Inferring Parameters for an Elementary Step Model of DNA Structure Kinetics with Locally Context-Dependent Arrhenius Rates

Sedigheh Zolaktaf[1], Frits Dannenberg[2], Xander Rudelis[2], Anne Condon[1], Joseph M Schaeffer[3], Mark Schmidt[1], Chris Thachuk[2], and Erik Winfree[2]

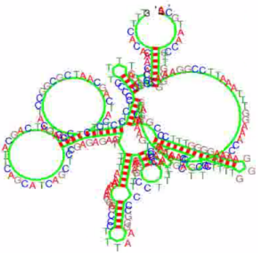
[1]University of British Columbia, Vancouver, BC, Canada

[2]California Institute of Technology, Pasadena, CA, USA

[3]Autodesk Research, San Francisco, CA, USA

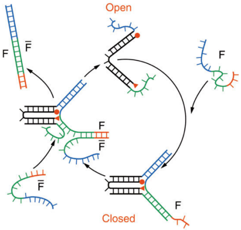
Nucleic Acid Kinetic Simulators

- Kinetic simulators aim to predict the kinetics of interacting nucleic acid strands
  - e.g., reaction rate constants.



Nucleic Acid Kinetic Simulators

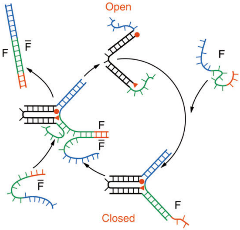
- Kinetic simulators aim to predict the kinetics of interacting nucleic acid strands
    - e.g., reaction rate constants.
- Useful for biological and biotechnological applications.



[Yurke et al., 2000]

Nucleic Acid Kinetic Simulators

- Kinetic simulators aim to predict the kinetics of interacting nucleic acid strands
  - e.g., reaction rate constants.
- Useful for biological and biotechnological applications.



[Yurke et al., 2000]

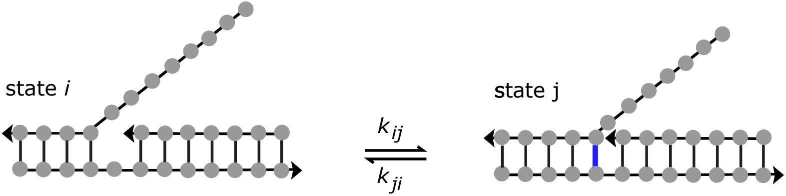
- Predicting kinetics is difficult (dependent on sequence, temperature, ...).
  - Accurate models of nucleic acid kinetics are required.

Contributions

- We introduce an Arrhenius kinetic model.
- We train kinetic models.
    - We collect a dataset of experimentally determined reaction rate constants.
    - We introduce a computational framework for predicting reaction rate constants.
- Our Arrhenius model performs better than an existing model.

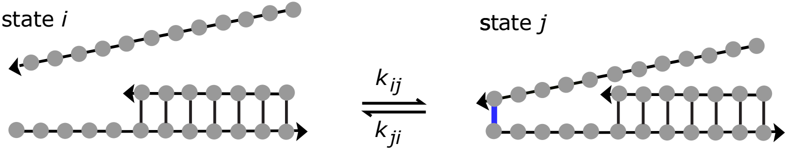
Modelling Kinetics of Interacting Nucleic Acid Strands

- Kinetics of interacting strands are modelled as continuous time Markov chains (CTMC) [Schaeffer et al., 2015].
  - The state space represents non-pseudoknotted secondary structures.



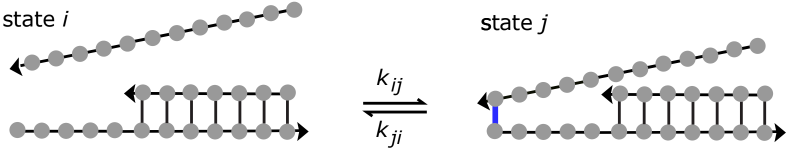
Modelling Kinetics of Interacting Nucleic Acid Strands

- Kinetics of interacting strands are modelled as continuous time Markov chains (CTMC) [Schaeffer et al., 2015].
    - The state space represents non-pseudoknotted secondary structures.



Modelling Kinetics of Interacting Nucleic Acid Strands

- Kinetics of interacting strands are modelled as continuous time Markov chains (CTMC) [Schaeffer et al., 2015].
  - The state space represents non-pseudoknotted secondary structures.



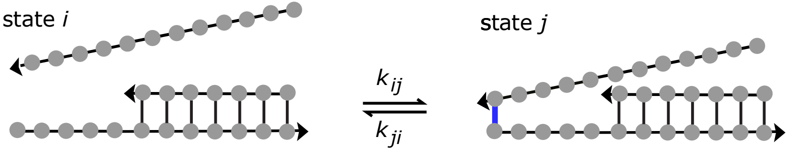
- The transition rates are determined by a kinetic model and obey detailed balance:

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G^{0}(j) - \Delta G^{0}(i)}{RT}}$$

$\Delta G^0(i)$: free energy of state i, R: gas constant, T: temperature.

Modelling Kinetics of Interacting Nucleic Acid Strands

- Kinetics of interacting strands are modelled as continuous time Markov chains (CTMC) [Schaeffer et al., 2015].
    - The state space represents non-pseudoknotted secondary structures.



- The transition rates are determined by a kinetic model and obey detailed balance:

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G^{\mathbf{0}}(j) - \Delta G^{\mathbf{0}}(i)}{RT}}$$

$\Delta G^0(i)$: free energy of state i, R: gas constant, T: temperature.

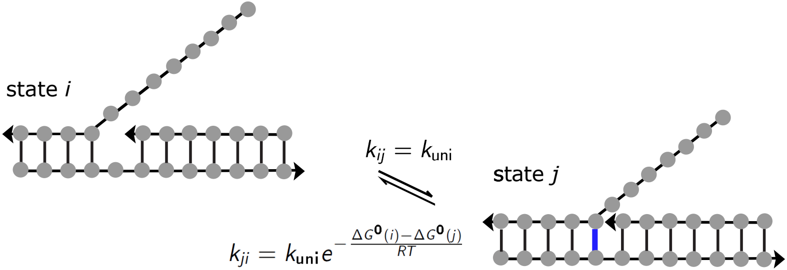
- Estimate reaction rate constants from mean first passage times (MFPTs).
    - The MFPT of a CTMC is the average time it takes to reach one of a set of final states from an initial state.

The Metropolis Kinetic Model

- A kinetic model is the Metropolis model [Schaeffer et al., 2015].

The Metropolis Kinetic Model

- A kinetic model is the Metropolis model [Schaeffer et al., 2015].
- For states i and j, if $\Delta G^0(i) > \Delta G^0(j)$, the unimolecular transition rates are:



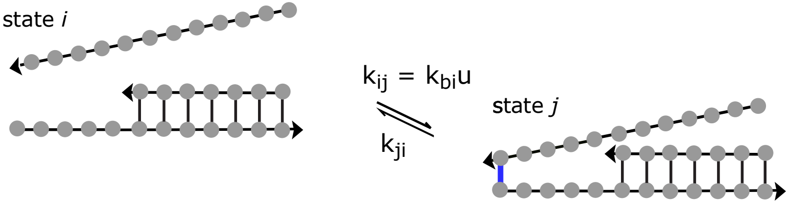
$$k_{ij} = k_{\mathbf{uni}}$$

$$k_{ji} = k_{\mathbf{uni}} e^{-\frac{\Delta G^{\mathbf{0}}(i) - \Delta G^{\mathbf{0}}(j)}{RT}}$$

- $k_{\mathbf{uni}} > 0$: unimolecular rate constant

The Metropolis Kinetic Model

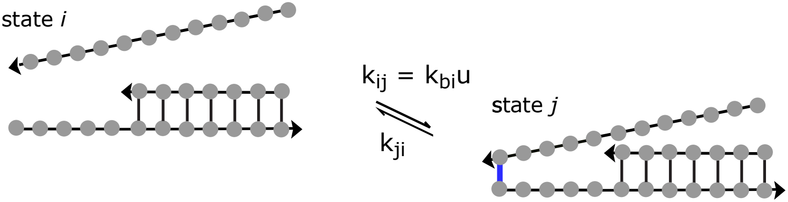
- A kinetic model is the Metropolis model [Schaeffer et al., 2015].
- Bimolecular transitions are given by:



- $k_{\text{bi}} > 0$: bimolecular rate constant
- u: initial concentration of the reactants

The Metropolis Kinetic Model

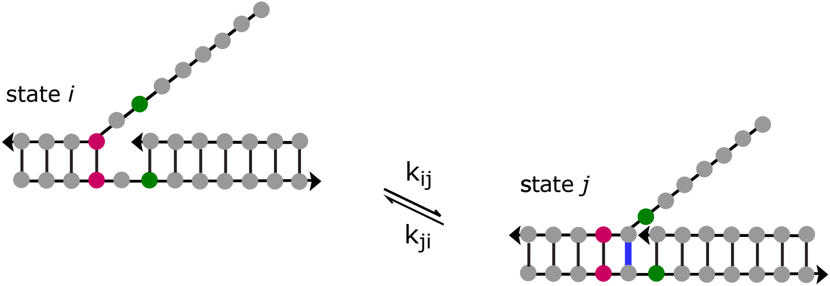
- A kinetic model is the Metropolis model [Schaeffer et al., 2015].
- Bimolecular transitions are given by:



  - $k_{bi} > 0$: bimolecular rate constant
  - u: initial concentration of the reactants
- The model predictions are off by several orders of magnitude when we train it with our framework.

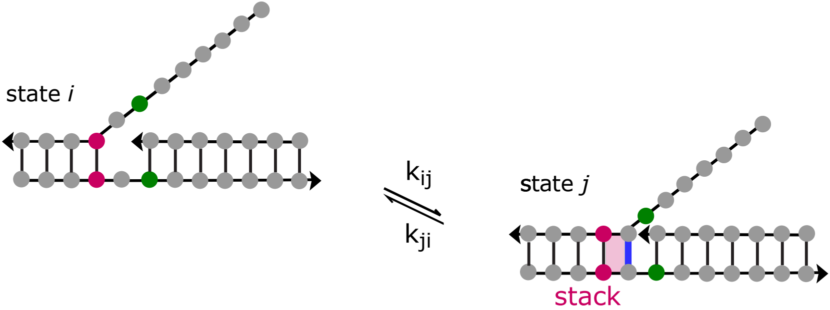
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



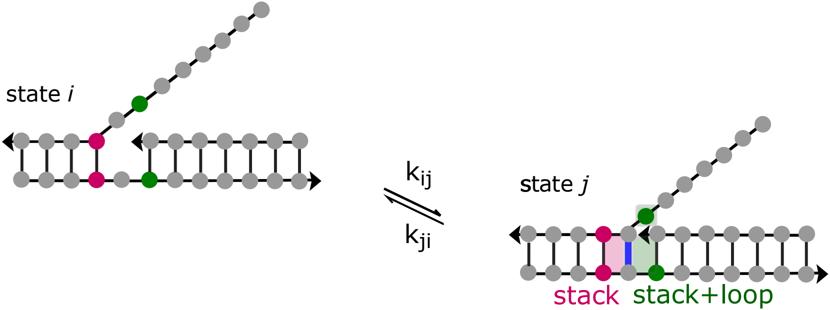
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



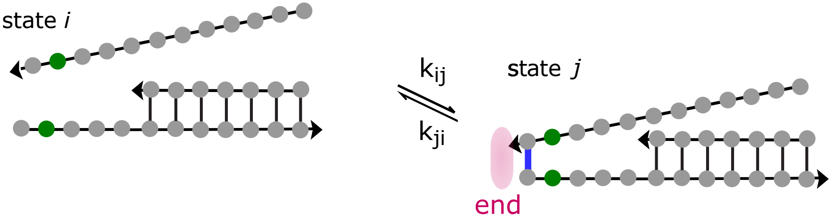
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



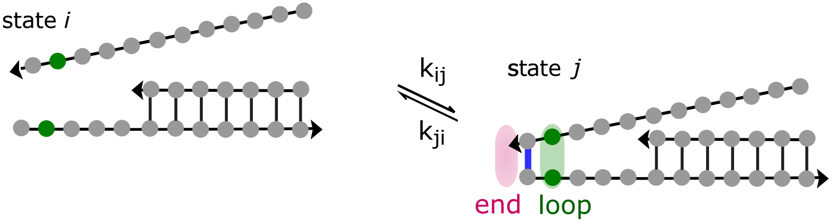
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



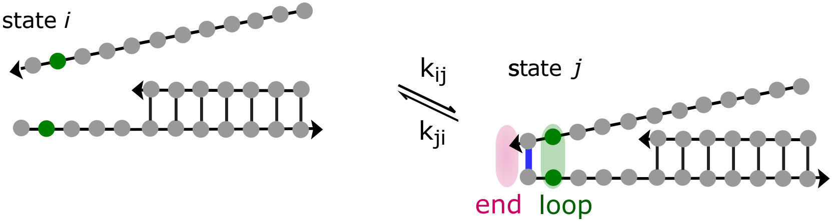
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on the pairing status of the bases immediately to **left** and the **right** side of the base pair forming or breaking.



- Our model differentiates between seven different half contexts:

$$\mathcal{C} = \{\text{stack}, \text{loop}, \text{end}, \text{stack+loop}, \ldots\}$$

The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
    - Transition rates depend on activation energy.

The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on activation energy.
- Based on the set of half contexts $\mathcal{C}$, we redefine $k_{\mathsf{uni}} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_{>0}$:

$$k_{\mathsf{uni}}(left, right) = A_{left} e^{-\frac{E_{left}}{RT}} A_{right} e^{-\frac{E_{right}}{RT}}$$

E_{left}, E_{right}: activation energies A_{left}, A_{right}: rate constants

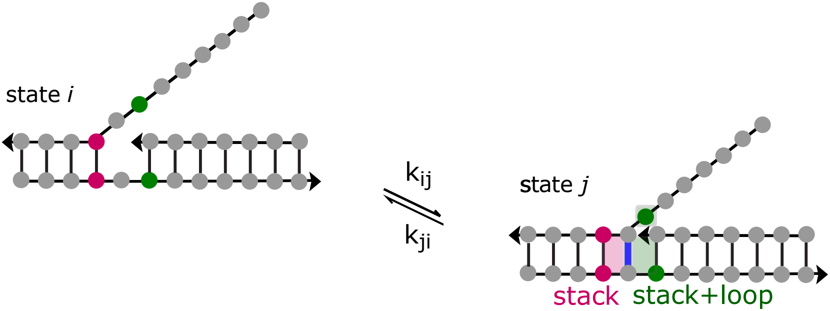
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on activation energy.
- Based on the set of half contexts $\mathcal{C}$, we redefine $k_{\textsf{uni}} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_{>0}$:

$$k_{\textsf{uni}}(\mathit{left}, \mathit{right}) = A_{\mathit{left}}\, e^{-\frac{E_{\mathit{left}}}{RT}}\, A_{\mathit{right}}\, e^{-\frac{E_{\mathit{right}}}{RT}}$$

E_{left}, E_{right}: activation energies A_{left}, A_{right}: rate constants

- $k_{\textsf{uni}}(\textsf{stack}, \textsf{stack+loop}) = A_{\textsf{stack}}\, e^{-\frac{E_{\textsf{stack}}}{RT}}\, A_{\textsf{stack+loop}}\, e^{-\frac{E_{\textsf{stack+loop}}}{RT}}$



The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on activation energy.
- We redefine $k_{\mathsf{bi}} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_{>0}$:

$$k_{\mathsf{bi}}(\mathit{left}, \mathit{right}) = \alpha k_{\mathsf{uni}}(\mathit{left}, \mathit{right})$$

α: bimolecular scaling constant

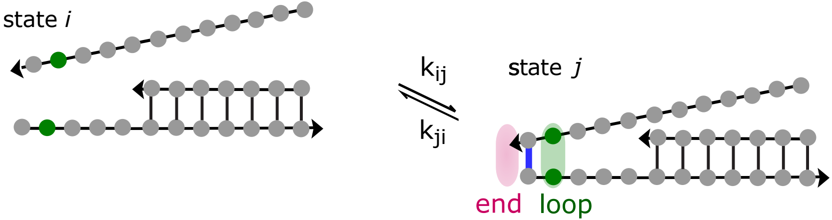
The Arrhenius Kinetic Model

- We introduce a new model that has locally context-dependent Arrhenius rates.
  - Transition rates depend on activation energy.
- We redefine $k_{\text{bi}} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_{>0}$:

$$k_{\text{bi}}(\textit{left}, \textit{right}) = \alpha k_{\text{uni}}(\textit{left}, \textit{right})$$

α: bimolecular scaling constant

- $k_{\text{bi}}(\text{end}, \text{loop}) = \alpha A_{\text{end}} e^{-E_{\text{end}}/RT} A_{\text{loop}} e^{-E_{\text{loop}}/RT}$

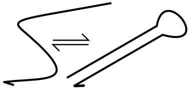


Dataset of Experimental Reaction Rate Constants

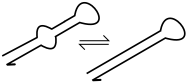
- $|\text{Training set}| = 320$, $|\text{Testing set}| = 56$

Dataset of Experimental Reaction Rate Constants

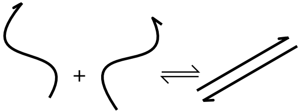
- |Training set| = 320, |Testing set| = 56



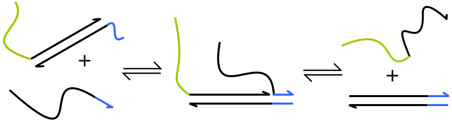
Hairpin closing and opening [Bonnet et al., 1998], [Bonnet, 2000], [Kim et al., 2006]



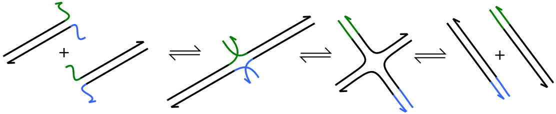
Bubble closing [Altan-Bonnet et al., 2003]



Helix association and dissociation [Morrison and Stols, 1993], [Reynaldo et al., 2000]

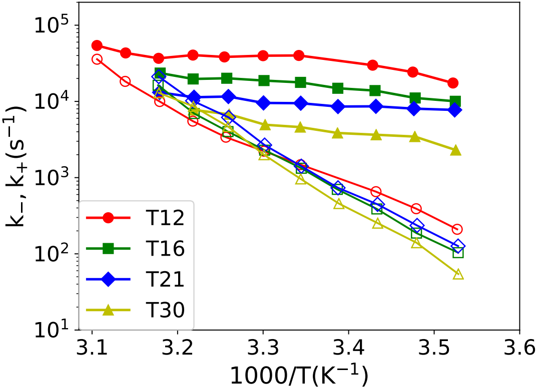


Toehold-mediated 3-way strand displacement [Reynaldo et al., 2000], [Zhang and Winfree, 2009], [Machinek et al., 2014]

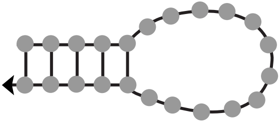


Toehold-mediated 4-way strand exchange [Dabby, 2013]

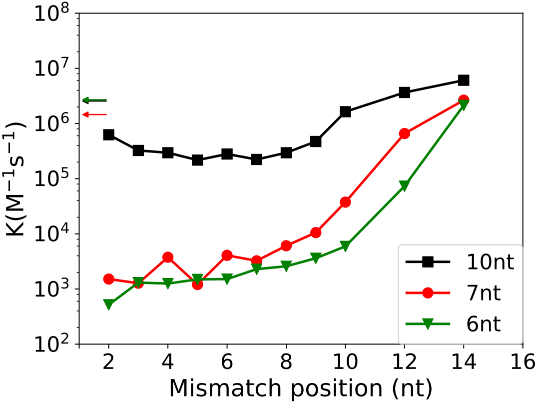
Example: Hairpin Closing and Opening

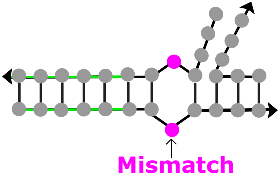


Hairpin closing (solid) and opening (open) [Bonnet et al., 1998]. The legend shows the hairpin loop length.



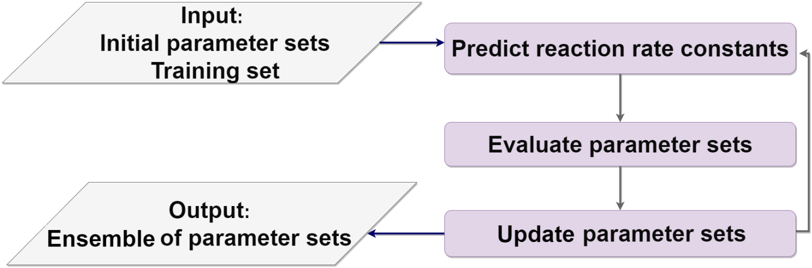
Example: Toehold-mediated 3-way Strand Displacement With Mismatches





Toehold-mediated 3-way strand displacement with mismatches [Machinek et al., 2014]. The legend shows the length of the toehold domain.

Framework Overview



Input:
Initial parameter sets
Training set

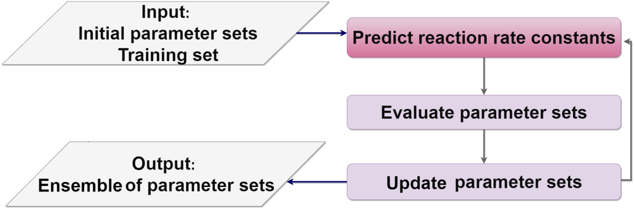
Predict reaction rate constants

Evaluate parameter sets

Update parameter sets

Output:
Ensemble of parameter sets

Predict Reaction Rate Constants



Input:
Initial parameter sets
Training set

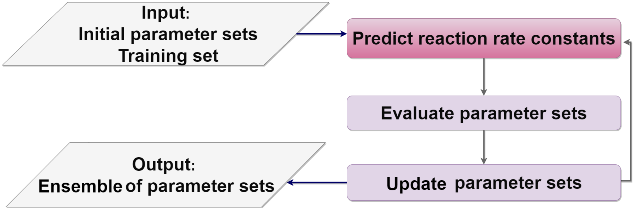
Predict reaction rate constants

Evaluate parameter sets

Update parameter sets

Output:
Ensemble of parameter sets

Predict Reaction Rate Constants



Input:
Initial parameter sets
Training set

Predict reaction rate constants

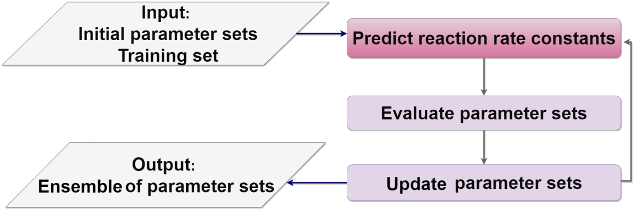
Evaluate parameter sets

Update parameter sets

Output:
Ensemble of parameter sets

- Instead of stochastic simulations, we estimate mean first passage times (MFPTs) with exact solvers [Suhov and Kelbert, 2008].

Predict Reaction Rate Constants



Input:
Initial parameter sets
Training set

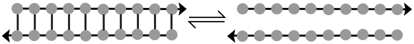
Predict reaction rate constants

Evaluate parameter sets

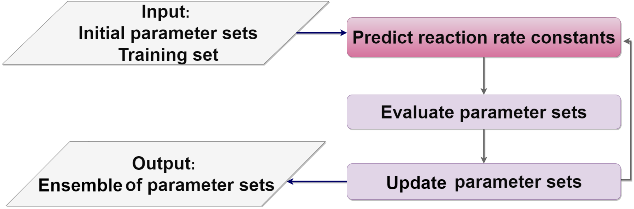
Update parameter sets

Output:
Ensemble of parameter sets

- Instead of stochastic simulations, we estimate mean first passage times (MFPTs) with exact solvers [Suhov and Kelbert, 2008].
  - Can estimate the MFPT of slow reactions efficiently.



Predict Reaction Rate Constants



Input:
Initial parameter sets
Training set

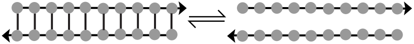
Predict reaction rate constants

Evaluate parameter sets

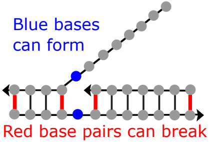
Update parameter sets

Output:
Ensemble of parameter sets

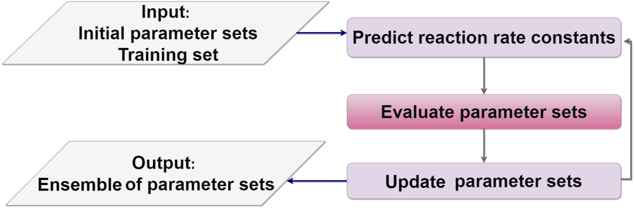
- Instead of stochastic simulations, we estimate mean first passage times (MFPTs) with exact solvers [Suhov and Kelbert, 2008].
  - Can estimate the MFPT of slow reactions efficiently.



  - We use a reduced state space approach to enable sparse matrix computations.



Evaluate Parameter Sets



Input:
Initial parameter sets
Training set

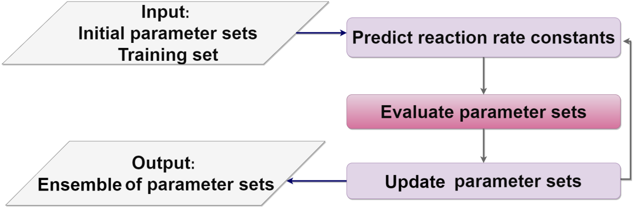
Predict reaction rate constants

Evaluate parameter sets

Update parameter sets

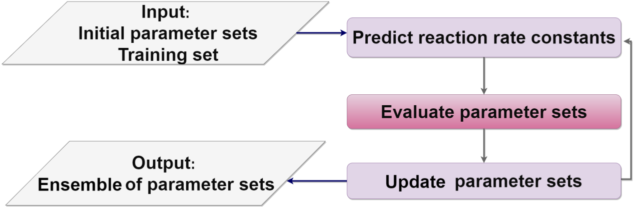
Output:
Ensemble of parameter sets

Evaluate Parameter Sets



- Let θ be the set of parameters in a kinetic model.
  - For the Metropolis model, $\theta = \{\ln k_{\mathbf{uni}}, \ln k_{\mathbf{bi}}\}$.
  - For the Arrhenius model, $\theta = \{\ln A_l,\ E_l \mid \forall l \in \mathcal{C}\} \cup \{\alpha\}$.
- Let k_r and $\hat{k}_r$ be the experimental and predicted reaction rate constant of reaction r, respectively.

Evaluate Parameter Sets



Input:
Initial parameter sets
Training set

Predict reaction rate constants

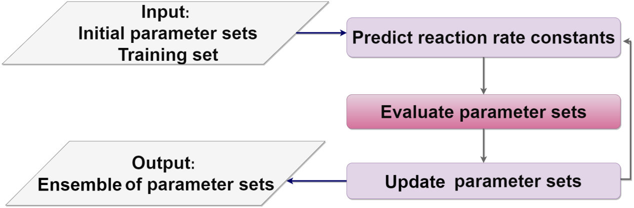
Evaluate parameter sets

Update parameter sets

Output:
Ensemble of parameter sets

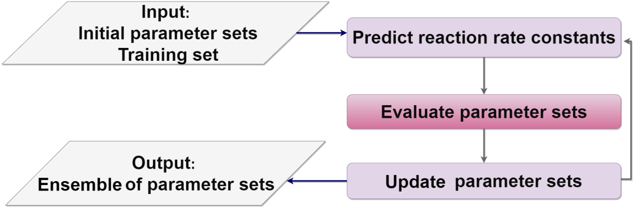
- Let θ be the set of parameters in a kinetic model.
  - For the Metropolis model, $\theta = \{\ln k_{\mathbf{uni}}, \ln k_{\mathbf{bi}}\}$.
  - For the Arrhenius model, $\theta = \{\ln A_l, E_l \mid \forall l \in \mathcal{C}\} \cup \{\alpha\}$.
- Let k_r and $\hat{k}_r$ be the experimental and predicted reaction rate constant of reaction r, respectively.
- The prediction error, ϵ_r, is the difference between $\log_{10} k_r$ and $\log_{10} \hat{k}_r$
  - $\epsilon_r \sim N(0, \sigma^2)$.

Evaluate Parameter Sets



- Let θ be the set of parameters in a kinetic model.
  - For the Metropolis model, $\theta = \{\ln k_{\mathrm{uni}}, \ln k_{\mathrm{bi}}\}$.
  - For the Arrhenius model, $\theta = \{\ln A_l,\ E_l \mid \forall l \in \mathcal{C}\} \cup \{\alpha\}$.
- Let k_r and $\hat{k}_r$ be the experimental and predicted reaction rate constant of reaction r, respectively.
- The prediction error, ϵ_r, is the difference between $\log_{10} k_r$ and $\log_{10} \hat{k}_r$
  - $\epsilon_r \sim N(0, \sigma^2)$.
- We also use priors.

Evaluate Parameter Sets



Input:
Initial parameter sets
Training set

Predict reaction rate constants

Evaluate parameter sets

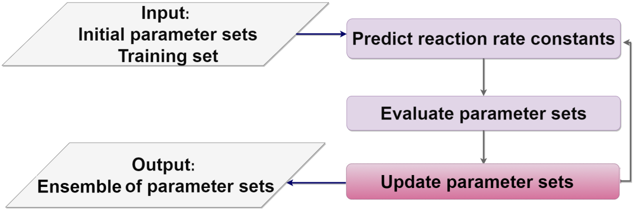
Output:
Ensemble of parameter sets

Update parameter sets

- Let θ be the set of parameters in a kinetic model.
  - For the Metropolis model, $\theta = \{\ln k_{\text{uni}}, \ln k_{\text{bi}}\}$.
  - For the Arrhenius model, $\theta = \{\ln A_l,\ E_l \mid \forall l \in \mathcal{C}\} \cup \{\alpha\}$.
- Let k_r and $\hat{k}_r$ be the experimental and predicted reaction rate constant of reaction r, respectively.
- The prediction error, ϵ_r, is the difference between $\log_{10} k_r$ and $\log_{10} \hat{k}_r$
  - $\epsilon_r \sim N(0, \sigma^2)$.
- We also use priors.
- Thus, the log of the posterior distribution on the training set $\mathcal{D}_{\text{train}}$ is:

$$\log P(\theta, \sigma | \mathcal{D}_{\text{train}}) \approx -\frac{1}{2\sigma^2} \sum_{r \in \mathcal{D}_{\text{train}}} \left(\log_{10} k_r - \log_{10} \hat{k}_r\right)^2 - (n+1)\log\sigma - \frac{\lambda}{2}\|\theta\|_2^2$$

Sample Parameter Sets


Input:
Initial parameter sets
Training set

Predict reaction rate constants

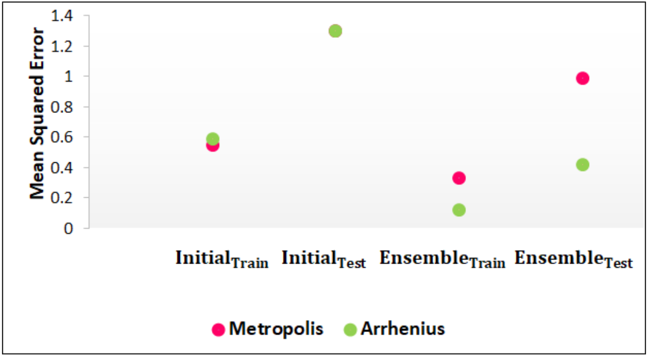
Evaluate parameter sets

Update parameter sets

Output:
Ensemble of parameter sets

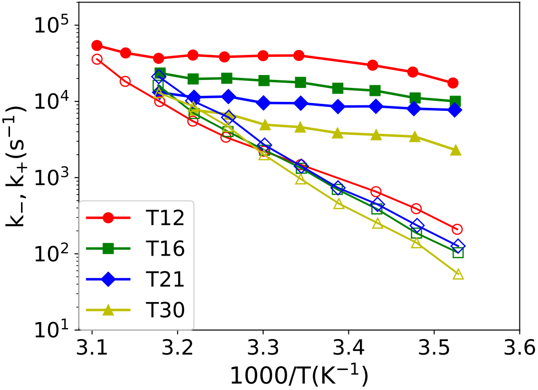
- We draw samples from the posterior distribution of the parameters.
- We approximate the expected value of a reaction rate constant by averaging the predictions of all samples.
- We use the emcee software package [Foreman-Mackey et al., 2013], a Markov chain Monte Carlo (MCMC) ensemble sampler.

Results

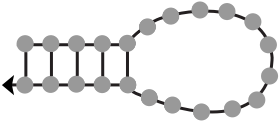


- Initial is an initial parameter set.
- Ensemble is the MCMC ensemble approach.

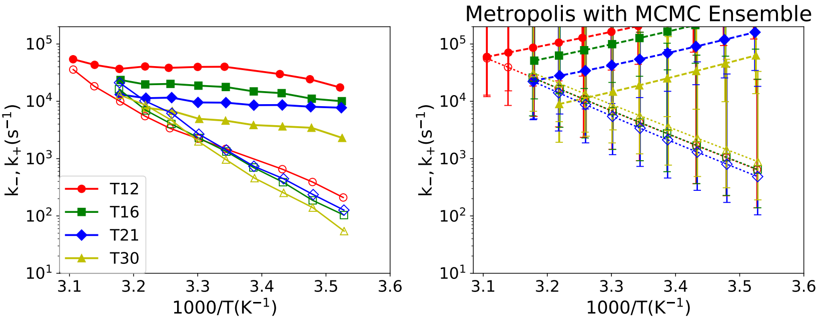
Hairpin Closing and Opening



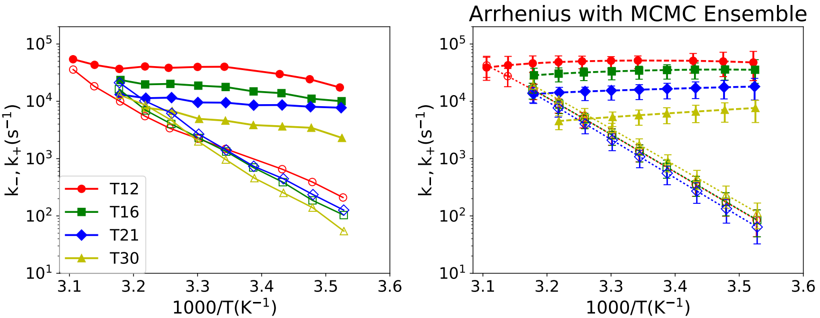
Hairpin closing (solid) and opening (open) [Bonnet et al., 1998]. The legend shows the hairpin loop length.



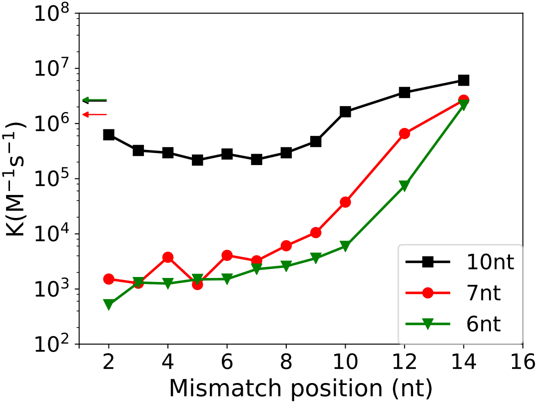
Metropolis Model Fitting for Hairpin Closing and Opening

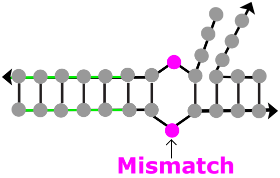


Arrhenius Model Fitting for Hairpin Closing and Opening



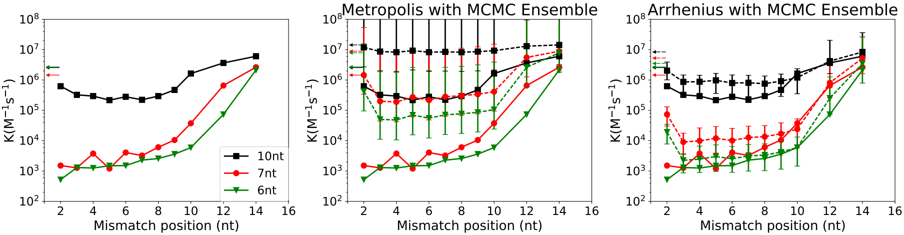
Example: Toehold-mediated 3-way Strand Displacement With Mismatches





Toehold-mediated 3-way strand displacement with mismatches [Machinek et al., 2014]. The legend shows the length of the toehold domain.

Model Predictions for Toehold-mediated 3-way Strand Displacement With Mismatches



Summary

- We introduce an Arrhenius kinetic model.
- We train kinetic models.
  - We collect a dataset of experimentally determined reaction rate constants.
  - We introduce a computational framework for predicting reaction rate constants.
- Our Arrhenius model performs better than an existing model.

Thank You!


Frits Dannenberg


Xander Rudelis


Anne Condon


Joseph M Schaeffer


Mark Schmidt


Chris Thachuk


Erik Winfree