

The Mathematics of gene descrambling

Mark Daley

Depts. of Computer Science & Biology

University of Western Ontario

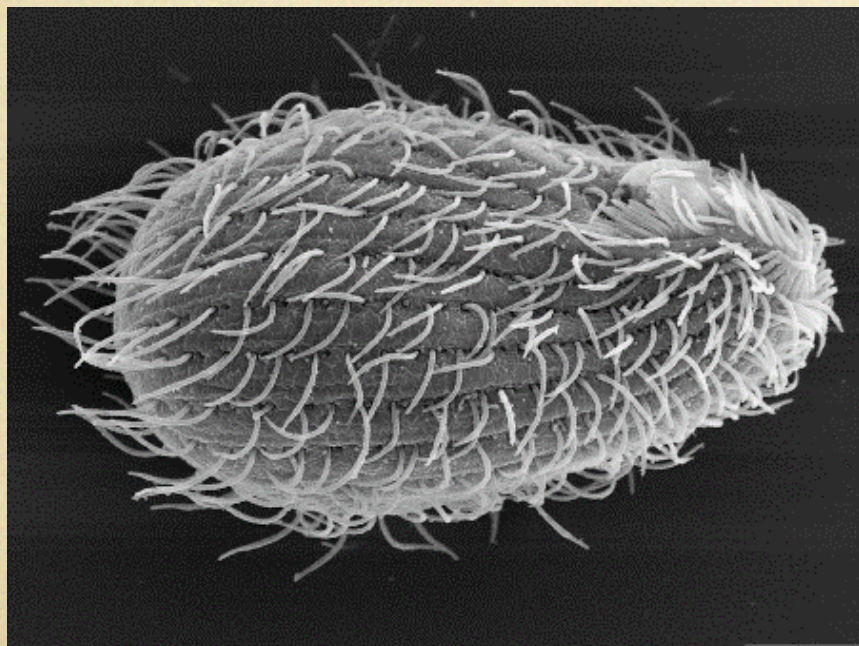
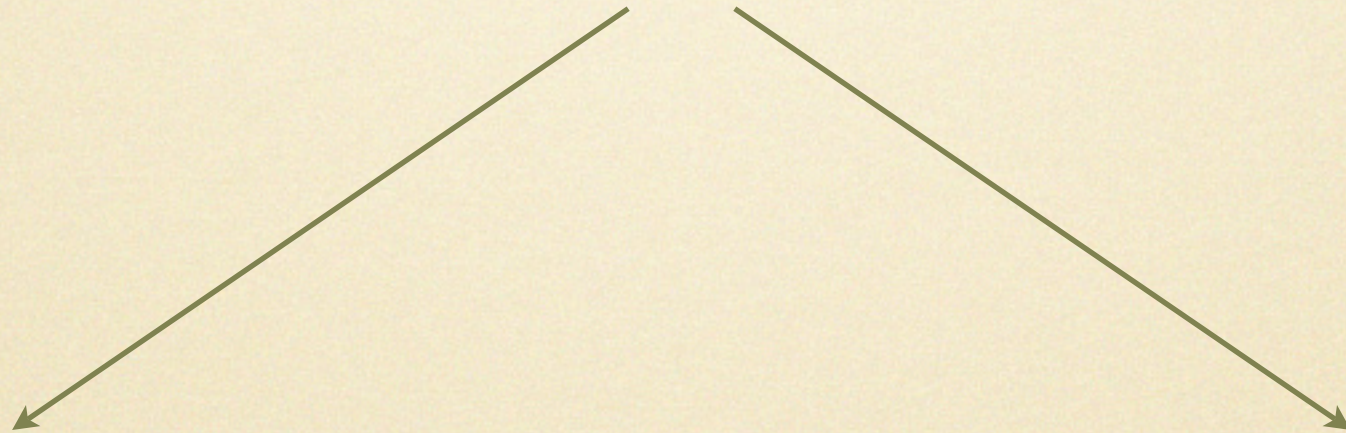
A CS problem c. the 1980s...

How do we increase processing power?

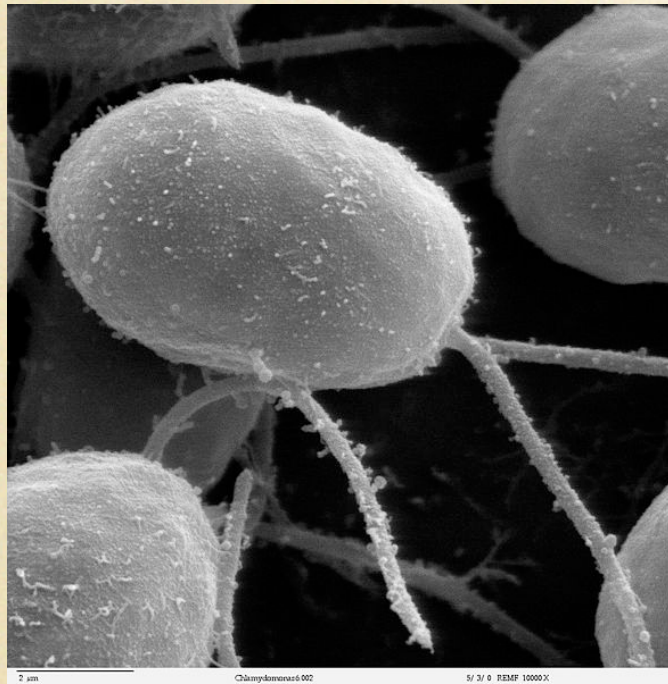


An evolutionary problem c. 1 billion ya

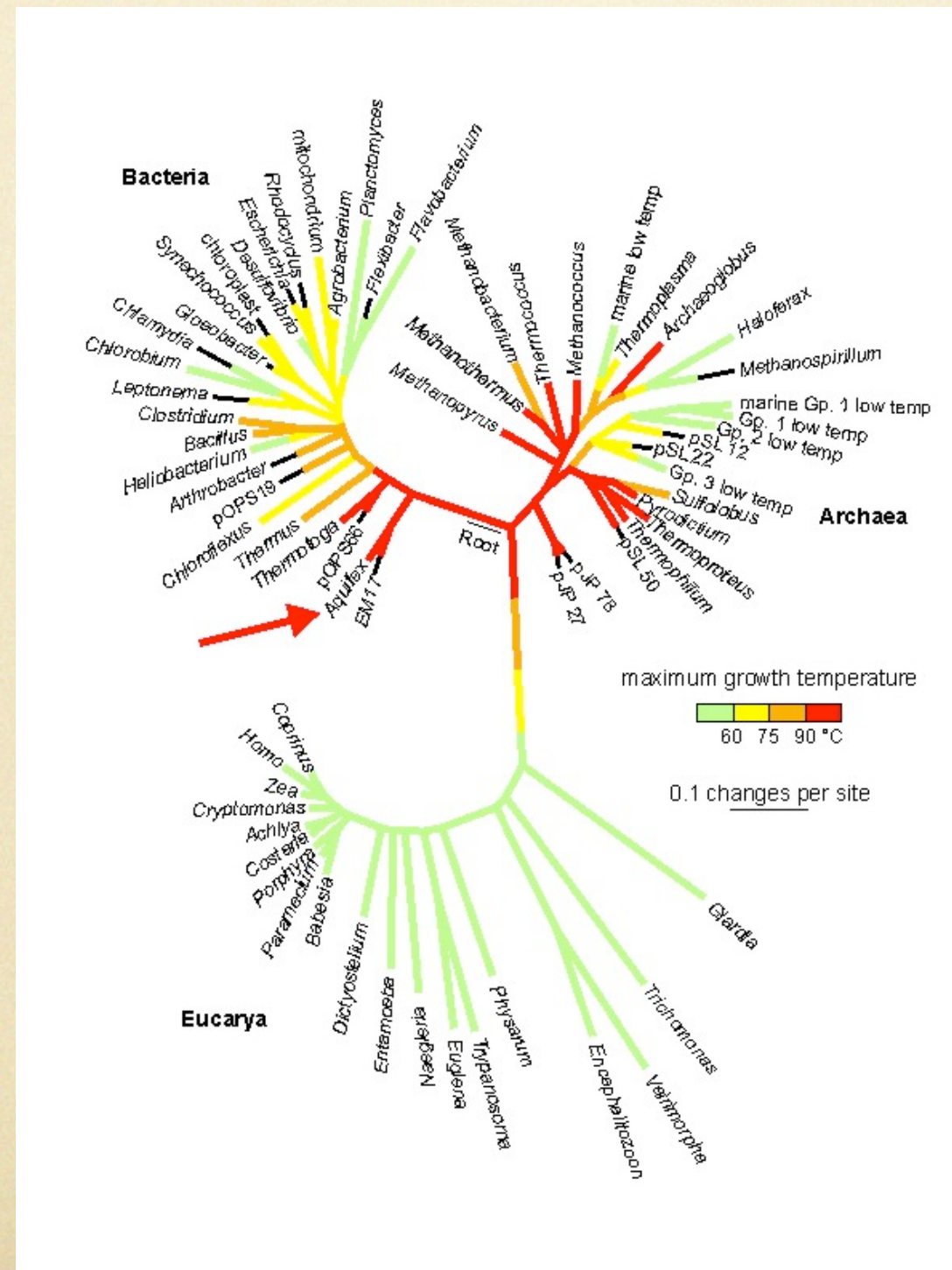
How do we increase “processing” power?



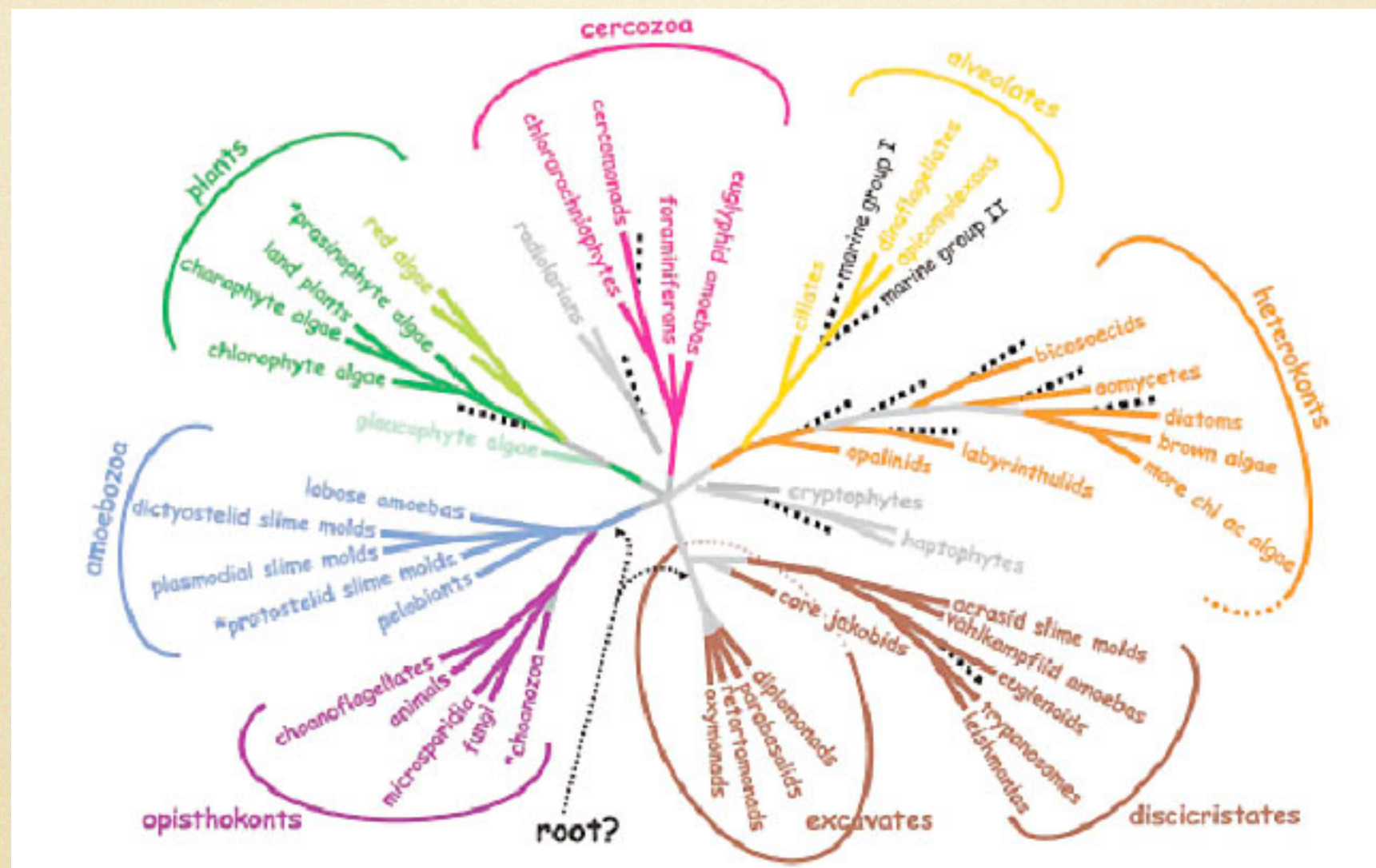
Protozoa



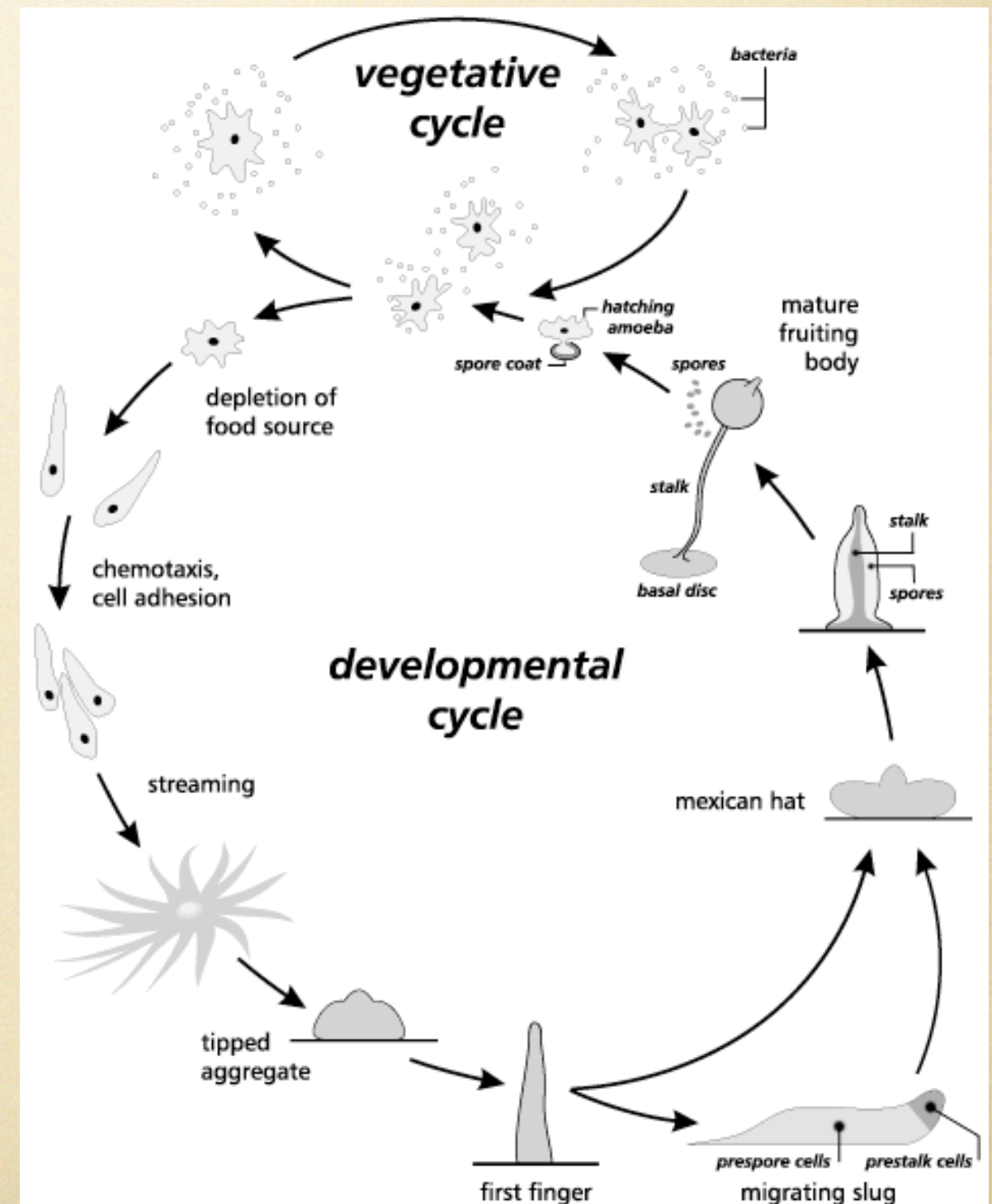
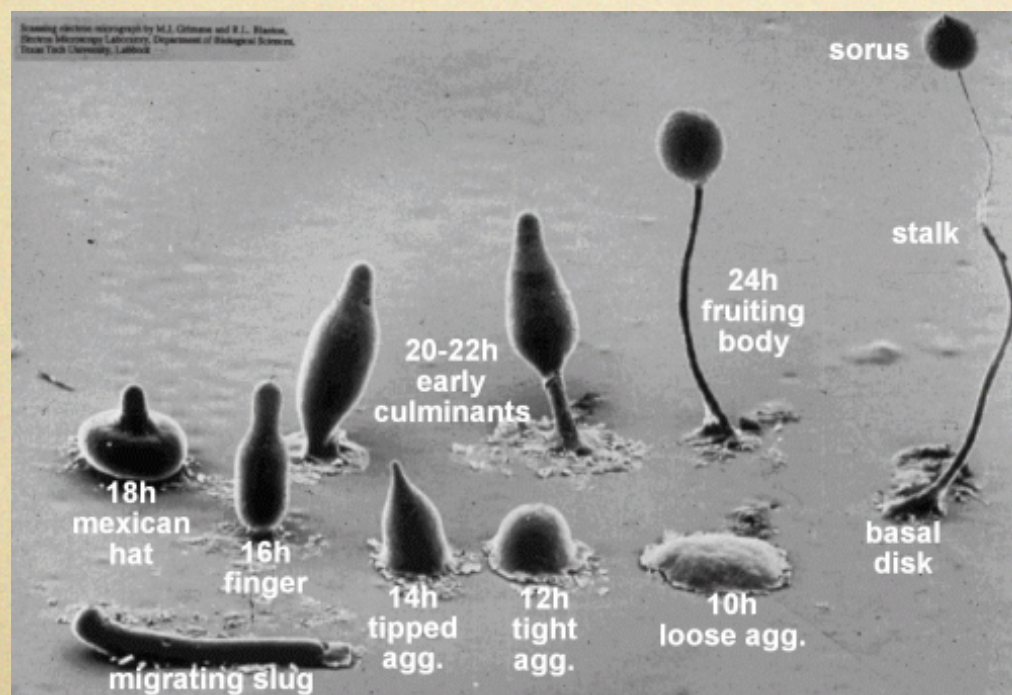
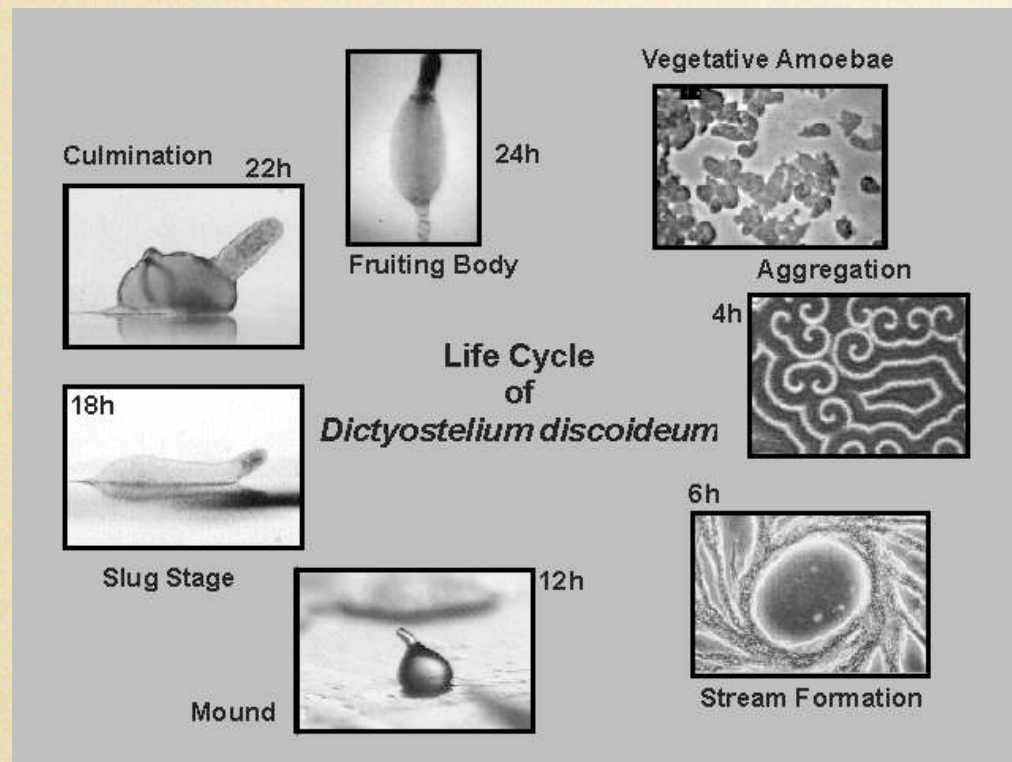
Tree of life



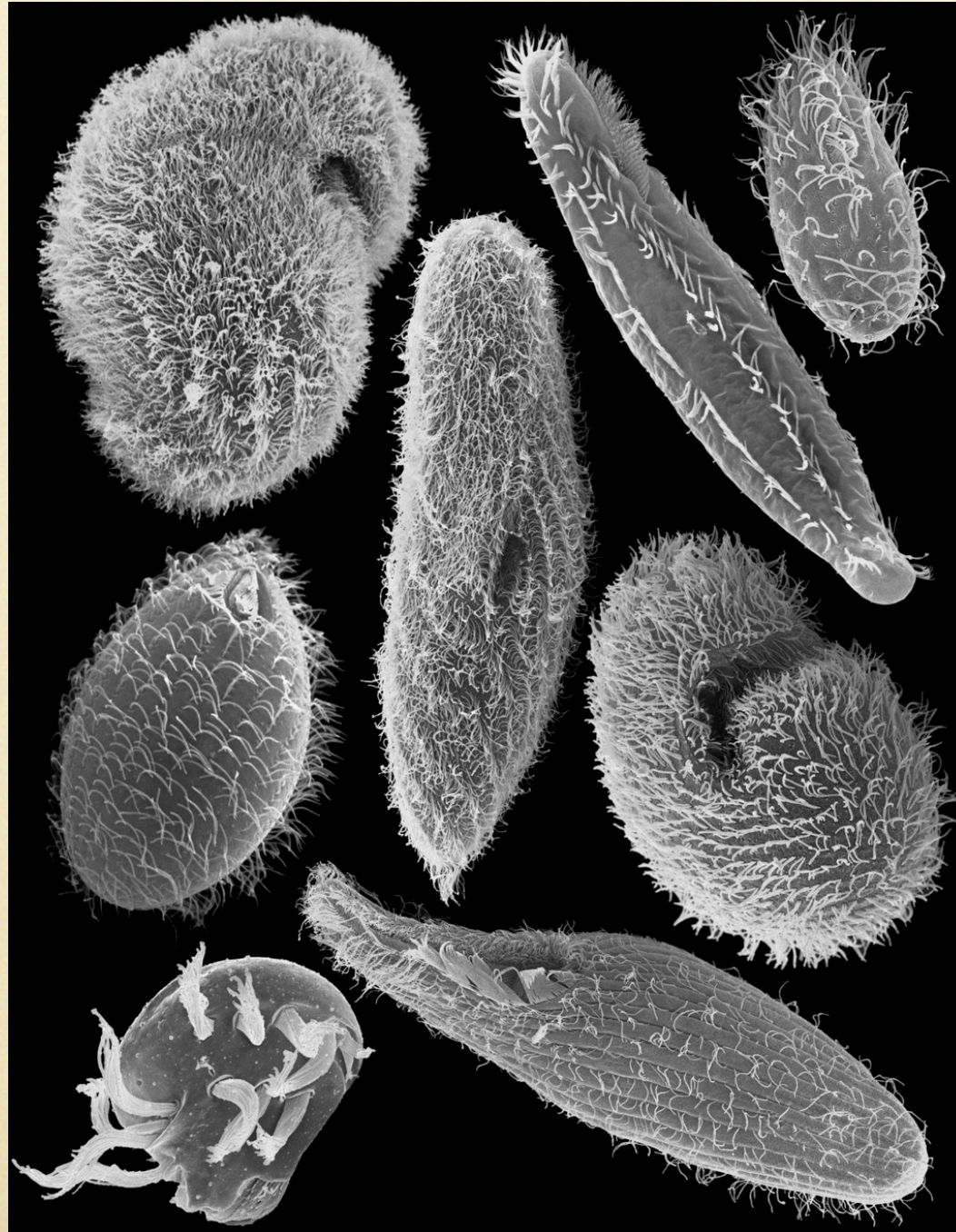
Eukaryotic tree of life



Aside: "Missing link"



Ciliates



Habitat



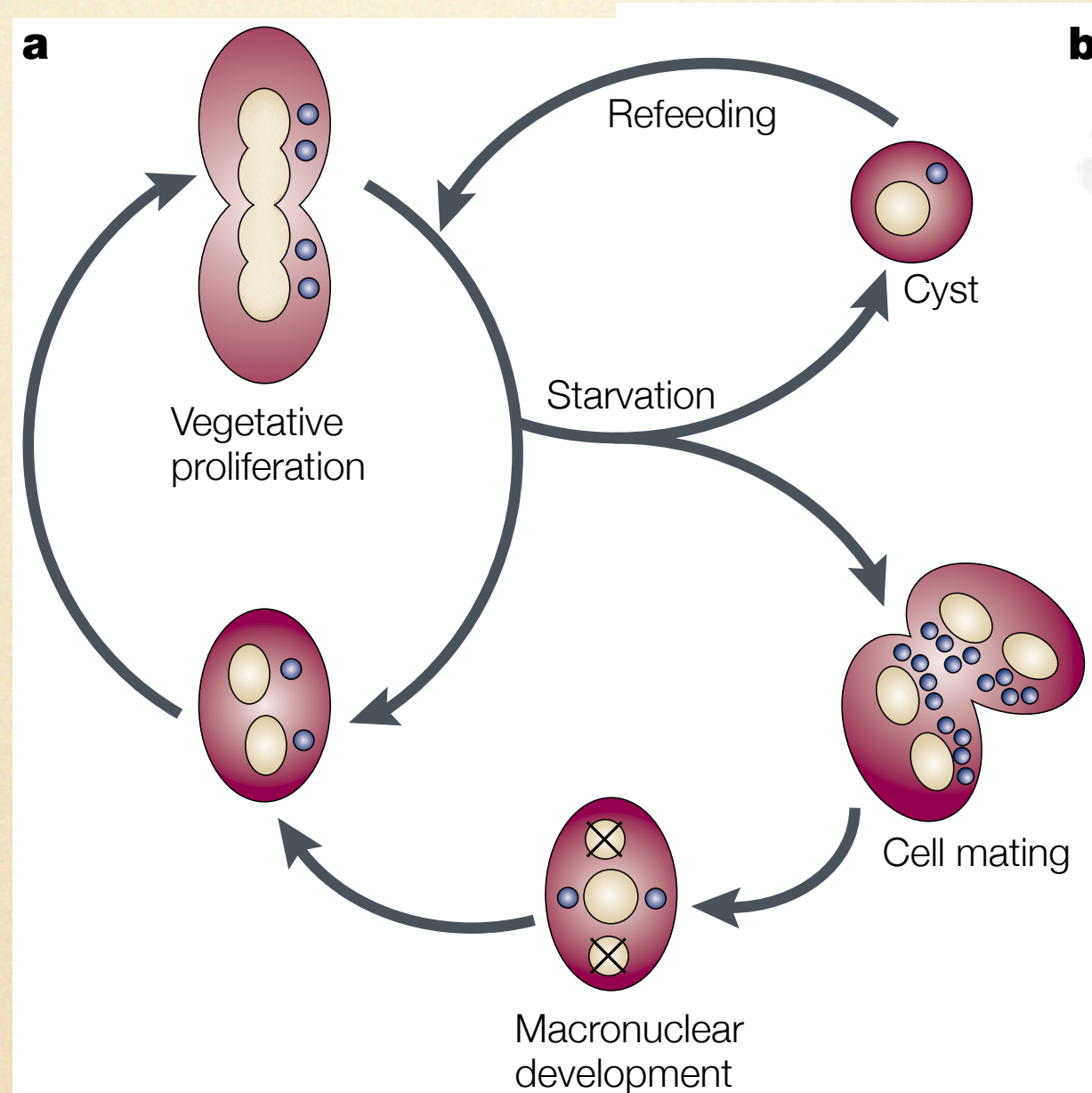
Puddle dwellers

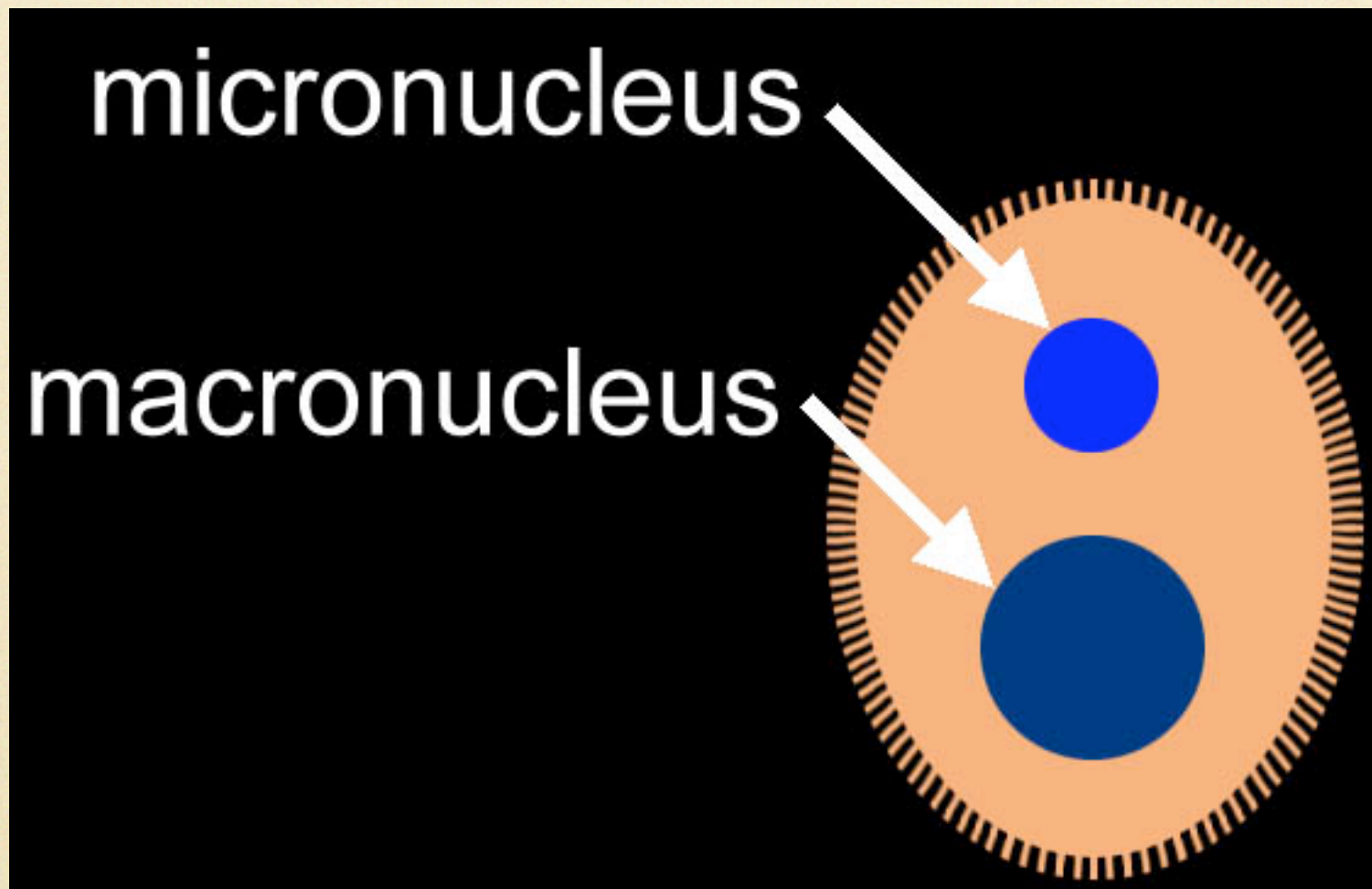


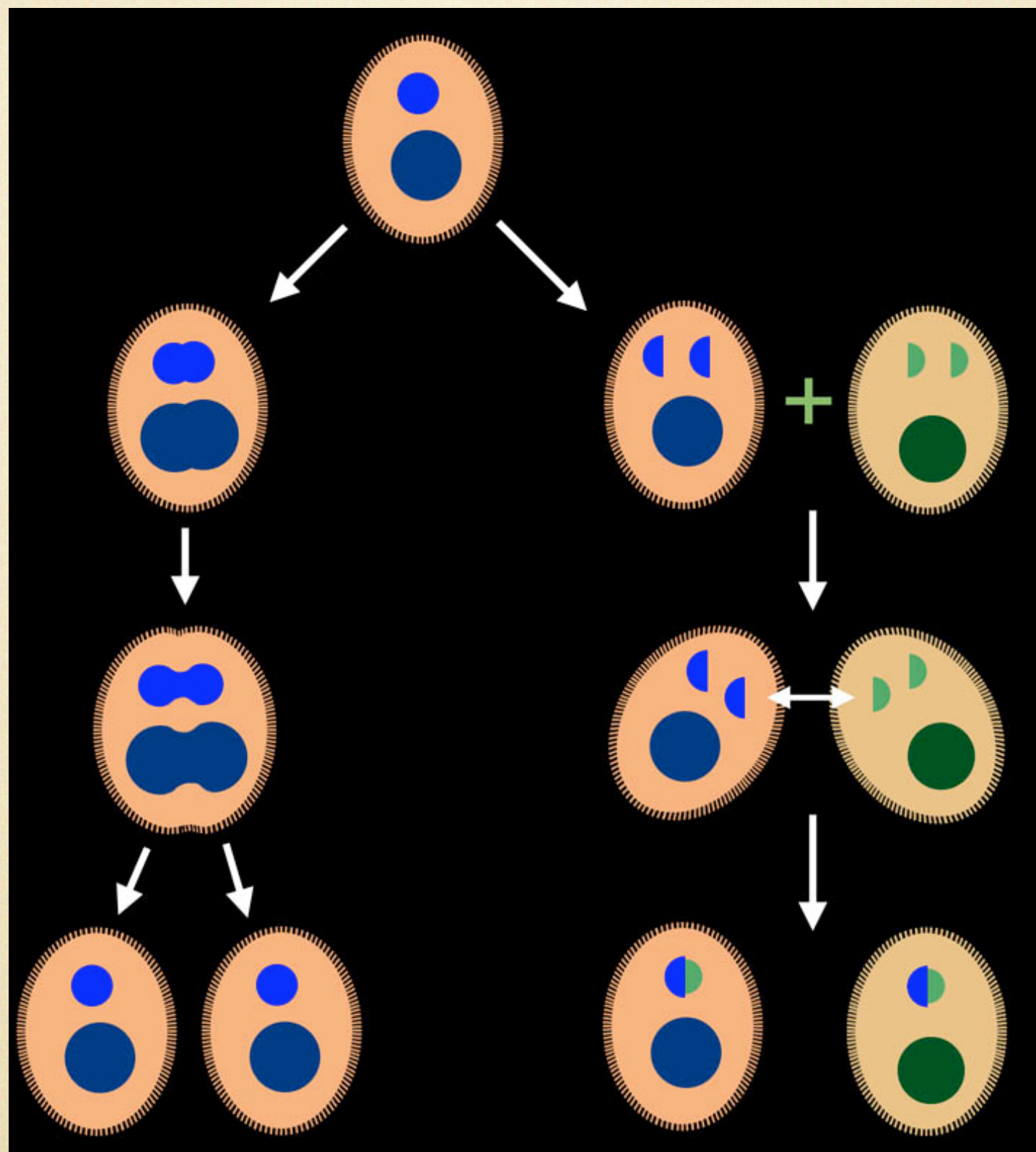
Fish Tank Dweller

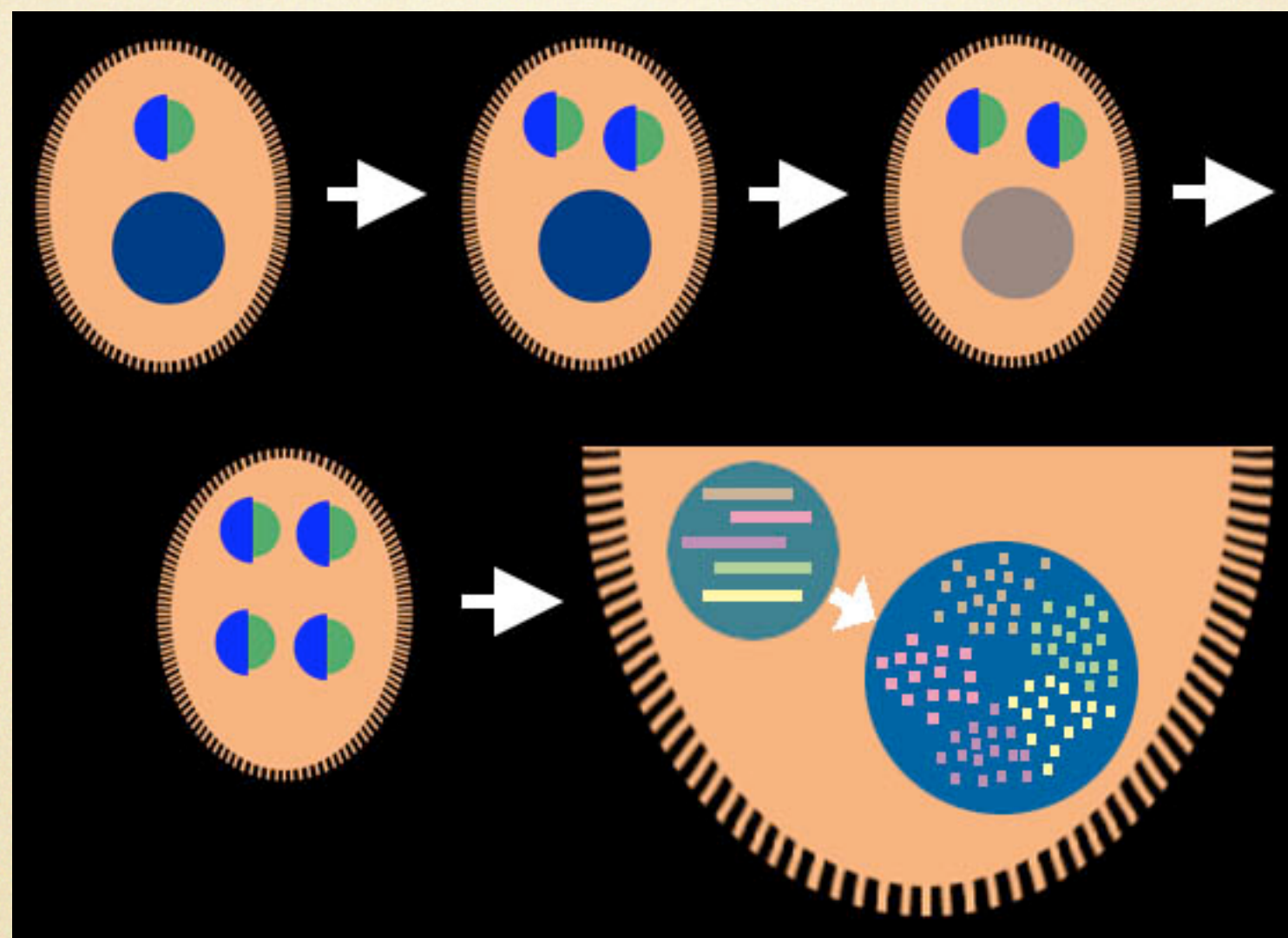


Ciliate Life Cycle

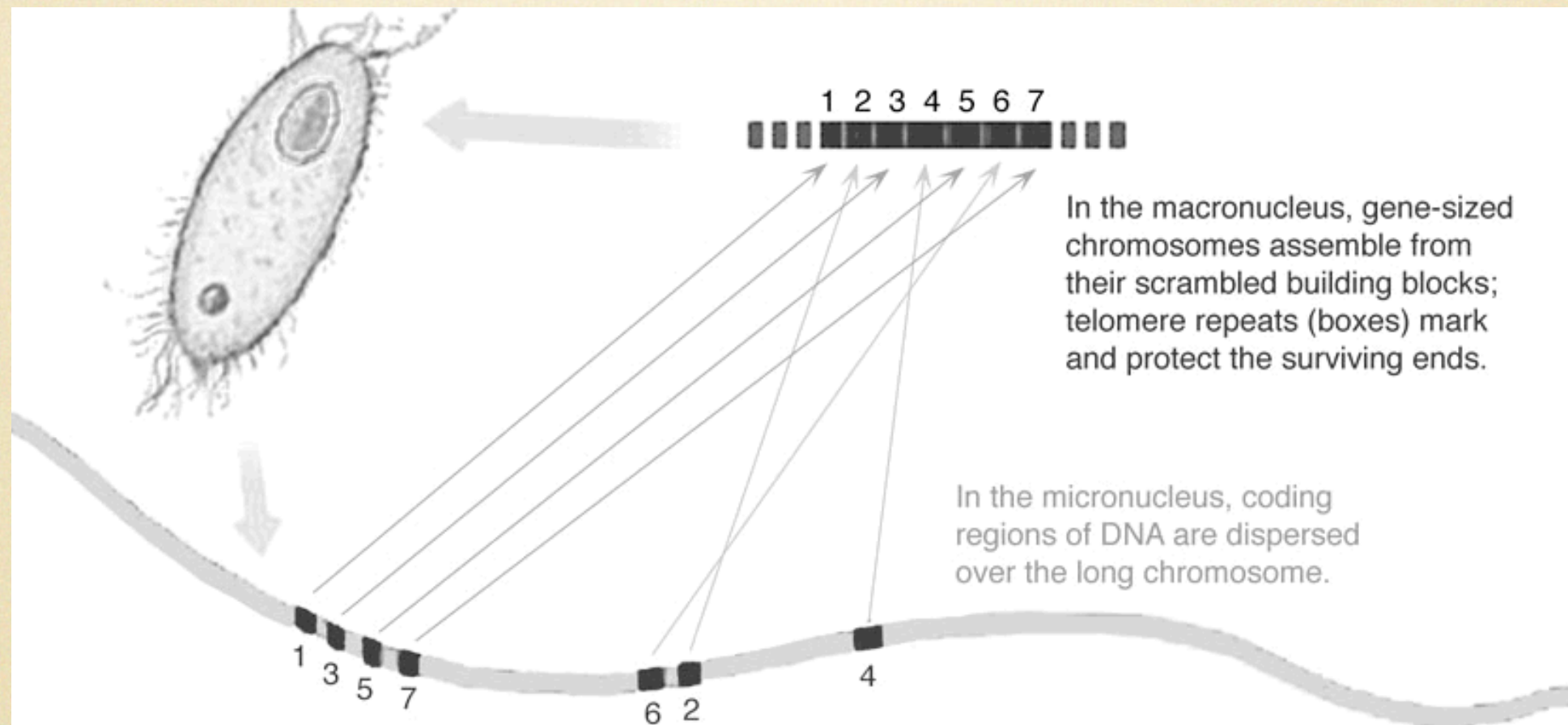




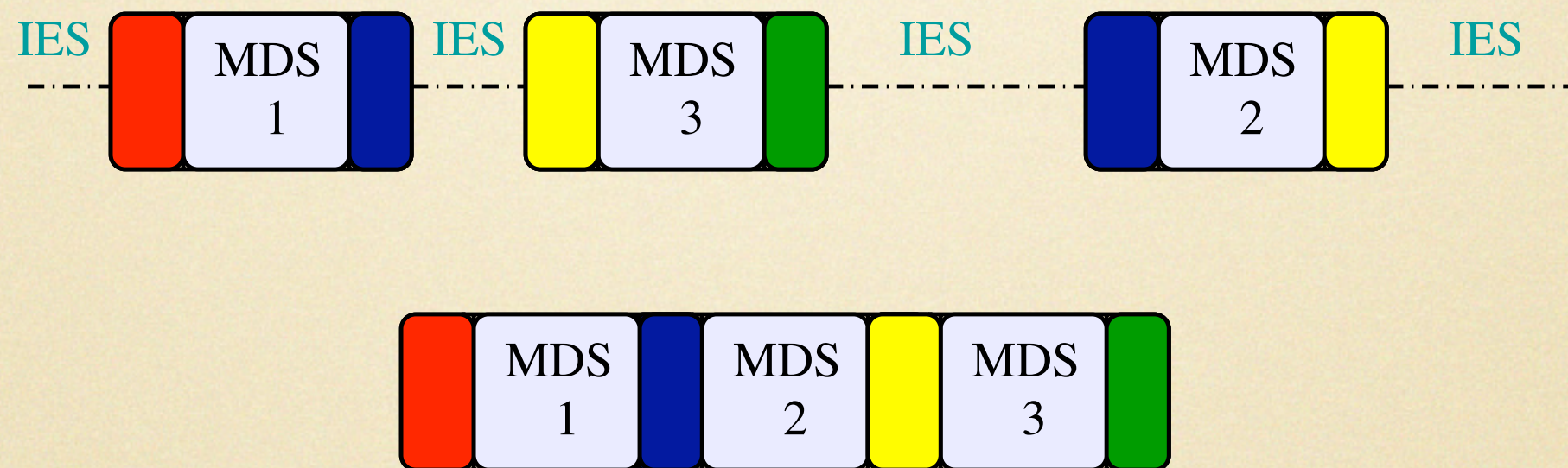




Stichotrich Subtlety



A Scrambled Gene



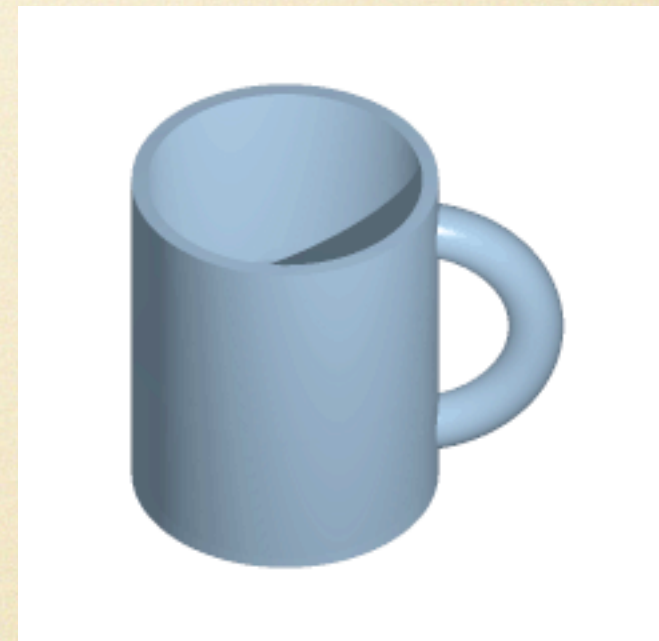
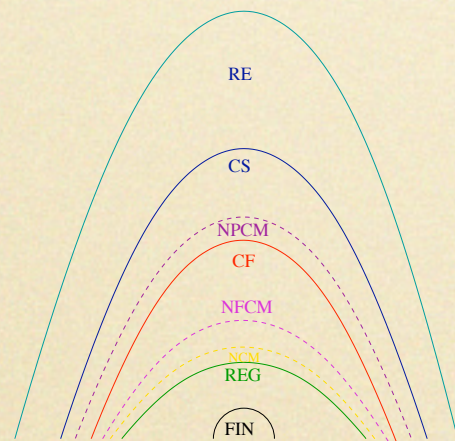
- How does the ciliate turn a scrambled micronuclear gene into a functional macronuclear gene?

My Tools

$$\Sigma = \{a, b, c, d\}$$

$$L = \{a, aab, bdcc, ddaa, cc\} \subseteq \Sigma^+$$

$$L = \{w \mid w \text{ has an even number of 'a's'}\} \subseteq \Sigma^+$$



Formal Language Theory

Topology

Prop $\deg Sf = \deg f$, where $Sf: S^{\infty} \rightarrow S^{\infty}$ is a separable of the map $f: S^n \rightarrow S^n$.

Proof:

Let CS denote the cone $(S^n \times I)/(S^n \times I) \cong \text{base}$
 $S^n = S^n \times 0 \subset CS^n$, so CS^n/S^n is a sep. of S^n . The
 map f induces $Cf: (CS^n, S^n) \rightarrow (CS^n, S^n) \cong \text{quotient}$
 Sf . Prop

Boundary maps in the l.s.s. of $(CS^n, S^n) \Rightarrow \text{diag commutes.}^{(4)}$
 Thus, if f_* is nul. by d, so is Sf_* .

$$(1) \quad \begin{array}{ccc} \tilde{H}_{n+1}(S^{n+1}) & \xrightarrow{\partial} & \tilde{H}_n(S^n) \\ \downarrow Sf_* & & \downarrow f_* \\ \tilde{H}_{n+1}(S^{n+1}) & \xrightarrow{\partial} & \tilde{H}_n(S^n) \end{array}$$

That is: $\partial \circ Sf_* \Rightarrow f_* \circ \partial$.

Smoothings:

$L1 \rightarrow \infty$ inverts
 $L1 \rightarrow 0$ breaks

$a \mid x \mid b$
 $\downarrow$
 $a \mp b$

$L1 \rightarrow L\infty$

$a \mid b \mid 2^+ c \mid 2^- d \mid e$

$L1 \rightarrow \infty$:

$T_1, abT_2cd \Leftrightarrow T_1, a\bar{c}\bar{T}_2\bar{b}d$

$L1 \rightarrow 0$:

$T_1, abT_2cd \Leftrightarrow T_1, ad + bT_2c$

$L0 \rightarrow \infty$:

$T_1, ad + T_2cb \Leftrightarrow T_1, a\bar{c}\bar{T}_2\bar{b}d$

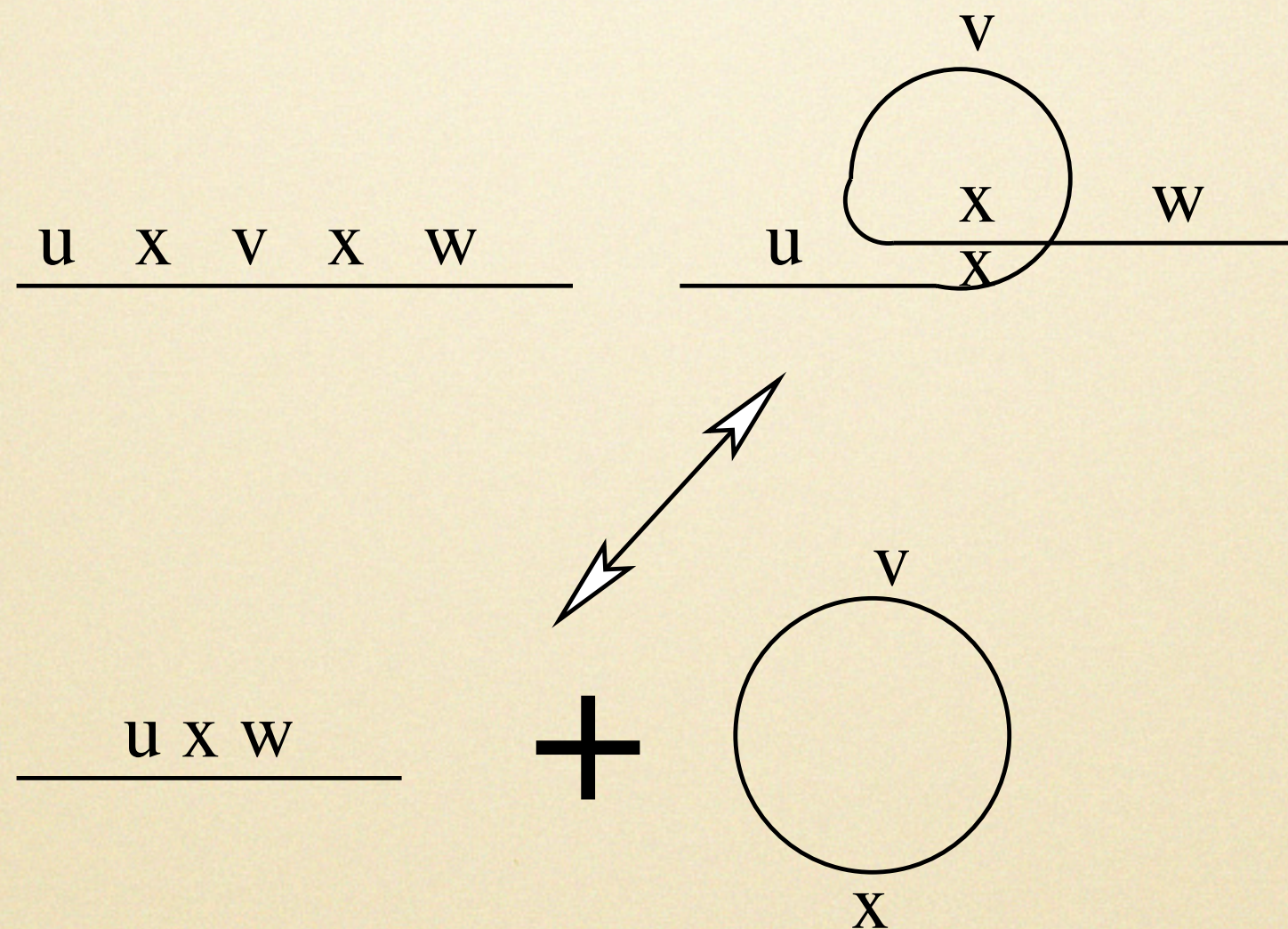
How?

The naïve model

$$\prod_{i=1}^n C(\gamma(i), 2)$$

- *S. lemnae* DNA pol. α gene, on average:
- **14,445,331,655,040**
attempts to get a functional protein.

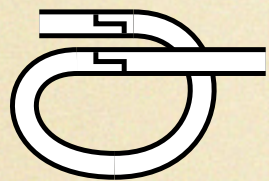
Synchronized Insertion & Deletion



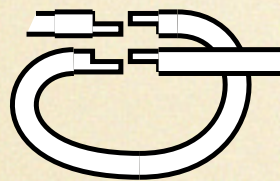
Kari & Landweber 1999

P = ordered set of all pointers
while G_{seq} is not descrambled **do**
 Remove from P any pointers that appear in
 • G_{seq} in the form $p_n p_n$
 $p = P_0$
 excise(p)
 Find minimal i such that $p_i \in P - \{p\}$, $p_i \in G_{seq}$
 and $p_i \in G_{plasmid}$
 insert(p_i)
 append (p, p_i) to d-vector
 $P = P - \{p, p_i\}$
end while

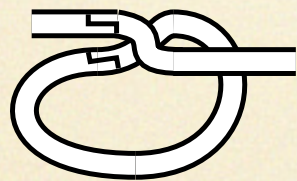
hi/dlad/ld



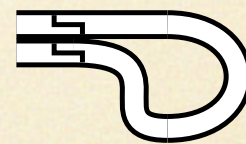
ld(*i*)



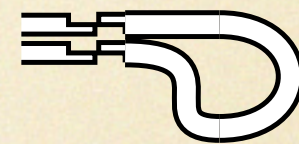
ld(*ii*)



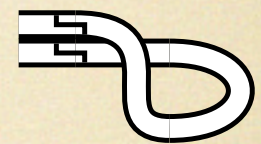
ld(*iii*)



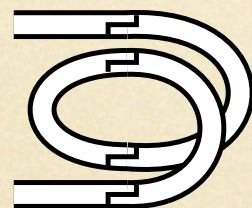
hi(*i*)



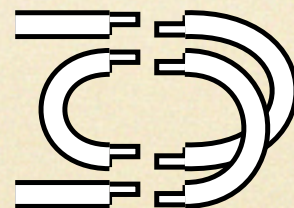
hi(*ii*)



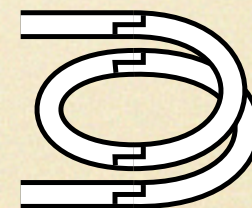
hi(*iii*)



dlad(*i*)

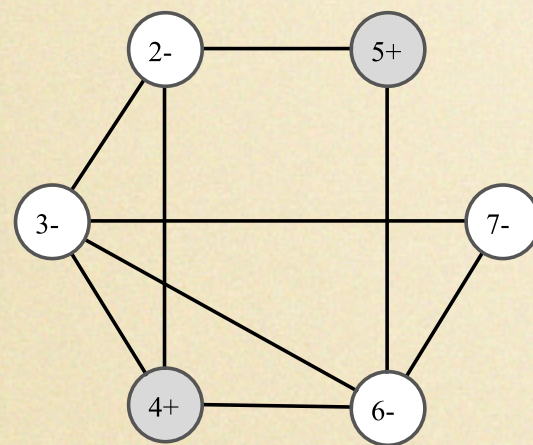


dlad(*ii*)

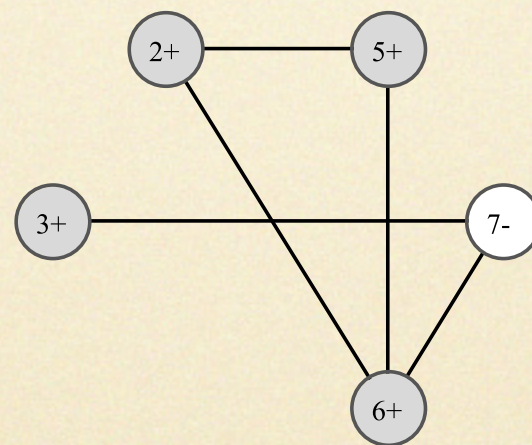


dlad(*iii*)

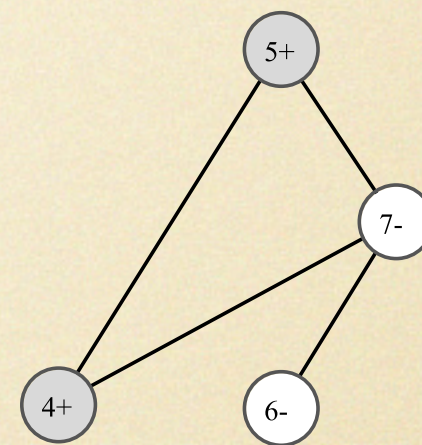
Prescott, Ehrenfeucht & Rozenberg 1999



(a)



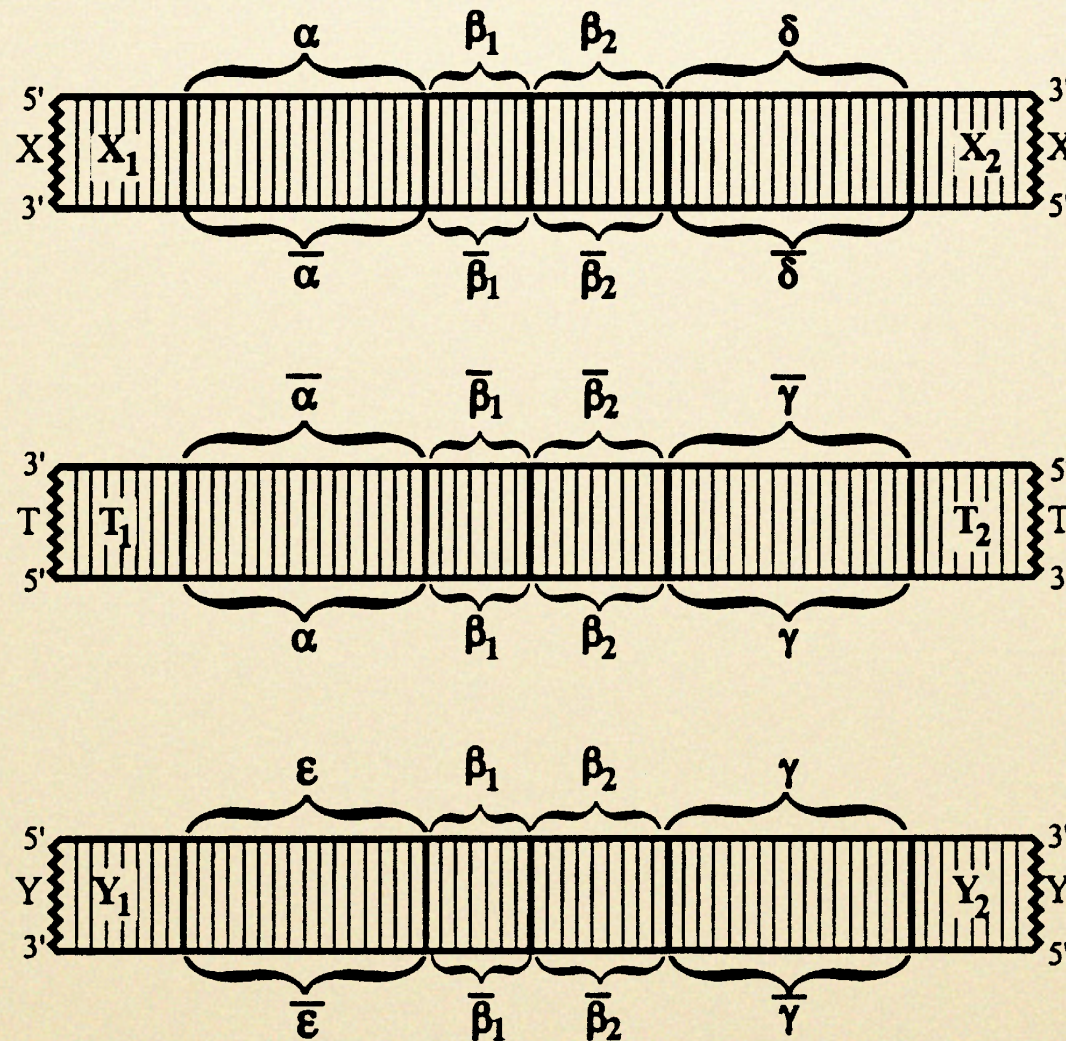
(b)



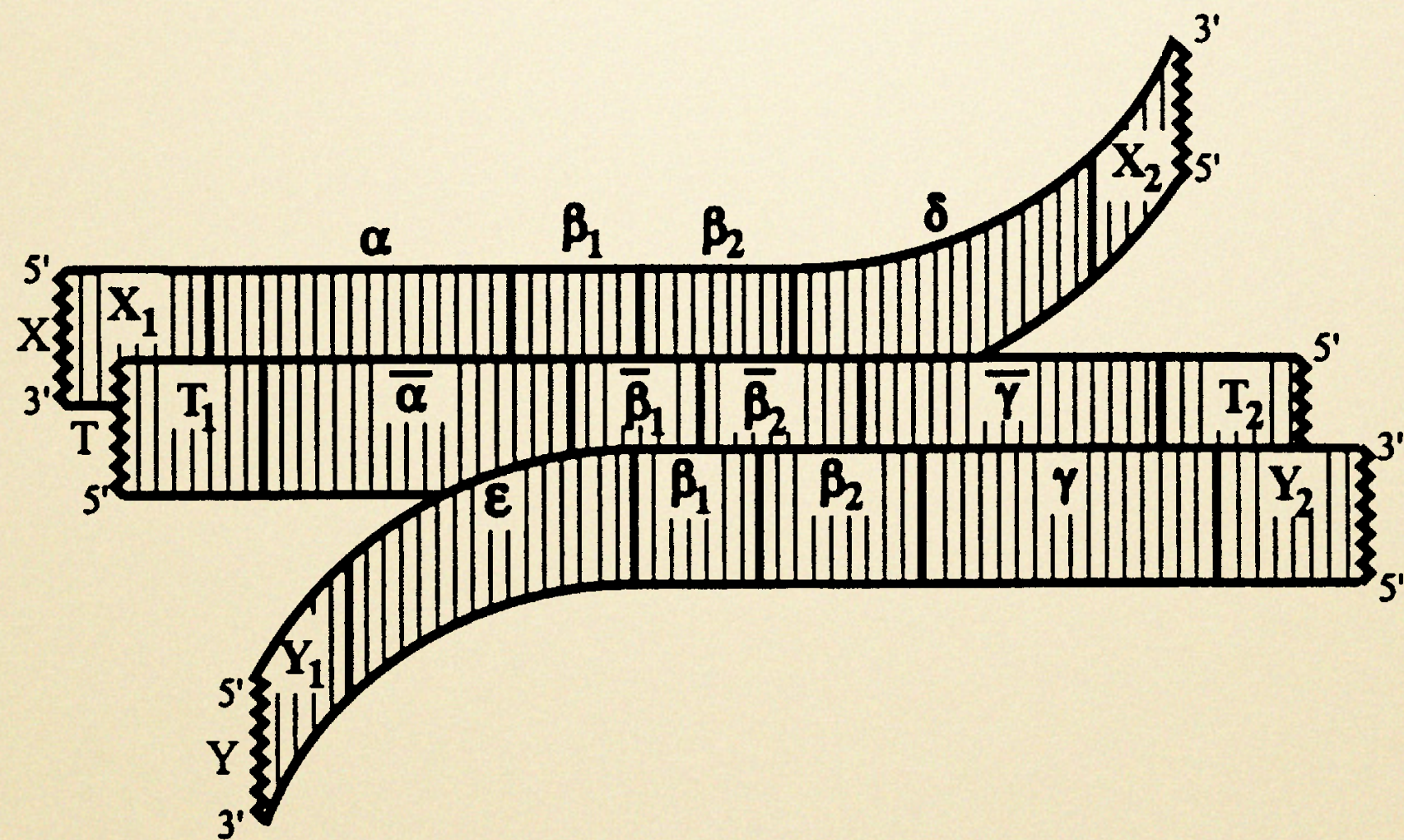
(c)

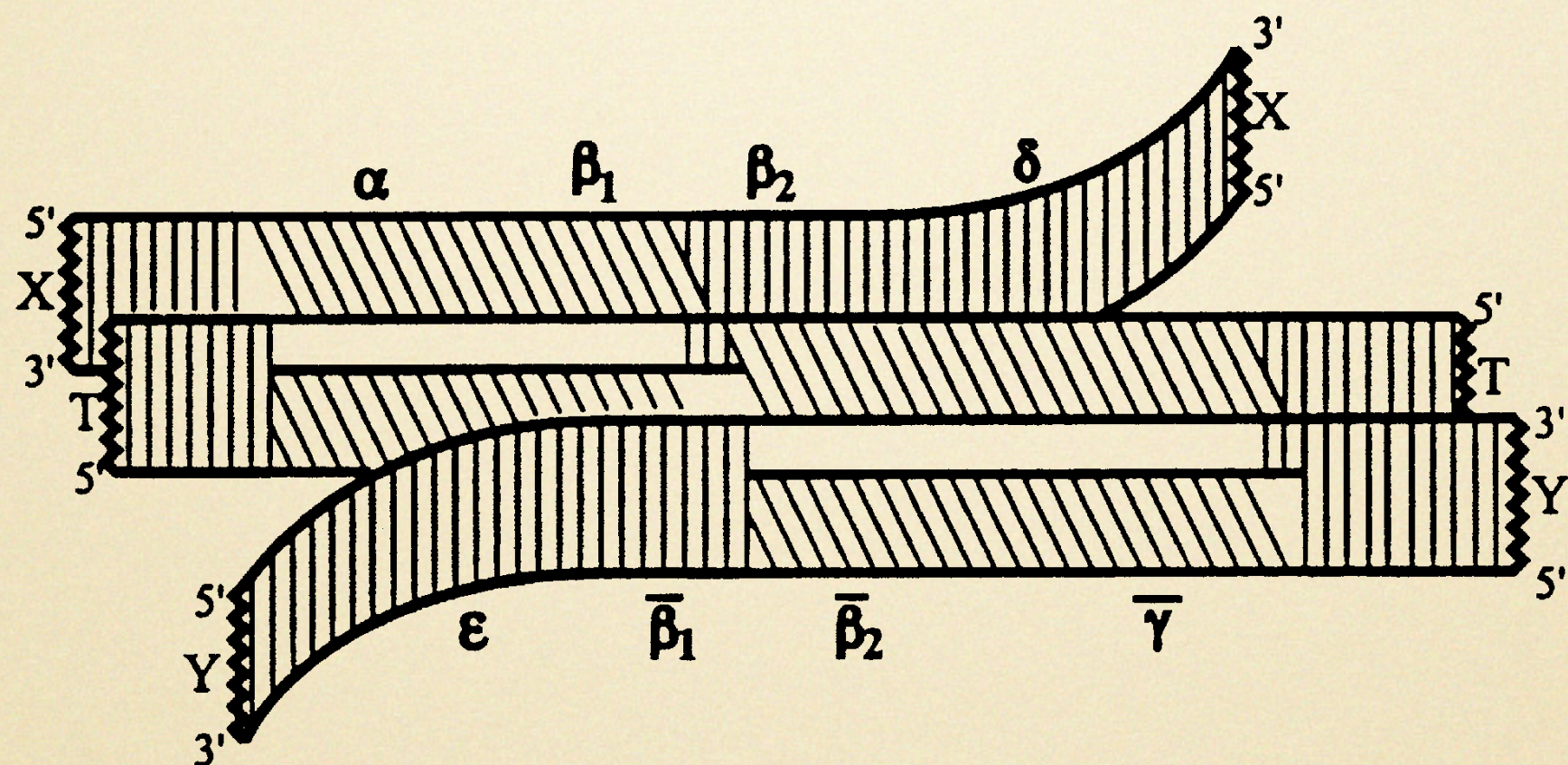
Figure 5: (a) The graph G_u in Example 2; (b) $hi_4(G_u)$; (c) $dlad_{2,3}(G_u)$

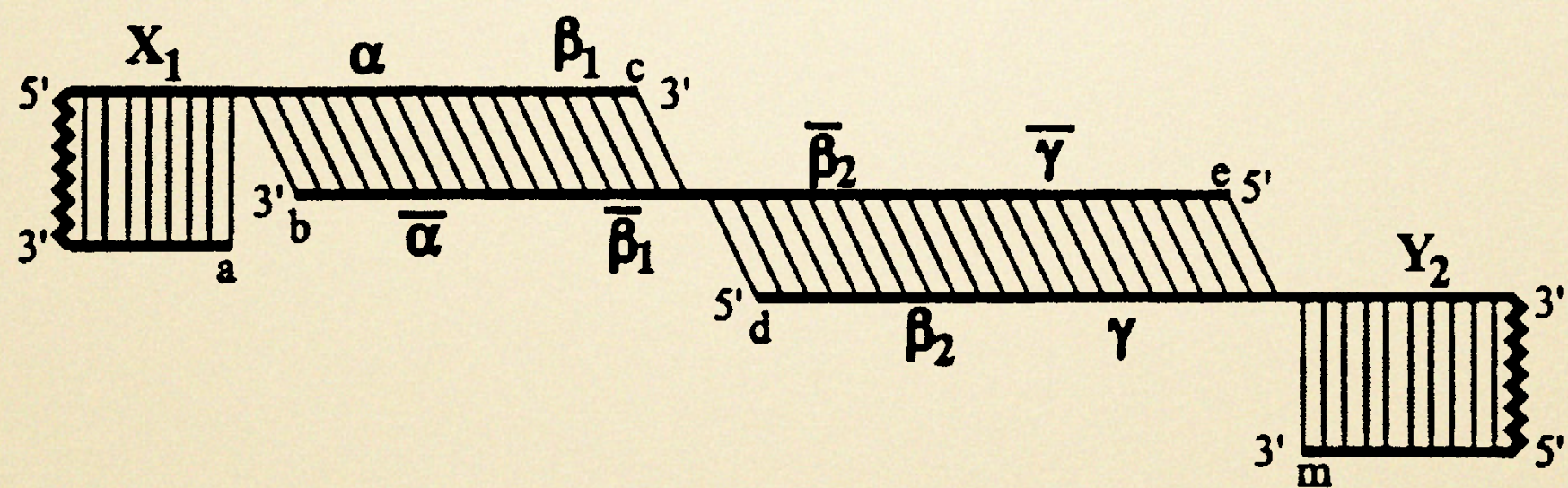
Template-guided Recombination



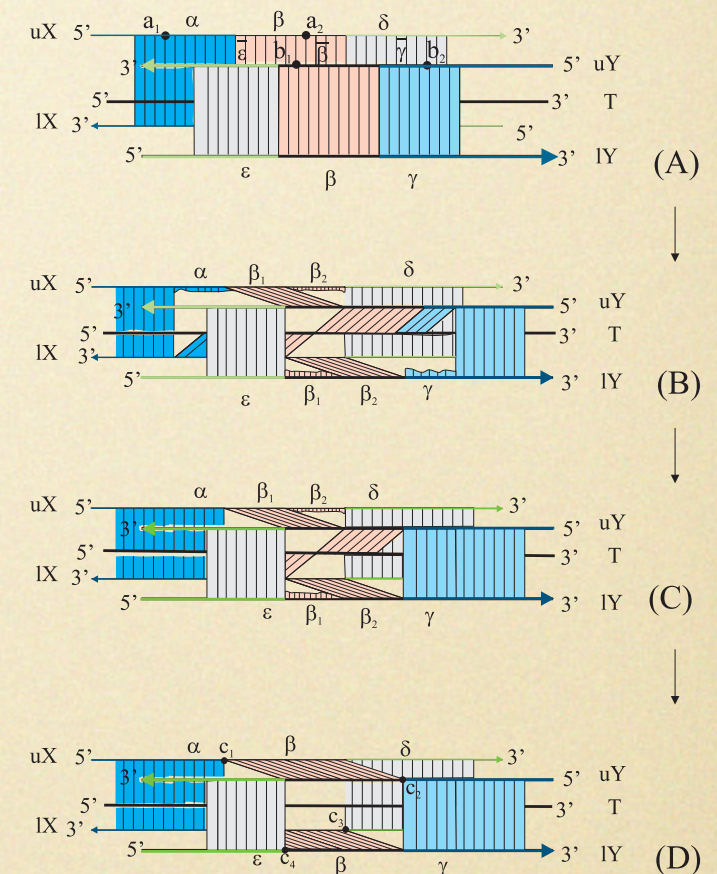
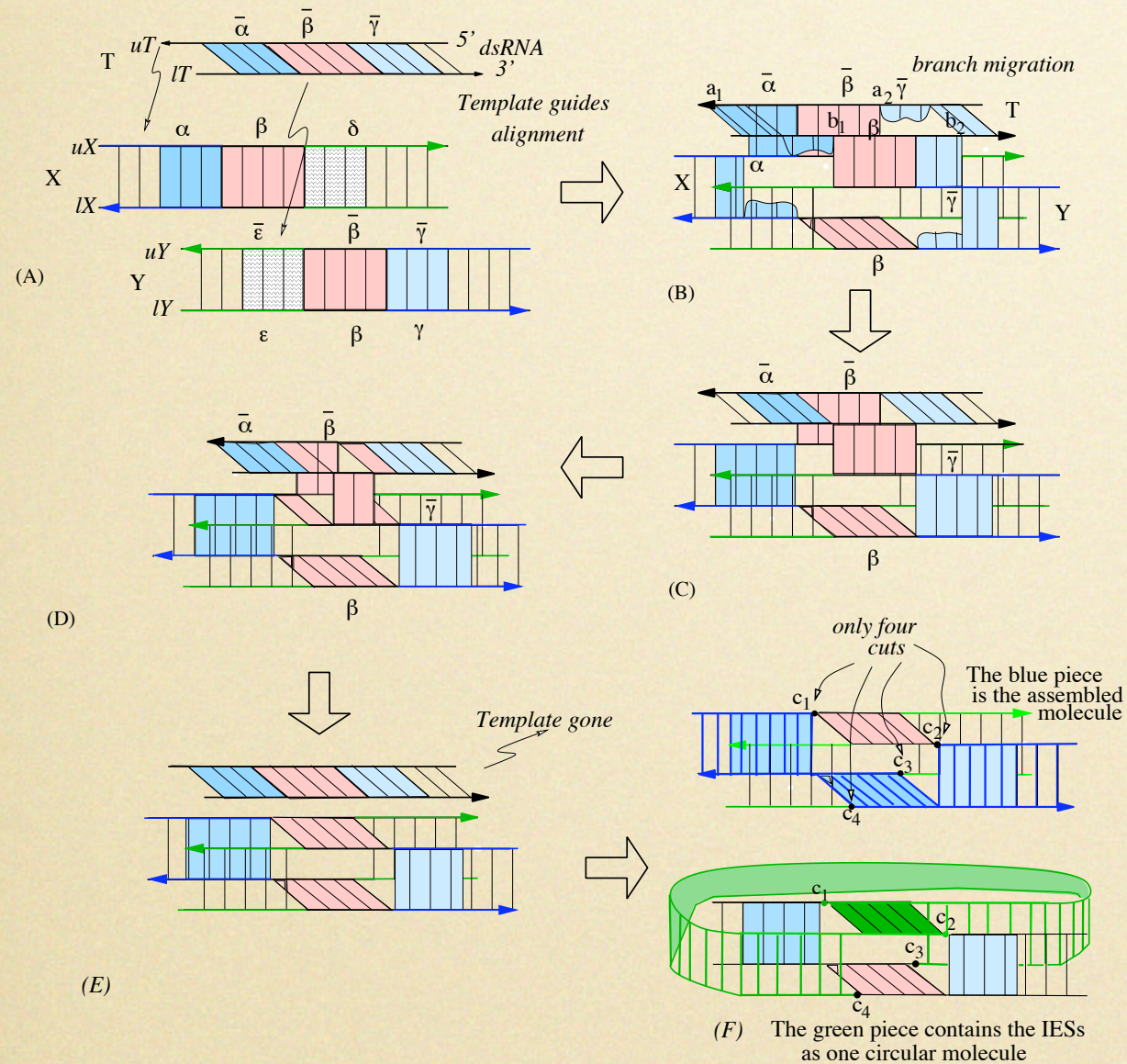
Prescott, Ehrenfeucht & Rozenberg 2003







RNA templates



Angeleska, Jonoska, Saito & Landweber 2007

Formal TGR

Let $x, y \in \Sigma^*$. We define

$$x \dot{\cup}_T y = \{u\alpha\beta\gamma v \mid x = u\alpha\beta d, y = e\beta\gamma v, \alpha\beta\gamma \in T, u, v, d, e \in \Sigma^*, \alpha, \beta, \gamma \in \Sigma^+\}.$$

$$\begin{array}{cccccc} u & \alpha & \beta & d & & \\ & \alpha & \beta & \gamma & & \\ & e & \beta & \gamma & v & \\ \hline u & \alpha & \beta & \gamma & v & \end{array}$$

(with I. McQuillan & M. Domaratzki)

We generalize $\dot{\cup}_T$ to languages by $L_1 \dot{\cup}_T L_2 = \{x \dot{\cup}_T y \mid x \in L_1, y \in L_2\}$.

Let $\mathcal{L}_1, \mathcal{L}_2$ be language families. We denote by $\mathcal{L}_1 \dot{\cup}_{\mathcal{L}_2} \mathcal{L}_1 = \{L_1 \dot{\cup}_T L_2 \mid L_1, L_2 \in \mathcal{L}_1, T \in \mathbb{L}L_2\}$.

We say that a language family $\mathcal{L}_1$ is closed under $\dot{\cup}_{\mathcal{L}_2}$ if and only if $\mathcal{L}_1 \dot{\cup}_{\mathcal{L}_2} \mathcal{L}_1 \subseteq \mathcal{L}_1$.

Every concatenation-closed full trio L_1 is closed under $\cap_{L_2}$ iff L_1 is closed under $\cap$ with L_2 (a trio).

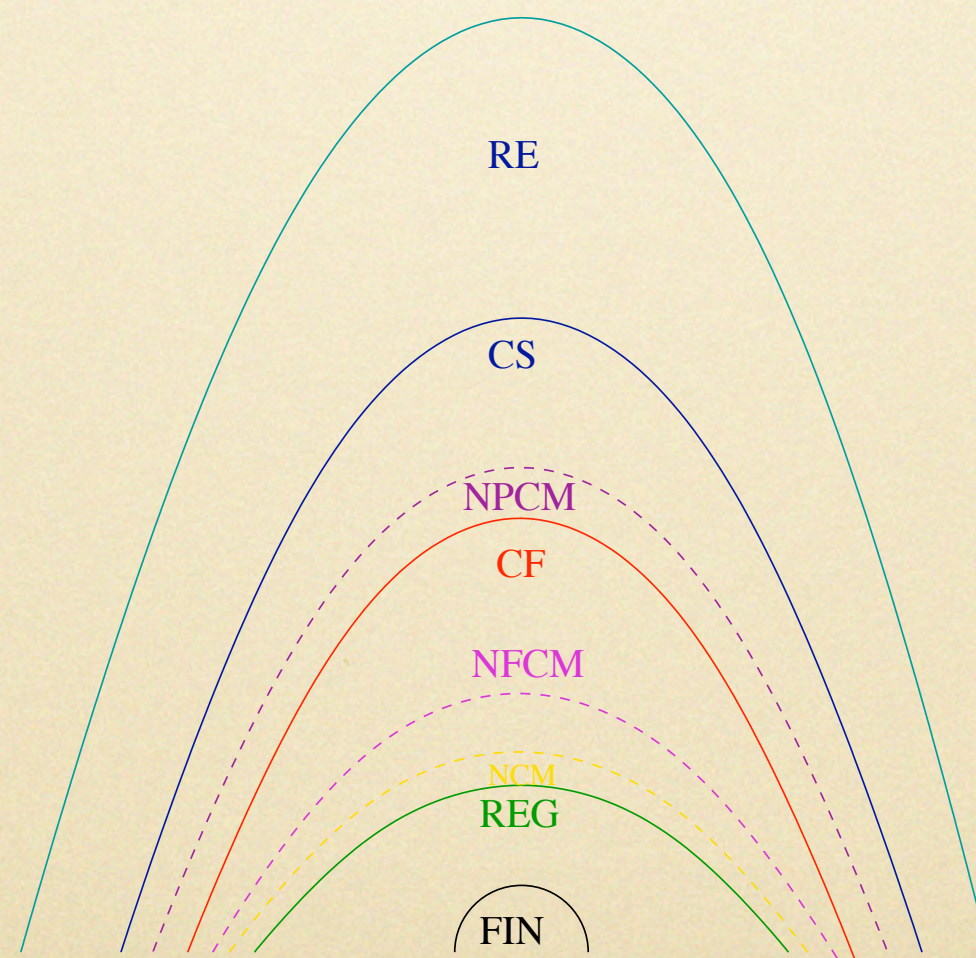
$\mathcal{L}_1$ $\mathcal{L}_2$	FIN	REG	LIN	CF	CS	RE
FIN	✓	✓	✓	✓	✓	✓
REG	✓	✓				
LIN						
CF	✓	✓				
CS						
RE	✓	✓	✓	✓	✓	✓

Table 1: Closure properties of $\mathcal{L}_1$ under $\cap_{\mathcal{L}_2}$.

Can ciliates be biocomputers?

- We proved that the TGR operation is actually *very weak* computationally.

- Starting with regular languages for both the 'base' and 'template' sets, you can *only* generate more regular languages. You get nowhere!



- (We showed that it *can* descramble MIC genes though).

Why?

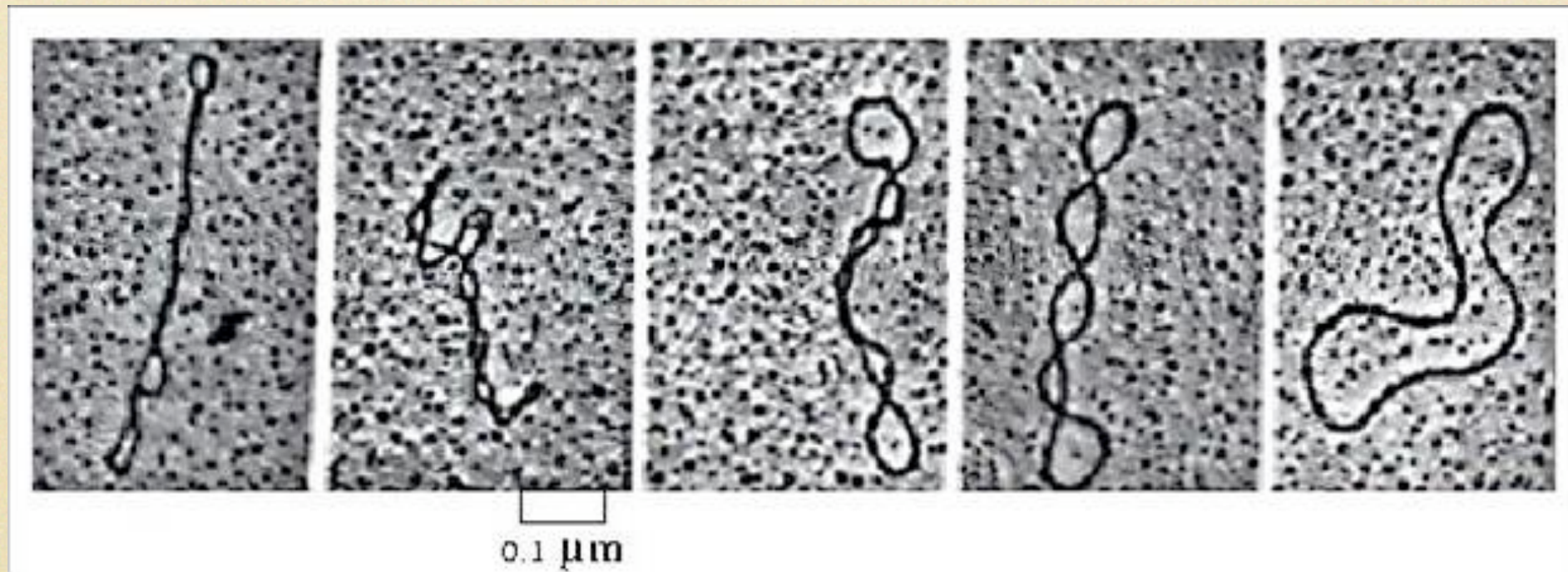
Show me the
thermodynamics!



Knot a problem



Supercoiled DNA



Rational Tangles

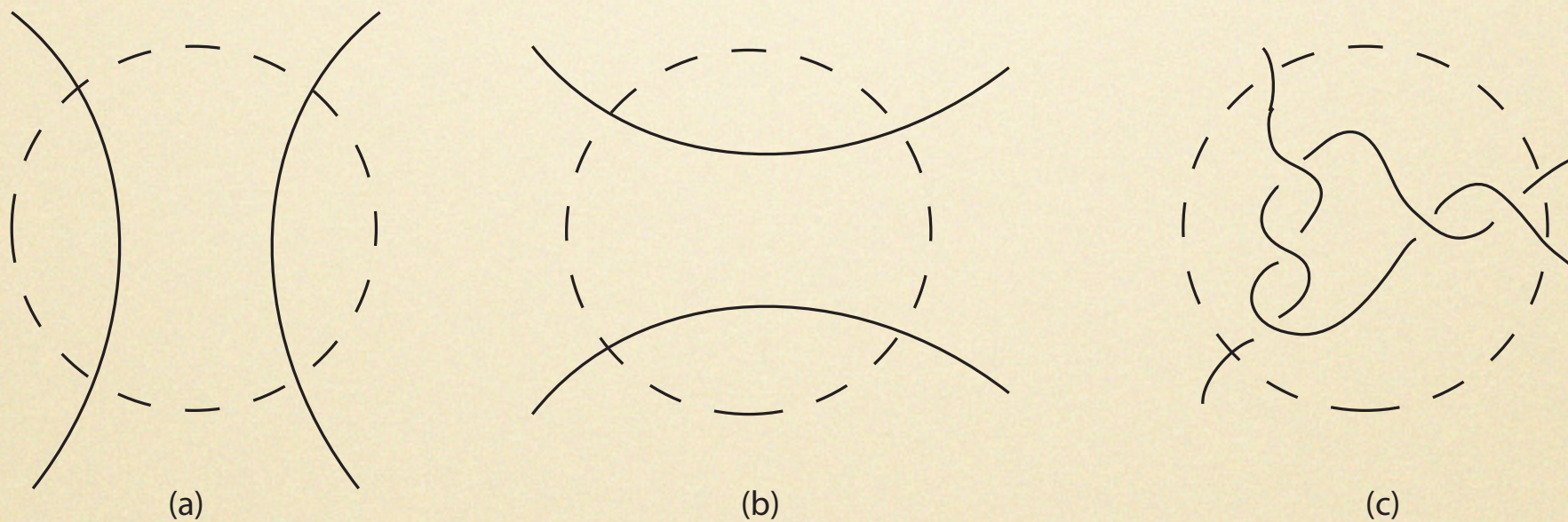
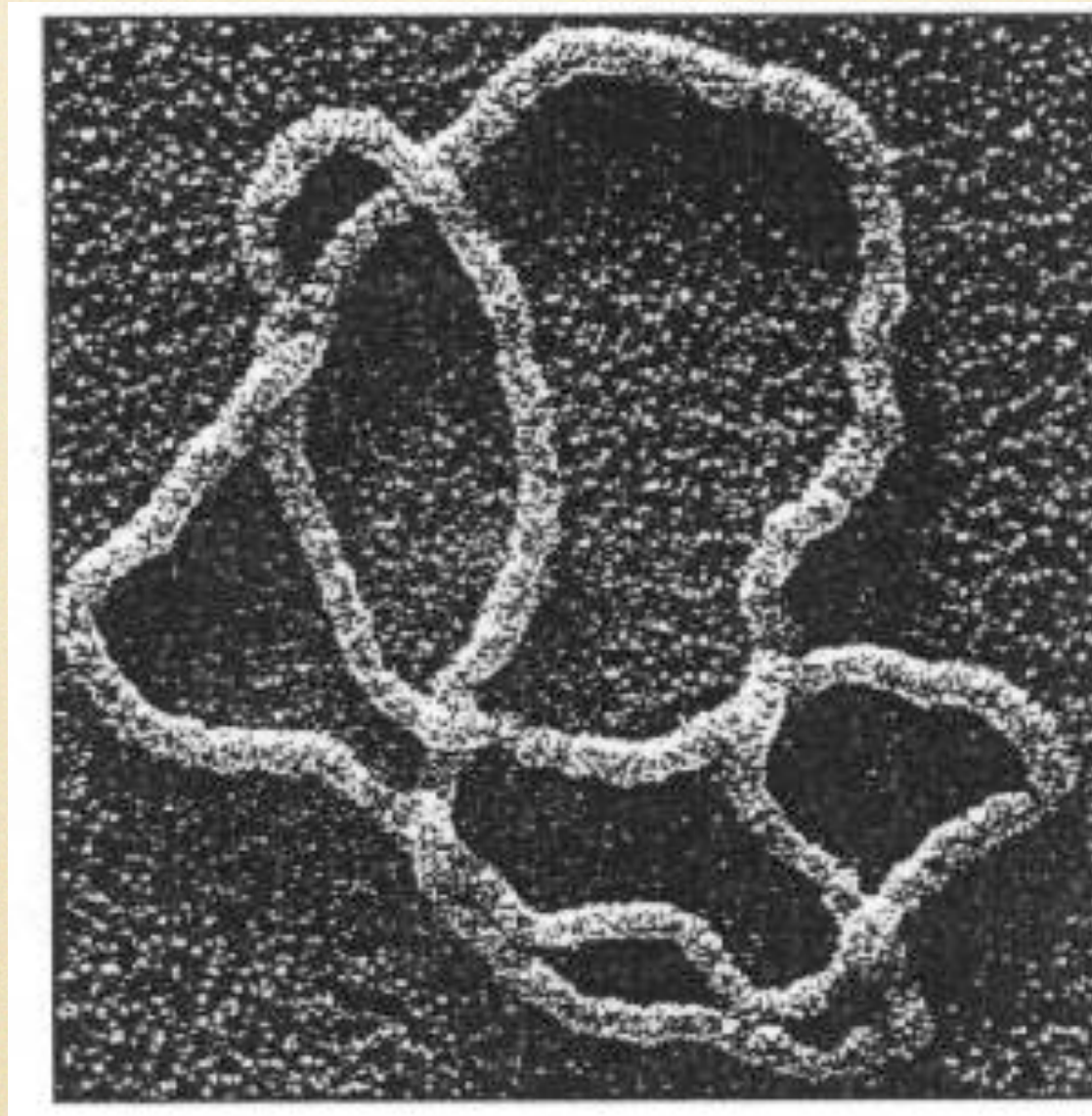


Figure 1: Diagrams of the tangles (a) $[\infty]$, (b) $[0]$ and (c) $\frac{1}{[3]} + [2]$

Recombination

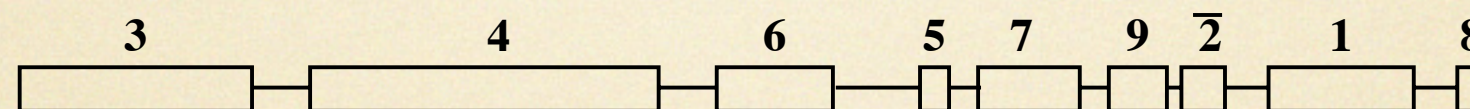


Sumners, D. 1995. Lifting the curtain: Using topology to probe the hidden action of enzymes. Notices of the AMS 42:528-537.

- What if the *topology* of the molecule juxtaposed the “next” MDSs to be descrambled?

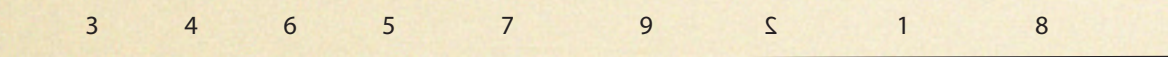
Actin I

(*O. trifallax*)

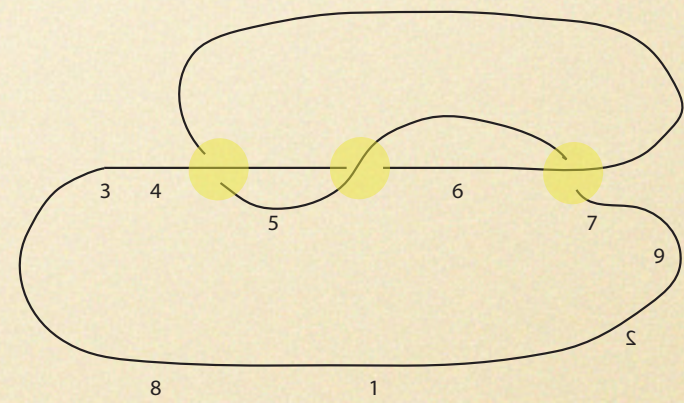
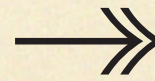


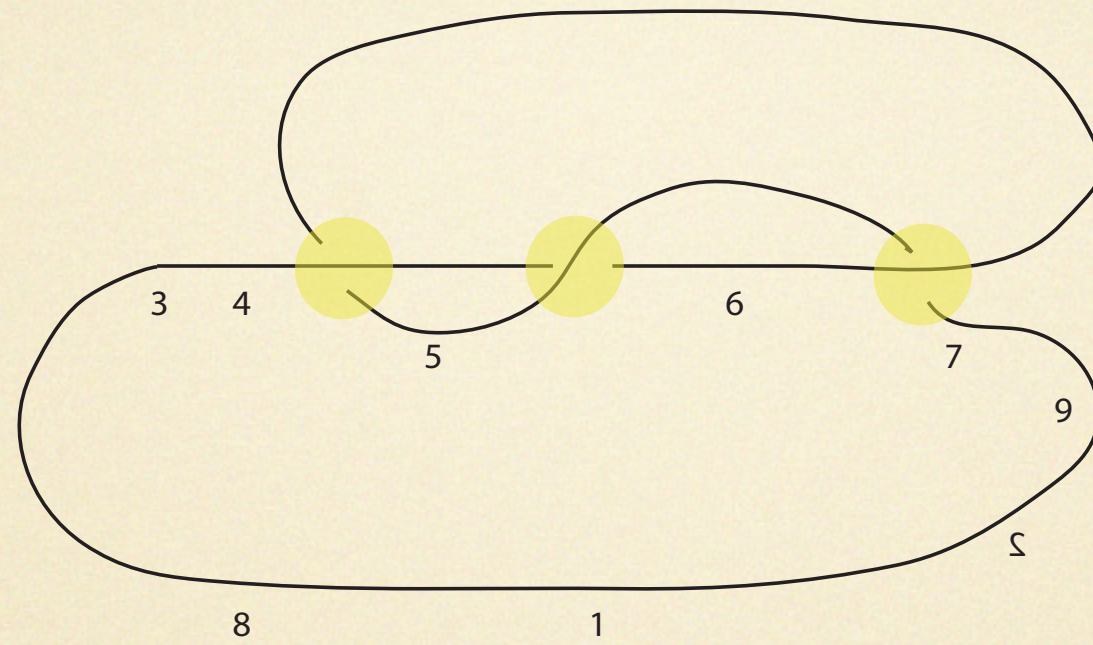
3 4 6 5 7 9 5 1 8

(a)

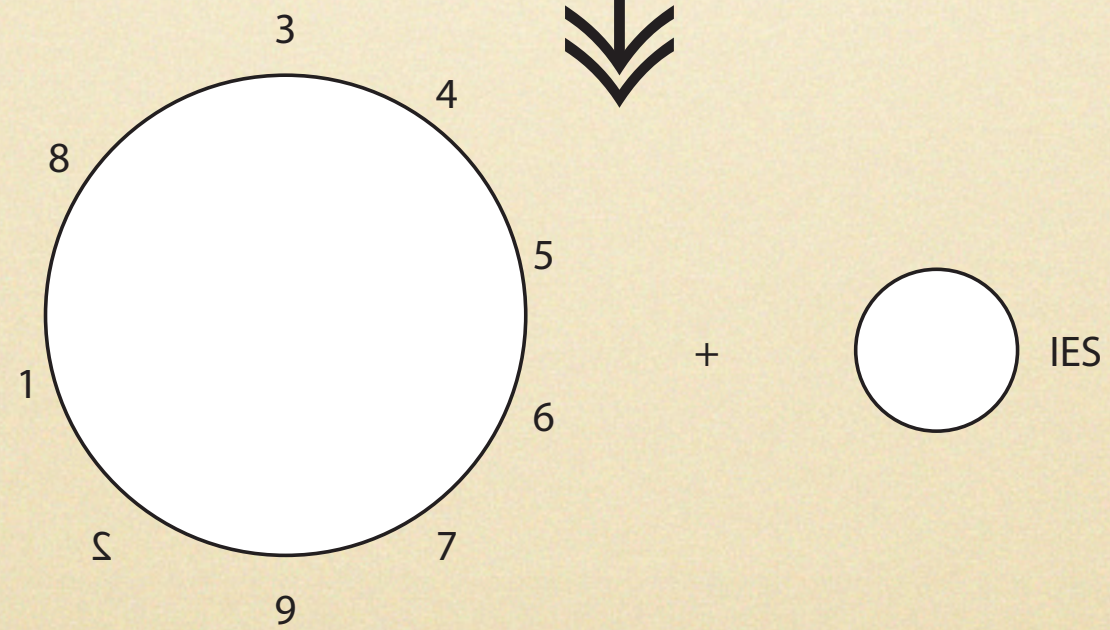


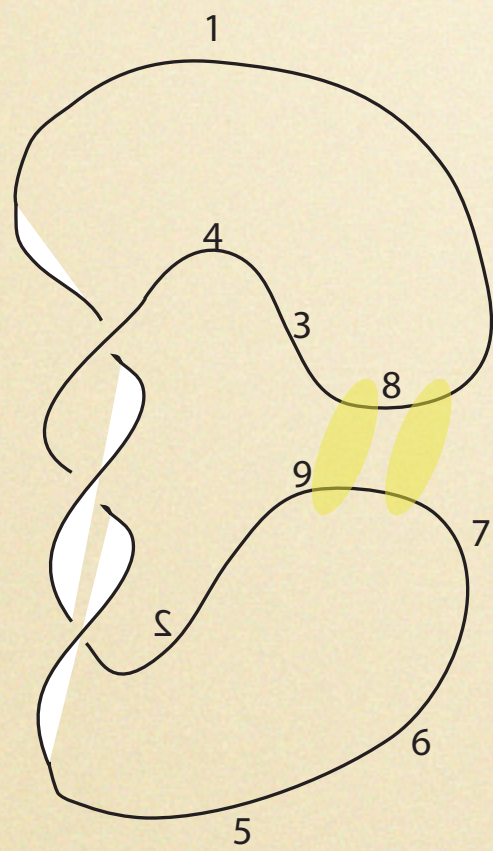
(a)



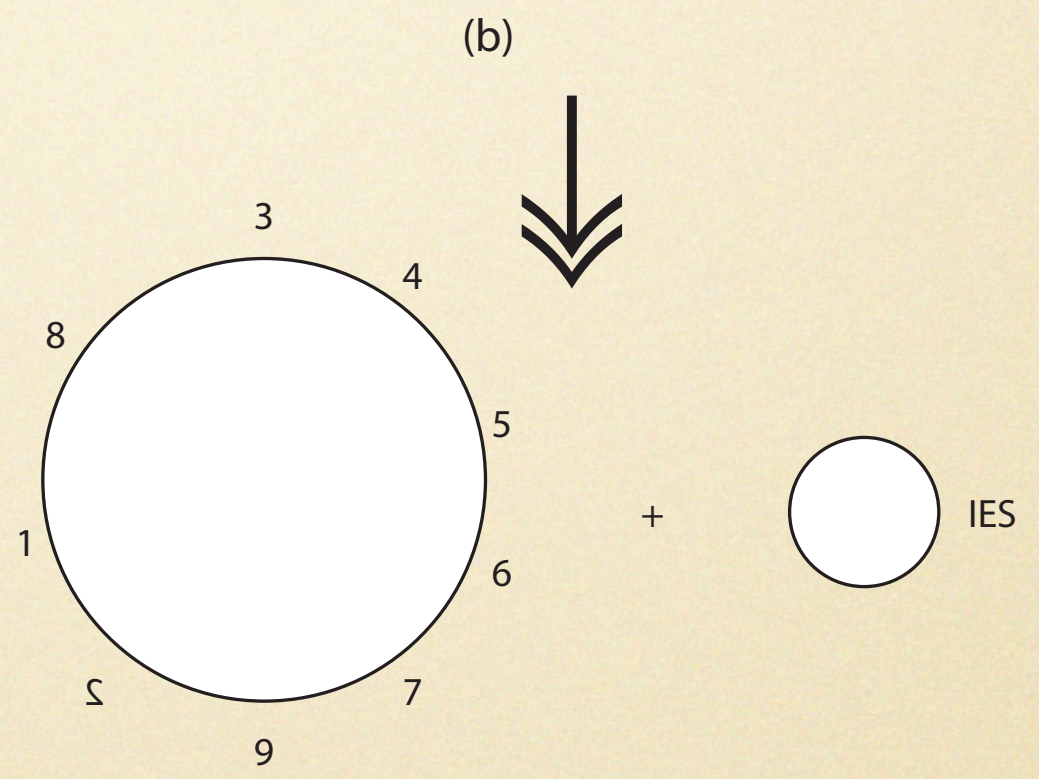


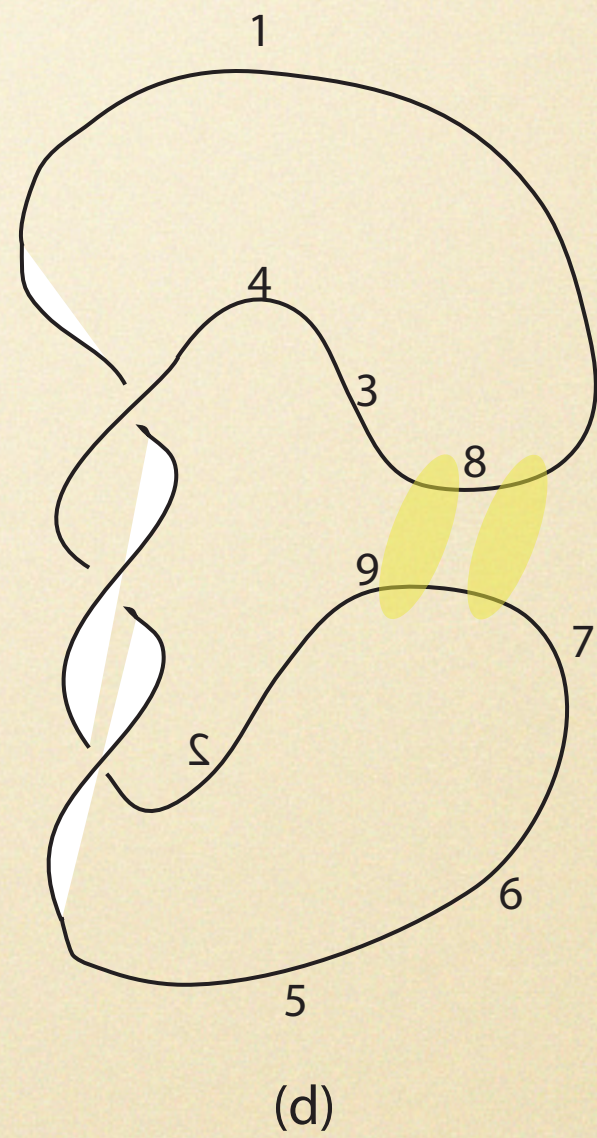
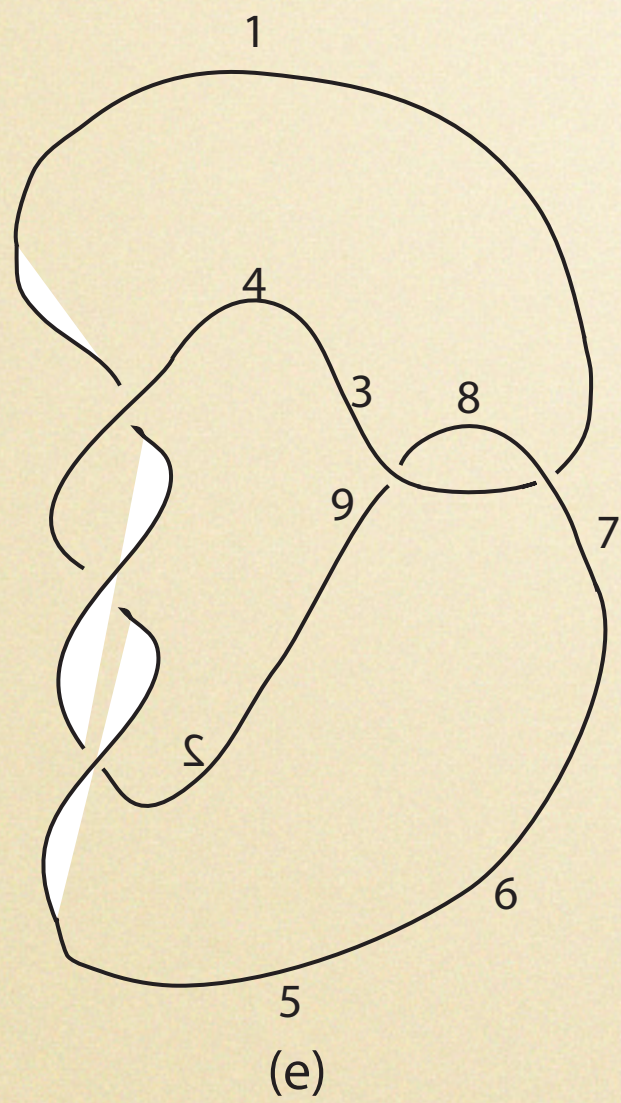
(b)

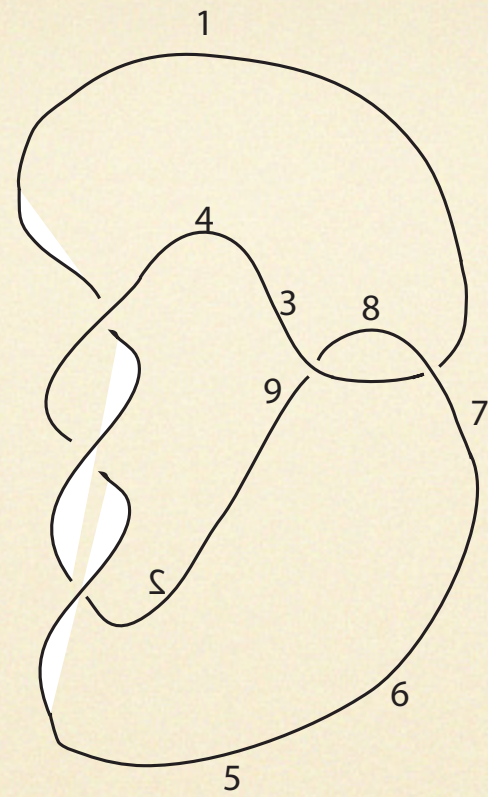




~

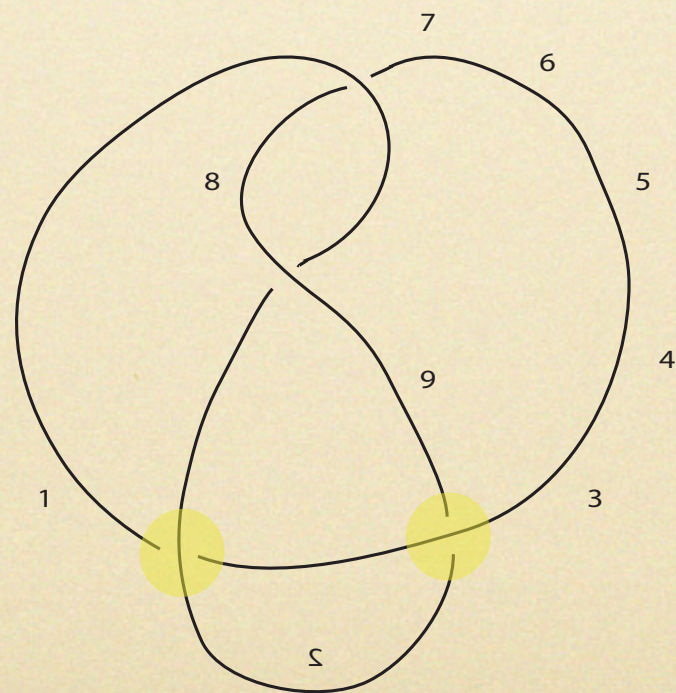




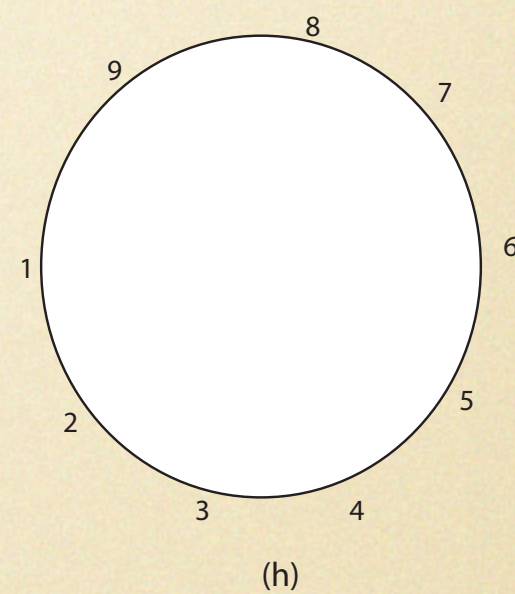
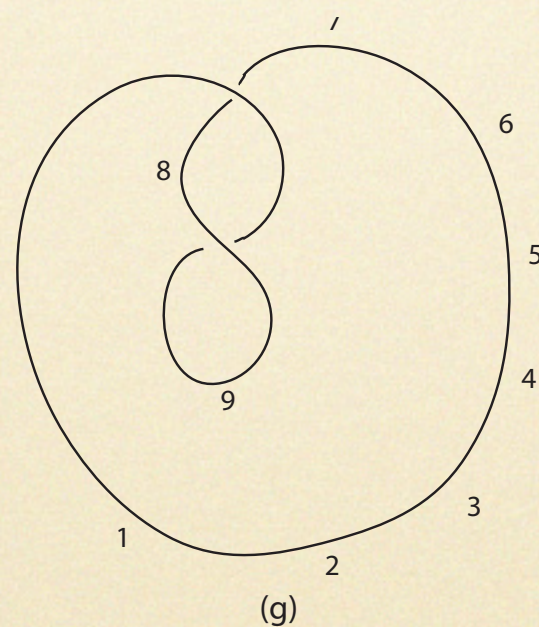
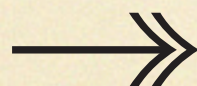
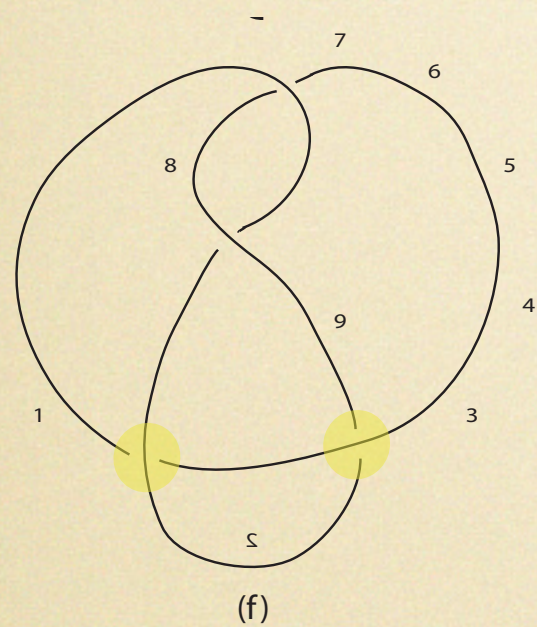


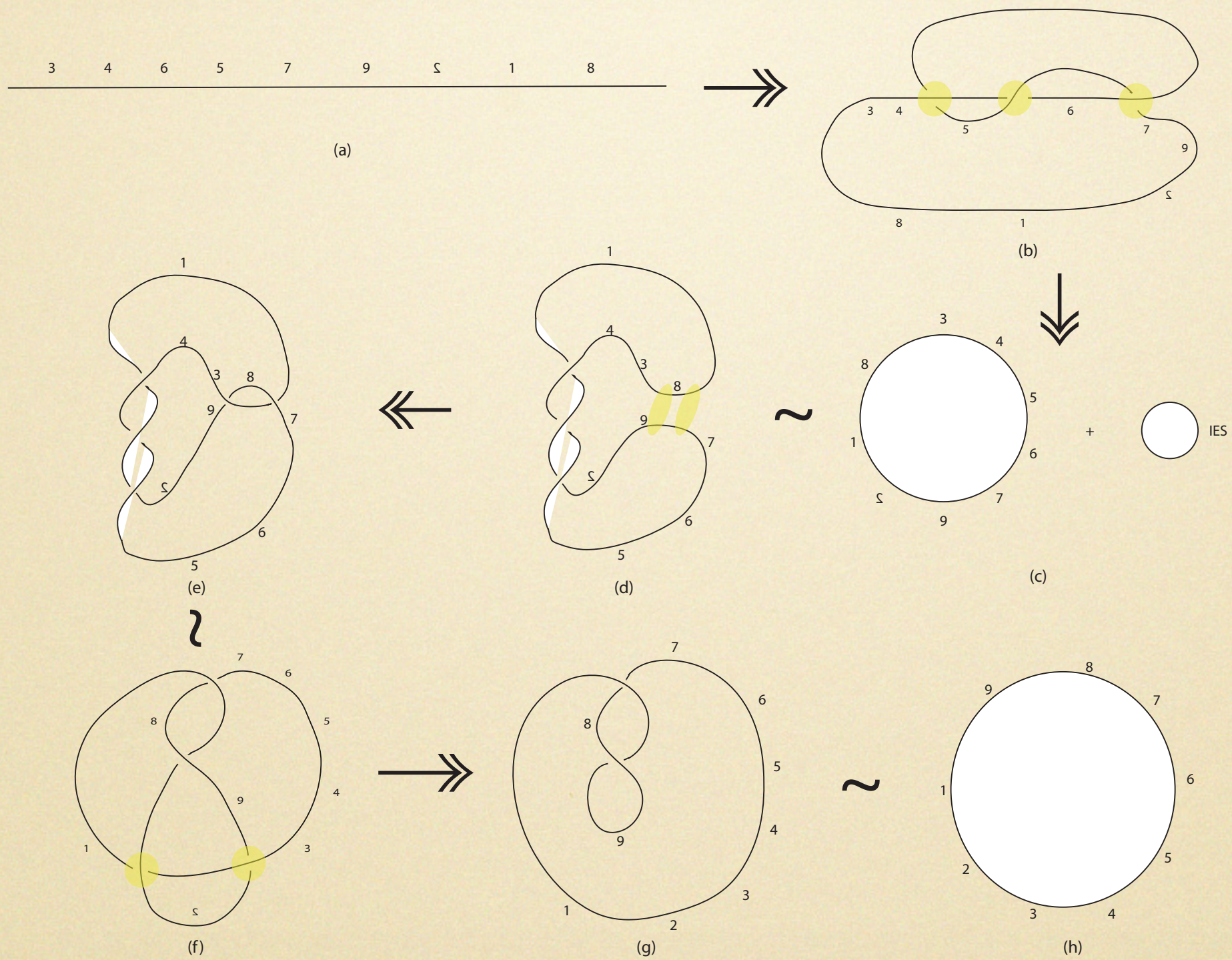
(e)

2



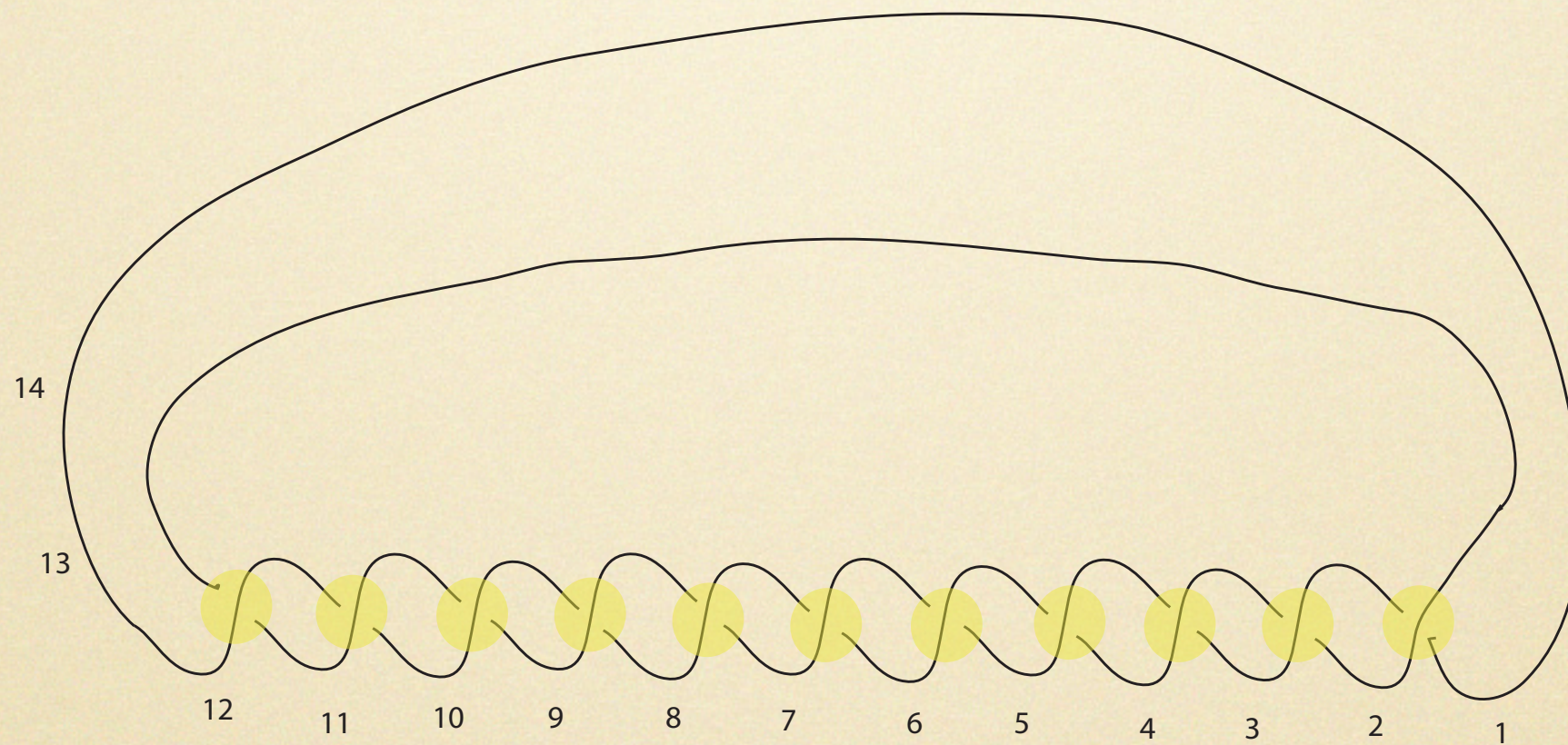
(f)



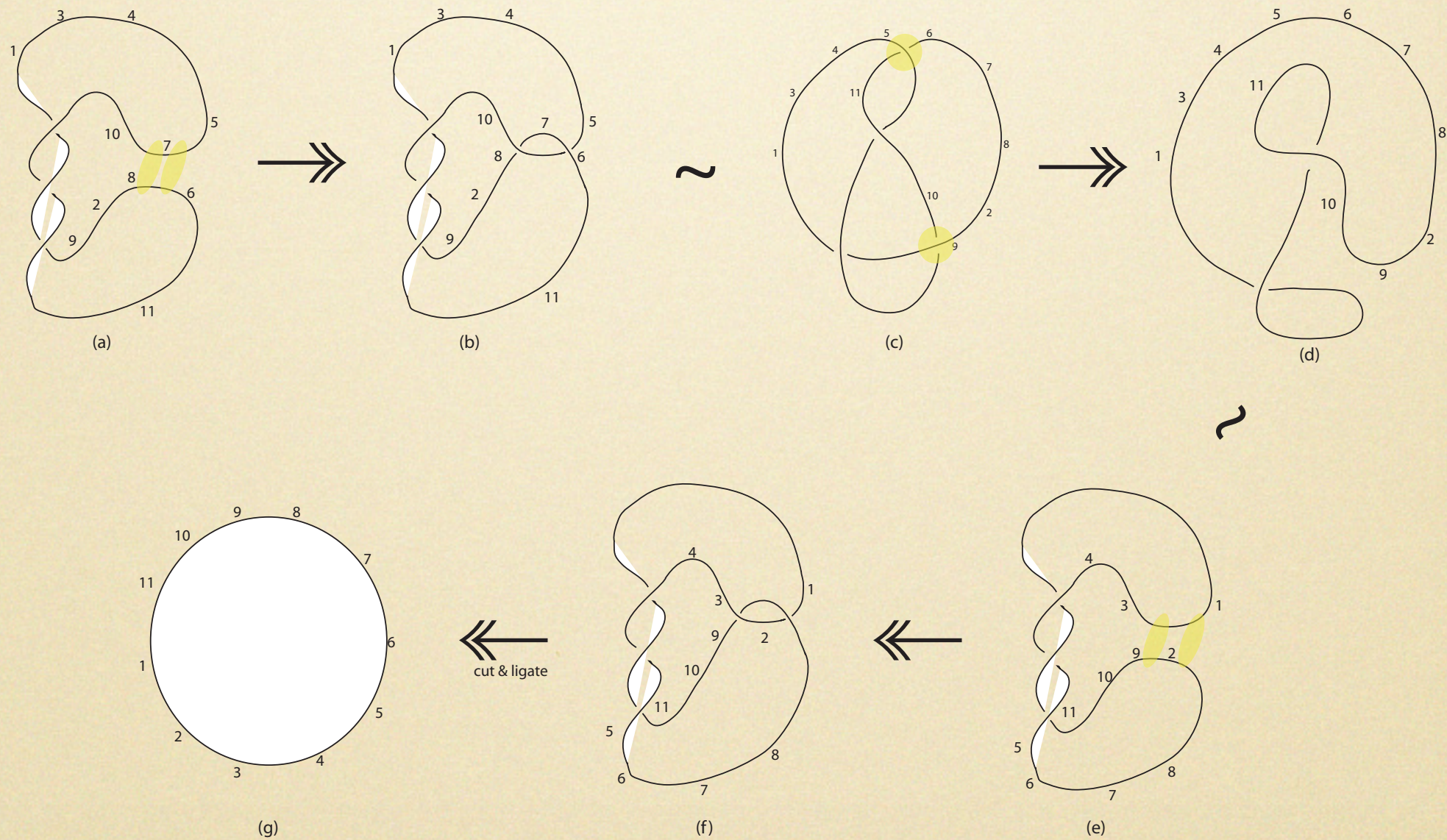


α -TBP

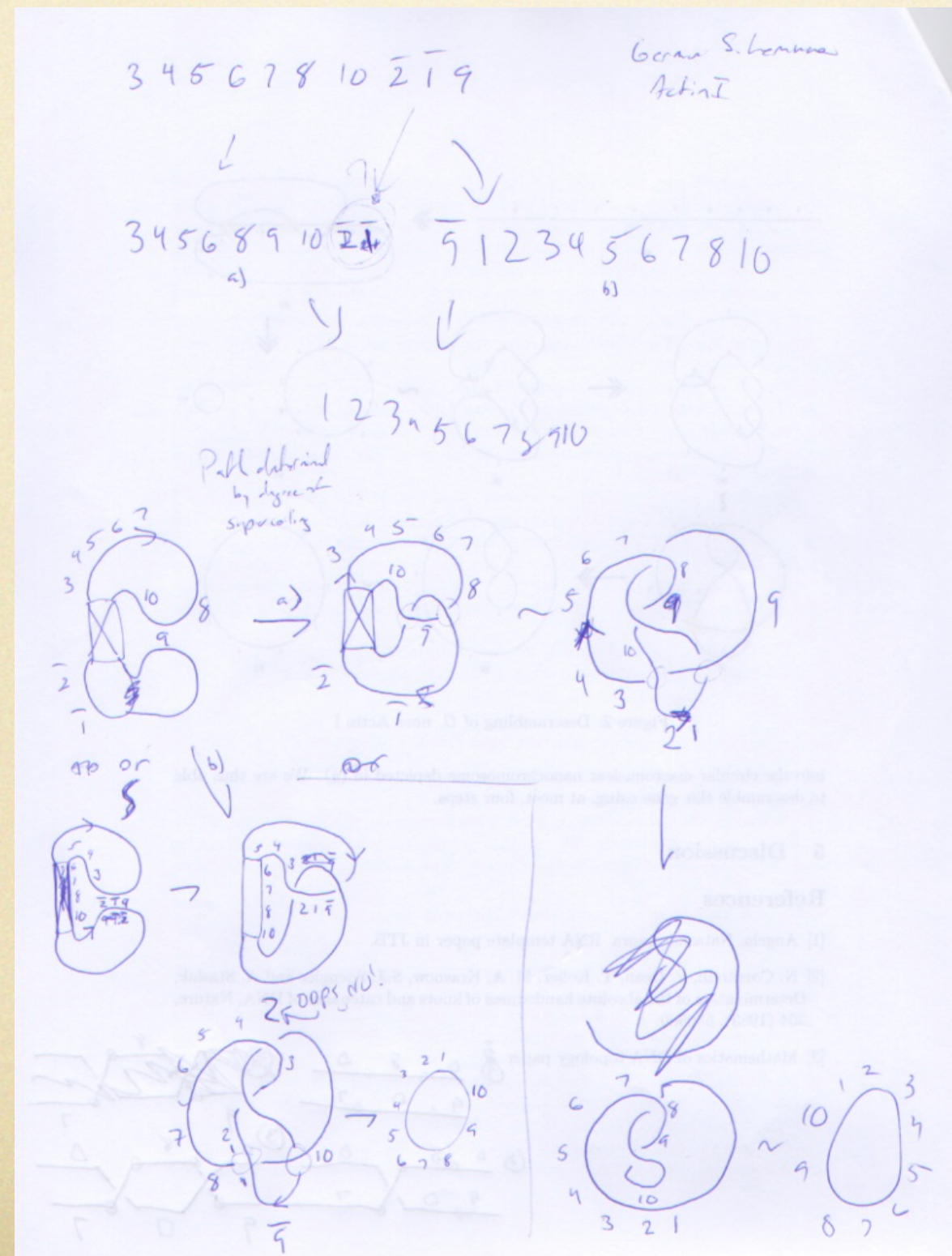
(*O. trifallax*)



Uroleptus Gene

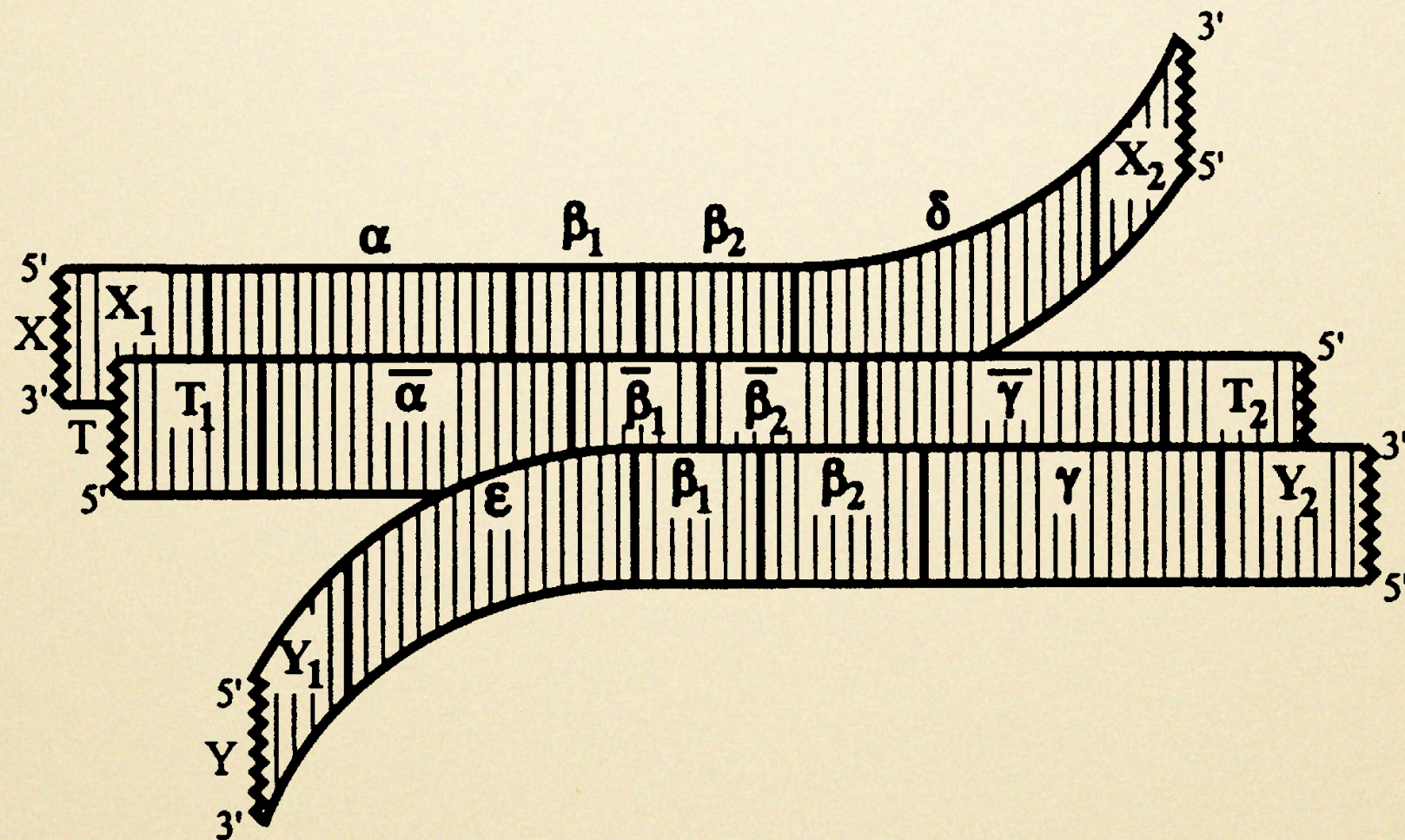


Bonus Actin I:

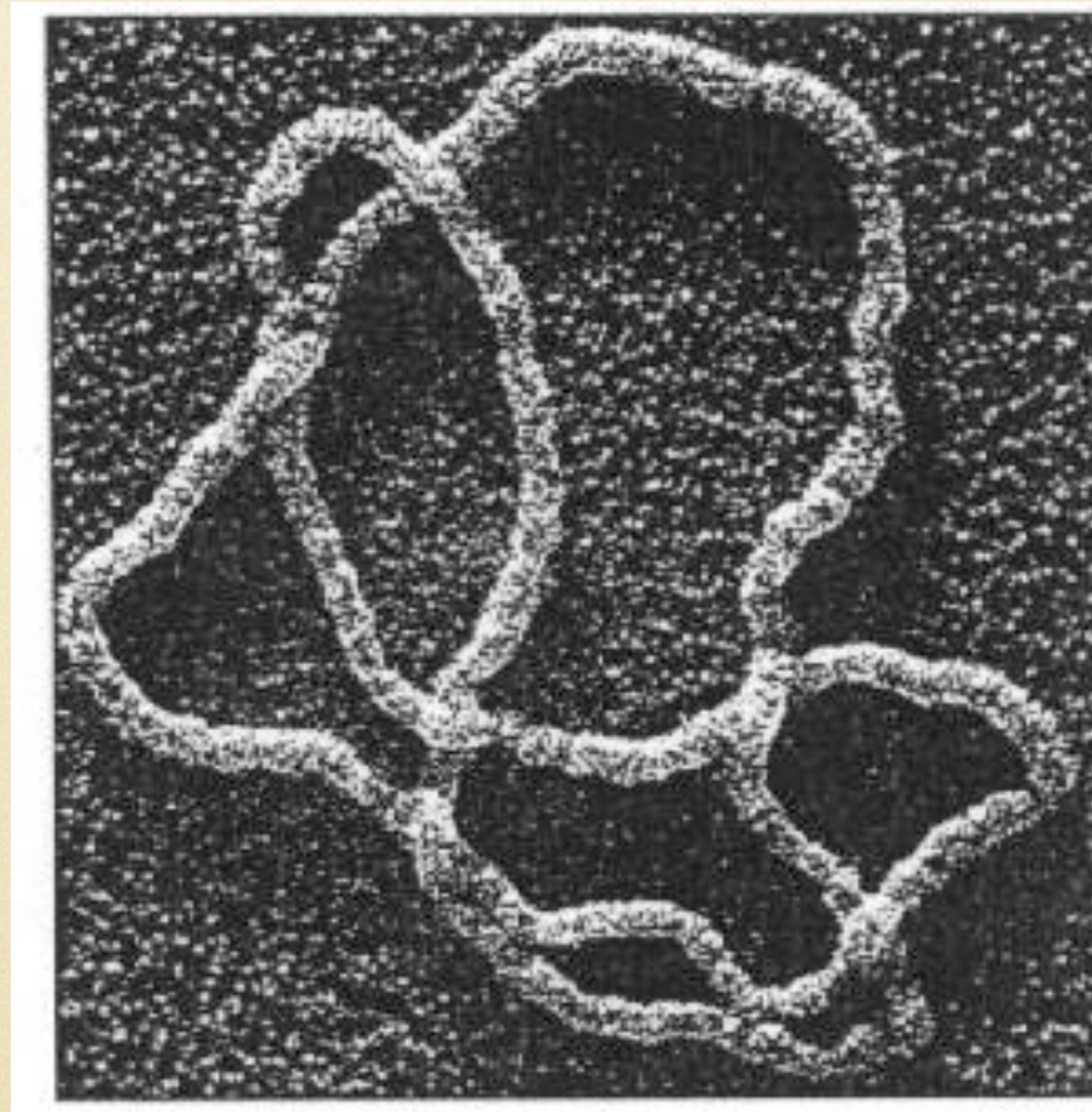


A modest proposal

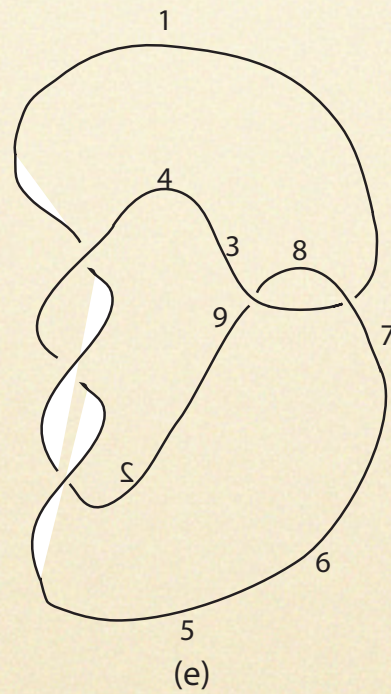
- Templates guide recombination.



- Recombination induces supercoiling.

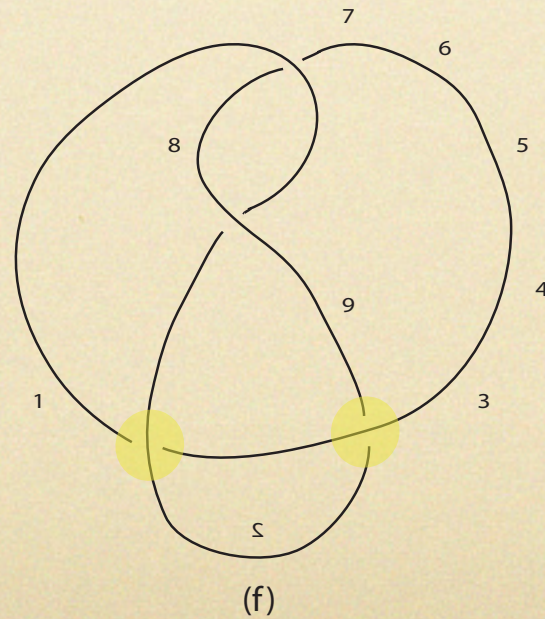


- Supercoiling induces topology.



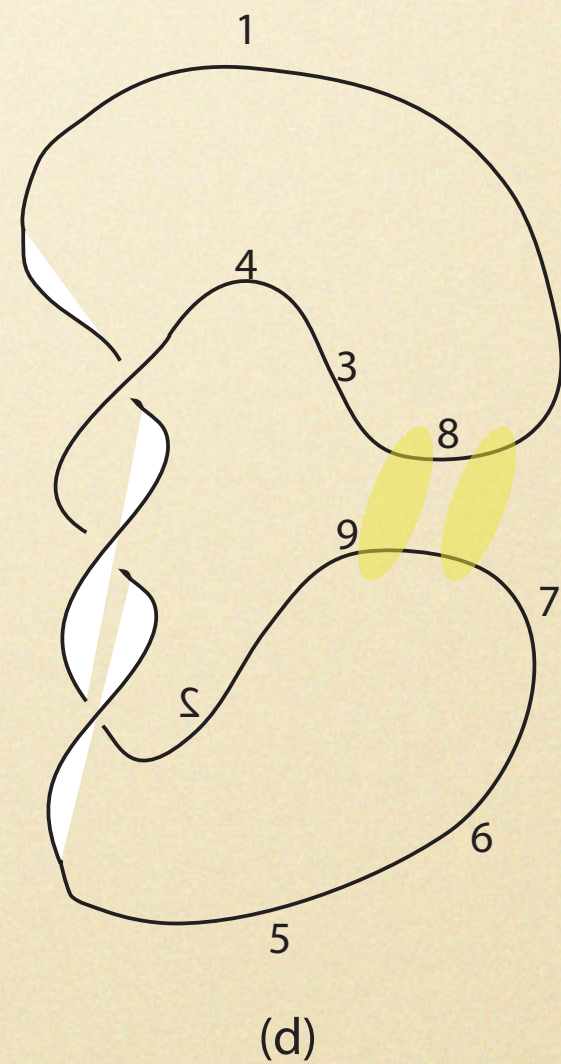
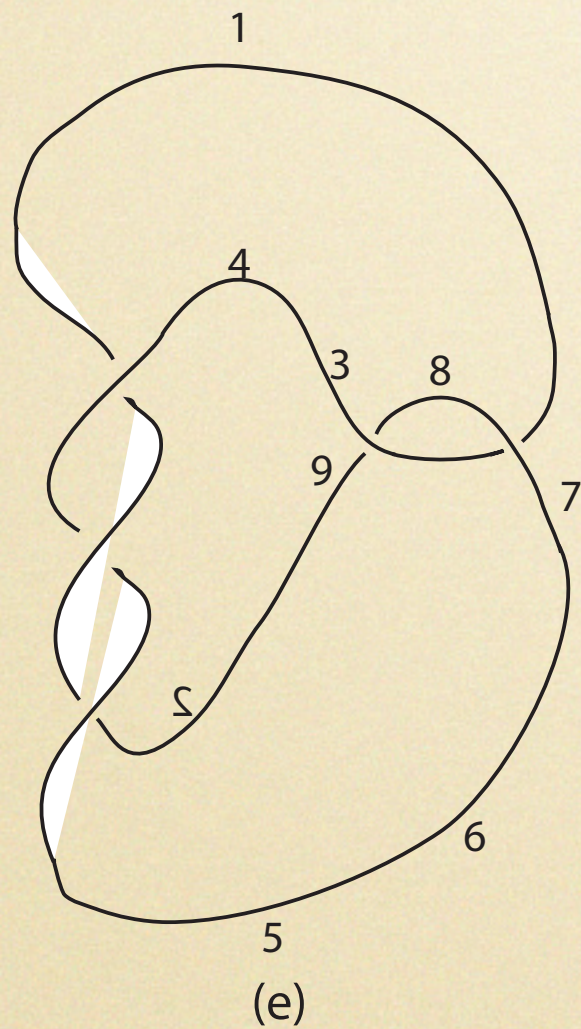
(e)

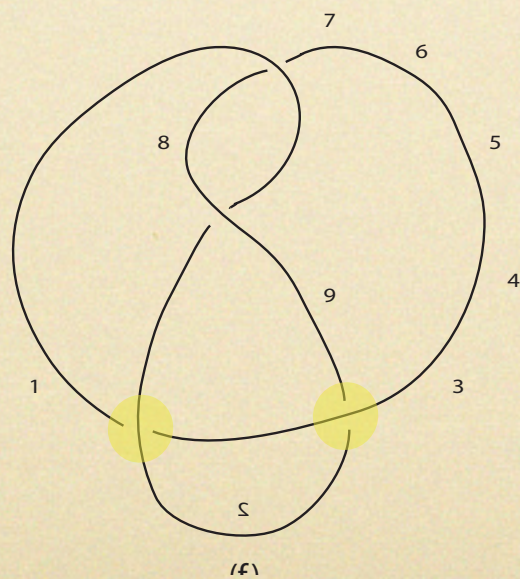
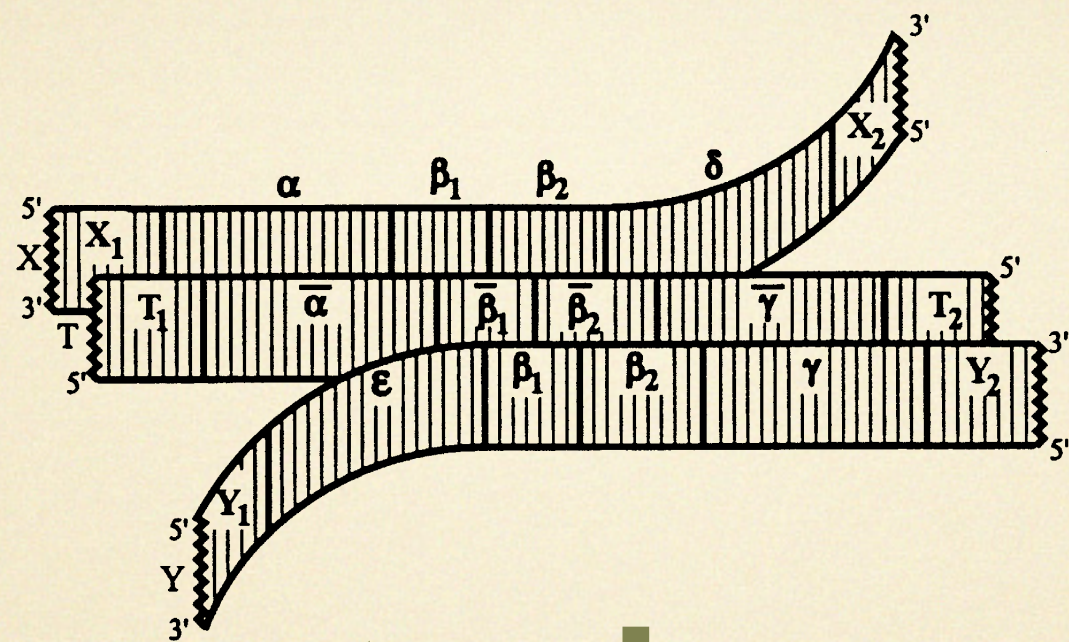
?



(f)

- Topology guides recombination.





Thank you