GENE ASSEMBLY IN CILIATES - A splendid example of NATURAL COMPUTING

G. ROZENBERG

- University of Leiden, The Netherlands
- University of Colorado at Boulder, USA

NATURAL COMPUTING

1) Computing Going On In Nature

or

2) Human-Designed Computing Inspired By (Gleaned From) Nature

……

2:

* Evolutionary Computing
* Neural Computing
* DNA Computing
* Quantum Computing

Volume 1 No. 1 2002 ISSN 1567-7818

Natural Computing

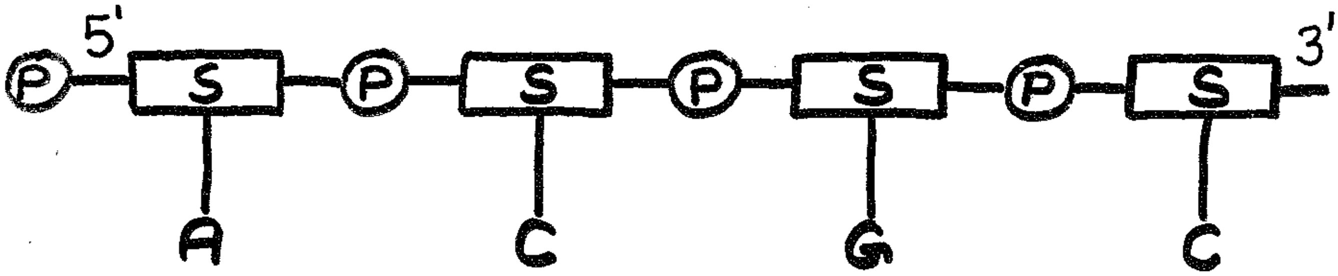
An International Journal

Kluwer Academic Publishers

DNA MOLECULES

Basic Structure

Single Stranded DNA



5'
3'
P
S
A
C
G
C

backbone

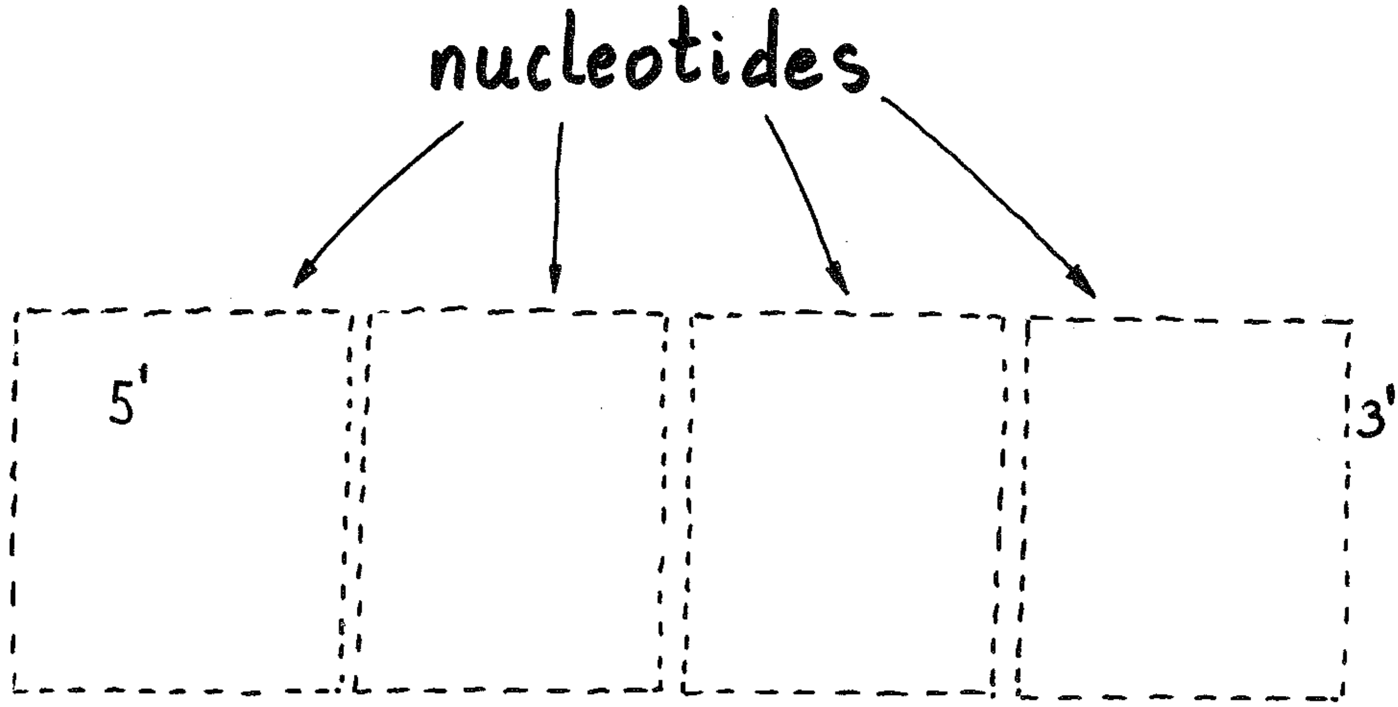
bases

4 possible bases: A, C, G, T

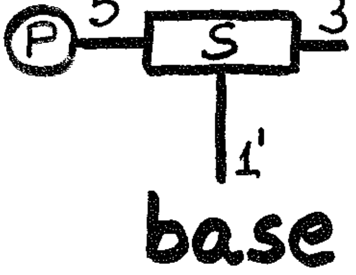
A ↔ T   C ↔ G

complementary pairs

Watson-Crick complementarity

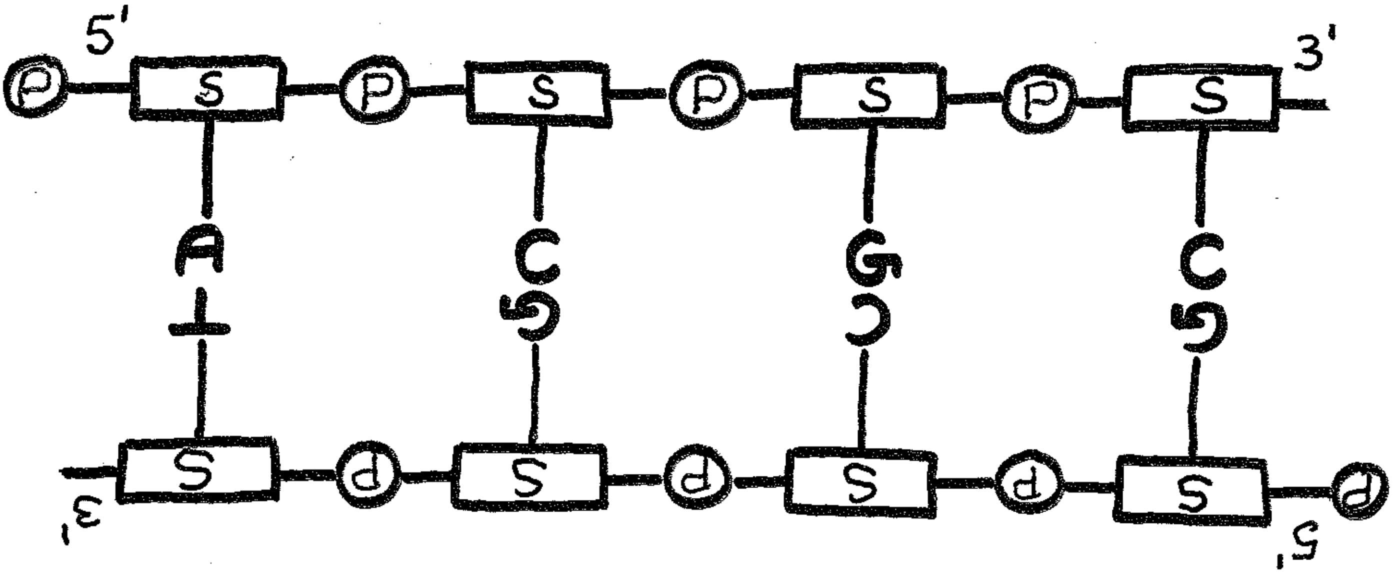


nucleotide:



P	Phosphate
S	Sugar

Double Stranded DNA

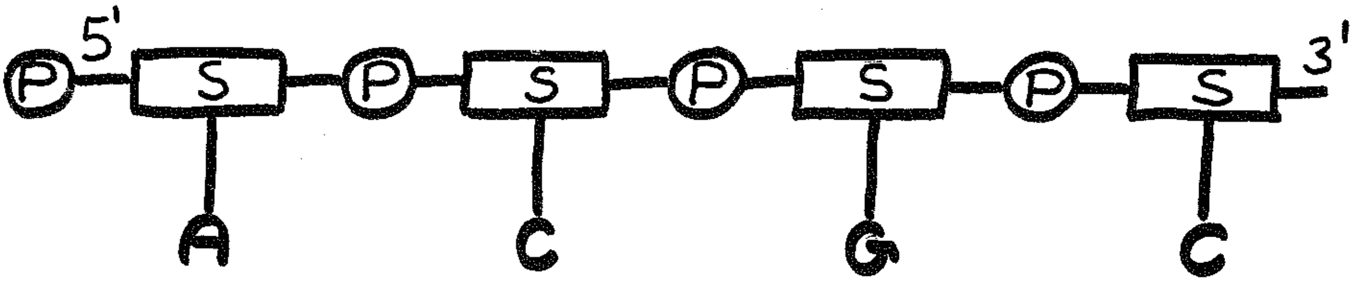


vertically standing "picket fence":

- pickets ~ paired bases
- top and bottom bars ~ backbones

NOTATION

Single Stranded DNA



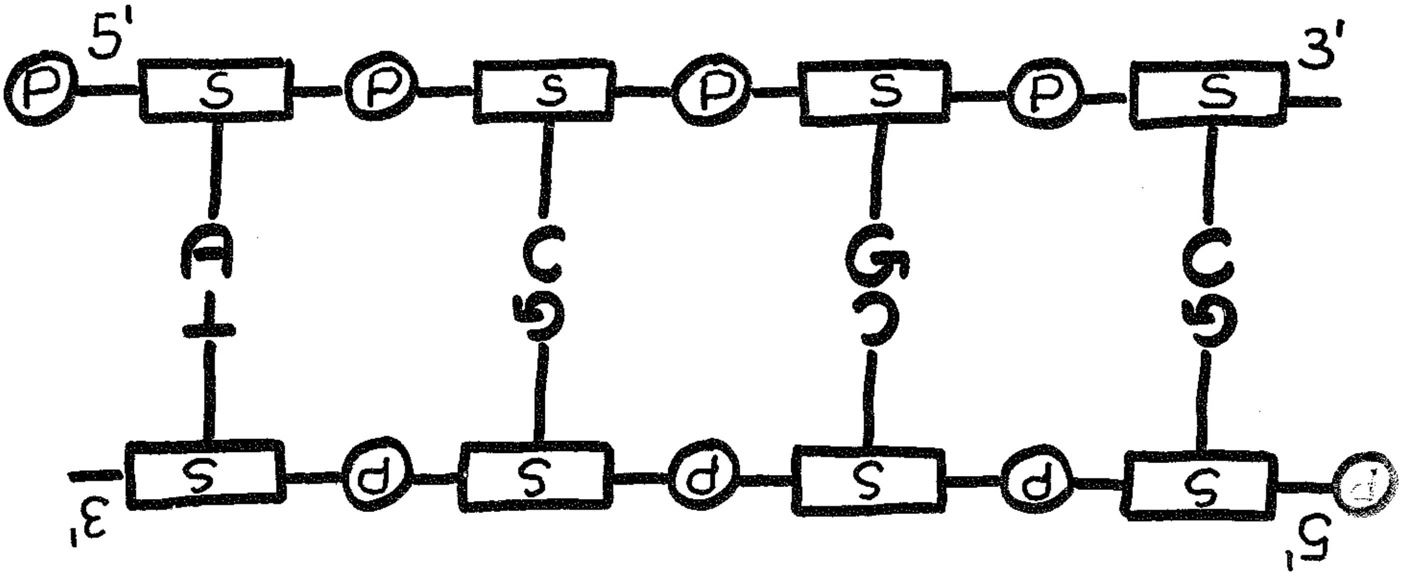
- 5'- A C G C -3'
- $\overrightarrow{\text{A C G C}}$
- A C G C

left → right

corresponds to

5' — 3'

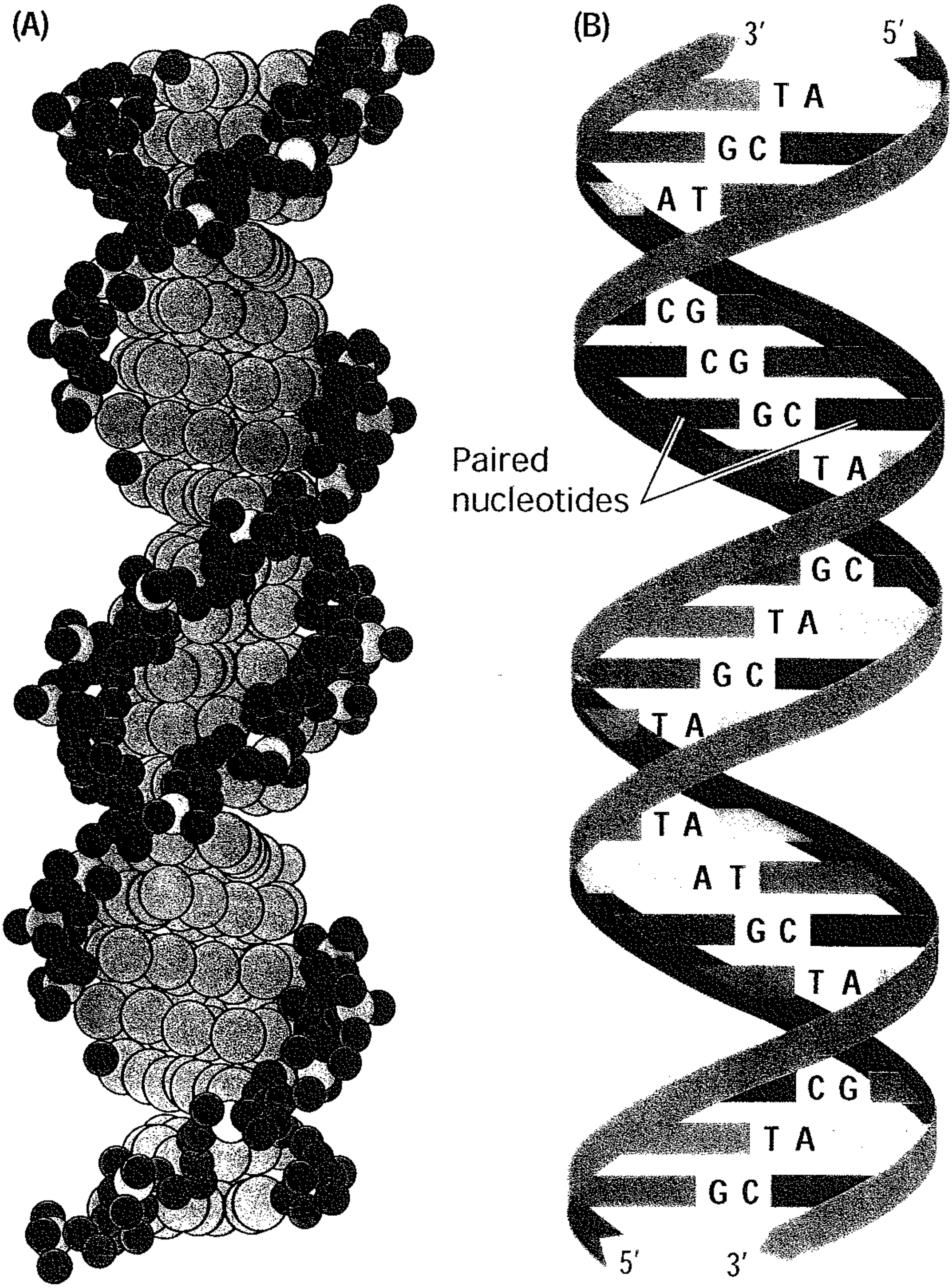
Double Stranded DNA



- 5'-A C G C
  T G C G -5'
- $\overrightarrow{\text{A C G C}}$
  T G C G
- A C G C
  T G C G

left → right
in the upper string
corresponds to
5'-3'

10 8

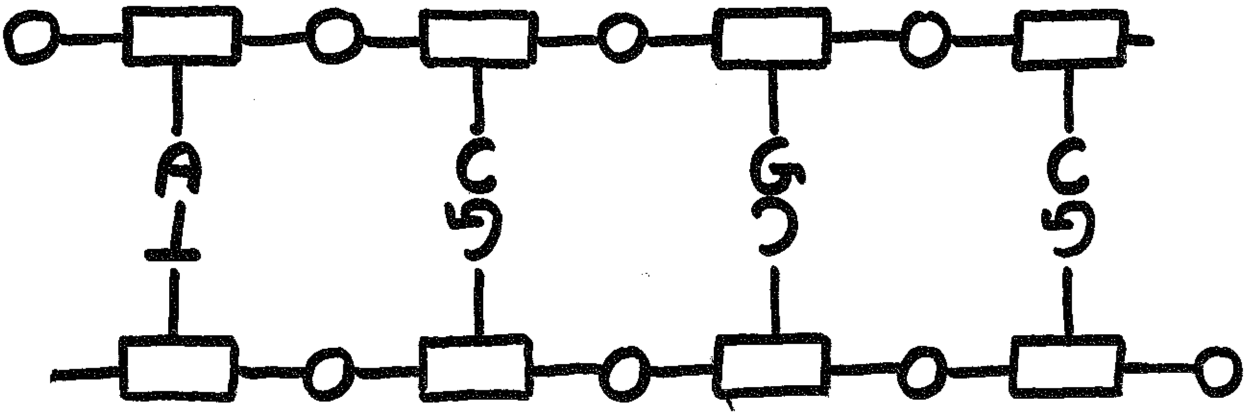


CH 1, FIG 5
Hartl, Genetics

JNVERSION

double stranded

- intuition - another notation for the same molecule

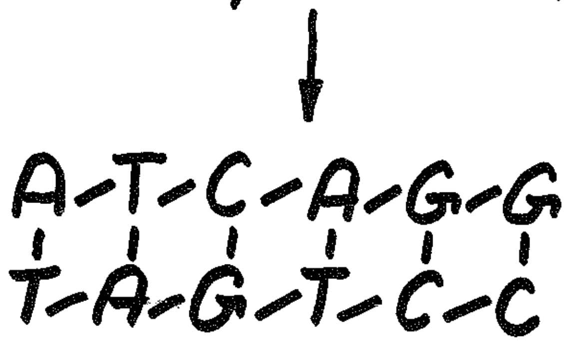


ACGC $\xrightarrow{inv}$ GCGT
TGCG CGCA

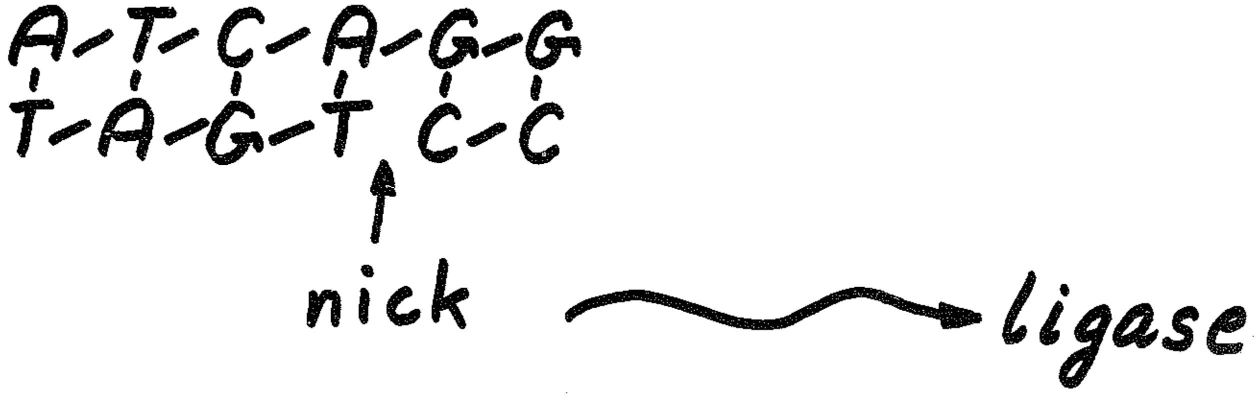
180° point rotation

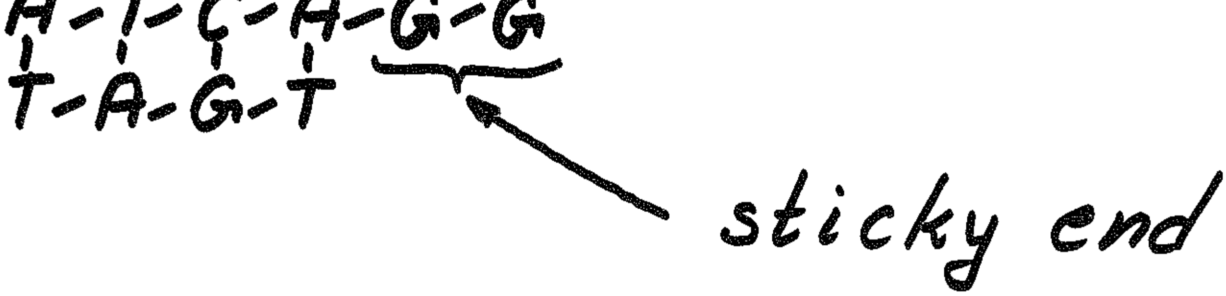

A C G C
G C G T

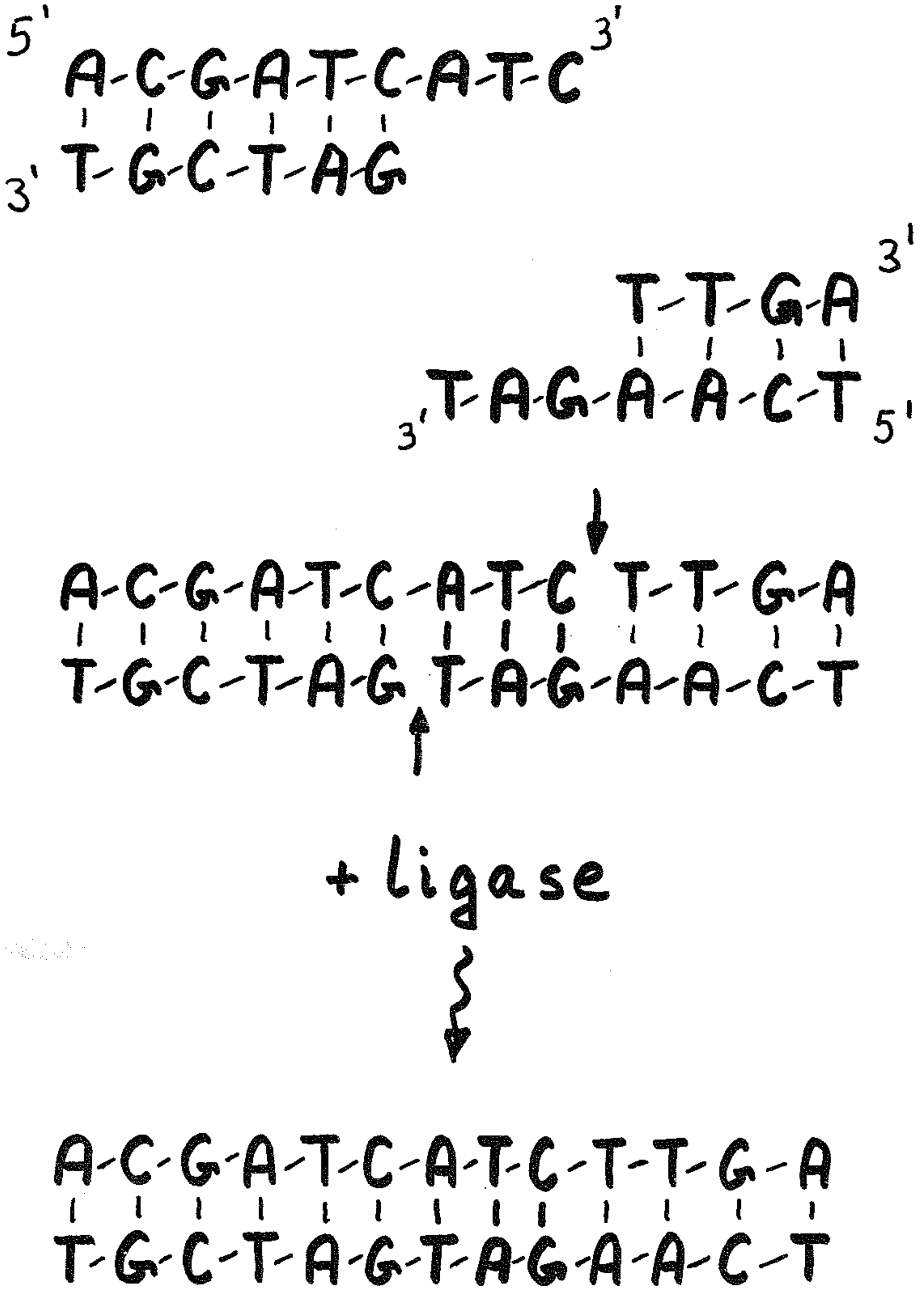
ATCAGG
TAGTCC ← perfect duplex



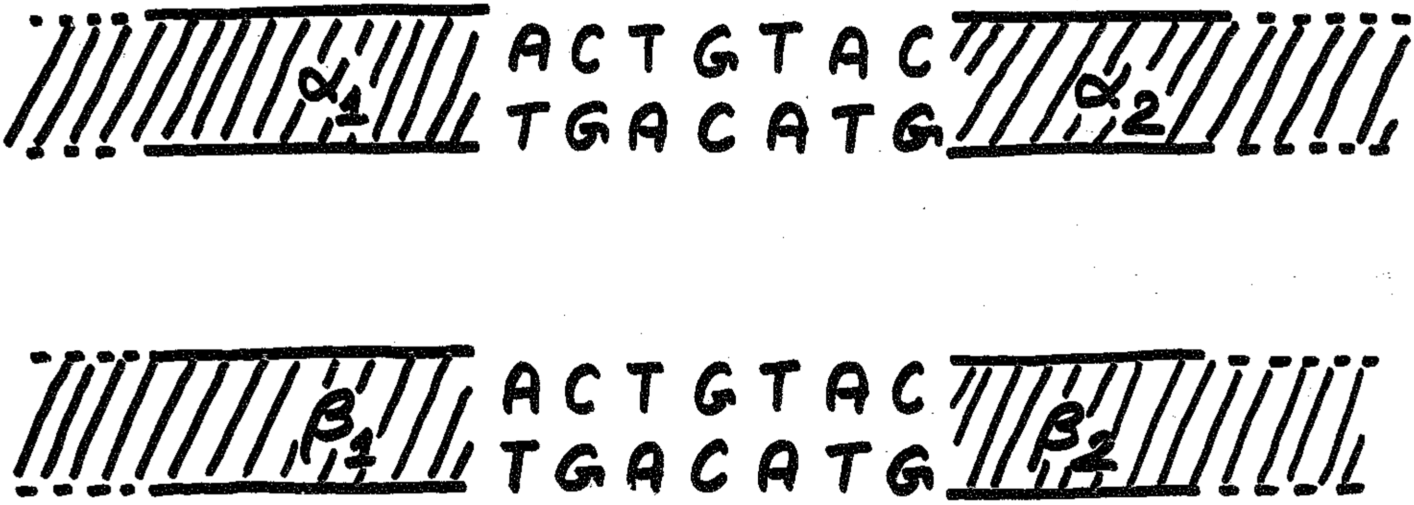
All kinds of imperfections:

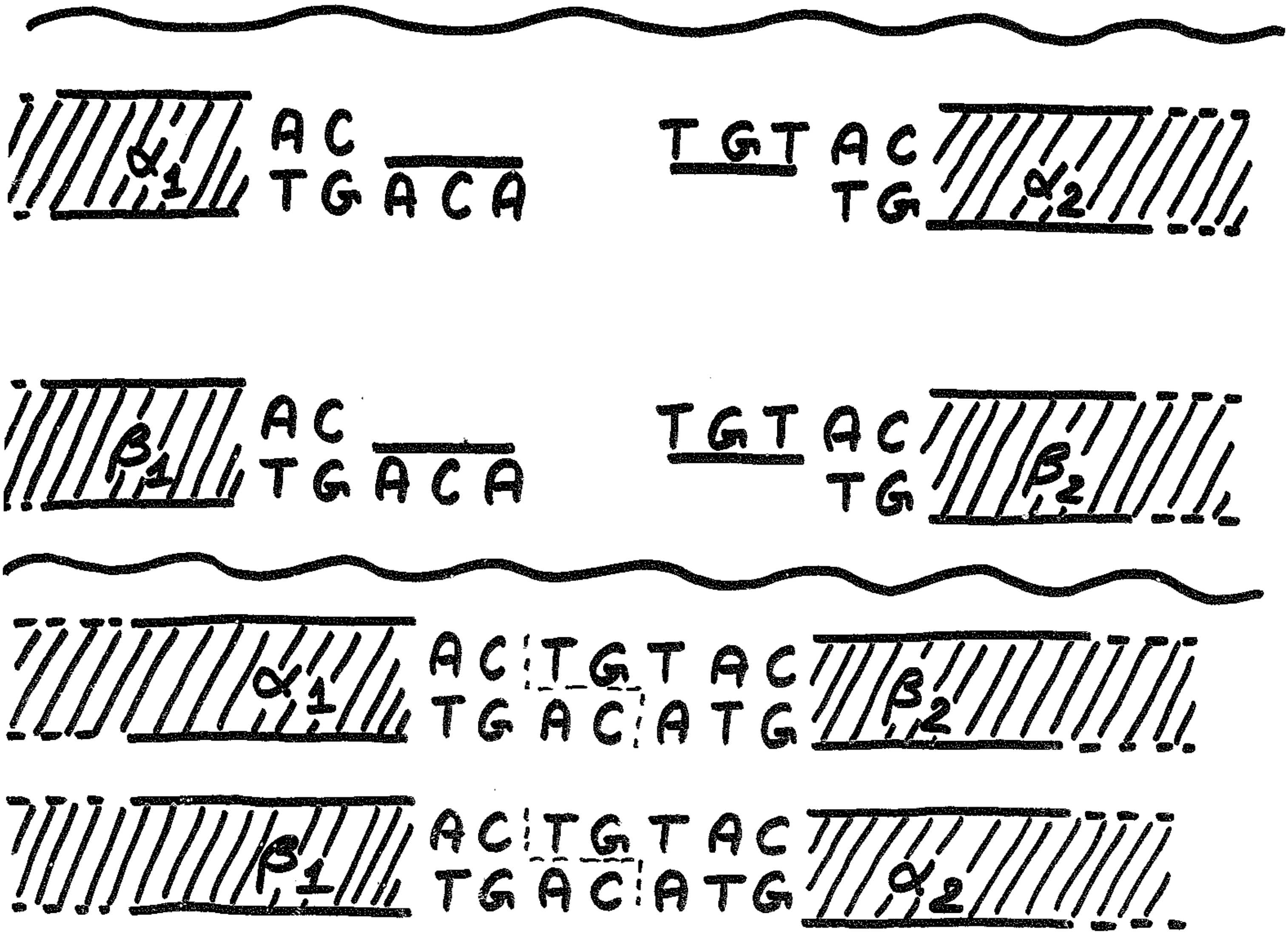





5'
A-C-G-A-T-C-A-T-C
3'
T-G-C-T-A-G
3'
T-T-G-A
3'
3'
T-A-G-A-A-C-T
5'
A-C-G-A-T-C-A-T-C T-T-G-A
T-G-C-T-A-G T-A-G-A-A-C-T
+ ligase
A-C-G-A-T-C-A-T-C-T-T-G-A
T-G-C-T-A-G-T-A-G-A-A-C-T

Homologous Recombination




α1
AC
TGACA
TGTAC
TG
α2
β1
AC
TGACA
TGTAC
TG
β2
α1
ACTGTAC
TGACATG
β2
β1
ACTGTAC
TGACATG
α2

Ciliates

single cell eukaryotes

very ancient (~ 2×10^9 years old)

very rich group (~ 10000 genetically different organisms)

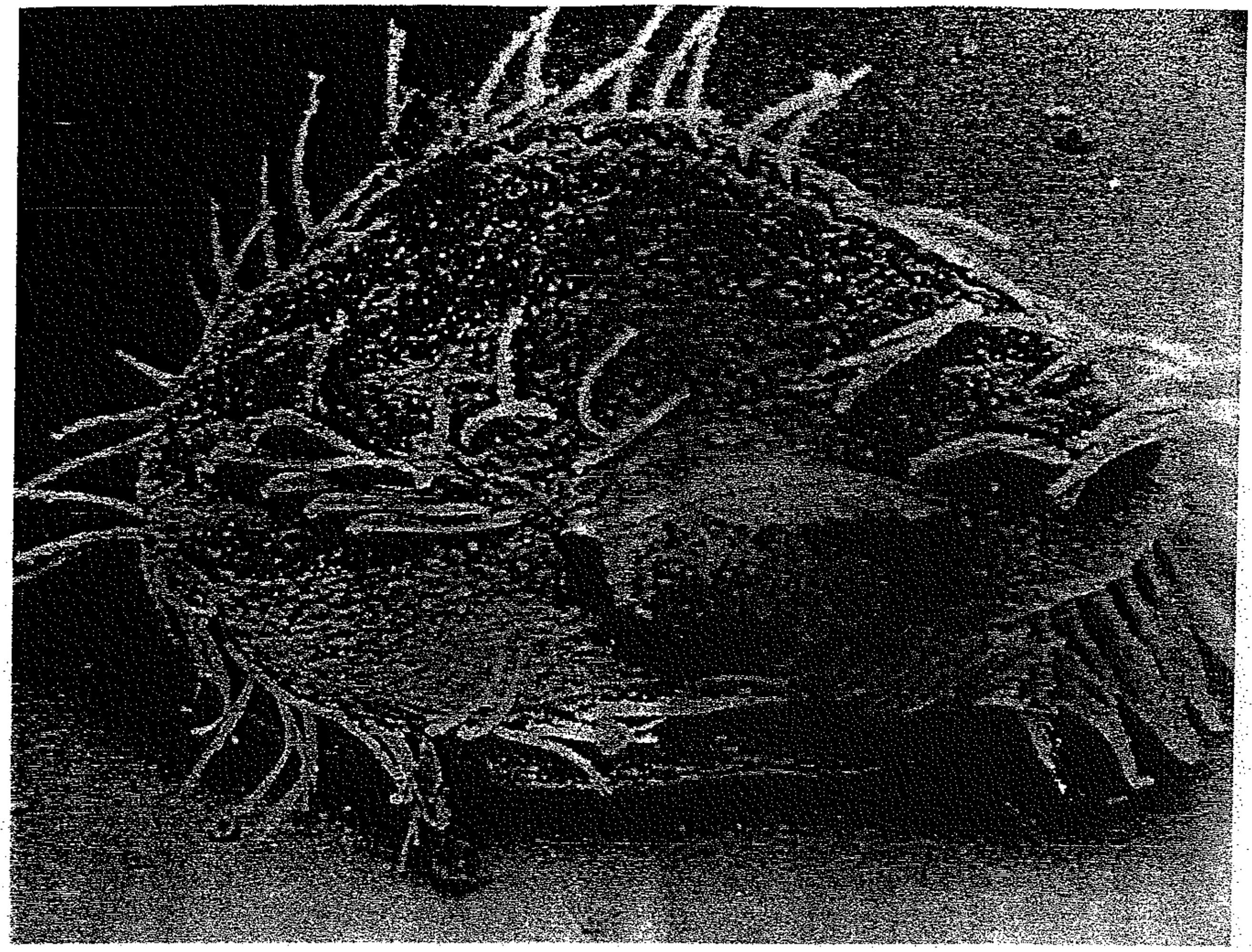
very important from the evolutionary point of view

Stichotrichs

A. Ehrenfeucht , Boulder, CO
T. Harju , Turku, Finland
I. Petre , Turku , Finland
D. M. Prescott , Boulder, CO
G. Rozenberg , Leiden + Boulde

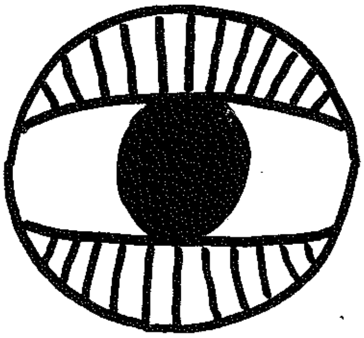
20 16

49



Ciliate

2 146



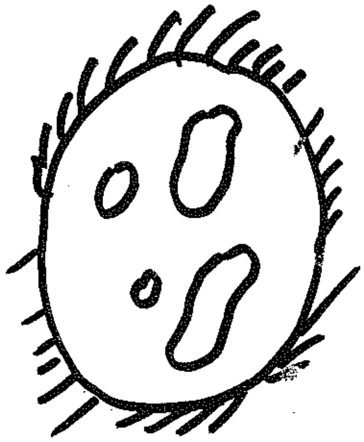
G.R. as seen by ciliates through the other end of a microscope

Unique feature of ciliates:

NUCLEAR DUALISM

They have <u>two</u> <u>functionally</u> <u>different</u> nuclei:

- **micronucleus** - dormant, activated only during sexual reproduction
- **macronucleus** - the "working household nucleus producing RNA transcripts



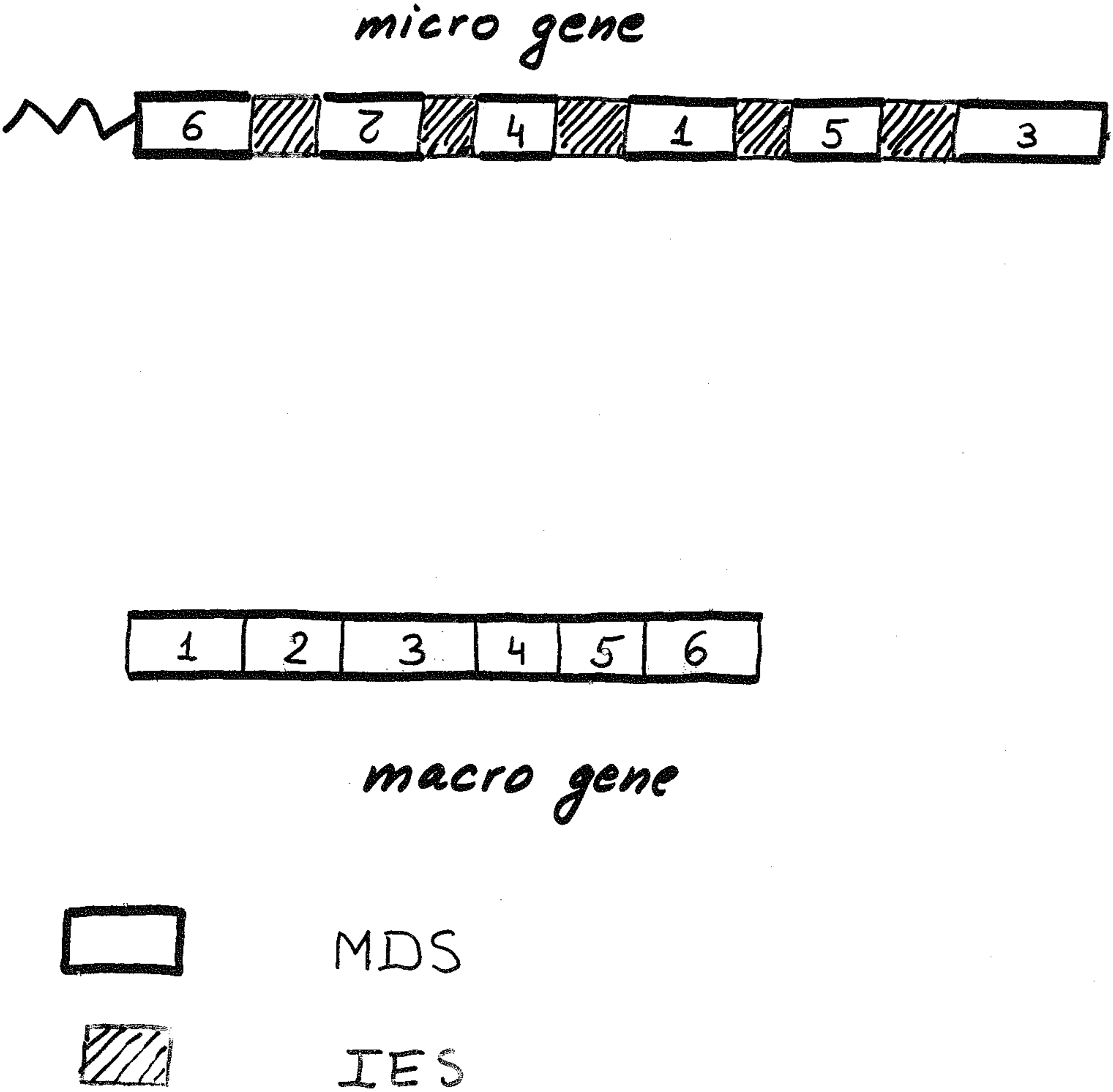
When you starve ciliates they

- either form a cyst
- or they eat each other
- or they go to sexual reproduction

At some stage of sexual reproductio a micronucleus gets transformed into macronucleus - this process is called **gene assembly.**

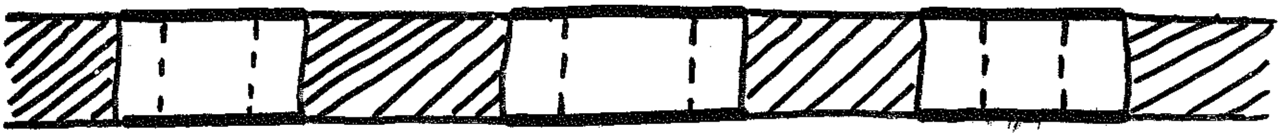
This is the most involved DNA processing known in living organisms, because .... **the micronuclear genom is dramatically different from the macronuclear genome!!**

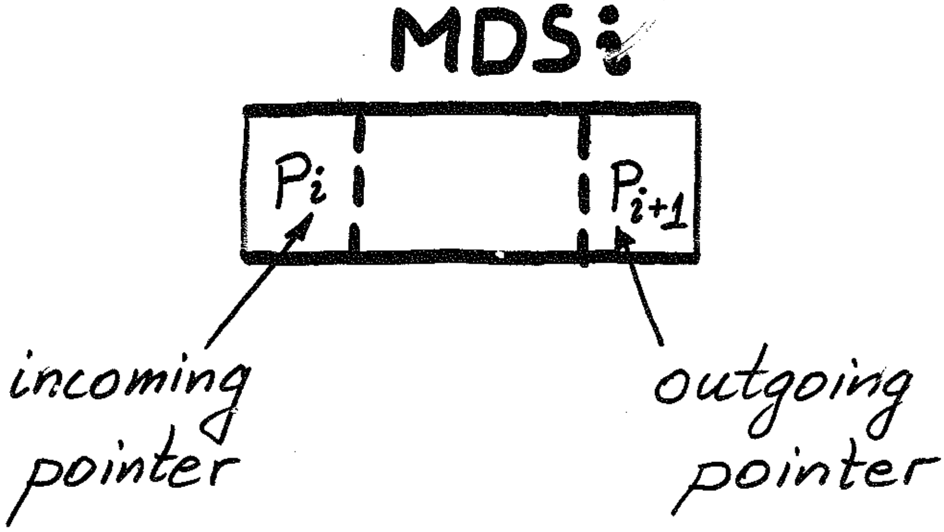
LOCAL difference:



During gene assembly IESs must be excised and **MDSs** must be spliced in the macronuclear (ortodox) order.

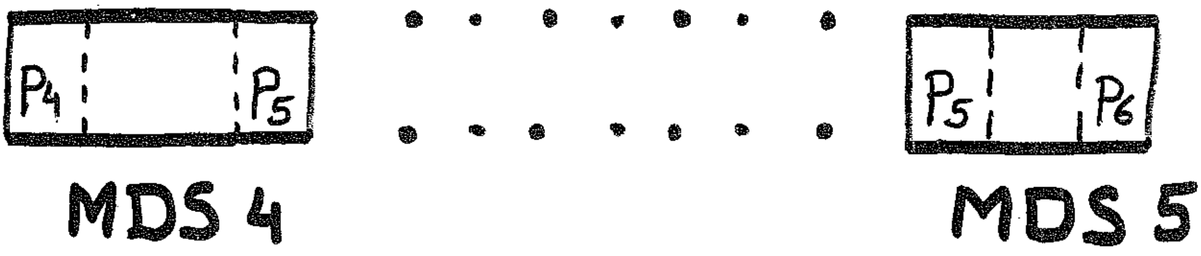
The structure of MDSs is of crucial importance here

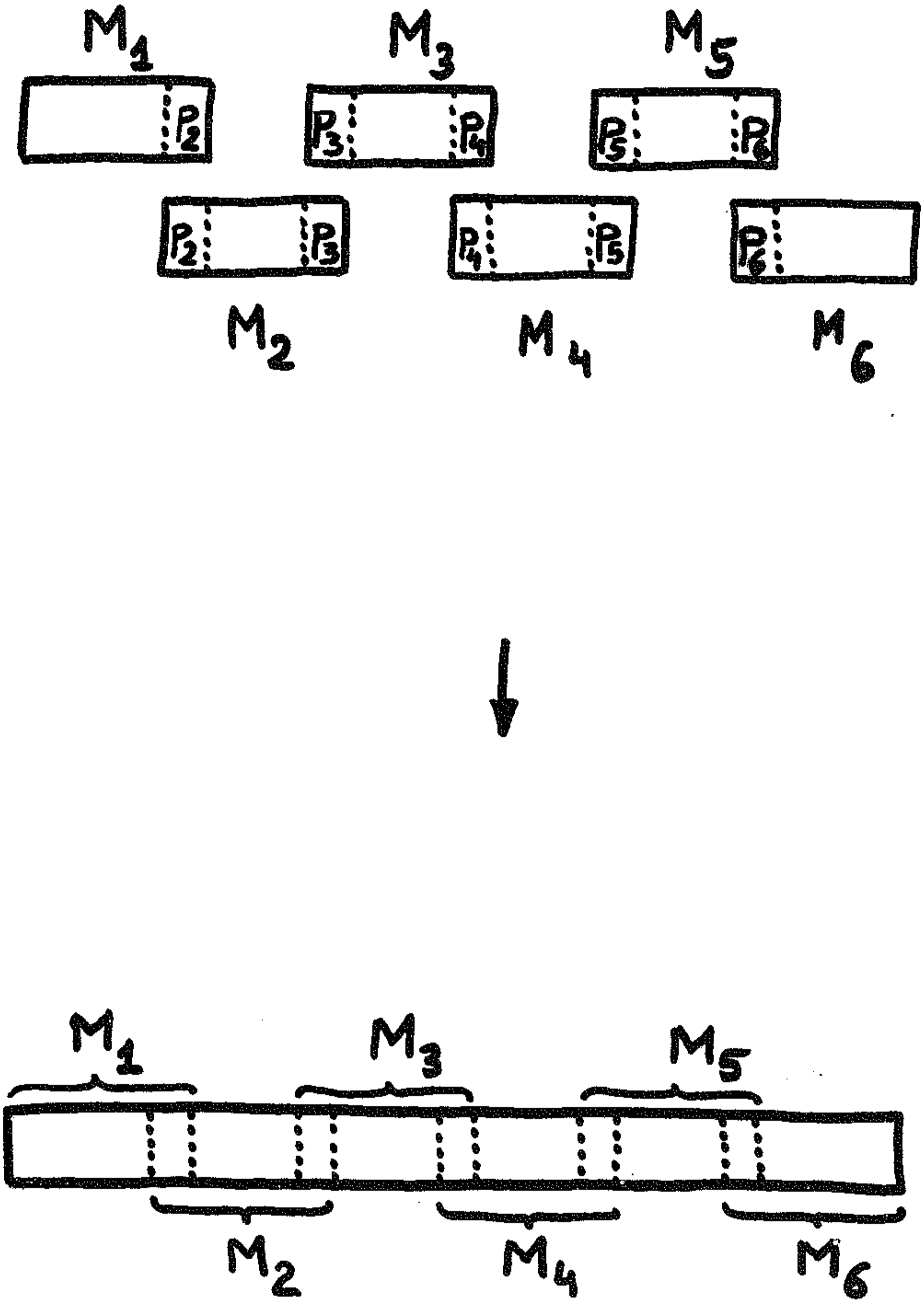




(P_i, P_{i+1}) unique for MDS i

outgoing of MDS i
= incoming of MDS i+1




M1
M3
M5
P2
P3
P4
P5
P6
P2
P3
P4
P5
P6
M2
M4
M6
M1
M3
M5
M2
M4
M6

IMPORTANT

Biological Questions:

- How are pointers aligned?
- How are pointers recognized?
- What is exactly the role of pointers?

Pointers may be very short!!

Pointers alone cannot guide DNA folding and alignment of true pointers

Solution

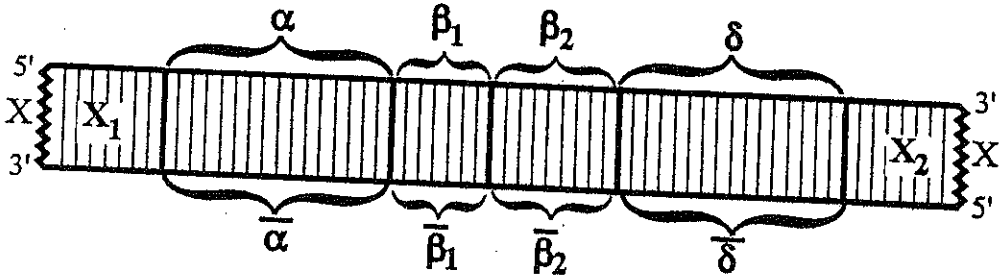
<u>Template</u>-<u>guided</u> <u>recombination</u> in which a DNA molecule from the old macronucleus enters the newly developing macronucleus and acts as a template to guide pointer alignment

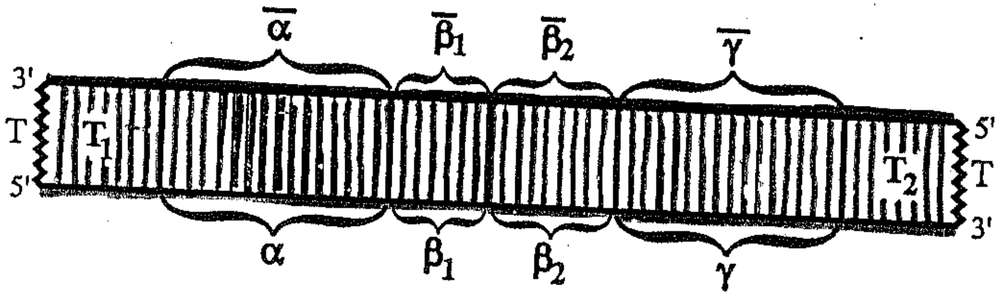

MIC
MAC

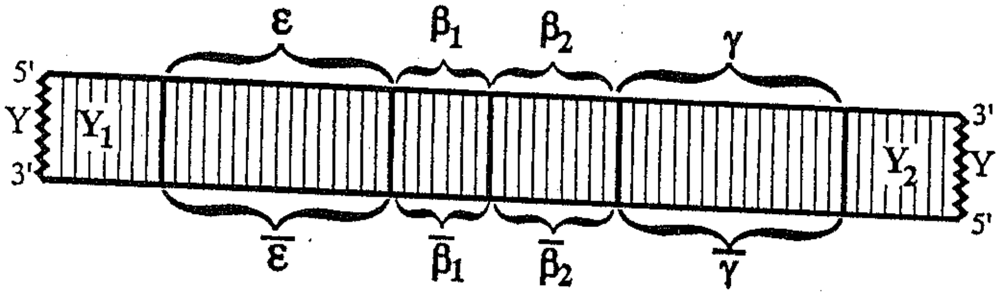
Geometry
of
Template-Guided Recombination

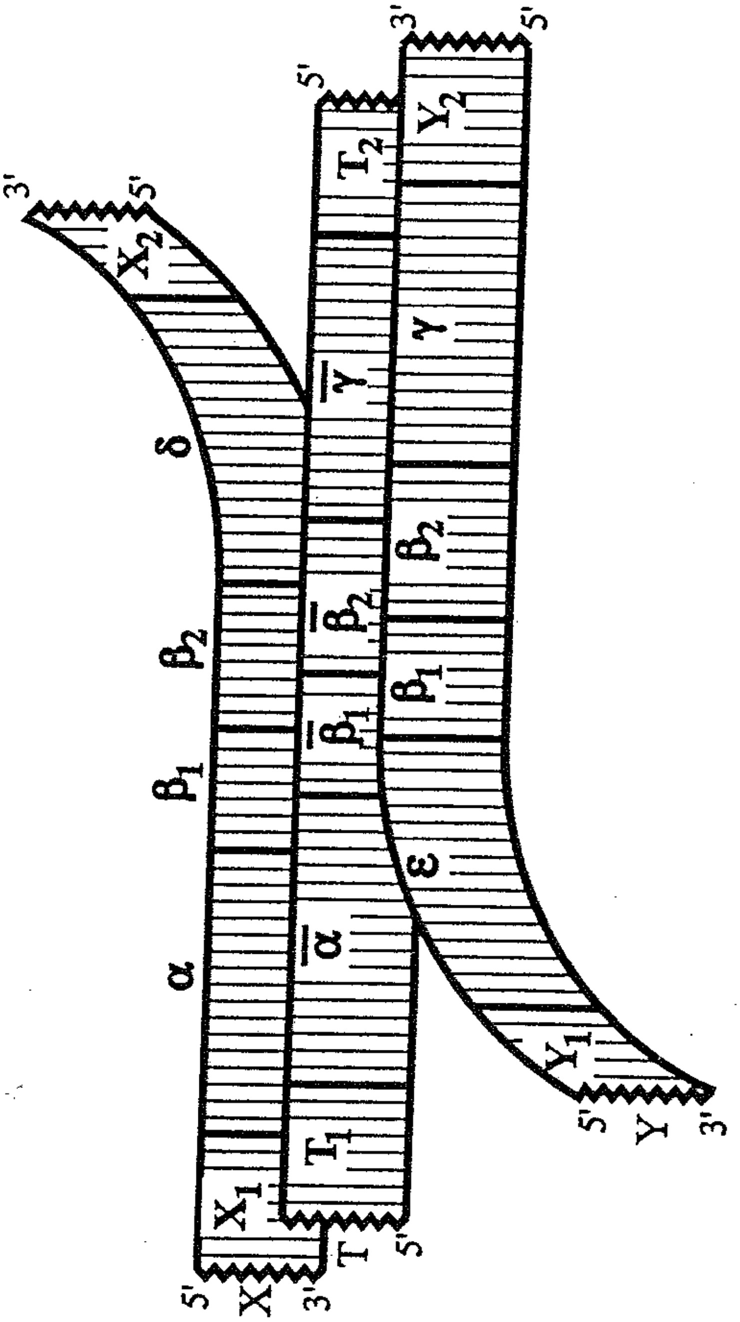
complementary: $\alpha \sim \overline{\alpha}$, $\beta_1 \sim \overline{\beta_1}$, $\beta_2 \sim \overline{\beta_2}$

not complementary: $\delta - \overline{\gamma}$ $\varepsilon - \overline{\alpha}$









6.

Fig. 4

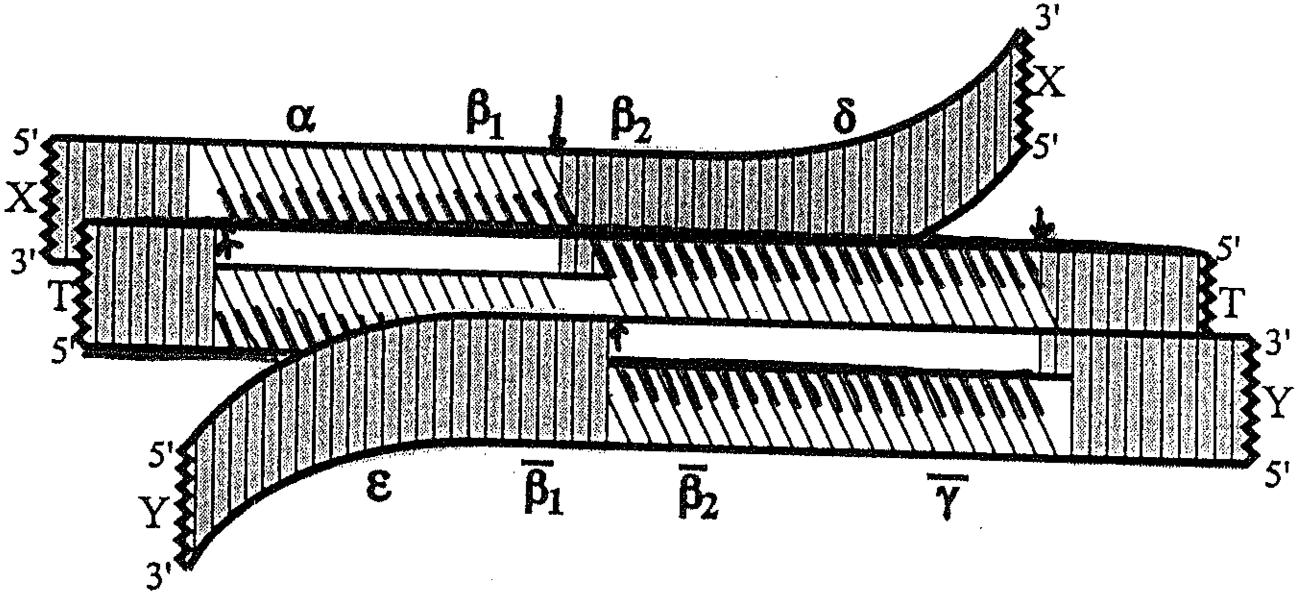
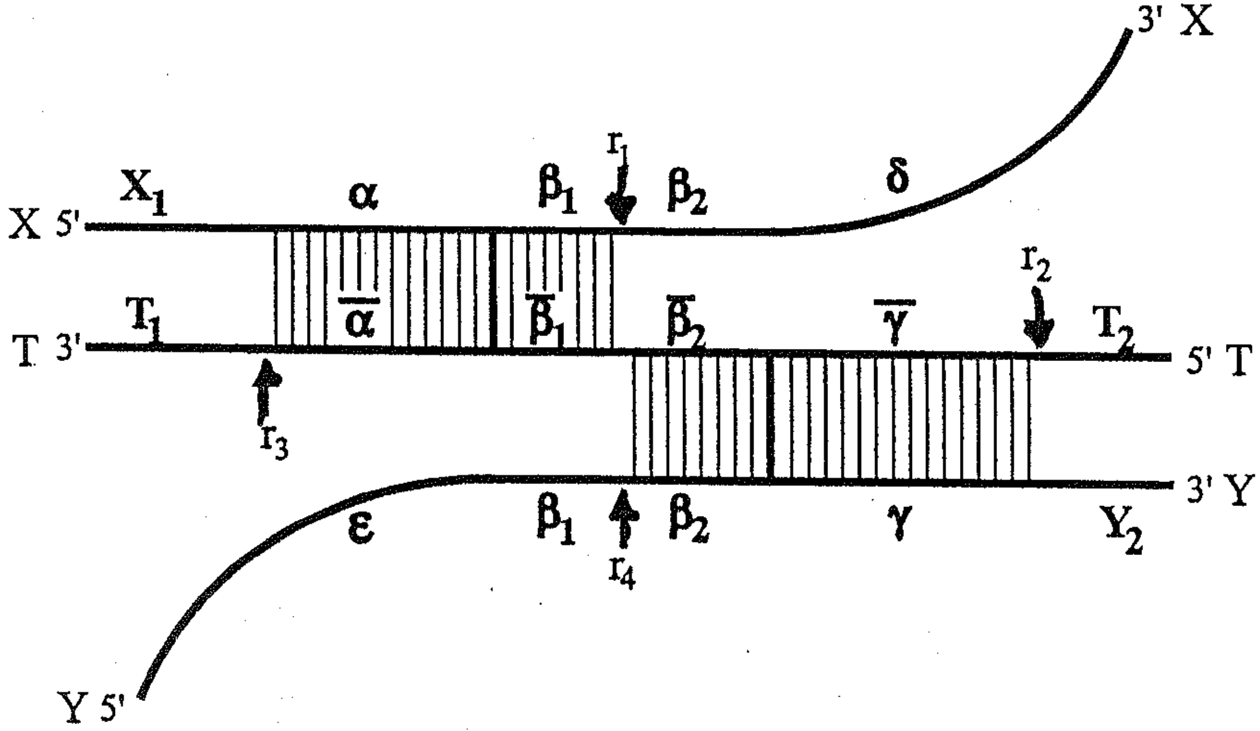

α
β_1
β_2
δ
3'
X
5'
5'
X
3'
T
5'
5'
T
3'
Y
5'
5'
Y
3'
ε
$\overline{\beta}_1$
$\overline{\beta}_2$
$\overline{\gamma}$

Fig. 7.

10.



Fi

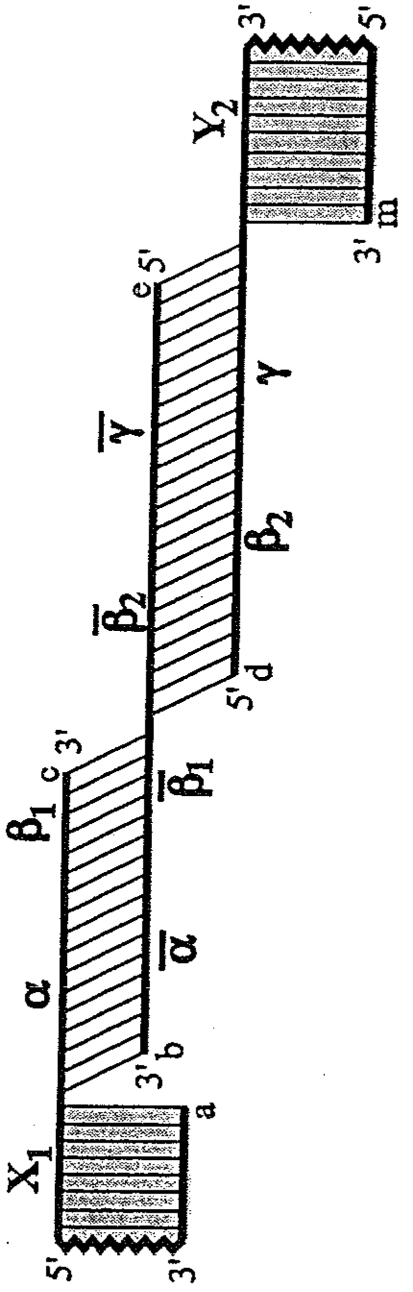

5'
3'
X_1
a
3'
b
α
β_1
c
3'
$\bar{\alpha}$
$\bar{\beta}_1$
5'
d
$\bar{\beta}_2$
β_2
$\bar{\gamma}$
γ
e
5'
3'
m
Y_2
3'
5'

Fig.

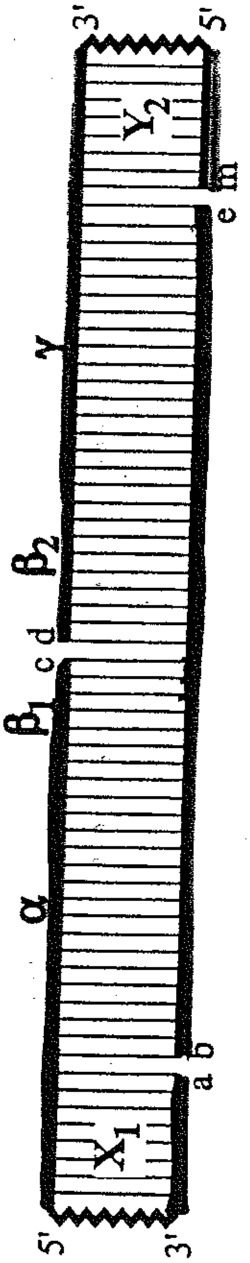
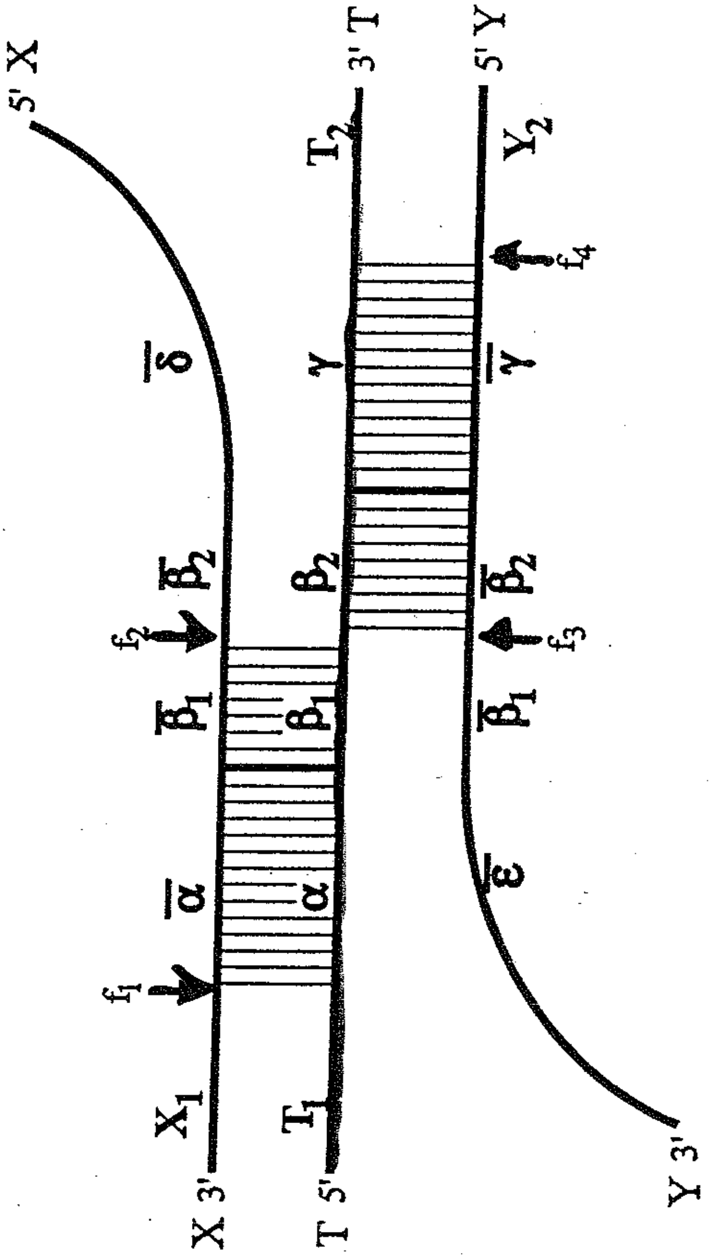


Fig. 1


5' X
3' T
5' Y
T_2
Y_2
f_4
$\bar{\delta}$
γ
$\bar{\gamma}$
$\bar{\beta}_2$
β_2
$\bar{\beta}_2$
f_2
f_3
$\bar{\beta}_1$
β_1
$\bar{\beta}_1$
$\bar{\varepsilon}$
$\bar{\alpha}$
α
f_1
X_1
T_1
X 3'
T 5'
Y 3'

13.

Fig.

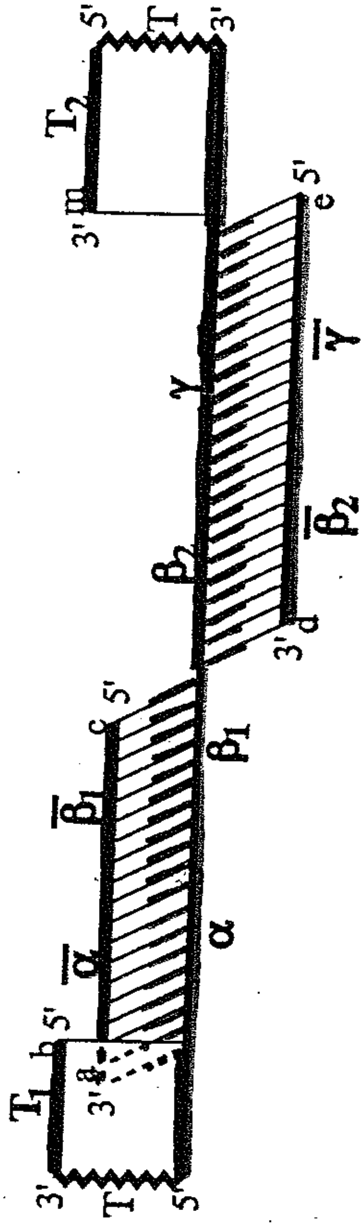

5'
T
3'
T_2
3' m
5'
e
γ
$\bar{\gamma}$
β_2
$\bar{\beta}_2$
3' d
5'
c
β_1
$\bar{\beta}_1$
α
$\bar{\alpha}$
b 5'
a
3'
T_1
3'
T
5'

Fig. 1

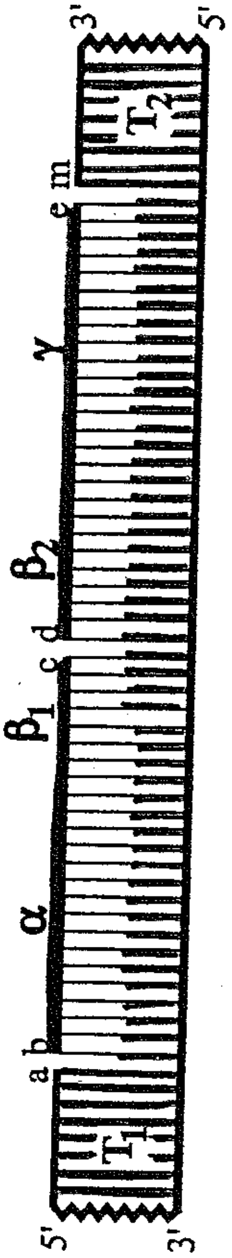


Fig.

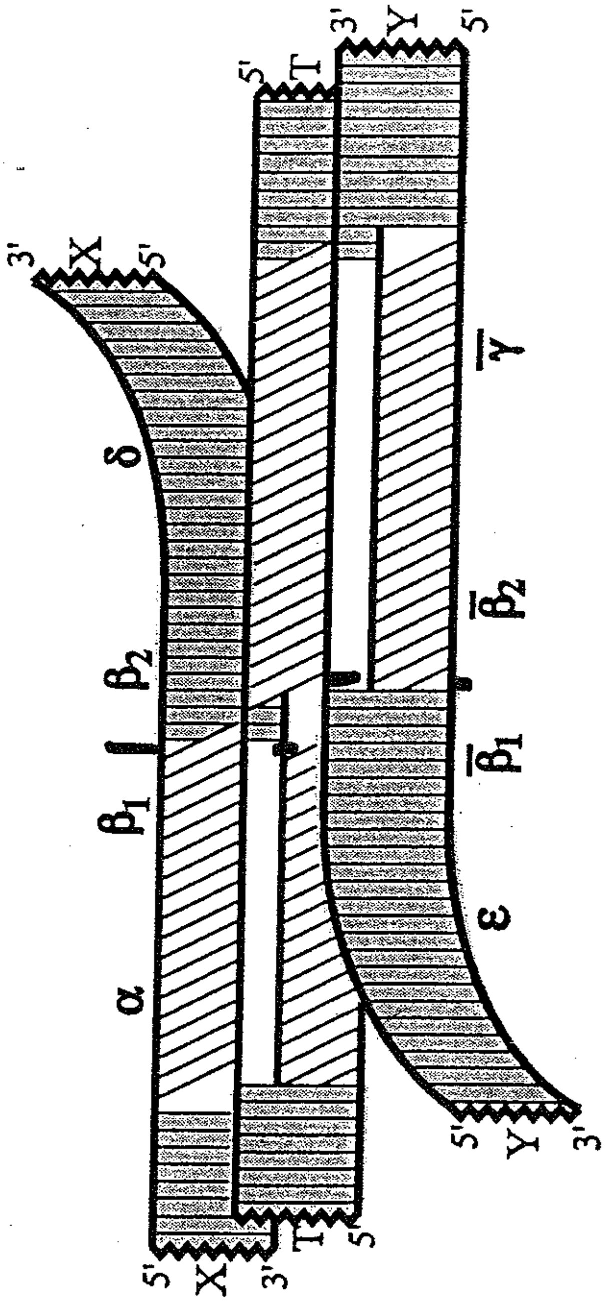
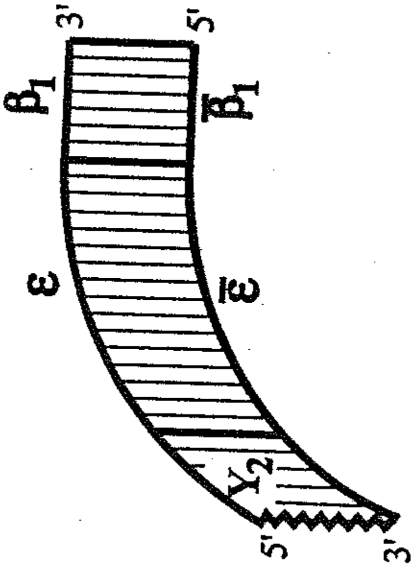


Fig. 16.



17.

Fig.

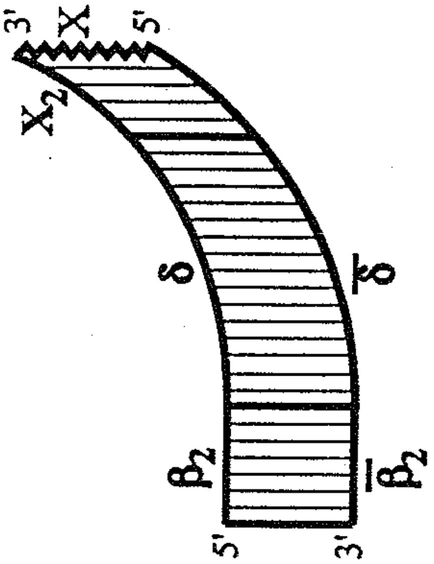
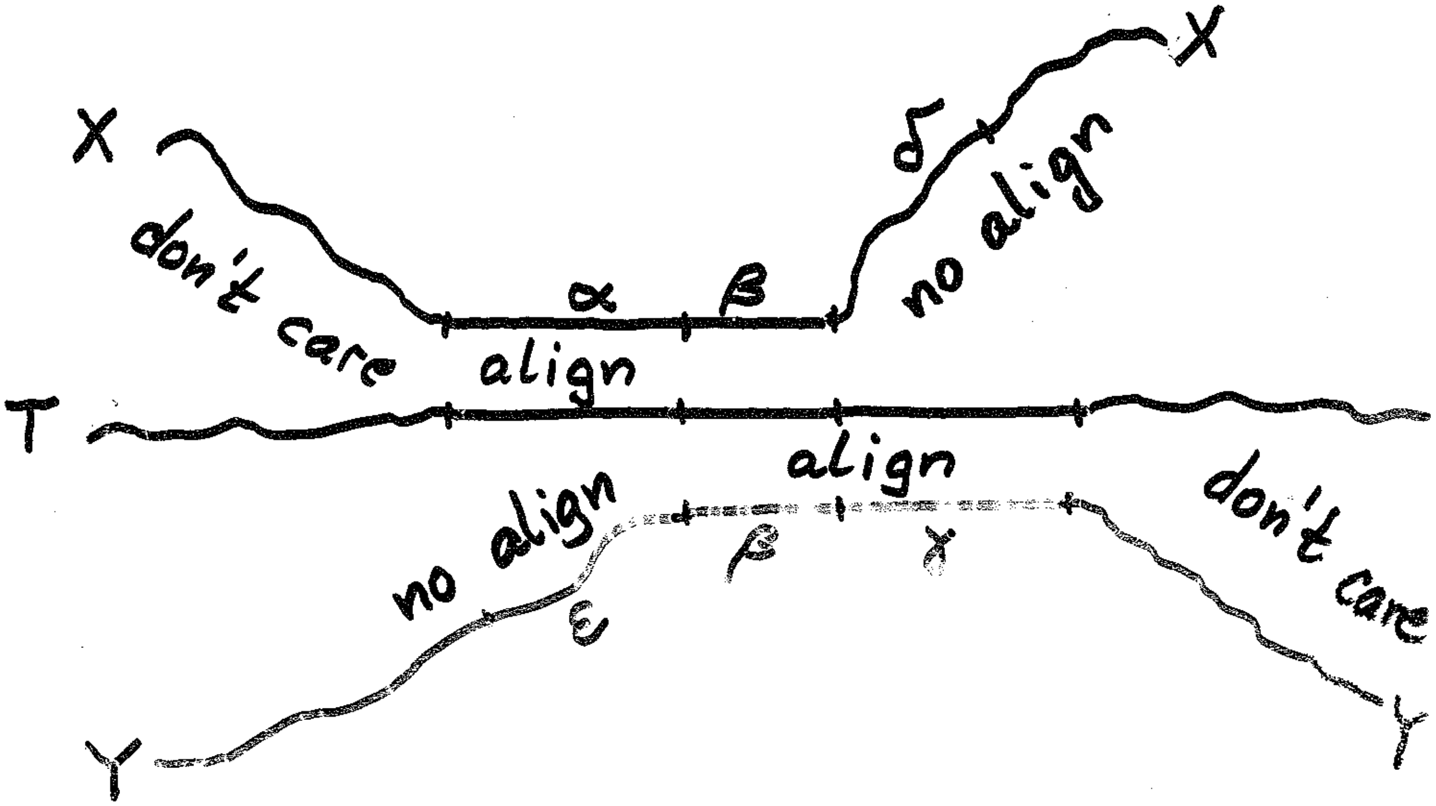
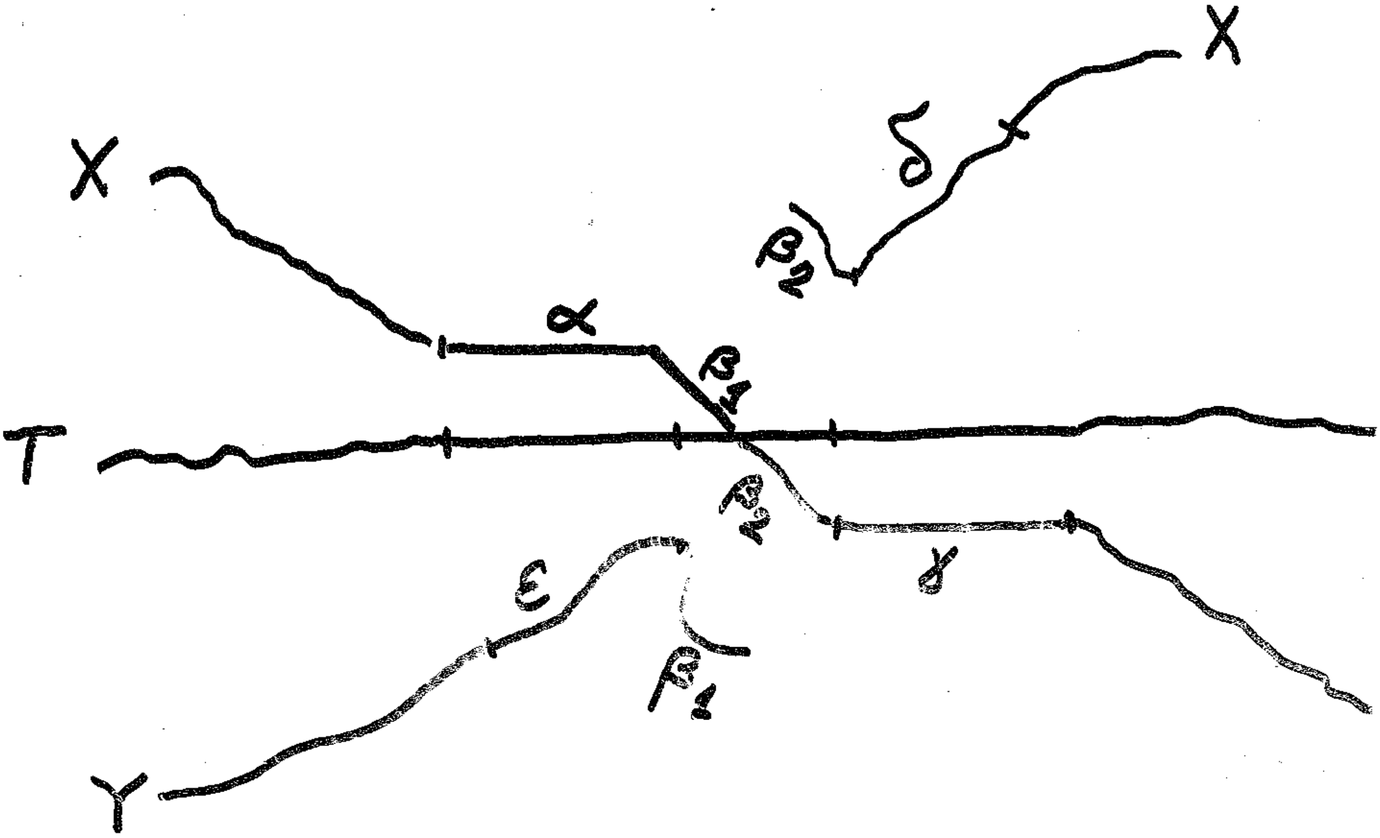


Fig. 9


X
X
don't care
α
β
δ
no align
align
T
align
no align
β
γ
ε
don't care
Y
Y

9.


X
X
δ
β2
α
β1
T
β2
γ
ε
β1
Y

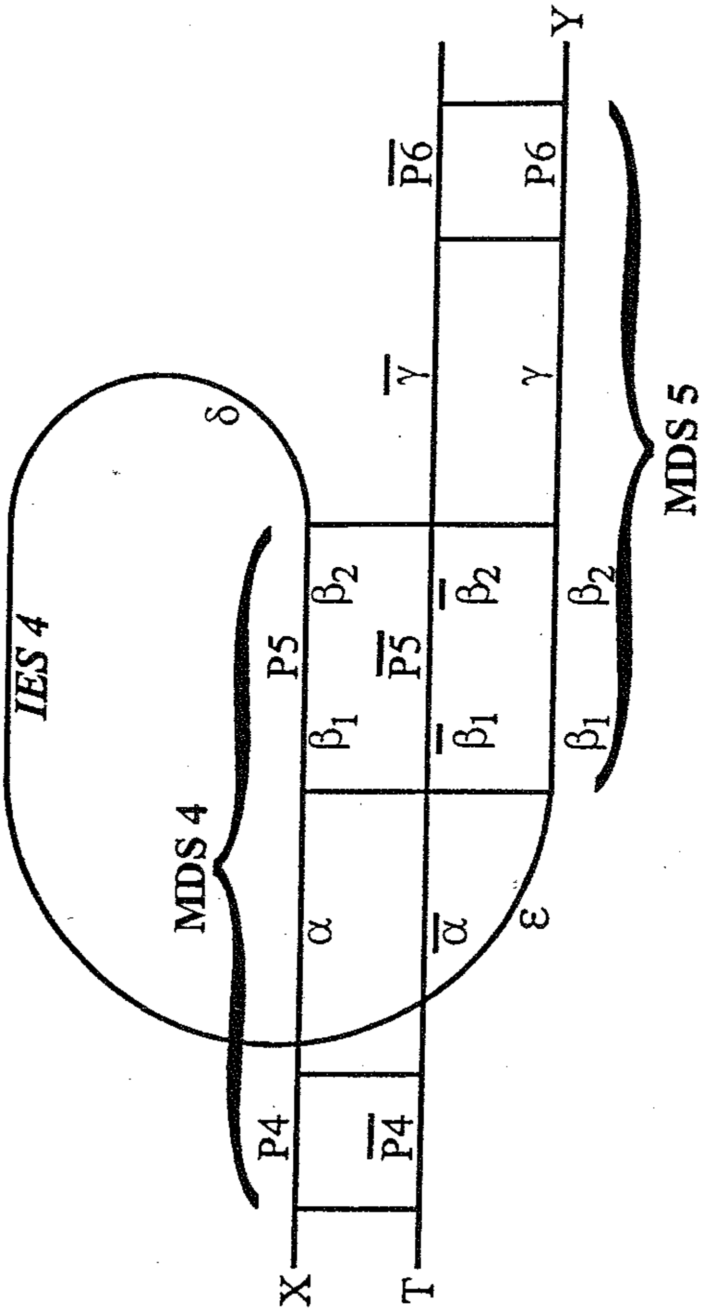

X
T
Y
P4
$\overline{P4}$
α
$\overline{\alpha}$
MDS 4
IES 4
β_1
P5
β_2
$\overline{\beta_1}$
$\overline{P5}$
$\overline{\beta_2}$
β_1
β_2
δ
ε
$\overline{\gamma}$
γ
$\overline{P6}$
P6
MDS 5

Fig. 14

Symbolic Notation for TGR

$$\overbrace{\alpha\beta\delta}^{X} + \overbrace{\alpha\beta\gamma}^{T} + \overbrace{\varepsilon\beta\gamma}^{Y} \Longrightarrow$$

$$\underbrace{\alpha\beta\gamma}_{XrY} + \underbrace{\alpha\beta\gamma}_{T} + \varepsilon\beta_1 + \beta_2\delta$$

$$\beta_1\beta_2 = \beta$$

!! Template becomes self-propagato

- *None of the 4 resulting products can serve as a template to recombin any 2 other molecules →*

TGR is irreversible !!

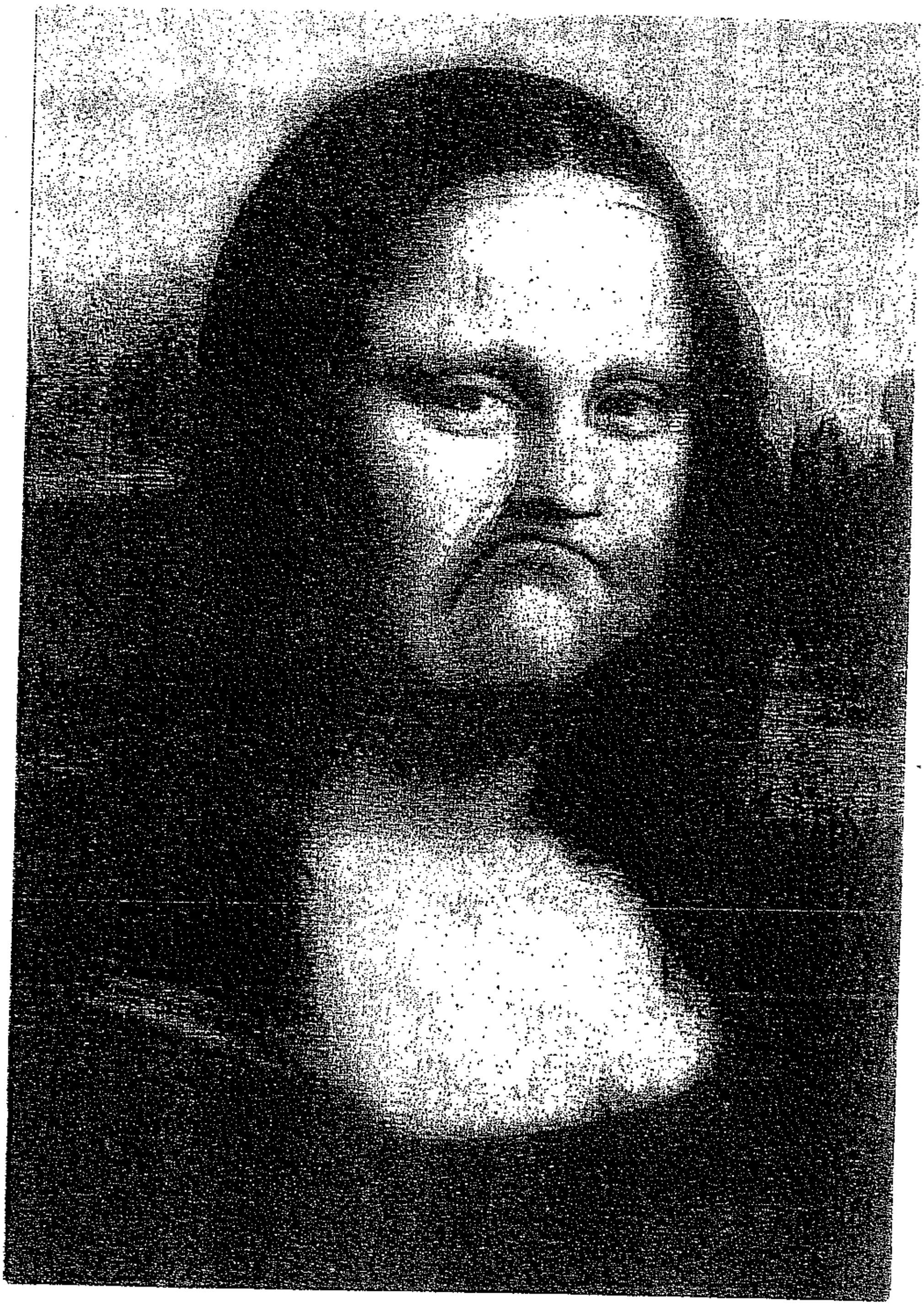
- The role of pointers:
  to point out where to cut
- The role of MDSs
  to cause the alignment

Summary:

The template guided recombination model provides satisfactory solutions to several crucial problems:

- identifies correct copies of pointer sequences
- defines precisely the boundary between MDSs and IESs
- provides for irreversible recombination
- self-propagation of template !!
- the whole mechanism is sequence aspecifi

6t



Template-Guided Recombination provides a solution of a problem crucial for understanding the biology of gene assembly; the problem was open for > 25 years!

* * *

Better understanding of bioware used to implement basic data structures

* * *

It is an interesting source for formulating theoretical models

COMPUTER SCIENCE

1° Mathematical theory of TGR → fundamental results on combinatorics of words

2° Many novel, interesting and challenging questions and models

3° New computing paradigms "Computing by Folding and Recombination"

4° Broader and novel understanding of the notion of computation

BIOLOGY

1° A formal system for reasoning about gene assembly in ciliates that leads to new <u>conjectures</u> and <u>experiments</u>
also: computer implementation

2° A <u>uniform</u> explanation of gene assembly for <u>all</u> experimental data

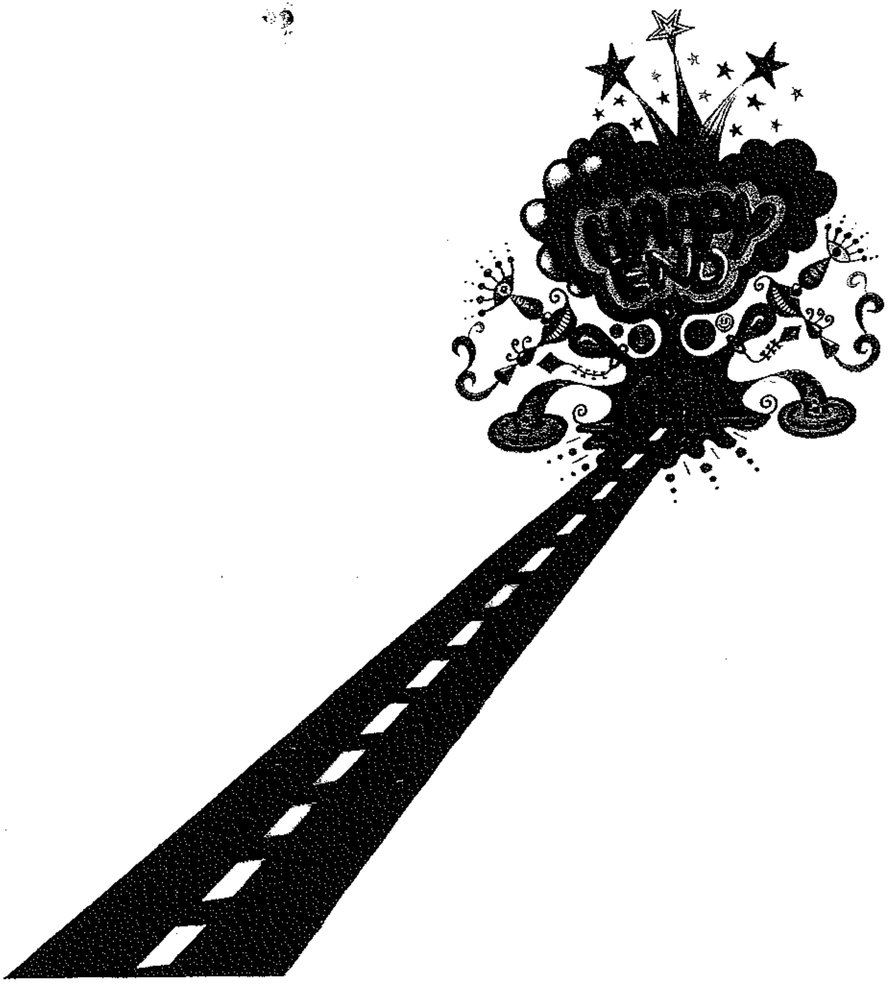
3° A solution to a problem open in biology for ~ <u>25 years</u> -
a progress in the quest to discover molecular hardware (bioware)

D.M. Prescott, A.Ehrenfeucht,
G. Rozenberg
"Template-guided recombination for IES elimination and unscrambling of genes in stichotrichous ciliates"
Journal of Theoretical Biology 222,
323-330, 2003

- - - - - - - - - - - - - -

A.Ehrenfeucht, T. Harju, I. Petre,
D.M. Prescott, G. Rozenberg
"Computation in Living Cells:
Gene Assembly in Ciliates"

Springer Verlag, 2004



Thank you for your attention!

www.liacs.nl/~rozenber/

rozenber@liacs.nl