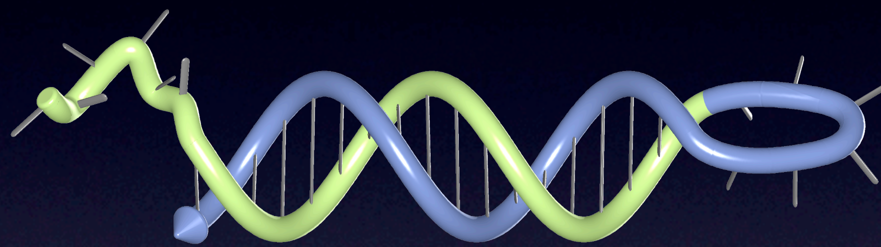
Molecular Instruments



Niles A. Pierce

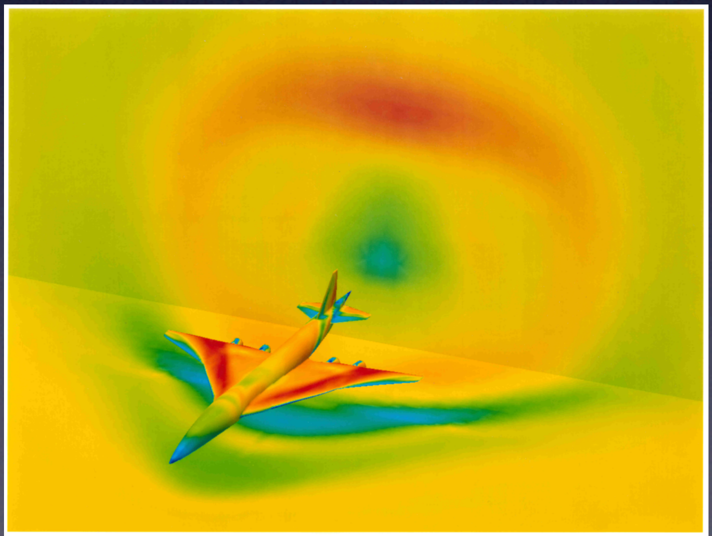
California Institute of Technology

In Honor of Antony Jameson
Stanford University
November 21, 2014

Antony Jameson

Analysis & Design

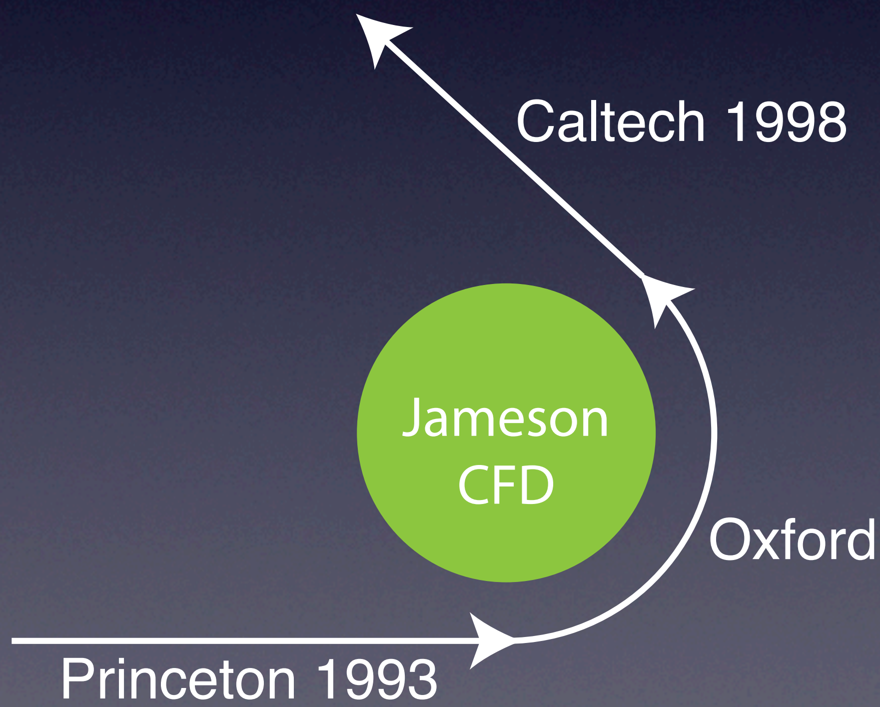
- Defining challenges
- Creative, elegant, practical solutions
- Diverse mathematical and algorithmic principles
- Relentless pursuit of efficiency

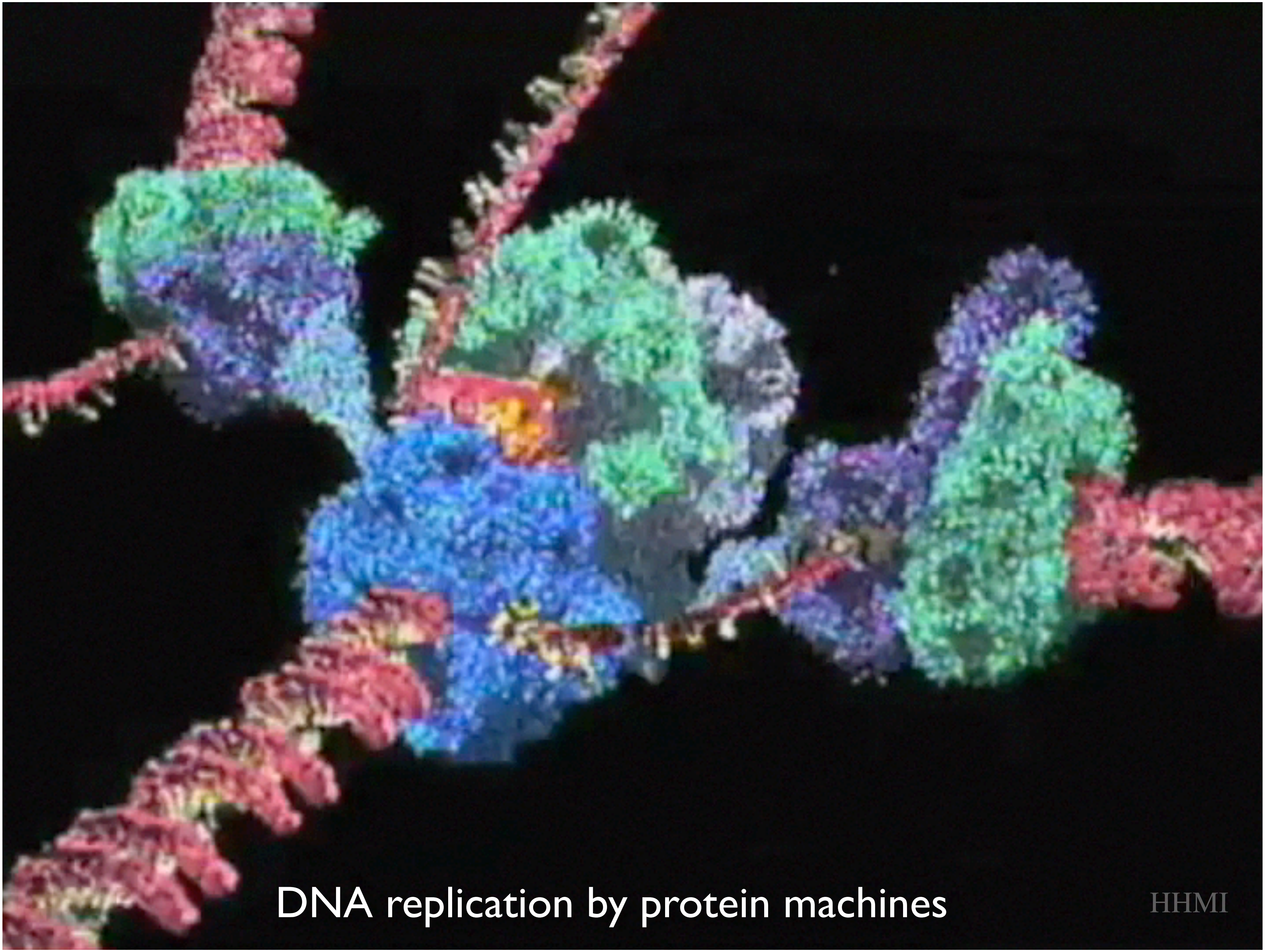


Antony Jameson

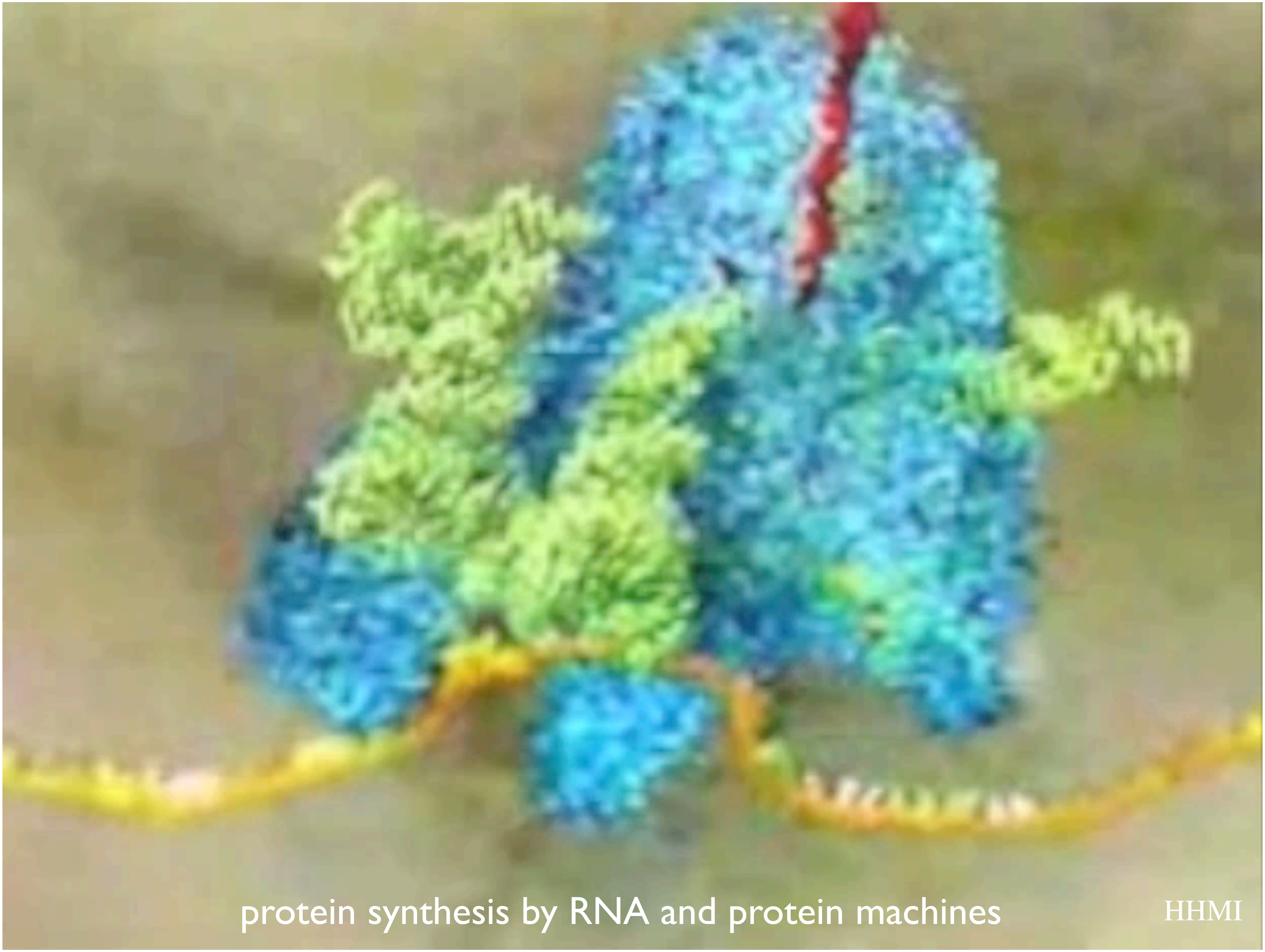
Analysis & Design

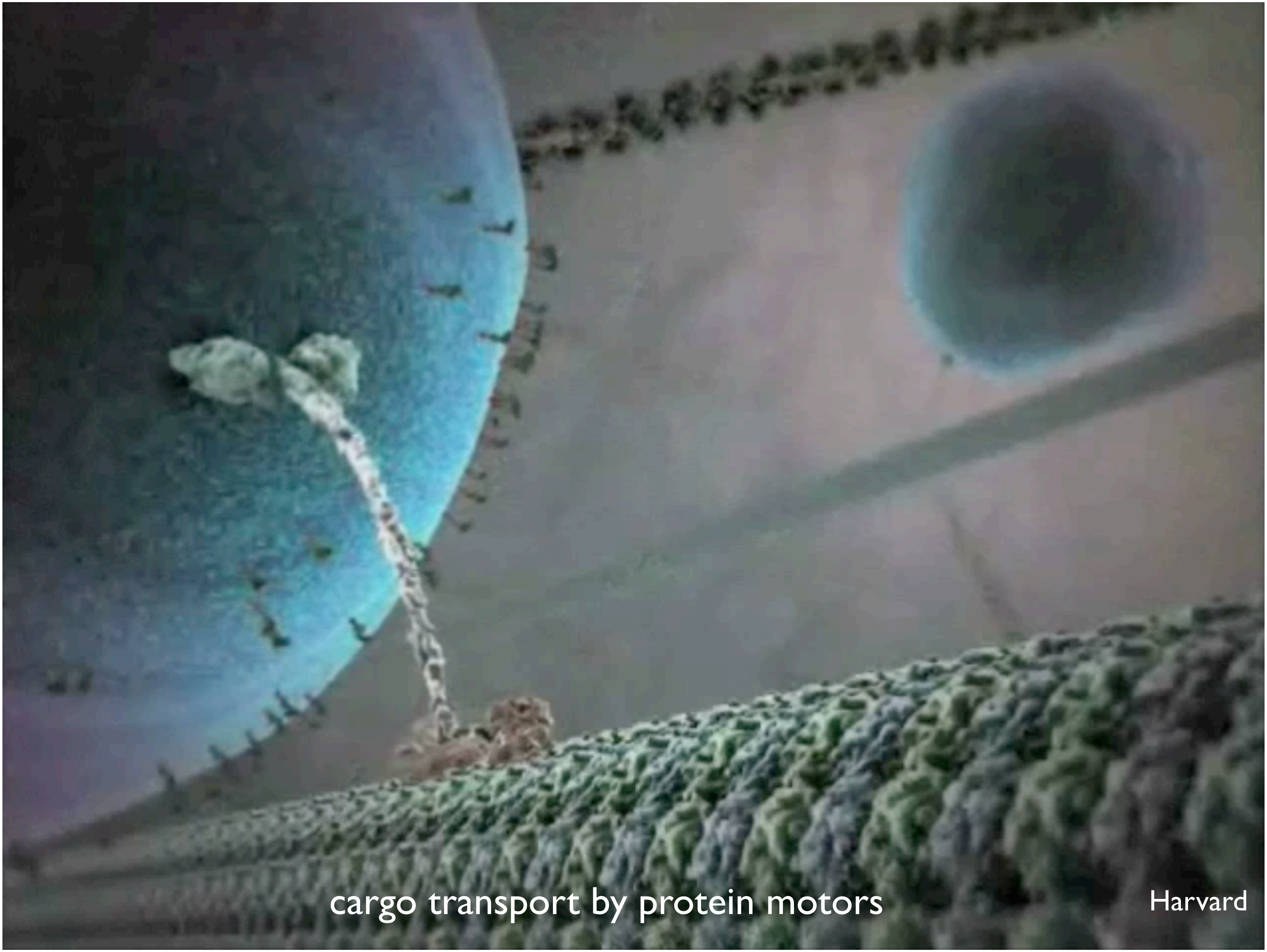
- Defining challenges
- Creative, elegant, practical solutions
- Diverse mathematical and algorithmic principles
- Relentless pursuit of efficiency





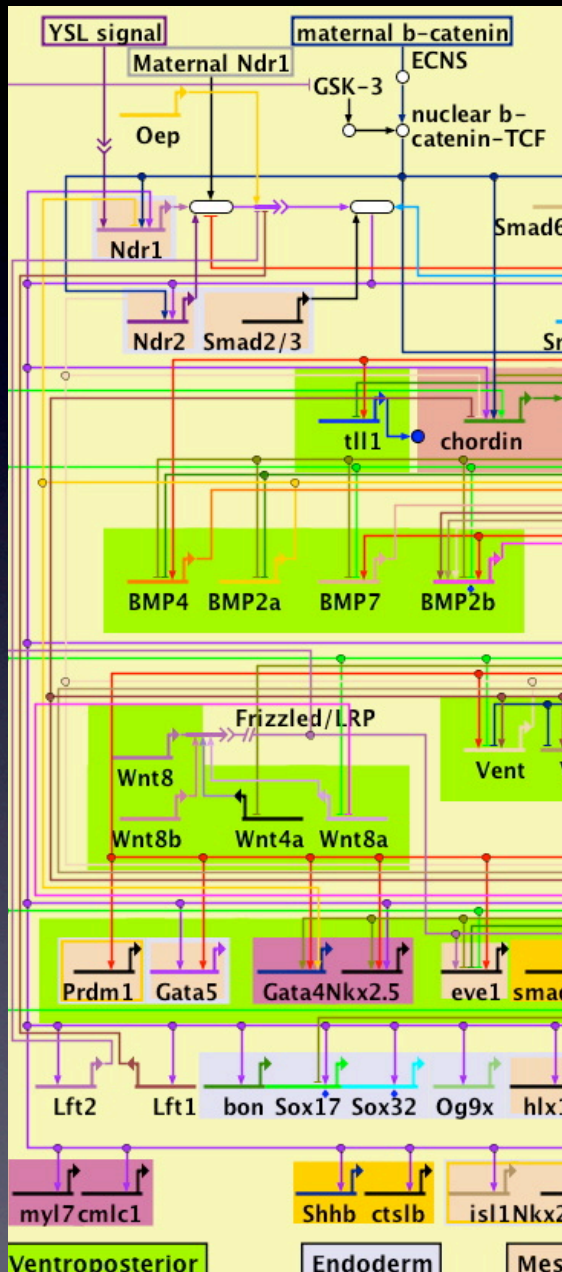
DNA replication by protein machines


protein synthesis by RNA and protein machines
HHMI


cargo transport by protein motors
Harvard

DNA Genome | Regulatory Circuitry | Zebrafish Development

...GTGGTACAGGT
GAATTTGGGTAGGC
TAAATTGTCCATAG
TTTATGTGTGTGAA
TGAGGGTGTATGGA
TGTTTCTCAGAGAT
GGGTTGCAGCTGGA
AGGGCGTCCATTGT
GCAAAACATATGCT
GGAGAAGTTGCCGG
TTCATTCTGCTGTG
GCGACCCCAGATTA
ATAAAAGGACTAAG
CCGAAAAGAAAATG
AAACATATATATAT
ATATATATATATAT
ATATATATATA...



Movie: Sean Megason (Harvard)

Life is orchestrated by programmable biomolecules

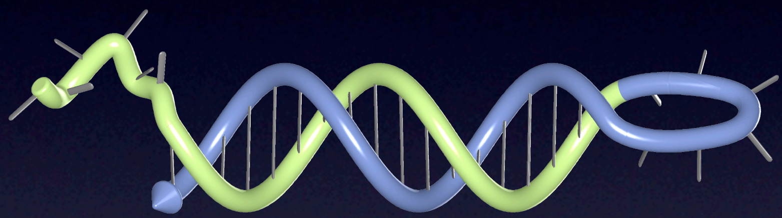
DNA	GTGGTACAGGTGAATTTGGGTAGGCTAAATTGTCCATA
RNA	GGGCUGUUUUCUCGCUGACUUUCAGCCCAAACAAAA
Protein	MTYRLELNGKTLKGRTRTEAVDAATAEKVFKQYANDNG

Sequencing a genome provides a parts list but no manual

Remains challenging to interrogate and perturb the state of endogenous biological circuitry within intact organisms

Molecular Instruments

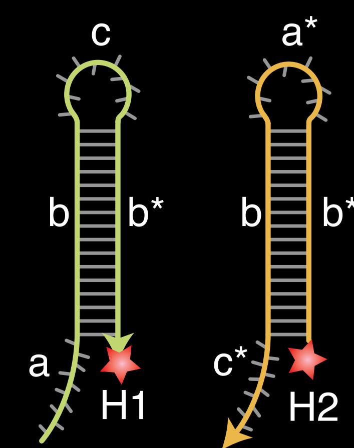
Exploit the very programmability that biological organisms exploit themselves



Engineer small conditional DNAs and RNAs (scDNAs and scRNAs) that interact and change conformation to read out or regulate the state of endogenous biological circuitry within intact organisms

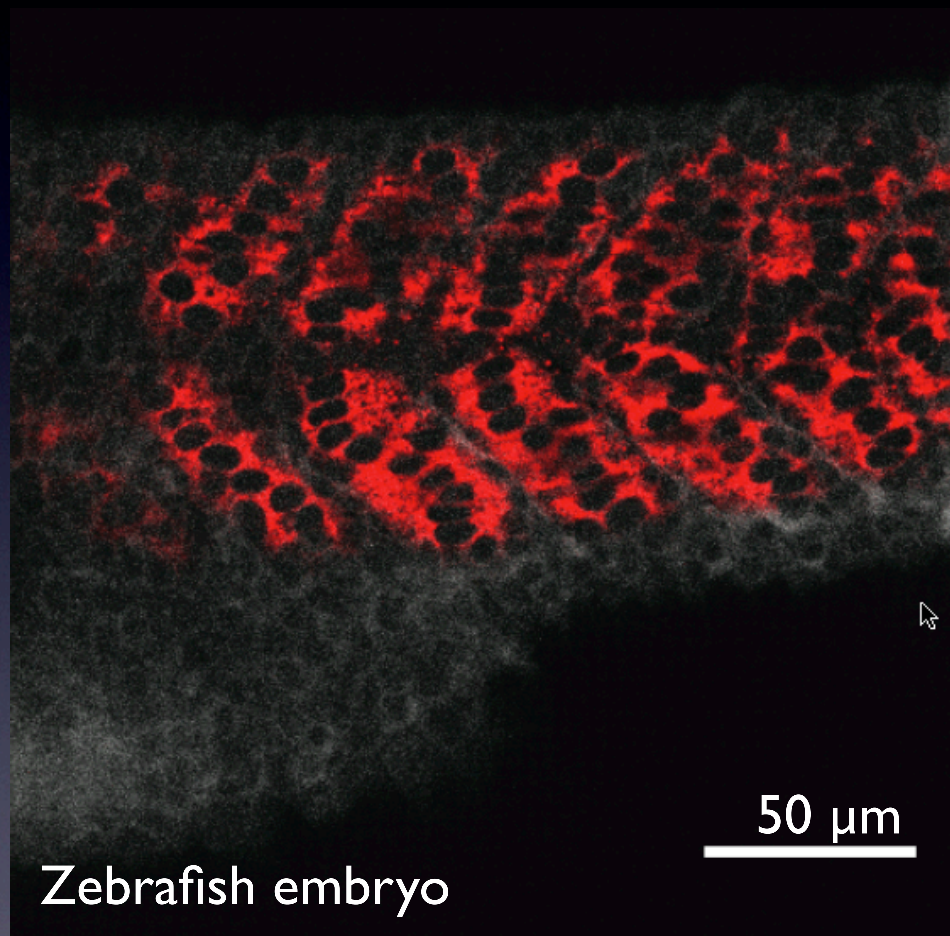
Why Nucleic Acids?

Molecular Instruments:
Mapping Biological Circuits within Organisms



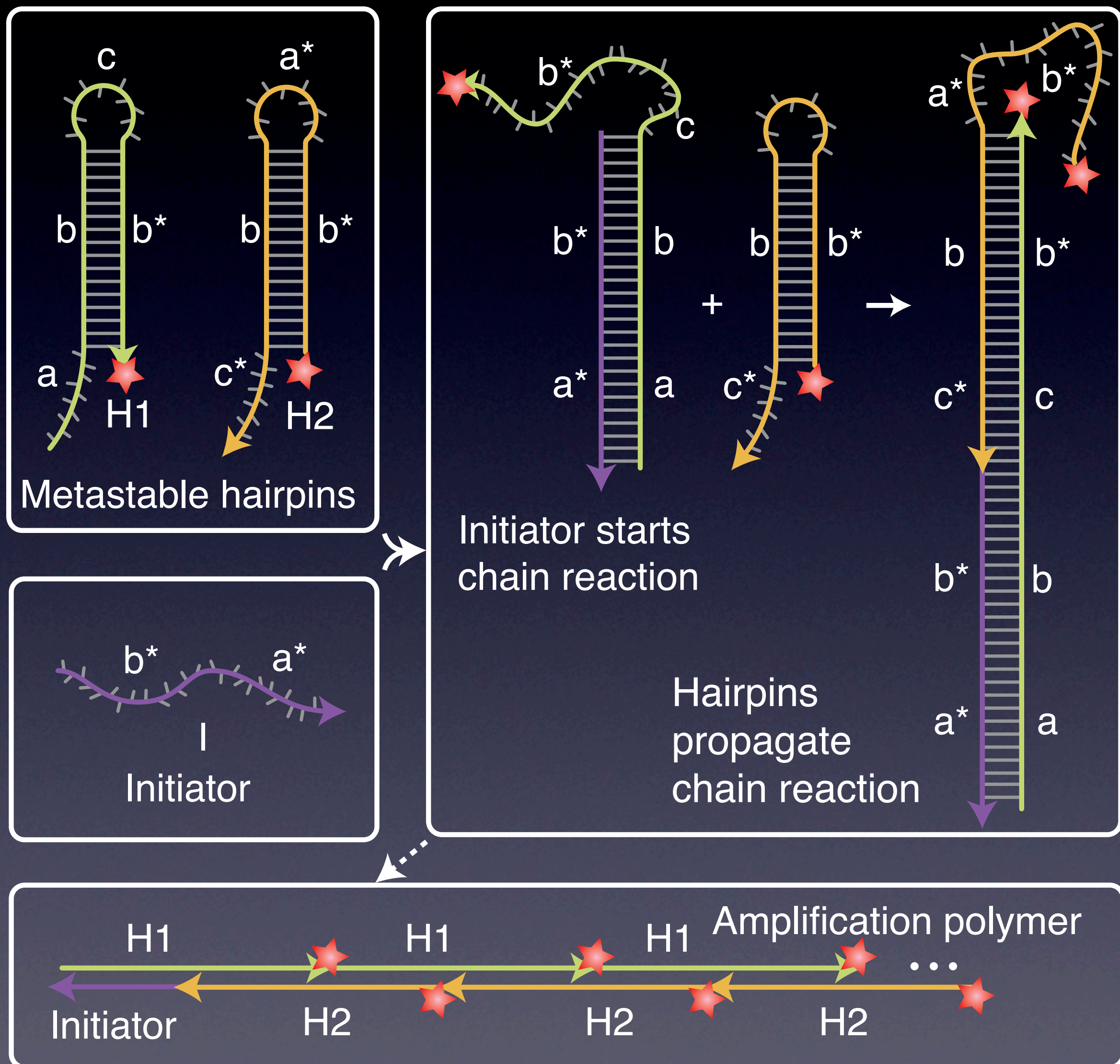
Mapping mRNA expression within intact vertebrate embryos

Goal: accurately map spatial relationships between the regulatory loci of multiple genes



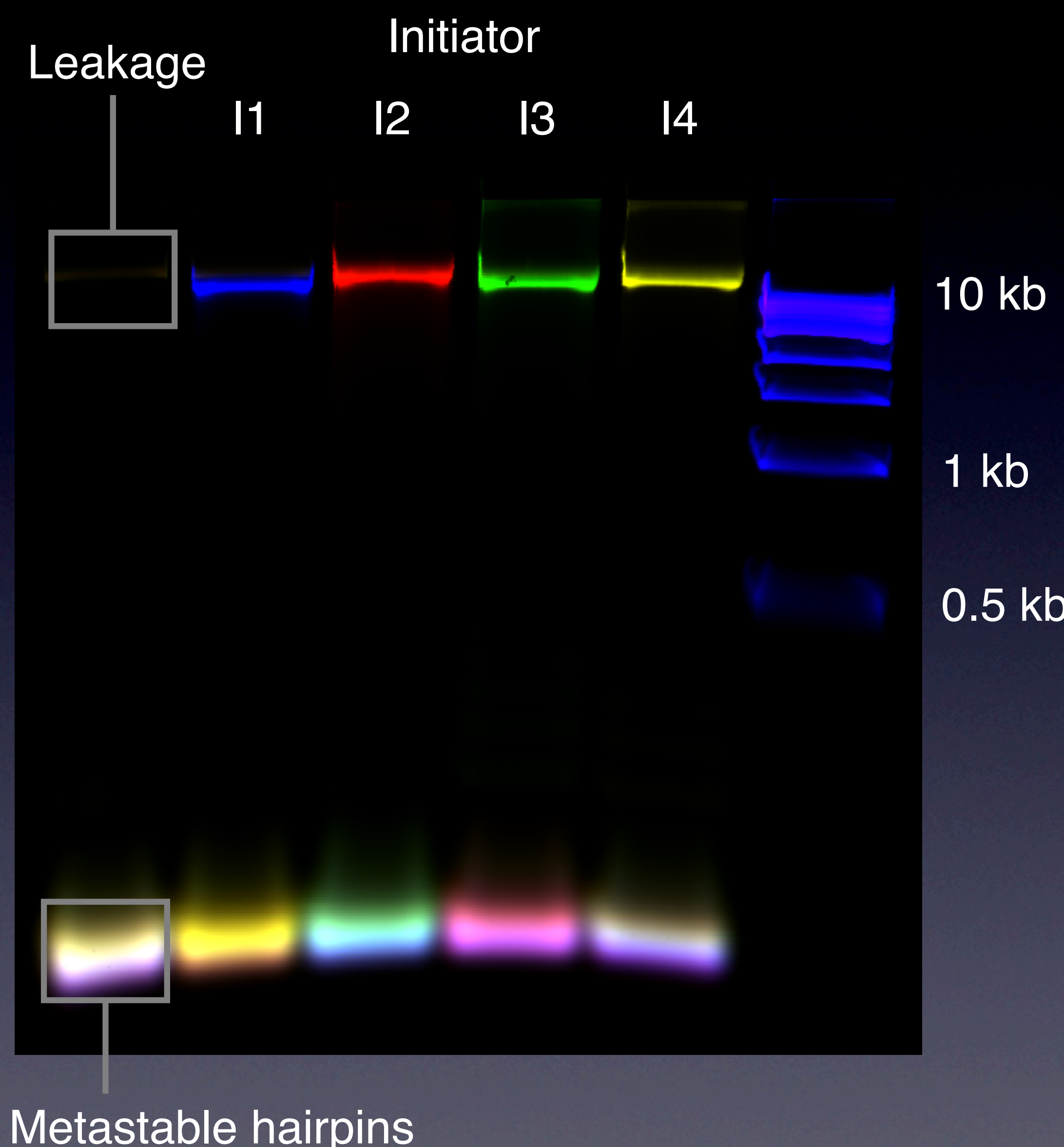
40-year challenge: lack of orthogonal in situ amplifiers leads to cumbersome serial multiplexing approaches

Conditional Self-Assembly via Hybridization Chain Reaction (HCR)

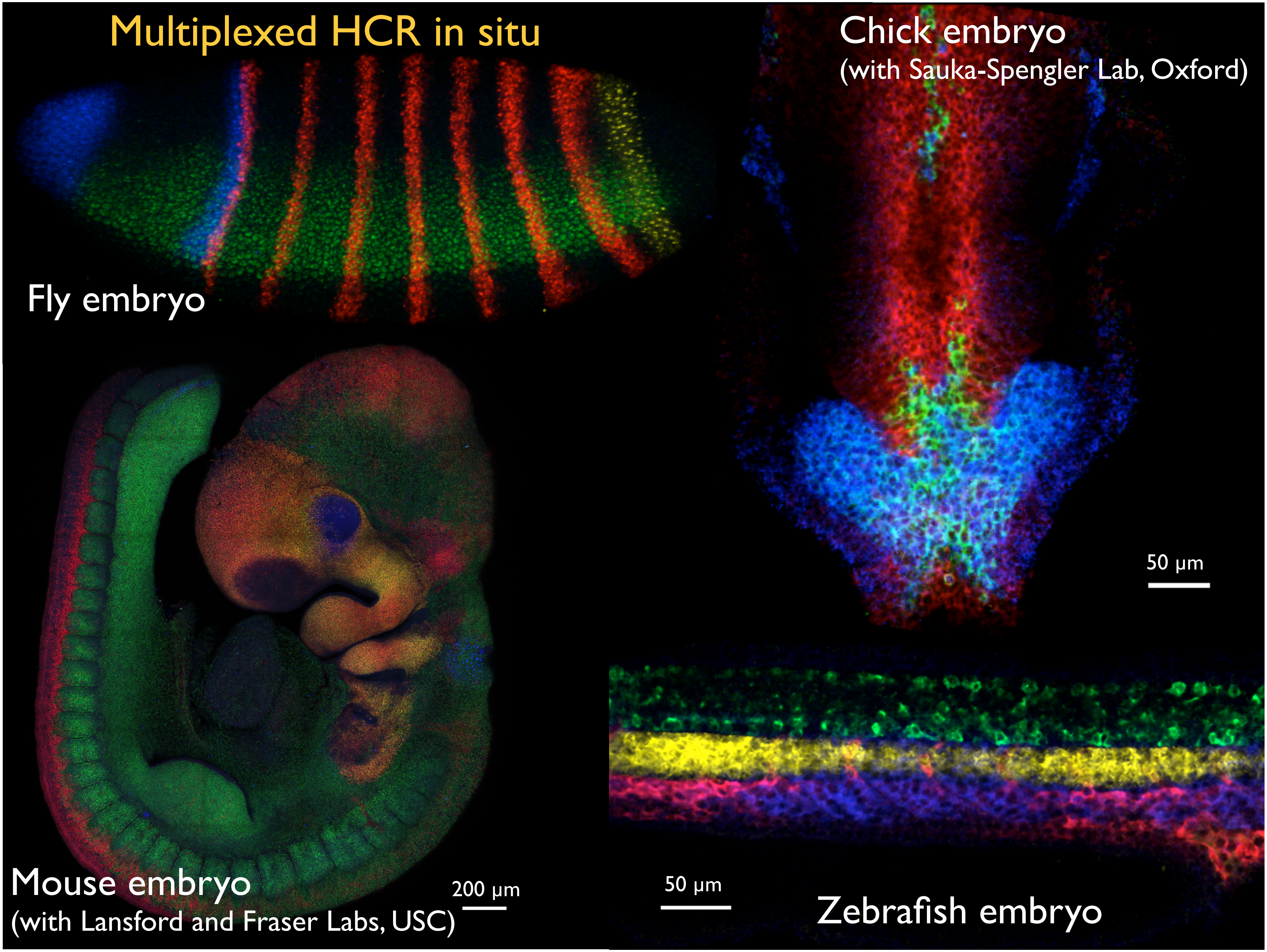


Dirks and Pierce, *Proc Natl Acad Sci USA*, 2004

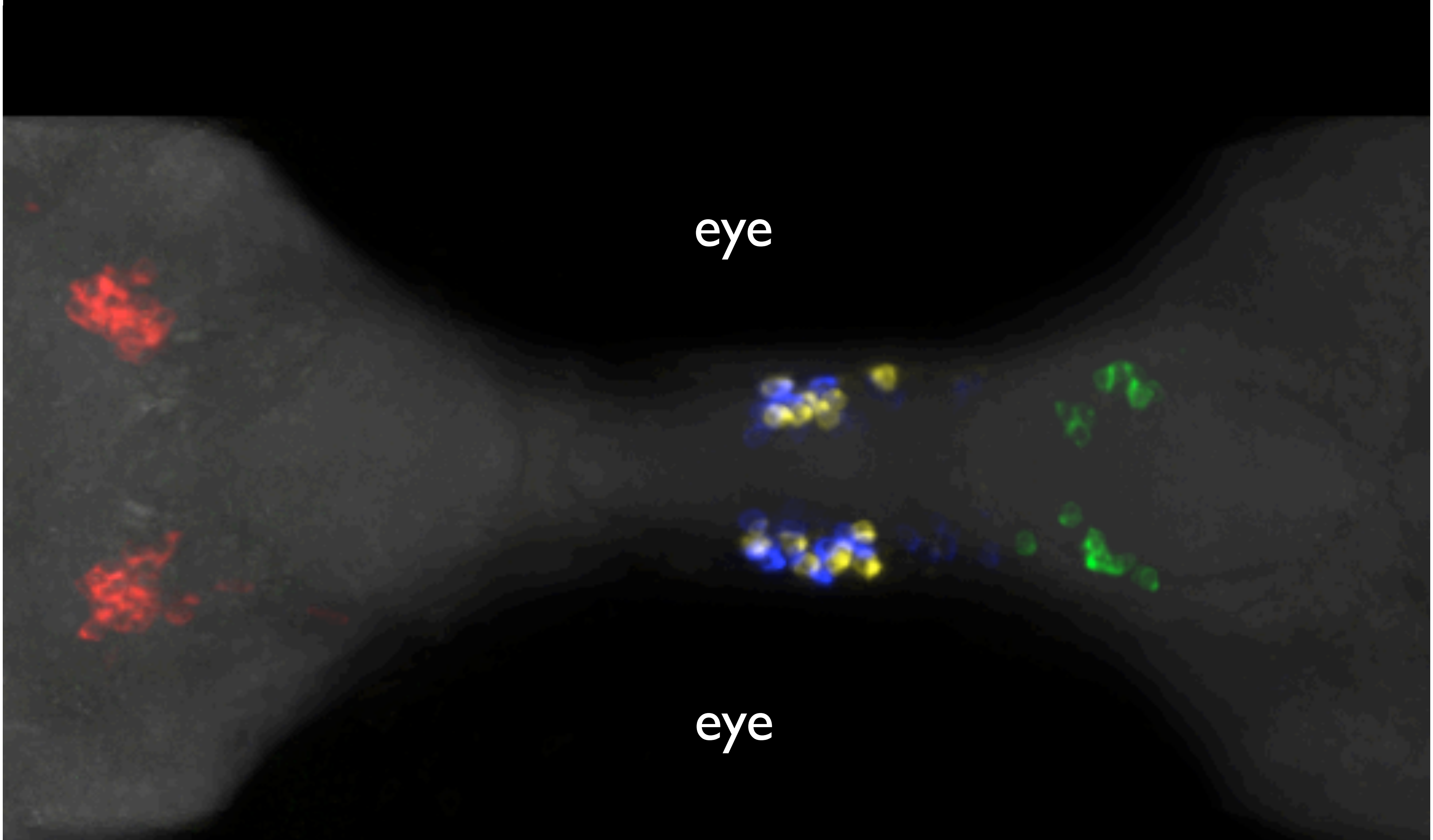
Multiplexed Signal Amplification using HCR



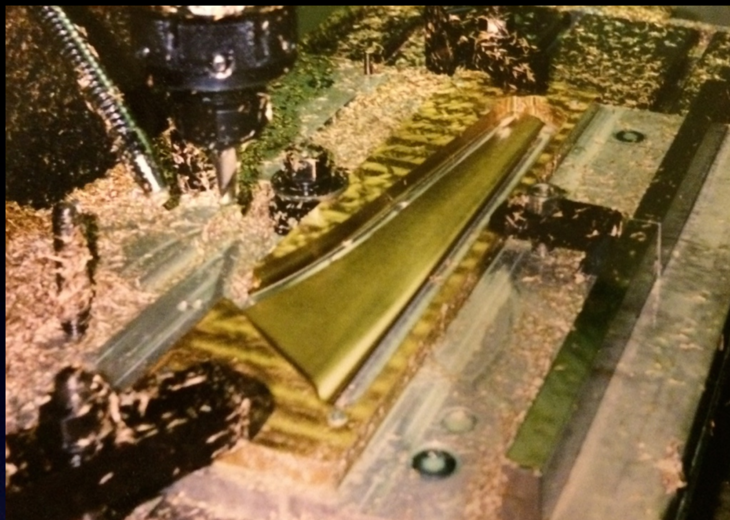
Choi, Beck, and Pierce, *ACS Nano*, 2014


Multiplexed HCR in situ
Chick embryo
(with Sauka-Spengler Lab, Oxford)
Fly embryo
50 μm
Mouse embryo
(with Lansford and Fraser Labs, USC)
200 μm
50 μm
Zebrafish embryo

mRNAs map to specific cells within the developing zebrafish brain



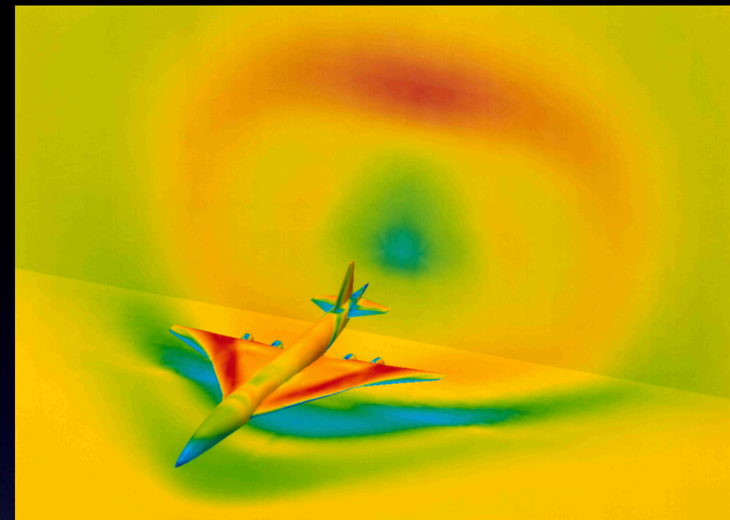
Shape



Analysis →

← Design

Flow



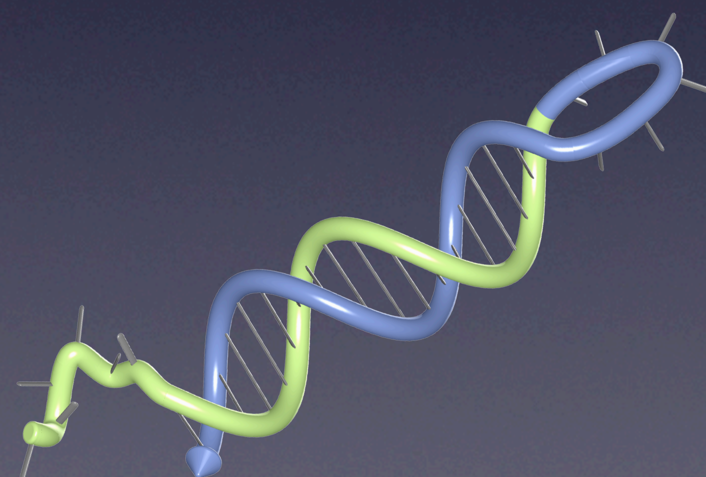
Sequence

GTGGTACAGGTGAATTTG
GGTAGGCTAAATTGTCCA
TAGTTTATGTGTGTGAAT
GAGGGTGTATGGATGTTT
CTCAGAGATGGGTTGCAG
CTGGAAGGGCGTCCATTG
TGCAAAACATATGAGGGT

Analysis →

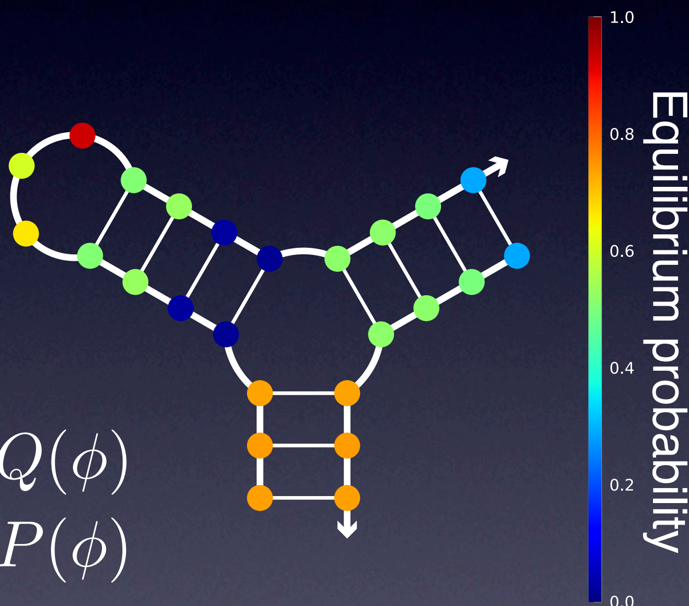
← Design

Shape



Analysis of a Complex

- Arbitrary number of strands
- Dynamic programming
- Graph theory
- Group theory



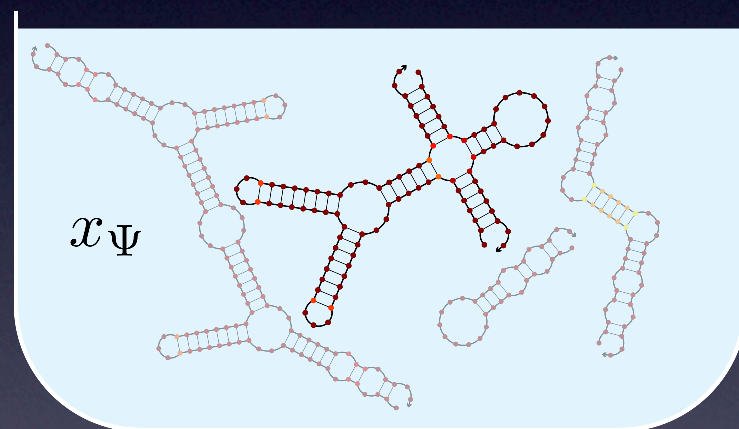
$Q(\phi)$

$P(\phi)$

Complexity: $\Theta(N^3)$

Analysis of a Test Tube

- arbitrary complex species Ψ
- convex programming
- Lagrange duality



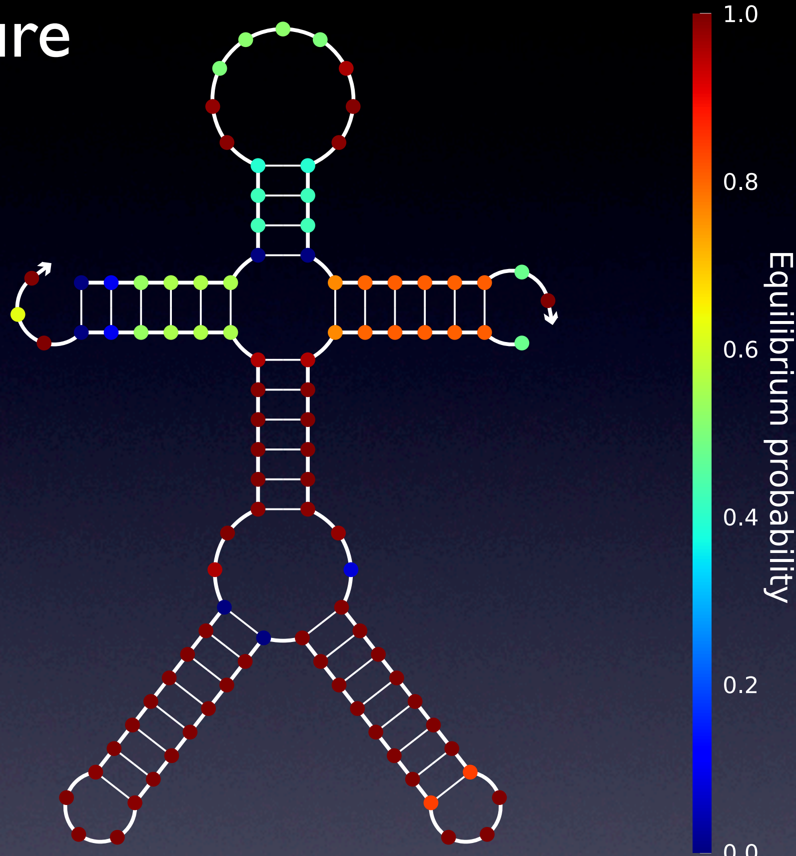
Complexity: $O(|\Psi|N^3_{\max})$

R.M. Dirks, J.S. Bois, J.M. Schaeffer, E. Winfree, N.A. Pierce, *SIAM Rev*, 2007

Demo

Complex Design

Target structure



Objective Function

Complex ensemble defect: average number of incorrectly paired nucleotides at equilibrium over the ensemble of the complex

Dirks, Lin, Winfree, Pierce, *Nucl Acids Res*, 2004

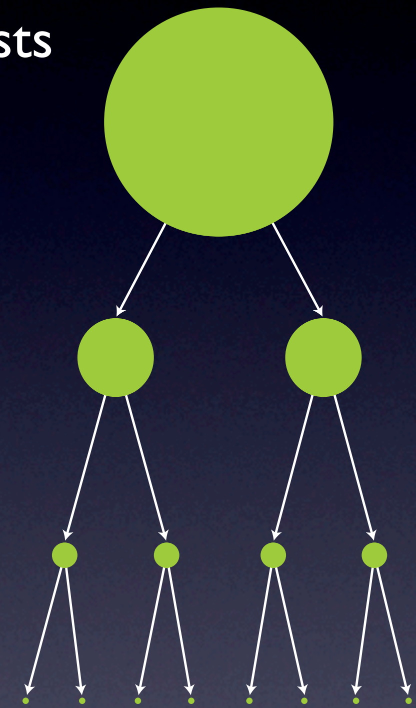
Cost of Design

Goal: Mutate sequence to reduce objective function

Question: If each evaluation of the objective function costs

$$c_{\text{eval}}(N) = \Theta(N^3)$$

what is the cost of designing a sequence from scratch?



Approach: hierarchical ensemble decomposition

Optimality bound:

$$c_{\text{des}}(N) \geq c_{\text{eval}}(N) \left[1 + 2\left(\tfrac{1}{2}\right)^3 + 4\left(\tfrac{1}{4}\right)^3 + 8\left(\tfrac{1}{8}\right)^3 + \ldots\right]$$

Zadeh, Wolfe, Pierce, *J Comput Chem*, 2011

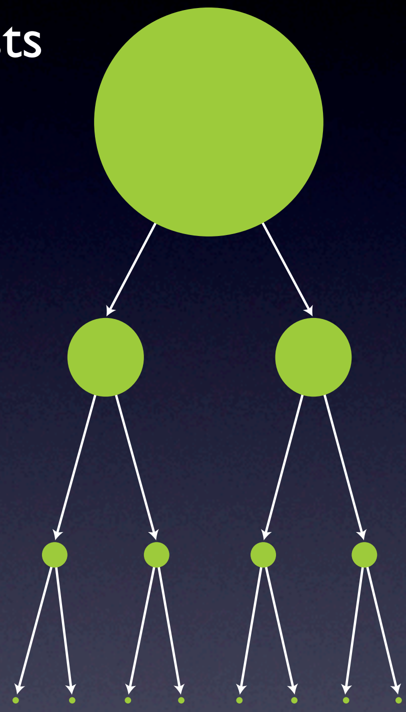
Cost of Design

Goal: Mutate sequence to reduce objective function

Question: If each evaluation of the objective function costs

$$c_{\mathrm{eval}}(N) = \Theta(N^3)$$

what is the cost of designing a sequence from scratch?



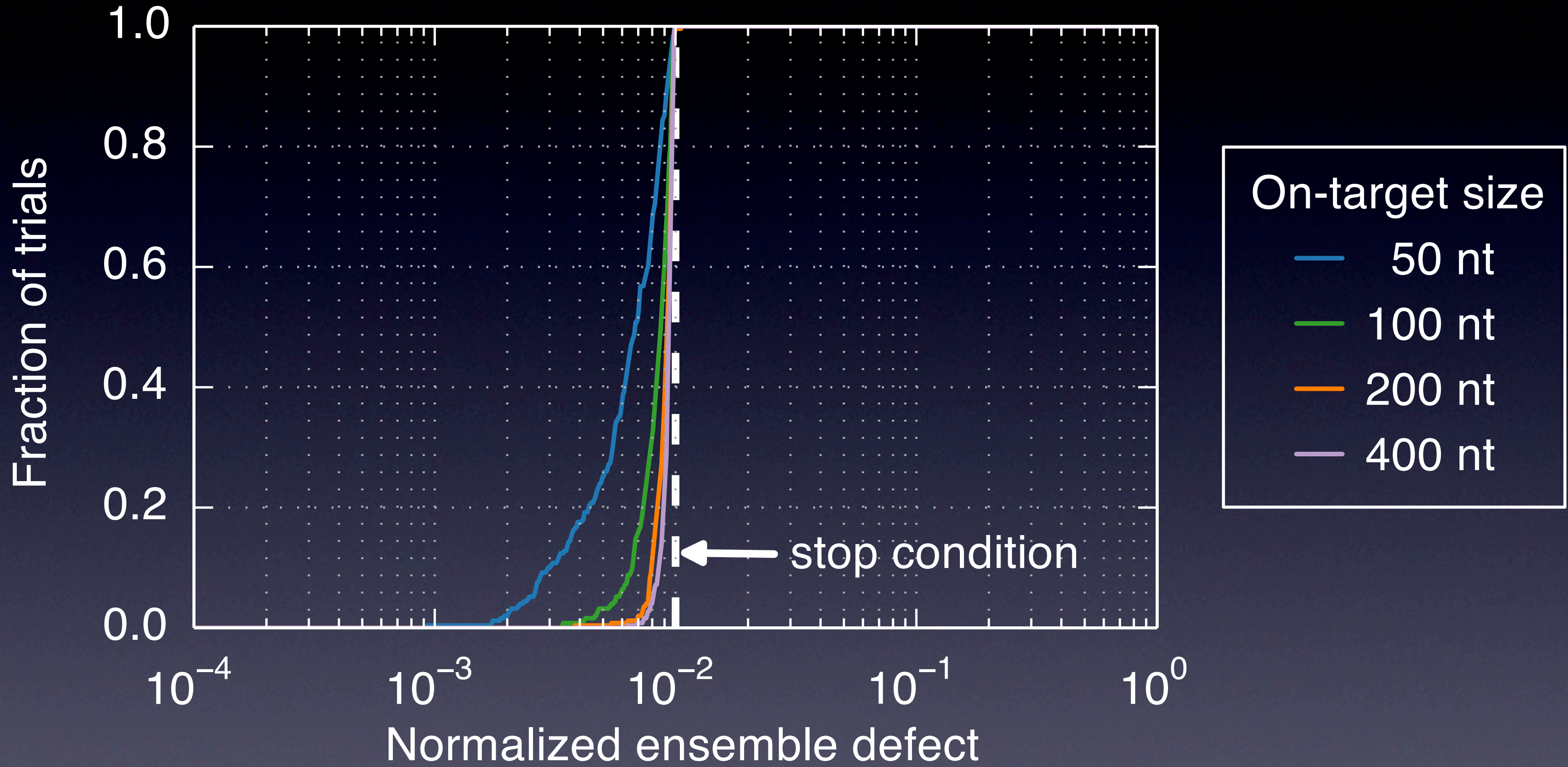
Approach: hierarchical ensemble decomposition

Optimality bound:

$$c_{\mathrm{des}}(N) \geq \frac{4}{3} c_{\mathrm{eval}}(N)$$

Zadeh, Wolfe, Pierce, *J Comput Chem*, 2011

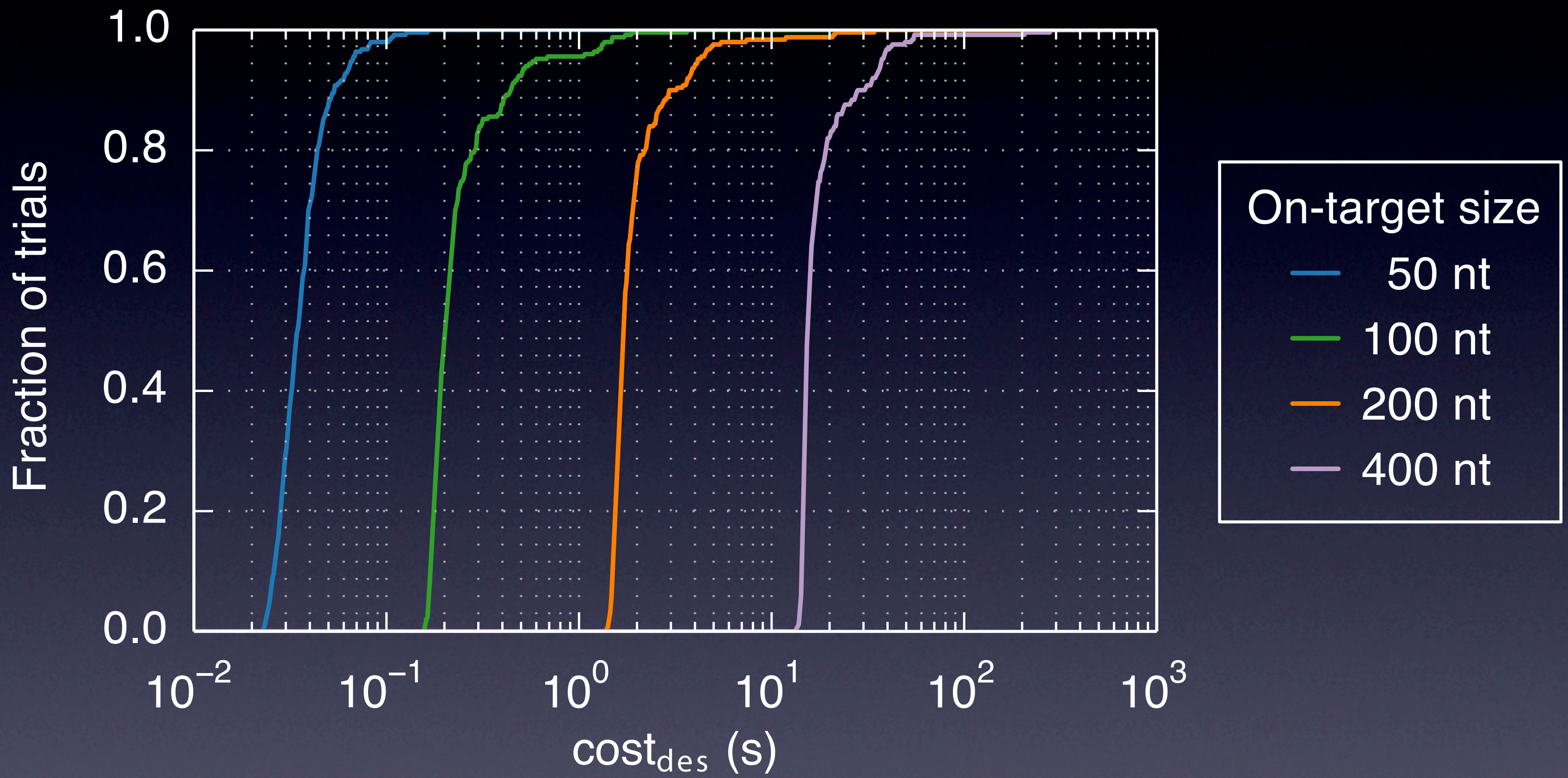
Design Quality



Randomly generated target structures

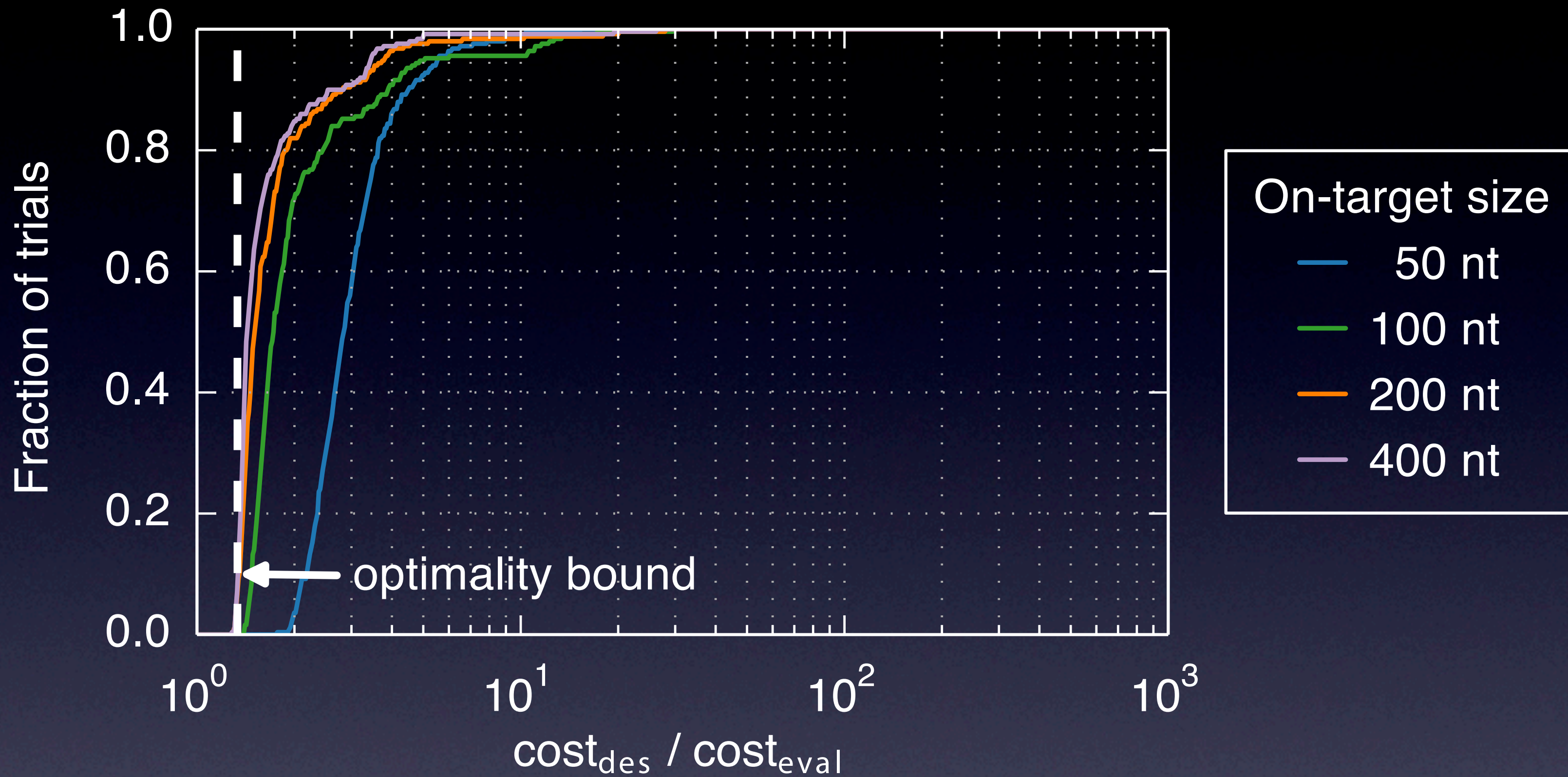
Typically <1% of nucleotides incorrectly paired at equilibrium

Design Cost



Typical design times ranging from fraction of second to seconds

Relative Design Cost

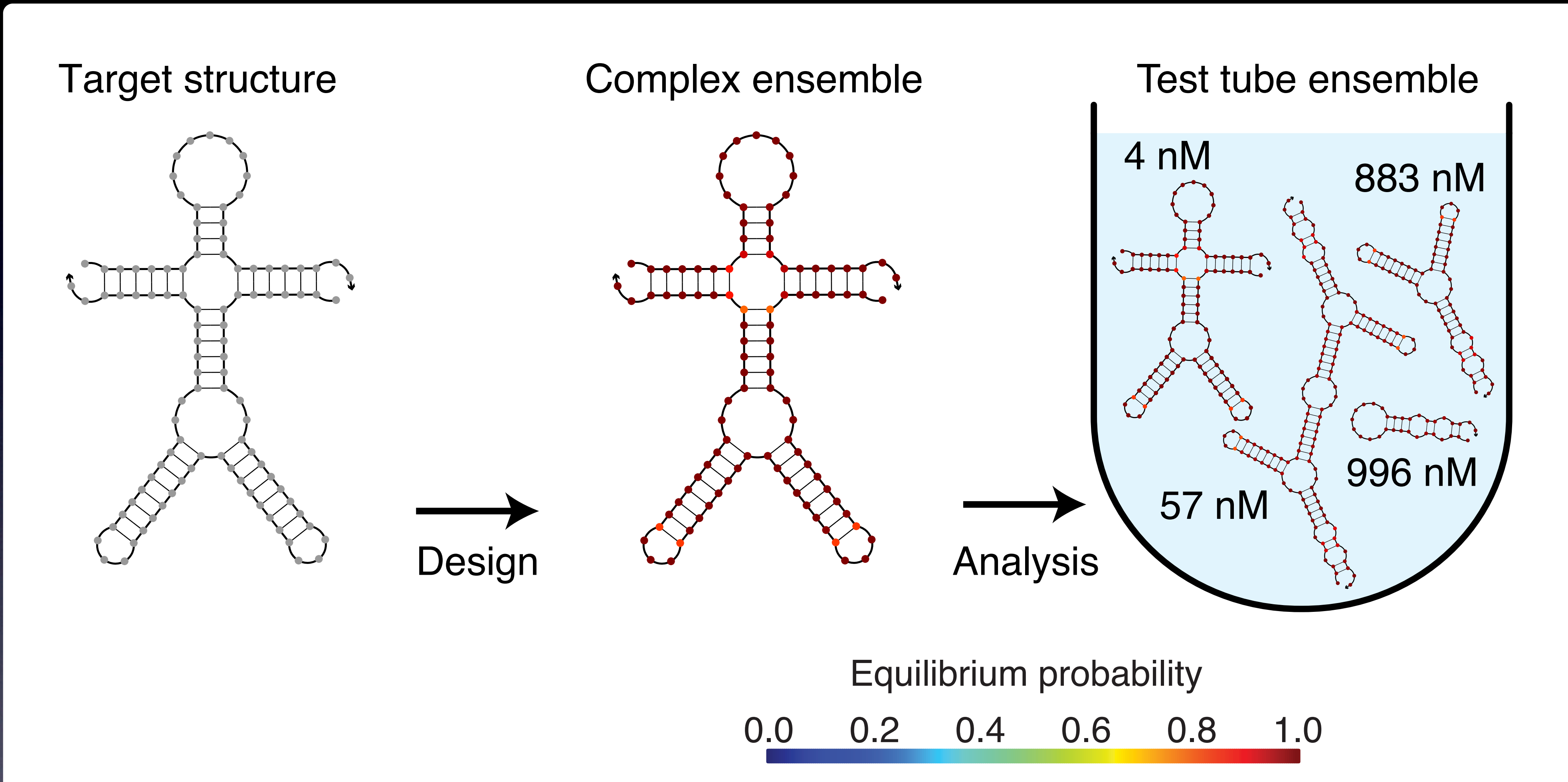


As complex size increases, typical cost of design decreases to 4/3 cost of analysis!

Empirically, algorithm exhibits asymptotic optimality

Zadeh, Wolfe, Pierce, *J Comput Chem*, 2011

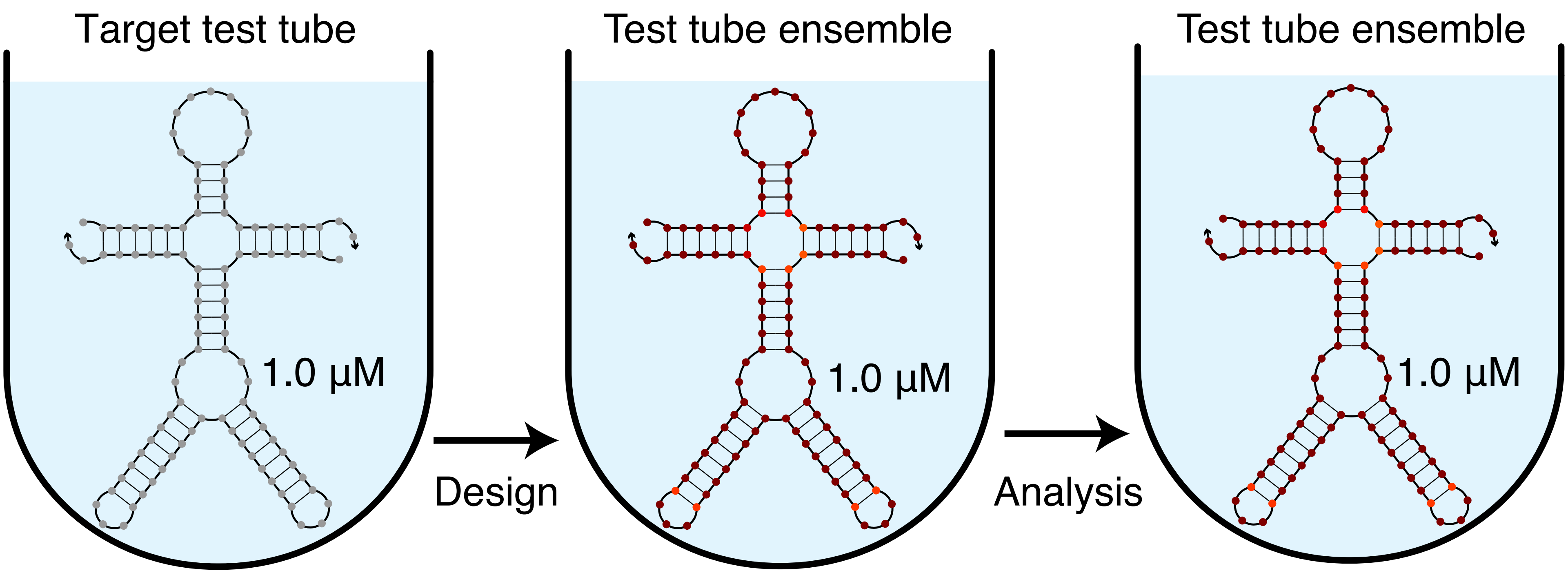
Complex Design: A Cautionary Tale



Complex design does not take into consideration:

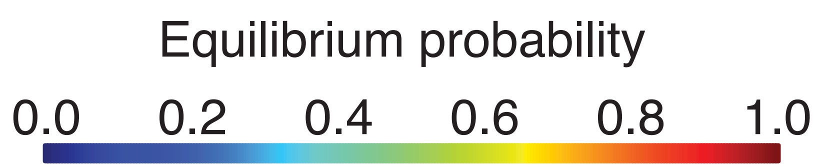
- concentration of desired complex
- concentration of competing off-target complexes

Test Tube Design



On-target complexes:

- target structure
- target concentration



Off-target complexes:

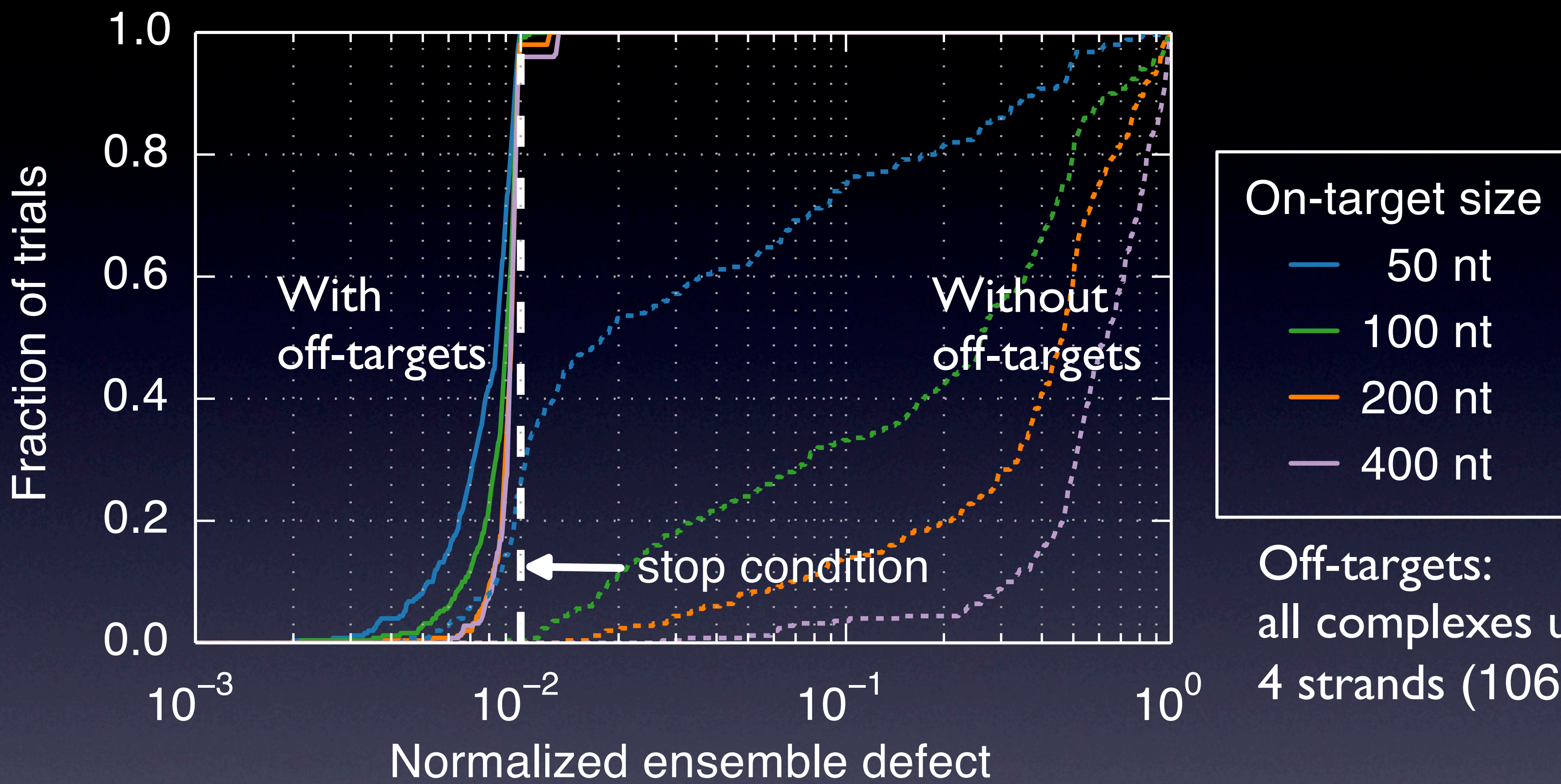
- vanishing target concentration

Wolfe and Pierce, *ACS Synth Biol*, 2014

Include 100s to 1000s of off-target complexes

Typical design cost 2x-10x analysis cost

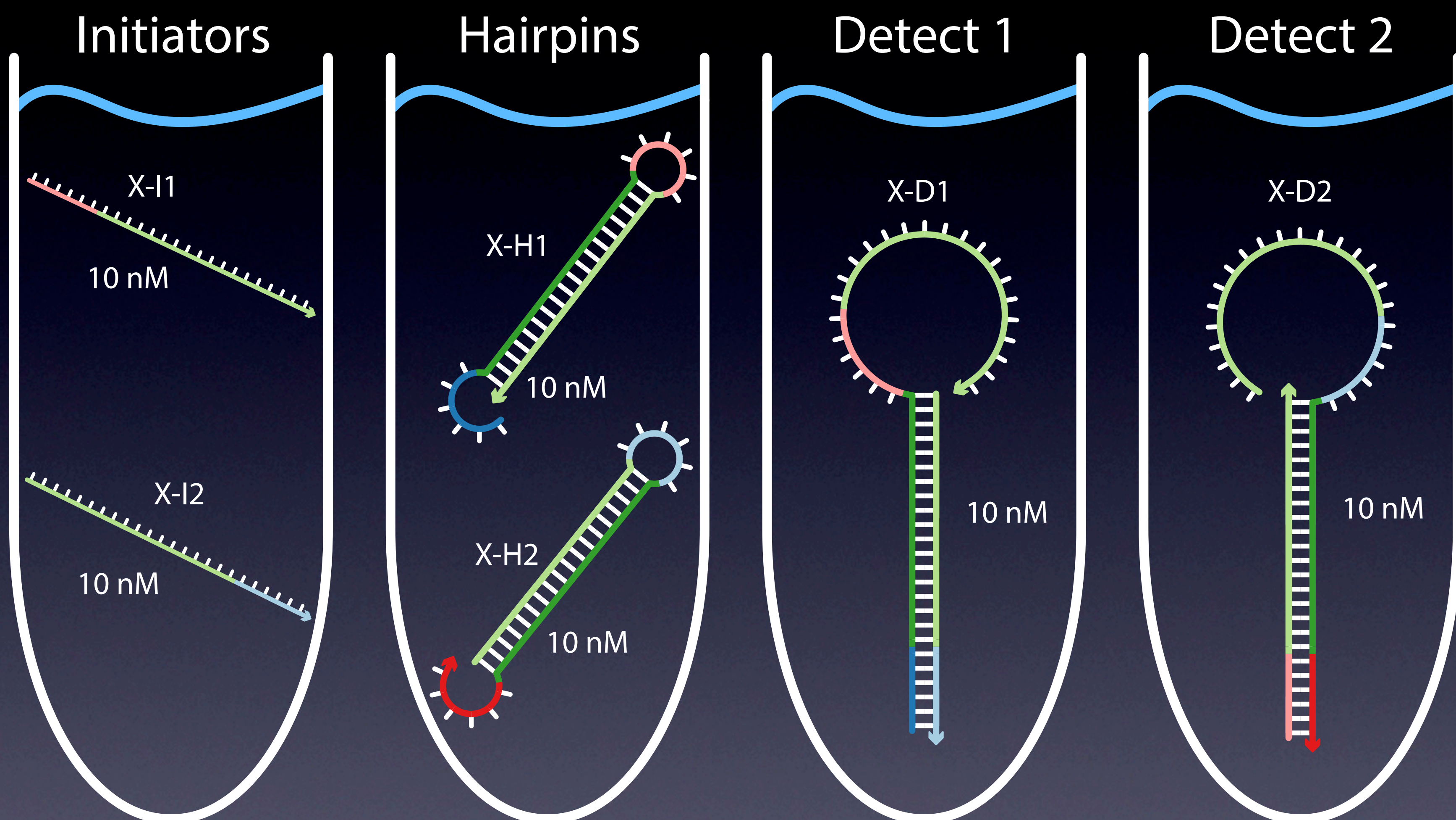
Importance of Designing Against Off-Targets



Off-targets:
all complexes up to
4 strands (106 total)

Designing without off-targets in the design ensemble leads to poor sequence quality

Multistate Test Tube Design



Target test tubes represent reactant, intermediate, and product states

Wolfe, Zadeh, Dirks, Pierce, *unpublished*

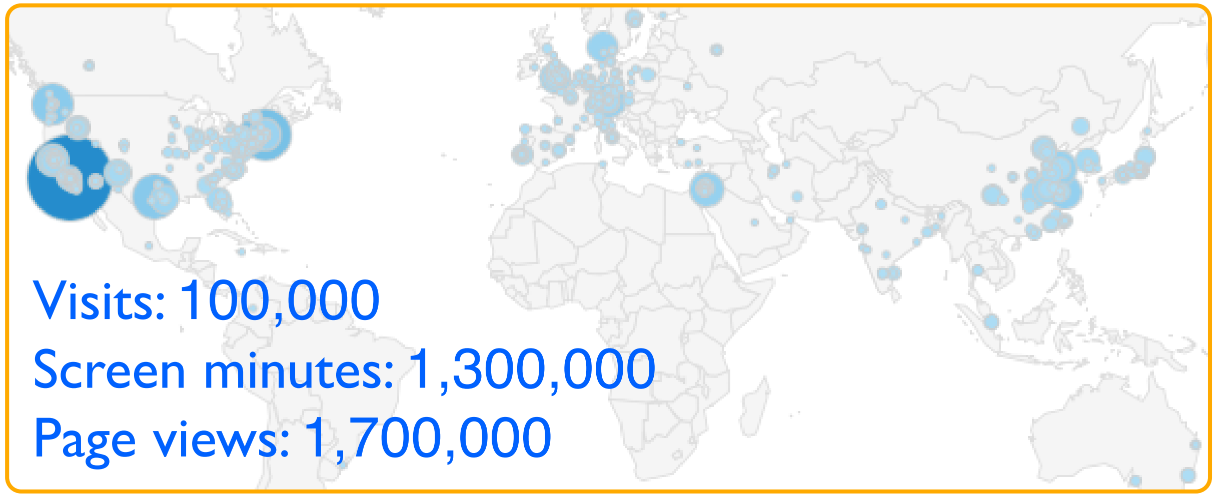
Demo

Non-Profit Academic Resources

NUPACK
nupack.org

Provide researchers with algorithms for the analysis and design of nucleic acid molecules, devices, and systems

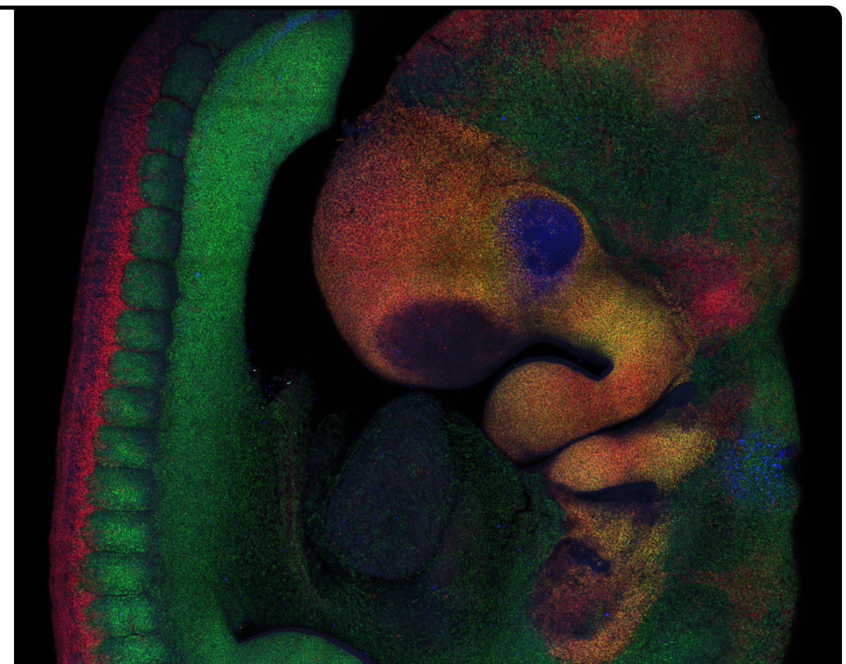
NUPACK usage during the last 3 years

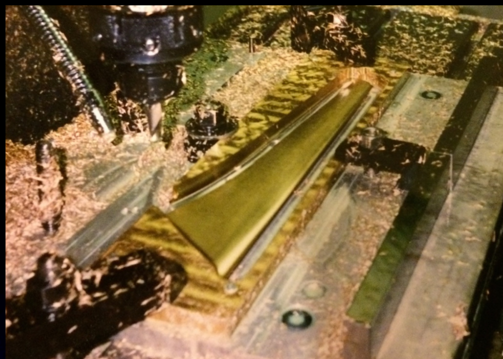


Molecular Instruments
molecularinstruments.org

Provide researchers with programmable molecular instruments for readout and regulation

Custom molecular instruments synthesized for over 70 labs





GTGGTACAGGTGAATTTGGG
TAGGCTAAATTGTCCATAGT
TTATGTGTGTGAATGAGGGT
GTATGGATGTTTCTCAGAGA
TGGGTTGCAGCTGGAAGGGC
GTCCATTGTGCAAAACATAT

Analysis ↓ ↑ Design

Analysis ↓ ↑ Design