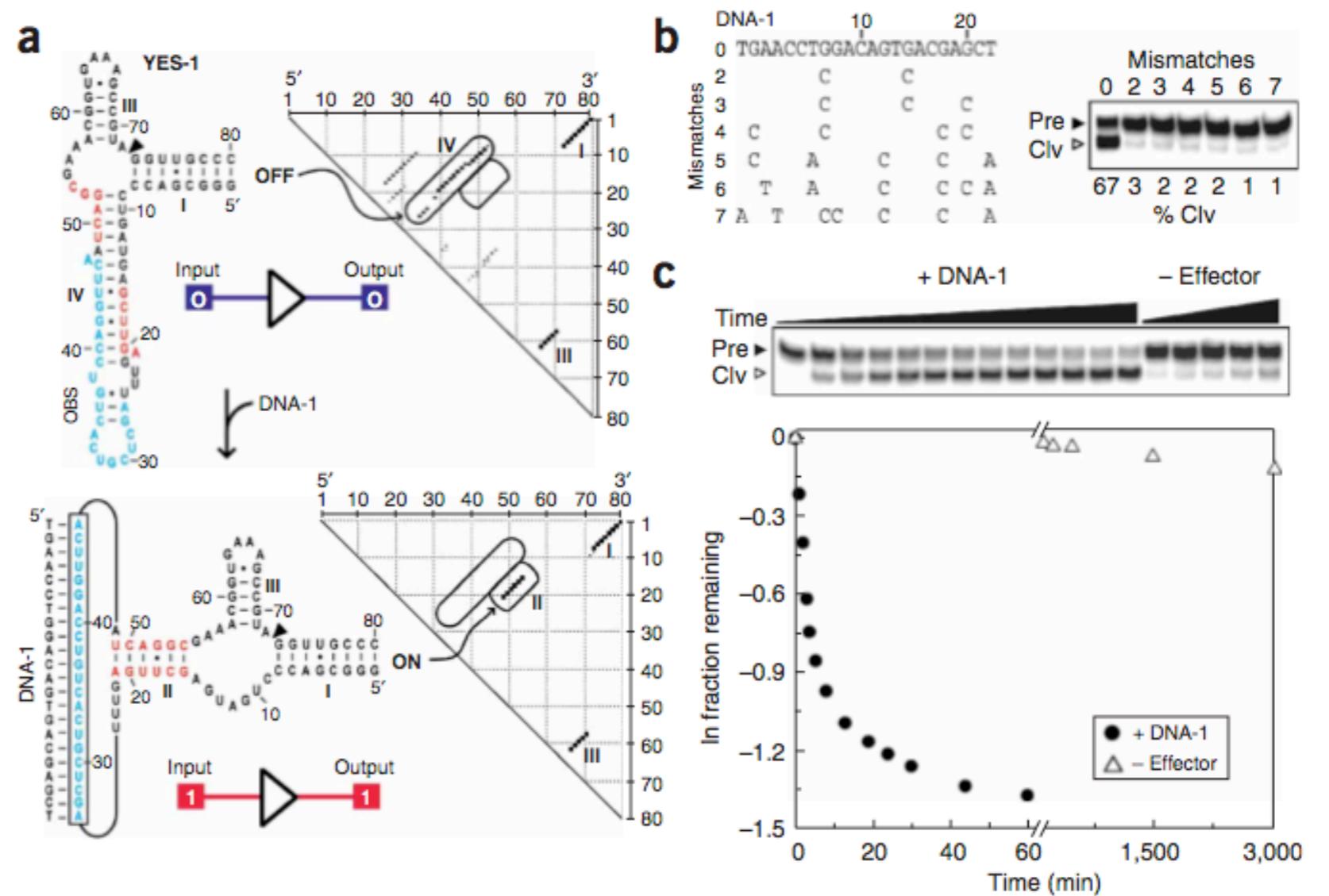


**Figure 1** Design architectures for the construction of oligonucleotide-responsive hammerhead ribozymes. (a) Parental hammerhead ribozyme sequence used previously<sup>33–36</sup> and in the current study to create ligand-responsive ribozyme switches. Numbering systems for the hammerhead is as described elsewhere<sup>52</sup>, where stems I through III represent base-paired structures that are essential for ribozyme function. The arrowhead identifies the site of ribozyme self-cleavage. (b) Integration of one (left) or two (right) oligonucleotide binding sites (OBS) into stem II of the parent hammerhead depicted in a. (c) Extended hammerhead ribozyme that exhibits faster RNA cleavage rates with low Mg<sup>2+</sup> concentrations<sup>53</sup>. (d) Integration of an OBS in stem I of the extended ribozyme to create RNA switches with NOT function.

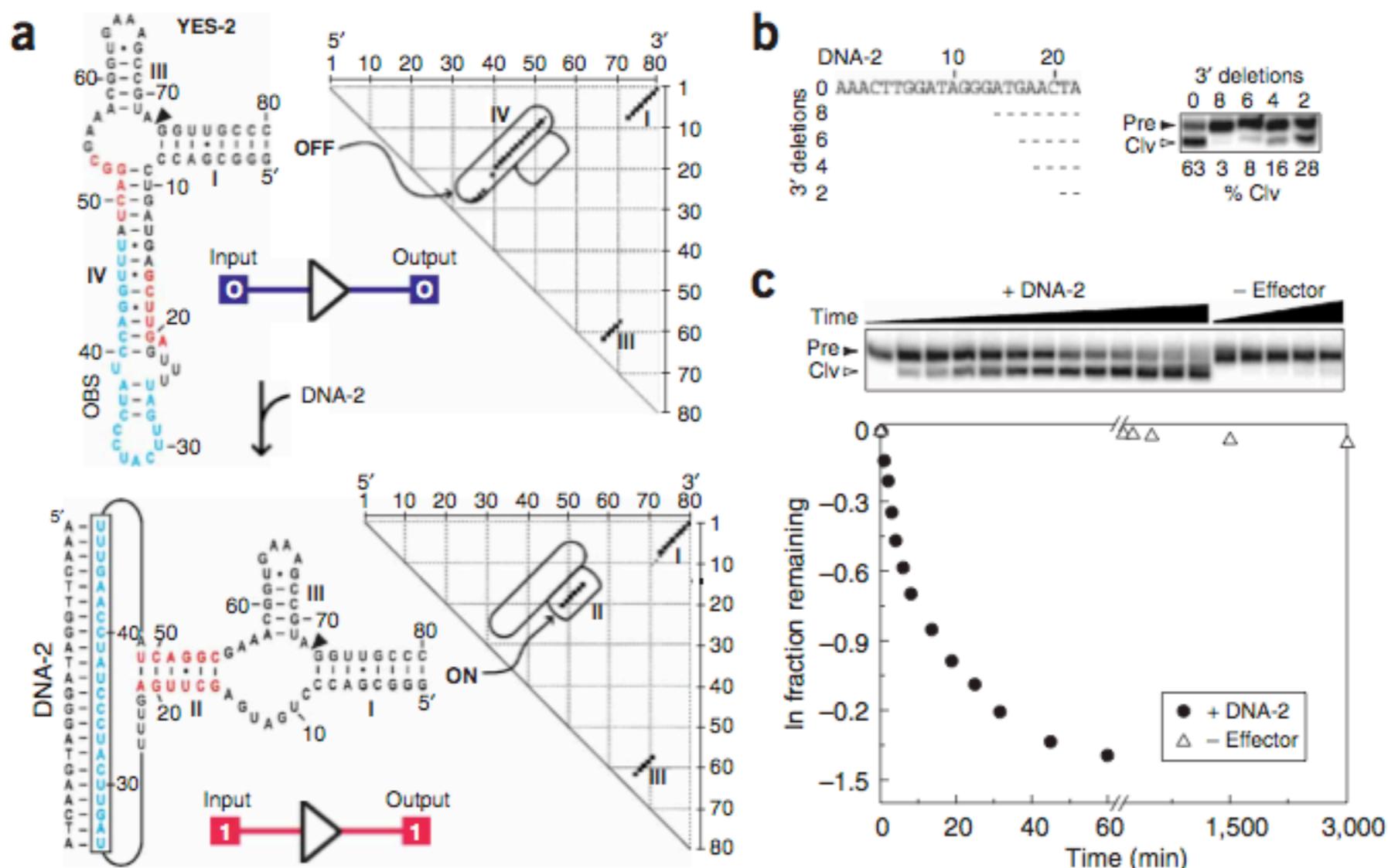
OBS =  
oligonucleotide  
binding site

# Penchovsky & Breaker, 2005

**Figure 2** Design and characterization of an oligonucleotide-specific RNA switch possessing YES logic function. (a) Secondary structure models for the most stable conformers as computed using the partition function algorithm in the absence (OFF) or presence (ON) of a 22-nucleotide DNA effector. The effector-binding site (light blue) is joined to nucleotides 10.1 and 11.1 of the hammerhead core via eight- and six-nucleotide linkers. In the ON state, most of these linker nucleotides are predicted to form an extended stem II structure (red). To the right of each model is a dot matrix plot wherein larger points reflect greater probability of base pairing. Encircled points reflect the main differences in predicted structures between the OFF (stem IV) and ON (stem II) states. Nucleotides 1 through 79 are numbered from 5' to 3' across the top and right of the plots. Schematic representations of the logic states of the constructs are shown in this and subsequent figures. (b) Selective activation of ribozyme self-cleavage by an effector DNA complementary to the OBS. Radiolabeled ribozymes (5'  $^{32}\text{P}$ , Pre) undergo self-cleavage only with the perfectly matched DNA effector (0 mismatches) and the resulting radiolabeled cleavage fragment (Clv) is separated from the precursor by denaturing 10% PAGE. Products were visualized and cleavage yields were quantified by PhosphorImager. (c) Kinetics of ribozyme (1  $\mu\text{M}$ ) self-cleavage in the presence (+) of perfectly matched 22-nt effector DNA (3  $\mu\text{M}$ ) and in the absence (-) of effector DNA. Gel image is as described in b. Plot using data derived from the gel depicts the natural logarithm of the fraction of RNA remaining uncleaved versus time.

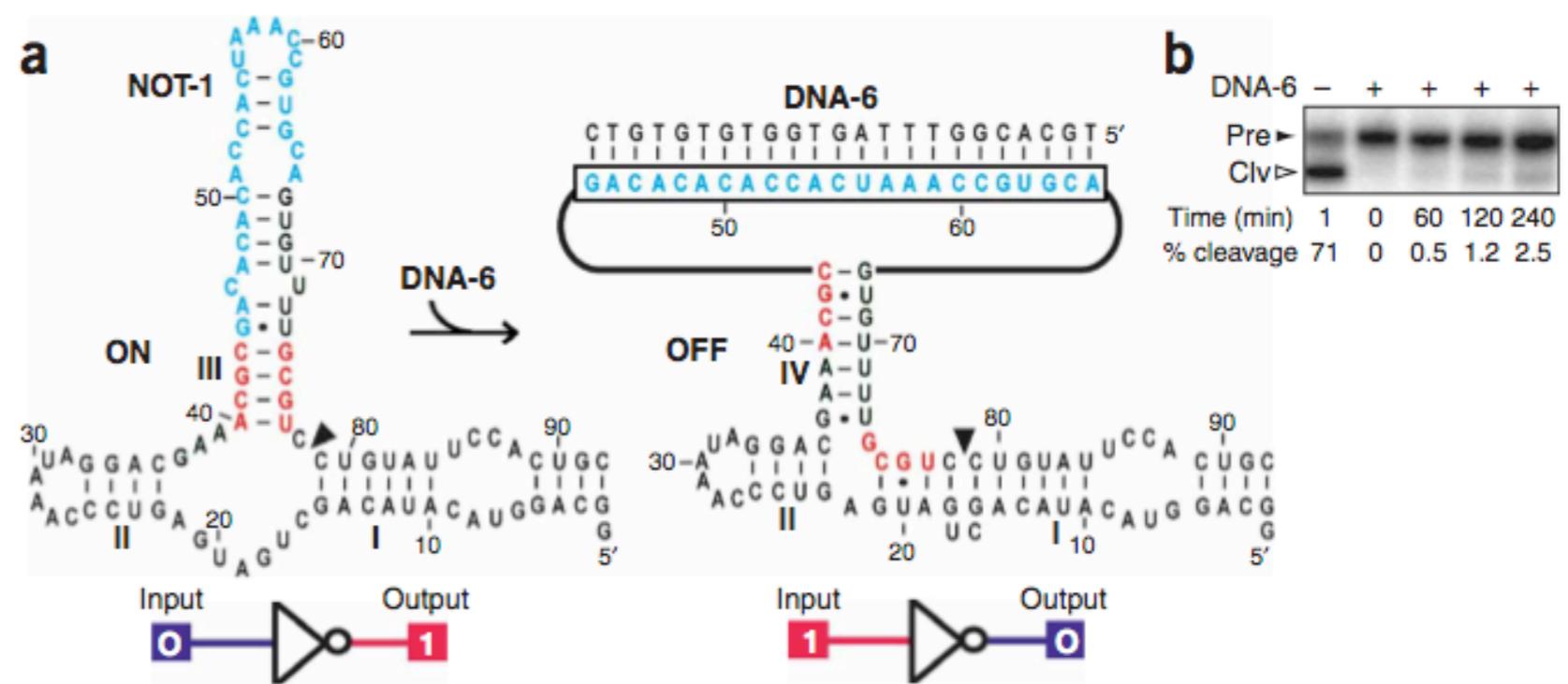


# Penchovsky & Breaker, 2005



**Figure 3** Design and characterization of YES-2, a variant of YES-1 that exhibits altered oligonucleotide specificity. **(a)** Secondary structure models for the most stable conformers in the absence (OFF) or presence (ON) of a 22-nucleotide DNA effector complementary to the changed OBS. **(b)** Selective activation of YES-2 self-cleavage by effector DNAs complementary to the OBS. **(c)** Kinetics of YES-2 self-cleavage in the presence of perfectly matched 22-nt effector DNA and in the absence of effector DNA. For additional details, see the legend to **Figure 2**.

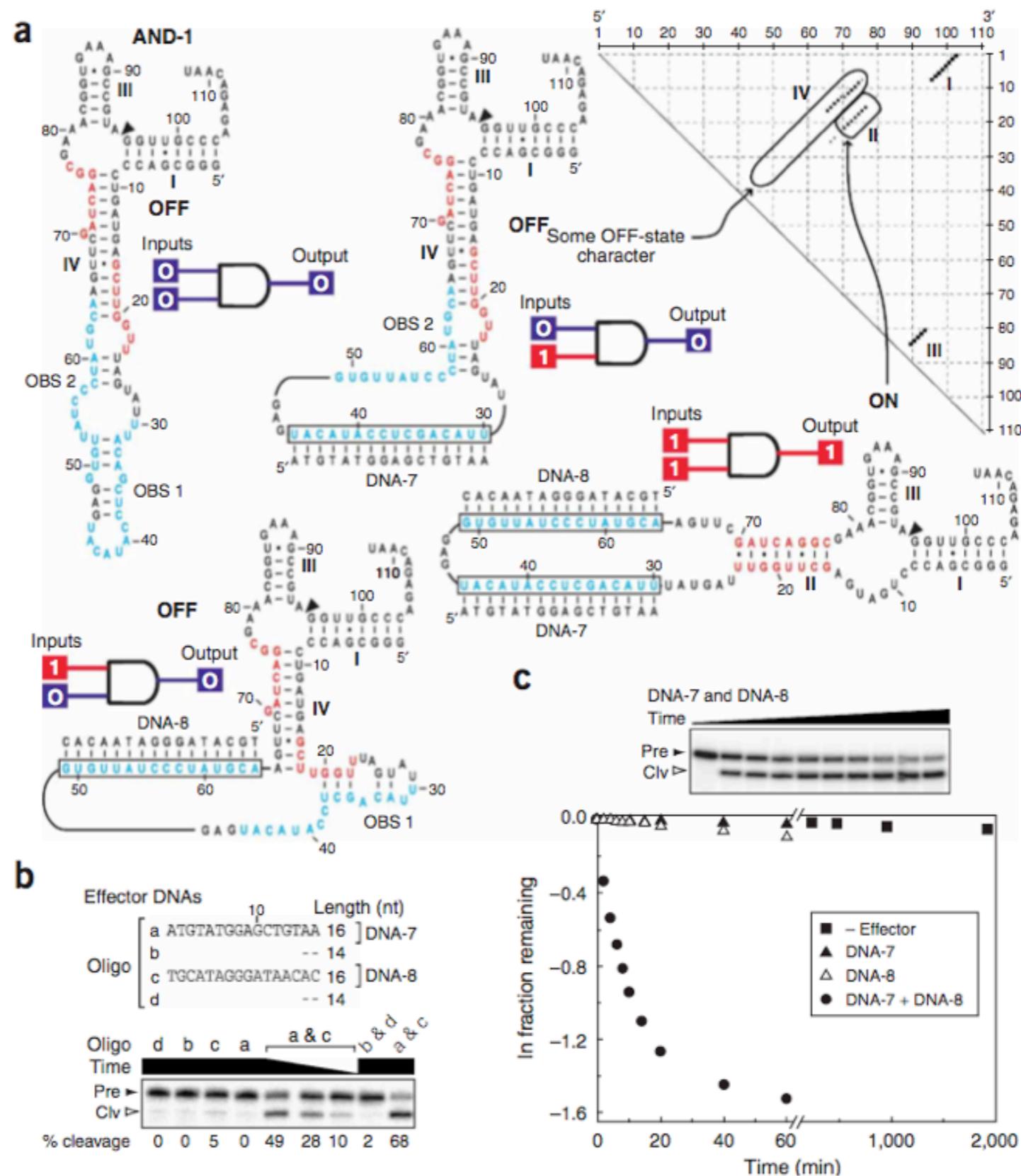
# Penchovsky & Breaker, 2005



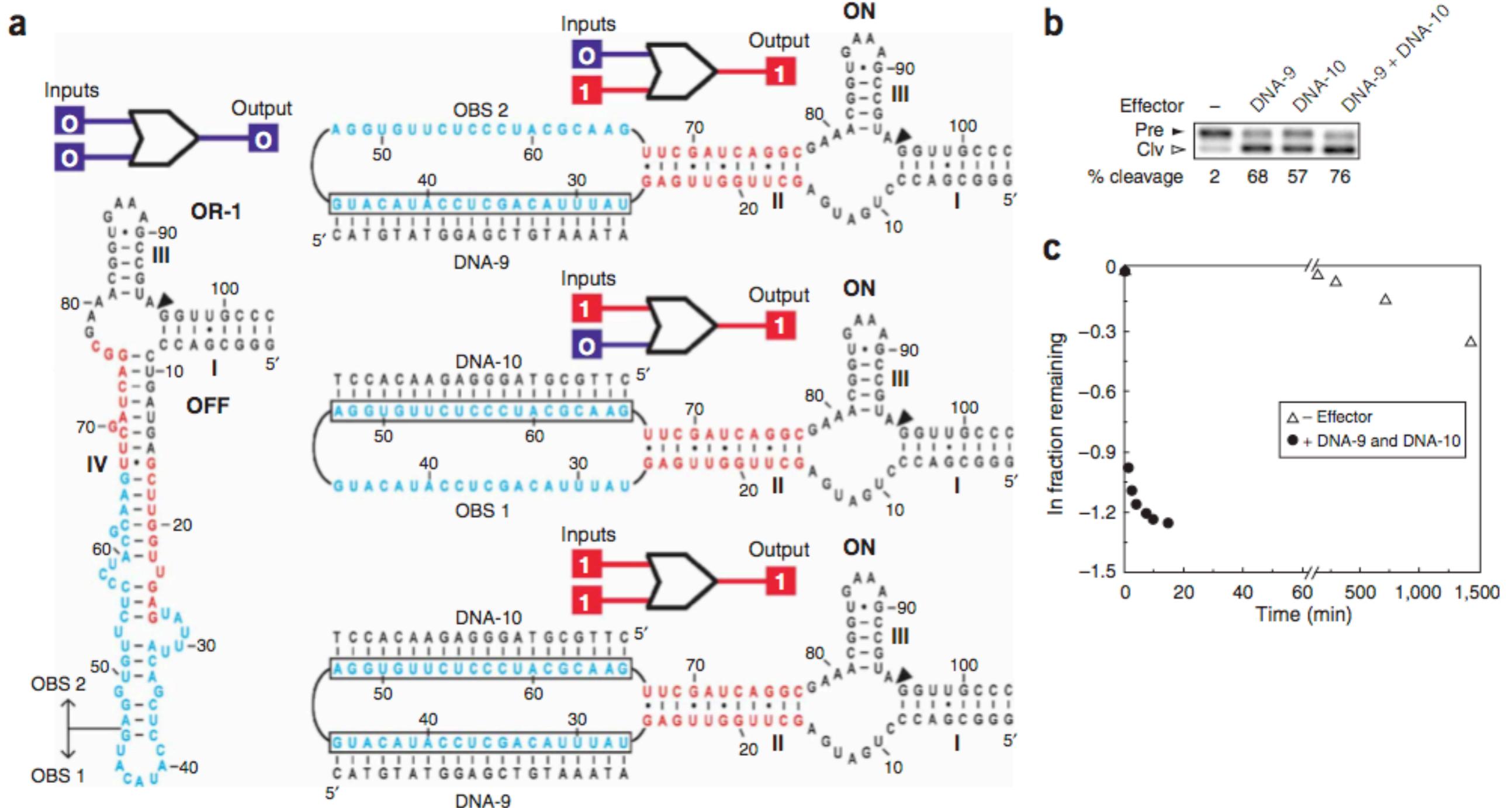
**Figure 4** Design and characterization of NOT-1 based on an extended hammerhead ribozyme. (a) Secondary structure models for the most stable conformers predicted in the absence (ON) and presence (OFF) of a 23-nucleotide effector. Dot matrix plots for the construct are presented in **Supplementary Figure 5** online. (b) Deactivation of NOT-1 by a DNA complementary to the OBS.

# Penchovsky & Breaker, 2005

**Figure 5** Design and characterization of AND-1, an oligonucleotide-specific molecular switch that possesses AND logic function. (a) AND-1 is designed to form the active hammerhead structure and self-cleave only when presented simultaneously with its two corresponding effector DNAs (DNA-7 and DNA-8). The dot matrix plots for the ON state showing some character of the OFF states (stem IV) is depicted. Dot matrix plots for the three OFF states are presented in **Supplementary Figure 7** online. (b) Activation of AND-1 self-cleavage requires both full-length DNA-7 and DNA-8 effectors. Maximum incubation time is 60 min. (c) Kinetics of AND-1 self-cleavage under various combinations of effector DNAs. Details are as described in the legend to **Figure 2**.

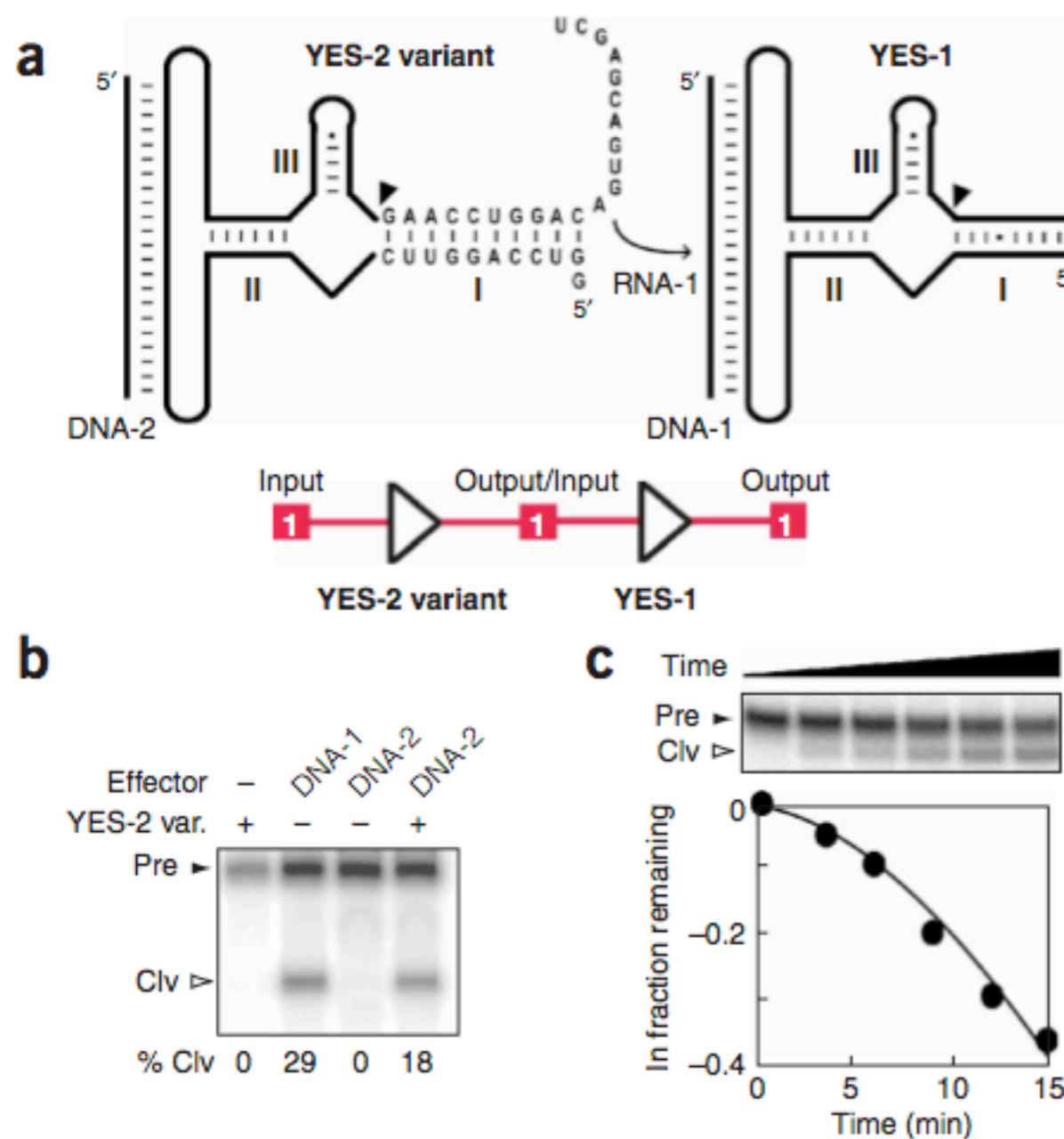


# Penchovskiy & Breaker, 2005



**Figure 6** Design and characterization of OR-1, an oligonucleotide-specific molecular switch that possesses OR logic function. (a) OR-1 is designed to trigger self-cleavage when either effector (DNA-9 or DNA-10) or both effectors are present. Dot matrix plots for each of the four structures are presented in **Supplementary Figure 7** online. (b) Activation of OR-1 self-cleavage occurs when either or both effectors are present when incubated for 5 min under standard assay conditions. (c) Kinetics of OR-1 self-cleavage in the absence of effector and in the presence of both effector DNAs. Details are as described in the legend to **Figure 2**.

# Penchovsky & Breaker, 2005



**Figure 7** A two-step ribozyme signaling pathway constructed using YES-1 and a variant of YES-2. (a) Nucleotides of the YES-2 variant RNA that differ from YES-2 are depicted. Upon activation of YES-2 variant by effector DNA-2, the 3' cleavage fragment (RNA-1) is released and serves as an effector for YES-1 activation. (b) Assay depicting function of a ribozyme-signaling pathway. YES-1 RNAs are radiolabeled in all lanes. Ribozymes and effector DNAs are present as defined at concentrations of 1 and 3  $\mu$ M, respectively, and reactions were incubated at 23 °C for 5 min. Other details are as described for **Figure 2**. (c) Kinetic analysis of YES-1 self-cleavage in the presence of YES-2 variant and its effector DNA-2. Details are as described in b.

# Design of YES gate

## I.1 Generate random 16-22nt OBS

## 1.2 Insert OBS sequence into hammerhead template

### ► I.3 Compute partition function at 37°C in OFF state

#### 1.4 Check that catalytic core is base-paired in OFF state

### 1.5 Flag OBS segment as being unable to form basepairs (ON state)

## 1.6 Recompute partition function at 37°C in ON state

I.7 Check that all three stems in catalytic structure are present in ON state

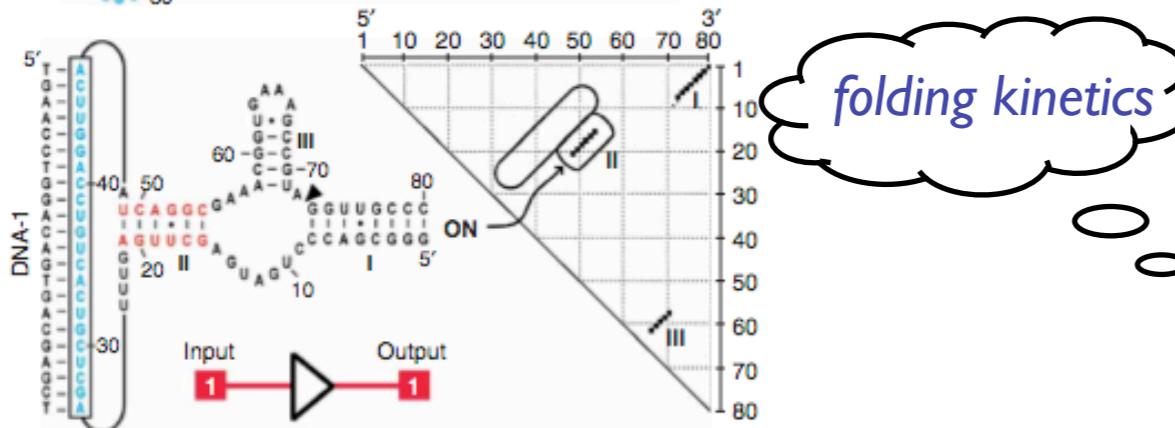
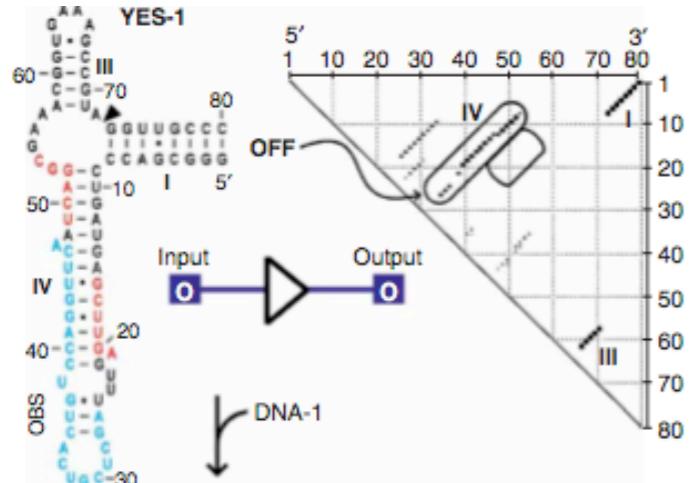
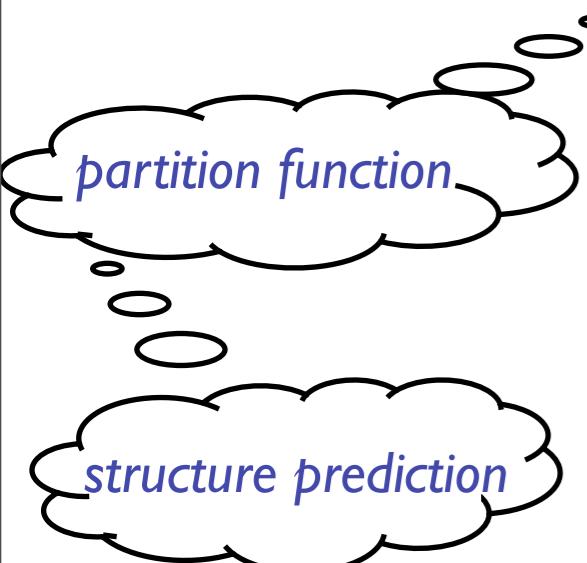
#### 1.8 Check that 30-70% of OBS is base-paired in OFF state

1.9 Check that energy gap between OFF and ON is 6-10 kcal/mol

I.10 Check that structures for both states are stable from 20-40°C

I.11 Check that structure ensemble diversity for both states is <9 um

0 50 60 70 80 3'



## 2.1 Re-randomize the OBS using *inverse folding* from the OFF state

## 2.2 Repeat step 1.4 for new OBS

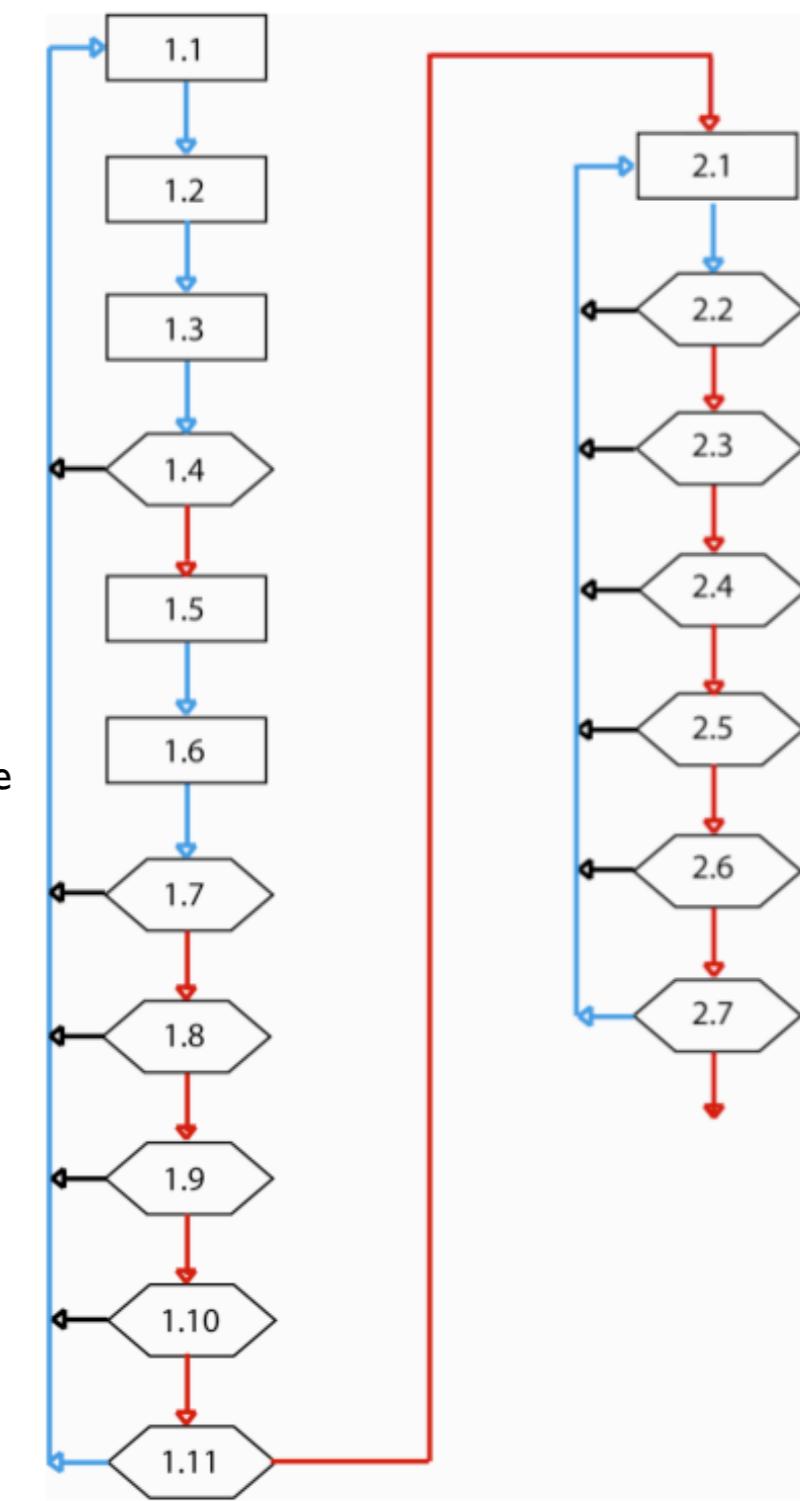
### 2.3 Compare stability of OFF state to candidate from stage I

#### 2.4 Compare stability of OBS-cognate duplex to candidate from stage I

## 2.5 Repeat step 1.10 for new OBS

2.6 Simulate RNA folding kinetics from OFF to ON. Reject if time>480

2.7 Compute OFF-ON energy gap; reject if  $>2^*$  gap from stage 1.9

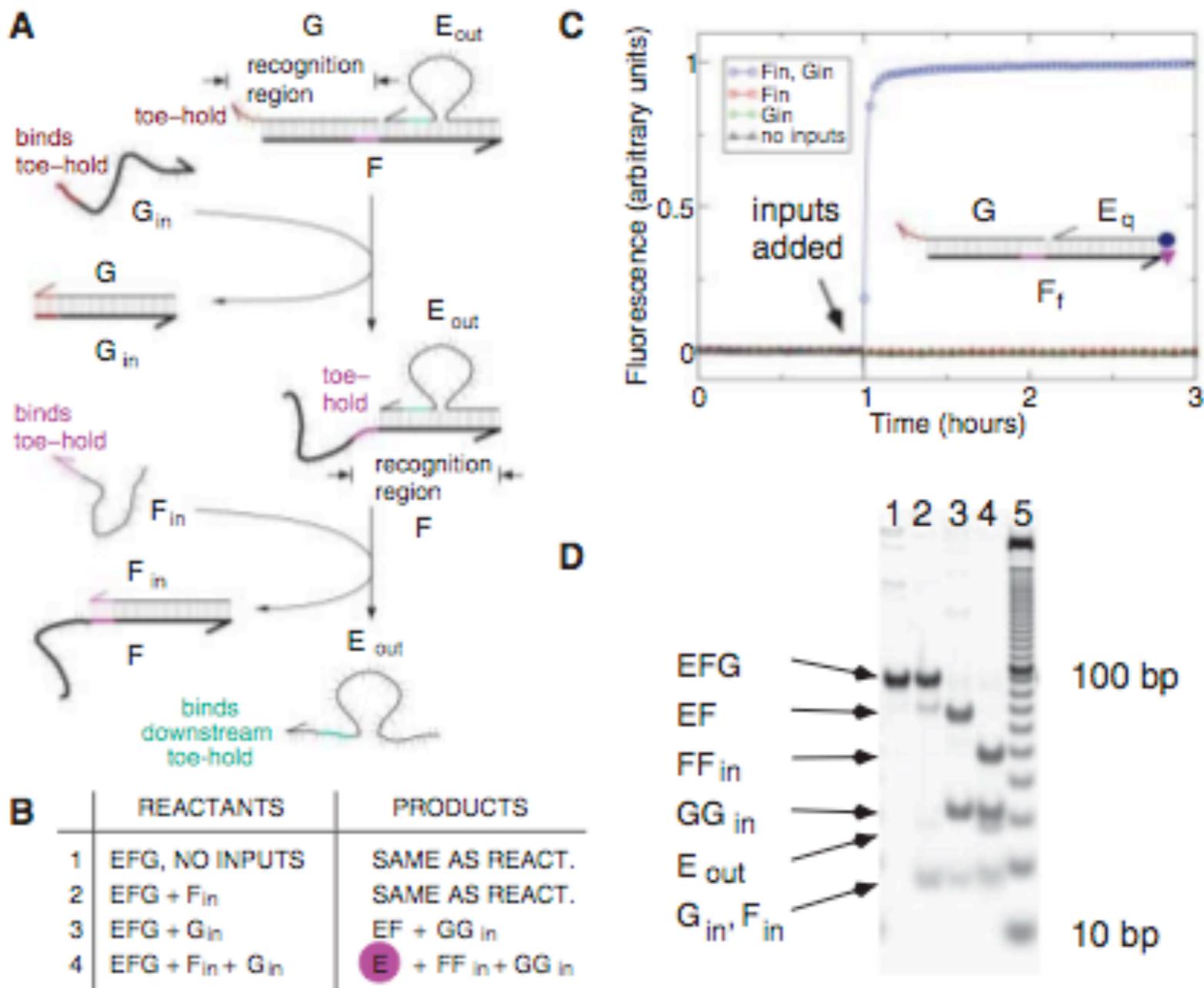


# Seelig *et al*, 2006

- Single-stranded nucleic acids as inputs and outputs
- Mechanism relies entirely on sequence recognition and strand displacement
- AND, OR, NOT gates
- Signal restoration, cascading, fan-out, modularity

# AND

# Seelig et al, 2006



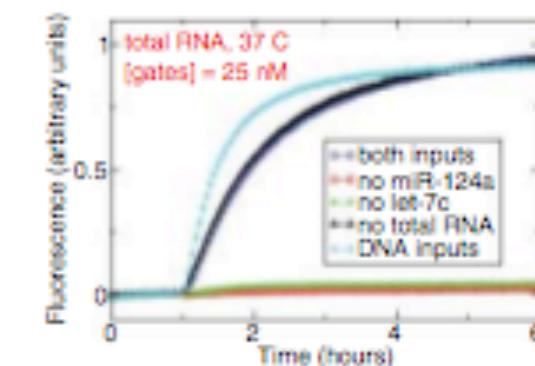
**Fig. 1.** Two-input AND gate. **(A)** The gate consists of three DNA strands,  $E_{out}$  [57 nucleotides (nt)],  $F$  (60 nt), and  $G$  (36 nt). The 3' ends are marked by arrows. Toeholds and toehold binding regions (all six nucleotides) are indicated in color. Input strands  $F_{in}$  and  $G_{in}$  (36 nt) are complementary to recognition regions within the corresponding gate strands  $F$  and  $G$ . **(B)** Truth table for the two-input AND gate. The released output strand is highlighted. **(C)** In fluorescence experiments, strands  $F_t$  [carboxytetramethylrhodamine (TAMRA) fluorophore at the 3' end] and  $E_q$  (Iowa Black RQ quencher at the 5' end, without bulge loop) were used instead of  $F$  and  $E_{out}$  (see inset). Release of output strand results in increased fluorescence. Experiments were conducted at 25°C with gate concentrations of 250 nM and input concentrations of 300 nM in a Tris-acetate-EDTA buffer containing 12.5 mM Mg<sup>++</sup>. **(D)** Nondenaturing gel electrophoresis directly confirms reaction intermediates and waste products for each possible input combination. Lanes 1 to 4: The samples are as described in entries 1 to 4 of the truth table. The gate used in this experiment is as shown in (A). Lane 5: 10-base pair (bp) ladder.

# “Translator”

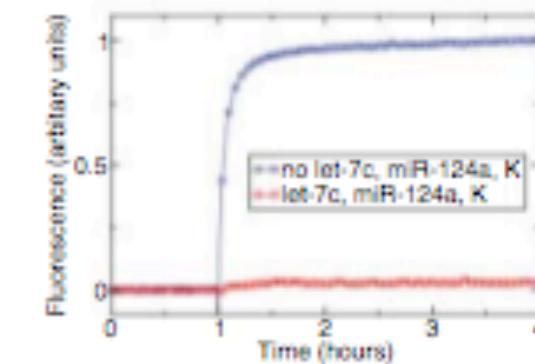
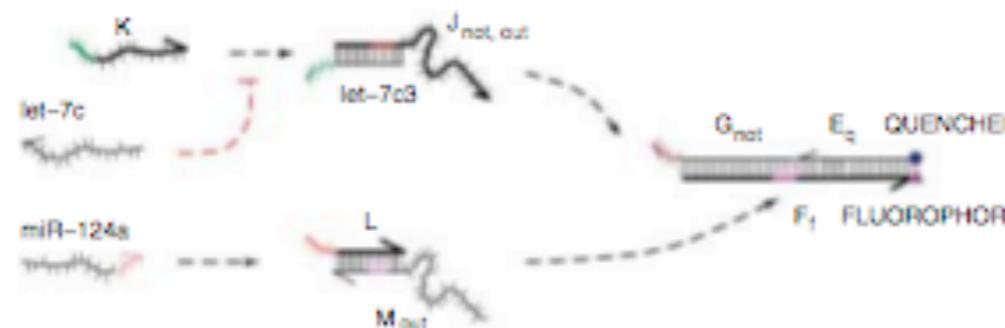
NOT, AND  
Seelig et al,  
2006

(3-input AND where inputs 1&3 are identical; input 2 (Th2) is always present & required only for structural purposes)

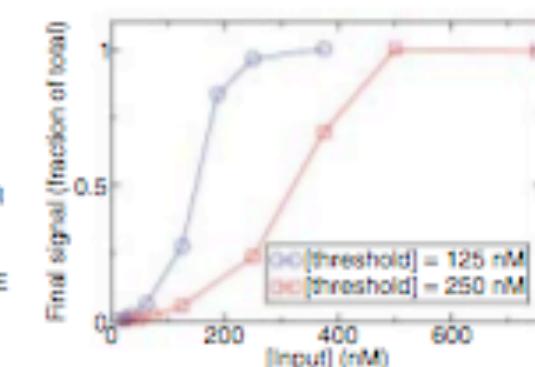
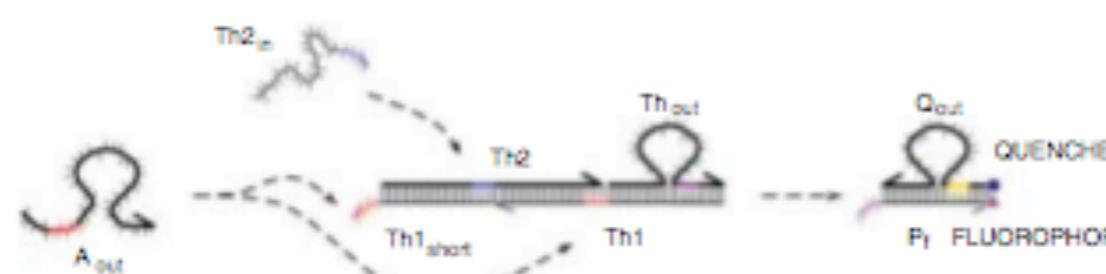
## A CIRCUIT DIAGRAM FOR: let-7c AND miR-124a



## B CIRCUIT DIAGRAM FOR: (NOT let-7c) AND miR-124a



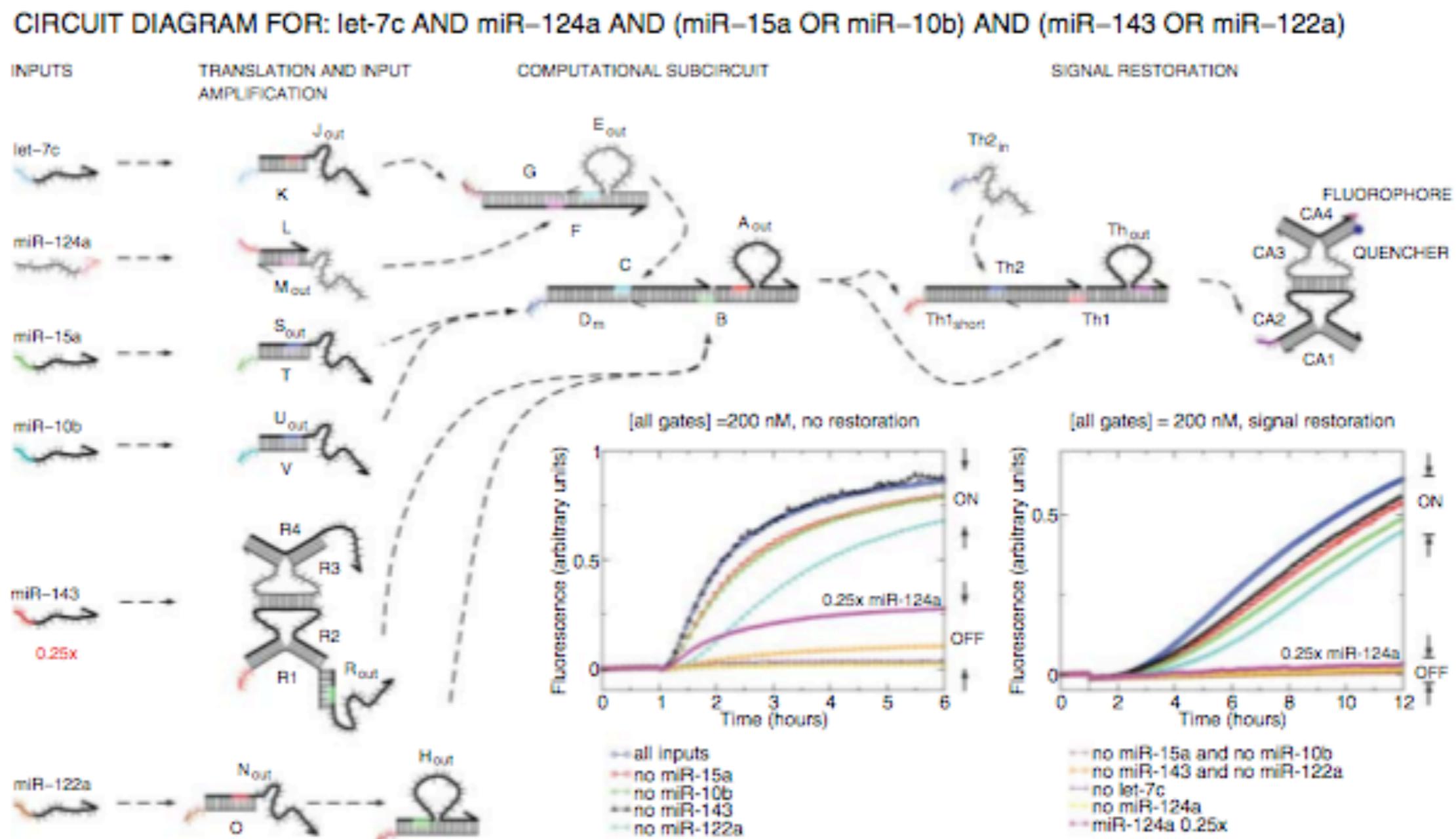
## C CIRCUIT DIAGRAM FOR THRESHOLDING



**Fig. 2.** Translator gates, NOT operation, and signal restoration. Dashed arrows indicate where input or output strands can serve as inputs to downstream gates. (A) Circuit operation at 37°C with RNA inputs and DNA gates in a total RNA background. All gates are at 25 nM, synthetic RNA inputs are at 30 nM, and total RNA (mouse brain) is at a concentration of 200 µg/ml. Proper function is observed. For comparison, experiments with no total RNA were performed, using either both RNA inputs or both DNA inputs. (B) The NOT gate consists of a translator gate and an inverter strand complementary to let-7c. Gate, inverter strand, and input concentrations are 250 nM, 300 nM, and 300 nM, respectively. Here and in all subsequent experiments, the temperature was 25°C and DNA equivalents of the biological microRNAs were used. If let-7c was present, inverter strand K preferentially hybridized to let-7c. Otherwise, inverter strand K triggered the translator. (C) The thresholding gate, using a dye/quencher-labeled readout gate to monitor the output. Strand Th2<sub>in</sub> is part of the thresholding unit and was added before the start of the experiment. The final fluorescence is plotted against the input concentration for two different concentrations of the threshold gate.

# Seelig et al, 2006

**Fig. 3.** Signal propagation through a complex chemical circuit combining AND, OR, sequence translation, input amplification, and signal restoration. The five-layer circuit consists of a total of 11 gates and accepts six inputs. With the exception of the threshold gate, which was at 100 nM with its  $\text{Th2}_{\text{in}}$  strand at 150 nM, all gates were at 200 nM (1x) per gate. Unless otherwise specified, inputs were added at 250 nM (1.25x). miR-143 was added at 50 nM (0.25x) and subsequently amplified by the input amplifier. (Inset) Fluorescence traces of circuit operation without and with the signal restoration module (threshold plus amplifier). The traces for input conditions corresponding to a logical TRUE output (ON) are clearly distinguishable from the logical FALSE output (OFF). Cases tested include when all inputs are present, all cases in which exactly one input is missing, and combinations of inputs that turn off an



OR clause. Assuming monotonicity, withholding additional inputs will never lead to a logical TRUE output. To determine the response of the circuit to a leaky OFF signal, input miR-124a was added at 50 nM (0.25x) while all other inputs were added normally.

## Multi-gate circuit (3\*AND, 2\*OR, amplifier)

# Further applications of RNA design

- Ribozyme alteration of DNA sequence
- Antisense oligonucleotides; RNAi; siRNAs
- mRNA cleavage by (deoxy)ribozymes
- Ribozyme repair of mRNA
- Nucleoside analogue drugs (HIV rev inhibitors)
- Virus engineering...

# Virus design

- Applications of virus design:
  - Transformation of cells
  - Lysing of cells (e.g. cancer)
  - Delivery of genetic material

# Ideal properties of therapeutic virus

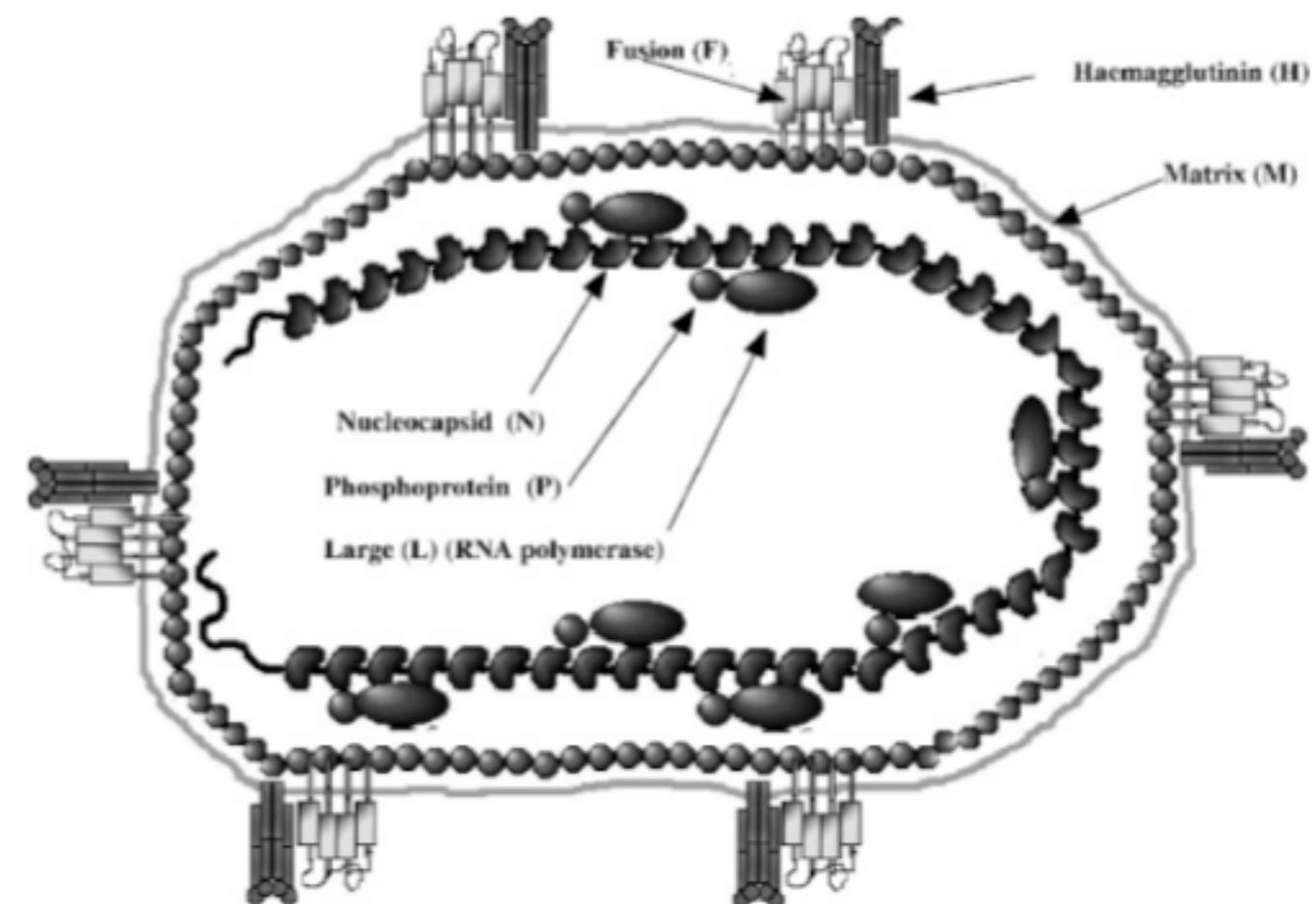
- Accurately **targeted** to specific cells/tissues
  - Untargeted cells/tissues should be unharmed
- Safety issues:
  - Associated with very mild, self-limited, or no human **disease**
  - Availability of **effective antiviral treatment** if necessary
  - **“Fail-safes”**; e.g. reduced competence, requirement for externally administered competence factor
- Genetic **stability** (no environmental recombination)
- Robust **manufacturing system**
- Ease of **transformation** to incorporate payload



# Measles

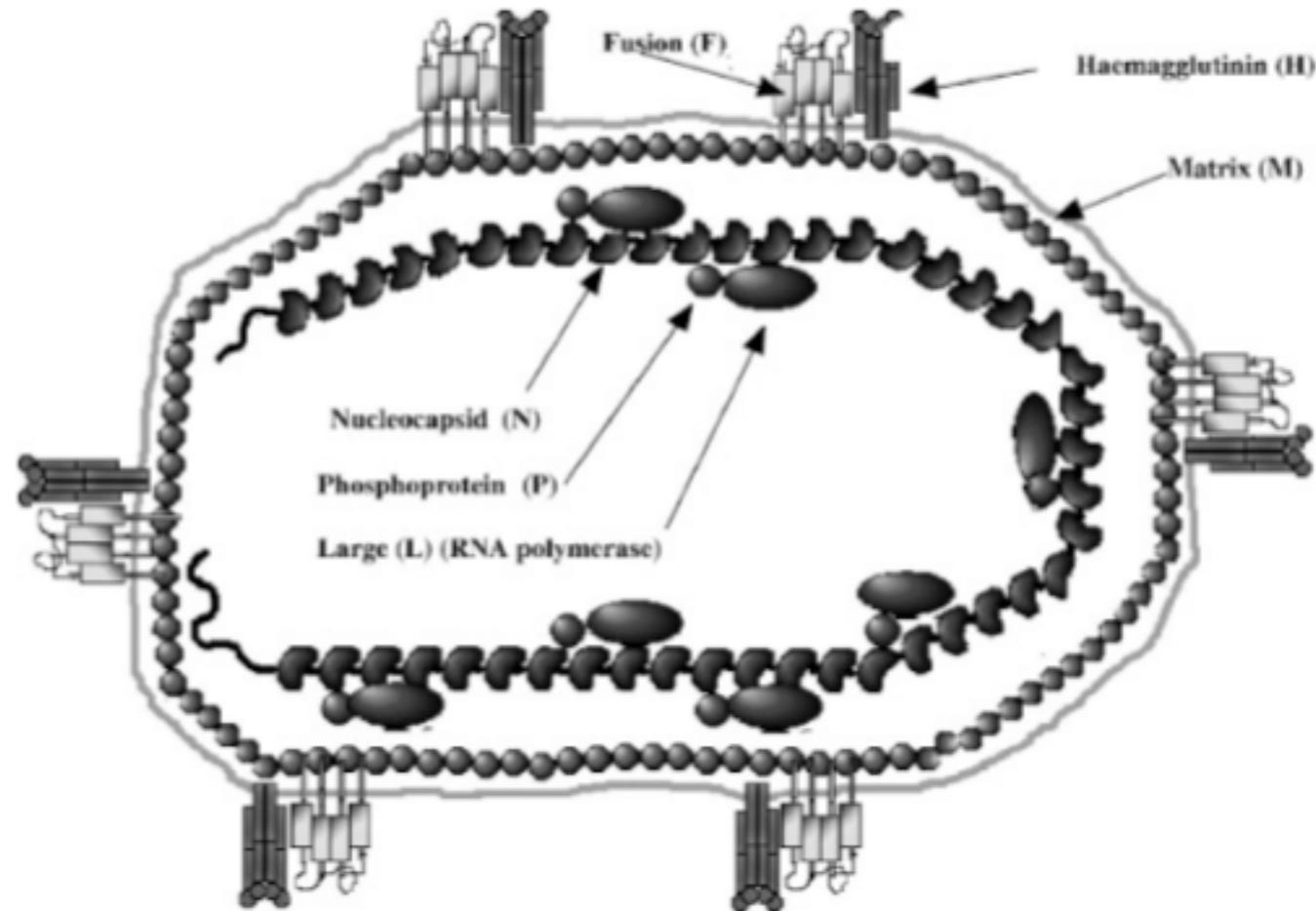
(Fielding, Rev.Med.Virol. 2005)

1	NP_056918.1	nucleocapsid protein	525.aa	<a href="#">1T6O_A</a>
2	NP_056919.1	phosphoprotein	507.aa	<a href="#">1OKS_A</a>
3	NP_056920.1	C protein	186.aa	
4	NP_056921.1	matrix protein	335.aa	
5	NP_056922.1	fusion protein	550.aa	
6	NP_056923.1	haemagglutinin protein H	617.aa	
7	NP_056924.1	L protein	2183.aa	



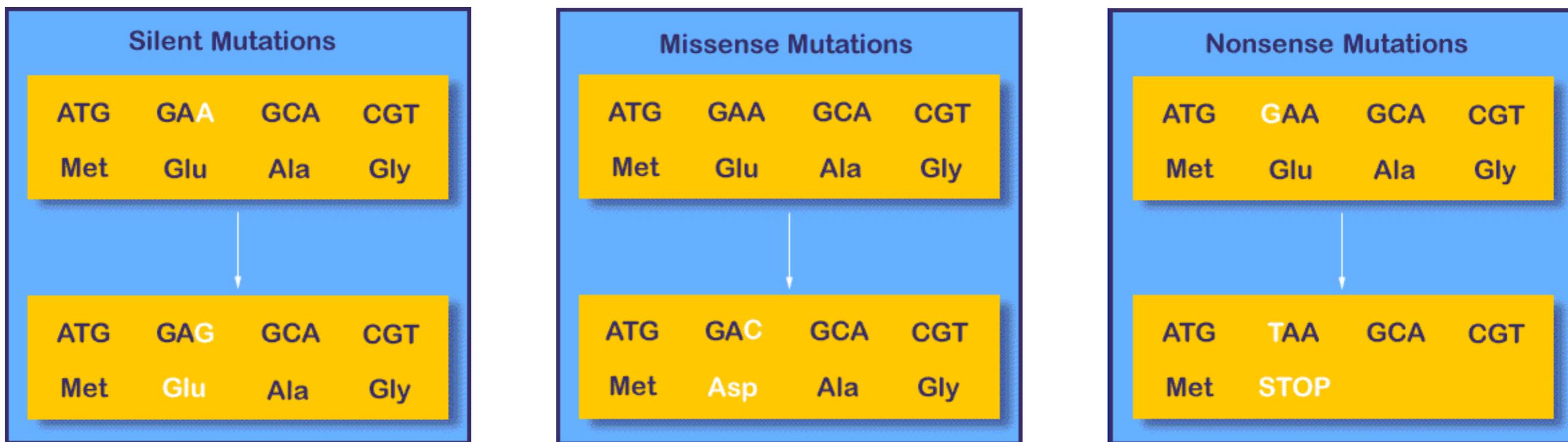
# Oncolytic viruses

# Measles



- Safety issues
  - Known to be genetically stable
  - No human or animal reservoirs
  - Low rate of “nonsynonymous” codon mutations suggests diversity is not driven by immune selection

# What's a “nonsynonymous substitution”?



Synonymous

Non-synonymous

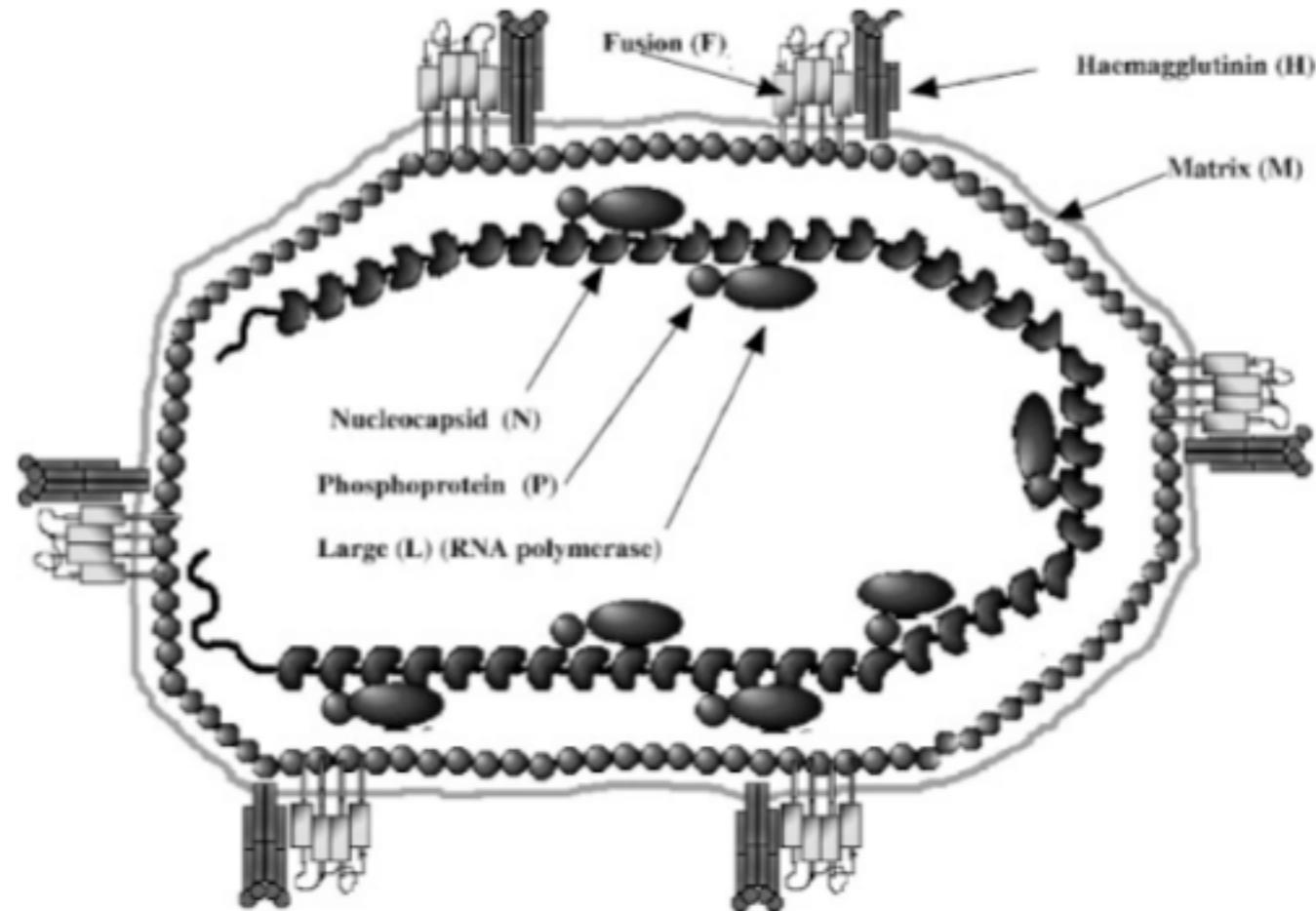
Nonsense

Ratio of nonsyn/syn is written  $K_a/K_s$  or  $dN/dS$

$K_a/K_s > 1$	“Diversifying selection” (e.g. immune evasion)
$K_a/K_s \approx 1$	“Neutral selection”
$K_a/K_s < 1$	“Purifying selection” (e.g. “housekeeping” genes)

...can be measured using **molecular evolution software** (e.g. PAML)

# Measles



- Engineering pros/cons
  - No known size constraint on genome that can be encapsidated
  - Some progress in targeting entry to host cells by C-terminal ligand extensions to H protein (e.g. growth factors)
  - Immune-mediated elimination is a problem (most adults are immune to measles)

# Wang & White, 2007

RNA regulatory element  
(including stem-loop “SL5”)

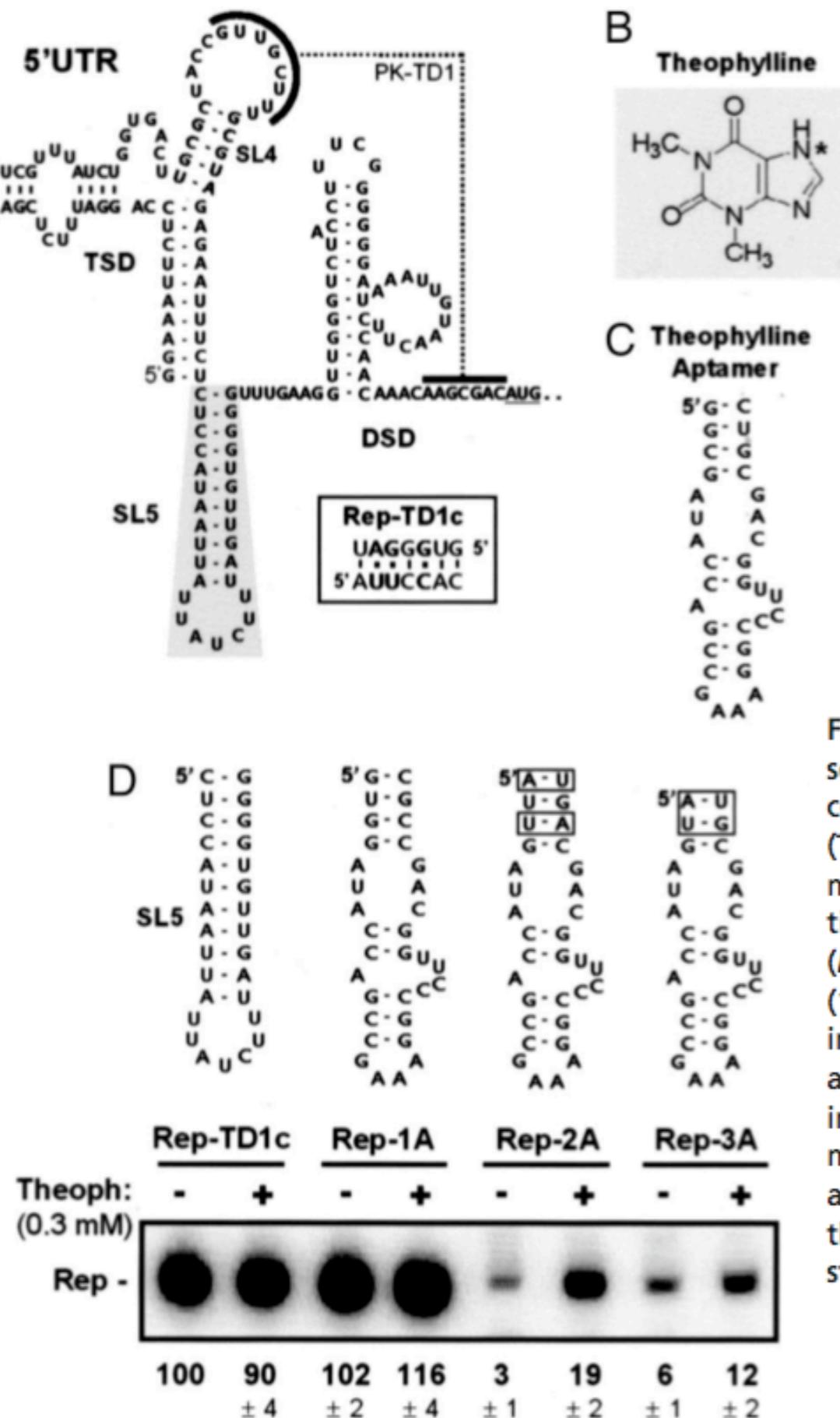
- Example of a “fail-safe”
- Tomato bushy stunt virus (TBSV) contain an RNA regulatory element that is required for replication
- Wang & White replaced this with a theophylline aptamer



Fig. 1. Schematic structures of the TBSV genome and a viral replicon. (A) Linear representation of the TBSV RNA genome. Encoded viral proteins are depicted as boxes with their molecular masses (in thousands) prefixed by “p.” The 5’ UTR is delineated. Initiation sites for sg mRNA transcription are labeled sg1 and sg2, and corresponding structures of the two sg mRNAs are represented by arrows above the genome. (B) The TBSV-derived replicon shown is composed of four noncontiguous regions (I–IV) that correspond to different segments of the viral genome (delineated by vertical dotted lines). Note that region I in the replicon is the 5’ UTR from the viral genome. The horizontal line joining the four regions represents genomic segments that are not present in the replicon.

Many viruses contain RNA regulatory elements critical for replication, packaging or other processes

# Wang & White, 2007

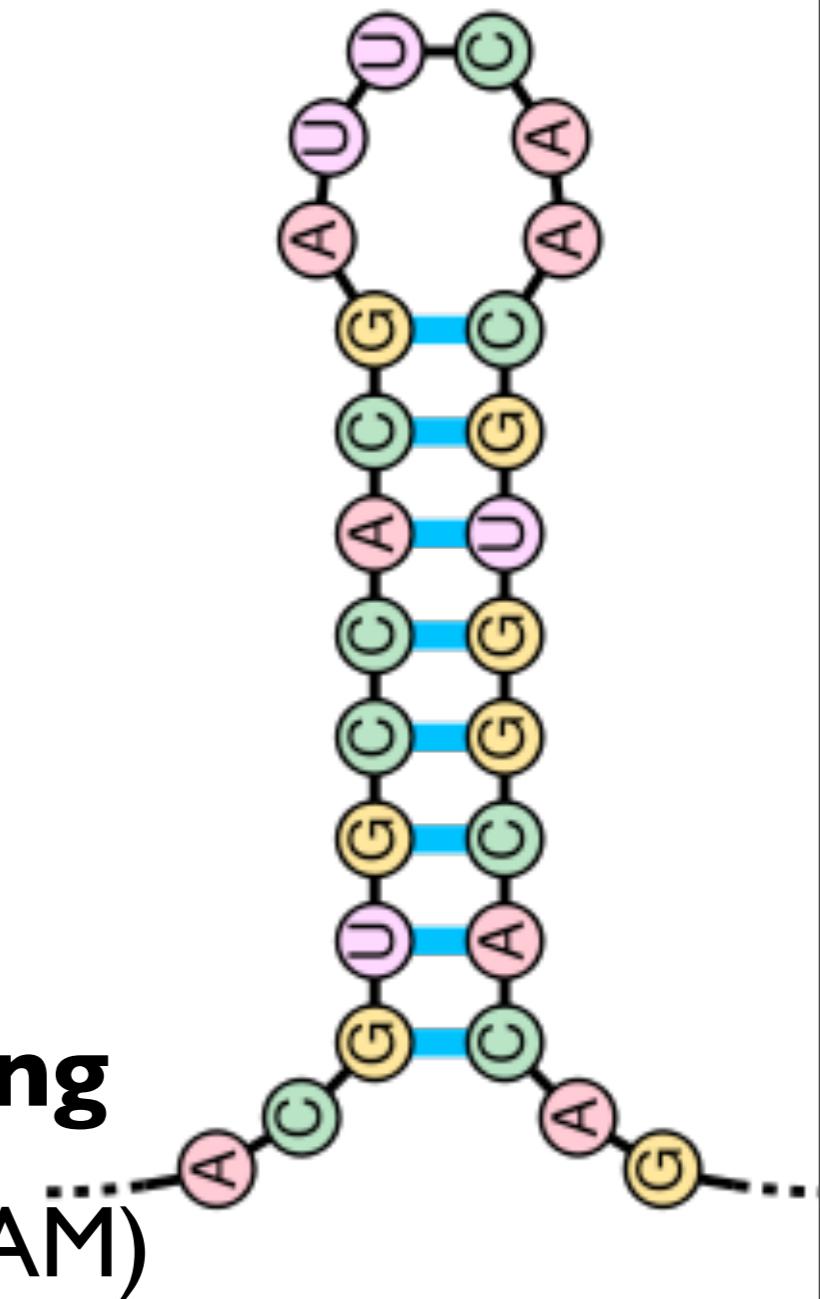


**Fig. 2.** Designing and testing a ligand-inducible viral replicon. (A) RNA secondary structure of the 5' UTR of the TBSV RNA genome (which also corresponds to region I of the replicon). The 5'-proximal T-shaped domain (TSD) and 3'-proximal downstream domain (DSD) are indicated. These domains are separated by SL5, which is outlined by shading. A tertiary interaction, PK-TD1, is indicated, and substitutions in Rep-TD1c are shown in the box. (B) Structure of theophylline (\*, N7). (C) A theophylline-binding RNA aptamer (1). (D) The structures of putative RE riboswitches that were used to replace SL5 in Rep-TD1c are shown at the top. Sequence differences in the closing stems are boxed. Replicons were cotransfected with helper TBSV genome sg1T100 into plant-cell protoplasts, incubated in the absence (–) or presence (+) of 0.3 mM theophylline, and their accumulation levels quantified by Northern blot analysis at 22 h after cotransfection. The values at the bottom represent means that were normalized to that for uninduced Rep-TD1c (set at 100%) with standard deviations from three separate experiments.

Stable stem-loop SL5 is replaced by an unstable stem-loop which is stabilized by theophylline binding

# Tools for RNA design & discovery

- **Structure prediction** (*folding, inverse folding, partition function*)
- Prediction of RNA **folding kinetics**
- Analysis of RNA **interactions**
- **Visualization** of RNA structure
- Tools specifically for **designing** RNA
- Tools for general RNA informatics:
  - RNA **gene prediction**
  - Homology search, **alignment, profiling**
  - Databases of RNA gene alignments (RFAM)



Rfam

Rfam: Corona\_package (RF00182)

http://www.sanger.ac.uk/cgi-bin/Rfam/getacc?RF00182

Google

BioWiki PubMed FaxItNice info/eloise News Printer NetflixQ Mapquest Yam Zabo TWikiPlugins Shortcuts kennarts RT

**Rfam** RNA families database of alignments and CMs Wellcome Trust Sanger Institute

Home Keyword Search Sequence Search Browse Rfam Genomes ftp Help miRNA Corona\_package family Corona\_package (RF00182)

**Wikipedia annotation**

**Coronavirus packaging signal**

[Edit Wikipedia Entry](#)

Please [read the notes](#) on editing Wikipedia annotations before making any changes.

The **Coronavirus packaging signal** is a conserved *cis*-regulatory element found in [Coronavirus](#) which has an important role in regulating the packaging of the viral genome into the capsid.

As part of the viral life cycle, within the infected cell, the viral genome becomes associated with viral [proteins](#) and assembles into new infective [progeny](#) viruses. This process is called packaging and is vital for viral replication. This virus has [positive-sense](#) single-stranded [RNA](#) genome. A short region (190 base pairs) in the viral genome was identified that interacts with a [viral envelope protein](#) (protein M) and enables the viral RNA to be specificity packaged into virions [1][2].

**References**

1. Qin L, Xiong B, Luo C, et al (2003). "Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis". *Acta Pharmacol. Sin.* **24** (6): 489-96. PMID 12791173.
2. Narayanan K, Makino S (2001). "Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging". *J. Virol.* **75** (19): 9059-67. DOI:10.1128/JVI.75.19.9059-9067.2001 PMID 11533169.

**Coronavirus packaging signal**

Type: Cis-reg; 2 $\text{\AA}$  structure: Published; PubMed

Seed alignment: PubMed

Avg length: 129.0 nucleotides

Avg identity: 82%

**Alignment**

Seed (15)  Full (23)

Format:

Coloured Blocked alignment

[Get alignment](#)

Help relating to Rfam alignments [here](#)

**Member sequences**

Seed (15)  Full (23)

[View Members](#)

**Species distribution**

Tree depth : Show all levels

[View Species Tree](#)

This molecular or cell biology article is a **stub**. You can help Wikipedia by [expanding it](#).

Retrieved from "[http://en.wikipedia.org/wiki/Coronavirus\\_packaging\\_signal](http://en.wikipedia.org/wiki/Coronavirus_packaging_signal)"

Categories: Cis-regulatory RNA elements | Molecular and cellular biology stubs

All Wikipedia text is available under the terms of the [GNU Free Documentation License](#).

# A few challenges of ribo-engineering

- Systematic discovery of natural RNAs
- Increased automation of directed evolution
- Broader chemical repertoire for RNA
- Modification of DNA/RNA/protein by RNA
- Mature theory of ribocircuit design

# Viral structure

(Schaffer, Koerber & Lim, 2008)

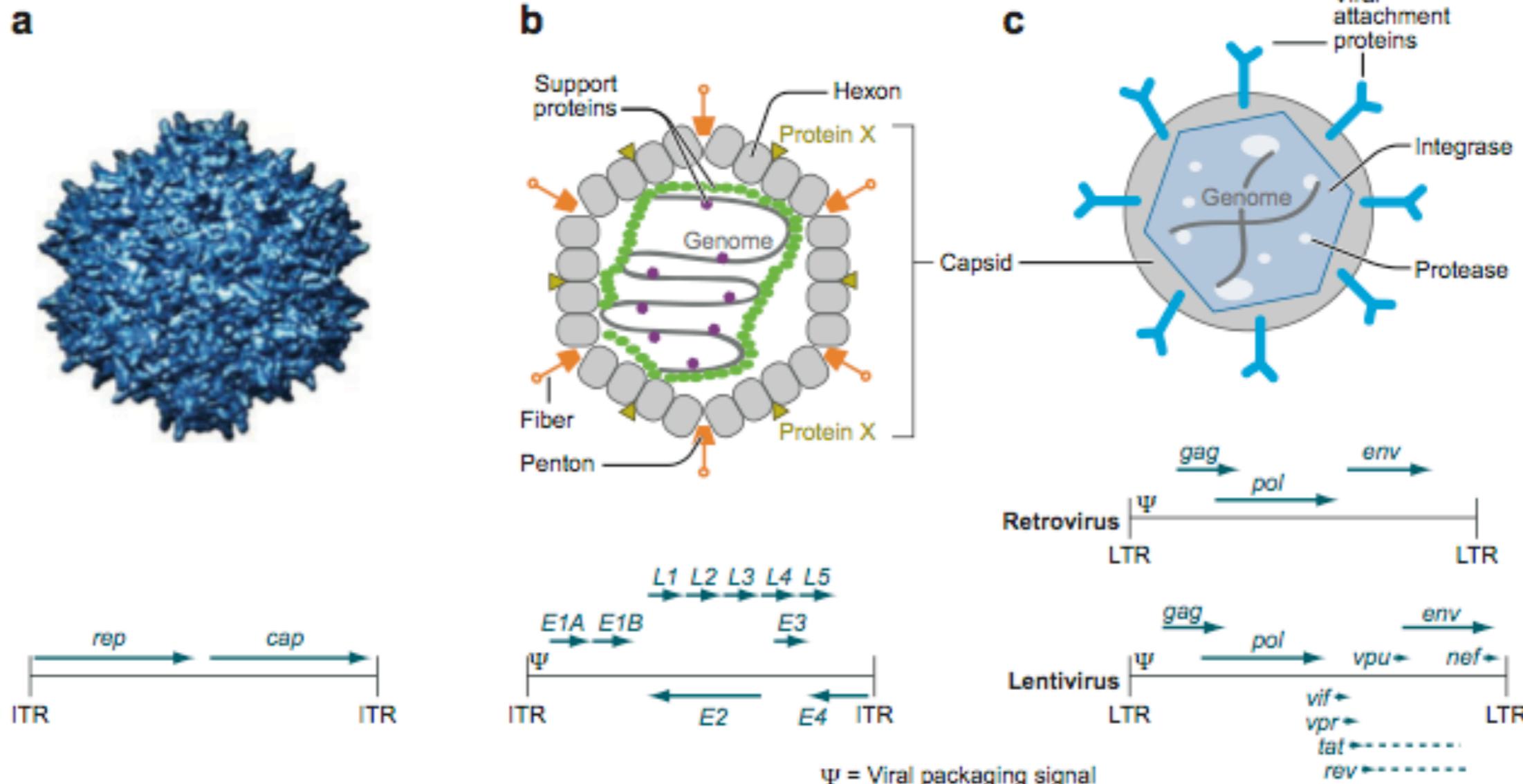
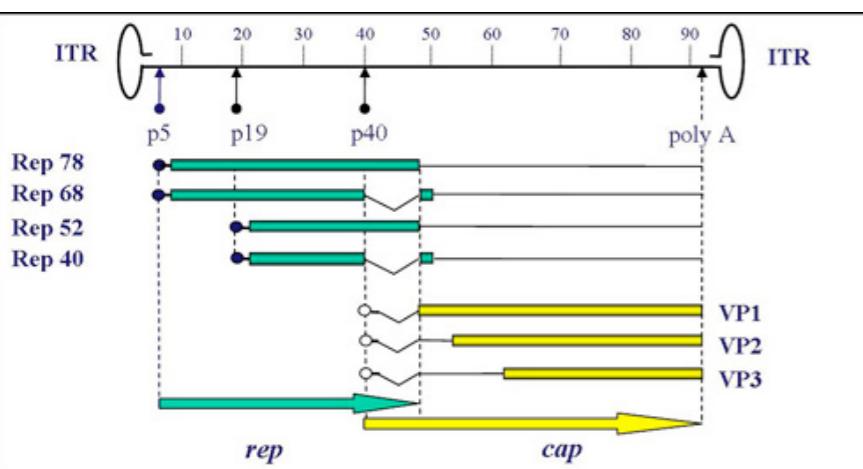
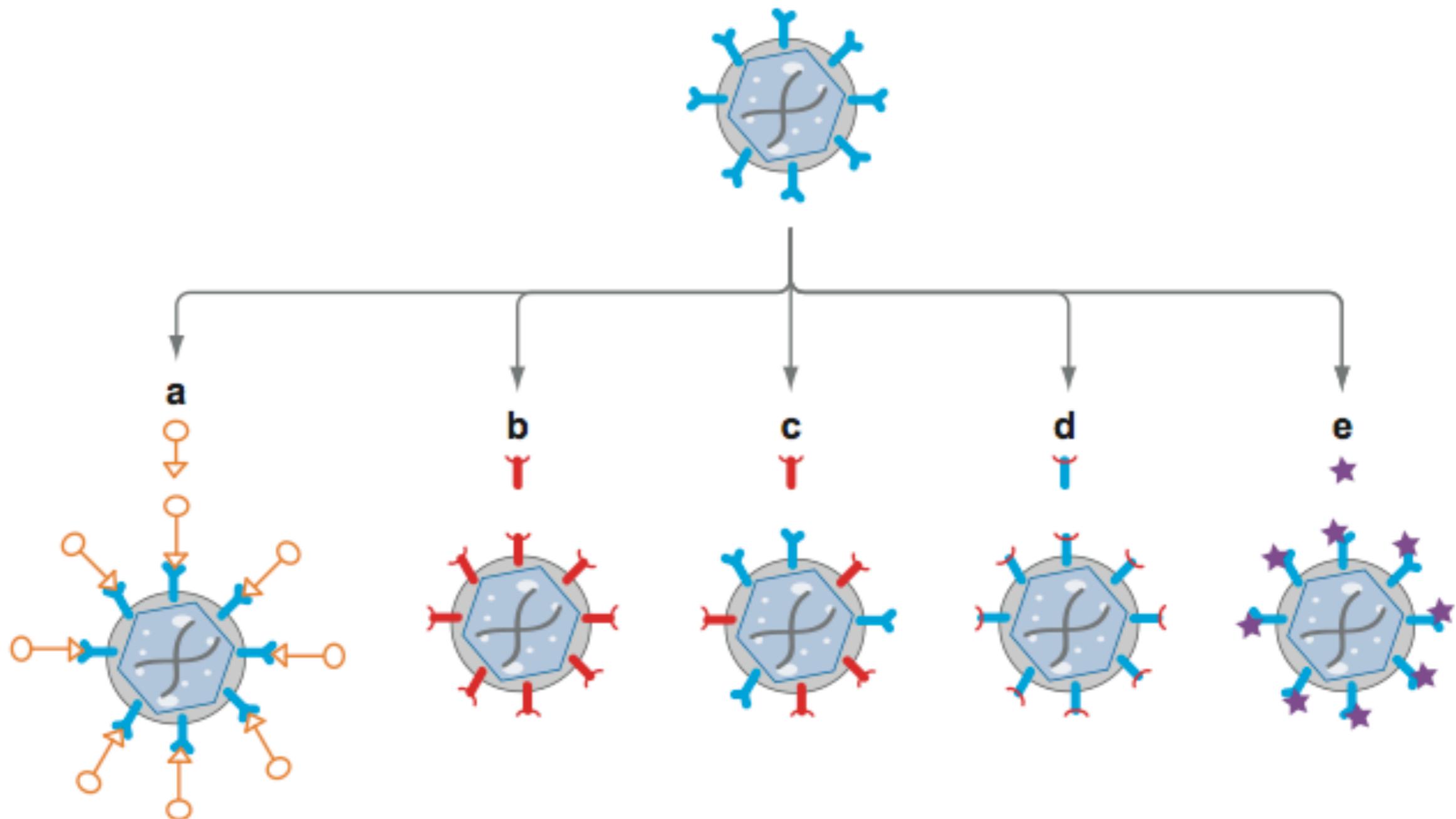


Figure 1

Schematic of the structure of viral particles and organization of the viral genomes. (a) Three-dimensional representation of the AAV capsid from VIPER database (<http://viperdb.scripps.edu/>) and schematic of AAV genome. (b) Representation of key components of the adenoviral capsid and genome organization. (c) Representation of key components of retroviral and lentiviral particles along with the genome organization of each virus.

# “Rational” virus design



**Figure 2**

Overview of rational protein engineering strategies for viral vectors. Rational design methods include (a) the use of a bispecific adaptor, (b) pseudotyping with an alternate capsid or VAP, (c,d) the generation of mosaic or chimeric particles, and (e) genetic engineering of VAP to insert peptide or point mutations.

# *In vitro* virus evolution

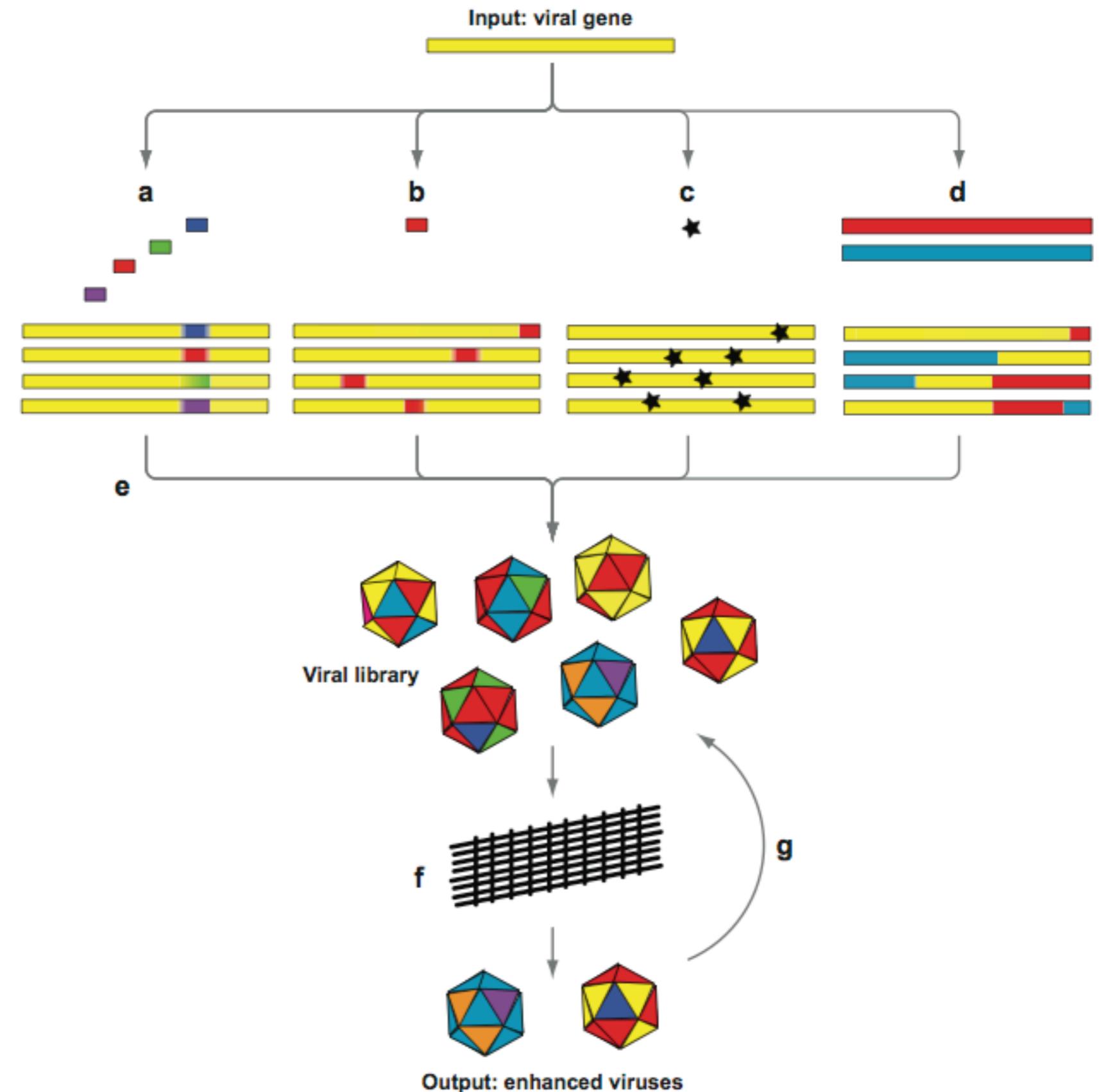


Figure 3

Overview of library protein engineering strategies for viral vectors. Library generation and selection methods include (a) display of random peptide in defined location, (b) random insertional mutagenesis, (c) random point mutagenesis, and (d) in vitro recombination. Upon DNA library generation, a highly diverse viral library is produced (e). Finally, (f) high-throughput selection followed by (g) recovery of successful variants and iteration with steps (a)–(d) and (e) employs directed evolution to enhance desired gene delivery properties.

# Tools for viral bioinformatics

- Sequence analysis
  - Gene-finders (overlapping ORFs)
  - Repeat-finders (LTRs, TIRs)
  - Motif-finders (binding sites)
  - DNA & RNA secondary structure prediction
  - Protein domain analysis tools (e.g. InterPro)
  - Molecular evolution tools (e.g. PAML, recombinant subtyping)
- Visualization
  - Genome browsers
  - Protein 3D structure viewers

