Conformational Proofreading:
Enhancing molecular recognition by structural changes

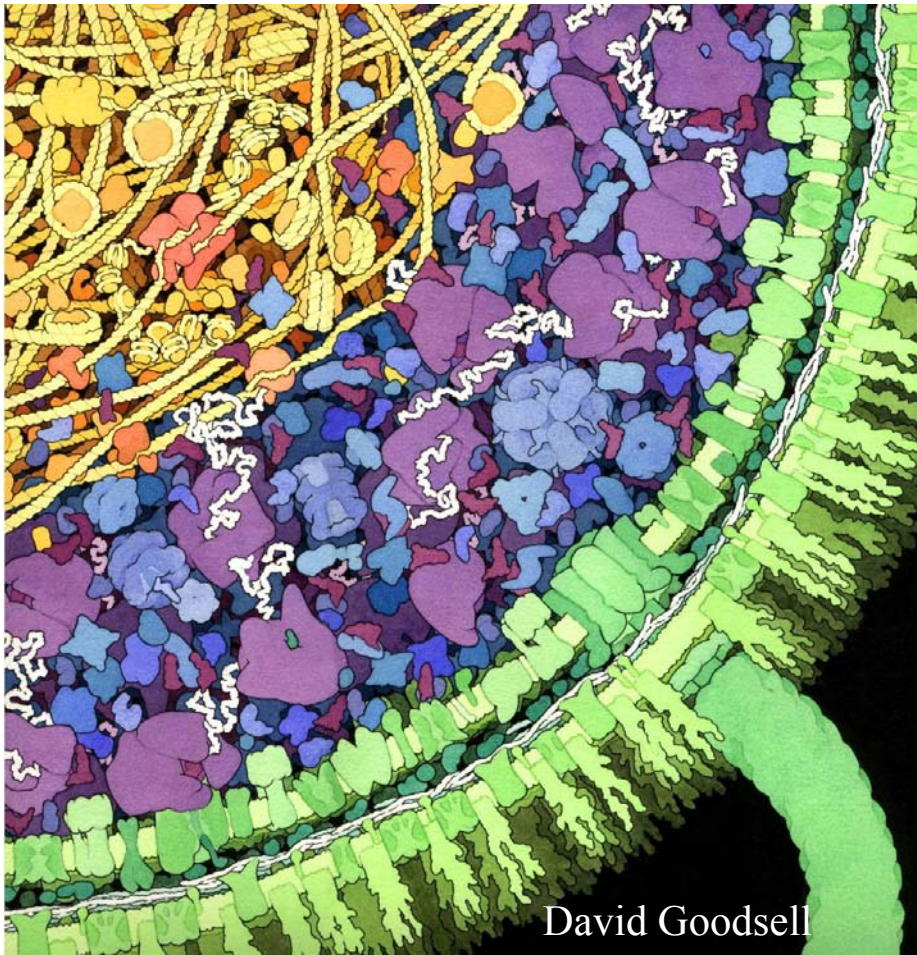
Yoni Savir and Tsvi Tlusty
Weizmann Institute

- Molecular recognition in the presence of competition and noise.
- Conformational changes (**induced fit**) enhance recognition.
- Case study: Homologous Recombination.
- Suggested mechanism: **Conformational Proofreading.**

Challenge of molecular recognition

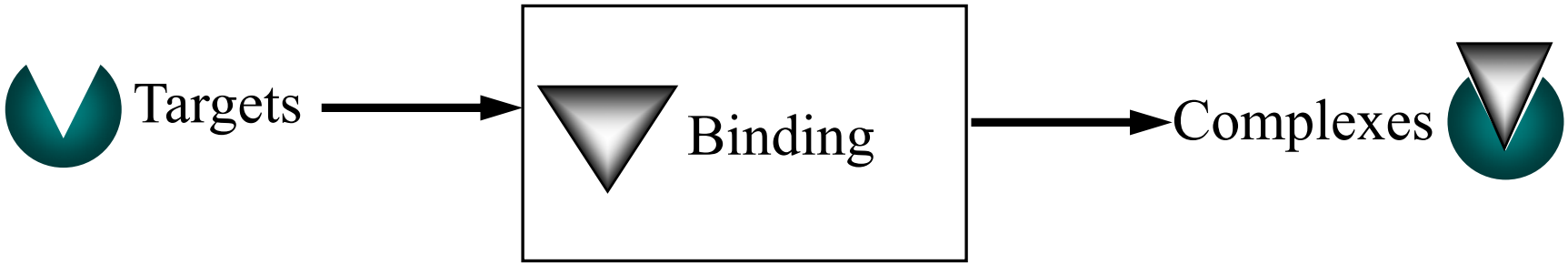
- *Noisy*, *crowded* milieu.
- Recognizer and target fluctuate.
- Many competing *lookalikes*.
- Relatively weak recognition interactions.
- Challenge:

How to find and identify targets ?



Designing noisy molecular information channels

- Molecular information channels rely on molecular recognition.



- Large-scale: *channel optimization*

  Assigning outputs to inputs to minimize the impact of errors.

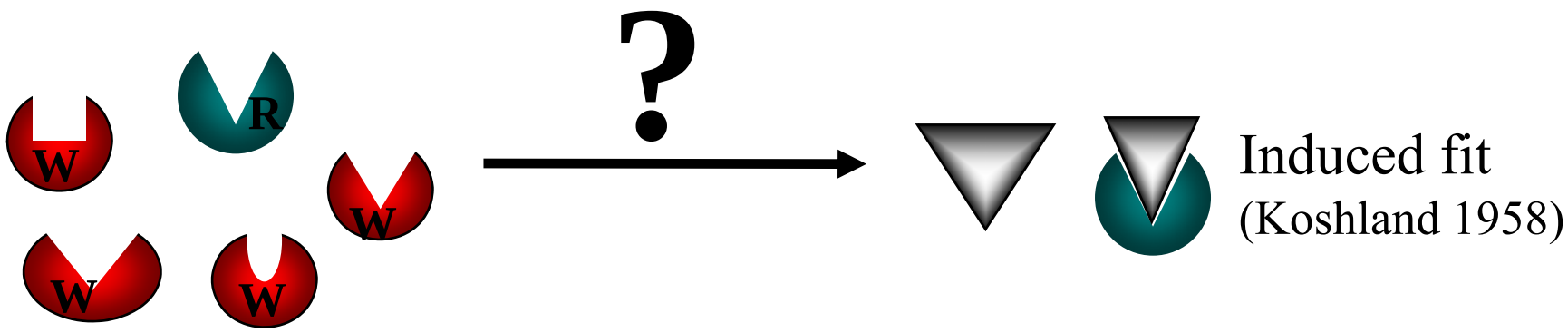
[Tlusty, Phys Rev Lett 08, J Th Bio 07, Phys Bio 08]

- Small-scale: Improving *accuracy* of each recognition event by conformational changes.

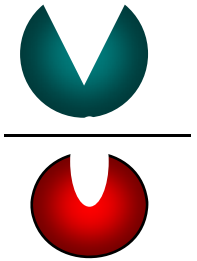
[Savir &Tlusty, PLoS ONE 2007, IEEE J Signal Processing 2008]

Why induced fit?

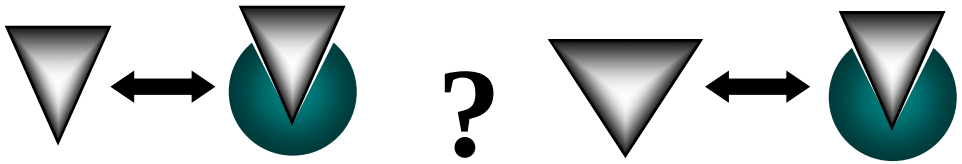
- **Induced fit**: Recognizers change their shape upon binding.
- Can molecular recognition gain from induced fit?



- Quality measures: *specificity* = [Right]/[Wrong] =
- What is the optimal recognition strategy?

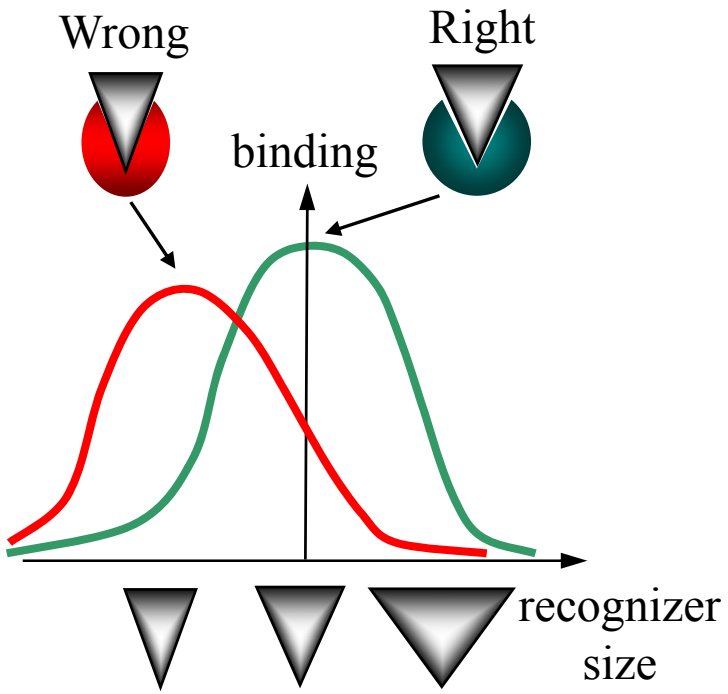


Lock-and-key or *induced fit ?*

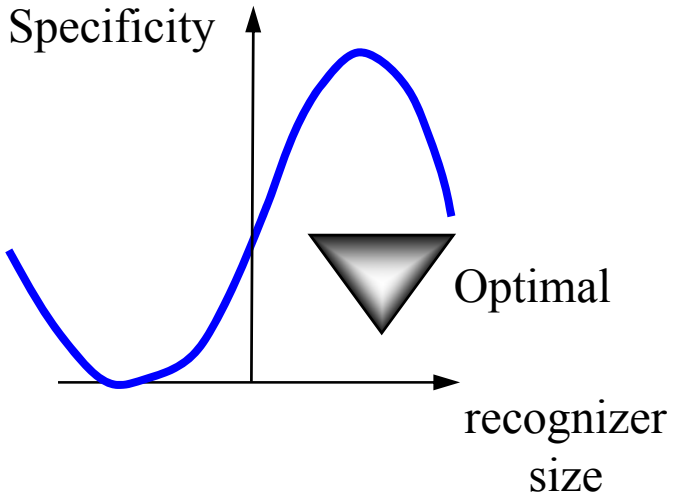


Conformational Proofreading: When off-target is right on

- Structural *mismatch* reduces Right, *but* also reduces Wrong even more.
- Result: Enhancement of specificity and other quality measures.
- Optimal specificity at finite mismatch.



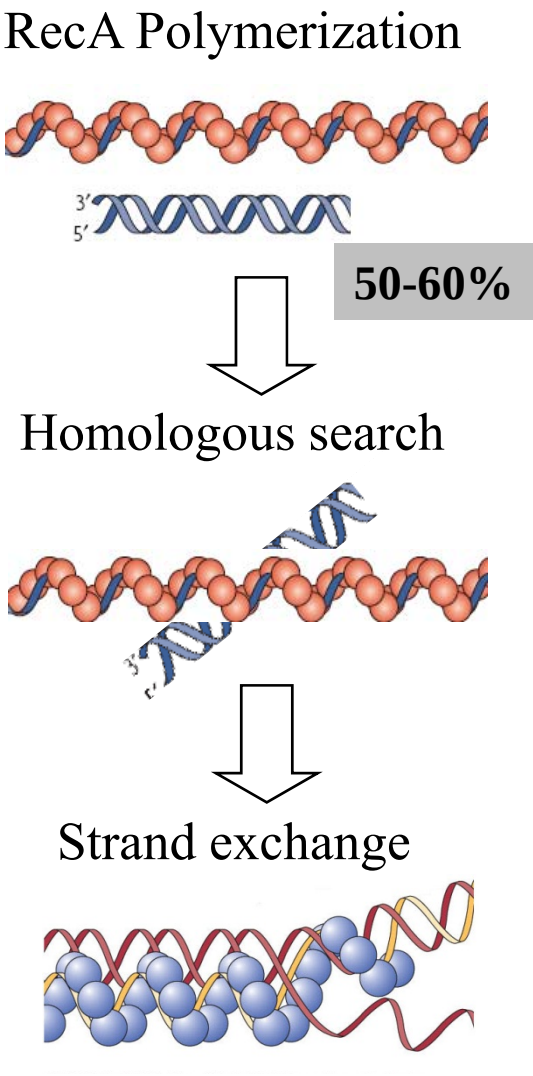
- ***Induced fit*** enhances recognition.
- Optimal recognizer is ***off-target***
- Not lock-and-key.



- Quantitative example: *Homologous Recombination*.

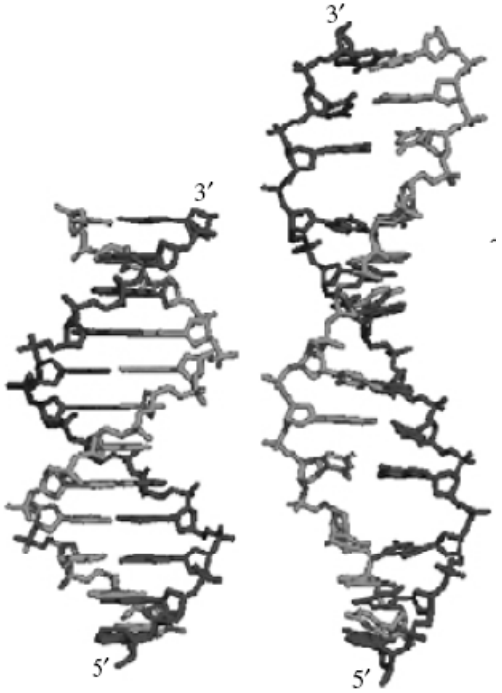
Case study: Homologous Recombination

- Essential for genome repair and genetic diversity by crossover and horizontal transfer.
- Mediated by proteins of RecA superfamily.
- Structure and function conserved from bacteria to human (RecA, Rad51, hRad51).
- RecA extends DNA by **50-60%** - *conserved*.
- Homologous search: recognition of homologous sequence within lookalikes.
- Homologous search occurs *without* ATP hydrolysis.

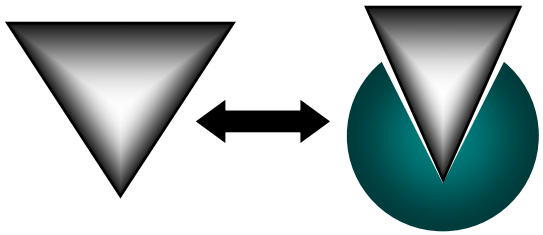


Role of DNA extension in homologous search?

- Costs significant deformation energy,
  3-4 k_BT per base-pair (k_BT=0.6 kcal/mol).
- Structural reason: exposing bases?
- Increasing effective target size?

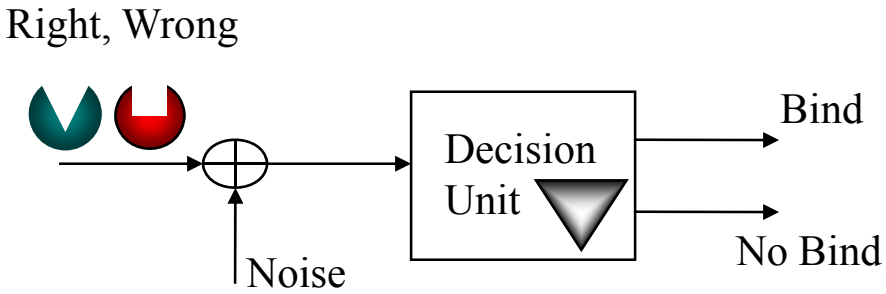


- **Conformational Proofreading**?
  mechanism to detect Right DNA in large pool of similar targets.



Molecular recognition as a decision problem

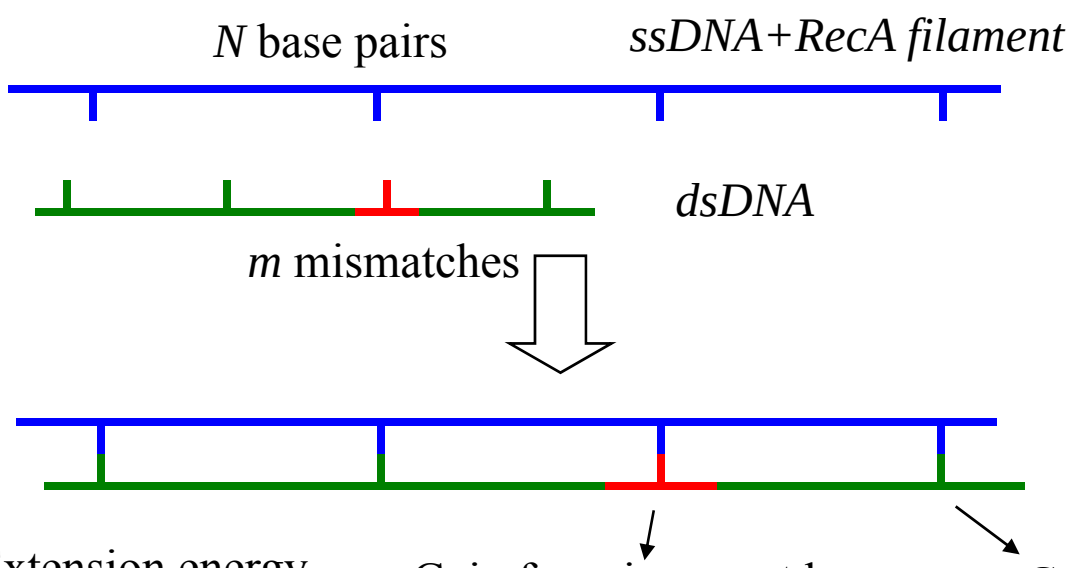
- Natural measure for the performance of molecular recognition .
- Each possible binding event has a cost/benefit of identification and misidentification.

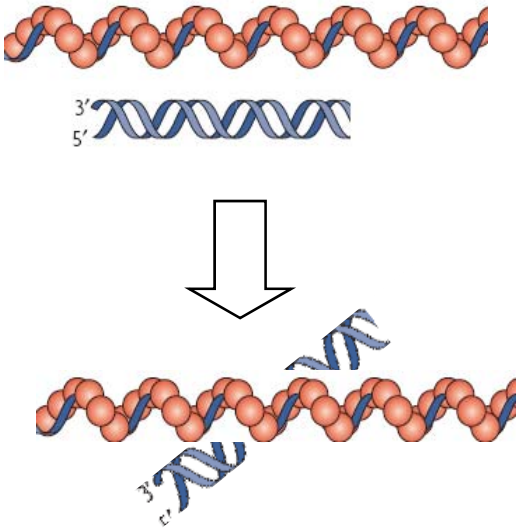


Target / Decision	Right	Wrong
Bind	Correct, C_{RB}	False Alarm C_{WB}
No Bind	Miss, C_{RN}	Correct, C_{WN}

- *Cost* = Cost(event)×Prob(event) = correct + mis + false-alarm.
- *Cost depends on structural parameters and can be optimized.*

Cost depends on structural parameters





Extension energy

ΔG_{ext}

Gain form incorrect bp: nonspecific interactions

ΔG_{ns}

Gain from correct bp: specific interactions

ΔG_{s}

- Binding probability for *N* base pairs and *m* mismatches :

$$P_{binding}(N,m) = F(N,m,\Delta G_{ext},\Delta G_{s},\Delta G_{ns})$$

Cost balances Right and Wrong binding

Maximize Right detection + Minimize Wrong detection

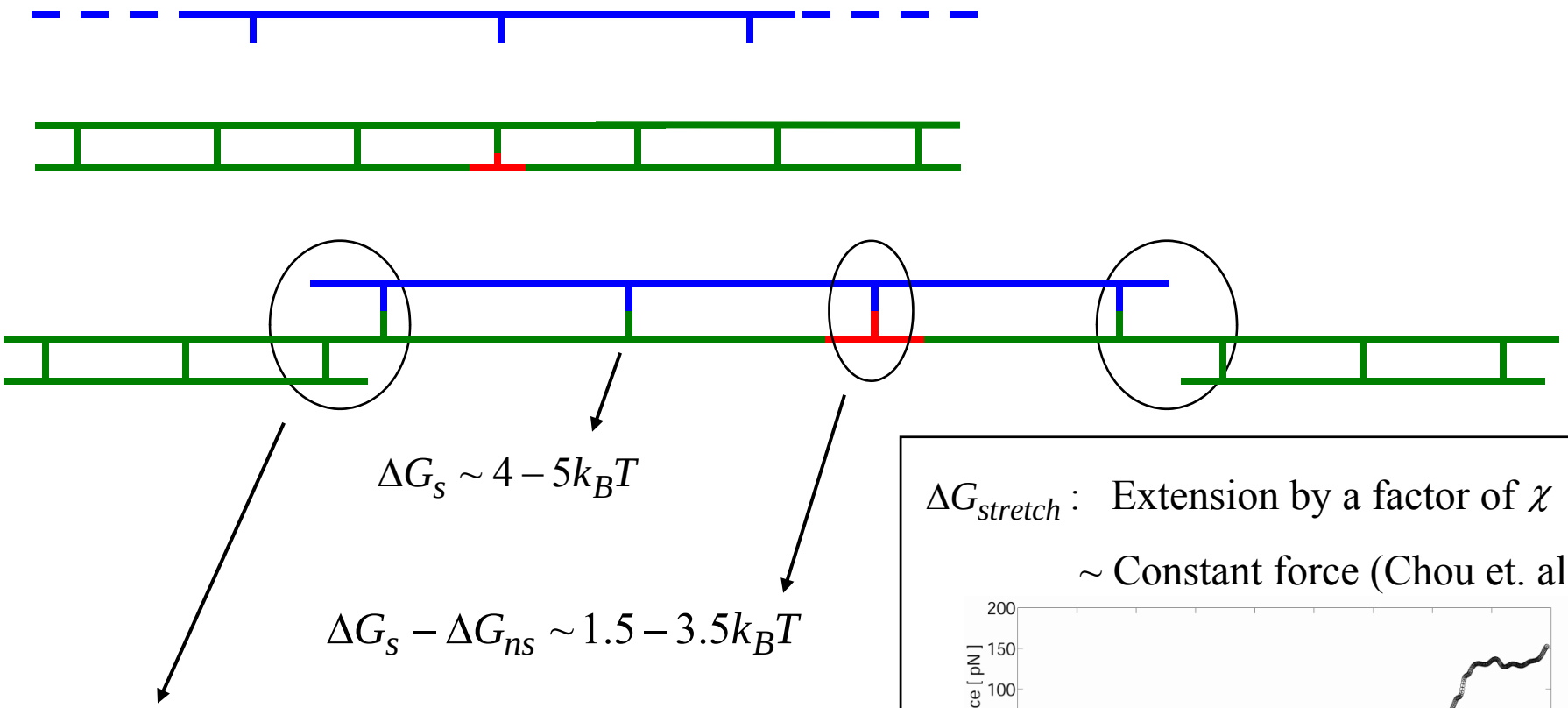
$$Cost = -P_{binding}(\text{hom}) + t \times P_{binding}(non - \text{hom})$$

- Tolerance *t* measures relative cost of error, accounts for abundance, functionality.
- *t* increases → system less tolerant to errors

- Cost depends on structural parameters and binding energetics:

ΔG_{ext} extension
ΔG_s specific interactions
ΔG_{ns} nonspecific interactions
$\Delta\Delta G_{mis} = \Delta G_s - \Delta G_{ns}$

Structural parameters are measured



$\Delta G_s \sim 4 - 5k_BT$

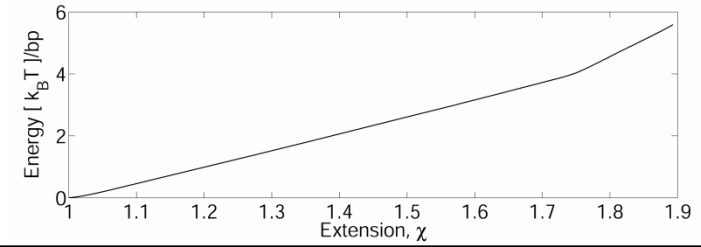
$\Delta G_s - \Delta G_{ns} \sim 1.5 - 3.5k_BT$

Interfacial energy

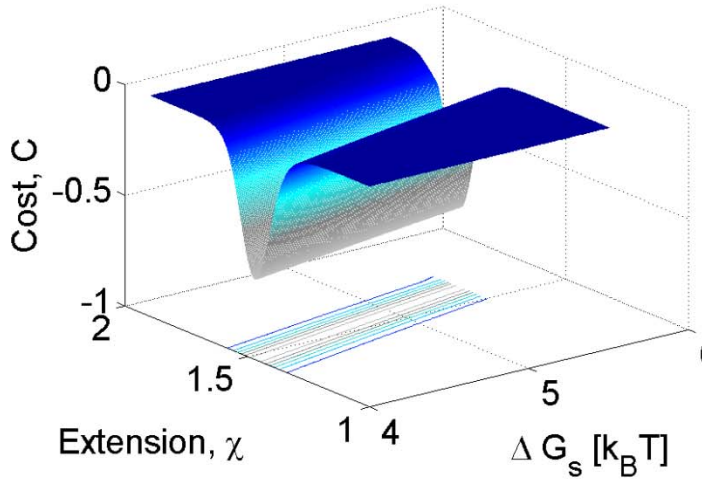
$\Delta G_{int} \sim 3.6k_BT$

$\Delta G_{stretch}$: Extension by a factor of χ

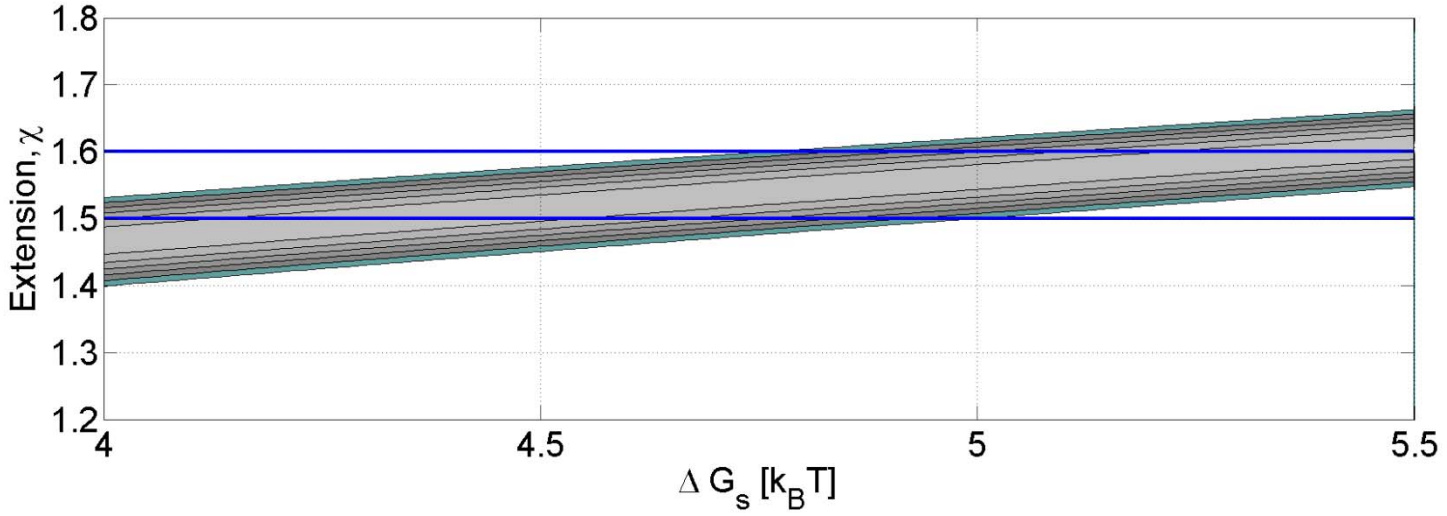
~ Constant force (Chou et. al.)



Cost exhibits minimum at optimal extension



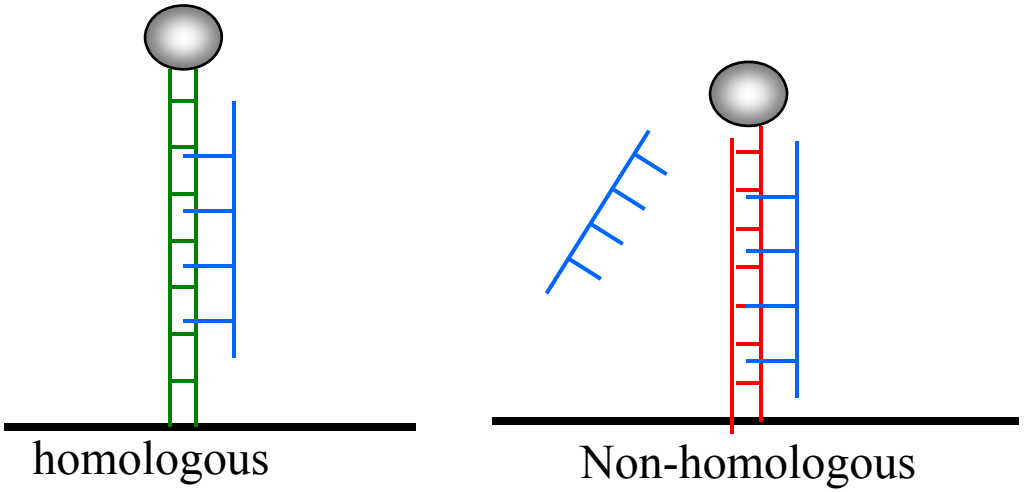
- Well-defined valley ~ 50-60%

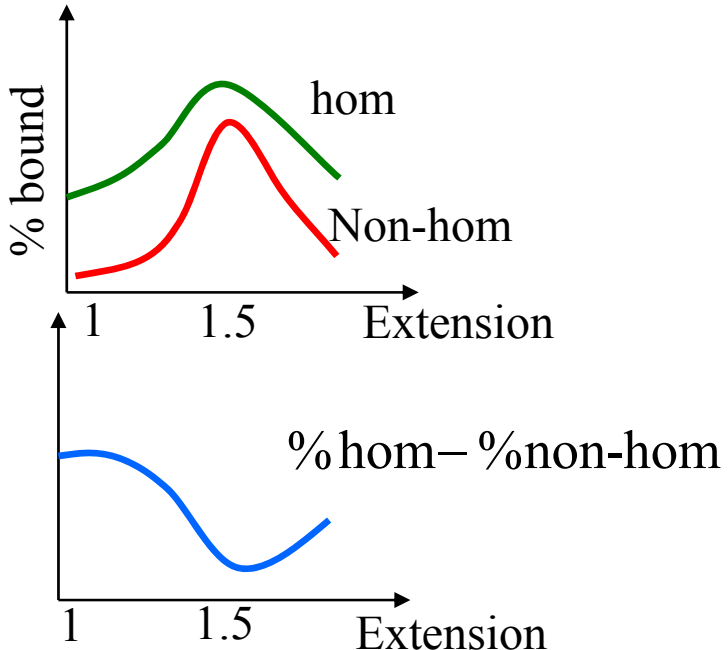


- **One** RecA monomer, one mismatch, symmetric: $N = 3,\ m = 1,\ \Delta\Delta G_{mis} = 3k_BT,\ t = 1$

Possible experimental tests

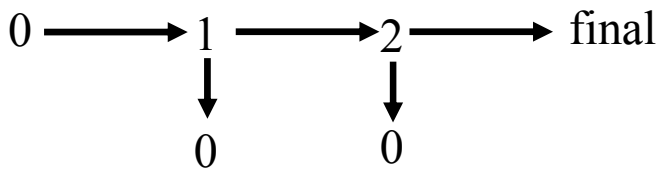
- Binding curves as a function of controlled extension (e.g. Viovy et al.,Curie)
- Predicting increase of binding fraction with extension.
- But optimal cost at zero extension.





Conformational and Kinetic Proofreading

- **Kinetic Proofreading** (Hopfield, Ninio):
- Intermediate *irreversible* steps.
- Reduce production of *both* right and wrong. Reduction of the wrong product is *larger* and specificity improves.
- Recent evidence in recombination.



Sagi, Tlusty, Stavans (2006) *NAR*

Kinetic Proofreading:

Time delay (additional steps)

Energy-consuming non-equilibrium.

Conformational Proofreading:

Spatial mismatch.

Quasi-equilibrium.

Summary and outlook

- **Conformational Proofreading:** conformational changes (induced fit) may enhance the quality of molecular recognition.
- RecA-mediated recombination involves extreme *DNA extension*.
- Analogy between molecular recognition and *decision problem* with a natural "fitness" reveals optimal extension close to observed extension of 50-60%.
- Homologous recombination combines:

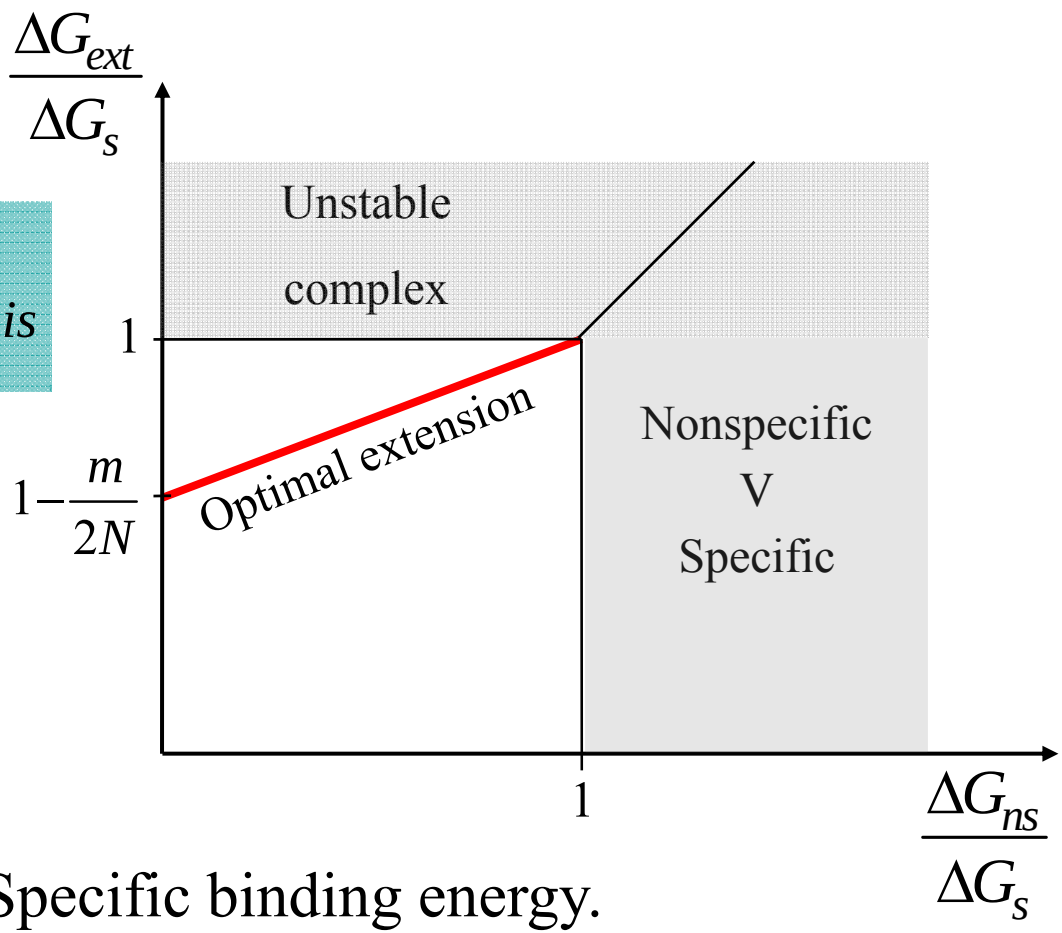
  Kinetic + Conformational Proofreading.
- Outlook: Application of Conformational Proofreading to other systems: tRNA, transcription, enzymes… (in progress).

• **Savir &Tlusty, PLoS ONE 2007; IEEE J Signal Processing 2008; RecA – submitted.**

Optimal extension minimizes detection cost

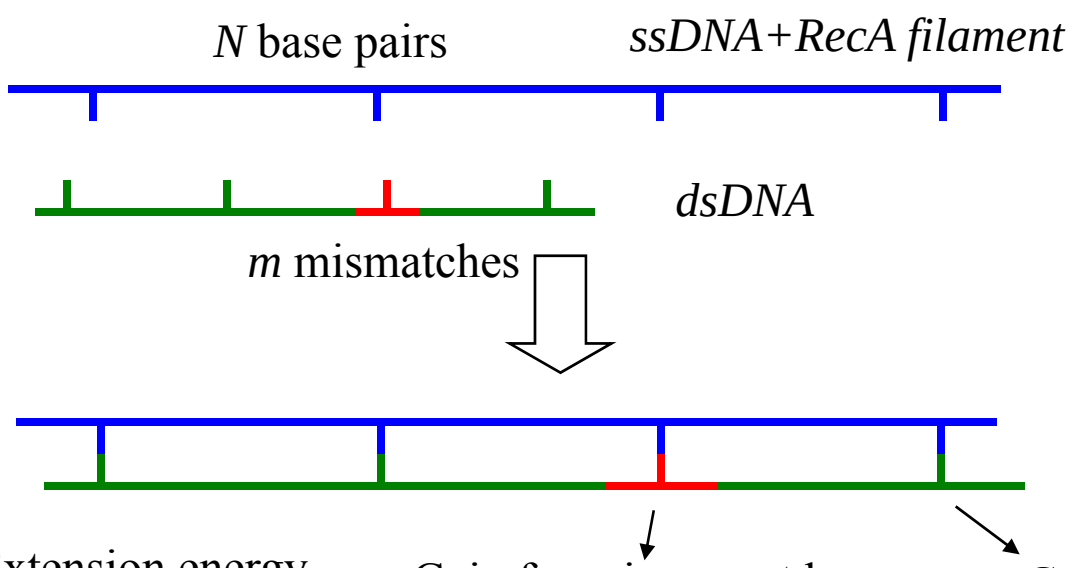
- Minimization of the cost reveals an optimal extension ($t = 1$):

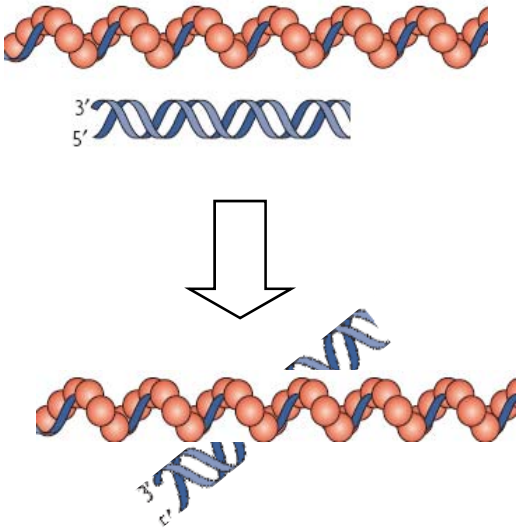
$$\overline{\Delta G_{ext}} = \Delta G_s - \frac{m}{2N}\Delta\Delta G_{mis}$$



- Optimal extension energy ~ Specific binding energy.

Cost depends on structural parameters





Extension energy

ΔG_{ext}

Gain form incorrect bp: nonspecific interactions

ΔG_{ns}

Gain from correct bp: specific interactions

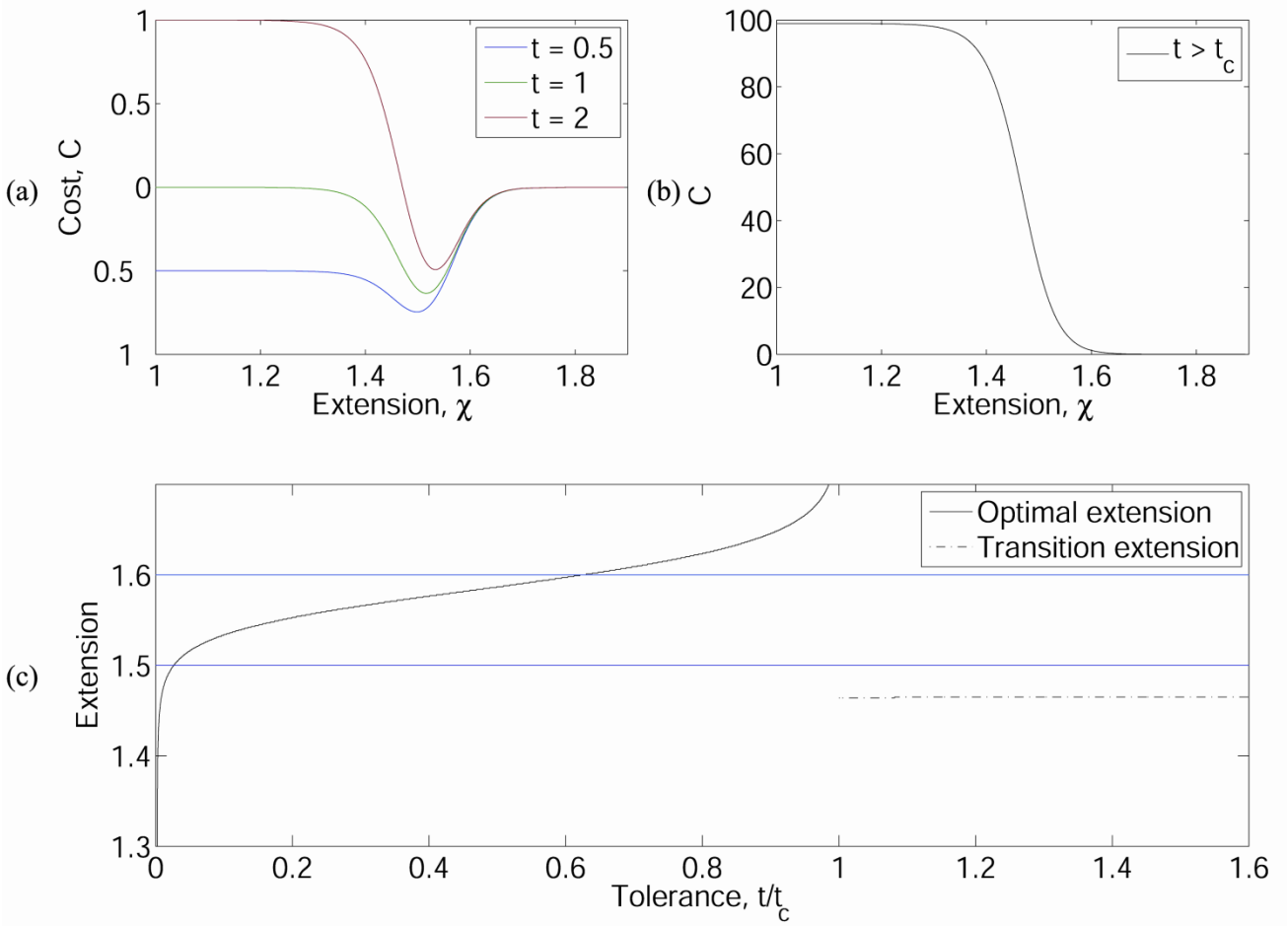
ΔG_s

- Binding probability for N base pairs and m mismatches :

$$P_{binding}(N,m) = \frac{1}{1 + exp(N \cdot \Delta G_{ext} - (N-m) \cdot \Delta G_s - m \cdot \Delta G_{ns})}$$

Optimal extension is not sensitive to tolerance

- t increases → system less tolerant to errors

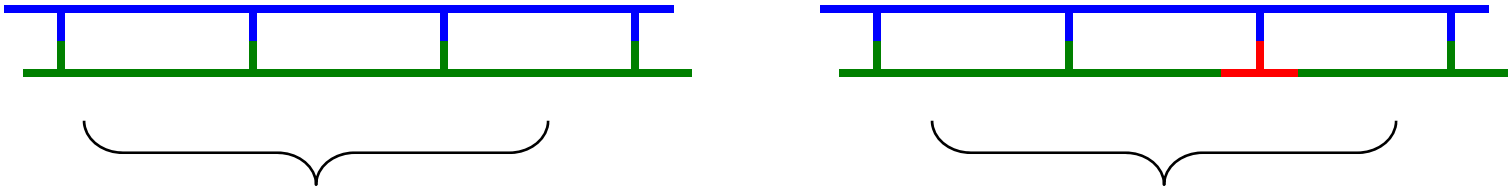


$$t_c \cong e^{m\Delta\Delta G_{mis}}$$

$$N = 3,\ m = 1,\ \Delta G_s = 1.5 k_B T,\ \ \Delta\Delta G_{mis} = 3 k_B T$$

Cost balances correct and incorrect binding

$$Cost = \frac{-1}{1+exp(N\cdot\Delta G_{ext} - N\cdot\Delta G_s)} + \frac{t}{1+exp(N\cdot\Delta G_{ext} - N\cdot\Delta G_s + m\cdot\Delta\Delta G_{mis})}.$$



Maximize correct detection + Minimize incorrect detection

ΔG_{ext} extension
ΔG_s specific interactions
ΔG_{ns} nonspecific interactions
$\Delta\Delta G_{mis} = \Delta G_s - \Delta G_{ns}$

- t – *Tolerance* of the System

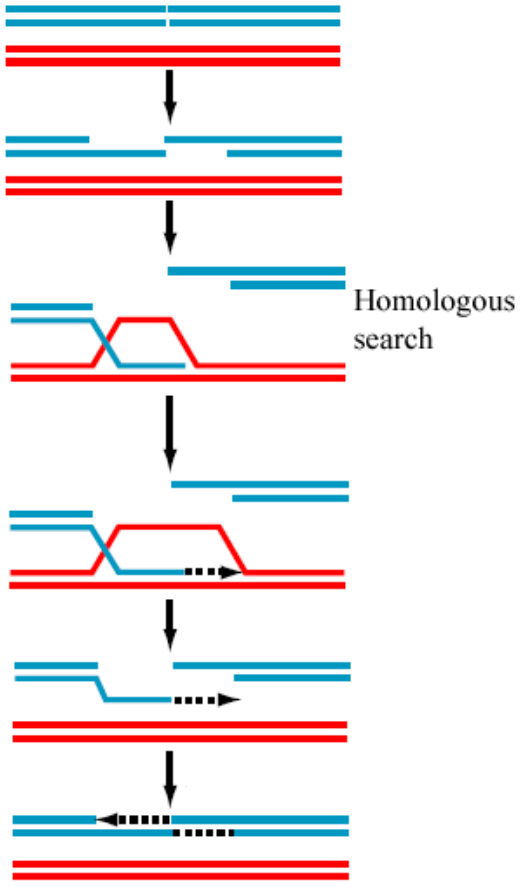
Cost ← Occurrence ↑ functionality

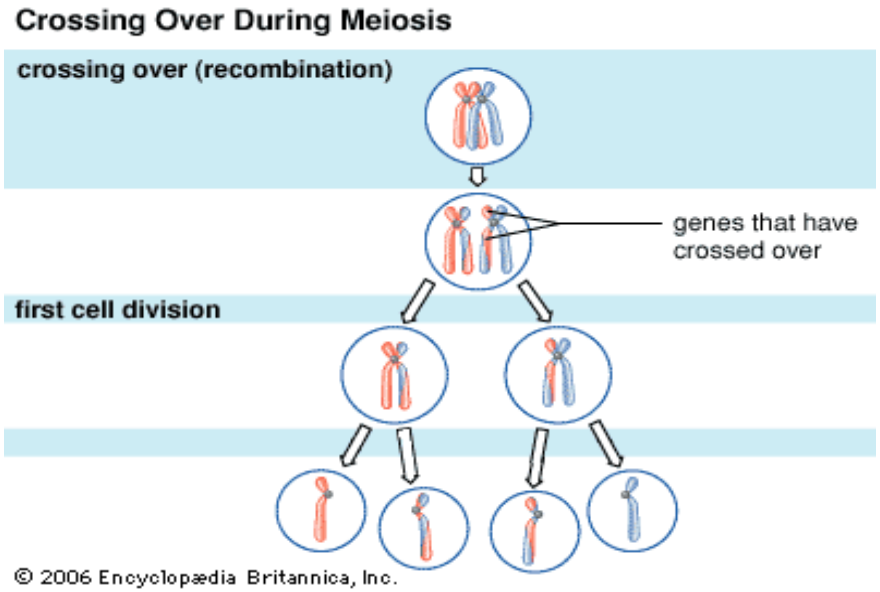
$$t = -\frac{c_{\text{correct}}\cdot p_h(\text{hom})\cdot p_f(\text{hom})}{c_{\text{incorrect}}\cdot p_h(\text{non-hom})\cdot p_f(\text{non-hom})}$$

t increases → the system is less tolerant to errors

Homologues Recombination

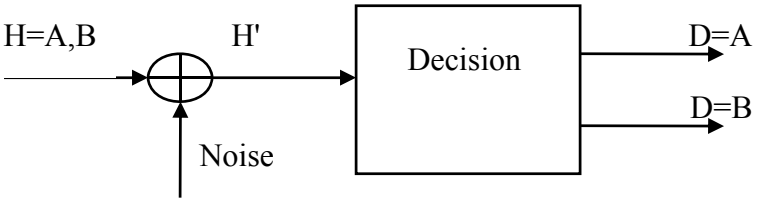
- Genome repair
- Generating genetic diversity





Bugreev et al, *Nature Structural & Molecular Biology 14, 746 - 753 (2007)*

Molecular recognition as signal detection



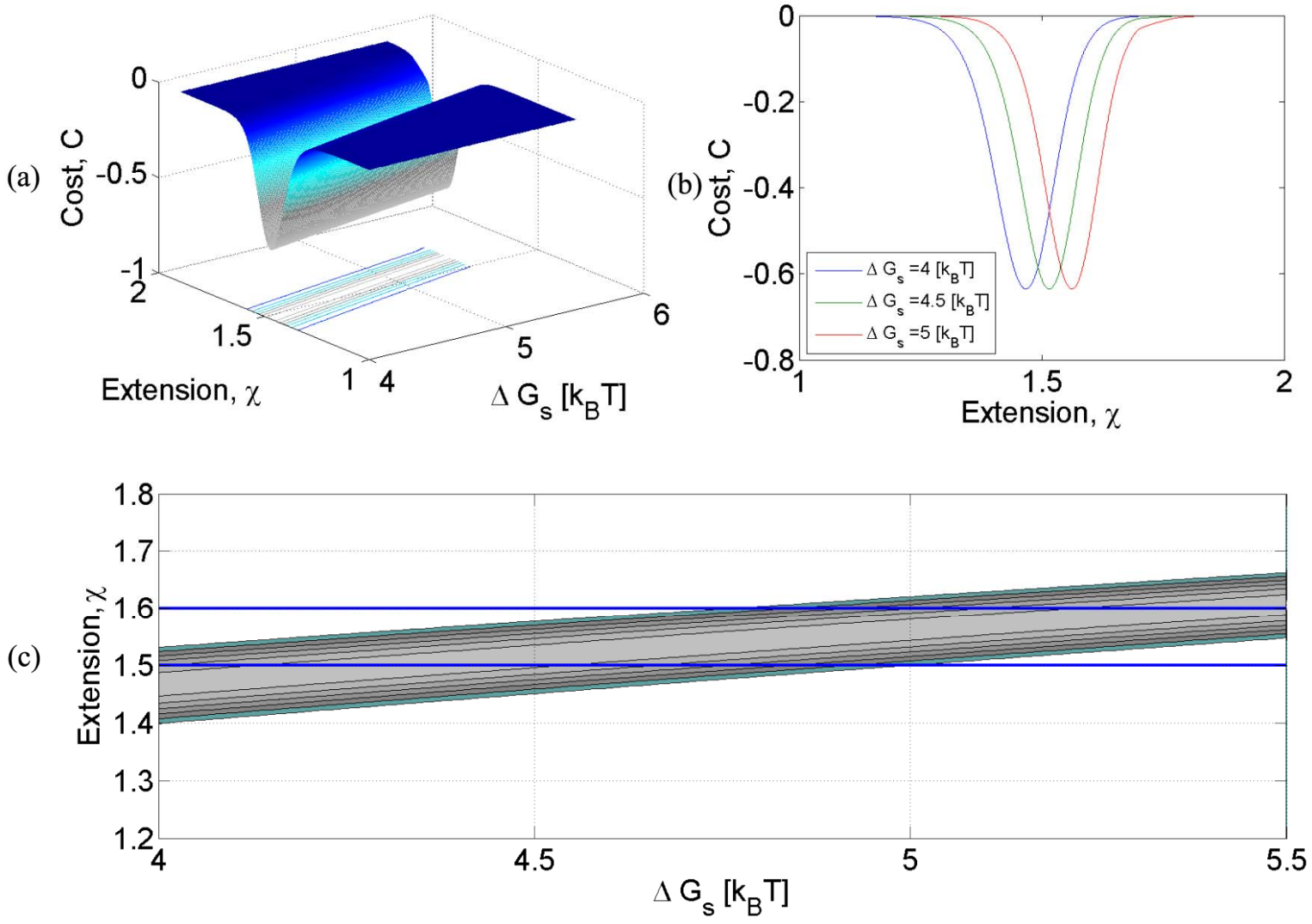
In some systems, misidentification may cause no harm while in others it may be fatal.

Decision, D	True hypothesis, H	
	A	B
A	Correct, C_{AA}	Type I ("False Alarm"), C_{AB}
B	Type II ("Miss"), C_{BA}	Correct, C_{BB}

$$Ec = \sum_{i,j} c_{ij} p_{ij} = \sum_{i,j} c_{ij} p_h(j) p_d(i \mid j)$$

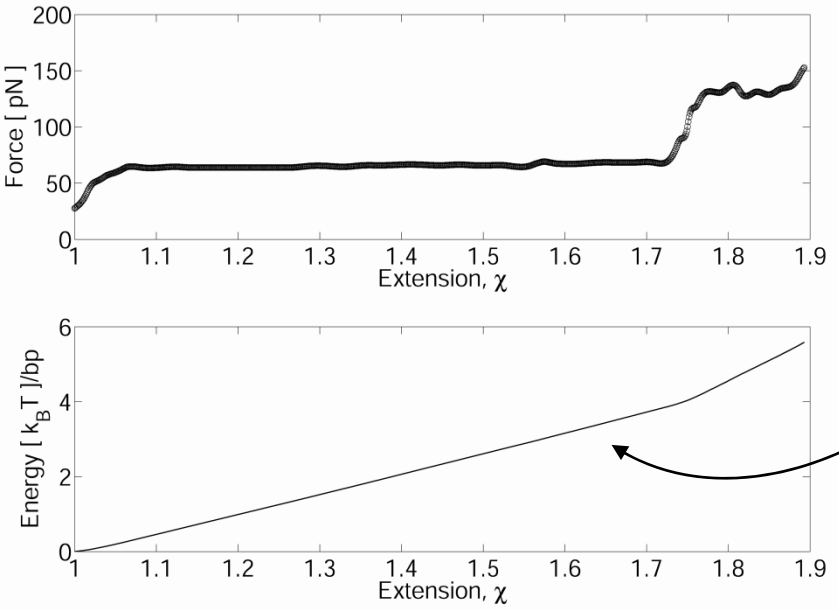
- The *average cost:*

Cost exhibits minimum at optimal extension



- **One** RecA monomer, one mismatch, symmetric: $N = 3,\ m = 1,\ \Delta\Delta G_{mis} = 3k_BT,\ t = 1$

Analytic approximation



linear with good approximation

$$\Delta G_{ext} = N \cdot \Delta G_{str}(\chi - 1) + \Delta G_{int}(\chi - 1)^2$$

specific interactions

stretching interaction

$$\overline{\chi} = \frac{N \cdot \Delta G_{str}}{2\Delta G_{int}}\left[\sqrt{1 + \frac{4\Delta G_{int}(\Delta G_s - \frac{m}{2N}\Delta\Delta G_{mis})}{N \cdot \Delta G_{str}^2}} - 1\right] \Rightarrow \overline{\chi} \cong \frac{N}{6}\left[\sqrt{1 + \frac{1}{N}(10 - \frac{3m}{N})} - 1\right]$$

interfacial interaction

destabilization due to mismatch