What does a protein need to work?

Leonid Mirny
leonid@mit.edu
What does a protein need to work?

1. Stable structure
2. Specific active/binding sites.
What does a protein need to work?

1. Stable structure
2. Specific active/binding sites.
3. Somewhat unstable structure
4. Non-specific binding site
A simple DNA-binding protein

Function
1. Find its site on DNA
2. Bind it tightly
3. IF [ligand] > 0
   leave the site
ELSE
   goto step 1.
END
A simple DNA-binding protein

**Function**

1. Find its site on DNA
2. Bind it tightly
3. IF [ligand] > 0
   leave the site
   ELSE
   goto step 1.
   END
Problem 1:
find cognate site among $10^6$-$10^9$ non-cognate sites
Experiment:
Riggs et al 1970
Diffusion-limited association

Theory:
Richter and Eigen 1974,
Berg, Winter, von Hippel 1981

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

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

WATER
\[ t_{a} \sim 1 - 10 \text{ sec} \]

CELL
\[ t_{a} \sim 10^2 - 10^3 \text{ sec} \]
Model: 1D+3D

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

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

A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA
Our model: 1D+3D

Energy landscape of 1D sliding
Protein-DNA interaction energy

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

Energy is strongly sequence dependent

Energy landscape

NO ENERGY GAP between cognate and random sites
Results

Fast sliding requires smooth landscape

\[ D_{1d} \sim e^{-\frac{7}{4} (\sigma / k_B T)^2} \]

Roughness of the energy landscape
Results

Specific recognition requires **rough landscape**

\[ \sigma \]

\[ \text{FRACTION OF TIME SPENT ON THE COGNATE SITE} \]

Roughness of the energy landscape
Speed-stability paradox

Either speed or stability 
but not both!

Proposed Mechanism
Proposed Mechanism

SEARCH MODE

RECOGNITION MODE
Experiments

Kalodimos et.al Science.2004
Structure and animation by Babis Kalodimos et al
Landscape model

\[ U(x, \ z) = U_{\text{spec}}(x, \ z) + U_{\text{non-spec}}(z) \]

\[ U(x, \ z) = U_s(x)e^{-z} + \frac{\alpha}{2}(z - z_0)^2 \]
Meso-scale dynamics of landscape model

\[ \log(t(x,z)) \]
Macroscale dynamics

Experimental folding rate

Flat landscape model

Correlated landscapes
Summary

1. 1D+3D search is fast if the protein-DNA complex is **FLEXIBLE**.

2. Conformational transition in the DNA-binding protein controls the search time.
Somewhat unstable structure is needed for …

**Function**

1. Find its site on DNA
2. Bind it tightly
3. IF \([\text{ligand}] > 0\) leave the site
   ELSE
   goto step 1.
4. END
What does a protein need to work?

1. Stable structure
2. Specific active/binding sites.
3. Somewhat unstable structure
4. Non-specific binding site
What does a protein need to work?

1. Stable structure
2. Specific active/binding sites.
3. Somewhat unstable structure
4. Non-specific binding site
Model: 1D+3D

..CATGTTCAAGGCAACGTAAGC...

\( \tau_{1d} \)

\( \tau_{3d} \)
Model: 1D+3D

Energy landscape of 1D sliding
Results

Fast sliding requires optimal non-specific binding

Free energy of non-specific binding
Non-specific binding is needed for …

Function

1. Find its site on DNA
2. Bind it tightly
3. IF [ligand] > 0
   leave the site
   ELSE
   goto step 1.
END
Non-specific DNA Binding of Genome Regulating Proteins as a Biological Control Mechanism: I. The lac Operon: Equilibrium Aspects

(lac repressor/DNA–protein interactions/RNA polymerase/repressor–inducer complexes)

PETER H. VON HIPPEL, ARNOLD REVZIN, CAROL A. GROSS*, AND AMY C. WANG

$M \approx 10^6$ - non-specific sites

$m \approx 10$ - number of LacI proteins per cell

$$P = \frac{1}{1 + \frac{M}{m} \frac{K_d^S}{K_d^{ns}}}$$

$K_d^{ns} \approx 10^{-6} M$

$K_d^{s,NO-LIGAND} \approx 10^{-12} M$

$K_d^{s,LIGAND} \approx 10^{-9} M$

Fraction of time the site is bound

$P_{NO-LIGAND}^{NO-LIGAND} = \frac{1}{1 + 0.1} = 0.9$

$P_{LIGAND}^{LIGAND} = \frac{1}{1 + 100} = 0.01$
Non-specific binding is needed for …

Function

1. Find its site on DNA
2. Bind it tightly
3. IF [ligand]>0
   leave the site
ELSE
   goto step 1.
END
1. 1D+3D search is fast if the protein-DNA complex is **FLEXIBLE**.

2. Conformational transition in the DNA-binding protein controls the search time.

3. Non-specific binding is essential for protein function.
What does a protein need to work?

1. Stable structure
2. Specific active/binding sites.
3. Somewhat unstable structure
4. Non-specific binding site
Testable predictions

1. Diffusion of a protein on DNA is sequence-dependent.

2. DNA sequences can influence
   - folded/unfolded equilibrium
   - rate of conf. transition in the protein
     (nucleate folding on the target site)

3. Mutations that change the stability and rate can have affect on the total search time and timing of gene expression.
Acknowledgements

Michael Slutsky, MIT Physics