Coupling of folding and binding in protein-DNA interactions

Leonid Mirny

leonid@mit.edu
Problem:

find the target site among $10^6$-$10^9$ decoy sites (by Brownian motion)
Experiment (in vitro):
Experiment (*in vitro*):
Experiment (in vitro):

in water (diffusion limited)

\[ k_{3D} = 4\pi D_{3D} ba \approx 10^7 - 10^8 \text{M}^{-1}\text{s}^{-1} \]
Experiment (in vitro):

\[ k_{3D} = 4\pi D_{3D}ba \approx 10^7 - 10^8 \text{M}^{-1}\text{s}^{-1} \]

Experimental

\[ k_{\text{exp}} \approx 10^{10} \text{M}^{-1}\text{s}^{-1} \]

Riggs et al. 1970
Experiment (in vitro):

Conclusion:
Proteins can bind target sites $\sim 10^2 - 10^3$ times faster than diffusion limit

\[ k_{3D} = 4\pi D_{3D} ba \approx 10^7 - 10^8 \text{M}^{-1}\text{s}^{-1} \]

Experimental
\[ k_{\text{exp}} \approx 10^{10} \text{M}^{-1}\text{s}^{-1} \]

Riggs et al 1970
Experiment: Association rate

Riggs et al 1970

$\begin{align*}
& k_{\text{on}} \approx 10^{10} \text{ M}^{-1} \text{s}^{-1} \\
& P + D \xrightarrow{k_{\text{on}}} PD \\
& \frac{d[PD]}{dt} = k_{\text{on}}[P][D]
\end{align*}$

Muller-Hill Lab 1998

$\begin{align*}
& \text{Dimer} \\
& 2,455 \text{ bp} \quad 49,000 \text{ bp} \\
& Y_{17}Q_{18} \quad 5'\text{TGTGAGC-GCTCACA (O}_{id})
\end{align*}$

Matthews Lab 1998

$\begin{align*}
& \begin{array}{cccc}
  & k_{\text{dissociation}} & k_{\text{association}} & k_{\text{dissociation}}/k_{\text{association}} & K_d^o \\
  & \text{ s}^{-1} & \text{ M}^{-1} \text{s}^{-1} & \text{ M} & \text{ M} \\
\text{PurR} & 2.8 \pm 0.4 \times 10^{-2} & 3.2 \pm 0.6 \times 10^5 & 8.8 \times 10^{-8} & 4.0 \times 10^{-3} \\
\text{PurR} + \text{guanine} & 1.2 \pm 0.2 \times 10^{-3} & 1.5 \pm 0.2 \times 10^7 & 0.8 \times 10^{-10} & 1.9 \times 10^{-10} \\
\text{LacI} & 3.7 \pm 1.3 \times 10^{-2} & [5.1 \times 10^6]^b & 0 & 7.2 \times 10^{-9}
\end{array}
\end{align*}$
Theory of facilitated diffusion:

Max Delbrück
Adam & Delbruck 1968
- reduction of dimensionality in diffusion

O. Berg and P. von Hippel 1981
- PROTEINS CAN SLIDE ALONG DNA
Theory of facilitated diffusion:

Max Delbrück
Adam & Delbruck 1968
- reduction of dimensionality in diffusion

O.Berg and P.von Hippel 1981
- PROTEINS CAN SLIDE ALONG DNA

Facilitated diffusion: 1D diffusion along DNA ("sliding") + 3D

theory: Halford & Marko; R.Bruinsma; M.Moreau, Slutsky & Mirny;
Hu,Grosberg,Shklovskii; J.Langowski; T.Hwa; P.Wolynes
2002-2006

experiments: R.Austin; X.S.Xie; S.Halford
2005-2006
Single-molecule experiments

Fluorescently-labeled transcription factor flow DNA microscope objective

p53 constructs courtesy of Alan Fersht

Anahita Tafvizi

Antoine van Oijen
Single-molecule experiments

Fluorescently-labeled transcription factor

DNA

flow

microscope objective

p53 constructs courtesy of Alan Fersht

Anahita Tafvizi

Antoine van Oijen
Theory of facilitated diffusion:

Max Delbrück
Adam & Delbruck 1968
- reduction of dimensionality in diffusion

O.Berg and P.von Hippel 1981
- PROTEINS CAN SLIDE ALONG DNA

Facilitated diffusion: 1D diffusion along DNA ("sliding") + 3D

theory: Halford & Marko; R.Bruinsma; M.Moreau, Slutsky & Mirny; 2002-2006
Hu,Grosberg,Shklovskii; J.Langowski; T.Hwa

experiments: R.Austin; X.S.Xie; S.Halford 2005-2006
Model: $1D + 3D$

\[
\bar{t}_s(n, M) = \frac{M}{n(\tau_{1d})} \left[ \tau_{1d} + \tau_{3d} \right]
\]

$\bar{t}_s$ – search time

$M$ – genome size
Model: 1D+3D

1. Optimal search time reached at \((\tau_{1D} = \tau_{3D})\)

\[
\tau_s = \frac{2M}{n} \tau_{3D}
\]

2. Mean number of bp scanned in one round

\(n \approx 100 - 500 \text{ bp}\)

3. Optimal 1D/3D is

\(n \ (\sim 100-500) \text{ times faster than 3D only}\)

\(\frac{M}{n} \ (\sim 10^5) \text{ times faster than 1D only}\)
Model: 1D+3D

1. Optimal search time reached at \( \tau_{1D} = \tau_{3D} \)

\[
t_s = \frac{2M}{n} \tau_{3D}
\]

**PROBLEM**

\[
\frac{\tau_{1D}}{\tau_{3D}} = \frac{K_d}{[DNA]}
\]

- \( t_s \) – search time
- \( M \) – genome size
Model: 1D+3D

1. Optimal search time reached at \( \tau_{1D} = \tau_{3D} \)

\[
    t_s = \frac{2M}{n} \tau_{3D}
\]

PROBLEM

\[
    \frac{\tau_{3D}}{\tau_{1D}} = \frac{K_d}{[DNA]} = \frac{10^{-6}M}{10^7 \text{bp} \cdot 10^{-9}M} = 10^{-4}
\]

“...theory isn't worth a damn unless you put in numbers...”

attributed to Robert Austin

by Rob Phillips
Sliding on the energy landscape

Energy landscape of 1D sliding

..CATGTTCAGGCCACGTAGC...

solvent

Energy landscape of 1D sliding

$E_{\text{ns}}$

$\sigma$
Energy is strongly sequence dependent.

\[ E = \sum_{i=1}^{L} e(i, b_i) \]

Energy \( E \) is strongly sequence dependent.

Landscape

NO ENERGY GAP between target and random sites.

\[ Z = \frac{E_N - \langle E \rangle}{\sigma(E)} > Z_{\min} = -\sqrt{3L} \]
Protein-DNA interaction energy

\[ E = \sum_{i=1}^{l} e(i, b_i) \]

Distribution of energy of random (genomic) sites is GAUSSIAN

NO ENERGY GAP between cognate and random sites

FIGURE 1  Spectrum of binding energy for three different transcription factors and the Gaussian approximation (solid line).
Results

Fast sliding requires smooth landscape (small $\sigma$)

1. It’s DIFFUSION
   \[ t \sim L^2/2D_{1d} \]

2. DIFFUSION COEF.
   \[ D_{1D} \sim e^{-\gamma \beta^2 \sigma^2} \]
   \[ \gamma \sim 1 \]

Roughness of the energy landscape

\[ \sigma \]

smooth \hspace{2cm} rugged
Results

Stability of the ground state requires large $\sigma$

Roughness of the energy landscape

NO ENERGY GAP between target and random sites

$Z_{\text{target}} \sim \sqrt{L}$

FRACTION OF TIME SPENT ON THE TARGET SITE

Smooth vs. rugged

Roughness of the energy landscape
Either speed or stability but not both!

FRACTION OF TIME SPENT ON TARGET SITE

STABILITY: $\sigma > 5kT$

SPEED $\sigma < 2kT$

Either speed or stability

*but not both!*
Two-state mechanism
Two-state mechanism

Coupling of binding and folding

SEARCH MODE

RECOGNITION MODE

$\Delta G_{\text{fold}}$
NMR of non-specific complex

Kalodimos et.al Science.2004
Structure and animation by Babis Kalodimos et al
Two-state mechanism

SEARCH MODE

RECOGNITION MODE

$\Delta G_{\text{fold}}$

$T$
Effective landscape
Effective landscape
Effective landscape
Effective landscape:

use Random Energy Model to get

1. kinetics (fraction of time in the S state)
2. stability on the target site
Two-state mechanism

Prob. to fold on the site during a round of sliding

Fraction of time in S state

$$t_s = \frac{1}{P_f n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right)$$

$$f = \frac{1}{K + 1}$$

$$K(i) = \frac{k_F}{k_U(i)}$$
Two-state mechanism

\[ t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right) \]

Proba. to fold on the site during a round of sliding

\[ P_f = \frac{k_F}{k_F + \tau_{res}^{-1}} \]
Two-state mechanism

\[ t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right) \]

Prob. to fold on the site during a round of sliding

\[ P_f = \frac{k_F}{k_F + n / \tau_{1D}} \]
Two-state mechanism

Two effects:
1. Speed up due to \( S \) state.
2. Slow down due to a possibility to miss the site (i.e. come to the site and leave it before going into \( R \)).

\[
t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right)
\]

\[
K = \frac{k_F}{k_U}
\]
Two-state mechanism

Two effects:
1. Speed up due to $S$ state.
2. Slow down due to a possibility to miss the site (i.e. come to the site and leave it before going into $R$)

$$t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right)$$

$$K = \frac{k_F}{k_U}$$
Two-state mechanism

Two effects:
1. Speed up due to S state.
2. Slow down due to a possibility to miss the site (i.e. come to the site and leave it before going into R)

\[ t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right) \]

solvent

S

R

K = \frac{k_F}{k_U}
Two-state mechanism

Two effects:
1. **Speed up** due to S state.
2. **Slow down** due to a possibility to miss the site (i.e. come to the site and leave it before going into R)

\[ t_s = \frac{1}{P_F} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right) \]

- depends on rate of folding
- depends on \( \Delta G \) between R and S

\[ K = \frac{k_F}{k_U} \]
Two-state mechanism

Two effects:
1. Speed up due to S state.
2. Slow down due to a possibility to miss the site (i.e. come to the site and leave it before going into R)

\[ t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D} (1 + K) \right) \]

rate of folding

\( \Delta G \) between R and S
Two-state mechanism

\[ P_f = \frac{k_F}{k_F + n/\tau_{1D}} \]
Q: What if I make the protein-DNA complex more rigid?

A: The search becomes 10-1000 times slower!

SMALL CHANGE IN RATES LEADS TO BIG SLOW-DOWN OF THE SEARCH!
Two-state mechanism

Coupling of binding and folding

\[ P_f = \frac{k_F}{k_F + n / \tau_{1D}} \]

\[ \frac{\tau_{1D}}{n} = C \frac{n}{4D_{1D}} \approx \frac{10^2}{4 \cdot 2 \cdot 10^6} \approx 10^{-5} \text{sec} \]
Two-state mechanism

Coupling of binding and folding

PROBLEM
Requires fast SR transition

Experimental folding time $\sim 10^{-4}$ s
**Idea:** It would be nice to have a protein that undergoes S-to-R transition MOSTLY on the cognate sites.

**IDEA:** CORRELATED LANDSCAPES:
Kinetic pre-selection: correlated landscapes

\[ t_s = \frac{1}{P_f} \frac{M}{n} (\tau_{3D} + \tau_{1D}(1 + K)) \]

\[ P_f = \frac{k_F}{k_F + \tau_{\text{res}}^{-1}} \]

Prob. to fold on the site during a round of sliding
Kinetic pre-selection: correlated landscapes

\[ t_s = \frac{1}{P_f} \frac{M}{n} \left( \tau_{3D} + \tau_{1D}(1 + K) \right) \]

\[ P_f = \frac{k_F}{k_F + \tau_{res}^{-1}} \]

\[ \tau_{res}(i) \approx \tau_{1D}/n \cdot \exp\left[ \frac{\beta^2 \sigma_s^2}{4} - E_S(i)/kT \right] \]

**coupling between folding and binding**
Kinetic pre-selection: correlated landscapes

\[ \frac{\tau_{1D}}{n} \]

Experimental folding rate
Kinetic pre-selection: correlated landscapes

coupling between folding and binding

\[ \frac{\tau_{1D}}{n} \]

Total search time (sec)

Folding time $1/k_F$ (sec)

Correlated landscapes

uncorrelated landscape
Kinetic pre-selection: correlated landscapes

coupling between folding and binding

![Graph showing correlated landscapes and experimental folding rate](image_url)
Coupling of folding and binding

1. allow to get both fast binding and stability;
2. kinetic pre-selection couple binding and folding (consistent with structures);
3. R state is “strained” (e.g. bent DNA) and $\Delta G$ is optimal (~ 15 Kcal/mol)
4. DNA-binding proteins are fine-tuned:
   Mutations that make folding slower or stabilize/destabilize R state CAN KILL PROTEIN-DNA RECOGNITION.
Requirements for $\Delta G$
Requirements for $\Delta G$
Requirements for $\Delta G$

Conclusion: narrow range of $\Delta G$ works
Comparison with experiment?
Comparison with experiment?

"Measured" energy

$E(i) = -\log P(i)$

Energy of the site

Fig. 2. Binding affinities of C-terminally tagged TFs
Coupling of folding and binding

1. allow to get **both** fast binding and stability;
2. **kinetic pre-selection** couples binding and folding (consistent with structures);
3. R state is “strained” (e.g. bent DNA) and ΔG is optimal (~ 15 Kcal/mol)
4. DNA-binding proteins are fine-tuned: **Mutations** that make folding slower or stabilize/destabilize R state **CAN KILL PROTEIN-DNA RECOGNITION.**
Summary

0. Theory isn’t worth a damn unless you put in numbers.

1. To search fast and bind tightly the protein needs to have (at least) two states (S and R).

2. DNA-binding proteins must be fine-tuned.

3. Facilitated (1D/3D) diffusion may not be so facilitated (see 0. above) eukaryotes?
Michael Slutsky, MIT Physics

Zeba Wunderlich, Harvard Biophysics