Generalized Algebraic Dynamic Programming: ADP over General Data Structures

Bled 2015

Christian Höner zu Siederdissen
Sonja J. Prohaska
Peter F. Stadler

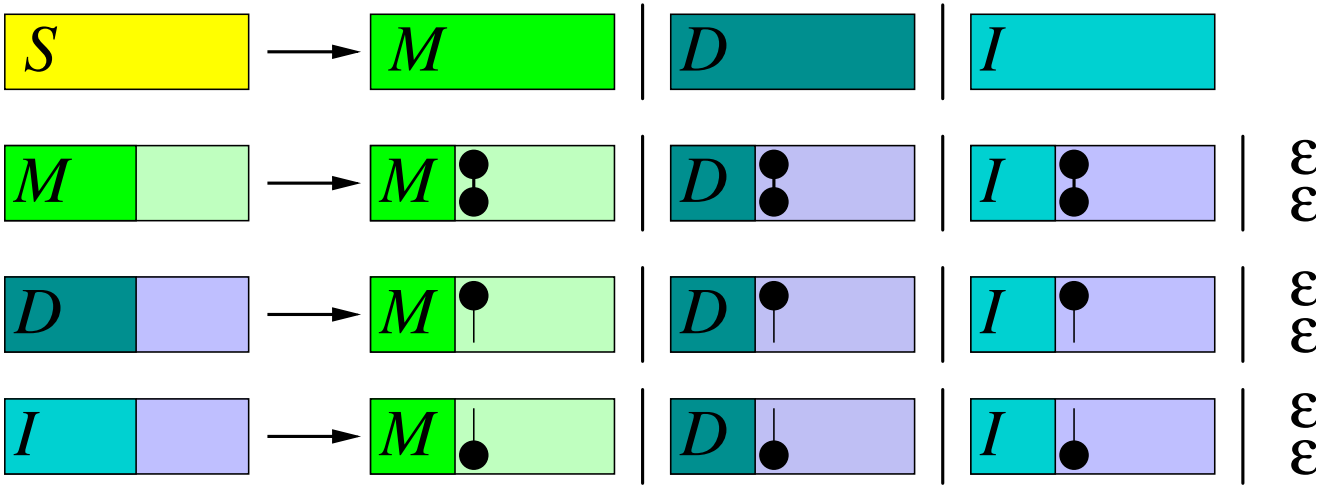
Bioinformatics Group, Dept. of Computer Science, University of Leipzig

Institute for Theoretical Chemistry, University of Vienna

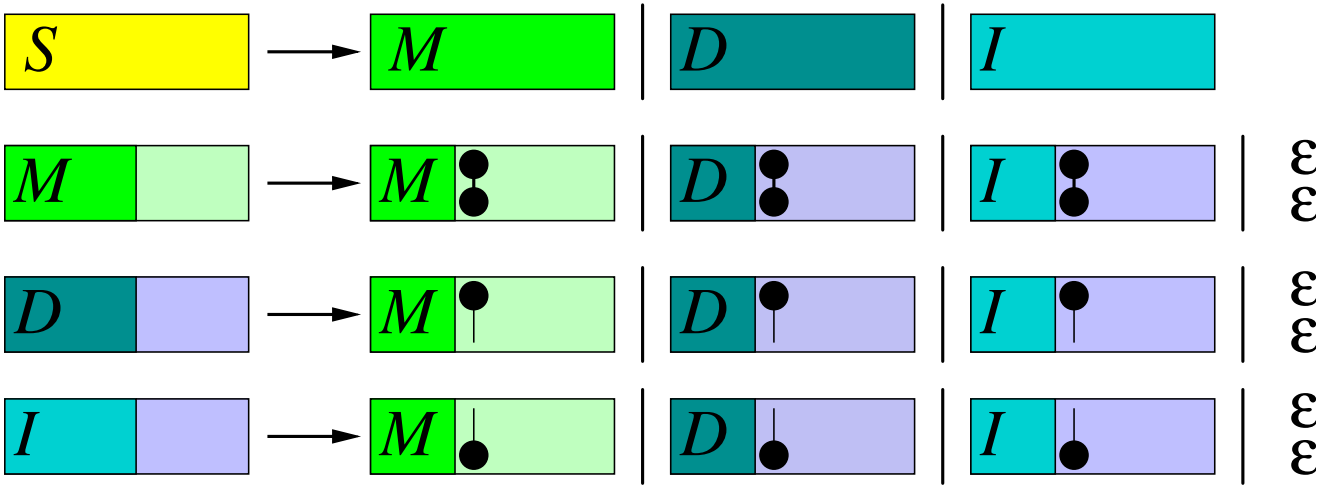
The Church of

Feb 15, 2015 – Feb 22, 2015

Alignment with Affine Gap Costs (Gotoh)



Alignment with Affine Gap Costs (Gotoh)



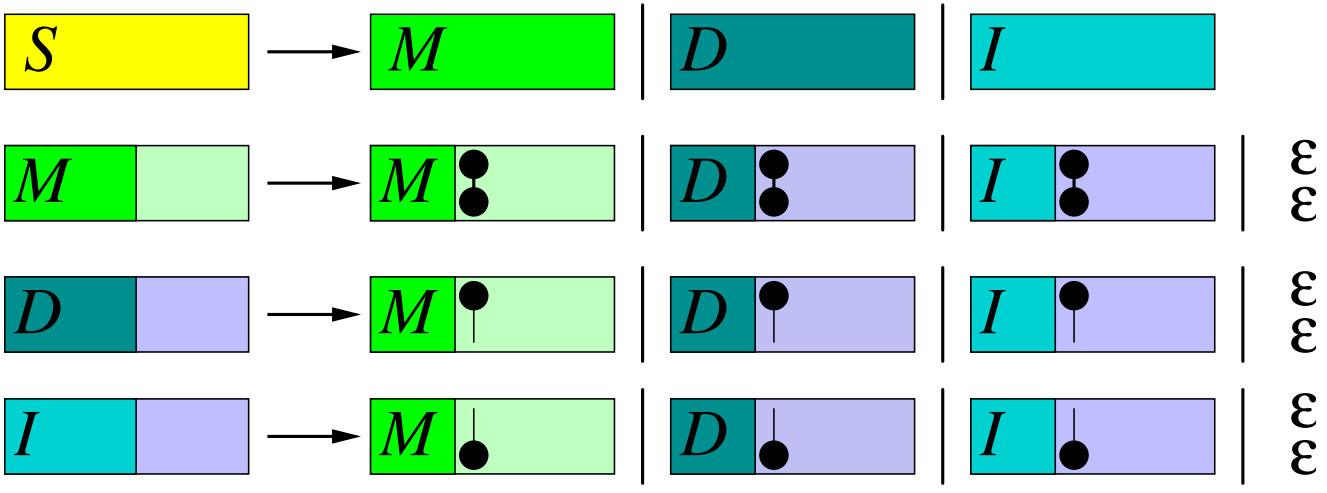
$$
\begin{array}{rcl|c|c|c|c}
S & \to & M & D & I & \\
M & \to & M\binom{u}{v} & D\binom{u}{v} & I\binom{u}{v} & \binom{\varepsilon}{\varepsilon} \\
D & \to & M\binom{u}{-} & D\binom{u}{\cdot} & I\binom{u}{-} & \binom{\varepsilon}{\varepsilon} \\
I & \to & M\binom{-}{v} & D\binom{-}{v} & I\binom{\cdot}{v} & \binom{\varepsilon}{\varepsilon}
\end{array}
$$

Alignment with Affine Gap Costs (Gotoh)

$$
\begin{array}{lcl|l|l|l}
S_{m,n} & \to & M_{m,n} & D_{m,n} & I_{m,n} & \\
M_{i,j} & \to & M_{i-1,j-1}\binom{u_i}{v_j} & D_{i-1,j-1}\binom{u_i}{v_j} & I_{i-1,j-1}\binom{u_i}{v_j} & \binom{\varepsilon_0}{\varepsilon_0} \\
D_{i,j} & \to & M_{i-1,j}\binom{u_i}{-} & D_{i-1,j}\binom{u_i}{\cdot} & I_{i-1,j}\binom{u_i}{-} & \binom{\varepsilon_0}{\varepsilon_0} \\
I_{i,j} & \to & M_{i,j-1}\binom{-}{v_j} & D_{i,j-1}\binom{-}{v_j} & I_{i,j-1}\binom{\cdot}{v_j} & \binom{\varepsilon_0}{\varepsilon_0}
\end{array}
$$

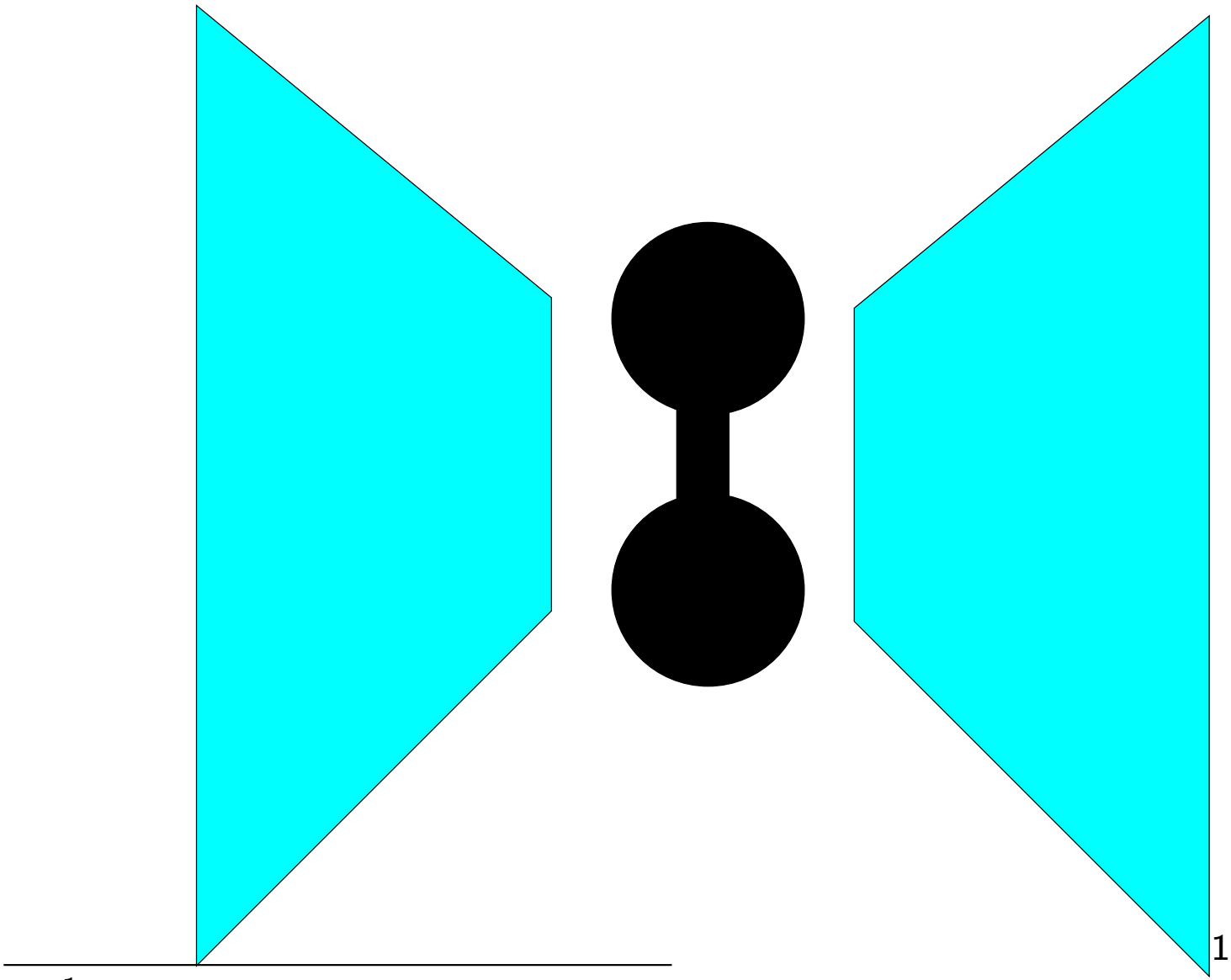
$$
\begin{array}{lcl|l|l|l}
S & \to & M & D & I & \\
M & \to & M\binom{u}{v} & D\binom{u}{v} & I\binom{u}{v} & \binom{\varepsilon}{\varepsilon} \\
D & \to & M\binom{u}{-} & D\binom{u}{\cdot} & I\binom{u}{-} & \binom{\varepsilon}{\varepsilon} \\
I & \to & M\binom{-}{v} & D\binom{-}{v} & I\binom{\cdot}{v} & \binom{\varepsilon}{\varepsilon}
\end{array}
$$

Alignment with Affine Gap Costs (Gotoh)



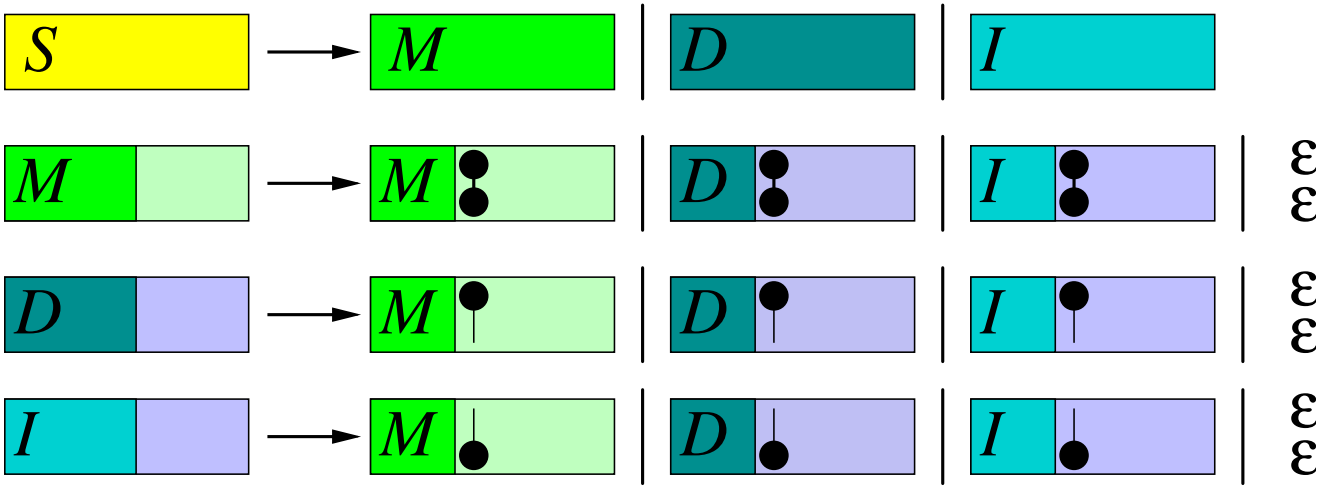
$$
\begin{array}{rcl|c|c|c|c}
S & \to & M & D & I & \\
M & \to & M\binom{u}{v} & D\binom{u}{v} & I\binom{u}{v} & \binom{\varepsilon}{\varepsilon} \\
D & \to & M\binom{u}{-} & D\binom{u}{\cdot} & I\binom{u}{-} & \binom{\varepsilon}{\varepsilon} \\
I & \to & M\binom{-}{v} & D\binom{-}{v} & I\binom{\cdot}{v} & \binom{\varepsilon}{\varepsilon}
\end{array}
$$

Inside and Outside Combined

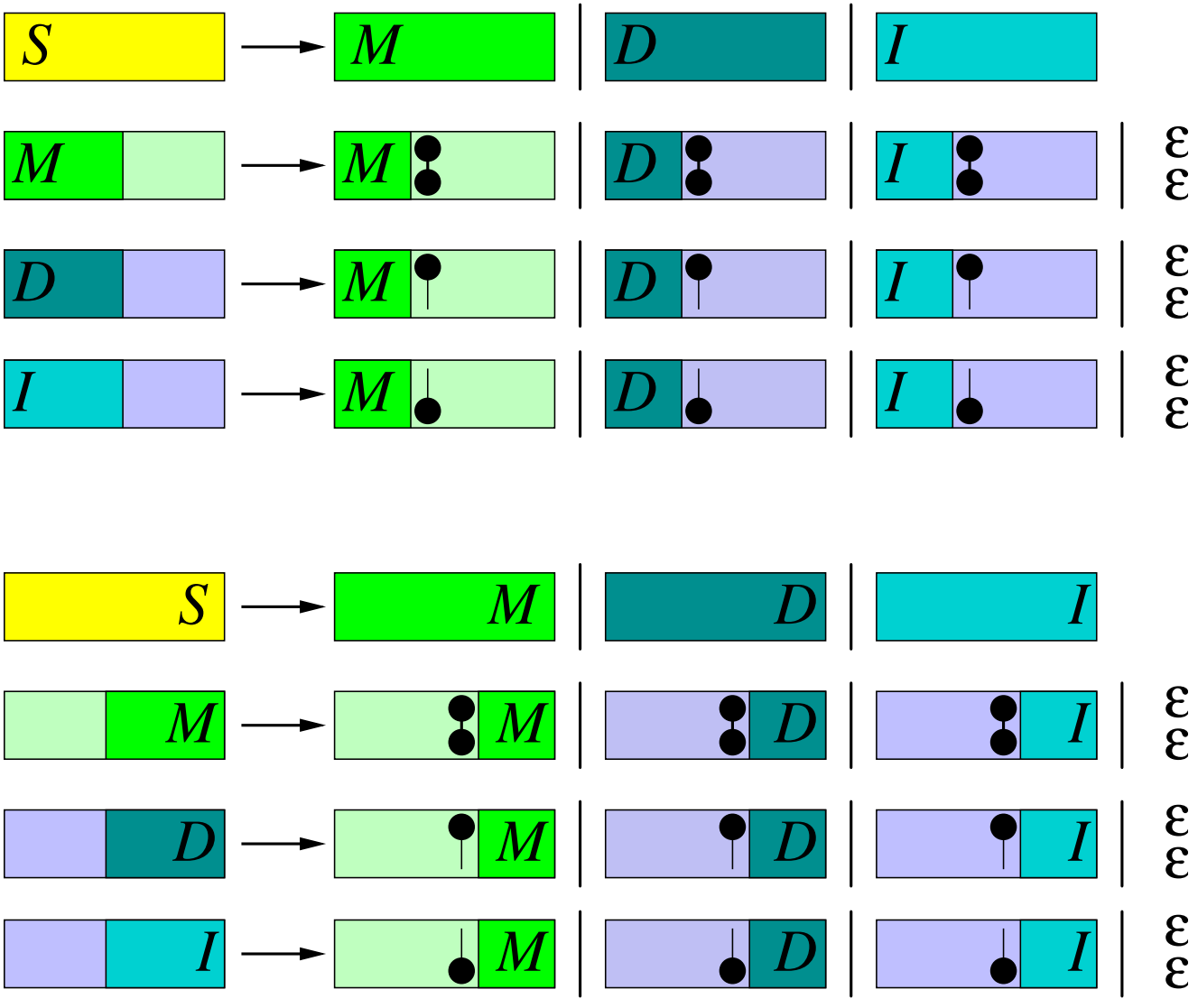


[1] abstract art

Outside Gotoh: Just Mirror the Grammar?

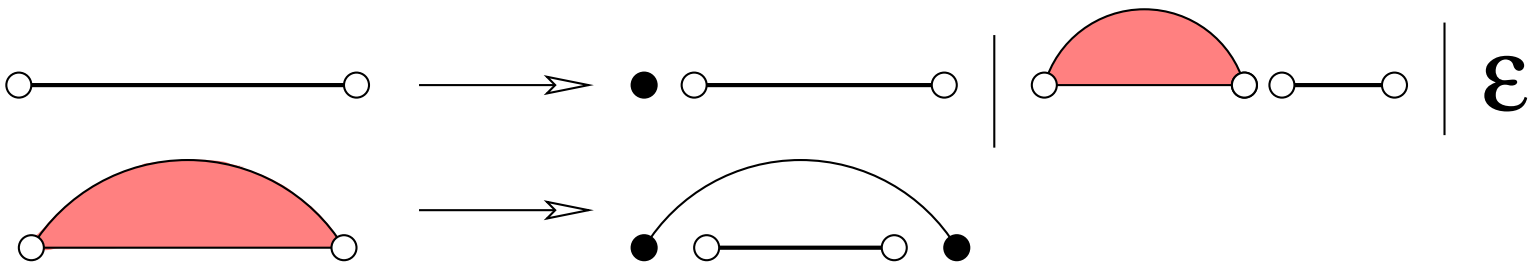


Outside Gotoh: Just Mirror the Grammar?



Unfortunately only a Special Case

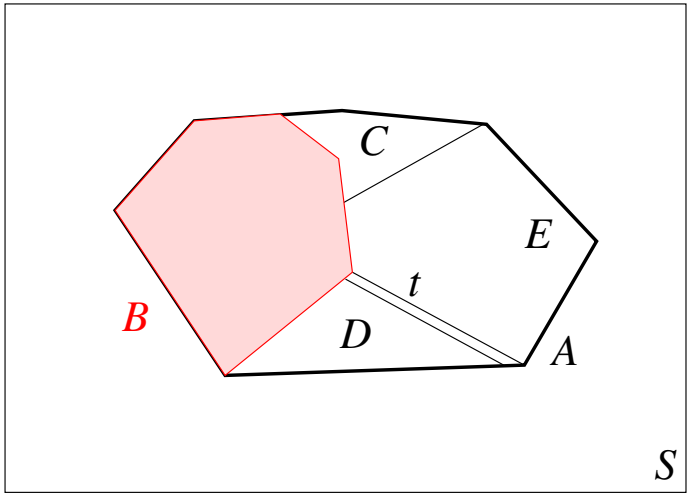
how would you mirror this RNA folding grammar?



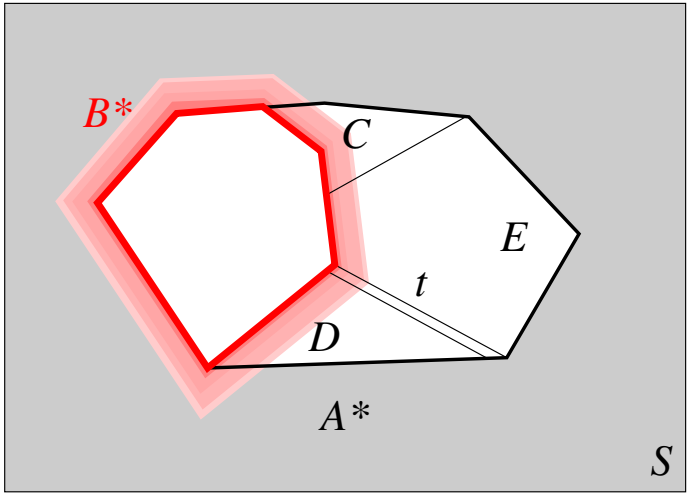
"mirroring" doesn't generalize!

Inside to Outside Productions

Inside



Outside

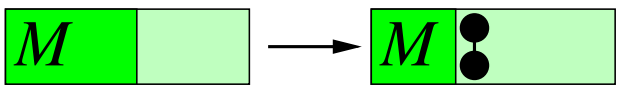


$$A \cup A^* = S$$

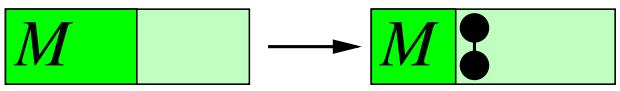
$$\begin{aligned} A &\rightarrow BCDEt \\ B^* &\rightarrow A^*CDEt \end{aligned}$$

$$\begin{aligned} C^* &\rightarrow BA^*DEt \\ &\vdots \end{aligned}$$

Outside Rules Step by Step

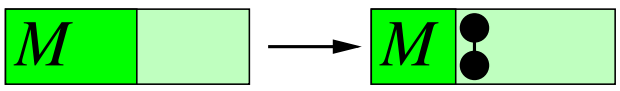


Outside Rules Step by Step

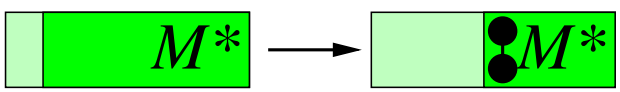


$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$

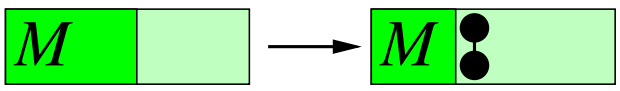
Outside Rules Step by Step



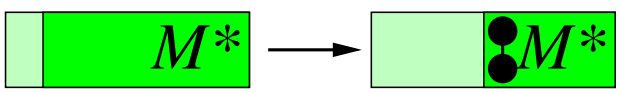
$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$



Outside Rules Step by Step

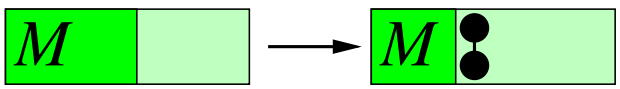


$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$

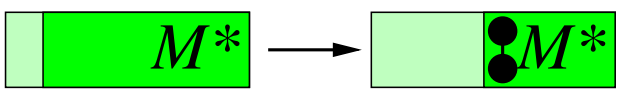


$$M^*_{i-1,j-1} \rightarrow M^*_{i,j}\binom{u_i}{v_j}$$

Outside Rules Step by Step



$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$

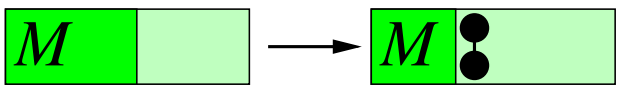


$$M^*_{i-1,j-1} \rightarrow M^*_{i,j}\binom{u_i}{v_j}$$

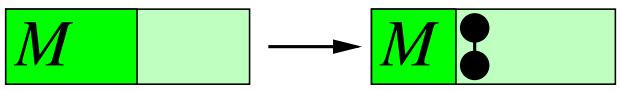
`rewrite indices:  +1/+1`

$$M^*_{i,j} \rightarrow M^*_{i+1,j+1}\binom{u_{i+1}}{v_{j+1}}$$

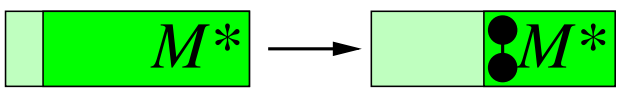
Outside Rules Step by Step



$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$



$$M \rightarrow M\binom{u}{v}$$

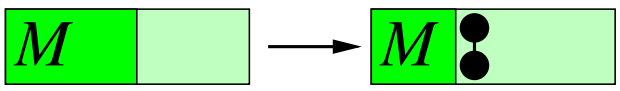


$$M^*_{i-1,j-1} \rightarrow M^*_{i,j}\binom{u_i}{v_j}$$

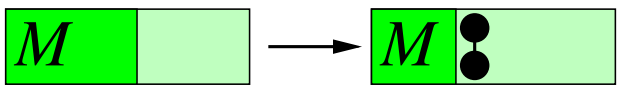
```
rewrite indices:  +1/+1
```

$$M^*_{i,j} \rightarrow M^*_{i+1,j+1}\binom{u_{i+1}}{v_{j+1}}$$

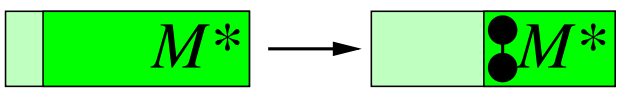
Outside Rules Step by Step



$$M_{i,j} \rightarrow M_{i-1,j-1}\binom{u_i}{v_j}$$



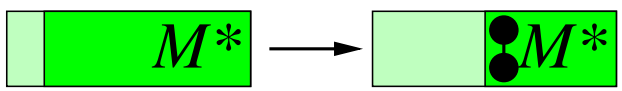
$$M \rightarrow M\binom{u}{v}$$



$$M^*_{i-1,j-1} \rightarrow M^*_{i,j}\binom{u_i}{v_j}$$

rewrite indices: +1/+1

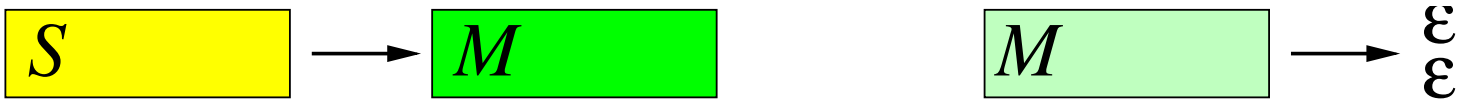
$$M^*_{i,j} \rightarrow M^*_{i+1,j+1}\binom{u_{i+1}}{v_{j+1}}$$



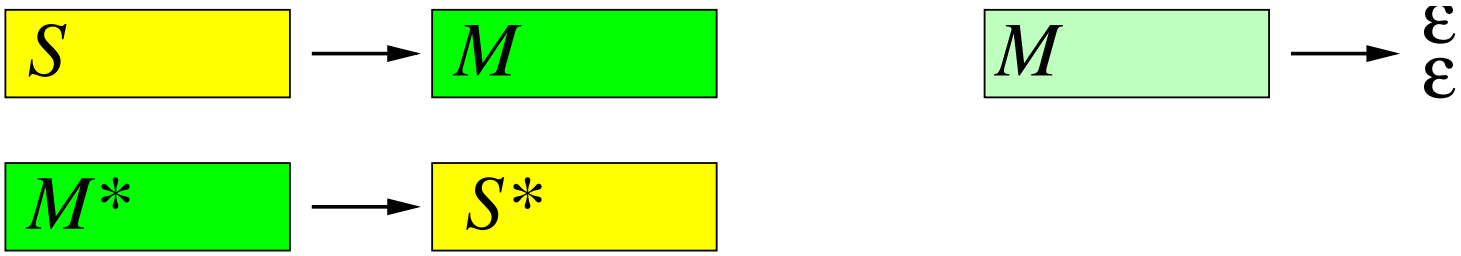
$$M^* \rightarrow M^*\binom{u}{v}$$

no rewrite necessary, let ADPfusion figure it out

Start and End Symbols

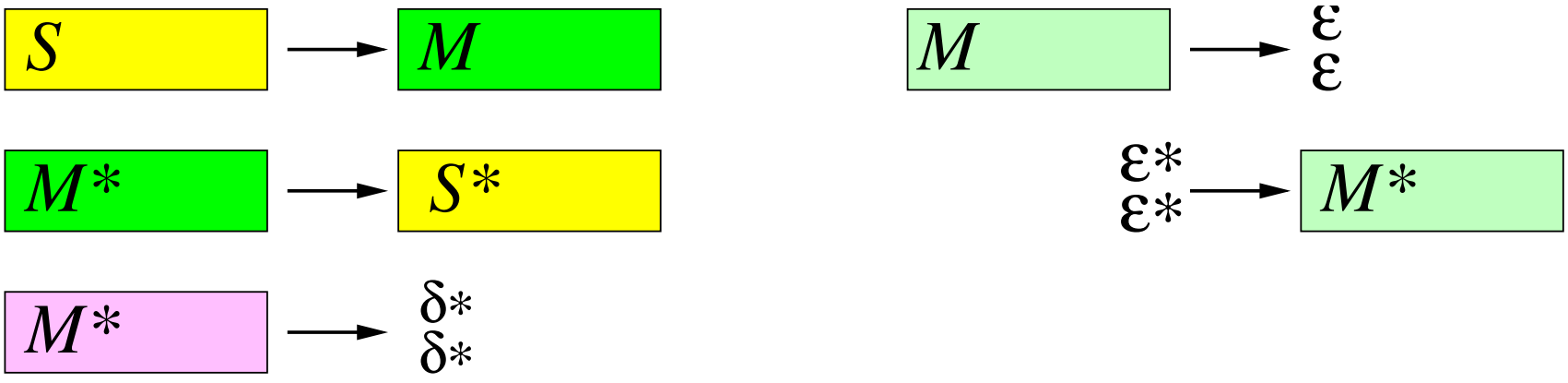


Start and End Symbols

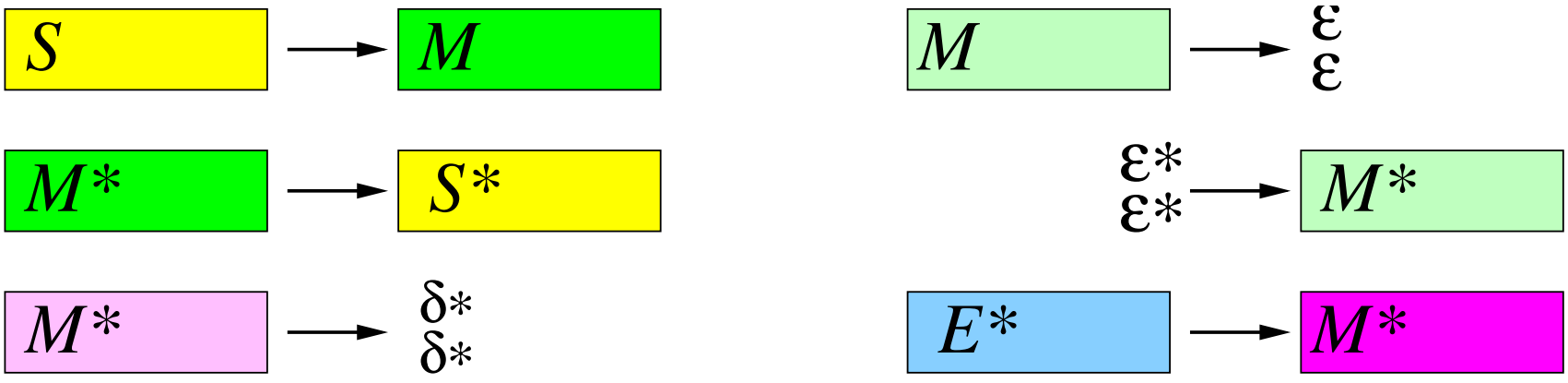


Start and End Symbols

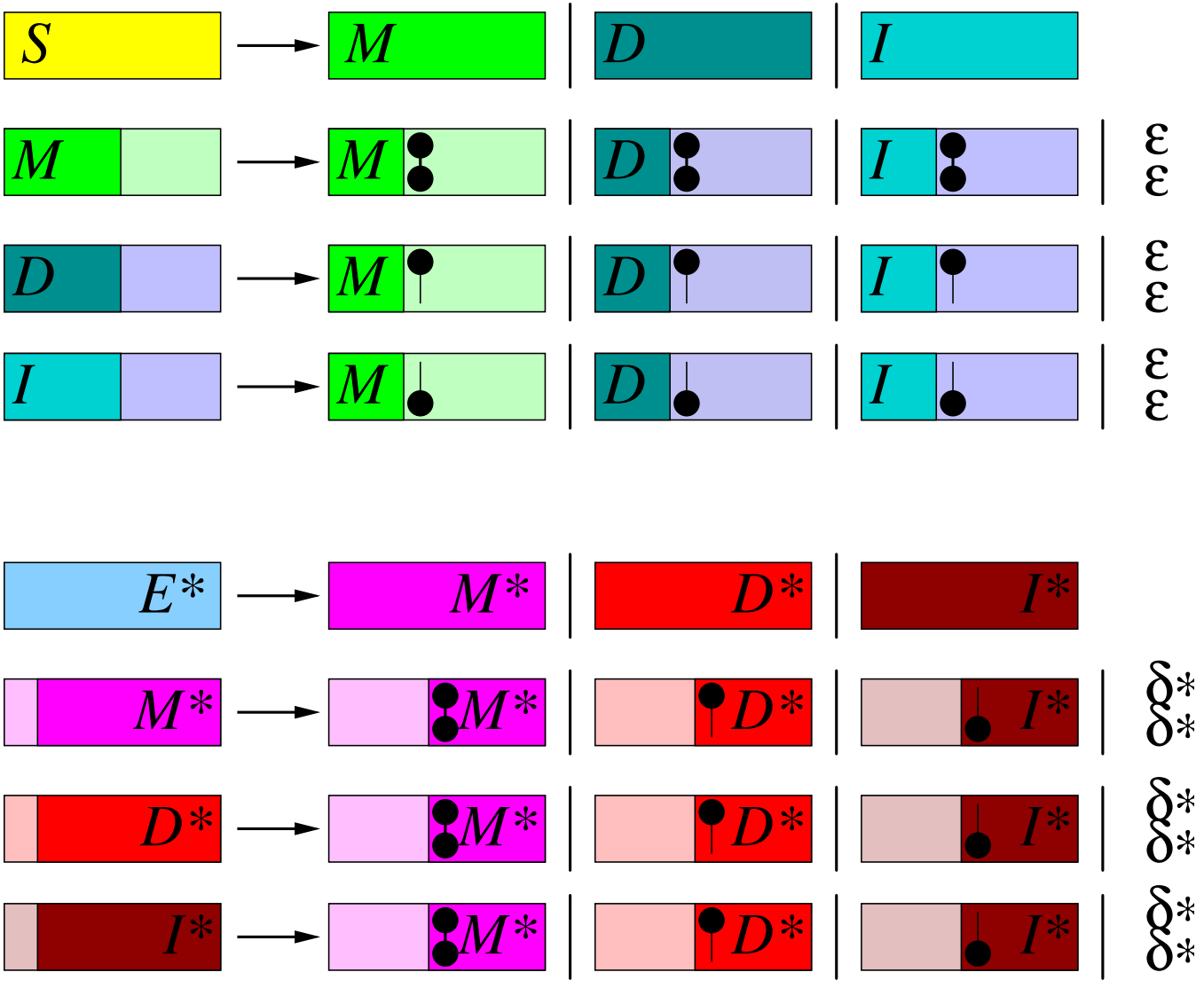
Start and End Symbols



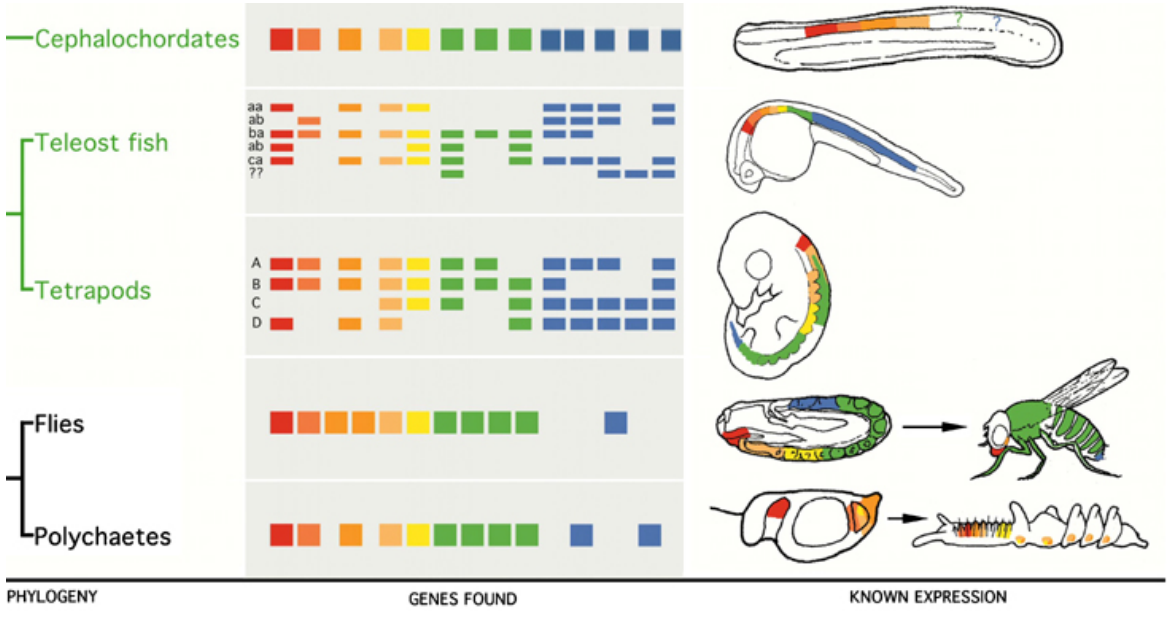
Start and End Symbols

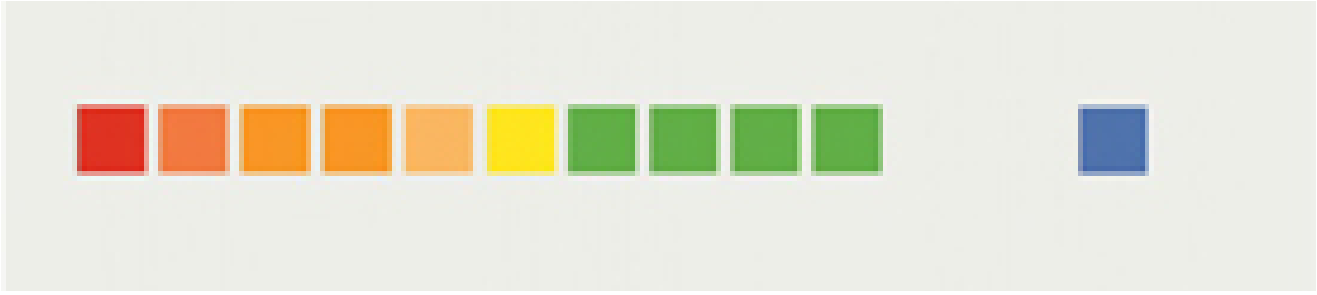


Inside & Outside Gotoh

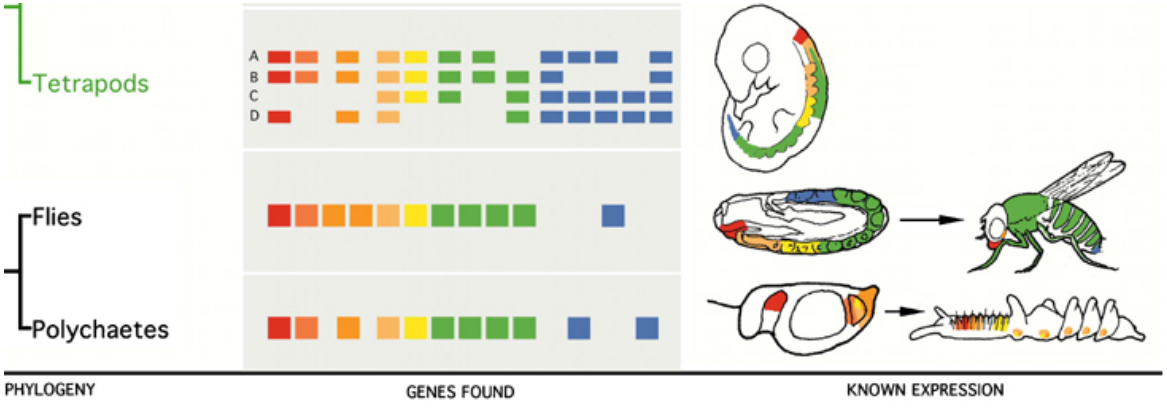


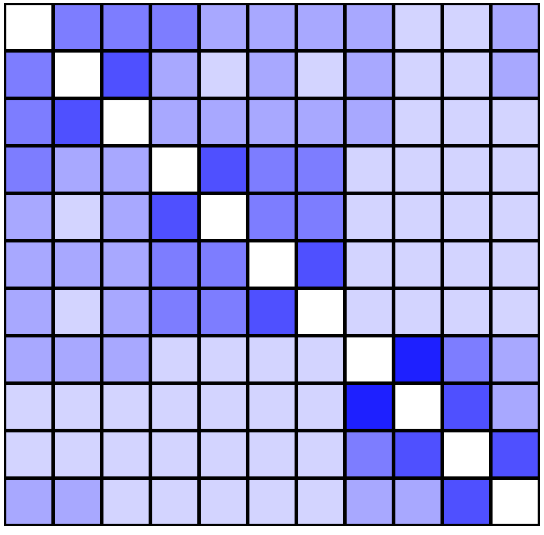
Hox Cluster History

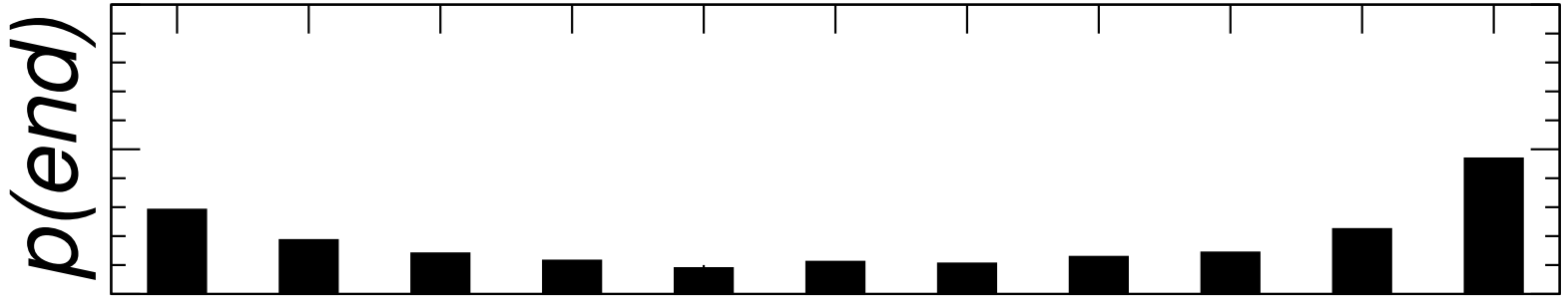




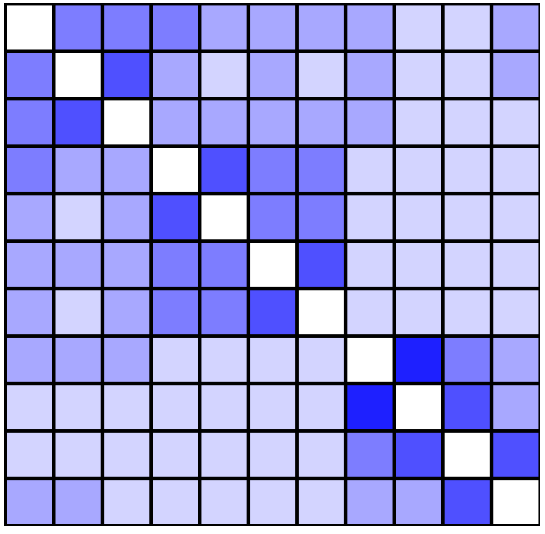
Extracting Summary Information







Calculate Edge Probabilities



$$P(i \sim j) = \frac{1}{Z(S)} \sum_{p,q}$$

$$\sum_{A \subset S} Z([p, A, i]) Z(\langle i, j \rangle) Z([j, S \setminus A, q]) \,.$$

$$S_{p,q} \to A_{p,q}$$

$$A_{p,q} \to \varepsilon$$

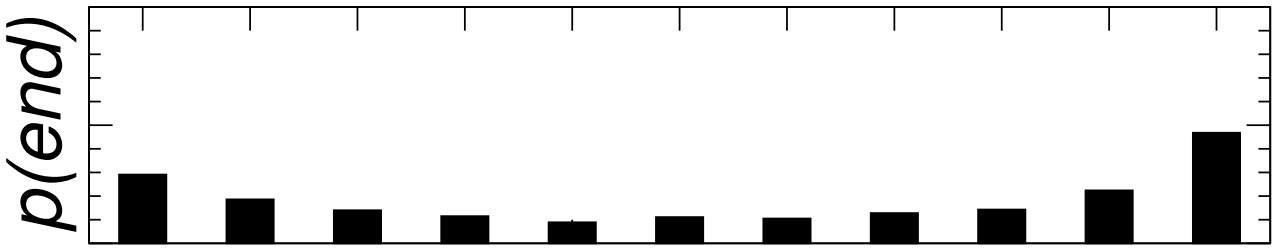
$$A_{p,q} \to A_{p,k}(k, q)$$

$$S \to A_{p,i}(i, j) A^*_{j,q}$$

$A_{p,q}$ `as left`

$A^*_{p,q}$ `mechanically derived`

Calculate End Probabilities



$$P(q \text{ is end}) = \frac{1}{Z(S)} \sum_{p,A \in S} Z([p, A, q])\,.$$

$S_{p,q} \to A_{p,q}$

$A_{p,q} \to \varepsilon$

$A_{p,q} \to A_{p,k}(k, q)$

$S_i \to A_{p,i}$

$A_{p,q}$ `as left`

Generalized ADP

Outside Grammars automatically derived from Inside grammars

Type-Rich DP index-type generic framework (with sequences and sets as examples)

- efficient algorithms as usual
- compatible with grammar products (grammars are objects)

Höner zu Siederdissen, Prohaska, Stadler. 2015
Algebraic Dynamic Programming over General Data Structures.

Hox figure:Building divergent body plans with similar genetic pathways; B J Swalla; Heredity