

# The irony of molecular biology: informatics replaces mechanics

James A. Shapiro

University of Chicago

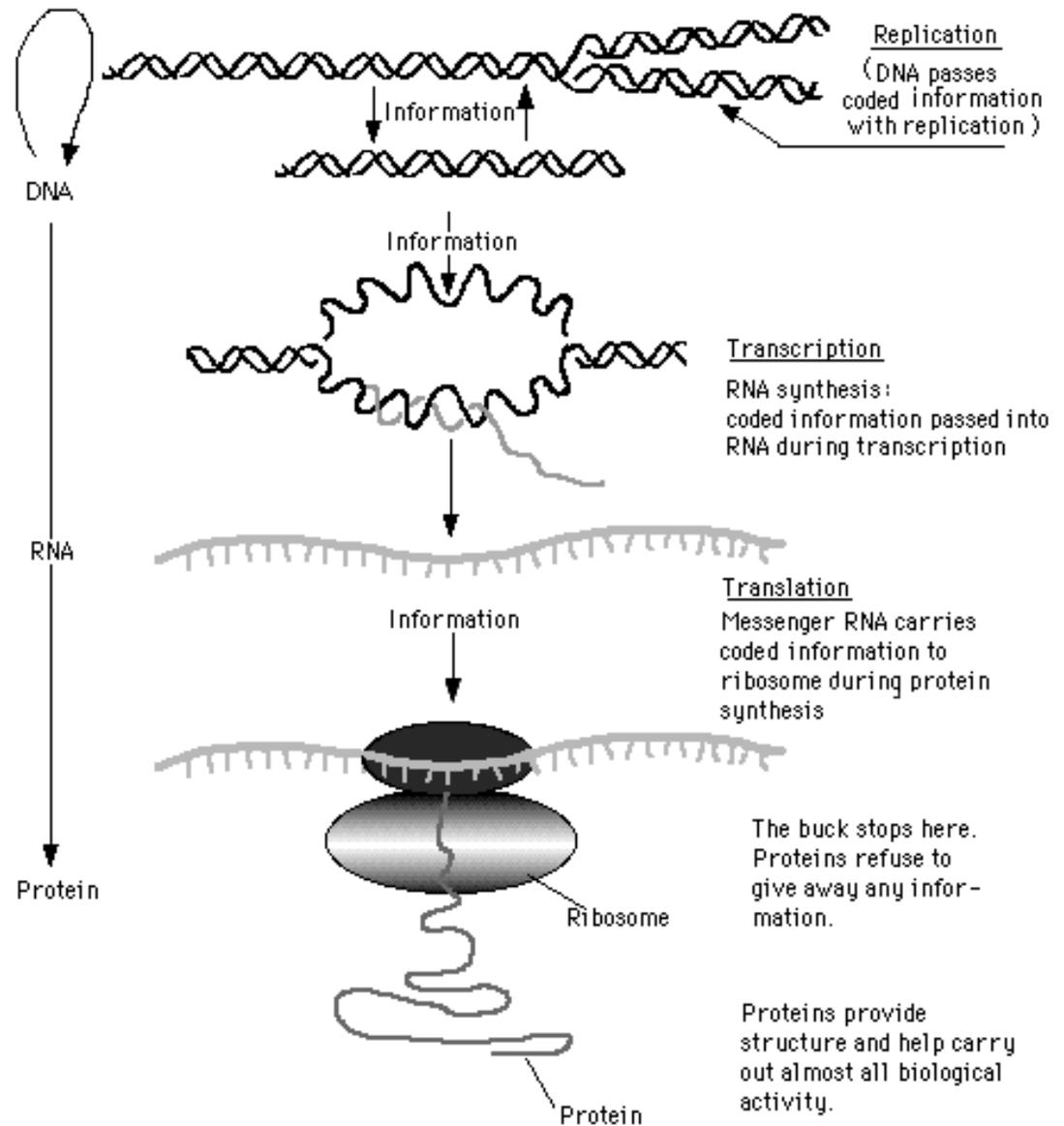
[jsha@uchicago.edu](mailto:jsha@uchicago.edu), <http://shapiro.bsd.uchicago.edu>

# Historical Context of Molecular Genetics

1. Early 20th Century: triumph of mechanism over vitalism, gene-directed view of heredity
2. 1940s: Modern Synthesis of Darwinism & Mendelism
3. 1953: Structure of DNA, chemical basis for heredity
4. 1962: Nobel Prize for DNA structure
5. 1970: “Central Dogma of Molecular Biology”
6. 1975: Nobel Prize for reverse transcription
7. 1983: Nobel Prize for Mobile Genetic Elements
8. 1993: Nobel Prize for Splicing
9. 1995: Nobel Prize for Embryonic Control Circuits
10. 2001: Nobel Prize for Cell Cycle Regulation

# Conventional View

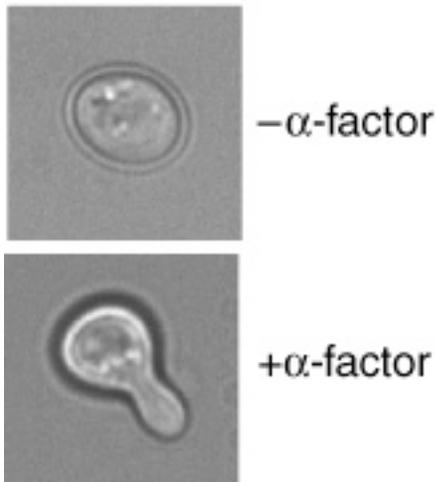
## The Central Dogma of Molecular Biology



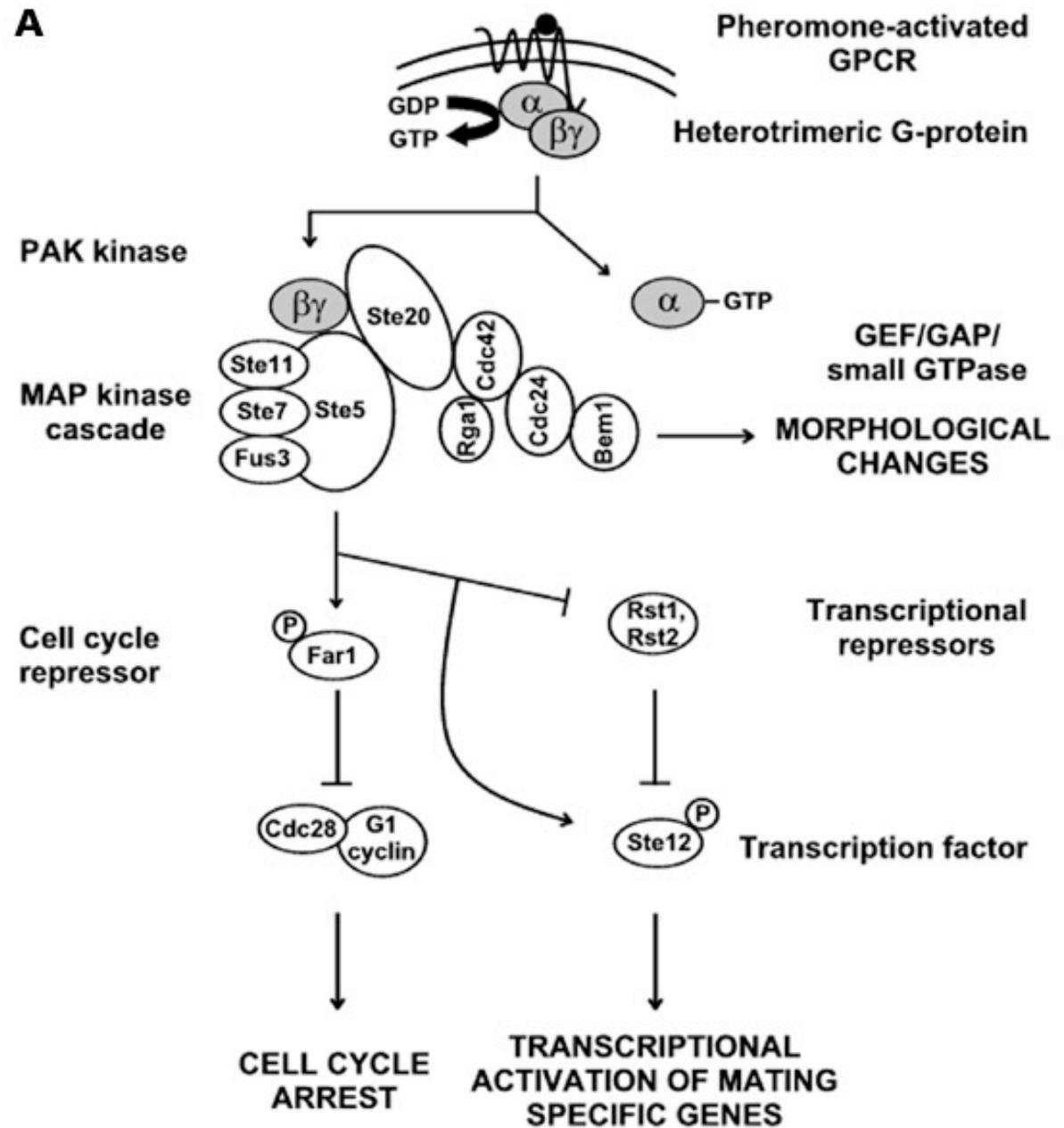
# Molecular Influences on the Genome that are more Informational than Mechanical

- Signaling molecules (hormones, cytokines, second messengers)
- Cell surface receptors for nutrients, signals, surfaces, neighboring cells
- Internal monitors for error & damage repair, checkpoint control
- Signal transduction networks to process information from receptors & monitors (e.g. kinase cascades, cell cycle control circuits)

# Signal Transduction in Yeast Mating

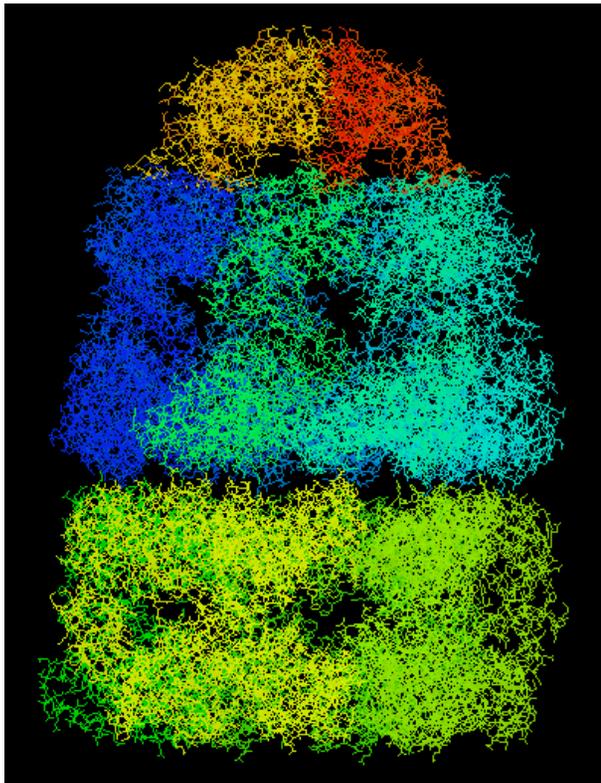


Pheromone Response  
Frank van Drogen et al. MAP kinase dynamics in response to pheromones in budding yeast. *Nature Cell Biology* 3, 1051 - 1059 (2001).

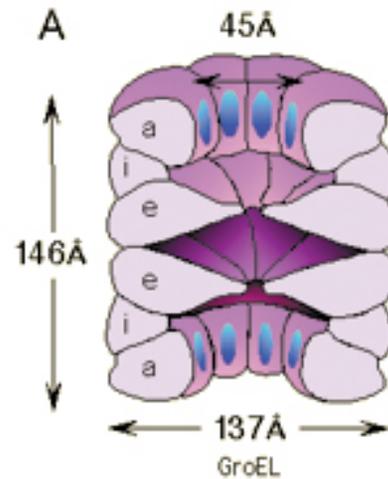


Mary J. Cismowski et al. Genetic screens in yeast to identify mammalian nonreceptor modulators of G-protein signaling. *Nature Biotechnology* 17, 878 - 883 (1999).

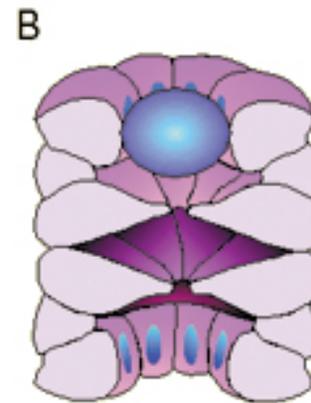
# A Protein Machine for Protein Folding



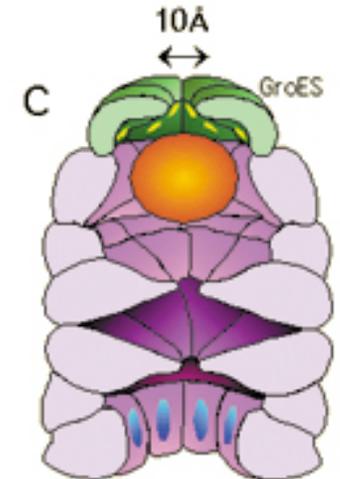
GroEL



GroEL with  
substrate bound



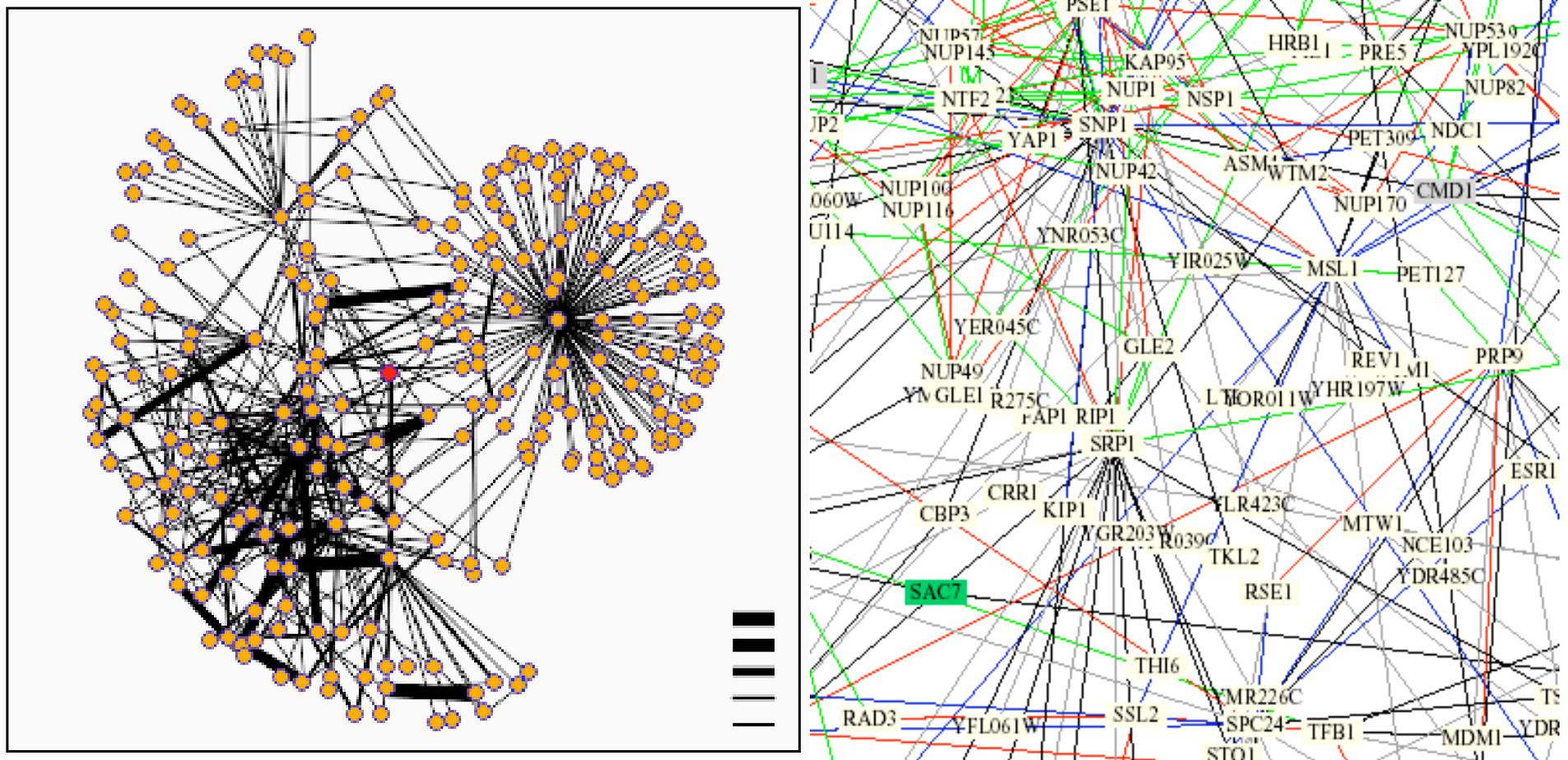
GroEL:GroES with  
enclosed substrat



<http://parasol.tamu.edu/~sthomas/>

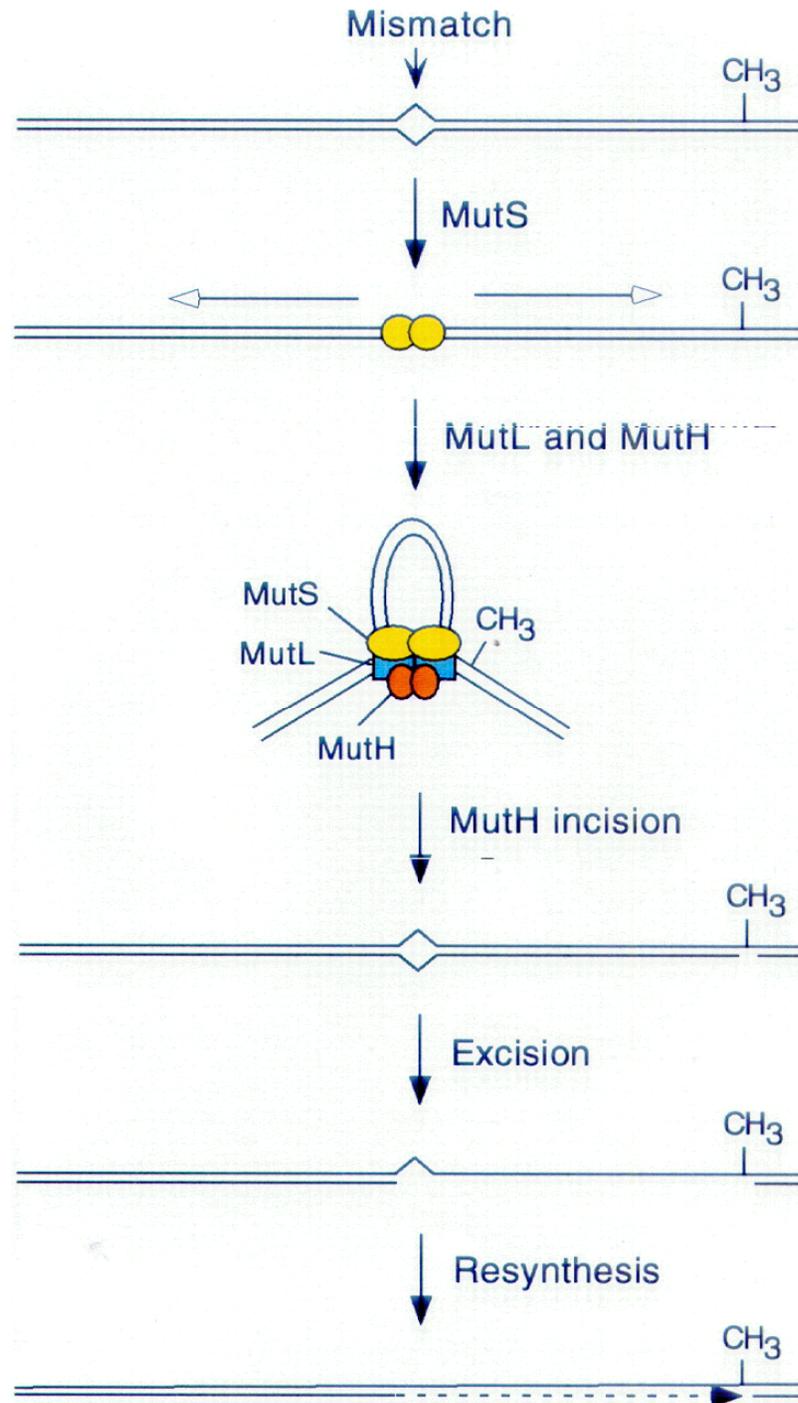
<http://www.biochem.mpg.de/en/research/rg/hayer-hartl/index.html>

# A Protein Network in Yeast Cells: Proteins interacting with yeast SSL2, a DNA repair helicase



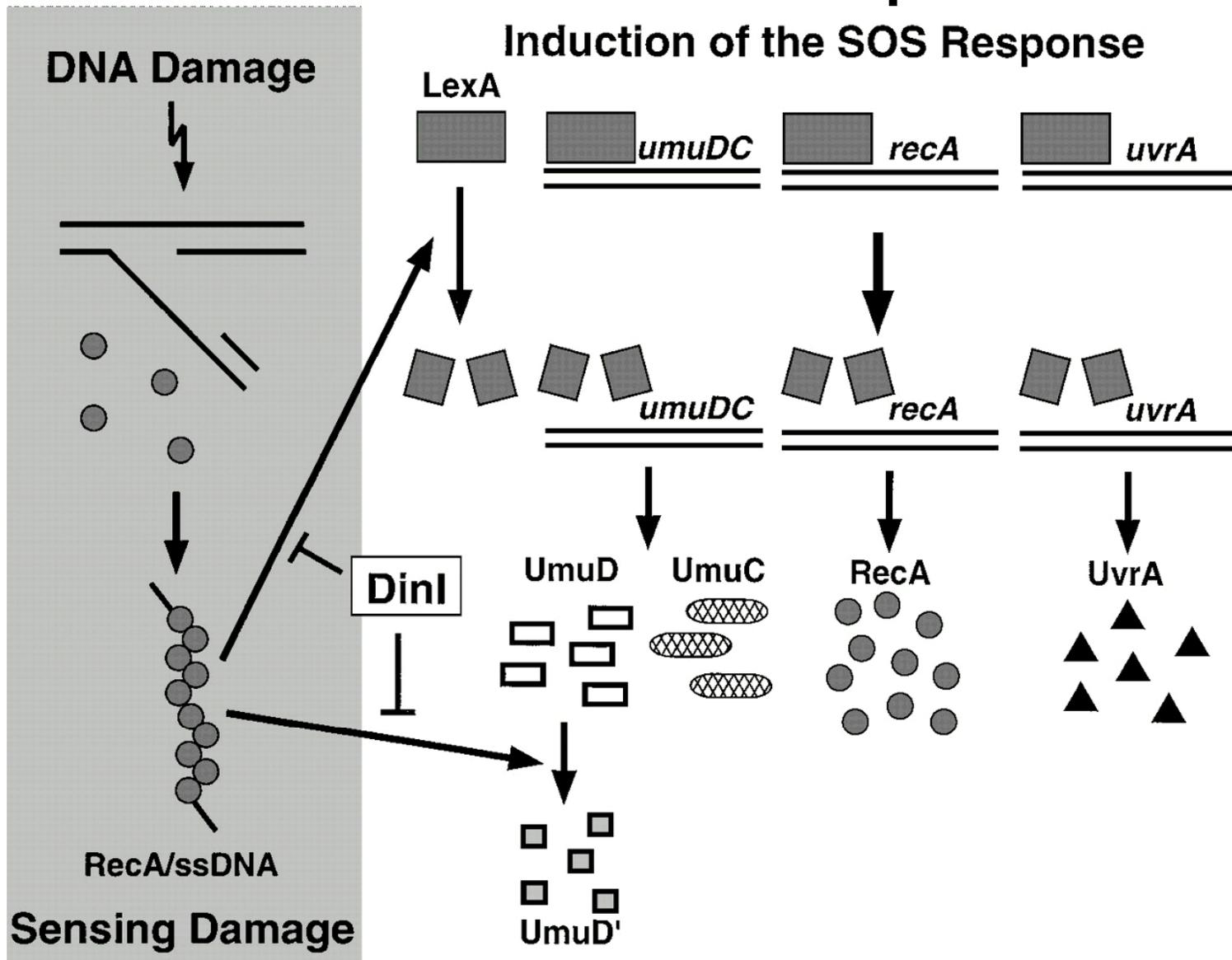
<http://www.bio.davidson.edu/courses/genomics/2002/James/FavoriteYeastProteins.htm>

# The Cognitive Nature of DNA Repair: Monitors and Enzymes



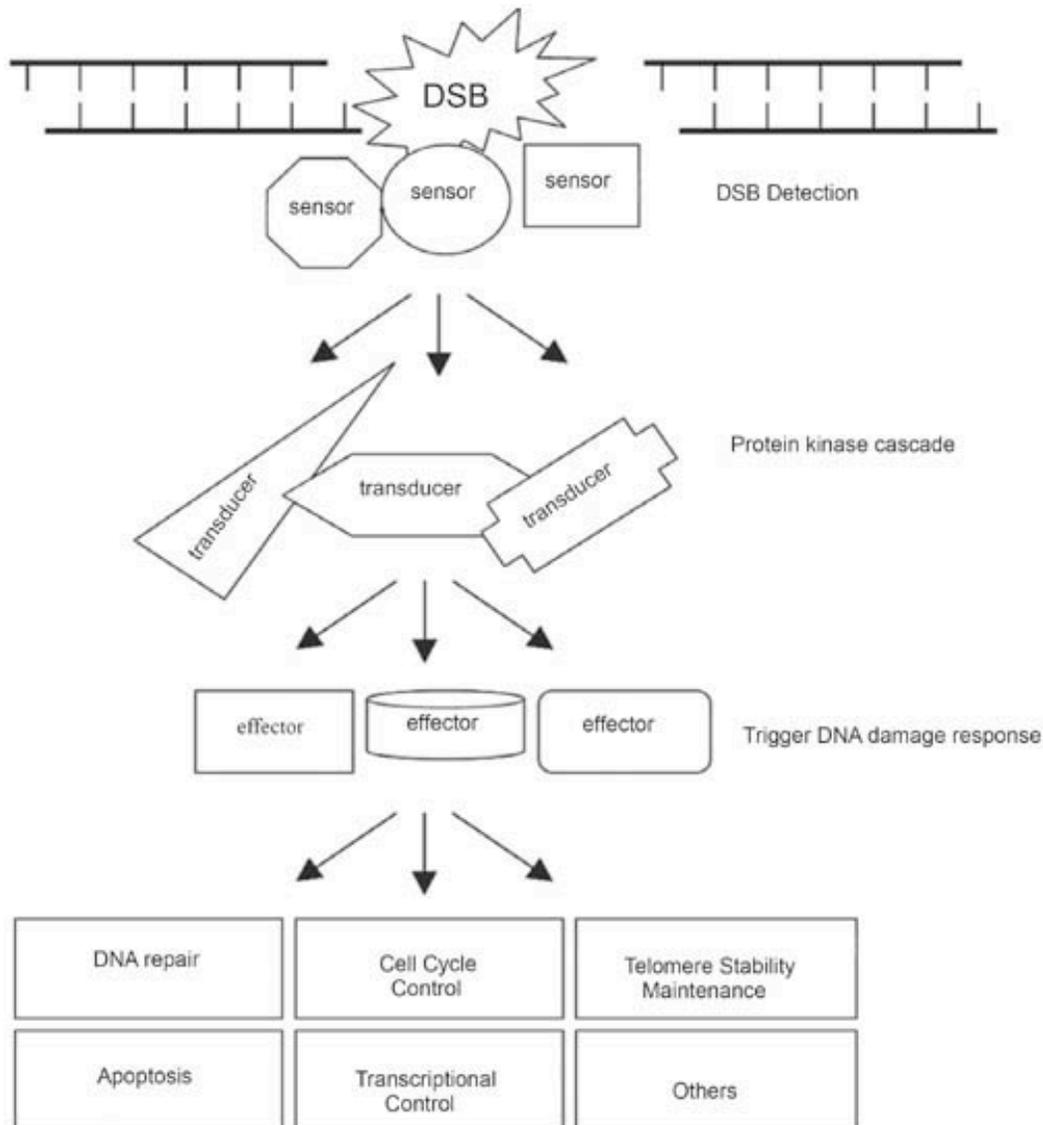
B. Harfe and S. Jinks-Robertson. DNA MISMATCH REPAIR AND GENETIC INSTABILITY. Annu. Rev. Genet. 2000. 34:359-399

# The Cognitive Nature of DNA Repair: the SOS response



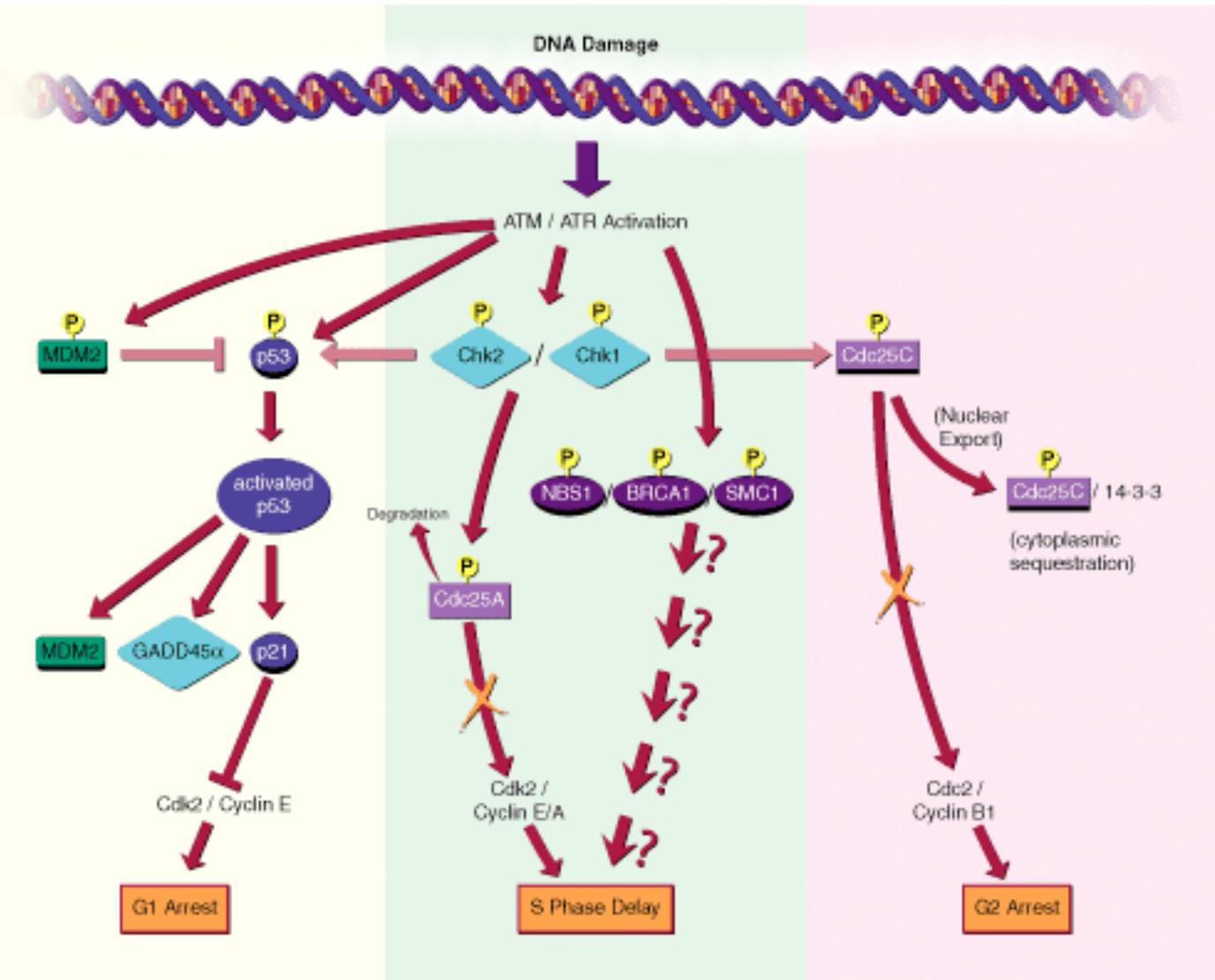
Sutton, et al. THE SOS RESPONSE: Recent Insights into umuDC- Dependent Mutagenesis and DNA Damage Tolerance. Annu. Rev. Genet. 2000. 34:479-497

# The Cognitive Nature of DNA Repair: mammalian responses to damage



Ying Zhang,  
Junqing Zhou and  
Chang UK Lim.  
The role of NBS1  
in DNA double  
strand break  
repair, telomere  
stability, and cell  
cycle checkpoint  
control. *Cell  
Research* (2006)  
**16**: 45–54. 2006.

# Checkpoints: self-monitoring and decision-making



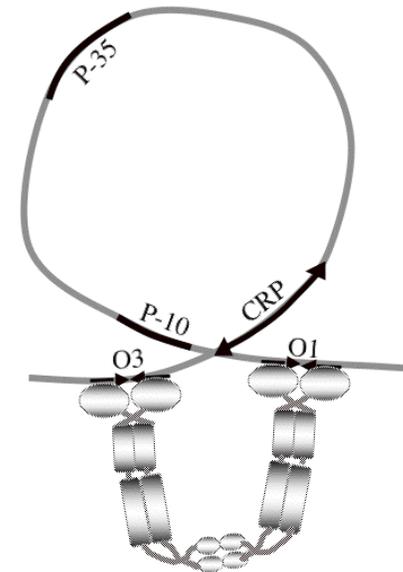
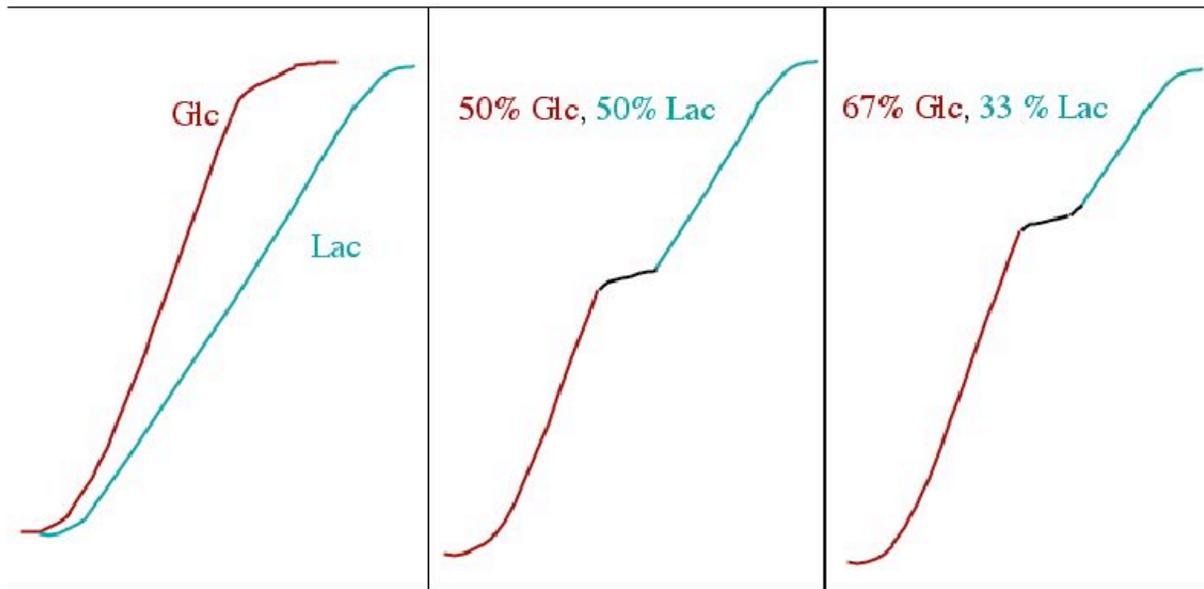
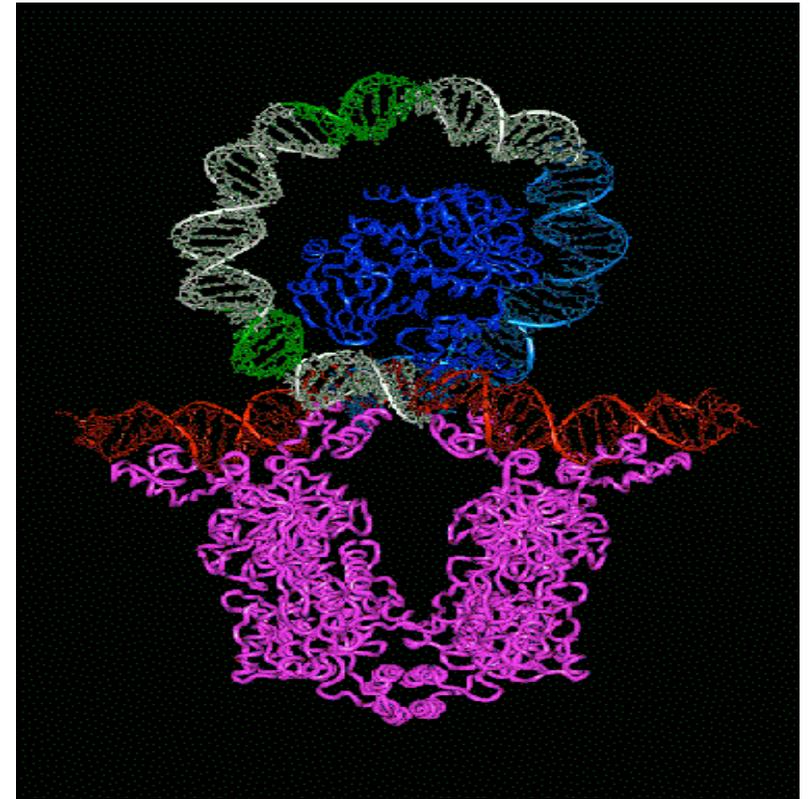
[http://www.rndsystems.com/mini\\_review\\_detail\\_objectname\\_MR03\\_DNA Damage Response.aspx](http://www.rndsystems.com/mini_review_detail_objectname_MR03_DNA Damage Response.aspx)

# The Genome as a RW Memory

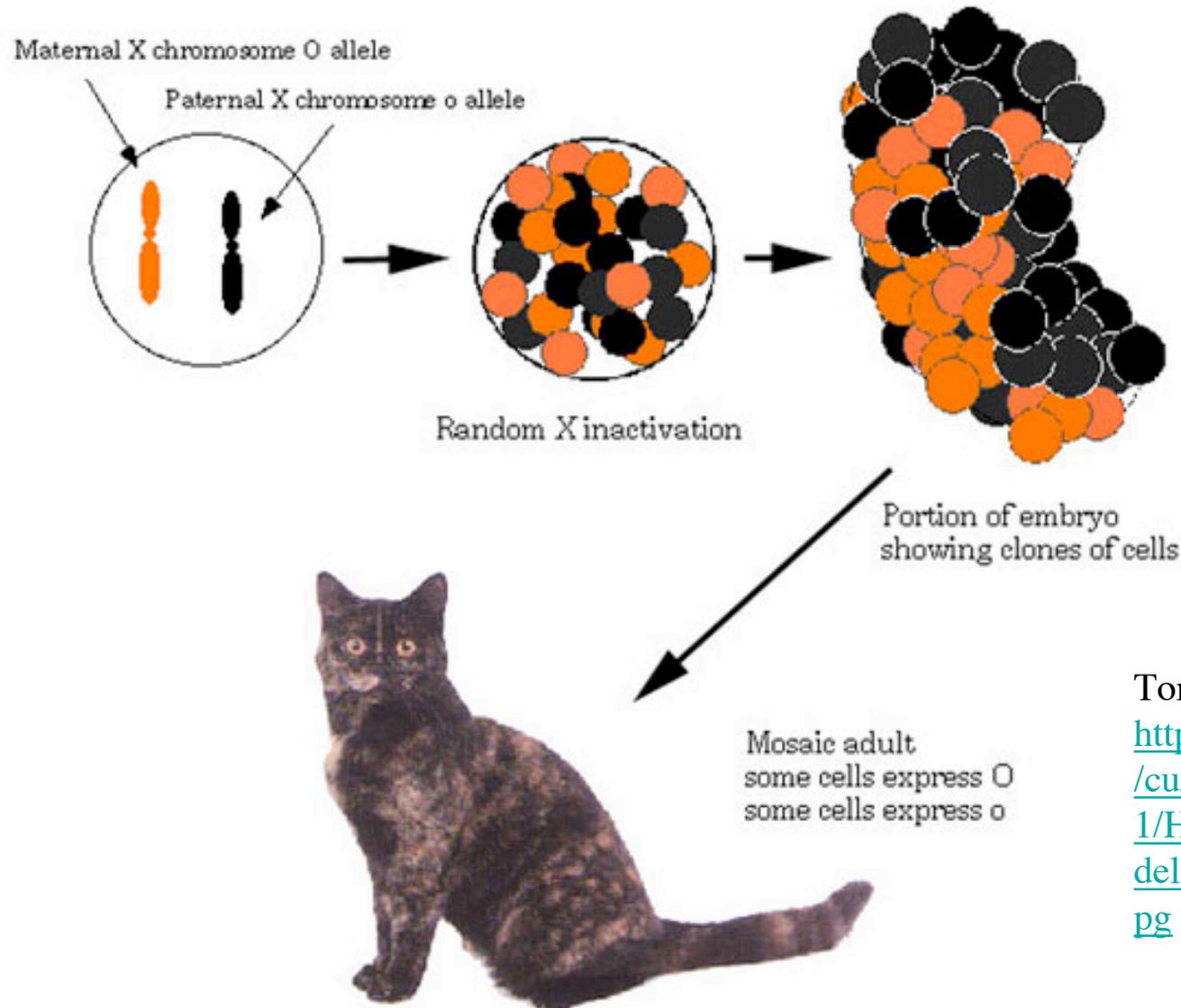
- Responses within a single cell cycle  
(computational memory)
- Responses that last many cell cycles  
(epigenetic memory)
- Responses that last many organismal  
generations (DNA sequence memory)

# Computational Memory: bacterial sugar metabolism

Crystal Structure of the  
Lactose Operon  
Repressor and Its  
Complexes with DNA  
and Inducer. Mitchell  
Lewis, et al. Science,  
Vol. 271(Mar. 1, 1996),  
pp. 1247-1254.



# Epigenetic Memory: X chromosome inactivation



Tortoise-shell Cat

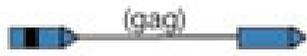
<http://www.columbia.edu/cu/biology/courses/w3031/HOandOverheads/mendelian%20folder/XInact.jpg>

# DNA Sequence Memory:

## dispersed repeats in the human genome

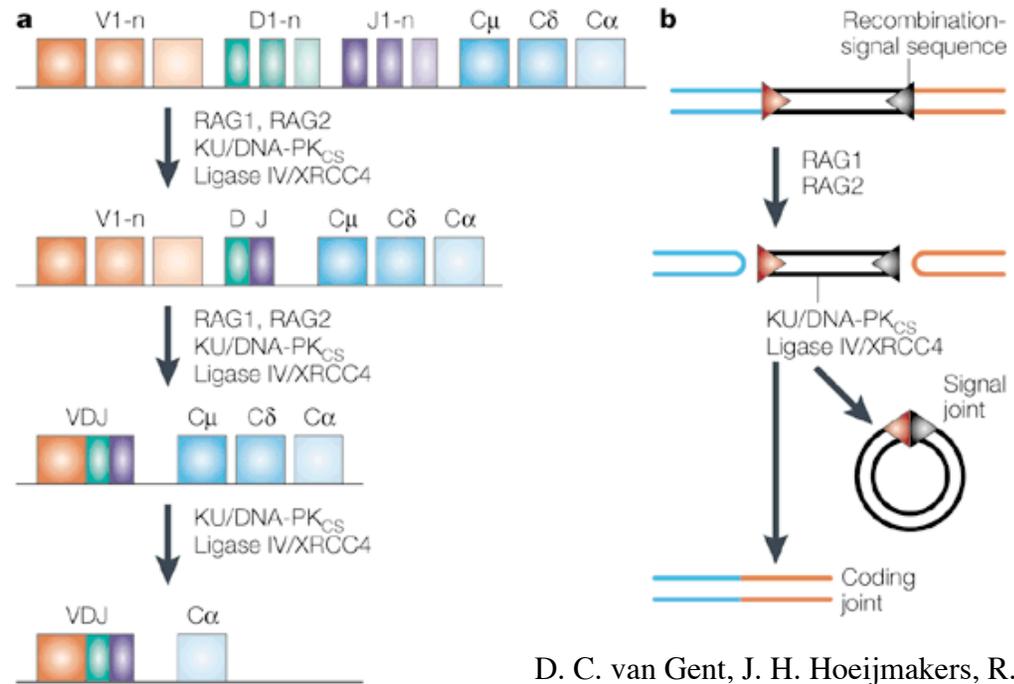
- coding sequences (data files): ~1.5 - 2% of human genome
- intervening sequences
- repetitive sequences: > 50% of human genome

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	<b>Autonomous</b>		6–8 kb	850,000	21%
	<b>Non-autonomous</b>		100–300 bp		
Retrovirus-like elements	<b>Autonomous</b>		6–11 kb	450,000	8%
	<b>Non-autonomous</b>		1.5–3 kb		
DNA transposon fossils	<b>Autonomous</b>		2–3 kb	300,000	3%
	<b>Non-autonomous</b>		80–3,000 bp		

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409, 860 - 921 (2001)

# Natural Genetic Engineering: the mammalian immune system

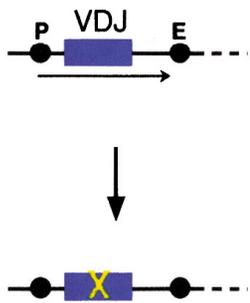


D. C. van Gent, J. H. Hoeijmakers, R. Kanaar, Chromosomal Stability And The DNA Double-Stranded Break Connection 2, 196 (2001) *Nature Reviews | Genetics*

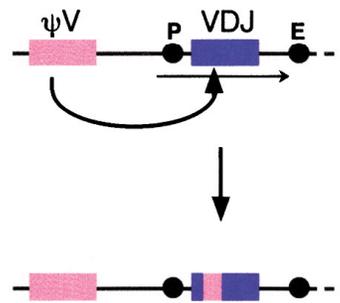
A. Ig Heavy Chain Locus



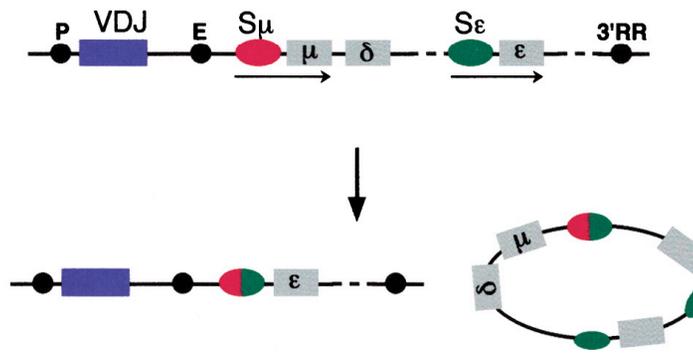
B. Hypermutation



C. Gene Conversion



D. Class Switch Recombination



Somatic hypermutation & class switch recombination - transcription directed

Tasuko Honjo, Kazuo Kinoshita, and Masamichi Muramatsu. 2001. Molecular Mechanism of Class Switch Recombination: Linkage with Somatic Hypermutation. *Annu. Rev. Immunol.*;

# Natural Genetic Engineering in Evolution: regulatory signals in the human genome

A

```

AluY      TTTTTTTTTGAGACGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTG
Hs CD8α   TTTTTTTTTGAGACGGAGTCTCGCTATCTCGCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTG

AluY      TCCTGCCTCAGCCTCCCGAGT-----AGCTGGACTACAGGCGCCGCCACCACGCCCGGCTAATTTTTTGTATTTTGTAGT
Hs CD8α   TCCTGCCTCAACCTCCAAGTAGCTGGGACTACAGGCACCCGCCAGCTGGGACTACAGGCACCCGCCACAACACCCGGCTAATTTTTTGTATTTTGTAGT

AluY      AGAGACGGGGTTTACCGTGTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCTCCCAAAGTGTGGGATTACAGGCGTGAG
Hs CD8α   AGAGACGGGGTTTACCGTGTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTTGGCCTCCCAAAGTGTGGGATTATAGGCGTGAG
  
```

B

```

L2        CAATCCATCAGCAAATCTGTGGCTCTACCT-TCAA-AATATATCCAGAA---TCCGACCACTTCTCACCACCTCC---ACTGCCACCACCCTGGTCCA
Hs GPIIb  -AGTTTATTACCATGTCCTATTGGTTCTGCCTATCAATGATGTCTCTTGAA---TC-----TTTACCCCATCTCT---ACTACCACCGTGTAGTCCA
Mm GPIIb  -AGCTCACCTCCATGTCCTGTGTCTCTGTTTCTCAGCCACAAGCCCGCAATCCC-----CTTCTCTTTCTCTTCTCTGTCACTGTGAAAGTCC-

L2        AGCCACCATCATCTCTGCCTGGATTACTGCAATAGCCTCCTAA-----CTGGTCTCCCTGCTT--CCACCCTTGCCCCCTNCAGT-CTA-TTCTC
Hs GPIIb  AGCCACCATCACTTTCTGCTTGGGATAGCGTGAT---TGTGAA-----CTGGTC--CATACTGTCTACTCTAGCC---TACAGTTCTA-ATCTC
Mm GPIIb  AGCCACCATCCCTTTCTGCCTGGGAAAGTATAAC---CATGAATTCGCTGCCTGGCT--CAC-----TTTGCCATGGCC---CACAGTTGTACAGCTC
  
```

C

```

MaLR      TAATCCCAATGTGATGGTATTAGGAGGTGGGGCCTTTGGGAGGTGATTAGGATTAGATGAGGTCATGAGGGCGGGGCCCTCATAATGGGATTAGTGCCC
Hs globin CAACTCCCAAC-TGACCTTATCTGTGGGGAGGCTTTTGAAGAAAGTAATTAGGTTTAGCTGAGCTCATAAGAGCAGATCCC-CATCATAAAATTATTTTCC

MaLR      TTAT-AAAAGAGACCYCAGAGAGCT--CCCTTGCCCTTCCGCCATGTGAGGACACAGTGAGAGGC-GCCGTCTACGAACCAGGGGAATGAGCCCTCAC
Hs globin TTATCAGAAG-----CAGAGAGACAAGCCATTTCTCTTTCTCCCGGTGAGGACACAGTGAGAGAGTCCGCCATCTGCAATCCAGGAAGAGAACCCGTGAC
  
```

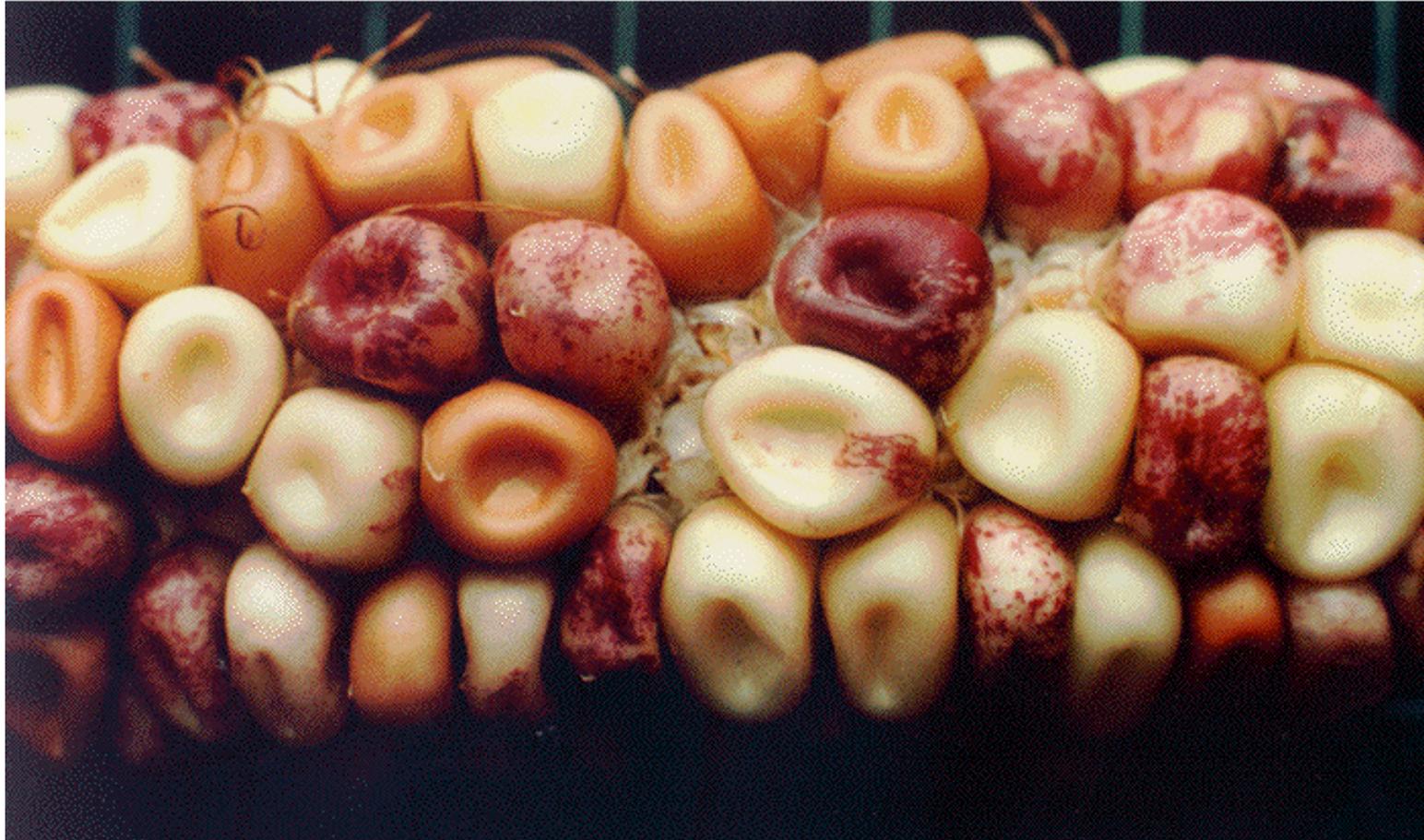
Fig. 3. Sequence alignments that show the relationship of TE-derived sequences, host promoter sequences and experimentally characterized cis-binding sites. TE family consensus sequences are aligned with host genome sequences. Cis-binding sites are characterized for human sequences and their locations in the alignments are boxed. (A) An Alu element that inserted after the diversification of the human and mouse lineages donated three cis-binding sites to human (Hs) CD8alpha gene regulatory sequences. (B) A L2 element that inserted prior to the diversification of the human (Hs) and mouse (Mm) lineages, and was then conserved, donated three cis-binding sites to the GPIIb gene regulatory region. (C) A MaLR element that inserted prior to the diversification of the human and mouse lineages but was only conserved in the human (Hs) lineage donated four cis-binding sites to the gammaA-globin enhancer region.

Mariño-Ramírez, K.C. Lewis, D. Landsman, I.K. Jordan. **Transposable elements donate lineage-specific regulatory sequences to host genomes.** *Cytogenetic and Genome Research* 2005;110:333-341

# What Does Natural Genetic Engineering Mean in Thinking About Evolution?

- Because genome change is a biological activity, it can be turned on and off; this means that genomic novelties can arise suddenly and involve multiple DNA changes.
- Because natural genetic engineering rearranges proven components non-randomly, the chances of successful inventions are greater.
- Because cells regulate natural genetic engineering activities, the potential exists for biologically significant feedback into genome evolution.

# Thinking About Genome Reorganization



"...variations which seem to us in our ignorance to arise spontaneously. It appears that I formerly underrated the frequency and value of these latter forms of variation, as leading to permanent modifications of structure independently of natural selection." (C. Darwin, *Origin of Species*, 6th edition, Chapter XV, p. 395).