

On the Use of Coarse-Grained Thermodynamic Landscapes to Efficiently Estimate Folding Kinetics for RNA Molecules

Evan Senter

2015

Outline

Overview

The really big picture

ncRNA background

Kinetic justification

The road to Hermes

Kinetics with Hermes

About me

My background

- ▶ B.A. in Computer Science, Computational Biology
- ▶ Worked in software engineering for ≈ 2 years after
- ▶ Started at Boston College in Fall, 2011
- ▶ Joined the Clote Lab focusing on Computational RNA Biology



University of California, Santa Barbara

Goal of this talk

Primary aim

Present research on rapidly estimating RNA folding kinetics *in silico*

1. Motivate interest in the study of RNA
2. Highlight interesting roles of non-coding RNAs (ncRNA)
3. Identify biological relevance of folding kinetics
4. Present overview of findings
5. Explain research leading to these findings

What's the takeaway?

- ▶ A thesis?...

What's the takeaway?

- ▶ A thesis?...

How biologists see bioinformagicians



A biologist when stumbling into a math-heavy talk...



What we aim for...



Outline

Overview

The really big picture

ncRNA background

Kinetic justification

The road to Hermes

Kinetics with Hermes

Organization of Hermes

Hermes

Collection of kinetics algorithms based on transition matrices derived from energy landscapes

Thermodynamics-Based Approach

FFTbor2D

2D Energy Landscape from input RNA
and Starting / Target Structures

Kinetics-Based Approach

RNAmfpt

Average Folding Time from A to B,
from input Probability Matrix

RNAeq

Population Proportion and Equilibrium
Time from input Probability Matrix

FFTmfpt

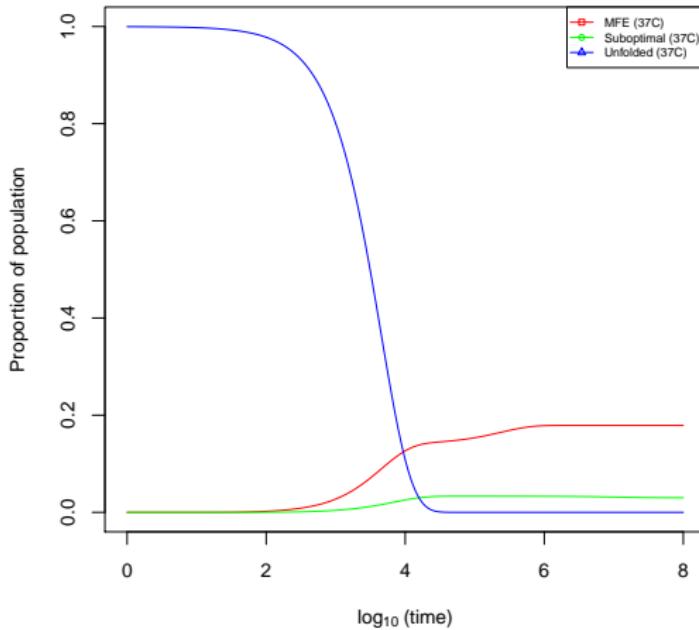
Mean First Passage Time from input RNA
and Starting / Target Structures

FFTeq

Population Proportion / Equilibrium Time
from input RNA and Starting / Target
Structures

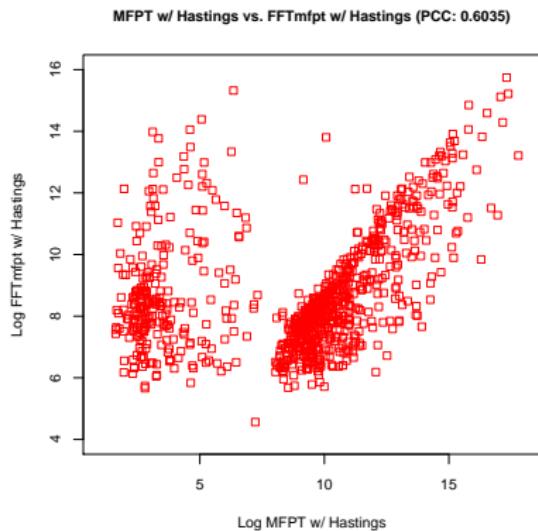
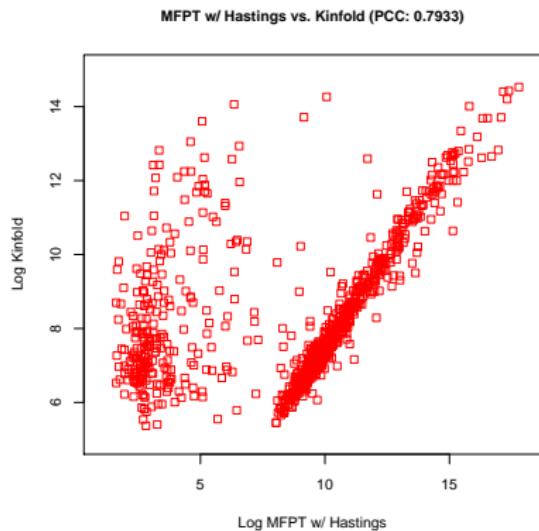
Population occupancy curves

FFTbor2D folding kinetics for *L. collosoma* spliced leader RNA



Population occupancy curves computed with FFTeq for the 56 nt conformational switch *L. collosoma* spliced leader RNA

How approximate are the kinetics?



Comparison of various kinetics programs

Hastings (Yes\No)	RNAmfpt	RNAeq	Kinfold	FFTmfpt	RNA2Dfold	FFTbor	BarriersEq	FFTeq
RNAmfpt	1	0.5683	0.7945	0.5060	0.5110	0.5204	0.5280	0.4472
RNAeq	0.5798	1	0.7814	0.7043	0.7025	0.5080	0.5979	0.6820
Kinfold	0.7933	0.7507	1	0.7312	0.7358	0.6241	0.6328	0.6445
FFTmfpt	0.6035	0.7935	0.7608	1	0.9980	0.5485	0.8614	0.9589
RNA2Dfold	0.6076	0.7919	0.7655	0.9983	1	0.5584	0.8538	0.9515
FFTbor	0.5416	0.5218	0.6241	0.5748	0.5855	1	0.3450	0.4229
BarriersEq	0.6346	0.6578	0.6328	0.8310	0.8217	0.3450	1	0.9149
FFTeq	0.5614	0.7916	0.6966	0.9670	0.9590	0.4757	0.8940	1

- ▶ RNAmfpt, FFTmfpt, RNAeq, and FFTeq included in the **Hermes** package
- ▶ RNA2Dfold (Lorenz *et. al.*, 2009), BarriersEq (Flamm *et. al.*, 2002), and FFTbor (Senter *et. al.*, 2012) kinetics computed with **Hermes**

Outline

Overview

The really big picture

ncRNA background

Kinetic justification

The road to Hermes

Kinetics with Hermes

Why do we care about RNA?

NATURE VOL. 227 AUGUST 8 1970

- ▶ Phrase ‘junk DNA’ pigeonholed RNA into predetermined roles
 - ▶ Messenger RNA (mRNA)
 - ▶ Transfer RNA (tRNA)
 - ▶ Ribosomal RNA (rRNA)
- ▶ Diverse roles for ncRNA beyond tRNA and rRNA

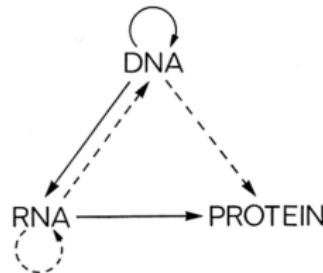


Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

Crick, F. (1970). Central dogma of molecular biology. Nature.

Why do we care about RNA?

NATURE VOL. 227 AUGUST 8 1970

- ▶ Phrase ‘junk DNA’ pigeonholed RNA into predetermined roles
 - ▶ Messenger RNA (mRNA)
 - ▶ Transfer RNA (tRNA)
 - ▶ Ribosomal RNA (rRNA)
- ▶ Diverse roles for ncRNA beyond tRNA and rRNA

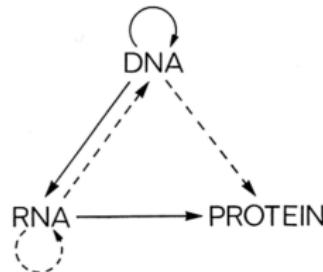


Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

Crick, F. (1970). Central dogma of molecular biology. Nature.

Why do we care about RNA?

NATURE VOL. 227 AUGUST 8 1970

- ▶ Phrase ‘junk DNA’ pigeonholed RNA into predetermined roles
 - ▶ Messenger RNA (mRNA)
 - ▶ Transfer RNA (tRNA)
 - ▶ Ribosomal RNA (rRNA)
- ▶ Diverse roles for ncRNA beyond tRNA and rRNA

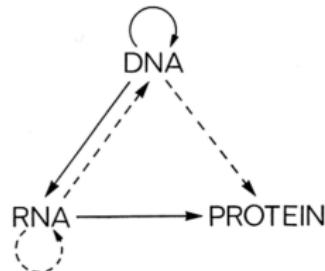


Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

Crick, F. (1970). Central dogma of molecular biology. Nature.

ncRNAs—what are they good for?

The reality

We were not wrong in assigning importance to the aforementioned roles of RNA, but...

ncRNAs—what are they good for?

We have since found a diverse set of roles for RNA, including...

- ▶ Peptide bond catalysis

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science (New York, N.Y.)*, 289(5481), 920–930.

- ▶ Intron splicing

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., et al. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science (New York, N.Y.)*, 273(5282), 1678–1685.

- ▶ Post-transcriptional gene regulation via RNAi

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811.

ncRNAs—what are they good for?

We have since found a diverse set of roles for RNA, including...

- ▶ Peptide bond catalysis

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science (New York, N.Y.)*, 289(5481), 920–930.

- ▶ Intron splicing

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., et al. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science (New York, N.Y.)*, 273(5282), 1678–1685.

- ▶ Post-transcriptional gene regulation via RNAi

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811.

ncRNAs—what are they good for?

We have since found a diverse set of roles for RNA, including...

- ▶ Peptide bond catalysis

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science (New York, N.Y.)*, 289(5481), 920–930.

- ▶ Intron splicing

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., et al. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science (New York, N.Y.)*, 273(5282), 1678–1685.

- ▶ Post-transcriptional gene regulation via RNAi

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811.

ncRNAs—what are they good for?

We have since found a diverse set of roles for RNA, including...

- ▶ Peptide bond catalysis

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science (New York, N.Y.)*, 289(5481), 920–930.

- ▶ Intron splicing

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., et al. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science (New York, N.Y.)*, 273(5282), 1678–1685.

- ▶ Post-transcriptional gene regulation via RNAi

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811.

ncRNAs—what are they good for?

- ▶ Xist ncRNA for suppression of inactive X chromosome

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137.

- ▶ Retranslation events (SECIS elements)

Walczak, R., Westhof, E., Carbon, P., & Krol, A. (1996). A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *Rna*, 2(4), 367–379.

- ▶ Ribosomal frameshift events

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331(6153), 280–283.

Ofori, L. O., Hilimire, T. A., Bennett, R. P., Brown, N. W., Smith, H. C., & Miller, B. L. (2014). High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting in vitro and interferes with HIV-1 infectivity. *Journal of Medicinal Chemistry*, 57(3), 723–732.

ncRNAs—what are they good for?

- ▶ Xist ncRNA for suppression of inactive X chromosome

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137.

- ▶ Retranslation events (SECIS elements)

Walczak, R., Westhof, E., Carbon, P., & Krol, A. (1996). A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *Rna*, 2(4), 367–379.

- ▶ Ribosomal frameshift events

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331(6153), 280–283.

Ofori, L. O., Hilimire, T. A., Bennett, R. P., Brown, N. W., Smith, H. C., & Miller, B. L. (2014). High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting in vitro and interferes with HIV-1 infectivity. *Journal of Medicinal Chemistry*, 57(3), 723–732.

ncRNAs—what are they good for?

- ▶ Xist ncRNA for suppression of inactive X chromosome

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137.

- ▶ Retranslation events (SECIS elements)

Walczak, R., Westhof, E., Carbon, P., & Krol, A. (1996). A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *Rna*, 2(4), 367–379.

- ▶ Ribosomal frameshift events

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331(6153), 280–283.

Ofori, L. O., Hilimire, T. A., Bennett, R. P., Brown, N. W., Smith, H. C., & Miller, B. L. (2014). High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting in vitro and interferes with HIV-1 infectivity. *Journal of Medicinal Chemistry*, 57(3), 723–732.

ncRNAs—what are they good for?

And finally...

- ▶ Transcriptional and translational regulation via riboswitches

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., & Breaker, R. R. (2002). Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology*, 9(9), 1043–1049.

- ▶ Spliced leader (SL) trans-splicing events in *L. collosoma*

LeCuyer, K. A., & Crothers, D. M. (1993). The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry*, 32(20), 5301–5311.

- ▶ hok/sok postsegregational killing mechanism in *E. coli*

Gerdes, K., Rasmussen, P. B., & Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences*, 83(10), 3116–3120.

Nagel, J. H., Gulyaev, A. P., Gerdes, K., & Pleij, C. W. (1999). Metastable structures and refolding kinetics in hok mRNA of plasmid R1. *Rna*, 5(11), 1408–1418.

ncRNAs—what are they good for?

And finally...

- ▶ Transcriptional and translational regulation via riboswitches

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., & Breaker, R. R. (2002). Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology*, 9(9), 1043–1049.

- ▶ Spliced leader (SL) trans-splicing events in *L. collosoma*

LeCuyer, K. A., & Crothers, D. M. (1993). The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry*, 32(20), 5301–5311.

- ▶ hok/sok postsegregational killing mechanism in *E. coli*

Gerdes, K., Rasmussen, P. B., & Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences*, 83(10), 3116–3120.

Nagel, J. H., Gulyaev, A. P., Gerdes, K., & Pleij, C. W. (1999). Metastable structures and refolding kinetics in hok mRNA of plasmid R1. *Rna*, 5(11), 1408–1418.

ncRNAs—what are they good for?

And finally...

- ▶ Transcriptional and translational regulation via riboswitches

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., & Breaker, R. R. (2002). Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology*, 9(9), 1043–1049.

- ▶ Spliced leader (SL) trans-splicing events in *L. collosoma*

LeCuyer, K. A., & Crothers, D. M. (1993). The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry*, 32(20), 5301–5311.

- ▶ hok/sok postsegregational killing mechanism in *E. coli*

Gerdes, K., Rasmussen, P. B., & Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences*, 83(10), 3116–3120.

Nagel, J. H., Gulyaev, A. P., Gerdes, K., & Pleij, C. W. (1999). Metastable structures and refolding kinetics in hok mRNA of plasmid R1. *Rna*, 5(11), 1408–1418.

ncRNAs—what are they good for?

Summary

ncRNAs have diverse cellular responsibilities, beyond the canonical tRNA and rRNA examples

Outline

Overview

The really big picture

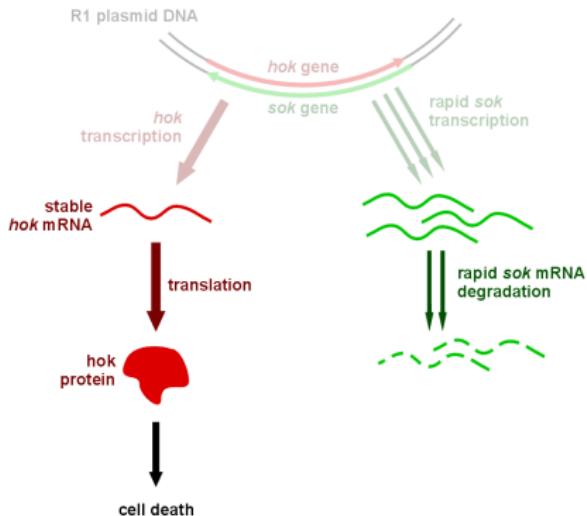
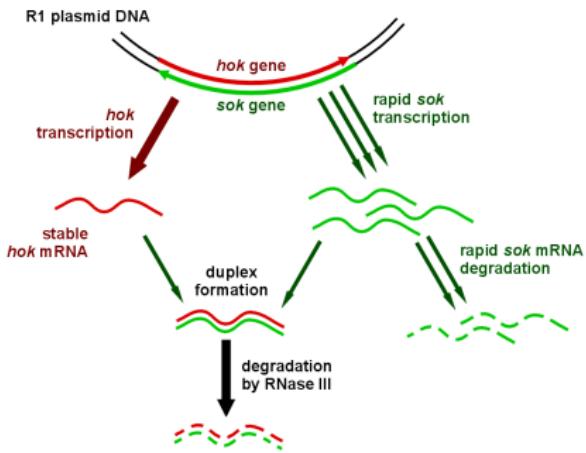
ncRNA background

Kinetic justification

The road to Hermes

Kinetics with Hermes

hok/sok and kinetics



https://en.wikipedia.org/wiki/File:Hok_sok_system_R1_plasmid_present.gif
https://en.wikipedia.org/wiki/File:Hok_sok_system_R1_plasmid_absent.gif

hok folding kinetics

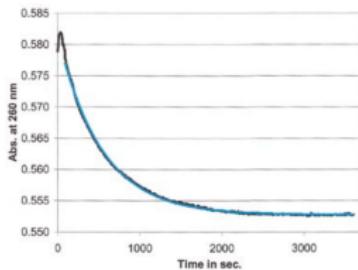
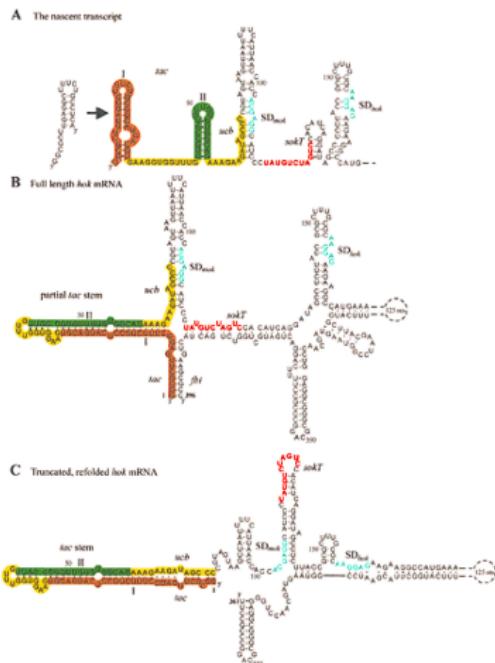


Figure 2. The kinetic refolding from the metastable to the stable conformation in the *hok*³⁴ RNA fragment. The metastable conformation was trapped by the heating/cooling cycle and the kinetics monitored at 260 nm in a UV spectrophotometer at 37°C in 950 mM NaCl and 50 mM Na cacodylate buffer, pH 7.2. The measured real-time curve (in black) and the first-order exponentially fitted curve is indicated in light blue (fitted parameters $t_{1/2} = 669 \pm 1$ s).

Nagel, J. H. A., Gulyaev, A. P., Oistämö, K. J., Gerdes, K., & Pleij, C. W. A. (2002). A pH-jump approach for investigating secondary structure refolding kinetics in RNA. *Nucleic Acids Research*, 30(13), e63.

Adapted from Nagel, J. H., Gulyaev, A. P., Gerdes, K., & Pleij, C. W. (1999). Metastable structures and refolding kinetics in *hok* mRNA of plasmid R1. *RNA*, 5(11), 1408–1418.

How are RNA kinetics experimentally measured?

Experimental protocols include...

- ▶ Temperature-jump experiments

Cole, P. E., & Crothers, D. M. (1972). Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, 11(23), 4368–4374.

- ▶ pH-jump experiments

Bina-Stein, M., & Crothers, D. M. (1974). Conformational changes of transfer ribonucleic acid. The pH phase diagram under acidic conditions. *Biochemistry*, 13(13), 2771–2775.

- ▶ Single molecule mechanical tension

Vieregg, J. R., & Tinoco, I. (2006). Modelling RNA folding under mechanical tension. *Molecular Physics*, 104(8), 1343–1352.

- ▶ Fluorescence resonance energy transfer (FRET)

Zhuang, X., & Rief, M. (2003). Single-molecule folding. *Current Opinion in Structural Biology*, 13(1), 88–97.

How are RNA kinetics experimentally measured?

Experimental protocols include...

- ▶ Temperature-jump experiments

Cole, P. E., & Crothers, D. M. (1972). Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, 11(23), 4368–4374.

- ▶ pH-jump experiments

Bina-Stein, M., & Crothers, D. M. (1974). Conformational changes of transfer ribonucleic acid. The pH phase diagram under acidic conditions. *Biochemistry*, 13(13), 2771–2775.

- ▶ Single molecule mechanical tension

Vieregg, J. R., & Tinoco, I. (2006). Modelling RNA folding under mechanical tension. *Molecular Physics*, 104(8), 1343–1352.

- ▶ Fluorescence resonance energy transfer (FRET)

Zhuang, X., & Rief, M. (2003). Single-molecule folding. *Current Opinion in Structural Biology*, 13(1), 88–97.

How are RNA kinetics experimentally measured?

Experimental protocols include...

- ▶ Temperature-jump experiments

Cole, P. E., & Crothers, D. M. (1972). Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, 11(23), 4368–4374.

- ▶ pH-jump experiments

Bina-Stein, M., & Crothers, D. M. (1974). Conformational changes of transfer ribonucleic acid. The pH phase diagram under acidic conditions. *Biochemistry*, 13(13), 2771–2775.

- ▶ Single molecule mechanical tension

Vieregg, J. R., & Tinoco, I. (2006). Modelling RNA folding under mechanical tension. *Molecular Physics*, 104(8), 1343–1352.

- ▶ Fluorescence resonance energy transfer (FRET)

Zhuang, X., & Rief, M. (2003). Single-molecule folding. *Current Opinion in Structural Biology*, 13(1), 88–97.

How are RNA kinetics experimentally measured?

Experimental protocols include...

- ▶ Temperature-jump experiments

Cole, P. E., & Crothers, D. M. (1972). Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry*, 11(23), 4368–4374.

- ▶ pH-jump experiments

Bina-Stein, M., & Crothers, D. M. (1974). Conformational changes of transfer ribonucleic acid. The pH phase diagram under acidic conditions. *Biochemistry*, 13(13), 2771–2775.

- ▶ Single molecule mechanical tension

Vieregg, J. R., & Tinoco, I. (2006). Modelling RNA folding under mechanical tension. *Molecular Physics*, 104(8), 1343–1352.

- ▶ Fluorescence resonance energy transfer (FRET)

Zhuang, X., & Rief, M. (2003). Single-molecule folding. *Current Opinion in Structural Biology*, 13(1), 88–97.

Outline

Overview

The really big picture

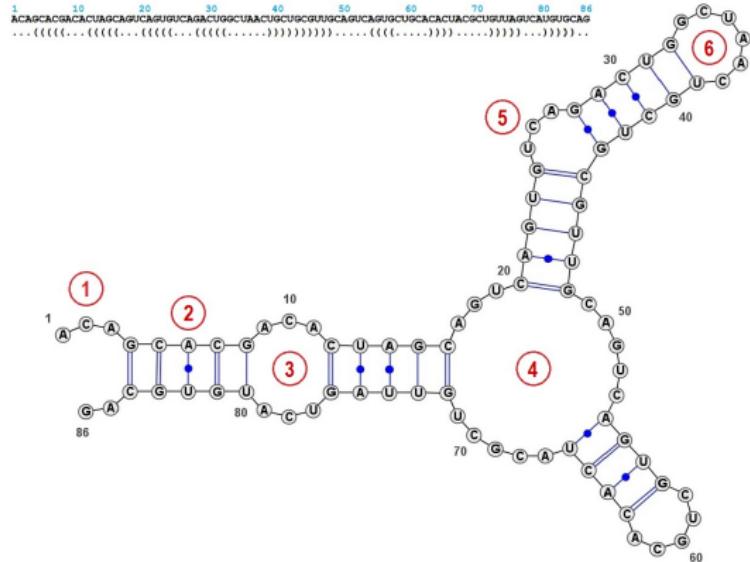
ncRNA background

Kinetic justification

The road to Hermes

Kinetics with Hermes

Structural motifs

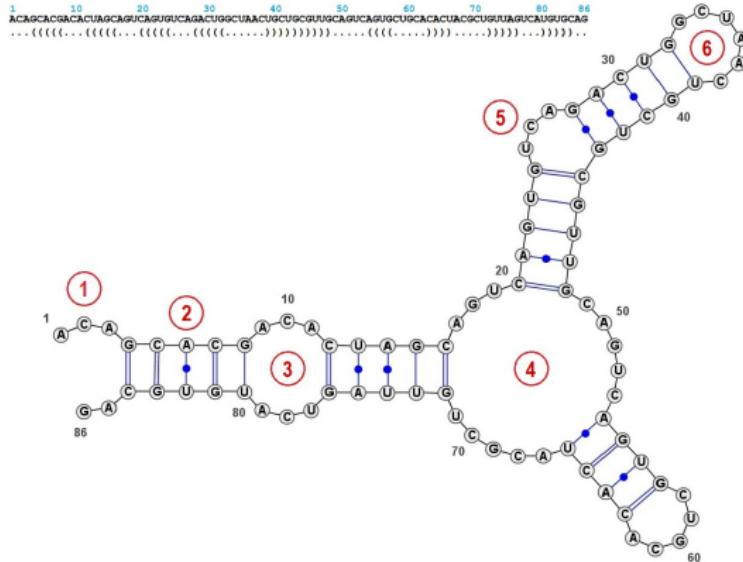


Structural Motifs

1. Exterior loop
2. Stack
3. Interior loop
4. Multiloop
5. Bulge
6. Hairpin

Lu, X.-J., Bussemaker, H. J., & Olson, W. K. (2015). DSSR: an integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Research.

Secondary structure of RNA



Symmetric distance

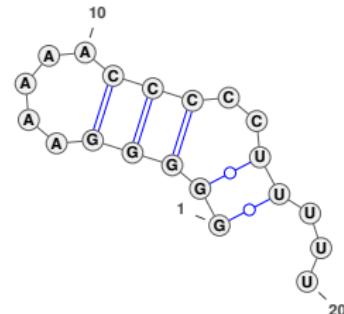
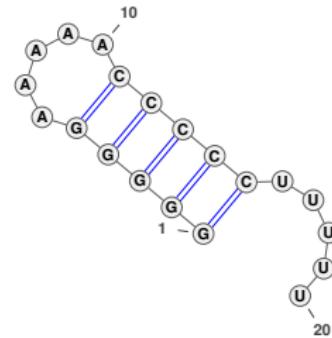
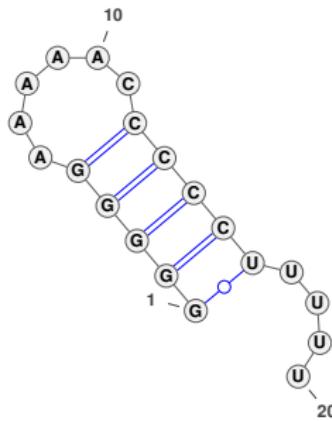
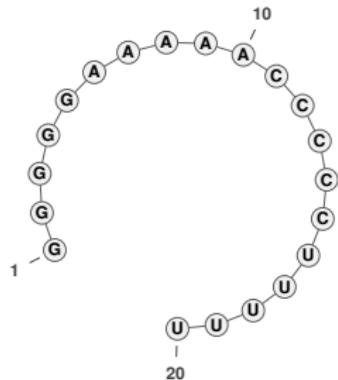
$$d_{BP}(\mathcal{S}, \mathcal{T}) = |\mathcal{S} \cup \mathcal{T}| - |\mathcal{S} \cap \mathcal{T}|$$

Distance between
two structures

$$\begin{aligned} d_{BP}(\mathcal{S}_{[i,j]}, \mathcal{T}_{[i,j]}) = \\ |\{(x,y) : i \leq x < y \leq j, \\ (x,y) \in \mathcal{S} - \mathcal{T} \text{ or} \\ (x,y) \in \mathcal{T} - \mathcal{S}\}| \end{aligned}$$

Lu, X.-J., Bussemaker, H. J., & Olson, W. K. (2015). DSSR: an integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Research.

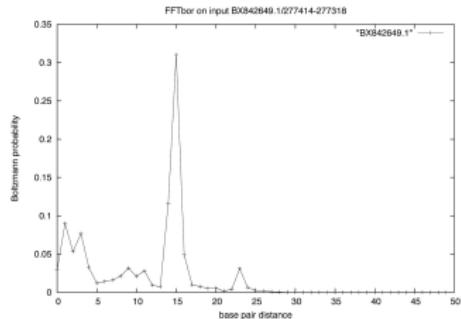
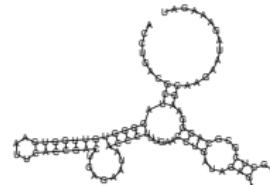
Base pair distance example



FFTbor

Objective

FFTbor computes for each integer k , the Boltzmann probability $p_k = Z^k / Z$ of the subensemble of structures whose base pair distance to a given *initial* structure S^* is k



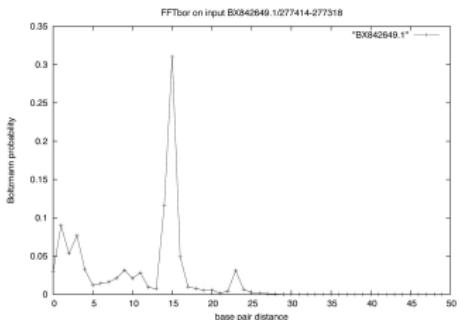
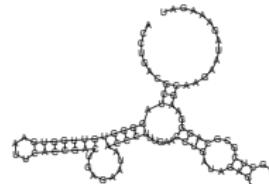
(Top) MFE structure for the 97 nt thiamine pyrophosphate (TPP) riboswitch BX842649.1 277414–277318. (Bottom) BX842649.1 277414–277318 appears to have a rugged energy landscape near its minimum free energy structure, with distinct low energy structures that may compete with the MFE structure during the folding process

FFTbor

$$p(k) = \frac{\mathbf{Z}^k}{\mathbf{Z}} \text{ for } 0 \leq k \leq n$$

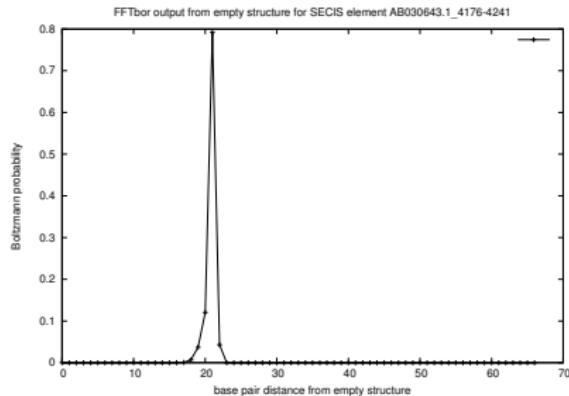
$$\mathbf{Z}^k = \sum_{\substack{\mathcal{S} \text{ such that} \\ d_{BP}(\mathcal{S}, \mathcal{S}^*)=k}} e^{\frac{-E(\mathcal{S})}{RT}}$$

$$\mathbf{Z} = \sum_{k=0}^n \mathbf{Z}^k$$

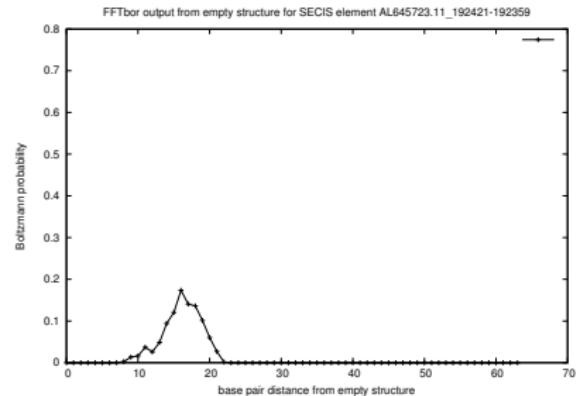


(Top) MFE structure for the 97 nt thiamine pyrophosphate (TPP) riboswitch BX842649.1 277414–277318. (Bottom) BX842649.1 277414–277318 appears to have a rugged energy landscape near its minimum free energy structure, with distinct low energy structures that may compete with the MFE structure during the folding process

Examples of correlation with SECIS



For SECIS element AB030643.1 4176–4241, the standard deviation σ of the FFTbor distribution is 0.7276, while the logarithm base 10 of the mean first passage time is 4.75



For SECIS element AL645723.11 192421–192359, the standard deviation σ of the FFTbor distribution is 2.6794, while the logarithm base 10 of the mean first passage time is 5.69

Analysis of distributions generated by FFTbor

	μ	σ	σ/μ	n	MFЕ	$\log_{10}(\text{MFPT})$
μ	1					
σ	-0.4372	1				
σ/μ	-0.6914	0.9437	1			
n	0.7077	-0.1590	-0.3646	1		
MFЕ	-0.5695	0.7395	0.7596	-0.3685	1	
$\log_{10}(\text{MFPT})$	-0.0363	0.4844	0.3762	0.4059	0.3990	1

Pearson correlation between various aspects of selenocysteine insertion sequences from the seed alignment of Rfam family RF00031. For each of the 61 RNA sequences, we ran FFTbor, starting from empty initial structure S^* , and we ran a Monte Carlo folding algorithm, developed by E. Freyhult and P. Clote (unpublished).

FFTbor vs. RNAbor: what changed?

Where did performance gain come from?

Rather than compute values of \mathbf{Z}^k for $0 \leq k \leq n$ to generate $p(k) = \mathbf{z}^k / \mathbf{z}$, noting that $\mathbf{Z} = \sum_{k=0}^n \mathbf{Z}^k$
 $O(n^5)$ speed, $O(n^3)$ space

$$\mathbf{Z}_{i,j}^k = \mathbf{Z}_{i,j-1}^{k-b_0} + \sum_{\substack{(s_r, s_j) \in \mathbb{B}, \\ i \leq r < j}} \sum_{w+w'=k-b(r)} \exp(-E_0(r, j)/RT) \cdot \mathbf{Z}_{i,r-1}^w \mathbf{Z}_{r+1,j-1}^{w'}$$

$b_0 = 1$ if j is base-paired in $\mathcal{S}_{[i,j]}^*$ and 0 otherwise

$$b(r) = d_{\text{BP}}(\mathcal{S}_{[i,j]}^*, \mathcal{S}_{[i,r-1]}^* \cup \mathcal{S}_{[r+1,j-1]}^* \cup \{(r, j)\})$$

FFTbor vs. RNAbor: what changed?

Where did performance gain come from?

Rather than compute values of \mathbf{Z}^k for $0 \leq k \leq n$ to generate $p(k) = \mathbf{z}^k / \mathbf{z}$, noting that $\mathbf{Z} = \sum_{k=0}^n \mathbf{Z}^k$
 $O(n^5)$ speed, $O(n^3)$ space

$$\mathbf{Z}_{i,j}^k = \mathbf{Z}_{i,j-1}^{k-b_0} + \sum_{\substack{(s_r, s_j) \in \mathbb{B}, \\ i \leq r < j}} \sum_{w+w'=k-b(r)} \exp(-E_0(r, j)/RT) \cdot \mathbf{Z}_{i,r-1}^w \mathbf{Z}_{r+1,j-1}^{w'}$$

$b_0 = 1$ if j is base-paired in $\mathcal{S}_{[i,j]}^*$ and 0 otherwise

$$b(r) = d_{BP}(\mathcal{S}_{[i,j]}^*, \mathcal{S}_{[i,r-1]}^* \cup \mathcal{S}_{[r+1,j-1]}^* \cup \{(r, j)\})$$

FFTbor vs. RNAbor: what changed?

Where did performance gain come from?

Evaluate the polynomial $\mathcal{Z}(x)$ for values $e^{2\pi i k/n}$, where $0 \leq k \leq n$ to generate y_0, y_1, \dots, y_n

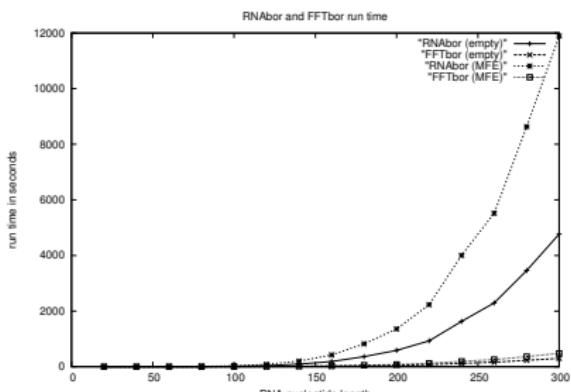
Use polynomial interpolation to find the coefficients of the polynomial $\mathcal{Z}(x) = \sum_{k=0}^n z_k x^k$, where $z_k = \mathbf{Z}^k$
 $O(n^4)$ speed, $O(n^2)$ space

$$\mathcal{Z}_{i,j}(x) = \mathcal{Z}_{i,j-1}(x) \cdot x^{b_0} + \sum_{\substack{(s_r, s_j) \in \mathbb{B}, \\ i \leq r < j}} \left(e^{\frac{-E_0(r,j)}{RT}} \cdot \mathcal{Z}_{i,r-1}(x) \cdot \mathcal{Z}_{r+1,j-1}(x) \cdot x^{b(r)} \right)$$

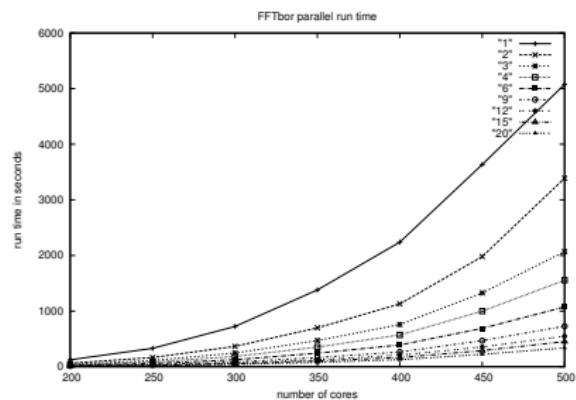
$b_0 = 1$ if j is base-paired in $\mathcal{S}_{[i,j]}^*$ and 0 otherwise

$$b(r) = d_{BP}(\mathcal{S}_{[i,j]}^*, \mathcal{S}_{[i,r-1]}^* \cup \mathcal{S}_{[r+1,j-1]}^* \cup \{(r,j)\})$$

Performance characteristics



RNAbor vs. FFTbor benchmarking



Parallelized FFTbor benchmarking

Where does this take us?

MFPT and σ

Correlation between folding time and σ suggests the coarse-grained landscape can be used to approximate kinetics

FFTbor2D

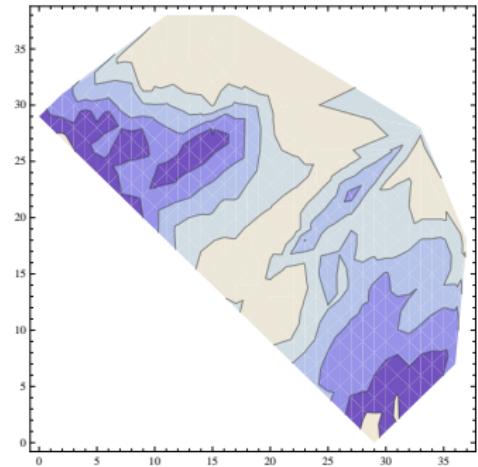
$$\mathcal{Z}_{i,j-1}(x) \cdot x^{\alpha_0 n + \beta_0} + \\ \sum_{\substack{(s_k, s_j) \in \mathbb{B}, \\ i \leq k < j}} \left(e^{\frac{-E_0(k,j)}{RT}} \cdot \mathcal{Z}_{i,k-1}(x) \cdot \mathcal{Z}_{k+1,j-1}(x) \cdot x^{\alpha(k)n + \beta(k)} \right)$$

$\alpha_0 = 1$ if j is base-paired in $\mathcal{A}_{[i,j]}$, 0 otherwise

$\beta_0 = 1$ if j is base-paired in $\mathcal{B}_{[i,j]}$, 0 otherwise

$\alpha(k) = d_{BP}(\mathcal{A}_{[i,j]}, \mathcal{A}_{[i,k-1]} \cup \mathcal{A}_{[k+1,j-1]} \cup \{(k,j)\})$

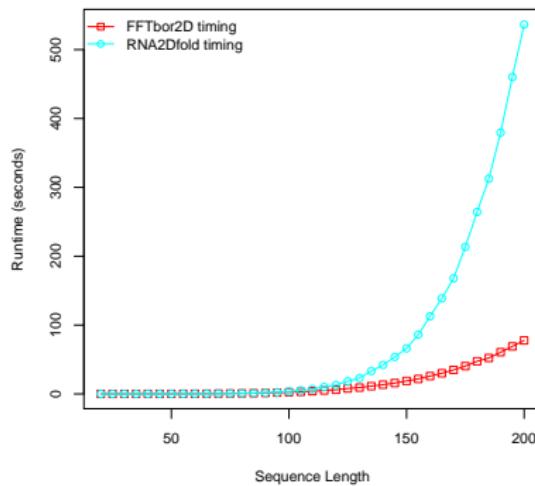
$\beta(k) = d_{BP}(\mathcal{B}_{[i,j]}, \mathcal{B}_{[i,k-1]} \cup \mathcal{B}_{[k+1,j-1]} \cup \{(k,j)\})$



2D projection of energy landscape for SL RNA from *L. collosoma*, as computed by FFTbor2D

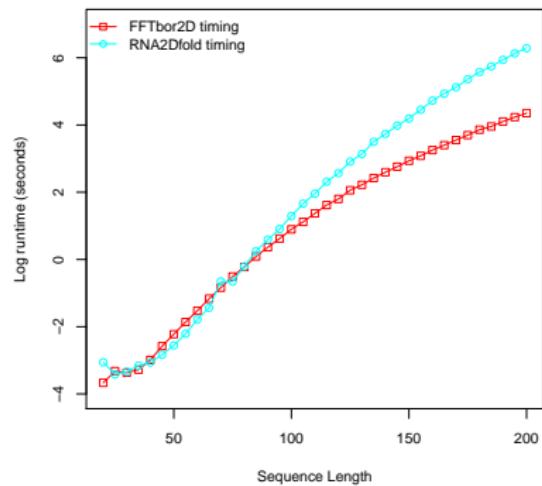
Performance characteristics

Time benchmarking (each point is the average of 100 sequences)



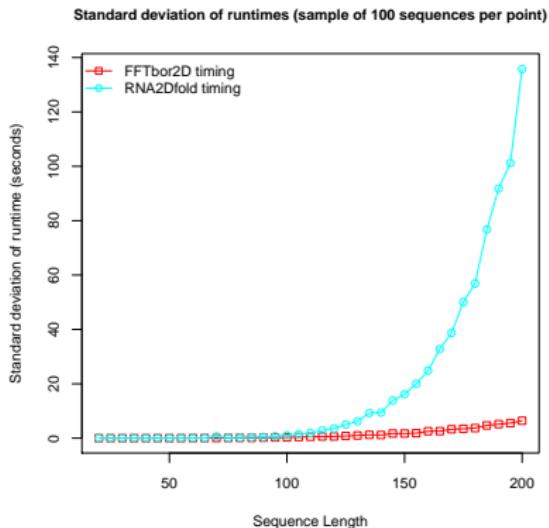
FFTbor2D vs. RNA2Dfold benchmarking

Time benchmarking (each point is the log average of 100 sequences)



FFTbor2D vs. RNA2Dfold benchmarking (log scale)

Performance characteristics



- ▶ Approach using FFT goes from $O(n^7)$ to $O(n^5)$
- ▶ We observe a real performance gain in line with 100x speedup
- ▶ Memory requirements drop from $O(n^4)$ to $O(n^2)$
- ▶ More consistent performance characteristics

Outline

Overview

The really big picture

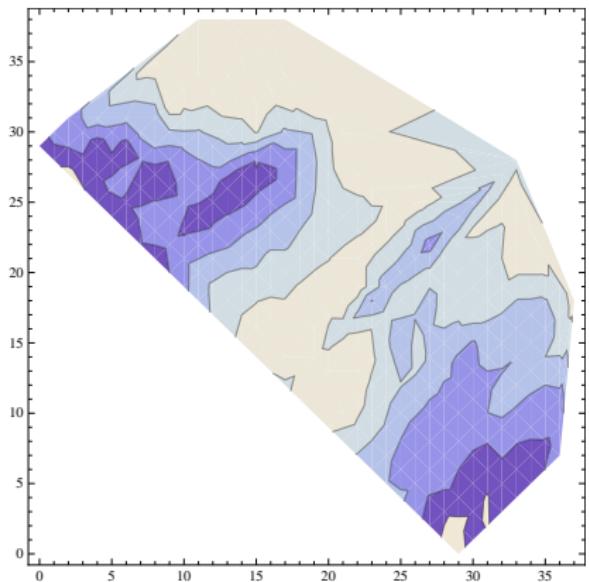
ncRNA background

Kinetic justification

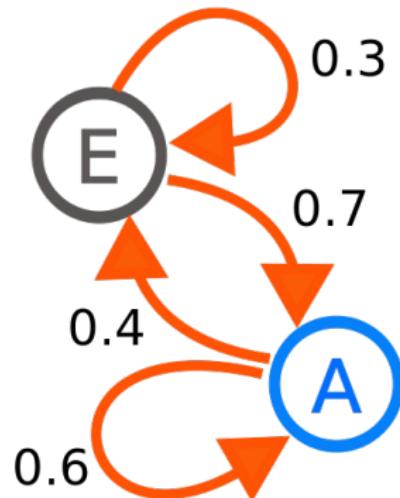
The road to Hermes

Kinetics with Hermes

Using FFTbor2D as input for Hermes



2D projection of energy landscape for SL RNA from *L. collosoma*, as computed by FFTbor2D



"Markovkate 01" by Joxemai4 - Own work. Licensed under CC BY-SA 3.0 via Commons

Organization of Hermes

Hermes

Collection of kinetics algorithms based on transition matrices derived from energy landscapes

Thermodynamics-Based Approach

FFTbor2D

2D Energy Landscape from input RNA
and Starting / Target Structures

Kinetics-Based Approach

RNAmfpt

Average Folding Time from A to B,
from input Probability Matrix

RNAeq

Population Proportion and Equilibrium
Time from input Probability Matrix

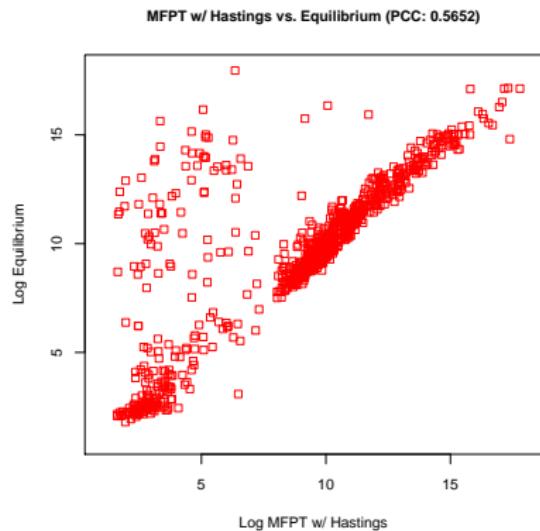
FFTmfpt

Mean First Passage Time from input RNA
and Starting / Target Structures

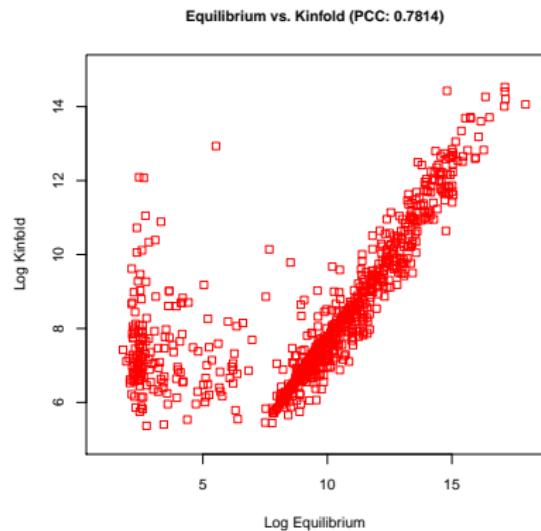
FFTeq

Population Proportion / Equilibrium Time
from input RNA and Starting / Target
Structures

Comparison of various kinetics programs

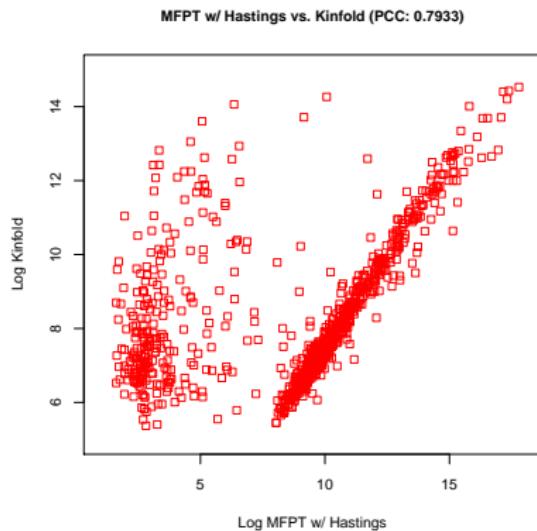


Scatter plots of the natural logarithm of times from RNAmfpt versus RNAeq

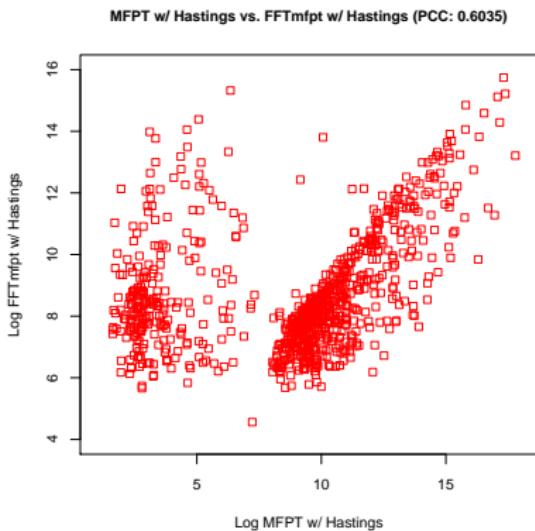


Scatter plots of the natural logarithm of times from RNAeq versus Kinfold

Comparison of various kinetics programs

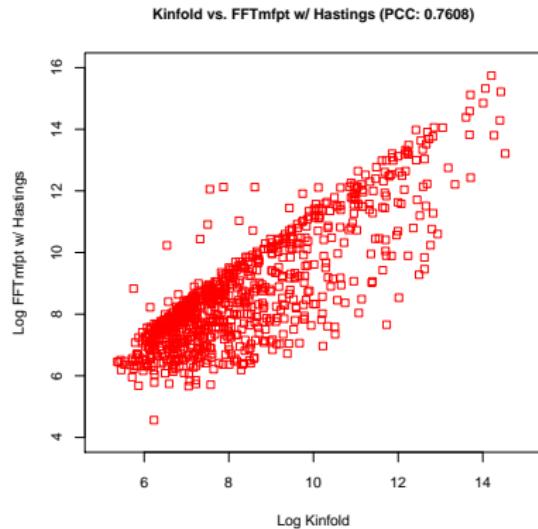


Scatter plots of the natural logarithm of times from RNAmfpt versus RNAeq

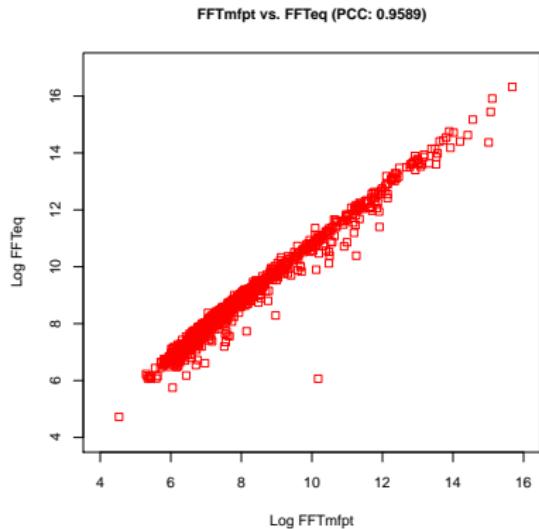


Scatter plots of the natural logarithm of times from RNAmfpt versus FFTmfpt

Comparison of various kinetics programs



Scatter plots of the natural logarithm of times from Kinfold versus FFTmfpt



Scatter plots of the natural logarithm of times from FFTmfpt versus FFTeq

Comparison of various kinetics programs

Hastings (Yes\No)	RNAmfpt	RNAeq	Kinfold	FFTmfpt	RNA2Dfold	FFTbor	BarriersEq	FFTeq
RNAmfpt	1	0.5683	0.7945	0.5060	0.5110	0.5204	0.5280	0.4472
RNAeq	0.5798	1	0.7814	0.7043	0.7025	0.5080	0.5979	0.6820
Kinfold	0.7933	0.7507	1	0.7312	0.7358	0.6241	0.6328	0.6445
FFTmfpt	0.6035	0.7935	0.7608	1	0.9980	0.5485	0.8614	0.9589
RNA2Dfold	0.6076	0.7919	0.7655	0.9983	1	0.5584	0.8538	0.9515
FFTbor	0.5416	0.5218	0.6241	0.5748	0.5855	1	0.3450	0.4229
BarriersEq	0.6346	0.6578	0.6328	0.8310	0.8217	0.3450	1	0.9149
FFTeq	0.5614	0.7916	0.6966	0.9670	0.9590	0.4757	0.8940	1

- ▶ RNAmfpt, FFTmfpt, RNAeq, and FFTeq included in the **Hermes** package
- ▶ RNA2Dfold (Lorenz *et. al.*, 2009), BarriersEq (Flamm *et. al.*, 2002), and FFTbor (Senter *et. al.*, 2012) kinetics computed with **Hermes**

The big picture

What's the big picture?

- ▶ **Hermes** presents a collection of tools that allow one to quickly estimate RNA kinetics
- ▶ Deterministic approach (doesn't use sampling)
- ▶ Performance is in a class of its own thanks to FFTbor2D

The big picture

What's the big picture?

- ▶ **Hermes** presents a collection of tools that allow one to quickly estimate RNA kinetics
- ▶ Deterministic approach (doesn't use sampling)
- ▶ Performance is in a class of its own thanks to FFTbor2D

The big picture

What's the big picture?

- ▶ **Hermes** presents a collection of tools that allow one to quickly estimate RNA kinetics
- ▶ Deterministic approach (doesn't use sampling)
- ▶ Performance is in a class of its own thanks to FFTbor2D

Acknowledgements

- ▶ Prof. Peter Clote
- ▶ Committee members: Profs. Michelle Meyer, Welkin Johnson, Michael Zuker, and François Major
- ▶ Department faculty and administration (especially Peter Marino, I don't think I ever had to enroll in a class myself)
- ▶ Family and friends

And of course my labmates and fellow grad students!



Questions?

