Maladies chroniques et Tangrams moléculaires

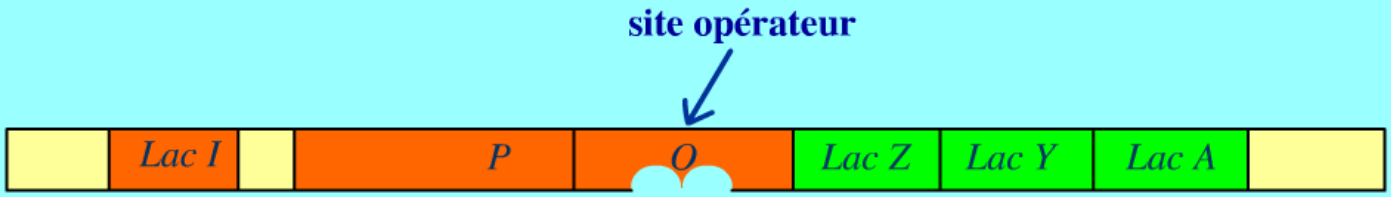
LAPTH
9 juin 2011

Jean-Marc Victor, CNRS, UPMC, Paris

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose

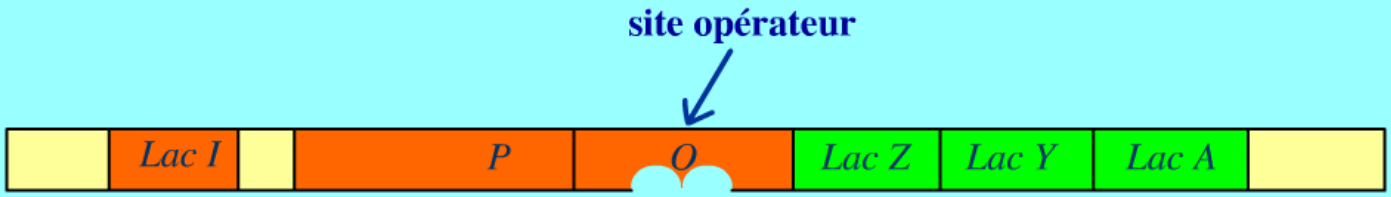


Répresseur (tétramère) à haute affinité pour l'opérateur

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose

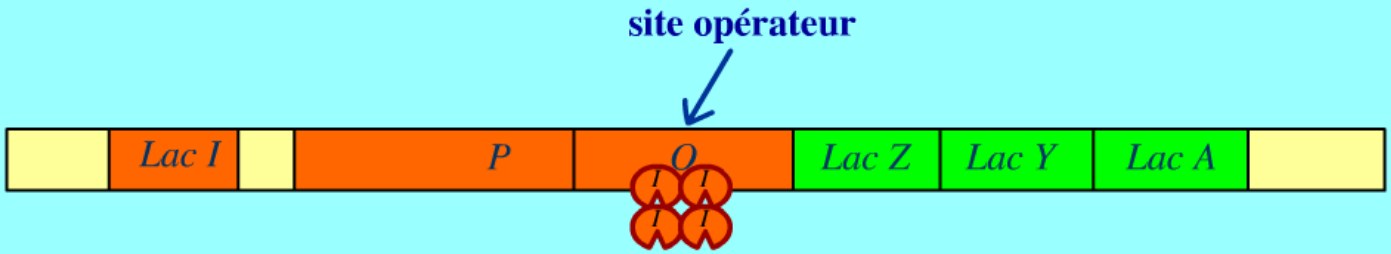


La protéine inhibitrice (le répresseur) forme un tétramère qui va se fixer sur un site spécifique au niveau de l'opérateur

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose

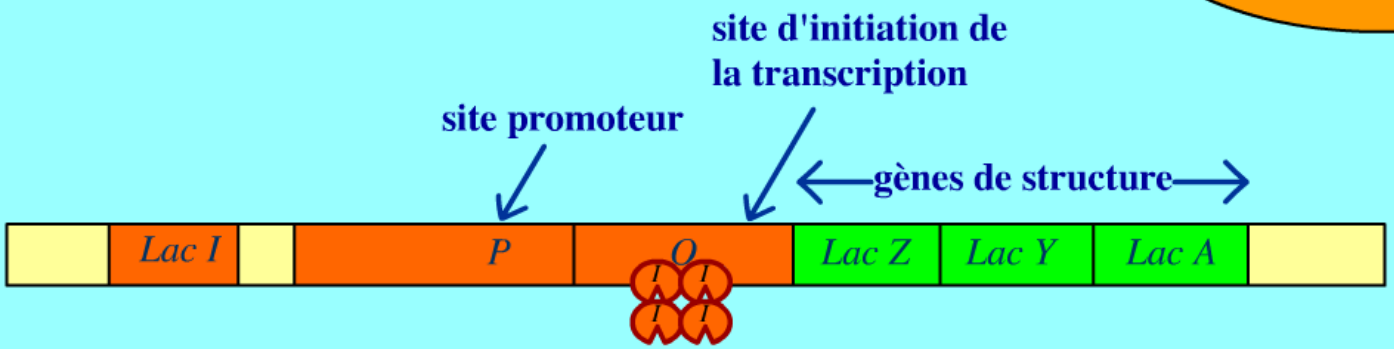


La protéine inhibitrice (le répresseur) forme un tétramère qui va se fixer sur un site spécifique au niveau de l'opérateur

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose



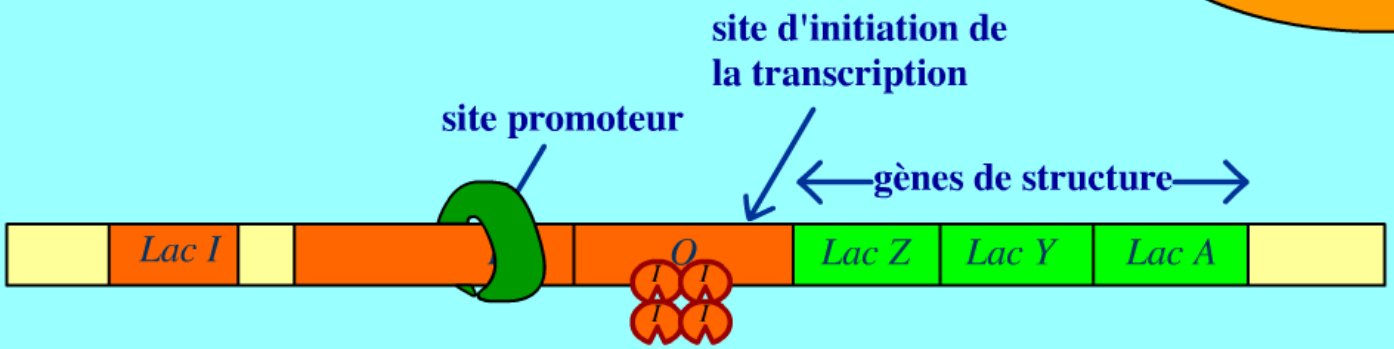
ARN polymérase

L'ARN polymérase peut se lier au promoteur mais elle est bloquée au niveau de l'opérateur et ne peut pas atteindre le site d'initiation de la transcription

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose

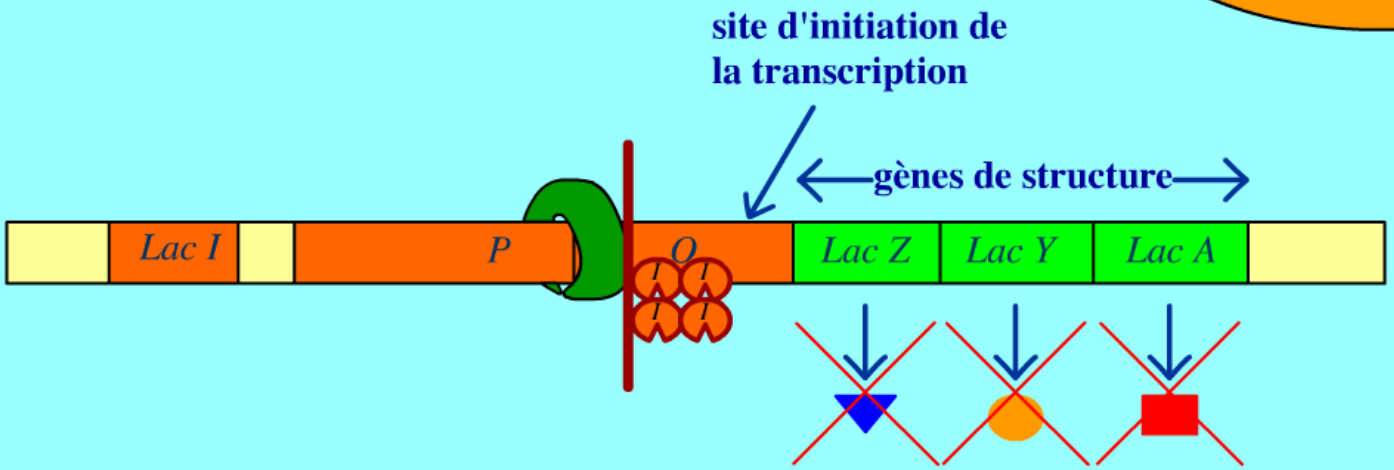


L'ARN polymérase peut se lier au promoteur mais elle est bloquée au niveau de l'opérateur et ne peut pas atteindre le site d'initiation de la transcription

L'Opéron lactose

Régulation de l'opéron lactose

En absence de lactose



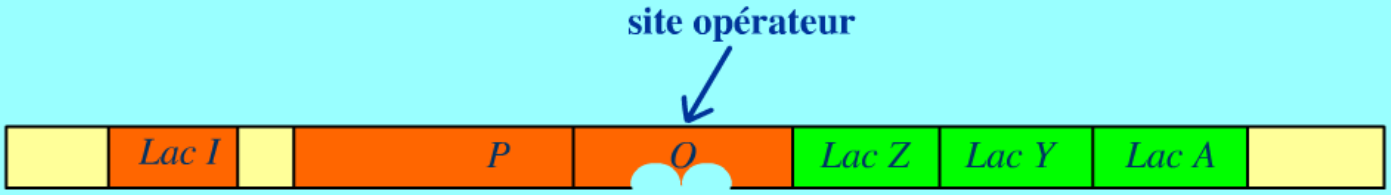
Les gènes structuraux ne sont pas exprimés

L'ARN polymérase peut se lier au promoteur mais elle est bloquée au niveau de l'opérateur et ne peut pas atteindre le site d'initiation de la transcription

L'Opéron lactose

Régulation de l'opéron lactose

En présence de lactose



Répresseur (tétramère) à haute affinité pour l'opérateur

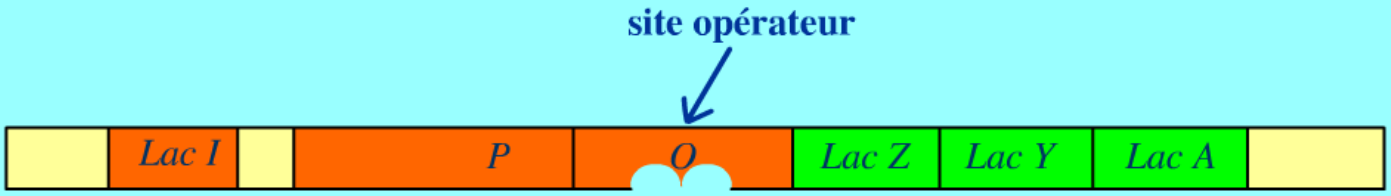
allolactose

Le lactose est métabolisé en allolactose.

L'Opéron lactose

Régulation de l'opéron lactose

En présence de lactose



L'allolactose se lie au répresseur

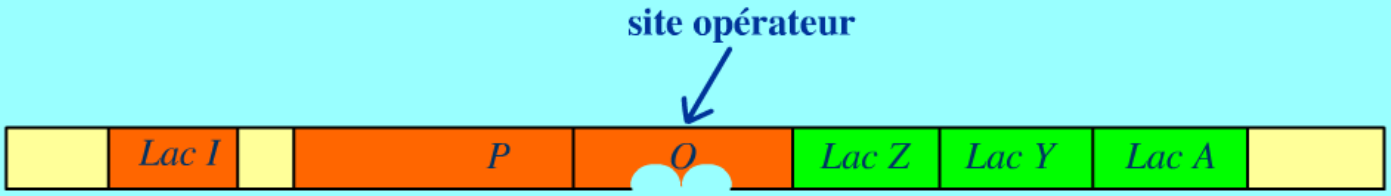
Répresseur (tétramère) à haute affinité pour l'opérateur

allolactose

L'Opéron lactose

Régulation de l'opéron lactose

En présence de lactose



Le répresseur combiné à l'allolactose perd son affinité pour l'opérateur

Allostérie

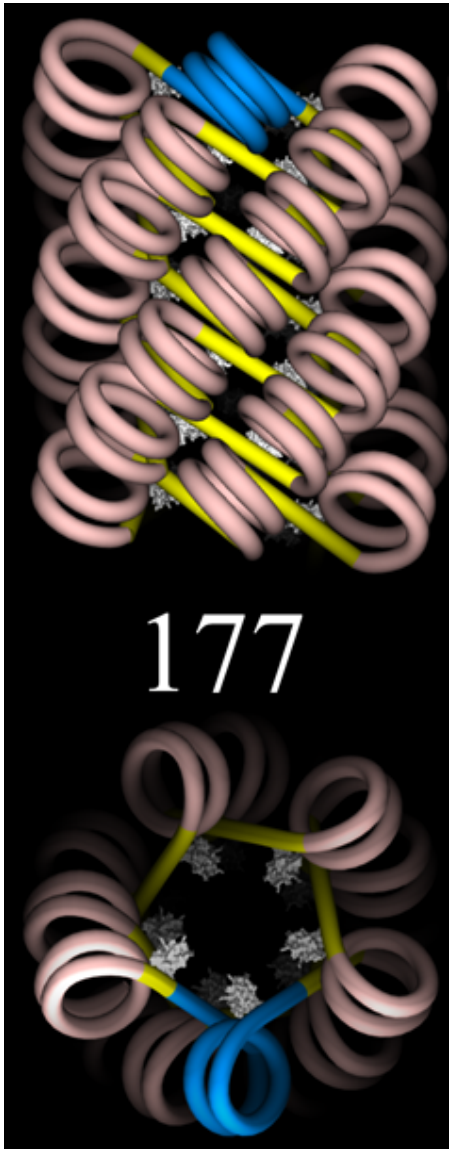
Interrupteur moléculaire

Switch: spécifique <--> non-spéciique

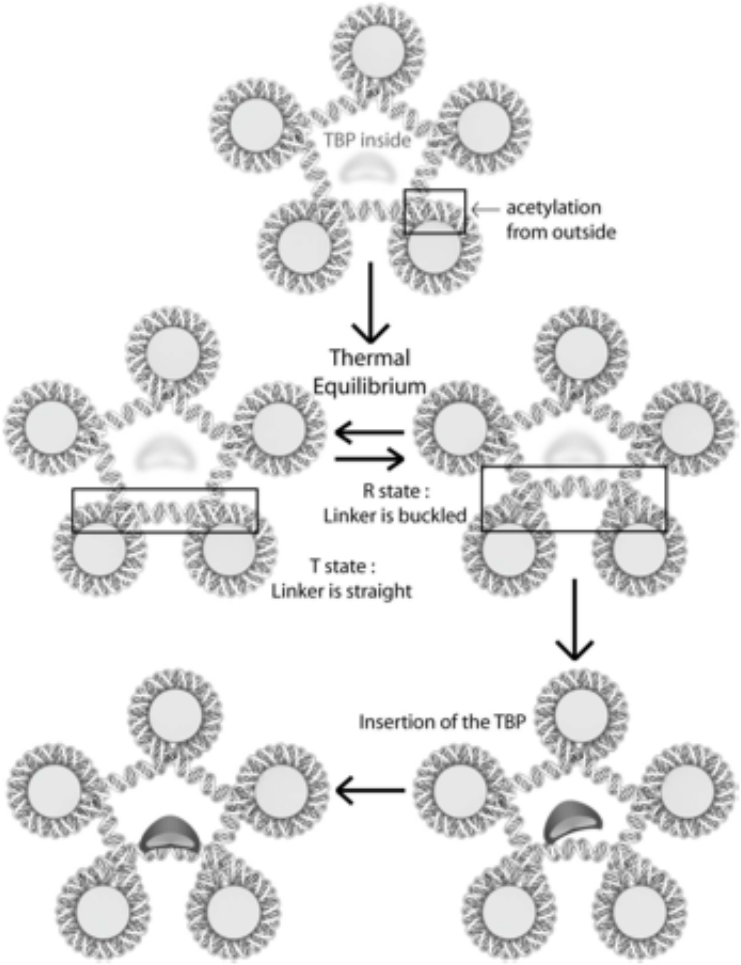
Bricolage

Et l'ADN ?

Fibre de chromatine



Fixation de la TBP par flambage des ADN de liaison



(Lesne & Victor, EPJE 2006)

Interaction ADN-Protéine

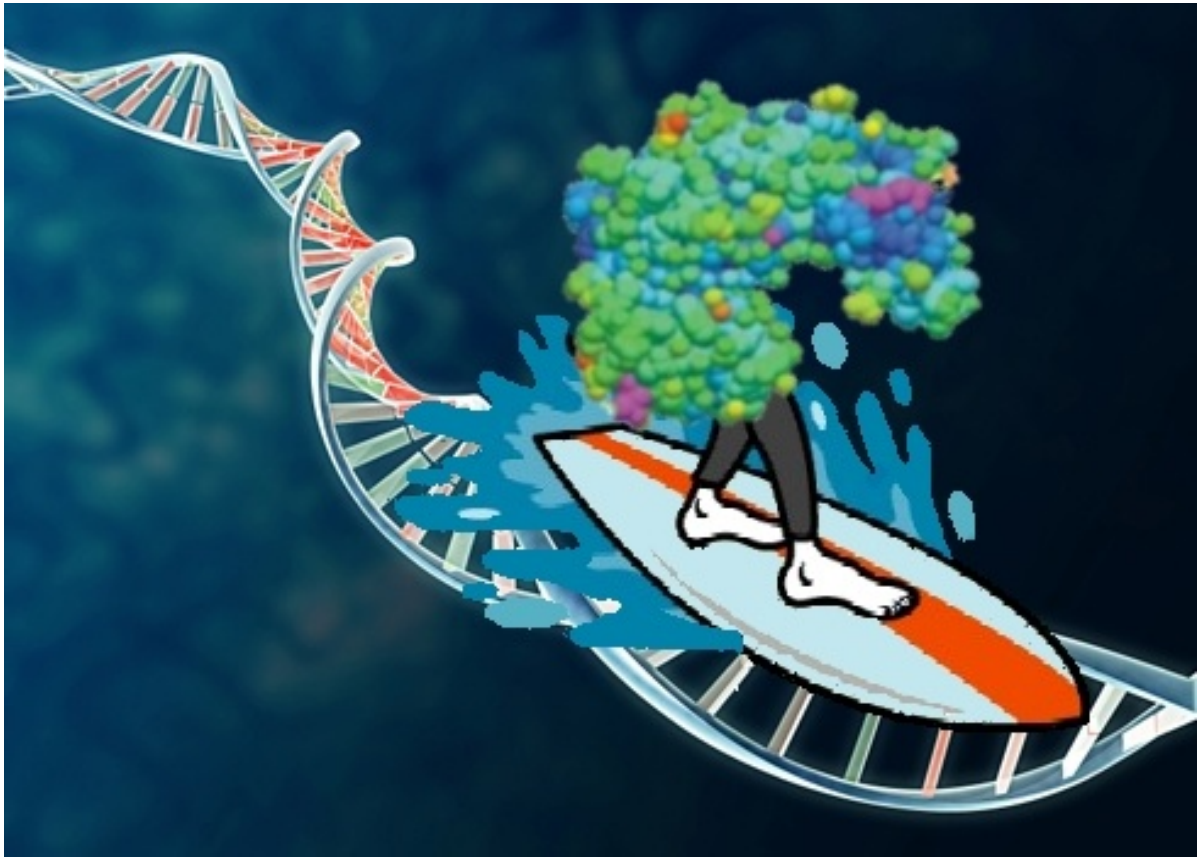
Spécifique / non-spécifique

Synergie

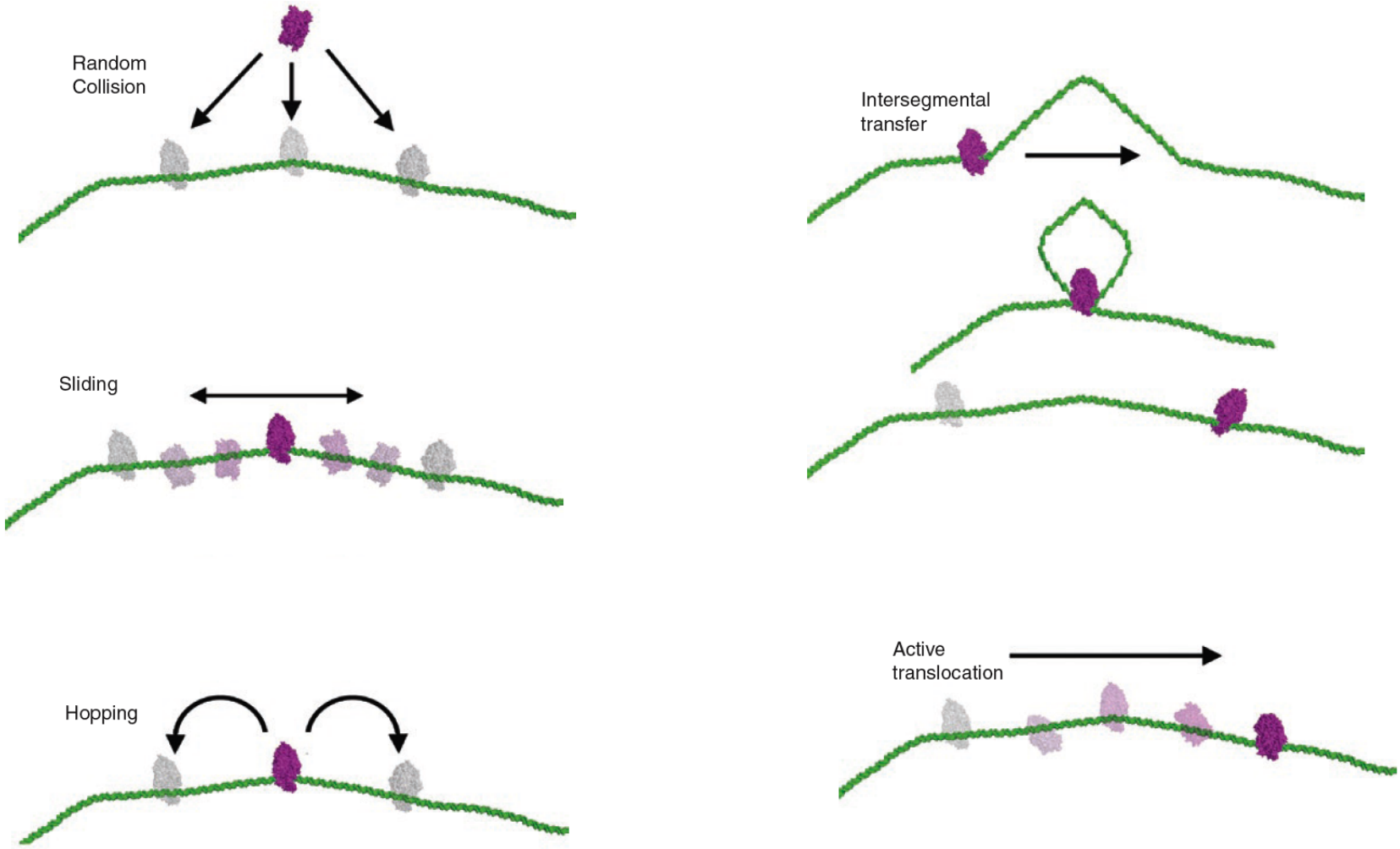
Co-évolution

Principes ?

Why proteins slide on DNA
and
How they find their target



Potential modes of target-site location

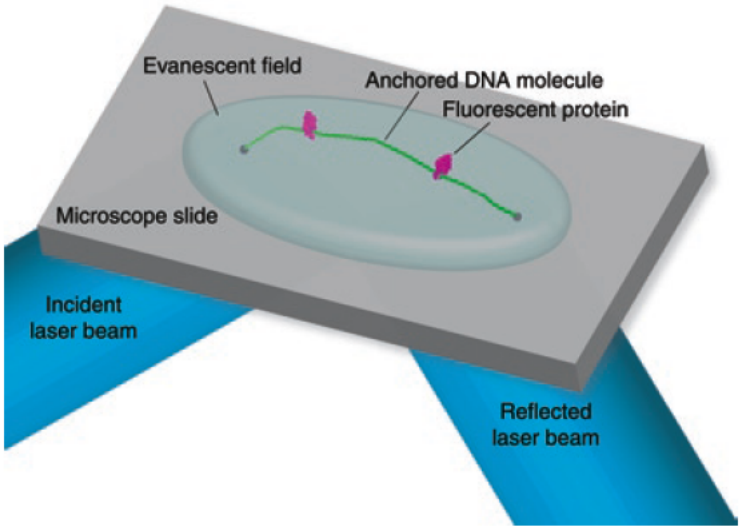


REVIEW

nature
structural &
molecular biology

Visualizing one-dimensional diffusion of proteins along DNA

Jason Gorman[1] & Eric C Greene[2]



TIRFM (Total Internal Reflection Fluorescence Spectrroscopy)

Mind the shape !

nature Vol 461 29 October 2009 doi:10.1038/nature08473

ARTICLES

The role of DNA shape in protein–DNA recognition

Remo Rohs[1]*, Sean M. West[1]*, Alona Sosinsky[1]†, Peng Liu[1], Richard S. Mann[2] & Barry Honig[1]

The recognition of specific DNA sequences by proteins is thought to depend on two types of mechanism: one that involves the formation of hydrogen bonds with specific bases, primarily in the major groove, and one involving sequence-dependent deformations of the DNA helix. By comprehensively analysing the three-dimensional structures of protein–DNA complexes, here we show that the binding of arginine residues to narrow minor grooves is a widely used mode for protein–DNA recognition. This readout mechanism exploits the phenomenon that narrow minor grooves strongly enhance the negative electrostatic potential of the DNA. The nucleosome core particle offers a prominent example of this effect. Minor-groove narrowing is often associated with the presence of A-tracts, AT-rich sequences that exclude the flexible TpA step. These findings indicate that the ability to detect local variations in DNA shape and electrostatic potential is a general mechanism that enables proteins to use information in the minor groove, which otherwise offers few opportunities for the formation of base-specific hydrogen bonds, to achieve DNA-binding specificity.

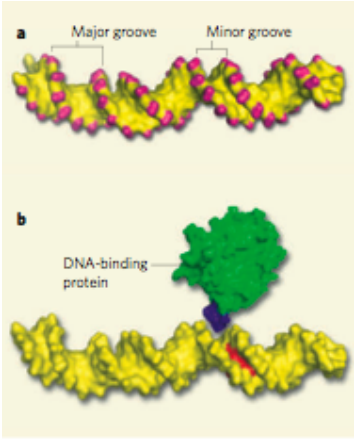


Figure 1 | Getting into the groove. Rohs *et al.*[1] report that the shape of the minor groove of DNA can direct the binding of proteins to specific sites. **a**, Negatively charged phosphate groups (magenta) are arrayed along the outer edge of the DNA major and minor grooves that spiral around the axis of the double helix. The width of the minor groove varies depending on the sequence of nucleotides. This variation leads to differences in the distance between phosphates across the groove, which in turn lead to variation in the negative electrostatic potential along the minor groove. **b**, A representation of a DNA-binding protein (green) that has a positively charged side chain on its surface, for example arginine (purple), is shown. The protein may recognize a binding site on DNA by its electrostatic potential. The protein is about to bind to the segment of the DNA minor groove that has the optimum groove width and electrostatic potential for binding (red). The DNA structure in **a** and **b** is derived from a structure in the RCSB Protein Data Bank, accession number 2O61. The illustration of a DNA-binding protein in **b** is hypothetical.

Electrostatic hallmark of the sequence?

DNA or protein shape?

MODELING NON-SPECIFIC DNA-PROTEIN INTERACTION

V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, J.M. Victor

PRL 102, 228101 (2009)

Poisson Boltzmann solution for oppositely charged plates

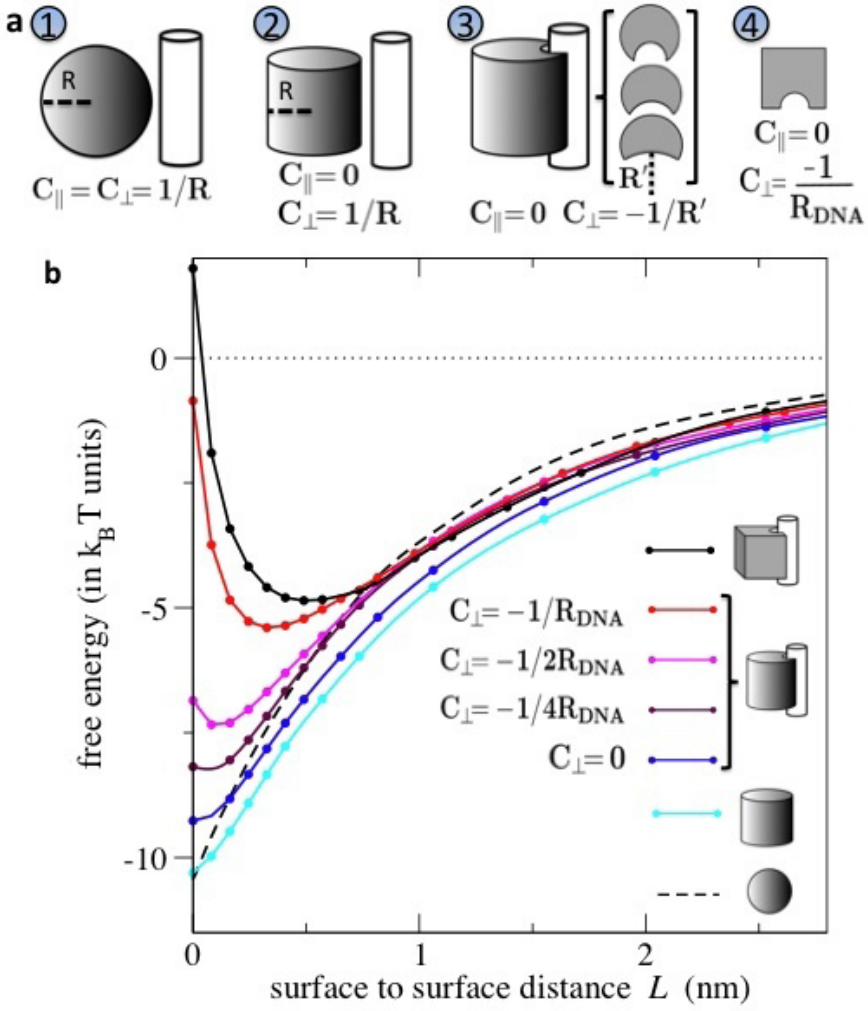
F. Pailluson, M. Barbi , J.M. Victor (Mol Phys 2009)

$$\sigma_{\mathrm{DNA}} = -1.0\, e\ \mathrm{nm}^{-2} \qquad \lambda_D = 1\ \mathrm{nm}$$

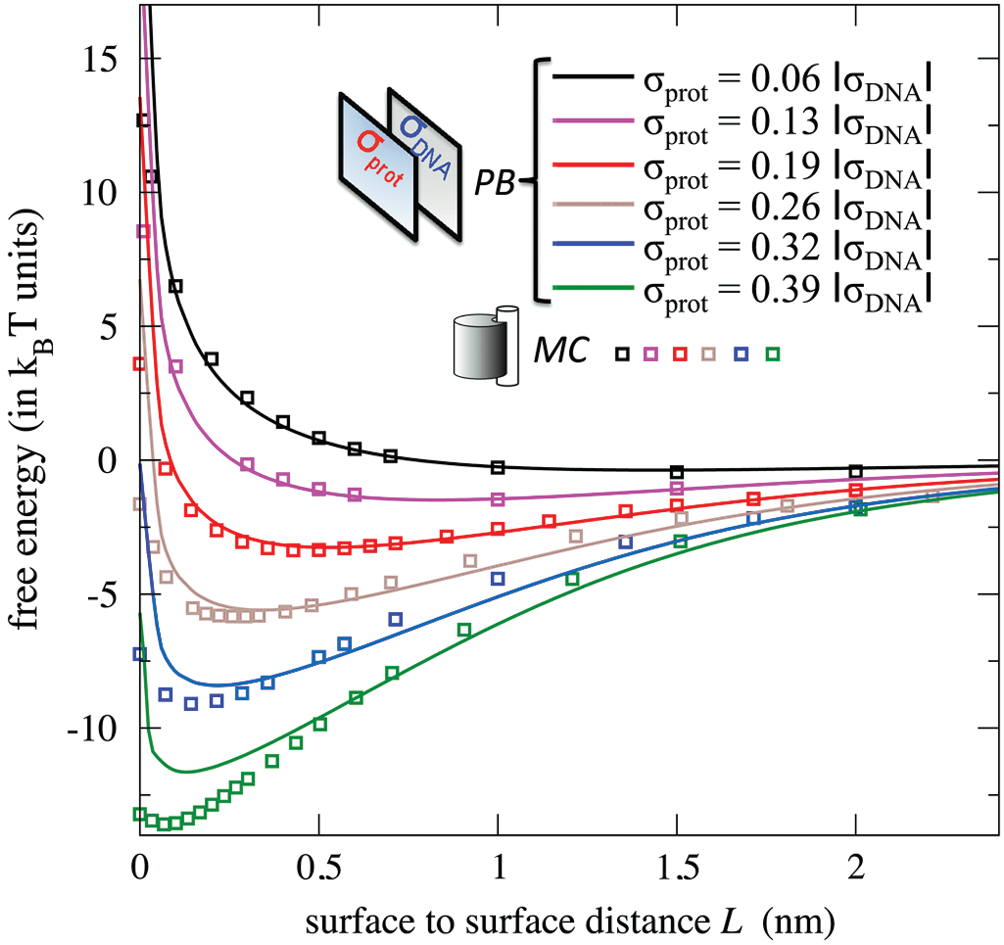
$$L_{\min} = \lambda_D \left| \ln \left(\frac{|\sigma'_{\mathrm{DNA}}|(2+\sqrt{\sigma'^2_{\mathrm{prot}}+4})}{|\sigma'_{\mathrm{prot}}|(2+\sqrt{\sigma'^2_{\mathrm{DNA}}+4})} \right) \right|$$

$$\text{where } \sigma'_X = 4\pi \ell_B \lambda_D \sigma_X$$

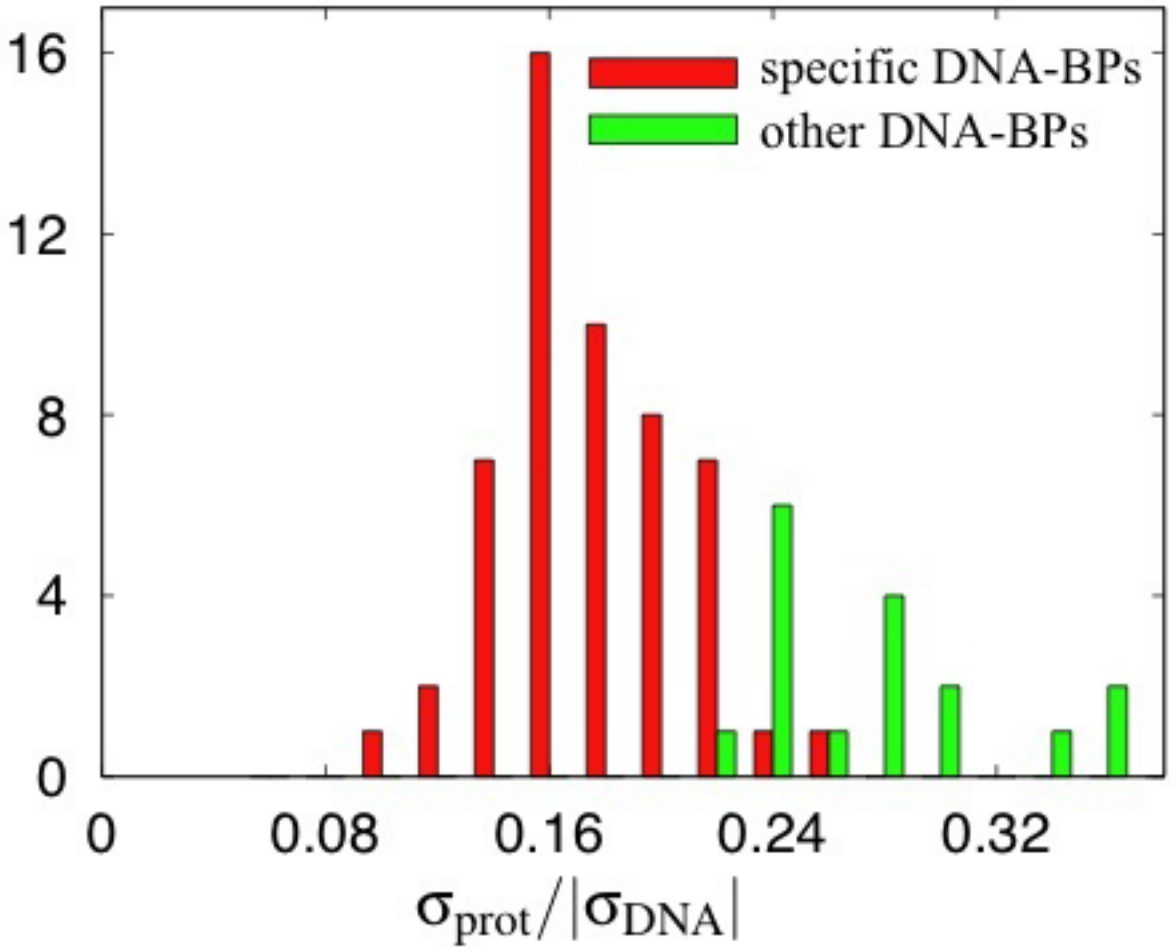
Influence of the protein shape



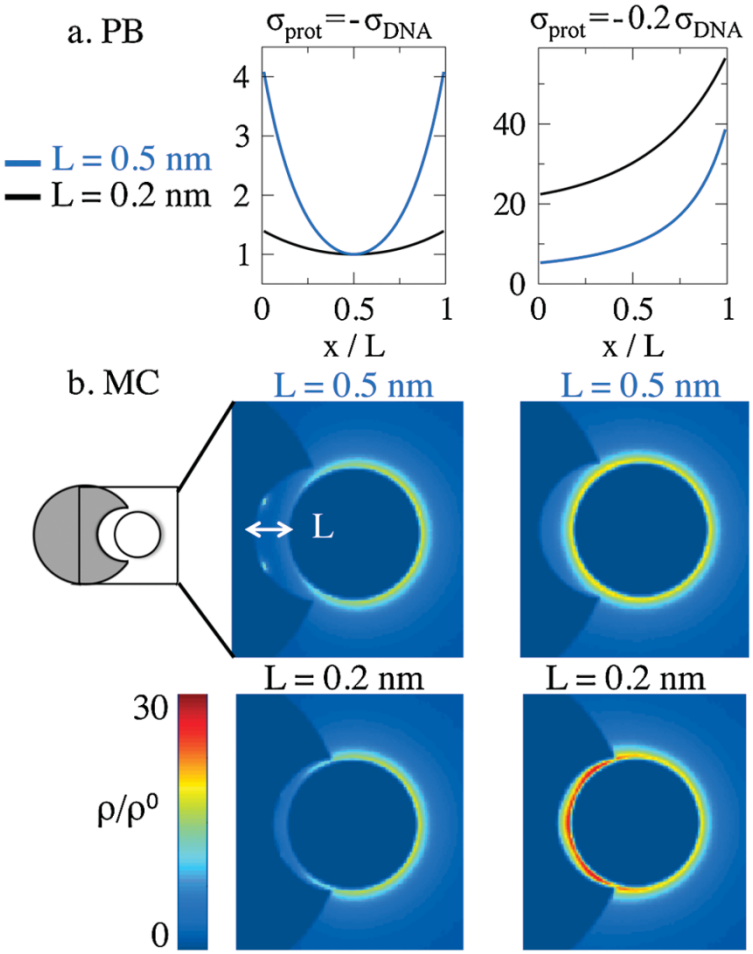
Influence of the protein charge density



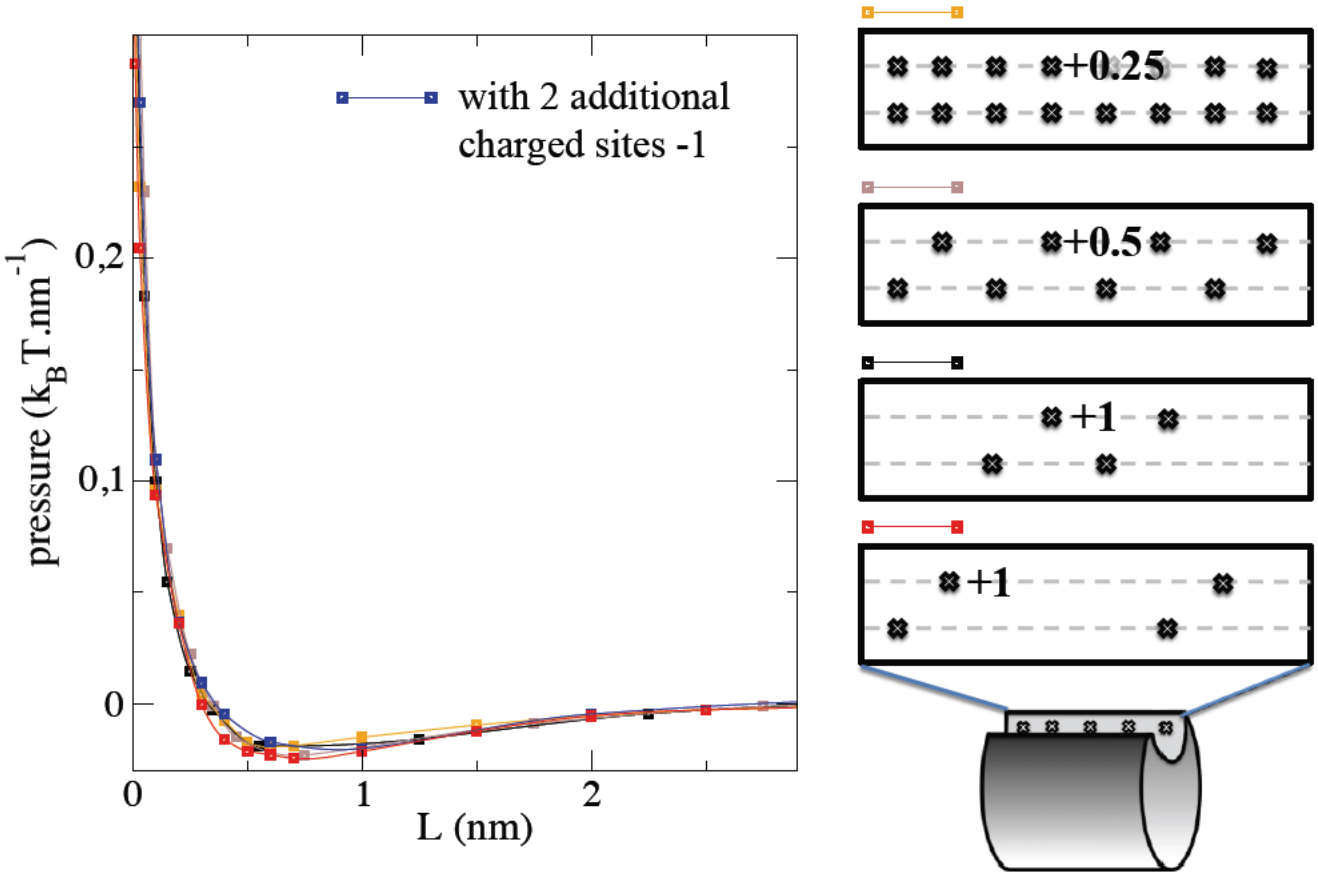
Specific vs Non-specific DNA-Binding Proteins



Visualizing ionic densities



No influence of the protein charge distribution



Modeling Hydrogen bonds

Morse potential $V_M(L) = D[(e^{-\alpha L} - 1)^2 - 1]$

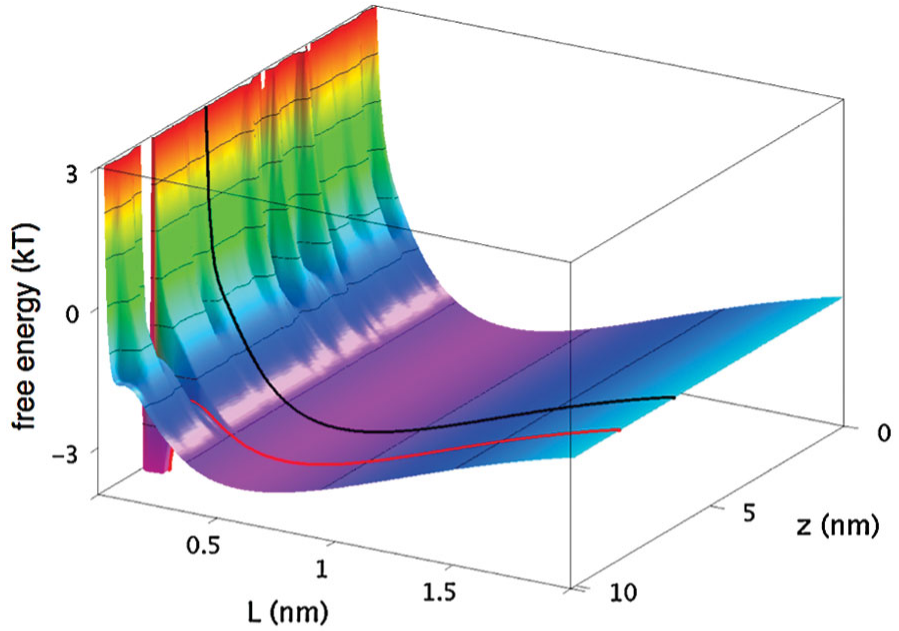
$D = 0.5\ \mathrm{k_B T}$ and $\alpha = 20\ \mathrm{nm}^{-1}$

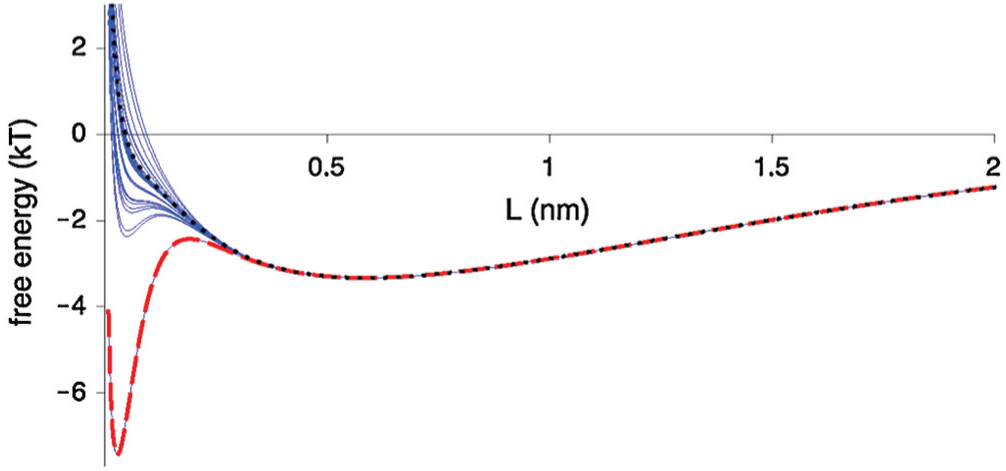
Gaussian distribution : $\langle n \rangle = n_{\max}/3$

$n_{\max} = 30$

$\sigma_n = \sqrt{n_{\max}}$

Search… and Find!





Mind the shape !!!

nature Vol 461 29 October 2009 doi:10.1038/nature08473

ARTICLES

The role of DNA shape in protein–DNA recognition

Remo Rohs[1]*, Sean M. West[1]*, Alona Sosinsky[1]†, Peng Liu[1], Richard S. Mann[2] & Barry Honig[1]

The recognition of specific DNA sequences by proteins is thought to depend on two types of mechanism: one that involves the formation of hydrogen bonds with specific bases, primarily in the major groove, and one involving sequence-dependent deformations of the DNA helix. By comprehensively analysing the three-dimensional structures of protein–DNA complexes, here we show that the binding of arginine residues to narrow minor grooves is a widely used mode for protein–DNA recognition. This readout mechanism exploits the phenomenon that narrow minor grooves strongly enhance the negative electrostatic potential of the DNA. The nucleosome core particle offers a prominent example of this effect. Minor-groove narrowing is often associated with the presence of A-tracts, AT-rich sequences that exclude the flexible TpA step. These findings indicate that the ability to detect local variations in DNA shape and electrostatic potential is a general mechanism that enables proteins to use information in the minor groove, which otherwise offers few opportunities for the formation of base-specific hydrogen bonds, to achieve DNA-binding specificity.

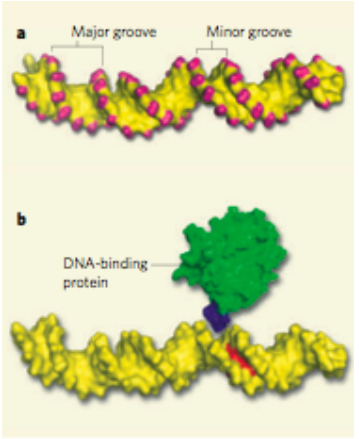


Figure 1 | Getting into the groove. Rohs *et al.*[1] report that the shape of the minor groove of DNA can direct the binding of proteins to specific sites. **a**, Negatively charged phosphate groups (magenta) are arrayed along the outer edge of the DNA major and minor grooves that spiral around the axis of the double helix. The width of the minor groove varies depending on the sequence of nucleotides. This variation leads to differences in the distance between phosphates across the groove, which in turn lead to variation in the negative electrostatic potentia along the minor groove. **b**, A representation of a DNA-binding protein (green) that has a positively charged side chain on its surface, for example arginine (purple), is shown. The protein may recognize a binding site on DNA by its electrostatic potential. The protein is about to bind to the segment of the DNA minor groove that has the optimum groove width and electrostatic potential for binding (red). The DNA structure in **a** and **b** is derived from a structure in the RCSB Protein Data Bank, accession number 2O61. The illustration of a DNA-binding protein in **b** is hypothetical.

« The Electric Slide » (Phys Rev Focus)

Protein shape matters!

Moonlighting Proteins

Multifunctional Proteins

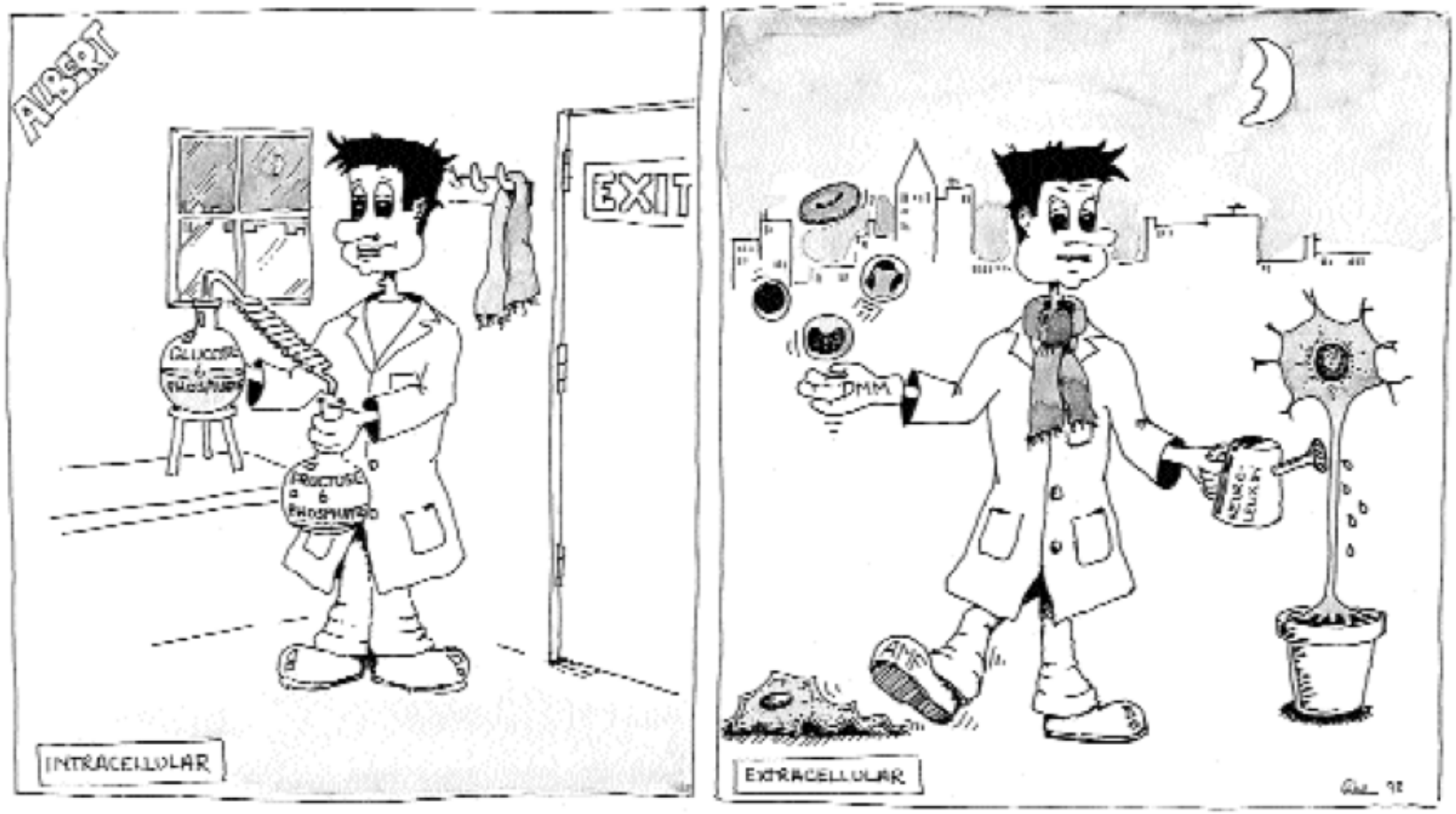
Hub proteins

Much more frequent in eukaryotes

Regulatory proteins

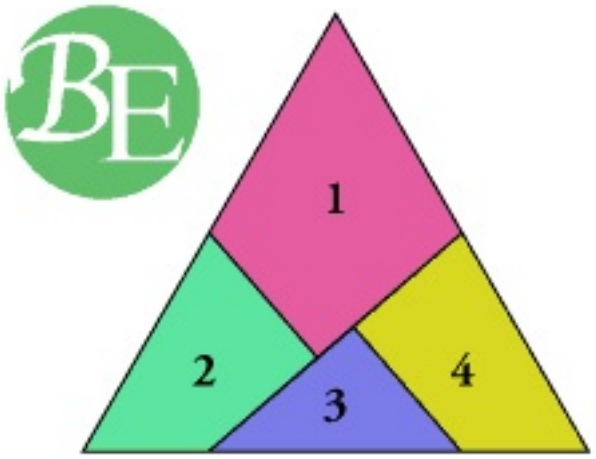
Switching mechanisms

TALKING POINT

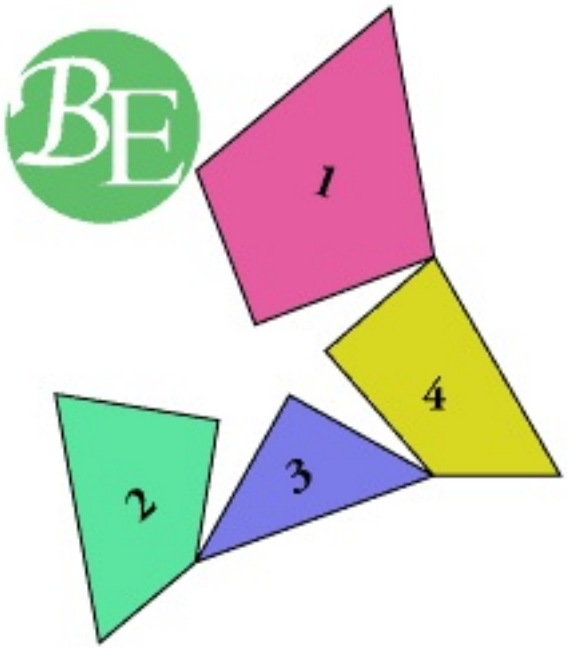


ALBERT GOES MOONLIGHTING AS PHOSPHOGLUCOSE ISOMERASE.

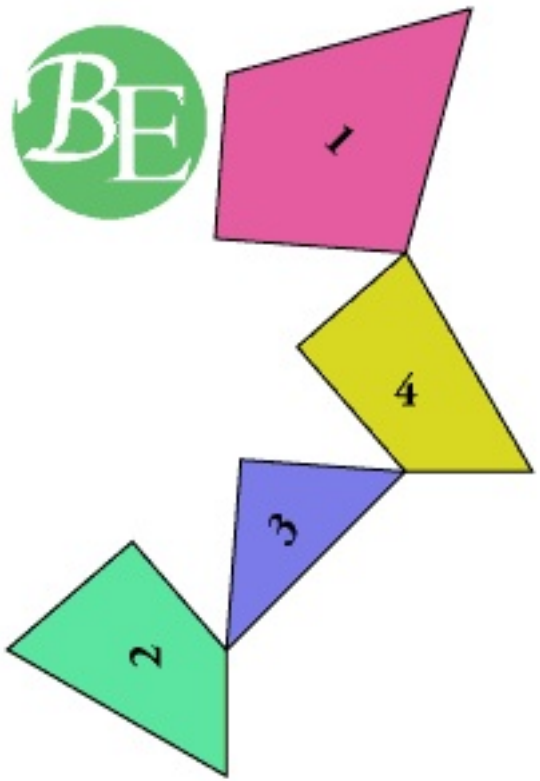
Jigsaw puzzles



The Haberdasher's Puzzle

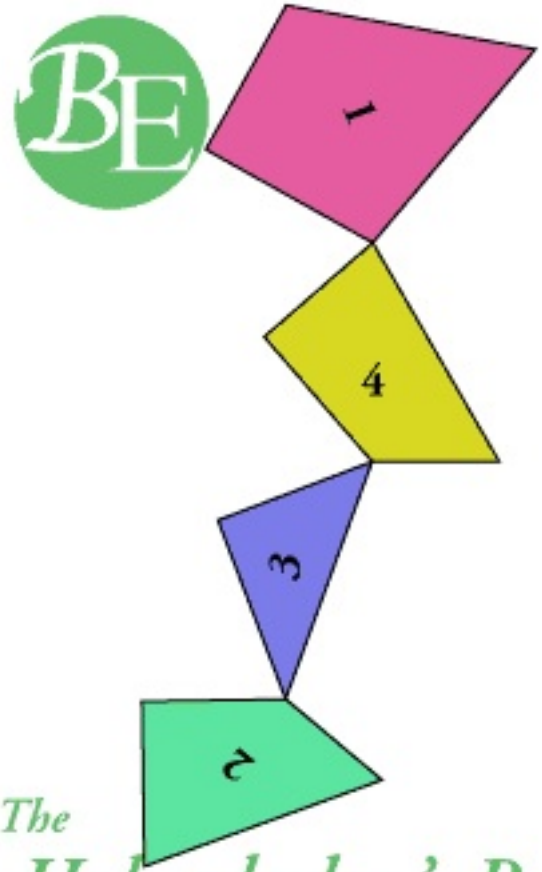


The Haberdasher's Puzzle

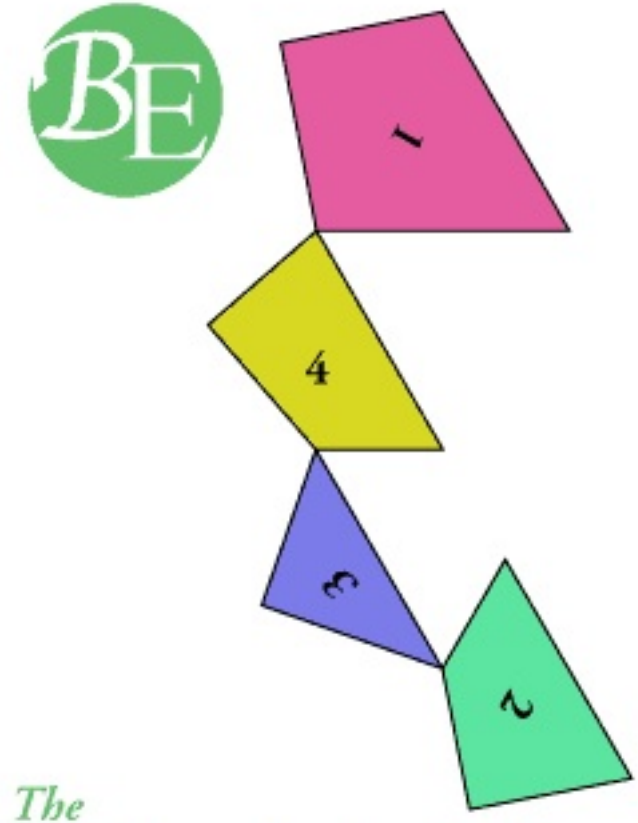


The

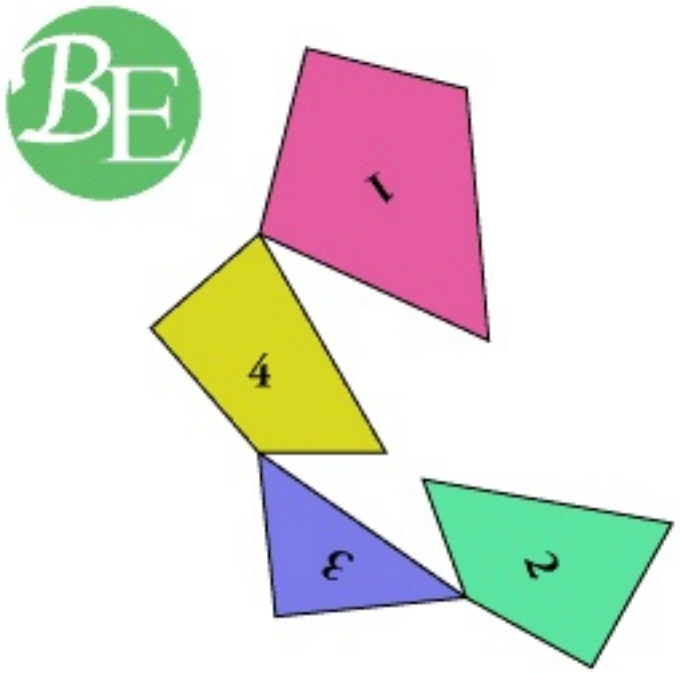
Haberdasher's Puzzle



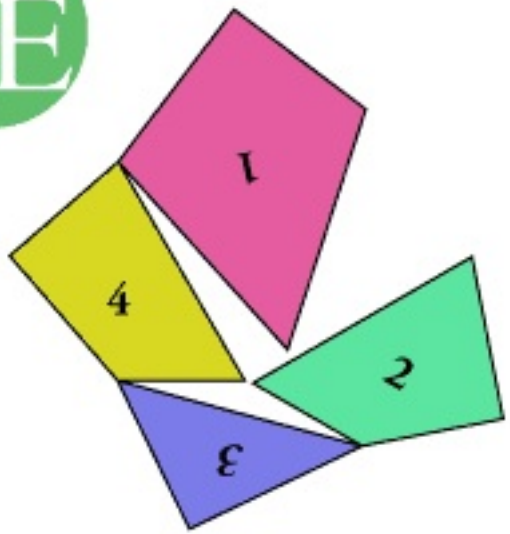
The Haberdasher's Puzzle



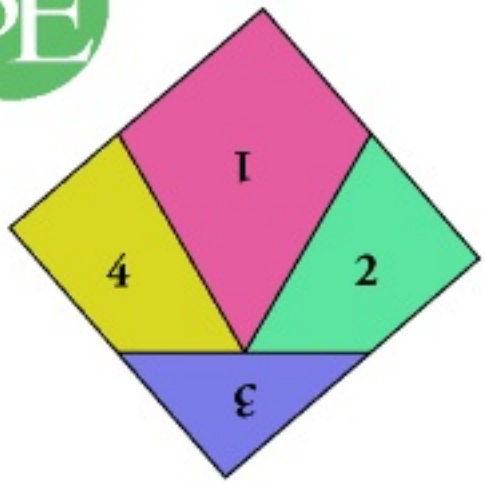
The Haberdasher's Puzzle



The Haberdasher's Puzzle

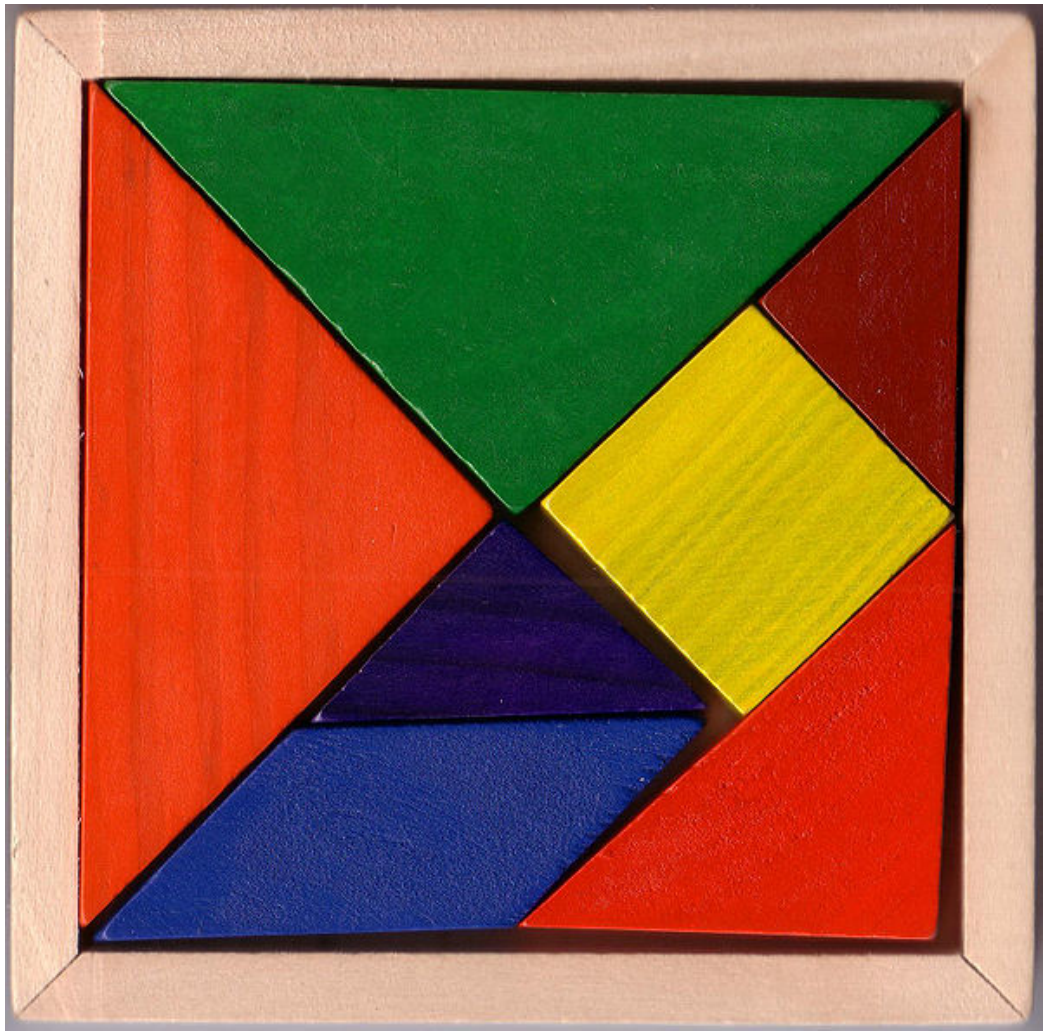


The Haberdasher's Puzzle

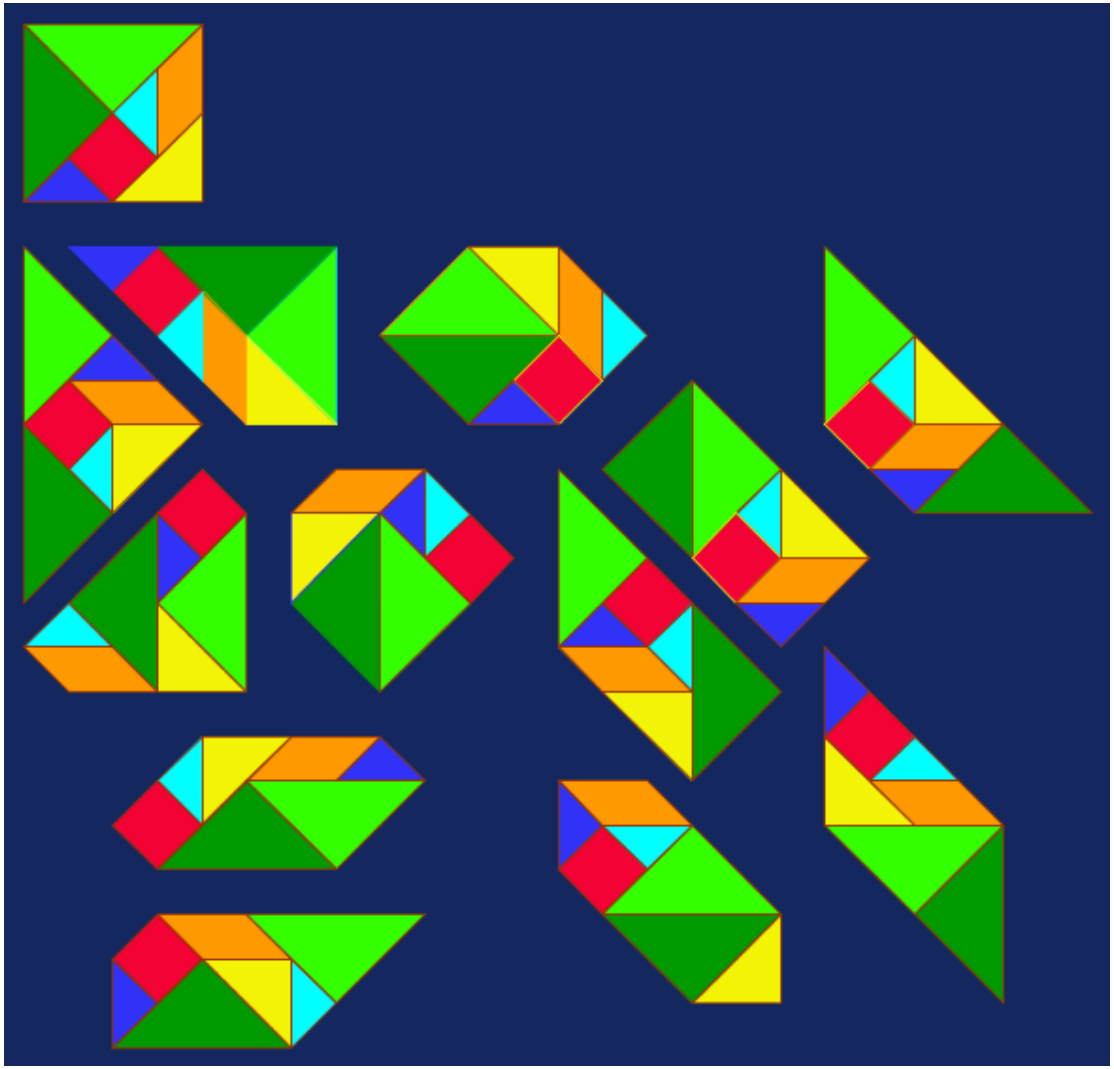


The Haberdasher's Puzzle

The Tangram dissection puzzle

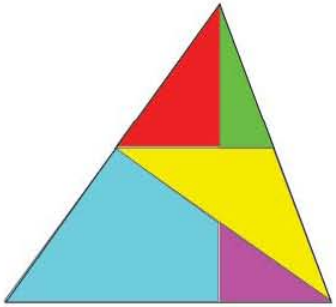


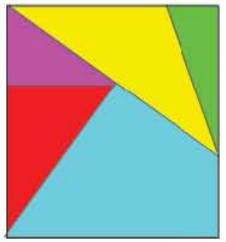
The 13 convex shapes



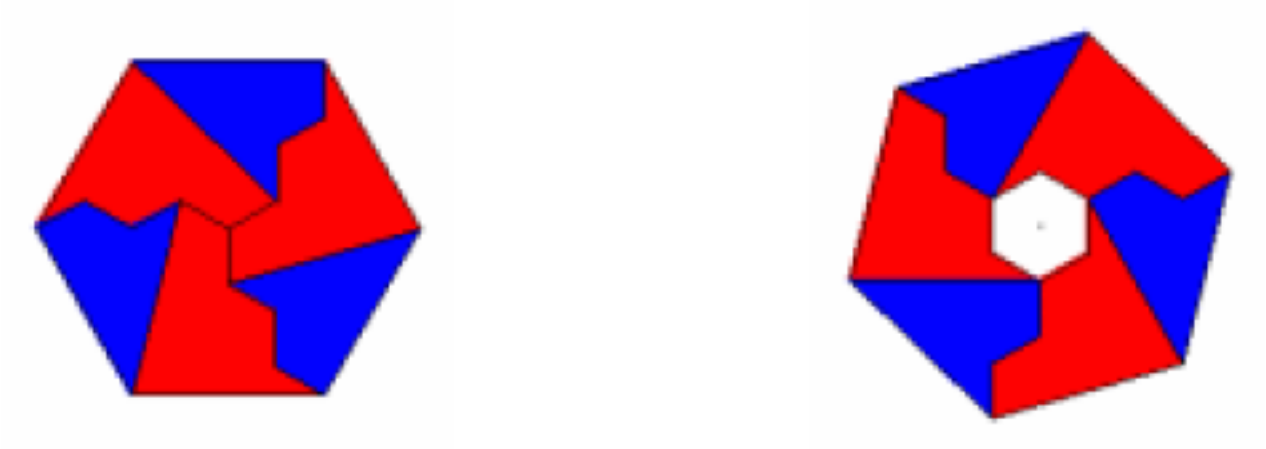
Dissection géométrique

Théorème (Hadwiger et Gur, 1954). – *Deux polygones de même aire se décomposent l'un en l'autre. De plus on peut toujours effectuer le découpage de telle façon que, lors du réassemblage, on ne procède qu'à des translations et des demi-tours.*





An oligomeric Tangram



Et les maladies chroniques?!

Le “paradoxe des jumeaux”

Vol 464|29 April 2010|doi:10.1038/nature08990 nature

LETTERS

Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis

Sergio E. Baranzini[1], Joann Mudge[2], Jennifer C. van Velkinburgh[2], Pouya Khankhanian[1], Irina Khrebtukova[3], Neil A. Miller[2], Lu Zhang[3], Andrew D. Farmer[2], Callum J. Bell[2], Ryan W. Kim[2], Gregory D. May[2], Jimmy E. Woodward[2], Stacy J. Caillier[1], Joseph P. McElroy[1], Refujia Gomez[1], Marcelo J. Pando[4], Leonda E. Clendenen[2], Elena E. Ganusova[2], Faye D. Schilkey[2], Thiruvarangan Ramaraj[2], Omar A. Khan[5], Jim J. Huntley[3], Shujun Luo[3], Pui-yan Kwok[6,7], Thomas D. Wu[8], Gary P. Schroth[3], Jorge R. Oksenberg[1,7], Stephen L. Hauser[1,7] & Stephen F. Kingsmore[2]

Monozygotic or ‘identical’ twins have been widely studied to dissect the relative contributions of genetics and environment in human diseases. In multiple sclerosis (MS), an autoimmune demyelinating disease and common cause of neurodegeneration and disability in young adults, disease discordance in monozygotic twins has been interpreted to indicate environmental importance in its pathogenesis[1–8]. However, genetic and epigenetic differences between monozygotic twins have been described, challenging the accepted experimental model in disambiguating the effects of nature and nurture[9–12]. Here we report the genome sequences of one MS-discordant monozygotic twin pair, and messenger RNA transcriptome and epigenome sequences of $CD4^+$ lymphocytes from three MS-discordant, monozygotic twin pairs. No reproducible differences were detected between co-twins among ~3.6 million single nucleotide polymorphisms (SNPs) or ~0.2 million insertion-deletion polymorphisms. Nor were any reproducible differences observed between siblings of the three twin pairs in HLA haplotypes, confirmed MS-susceptibility SNPs, copy number variations, mRNA and genomic SNP and insertion-deletion genotypes, or the expression of ~19,000 genes in $CD4^+$ T cells. Only 2 to 176 differences in the methylation of ~2 million CpG dinucleotides were detected between siblings of the three twin pairs, in contrast to ~800 methylation differences between T cells of unrelated individuals and several thousand differences between tissues or between normal and cancerous tissues. In the first systematic effort to estimate sequence variation among monozygotic co-twins, we did not find evidence for genetic, epigenetic or transcriptome differences that explained disease discordance. These are the first, to our knowledge, female, twin and autoimmune disease individual genome sequences reported.

We sought to assess the magnitude of genetic, epigenetic and transcriptomic differences in $CD4^+$ lymphocytes from MS-affected and pair 041896 was female, of Ashkenazi Jewish origin and beyond the susceptibility age-range for MS at the time of study (Supplementary Table 1). Twin pair 230178 was female and African-American, whereas twins 041907 were white males. Individual 041896-001 had an onset of MS at age 30 years, and is at present in the secondary progressive phase; individuals 230178-001 and 041907-001 had MS onset at ages 38 and 13, respectively, and have relapsing-remitting disease. Molecular typing of HLA loci showed identical genotypes within the three twin pairs (Supplementary Table 1). Only co-twins 041907 had DRB1*1501, the strongest genetic susceptibility factor for MS[14].

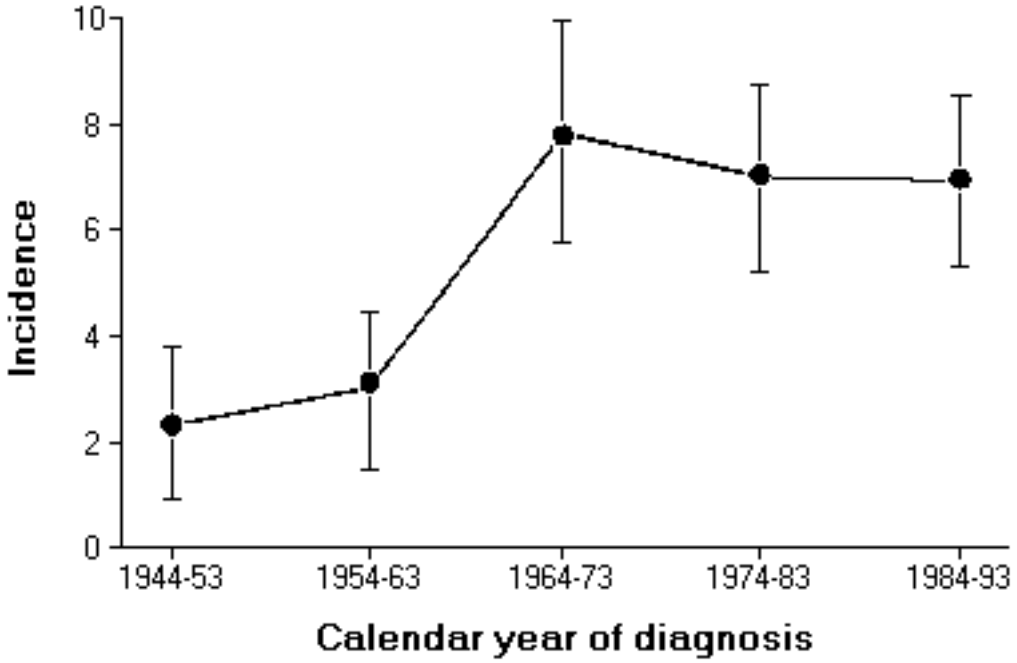
Nucleic acid samples were sequenced by sequencing-by-synthesis with reversible-terminator chemistry[15–18]. mRNA was prepared from blood samples drawn on different days from twin pair 041896 to ascertain sampling variance. A total of 50–68-million, high-quality, 36–44-nucleotide, singleton sequences from each of eight mRNA samples were aligned to the NCBI human genome reference, and read-counts per gene were calculated[18–20] (Supplementary Table 2). Sequencing to this depth (median relative transcript coverage of 5.0-fold and 6.4-fold for 041896-001 and 041896-101, respectively) allowed the determination of the diversity of the polyadenylated transcriptome in $CD4^+$ lymphocytes: ~92% of 20,601 genes with exon annotations were expressed, as assessed by aligned reads and the upper asymptote of the best-fit sigmoid curve (Supplementary Table 2 and Supplementary Fig. 2). The distribution of transcript abundance was a left-skewed, bell-shaped curve with >7 $\log_{10}$ dynamic range (Supplementary Fig. 2), in agreement with a previous study[17]. Digital gene expression values correlated well with exon-resolution array hybridization results (Supplementary Fig. 3), in agreement with another report[21]. Surprisingly, diagnosis or treatment of MS accounted for only 9.4% of variance in transcript abundance in T cells of monozygotic twins, compared with 57.3% being attributable to twin-pair-to-twin-pair differences, 23.6% to day-to-day variation (as assessed in

1. Epidémiologie

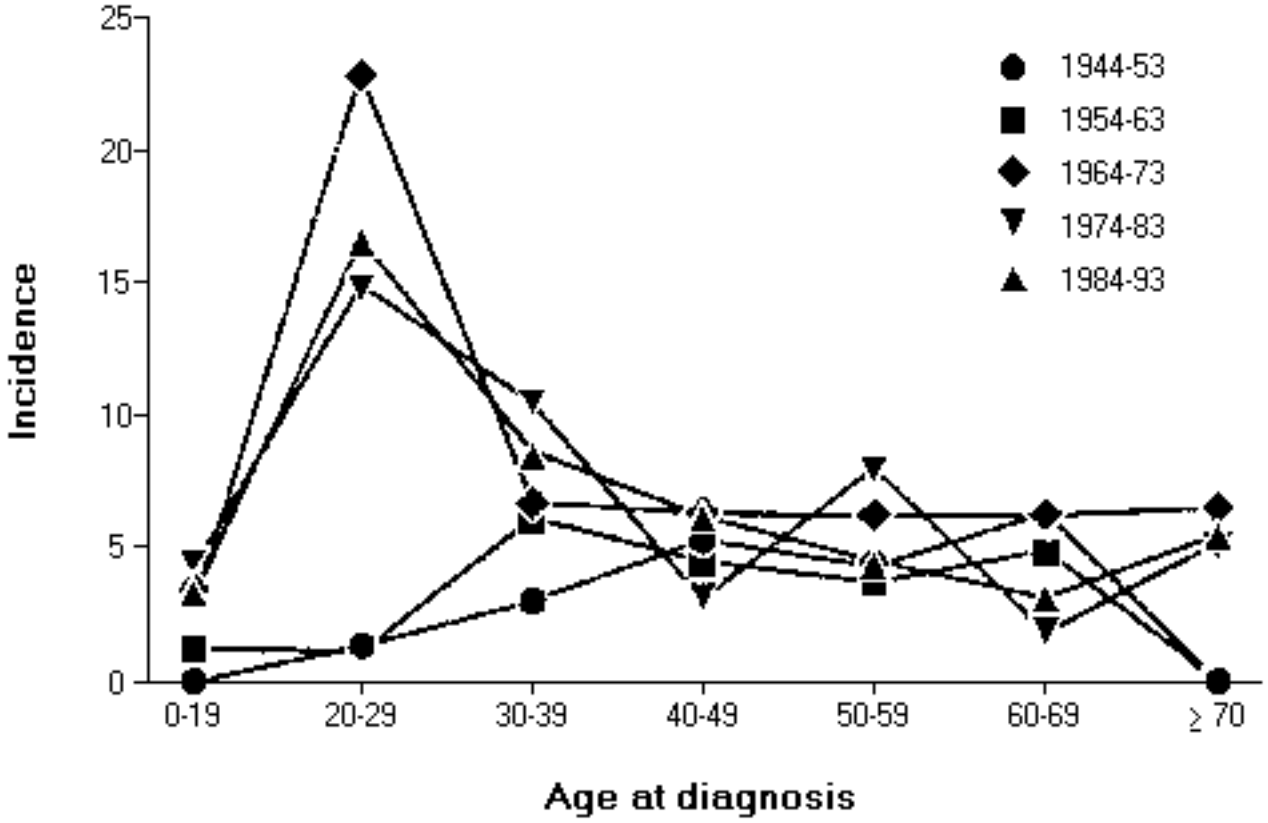
2. Génétique

3. Modèle

Incidence de la maladie de Crohn Olmsted County 1940-1993



Incidence de la maladie de Crohn en fonction de l'âge de début



Pourquoi un délai d’apparition ?

Comment l’environnement agit sur ce délai?

1. Epidémiologie

2. Génétique

3. Modèle

Les variants génétiques à risque

$$\textbf{Risque relatif : RR} = \frac{\textbf{Pr(malade} \mid \textbf{var1)}}{\textbf{Pr(malade} \mid \textbf{var2)}}$$

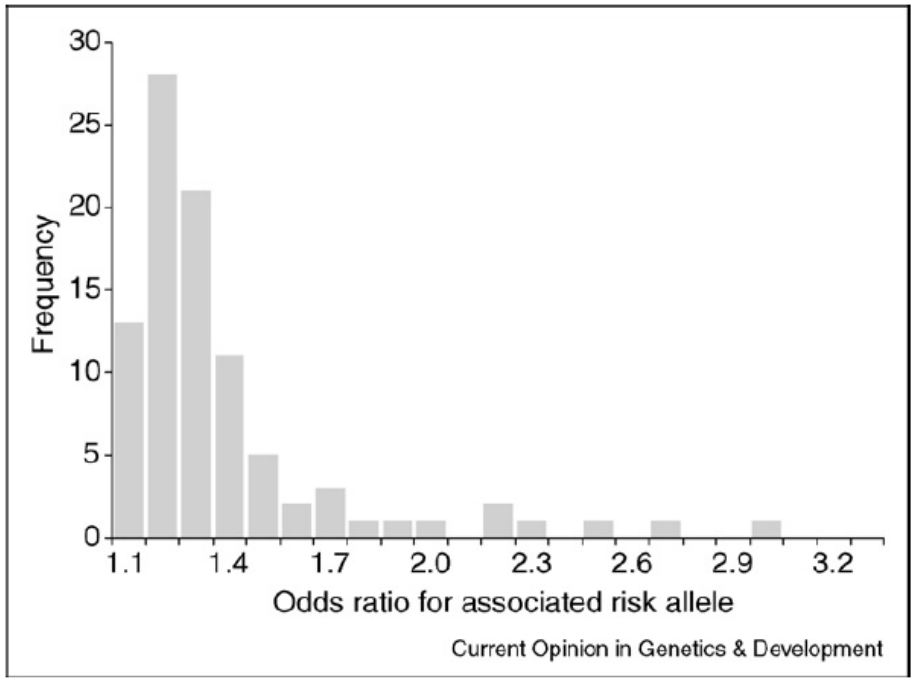
$$\textit{Théorème de Bayes}: \quad P(\textit{malade} \mid \text{var}).P(\text{var}) = P(\text{var} \mid \textit{malade}).P(\textit{malade})$$

$$\textbf{RR} = \frac{\textbf{Pr(var1 dans le groupe de malades)}}{\textbf{Pr(var1 dans le groupe contrôle)}} \Big/ \frac{\textbf{Pr(var2 dans le groupe de malades)}}{\textbf{Pr(var2 dans le groupe contrôle)}}$$

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development

Prediction of individual genetic risk of complex disease

Naomi R Wray[1], Michael E Goddard[2] and Peter M Visscher[1]

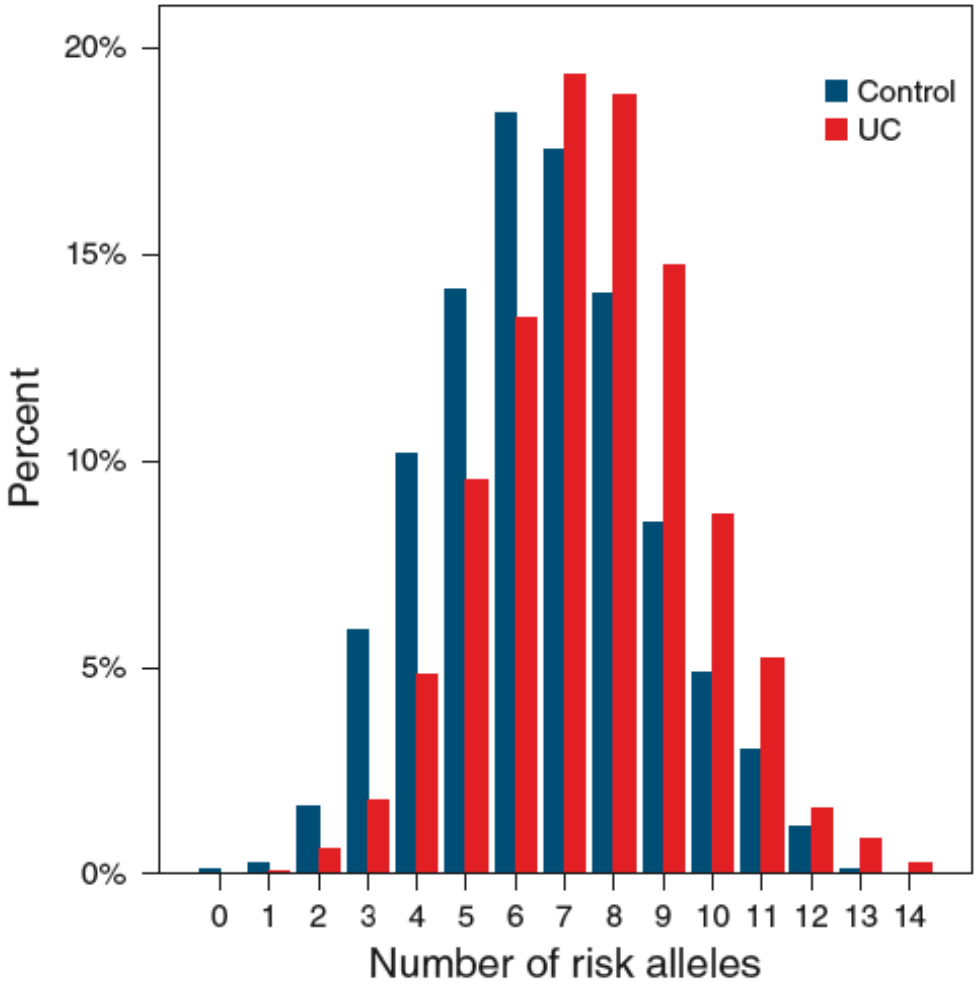


Frequency distribution of effect sizes expressed as Odds ratio for the risk allele of 92 validated associated SNPs identified from GWAS. These SNPs represent associations with one of 16 disorders (listed in Appendix A). The power of the GWAS to detect variants with effect size of 1.1 or smaller was low.

Appendix A

References for the 16 disorders used in Figure 2:

1. Abdominal aortic aneurysm [43].
2. Age-related macular degeneration [44].
3. Amyotrophic lateral sclerosis [45].
4. Ankylosing spondylitis [46].
5. Asthma [47].
6. Autoimmune thyroid disease [46].
7. Breast cancer [48].
8. Childhood obesity [49].
9. Colorectal cancer [50].
10. Coronary heart disease [51,52].
11. Crohn's disease [53–56].
12. Prostate cancer [57,58].
13. Rheumatoid arthritis [59,60].
14. Systemic lupus erythematosus [61–63].
15. Type 1 diabetes [64,65].
16. Type 2 diabetes [28].



Am J Gastroenterol 2010; 105:395–402; doi:10.1038/ajg.2009.576; published online 27 October 2009

Le risque relatif en questions

1. Pas de bons ni de mauvais variants, mais des variants adaptés à leur environnement (Evolution)

2. Risques absolus (très) faibles et (presque) toujours comparables

1. Epidémiologie

2. Génétique

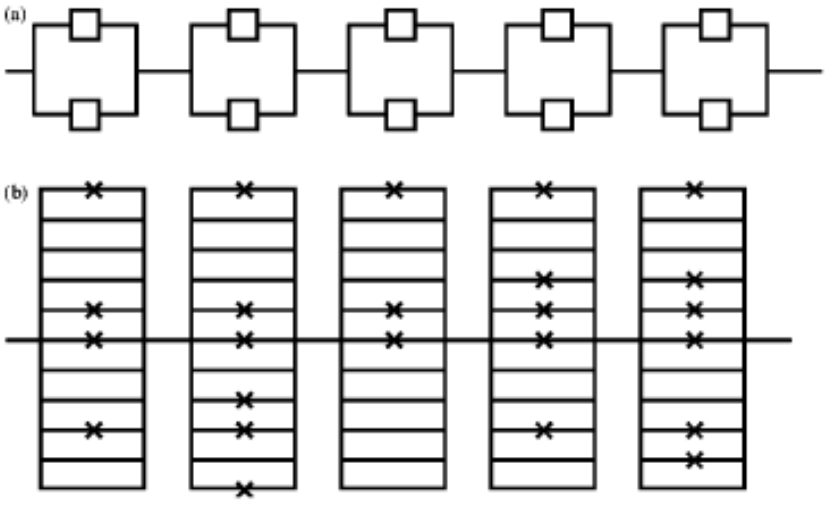
3. Modèle

J. theor. Biol. (2001) **213**, 527–545
doi:10.1006/jtbi.2001.2430, available online at http://www.idealibrary.com on IDEAL®

The Reliability Theory of Aging and Longevity

Leonid A. Gavrilov* and Natalia S. Gavrilova

Center on Aging, NORC/University of Chicago, 1155 *East 60th Street, Chicago, IL* 60637, *U.S.A.*



$$F_H(n,k,x) = P(X \leqslant x) = (1 - e^{-kx})^n.$$

Une collaboration interdisciplinaire

Hôpital Robert Debré: Camille Jung
Jean-Pierre Hugot

CEPH (Jean Dausset): Leigh Pascoe

PECSA (CNRS, UPMC): Vincent Dahirel
Marie Jardat

LPTMC (CNRS, UPMC): Maria Barbi
Gaëlle Debret
Annick Lesne
Fabien Paillusson
Jean-Marc Victor