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

Jameson
CFD
Oxford

Caltech 1998

Princeton 1993
DNA replication by protein machines
protein synthesis by RNA and protein machines
cargo transport by protein motors
DNA Genome

...GTGGTACAGGT
GAATTTGGGTAGGC
TAAATTGTCCATAG
TTTATGTGTGTGAA
TGAGGATGATAGGA
TGTTTCTCAGAGAT
GGGTTGCAGCTGGA
AGGGCGTCCATTGT
GCAAAAACATATGCT
GGAGAAGTTGCCGG
TTCATTCTGCTGTG
GCGACCCCAGATTA
ATAAAAGGACTAAG
CCGAAAAAGAAAAATG
AAACATATATATAT
ATATATATATATAT
ATATATATATATA...

Regulatory Circuitry

Zebrafish Development

Movie: Sean Megason (Harvard)
Life is orchestrated by programmable biomolecules

- **DNA**
  GTGGTACAGGTGAATTTGGGTAGGCTAAATTGTCCATA

- **RNA**
  GGGCUGUUUUUUCUCGCUGACUUUCAGCCCCAAACAAAA

- **Protein**
  MTYRLELNGKTLKLGTRTRTEAVDAATAEKFQYANDNG

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
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?

\[ \Delta G(\phi, s) \]
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)

Metastable hairpins

Initiator starts chain reaction

Hairpins propagate chain reaction

Amplification polymer

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

Fly embryo

Chick embryo
(with Sauka-Spengler Lab, Oxford)

Mouse embryo
(with Lansford and Fraser Labs, USC)

Zebrafish embryo
mRNAs map to specific cells within the developing zebrafish brain
Analysis of a Complex

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

Equilibrium probability:

\[ Q(\phi) \quad P(\phi) \]

Complexity: \( \Theta(N^3) \)

Analysis of a Test Tube

- arbitrary complex species \( \Psi \)
- convex programming
- Lagrange duality

Complexity: \( O(\|\Psi\|N_{\text{max}}^3) \)

Complex Design

Target structure

Objective Function

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

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:**

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

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:**

\[ c_{\text{des}}(N) \geq \frac{4}{3} c_{\text{eval}}(N) \]

Randomly generated target structures
Typically <1% of nucleotides incorrectly paired at equilibrium
Design Cost

Typical design times ranging from fraction of second to seconds
As complex size increases, typical cost of design decreases to 4/3 cost of analysis!

Empirically, algorithm exhibits asymptotic optimality

Complex design does not take into consideration:
- concentration of desired complex
- concentration of competing off-target complexes
Test Tube Design

Target test tube

Design

Test tube ensemble

Analysis

Equilibrium probability

0.0 0.2 0.4 0.6 0.8 1.0

On-target complexes:
- target structure
- target concentration

Off-target complexes:
- vanishing target concentration


Include 100s to 1000s of off-target complexes
Typical design cost 2x-10x analysis cost
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
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

Visits: 100,000
Screen minutes: 1,300,000
Page views: 1,700,000

Molecular Instruments

molecularinstruments.org

Provide researchers with programmable molecular instruments for readout and regulation

Custom molecular instruments synthesized for over 70 labs