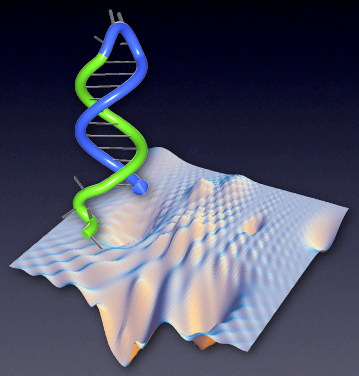
Analysis of Coarse-Grained Nucleic Acid Free Energy Landscapes



Niles Pierce
Erik Winfree
Caltech

Hideo Mabuchi
Stanford

Bernard Yurke
Boise State

August 4, 2008
Montreal

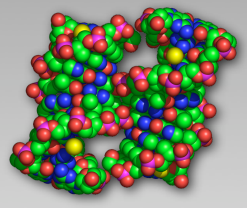
Research Goals

Develop algorithms to map coarse-grained nucleic acid free energy landscapes – identify the kinetic and equilibrium properties that dominate experimental observables

Computational and experimental analysis of free energy landscapes with metastable states

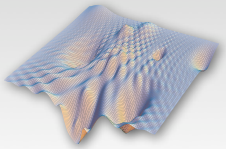
Free Energy Landscape

Tertiary Structure



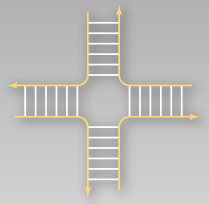
List of atomic coordinates and bonds

Landscape



Continuous representation

Secondary Structure



List of base pairs

Landscape

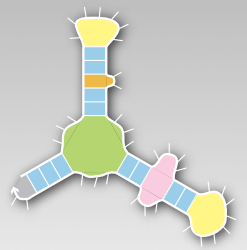
Graph representation

Each node is a secondary structure
Each edge is an elementary kinetic step

Thermodynamic Model

Free Energy

Empirical loop-based free energy model for secondary structure s



$$\Delta G(s) = \sum_{\text{loop} \in s} \Delta G(\text{loop})$$

Turner and SantaLucia

Partition Function

The partition function over the ensemble of secondary structures Ω

$$Q = \sum_{s \in \Omega} e^{-\Delta G(s)/kT}$$

can be used to calculate the equilibrium probability of any secondary structure $s \in \Omega$

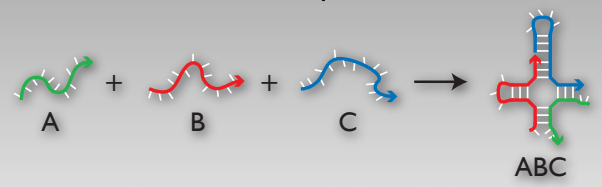
$$p(s) = \frac{1}{Q} e^{-\Delta G(s)/kT}$$

$|\Omega|$ grows exponentially with strand length

Zuker and McCaskill

Project 1: Thermodynamic Analysis of Interacting Strands

Partition function for a complex



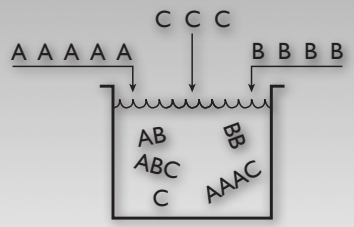
New algorithm for calculating the partition function of an unpseudoknotted complex of multiple interacting strands

Rigorous treatment of representation and distinguishability issues that do not arise in the single-stranded case

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

Project 1: Thermodynamic Analysis of Interacting Strands

Concentration determination in the thermodynamic limit

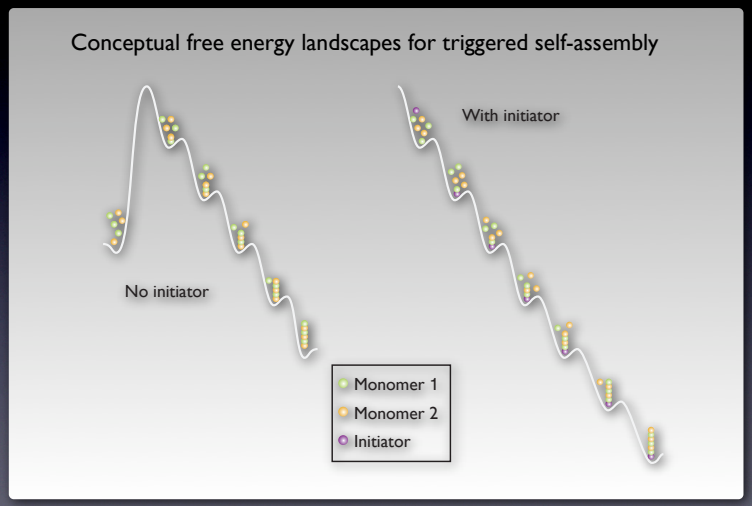


For large systems (e.g. a test tube), the equilibrium concentration of each species of complex can be obtained by solving a convex programming problem

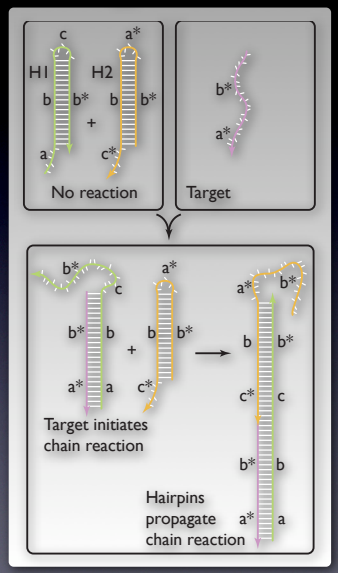
Efficient solution of the (high-dimensional) primal problem is achieved via solution of the (low-dimensional) dual problem

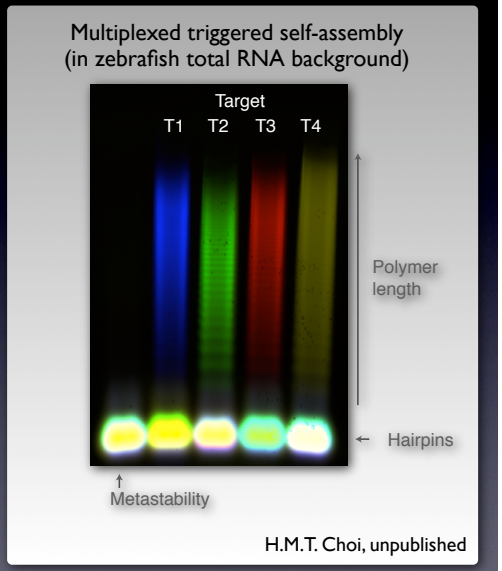
R.M. Dirks, J.S. Bois, J.M. Schaeffer, E. Winfree, N.A. Pierce, *SIAM Rev*, 2007

Project 2: Probing a free energy landscape with metastable states



Triggered self-assembly of metastable DNA hairpins





R.M. Dirks, N.A. Pierce, *Proc Natl Acad Sci USA*, 2004

Probing hairpin metastability

Are the hairpins metastable in the absence of initiator?

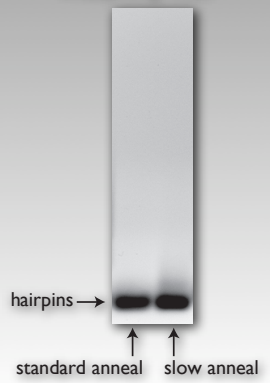
Experiment: anneal hairpins without initiator to see if they polymerize

No polymerization observed with 36-hour anneal

Competing hypotheses:

- Hairpins are metastable and the landscape is difficult to navigate via annealing
- Hairpins are stable

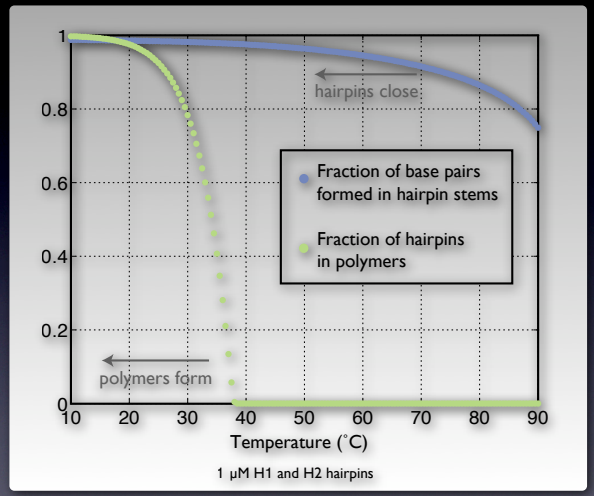
A surprising result



Standard anneal: 90s at 90°C
cool on bench top
Slow anneal: 95°C - 20°C
cool 30 min/°C

V. Beck, J.S. Bois, R.M. Dirks, N.A. Pierce, in preparation

Predicted annealing behavior



As cooling proceeds, hairpins are predicted to close before polymerization becomes energetically favorable

Tunneling through the landscape

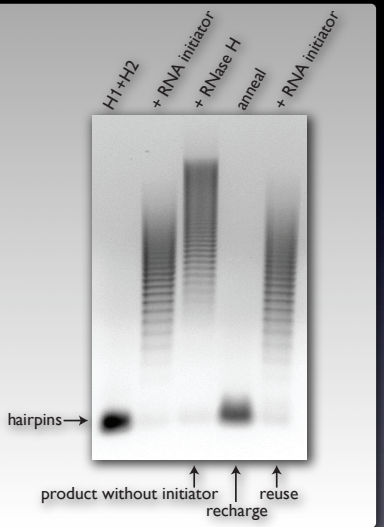
Trigger polymerization using an RNA initiator and then digest to obtain products with no bound initiator

Do they break up into hairpins?

No – the polymers increase in length

Supports the hypothesis that hairpins are metastable

Annealing recharges the hairpins



Kinetic Model

Model the time-varying probability $p_i(t)$ of secondary structure i via the master equation

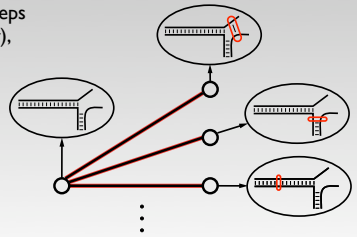
$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{|\Omega|} \left[k_{j \to i} p_j(t) - k_{i \to j} p_i(t) \right]$$

N.G. van Kampen, 1992

Rate constants are positive for elementary steps (e.g., addition or deletion of a single base pair), zero otherwise

Recall, $|\Omega|$ grows exponentially with strand length (too many states to write down)

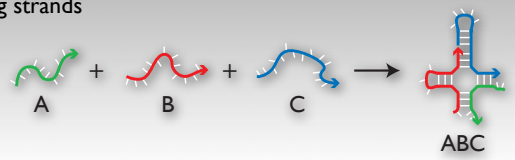
Simulate continuous-time Markov trajectories – generate move set on the fly based on current state



D.T. Gillespie, *J Comput Phys*, 1976
C. Flamm, W. Fontana, I.L. Hofacker, P. Schuster, *RNA*, 2000

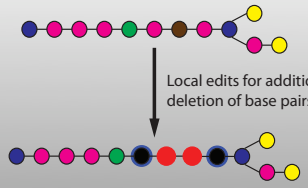
Project 3: Fast Stochastic Kinetics for Interacting Strands

New algorithm for stochastic kinetic simulations of unpseudoknotted complexes of interacting strands

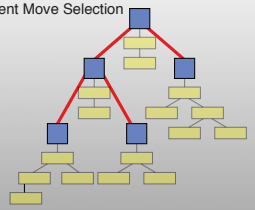


Lower complexity formulation with efficient data structures

Structures stored as loop tree

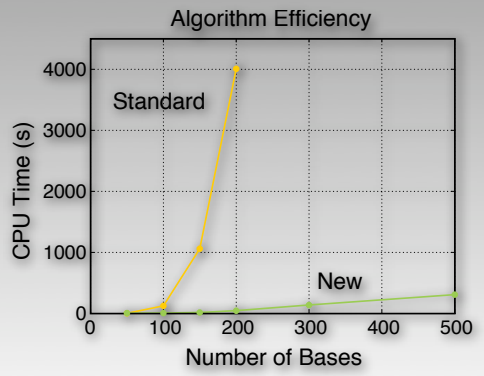


Efficient Move Selection



J.M. Schaeffer, J.S. Bois, R.M. Dirks, N.A. Pierce, E. Winfree, in preparation

Project 3: Fast Stochastic Kinetics for Interacting Strands



Fixed simulated time for random sequences from unstructured initial conditions

Project 4: Trajectory-Based Coarse-Graining

Goal

Discover a macrostate master equation to identify landscape features that dominate experimental measurements

Difficulty

The landscape is too large to enumerate the secondary structures – cannot write down the 'microstate' model

Approach

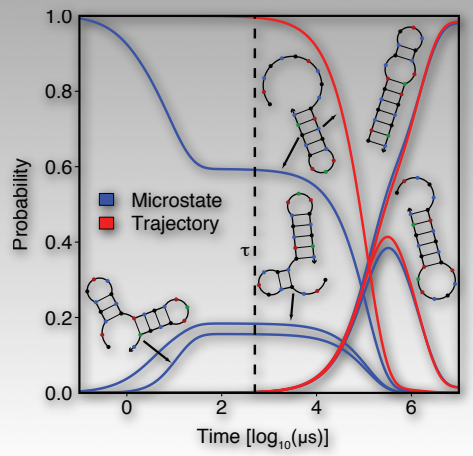
Use stochastic trajectories to explore the landscape (from specified initial conditions)

Identify transitions between macrostates by comparing occupancy distributions along trajectories (for a specified relaxation time scale τ)

Estimate macrostate partition functions and transition rates (with confidence intervals) using trajectories

J.A. Othmer and N.A. Pierce, in preparation

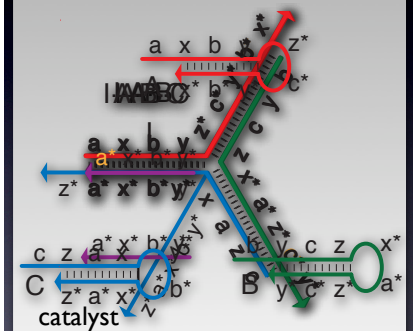
Testing Trajectory-Based Coarse-Graining

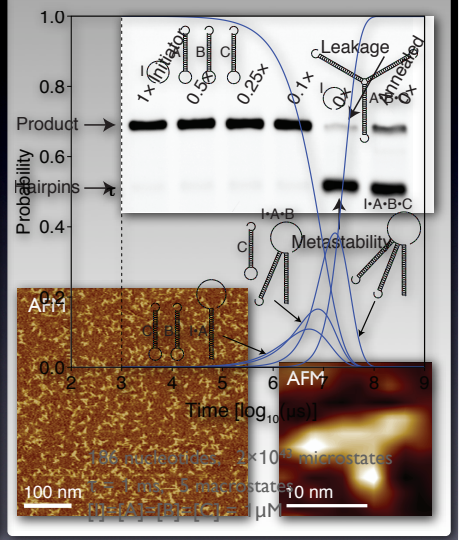


23 nucleotides
2×10^4 microstates
τ = 0.5 ms
3 macrostates

Trajectory-based coarse-graining example

Catalytic self-assembly of branched junctions





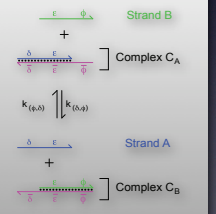
P.Yin, H. M.T. Choi, C.R. Calvert, N.A. Pierce, *Nature*, 2008

Project 5: Quantitative characterization of barrier states in catalytic pathways

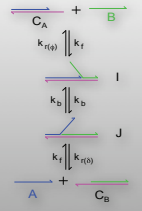
Toehold mediated strand displacement and toehold exchange are fundamental mechanisms in the design of DNA reaction pathways

Toehold sequences control reaction rates by determining barrier state energetics

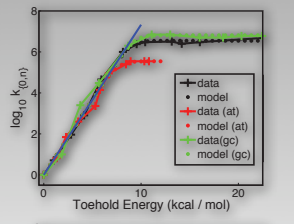
Experimentally measurable net reaction rates

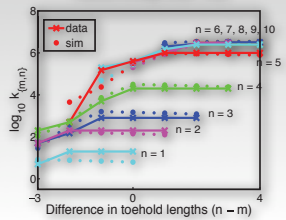


Unified macrostate model across toehold length and sequence composition



Remarkable agreement of experiment and theory





D.Y. Zhang, E. Winfree, in preparation

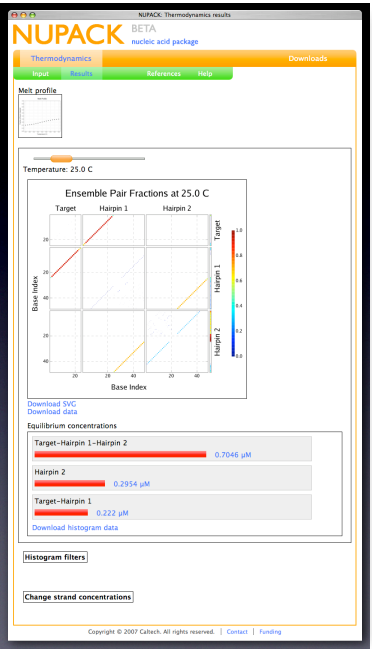
Project 6: NUPACK Nucleic Acid Package

www.nupack.org

A growing software suite for the analysis and design of nucleic acid systems

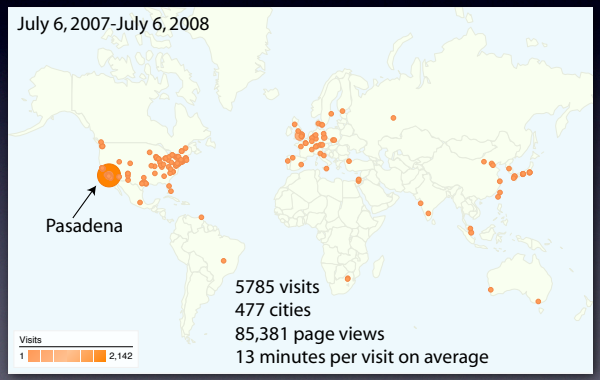
Web-based interface for efficient hypothesis testing and easy access

Downloadable source code for customization and batch processing

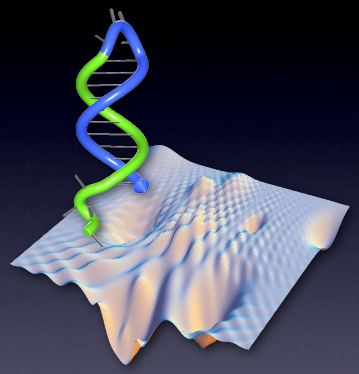


Project 6: NUPACK Nucleic Acid Package

www.nupack.org



J.N. Zadeh, J.S. Bois, C. Steenberg, M.B. Pierce, R.M. Dirks, J.A. Othmer and N.A. Pierce, in preparation



Niles Pierce
Erik Winfree
Caltech

Hideo Mabuchi
Stanford

Bernard Yurke
Boise State

nupack.org
Software

NSF
Funding