10th International Conference on
Unconventional Computation

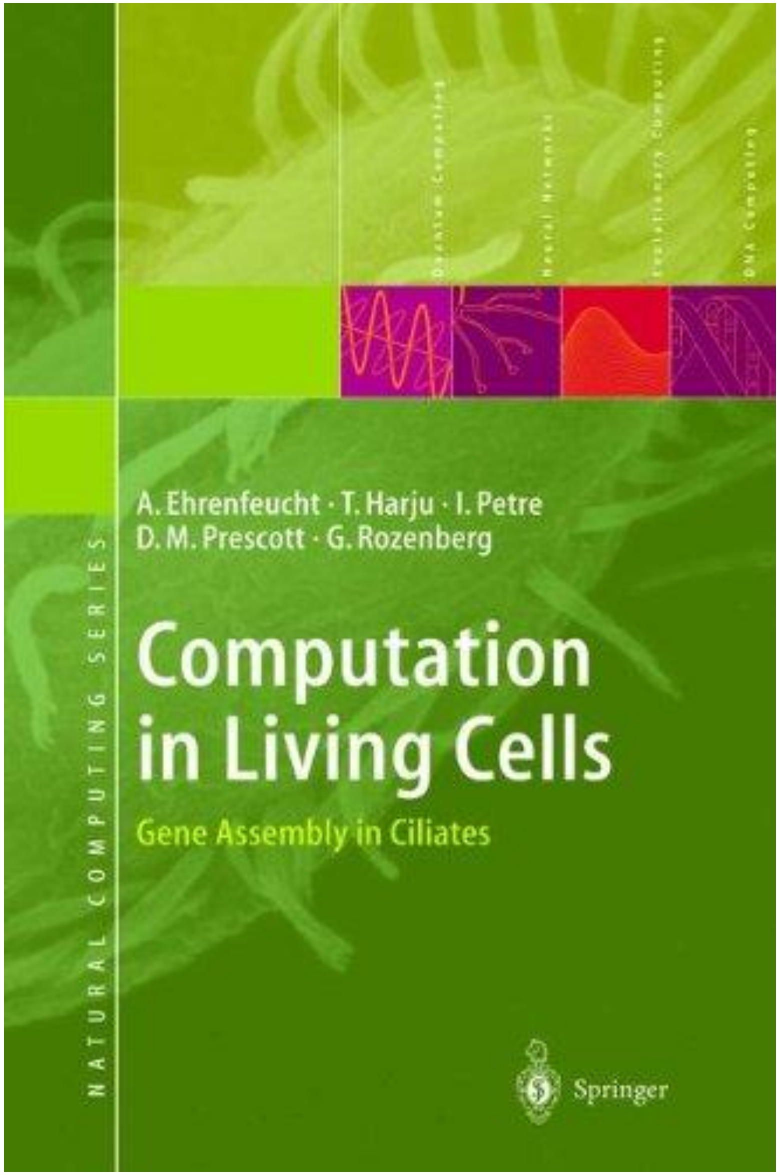
Workshop on
Language Theory in Biocomputing
Turku, June 9, 2011

The Algebra of Ciliates

Robert Brijder Hasselt

Hendrik Jan Hoogeboom
Leiden

image: http://www.depts.ttu.edu/hillcountry/research/protozoa.php

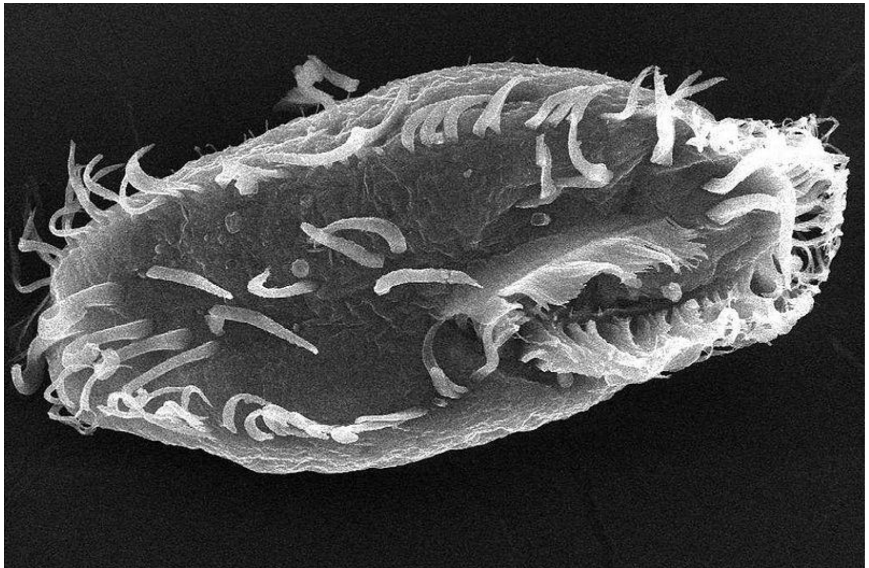


Computation in Living Cells
Gene Assembly in Ciliates

A. Ehrenfeucht, T. Harju, I. Petre,
D.M. Prescott, G. Rozenberg

Natural Computing Series, Springer,
2004.

ciliates



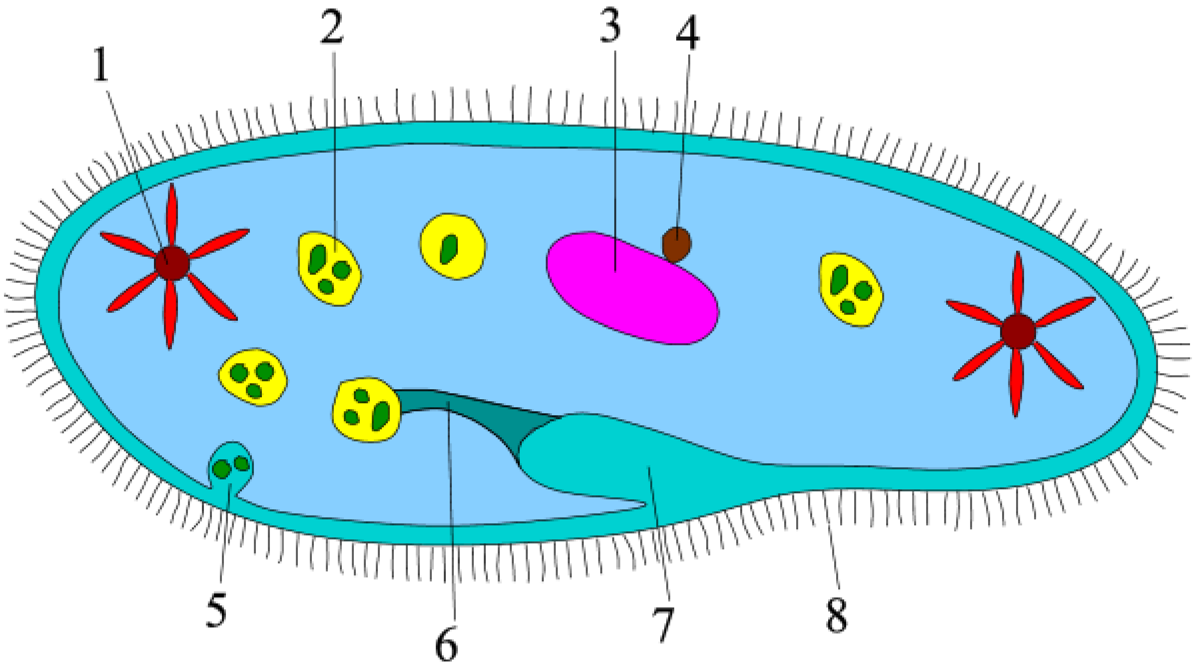
Oxytricha trifallax

The **ciliates** are a group of protozoans characterized by the presence of hair-like organelles called **cilia**, [...] variously used in swimming, crawling, attachment, feeding, and sensation.

Ciliates are one of the most important groups of protists, common almost everywhere there is water — in lakes, ponds, oceans, rivers, and soils. Ciliates have many ectosymbiotic and endosymbiotic members, as well as some obligate and opportunistic parasites. Ciliates tend to be large protozoa, a few reach 2 mm in length, and are some of the most complex protozoans in structure

http://en.wikipedia.org/wiki/Ciliate

micro and macro



cell structure:

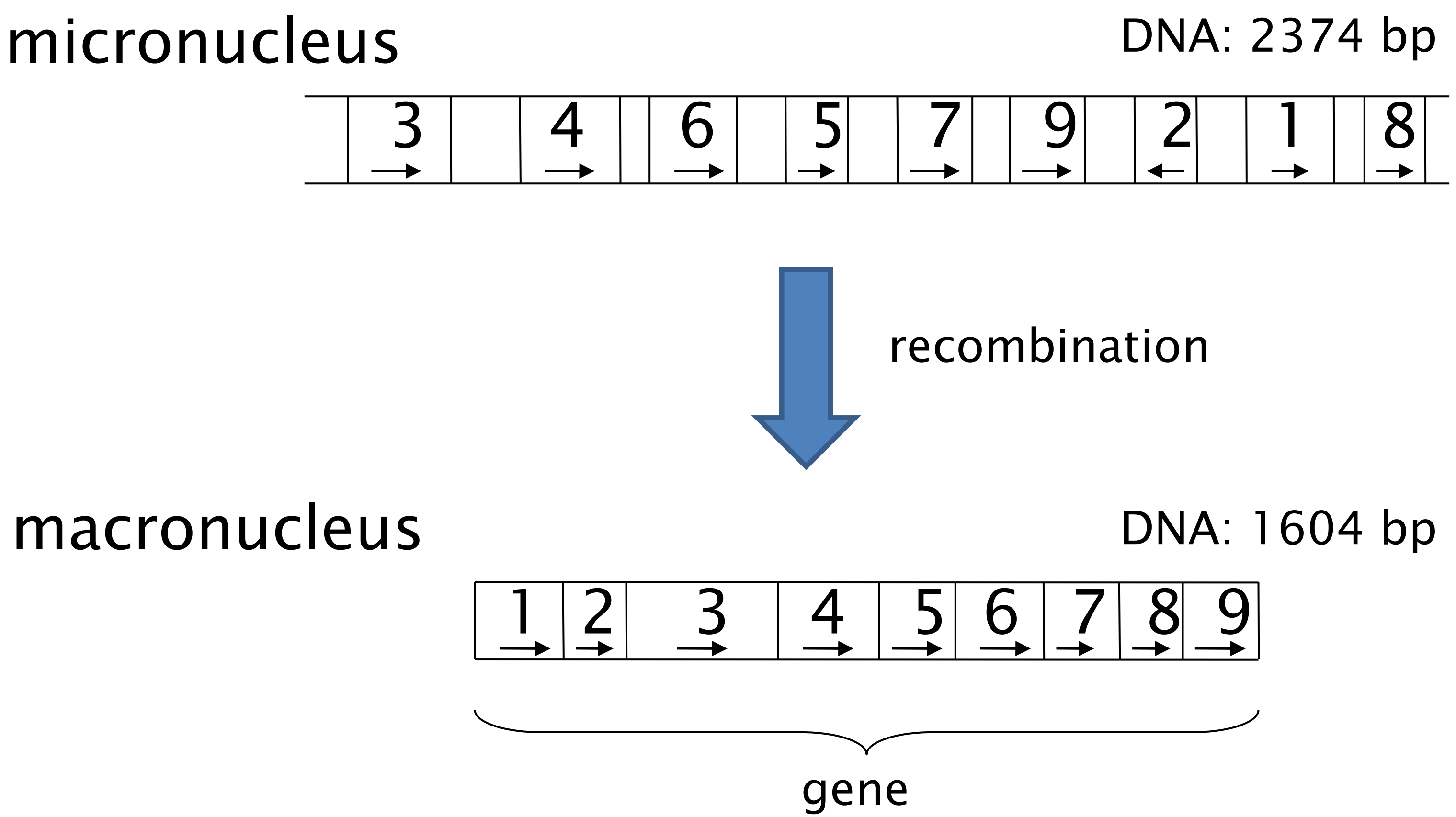
3. macronucleous
4. micronucleous
8. cilium

Unlike most other eukaryotes, ciliates have two different sorts of nuclei: a small, diploid **micronucleus** (reproduction), and a large, polyploid **macronucleus** (general cell regulation). The latter is generated from the micronucleus by amplification of the genome and heavy editing.

http://en.wikipedia.org/wiki/Ciliate

from micro to macro

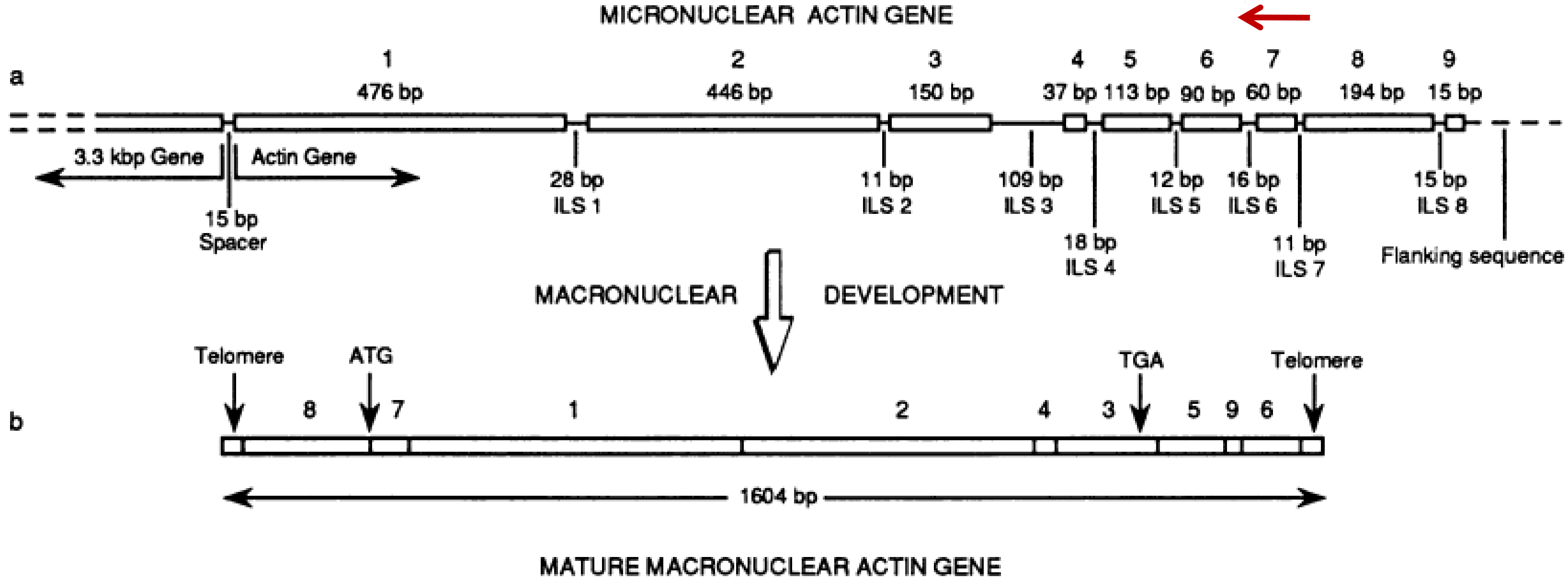
http://oxytricha.princeton.edu/cgi-bin/get_MDS_IES_Info.cgi?num=38



here: segment numbers in sorted order

micronucleus

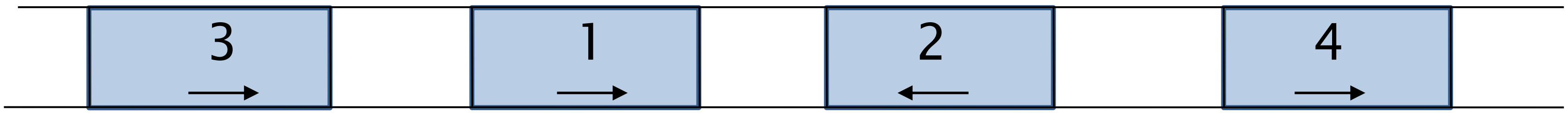
9 exons



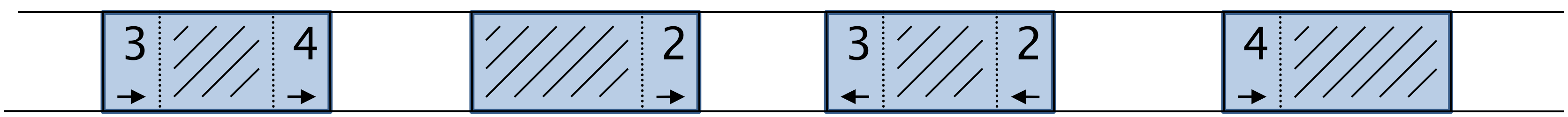
macronucleus

Greslin, Prescott etal. Reordering of nine exons is necessary to form a functional actin gene in Oxytricha nova. PNAS 86, 6264-6268, Aug 1989.

pointers

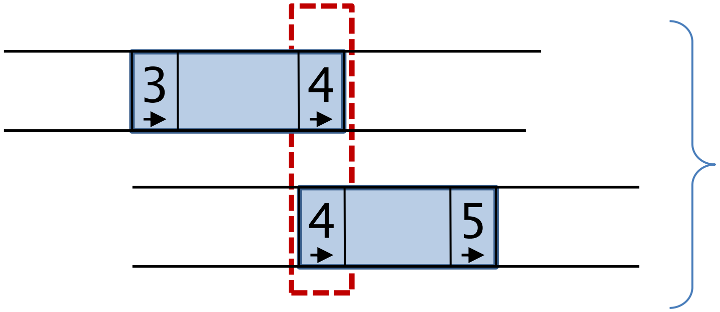


pointers – overlapping segments (for glueing)

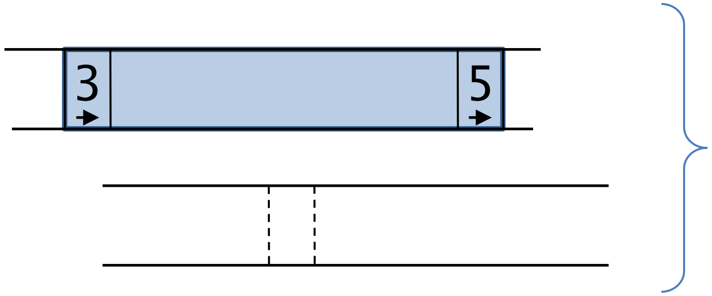


e.g., pointer 5 of actin gene: 13 bp

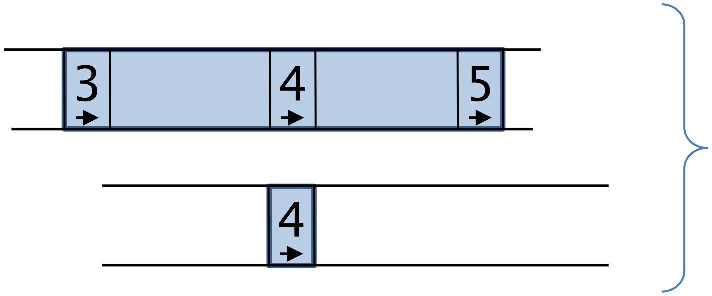
rc_4 recombination on pointer 4 'generic'



before



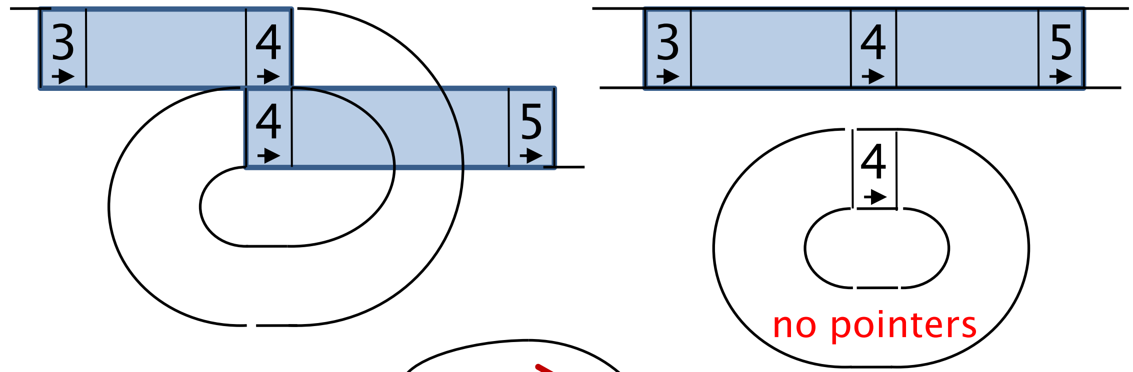
after 'ciliate view'



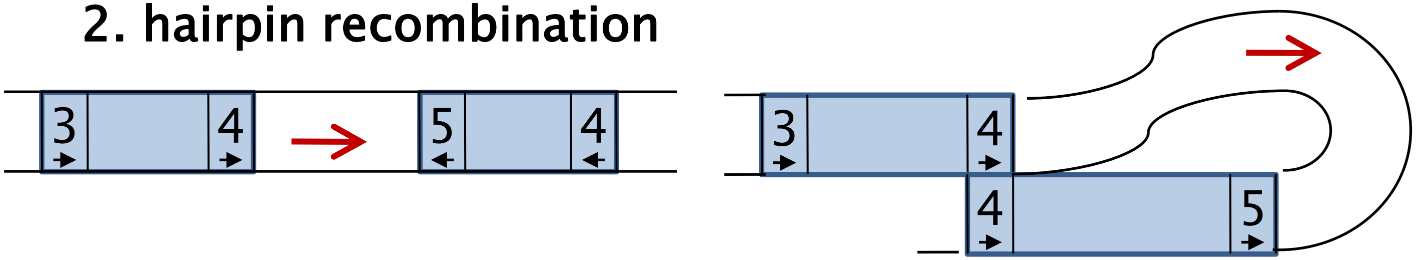
after 'math view'

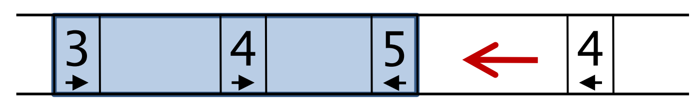
recombination on pointers

1. loop recombination

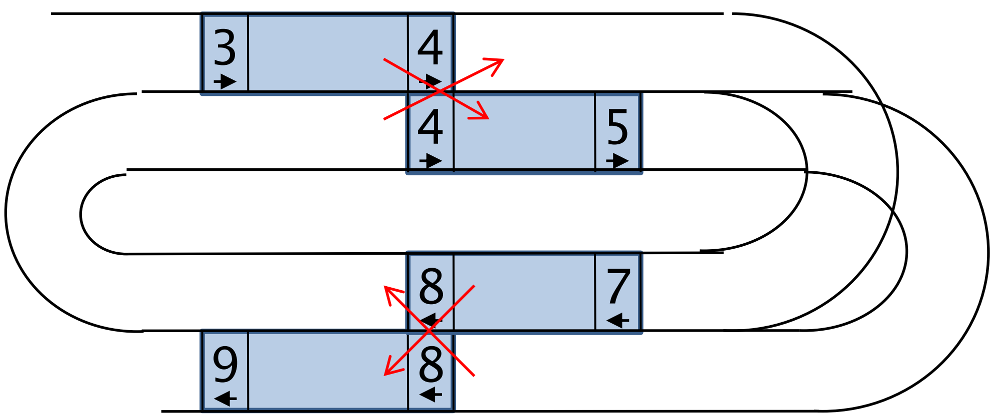


2. hairpin recombination





3. double-loop recombination

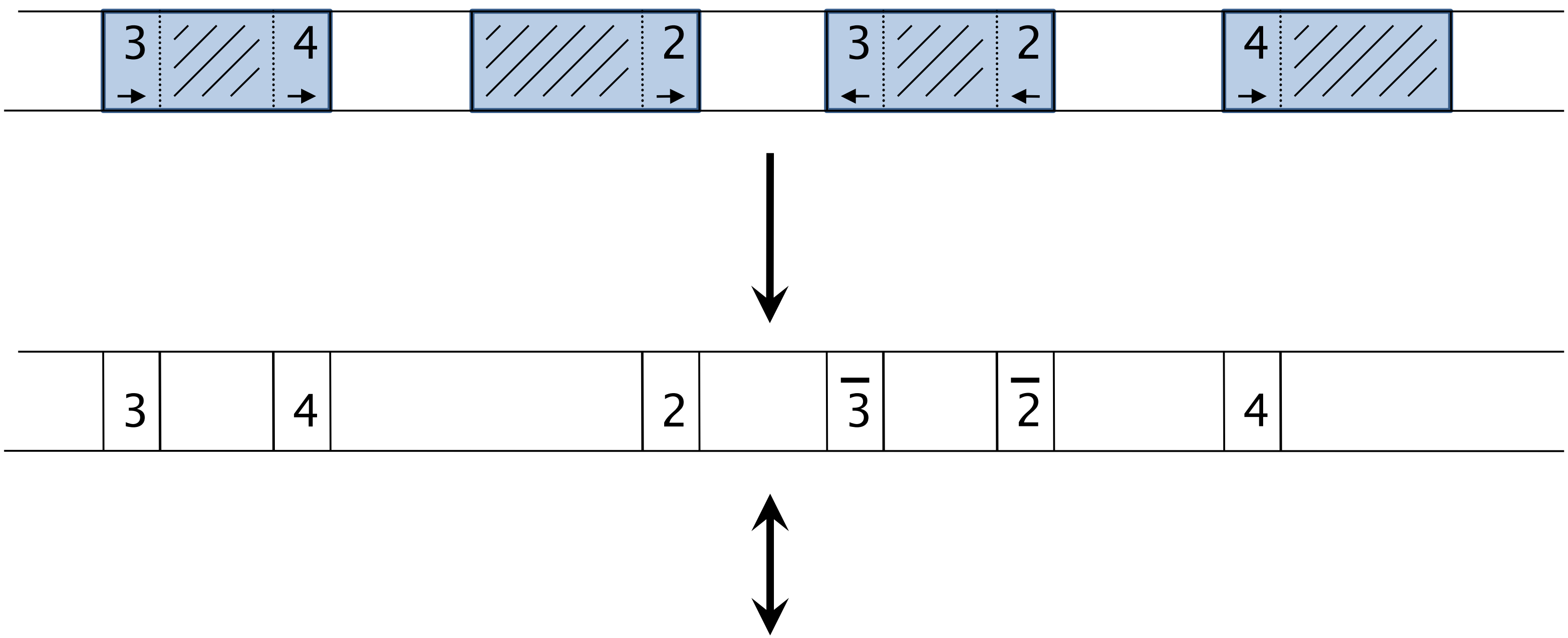


quest for the "right" model

- strings
- graphs
- matrices
- set systems

1

abstraction: pointers

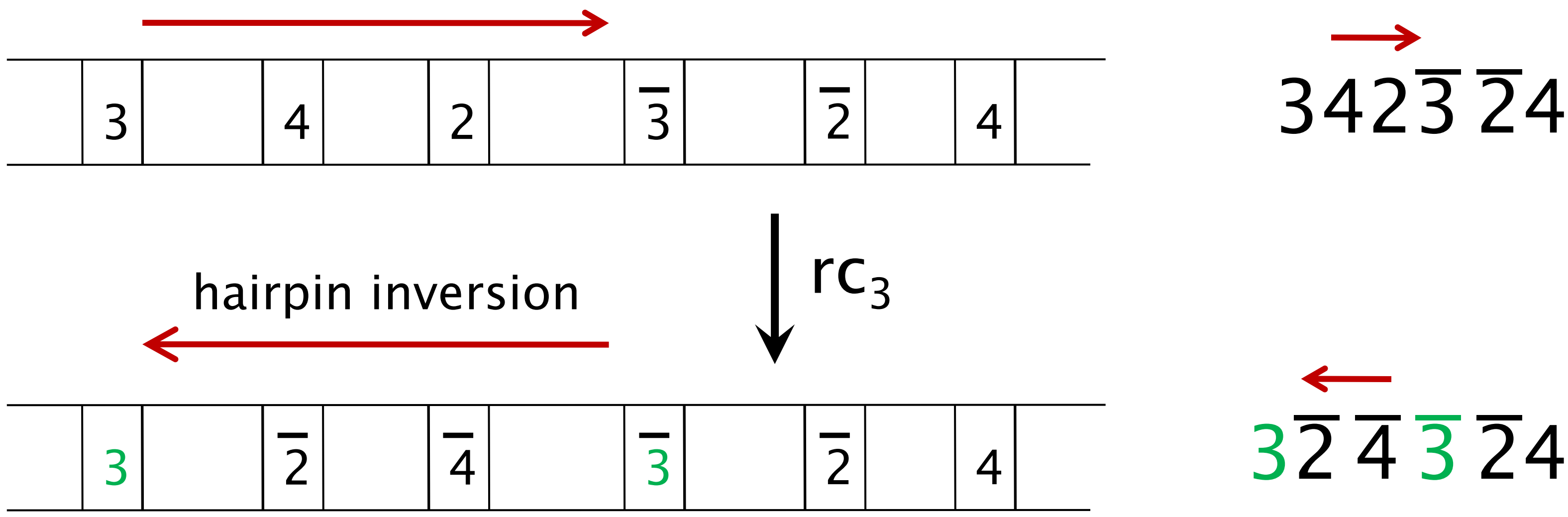


$342\bar{3}\bar{2}4$ ‘legal’ string

... 4774 ... realistic strings vs. generalizations

string positive rule

translating recombinations into string operations



$$rc_p(\ u_1 p u_2 \overline{p} u_3\ ) = u_1 p \overline{u}_2 \overline{p} u_3$$

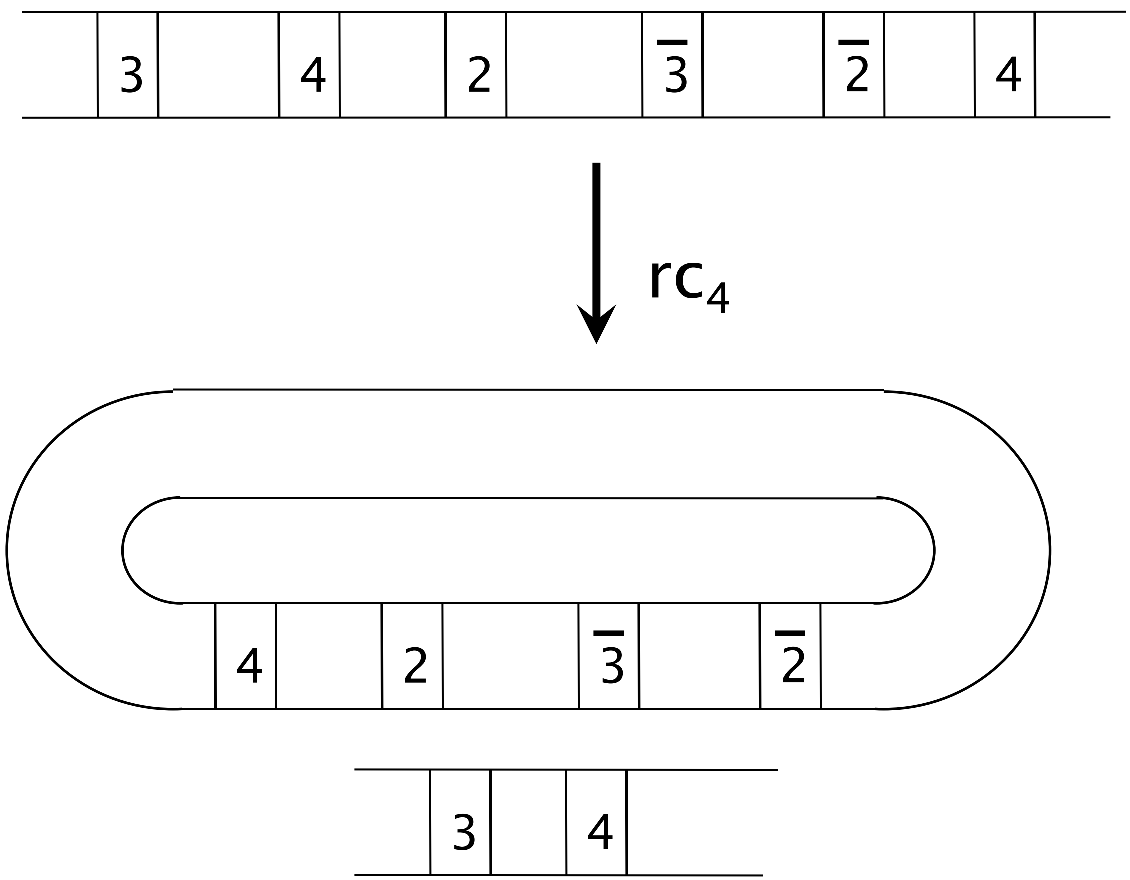
string pointer reduction systems

$$rc_p(\ u_1 pp u_2\ ) = u_1 pp u_2$$

no rearrangement
excision circular molecule

$$rc_p(\ u_1 p u_2 \overline{p} u_3\ ) = u_1 p \overline{u}_2 \overline{p} u_3$$

$$rc_{p,q}(\ u_1 p u_2 q u_3 p u_4 q u_5\ ) = u_1 p u_4 q u_3 p u_2 q u_5$$



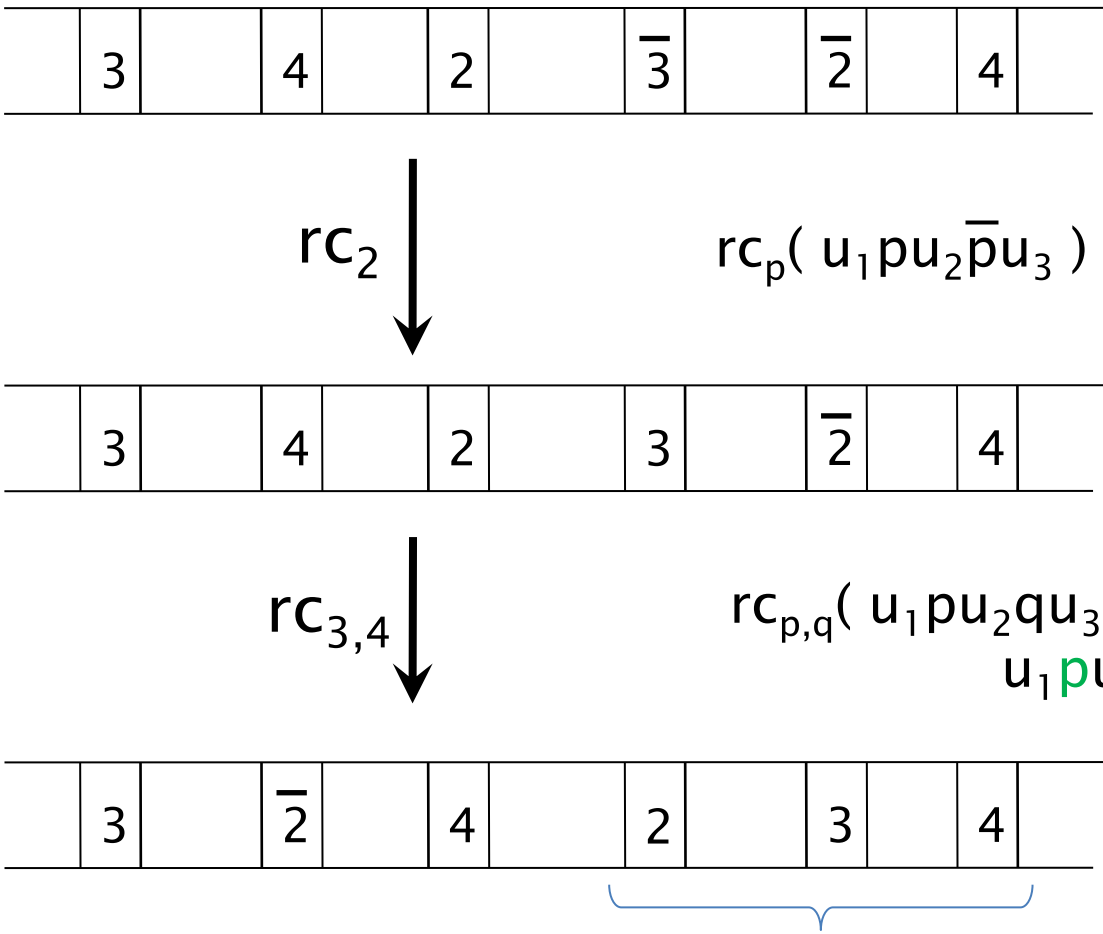
$34 2\bar{3}\,\bar{2}4$

undefined

(we will come beack to this)

sorting = reduction

Micronuclear DNA



$3 4 2 \bar{3}\, \bar{2} 4$

$rc_p(\, u_1 p u_2 \bar{p} u_3 \,) = u_1 p \bar{u}_2 \bar{p} u_3$

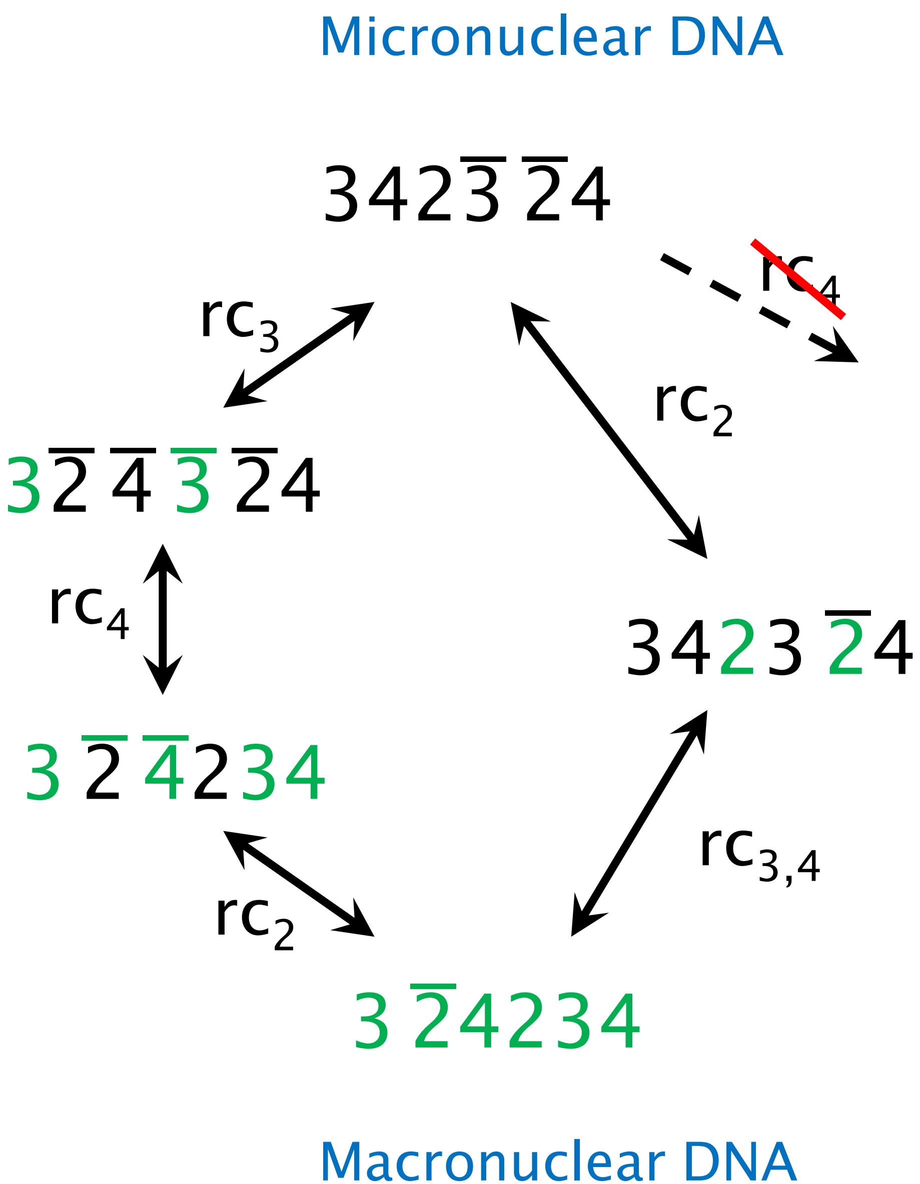
$3 4 2 3\, \bar{2} 4$

$rc_{p,q}(\, u_1 p u_2 q u_3 p u_4 q u_5 \,) =$
$u_1 p u_4 q u_3 p u_2 q u_5$

$3\, \bar{2} 4 2 3 4$

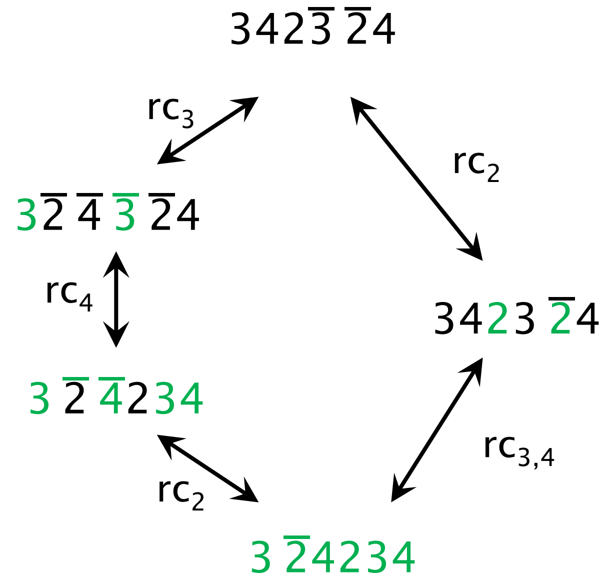
Macronuclear DNA

nondeterministic



$$rc_p(\, u_1 p u_2 \bar{p} u_3 \,) = u_1 p \bar{u}_2 \bar{p} u_3$$

$$rc_{p,q}(\, u_1 p u_2 q u_3 p u_4 q u_5 \,) = u_1 p u_4 q u_3 p u_2 q u_5$$



question:

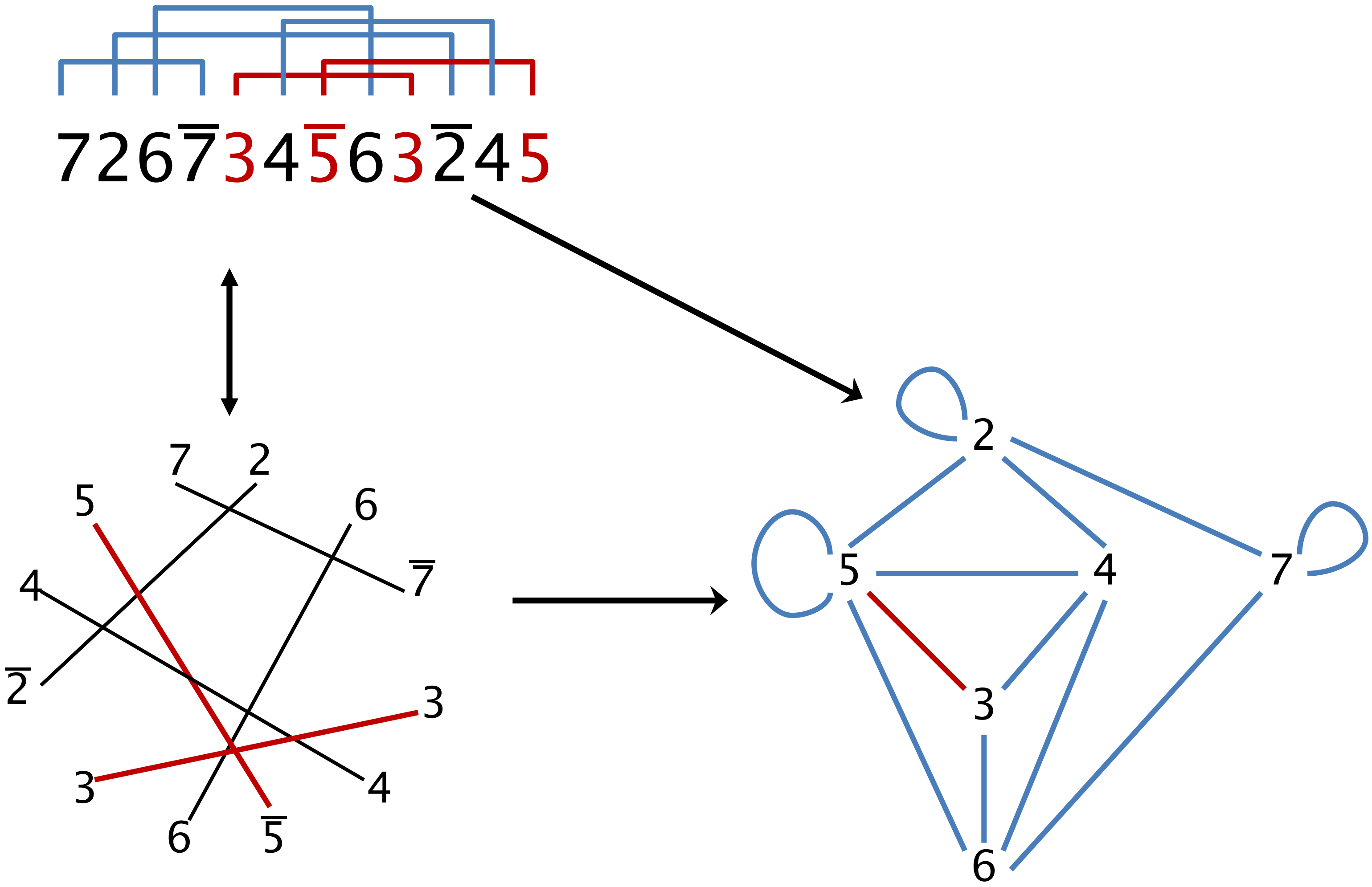
is the result of reductions independent of operations chosen?

2

quest for the "right" model

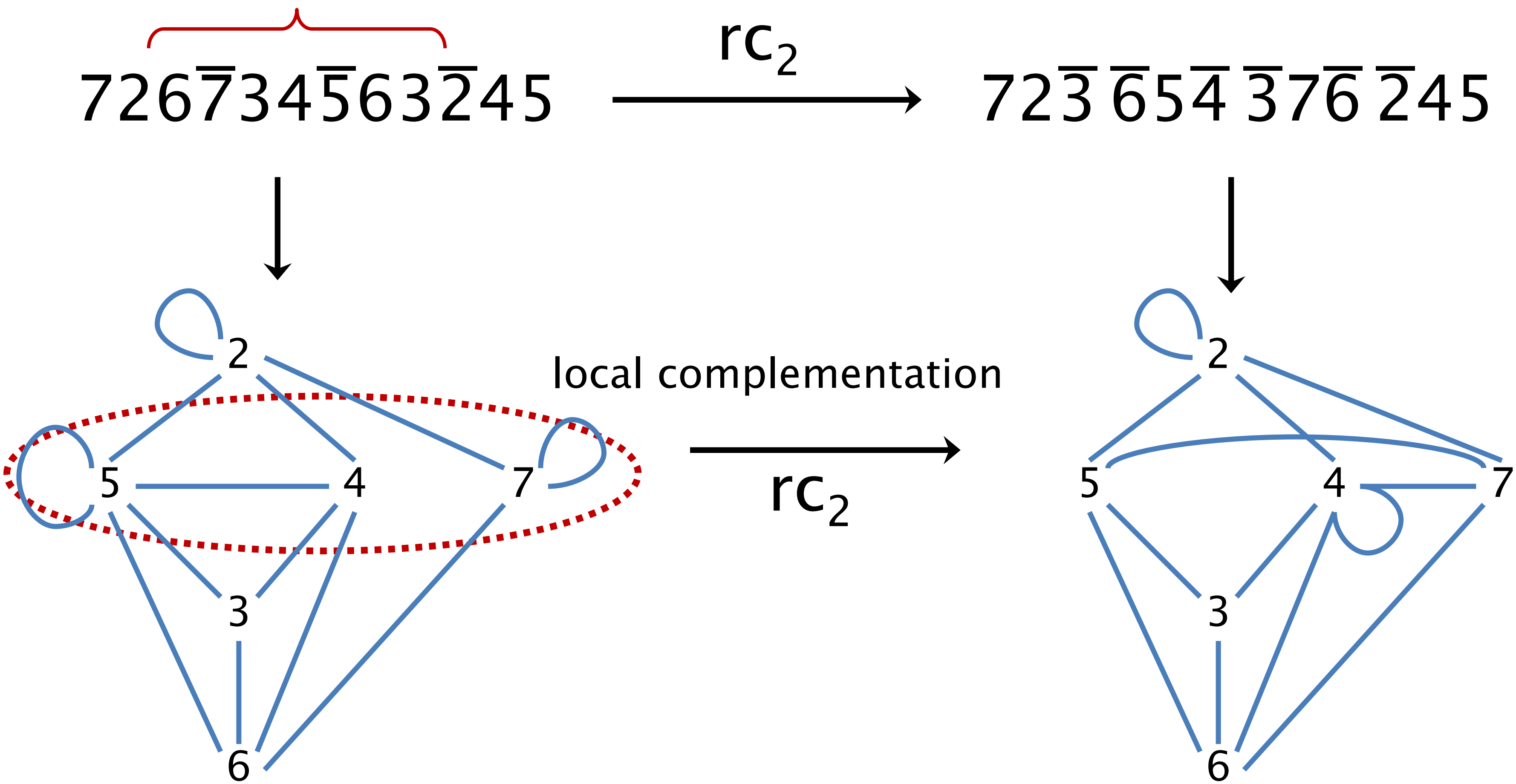
- strings
- graphs
- matrices
- set systems

circle & overlap graph



string to overlap graph

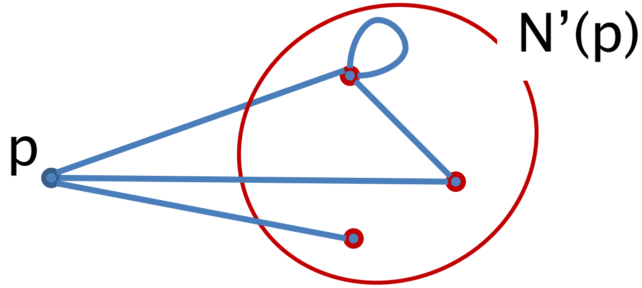
real generalization

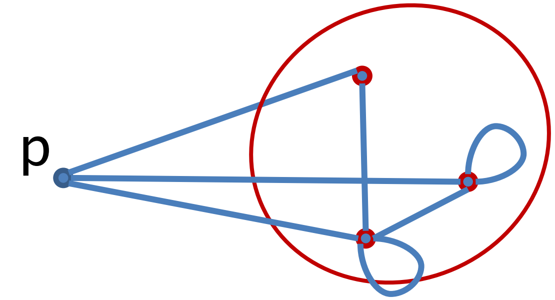


Ehrenfeucht etal, *Theor. Comp. Sci.*, 2003

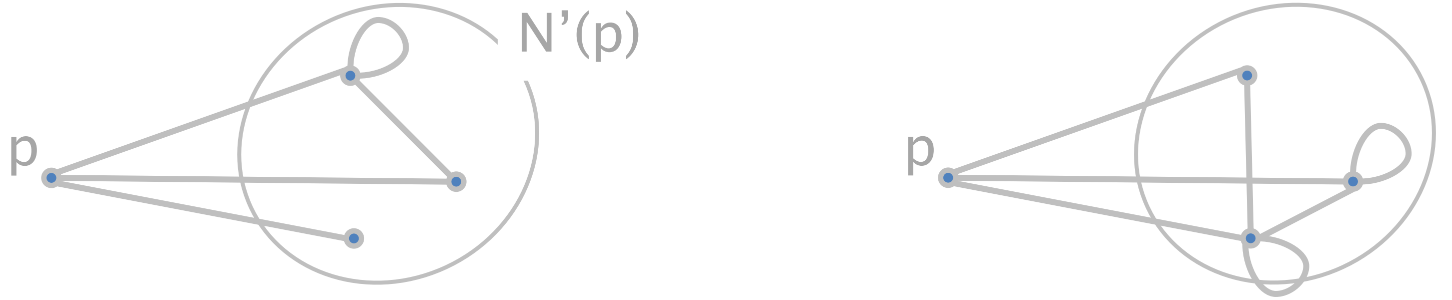
(for signed graphs instead of looped graphs)

rc_p *local complementation* looped vertex p

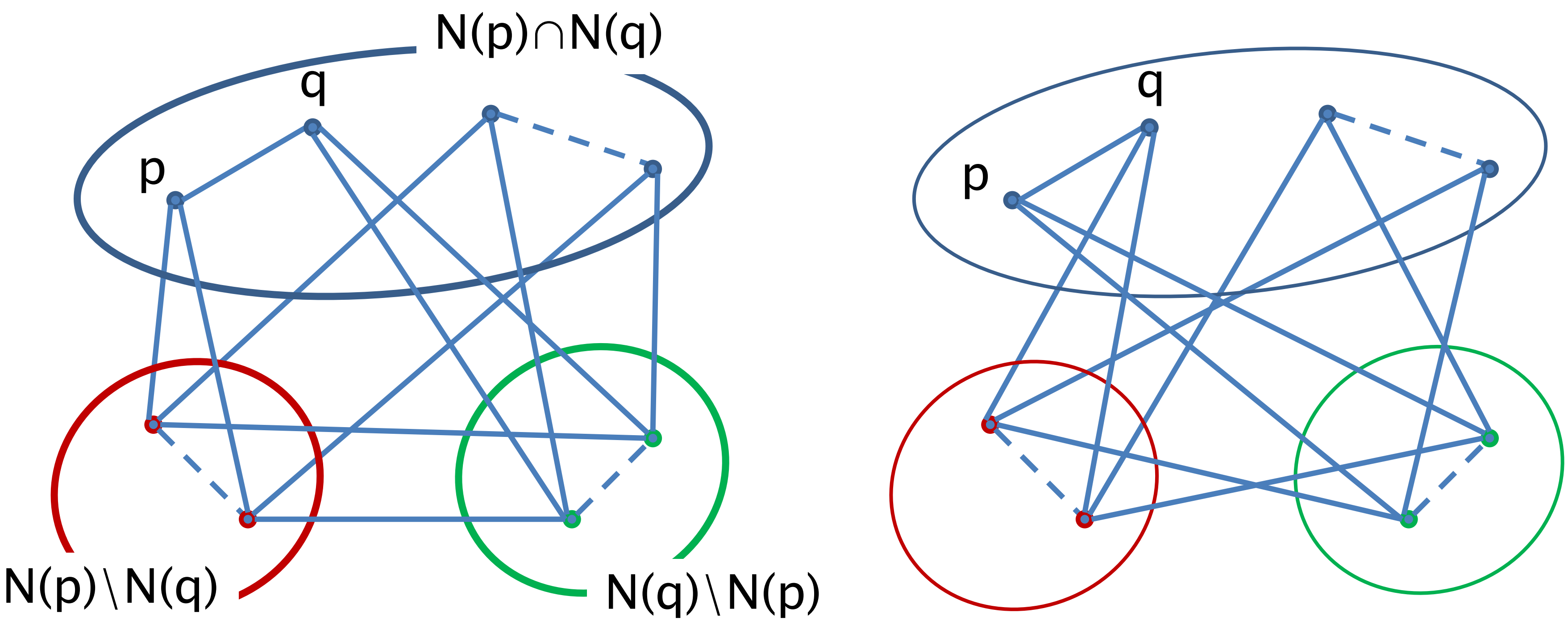




rc_p *local complementation* looped vertex p

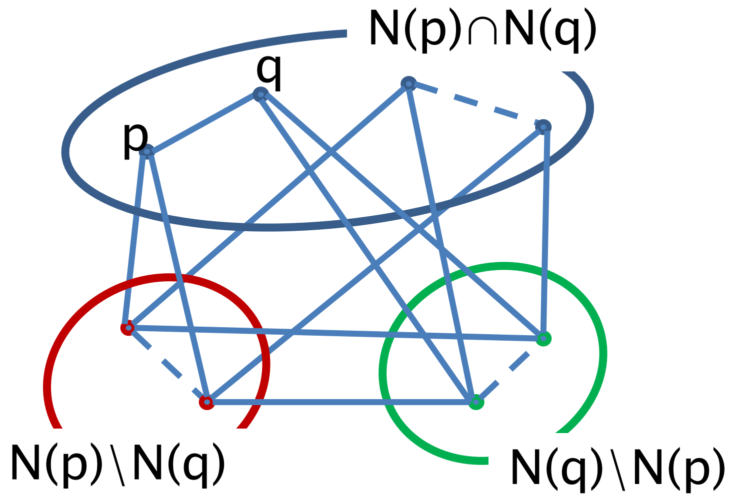


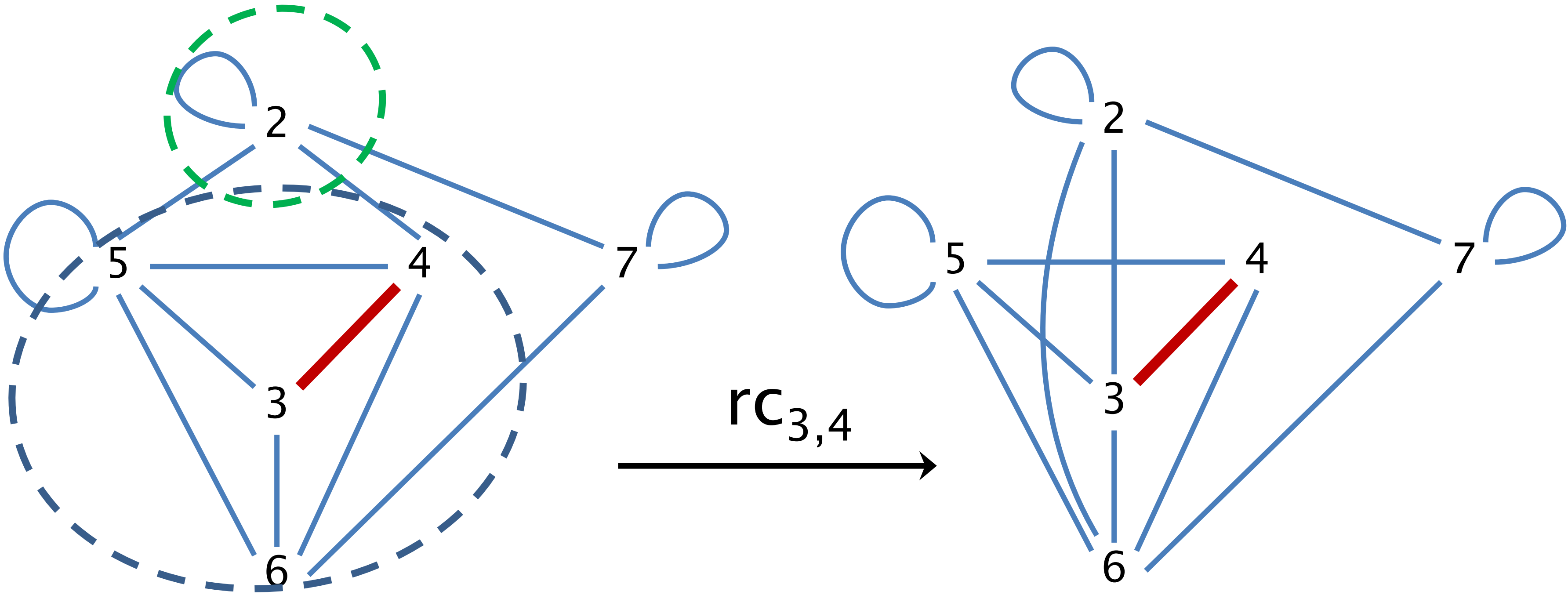
$rc_{p,q}$ *edge complementation* unlooped edge pq



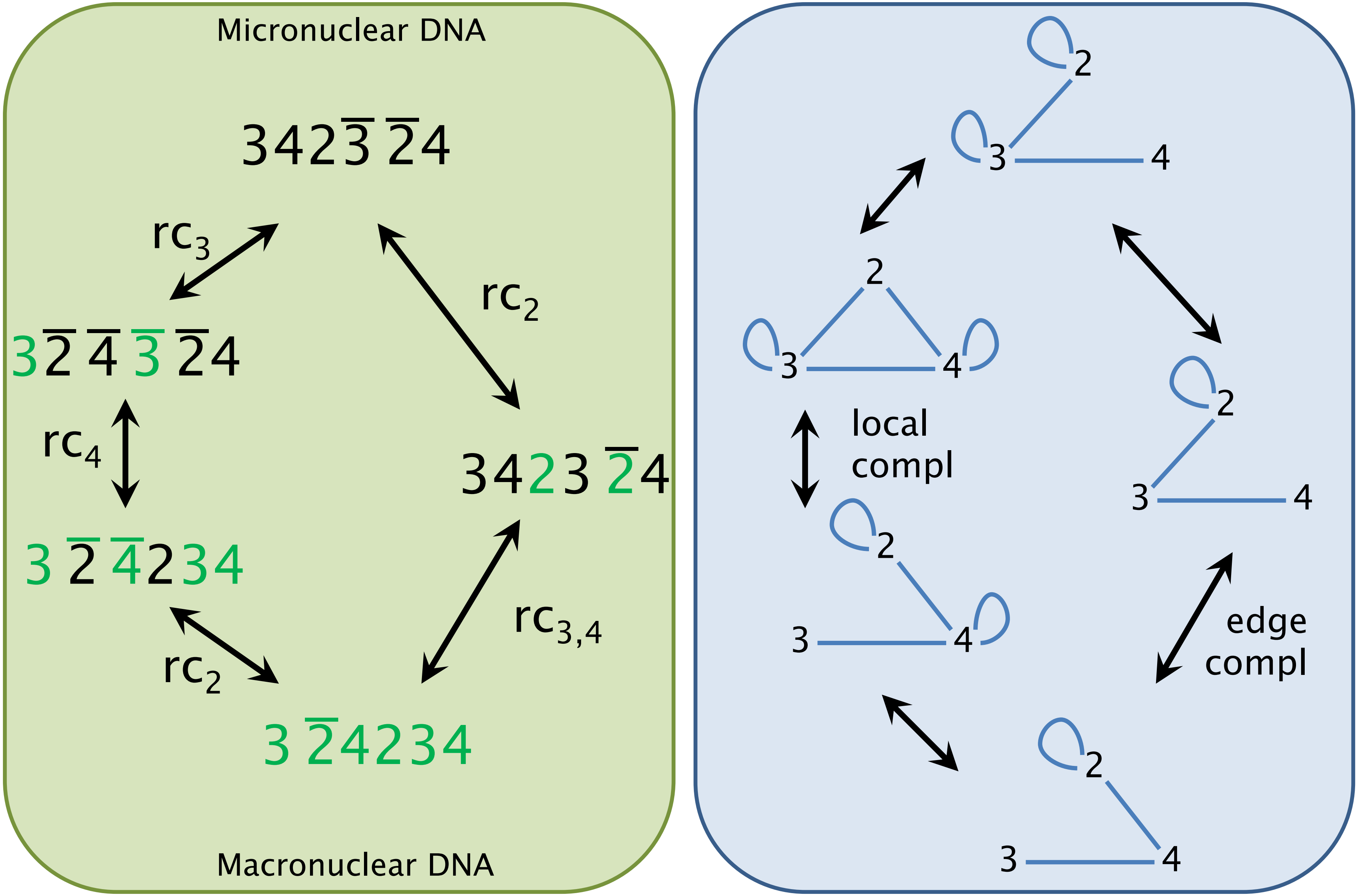
example edge complement

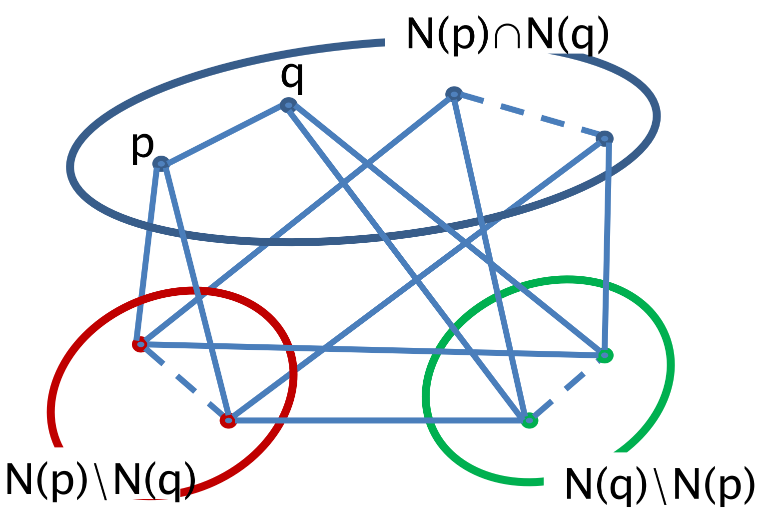
$rc_{3,4}$ on edge 3,4





two worlds


Micronuclear DNA
$3 4 2 \bar{3} \bar{2} 4$
rc3
rc2
$3 \bar{2} \bar{4} \bar{3} \bar{2} 4$
rc4
$3 4 2 3 \bar{2} 4$
$3 \bar{2} \bar{4} 2 3 4$
rc3,4
rc2
$3 \bar{2} 4 2 3 4$
Macronuclear DNA
2
3
4
local
compl
edge
compl



question:

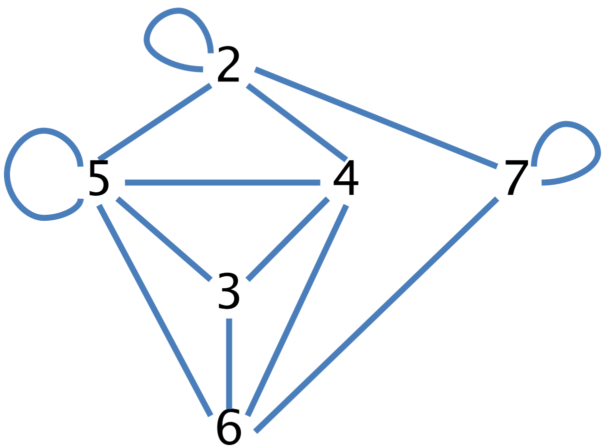
how do $rc_{p,q}$ and $rc_{p',q'}$ interact ?

3

quest for the “right” model

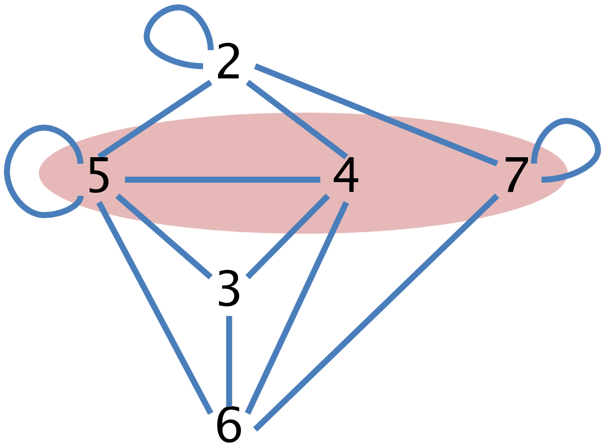
- strings
- graphs
- matrices
- set systems

graphs and matrices



	2	3	4	5	6	7
2	1	0	1	1	0	1
3	0	0	1	1	1	0
4	1	1	0	1	1	0
5	1	1	1	1	1	0
6	0	1	1	1	0	1
7	1	0	0	0	1	1

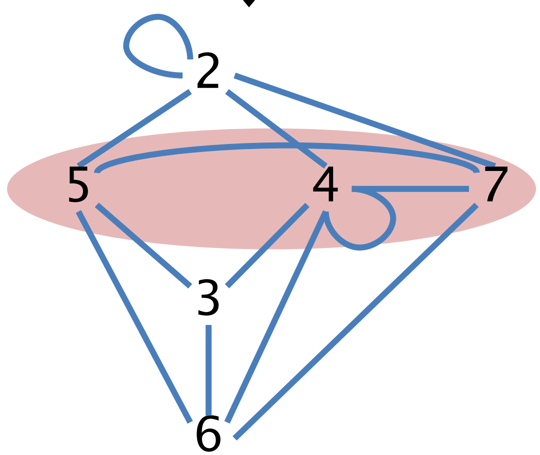
reconsider local/edge complementation



	2	3	4	5	6	7
2	1	0	1	1	0	1
3	0	0	1	1	1	0
4	1	1	0	1	1	0
5	1	1	1	1	1	0
6	0	1	1	1	0	1
7	1	0	0	0	1	1

rc_2

rc_2



	2	3	4	5	6	7
2	1	0	1	1	0	1
3	0	0	1	1	1	0
4	1	1	1	0	1	1
5	1	1	0	0	1	1
6	0	1	1	1	0	1
7	1	0	1	1	1	0

what *is* happening?

$3 4 2 \bar{3} \bar{2} 4$ $\xrightarrow{rc_3\ rc_4\ rc_2}$ $3 \bar{2} 4 2 3 4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \qquad \xrightarrow{rc_3\ rc_4\ rc_2} \qquad \begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 4 & 1 & 1 & 0 \end{array}$$

multiply (over the binary numbers)

$$\underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{\text{micro}} \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}}_{\text{macro}} = \begin{pmatrix} & & \\ & & \\ & 0 & \end{pmatrix}$$

$+$ xor $\oplus$ $1+1=0$

$*$ and $\wedge$

what *is* happening? inversion

$3\,4\,2\,\bar{3}\,\bar{2}\,4 \xrightarrow{rc_3\ rc_4\ rc_2} 3\,\bar{2}\,4\,2\,3\,4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \quad \xrightarrow{rc_3\ rc_4\ rc_2} \quad \begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 4 & 1 & 1 & 0 \end{array}$$

multiply (over the binary numbers)

$$\underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{\text{micro}} \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}}_{\text{macro}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$$

sorting DNA = computing the inverse

partial inverse

$A\,x = y$ iff $A^{-1}\,y = x$

principal pivot transform

$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \text{iff} \quad A{*}X\begin{bmatrix} y_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_2 \end{bmatrix}$$

← X pointers

← other

$A * X$ is defined iff $A[X]$ is invertible

real recipe (which we do not need)

$$A = \begin{matrix} & X & \\ X & \left(\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right) \end{matrix} \qquad A * X = \left(\begin{array}{c|c} P^{-1} & -P^{-1}\,Q \\ \hline R\,P^{-1} & S - R\,P^{-1}\,Q \end{array}\right)$$

$P = A[X]$ invertible / nonsingular i.e. $\det P \neq 0$

M.J. Tsatsomeros. Principal pivot transforms: properties and applications.
Linear Algebra and its Applications, 307(1-3):151-165, 2000

principal pivot transform

$$A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \text{iff} \quad A{*}X \begin{pmatrix} y_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$$

using partial inversion

xor

$$(A * X) * Y = A * (X \oplus Y)$$ (when defined)

$$A * \{p_1, p_2\} \ldots * p_n = A * V = A^{-1}$$ (all pointers)

this shows

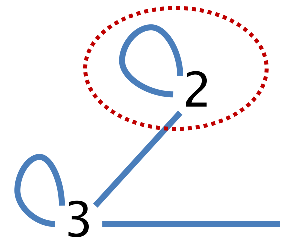
- how the rc_p and $rc_{p,q}$ interact
- result does not depend on order of operations

applicability

A * X is defined iff A[X] is invertible

rc_2

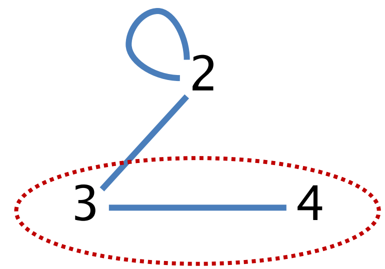
$3\,4\,2\,\overline{3}\,\overline{2}\,4$



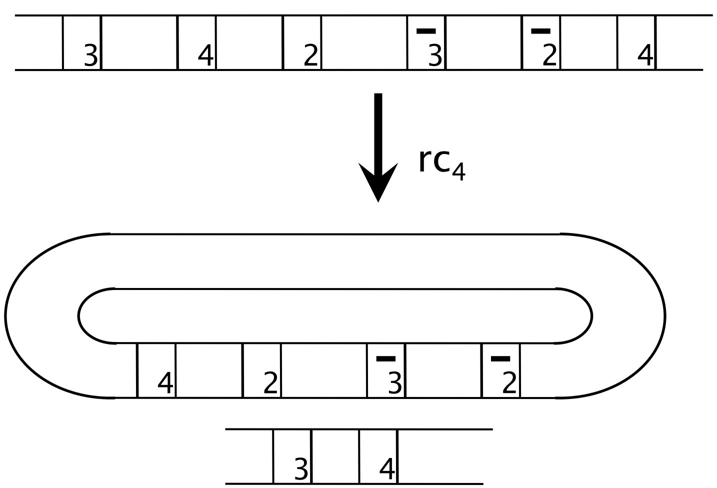
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$rc_{3,4}$

$3\,4\,2\,3\,\overline{2}\,4$

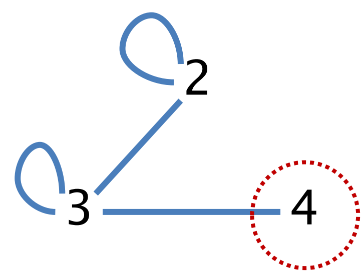


$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



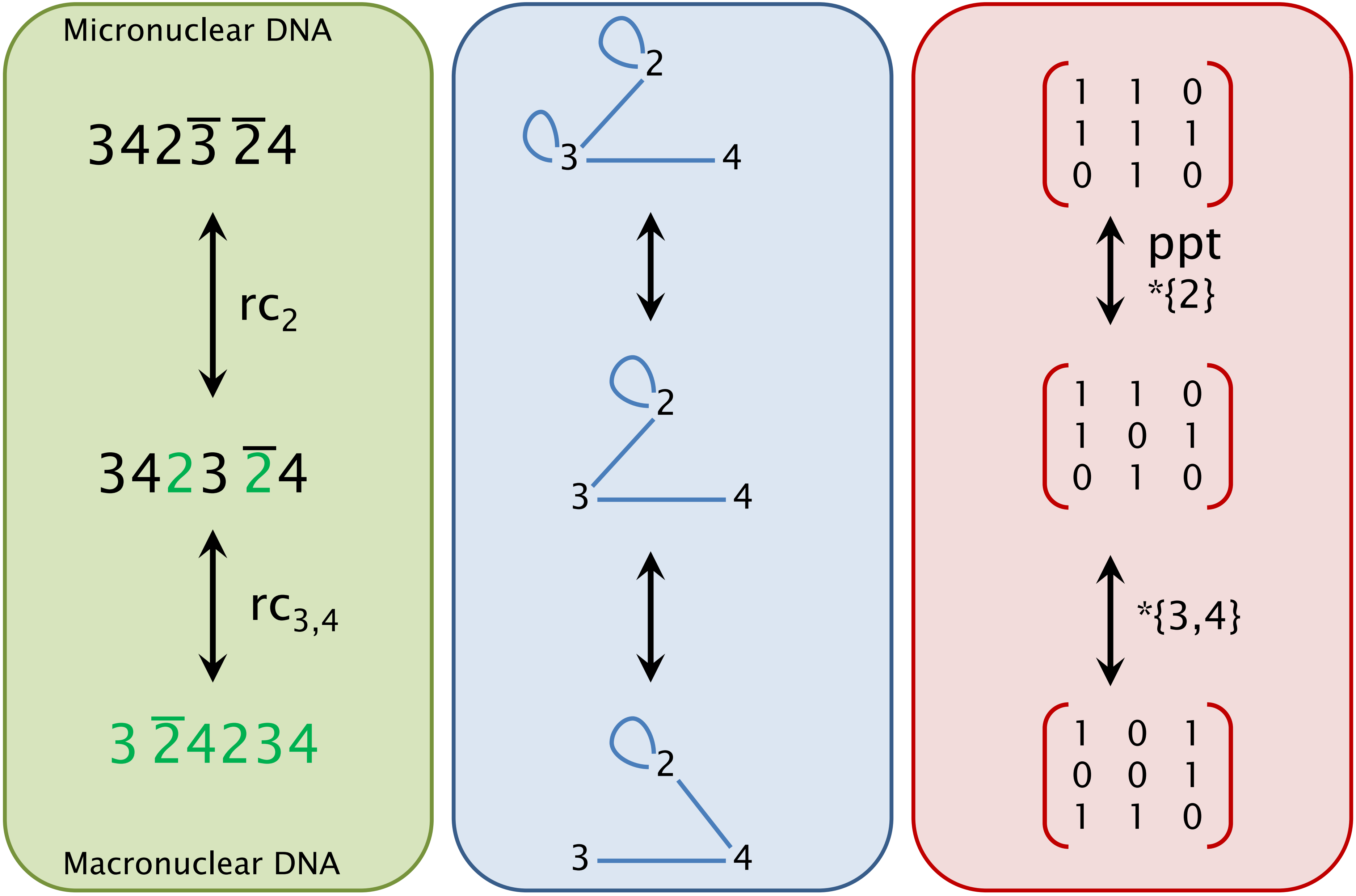
$3\,4\,2\,\overline{3}\,\overline{2}\,4$

undefined



	2	3	4
2	1	1	0
3	1	1	1
4	0	1	0

three worlds



conclusion (for now)

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt)
- we can use results about ppt to know more about gene assembly
  - independent order operations
  - interaction operations

quest for the "right" model

- strings
- graphs
- matrices
- set systems

the most elegant model was hidden

4

applicable sets

A * X is defined iff A[X] is invertible

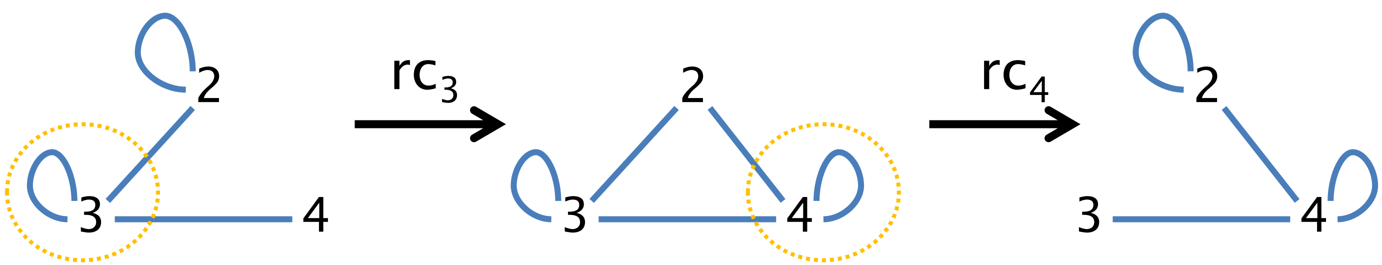
$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \quad \longleftrightarrow$$

A[{3,4}]

set system

V = {2,3,4}

D = { ∅, {2}, {3}, {2,4}, {3,4}, {2,3,4} }



which operation ?

graphs $\subseteq$ set systems (strict)

$342\bar{3}\bar{2}4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \quad \longleftrightarrow \quad \begin{array}{l} V = \{2,3,4\} \\ D = \{\, \emptyset, \{2\}, \{3\}, \{2,4\}, \{3,4\}, \{2,3,4\} \,\} \end{array}$$

$\updownarrow$ rc_3 $\qquad\qquad$ $\updownarrow$?

$3\bar{2}\bar{4}\bar{3}\bar{2}4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 0 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 \end{array} \quad \longleftrightarrow \quad \begin{array}{l} V = \{2,3,4\} \\ D' = \{\, \emptyset, \{3\}, \{4\}, \{2,3\}, \{2,4\}, \{2,3,4\} \,\} \end{array}$$

graphs $\subseteq$ set systems (strict)

$3\,4\,2\,\overline{3}\,\overline{2}\,4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \longleftrightarrow \begin{array}{l} V = \{2,3,4\} \\ D = \{ \emptyset, \{2\}, \{3\}, \{2,4\}, \{3,4\}, \{2,3,4\} \} \end{array}$$

$\updownarrow$ rc_3 $\qquad\qquad$ $\updownarrow$?

$3\,\overline{2}\,\overline{4}\,\overline{3}\,\overline{2}\,4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 0 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 \end{array} \longleftrightarrow \begin{array}{l} V = \{2,3,4\} \\ D' = \{ \{3\}, \{2,3\}, \emptyset, \{2,3,4\}, \{4\}, \{2,4\} \} \end{array}$$

graphs $\subseteq$ set systems (strict)

$3\,4\,2\,\overline{3}\,\overline{2}\,4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \quad \longleftrightarrow \quad \begin{array}{l} V = \{2,3,4\} \\ D = \{\, \emptyset, \{2\}, \{3\}, \{2,4\}, \{3,4\}, \{2,3,4\} \,\} \end{array}$$

rc_3 $\updownarrow$ $\oplus 3$ $\updownarrow$ xor 3

$3\,\overline{2}\,\overline{4}\,\overline{3}\,\overline{2}\,4$

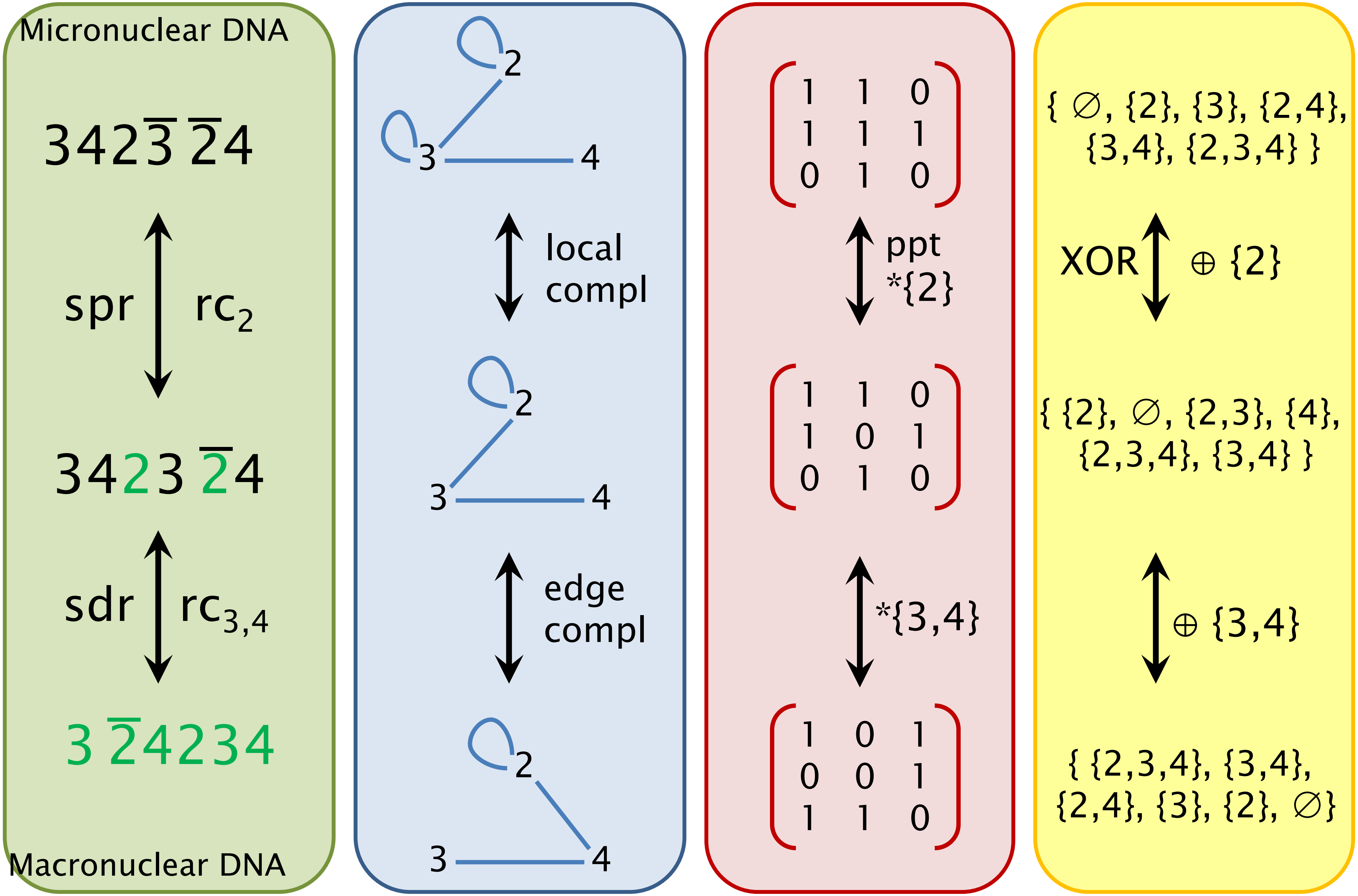
$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 0 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 \end{array} \quad \longleftrightarrow \quad \begin{array}{l} V = \{2,3,4\} \\ D' = \{\, \{3\}, \{2,3\}, \emptyset, \{2,3,4\}, \{4\}, \{2,4\} \,\} \end{array}$$

applicability (!)

XOR {4} is defined, while rc_4 is not

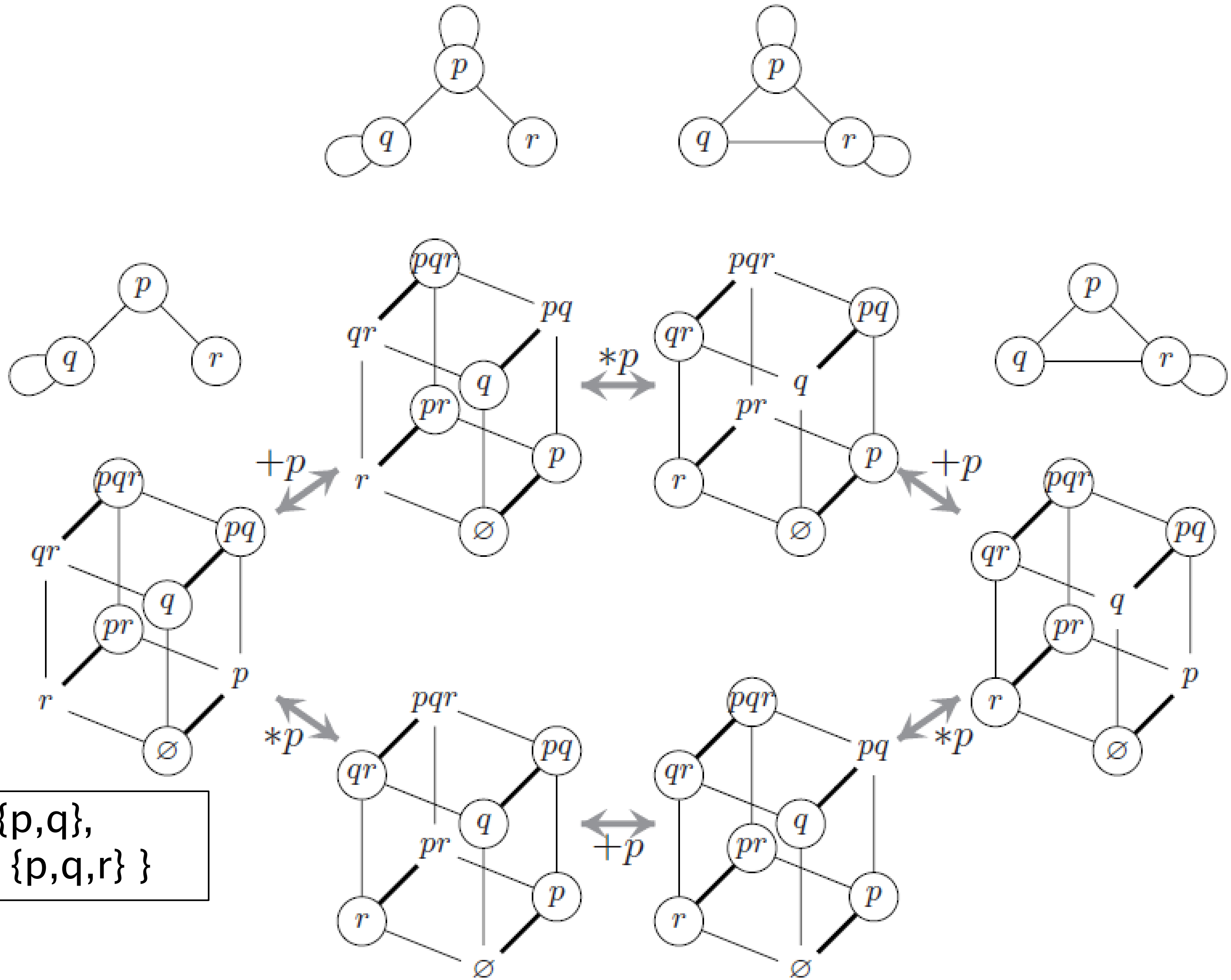
nb. {4} not in D

four worlds



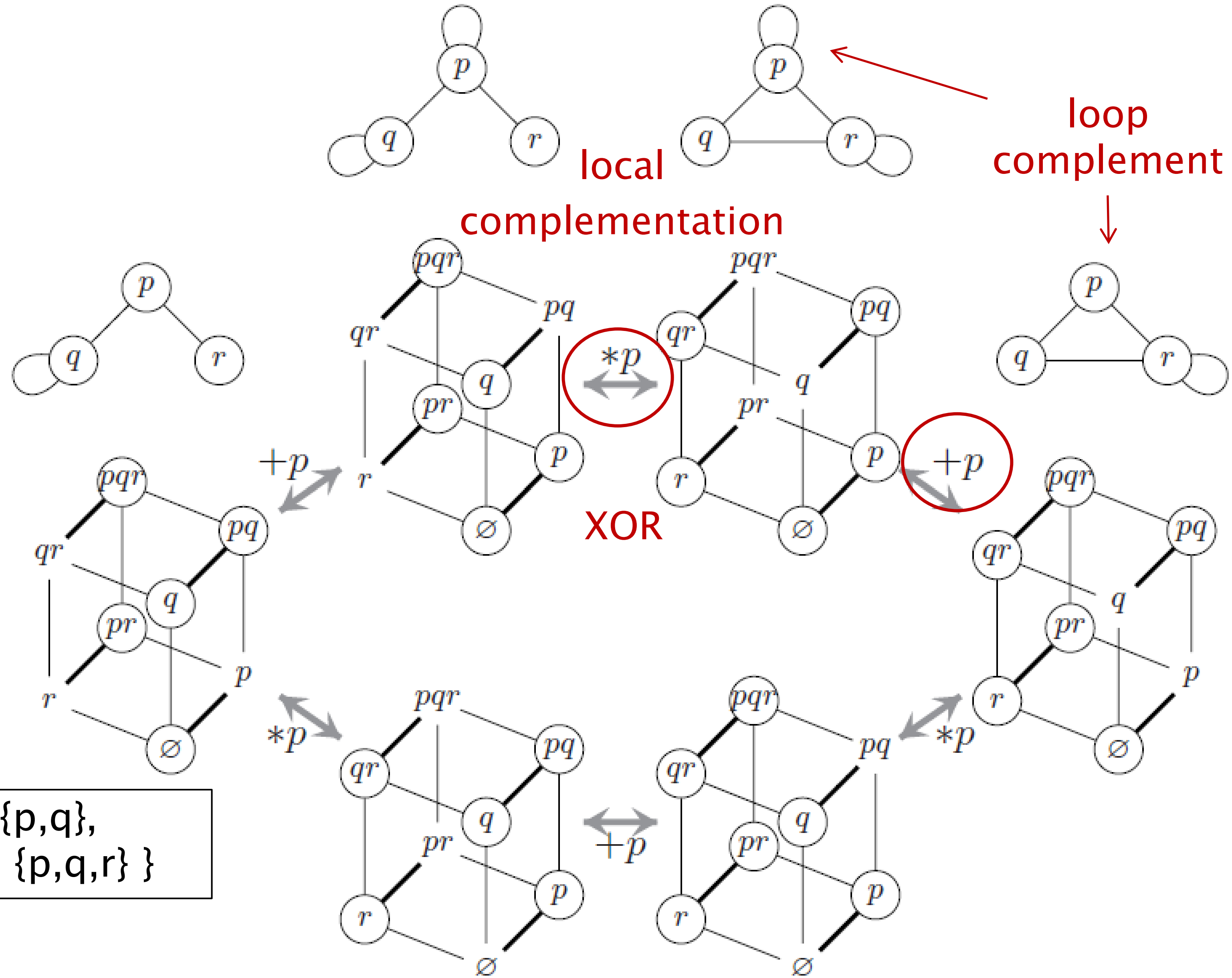
Micronuclear DNA
$342\overline{3}\,\overline{2}4$
spr
rc_2
$3423\overline{2}4$
sdr
$rc_{3,4}$
$3\overline{2}4234$
Macronuclear DNA
2
3
4
local
compl
2
3
4
edge
compl
2
3
4
1 1 0
1 1 1
0 1 0
ppt
*{2}
1 1 0
1 0 1
0 1 0
*{3,4}
1 0 1
0 0 1
1 1 0
{ ∅, {2}, {3}, {2,4}, {3,4}, {2,3,4} }
XOR
⊕ {2}
{ {2}, ∅, {2,3}, {4}, {2,3,4}, {3,4} }
⊕ {3,4}
{ {2,3,4}, {3,4}, {2,4}, {3}, {2}, ∅}

algebra of set systems



{ ∅, {q}, {p,q},
{p,r}, {p,q,r} }

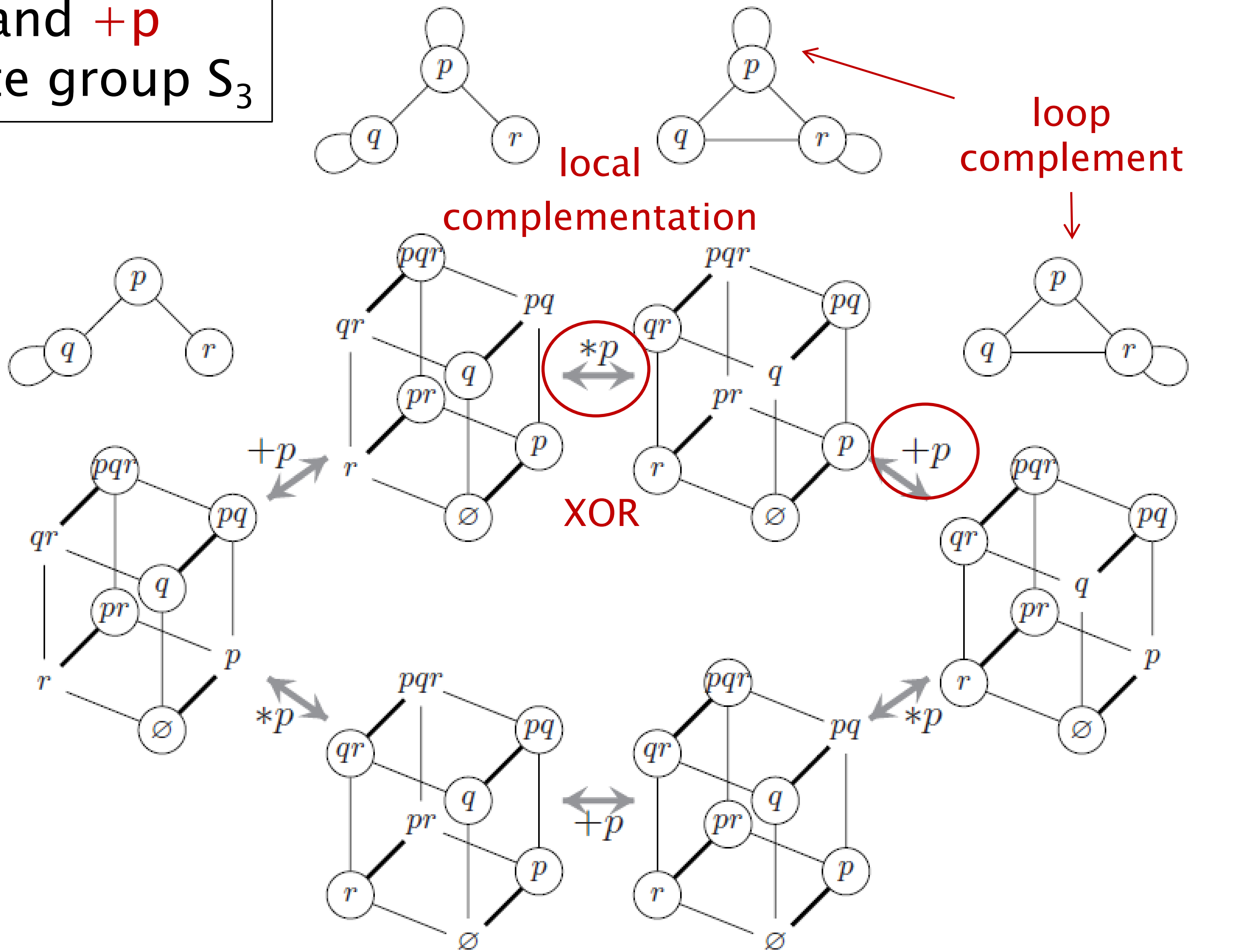
algebra of set systems



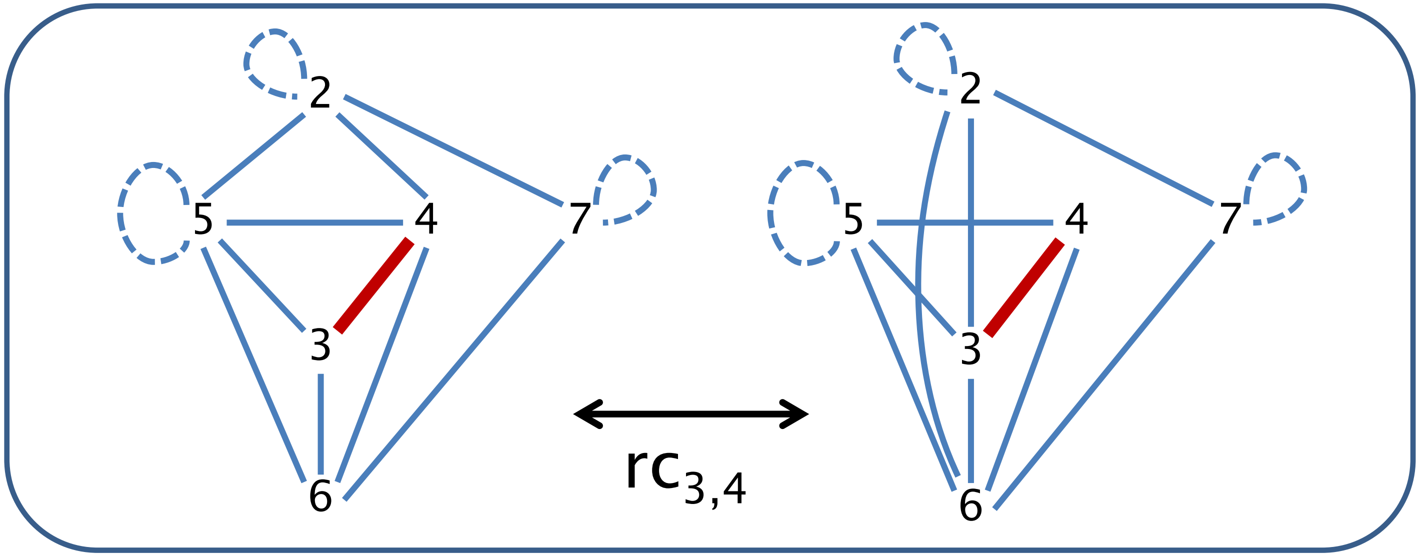
{ ∅, {q}, {p,q},
{p,r}, {p,q,r} }

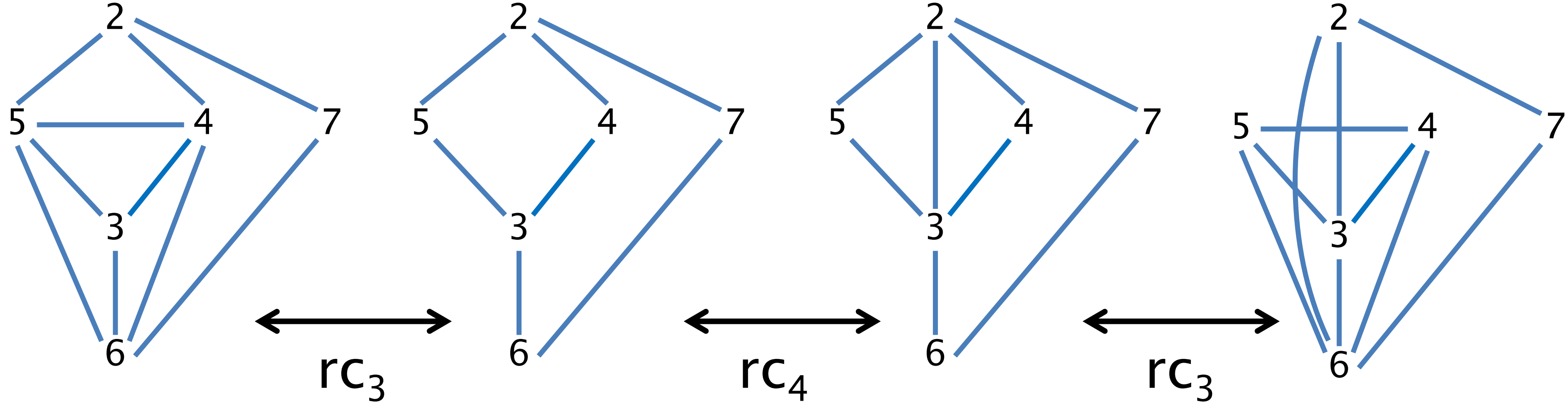
algebra of set systems

*p and +p generate group S_3



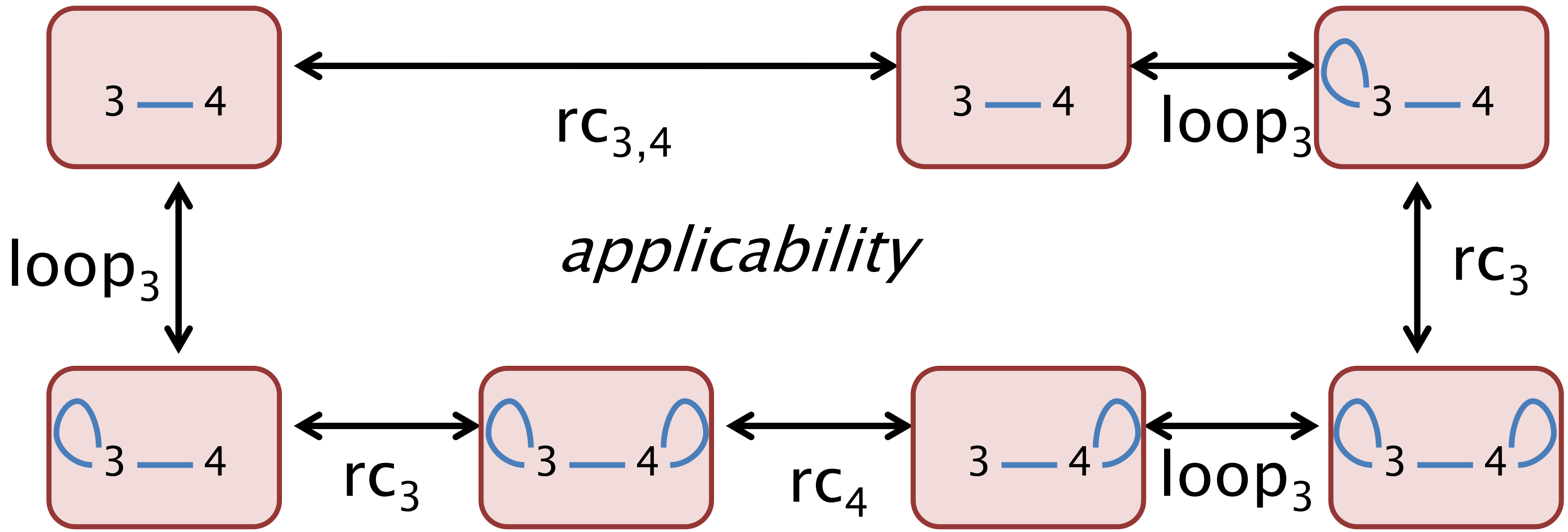
edge complement vs. local complement





ignoring loops

edge complement vs. local complement



basic algebra S_3

*3 *4 = *4 *3

*3 *3 = id = +3 +3

+3 *3 +3 = *3 +3 *3

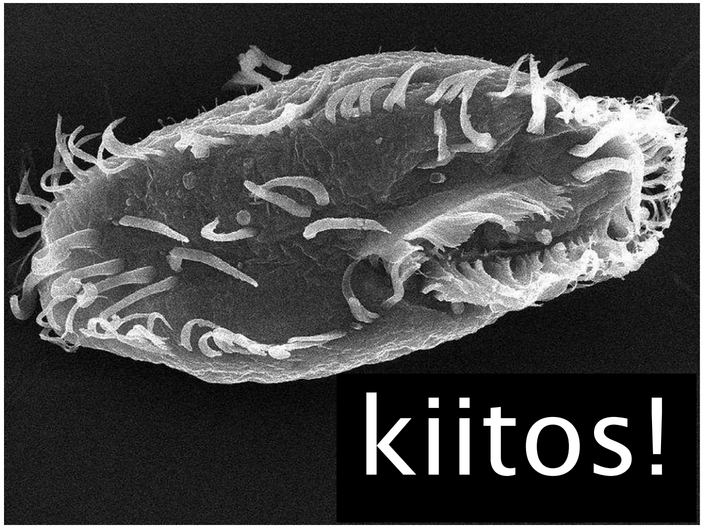
+3 *3 *4 +3 *3 +3 =
+3 *3 +3 *3 +3 *4 =
+3 *3 *3+3 *3 *4 =
+3 +3 *3 *4 =
*3 *4 =
*{3,4}

conclusion (updated)

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt) and XOR
- we can use results about ppt (on matrices) and XOR (on set systems) to know more about gene assembly
- but also inspiration the other way around …

however …

- parallellism
- ‘simple’ operations



references (to self)

R. Brijder, H.J. Hoogeboom. The Group Structure of Pivot and Loop Complementation on Graphs and Set Systems. Eur.J.Comb. (2011).

R. Brijder, H.J. Hoogeboom. Maximal Pivots on Graphs with an Application to Gene Assembly. Discr.Appl.Math. 158 (2010) 1977–1985.

R. Brijder, H.J. Hoogeboom. Reality-and-Desire in Ciliates.
In: *Algorithmic Bioprocesses* (Condon etal, eds.), Natural Computing Series, Springer (2009) pp.99–115.

R. Brijder, T. Harju, H.J. Hoogeboom, Pivots, determinants, and perfect matchings of graphs (2008)
submitted for publication – *really a long time ago now* [arXiv:0811.3500]

A. Ehrenfeucht, T. Harju, I. Petre, D. Prescott, G. Rozenberg, *Computation in Living Cells: Gene Assembly in Ciliates*, Natural Computing Series, Springer (2004)

(*this one you know, of course*)