



**HAL**  
open science

# LA FABRIQUE DES ORGANISMES

Jean Francois Daniel Nicolas

► **To cite this version:**

Jean Francois Daniel Nicolas. LA FABRIQUE DES ORGANISMES : Le Développement embryonnaire vu par la Génétique moléculaire et l'Évolution. 2020. hal-02890738v1

**HAL Id: hal-02890738**

**<https://hal.science/hal-02890738v1>**

Preprint submitted on 6 Jul 2020 (v1), last revised 23 Jul 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

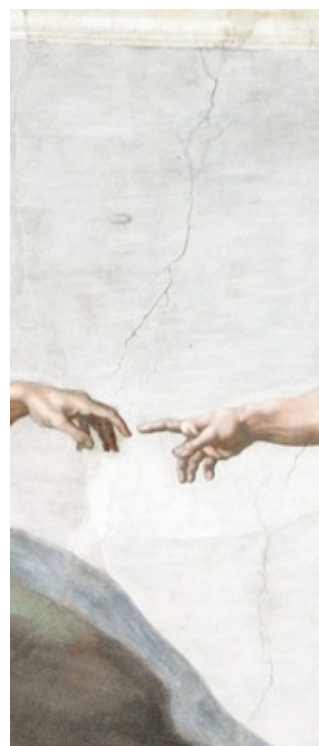
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JEAN-FRANÇOIS  
NICOLAS

# LA FABRIQUE DES ORGANISMES

Le Développement embryonnaire vu par la Génétique  
moléculaire et l'Évolution

---



JEAN-FRANÇOIS NICOLAS

# LA FABRIQUE DES ORGANISMES

Le Développement embryonnaire vu par la Génétique  
moléculaire et l'Évolution

***Illustration de couverture : La création, Michel-Ange, ici symbolisant l'importance des interactions (entre cellules dans l'embryon, entre molécules dans les cellules et les réseaux de régulations) dans l'émergence des propriétés biologiques et la complexification des organismes.***



JEAN-FRANÇOIS NICOLAS

# LA FABRIQUE DES ORGANISMES

# LA FABRIQUE DES ORGANISMES

**Quand en 1995, Walter Gehring obtient des yeux sur les pattes d'une mouche en forçant l'expression d'un seul gène dans quelques cellules de l'embryon, notre compréhension de la logique du développement et des moteurs de l'évolution est bouleversée. Plus généralement, avec la révolution de la génétique moléculaire, en particulier la possibilité de cloner et de réarranger les gènes, de séquencer et de comparer les génomes entiers, les fondements moléculaires et cellulaires du développement, sans doute le plus complexe de tous les processus connus dans notre univers, ont été, ces trente dernières années, élucidés. Une des grandes surprises fut de trouver que les développements d'organismes aussi différents que la mouche et l'homme sont organisés à partir de mêmes gènes qui étaient déjà présents chez leur ancêtre commun (qui n'était ni un insecte ni un vertébré), il y a plus d'un demi-milliard d'années. Dans l'évolution la clé de l'émergence des développements fut l'apparition des communications entre cellules. Elles sont à l'origine de la complexification des embryons et des réseaux de régulation des gènes desquels émergent les propriétés des cellules et des groupes de cellules utilisées par les développements. En retour, l'analyse des processus du développement a renouvelé la génétique et notre conception du génome. Ainsi, on ne considère plus actuellement ni les gènes ni les protéines individuellement mais gènes et protéines dans les systèmes, les réseaux d'interactions et de régulation qu'ils constituent. Tout cela permet de comprendre comment on envisage actuellement la fabrication des organismes et comment cela éclaire la nature, et du développement et de l'évolution.**



**Biological research is in crisis, .... Although many believe that ‘more is better’, history tells us that ‘least is best’.**  
**Sydney Brenner, 2012<sup>1</sup>.**

# Contenu

## La fabrique des organismes

- 1 Les trois problèmes majeurs du développement
- 2 La différenciation des types cellulaires
- 3 La formation des organes
- 4 Le positionnement des organes
- 5 La fabrique d'un organisme
- 6 Conclusion

## Annexes

- 1 Le crible génétique des mutants létaux embryonnaires chez la drosophile
- 2 L'activation distale de l'expression des gènes.
- 3 Les réseaux génétiques de régulation
- 4 Les communications entre cellules par les voies de signalisation
- 5 Une brève comparaison des génomes des choanoflagellés aux néphrozoaires
- 6 Les legs de notre ancêtre unicellulaire
- 7 L'origine de l'œil ou la naissance par chance d'une fonction.
- 8 La place du développement dans le monde du vivant





Comprend-t-on aujourd'hui le développement ? Qu'est-ce qui ne fait plus débats ? Reste-t-il des pans de ce processus dont on sait que l'on ne sait rien ? A quel degré de précisions en est-on ? Est-on proche d'une modélisation ?

Si comprendre c'est être capable de reproduire le processus, alors nous sommes loin de comprendre le développement. Le degré de précisions que l'on a du développement est très faible. Une modélisation du développement du plus simple des animaux n'est pas envisageable.

Mais si comprendre c'est avoir le sentiment que l'on n'aura pas à rajouter d'ingrédients majeurs à ceux que l'on connaît lorsqu'une modélisation d'un développement sera tentée, alors oui l'on a le sentiment de comprendre les grandes lignes du développement.

Il y a quarante ans (1977) presque qu'aucun de ces ingrédients n'étaient connus. Aucune des molécules impliquées n'était identifiée ; ni le rôle du génome ni celui des interactions cellulaires n'était compris ; l'unité génétique des développements de tous les animaux n'était pas soupçonnée. Actuellement, on a l'impression que l'on peut proposer des solutions à presque tous les problèmes qui se posent, même si ces solutions se limitent aux principes et ne peuvent préciser avec certitude ni quels gènes, ni quels types de communications entre cellules, ni quelles séquences d'événements sont impliquées.

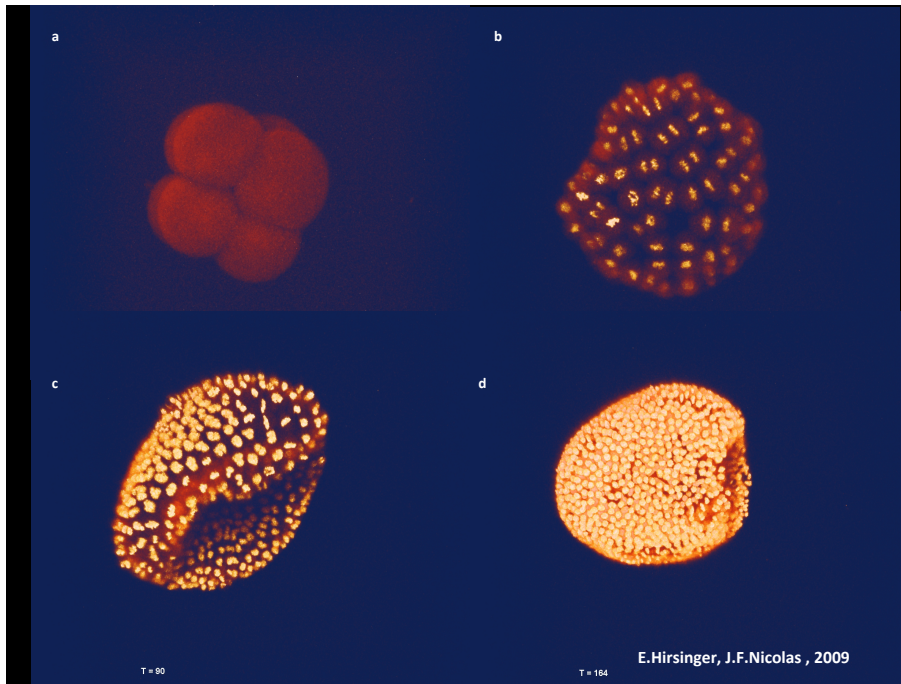
Ce qui est rapporté ici ce sont ces principes qui ne font plus débats<sup>1</sup>.

---

<sup>1</sup> Cet ouvrage, rédigé en 2018-19, est basé sur des conférences données en contexte scientifique pour des laboratoires de recherche, écoles doctorales et pour des associations de diffusion de la science depuis 2010. Il témoigne du cheminement des idées des biologistes « moléculaires » sur le développement et l'évolution au cours des dernières quarante années.

# I. LES TROIS PROBLÈMES DU DÉVELOPPEMENT

Les développements se fondent sur des opérations cellulaires qui utilisent des comportements de base des cellules. Les développements sont très différents les uns des autres, mais ils aboutissent à des organismes qui partagent beaucoup de traits. Tous les animaux actuels sont reliés entre eux par leurs ancêtres communs. Si c'est la raison de ces ressemblances, alors l'évolution est le fil conducteur pour comprendre les développements. Les organismes sont des ensembles d'organes. Les organes sont des ensembles de cellules, presque toujours de plusieurs types. Durant l'évolution des types cellulaires sont présents dès les premiers animaux. Les organes ne deviennent prévalents que chez les bilatériens. Les trois opérations fondamentales du développement sont la différenciation des types cellulaires, la formation des organes, et le positionnement des organes dans l'organisme. En étudiant ces opérations nous découvrirons qu'ils ont les mêmes exigences et se déduisent les uns des autres. Cela jette une vive lumière sur la nature, et du développement et de l'évolution.



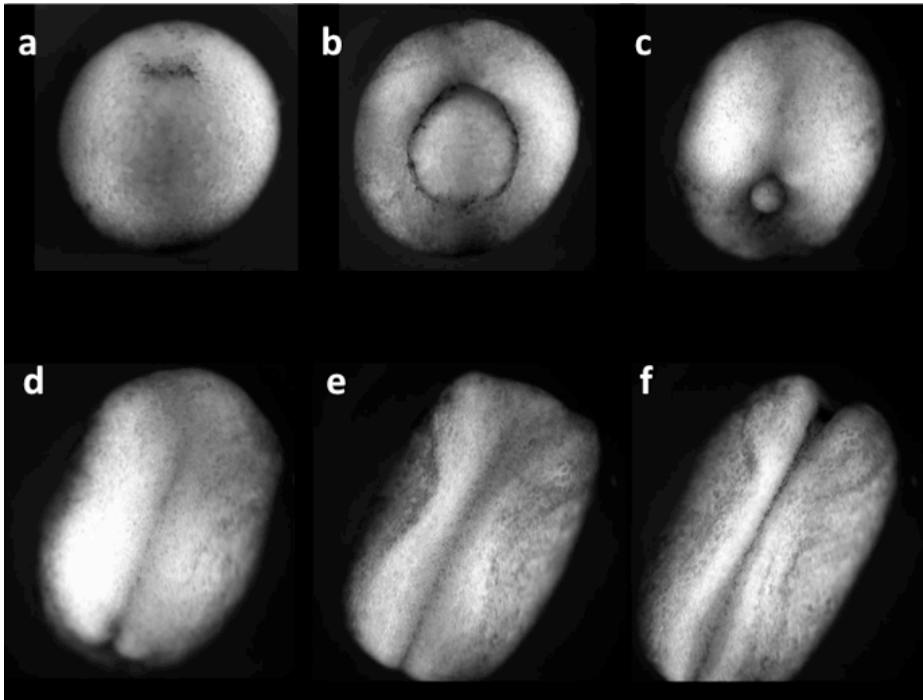
La fabrique des organismes se fait durant l'embryogenèse, la période de la vie des organismes qui démarre après la fusion des gamètes males et femelles (l'œuf fécondé) et se termine avec l'apparition des organes. Les technologies

d'imagerie actuelles permettent de filmer en vivant cette période. Examinons donc de cette manière les développements de quatre organismes très étudiés dans les laboratoires.

**Figure 1 : les 12 premières heures du développement de l'amphioxus.** a) le stade 8 cellules ; b) le stade 128-254 cellules (beaucoup des cellules sont en division) ; c) début de l'invagination ; d) fin de l'invagination, à droite l'ouverture de la cavité interne. Référence du film Estelle Hirsinger et al.

Suivons les 12 premières heures du développement du représentant actuel de l'ancêtre de tous les vertébrés, l'amphioxus (Figure 1). Le développement démarre d'une cellule unique, qui se divise (Figure 1a) et donne rapidement 1000 cellules qui forment une sphère creuse (Figure 1b, blastula) dont un hémisphère s'enfonce dans l'autre (Figure 1c, invagination). La cavité interne ainsi formée se referme incomplètement tandis que l'embryon s'allonge (Figure 1d). L'invagination met en contact les cellules des deux hémisphères de l'embryon. Il y a un feuillet externe et un feuillet interne. C'est la gastrulation une opération générale des développements. Le développement se poursuit après ce stade mais n'est plus facilement filmable<sup>2</sup>.

Suivons maintenant la deuxième phase de la gastrulation chez un vertébré, la grenouille (Figure 2). Les cellules s'internalisent comme chez

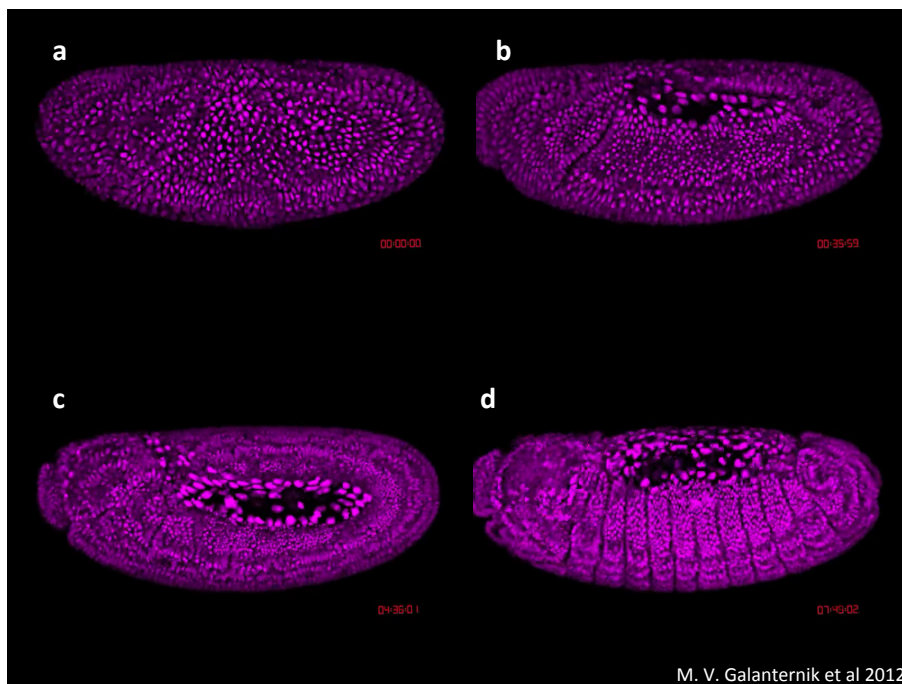


l'amphioxus mais avec une modalité différente et l'embryon s'allonge (Figure 2a-c). Puis le feuillet externe se transforme, les cellules convergent symétriquement et forment le tube neural qui se referme complètement (Figure 2d-f). Ce tube donnera la

moelle épinière et le cerveau.

**Figure 2 : la gastrulation et la neurulation chez la grenouille (*Xenopus laevis*).** a-c) internalisation des cellules ; d-f) formation du tube neural <sup>3</sup>. Référence du film [https://www.youtube.com/watch?v=ZEGvhkoVu\\_g](https://www.youtube.com/watch?v=ZEGvhkoVu_g)

Voici maintenant le développement d'un insecte, la mouche drosophile entre 3h et 10 heures après la fécondation de l'œuf, d'emblée oblongue (Figure 3). A 3h il y a déjà 6000 cellules (Figure 3a, blastoderme).

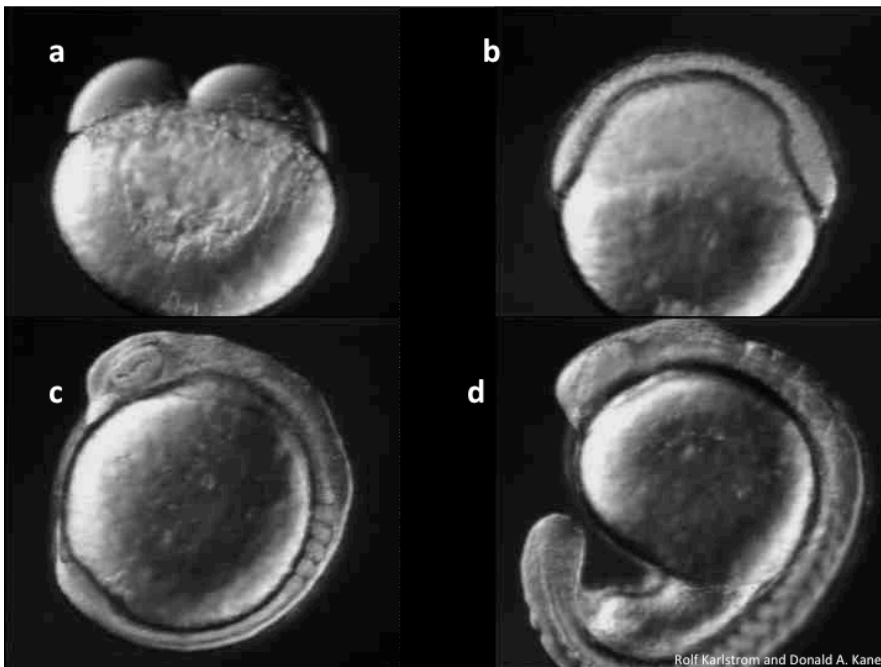


apparaissent des éléments répétés, appelés segments, une partie rostrale la future tête formée de plusieurs segments, une partie caudale (Figure 3d). C'est la structure de la première larve à la fin de gastrulation.

**Figure 3 : la**

**gastrulation et la formation des segments chez la mouche drosophile.** a, b) gastrulation (sillon céphalique à gauche) ; c, d) formation des segments<sup>3</sup>. Référence du film, Marina Venero Galanternik et al., 2012, <https://www.youtube.com/watch?v=SDMFsKrUCvA>

Enfin les 17 premières heures du développement du poisson zébré (Figure 4). L'œuf est beaucoup plus gros. Une partie seulement se divise dans une calotte (Figure 4a). Les quelques milliers de cellules convergent (Figure 4b) et l'embryon s'allonge (Figure 4c et d) tandis qu'apparaissent la tête, les yeux, les muscles, la partie caudale, d'autres organes non visibles. En 15 heures l'organisme est quasiment formé.



**Figure 4 : les 17 premières heures du développement d'un poisson (*Danio rerio*).** a) stade 2 cellules ; b) convergence ; c, d) formation des organes.<sup>3</sup>. Référence du film, Rolf Karlstrom and Donald A. Kane <https://www.youtube.com/watch?v=hSMgt5gV-8s>

Toutes ces actions sont les opérations cellulaires du développement. Elles sont multiples et sont basées sur les comportements des cellules. Les principaux comportements des cellules étant : les divisions (orientées, asymétriques, inégales, ou non), leur rythme ; la mobilité des cellules (adhésivité) ; les mouvements, (dispersion dans une ou toutes les directions à des vitesses variables ; leur changement de forme.

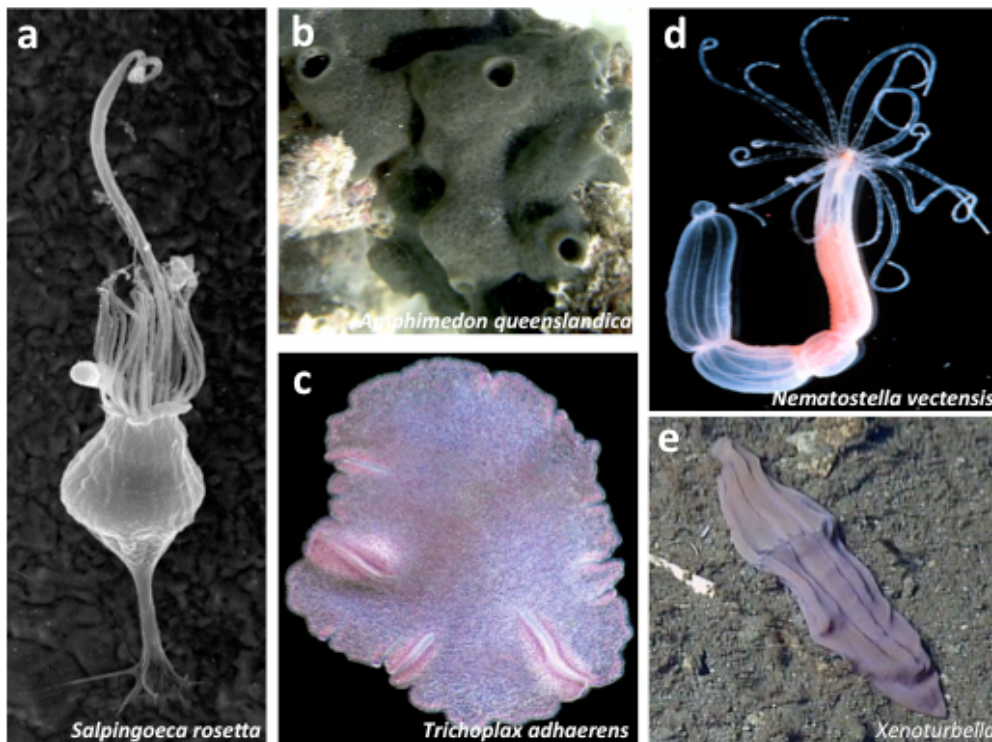
Ces actions aboutissent à des structures, comme la sphère formée de deux couches de cellules et une ouverture au début du développement chez l'amphioxus (Figure 1d) ; comme la plaque neurale puis un tube neural chez la grenouille, à la fin de ce qu'on appelle la neurulation (Figure 2d) ;

comme les segments chez la larve de la drosophile (Figure 3d) ; comme toutes sortes d'organes complexes chez le poisson (Figure 4d). Ces opérations cellulaires ont été décrites en grands détails par les embryologistes. Elles sont réexaminées actuellement avec l'analyse clonale et l'imagerie 4D et de nouveaux aspects sont découverts.

Ces développements sont extrêmement différents les uns des autres, difficilement comparables. Étonnamment, ils aboutissent à des animaux qui partagent beaucoup de traits communs. La souris comme la drosophile et le poisson présentent un dos un ventre, un avant un arrière, un côté droit et un côté gauche (une symétrie bilatérale, ce sont des « bilatériens »). Ils exhibent des organes qui se ressemblent, têtes, yeux, pattes/nageoires, auxquels une fonction est associée : la vue, le mouvement etc. Ces ressemblances ne sont pas anecdotiques.

Comment se fait-il alors qu'avec des débuts si différents ces développements convergent vers les mêmes formes ? Les évolutionnistes nous disent depuis le 19ème siècle que ces animaux sont reliés, qu'ils ont des ancêtres communs mais forts lointains. Si nous suivions les ancêtres communs entre la drosophile et la souris, nous remonterions d'un organisme unicellulaire jusqu'à un bilatérien. Quelques-uns de leurs représentants actuels (Figure 5) sont, pour l'organisme unicellulaire, un choanoflagellé (*Salpingoeca rosetta* Figure 5a). Pour les animaux multicellulaires, les éponges (*Amphimedon queenslandica*, Figure 5b), puis des animaux sans aucune symétrie comme les placozoaires (*Trichoplax adhaerens* Figure 5c) puis les hydres (*Nematostella vectensis*, Figure 5d) et enfin un bilatérien, un Xenacoelomorphe, une sorte de vers avec un tube digestif aveugle c'est-à-dire qui ne présente qu'une seule ouverture (*Xenoturbella*, Figure 5e).

Vous, moi, de tous ces animaux, des éponges, de ce bilatérien bizarre, nous en dérivons.



**Figure 5 :**  
**Quelques uns des représentants actuels de nos lointains ancêtres.** a) du dernier ancêtre unicellulaire, un choanoflagellé, *Salpingoeca rosetta* ; b) du plus ancien ancêtre multicellulaire, une éponge, *Amphimedon queenslandica* ; c) du premier

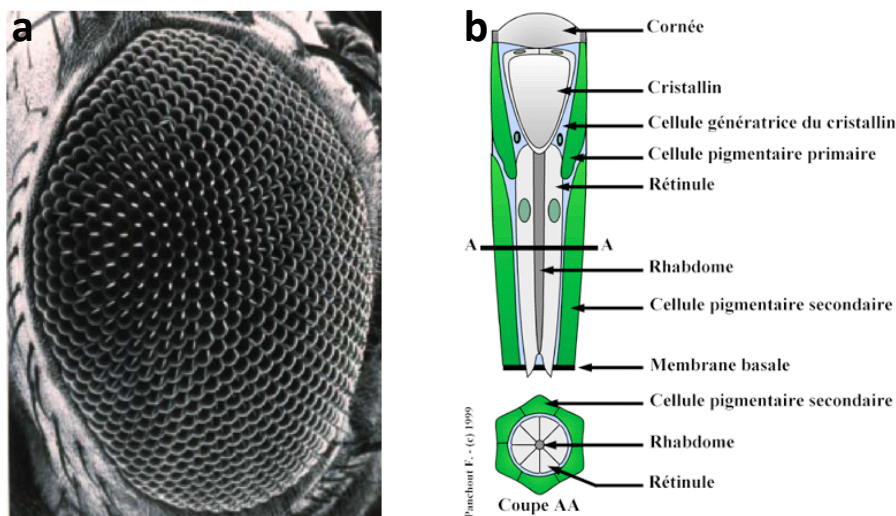
ancêtre présentant une spécialisation cellulaire significative, *Trichoplax adhaerens*, un placozoaire ; d) du premier ancêtre effectuant la gastrulation (encore sans véritables organes), une anémone de mer, *Nematostella vectensis* (cnidaire) ; et enfin e) du premier ancêtre bilatérien, *Xenoturbella* (sans organes). Pour leur position dans l'arbre du vivant voir la Figure 7.

Ces lointains ancêtres communs sont-ils la raison de ces ressemblances ? Si oui, l'évolution est un extraordinaire fil conducteur qu'il faut suivre pour comprendre le développement. C'est l'option de ce livre.

Les développements ont l'air de tendre vers la formation des organes, qui caractérise les organismes et sont liés à une fonction. Dans notre perspective la première question est donc, de quoi les organes sont-ils constitués ?

Prenons l'exemple de l'œil de la drosophile (Figure 6). Il est formé de plusieurs centaines d'unités, les ommatidies (Figure 6a). Mais ces ommatidies ne sont pas l'unité ultime. Chaque ommatidie est constituée de cellules. Des cellules photoréceptrices qui captent des photons, des cellules pigmentées qui forment une chambre noire et d'autres cellules encore qui constituent un cristallin.

Qu'est ce qu'un Organe ?



**Figure 6 : Les organes, des ensembles de cellules.** a) l'œil de la drosophile est formé de 700 ommatidies ; b) une ommatidie, formée d'une dizaine de types cellulaires.

Les organes, aussi complexes soient-ils, sont des ensembles de cellules, presque

toujours de plusieurs types. A ces types cellulaires et aux organes sont associées des fonctions.

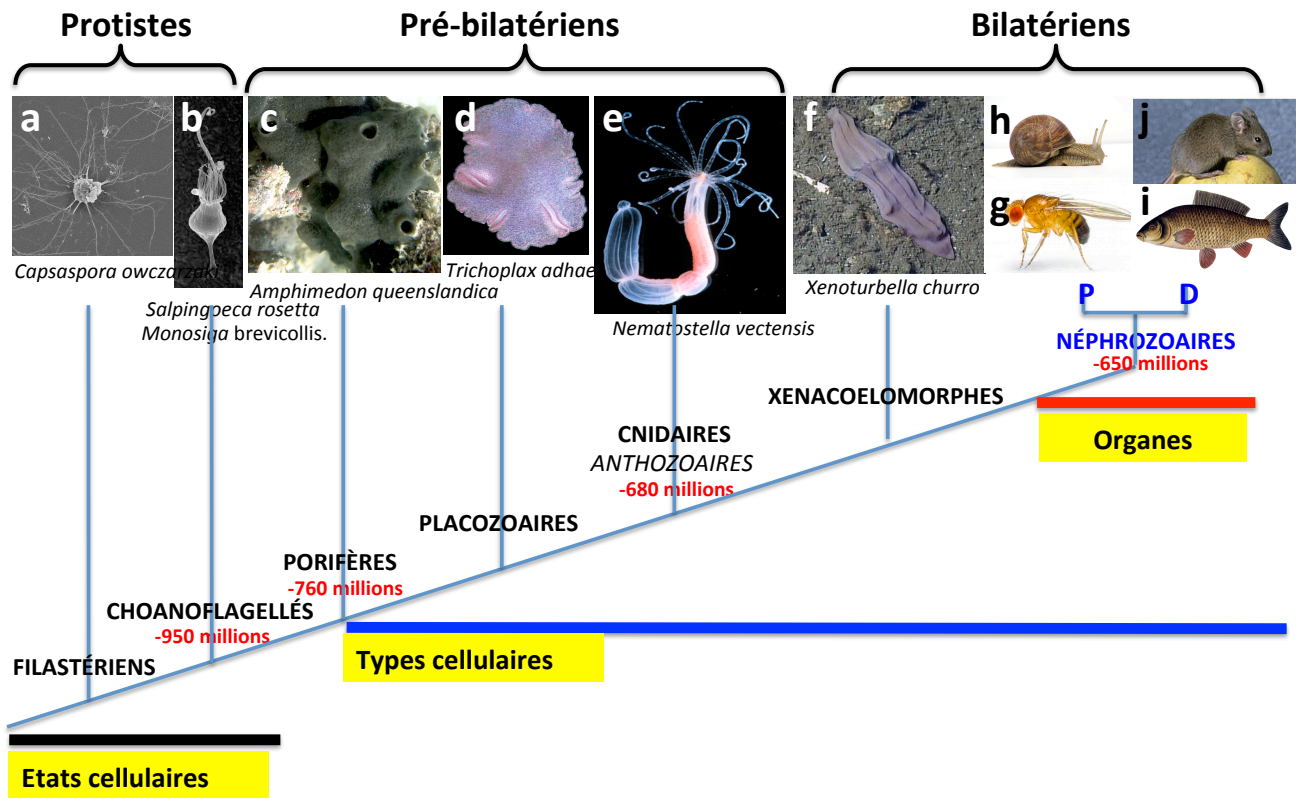
En fait, les organismes sont des ensembles d'organes. Tout dans un organisme est organe.

Mais alors quand dans l'évolution apparaissent ces organes et les types cellulaires ?

Appuyons-nous sur les dernières classifications du vivant qui intègrent les comparaisons des séquences d'ADN et de leurs arrangements, des génomes entiers de plus de 80 animaux<sup>4</sup>.

Les types cellulaires apparaissent chez les premiers animaux (Figure 7, la ligne bleue). Chez les animaux les plus simples, on trouve des types cellulaires mais pas vraiment d'organes. C'est le cas chez *Trichoplax adhaerens*<sup>5</sup>, un placozoaire (sans avant ni arrière, sans directionnalité dans ses mouvements, avec pour seul élément saillant un bord qui délimite un dessus et un dessous) ou chez les éponges, les cnidaires (hydres, méduses) et même chez les plus simples des animaux à symétrie bilatérale comme les xenacoelomorphes<sup>6</sup>.





**Figure 7 : D'où viennent les types cellulaires et les organes ?** Position dans l'arbre du vivant de quelques-uns de nos ancêtres depuis les protistes (filastériens et choanoflagellés, a et b) jusqu'aux bilatériens complexes, les néphrozoaires (P, protostomiens g et h et D, deutérostomiens, i et j). Leurs représentants actuels sont figurés en haut (a-j). La ligne bleue céruléen oblique figure le temps ; sous cette ligne sont indiqués les dates d'apparition de quelque uns des groupes d'animaux mentionnés. La ligne noire marque l'existence d'états cellulaires (chez les protistes, voir l'annexe 5), la ligne bleue l'existence de types cellulaires différenciés (chez les éponges) et la ligne rouge celle d'organes (chez les néphrozoaires).

Les organes ne deviennent prévalents que plus tard chez les bilatériens que l'on appelle aujourd'hui les néphrozoaires (Figure 7, la ligne rouge). L'amphioxus, les vertébrés, les insectes, les mollusques et bien d'autres sont des néphrozoaires.

Nous suivrons donc l'ordre d'apparition des grands processus que l'évolution nous indique. Nous discuterons de la manière dont se fait la différenciation des types cellulaires qui apparaissent en premier dans l'évolution (-760 millions d'années environ), puis de la formation des organes, qui émergent en second (-650 millions d'années environ) et enfin

du positionnement des organes dans l'organisme. Nous verrons que la formation et le positionnement des organes sont indépendants et nous apprendrons que ces trois processus fondamentaux peuvent être déduits les uns des autres car ils utilisent exactement les mêmes éléments.

Étonnamment cet ordre est inversé dans le développement : le positionnement des organes en premier puis leur formation et en dernier les types cellulaires. Cela donne l'étrange impression que l'embryon anticipe l'organisation de l'organisme. Nous verrons aussi comment ce qui fut jusqu'à récemment un mystère se résout aujourd'hui.

### **L'essentiel**

**Tous les animaux sont reliés entre eux par leurs ancêtres communs et tous les animaux dérivent du même unicellulaire. Les organismes les plus simples sont formés de quelques types cellulaires. Les organismes complexes sont formés d'organes. Et les organes sont formés de types cellulaires. Chez la plupart des organismes complexes, les organes présentent des dispositions similaires. Les développements semblent très différents les uns des autres. L'analyse au niveau moléculaire des trois opérations majeures du développement, dans l'ordre d'apparition dans l'évolution, va nous permettre, non seulement de comprendre le développement mais aussi comment les organismes ont évolués. C'est l'objet des chapitres suivants.**

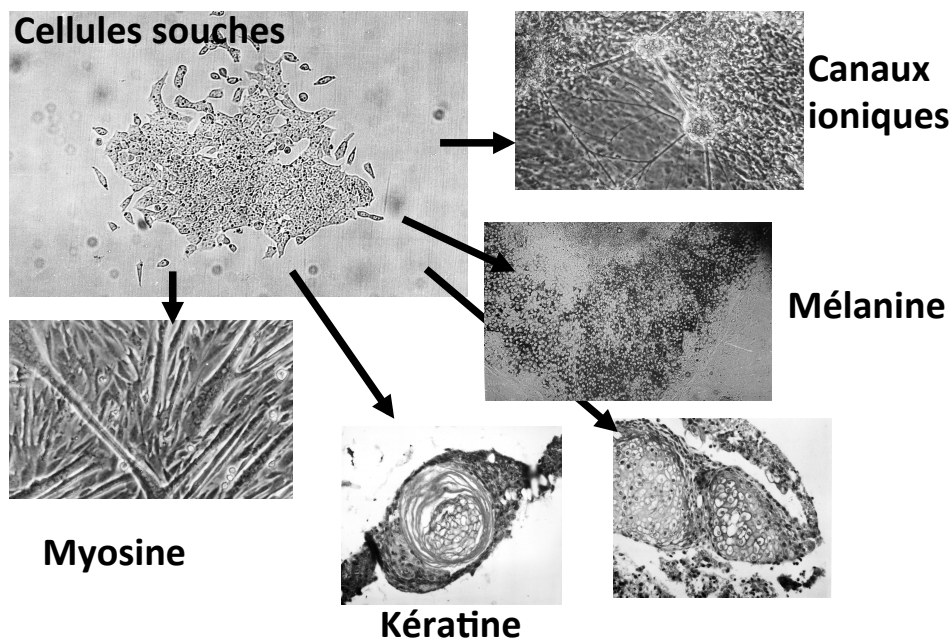


# II. LA DIFFÉRENCIATION DES TYPES CELLULAIRES

**La différenciation des types cellulaires implique deux niveaux d'interaction, celui des gènes entre eux et celui des cellules entre elles. Les interactions entre gènes se font par les facteurs de transcription (FTs) activateurs ou répresseurs. Les FTs s'activent ou se répriment entre eux aussi. De ces interactions entre FTs émergent des propriétés pour les cellules (mémoire, opérations logiques « et », « ou », et bien d'autres). Ces interactions entre gènes ne sont pas l'apanage des animaux. On les trouve chez les unicellulaires, protistes, bactéries et archées. Les interactions entre cellules créent des hétérogénéités entre elles, nécessaires à la différenciation. Les processus de communication des cellules entre elles apparaissent il y a 700 million d'années chez les premiers animaux qui cependant ne présentent pas d'organes.**

Qu'est-ce qui distingue les différents types cellulaires (Figure 8) ?

Les cellules se caractérisent toutes par des protéines particulières, comme la myosine pour les cellules musculaires qui les rendent contractiles. De même les cellules nerveuses se caractérisent par d'autres protéines comme les canaux ioniques dépendant du voltage, mais elles n'ont pas de myosine. Les cellules pigmentées rétinienne se caractérisent par de la mélanine qui est fabriquée par une protéine, la tyrosinase, mais elles n'ont ni myosine ni canaux ioniques. Les cellules de la peau font de la kératine etc. (Figure 8, référence du système de différenciation *in vitro*<sup>7</sup>).

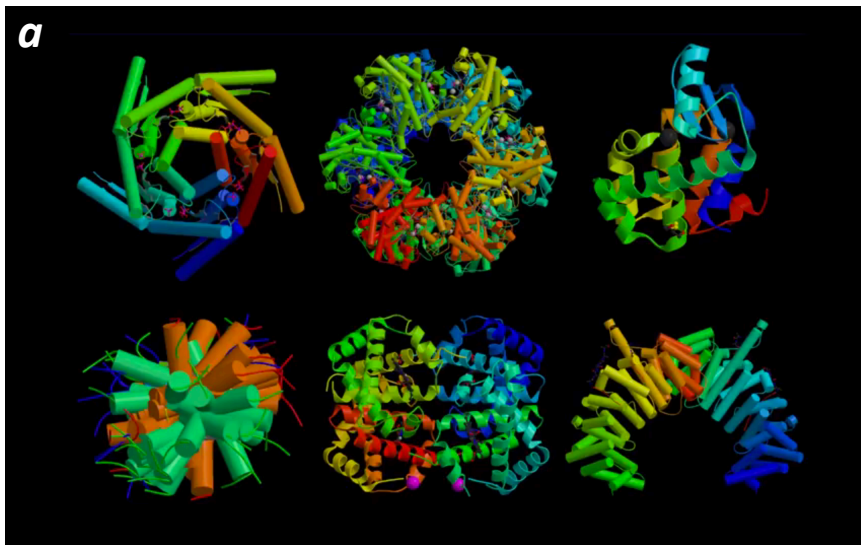


**Figure 8 : La différenciation cellulaire.** Les cellules différenciées présentent des caractéristiques différentes. Elles dérivent toutes des mêmes cellules, les cellules de l'embryon ou ici de cellules souches.

J.F.Nicolas, H.Jakob, F.Jacob, 1973

Les protéines sont des macromolécules. Elles diffèrent les unes des autres par leur forme (Figure 9a) qui est un facteur crucial pour leurs propriétés. Elles correspondent à des combinaisons différentes de quelques centaines d'éléments, les acides aminés, mis à la queue leu leu, et choisis parmi les 22 possibles. L'ordre des acides aminés dans les protéines est, avec leur forme, essentiel pour leurs propriétés.

**Figure 9 : Les protéines.** Les protéines sont constituées d'une suite d'acides aminés qui définit leur forme a) et leurs propriétés. A chaque protéine, b), correspond un gène (ADN), c). d).le gène est transcrit en ARN (voir figure 10) et l'ARN est traduit en protéine (voir figure 11).



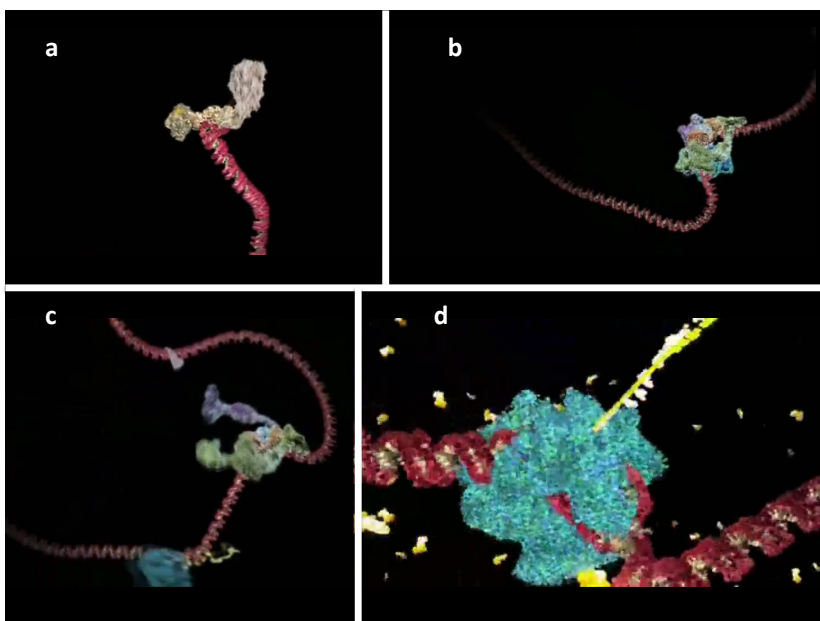
**b** Protéine (Acides aminés)                      **c** Gène, ADN (A, T, G, C)

**d** Gène (ADN) → ARN → Protéine

L'ADN est une macromolécule composée de 4 éléments, les bases puriques et pyrimidiques aussi mises à la queue leu leu, répétées quelques milliers de fois.

A une protéine correspond un gène (les gènes codent les protéines) et les gènes sont des séquences

Les biologistes moléculaires ont élucidé comment les gènes donnent les protéines. Ceci se fait en deux étapes (Figure 9b). Une copie ARN du gène ADN est produite qui sert ensuite à synthétiser la protéine. L'ARN est une macromolécule dont la structure est très proche de celle de l'ADN.



la structure est très proche de celle de l'ADN.

Voyons sur des animations ces deux étapes.

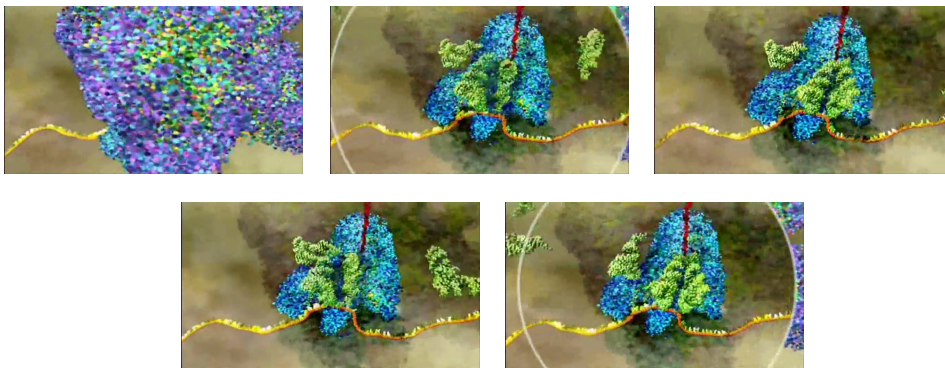
**Figure 10 : La transcription.** Un gène est transcrit en des milliers d'ARNs. Référence du film : <https://www.youtube.com/watch?v=D3fOXt4MrOM>, <https://www.youtube.com/wa>

tch?v=k04qdhPGDMM&t=3541s de 14' 40'' à 16'.

La transcription (Figure 10). Un facteur (en gris) est fixé sur le début du gène (en rouge), (a). Ce facteur recrute quelques 85 autres protéines, qui associées forment une machine moléculaire (en bleu-vert, b et c) qui va synthétiser l'ARN (en jaune) à partir des bases U, T, G, C du milieu en copiant l'un des deux brins de l'ADN (d).

Après la transcription, la traduction de l'ARN en protéine (Figure 11) se fait avec une autre énorme machine moléculaire. Des molécules, les tARN (en vert), se fixent sur 3 bases de l'ARN (un codon) auxquelles correspond l'acide aminé particulier qu'ils portent. Ces acides aminés sont assemblés selon la suite des codons, la suite des triplets de bases de l'ARN. Les tARN

aa2, aa3, aa4, aa5, aa6, aa7, aa8, aa9, aa10, aa11, aa12, aa13, aa14, aa15



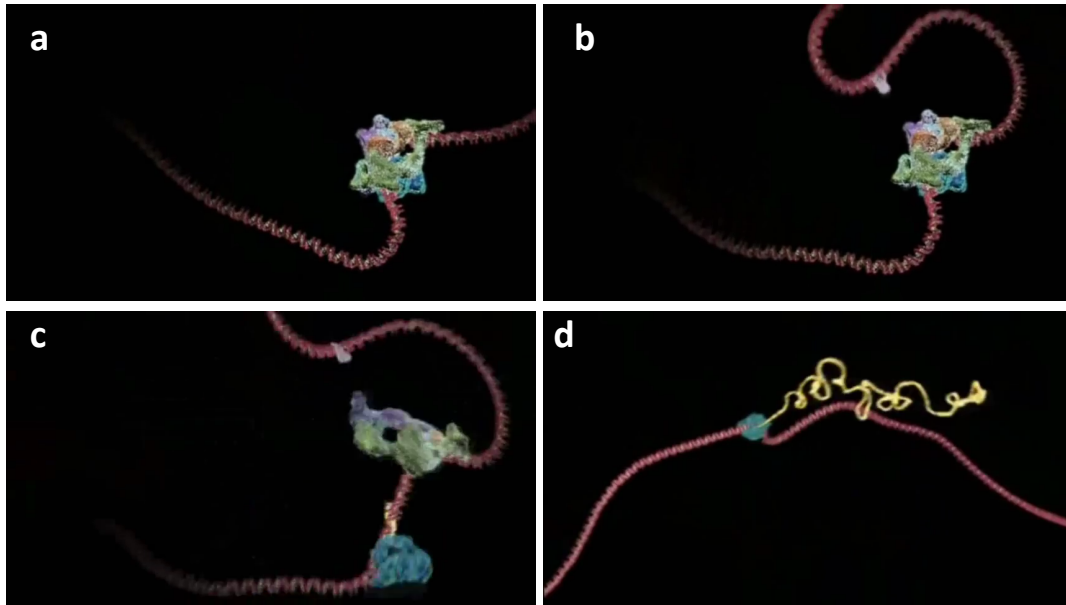
assurent la correspondance entre le gène  
 TGGCAACG  
 GTACG....  
 (suite de codons) et la protéine aa1, aa2, aa3,...  
 (suite d'acides

GGCAGACUGACUCAAGCCUAGCCCGUUCAGGAAUGCCUUUACGAGCCGUC  
 PBS program "DNA The Secret of Life »

aminés).

**Figure 11 : La traduction.** Les ARNs sont traduit en des protéines. Référence du film : <https://www.youtube.com/watch?v=D3fOXt4MrOM>, ou <https://www.youtube.com/watch?v=k04qdhPGDMM&t=3541s> de 16' à 17' 20''.

Si j'ai pris ce temps pour rappeler ce qu'on appelle le dogme fondamental de la biologie moléculaire, c'est que s'y glisse l'un des secrets de l'évolution, l'activation transcriptionnelle (Figure 12).



**Figure 12 :  
L'activation  
transcrip-  
tionnelle par  
les facteurs  
transcrip-  
tionnels  
(FTs).**

Référence du  
film :

[https://www.  
youtube.com/  
watch?v=D3f  
OXt4MrOM,](https://www.youtube.com/watch?v=D3fOXt4MrOM)

ou

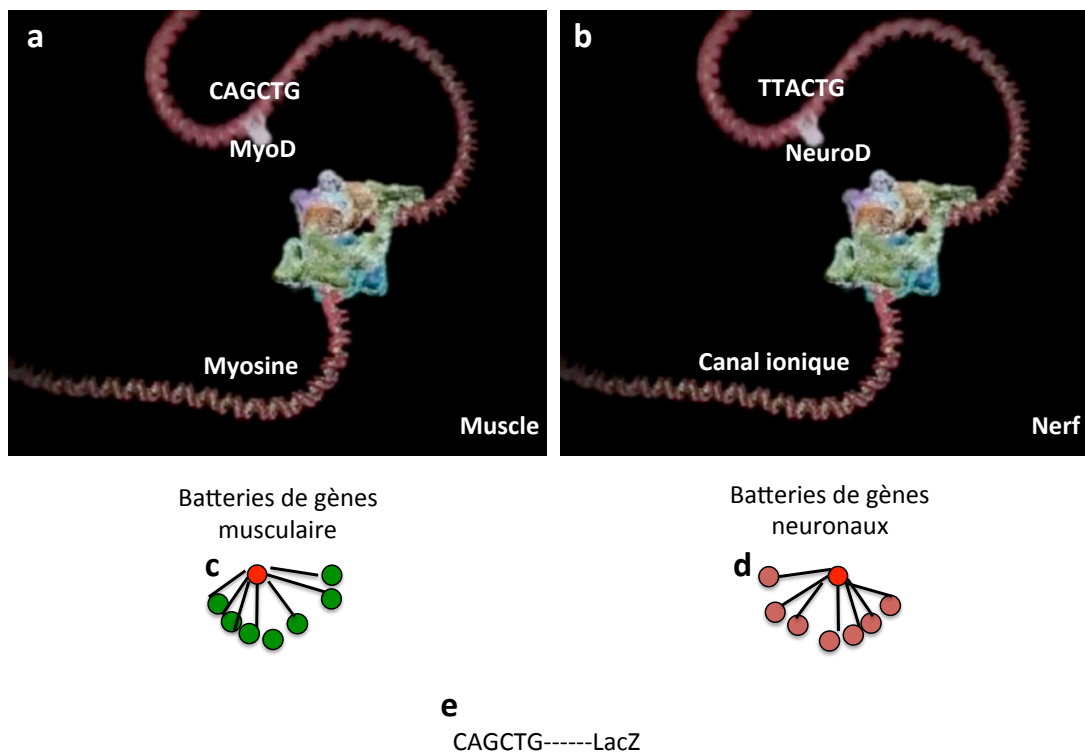
<https://www.youtube.com/watch?v=k04qdhPGDMM&t=3541s> de 17' 20'' à 18'40''.

Pour que la transcription débute (Figure 12c et d) il faut qu'un facteur fixé sur l'ADN (le petit élément gris en haut à droite, Figure 12b) vienne modifier chimiquement la machine moléculaire. Ce facteur est un facteur transcriptionnel (FT), une protéine, qui reconnaît sur l'ADN une séquence spécifique de 5 à 10 bases sur laquelle il se fixe. L'ADN fait une boucle pour mettre en contact le facteur et la machine moléculaire (Figure 12b et c).

Et voici pourquoi ceci est si important : dans les cellules musculaires un facteur transcriptionnel appelé MyoD reconnaît une séquence, CAGCTG, de l'ADN. MyoD<sup>8</sup> active la transcription du gène de la myosine en interagissant avec la machine moléculaire (Figure 13a).

Dans les neurones on trouve un autre facteur transcriptionnel appelé NeuroD qui reconnaît une autre séquence. Et NeuroD active le gène « canal ionique » (Figure 13b). MyoD n'est pas présent dans les neurones ; NeuroD n'est pas présent dans les cellules musculaires.

**Figure 13 : Des FTs spécifiques pour chaque type cellulaire.** a, c) MyoD contrôle l'expression de la myosine et de batteries de gènes musculaires ; b, d) NeuroD contrôle l'expression du canal ionique et de batteries de gènes neuronales ; e) le gène bactérien LacZ associé au site de fixation du FT MyoD.



Ce sont les facteurs transcriptionnels qui assurent la spécificité de l'expression des gènes.

MyoD active aussi

d'autres gènes que la myosine, ceux qui sont flanqué de la séquence CAGCTG qu'il reconnaît. On parle de la batterie de gènes activée par MyoD (Figure 13c). De même, NeuroD active des batteries de gènes neuronales (Figure 13d). Un seul facteur de transcription est capable d'activer des dizaines voire des centaines de gènes.

Nous allons introduire une notation pour cela. Des nœuds, en rouge ou en vert, représentent des gènes, et/ou leur produit, ça n'a pas d'importance. MyoD le facteur de transcription est en rouge, les gènes qu'il active en vert. Des flèches vont représenter les liens entre les gènes ; ici ils indiquent une



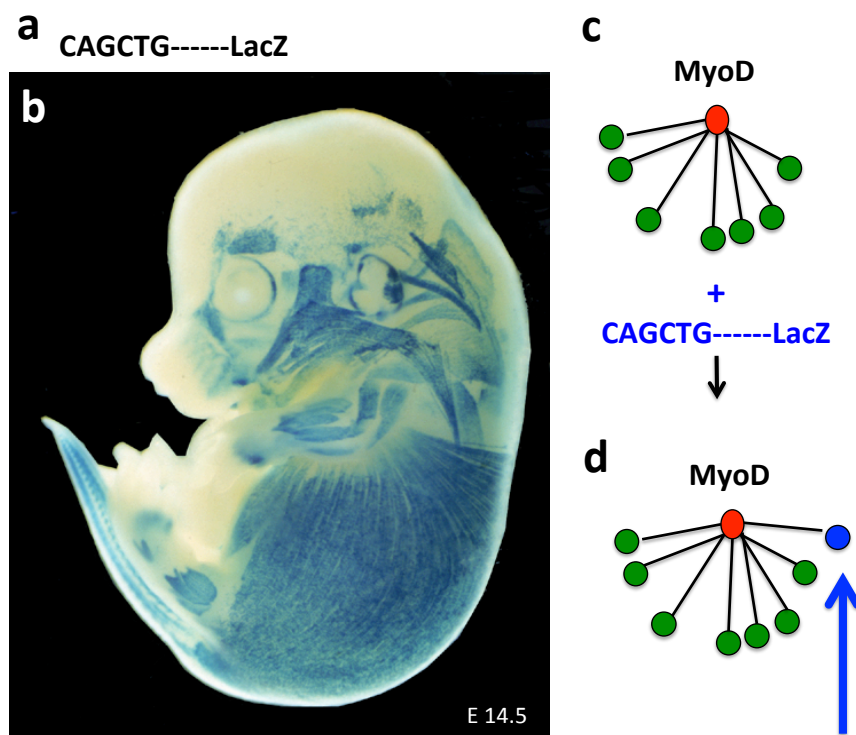


activation des gènes verts par MyoD (Figure 13c). Avec cette notation on voit qu'il existe un arrangement particulier entre ces gènes. Le même type d'arrangement pour NeuroD avec d'autres gènes activés représentés en marron (Figure 13d).

La souplesse de ce système vient de ce que chez les néphrozoaires (chez la souris ou la drosophile et donc sans doute aussi chez leur dernier ancêtre commun) les séquences sur l'ADN que reconnaissent MyoD ou NeuroD ou d'autres FT peuvent être presque n'importe où autour du gène. En amont, en aval ou même dans la partie qui code la protéine !

Faisons une expérience pour comprendre l'avantage énorme que cela représente. Posons-nous la question de savoir ce qu'il advient si l'on ajoute la séquence que reconnaît MyoD à un gène quelconque, par exemple un gène d'une bactérie (Figure 13e).

Choisissons le gène, LacZ<sup>9</sup> qui rend les cellules bleues et flanquons n'importe où la séquence que reconnaît MyoD (Figure 14a). Puis introduisons le tout dans le génome de la souris. Cela les biologistes moléculaires savent le faire depuis les années quatre-vingt<sup>10</sup>. On constate que le gène LacZ, quelle que soit la position où l'on a mis la séquence que reconnaît MyoD et quelle que soit sa position dans le génome<sup>11</sup>, est alors exprimé dans les muscles de la souris (Figure 14b). Il fait partie de la batterie de gènes MyoD (Figure 14c, la flèche bleue). Il a été adopté, annexé (on dit coopté).



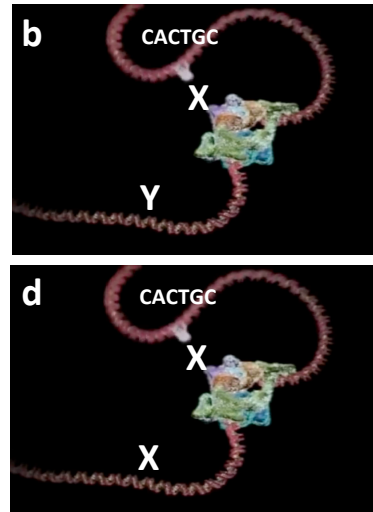
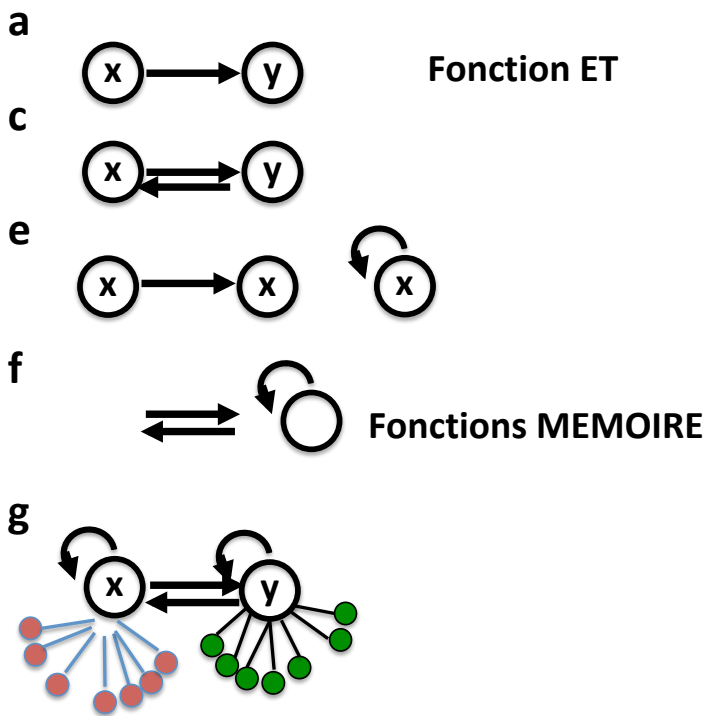
**Figure 14 : La cooptation du gène LacZ par MyoD.** a) le gène bactérien LacZ associé au site de fixation du FT MyoD ; b) après introduction dans le génome de la souris on constate qu'il est exprimé dans tous les muscles ; c) le gène LacZ (le rond bleu) a donc été coopté par MyoD (le rond rouge), il fait partie de la batterie de gènes (les ronds vert) que ce FT contrôle.

Des cooptations de ce type se produisent encore et encore au cours de l'évolution. Elles modifient les batteries de gènes que contrôlent les FTs. Elles ajoutent (ou retirent) à ces batteries des gènes qui peuvent être n'importe où dans le génome. Elles sont grandement facilitées par le fait que les séquences que reconnaissent les facteurs transcriptionnels n'ont pas de contrainte de position par rapport aux gènes.

Jusqu'où vont ces cooptations qui font que, brusquement, un gène en contrôle un autre ? Et si le gène contrôlé était un facteur de transcription, que se passerait-il ? Voyons cela sur deux exemples.

Soit Y un facteur de transcription ; Y est, tout comme toute protéine codée par un gène. Y peut donc être régulé par un autre facteur de transcription, tel X (Figure 15a). Pour cela il suffit qu'au gène Y se trouve associé la séquence ADN sur laquelle X se fixe (Figure 15b). Les facteurs de transcription peuvent se réguler les uns les autres<sup>12-14</sup>.

Deuxième exemple : X, un FT active Y un autre FT, et Y active X (Figure 15c).



**Figure 15 :**  
**Les motifs de réseau.**  
 Certains arrangements de gènes entre eux dotent les cellules de nouvelles propriétés.

Et là, brusque-

ment, il se produit pour la cellule un prodige. Les expressions de X et de Y deviennent permanentes. Y entretient l'expression de X qui entretient l'expression de Y. Sans cet arrangement, X et Y ne seraient exprimées que transitoirement. En effet X et Y sont des protéines qui donc se dégradent assez rapidement. Cet arrangement, l'activation réciproque (ou encore double boucle de rétroaction positive) fait que leur expression va persister !

Autre arrangement qui confère persistance à l'expression : la séquence de reconnaissance de X flanque le gène X (Figure 15d), X s'active lui-même (Figure 15e). C'est l'autorégulation positive.

Ces propriétés qui émergent de certains arrangements entre gènes ne peuvent être prédites et ont des conséquences inouïes. Ici (activation réciproque et autorégulation positive, Figure 15f) X et Y et les batteries de gènes qui leur sont associées sont co-exprimés indéfiniment ! Et justement c'est une des propriétés essentielles des cellules différenciées : un neurone, une cellule musculaire le restent tout au long de la vie de l'organisme qu'ils constituent.

Les propriétés émergentes, qui ne peuvent être prédites, et apparaissent brusquement (indépendamment de toute pression de sélection) sont capitales en biologie, présentes à tous les niveaux d'interactions.

L'activation réciproque est un arrangement qui confère non seulement persistance mais aussi une fonction « et ». *X et Y* et les batteries de gènes qu'ils contrôlent vont être activées de façon permanente. Dans une cellule différenciée plusieurs batteries de gènes sont activées par des facteurs de transcription différents (Figure 15g).

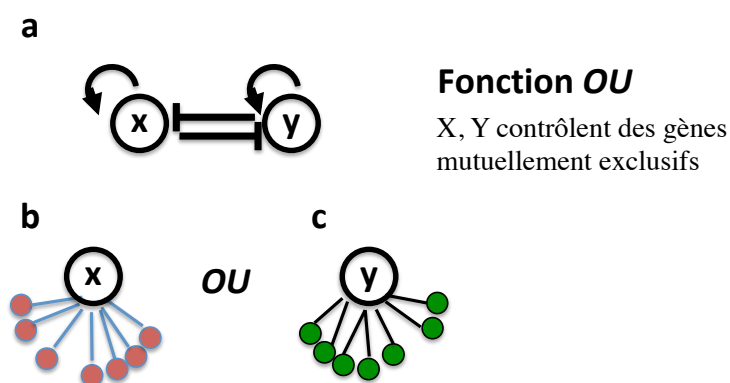
Tout cela vaut pour la situation inverse, la répression de l'activité des gènes. Elle se fait de la même manière que l'activation, avec des facteurs de transcription qui agissent à distance. C'est noté avec un signe différent (Figure 16a).

Dans le cas d'une double boucle de rétroaction négative (Figure 16a) si X est exprimé Y ne l'est pas (car il est réprimé par X) et réciproquement. A ce nouvel arrangement correspond la fonction « ou » (Figure, 16a). *Ou* bien la batterie de gènes contrôlée par X est exprimée (Figure 16b) et pas celle de Y, ou l'inverse (Figure 16c).

Nous saisissons un peu, au niveau des gènes, pourquoi une cellule musculaire et une cellule nerveuse sont différentes et le restent.

Est-ce suffisant pour expliquer la différenciation des types cellulaires ?

Reconsidérons un groupe de cellules dérivées par division d'une même cellule mère. Dans cette colonie les cellules sont toutes identiques (Figure 17, a). Comment alors des facteurs de transcription peuvent-ils se mettre à fonctionner dans seulement certaines cellules ? C'est là une question fondamentale.



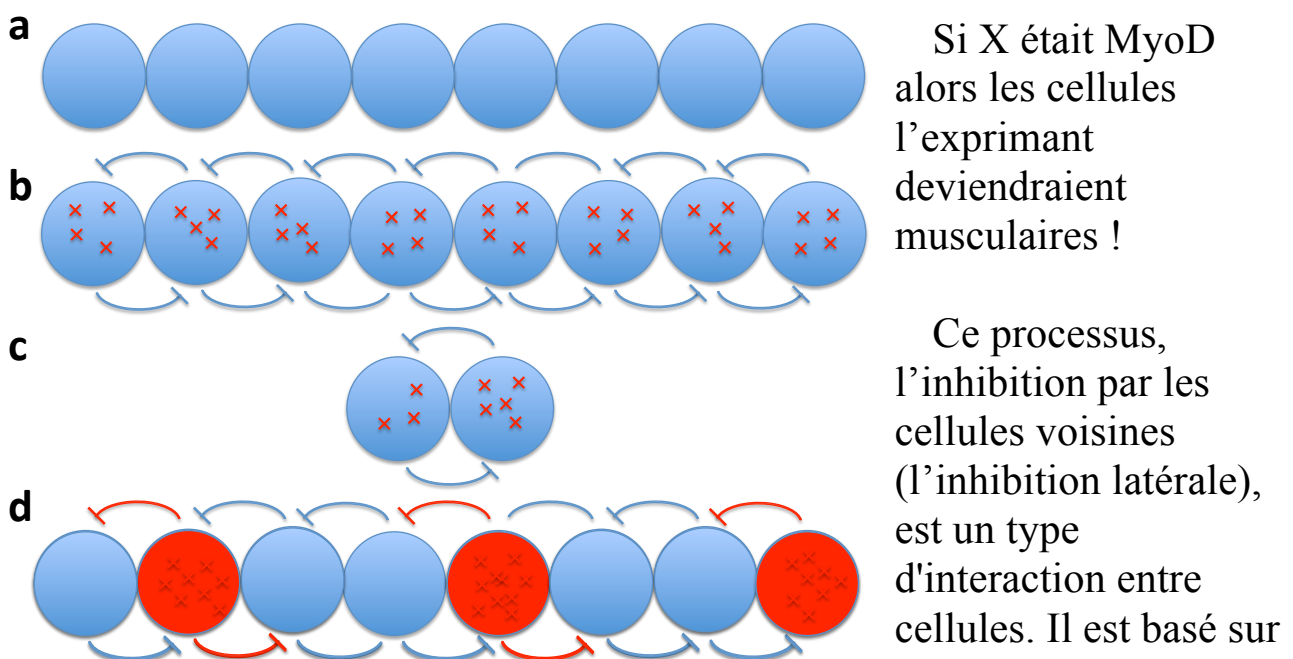
**Figure 16 : Les motifs de réseau :** les boucles de rétroaction négatives.

L'autre exigence pour la différenciation c'est un processus produisant une hétérogénéité entre cellules identiques. La solution retenue par l'évolution concerne non plus le niveau des gènes mais le niveau des cellules.

Voici l'exemple d'un tel processus utilisé par les organismes actuels.

Envisageons que les cellules produisent toutes une substance X dont la propriété est d'inhiber la production de X dans les cellules voisines (Figure 17b). Ces cellules sont identiques mais il est inéluctable qu'à un moment donné certaines cellules présentent un peu plus ou un peu moins de X. Ces fluctuations vont alors s'accroître. Dans les cellules avec plus de X (Figure 17c) X va encore augmenter car les cellules voisines vont moins l'inhiber ; dans les cellules avec moins de X, X va encore diminuer car les cellules voisines vont plus l'inhiber. Le processus va s'emballer et les cellules concernées vont devenir différentes (Figure 17d).

**Figure 17 : L'inhibition latérale, un processus qui crée de l'hétérogénéité entre les cellules.** Voir texte.



un événement aléatoire bien simple (la fluctuation de X) mais capable de générer des types différents dans un groupe homogène.

### **L'essentiel**

**La différenciation des types cellulaires nécessite des processus qui créent de l'hétérogénéité entre cellules identiques conduisant à une expression différentielle de FTs. Ces processus se sont les communications des cellules entre elles. Les FTs contrôlent les batteries de gènes spécifiques de chaque type cellulaire. Ils interagissent aussi entre eux et de ces interactions peuvent émerger des propriétés comme la persistance de l'expression.**



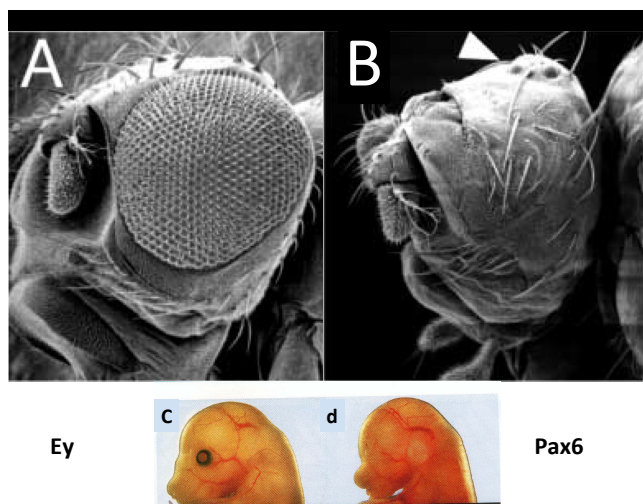
# III. LA FORMATION DES ORGANES.

La formation des organes implique, tout comme la différenciation cellulaire, deux niveaux d'interaction. D'abord les interactions entre cellules qui créent de multiples territoires cellulaires dans lesquelles des cascades différentes de FTs activent des opérations cellulaires particulières. Puis les interactions entre gènes qui forment des motifs de réseau plus complexes que ceux de la différenciation cellulaire, comme les cascades non linéaires et les boucles « feedforward », desquels émergent de nouvelles propriétés pour les cellules. Ces circuiteries génétiques confèrent à certains gènes un statut particulier, dit de gène maître, qui se renforce au cours de l'évolution par des processus comme l'intercalation. Ces gènes n'en restent pas moins des FTs comme les autres. La constitution de réseaux génétiques d'un tel niveau de complexité s'appuie sur un processus particulièrement efficace de cooptation des gènes, l'activation distale. Ce processus sous sa forme la plus élaborée (avec l'apparition de CTCF) coïncide avec l'explosion de la diversification des animaux.

Nous analysons l'exemple type d'un organe, l'œil (Figure 18A et c)<sup>15</sup>.

Cette histoire commence en 1915, quand une drosophile sans yeux fut décrite (Figure 18A, œil normal de la drosophile, B, le mutant, absence des yeux). Ce caractère est dû à la mutation d'un seul gène. Par la suite des mutations semblables furent décrites chez l'homme et chez la souris (Figure 18c et d) et de nouveau un seul gène est impliqué. Dans les 3 cas il s'agit d'une délétion quasi parfaite des yeux. Cependant, curieusement, ni les généticiens et ni les embryologistes n'ont vraiment élaboré là-dessus. Il a fallu attendre le temps du clonage des gènes pour que l'histoire se poursuive. Cloner un gène, c'est isoler et multiplier le fragment d'ADN qui lui correspond. En 1991 le gène responsable de la délétion de l'œil de la souris fut isolé, séquencé et reconnu comme étant un facteur de transcription.

L'importance de ces observations n'est apparue que quand Walter Gehring clona le gène de la drosophile en 1995 et annonça qu'il ressemblait comme deux gouttes d'eau à celui de la souris<sup>16</sup>.

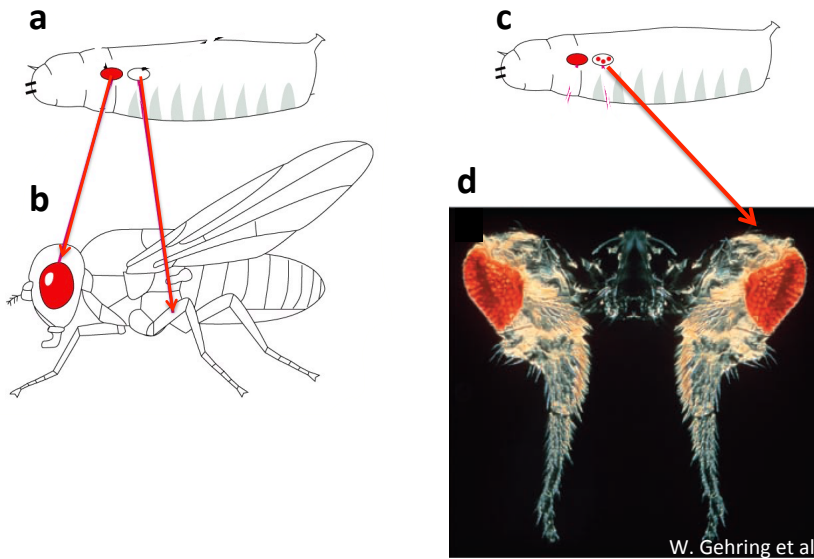


**Figure 18 : Les mutations « eyeless » chez la drosophile, B et Sey (Pax6) chez la souris, d. Des délétions parfaites des yeux**

Des mutations de gènes homologues, des facteurs de transcription, qui font disparaître tout un organe, quelle histoire étonnante ! Le gène de la souris est dénommé Pax6 et celui de la drosophile Ey.

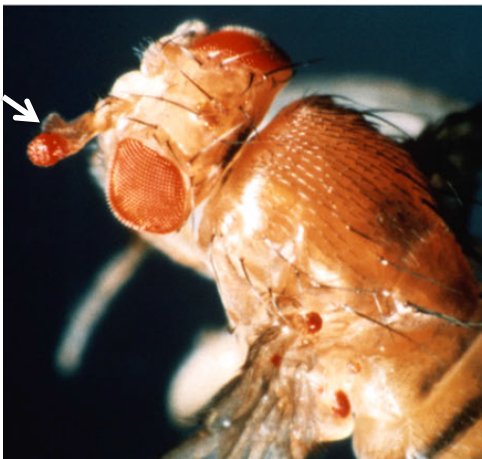
Mais Walter Gehring a été plus loin. Il a introduit ce gène dans les cellules du petit territoire de la larve de la drosophile à l'origine des pattes (Figure 19a). Ce territoire qui n'exprime pas normalement le gène va l'exprimer. Et lorsqu'on laisse se développer la larve le résultat est remarquable. Des yeux se forment sur les pattes<sup>17</sup> (Figure 19b).





**Figure 19 : Des yeux sur les pattes.** a, b) les territoires de la larve à l'origine de l'œil (en rouge) et de la patte (en blanc) de la mouche adulte ; c) si l'expression de Pax6 est forcée dans le territoire des pattes ceux-ci développent des yeux (sur les pattes d).

Et il a été encore plus loin, il a introduit le gène de la souris dans le territoire de la larve des antennes de la drosophile. Et voici le résultat : des yeux sur les antennes (Figure 20).



**Figure 20 : Des yeux sur les antennes.** Si l'expression de Pax6 est forcée dans le territoire des antennes celles-ci développent des yeux.

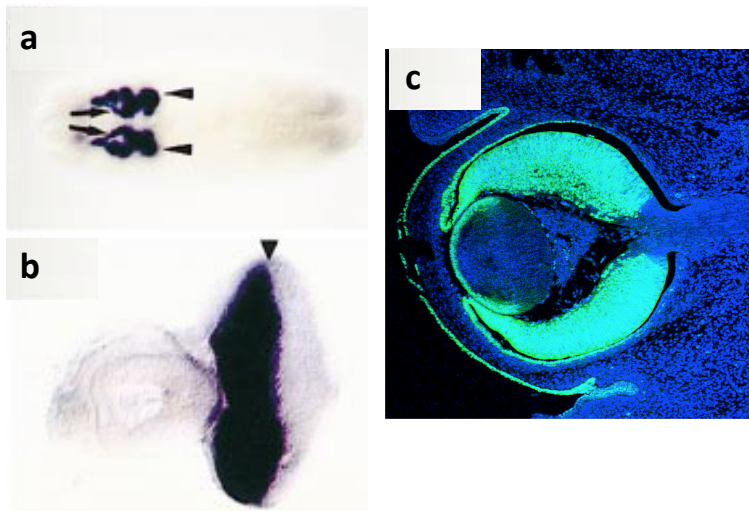
De ces remarquables expériences plusieurs conclusions peuvent être tirées :

Clairement le gène Pax6 ne contient pas l'information de l'œil, il ne code pas l'œil proprement dit, puisque dans la dernière expérience les yeux sur les antennes ne sont pas ceux de souris mais ceux de drosophile. Pax6 ne contient l'information que pour un facteur de transcription. C'est une conclusion que l'on peut généraliser : pas plus qu'il n'y a un gène de l'œil, il n'y a un gène de l'intelligence ou de l'habileté manuelle. Ceux qui prétendent cela vous disent des âneries. S'il est vrai qu'il y a beaucoup de pathologies qui ne dépendent que d'un gène qui muté produit une protéine défectueuse, les processus d'où émergent des propriétés (et l'intelligence ou l'habileté manuelle sont des propriétés émergentes) dépendent tous non pas des gènes mais d'interactions particulières (entre gènes, entre protéines, entre cellules, entre organes). Ils ne sont pas réductibles à un gène particulier.

Par contre ce facteur de transcription est capable de déclencher à lui seul le développement d'un organe complexe. Il est nécessaire (les mutants sont sans yeux) et suffisant (expression ectopique) et ce même dans des cellules destinées à toute autre chose. Comment cela se peut-il ?

Les patrons d'expression, c'est à dire quand et dans quelles cellules de l'embryon ces gènes sont exprimés, de Pax6 et de Ey nous renseignent.

**Figure 21 : Les patrons d'expression de Ey<sup>18</sup> et de Pax6.** L'expression de ces gènes démarre très tôt. Pour ey dès le stade 10 (voir figure 3c) et persiste jusqu'à la fin du développement de l'œil, a) stade 16, b) le disque imaginal de la troisième larve. Pour Pax6 dès l'ectoderme neural et jusque dans la rétine (en vert) c).

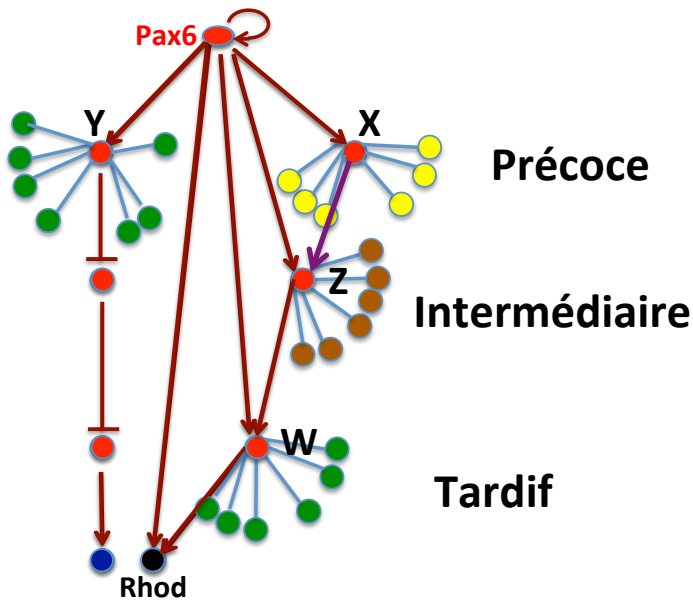


Ey, le gène de la drosophile, est exprimé très tôt, bien avant tout signe de développement de l'œil, dans le blastoderme de l'embryon. L'expression est maintenue (Figure 21a et b) jusqu'à ce que les structures de l'œil soient formées<sup>18</sup>. Chez la souris l'expression est aussi très précoce et implique plusieurs tissus, l'ectoderme

neural (la future rétine) et l'ectoderme de surface (le futur cristallin). Elle perdure jusqu'à la différenciation des cellules photoréceptrices (Figure 21c). Cette expression permanente suggère que Pax6 ou Ey font plus que d'enclencher le processus. Ils y participent.

Et de fait la génétique dévoile comment Pax6 intervient (Figure 22, notez que la circuiterie présentée est imaginaire mais illustre des propriétés réellement observées).

Pax6 active d'autres facteurs de transcription, X et Y. Puis ensemble Pax6 et X activent Z ; puis Pax6 et Z activent W. Enfin Pax6 et W activent les rhodopsines.



**Figure 22 : Pax6 contrôle des cascades de FTs impliqués dans la formation des yeux.** Remarquer aussi le motif de réseau boucle « feedforward » répété 3 fois, voir texte.

L'ensemble de ces activations-répressions forme des cascades non linéaires de facteurs de transcription. De cet arrangement émergent deux nouvelles propriétés : une

échelle de temps et une séquentialité. En effet chaque étape d'activation nécessite environ une génération cellulaire. Des gènes vont donc être d'expression précoce, d'autres intermédiaires et d'autres tardives (Figure 22). Et bien sûr les batteries de gènes qui exécutent des opérations cellulaires que ces facteurs de transcription contrôlent vont se déployer séquentiellement.

Apparaissent aussi de nouveaux motifs de réseau. Ainsi un motif à trois nœuds est utilisé trois fois. Pour que Z ou W ou les rhodopsines soient activés il faut deux conditions, Pax6 et X pour Z, Z et Pax6 pour W et W et Pax6 pour les rhodopsines. C'est la boucle « feedforward » (détecteur de persistance). Il protège le système d'activations inappropriées dues au bruit de fond ou à de petites fluctuations de X et de W etc. Des boucles feedforward de ce type ont été sélectionnées encore et encore au cours de l'évolution. C'est un motif majeur des réseaux de régulation.

Il devient évident que les facteurs de transcription sont les éléments qui forment un vaste réseau de régulation. L'importance de ce réseau vient des propriétés émergentes qu'il engendre. Le rôle des facteurs de transcription ou des gènes doit être interprété par rapport aux motifs de réseau et au réseau dans lequel ils se trouvent.

Ainsi si Pax6 a ce rôle si particulier c'est de par sa position hiérarchique dans le réseau et parce qu'il intervient à plusieurs niveaux en contrôlant

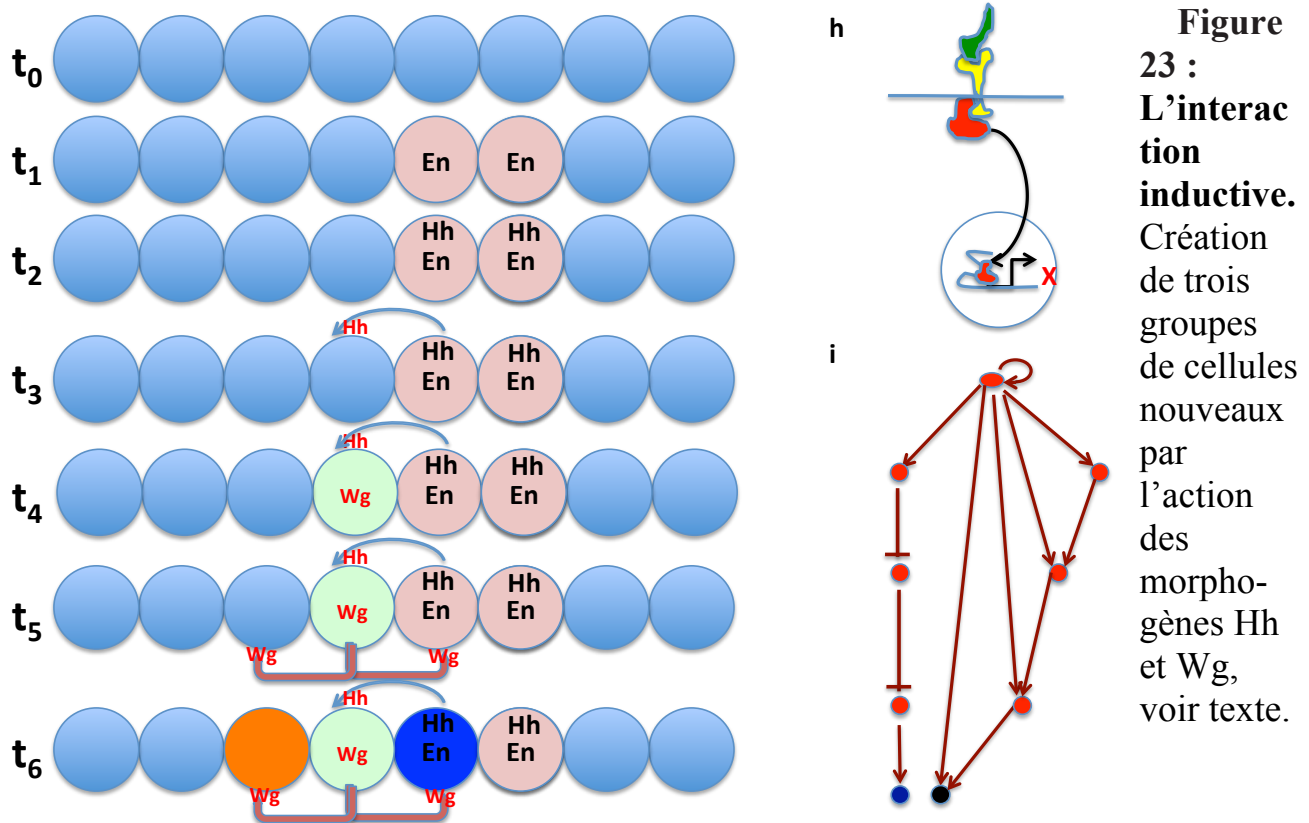
beaucoup d'autres facteurs de transcriptions du réseau de régulation. Ce gène a été comparé à un chef d'orchestre qui appelle successivement tel ou tel groupe d'instruments ou soliste à jouer ou à se taire. Pax6 est aussi appelé un gène maître. Cela ne doit pas cacher qu'il reste un facteur transcriptionnel comme les autres.

Voilà ce qu'il en est pour le premier niveau d'interactions qui concerne les gènes entre eux.

Mais un œil, comme tous les autres organes, se construit avec plusieurs types cellulaires (Figure 6). Nous avons vu que l'émergence d'un type cellulaire dans seulement certaines cellules d'un groupe de cellules identiques, nécessite un processus produisant une hétérogénéité entre ces cellules (Figure 17). C'est que, dans un groupe de cellules identiques, le facteur de transcription à l'origine du nouveau type cellulaire ne peut pas se mettre à fonctionner dans seulement certaines cellules de ce groupe. Un processus produisant une hétérogénéité entre les cellules du groupe doit intervenir. C'est cette hétérogénéité qui ouvre la possibilité au fonctionnement du nouveau facteur de transcription dans seulement certaines cellules.

Durant la formation de l'œil, il faut donc que des groupes de cellules se constituent (à partir d'autres groupes) d'ou émergeront les types cellulaires de l'organe. Les processus majeurs pour produire les hétérogénéités entre cellules sont les interactions cellulaires. C'est le deuxième niveau d'interaction mis en jeu dans la formation de organes. Il concerne les cellules entre elles.

Nous avons déjà vu un exemple de ces interactions, l'inhibition par les cellules voisines (l'inhibition latérale), basée sur les fluctuations de concentrations de certaines molécules (Figure 17). J'en mentionne ici un deuxième, l'interaction inductive (Figure 23) très utilisée par tous les développements. Il exploite la diffusion de molécules. Pour une discussion plus approfondie sur les interactions cellulaires, voir l'annexe 4.



L'interaction inductive démarre dans des cellules (Figure 23,  $t_0$ ) dans lesquelles a déjà été créé un groupe qui exprime un FT, nommé engrailed (En) (Figure 23,  $t_1$  les cellules roses). Ce facteur active une protéine, hedgehog (Hh) (Figure 23,  $t_2$ ), qui diffuse hors de ces cellules. Il se fixe sur les récepteurs membranaires des cellules voisines (Figure 23,  $t_3$ ) qui activent alors un autre facteur transcriptionnel : de nouveaux gènes dont Wg (wingless) sont activés faisant émerger un nouveau groupe, vert, dans le bleu (Figure 23,  $t_4$ ). Mais Wg est une molécule qui diffuse aussi hors des cellules et se fixe sur un autre récepteur (Figure 23,  $t_5$ ) qui active de nouveaux gènes et crée deux nouveaux groupes dans les territoires préexistants (Figure 23,  $t_6$ ). Ces trois nouveaux groupes sont à l'origine de la diversification des états d'activité des cellules. Dans les territoires générés par ces interactions cellulaires Pax6 contrôle des cascades différentes d'activation de gènes (Figure 23i, un exemple d'une telle cascade).

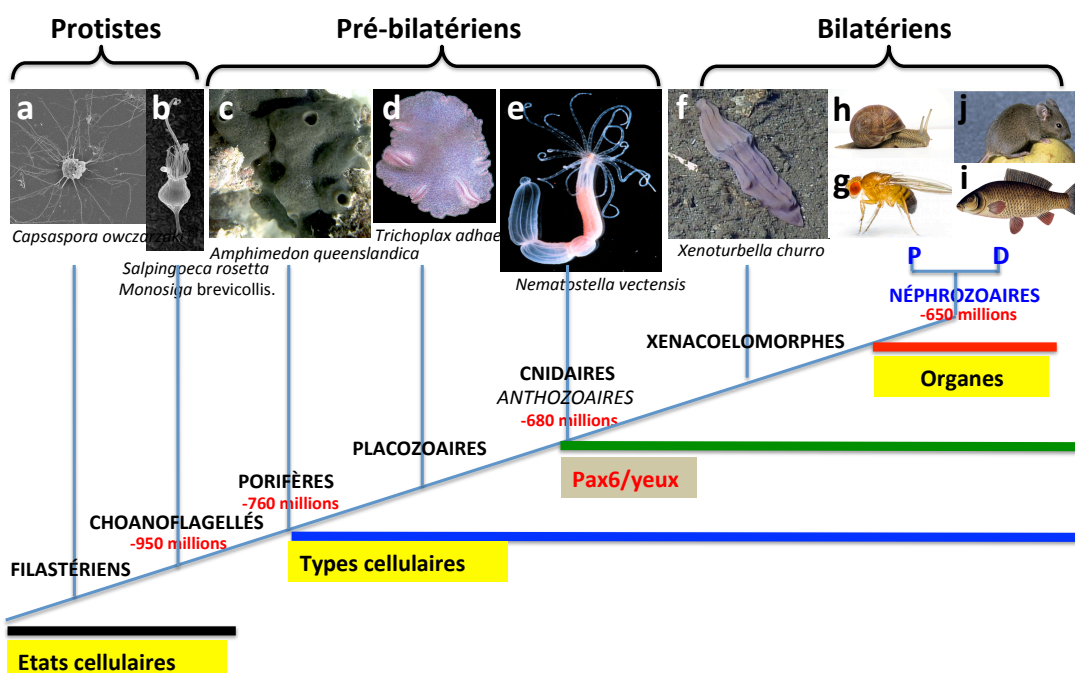
Hh et Wg, les molécules qui diffusent hors des cellules, sont ce que l'on appelle des morphogènes. Ce sont des éléments de voies de signalisation ici présentées d'une manière simplifiée (Figure 23h ; en vert, le morphogène ;

en jaune, le récepteur membranaire ; en rouge, le facteur de transcription modifié par le récepteur activé. Il active le gène X).

On pourrait penser que les développements ont multiplié les voies de signalisation en se complexifiant. Pour chaque nouveau type cellulaire, pour chaque nouvelle opération du développement, un nouveau signal, une nouvelle voie de signalisation avec son propre activateur transcriptionnel. Ça n'est pas le cas. Les développements utilisent une dizaine de voies de signalisation toujours les mêmes, que ça soit au travers d'un développement particulier ou au travers de tout l'arbre du vivant (les questions que cela soulève sont abordées dans l'annexe 4).

Voilà ce qu'il en est quant à la formation des organes. Nous comprenons un peu mieux le processus. Il s'apparente à celui de la différenciation cellulaire.

Il reste à expliquer pourquoi le même gène après 500 millions d'années sert chez la souris et la mouche et soit resté inchangé (rappelez-vous : quand le gène de la souris est introduit chez la drosophile dans le territoire des antennes, des yeux apparaissent sur les antennes, Figure 20). Mais tout d'abord, notons que l'association Pax6-œil ne concerne pas que la souris et la drosophile mais en fait tous les animaux de l'homme aux anémones de mer (Figure 24, la ligne horizontale verte).



**Figure 24 : L'association Pax6 yeux est universelle chez les animaux.** La ligne verte indique que l'association n'est détectée dès chez les cnidaires. D =

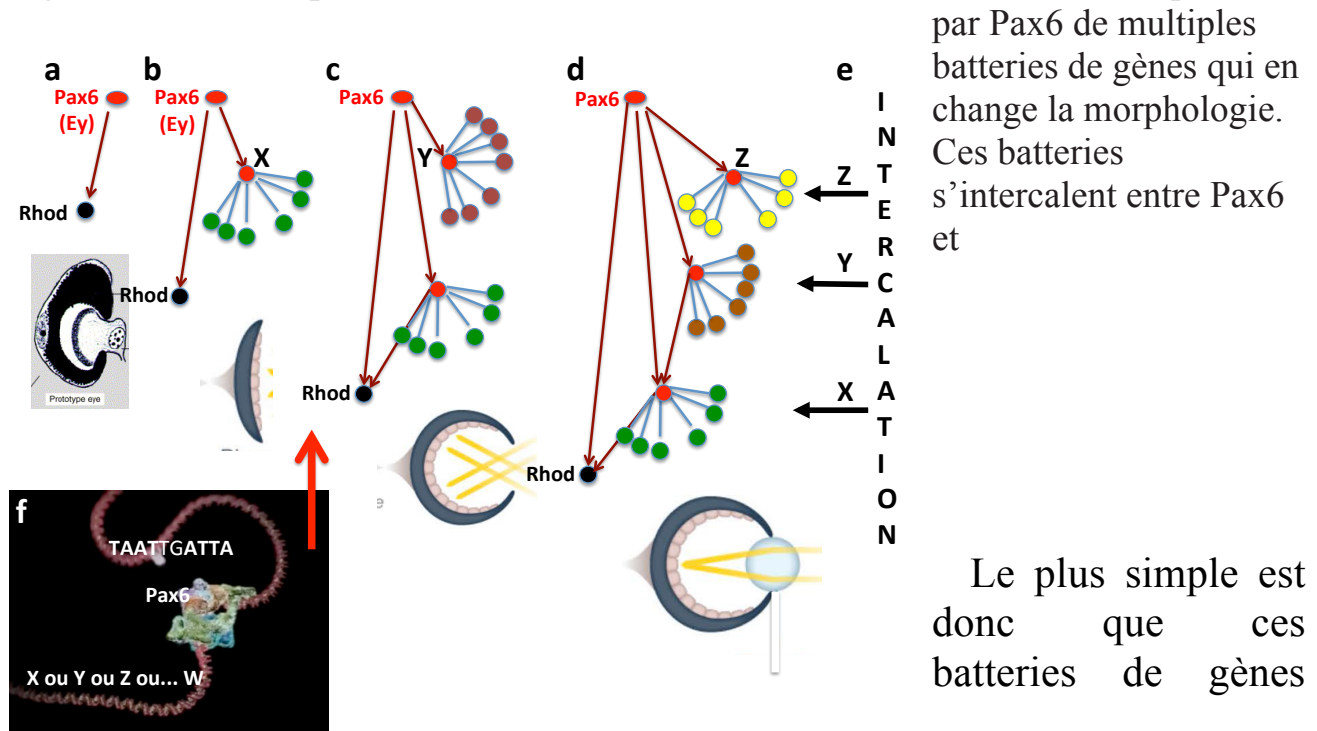
deutérostomes, P = protostomes.

Alors pourquoi l'association Pax6 yeux est-elle universelle ? Pourquoi Pax6 garde-t-il les mêmes caractéristiques ? Pourquoi d'autres facteurs de transcription ne sont-ils pas utilisés ? L'idée est la suivante<sup>19</sup> (pour plus de détails voir l'annexe 7) :

Avant les cnidaires, chez un animal comme Trichoplax, un œil très simple s'est, par pure chance, formé. Il était composé d'une cellule photoréceptrice et d'une cellule pigmentée (Figure 25a). Cela correspondant à de simples différenciations cellulaires, comme celles que nous avons déjà vu, Pax6 activant le gène de la rhodopsine.

La complexification de cet œil très simple passe par l'intervention de batteries de gènes. Ces batteries de gènes doivent être activées dans les bonnes cellules et au bon moment. Ces bonnes cellules et ce bon moment ce sont celles où Pax6 s'exprime déjà.

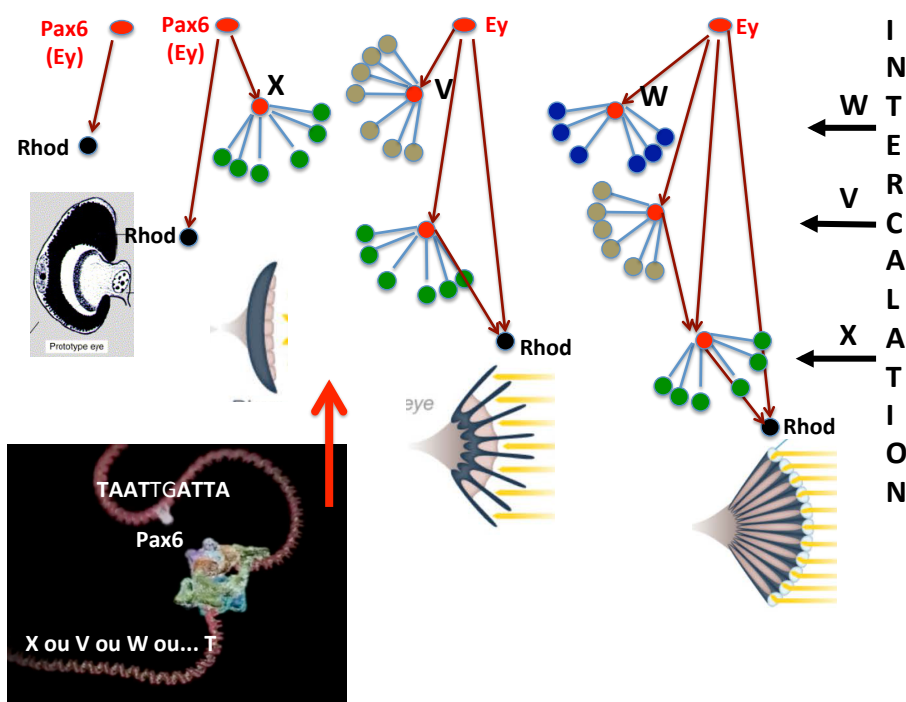
Figure 25 : La complexification de l'œil au cours de l'évolution. b à d) cooptations



tombent sous le contrôle de Pax6 (Figure 25b). Pour cela il suffit qu'aux alentours du FT X qui les contrôlent, la séquence que reconnaît Pax6 (5-TAATTGATTA-3 entre autres) apparaisse (Figure 25f). C'est une condition facilement réalisée, rappelez-vous les cellules bleues dans le muscle (Figure 14).

Plus de nouvelles batteries de gènes tombent de cette façon sous le contrôle de Pax6 (Figure 25c et d), et plus il devient difficile que l'œil soit contrôlé par un autre facteur de transcription. On parle d'intercalation de nouveaux gènes et de nouvelles batteries de gènes entre Pax6 et son contrôle initial de la rhodopsine (Figure 25e). Dès les cnidaires (Figure 25, la flèche rouge) le système comporte déjà beaucoup d'éléments et est, de ce fait, bloqué.

L'intercalation simplifie considérablement l'évolution. Elle est basée que sur des batteries de gènes qui existent déjà pour la différenciation cellulaire et sur un processus d'annexion de gènes assez simple,



**Figure 26 : La complexification de l'œil au cours de l'évolution.** Les batteries de gènes cooptées chez la mouche sont différentes de celles de chez la souris.



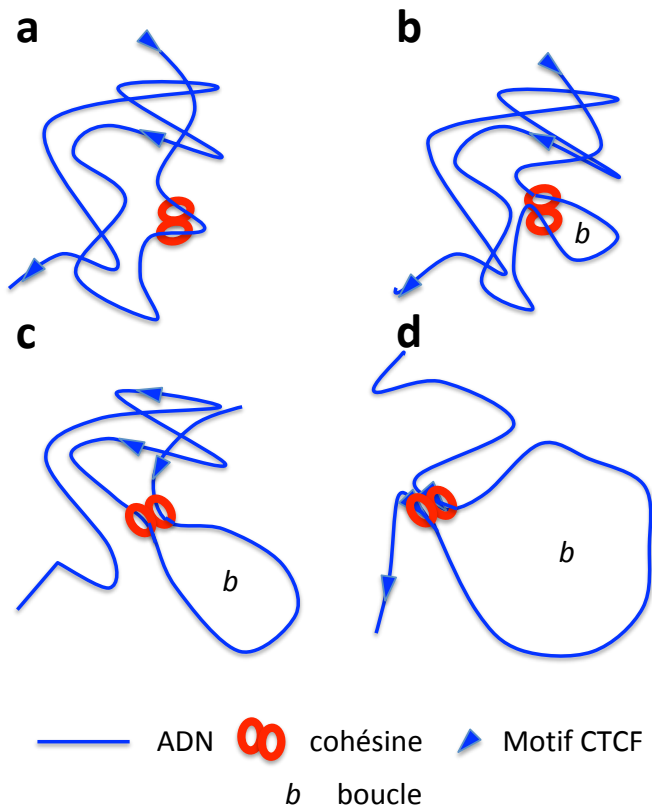
Et bien sûr les batteries de gènes qui ont été intercalées chez la souris et la drosophile sont différentes (Figure 26). Les yeux sont donc différents mais Pax6/Ey reste associé aux yeux chez la souris et la drosophile et dans toutes les espèces simplement et uniquement parce qu'il était le premier de cette chaîne chez l'ancêtre commun.

On passe insensiblement de la différenciation cellulaire (Figure 25a) à la formation d'organes très complexes (Figure 25d). Sans innovation particulière, sans création de nouveaux gènes. Pax6 était un gène de différenciation avant d'être un gène de formation d'organe.

Ces annexions de gènes par Pax6, se font facilement parce que la séquence de reconnaissance de Pax6 sur l'ADN peut être presque n'importe où autour du gène qui sera annexé (Figure 25f). Pax6 active le gène coopté par ce qu'on appelle l'activation distale de la transcription.

Et bien il est temps de dire que l'activation distale n'est apparue ni d'un coup ni très tôt dans l'évolution. Sous sa forme la plus élaborée elle n'est présente que chez l'ancêtre commun aux néphrozoaires (Figure 28).

Voici son mécanisme.



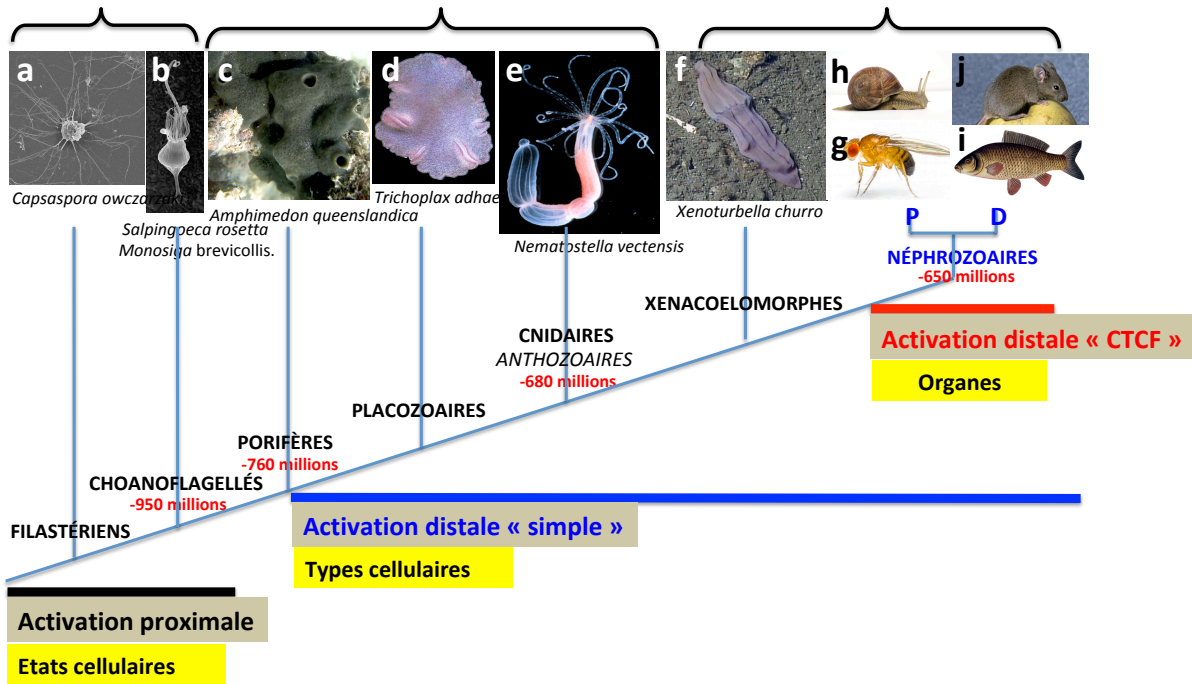
**Figure 27 : La formation d'une boucle par extrusion de l'ADN.** a) l'ADN s'enfile, et b-c), glisse dans les anneaux que forment les cohésines, des protéines, ce qui délimite des boucles (*b*, une seule est représentée). d) les cohésines sont bloquées par les complexes CTCF fixés sur leur motif (les pointes de flèches) Les régions adjacentes à ces motifs, initialement éloignées, se retrouvent proches l'une de l'autre.

La boucle qui rapproche le FT de la machinerie transcriptionnelle sur le début du gène (Figure 12) se fait activement par le biais d'un processus qui utilise deux ingrédients majeurs : la cohésine (ou plus exactement le complexe cohésine) et la protéine CTCF. Le complexe cohésine enfile l'ADN et glisse le long des brins de l'ADN, formant ainsi une boucle (Figure 27a-c). La protéine CTCF se fixe sur la séquence consensus, CCGCGNGGNGGCAG, de l'ADN. Lorsque le complexe cohésine rencontre ce complexe il est bloqué<sup>20 21</sup> (Figure 27d). On trouve toujours des séquences CCGCGNGGNGGCAG à proximité des activateurs distaux et des débuts des gènes<sup>22</sup>. La cohésine les rapprochent l'un de l'autre, au point d'ancrage d'une boucle de l'ADN (Figure 27d).

La cohésine est une très vieille innovation des cellules eucaryotes. Elle est impliquée dans beaucoup de processus de base comme la séparation des deux brins de l'ADN, après la réplication, durant la mitose, la recombinaison et la réparation de l'ADN et bien d'autres NOTE ?.

CTCF est une innovation plus récente. Elle apparaît chez l'ancêtre commun aux néphrozoaires. On compte 100 000 séquences CTCF dans le génome des mammifères, bien plus que de gènes. Ni les premiers bilatériens, les fameux xenacoelomorphes, ni *a fortiori* les méduses, Trichoplax ou les éponges ne l'ont<sup>23</sup>. L'activation transcriptionnelle y est distale mais ils utilisent une forme indépendante de CTCF, que l'on soupçonne être moins performante (le mécanisme n'est pas bien connu). Chez les protistes, l'activation distale n'est pas utilisée. Les sites de fixation des facteurs transcriptionnels sont proches du promoteur. Ils activent proximement la machinerie transcriptionnelle<sup>24</sup>.

Ces trois étapes de l'évolution de l'activation transcriptionnelle (Figure 28, les lignes noire, bleue et rouge) suivent les trois grandes étapes de complexification des organismes (Figure 28, les boîtes jaunes). Et si elles en étaient une des conditions ?



**Figure 28 :** Les étapes de l'évolution de l'activation transcriptionnelle par les FTs. Les FTs sont des innovations d'avant les protistes, de chez les bactéries et les archées. L'activation proximale existe chez les filastériens et les choanoflagellés, les groupes sœurs des animaux (la ligne noire). L'activation distale est une innovation des animaux (la ligne bleue). L'activation distale qui exploite le mécanisme d'extrusion de l'ADN et CTCF est une innovation des néphrozoaires qui vient après les bilatériens les plus simples, les xenacoelomorphes (la ligne rouge). Pour les autres éléments voir la figure 7.

### L'essentiel

La formation des organes nécessite, tout comme la différenciation cellulaire, des processus qui créent de l'hétérogénéité entre des cellules identiques, associés à une expression différentielle de FTs qui contrôlent des batteries de gènes et interagissent entre eux. Des arrangements des FTs entre eux émergent de nouvelles propriétés telles que séquentialité et temporalité de l'expression des gènes. Les organes se complexifient en utilisant certains FTs qui, de ce fait, acquièrent le statut de gène maître. L'élément majeur de la formation des organes est la cooptation de batteries de gènes. Un processus qui en augmente l'efficacité est apparu à la base des néphrozoaires.

# IV.

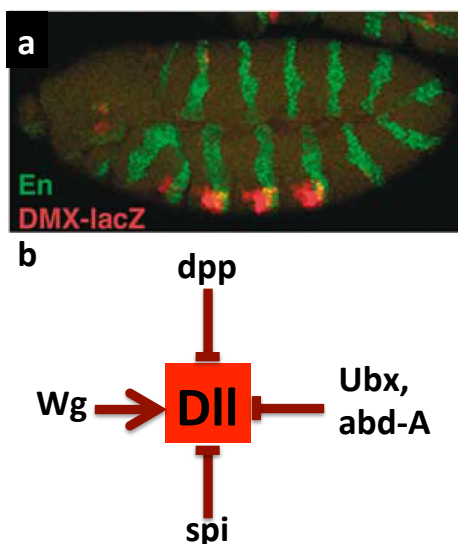
# LE POSITIONNEMENT DES ORGANES

**Au cours du développement, le positionnement des organes se fait par la localisation génétique de l'expression des gènes maîtres des organes : des facteurs de transcription activent et/ou inhibent ces gènes dans les territoires cellulaires de l'embryon. Parallèlement, dans ces territoires, de multiples opérations cellulaires forment les structures de l'embryon.**

**Les territoires de l'embryon ont été sélectionnés par l'évolution pour former les structures de l'embryon. Les gènes maîtres utilisent les territoires pour positionner les organes en cooptant certains de leur FTs. Les combinaisons de gènes maîtres particuliers, les gènes Hox, sont adaptées à la formation des séries d'éléments légèrement différents les uns des autres que sont les territoires répétés disposés antéro-postérieurement chez les bilatériens.**

L'un des enseignements des expériences de Walter Gehring est que des yeux peuvent se former un peu partout dans la drosophile, sur les antennes, les pattes, les ailes etc. Il suffit pour cela que le gène *Ey/Pax6* s'exprime dans les tissus à l'origine de ces organes. Il y a donc pour l'embryon un vrai problème de positionnement des organes.

Peut-on, avec les éléments que nous avons rencontrés, résoudre ce problème ? Une idée simple serait que l'activation de l'expression des gènes maîtres soit strictement limitée à une petite région de l'embryon. Cela ne nécessiterait que des facteurs de transcription activateurs et répresseurs et des interactions intercellulaires, des ingrédients que nous connaissons. Et bien de fait, c'est ce qui est observé.

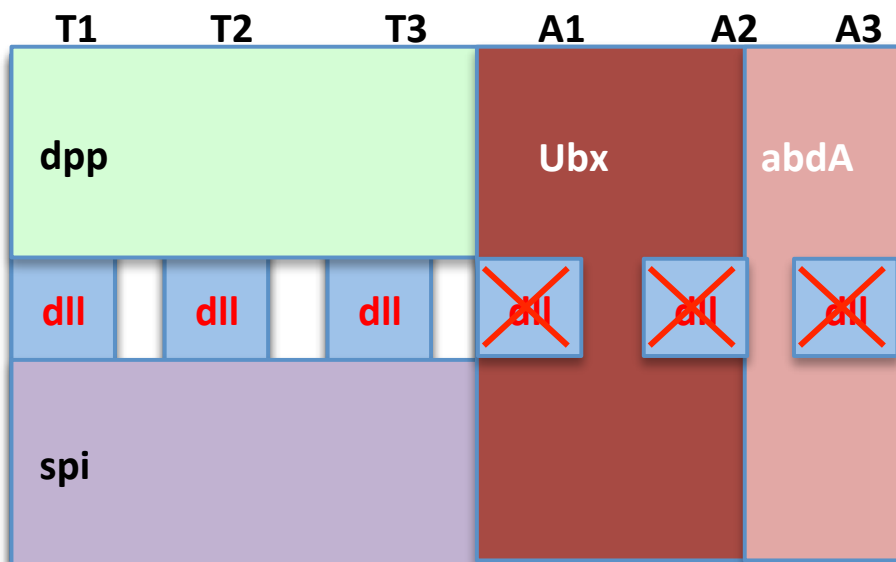


Pour illustrer cette idée prenons l'exemple d'un autre gène maître, celui des pattes de la drosophile, *Distal-less (Dll)* qui contrôle aussi beaucoup d'autres appendices chez les animaux<sup>25</sup>. C'est encore un exemple d'un gène conservé au cours de l'évolution. Dans la larve de la drosophile, il est exprimé exclusivement dans trois petits secteurs des trois segments thoraciques T1, T2 et T3 (Figure 29a, les trois spots rouges<sup>26</sup>).

**Figure 29 : Le contrôle de l'expression de Distal-less (Dll) dans les segments T1, T2, et T3 de la larve de la drosophile.** a) les trois spots rouges signalent l'expression de *Dll*. Les bandes transverses vertes signalent l'expression de *Wg* ; b) *Dll* est activé par *Wg* et réprimé par *dpp*, *Ubx*, *abd-A* et *spi*. Les domaines d'expression de ces gènes sont indiqués sur le schéma de la figure 30.

Les généticiens ont montré que *Dll* est activé dans ces trois segments par un morphogène, *Wingless* (Figure 29b, la flèche). Mais *Wingless* est exprimé dans tous les segments en bandes transverses (Figure 29a, le vert). En conséquence, *Dll* devrait être activé en bandes transverses dans tous les segments. Deux autres morphogènes, *dpp* (*Decapentaplegic*, homologue vertébré *BMP*) et *spi* interviennent. *Dpp* s'exprime dans un large secteur horizontal dorsal (Figure 30, le vert pâle exprimé dans tous les segments) et

inhibe Dll dans tous les segments. Spi, le récepteur d'un autre morphogène, s'exprime dans un large secteur horizontal ventral et inhibe aussi Dll dans tous les segments (Figure 30, le violet). Ne reste donc qu'un petit secteur intermédiaire, dans lequel ni Dpp ni spi ne s'expriment. Dll activé par Wg peut donc s'exprimer (les secteurs bleus). Dans les segments abdominaux deux facteurs de transcription vont réprimer directement dll : Ubx et AbdA (Figure 30, les secteurs marron foncé et marron clair et les secteurs bleus barrés). Ne restent donc activés que les trois petits secteurs des segments thoraciques (Figure 29a, les spots rouges, Figure 30, les trois secteurs bleu non barrés).



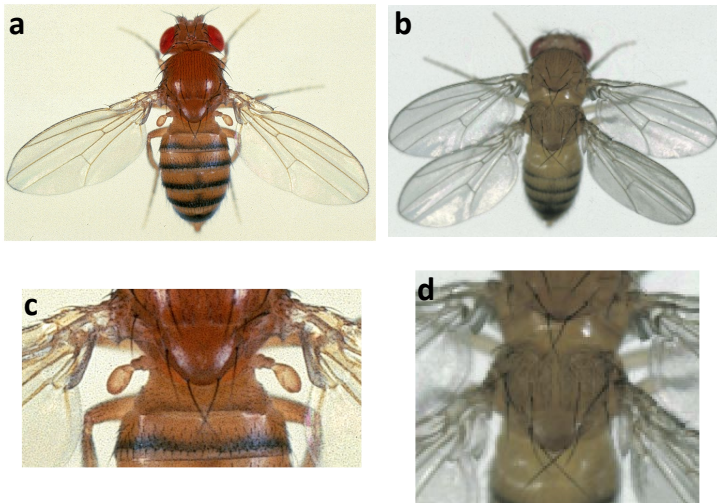
**Figure 30 : Les domaines d'expression des gènes qui positionnent Dll.** Représentation schématique des segments thoraciques (T1-T3) et des trois premiers segments abdominaux (A1-A3). Activation de Dll (par Wingless), inhibition

dorsale (par dpp), inhibition ventrale (par spi), inhibition dans les segments abdominaux (par Ubx et abdA)

La logique du positionnement des organes est donc celle-ci : l'expression des gènes maîtres des organes est localisée par une activation et de multiples répressions dans les territoires cellulaires de l'embryon. Pour dll trois morphogènes et deux FTs, Ubx et AbdA interviennent.

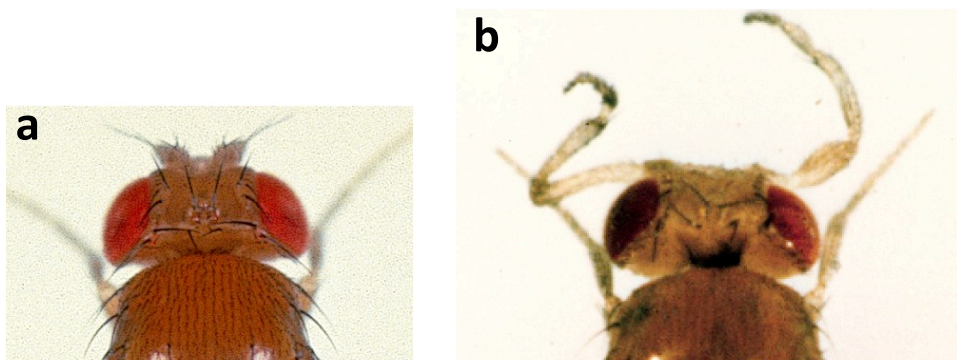
Que sont tous ces gènes ? A quoi servent-ils ?

Commençons par les FTs Ubx et AbdA.



**Figure 31 : Le mutant ultrabithorax.** a et c) la mouche normale ; b et d) le mutant à quatre ailes.

Ces gènes-là sont connus depuis longtemps<sup>27</sup>. Le mutant appelé ultrabithorax (Ubx), isolé en 1915, présente quatre ailes (Figure 31a, la mouche normale et 31b, le mutant). Le segment T3 qui, normalement, porte les balanciers (Figure 31c), des organes qui servent à la drosophile à équilibrer son vol, est chez le mutant transformé en un segment T2 qui porte des ailes (Figure 31b et d).



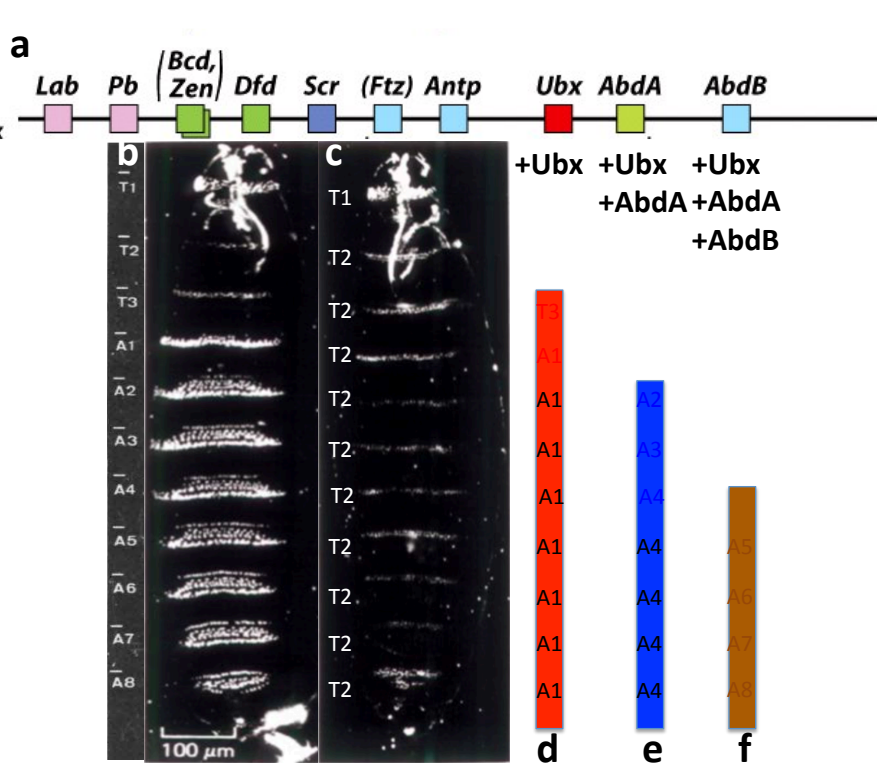
**Figure 32 : Le mutant antennapedia.** a) la mouche normale ; b) le mutant.

mutant.

D'autres mutants qui transforment un territoire de la mouche en un autre ont été décrits. Le mutant appelé antennapedia porte des pattes à la place de ses antennes (Figure 32a et b).

Les généticiens parlent d'un changement de l'identité d'une région du corps. C'est à première vue différent des gènes maîtres, Pax6 ou Dll, qui, mutés donne un organisme où l'organe n'est pas fait (absence des yeux, des pattes), et n'est pas remplacé.

Ces mutants ont fait couler beaucoup d'encre. Peut-être en avez-vous entendu parlé : ce sont les gènes Hox<sup>15 28</sup>. Mais, première bonne nouvelle,



on sait depuis 1983<sup>29</sup> (D. Hogness) que ce sont des facteurs transcriptionnels. Nous restons donc en terrain connu.

Chez la drosophile il y a dix gènes Hox (Figure 33a). On parle du « complexe Hox ».

**Figure 33 : Les gènes**

**Hox chez la drosophile.** a) le complexe Hox comprend dix gènes. b) la larve normale et les segments T1 à A8 ; c) la larve triple mutante Ubx et AbdA et B ; d) la larve restaurée pour Ubx AbdA et B ; e) la larve restaurée pour Ubx et AbdA ; f) la larve restaurée pour Ubx, AbdA et B. En rouge, bleu et marron les profils d'expression de Ubx, AbdA et AbdB.

Si ce dans complexe Hox l'on mute les gènes qui nous intéressent (Ubx, AbdA et AbdB) tous les segments abdominaux de la larve (Figure 33b) sont changés en segments thoraciques T2 (Figure 33c). Et donc effectivement, la mouche devrait avoir des pattes tout le long du corps.

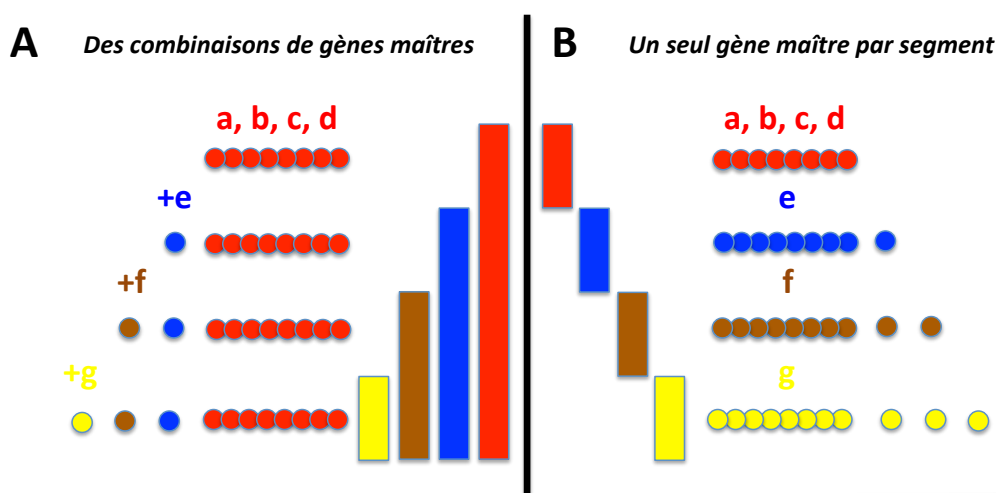
Chez ce mutant Si l'on restore le gène Ubx, alors les segments T3 et A1 réapparaissent mais les autres subissent encore la transformation en type A1 (Figure 33d). La mouche ne devrait plus avoir de pattes que dans les segments thoraciques. Si en plus l'on restore AbdA les segments 5 à 7, réapparaissent (A2, A3 et A4) mais les autres subissent encore une transformation en type A4 (Figure 33e) et enfin si on rajoute AbdB les segments A5 à A8 réapparaissent (Figure 33f).

Cela semble magique ! Qu'en est-il ?



Voyons les profils d'expression de ces gènes. Ubx s'exprime dans tous les segments postérieurs à partir de T3 (Figure 33d, en rouge) ; AbdA s'exprime dans tous les segments postérieurs à partir du A2 (Figure 33e, en bleu) ; et enfin AbdB s'exprime dans tous les segments postérieurs à partir de A5 (Figure 33f, en marron). En conséquence, dans T3 et A1, Ubx s'exprime seul. Dans A2, A3 et A4, Ubx et AbdA s'expriment tous les deux. Dans A5, A6, A7 et A8, les trois gènes, Ubx, AbdA et AbdB s'expriment ensemble. A chaque territoire sa propre combinaison. La clé de cette magie devient apparente : ces gènes agissent en association.

Détaillons cette hypothèse à l'aide de la figure 34A. En l'absence des gènes Ubx, AbdA et AbdB, les segments (représentés par les rectangles verticaux) présentent tous une même structure T2 (le rectangle rouge) contrôlée par les facteurs de transcription a, b, c, d qui s'expriment dans tous les segments (les 8 ronds rouges). Ubx rajoute des caractères (le rond bleu), contrôlés par e. Il les rajoute dans tous les segments où il s'exprime, c'est-à-dire tous les segments postérieurs au premier segment rouge. : il les transforme en segments bleus (le rectangle bleu) qui expriment a-d + e. AbdA rajoute encore d'autres caractères (le rond marron), contrôlés par f, qui les transforment en segments marron qui expriment a-d + e + f (le rectangle marron). Et AbdB rajoute encore d'autres caractères (le rond jaune), contrôlés par g, qui les transforment en segments jaune qui expriment a-d + e + f + g (le rectangle jaune). On voit bien que si l'on retire g les segments caudaux deviennent marron (a-d + e + f), si l'on retire f, ils deviennent bleus (a-d + e) et si l'on retire Ubx tous les segments deviennent rouges (a-d). Exactement comme ce qui est observé (Figure 33).



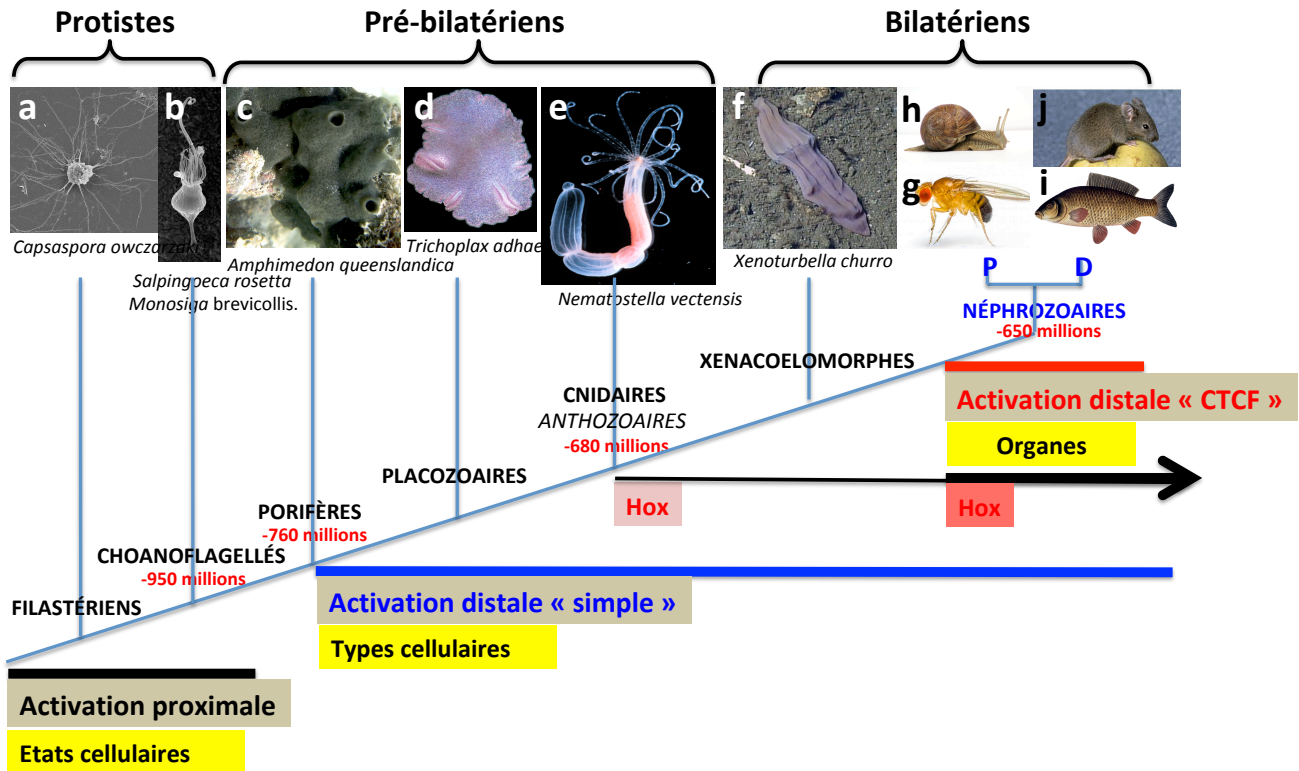
**Figure 34**  
Pour former une série d'éléments légèrement différents les uns des autres, le plus simple est de rajouter des

**facteurs de transcription à une combinaison de facteurs de transcription déjà existante (A).** (B) L'alternative : un gène maître par segments (voir le texte).

Un fait intéressant apparaît clairement : les gènes Ubx et AbdA et AbdB ont au final les caractéristiques de gènes maîtres, de respectivement e, f, et g. Tout comme la mutation de Pax6 fait disparaître les yeux et celle de Dll les pattes, les mutations de Ubx, AbdA ou AbdB font disparaître les caractères e, f ou g. Simplement, leur action s'inscrit dans une combinatoire : pour un segment donné, il y a plusieurs gènes maîtres.

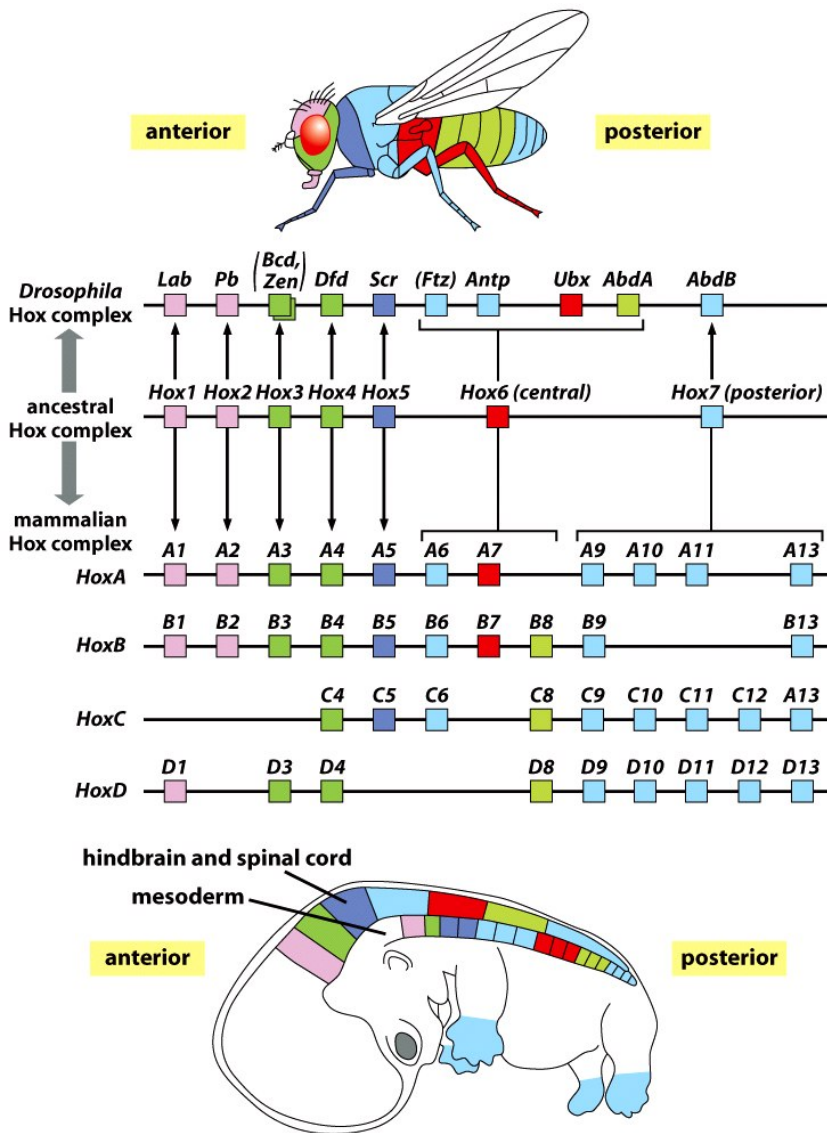
Apparemment, tout cela semble complexe. Pourquoi pas, plus simplement, un gène maître par segment ? Détaillons cette hypothèse à l'aide de la figure 34B en considérant les mêmes FTs a-g que précédemment. Avec un gène maître par segment, la situation serait la suivante (figure A, à gauche) : pour le premier segment, a-d contrôlent ces 8 ronds rouges, comme dans la situation combinatoire. Mais pour le segment bleu, le FT e devra contrôler, outre le rond bleu mais aussi les 8 ronds rouges ; e doit donc être raccordé aux 8 facteurs de transcriptions rouges. 9 événements indépendants doivent se produire contre un seul pour la situation combinatoire. Même chose pour le segment marron, 10 événements indépendants contre un pour la situation combinatoire et ainsi de suite. Clairement le système de type combinatoire est beaucoup plus simple à mettre en place ! Pour former une série d'éléments légèrement différents les uns des autres, le plus simple est de rajouter des facteurs de transcription à une combinaison de facteurs de transcription déjà existante. Et bien c'est le plus simple qui apparaît dans l'évolution.

Au regard d'un développement particulier cela pourrait sembler un détail, parmi les milliers d'opérations qui ont lieu. Est-ce bien le cas ? C'est, de nouveau, l'évolution qui nous donne la réponse. Ce système de spécification de territoires répétés, les segments chez la drosophile, est retrouvé chez tous les insectes, tous les arthropodes, chez l'amphioxus, les poissons, chez la souris et l'homme. En fait, il est déjà présent chez l'ancêtre commun aux bilatériens et aux cnidaires (Figure 35, la boîte rose).



**Figure 35 : Le complexe de gènes Hox apparait chez les cnidaires (la boîte rose).** Il se complexifie ensuite chez les bilatériens (la boîte rose foncée et voir figure 36). Pour les autres éléments voir les figures 7 et 28.

Ce complexe Hox est conservé, mais bien sûr modifié (Figure 35, la boîte rouge et Figure 36), jusqu'aux drosophiles et à nous, qui avons utilisé ces gènes pour spécifier, comme la souris, les régions de notre corps depuis la nuque jusqu'à la partie la plus caudale de notre corps<sup>30</sup> (Figure 36, le schéma du bas). Chez les deutérostomes, comme l'amphioxus, il est resté quasi inchangé, mais chez les vertébrés il a été dupliqué deux fois. Le problème de la localisation de l'expression des gènes Hox dans les segments est d'une redoutable complexité<sup>28,31</sup>.



**Figure 36 : Le complexe de gènes Hox ancestral aux bilatériens et son évolution chez la mouche (en haut) et la souris (en bas).** Sur les schémas de la mouche et de la souris les profils d'expression des gènes sont représentés suivant le code couleur.

La grande diversité des formes passées et actuelles d'animaux, (vers, mouches, vertébrés et autres) est en partie due à ce système de spécification somme toute assez simple.

Les autres facteurs impliqués dans la régulation de Dll sont Wg, Dpp et spi (Figure

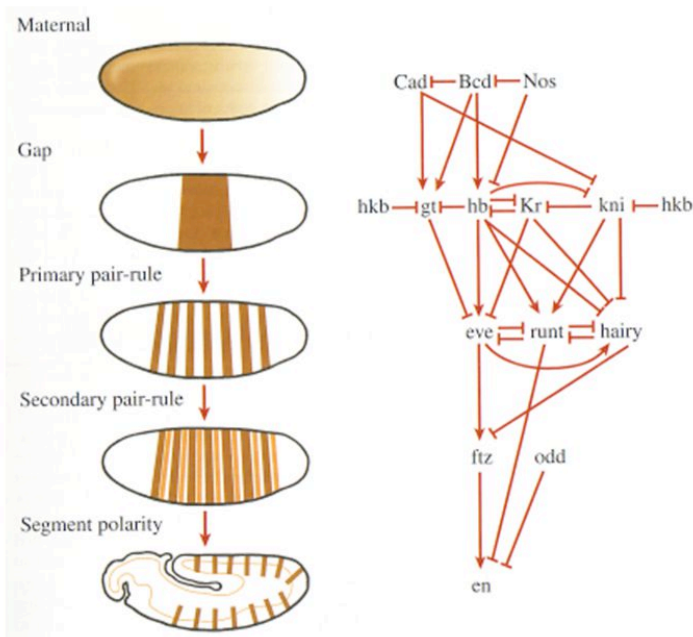
29b). Wg et Dpp sont des morphogènes. Nous avons vu comment les morphogènes créent des territoires (Figure 23). C'est encore le cas ici. Wg crée un large territoire transverse dans chacun des segments de la larve (Figure 29a, les bandes transverses vertes). Dans chacun de ces territoires transverses, Dpp crée un territoire dorsal et Spi crée un territoire ventral (Figure 30). Dans ces deux territoires Wg + dpp dorsal et Wg + spi ventral, l'expression de Dll est exclue. Les conditions pour l'expression de Dll ne sont donc réunies que dans le territoire Wg intermédiaire.

Cela ne signifie pas que l'embryon crée des territoires pour positionner les organes. Ces morphogènes ont leur propre rôle dans le développement : ils contrôlent les opérations cellulaires qui aboutissent à la structure de la larve. Wingless est impliqué dans les opérations antéro-postérieures du développement de l'embryon de la drosophile. Dpp est impliqué dans les

opérations dorso-ventrales. Les organes utilisent ces contrôles existants pour se positionner. Ils cooptent Wg, dpp, spi par le mécanisme que nous avons rencontré plusieurs fois, l'activation distale. Celles pour Wingless, l'activateur de Dll, tirent leur origine des hétérogénéités moléculaires de l'œuf de la drosophile avec Bcd, Cad et Nos et aboutissent à l'expression d'engrailed, le FT qui active Wingless, dans les 14 segments (Figures 23 et 29). L'annexe 1 explique comment tous ces mutants ont été découverts grâce à un crible génétique d'une remarquable simplicité. L'annexe 3 examine le réseau que ces gènes constituent.

Nous sommes remontés à l'origine du développement, l'œuf. Nous avons une idée de la continuité des événements qui va du positionnement du gène maître Dll à l'œuf (Figures 29 et 37) et l'exemple de Pax6 nous a donné une idée de la chaîne qui va du gène maître à l'organe (Figure 22). Les principes qui sous-tendent les trois opérations du développement, la différenciation des cellules, la formation des organes et leur positionnement, qui nous ont révélé cette continuité, nous sont aussi maintenant connus. Dans le dernier chapitre nous verrons comment ces principes s'articulent dans la fabrique des organismes à partir de l'œuf.

**Figure 37 : Les cascades de facteurs de transcription qui contrôlent les gènes des segments de la drosophile.** A droite le réseau génétique ; à gauche les stades de développement et les profils d'expression des gènes concernés, de l'œuf à la larve segmentée. Voir le texte et les annexes 1 et 3.



**L'essentiel**  
De multiples FTs localisent dans l'embryon l'expression des gènes maîtres des organes. Les éléments répétés des bilatériens (segments, territoires antéro-postérieurs) utilisent des combinaisons de gènes maîtres très particuliers, organisés en un complexe, conservé au cours de l'évolution, le complexe Hox.





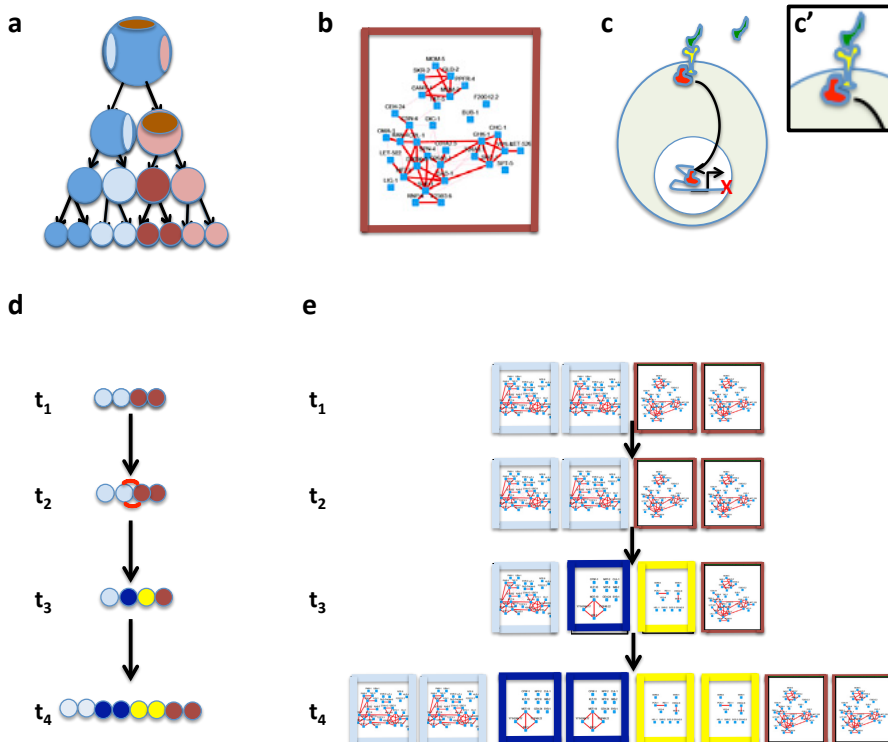
# V.

# LA FABRIQUE D'UN ORGANISME

L'œuf est, très généralement, hétérogène. La fécondation active les divisions cellulaires ce qui entraîne la ségrégation de certaines molécules dans les premiers territoires de l'embryon. Ces molécules activent des combinaisons différentes de FTs. Sont ainsi engagées des interactions intercellulaires entre les premiers territoires, ce qui génèrent de nouveaux territoires avec de nouvelles combinaisons de FTs. Les combinaisons de FTs définissent les opérations cellulaires qui forment les structures de l'embryon. Les gènes maîtres qui contrôlent la formation des organes et la diversification des éléments répétés sont activés durant ces phases. Le développement explore des milliers de combinaisons de FTs qui correspondent à autant d'états du réseau de régulation génétique. Leur déploiement temporel procède à la fois des interactions entre cellules et de la configuration du réseau génétique. La configuration du réseau de régulation donne une apparente direction au développement. Les interactions entre cellules sont la véritable nouveauté des développements. En créant de nouveaux territoires cellulaires elles complexifient et l'embryon et le réseau génétique.

Dans ce dernier chapitre nous montrons comment les quelques principes à l'œuvre dans la différenciation des cellules, la formation des organes et leur positionnement, s'articulent dans la fabrique des organismes. Déroulons l'histoire depuis l'œuf jusqu'aux organes, c'est à dire dans l'ordre, non pas de l'évolution, mais du développement.

Après la fécondation, l'œuf se divise selon le schéma indiqué (Figure 38a et 39, les flèches noires). Pour suivre facilement les filiations, les cellules de l'embryon, qui forment une structure 3D (Figure 1-4), ont été représentées selon une chaîne linéaire (les chaînes horizontales). En général l'œuf est hétérogène : certaines molécules sont localisées exclusivement à l'un des pôles (le marron foncé de la figure 38a), d'autres à « l'avant » ou à « l'arrière » (les couleurs bleu pâle et marron clair de la figure 38a). Ces hétérogénéités sont essentielles, elles préfigurent les premiers territoires cellulaires de l'embryon. En effet, les divisions cellulaires ségrègent ces molécules dans des groupes de cellules distincts de l'embryon (Figure 38a, les groupes de cellules de même couleur). Les premières combinaisons de facteurs de transcription (Figure 38b, les petits points bleu arrangés en réseau) émergent de l'activité de ces molécules (Figure 38e,  $t_1$ , Figure 39b ; les boîtes bordées de bleu, bleu pâle, marron foncé, marron clair).



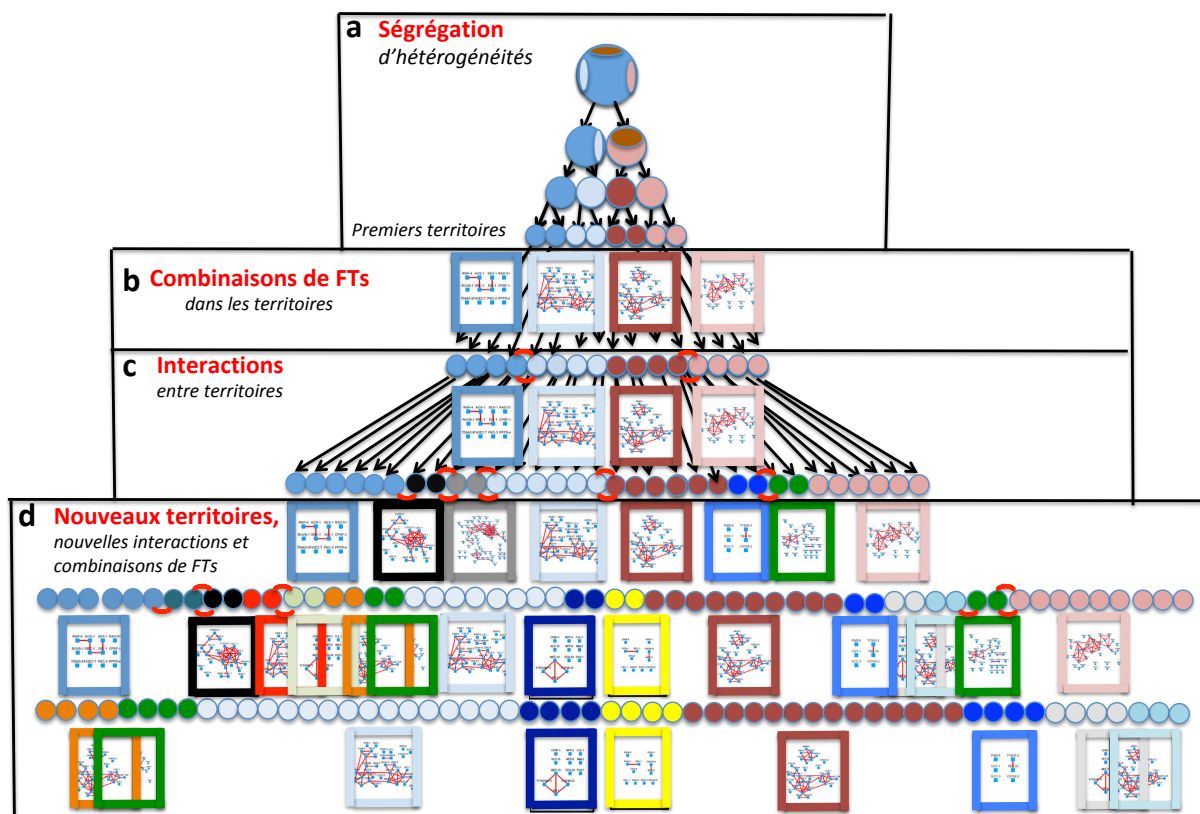
**Figure 38 : Les éléments majeurs des développements.** a) les divisions (les flèches) des cellules (les ronds colorés) ségrègent des molécules réparties de manière hétérogène dans l'œuf (les couleurs bleu, bleu pâle, marron, marron clair) générant ainsi les premiers territoires cellulaires de l'embryon; b) à chaque territoire correspond une combinaison particulière de FT (les points bleus et les flèches rouges



dans la boîte bordée de la couleur correspondant à celle des cellules des territoires, ici marron) ; c) et c') les interactions cellulaires via un morphogène (c', en vert) qui diffuse hors des cellules et se fixe sur les récepteurs membranaires des cellules voisines (c', en jaune), ce qui active alors un facteur transcriptionnel (c', en rouge) qui après transfert dans le noyau de la cellule (c, la flèche) active un ou plusieurs nouveaux gènes (X) ; d) ces interactions cellulaires (t<sub>2</sub>, les 2 flèches rouge entre les cellules qui marquent une interaction réciproque) génèrent 2 nouveaux territoires (en bleu foncé et jaune, t<sub>3</sub> puis t<sub>4</sub> après une division des cellules) ; e) les combinaisons de FTs qui caractérisent les cellules de d) avant (t<sub>1</sub> et t<sub>2</sub>) et après l'interaction (t<sub>3</sub> et t<sub>4</sub>).

Puis des interactions (Figure 38c) entre les cellules de ces premiers territoires (Figure 38d, les petites flèches rouges entre les cellules) génèrent de nouveaux territoires (Figure 38d, les couleurs bleu foncé et jaune). De nouvelles combinaisons de FTs les caractérisent (Figure 38e, t<sub>3</sub>, t<sub>4</sub>, les boîtes bordées de bleu foncé et jaune).

Ces combinaisons de facteurs de transcription déterminent les comportements cellulaires des territoires qui, ensemble, effectuent les multiples opérations cellulaires du développement. Les comportements cellulaires incluent des changements de forme ; la mobilité (adhésivité) ; les mouvements (dispersion, directionnalité, vitesse) ; les divisions, leur orientation (orientées, asymétriques, inégales, ou non), leur vitesse ; etc. Les opérations cellulaires incluent bien sûr la formation des organes, leur positionnement, la différenciation cellulaire, mais aussi la gastrulation, l'élongation de l'embryon et bien d'autres. De ces opérations du développement émergent des structures. La première visible, correspond à une sphère simple appelée blastula (Figure 1b). La deuxième visible, correspond à une sphère formée de deux couches de cellules et une ouverture, appelée gastrula (Figure 1d). Tous les développements depuis les cnidaires passent par ces stades blastula et gastrula. D'autres exemples de structures plus tardives : les segments chez la larve de la drosophile (figure 3d), la plaque neurale, le tube neural (figure 2d-f), les structures de toutes sortes d'organes (figure 4c, d).



**Figure 39 : La fabrique d'un organisme.** a) les divisions des cellules (les flèches noires) ségrègent des molécules réparties de manière hétérogène dans l'œuf (les couleurs bleu, bleu pâle, marron, marron foncé) dans les premiers territoires de l'embryon ; b) à chaque territoire correspond une combinaison particulière de FTs (les boîtes bordées de la couleur des cellules des territoires auxquelles elles correspondent) ; c) des interactions entre cellules différentes (les petites flèches rouges sous les ronds à partir du stade 16 cellules) créent de nouveaux territoires (le noir, le gris, le bleu, le vert) ; d) à ces nouveaux territoires correspondent de nouvelles combinaisons de FTs (les boîtes bordées de noir, gris, bleu et vert) qui permettent de nouvelles interactions. Les gènes maîtres qui contrôlent la formation des organes et la diversification des territoires répétés (les gènes Hox), sont activés durant ces phases selon les modalités de la figure 30.

Sans cesse de nouvelles interactions entre cellules créent de nouveaux territoires où sont activées d'autres combinaisons de facteurs de transcription qui déterminent encore d'autres opérations du développement (Figure 39d). Les gènes maîtres qui contrôlent la formation des organes et la diversification des territoires répétés (segments chez la drosophile) par les combinatoires des gènes Hox sont activés durant ces phases. Ils poursuivent le processus en utilisant les mêmes principes et souvent les mêmes gènes.



Les opérations du développement nous donnent l'impression que l'embryon devine le futur, place les organes aux bons endroits ; c'est le paradoxe levé au début de ce livre. On comprend maintenant que c'est le résultat de la structure du réseau de régulation acquise au cours de l'évolution et non d'une magie ou d'une force mystérieuse. Les cooptations y jouent un grand rôle. Le meilleur exemple est l'évolution des organes par l'intercalation (Figure 25). Entre des gènes clés comme les gènes maîtres et leurs contrôles initiaux sont cooptés toutes sortes d'autres gènes qui font ainsi partie de la hiérarchie qu'ils contrôlent. L'intercalation repousse la différenciation cellulaire à la fin du processus. La structure du réseau de régulation donne une direction (en fait des directions) au développement. C'en est même une des propriétés émergentes les plus notables.

Les successions de combinaisons de facteurs de transcription dans les territoires qui se forment (les successions verticales des boîtes de la figure 39) correspondent à autant d'états du réseau de régulation génétique. Le développement explore des centaines (plus probablement des milliers) d'états du réseau qui s'enchaînent. Ces suites d'états du réseau tirent leur dynamique des interactions entre cellules et de la structure du réseau génétique. La véritable nouveauté des développements, ce fut l'émergence des interactions entre cellules qui par les voies de signalisation relaient les signaux entre cellules au réseau génétique (Figure 38c). Les interactions entre cellules sont apparues avec les éponges il y a plus de 700 millions d'années avec les premières différenciation cellulaires (voir l'annexe 5).

Les voies de signalisation sont à l'origine de nouveaux territoires cellulaires dans lesquels se déploient de nouveaux arrangements du réseau de régulation des gènes (Figure 39d). Elles permettent donc la complexification et de l'embryon et du réseau génétique (Figure 39d).

Avec les développements des organismes, un niveau de complexité supplémentaire est apparu. Avec les nouveaux territoires et la différenciation cellulaire on passe du niveau des cellules individuelles au niveau des groupes de cellules hétérogènes ; puis avec la formation des organes, basée sur des opérations du développement et la différenciation

cellulaire, au niveau des organismes. C'est sur ces nouveaux niveaux que l'évolution joue.

Au réseau que constituent les 1000 à 2000 gènes de régulation chez les animaux, correspond un nombre d'états potentiels infinis ( $2^{2000} = 10^{600}$  ; rappel : nombre de galaxies dans l'univers,  $10^{11}$ , nombre de particules dans l'univers,  $10^{80}$ , nombres d'espaces de Calabi-Yau,  $10^{500}$ , théorie des cordes<sup>32</sup>) du réseau génétique.

Et bien vous, moi, les éponges, les mollusques, les poissons, sommes des réalisations d'un jeu de quelques-uns de ces états, parmi une infinité de possibles.

### **L'essentiel**

**Les interactions entre cellules sont la véritable nouveauté des développements. En créant de nouveaux territoires cellulaires elles complexifient et l'embryon et le réseau génétique.**



# CONCLUSION

**Ces principes qui ne font plus débats, les voici :**

Centrales au développement, les communications entre cellules multiplient les territoires. Elles complexifient et l'embryon et le réseau génétique.

Le développement requiert le réseau génétique. Ce réseau organise les batteries de gènes, les combine entre elles. Des arrangements de gènes entre eux, des circuiteries qui le constituent, émergent des propriétés cellulaires et supra-cellulaires essentielles au développement, notamment lorsque sont concernés les interactions des FTs entre eux.

Les territoires de l'embryon sont caractérisés par des états d'activité distincts du réseau génétique. Ils spécifient leurs comportements.

Le développement sollicite ces comportements. Ils sont constitués en batteries de gènes qui agissent ensemble et/ou séquentiellement. Ces actions affectent simultanément toutes les cellules d'un territoire et plus rarement des cellules isolées.

Les opérations du développement émergent de combinaisons de ces actions.

La dynamique du développement est celle des subdivisions des territoires, très souvent dues à des interactions entre cellules.

Si nous avons une description complète des interactions cellulaires, des circuiteries génétiques, des comportements cellulaires et des opérations du développement (les éléments du schéma esquissé dans la figure 39), alors nous serions proches d'une description du développement permettant les premières modélisations. Nous en sommes encore loin.

Mais cela ne nous donnerait pas la raison de l'utilisation de tel gène à tel moment. Cette information nécessite de placer les développements dans leur évolution. Cela requiert la comparaison de plusieurs des représentants actuels de certains de nos ancêtres lointains. Nous avons vu des exemples.

A partir des principes qui ont été exposés, on a l'impression que l'on peut proposer des solutions à presque tous les problèmes qui se posent. Mais il y a une limitation. Ces solutions ne peuvent préciser ni quels gènes, ni quels voies de communications entre cellules, ni quelles séquences d'événements sont effectivement impliquées. En effet comme nous l'avons vu, il y a une part d'arbitraire dans ces choix. Cet arbitraire vient de ce qu'un développement se construit au cours de l'évolution à partir d'événements qui se sont produits chez les ancêtres. Ces événements contraignent le développement.

### **Des développements ancrés dans l'évolution**

Une vision « évolution » du développement montre aussi comment ces éléments s'intègrent et quelles furent les innovations (Figure 40).

Les processus de la différenciation des types cellulaires viennent de ceux des états cellulaires de nos ancêtres unicellulaires protistes. (Figure 40, les boîtes jaunes). La fonction ancestrale des FTs et de l'activation transcriptionnelle, fut sans doute de produire des états cellulaires différents chez les protistes. Il est aussi plausible que des voies de signalisation des protistes, qui répondent à des signaux de l'environnement pour produire ces états cellulaires, ont évoluées en des voies de communications entre cellules des animaux qui répondent à des signaux des cellules pour déclencher la différenciation (voir annexe 4).

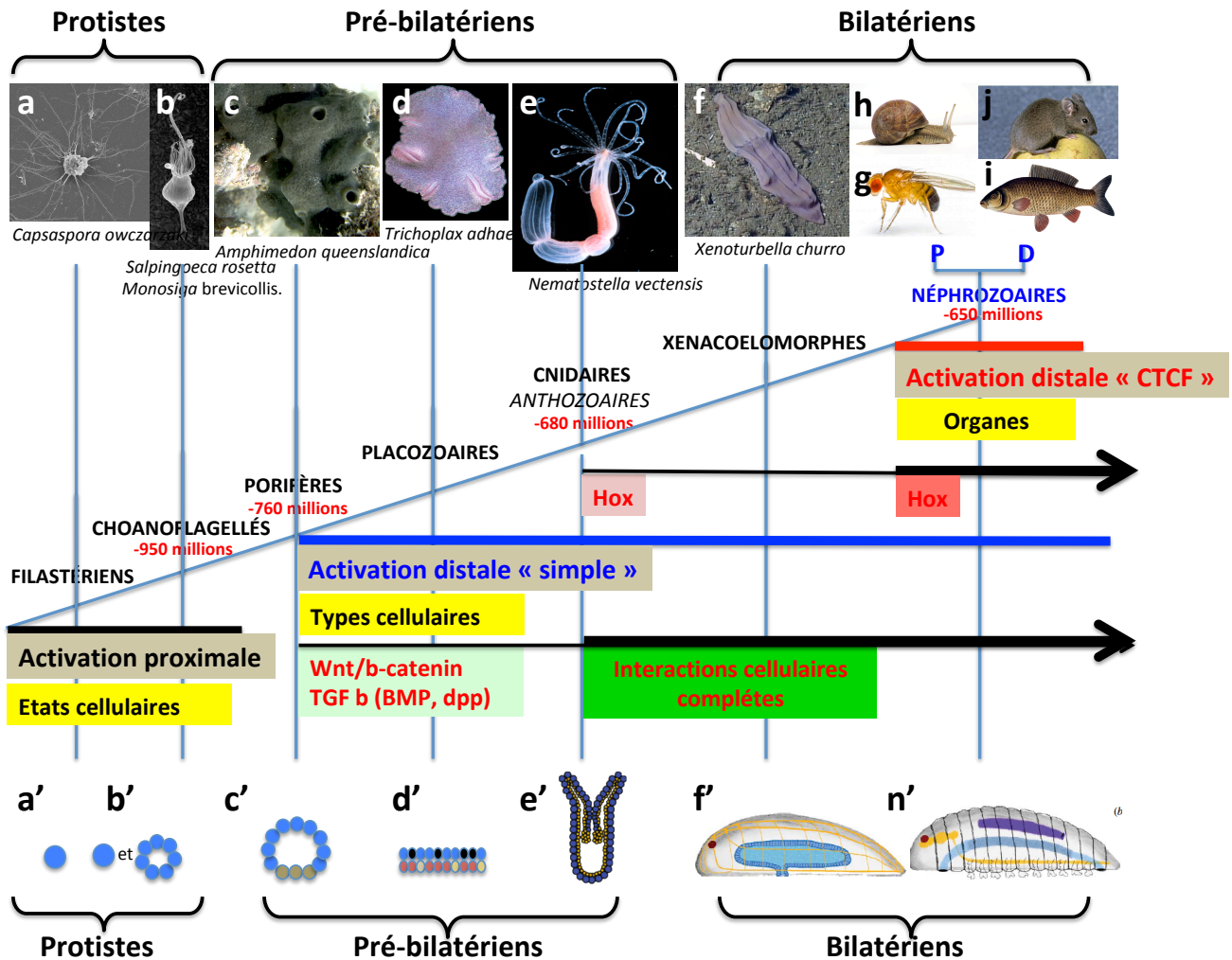
Les processus de la différenciation des types cellulaires évoluèrent ensuite en processus de la formation des organes et de leur positionnement dans les organismes (Figure 40, la boîte jaune « organes »).

Les innovations qui ont accompagnées ces évolutions, furent, la diversification et l'apparition de nouvelles voies de communications entre cellules (Figure 40, les boîtes vert pâle et vert foncé) qui peuvent créer de nouveaux territoires cellulaires dans l'embryon ; la facilitation de la cooptation des gènes (Figure 40, les boîtes Activation distale proximale, Activation distale simple, Activation distale « CTCF ») ; les duplications des gènes du complexe Hox (Figure 40, l'épaississement de la flèche noire Hox) impliqués dans la diversification des territoires répétés à la base des néphrozoaires ; et à coup sûr d'autres qu'il reste à découvrir.

Ces innovations furent nécessaires à la complexification des animaux. Sans les interactions cellulaires, les organismes forment au mieux des colonies de cellules identiques (Figure 40, a' et b'); avec seulement l'activation distale et des voies de signalisation rudimentaires les organismes comme les éponges et les placozoaires correspondent au mieux à des multicellulaires très simples assimilables à des blastula, ne présentant que quelques types cellulaires différenciés (Figure 40, c' et d') et les cnidaires et les xenacoelomorphes malgré une grande diversité d'apparence correspondent à des sortes de « gastrula étirées » (Figure 40, e' et f') ; par contre avec l'activation distale CTCF, la douzaine de voies de signalisation complètes, les duplications des gènes du complexe Hox et aussi l'émergence du mésoderme, les organismes ont acquis une structure de base complexe avec un tube digestif à deux ouvertures, des territoires répétés, des organes yeux, cœur, des appendices (Figure 40, n'). Y-eut-il une seule (CTCF ?) ou plusieurs conditions qui fût ou fûrent décisives reste incertain. Mais cet héritage d'un bilatérien simple (lui aussi encore mal identifié) a permis l'apparition des très nombreux nouveaux phylums qui constituent les protostomes et les deutérostomes.



**Figure 40 : Position dans l'arbre du vivant des innovations discutées dans ce livre.** *Sont figurés en haut* quelques-uns des représentants actuels de nos ancêtres depuis les protistes (filastériens et choanoflagellés, a et b) jusqu'aux bilatériens complexes, les néphrozoaires (P, protostomiens g et h et D, deutérostomiens, i et j). La ligne bleue céruléen oblique figure le temps ; au dessus de cette ligne sont indiquées les dates d'apparition de quelque uns des groupes d'animaux mentionnés. Les boîtes jaunes indiquent l'existence d'états cellulaires (chez les protistes), de types cellulaires différenciés (chez les éponges) et celle d'organes (chez les néphrozoaires). Les apparitions de l'activation proximale, l'activation distale simple et l'activation distale CTCF sont notées. L'origine des voies de signalisation, les boîtes vertes et du complexe génétique Hox, les boîtes rouges, est indiquée. (voir aussi les figures 7, 28 et 35 de l'opuscule). *Sont représentés en bas*, les degrés de complexité cellulaire atteint par chacun de ces ancêtres. a' et b' unicellulaires et phase d'agrégation ; pour c' et d' multicellulaires et types différenciés chez les éponges et les placozoaires ; pour e' et f' cavité digestive chez les hydres et anémones de mer et chez les bilatériens xenacoelomorphes; pour les bilatériens protostomiens et deutérostomiens (néphrozoaires), n', tube digestif à deux ouvertures, des territoires répétés, des



organes.

### **Y-a-t-il des principes majeurs manquants ?**

Le degré de précision que l'on a dans notre compréhension du développement est extrêmement faible. Une modélisation du développement du plus simple des animaux n'est pas encore envisageable. Sans modèle, il est téméraire d'estimer si des principes majeurs manquent. Cependant, les combinaisons infinies que permettent : l'activation distale de la transcription ; la diversification des réseaux génétiques par les communications des cellules ; les propriétés émergentes émanant des motifs de réseau et des combinaisons des comportements cellulaires, semblent suffisantes.

De plus, comme nous l'avons vu chez la drosophile, les grandes lignes de séquences significatives de son développement, comme celles qui vont de l'œuf à l'activation d'un gène maître, *dll*, ou la formation des organes des segments mésodermiques sont déjà esquissées et même en partie modélisées (voir les annexes 3 et 4) . Elles n'utilisent que ces éléments. Se pourrait-il qu'il n'y ait plus de grosses surprises à attendre de l'étude des développements ?

### **D'où viendront les surprises ?**

Il est certain que de grosses surprises continueront de venir de ce que nous apprendrons de l'évolution. La liste des problèmes sans solution est longue. On ne sait pas comment les 100 000 sites de fixation de CTCF se sont répandus dans le génome. Idem pour la constitution de nouveaux sous réseaux qui réclament la mobilisation de centaines de gènes. Il y a probablement d'autres accélérateurs de l'évolution à ajouter à la courte liste actuelle (l'activation distale transcriptionnelle, CTCF, etc.).

### **Un nouveau regard sur l'évolution et au-delà ?**

Un dernier mot, peut-être, sur les surprises. Les notions que nous avons discutées dans l'opuscule ont un impact qui va au-delà de la génétique et du développement. Ces cooptations génétiques, ces motifs de réseau, les interactions entre FTs, entre modules, entre cellules, entre territoires font apparaître *d'un coup* de nouvelles et étonnantes propriétés. En conséquence elles ne doivent rien à la sélection telle que nous l'entendons habituellement<sup>33</sup>. C'est de l'ordre fourni gratuitement, le « order for free »

de Kauffman<sup>34</sup>. En conséquence l'ordre des systèmes du développement vient *et* de l'évolution *et* du système lui-même (les propriétés émergentes, l'ordre spontané des systèmes). Et donc les systèmes évoluent en partie *spontanément* vers un ordre que simplement l'évolution retient. Le caractère historique et contingent de l'évolution en est diminué d'autant.

### **Pourra-t-on un jour prévoir l'évolution ?**

La réponse est certainement non. Et pour au moins deux raisons. D'une part, les paramètres pris en compte par l'évolution sont en nombre quasi infini. Il ne peut y avoir d'équations qui les contiennent tous. D'autre part, et cette raison est encore plus rédhibitoire, les propriétés émergentes non prévisibles ouvrent des voies que nous ne pouvons imaginer. Ce point ne peut pas être mieux résumé que par une citation de Stuart A. Kauffman, l'un des pionniers de l'étude des réseaux de régulation : « Not only do we not know what will happen, we often do not even know what can happen »<sup>35</sup> (« *Non seulement nous ne savons pas ce qui va arriver, mais souvent nous ne savons même pas ce qui peut arriver* »).

# ANNEXES

- 1 Le crible génétique des mutants létaux embryonnaires chez la drosophile
- 2 L'activation distale de l'expression des gènes
- 3 Les réseaux génétiques de régulation
- 4 Les communications entre cellules par les voies de signalisation
- 5 Une brève comparaison des génomes des choanoflagellés aux néphrozoaires
- 6 Les legs de notre ancêtre unicellulaire
- 7 L'origine de l'œil ou la naissance par chance d'une fonction.
- 8 La place du développement dans le monde du vivant

Les annexes sont des compléments à des points présentés dans le document principal. Elles sont destinées à ceux qui seraient désireux d'en apprendre un peu plus. Elles ont été conçues pour pouvoir être lues indépendamment les unes des autres. Elles peuvent aussi être lues comme un tout dont voici la logique.

On a voulu que le lecteur saisisse comment furent découverts les gènes du développement dont nous avons parlé. **C'est l'annexe 1 : « Le crible génétique des mutants létaux embryonnaires chez la drosophile et pourquoi il a réussi ».**

Tout est parti de mutants. Pour quelques cas, de mutants apparus spontanément dans les stocks de maintien des mouches (comme *Ey* et certains des gènes *Hox*) ou des souris (*Pax6*), et pour la plupart des autres cas de mutants issus de mutagenèses réalisées par les chercheurs.

Nous décrivons très exactement le crible génétique qui a abouti à isoler les gènes qui depuis l'œuf contrôlent la mise en place des segments de la larve de la drosophile. Ces résultats ont eu un impact extraordinaire sur toute la biologie du développement en montrant que le développement est basé sur des gènes organisés hiérarchiquement.

Cette histoire illustre la force de la génétique qui ne nécessite ni théorie ni même une quelconque hypothèse sur le processus analysé. C'est une approche qui va de l'expérience au modèle et non l'inverse.

Cette première annexe débouche sur l'idée comme quoi les gènes détectés par le crible se contrôlent hiérarchiquement. Plus tard, ces gènes furent clonés et l'on découvrit qu'ils codent des facteurs de transcription qui se contrôlent hiérarchiquement en se régulant les uns les autres par le mécanisme de l'activation distale de la transcription.

Nous donnons donc quelques détails sur ce mécanisme dans **l'annexe 2 : « L'activation distale de l'expression des gènes. Des cis-activateurs aux intégrateurs ».**

L'activation distale de l'expression des gènes est à la base des réseaux génétiques. Des arrangements des gènes entre eux émergent pour les cellules et l'embryon des propriétés essentielles au développement. C'est un des composants du développement, bien que nombre de ces arrangements soient aussi trouvés dans les réseaux des unicellulaires. Ceci est décrit dans **l'annexe 3 : « Les réseaux génétiques de régulation ».**

Mais comment un réseau génétique peut-il avoir plusieurs états d'activation, alors que les cellules qui se divisent donnent naissance à des cellules identiques ? Nous avons vu que des processus qui créent de l'hétérogénéité entre cellules sont mis en jeu

et que ceux-ci exploitent les communications intercellulaires. C'est peut-être le composant majeur du développement. On présente quelques informations sur la manière dont les voies de signalisation sont impliquées dans la génération d'états du réseau qui forment des organes différents à partir d'un même tissu. **C'est l'annexe 4 : « Les communications entre cellules par les voies de signalisation ».**

**Les annexes 5 et 6 « Une brève comparaison des génomes des choanoflagellés aux néphrozoaires » et « Les legs de notre ancêtre unicellulaire »,** détaillent un peu de quand sont apparus les ingrédients discutés précédemment et s'ils sont de véritables innovations.

Enfin après avoir tant discuté des relations logiques des gènes de régulation entre eux, nous discutons de l'origine des gènes princeps des organes qui ne sont pas des gènes de régulations mais leur cible. L'on verra que, au moins pour celles de la vue, les rhodopsines, ils remontent jusqu'aux procaryotes d'il y a plus d'un milliard d'années. Ils sont bien plus vieux que les gènes de régulation. **C'est l'annexe 7 : « L'origine de l'œil ou la naissance par chance d'une fonction ».**

Dans la dernière annexe : **« Sur la place du développement dans le monde du vivant »**, nous concluons en plaçant le développement dans le monde du vivant.

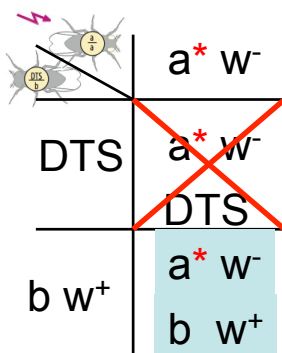
## ANNEXE 1

## LE CRIBLE GÉNÉTIQUE DES MUTANTS LÉTAUX EMBRYONNAIRES CHEZ LA DROSOPHILE ET POURQUOI IL A RÉUSSI.

Décrivons en détail le crible génétique qui a conduit à l'identification de mutants majeurs du développement de la drosophile sur lesquels est basée notre compréhension du positionnement des organes chez la drosophile. Ce crible (dû à C. Nüsslein-Volhard <sup>36</sup>) visait à identifier les gènes dont la mutation est létale pour l'embryon (il ne couvre donc pas les gènes de développement non létaux, beaucoup plus difficile à détecter). Le crible fut réalisé successivement pour les gènes portés par les quatre chromosomes de la drosophile. Choisissons l'exemple du chromosome II.

On utilise des drosophiles mâles dont le chromosome II, celui qui après mutagénèse sera analysé, porte un gène muté conférant aux mouches qui le portent un phénotype facilement identifiable à l'état homozygote. Cela sert à le repérer dans la descendance de ces mâles par un examen visuel simple et rapide. Le gène White qui change la couleur de l'œil de la drosophile de rouge à blanche remplit ces critères. Ces drosophiles mâles sont traitées par un agent mutagène. Elles vont donc fabriquer des spermatozoïdes dont le chromosome II porte des mutations induites et le gène white (chromosome mâle  $a^* w^-$ ).

### Croisement F0



	$a^* w^-$
DTS	<del><math>a^* w^-</math></del>
$b w^+$	$a^* w^-$ $b w^+$

**Figure 1 : Le tableau des gamètes du croisement F0.** Les gamètes de la drosophile mâle (la case du haut) portent sur le chromosome II le gène  $w^-$  et éventuellement un gène muté  $a^*$  ; les gamètes de la drosophile femelle (les cases à gauche) portent soit un chromosome II avec DTS (en haut à gauche) soit un chromosome II avec  $b$  et le gène  $w^+$  (en bas à gauche) ; les mouches issues de ce croisement (aux intersections des cases des gamètes) sont non viables (la case barrée, à cause du chromosome portant DTS,) ou viables (la case surlignée en

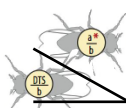
violet, elles portent un chromosomes II  $a^* w^-$  et un chromosomes II  $b w^+$ ).

Les mâles mutagénisés sont croisés avec des drosophiles femelles, elles aussi soigneusement choisies. Elles portent sur l'un de leur chromosome II une mutation, DTS, et sur l'autre une mutation, b. DTS est une mutation dominante létale dite sensible à la température car son effet létal n'est observé que quand les larves de la drosophile sont élevées à haute température (29°C). Les mouches portant ce chromosome, même à l'état hétérozygote, meurent quand elles sont élevées à 29°C. b est une mutation récessive létale. Toute mouche qui porte deux chromosomes marqués par b meurt. Grâce à ces deux marqueurs DTS et b, les drosophiles portant les chromosomes non mutagénisés seront automatiquement éliminées et seules, les drosophiles dont le chromosome II a été mutagénisé survivront. Voyons cela de plus près en suivant les croisements réalisés.

Croisement F0 (Figure 1). Les drosophiles mâles traitées par le mutagène sont croisées à 29° C avec des femelles portant les chromosomes II DTS et b. Dans la descendance, les mâles issus de ce croisement qui portaient un chromosome II DTS et un chromosome II mutagénisé ont donc été éliminés. Les mâles survivants portent tous un chromosome II b et un chromosome II mutagénisé portant éventuellement une mutation létale embryonnaire.

Croisement F1 (Figure 2). Ces mâles survivants sont croisés à 29° C avec des femelles portant encore les chromosomes II DTS et b. Seules survivent les drosophiles portant le chromosome II d'origine paternelle portant éventuellement la mutation létale embryonnaire et le chromosome d'origine maternelle portant le marqueur b.

### Croisement F1



	$a^* w^-$	$b w^+$
DTS	<del><math>a^* w^-</math></del>	<del><math>b w^+</math></del>
$b w^+$	$a^* w^-$ $b w^+$	<del><math>b w^+</math></del>


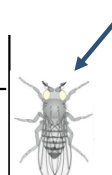

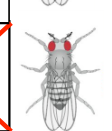
**Figure 2 : Le tableau des gamètes du croisement F1.** Les gamètes de la drosophile mâle (les 2 cases du haut) portent sur le chromosomes II le gène  $w^-$  et éventuellement un gène muté  $a^*$  (case du haut à gauche) ou les gènes b et  $w^+$  (case du haut à droite) ; les gamètes de la drosophile femelle (les cases à gauche) portent soit un chromosome II avec DTS (en haut à gauche) soit un chromosome II avec b et le gène  $w^+$  (en bas à gauche) ; les mouches issues de ce croisement (aux intersections



des cases des gamètes) sont non viables (à cause du chromosome portant DTS, ou à cause des deux gènes b, les cases barrées) ou viables (la case surlignée en violet) et portent un chromosomes II  $a^* w^-$  et un chromosomes II  $b w^+$ .




Croisement F2 (Figure 3). Ces drosophiles sont croisées entre elles. L'expérimentateur surveille la descendance de ces croisements. Il examine la couleur des yeux des drosophiles adultes. Il attend trois types de drosophiles, celles portant deux chromosomes II b qui meurent à cause de la mutation b, celles portant un chromosome mutagénisé d'origine paternelle et un chromosome II b qui doivent survivre et ont les yeux rouges, et celles portant deux chromosomes mutagénisés d'origine paternelle. Elles auront les yeux blancs (Figure 3, cas 1). Mais ces dernières ne sont pas observées si le chromosome d'origine paternelle porte une mutation létale embryonnaire (Figure 3, cas 2).

**Croisement F2**

	$a^* w^-$	$b w^+$	
$a^* w^-$	$a^* w^-$	$b w^+$	
	$a^* w^-$	$a^* w^-$	
$b w^+$	$a^* w^-$	<del><math>b w^+</math></del>	
	$b w^+$	<del><math>b w^+</math></del>	

Cas 1, pas de léthal embryonnaire.

**Croisement F2**

	$a^* w^-$	$b w^+$	
$a^* w^-$	<del><math>a^* w^-</math></del>	$b w^+$	
	<del><math>a^* w^-</math></del>	$a^* w^-$	
$b w^+$	$a^* w^-$	<del><math>b w^+</math></del>	
	$b w^+$	<del><math>b w^+</math></del>	

Cas 2, le chromosome 2 porte une mutation létale embryonnaire.

**Figure 3 : Le tableau des gamètes des croisements F2.** Les gamètes de la drosophile mâle (les 2 cases du haut) portent sur le chromosomes II le gène  $w^-$  et éventuellement un gène muté  $a^*$  (case du haut à gauche) ou les gènes b et  $w^+$  (case du haut à droite) ; les gamètes de la drosophile femelle (les cases à gauche) portent aussi soit un chromosome II avec le gène  $w^-$  et éventuellement un gène muté  $a^*$  (en haut à gauche) soit un chromosome II avec b et le gène  $w^+$  (en bas à gauche) ; les mouches issues de ce croisement (aux intersections des cases des gamètes) sont, **cas 1**, non viables (à cause des deux gènes b, la case barrée) ou viables (les 3 autres cases ; elles ont les yeux rouges ou blancs) ; **cas 2**, non viables (à cause des deux gènes b, la case barrée ou à cause gène muté  $a^*$  létal) ou viables (les 2 autres cases ; elles ont les yeux rouges).

Aussi, l'expérimentateur va trier les croisements en fonction de ce critère. Il ne gardera pas les croisements dans lesquels il observe des drosophiles aux yeux blancs (Figure 3, cas 1). Par contre, il gardera les croisements dans lesquels il n'observe que des drosophiles aux yeux rouges (Figure 3, cas 2). Il sait que les chromosomes II mutagénisés portent une mutation létale embryonnaire puisque les drosophiles de ces croisements qui portaient le chromosome II mutagénisé en deux exemplaires (à l'état homozygote) sont mortes.

Il va donc ensuite recroiser entre elles les mouches aux yeux rouges de ces croisements. Il examinera très attentivement les larves dont certaines vont être anormales.

Ce crible génétique est d'une grande simplicité d'exécution. Les croisements F1 et F2 sont réalisés avec les mouches survivantes sans tri ni examen particuliers, donc automatiquement, en quelque sorte. La détection des mutants dans la descendance des croisements F2 ne nécessite que de déterminer l'absence de mouches aux yeux blancs dans les descendances. Une opération extrêmement simple !

Ce crible génétique permet l'approche d'un des problèmes les plus complexes de la biologie en évitant tout a priori théorique sur la nature et la fonction des gènes de développement ! C'est l'extraordinaire force de la génétique. Mais encore faut-il qu'il y ait des gènes du développement dont la mutation est létale. C'est là tout le pari.

Si, à partir des drosophiles mâles mutagénisés, on réalise 5000 croisements F2 de ce type, on récupère de 20 à 50 croisements avec uniquement des drosophiles aux yeux rouges qui correspondent à tous les mutants embryonnaires létaux possibles.

C'est à partir de cribles génétiques de ce type que les généticiens ont isolé parmi les 20.000 gènes que cette espèce porte les quelques dizaines de gènes du développement de la drosophile.

Il y a trois raisons qui expliquent le succès de la drosophile dans ce crible.



Tout d'abord, comme nous venons de le voir, la drosophile se prête particulièrement bien aux expériences de mutagenèse. On produit très facilement des mutants. Ni la grenouille, ni le poulet, ni même la souris ne se prêtent à ce type d'approche.

D'autre part, le développement de la drosophile (comme celui des autres insectes) débute avec un embryon syncytial, c'est-à-dire dans lequel les noyaux se divisent sans formation des membranes cellulaires. Ce caractère syncytial de l'embryon perdure durant les 13 premières divisions. Ce n'est qu'à la quatorzième division que les membranes cellulaires se forment. En conséquence, jusqu'à ce stade, les molécules et en particulier les protéines et les ARN diffusent librement entre noyaux. L'échange d'information en est considérablement simplifié. C'est un gros avantage dans les études génétiques. Dans tous les autres types de développement, les membranes cellulaires font obstacle à la libre diffusion des molécules.

Le troisième avantage de la drosophile est la simplicité de son génome. Alors que le génome des vertébrés s'est, au cours de l'évolution, dupliqué deux fois, ce qui a multiplié par quatre beaucoup des gènes dont les gènes de développement (voir par exemple le cas des gènes Hox, Figure 36), celui de la drosophile, comme d'ailleurs celui de nombreux autres invertébrés, n'a pas subi ces duplications globales. Avoir quatre gènes presque identiques qui, souvent, gardent la même fonction, complique très sérieusement l'analyse génétique, car, muter un gène n'a, alors, qu'un effet mineur, les gènes homologues non mutés continuant d'assurer la fonction ; c'est ce qu'on appelle la redondance génétique. En l'absence de redondance génétique, la mutation du seul gène conduit à un phénotype franc. Si la drosophile avait subi des duplications globales de son génome, il est probable que le crible de mutants létaux n'aurait rien donné.

Au moment où les expérimentateurs se sont intéressés à la drosophile de ces trois raisons, seule la première leur était connue.

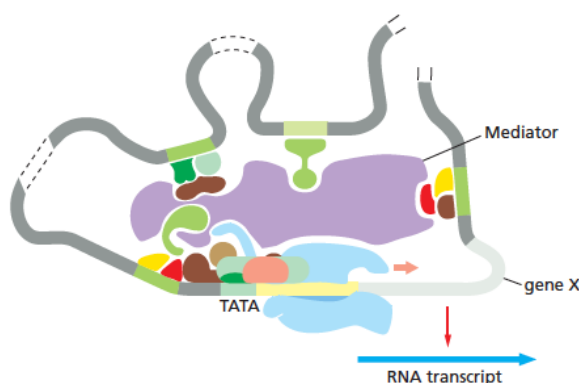
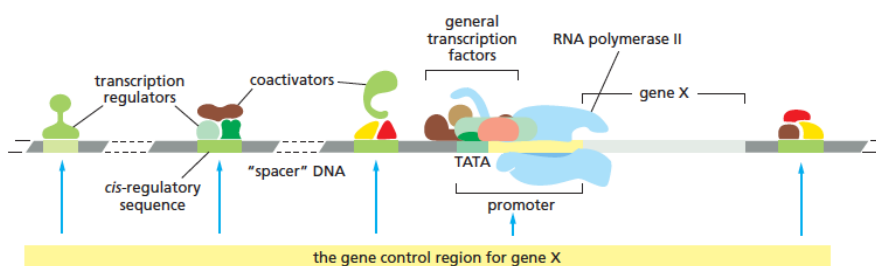
## ANNEXE 2

## L'ACTIVATION DISTALE DE L'EXPRESSION DES GÈNES. DES CIS-ACTIVATEURS AUX « INTÉGRATEURS ».

Les états cellulaires chez les protistes et les types cellulaires chez les animaux sont dus à des activations et répressions, au niveau transcriptionnel, de certains gènes (Figures 10 et 12 de l'opuscule). Une cellule musculaire ou nerveuse ou une cellule de l'ectoderme ou du mésoderme au cours du développement (Figure 8 de l'opuscule) produisent des ARNs différents.

C'est l'ARN polymérase II (ARN Pol II), une grosse protéine, qui produit une copie ARN de l'ADN (Figure 10). Elle se positionne sur l'ADN au niveau du promoteur avec plus de 80 autres protéines appelées facteurs de transcription généraux. Ensemble ils constituent la machinerie transcriptionnelle (Figure 1). Ces protéines sont nécessaires pour positionner très précisément l'ARN polymérase sur l'ADN (au niveau de la fameuse boîte TATA) et séparer les 2 brins de l'ADN. Cela donne à l'ARN polymérase accès au brin ADN qu'elle va copier. Tout cela constitue ce qu'on appelle la phase de l'initiation de la transcription. Ensuite des

phosphorylations de la molécule (de sa petite « queue, le CTD) préludent à la phase dite d'élongation de l'ARN, la lecture accélérée du brin de l'ADN).



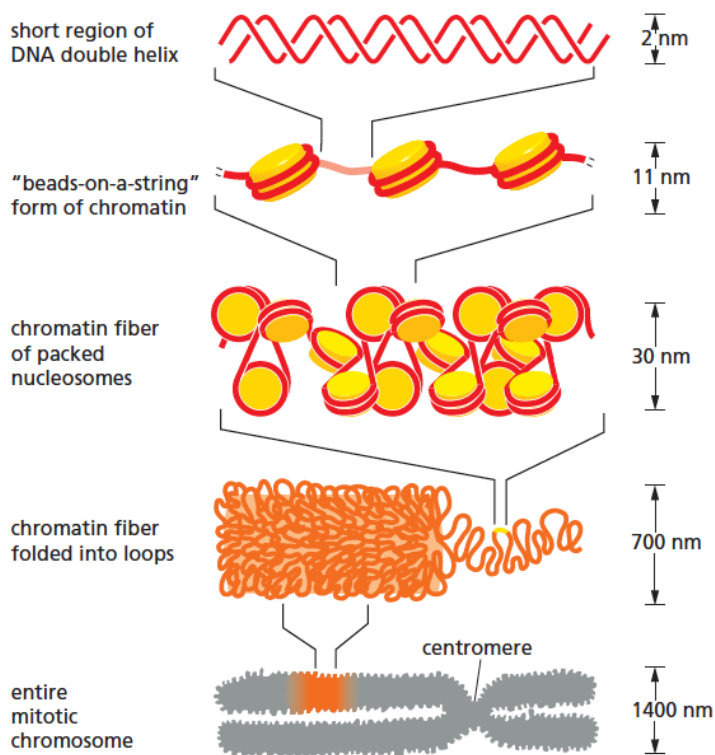
**Figure 1 : La région de contrôle d'un gène.** a) le gène X, sa boîte TATA et 4 cis-éléments sur lesquels

se fixent les FTs qui agrègent des coactivateurs. L'ARN polymérase II est représentée en bleue. b) des boucles de l'ADN rapprochent les cis éléments du promoteur. Le médiateur sert d'intermédiaire entre les cis-éléments et la machinerie transcriptionnelle. La transcription du gène dépend de l'intégration de multiples informations médiées par les FTs et la structure des cis-éléments.

Pour les gènes dont l'expression est différentielle l'assemblage de la machinerie transcriptionnelles sur le promoteur dépend de deux éléments, les FTs (les facteurs de transcription de l'opuscule encore appelés activateurs ou régulateurs de la transcription) et des éléments de l'ADN, les cis-éléments (ou enhancers ou « cis regulatory modules, CRM »), sur lesquels les FTs se fixent par leur domaine de liaison à l'ADN (DNA binding domain, DBD) (Figures 12 et 13 de l'opuscule). Leur site de fixation (ou leur « response element » ou leur cis-regulatory sequences) dans les cis-éléments correspond à des séquences spécifiques de 5 à 10 bases (Figures 12 et 13 de l'opuscule). Les cis-éléments ont une taille variable (de 50 à 1500bp) et peuvent être éloignées de plusieurs centaines à plusieurs milliers voir millions de paires de base du promoteur.

Les FTs attirent par un deuxième domaine, le domaine de transactivation (transactivation domain, TAD), d'autres protéines, les coactivateurs. Celles-ci forment des complexes qui remodelent la chromatine et modifient les histones par leurs activités catalytiques (acétylations/déacétylations ; méthylations/déméthylations ; kinases).

En effet l'ADN n'est jamais nu mais associés à des protéines, les histones. Ensemble ADN et histones forment les nucléosomes (Figure 2b). Les nucléosomes s'associent entre eux pour former des fibres qui caractérisent ce que l'on appelle la chromatine. La fibre sous laquelle se trouve généralement l'ADN est la fibre de chromatine de 30-nm (Figure 2c). L'ADN y est peu accessible. L'un des rôles des FTs est de le rendre accessible. Et tout d'abord au niveau des cis-éléments (enhancers) ou la fixation d'un FT va entraîner la modification de la chromatine ce qui va faciliter la fixation d'autres FTs. Souvent les cis-éléments associés à leurs FTs sont des régions dépourvues de nucléosomes.



**Figure 2 : L'ADN et la chromatine.** a) les deux brins de l'ADN forment une double hélice ; b) la double hélice est enroulée autour de sorte de perles que constituent les histones ; ce sont les nucléosomes c) ces nucléosomes sont eux-mêmes organisés en une fibre de 30 nm. d) ces fibres sont elles-mêmes organisées en boucles accrochées sur ???. e) un chromosome tel qu'on le voit à la mitose. Le facteur de compaction entre a) et d) est de plus de 10000 fois. (nm =  $10^{-9}$ m, micron  $10^{-6}$ m)

L'un des coactivateurs des FTs est un complexe de protéines

appelé médiateur (Figure 1b), car il sert de pont entre les FTs fixés à l'ADN et l'ARN polymérase et d'autres facteurs de transcription généraux dont là encore des enzymes de modification de la chromatine comme les histones acétyl-transférases, les histones kinases et les facteurs de remodelage des nucléosomes qui facilitent leur assemblage sur le promoteur et établissent un environnement chromatinien propice à la transcription.

Ensuite, pour commencer à transcrire l'ARN polymérase doit se libérer de cet énorme complexe. Souvent cela requiert la protéolyse du FT.

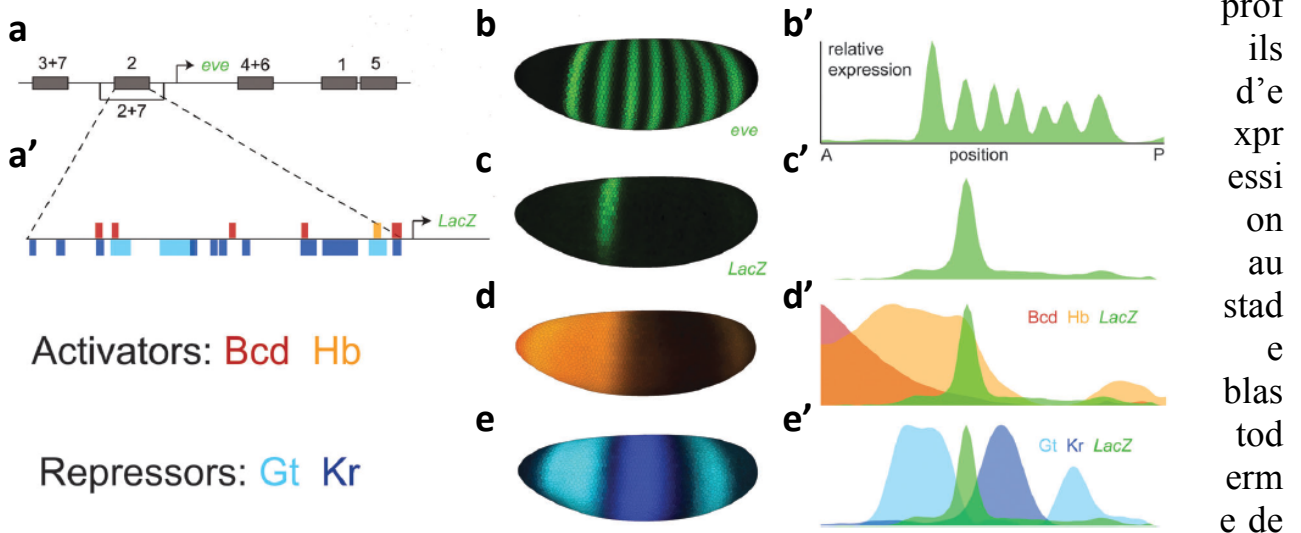
Les FTs seuls, non fixés sur les enhancers, n'ont pas d'activité. Les cis-éléments seuls n'ont pas non plus d'activité. Et de plus nous avons vu que l'action des FTs sur le complexe transcriptionnel nécessite une boucle de l'ADN qui rapproche les cis-éléments des promoteurs. Le processus qui fait cette boucle est connu (Figure 27 de l'opuscule) et son émergence pourrait avoir été d'une grande importance dans l'évolution (Figure 28 de l'opuscule).

La répression par les FTs est tout aussi, voir plus fréquente, que l'activation (Figures 29 et 37 de l'opuscule, par exemple). Les FTs

répresseurs peuvent agir de plusieurs manières : par compétition avec les activateurs pour leur fixation à l'ADN ; par masquage du site d'activation du FT activateur ; en bloquant l'assemblage de la machinerie transcriptionnelle sur le promoteur ; en recrutant des corepresseurs qui vont rétablir un état chromatinien incompatible avec la transcription etc.

Voyons sur un exemple, *Even skipped (eve)*, a quoi ressemble la région de contrôle d'un gène, <sup>37, 38, 39</sup>.

**Figure 3 : Le gène *even-skipped* de la drosophile** a) La région de contrôle du gène *even-skipped* de 20 000 paires de base englobe le gène. Les boîtes grises délimitent les cis-éléments qui contrôlent l'expression dans les bandes indiquées par les chiffres; a') l'élément qui contrôle l'expression de la bande 2 comprend des sites de liaison pour les FTs Bicoid, Hunchback, Giant, Kruppel (les boîtes de différentes couleurs) ; les



b) *even-skipped* (7 bandes transverses), c) *LacZ* dans une drosophile transgénique pour le transgène S2*LacZ* (a'); Bicoid et Hunchback, d); Giant et Kruppel e); b'-e') expression en en fonction de la position antéro-postérieure dans le blastoderme pour b-e <sup>38</sup>.

*Eve* est un des gènes, dit pair-rule, de la drosophile qui intervient dans la formation des segments et s'exprime en 7 bandes transverse dans la larve (Figure 2b et figures 3 et 37 de l'opuscule). Nous avons vu dans l'annexe 3 comment par le réseau de régulation on comprend son activation (annexe 3, le réseau de la mise en place des segments chez la drosophile). Voyons ici à quoi ressemblent au niveau moléculaire les interactions entre gènes qui constitue le réseau. *Eve* est l'un des gènes dont la région de contrôle est la mieux connue.

La région de contrôle d'*even skipped* s'étend sur 20 000 paires de base (20kb) qui englobent le gène (Figure 2a). On y a reconnu 5 cis-éléments distincts. L'un d'entre eux, le cis-élément 2, contrôle l'expression d'*even skipped* dans la bande 2. Les quatre autres contrôlent respectivement l'expression dans les bandes 3 et 7, 4 et 6, 1 et 5. Le cis-élément 2 contient 5 sites de fixation pour bicoid, 1 pour hunchback, 8 pour krupel et 3 pour giant. Ces quatre FTs sont des FTs du réseau de régulation qui interagissent avec *even skipped* (annexe 3, figure 7b ou *eve* serait « S »). Bicoid (Bcd) et hunchback (Hb) sont des activateurs, giant (Gt) and Kruppel (Kr) sont des répresseurs. Les profils d'expression de ces 4 gènes (Figure 2 b, d-e et b', d'-e') montrent que, logiquement, premièrement *even skipped* n'est exprimé (Figure 2, le pic vert dans d' et e') que dans la région où les deux répresseurs sont à un niveau très faible et les deux activateurs à un niveau très fort (Hb) et faible mais détectable (bicoid) et deuxièmement que les expressions des deux répresseurs dans les régions adjacentes à la bande S2 rendent compte de ses frontières. Ainsi est expliqué, qualitativement, l'expression spatiale d'*even skipped*.

Ce schéma répond qualitativement aux principales questions que l'on se pose. Cependant, il est incomplet. Ainsi, les tentatives pour synthétiser un cis-élément qui reproduirait correctement l'expression d'*even skipped* dans la bande 2 ont jusqu'à présent, échouées<sup>38</sup>. Et ce bien que ces tentatives tiennent compte de plusieurs découvertes récentes : 1) le fait que dans le cis-élément 2 l'action de Hb est contrecarrée par un autre FT, caudal<sup>39</sup>, 2) qu'un troisième FT répresseur, Sloppy-paired intervienne (REF) et 3) que le FT Zelda, un FT exprimé dans tout l'embryon, soit nécessaire à l'expression de *eve* (il y a d'ailleurs plusieurs sites de fixation pour Zelda dans le cis-élément 2 ; Zelda est requis pour rendre possible la fixation des autres FTs sur les cis-éléments). Clairement encore d'autres FTs non identifiés et/ou des conditions particulières inconnues interviennent. Tant que la biologie de synthèse n'aura pas reproduit le profil d'expression d'*eve*, on sait que des éléments nous échappent.

Un deuxième niveau où notre ignorance reste grande est le mode précis d'action des cis-éléments. On a vu que souvent les coactivateurs des FTs ont des activités enzymatiques de modification de la chromatine comme les



histones acétyl-transférases, les histones kinases et les facteurs de remodelage des nucléosomes. On ne peut rien dire de précis pour ce qui concerne aucun des cis-éléments d'even-skipped.

Il apparait donc que les régions de contrôle des gènes sont complexes, plus complexe que ne le laisse apparaître les réseaux de régulation, et que notre connaissance de leur fonctionnement n'est que qualitative et même souvent incomplète. Ce sont des centaines de FTs qui interviennent. On dénombre déjà plus de 20 sites de fixation pour au moins 6 FTs différents dans le cis-élément 2 d'even et il y a 5 cis-éléments de ce type dans les 20kb qui entoure le gène. Cette densité de FTs est sans doute nécessaire pour la fixation coopérative des FTs sur le cis-élément, pour agir sur la chromatine qui exerce naturellement une action répressive qui doit être levée et finalement pour modifier le complexe transcriptionnel.

Par contre, au final, la réponse est simple : le gène ne peut être que dans deux états : actif ou inactif. Ce sont les 0 et les 1 de la logique booléenne des modélisations des réseaux transcriptionnels (voir l'annexe 3).

L'intervention de plusieurs FTs différents, activateurs et inhibiteurs fait que la réponse doit intégrer de nombreux signaux. Le cis-élément peut donc être vu comme un intégrateur d'informations, les informations étant la présence des FTs. Ceux-ci constituent pour la cellule une sorte d'image interne de son histoire passée et de son voisinage mélangés. De cette intégration dépend la nature de l'interaction avec le complexe transcriptionnel. Dans ce cas précis l'expression ou non du gène even skipped dont le futur de la cellule dépend.





## ANNEXE 3

### LES RÉSEAUX GÉNÉTIQUES DE RÉGULATION

De certains des arrangements des FTs entre eux émergent des propriétés cellulaires que les éléments pris séparément ne peuvent donner. Le réseau de l'initiation du développement de l'œil chez la drosophile présente une belle simplicité structurale. Il n'a qu'un seul état stable. L'entrée du réseau est protégée des « erreurs » par un détecteur persistant. Les autres motifs de réseau fonctionnent de façon modulaire et renforcent mutuellement leur expression. Le réseau de la mise en place des segments chez la drosophile présente aussi une simplicité structurale. Le FT à l'entrée du réseau agit sur trois cibles selon sa concentration créant ainsi 3 territoires. Des étapes combinatoires entre FTs créent ensuite de nouveaux territoires. Ces interactions directes entre les FTs sont particulières aux insectes. Le réseau de l'initiation de la formation des organes dérivés des segments mésodermiques chez la drosophile implique trois éléments externes aux cellules (des morphogènes) et des éléments internes dont les trois FTs des voies de signalisation et les gènes maîtres des organes et tissus. Les éléments externes créent des territoires et leur géométrie. Les éléments internes du réseau gèrent la dynamique des opérations cellulaires des territoires. Ce type de situation est général aux développements. L'évolution joue sur les réseaux. Certaines mutations rajoutent ou retirent des liens aux éléments du réseau et peuvent ainsi avoir un effet sur le déroulement des opérations du développement, changer fondamentalement les types cellulaires, les organes, et leur disposition. Leur apparition, qui n'est ni sélectionnée ni prédictible, apporte inopinément de l'ordre au système. Stuart Kauffman parle « d'ordre gratuit » apporté au système. Le génome code les produits des gènes et leurs interactions. Le réseau associé à un génome comporte des milliers de nœuds et des dizaines de milliers de flèches. Ni le réseau de la drosophile, ni celui d'aucun autre l'animal ne sont actuellement déchiffrés.

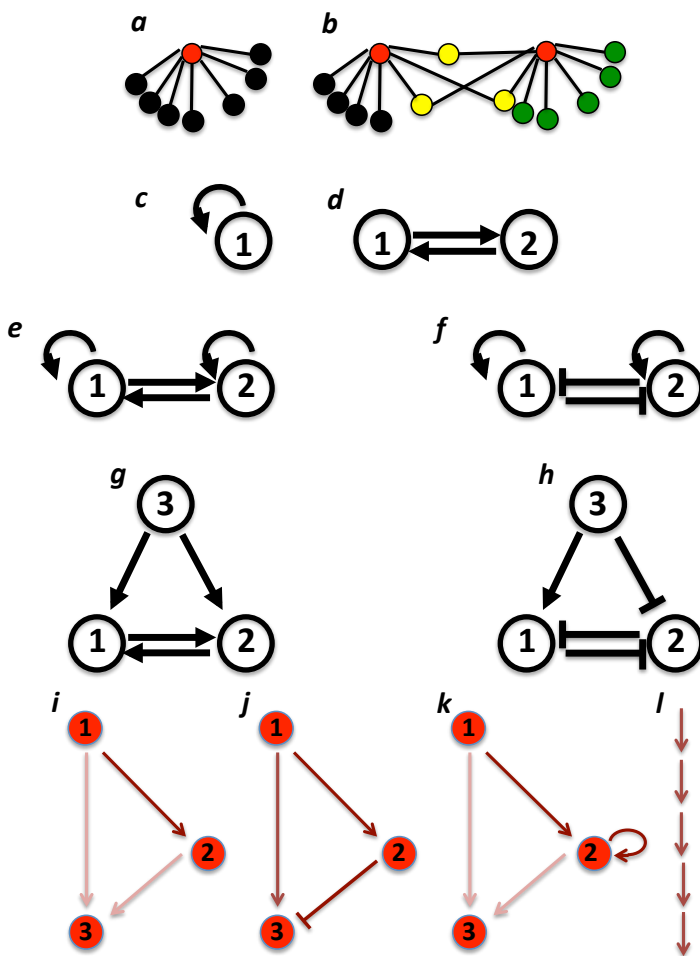
Nous analysons ici le fonctionnement des réseaux de régulations et discutons en quoi cela a changé notre manière de penser l'action des gènes, le développement et l'évolution.

Nous savons que les facteurs de transcription (FTs) non seulement régulent des gènes qui codent des protéines caractéristiques de type cellulaires mais qu'ils se régulent entre eux. Ces régulations se traduisent par des arrangements que l'on représente par des nœuds, les régulateurs, et des flèches, le type de régulation et sa nature, activation ou inhibition (voir Figures 15, 16, 22 ou 29 de l'opuscule). Les régulateurs peuvent ne pas être des facteurs de transcription (des ARN mais aussi des morphogènes peuvent être des régulateurs). En général on considère qu'aux nœuds correspondent deux états, inactif ou actif, (0,1), et qu'ils sont reliés ensemble par des relations logiques (du type si A actif alors B actif). On parle de variables booléennes et si l'on considère la dynamique du réseau l'état cellulaire au temps  $t$  dépend de l'état à  $t-1$  (on parle d'un réseau d'automates markoviens).

### **Les motifs de réseau <sup>14</sup>.**

A certains des arrangements des FTs entre eux correspondent des propriétés qu'aucun des éléments pris séparément n'a. Ces propriétés émergent de l'architecture de l'arrangement. Parmi ces arrangements, nous avons vu (Figure 1) : le single input module (SIM architecture) qui coordonne l'expression de toute une famille de gènes (Figure 1a) ; le multiinput motif (MIM architecture) encore appelé dense overlapping regulon (DOR architecture) ou un ensemble de régulateurs contrôle plusieurs combinaisons d'un ensemble d'autres gènes (Figure 1b) ; l'autorégulation positive (PAR architecture, mémoire) (Figure 1c) ; l'autorégulation négative (NAR architecture,) ; l'activation réciproque qui donne des motifs mémoire à deux régulateurs associés à une opération logique ET (Figure 1d et e) ; l'inhibition réciproque qui donne des motifs mémoire associés à une opération logique OU (choix binaire) (Figure 1f) ; ces deux mêmes motifs mais auxquels s'ajoute un opérateur qui agit sur 1 et 2 et ainsi définit le fonctionnement du motif qui une fois enclenché persiste même en l'absence de l'opérateur (mémoire d'un événement passé) (Figure 1g et h) ; les feedforward loop (FFL architecture) ou un gène régule un autre gène et ensemble, ces deux gènes en régulent un troisième (Figure

1i-k). Les types de FFL sont au nombre de huit et chacun a des propriétés différentes. Chacun peut de plus être associé à une logique OU (1 OU 2 nécessaire pour 3) ou ET (1 ET 2 nécessaire pour 3). Le FFL cohérent type 1 (Figure 1i) filtre les pulses de signaux : un signal sur 1 qui ne persiste pas ne se traduit pas l'expression de 3 car 2 qui est nécessaire n'a pas le temps de s'accumuler ; seuls les signaux persistant se traduisent par l'expression de 3. On parle de détecteur persistant. Ce motif est un motif majeur. Le FFL incohérent type 1 (Figure 1j) fonctionne en générant un pulse de 3 car 1 active 3 et 2 et après un délai 2 inhibe 3. Enfin les cascades d'activation ou d'inhibitions qui génèrent une échelle de temps en relation avec le délai d'activation ou d'inhibition des FTs (Figure 1l).



1 active 3 et 2 et après un délai 2 inhibe 3. Enfin les cascades d'activation ou d'inhibitions qui génèrent une échelle de temps en relation avec le délai d'activation ou d'inhibition des FTs (Figure 1l).

Figure 1 : Les motifs de réseau majeurs du développement. Voir le texte.

Tous ces arrangements sont appelés motifs de réseau car les relations entre tous les FTs et plus généralement entre tous les régulateurs du génome constituent un réseau. Ce réseau inclut tous les gènes du génome.

Une bactérie comme E. coli a un réseau à 4000 nœuds (ses gènes) 300 facteurs de transcription et 10000 flèches<sup>40</sup>. Le génome de l'homme code environ 30000 gènes dont 2500 facteurs de transcription et donc encore plus de régulateurs. On y a reconnu un certain nombre de motifs de réseau, comme nous verrons plus loin mais seule une très faible fraction de ce réseau nous est connu. De même que tout dans un organisme est organe, tout dans un réseau est motif de réseau. Et de même qu'un organisme est beaucoup plus que l'ensemble de ses organes, le réseau est plus que l'ensemble de ses motifs. C'est un des

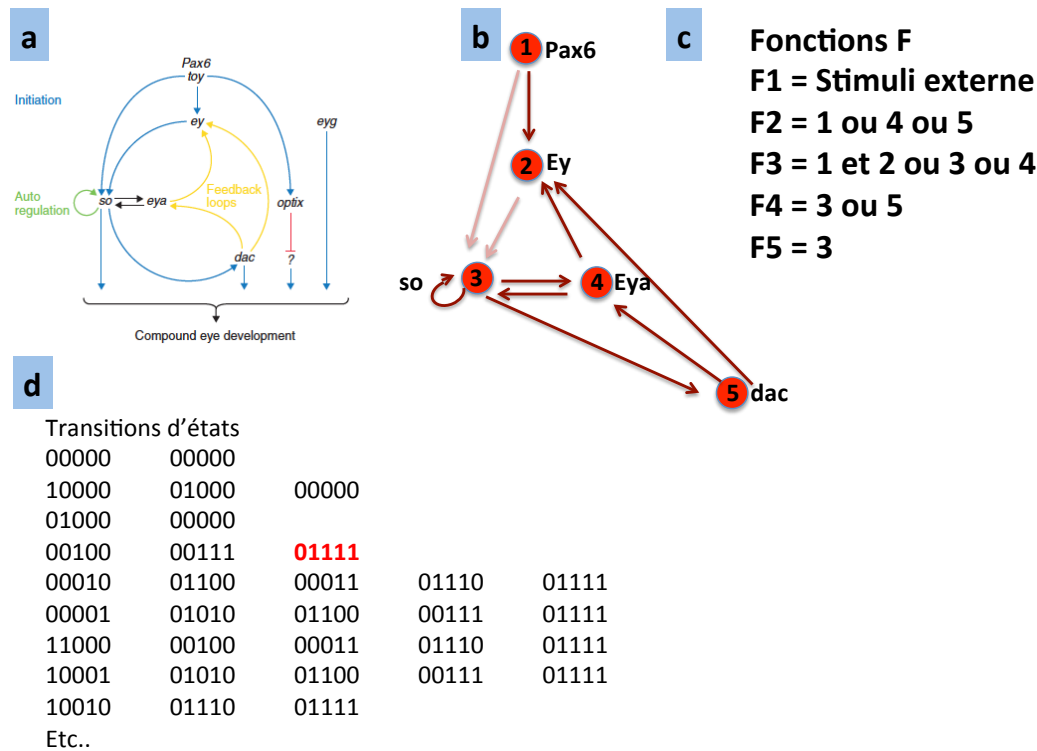
challenges actuels que de préciser cela et de voir jusqu'où mène cette constatation.

Une dernière précision sur les motifs de réseau. Les mutations dans un gène créent parfois un nouveau gène avec des propriétés nouvelles. Celui-ci peut alors être conservé quasiment inchangé au travers de l'évolution et être utilisé dans différents contextes. Par contre les motifs de réseau qui interviennent encore et encore au cours du développement sont des nouveautés indépendantes. Ils résultent d'évènements distincts et concernent des groupes de gènes différents. Ils doivent être « inventés » et « réinventés » encore et encore.

Voyons comment fonctionnent quelques ensembles de motifs de réseau et ce qu'ils nous expliquent du développement. Trois exemples seront traités : l'initiation du développement de l'œil par Pax6/toy chez la drosophile, la mise en place des fameux segments chez la larve de la drosophile et l'initiation de la formation des quatre organes dérivés des segments mésodermiques, le cœur, les muscles viscéraux, les muscles somatiques et le « fat body » là encore chez la drosophile. Le nombre de régulateurs impliqués dans le premier cas est de cinq, dans le deuxième cas de 7 et dans le dernier cas de 17. C'est un début mais le nombre de régulateurs actifs dans une cellule de l'embryon est énormément plus grand (ce nombre est connu mais leur relation entre eux ne l'est pas).

### **Le réseau de l'initiation du développement de l'œil chez la drosophile<sup>15</sup>.**

La fraction du réseau que nous considérons a cinq nœuds (ou régulateurs), (Figure 2a et b, 1 à 5) et dix interactions (Figure 2b, les flèches). A chaque nœud est associée une fonction F qui indique les conditions de son activité (0 inactif, 1 actif, Figure 2c). Pour 3 F3 indique que 3 est actif quand 1 et 2 sont actifs ou quand 3 ou 4 sont actifs. On identifie dans la Figure 2b cinq motifs de réseau : une double boucle de rétroaction positive (entre 3 et 4), une autorégulation positive (pour 3), un FFL cohérent type 1 (pour 3, avec F3 = 1 et 2) un autre FFL cohérent type 1 pour 2 (avec F2 = 4 ou 5) et un motif en boucle pour 3, 4, 5.



**Figure 2 : Le réseau de l'initiation du développement de l'œil chez la drosophile (voir le texte).**  
 a) le réseau établi par la génétique<sup>15</sup> ;  
 b) schéma pour rendre apparent les motifs de réseau ; les flèches en marron clair

indiquent une interaction conditionnelle c) les fonctions (conditions pour être dans l'état 1) associées à chaque nœuds ; d) quelques-unes des transitions d'états : état initial colonne de gauche, états instables les autres colonnes ; 00000 et 01111 sont les états stables.

Le nombre d'états possibles de ce système est de  $2^5$  soit 32. Quelques-uns sont indiqués en d, première colonne. 1 actif est noté 10000, 2 actif 02000 etc. Premier point à noter la plupart de ces états ne sont pas stables, ils évoluent vers un autre état, d deuxième colonne. Ainsi l'état 5 actif, 00001, se change en 01010 car 5 est un activateur de 2 et 4. Mais cet état n'est pas non plus stable car 4 va activer 2 et 3 ce qui donne l'état 01100 et ainsi de suite jusqu'à l'état 01111 qui lui n'évolue plus. Quelques-unes des autres transitions sont indiquées en d (suivre les lignes). L'analyse des 32 états possibles montre qu'il n'y a pas d'autre état stable en dehors de 01111 et 00000.

L'état stable 01111 correspond à l'expression de quatre des cinq régulateurs. Il exclut Pax6/toy. L'expression de chacun des quatre éléments est très stable. Chaque élément est maintenu actif par 2 autres éléments : 2 est maintenu actif par 4 et 5 ; 3 par une boucle d'autorégulation et 4 ; 4 par 5 et 3 ; 5 par 3 qui s'autorégule positivement. Les motifs de réseau identifiés agissent dans le même sens. Ils se renforcent. Il y a tout à parier

que l'activité de ces 4 éléments est requise à long terme dans le territoire de l'œil. La biologie dit que c'est effectivement le cas.

Le FFL cohérent type 1 pour 3 est positionné de façon à ce que des activités sporadiques de Pax6/toy et ou de Ey n'activent pas le système. Ce motif lui sert de filtre. On peut donc soupçonner que l'activation du circuit se fait par 1. C'est en effet le cas.

Ce petit réseau présente une belle simplicité structurale. Un des motifs à l'entrée du réseau protège des « erreurs ». Les autres motifs de réseau fonctionnent de façon modulaire et renforcent mutuellement leur expression. On ne détecte qu'un seul état stable pour ce réseau, une seule solution.

En isolant ainsi une micro fraction du réseau on pose qu'un grand nombre d'éléments, dont on ne tient pas compte mais qui sont nécessaires à l'expression de ces 5 FTs, sont dans l'état d'activité qui convient. Ne sont incluses non plus ni les étapes en amont de Pax6/toy, son expression dans un territoire de l'embryon, ni les cibles en aval des cinq FTs qui contrôlent les opérations cellulaires du développement de l'œil. Ceci ne retire rien à la pertinence de l'analyse.

### **Le réseau de la mise en place des segments chez la drosophile.**

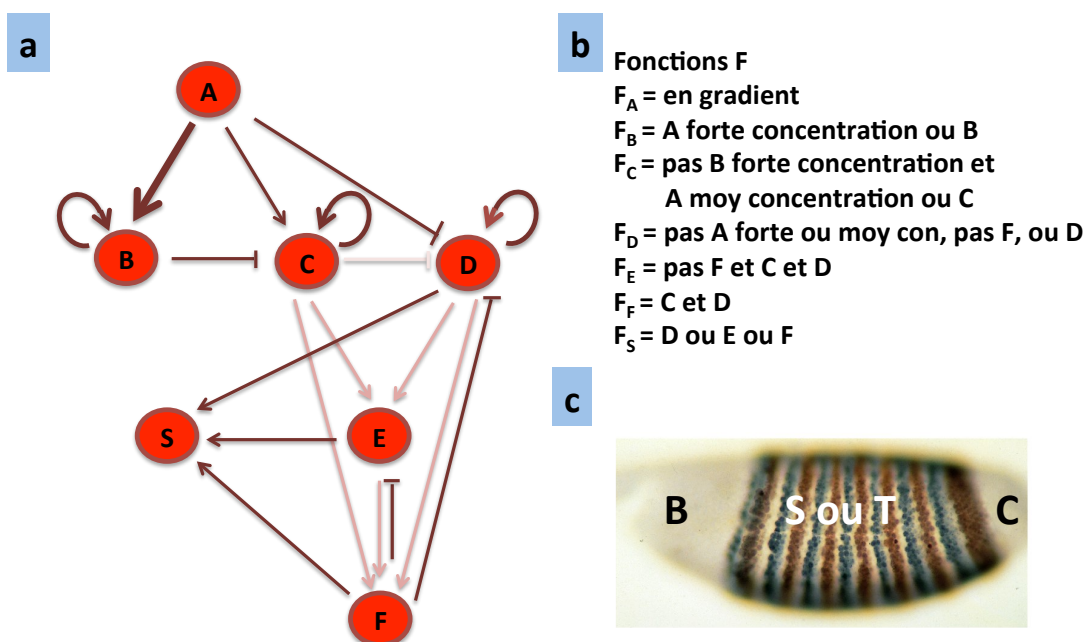
Avec le deuxième exemple nous verrons comment des interactions entre molécules peuvent créer plusieurs territoires. Ce réseau concerne des territoires qui ne sont pas constitués de cellules mais des noyaux du syncytium qui caractérise les premières heures du développement de la drosophile (et plus généralement des insectes). On parle d'embryons syncytiaux. L'absence de membranes cellulaires entre les noyaux permet les interactions directes entre les FTs et les activateurs distaux des gènes. De plus et cela est sans doute le plus remarquable cela permet aussi à certains des FTs d'agir en fonction de leurs concentrations sur leurs gènes cibles. Vont ainsi émerger des territoires dont la forme, celle de bandes transverses (figure 3, c les 7 bandes bleue et les 7 bandes marron), est due aux paramètres de la diffusion des molécules.



Cet exemple correspond à une situation plus complexe que la précédente. Il y a 2 états initiaux (éléments A et D exprimés) et 3 états finaux (éléments B, C, S exprimés). A ces trois états finaux correspondent dix territoires dont 2 répétés quatre fois (S et non S). Ce système groupe les interactions entre les FTs qui conduisent en 3 heures aux segments de la larve de la drosophile (revoir le film, Figure 3 de l'opuscule)<sup>41</sup>. La simplification (basée sur<sup>42</sup>) conserve l'essentiel du système.

La fraction du réseau que nous considérons a sept nœuds (ou régulateurs, figure 3, a, A à S) et dix-sept interactions (figure 3, a, les flèches). A chaque nœud est associée une fonction F qui indique les conditions de son activité (figure 3, b, fonctions F). Par rapport au cas précédent on voit apparaître dans les Fx *des conditions de concentrations*. Par exemple pour C, F<sub>C</sub> indique que C est actif quand on n'a pas B en forte concentration et que A est en concentration moyenne. Des conditions de concentrations concernent B, C, et D. Cette condition sur les concentrations crée des limites entre territoires dont la forme dépend de paramètres physiques (de gradients de concentration de molécules). Cela fait entrer la dimension spatiale dans le développement. Voyons cela.

On identifie 9 motifs de réseau : trois autorégulations positives (pour B, C, D), 6 motifs à trois nœuds (cohérents et incohérents), dont un FFL cohérent type 1 (pour S, avec D et F) (figure 3, a).

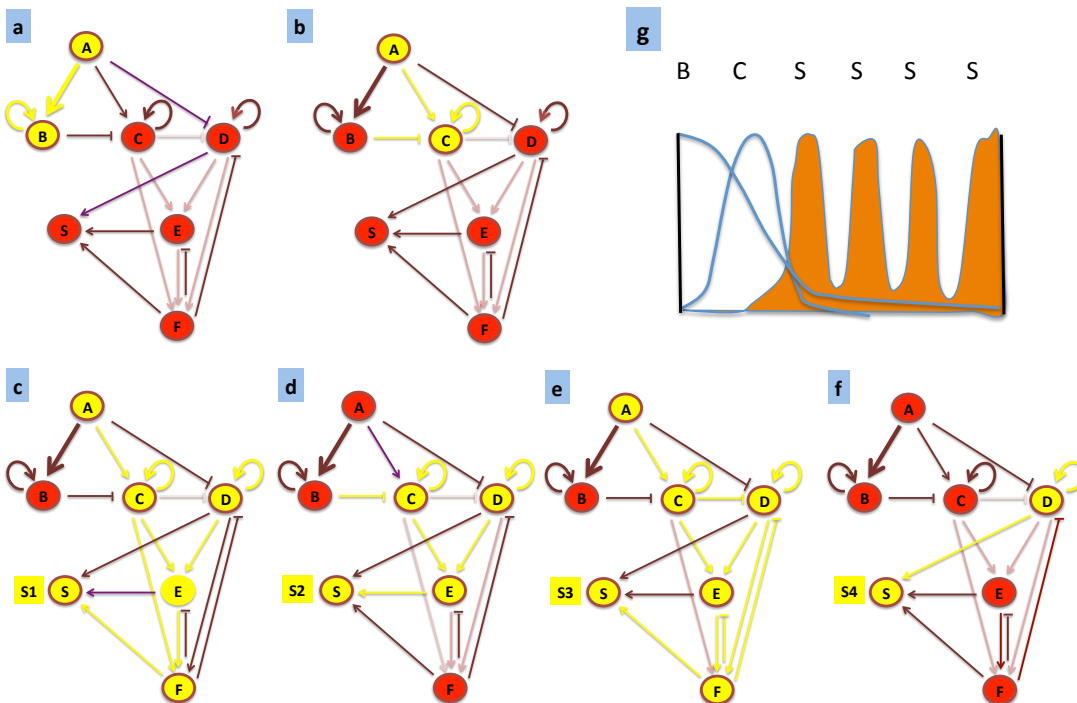


**Figure 3 :**  
**Le réseau de la mise en place des segments chez la drosophile.**  
 a) les régulateurs et les interactions, les flèches en marron

clair indiquent une interaction conditionnelle, par exemple C n'active F que si E est activé ; b) les fonctions associées à chaque nœud ; c) les segments de la larve, S s'exprime dans les bandes bleues, B à l'antérieur et C au postérieur ; la mise en place de l'expression de T utilise un système (non représenté) parallèle à celui montré en a.

Ce diagramme se déchiffre en suivant les activités des FTs qui forment des cascades. Par ailleurs ces cascades donnent au système sa temporalité.

Ainsi, si A est à forte concentration alors B est activé ; B actif active B (boucle d'autorégulation positive). Cela correspond à l'un des états stables du système. Il donnera un territoire « B » (figure 4 a, les nœuds et les flèches jaunes ; g, le pic B).



**Figure 4 :**  
**Les 6**  
**cascades**  
**du réseau.**  
 a à f, les  
 cascades  
 d'activation  
 du  
 réseau abou  
 tissant à des  
 états stables  
 ; g) les  
 activités  
 dans les  
 territoires  
 de B, C et

S, représentation schématique de la larve de la figure 3c.

Si A est en concentration moyenne alors C est activé ; C actif active C (boucle d'autorégulation positive). Cela correspond à un deuxième état stable du système. Il donnera un territoire « C » (figure 4 b les nœuds et les flèches jaunes ; g, le pic C).

Si C et D sont actifs alors ensemble ils activent E et ils s'auto activent. Puis d'une part C et E activent F qui va activer S, correspondant à un territoire « S1 » (figure 4 c les nœuds et les flèches jaunes ; g, le pic S avant). D'autre part E et D vont activer F qui va activer S, correspondant à



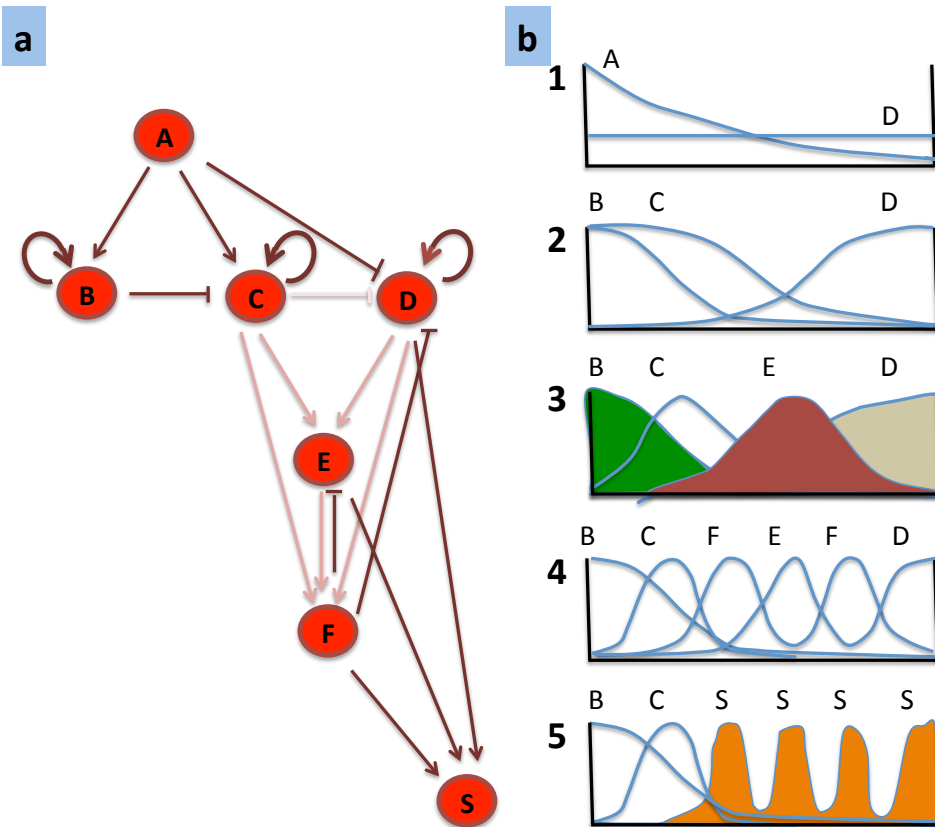
un territoire « S3 » (figure 4 e, les nœuds et les flèches jaunes ; g, le troisième pic S) et enfin E dont l'expression persiste active S, correspondant à un territoire « S2 » (figure 4 d, les nœuds et les flèches jaunes ; g, le deuxième pic S).

Si A est à faible concentration alors D reste activé ; puis il active S, correspondant à un territoire « S4 » (figure 4 f les nœuds et les flèches jaunes ; g, le quatrième pic S) et enfin est inactivé par F qui avait été activé par E et D.

Les bandes sont donc identifiables en tant qu'états stables du système. Ce système admet l'expression de B, C et S à  $t+4$  (la cascade qui le constitue a quatre flèches). S1, S2, S3, et S4 ont tous la même identité génétique S. Mais ils correspondent à des territoires différents dans l'embryon (figures 3c et 4g). Tous dépendent d'au moins une condition de concentration impliquant A. La biologie montre que A est positionné à l'un des pôles de l'œuf oblong de la drosophile. Au gradient de concentration de A correspondent donc des bandes transverses. Les chemins dans le réseau pour parvenir à B, C, S1, S2, S3, et S4 sont tous différents et concernent un nombre variable de nœuds, de 2 à 6 et de flèches. Les « histoires » de S1, S2, S3, et S4 sont différentes.

A chaque état stable un territoire ! C'est magique ! Comment cela se fait-il ? Pour bien le comprendre reprenons le raisonnement en suivant avec la figure 5b non seulement les activités mais aussi les positions des FTs dans l'embryon en fonction des étapes.

A la première étape A présent en gradient de concentration à l'extrémité antérieure de l'œuf active à forte concentration B et à moyenne concentration C (Figure 5b, le passage de 1 à 2 ; représentation de l'œuf, l'antérieur à gauche, l'arrière à droite). Ces facteurs transcriptionnels s'accumulent en pics aux positions correspondantes (Figure 5b, 2). De plus, A agit comme un répresseur de D. D ne s'accumule donc que dans la région postérieure de l'œuf (Figure, 5b, 2). A l'étape suivante, c'est-à-dire lorsque B, C et D ont atteint leur seuil d'activité, B réprime C, ce qui positionne son pic d'accumulation à l'arrière de B. Les deux facteurs C et D coopèrent



**Figure 5 : Le réseau, ordonné temporellement, de la mise en place des segments chez la drosophile.** a) les flèches en marron clair indiquent une interaction conditionnelle ; b) représentation de la larve, l'antérieur à gauche, l'arrière à droite ; les cinq étapes.

alors pour activer E. Comme ils sont tous les deux requis pour cette activation, E se localise à l'intersection des pics de C et de D (là où les deux facteurs sont simultanément présents) derrière C et devant D (Figure, 5b, 3). Durant cette étape, B et C activent leur propre production, ce qui maintient leur pic à leur position. Cette auto-activation perdurera jusqu'à l'étape 5 (Figure, 5b, 3 à 5). A la troisième étape c'est-à-dire de nouveau lorsque E a atteint son seuil d'activité, C et E coopèrent pour activer F qui se positionnera donc à l'arrière de C et à l'avant de E. De même E et D vont coopérer pour activer F, créant un second pic de F à l'arrière de E et à l'avant de D (Figure, 5b, 4). Puis F va réprimer E et D, ce qui va affiner les pics de ces deux facteurs transcriptionnels. A la quatrième étape, F, D et E vont activer S créant 4 pics séparés de S (Figure, 5b, 5).

En quatre étapes, sont ainsi définies de l'avant à l'arrière de l'embryon 6 domaines d'expression génique. Ces domaines ont soit une identité génétique propre (B et C), soit répétée (S).

Réitérons-le, ce qui crée les bandes S se sont les concentrations des FTs, d'abord le gradient de concentration de A, les trois entrées dans le réseau par B, C, et D. Et l'on joue encore sur les concentrations pour n'activer E que dans une région de l'embryon, là où les concentrations de C et D sont adéquates. Par adéquat il faut comprendre les concentrations qui font que l'activateur distal de E se met à fonctionner.

Ce réseau a été établi suite à une longue histoire qui commença par l'analyse de toutes sortes de mutants fort curieux dont le développement s'arrête précocement<sup>36</sup>. Ces mutants furent obtenus en réalisant ce que l'on appelle des cribles génétiques. La conception de ces cribles est décrite dans l'annexe 1 « le crible génétique des mutants létaux embryonnaires ».

L'isolement des gènes concernés en dévoila ensuite la nature, des FTs, et permis la mise au point de molécules (appelés sondes) permettant de décrire leur expression (c'est-à-dire la localisation des cellules qui contiennent leur ARN messenger et de leurs protéines). Cela a abouti au schéma que nous avons vu.

Eh bien voyons si à partir de cette représentation l'on peut retrouver le phénotype des mutants de ces gènes.

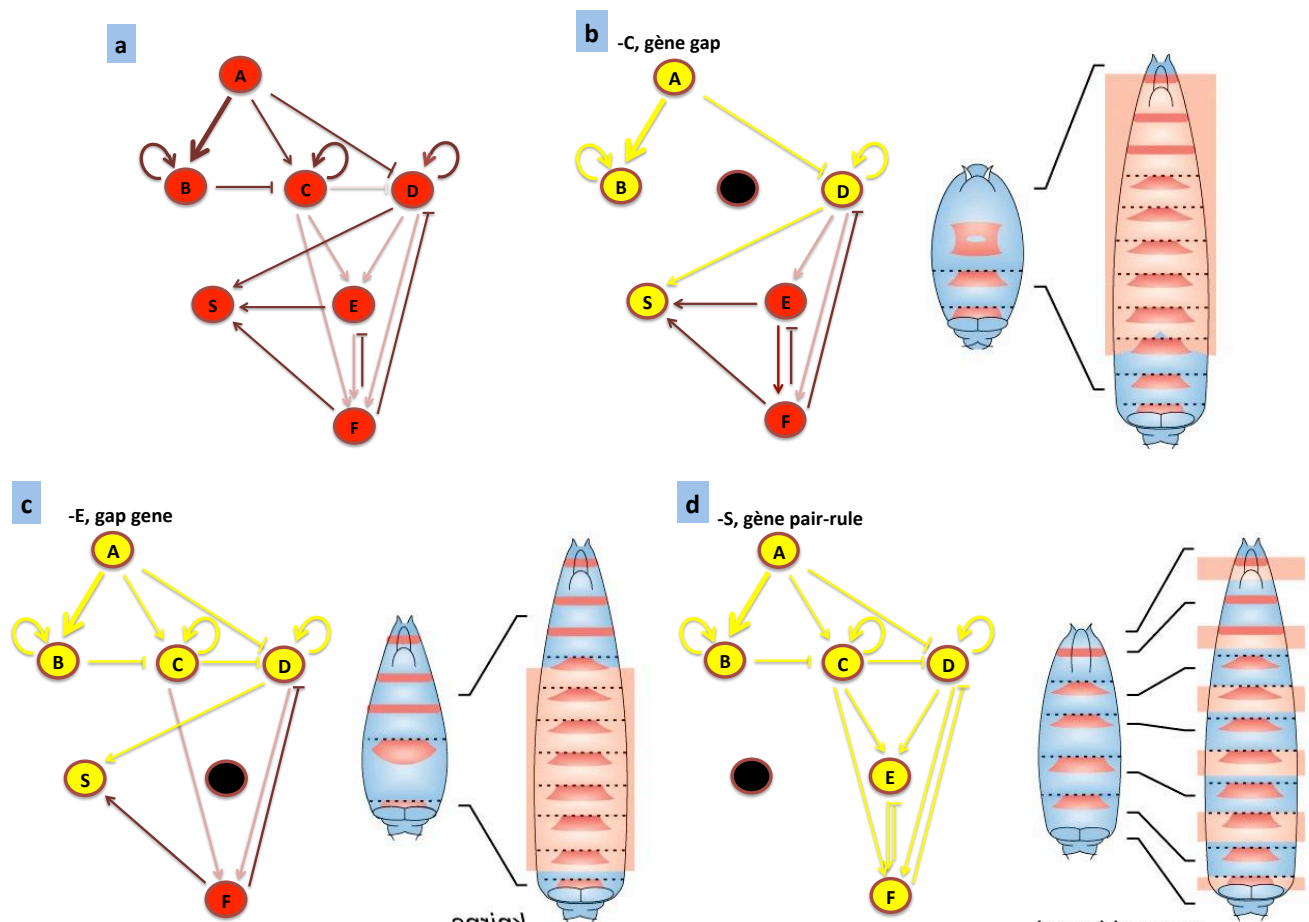
Par exemple supprimons C (Figure 6a et b). Alors ni E ni F ne sont plus activables. On perd donc C et S1, S2 et S3 (Figure 6b). Ceci correspond aux mutations dites « de type gap » où une grande région de la larve est absente ici presque tous les segments sauf le plus caudal.

Supprimons maintenant E alors F ne peut pas être activé. On garde donc B et C et l'on perd tous les segments qui nécessitent ou E ou F donc S1, S2 et S3 (Figure 6c). Ceci correspond à une autre mutation de type gap.

Enfin supprimons S. tous les éléments du réseau sont activables mais S ne peut plus être activé. On perd donc S1, S2, S3 et S4 (Figure 6d). Ceci correspond à un des mutants dit « pair-rule » où 1 segment sur deux est absent (en effet pour être plus complet on aurait dû rajouter une cinquième étape à la figure ou des segments T se seraient intercalés entre les segments

S pour donner les 14 segments de la larve de la drosophile, (fushi tarazu et even skipped 7 bandes chacun).

Le phénotype des mutants peut donc être prédit de cette représentation. On saisit qu'ils correspondent à une altération non d'un gène mais du fonctionnement du réseau. Le rôle des gènes prend son sens par rapport au réseau. Dans un autre contexte les gènes auraient un rôle différent.

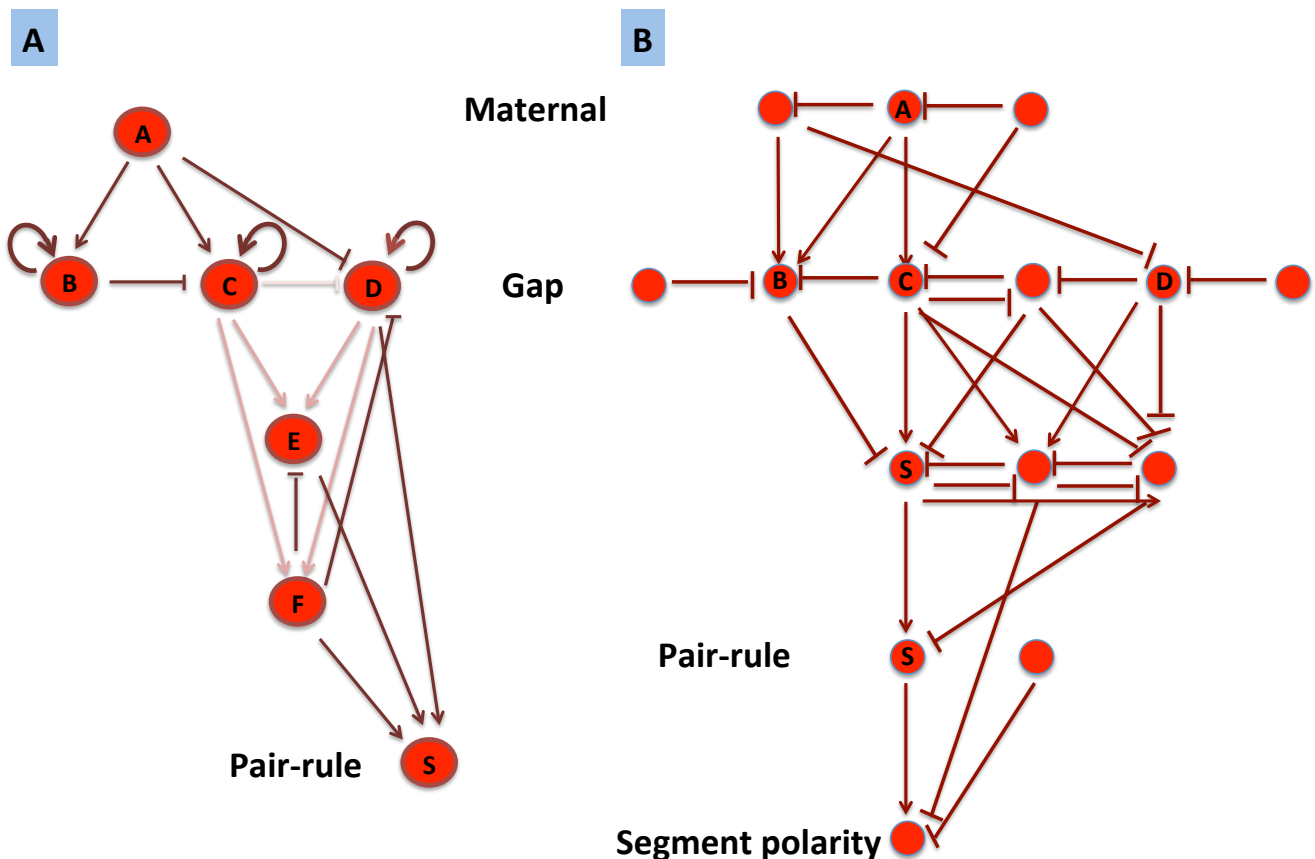


**Figure 6 : Phénotype de mutants des gènes du réseau.** a) le réseau ; b) simulation d'une mutation de C, le phénotype correspond à un mutant de type gap ou une partie des segments est manquante ; c) simulation d'une mutation de E ; d) simulation d'une mutation de S, le phénotype correspond à un mutant de type pair rule où un segment sur deux est absent (les segments « S »). b à d), en jaune les éléments du réseau qui restent actifs dans le contexte des mutants.

En résumé, ce réseau présente aussi une simplicité structurale. Le FT à l'entrée du réseau agit sur trois cibles selon sa concentration. Cela crée autant de territoires. Les autres étapes sont combinatoires (C+D, C+E, E+D) puis directes (F, E, et D activent S). Ce sont les combinatoires qui créent de nouveaux territoires. L'inhibition joue un rôle important. On détecte trois états stables (C, D et S) tous réalisés. La situation est particulière aux insectes car l'absence de membrane cellulaire entre les noyaux permet les interactions directes entre les FTs et les activateurs distaux des gènes.

Le graphe a de la figure 3 est une version très simplifiée des trois premières lignes du graphe b de la Figure 7 qui représente les éléments connus en 2001 qui rendent compte de l'activation du gène *Eve* un gène dit « pair rule » (inspiré <sup>43</sup> p. 59, fig. 3.5). Les deux autres lignes correspondent à la suite des interactions qui finalement activent le gène *Engrailed* dans la partie postérieure de chacun des 14 segments de la larve.

**Figure 7 : Le réseau de la mise en place des segments chez la drosophile.** a) le réseau simplifié ; b) le réseau plus complet, pour un segment et avec les étapes supplémentaires pour activer *Engrailed*. (Gène dit de « segment polarity »).



On a vu comment ensuite *Engrailed* active *Wg* dans la partie antérieure de chacun des 14 segments et comment ces gènes sont impliqués dans le positionnement de l'expression de gènes maîtres des organes comme *Dll* (Figures 23 et 29 de l'opuscule). Nous avons donc une idée de ce qui se passe du début du développement à l'initiation des organes chez la drosophile.

Le système génétique qui comprend tous les gènes dont les interactions mutuelles découpent l'embryon en onze territoires disposés de l'avant à l'arrière a été appelé le système "avant-arrière". Un autre système, très différent, le système "dorso-ventral" découpe parallèlement l'embryon en quatre territoires dorso-ventraux génétiquement différents.

Que contrôlent ces systèmes ? Eh bien, de nouveau, ils contrôlent les opérations cellulaires qui forment les segments. Ces opérations sont reliées à ces systèmes en conséquence (pas à n'importe quel FT). Le déploiement des activations, les chemins empruntés dans le réseau forment la dynamique du développement.

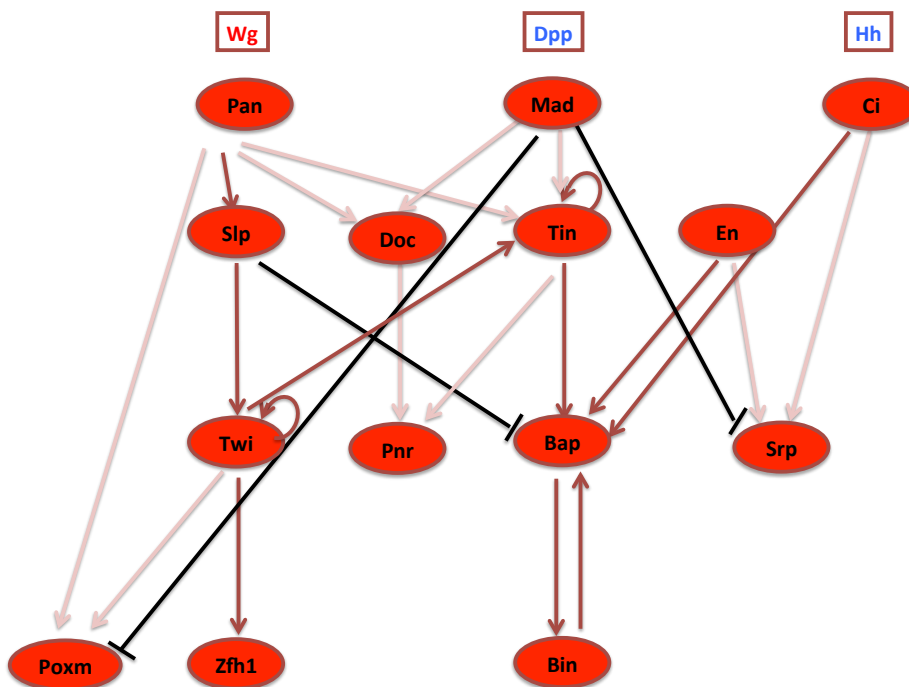
### **Le réseau de l'initiation de la formation des quatre organes dérivés des segments mésodermiques chez la drosophile<sup>44</sup>.**

Le troisième exemple que nous traiterons correspond à une situation où apparaissent les voies de signalisation. Elles seront des entrées dans un réseau complexe. Ces voies de signalisation vont jouer le rôle qu'ont joué les gradients de concentrations et les combinatoires dans le système précédent ou la structure est dépourvue de membranes cellulaires. Ici la structure est constituée de cellules. Le système donne 4 réalisations, 4 organes. Il comporte 3 voies de signalisation qui se combinent. Le circuit va-t-il encore être déchiffrable ?

Les tissus présomptifs du cœur, les muscles viscéraux, les muscles somatiques et le « fat body » sont formés dans les segments mésodermiques de la drosophile entre le stade 8 et 11. Ce sont des signaux émis par l'ectoderme qui activent le système dans le mésoderme. Ici sont envisagés les aspects « réseau ». Pour les aspects biologiques voir l'annexe 4 « les communications entre cellules par les voies de signalisation ».



Ce réseau a 14 nœuds correspondant à des facteurs de transcription et trois correspondant à des morphogènes (Wg, Dpp, et Hh, les signaux émis par l'ectoderme) (Figure 8). Il comporte 25 interactions (Figure 8, les flèches). Beaucoup correspondent à des motifs de réseau : une double boucle de rétroaction positive (entre Bap et Bin), deux autorégulations positives (pour Tin et Twi), cinq FFL cohérent type 1 (pour Doc, Tin, Pnr, Srp et Poxm) et 3 inhibitions directes (pour Poxm, Bap et Srp). On remarque 3 cascades linéaires directes.



**Figure 8 : Le réseau de l'initiation de la formation des quatre organes dérivés des segments mésodermiques chez la drosophile.** Le nom des régulateurs est indiqué dans les ellipses qui correspondent aux nœuds du réseau. Les interactions sont représentées par les flèches. Les flèches en marron clair indiquent une interaction

conditionnelle, par exemple Pan n'active Poxm que si Twi est activé

Si nous ne connaissons rien de ce système ces cascades linéaires directes et ces inhibitions nous aideraient à l'ordonner. Ces 3 cascades en effet agencent le système par rapport au temps et suggèrent d'emblée trois suites d'états probablement significatives. L'inhibition par Mad de la voie de droite et l'activation de Bap par cette voie de droite désignent une interaction entre Mad et Ci pour créer une voie incluant Bap et Bin. L'inhibition croisée entre Mad et Pan et le FFL cohérent type 1 sur Doc suggèrent une quatrième cascade, Doc Pnr, nécessitant ces deux régulateurs. Il est donc clair que Pan, Mad et Ci sont des éléments qui seuls

ou en combinaison génèrent des états du système auxquels vont correspondre des états des groupes de cellules qui répondent à ces signaux.

Décrivons ces états.

Si Ci est actif alors Srp est activé, puis ses cibles (puis retour à 0 pour Srp après disparition du signal qui active Ci). Or la signature du fat body, est l'expression de Srp (Srp est le gène maître du FB).

Si Pan est actif alors Slp, Twi, Zfh1 et Poxm sont activés dans cet ordre avec Twi, Zfh1 permanents. Si Pan est maintenu (par un signal externe ?) alors Slp, Twi, Zfh1 et Poxm restent actifs. Or la signature des muscles somatiques est Twi (niveau 2), Pox meso (Poxm), DSix4 and Zfh1 (ce sont les gènes maîtres des muscles somatiques).

Si Mad est actif aucun autre élément du système n'est activé car les cibles potentielle sont protégées par des FFL incohérent ET.

Si Pan et Mad sont actifs alors Doc et Tin sont activés et ensemble ils activent Pnr. Or la signature du cœur est expression of Tin (niveau 2), Doc, and Pannier (Pnr) ; de plus Mad inhibe le gène maître Poxm et Pan Slp inhibe le gène maître Bap.

Si Ci et Mad sont actifs alors Tin (?) Bap et Bin sont activés. Or la signature des muscles viscéraux est expression de Tin (niveau 1), Bap (niveau 2), and Bin.

On avait donc bien identifié intuitivement 4 états de ce système qui correspondent aux 4 états observés par les biologistes. Ce système n'est donc pas impénétrable, on peut le déchiffrer assez facilement. Il est même étonnamment simple. Les « accès » aux gènes maîtres sont tous protégés par des FFL incohérent type 1. Des inhibitions directes ciblées sur les gènes maîtres, contribuent à ce que les combinaisons de signaux ouvrent un nouvel état qui exclut les éléments des autres états possibles. Les motifs de réseau jouent leur rôle indépendamment les uns des autres. Leur comportement n'est pas affecté par le reste du réseau.



Testons le système comme nous l'avons fait pour le système précédant en fixant l'activité de certains nœuds.

Si les trois entrées Pan, Mad et Ci des signaux des cellules voisines sont actives, alors Slp, Twi, Zfh1 seront activés mais pas Poxm ; Doc, Pnr seront activés ; Tin mais pas Bap ni Bin ; pas Srp. Cette combinaison se traduit par une inhibition de trois des quatre voies du système.

Si les cibles directes des FTs activés par les signaux des voies de signalisation sont toutes actives (Slp etc.) alors Twi et Zfh1 seront activés mais pas Poxm ; Doc, Pnr seront activés ; mais pas Bap ni, Bin et pas Srp. Cette combinaison se traduit par une inhibition de trois des quatre voies du système. L'activation des cibles directes des FTs activés par les signaux ne se traduit pas par l'activation de tous les nœuds du système. Cela est dû à la fois aux fonctions inhibitrices de ces cibles (Slp) et aux FFL cohérent type 1. On peut y voir une protection du système contre les expressions inappropriées, ce qui est appelé le « bruit » du système.

De nouveau le système est compréhensible par ses éléments. Il présente une certaine simplicité structurale. Il est très protégé des « erreurs ». Il ne comporte que des solutions réalisées (apparemment pas d'attracteurs cryptiques).

Nous avons là un réseau qui comporte des éléments externes aux cellules, les morphogènes Wg, Dpp, Hh et des éléments internes, à commencer par les trois FTs des voies de signalisation, Pan, Mad et Ci et les autres, les gènes maîtres des organes et tissus. A ces deux niveaux sont associées des propriétés biologiques très différentes. Les éléments externes créent des territoires et de la géométrie dans l'embryon. Les éléments internes du réseau gèrent la dynamique des opérations cellulaires d'un et d'un seul des territoires créé par les morphogènes. Et bien sur les FTs activés dans les quatre territoires contrôlent beaucoup d'autres FTs.

### **Conclusion sur les réseaux booléens.**

Les réseaux que nous avons présentés sont une manière synthétique de représenter les résultats de l'analyse génétique. Les trois exemples discutés compilent toute la littérature ayant trait aux gènes impliqués dans les

processus étudiés. On en déduit un réseau d'interactions entre molécules (les gènes et/ou leurs produits), les variables, et les opérations logiques associées. Pour les réseaux booléens les variables ont deux états actif, inactif, (0,1). Elles sont reliées par des relations logiques de régulation (si A actif alors B actif), les fonctions F.

Ces représentations sont plus qu'un résumé ce sont des modèles simplifiés des interactions qui régulent une série d'opérations. On peut donc comme nous l'avons fait en simuler le fonctionnement. L'évaluation, nécessaire, du réseau va consister à vérifier si des phénotypes connus sont prédits correctement. On peut aussi en tester la cohérence. Reproduit-il tous les aspects qualitatifs observés ? Un défaut de cohérence conduit à déterminer logiquement l'interaction et/ou l'élément de régulation manquant, à les rechercher *in vivo* et à les introduire dans le modèle. On voit là l'un des intérêts majeurs de ces représentations en réseaux.

Ces représentations permettent aussi de repérer les motifs de réseau et de leur donner du sens dans l'ensemble du processus que contrôle le réseau. Regarder de nouveau la figure 7 et notamment les inhibitions réciproques. La complexité qui résulte de leur juxtaposition fait que l'on perd l'intuition de ce qu'ils signifient. Dans le cas présent la troisième ligne correspond à la mise en place des segments impairs (ftz, 1, 3, 5..) qui viennent s'intercaler entre les segments pairs (eve 2, 4, 6), précédemment créés. Seule une analyse de l'ensemble du système permet de le saisir puis d'en vérifier la cohérence. Souvent alors, ce qui à première vue pouvait apparaître comme superflu, anormalement compliqué ou même être interprété trop rapidement comme du bricolage prend son sens.

Les trois réseaux que l'on a vus sont tous compréhensibles par leurs éléments. Même le plus complexe présente à l'analyse une simplicité structurale certaine. Cela suggère que les réseaux même complexes sont essentiellement des groupements de plusieurs motifs de réseaux dont le rôle peut être compris.

Une fois bien établie la représentation peut servir à simuler le fonctionnement du système en présence de perturbations génétiques ou

pharmacologiques. Enfin elle sert à simuler les phénotypes de double ou triple mutants parfois impossibles à produire génétiquement.

Enfin n'oublions pas que sont reliés à certains des nœuds de ces réseaux les gènes qui effectuent les opérations cellulaires du développement. La dynamique des opérations, leur temps d'action, leur succession etc. dépendent en grande partie de la structure du réseau. Le résultat ce sont ces extraordinaires ballets de cellules et de tissus, ces changements de forme, ces séries de divisions que l'on voit dans les films. Et encore ils ne représentent qu'une fraction de ce qui se passe aux niveaux des cellules. Ainsi les quelques gènes de ces réseaux contrôlent des centaines voire des milliers d'autres, les gènes effecteurs (qui ne sont pas des FTs).

Le génome code les produits des gènes mais aussi leurs interactions entre eux. Le réseau associé à un génome comporte des milliers de nœuds et des dizaines de milliers de flèches. Si le réseau du génome d'un microbe, *Escherichia Coli* ou d'un champignon unicellulaire eucaryote, la levure (*Saccharomyces cerevisiae*), commencent à être aujourd'hui connus, ni le réseau de la drosophile, ni celui d'aucun autre l'animal ne le sont. Les analyses que nous avons discutées, ne concernent qu'une infime partie du réseau de la drosophile, celle accessible à l'analyse génétique et franchement impliquée dans le processus étudié. Elles ignorent l'état d'activité du réseau hors la fraction étudiée. Le réseau est pour l'instant, traité par micro-morceaux.

Réaliser que les réseaux représentent une réalité c'est reconnaître que c'est un niveau sur lequel la sélection joue. Certaines mutations vont optimiser des motifs de réseau ou le réseau (en jouant surtout sur les niveaux d'expression). D'autres mutations vont changer les liens des éléments du réseau entre eux, en créer ou en supprimer. Elles peuvent ainsi d'un coup faire apparaître un nouveau motif de réseau qui va changer les propriétés de l'ensemble ou même vont fusionner des « modules » entiers entre eux. Elles cooptent ainsi d'un coup toute une série d'opérations et l'affectent à une opération existante, voire en créent une nouvelle. Toutes ces mutations ont un effet direct sur le déroulement des opérations du développement et peuvent changer fondamentalement les types cellulaires, les organes, et leur disposition. Ce sont des accélérateurs de l'évolution.

On voit ainsi qu'elles sont très différentes des mutations qui affectent la structure des gènes comme les mutations ponctuelles qui changent un acide aminé pour un autre, les remaniements des gènes entre eux par associations de domaines (à l'origine de nouveaux gènes) ou les duplications des gènes (à l'origine des familles de gènes). Ces mutations ne modifient aucunement l'organisation de leur expression ni dans le temps ni dans l'espace. Elles ont peu de chance de changer l'organisme.

Mais, cela va plus loin. Rappelons que de certains des arrangements que produisent, par chance, les mutations dans le réseau émergent des fonctions. Ce sont les fonctions des motifs de réseau ou plus largement les différents types de dynamique liés au réseau. Leur apparition n'est pas sélectionnée. Elles surgissent d'elles-mêmes, ne sont pas prédictibles et apportent inopinément de l'ordre au système. C'est de l'organisation spontanée. Stuart Kauffman a popularisé cette notion en parlant « d'ordre gratuit » apporté au système. Dans les théories de l'évolution il faut marier organisation spontanée et sélection<sup>34</sup>. Cela implique que les systèmes biologiques (en fait tous les systèmes) sont contraints, et imprévisibles, car basé sur des émergences. Il faut abandonner l'idée d'une théorie du tout déduite d'une seule loi de la physique (<sup>35</sup> et quelques autres pour qui non seulement on ne sait pas ce qui arrivera mais on ne sait même pas ce qui peut arriver « we even don't know what can happen »).

### **Vers un modèle mathématique du développement embryonnaire**

Les réseaux booléens sont une simplification des réseaux de régulation. Souvent des étapes intermédiaires, qui n'altèrent pas la logique du système, ne sont pas représentées. Il n'est pas tenu compte des cinétiques de production et dégradation des protéines. Ces réseaux négligent donc toutes sortes de paramètres comme les paramètres de délai qui sont cruciaux pour le développement surtout si l'on songe à ce que les variables ne contrôlent pas seulement d'autres variables du réseau mais aussi des molécules impliquées dans des opérations cellulaires du développement qui n'apparaissent pas dans la schématisation. La dynamique que l'on peut en extraire est une dynamique ultra simplifiée.



La réintroduction de tous ces paramètres est une tâche lourde mais nécessaire. Les réseaux booléens sont l'étape qui prépare les modèles quantitatifs qui eux-mêmes préparent les modèles mathématiques du développement embryonnaire.

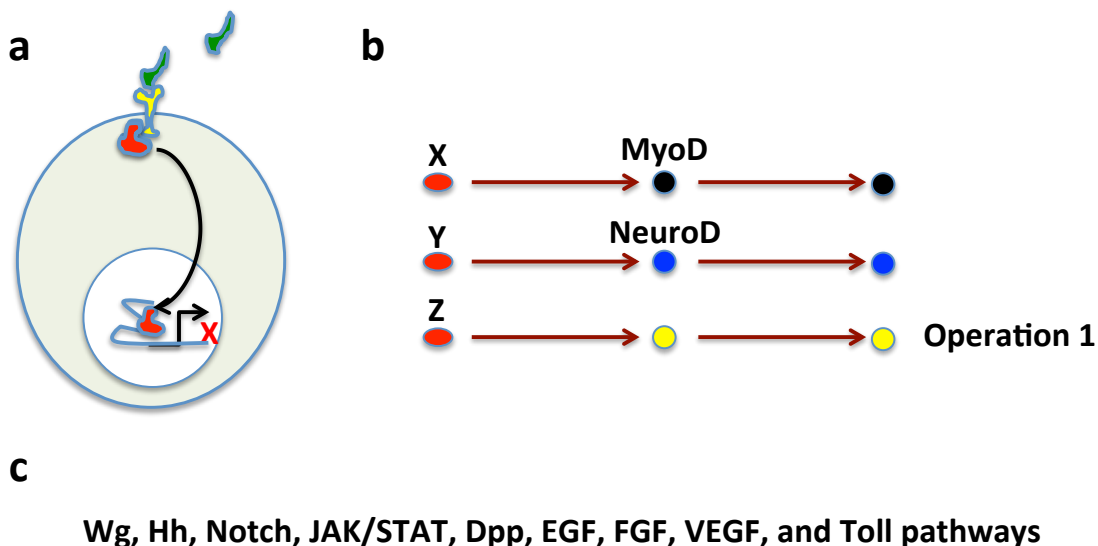
## ANNEXE 4

### LES COMMUNICATIONS ENTRE CELLULES PAR LES VOIES DE SIGNALISATION

La grande majorité des processus du développement nécessite des mécanismes de communication des cellules entre elles, c'est-à-dire des processus ou des messages émis par des cellules sont reçus par d'autres cellules qui en réponse activent de nouveaux gènes. Parfois au signal reçu les cellules répondent en modifiant directement des éléments impliqués dans les comportements cellulaires. Il n'y a qu'une dizaine de voies de signalisation utilisée par les développements. Mais les voies de signalisation se combinent entre elles et aux combinatoires de leurs FTs s'ajoutent des FTs spécifiques de la cellule qui reçoit les signaux ce qui multiplie leurs cibles. Les FTs des voies de signalisation ont presque directement des gènes maîtres d'organes ou de tissus pour cibles. La dynamique externe des interactions cellulaires gère la dynamique de l'apparition des territoires et dicte leur géométrie. La dynamique interne à la cellule du réseau transcriptionnel gère la dynamique des opérations cellulaires d'un territoire.



Nous analysons ici l'implication de la communication entre cellules dans le développement. Quel en est le rôle ? Comment seulement quelques voies de signalisation peuvent-elles générer le grand nombre de territoires nécessaires au développement ? Comment les FTs des voies de signalisation accèdent-ils aux gènes maîtres d'organes ou de tissus ? Comment les quelques FTs des voies de signalisation activent-ils autant de cascades différentes ? Et pourquoi si peu de voies de signalisation plutôt qu'une multitude ? Les réponses nous permettront d'envisager le pourquoi du succès des interactions cellulaires et non des arrangements d'éléments du génome entre eux.



**Figure 1 :**  
Les voies de signalisation. a) un message (en vert) émis par une cellule est reçu par le récepteur (en jaune) d'une

autre cellule qui modifie un FT (en rouge) qui après transfert dans le noyau active un gène X (ou Y ou Z) ; b) X, Y, ou Z activent des FTs spécifiques de tissus ou qui contrôlent des opérations cellulaires ; c) les 9 voies de signalisation les plus utilisées dans les développements.

Nous avons vu que le processus de la différenciation cellulaire et celui beaucoup plus élaboré de la formation des organes utilisent des mécanismes de communication entre cellules. Un message émis par une cellule est reçu par une autre cellule qui en réponse active de nouveaux gènes par démasquage d'un FT (Figure 1a, voir aussi la figure 23 de l'opuscule). Dans la différenciation cellulaire les gènes activés sont par exemple les gènes qui font qu'une cellule est musculaire (MyoD) ou qu'une autre est nerveuse (NeuroD) (Figure 1b). Dans la formation des organes c'est un FT

qui va être à l'origine de cascades d'activations de gènes dont dépend une série d'opérations cellulaires. Cela se produit plusieurs fois au cours du processus et dans plusieurs territoires.

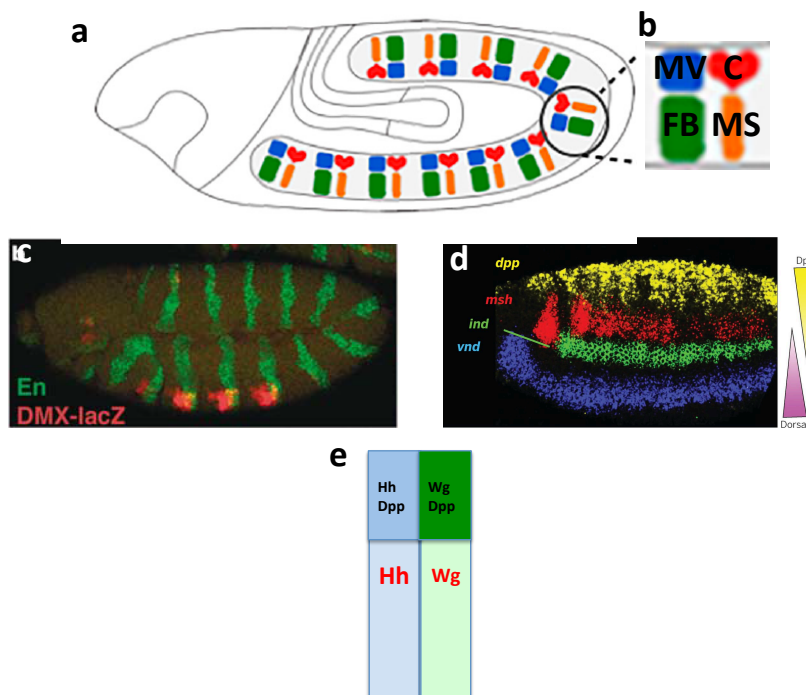
Notons que parfois au signal reçu la cellule répond en modifiant directement les éléments impliqués dans les comportements cellulaires (un exemple est la modification directe du cytosquelette via la voie Wnt Rok). Et alors un nouveau comportement cellulaire apparaît immédiatement, sans le délai dû à la transcription et à la traduction.

Enfin le problème du positionnement des organes se résout aussi à un problème d'activation et de répression des gènes, via les voies de signalisation. Rappelez-vous le contrôle de Dll par trois voies de signalisation chez la drosophile (Figure 29 de l'opuscule).

Les voies de signalisation sont donc centrales dans les développements. En conséquence elles sont une mesure de leur complexité. On pourrait penser qu'à la complexification des développements correspondrait la multiplication des signaux et des voies de signalisation. A chaque nouveau type cellulaire, à chaque nouvelle opération du développement, un nouveau signal, une nouvelle voie de signalisation. Mais ce n'est pas ce qui est observé. On constate que les développements quelles que soient leur complexité utilisent moins d'une dizaine de voies de signalisation, toujours les mêmes (figure 1c), que ça soit au travers d'un développement particulier ou au travers de tout l'arbre du vivant. En conséquence on pourrait alors imaginer qu'à une voie de signalisation correspondrait un nombre croissant de FTs. Mais de nouveau, ce n'est pas le cas. A une voie de signalisation ne correspond qu'un FT.

Comment si peu de voies de signalisation, si peu de FTs peuvent-ils contrôler autant d'événements différents ?

La réponse à ces questions va nous être donnée par l'analyse de l'initiation de la formation des organes des segments mésodermiques de la drosophile<sup>44</sup>, le cœur, les muscles viscéraux, les muscles somatiques et le « fat body ». Cela se passe très tôt, entre les stades 8 et 11, juste après la gastrulation. (Figure 2a, voir aussi les figures 3 et 29 de l'opuscule).



**Figure 2 : L'initiation de la formation des organes des segments mésodermiques de la drosophile.**

a) les 11 segments mésodermiques au stade 8 de la larve ; b) les secteurs du cœur (C), des muscles viscéraux (MV), du fat body (FB) et des muscles somatiques (MS) ; c) les 11 segments ectodermiques au stade 8, les spots rouges dans T1-T3 marquent l'expression de Dll ; les bandes vertes expression d'En ; d) stade 5 (blastoderme) en jaune l'expression de dpp ; e) les 4

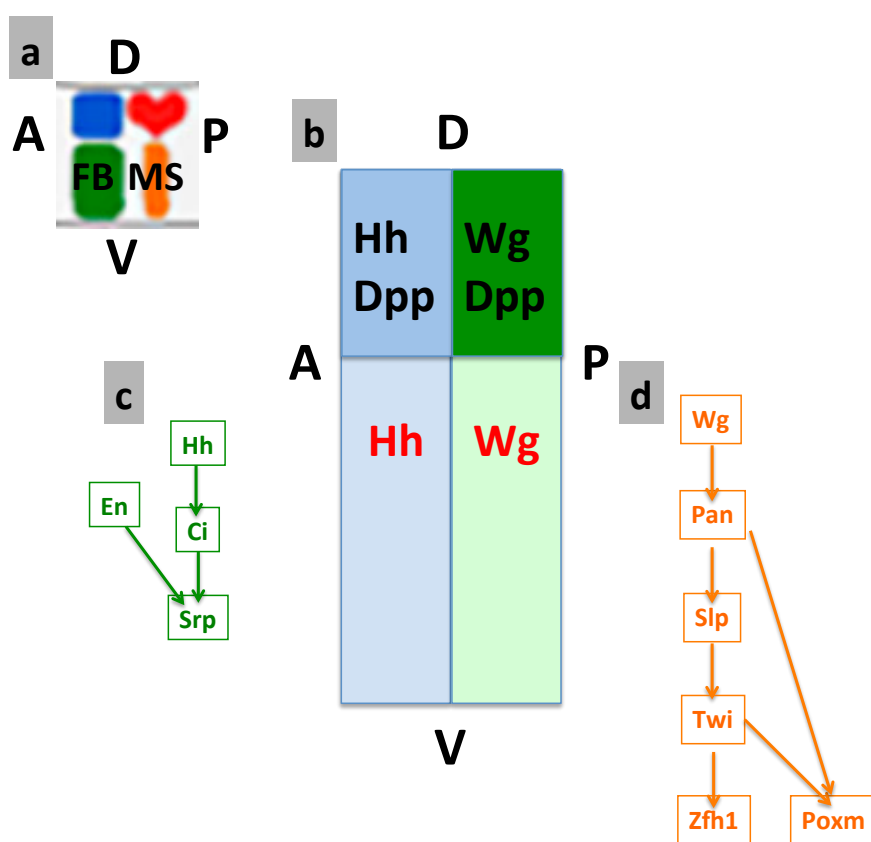
secteurs distincts des segments mésodermiques qui sécrètent Hh, ou Wg, ou Hh et Dpp, ou Wg + Dpp.

Durant cette même période se fait le positionnement de Dll, dans les petits territoires rouges des trois segments thoraciques (Figure 2c). Nous avons vu que ce gène est activé dans ces segments par un morphogène, wingless (Wg), exprimé en bandes transverses (Figure 2c, les bandes vertes), lui-même activé par Hh exprimé aussi dans des bandes transverses (c'est l'interaction inductive : En active Hh, Hh active Wg, Figure 23 de l'opuscule). Rappelez-vous aussi que Dll est inhibé par un autre un morphogène, Dpp qui s'exprime dans un large secteur horizontal dorsal de l'embryon (Figure 2d<sup>45</sup>). Dpp subdivise donc les bandes Hh et Wg en deux. Au final on a donc à ce stade 4 secteurs distincts (Figure 1e), antéro-ventral, qui sécrète Hh, postéro-ventral, qui sécrète Wg, antéro-dorsal, qui sécrète Hh et Dpp, et postéro-dorsal, qui sécrète Wg + Dpp. Ceci concerne le feuillet superficiel de cellules, l'ectoderme.

Ce sont ces mêmes signaux, Wg, Hh et Dpp, émis par l'ectoderme (dont on sait faire remonter l'histoire à l'œuf) qui vont agir sur le feuillet adjacent plus profond de cellules, le mésoderme.

Voyons cela<sup>44</sup>.

(1) Le développement du « fat body » est contrôlé par un facteur de transcription appelé Srp, que l'on peut considérer comme le gène maître de ce tissu. L'initiation de sa formation (Figure 3) va correspondre à l'activation de ce gène. Dans le secteur antéro-ventral du mésoderme de chaque segment le morphogène Hh émit par l'ectoderme active la voie Hh. Le FT Ci est alors activé (CiR est modifié en Ci Act). Ci Act et Engrailed, qui est exprimé dans ce secteur, activent alors Srp (Figure 3c). C'est l'étape essentielle à la formation du « fat body » puisque Srp contrôle les autres facteurs transcriptionnels impliqués.



Il faut deux conditions pour que Srp soit activé, Ci Act qui résulte de l'induction ET la présence d'En dans les cellules qui dépend de l'histoire de celles-ci. On a vu que En est exprimé dans tout le secteur antérieur du segment. Si une des deux conditions manque, Srp n'est pas activé. C'est le cas dans les deux secteurs postérieurs où En n'est pas exprimé. Dans le secteur antéro-dorsal En est présent et les cellules sont exposées à Hh. Cependant Srp n'est pas exprimé car la voie Dpp le réprime directement (Figure 3b, voir le paragraphe 3). On voit qu'au final les conditions

d'expression de *Srp* sont : *Ci* & *En* & *!(Mad & Med)*. 3 flèches, 3 conditions. Cela se traduit au niveau du gène par la présence de trois séquences de reconnaissance, pour *En*, *Ci Act* et *Mad*.

(2) Continuons avec le secteur *Wg* et le FT associé à la voie *Wg*, *Pan*. Le morphogène *Wg* active la voie *Wg* des cellules du mésoderme du secteur postéro-ventral.

Les FTs essentiels au développement des muscles somatiques, de nouveau des sortes de gènes maîtres, sont *Zfh1* et *Poxm* (*Pox meso*). Pour être activé *Zfh1* requiert *Twi* un facteur mésodermique assez général. Dans ce secteur *Twi* est contrôlé par *Slp* (*sloppy paired*) qui lui-même est contrôlé par *Pan*. C'est cette cascade linéaire *Pan*, *Slp*, *Twi* qui active *Zfh1* (Figure 3d). *Poxm* quant à lui requiert à la fois *Pan* et *Twi* tous deux présents quand la voie *Wg* est activée (Figure 3d).

La voie *Wg* contrôle donc directement les éléments essentiels à l'activation des deux gènes clés dans le développement des muscles somatiques. Ces conditions nécessitent une petite cascade linéaire d'activation, *Pan*, *Slp*, *Twi* pour activer *Zfh1* et une coopération entre deux éléments de cette même cascade, *Pan* et *Twi* pour activer *Poxm*. Cela se traduit pour le gène *Zfh1* par la présence d'une séquence de reconnaissance pour *Twi* et pour le gène *Poxm* par la présence de deux séquences de reconnaissance pour *Twi* et *Pan*. Ces séquences contrôlent distalement ces gènes comme vu au chapitre xx.

Voyons maintenant comment cela se passe quand deux voies de signalisation sont activées dans les mêmes cellules. En particulier comment évite-t-on les mélanges d'activations qui mélangeraient les organes. Comment l'activation du tissu ou de l'organe que contrôlent individuellement ces voies est-elle empêchée ?

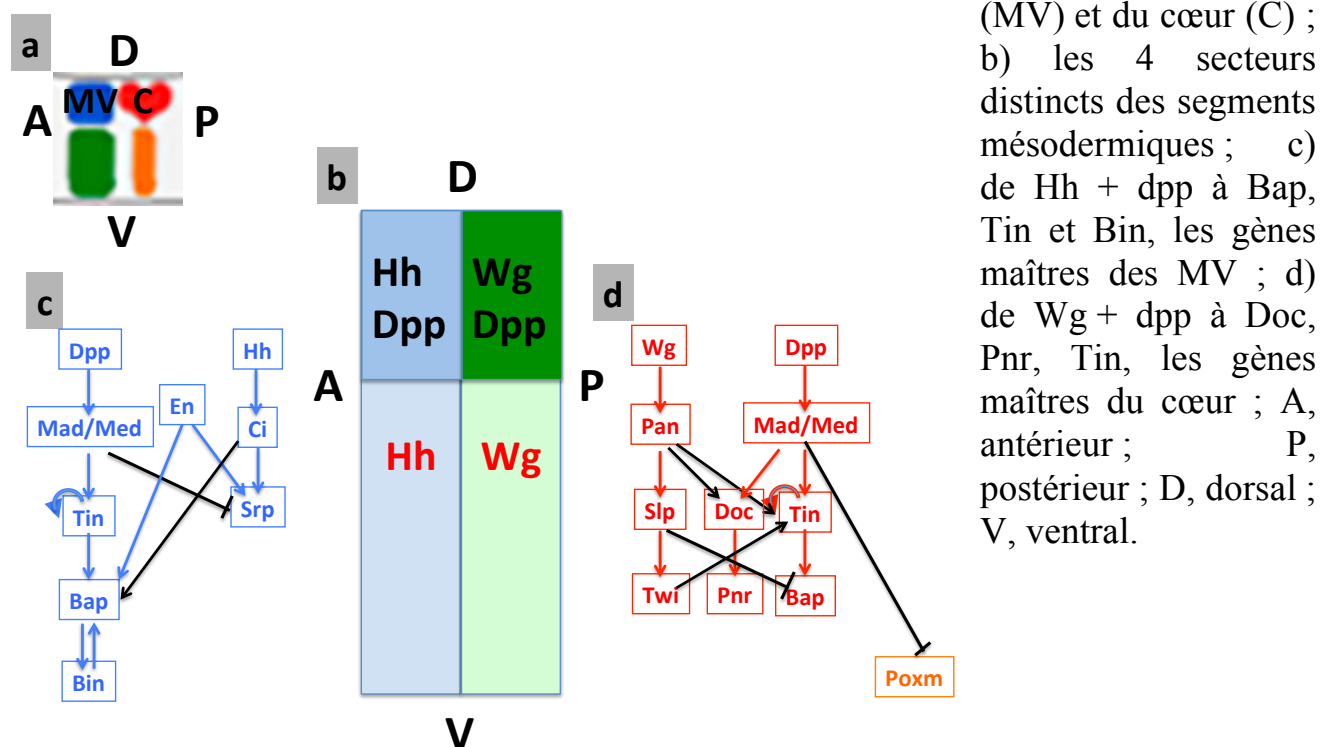
(3) Le secteur exposé à *Hh* et à *Dpp*.

On retrouve les mêmes états d'activation pour *En* et *Ci* que dans le secteur antéro-ventral (Figure 4c). Mais *Srp* n'est au final pas activé car le FT de la voie *Dpp*, *Mad/Med*, l'inhibe directement. Les cellules

n'expriment donc pas les éléments correspondant au « fat body » (Figure 4c, le trait noir).

Mad/Med active Tin. Ensuite Tin, Ci et En ensemble activent Bap qui active Bin. Bap, Tin et Bin sont les FTs clés des muscles viscéraux. Pour que ceux-ci se forment il faut donc quatre conditions, 2 voies de signalisation activées, une condition historique et une inhibition.

Figure 4 : Du morphogène aux gènes maîtres. a) les secteurs des muscles viscéraux (MV) et du cœur (C) ;



b) les 4 secteurs distincts des segments méso-dermiques ; c) de Hh + dpp à Bap, Tin et Bin, les gènes maîtres des MV ; d) de Wg + dpp à Doc, Pnr, Tin, les gènes maîtres du cœur ; A, antérieur ; P, postérieur ; D, dorsal ; V, ventral.

Notons la co-activation de Bap par trois éléments.

(4) Le secteur exposé à Wg et Dpp.

Les FTs essentiels à la formation du cœur sont Doc, Pnr et Tin. Tin est un élément commun au cœur et aux muscles viscéraux.

On retrouve le même état d'activation pour Pan, Slp et Twi (la voie Wg) que pour le secteur postéro-ventral. Mais Mad/Med inhibe Poxm ce qui empêche la formation des muscles somatiques (Figure 4d, le trait noir).

On retrouve aussi le même état d'activation Mad, Tin, Bap (la voie Dpp) que pour le secteur antéro-dorsal. Mais Slp activé par la voie Wg inhibe Bap ce qui empêche la formation des muscles viscéraux.

Ensemble, Pan et Mad/Med co-activent Doc qui active Pnr. Tin lui aussi est co-activé par les 2 voies (Figure 4d). Les trois FTs essentiels à la formation du cœur nécessitent donc les deux voies et leur activation ne nécessite pas d'intermédiaire.

### Conclusions

Le graphe de l'ensemble des régulations, de signalisation et transcriptionnelles, pour le mésoderme des segments est présenté dans la

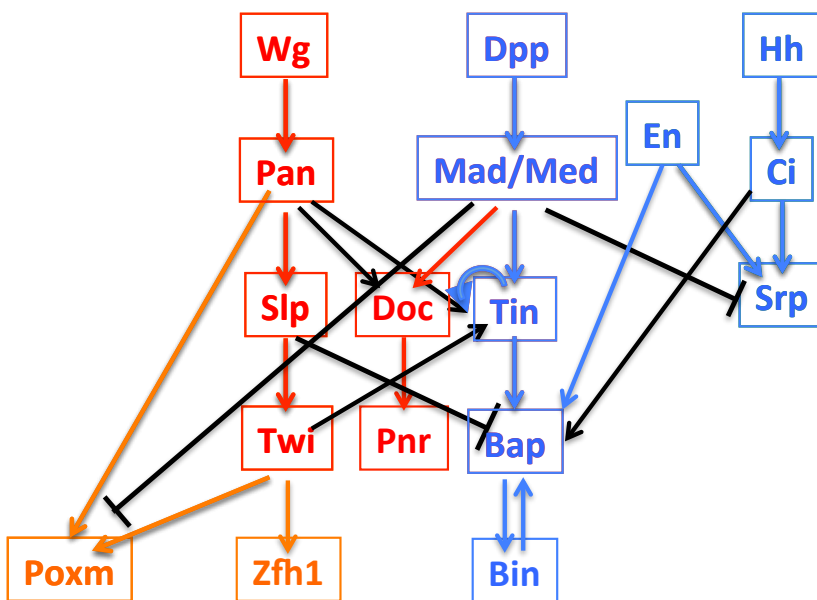


figure 5. Il comporte 17 nœuds majeurs (les intermédiaires implicites ont été ignorés, il y a 45 nœuds en tout) et 25 interactions (les flèches). Jusqu'à 5 (6) interactions vers Bap, 4 vers Tin et 3 vers Poxm et Srp. Ces interactions mobilisent autant d'activateurs transcriptionnels distaux. Les cibles de Pan sont au nombre de 4 (8), celles de

Mad 4 (8) et celles de Ci 2 (3).

**Figure 5 : La circuiterie dans les segments mésodermiques.** Noter la proximité entre les FTs de voies de signalisation et les gènes maîtres.

Les entrées dans ce réseau sont les trois morphogènes Hh, Wg, Dpp. Il faut donc que le ou les signaux soient reçus par les cellules et transférés jusqu'aux FTs des voies, Ci, Pan, Mad/Med.

Les FTs des voies ont presque directement des gènes maîtres d'organes ou de tissus pour cibles : Srp (fat body), Poxm (muscles somatiques), Pnr (cœur) et Bap, Tin et Bin (muscles viscéraux).

L'activation de ces gènes maîtres nécessite presque toujours deux voire trois FTs. Cela protège les cellules d'activation inappropriées. Ceux-ci peuvent dépendre des voies de signalisation ou d'un FT des cellules qui est activé indépendamment de ces voies (illustré ici par En). C'est ce qui peut être appelé une condition endogène. Cela fait que les cellules ne répondent pas de la même manière à un même signal.

Ces gènes maîtres contrôlent une ou plusieurs autres cascades de FTs (non représenté sur le graphe) comme on l'a vu avec Pax6 (Figure 22 de l'opuscule). Dès que ces cibles sont activées le futur des cellules est engagé.

On observe plusieurs motifs de réseaux : feedforward loop, double boucle de rétroaction (double feedback loop), autorégulation positive (voir annexe 3).

Et voici les réponses aux trois questions que nous nous posions :

Comment si peu de voies de signalisation peuvent-elles contrôler autant d'événements différents ? Eh bien ces voies se combinent entre elles, ce qui multiplie leurs cibles (Hh, Wg, Dpp, + Hh et Dpp, Wg et Dpp etc.).

Comme à une voie de signalisation ne correspond qu'un FT qui lui est propre (Pan, Mad, Ci), comment un seul FT peut-il activer autant de cascades différentes ? Eh bien il y a les combinaisons des FT des voies de signalisation entre elles auxquelles s'ajoutent des FT qui sont spécifiques de la cellule qui reçoit le ou les signaux. C'est ce qu'on a appelé le facteur endogène (Ci + Mad/Med + En pour Bap, par exemple). Cette condition endogène augmente considérablement le nombre d'interprétations d'un signal ou d'une combinaison de signaux (Ci + En = fat body, Ci + XX = y, Ci + WW = Z etc. Et ici on voit même Ci + En + Mad/Med = Bap).

Et pourquoi si peu de voies de signalisation ?





Eh bien parce qu'il est infiniment plus simple d'utiliser ces deux combinatoires que de créer une nouvelle voie. Créer une nouvelle voie est très difficile. Songez-y : beaucoup d'éléments internes, un FT masqué, un récepteur et un signal qui doit être exprimé par un autre groupe de cellules. Alors pourquoi ne pas plus simplement utiliser les voies existantes et coopter un FT ou coopter le FT d'une deuxième voie de signalisation. Et de nouveau avec l'évolution c'est toujours le plus simple qui se produit.

Enfin une dernière remarque. Il faut distinguer ici deux niveaux auxquels correspondent deux dynamiques. Il y a celle interne, du réseau transcriptionnel qui gère la dynamique des opérations cellulaires d'un territoire et d'un seul. Elle ne concerne à chaque fois qu'une partie du schéma et correspond aux quatre diagrammes des figures 3 et 4. Il y a celle externe, des interactions cellulaires, les signaux Wg, Hh et Dpp émis par l'ectoderme et reçus par le mésoderme. Ce sont ces interactions entre cellules qui créent les quatre territoires et dictent leur géométrie.

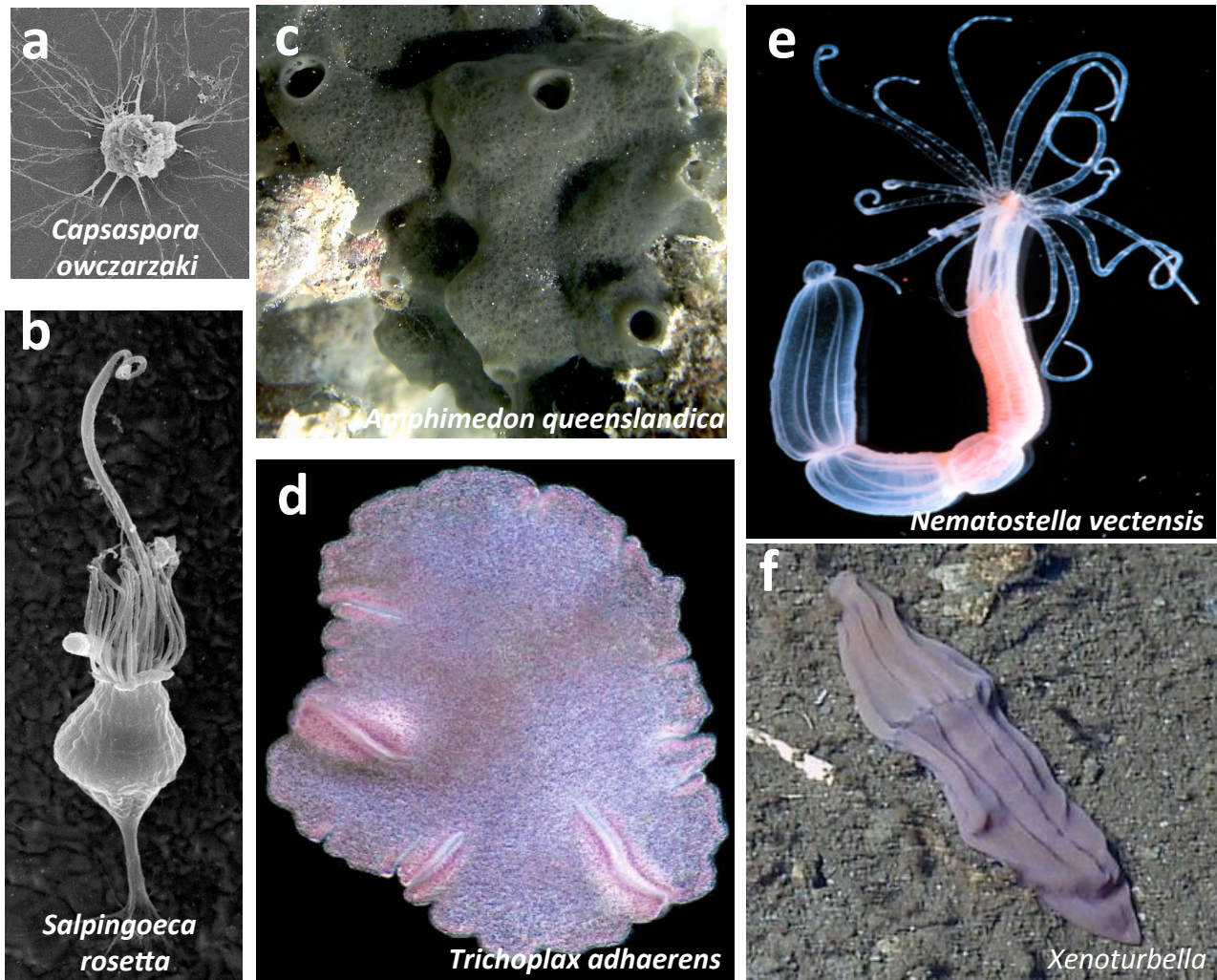
## ANNEXE 5

### UNE BRÈVE COMPARAISON DES GÉNOMES DES NÉPHROZOAIRE AUX CHOANOFLAGELLÉS.

Il s'agit ici de tracer l'origine et la fonction ancestrale des éléments du développement décrits dans l'opuscule : 1) différenciation cellulaire, 2) facteurs de transcription/circuiteries génétique/activation distale, 3) interactions cellulaire/voies de signalisation, 4) comportements cellulaires, et 5) opérations du développement.

La voie royale pour répondre à ces questions est l'examen des répertoires des gènes et de leurs éléments de régulations chez des représentants actuels de nos ancêtres de plus en plus lointains (Figure 1). Les répertoires des gènes sont donnés par le séquençage des génomes. Les éléments de régulations peuvent être détectés par les techniques de ChIP-seq (chromatin immunoprecipitation) et d'ATAC-seq (transposase accessible chromatin sequencing).

Les réponses indiquent si ces éléments sont des innovations ; s'ils émanent de créations de nouveaux gènes, du réarrangement d'anciens gènes par nouvelles combinaisons d'anciens domaines ou de cooptations de gènes existants ; s'ils sont apparus plusieurs fois indépendamment. Elles doivent permettre de saisir comment l'invention des développements s'est faite.



**Figure 1 : Quel héritage avons-nous reçu :** a, b) des derniers ancêtres protistes unicellulaires, a) des filastériens, le groupe sœur des choanoflagellés (génomme de *Capsaspora owczarzaki*<sup>46</sup> et b) des choanoflagellés (génomme de, *Salpingoeca rosetta*<sup>47</sup>, et *Monosiga brevicollis*<sup>48</sup> ; c-f) des derniers ancêtres multicellulaires simples, c) un porifère (génomme d'*Amphimedon queenslandica*, une éponge<sup>49</sup>) ; d) de *Trichoplax adhaerens*, un placozoaire<sup>5</sup> ; e) du dernier ancêtre présentant un développement effectuant la gastrulation mais toujours sans véritables organes, un cnidaire (génomme de *Nematostella vectensis*<sup>50</sup>, une anémone de mer) ; et enfin f) des bilatériens basaux sans organes (génomme de *Xenoturbella*,<sup>4 6</sup> un Xenacoelomorpha). Pour leur position dans l'arbre du vivant voir la Figure de l'opuscule.

Les résultats sont surprenants.

L'origine des gènes impliqués dans la différenciation cellulaire et dans les comportements cellulaires, utilisés par les opérations du développement, remonte presque toujours à notre dernier ancêtre unicellulaire.

Il n'a fallu que quelques gènes nouveaux pour faire les organismes. Mais, ces gènes ont fait émerger une nouvelle dimension pour l'embryon, celle de l'organisation spatiale des cellules et un nouveau niveau pour l'évolution, celui des réseaux d'interactions génétique. C'est surtout l'évolution de ces réseaux qui rend compte de la complexification des développements et des organismes.

### 1) Origine de la différenciation cellulaire

Commençons par l'activation des gènes par les facteurs de transcription. Nous savons que nous devons remonter loin puisque les placozoaires comme *Trichoplax adhaerens* ont déjà une douzaine de types cellulaires. De fait leur génome présente un grand nombre de facteurs de transcription qui régulent la spécification de leurs types cellulaires. Cela inclut des FTs spécifiques des neurones (domaine homéo-LIM, domaine HLH), des muscles (un autre domaine HLH), des cellules endocrines (domaine « pair of POU ») et des cellules sanguines (domaine GATA doigt de zinc)<sup>5 51</sup>.

Qu'en est-il alors des éponges ?

Eh bien elles présentent aussi plusieurs types cellulaires et possèdent aussi un riche répertoire de FTs dont certains spécifient les muscles et les neurones (domaines ou FTs PaxB, Lhx, SoxB, Msx, Mef2, Irx et bHLH)<sup>49</sup>.

Alors il nous faut examiner les unicellulaires !<sup>24,52</sup>.

Chez les unicellulaires les plus proches des animaux on dénombre une centaine de FTs (pour environ 10 000 gènes), dont beaucoup sont retrouvés chez les animaux (domaines ou FTs homeo, T, TALEs et non-TALEs, MYC, Brachyury, CSL, RUNX, STAT, NF-κB, p53, Sox/TCF, MEF2, RFX etc.). En outre, le MYC de *Monosiga brevicollis* interagit avec les mêmes partenaires (comme bHLH ou MAX) et utilise la même séquence de liaison à l'ADN (les boîtes E) que les métazoaires. Enfin, chez *Capsaspora owczarzaki* les orthologues de Brachyury and MYC sont impliqués dans le contrôle de la motilité et de la prolifération des fonctions similaires à celles de leurs orthologues métazoaires<sup>24</sup>. Les premières traces de la



différenciation cellulaire, les FTs, sont donc déjà chez nos ancêtres protistes unicellulaires.

Mais à quoi ces FTs leur servent-ils ?

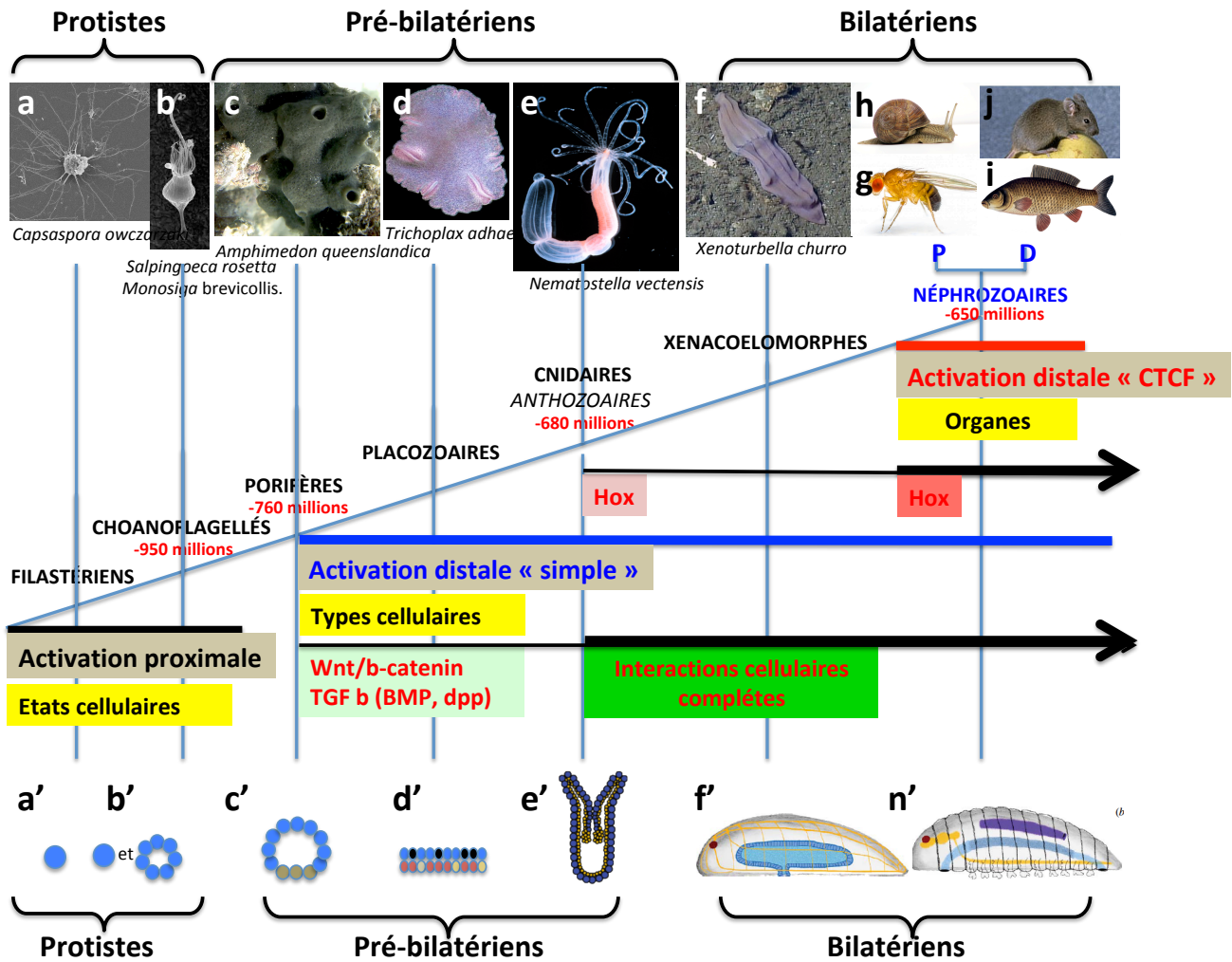
Ces unicellulaires ont des styles de vie qui comportent plusieurs états cellulaires. Ainsi les choanoflagellés passent par des stades de cellules isolées nageantes ou fixées et des stades coloniaux où les cellules sont reliées par des ponts intercellulaires. Ils présentent aussi un cycle sexuel (haploïde, diploïde) avec méiose, division et fusion des cellules haploïdes (sans gamètes). Autant d'états cellulaires différents qui se succèdent temporellement.

Il a été montré qu'à chacun de ces états correspondent des assortiments d'ARN spécifiques et que des FTs activent certains gènes (comme par exemple les septines impliquées dans les jonctions entre cellules au stade colonial), et en répriment d'autres. Ceci vaut pour les choanoflagellés comme pour les représentants du groupe voisin (*Capsaspora owczarzaki*)<sup>24</sup>.

Chez *Capsaspora owczarzaki* il a de plus été montré que nombre de gènes sont contrôlés par l'action de FTs sur des sites de liaison à l'ADN proche du promoteur.

Il est donc clair que les premières traces de la différenciation cellulaire des animaux sont les états cellulaires des protistes. La fonction ancestrale des FTs et de l'activation transcriptionnelle, fut de produire des états cellulaires différents chez les protistes (Figure 2, la ligne noire).

La deuxième condition à la différenciation cellulaire, c'est un processus qui crée de l'hétérogénéité entre cellules. Il n'y a pas de tel processus connu chez les unicellulaires. Les signaux qui déclenchent les changements d'états chez les protistes sont externes. Ils passent souvent par des récepteurs membranaires que l'on ne retrouve pas chez les animaux. Le processus de différenciation cellulaire s'est donc constitué en au moins deux étapes, celle des états cellulaires impliquant les FTs, chez les unicellulaires, et celle de l'hétérogénéité entre cellules, plus tardive.



**Figure 40 : Position dans l’arbre du vivant des innovations discutées dans ce livre.** Quelques-uns de nos ancêtres. Leurs représentants actuels sont figurés en haut depuis les protistes (filastériens et choanoflagellés, a et b) jusqu’aux bilatériens complexes, les néphrozoaires (P, protostomiens g et h et D, deutérostomiens, i et j). La ligne bleue céruleen oblique figure le temps ; au dessus de cette ligne sont indiquées les dates d’apparition de quelque uns des groupes d’animaux mentionnés. Les boîtes jaunes indiquent l’existence d’états cellulaires (chez les protistes), de types cellulaires différenciés (chez les éponges) et celle d’organes (chez les néphrozoaires). Les apparitions de l’activation proximale, l’activation distale simple et l’activation distale CTCF sont notées. L’origine des voies de signalisation, les boîtes vertes et du complexe génétique Hox, les boîtes rouges, est indiquée. (voir aussi les figures 7, 28 et 35 de l’opuscule). En bas, sont représentés les degrés de complexité cellulaire atteint par chacun de ces ancêtres. a’ et b’ unicellulaires et phase d’agrégation ; pour c’ et d’ multicellulaires et types différenciés chez les éponges et les placozoaires ; pour e’ et f’ cavité digestive chez les hydres et anémones de mer et chez les bilatériens xenacoelomorphes; pour les bilatériens protostomiens et deutérostomiens (néphrozoaires), n’, tube digestif à deux ouvertures, des territoires répétés, des organes.



Par contre ces signaux externes sont transmis au noyau par des voies de signalisation dont des éléments se retrouvent chez les animaux. De fait, plusieurs des tyrosines kinases des unicellulaires se retrouvent dans beaucoup de voies de signalisation des animaux.

Une hypothèse plausible est qu'aux signaux externes qui déclenchent chez les unicellulaires des changements d'états se soient substitués chez les multicellulaires des signaux intercellulaires qui déclenchent la différenciation.

La différenciation cellulaire serait alors le changement de récepteurs membranaires. Ce changement aurait transformé un processus temporel chez les unicellulaires (la succession des états cellulaires) en un processus spatial chez les multicellulaires.

## **2) Origine des réseaux de régulation : dès les unicellulaires ?**

Dès lors que des FTs sont actifs un réseau de régulation émerge. Il concerne à minima ces FTs et leurs cibles. C'est lorsque des FT contrôlent d'autres FTs que les réseaux deviennent vraiment complexes.

Chez *Capsaspora* les gènes qui codent des FTs montrent plus de sites de régulation que les autres. Et ces sites de régulation sont enrichis pour les autres FTs. Il y a donc chez *Capsaspora* des interactions FT-FT et un réseau de régulation hiérarchisé complexe. Les gènes à boîtes T, bHLH et bZIP ont le plus grand nombre de sites de régulation.

Certains de ces arrangements de ce réseau ont-ils été hérités par les animaux ? La réponse ne peut bien sûr n'être que très spéculative. Voici ce que l'on en sait. 63 des gènes cible de *Capsaspora*-Bra sont des orthologues de gènes aussi contrôlés par Bra chez la souris. Ce sont des gènes impliqués dans les fonctions de mobilité amiboïdale et dans la constitution du cytosquelette d'actine. Ceci suggère qu'il y avait un réseau de régulation contrôlé par Brachyury déjà présent chez les pré-métazoaires et impliqué dans la migration cellulaire, une fonction cellulaire utilisée plus tard pour la gastrulation. Ce réseau pourrait être à l'origine de celui qui contrôle certaines des opérations cellulaires du développement. La même chose

s'applique à Myc qui contrôle la prolifération cellulaire<sup>24</sup> là aussi un processus très important pour le développement.

Des réseaux de régulation complexes existent dès chez les unicellulaires. Des éléments des développements, et des circuits contrôles de la motilité et de la prolifération cellulaire sont déjà en partie constitués chez notre ancêtre unicellulaire et les premiers animaux multicellulaires en ont peut-être hérité. Cependant la masse des éléments des développements est due aux animaux.

### **3) Origine des interactions cellulaire et des voies de signalisation**

Les communications entre cellules sont centrales pour le développement et la différenciation cellulaire (l'opuscule).

Chaque voie de signalisation est une innovation dont l'apparition n'est en rien aidée par les voies de signalisation existantes. C'est donc l'origine des communications entre cellules effectivement utilisées dans la différenciation cellulaire et les développements qui nous intéresse ici.

Actuellement on ne peut pas dire quelles voies de signalisation furent associées aux premiers types cellulaires. Je renvoie au paragraphe 1 pour la discussion de ce problème.

Pour les développements les voies Wnt et BMP sont utilisées pour la première opération majeure et universelle, la gastrulation. Les voies Hedgehog (Hh), Notch/Delta, et celles dérivées des Récepteur Tyrosine Kinase (RTK) sont utilisées ensuite<sup>53</sup>.

Analysons donc les animaux qui présentent des états ou des types cellulaires avant leur constitution en organes : protistes, éponges, placozoaires, cnidaires.

Chez les protistes on détecte certains des éléments de ces voies mais aucune complètes, en particulier leurs ligands, leurs récepteurs et leurs FTs associés sont absents, à l'exception de Hh. Ceux-ci sont tous des innovations des métazoaires.



Chez les éponges comme *Amphimedon queenslandica* on trouve des éléments des voies de signalisation, Wnt, TGF- $\beta$ , Notch et Hedgehog et leurs ligands et récepteurs. Il est probable que ces voies de signalisation se sont constituées à partir de celles des unicellulaires en engageant de nouveaux récepteurs et ligands. Mais ces derniers ne correspondent pas aux familles des autres animaux<sup>49</sup>.

Si l'on regarde le génome de *Trichoplax*<sup>5</sup> on trouve des voies de signalisation complètes Wnt/ $\beta$ -catenin et TGF- $\beta$ . Les voies de signalisation Hh, Notch et JAK/STAT sont incomplètes. Il est possible que ces gènes aient été perdus dans le lignage des placozoaires.

C'est chez les cnidaires comme *Nematostella vectensis* que l'on trouve tous les éléments des voies Wnt et TGF $\beta$  ligands et récepteurs inclus (Wnt, secreted frizzled-related protéine (SFRP), Dpp/BMP, activin, and chordin). Ces ligands et les récepteurs furent ensuite conservés chez les bilatériens. Ce sont donc peut-être des innovations des cnidaires<sup>50</sup>.

Il est remarquable que ce soit précisément chez les cnidaires qu'apparaît la gastrulation qui nécessite des comportements et des opérations cellulaires intégrés (voir le paragraphe 4).

En résumé les voies de signalisation utilisées dans les développements sont absentes ou incomplètes chez les protistes. Elles apparaissent complètes dès les plus simples des animaux (Figure 2). Elles comportent alors des gènes nouveaux et des gènes réarrangés correspondant à des assemblages d'anciens domaines présents chez les protistes. Les voies de signalisation des animaux se sont probablement constituées en engageant de nouveaux récepteurs et ligands à des voies de signalisation qui chez les protistes servaient à la communication avec le milieu extérieur.

On ne sait pas si les premières voies de signalisation utilisées dans la différenciation cellulaire furent celles utilisées par les premiers développements.

#### **4) Origine du développement**

Qu'en est-il du développement<sup>54</sup> ?

Chez tous les animaux le développement débute par la fusion de deux cellules haploïdes, se poursuit par des divisions cellulaires puis par la formation d'une cavité digestive (la gastrulation).

La méiose et l'alternance d'un état diploïde et d'un état haploïde sont des inventions qui prédatent les protistes. Les animaux les doivent aux choanoflagellés.

La première invention majeure du développement fut la gastrulation qui crée une cavité digestive. Rappelez-vous le film sur l'amphioxus où l'on voit l'un des hémisphères de l'embryon s'enfoncer dans l'autre pour former une gastrula simple (Figure 1 de l'opuscule). Cela augmenta l'efficacité de la digestion, (avec peut-être comme intermédiaire un stade ressemblant aux placozoaires avec les différents types cellulaires dont ceux dédiés à la digestion et à l'excrétion sur une face ventrale). De plus, la gastrulation met en contact deux feuillets de cellules ce qui ouvre l'opportunité d'échanges de signaux.

L'organisation des cnidaires (les hydres et les méduses) correspond à un tube ouvert, c'est-à-dire à une gastrula étirée. L'axe de l'animal (dit oral-aboral) est celui de l'embryon (dit animal-végétal). L'ouverture orale unique sert à la fois à la capture de la nourriture et à l'évacuation des résidus de la digestion. Cette organisation est juste un peu plus complexe que celle des placozoaires (Figure 2, les figurés sous la ligne oblique qui représente le temps).

La gastrulation des cnidaires fait intervenir la voie de signalisation Wnt/ $\beta$ -caténine (qui donc crée de l'hétérogénéité cellulaire). L'expression des gènes Wnt marque l'ouverture orale chez les cnidaires. La suppression de cette voie se traduit par l'absence de la gastrulation et une expression ectopique par des ouvertures orales additionnelles (discuté dans <sup>54</sup>). Wnt fonctionne dès les cnidaires comme un morphogène.

Il est extraordinaire de constater que, des cnidaires à l'amphioxus et aux vertébrés et même probablement aux insectes et autres protostomiens, ce



même morphogène est impliqué dans leur gastrulation. Par exemple chez les bilatériens l'expression de Wnt marque le blastopore et son expression ectopique se traduit par un doublement de l'embryon. Et de fait, le répertoire des gènes Wnt est incroyablement conservé des cnidaires aux bilatériens avec 12 des 13 orthologues bilatériens présents chez *Nematostella vectensis*.

D'autre part, l'autre voie de signalisation qui, chez les bilatériens, a un rôle majeur dans l'organisation des structures selon l'axe dorsal-ventral, Dpp (exprimé en un secteur longitudinal dans la larve de la drosophile, Fig. 30 de l'opuscule) est aussi retrouvée chez les cnidaires les plus proches des bilatériens, les anthozoaires (comme *Nematostella*). Elle intervient sous le nom de BMP dans l'agencement des structures du feuillet interne (les plis endodermiques ou mésenteries).

Les voies de signalisation Wnt/ $\beta$ -caténine et TGF- $\beta$  (=Dpp ou BMP) sont donc déjà présentes chez nos ancêtres animaux parmi les plus anciens et les plus simples. Elles y jouent déjà un rôle dans le développement qui semble bien être celui sur lequel les développements plus complexes ont ensuite été élaborés.

Nous avons vu (paragraphe 3) que l'on trouve des traces de ces voies de signalisation chez nos ancêtres encore plus lointains, chez les éponges. Leur expression suggère qu'ils sont impliqués dans la définition d'un territoire « arrière ». Le rôle le plus ancestral que l'on puisse attribuer à la voie Wnt est donc celui de « marquer » un arrière<sup>55 56</sup>.

Après la gastrulation, l'une des étapes suivantes des développements fut l'émergence d'un troisième feuillet, le mésoderme. Chez les éponges et les cnidaires ce feuillet est absent que ce soit morphologiquement ou au niveau moléculaire (absence des gènes majeurs du mésoderme comme Fkh, Gsc, Twist, Snail). Il apparaît chez les xenacoelomorphes<sup>57</sup>. Puis avec les néphrozoaires on assiste à une explosion des développements complexes. Cette dernière étape ne correspond pas à l'apparition des bilatériens puisque les xenacoelomorphes ont un mésoderme mais encore un tube digestif à un seul orifice comme les cnidaires. L'émergence de la bilatéralité n'a pas été suffisante en soi pour sous-tendre l'explosion des néphrozoaires.

Notons enfin que chez les vertébrés comme chez les insectes et autres protostomiens ces voies de signalisation Wnt/ $\beta$ -caténine et BMP sont utilisées au-delà de la gastrulation encore et encore dans d'autres processus. Nous les avons rencontrées plusieurs fois dans l'opuscule notamment pour le positionnement des organes (Wnt/Wg est exprimée dans des bandes transverses dans la larve de la drosophile, BMP dans des bandes longitudinales, l'opuscule, Figure 30 de l'opuscule). La cooptation encore et toujours.

En conclusion la constitution des voies de signalisation complètes Wnt et BMP rendent compte de l'émergence des développements. Dès les éponges Wnt définit un territoire « arrière ». Chez les cnidaires Wnt définit l'hémisphère de l'embryon qui s'enfonce dans l'autre pour former une gastrula et BMP dans l'agencement des structures du feuillet interne. Les étapes suivantes se sont faites sur ces bases puisque depuis les cnidaires jusqu'aux bilatériens on retrouve ces molécules dans les mêmes processus.

Les plus simples et les plus anciens des bilatériens, les xenacoelomorphes, sont comme les cnidaires des sortes de gastrula étirées (Figure 2, les figurés sous la ligne oblique qui représente le temps). Ces phylums comportent peu d'espèces comparées aux néphrozoaires. Les néphrozoaires dérivent tous d'un unique ancêtre commun qui sans doute ressemblait à un xenacoelomorphes, sans organes, mais avec un dos, un ventre, un avant, un arrière, un peu comme les cnidaires. Qu'a-t-il fallu de plus pour qu'explorent avec les néphrozoaires les développements ? Une nouvelle voie de signalisation ? Une nouvelle famille de FTs ? Une émergence dans les arrangements de la circuiterie génétique ? Ceci est brièvement discuté en conclusion de cette annexe.

## 5) Origine des comportements cellulaires

Nous discutons ici des comportements (propriétés) cellulaires qui sont impliqués dans les opérations du développement. Nous en avons fait un listing partiel. Ils incluent : la mobilité ; les mouvements (dispersif, directionnel, vitesse) ; des changements de forme ; l'adhésivité des cellules entre elles ou à une matrice ; les divisions, leur orientation (orientées, asymétriques, inégales, ou non), leur vitesse ; etc. A chacun correspond



souvent une protéine majeure, actine, myosine, tubuline, cadhérine (l'adhésivité cellule-cellule), intégrine (l'adhésivité cellule matrice).

On peut souvent discerner un processus subcellulaire associé comme, la polymérisation de l'actine, la contractilité de l'actomyosine, les polarités apicobasale, planaire ou radiale, l'adhésion des cellules entre elles ou à une matrice etc. Les combinaisons de divers modes d'adhésions avec des réseaux distincts d'actomyosine diversifient les comportements cellulaires.

Tout cela se traduit souvent par une transition de forme des cellules, d'épithéliale à mésenchymale ou l'inverse avec des intermédiaires.

Les systèmes moléculaires ont une organisation modulaire qui sous-tend les comportements cellulaires<sup>58</sup>.

Eh bien toutes ces molécules, actine, myosine, tubuline, cadhérine, intégrine, étaient déjà présentes chez notre ancêtre unicellulaire comme l'étaient les processus subcellulaires associés comme, la polymérisation de l'actine, la contractilité de l'actomyosine, les polarités apicobasale, planaire ou radiale, l'adhésion des cellules entre elles ou à une matrice.

## ANNEXE 6

### LES LEGS DE NOTRE ANCÊTRE UNICELLULAIRE

Notre dernier ancêtre unicellulaire c'est l'ancêtre commun aux choanoflagellés et aux bilatériens d'il y a presque 1 milliard d'années. C'était un protiste (ou encore un infusoire, un animalcule, un protozoaire, tous ces noms sont équivalents), unicellulaire. Certaines de ses caractéristiques ont rendu possible le passage aux animaux et nul doute aussi qu'elles ont aussi exercé des contraintes sur la forme et les fonctions de ces premiers animaux. Elles ont limité « le jeu des possibles » (François Jacob).

Un mot sur sa morphologie puisque nous y recherchons les fonctions ancestrales des gènes utilisées actuellement dans le développement des animaux.

Ces protistes sont polarisés. Ils nagent grâce à un flagelle entouré d'une collerette composée de microvilli (microvillosités). Ce flagelle sert aussi à créer des courants pour attrapper les bactéries dans les microvilli et les phagocyter. Antérieurement ils présentent des filopodes qui servent à la fixation sur un substratum (Figure 1, annexe 5).

Selon les conditions du milieu ils modulent leurs propriétés. Leur cycle de vie inclut des stades de cellule isolée, nageante ou fixée et des stades coloniaux où les cellules sont reliées par des ponts intercellulaires. Ils présentent aussi un cycle sexuel avec des stades méiose (commun chez quasiment tous les protistes<sup>59</sup>), division des cellules haploïdes et fusion entre elles. Il n'y a pas de gamètes.

La liste non exhaustive des principaux gènes correspondant à ces structures et caractéristiques est la suivante :

Au flagelle correspondent des microtubules (protéine, la microtubuline) arrangés et une dynéine un complexe protéique qui fait battre le cil en

faisant glisser les microtubules entre eux. Les microvilli et filopodes, correspondent à des protrusions d'actine. Les interactions avec les bactéries, l'adhésion entre eux des microvilli de la collerette et sans doute d'autres interactions avec l'environnement extracellulaire sont réalisées par les cadhérines et les intégrines alpha. L'attachement au substrat se fait par 5 domaines « collagène », et des domaines associés au collagène, un domaine laminine, le domaine reeler, le domaine épendyme (chez *Monosiga brevicollis*). La reconnaissance des proies et la détection des changements dans l'environnement se font par des voies de signalisation tyrosine kinase (TK)<sup>52</sup>. Le génome de *Monosiga brevicollis* contient 120 gènes contenant un domaine TK, 30 contenant un domaine pTyr phosphatase et 80 un domaine SH2. La signalisation tyrosine kinases nécessite ces trois éléments.

Le passage aux stades coloniaux, la formation des rosettes implique un mode de division polarisé, une matrice extracellulaire sur laquelle les cellules sont attachées et une lectine de type C.

Le cycle méiotique implique plusieurs dizaines de gènes spécifiques (voir <sup>59</sup>).

Les correspondances de ces structures et de ces gènes chez les animaux sont les suivantes.

Les gènes spécifiques de la méiose se retrouvent chez tous chez les animaux.

Les éléments de la formation des rosettes, matrice extracellulaire, C-lectin, filopodes se retrouvent chez les blastula des éponges, des cnidaires et même de l'amphioxus (références dans <sup>60</sup>) dont elles assurent la cohésion des cellules.

La microtubuline et les flagelles se retrouvent inchangées chez les animaux dans toutes sortes d'épithélium, ce sont les cils du pôle apical de la cellule, le cil primaire sensoriel, le cil vibratil et le cil nodal. Les tyrosine-kinases (reconnaissance des proies chez les choanoflagellés) sont des composants de cascades de signalisation impliquées dans la communication

entre cellules. Ces voies de signalisation qui détectaient des molécules de l'environnement ont été cooptées pour détecter des signaux entre les cellules. Quatre des neuf voies de signalisation impliquées dans le développement<sup>53</sup>, font partie de la famille des TK (JAK/STAT, EGF, FGF et VEGF). L'actine, les molécules des microvilli et des filopodes, et leurs molécules associées se retrouvent dans toutes sortes de cellules ce sont les bordures en brosse des épithéliums. Les cadhérines et les intégrines alpha sont prépondérantes dans l'adhésion des cellules entre elles. Enfin les collagènes et molécules associés forment les matrices extracellulaires ou encore les membranes basales chez les animaux.

Et nous avons vu que ces molécules sont les mêmes chez les protistes et tous les animaux depuis les éponges et les méduses.

Outre les correspondances notées ici, comme nous l'avons vu aussi, beaucoup de ces molécules sont aussi à la base de comportements cellulaires variés notamment de motilité et de communication entre cellules si importants dans les opérations du développement.

L'héritage de notre ancêtre unicellulaire commun avec les choanoflagellés est immense !

Pour rappel (voir l'annexe 5) :

1) Sur la centaine de TFs (sur 10000 gènes) que les protistes présentent, un nombre significatif est partagé avec les animaux (voir le paragraphe 2). Là encore l'héritage est immense.

Rappelons par contre que si l'activation proximale des gènes par les FTs est en place chez nos ancêtres unicellulaires, l'activation à distance ne l'est pas. Les éponges et Trichoplax ne fonctionnent aussi qu'avec l'activation proximale<sup>51</sup>. Il faut attendre les cténophores et les cnidaires pour l'activation à distance<sup>24</sup> et ce n'est qu'à la base des néphrozoaires qu'a émergé l'activation distale basée sur CTCF (voir<sup>24,46</sup>).

2) Il n'y a pas de preuve qu'aucune des voies de signalisation des animaux soit complète chez les choanoflagellés. Aucun récepteur ni ligand des voies Wnt, TGF- $\beta$  ou NHR (nuclear hormone receptor) correspondant à



ceux des animaux n'est présent<sup>47</sup>. C'est chez les éponges que l'on trouve les voies Wnt (la bêta-catenine) et TGF- $\beta$ . Les ligands Wnt reliés à ceux des bilatériens apparaissent avec les cnidaires.

De même la voie Notch est incomplète. Mais les domaines protéiques des récepteurs Notch (EGF, NL et ANK, ankyrin repeats) sont présents bien que dans différents gènes. C'est de nouveau chez les cnidaires qu'on les trouve assemblés et c'est le récepteur qu'utilisent les animaux. Plus généralement les arrangements entre domaines protéiques différent souvent entre les choanoflagellés et les métazoaires. De nombreux réarrangements de domaines protéiques ont eu lieu au début de l'évolution des métazoaires.

En conclusion : bien que quelques éléments des voies de signalisation soient présents chez l'ancêtre unicellulaire, leur récepteurs et ligands évoluèrent après la divergence (références dans <sup>48</sup>).

3) De même pas de gènes Hox ; ceux-ci apparaissent chez les cnidaires<sup>61</sup>  
<sup>62</sup>.

4) Enfin grosse surprise plusieurs protéines des synapses des neurones tirent leur origine des choanoflagellés. Les canaux calciques voltage-dépendants, les canaux sodiques voltage-dépendants (important pour générer des potentiels d'action chez les cellules animales) et plusieurs protéines que les cellules nerveuses des animaux utilisent pour attacher les neurotransmetteurs à la membrane post synaptique (les postsynaptic scaffolding proteins comme PSD-95), sont présentes chez *M. brevicollis*. De plus un ensemble de protéines de l'appareil neurosécrétoire présynaptique appelé SNARE interagissent chez *M. brevicollis* de la même façon que chez les vertébrés. Ces protéines sont localisées à la membrane apicale (du côté de la collerette) et sont impliquées dans le relargage de molécules de vésicules. On ne connaît pas le rôle de ces canaux ioniques (références dans <sup>63</sup>).

## Conclusion

Les cycles méiotique, diploïde haploïde, et le cycle de vie où ils forment des rosettes de cellules accolées les unes aux autres après division de notre ancêtre unicellulaire, furent sans doute le point d'ancrage des développements des animaux.

La sorte de blastula qui correspond aux rosettes a vraisemblablement permis l'émergence de la diversification des cellules et des opérations du développement par les voies de signalisation. L'apparition de territoires non dédiés à l'ingestion des bactéries et l'internalisation des cellules s'en est probablement suivie. Cela mène aux cnidaires qui correspondent à des gastrula étirées. La suite découle de cette structure initiale que l'invention des communications entre cellules et l'émergence de quelques accélérateurs de l'évolution à la base des cnidaires, et à la base des néphrozoaires, entre autres, ont complexifié.

Nous dérivons des choanoflagellés d'il y a 1 milliard d'années car ils réunissaient les conditions pour que ces étapes puissent se faire.

## ANNEXE 7

**L'ORIGINE DE L'ŒIL OU LA NAISSANCE PAR CHANCE D'UNE FONCTION.**

La vue dépend des rhodopsines, des protéines associées au rétinol qui détecte la lumière. Elles convertissent l'énergie lumineuse en signal neurochimique. L'œil du représentant actuel de l'ancêtre de tous les chordés, l'amphioxus, consiste en un calice pigmentaire et un groupe de neurones photorécepteurs. Bien que beaucoup plus simple que les yeux des vertébrés il utilise les mêmes rhodopsines, les opsines-c. Tous les chordés (vertébrés, deutérostomiens) utilisent ces opsines-c pour la vue. Les insectes, plus généralement les protostomiens, ont développé leurs yeux à partir des opsines-r. C'est chez l'ancêtre commun aux bilatériens que le gène codant une opsine s'est dupliqué et a donné les opsines de type c et les opsines de type r. Chez le groupe ancestral aux bilatériens, les cnidaires, on trouve des cellules photoréceptrices qui détectent la lumière grâce au rétinol et à des opsines voisines du type c. Ils provoquent le battement du cil de la larve. L'opsine à l'origine des opsines des cnidaires et des types c et r est l'origine moléculaire de cet organe. Comment une molécule peut être à l'origine d'un organe a été expliqué dans l'opuscule (Figures 25 et 26). En dehors des animaux certaines algues vertes sont phototactiques. Leur chloroplaste abrite des rhodopsines qui détectent la lumière via un rétinol et déclenchent un signal de dépolarisation de la membrane qui se propage à un flagelle qui se met à battre. Ces rhodopsines ont la même structure que celle des animaux. Cette évolution parallèle à celle des animaux n'a pas eu chez les végétaux le succès qu'elle a eu chez les animaux. Elle disparaît chez les plantes. Les chloroplastes dérivent des cyanobactéries. Les cyanobactéries phototactiques hébergent des rhodopsines photosensibles pompes à sodium qui leur servent à faire battre un flagelle. Les autres rhodopsines bactériennes et archéales captent des photons par un rétinol et convertissent cette énergie en énergie chimique grâce aux ATP synthases. Cette phototrophie a précédé la photosynthèse.

**Le scénario de la naissance d'une fonction.**

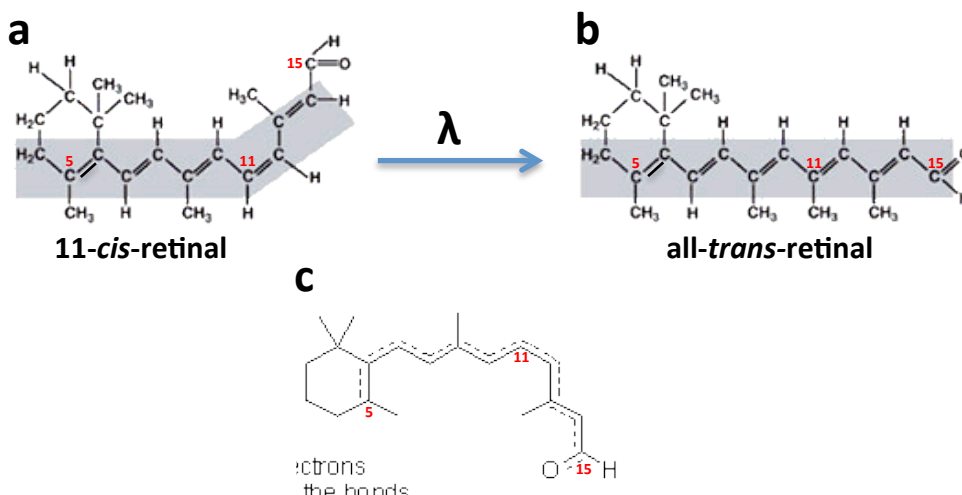
L'histoire des rhodopsines nous fait remonter jusqu'aux bactéries primitives. Chez ces bactéries la structure de ces protéines - sinon leur séquence - présente des homologues avec nos rhodopsines. Ce sont des pompes à ions, sodium ou protons. Elles servent à la phototrophie qui présente un énorme avantage pour satisfaire les besoins énergétiques. Certaines des rhodopsines deviennent photosensibles lorsqu'une poche pour le rétinol apparaît. C'est l'émergence du phototactisme, de nouveau énorme avantage pour les besoins énergétiques. Ce sont les plus anciennes molécules photosensibles. L'ancêtre unicellulaire des animaux, LECA, a hérité de ces pompes. Une transition entre rhodopsines

bactériennes et eucaryotes a possiblement eu lieu. La partie cytoplasmique de la rhodopsine aurait établi des contacts avec d'autres protéines de la membrane plasmique, les protéines G. devenant l'ancêtre de la famille des GPCRs. Une de ces GPCRs aurait regagné le rétinale devenant l'ancêtre des opsines visuelles à l'origine des opsines c et r des animaux. Les insectes, plus généralement les protostomiens, ont développé leurs yeux à partir des opsines-r. Les vertébrés, plus généralement les deutérostomiens, ont développé leurs yeux à partir des opsines-c. Les opsines-r servent chez les vertébrés une fonction « horloge », exemple de l'émergence à partir d'une même propriété (la capacité à détecter la lumière) de deux fonctions différentes, facilitée par la duplication de gènes. L'ancêtre de la famille des GPCRs a été à l'origine de multiples nouvelles fonctions chez les bilatériens, dont chez les vertébrés les molécules de la détection des odeurs et du goût. Ainsi la vision, l'odorat et le goût ont émergé de la même famille de molécules et sont liés par cette même origine moléculaire. Ce survol de la vision montre comment à partir d'une molécule émerge d'un coup, comme par magie, une fonction comme la phototrophie ou le phototactisme. Et aussi comment l'émergence d'une fonction favorise l'émergence d'autres fonctions qui, de ce fait, se trouvent apparentées. Et encore, comment les fonctions en se combinant font émerger d'autres fonctions.

## La naissance d'une fonction

Prenons comme exemple la vue. C'est clairement un processus complexe qui s'appuie sur un organe complexe, l'œil (Figures 6, 25 et 26 de l'opuscule). Songeons, de plus, à tous les éléments qui aboutissent à la perception visuelle, comme le traitement de l'information par le cerveau. Cependant, bien que complexe, la vue dépend en premier lieu d'une manière de détecter la lumière. Pour l'homme, cette manière c'est une molécule, le rétinale, la vitamine A, associée de manière covalente à des protéines, les rhodopsines (des opsines dites de type c). Les rhodopsines se trouvent dans les membranes des neurones (les cônes et bâtonnets) de notre rétine. Il y en a quatre qui sont un tout petit peu différentes les unes des autres. Le rétinale en absorbant la lumière s'isomérise (Figure 1a et b). Ce changement de forme enclenche une suite de signaux, qui convertit

l'énergie lumineuse en signal neurochimique. Ce processus passe par l'activation par les rhodopsines d'une protéine du cytoplasme, une G-protéine (la transducine).



**Figure 1 : Le rétinale.** a, b) Cette molécule en absorbant la lumière ( $\lambda$ ) change de forme (s'isomérise) : l'énergie du photon est convertie en mouvement ; c) sa capacité à absorber des photons est due à ce que les électrons sont délocalisés du carbone 5 à l'oxygène lié au carbone 15.

Dans ce chapitre nous chercherons à savoir : à quoi ont ressemblé les événements d'où ont émergé ces protéines photosensibles ? Combien d'autres molécules photosensibles qui ne seraient pas des rhodopsines, sont utilisées pour cette fonction chez les animaux ? Ont-elles émergé d'un coup ou bien y-a-t-il eu un passage du non visuel au visuel pour les rhodopsines ? Quelle était leur fonction avant ? Le scénario qui a mené à la

vue est-il direct et logique ? Cette protéine a-t-elle été à l'origine, outre du visuel, d'autres fonctions ?

Pour le savoir il nous faut reconstituer l'histoire de cette molécule, rechercher quand, dans le passé, ont eu lieu les mutations qui ont créé cette fonction. Nous allons donc suivre cette protéine dans l'arbre du vivant<sup>15,64,65</sup>).

### **Chez les animaux**

Tout d'abord on retrouve ces mêmes quatre rhodopsines, des opsines de type c, avec la même fonction dans les mêmes cellules chez les autres mammifères et aussi chez les oiseaux, les grenouilles et les poissons (mais chez ces derniers dans certaines opsines le déhydrorétinal remplace le rétinol, une adaptation à la vision dans l'eau). Il n'y a pas plusieurs molécules de la vision chez les vertébrés, tous utilisent la même. La ressemblance des yeux des vertébrés ne trompe pas. Elle va jusqu'au niveau moléculaire.

Qu'en est-il quand on s'éloigne de 500 millions d'années en arrière et que l'on considère le représentant actuel de l'ancêtre de tous les vertébrés, l'amphioxus, un petit ver marin ? L'amphioxus a plusieurs organes de la vue : un œil frontal et des ocelles dorsales, les organes de Hesse et les cellules de Joseph<sup>66</sup>. L'œil frontal de l'amphioxus est beaucoup plus simple que les yeux des vertébrés. Il consiste en un calice pigmentaire, un groupe de neurones photorécepteurs et d'autres neurones. Eh bien, on détecte dans les neurones photorécepteurs une opsine-c qui est très proche des opsines-c des cônes et bâtonnets des vertébrés. Ce sont donc les mêmes molécules qui sont utilisées pour la vue chez tous les chordés. Or ces neurones photorécepteurs sont connectés au centre de la locomotion de l'amphioxus<sup>67</sup>. Cela nous apprend que le rôle originel de la vue est d'orienter les mouvements de ce petit vers, de le diriger vers les zones plus éclairées où se trouvent sa nourriture le plancton. Vue et mouvement sont originellement connectés.

Par contre les cellules photoréceptrices des ocelles dorsales de l'amphioxus contiennent, elles, de la mélanopsine. Cette mélanopsine est une opsine, dite de type r, une molécule sœur des opsines-c. On entend par

là que ces deux molécules viennent de la duplication d'un gène plus ancestral dans le génome d'un des ancêtres de l'amphioxus. Les deux gènes ont ensuite divergé, « évolué », indépendamment l'un de l'autre. L'intéressant est que cette classe d'opsines-r existe aussi chez les vertébrés. Ainsi chez les mammifères elle se trouve dans les cellules ganglionnaires de la rétine et sert tout comme les opsines-c à détecter la lumière. Mais ces cellules sont impliquées non dans la vision mais dans une horloge interne impliquée dans les rythmes circadiens. La mélanopsine activée par la lumière, active dans le cytoplasme non pas une transducine mais d'autres G-protéines. Une fonction « horloge » qui réclame la détection de la lumière s'est donc ajoutée, chez les vertébrés, à celle de la vision, et c'est une fonction qui a été intégrée par les rythmes circadiens. On a là un superbe exemple de l'émergence à partir d'une même propriété (la capacité à détecter la lumière) de deux fonctions différentes facilitées par la duplication de gènes. Mais bien sûr les opsines, qu'elles soient de type -c ou -r, représentent la même « invention ».

Comme nous n'avons toujours pas trouvé la transition entre une rhodopsine visuelle et non visuelle faisons d'un coup un très grand saut et demandons-nous à quoi ressemble la molécule photoréceptrice des yeux d'animaux d'une division du vivant très éloignée des vertébrés. Retrouvons la *Drosophile*. Son œil est très différent de ceux des vertébrés, il est composé non pas d'une rétine mais de 800 ommatidies. Chacune de ces ommatidies contiennent des neurones photorécepteurs (Figure 6 de l'opuscule). La question est donc de savoir quelle molécule photosensible ces cellules utilisent. Eh bien les photorécepteurs de ces neurones sont encore des rhodopsines liées au rétinale.

Mais si l'on y regarde d'un peu près l'on découvre que ces photorécepteurs sont beaucoup plus proches des mélanopsines des mammifères que des opsines-c des cônes et bâtonnets. On vient de voir que ces mélanopsines sont, chez les vertébrés, impliqués dans l'horloge circadienne et chez l'amphioxus dans les ocelles dorsales. Cela montre tout d'abord que les deux classes d'opsines (-c et -r) liées au rétinale existaient avant même la séparation des insectes et des vertébrés, c'est-à-dire chez leur ancêtre commun. Cet ancêtre commun c'est l'ancêtre commun aux protostomiens et aux deutérostomiens des figures 28 et 35 de l'opuscule.

Et, clairement, les insectes ont développé leurs yeux à partir de l'une des classes, les opsines-r, et les vertébrés de l'autre, les opsines-c. La même famille de molécules et pourtant des yeux radicalement différents.

L'étude du groupe voisin, à la base de tous les animaux métazoaires, le phylum des cnidaires (des coraux, les méduses comme *Tripedalia*), va nous renseigner sur une période encore plus lointaine, il y a 600-700 millions d'années. On trouve chez la larve de ces animaux des cellules photoréceptrices. Elles présentent sur leur membrane des microvillosités à proximité de granules de pigment et d'un cil (Figure 5d). Voici donc une structure impliquée dans la détection de la lumière mais qui est beaucoup plus simple que nos yeux et même que l'œil frontal de l'amphioxus, puisqu'elle ne consiste qu'en une seule cellule. Eh bien, ces microvillosités abritent aussi des rhodopsines qui détectent la lumière aussi grâce au rétinol et cette détection provoque le battement du cil et donc le mouvement de la larve. Par ailleurs la méduse *Tripedalia* présente des yeux plus compliqués et encore avec des cellules photoréceptrices de type ciliées. Les rhodopsines des cnidaires sont proches des opsines-c. Elles dérivent d'un gène unique à quatre exons présent chez l'ancêtre commun à tous les eumétazoaires.

Ainsi la conclusion de ce survol des molécules photosensibles chez les animaux est plutôt surprenante : bien que leurs yeux présentent des structures extrêmement différentes, bien que tout semble pouvoir changer dans l'œil, tous les animaux utilisent et ont toujours utilisé pour « voir », même il y a 700 millions d'années, la même molécule. Chez l'ancêtre des métazoaires il y avait une opsine et chez l'ancêtre des bilatériens, il y avait deux familles de telles molécules, les opsines-c et les opsines-r (les mélanopsines). Les vertébrés ont préféré pour la vision les opsines-c et s'y sont tenus, les insectes les opsines-r. Clairement la vue a été originellement couplée avec le mouvement via un cil comme chez les méduses ou via une connexion, par des signaux neurochimiques, aux centres de locomotion, comme chez l'amphioxus.

Cette unique origine moléculaire pour les yeux des animaux, à partir d'une unique molécule à l'origine des trois classes d'opsines (c, r et cnidaires), est un fait remarquable du point de vue de notre compréhension



de l'évolution en général et de l'évolution des organes en particulier. Y-aurait-il une origine moléculaire aux organes ? Ceci est discuté dans l'opuscule.

### **Chez les unicellulaires, Eukaryota et bactéries**

Nous sommes arrivés à la base des animaux et n'avons pas encore rencontré le moment où cette molécule devint photosensible. Nous devons donc remonter encore plus dans le temps et considérer les unicellulaires, les Eukaryota. Eh bien, on trouve encore, chez par exemple, certaines algues vertes, comme *Chlamydomonas*, (un chlorophyte, Figure 4, annexe 8), une rhodopsine avec la même structure que celle des animaux. En effet cette molécule tout comme les opsines-c et -r ont sept domaines transmembranaires et un rétinol lié à une lysine du 7ème domaine transmembranaire. A quoi sert cette molécule chez cette algue ? *Chlamydomonas* possède un organite le stigma, aussi appelé tache oculaire. Celle-ci concentre la lumière qu'elle envoie sur une membrane cellulaire et c'est cette membrane qui abrite les rhodopsines. Grâce au rétinol ces rhodopsines détectent la lumière, ce qui déclenche un signal de dépolarisation de la membrane. Ce signal se propage à un flagelle qui se met à battre. Détection de la lumière et mouvement sont ainsi couplés. On appelle cela phototactisme. Le signal à la membrane est direct et passe par le canal que forment les 7 domaines membranaires. Il n'utilise pas de relais cytoplasmique. La différence d'avec les opsines animales est qu'il n'y a pas de connexion à des protéines G.

Mais ici, grosse surprise : la membrane qui abrite les rhodopsines n'est pas celle de la cellule, c'est celle du chloroplaste de l'algue ! Or, souvenons-nous, ces chloroplastes dérivent des cyanobactéries, ces bactéries photosynthétiques que nous avons déjà croisées plusieurs fois. Cette observation a deux conséquences. Premièrement regardez dans la figure 4 de l'annexe 8 la place des chlorophytes. Ils appartiennent au grand groupe des chlorobiontes dans lequel on trouve aussi les plantes. Les taches oculaires et les rhodopsines des chlorophytes ne sont pas dans le lignage animal. Les dispositifs et les rhodopsines des animaux ne descendent pas des taches oculaires et des rhodopsines du chloroplaste d'une algue. La bifurcation entre animaux et végétaux s'est faite avant leur évolution (regarder la grosse flèche violette sur la figure 4 de l'annexe 8, là où

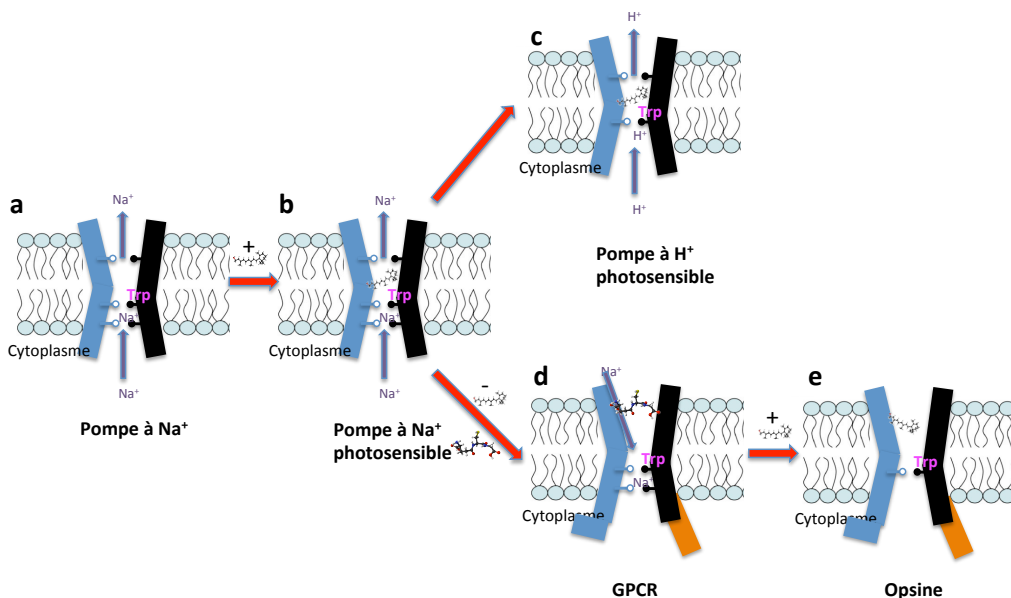
l'endosymbiose primaire s'est produite). Le stigma des algues vertes est une évolution parallèle. L'œil des végétaux en quelque sorte. Mais force est de constater qu'il n'a pas, dans ce règne, eu le succès qu'il a eu chez les animaux. Il a peu évolué et a même disparu chez les plantes (les embryophytes). Cependant tout comme chez les cnidaires, détection de la lumière et mouvement ont été couplés.

Deuxième conséquence, on peut immédiatement suspecter que les cyanobactéries elles-mêmes utilisent ces rhodopsines. Nous faisons donc d'un coup un énorme saut en arrière qui nous amène avant les eucaryotes. Et, effectivement on trouve chez certaines cyanobactéries actuelles comme *Leptolyngbya* (qui forme des filaments) un spot orange riche en caroténoïdes qui rappelle le stigma des algues et tout porte à croire que ces bactéries ont des organites qui correspondent à un photorécepteur qui utilise les rhodopsines qui les aide à se mouvoir très lentement vers une source lumineuse. Il n'est donc pas impossible que chez les bactéries existent aussi des dispositifs de détection de la lumière couplés à un mouvement, du phototactisme.

Y-a-t-il, en dehors des cyanobactéries, des rhodopsines chez les bactéries ? Eh bien oui, on trouve de nombreuses rhodopsines bactériennes et archéales et elles captent des photons par un rétinol. Ces rhodopsines microbiennes ont des fonctions qui s'éloignent de celles que nous avons jusqu'à présent rencontrées. Leurs fonctions concernent le métabolisme de l'énergie. Certaines de ces rhodopsines sont des pompes à sodium ou à protons qui ont une fonction phototrophique. Elles transfèrent l'énergie des photons en énergie chimique qu'elles stockent dans de l'ATP (grâce aux ATP synthases, des enzymes  $\text{Na}^+$  et  $\text{H}^+$  dépendant, responsables de la synthèse d'ATP). D'autres rhodopsines sont des pompes à cations tel le chlore, elles ont une fonction de transport d'ions. Elles font rentrer activement ces ions (surmontant le potentiel de membrane) en utilisant l'énergie des photons et non l'ATP ce qui réduit considérablement le coût énergétique. Ces transporteurs d'ions dépendant de la lumière représentent le premier moyen par lequel la vie (les procaryotes) a exploité la lumière comme source d'énergie, avant même la photosynthèse. Les rhodopsines du phototactisme ont probablement évolué à partir des rhodopsines pompes à

ions<sup>64,68</sup>. Ce sont elles qui sont, sans doute, les plus anciennes molécules photosensibles<sup>64,65</sup>).

Voilà, nous sommes arrivés au bout de notre chasse. Nous avons identifié des molécules qui s'apparentent aux rhodopsines visuelles des cônes et bâtonnets de nos yeux, mais ont une fonction non visuelle. Mais, et c'est là sans doute le plus extraordinaire, il a fallu remonter jusqu'aux bactéries des plus primitives. Et chez ces bactéries la structure de ces protéines - sinon leur séquence - présente encore des homologies avec nos rhodopsines : toutes deux sont des protéines à 7 domaines transmembranaires qui forment un canal avec un site de fixation du Na<sup>+</sup> conservé et dans lequel est logé un rétinol sur une lysine du 7ème domaine.



### Le scénario qui a mené à la vision

Alors quel est le scénario qui a mené à la vision ? Il nous est donné par l'histoire de cette molécule<sup>64</sup> (Figure 2).

**Figure 2 : le scénario qui a mené à la vision.** a-c) étapes chez les bactéries ; d, e) étapes chez les eucaryotes. Modifié de <sup>64</sup>.

Tout d'abord on peut envisager que l'étape que nous recherchons, le changement dans la protéine qui la rend « visuelle », a eu lieu il y a 4 milliards d'années. Chez des bactéries anaérobiques, des mutations ponctuelles dans un gène codant une pompe à sodium (Figure 2a) auraient, par chance, créé une poche complémentaire au rétinol rendant cette protéine susceptible de se lier de manière covalente au groupe ε-amino d'une lysine de l'α-hélice 7 (Figure 2b). La pompe devint ainsi photosensible. En

s'isomérisant à la lumière le rétinol provoque (par le transfert d'un proton sur un acide aspartique) l'expulsion de  $\text{Na}^+$  en dehors de la cellule. Cela crée un flux de  $\text{Na}^+$  qui se traduit par un gradient. Or ce gradient de  $\text{Na}^+$  peut être utilisé tel quel par des ATP synthase  $\text{Na}^+$  (des protéines membranaires aussi) pour fabriquer de l'ATP, c'est-à-dire, pour stocker l'énergie (vidéo « transforming light-bacteriorhodopsin »). Cela ne nécessite aucun couplage physique entre l'ATP synthase et la nouvelle protéine ! C'est gagné ! Surtout pour des bactéries anaérobiques dans ces temps de famine, (on est avant l'émergence de la photosynthèse). La première forme de phototrophie venait, en peu d'étapes, d'apparaître ! Notons que cette fonction nouvelle résulte, bien sûr, de l'interaction entre le rétinol et la lumière, mais implique aussi les pompes à  $\text{Na}^+$  et l'ATP synthase. Ce sont ces éléments préexistants qui font que les mutations qui combinent le rétinol à la pompe à  $\text{Na}^+$  furent une chance. Cet événement n'a certes pas créé la vision telle qu'on la conçoit aujourd'hui mais un système qui réagit à la lumière. Ce fut la première marche.

Cette pompe à  $\text{Na}^+$  photosensible se serait ensuite diversifiée en pompes à  $\text{H}^+$  photosensibles (Figure 2c), qui couplées aux ATP synthase  $\text{H}^+$  dépendantes constituaient de nouvelles machines encore plus performantes pour la phototrophie, et en pompes à autres cations. Pour ces transformations il ne faut guère que quelques mutations qui changent quelques acides aminés le long du canal pour faire circuler le cation.

Chez les cyanobactéries et donc beaucoup d'algues (par leurs chloroplastes) ces pompes photosensibles rencontrèrent un système complexe qui avait évolué en parallèle, le flagelle. Et là encore c'est le gradient de cations créé par ces pompes qui est directement utilisé pour activer le flagelle (cation dépendant). De nouveau une petite transformation, due là encore à seulement quelques mutations aléatoires (une pompe à proton devient une pompe à cations) mais qui crée d'un coup la fonction du phototactisme en additionnant détection de la lumière et mouvement (association rhodopsine-flagelle). Le phototactisme permet à ces cellules de contrôler leur exposition à la lumière que par ailleurs elles utilisent déjà pour la photosynthèse. Mais cette voie n'ira pas plus loin que les algues dans le grand règne des végétaux. Et chez les plantes ces molécules seront perdues et jamais réinventées.

La voie pour arriver aux rhodopsines animales est parallèle à ce que nous venons de décrire et nous fait repartir de la première marche, les pompes  $\text{Na}^+$  photosensibles bactériennes (Figure 2b), pour aller à LECA, -1500 millions d'années en arrière. Le raisonnement est le suivant : les membranes des cellules eucaryotes primitives comme LECA étaient très probablement plutôt imperméables au  $\text{Na}^+$  mais beaucoup moins à  $\text{H}^+$ . Aussi les pompes  $\text{Na}^+$  (Figure 2b) héritées de l'un de leurs deux ancêtres microbiens (archée ou bactérie) étaient pour ces cellules un avantage immédiat évident car elles leur permettaient de pomper contre l'imperméabilité. Elles les ont donc conservées. Ces pompes se seraient ensuite diversifiées en récepteurs membranaires à des molécules externes, mais après perte du rétinol et évolution de la cavité ainsi libérée en une poche de liaison  $\text{Na}^+$  pour de petites molécules organiques (Figure 2d). La perte du rétinol est chose facile il suffit d'une mutation qui change la fameuse lysine en un autre acide aminé. L'évolution en poche pour de petites molécules l'est aussi.

A ce stade les transformations des rhodopsines concernent aussi la transmission du signal vers l'intérieur de la cellule. La partie cytoplasmique de la rhodopsine va établir des contacts avec une autre protéine de la membrane plasmique, une protéine G. Nous ne les discuterons pas. Elles seraient à ce moment-là devenues l'ancêtre de la famille des GPCRs (Figure 2d). Tout cela s'est passé au plus tard chez l'ancêtre commun aux animaux et végétaux. En effet on trouve une GPCR chez les plantes. Ces GPCRs sont les éléments de voies de signalisation, par lesquelles un message reçu de l'extérieur est communiqué et interprété à l'intérieur de la cellule. Mais à ce stade ces GPCRs ont perdu la photosensibilité ! Cependant elles sont bien la deuxième marche vers les rhodopsines animales. En effet et c'est là une grosse surprise, c'est l'une de ces GPCRs qui a perdu la photosensibilité, et non pas une rhodopsine photosensible de type microbien, qui est à l'origine des opsines visuelles. Cette GPCR aurait regagné le rétinol, en refaisant ce qui avait été fait chez les bactéries, en recréant une poche avec une lysine pour fixer le rétinol (Figure 2e). Cela aurait eu lieu au plus tard chez l'ancêtre commun aux eumétazoaires (-1 milliard d'années environ). C'est la troisième marche qui sans doute a nécessité moins d'étapes que la deuxième. Les circonstances de ce tour de

passé-passe (perte et gain du rétinol) ne sont, elles non plus, pas connues. Mais il est certain que l'évolution du mode de transmission du signal était un prérequis chez ces Eukaryota qui sont 1000 fois plus grandes que les microbes et chez lesquels une amplification des signaux est nécessaire. Si ce scénario est le bon (les arguments moléculaires sont sérieux mais cela reste à confirmer) le recrutement du rétinol aurait eu lieu deux fois dans l'histoire des molécules de la vision, une fois chez les bactéries et une autre fois chez les Eukaryota. Mais ces deux fois par des molécules apparentées. L'histoire des rhodopsines visuelles aurait commencé par celle qui va de la protéine membranaire des pompes microbiennes aux récepteurs membranaires à des molécules externes, les GPCRs chez LECA, et aurait été suivie par celle de l'acquisition du rétinol chez ces dernières.

Les animaux les plus primitifs, les cnidaires (méduses) ont utilisé le gène de cette GPCR visuelle (après retro transposition et perte des 4 introns) pour les premiers véritables yeux apparus chez le vivant<sup>69</sup>. Chez l'ancêtre commun aux bilatériens ce gène s'est dupliqué et les deux exemplaires se sont diversifiés. L'un d'entre eux a donné les opsines de type c (4 introns et ça ne bougera plus) associées aux photorécepteurs ciliaires, le signal après transduction conduisant à une hyperpolarisation de la cellule. L'autre a donné les opsines de type r (10 introns et ça ne bougera pas non plus) associées aux photorécepteurs rhabdomériques, le signal après transduction conduisant à une dépolarisation de la cellule (voir figure 8 in<sup>15</sup>). Les deux familles de molécules dérivées de ces deux gènes sont celles qui sont utilisées pour la vision chez l'ensemble des animaux. Et malgré la diversité des métazoaires il n'y a pas eu, chez aucun d'entre eux et pour aucun des systèmes de la vision, une autre origine. Cette observation est très surprenante. Son explication nous est donnée dans l'opuscule. Elle a trait à la façon dont les organes sont construits et constitue la clé du développement et de l'évolution. On peut dire qu'il y a une origine moléculaire aux organes.

En parallèle, l'ancêtre de la famille des GPCRs apparu chez les Eukaryota (Figure 2e) a été à l'origine de multiples nouvelles fonctions. C'est que, en tant que récepteurs membranaires à des molécules externes, les mutations qui peuvent se traduire par la détection de nouvelles molécules sont fréquentes et donc le répertoire des molécules détectées (des



peptides, des acides aminés des protéines, des acides gras etc.) peut augmenter rapidement. Plus le répertoire est large plus la cellule est renseignée sur le monde extérieur et sur les autres cellules. Les GPCRs incluent les molécules de la détection des odeurs (700 gènes chez l'homme), du goût (26 gènes chez l'homme) et de bien d'autres fonctions (84 autres gènes chez l'homme). Pour l'odorat et le goût sont associés des organes bien définis et pour les autres à tout le moins un tissu. La communication entre cellules en particulier chez les organismes multicellulaires passe aussi par ce type de molécules.

Voilà la boucle est bouclée. Il existe, chez les animaux une molécule de la vision et une seule. La séquence de son apparition n'est pas la plus directe qu'on puisse concevoir. Si l'événement princeps, l'association de la protéine et du rétinol est simple, les événements qui ont menés à la protéine à 7 domaines transmembranaires, chez les bactéries, et ceux qui ensuite lui ont conféré le caractère de récepteur capable de transférer le message aux G protéines, chez les Eukaryota, le sont moins. Cette molécule complexe est apparue dans l'histoire du vivant avant la multicellularité et le développement, et donc avant les organes. Elle nous vient des bactéries. Le fait que ça soit une molécule complexe rend plus probable les évolutions ultérieures, en particulier les différentes manières de processor le signal via les G protéines. Cela augmente le jeu des possibles. Ce survol de la vision nous a montré comment à partir d'une molécule une fonction émerge d'un coup (la phototrophie, le phototactisme chez les bactéries) et là vraiment comme par magie, mais par rencontre des conditions qui donnent un « sens » à l'événement fortuit. Et « sens » veut dire, et ne veut que dire, avantage sélectif.

Ce survol montre aussi comment l'émergence d'une fonction favorise l'émergence d'autres fonctions qui de ce fait se trouvent apparentées. Ainsi la vision, l'odorat et le goût ont émergé de la même famille de molécules et sont ainsi de manière insoupçonnée profondément, matériellement, liés par cette même origine moléculaire. L'émergence par chance d'une molécule transmembranaire avec une poche pour le rétinol, une rhodopsine photosensible bactérienne a donné naissance à des molécules « eucaryotes » capables de reconnaître, après perte du rétinol, toutes sortes de composés chimiques. Et ces molécules des sens réévoluent en une molécule de la

vision en se combinant une deuxième fois à ce même rétinale. L'émergence précipite l'émergence de tous les organes des sens. Mécaniquement. Et dans un premier temps non par gain mais par perte.

A ce stade un autre enseignement général peut être tiré, qui est que les nouvelles fonctions correspondent toujours à de nouvelles interactions, des additions de chances. La conséquence purement mécanique mais importante est que plus il y a de fonctions existantes dans un système, plus il y a de possibilités de nouvelles fonctions. La fonction appelle la fonction.

Il nous reste aussi à comprendre comment à partir de cette molécule les organes ont émergé (ceci est développé dans l'opuscule), mais l'on sait déjà que leur fonction existe en miniature simplifiée dans les organites des unicellulaires. Et la diversité des organites des unicellulaires excède de beaucoup la diversité des organes des métazoaires.





## ANNEXE 8

### LA PLACE DU DÉVELOPPEMENT DANS LE MONDE DU VIVANT.

Dans l'opuscule j'ai présenté les principes qui permettent de comprendre le développement embryonnaire, c'est-à-dire la formation d'un organisme multicellulaire complexe à partir d'une cellule unique, un processus connu sous le nom d'ontogenèse.

La fascination que ce processus engendre vient, en partie, de l'extrême rapidité, de quelques heures à quelques jours suivant les espèces, avec laquelle des structures très élaborées et très différentes apparaissent et forment un organisme autonome (Figures 1-4 de l'opuscule). L'idée comme quoi ce processus doit être d'une grande complexité paraît incontournable. Et, de fait, l'une des conclusions de cet opuscule est que le développement embryonnaire est sans doute le processus le plus complexe (la complexité étant mesurée par le nombre d'interactions différentes entre entités différentes) de tout l'univers. Même une comparaison avec le « big-bang » duquel les photons, les neutrinos, les premiers atomes, la matière noire, l'énergie noire et en conséquence notre univers a émergé, si elle pouvait être faite, dirait sans doute que ce dernier n'a pas un tel niveau de complexité. La cosmogénèse dans son ensemble, la formation des galaxies, des étoiles et des planètes, sont de loin plus simple que l'ontogenèse qui, cependant, en est le prolongement. Mais une autre des conclusions de cet opuscule sera que, bien que complexe, ce processus est compris et est, dans ses principes, compréhensible par tous.

Replaçons le développement dans le monde du vivant (Figure 1). Premières constatations : le développement est apparu chez les eucaryotes (nous définissons plus loin ce terme) ; il n'a pas concerné les

deux autres divisions du monde du vivant, bactéries et archées qui composent les procaryotes (Figure 2) ; il est apparu indépendamment au moins deux fois, d'une part chez les animaux et d'autre part chez les plantes (Figure 1, les rectangles vert et marron) ; cette apparition a été plutôt tardive à l'échelle de l'évolution (-550 millions d'années contre quatre milliards d'années) ; chez les animaux elle a accompagné ce qui est connu comme l'explosion cambrienne durant laquelle en quelques millions d'années seulement presque tous les phylums animaux actuels (et quelques autres qui ont rapidement disparu) ont pris naissance. Par phylum, il faut comprendre un groupement d'êtres vivants caractérisés par une organisation particulière (nous y reviendrons longuement) celle par exemple, des vertébrés, comme les poissons ou les singes, très différente de celles des mollusques, comme l'huître ou des arthropodes comme la mouche ou bien encore des plantes. Ces premières constatations appellent déjà plusieurs questions. Pourquoi une apparition si tardive ? Pourquoi le développement n'est-il pas apparu dans les deux autres grandes divisions du vivant pourtant beaucoup plus anciennes ? Le développement serait-il à l'origine de l'explosion cambrienne ? Si oui quelle propriété de ce processus pourrait en rendre compte ? Pour répondre à ces questions il faut en savoir un peu plus sur ce qui s'est passé avant l'apparition des développements.

Ont précédé l'émergence des développements : cinq cent millions d'années d'évolution des molécules (Figure 1, rectangle noir), avant l'apparition des premières bactéries et archées, les deux premières grandes divisions du vivant, à partir d'une cellule ancestrale commune appelée LUCA, pour last universal common ancestor ; puis deux milliards d'années d'évolution de ces organismes unicellulaires avant l'apparition des Eukaryota (Figure 1, rectangle bleu) à partir d'une cellule appelée LECA, pour last eukaryote common ancestor ; et encore un milliard d'années avant que les développements apparaissent chez les eucaryotes. A s'en tenir à ce constat, l'émergence du développement semblerait avoir été beaucoup plus difficile que l'émergence de la première cellule à partir de molécules ! Et encore une fois il semblerait aussi que les conditions requises n'aient été réunies que chez certains eucaryotes. Mais alors que sont les eucaryotes, d'où viennent-ils ?

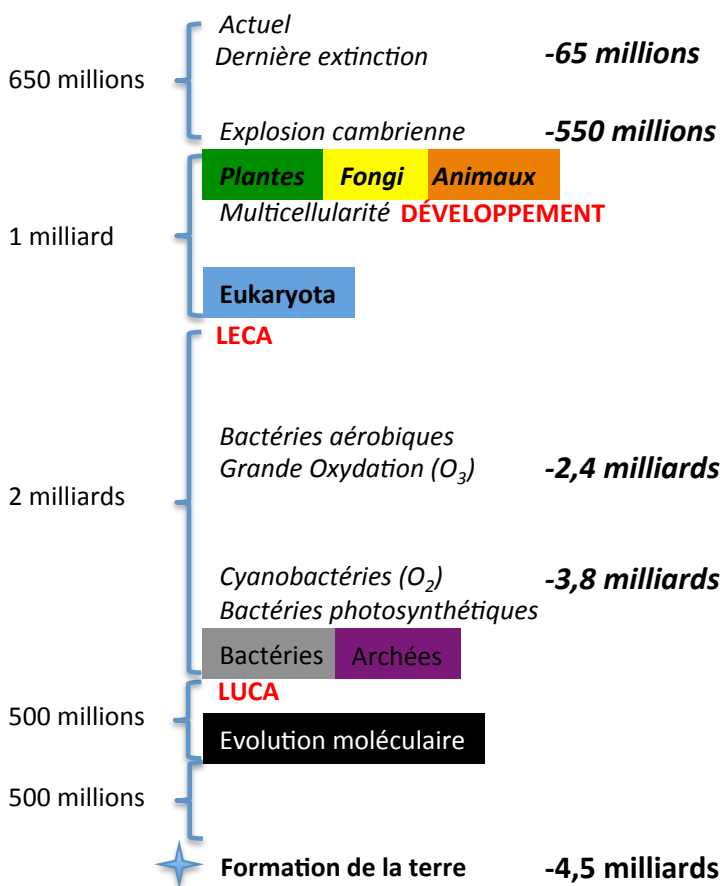


Figure 1. Quelques repères dans l'histoire du vivant.

Les Eukaryota regroupent les champignons (fongi), les animaux, les plantes et de nombreux autres embranchements constitués d'organismes monocellulaires (Figure 4) qui ont longtemps été regroupés sous les noms de protistes ou de protozoaires. Ensemble, ils dérivent tous de LECA. Les cellules eucaryotes sont caractérisées par différents traits. Tout d'abord dans

leurs cellules un noyau, l'organite qui contient l'ADN, entouré d'une membrane, d'où son nom eucaryote. Plus généralement des membranes internes compartimentent la cellule et cette compartimentation permet d'avoir différents types d'espaces (le reticulum endoplasmique, l'appareil de Golgi, les peroxysomes). Les autres traits des cellules eucaryotes, tout aussi distinctifs, sont la présence d'un « squelette » interne (le cytosquelette formé d'actine et des microtubules qui confèrent une rigidité à la cellule), de flagelles, d'un centriole et d'organites très élaborés, les mitochondries - productrices de l'énergie de la cellule en condition aérobie - et pour les plantes des chloroplastes - qui réalisent la photosynthèse.

En comparaison les procaryotes n'ont pas de noyau, leur ADN est dans leur cytoplasme qui n'est pas compartimenté et ne présente ni organisation ni organite. Ils sont dix fois plus petits (mille fois plus en volume) que les eucaryotes. Par contre ils présentent une extraordinaire diversité métabolique que ne présentent plus les eucaryotes. Ils sont chimiotrophes (utilisent des composés chimiques) ou phototrophes (utilisent la lumière) pour produire leur énergie ; lithotrophes (utilisent des composés minéraux)

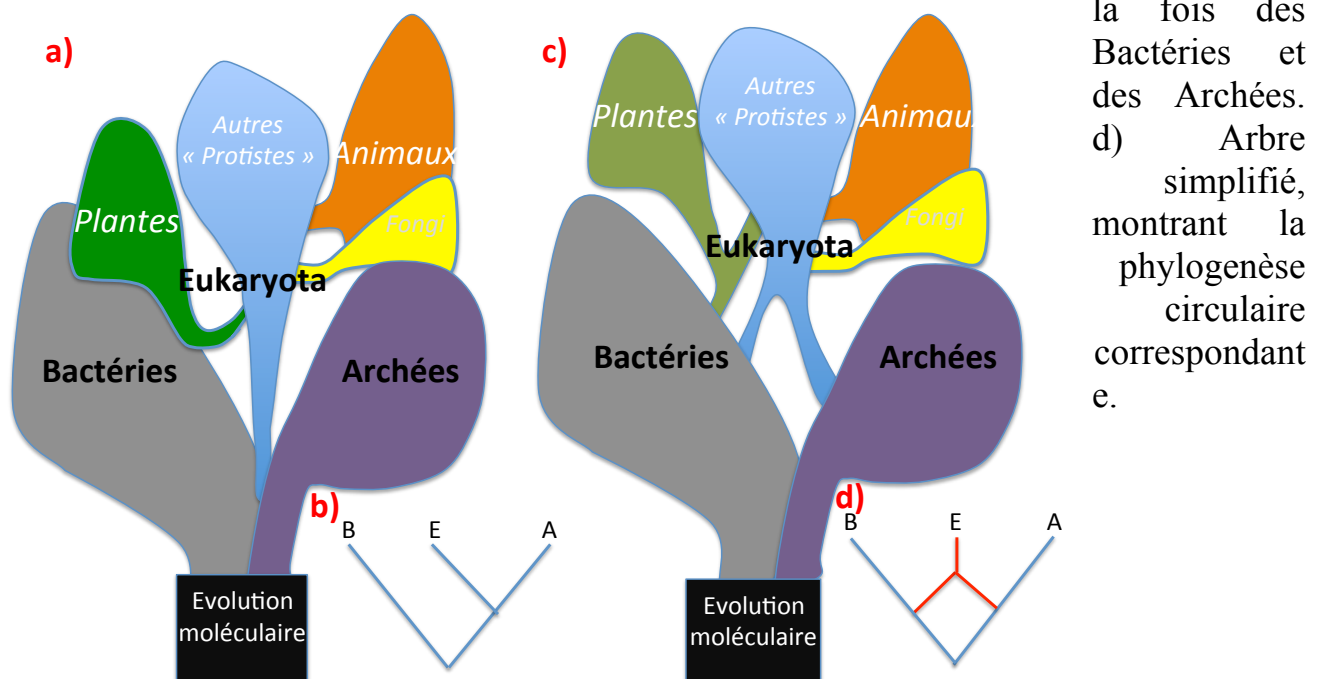
ou organotrophes (utilisent des composés organiques) pour la synthèse de leurs molécules. Et enfin, autotrophe ( $\text{CO}_2$  comme source de carbone) ou hétérotrophes (source de carbone organique) pour l'origine de leur carbone. Et les huit combinaisons sont rencontrées. Par exemple les cyanobactéries, qui ont joué un rôle très important dans l'évolution, nous verrons cela, sont des photoorganohétérotrophe. Les archées méthanogènes sont des chimioautotrophe anaérobies qui vivent dans des conditions extrêmes (comme beaucoup des archées) près des sources hydrothermales. Notons au passage qu'on estime que plus de 99% des procaryotes nous sont encore totalement inconnus.

Pendant longtemps on a pensé que les eucaryotes étaient le troisième domaine de l'arbre du vivant, et que dès le début de l'évolution trois branches se seraient séparées - bactéries, archées, eucaryotes - puis auraient évolué indépendamment (Figure 2a et b). Mais, on sait maintenant grâce aux comparaisons des séquences ADN de nombreux génomes de ces trois groupes qu'il n'en est rien. L'histoire est beaucoup plus surprenante. Les eucaryotes ne sont pas nés il y a 3,8 milliards d'années, LECA n'est pas le groupe sœur des archées, ni d'ailleurs des bactéries. LECA est née bien après les bactéries et les archées, il y a 1,5 milliard d'années. Il n'y a que deux domaines à l'arbre de la vie. Et, plus surprenant encore les comparaisons des séquences montrent que LECA n'est fille ni de l'un ni de l'autre ! Comment est-ce possible ? Voyons cela.

Les premiers procaryotes apparus sur terre tiraient leur énergie et leur nourriture (les éléments nécessaires à la fabrication de leurs constituants) des rares molécules organiques du milieu. Ils étaient chimiotrophes (auto- ou hetero-trophes) anaérobiques (et pour cause, il n'y avait pas d'oxygène dans l'atmosphère) et avaient du mal à produire leur énergie. Puis sont apparues des bactéries qui tiraient leur énergie non plus de ces rares molécules organiques du milieu mais de l'énergie solaire. Elles étaient phototrophes. Cela leur permettait de synthétiser leurs propres molécules directement à partir de l'azote et du gaz carbonique atmosphériques. Un avantage sélectif majeur ! Ces bactéries utilisèrent d'abord comme donneurs d'électrons dans ces réactions chimiques, de l'hydrogène sulfuré,  $\text{H}_2\text{S}$ , puis avec les cyanobactéries l'eau,  $\text{H}_2\text{O}$ , avec pour conséquence la libération d'oxygène,  $\text{O}_2$ , dans le milieu. La photosynthèse avait fait son

apparition. L'impact pour le milieu extérieur a été très lent mais phénoménal. En effet l'oxygène accumulé dans l'atmosphère a progressivement augmenté au cours des centaines de millions d'années qui ont suivi. La vie aérobie est alors apparue, encore une fois chez les bactéries. Elle leur permettait de produire leur énergie avec une grande efficacité car grâce à l'oxygène elles brûlaient complètement les composés organiques.

**Figure 2 : L'origine des Eukaryota.** a) L'arbre à trois domaines. Les lignages des Archées et des Eukaryota sont des groupes sœurs aussi vieux l'un que l'autre. b) Arbre simplifié, montrant la phylogénèse linéaire correspondante. c) Les bactéries et les archées sont les lignages primaires. Les Eukaryota apparaissent tardivement à partir à la fois des Bactéries et des Archées. d) Arbre simplifié, montrant la phylogénèse circulaire correspondante e.



Parallèlement, la nourriture étant, grâce à la photosynthèse, beaucoup plus abondante, des cellules plus grandes, environ dix fois plus grandes, sont apparues et ceci s'est produit du côté des archées. Certaines ont développé un processus d'ingestion de particules et même de bactéries. La phagocytose s'élaborait. Les premiers prédateurs étaient nés. Le vivant se nourrissait de plus en plus directement du vivant. Mais la libération d'oxygène dans le milieu de par son pouvoir oxydant représentait un danger mortel pour ces prédateurs comme pour les autres bactéries et archées anaérobies.

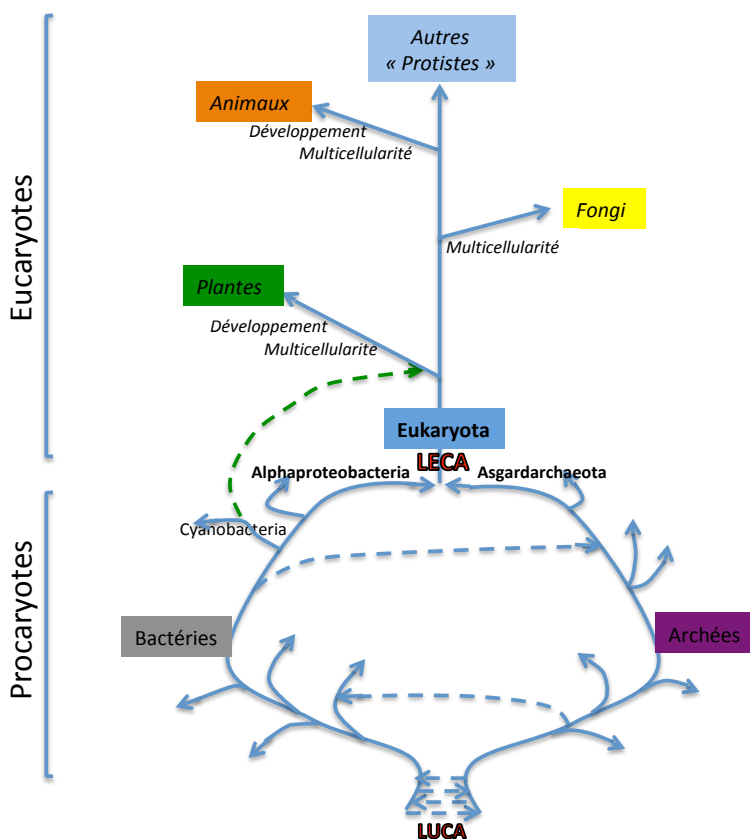
C'est alors qu'eut lieu un événement très singulier, à première vue un peu en dehors des manières habituelles de l'évolution, qui a donné naissance à LECA. Nous avons mentionné que les cellules eucaryotes contiennent certains organites, les mitochondries - productrices de l'énergie de la cellule en condition aérobie. LECA aurait pu inventer elle-même cette machinerie. Mais en fait son origine est tout autre. L'apparition dans le cytoplasme de LECA des mitochondries n'est pas dû à des modifications d'éléments existants internes des cellules - comme sait le faire l'évolution - mais à une annexion d'un organisme étranger entier, une bactérie aérobie proche des Alphaproteobacteria actuelles. Ce sont les archées prédatrices, proche du groupe actuel des Lokiarchaeota, qui grâce à la phagocytose (mais bien sûr sans digestion de la bactérie) ont accompli cette prouesse (Figure 3, les deux flèches qui convergent vers LECA). Ainsi, en une étape, l'ancêtre des cellules eucaryotes s'appropriait la manière la plus efficace de produire l'énergie, manière qui avait nécessité chez la bactérie un processus d'évolution de plusieurs dizaines de millions d'années. Et du même coup, et c'est sans doute le plus surprenant, le génome de la bactérie ingérée fut massivement transféré vers le noyau de LECA. Deuxième incongruité par rapport à l'évolution : l'appropriation d'un génome entier. Et à vrai dire qui s'appropriera le génome de l'autre est impossible à dire. Le génome des eucaryotes est hybride. La contribution des bactéries au génome de LECA fut égale à celle des archées. LECA n'est la fille ni de l'un ou de l'autre mais des deux (Figure 2c et d, figure 3).

Les gènes apportés par les archées furent, entre autres, les éléments du cytosquelette, de la machinerie transcriptionnelle, du trafic vésiculaire et de la mitose. Ceux apportés par les bactéries furent ceux de la bêta-oxydation, de la chaîne des transporteurs d'électrons, du métabolisme des sucres, des voies de biosynthèse des lipides, de la biotine, de l'hème et des groupes fer-sulfure. Et aussi les introns de groupe II qui évoluèrent en introns épissés (voir plus loin).

Ce pourquoi une archée anaérobie et une bactérie aérobie se sont ainsi associées doit bien être clair : l'archée apportait en abondance à la bactérie les produits de sa glycolyse (lactate et pyruvate), qui ne lui fournissait que bien peu d'ATP (2 par molécule de glucose). La bactérie les convertissait en CO<sub>2</sub> en produisant beaucoup d'ATP (une trentaine par molécule de

glucose). Or cet ATP, est un véritable concentré d'énergie (l'énergie est stockée dans sa liaison phosphate-phosphate). Il était ainsi fourni en retour à la cellule. Cette association avait donc la forme d'une endosymbiose rapidement poussée à son extrême avec la transformation de la bactérie phagocytée en une mitochondrie et la mise en commun des deux génomes. Avantages réciproques tellement énormes que cette « alliance » de deux représentants des deux grands domaines du vivant de l'époque a produit le premier représentant du troisième domaine du vivant, les Eukaryota !

Un scénario similaire expliquera, plus tard dans l'évolution, l'origine des chloroplastes par l'ancêtre commun aux plantes par l'annexion de bactéries photosynthétiques, des cyanobactéries (Figure 3, la flèche qui part des cyanobactéries et croise la ligne des eucaryotes). Notons que quels que soient les détails de ces annexions, il n'y a aucun doute sur l'origine bactérienne des mitochondries et des chloroplastes. Tout ceci a des répercussions majeures sur notre manière de penser l'évolution. A ces deux endroits l'arbre de la vie n'est pas linéaire mais circulaire (Figure 2C et D) ! Plus généralement d'ailleurs on sait depuis peu qu'entre archées et bactéries il y eut plusieurs transferts massifs de gènes dans le sens bactéries vers archées, mais aussi dans l'autre. On sait aussi que l'échange de matériel génétique a été et est encore fréquent chez les bactéries. On parle de transfert horizontal (c'est-à-dire entre individus d'une même génération) en contraste avec les transferts verticaux (c'est-à-dire d'une génération à la suivante) qui nous sont familiers. Les filiations directes ne rendent pas compte des transferts horizontaux, qui font qu'une espèce donnée ne dérive plus d'une seule espèce mais de plusieurs. Et parfois même comme pour les eucaryotes, d'espèces qui appartiennent aux deux domaines du vivant ! La phylogenèse n'est plus directionnelle mais pleine de « ronds ». Elle est réticulée. Plutôt qu'à un arbre, on a affaire à un réseau phylogénétique.



**Figure 3 : Arbre phylogénétique montrant les transferts horizontaux majeurs qui ont eu lieu au cours de l'évolution.** Des transferts horizontaux (les flèches en pointillées) entre bactéries et archées ont eu lieu à plusieurs reprises. Le dernier ancêtre commun des eucaryotes, LECA, provient de l'endosymbiose entre une archée proche des Asgardarchaeota actuels<sup>70</sup> et une bactérie proche des Alphaproteobacteria actuels (les mitochondries de nos cellules). Le dernier ancêtre commun des plantes provient de l'endosymbiose entre une cyanobactérie et un eucaryote

bikonte (la flèche pointillée verte).

Quelles constatations étonnantes ! Nos cellules, mais aussi celles des mouches ou même des végétaux sont habitées par des bactéries, certes modifiées, mais qui ont gardé une certaine autonomie (elles ont leur propre génome et se divisent indépendamment de la cellule). Et ces bactéries dans ces divers organismes ont toutes une même et unique origine (une  $\alpha$ -protéobactérie, proche des Rickettsia actuelles). Ces bactéries modifiées en mitochondries sont indispensables au métabolisme de nos cellules. De plus, les cellules de toutes les plantes hébergent en plus des mitochondries une autre bactérie transformée en chloroplastes, eux aussi autonomes, qui leur confèrent ce qui constitue leur caractère le plus discriminatif, celui de capter et de stocker l'énergie solaire ! Et là encore, tous les chloroplastes de toutes les plantes ont une même et unique origine : une cyanobactérie qui vivait il y a un milliard et demi d'années.

Après LECA, les cellules eucaryotes se sont encore diversifiées, ont acquis d'autres organites achevant de leur conférer un niveau d'organisation infiniment supérieur à celui des bactéries. Le centriole et le réseau de microtubules ouvraient la possibilité de la division des cellules

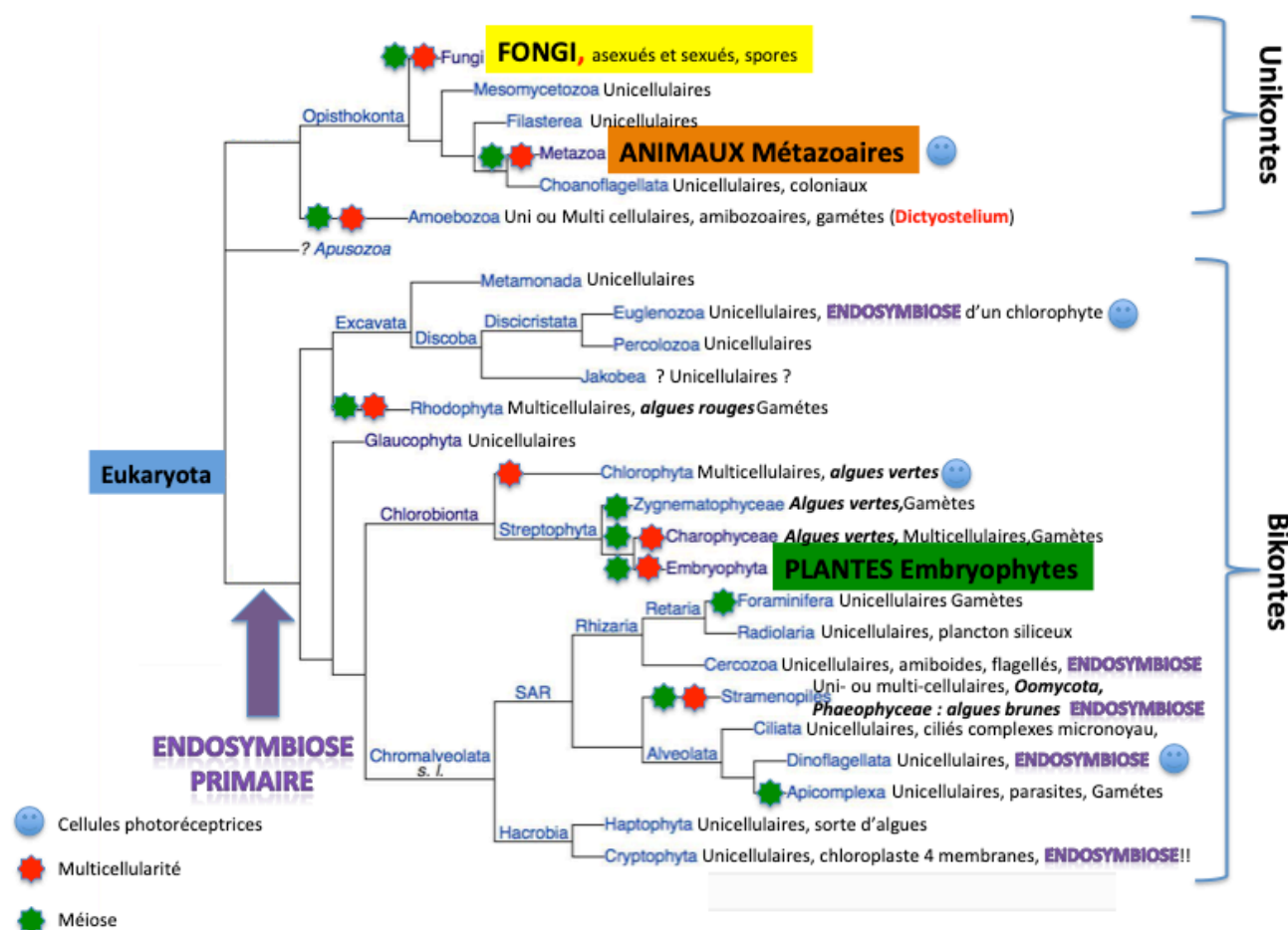


par la mitose puis par la méiose. La méiose fut encore une de ces inventions « inimaginables » aux conséquences incalculables. Ce mode de division de la cellule distribue dans les cellules filles (gamètes ou spores) des génomes différents créant ainsi de la diversité génétique dans la population et accélérant donc la vitesse d'évolution des cellules. Encore un palier atteint et conservé par quasiment tous les animaux, plantes et fungi. Notons enfin un autre accélérateur de l'évolution et non des moindres. Les gènes des eucaryotes sont morcelés et cela les différencie de ceux de procaryotes. Un gène typique d'eucaryote est une suite d'exons et d'introns non codants (Exon1-Intron1-Exon2-Intron2-Exon3-Intron3- etc.). Une machinerie particulière, le splicéosome, produit à partir de l'ARN transcrit l'ARN mature en épissant (éliminant) les introns. Nous y reviendrons (voir le chapitre 2 de l'opuscule) mais ceci est mentionné ici car cette circonstance permet de coder plusieurs protéines par un seul ARN (ici par exemple Exon1-Exon2-Exon3 et Exon1-Exon3) et aussi, au cours de l'évolution, de combiner les exons de différents gènes pour fabriquer de nouvelles protéines. Les gènes morcelés épissés par le splicéosome sont spécifiques des eucaryotes. Ils ont évolué très tôt après LECA à partir des introns qui s'auto-épissent et qui sont communs chez les bactéries et les archées. Sans détailler plus, durant le milliard d'années qui a précédé l'explosion cambrienne l'évolution a progressivement inventé tous ces ingrédients cellulaires nouveaux et une multitude d'autres (des vacuoles pulsatiles, des cils, des flagelles, des cils sensoriels, des sortes d'appendices, des pièces buccales etc.). En témoigne l'étonnante diversité des 300 000 espèces d'eucaryotes unicellulaires actuels (Figure 4).

Mais ni les bactéries, ni les archées, ni la majorité des eucaryotes unicellulaires qui ensemble, encore actuellement, regroupent la quasi-totalité de la diversité du vivant (en terme de diversité génétique et de mécanismes élémentaires différents) n'ont de processus qui s'apparenterait au développement embryonnaire. Ces êtres unicellulaires se perpétuent par divisions simples (fission, mitose, ou, mais plus rarement, méiose, les étoiles vertes, Figure 4) qui génèrent des cellules identiques et autonomes. Il a fallu attendre le palier de la multi-cellularité (les étoiles rouges, Figure 4) et surtout des spécialisations cellulaires qui ont pu s'ensuivre pour qu'apparaissent les développements embryonnaires indépendamment chez les animaux et chez les plantes. Et précisons-le cette invention n'a eu lieu

qu'une fois chez les animaux et une fois chez les plantes. Par là il faut comprendre que, par exemple, le développement des poissons et celui de l'homme aussi différents qu'ils semblent ont une même origine. Et de même le développement des mousses (bryophytes) et des plantes à fleurs (angiospermes) comme de tous les embryophytes ont une même origine.

**Figure 4 : Classification phylogénétique des Eukaryota**, basée sur des comparaisons moléculaires et cladistiques et montrant les relations de parenté entre les



différents groupes. Cette classification modifie considérablement (voyez les emplacements des algues) celle actuellement abandonnée des protistes-protistes. Le groupe des Opisthocontes, s'est séparé des autres eucaryotes il y a environ un milliard d'années. Il correspond aux Unikontes, le superrègne « animal », les bikontes correspondant au superègne « végétal » (avec l'acquisition du chloroplaste au moment de l'endosymbiose primaire (la flèche violette). C'est parmi les Opisthocontes que sont apparus les champignons (fongi) et les animaux (métazoaires). Les Plantes (embryophytes) avec les algues vertes composent les Chlorobiontes. Notez l'extraordinaire diversité des groupes. Ainsi chez les Cryptophytes le chloroplaste pourvu de quatre membranes dérive d'une endosymbiose d'un eucaryote

photosynthétique proche des Rhodophyta comme l'a montré leur nucléomorphe, qui est le vestige du noyau de l'eucaryote capturé. Une autre endosymbiose a eu lieu chez les Chlorarachniophytes (Cercozoa), l'endosymbiote étant un Chlorophyte. Et encore d'autres chez les algues brunes, les dinoflagellés et les euglènes. Les groupes qui comportent des organismes multicellulaires sont indiqués par une étoile rouge et ceux qui produisent de spores ou des gamètes par une étoile verte. Les positions des fongi, plantes et animaux ont été surlignées. Arbre, Nozaki et al. modifié <sup>71</sup> (origine wikipédia), annotations, Jean-François Nicolas.

A la question : quelle place le développement occupe-t-il dans le monde du vivant, nous pouvons donner une première réponse. C'est à n'en pas douter une des grandes transitions de l'histoire du vivant, de laquelle a dépendu l'expansion des animaux et des plantes. Elle se place au même rang que celle de LECA qui elle-même rend compte de l'expansion des protistes et bien sûr que celle des procaryotes et de LUCA. Et notons que ces grandes transitions s'accompagnent aussi d'un déplacement de ce sur quoi la sélection va principalement porter ensuite. Après une sélection sur les molécules qui menèrent à LUCA, ce fut une sélection sur les voies de biosynthèse et les manières de produire de l'énergie avec les bactéries et les archées, puis avec LECA sur les organites et autres moteurs moléculaires qui menèrent au seuil des développements. Et ensuite ce fut une sélection sur les organismes.

C'est ce dernier niveau qui nous a intéressé dans l'opuscule. Nous avons identifié quels sont les éléments qui rendent compte de l'émergence de ce processus si complexe qu'est le développement, duquel découle cette sélection sur les organismes et la diversification des animaux. Ceux qui imaginaient que ces éléments correspondaient à l'apparition de gènes aux propriétés extraordinaires, ou bien de nouveaux organites ou encore de nouvelles opérations cellulaires ont été déçus ! Les gènes et les opérations cellulaires qui sont mobilisés au cours des développements étaient toutes déjà présentes chez notre premier ancêtre multicellulaire, une éponge. Nous avons découvert que ce sont essentiellement des modifications dans la régulation des génomes qui ont tout changé. Nous avons vu aussi que d'avoir identifié la régulation des génomes comme le niveau clé dans l'émergence puis dans la complexification du développement nous a permis d'aborder sous un angle nouveau certaines des questions parmi les plus

mystérieuses, non seulement l'explosion de la diversité animale au cambrien mais, aussi, cette irritante question de la finalité que les développements embryonnaires semblent avoir.



## REMERCIEMENTS

Un merci particulier à Pierre-André Nicolas qui m'a aidé à beaucoup améliorer texte et figures et à Dominique et Marie-Dominique Nicolas, les premières lectrices. Mille mercis aussi à Claire Bonnerot, Pascale Briand, Suzanne Capgras, Christine Chevallier, André Choulika, Pascal Dardenne, Jacqueline Deschamps, Morgane Dolez, Sophie Eloy-Trinquet, Sylvie Forlani, Stephan Gasca, François Hyafil, Thierry Heidmann, Odile Heidmann, Estelle Hirsinger, Françoise Kamel, Emilie Legué, Claude Legrand, Luc Mathis, Lucile Montfort, John Muschler, Anne-Cécile Petit, Nadine Peyriéras, Didier Rocancourt, Isabelle Rosko, John Rubenstein, Josh Sanes, Inês Sequiera, Elena Tzouanacou, Muriel Vernet, Amélie Wegener, Valérie Wilson, Jean Deutsch, Gérard Buttin, Gabriel Gachelin, Marie-Thérèse Schnebellen, Philippe Brûlet, François Hyafil, Denise Paulin, Margaret Buckingham, Sigolène Meilhac, Edwig Jakob, Paul Berg, François Jacob et à tous les autres.

## REFERENCES DES FIGURES

- Figure 1 : Estelle Hirsinger, Nadine Peyri ras et Jean-Fran ois Nicolas. Film visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre 4'30 et 5'07.
- Figure 2 : R f rence du film [https://www.youtube.com/watch?v=ZEGvhkoVu\\_g](https://www.youtube.com/watch?v=ZEGvhkoVu_g)
- Figure 3 : R f rence du film, Marina Venero Galanternik et al., 2012, <https://www.youtube.com/watch?v=SDMFsKrUCvA> . Visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre 5'40 et 6'10.
- Figure 4 : R f rence du film, Rolf Karlstrom and Donald A. Kane <https://www.youtube.com/watch?v=hSMgt5gV-8s>. Visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre 6'30 et 7'10.
- Figure 5 : a : <https://instruct.uwo.ca/biology/489a/Choanoflagellates.html> ; b : Wikipedia; c : Ana Signorovitch/Yale University; d : <https://genome.jgi.doe.gov/Nemve1/Nemve1.home.html> ; et e : <https://scripps.ucsd.edu/news/churro-marine-worm-discovered-scripps-scientists-one-top-10-species-2017>; composition et annotations JFN.
- Figure 6 : a et b : r utilisation autoris e.
- Figure 7 : Composition et annotations JFN ; photos, a : <https://www.flickr.com/photos/146564503@N06/34619170301>; b-f voir figure 5 ;
- Figure 8 : JFN, H. Jakob et F. Jacob.
- Figure 9 : Photo : r utilisation autoris e, annotations JFN.
- Figure 10 : R f rence du film : <https://www.youtube.com/watch?v=D3fOXt4MrOM>, visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre de 14' 40''   16'.
- Figure 11 : R f rence du film : <https://www.youtube.com/watch?v=D3fOXt4MrOM>, ou visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre 16'   17' 20''.
- Figure 12 : R f rence du film : <https://www.youtube.com/watch?v=D3fOXt4MrOM>, ou visible   <https://hal.archives-ouvertes.fr/medihal-02008863v1>, entre 17' 20''   18'40''.
- Figure 13 : JFN. Photos a et b voir figure 12.
- Figure 14 : a et c, JFN ; b, C. Bonnerot, S. Eloy-Trinquet et JFN.
- Figure 15 : JFN.
- Figure 16 : JFN.
- Figure 17 : JFN.
- Figure 18 : a et b voir <sup>72</sup> ; c et d voir ??
- Figure 19 : voir <sup>73</sup> (free sharing), modifi .
- Figure 20 : voir <sup>74</sup> (open access).
- Figure 21 : a et b voir <sup>14</sup> ; c voir Wikipedia. [https://commons.wikimedia.org/wiki/File:Maus-Auge\\_Pax6-2.jpg?uselang=frs](https://commons.wikimedia.org/wiki/File:Maus-Auge_Pax6-2.jpg?uselang=frs)
- Figure 22 : JFN.
- Figure 23 : JFN.
- Figure 24 : Arbre phylog n tique voir <sup>54</sup> ; Photos, domaine public ; annotations et composition JFN.
- Figure 25 : Photos des yeux Dan-Eric Nilsson <sup>75</sup>, composition et annotations JFN.

- Figure 26 : Photos des yeux Dan-Eric Nilsson <sup>75</sup>, composition et annotations JFN.  
 Figure 27 : JFN. Pour le processus d'extrusion de l'ADN voir <sup>20 et 21</sup>.  
 Figure 28 : Photos voir figure 5; annotations JFN.  
 Figure 29 : a : voir <sup>26</sup> ; b : JFN.  
 Figure 30 : JFN.  
 Figure 31 : [https://www.mun.ca/biology/scarr/Bithorax\\_Drosophila.html](https://www.mun.ca/biology/scarr/Bithorax_Drosophila.html)  
 Figure 32 : Courtesy of Matthew Scott.  
 Figure 33 : a : voir figure 36 ; b : voir <sup>33</sup> (avec permission); d-f : JFN ; composition et annotations de la figure JFN.  
 Figure 34 : JFN.  
 Figure 35 : Photos voir figure 5; annotations JFN.  
 Figure 36 : courtesy of William McGinnis.  
 Figure 37 : voir <sup>28</sup>.  
 Figure 38 : JFN.  
 Figure 39 : JFN.  
 Figure 40: JFN.

### **Annexe 1**

- Figure 1 : JFN.  
 Figure 2 : JFN.  
 Figure 3 : JFN.

### **Annexe 2**

Figures 1 et 2: Molecular biology of the cell, Bruce Alberts et al, Sixth edition. 2015, ISBN 978-0-8153-4432-2

- Figure 3 : voir <sup>38</sup>.

### **Annexe 3**

- Figure 1 : JFN.  
 Figure 2 : a : voir <sup>13</sup> ; b-d : JFN.  
 Figure 3 : a et b : JFN ; c : [https://www.mun.ca/biology/scarr/Pair-Rule\\_Gene\\_expression.html](https://www.mun.ca/biology/scarr/Pair-Rule_Gene_expression.html).  
 Figure 4 : JFN.  
 Figure 5 : a : JFN ; b : JFN, inspiré de <sup>42</sup>.  
 Figure 6 : Photos : voir <sup>33</sup> (avec permission); diagrammes et composition JFN.  
 Figure 7 : JFN.  
 Figure 8 : JFN.

### **Annexe 4**

- Figure 1 : JFN.  
 Figure 2 : a et b voir <sup>44</sup> (open access); c : voir <sup>26</sup> (open access); d : voir <sup>45</sup> (open access); e : JFN.  
 Figure 3 : JFN.  
 Figure 4 : JFN.  
 Figure 5 : JFN.

**Annexe 5**

Figure 1 : Photos voir figure 7 de l'opuscule. Composition JFN.

Figure 2 : Photos voir figure 7 de l'opuscule. Figurés des xenocoelomorphes et de l'ancêtre commun aux néphrozoaires voir <sup>76</sup> ; Composition et annotations JFN.

**Annexe 7**

Figure 1 : composition JFN, dessin des molécules Wikipédia.

Figure 2 : JFN, Modifié de <sup>64</sup>.

**Annexe 8**

Figure 1 : JFN.

Figure 2 : JFN.

Figure 3 : JFN.

Figure 4 : Arbre phylogénétique Nozaki et al. <sup>71</sup> modifié (origine Wikipédia) ; composition et annotations JFN.



## REFERENCES BIBLIOGRAPHIQUES

- 1 Brenner, S. Turing centenary: Life's code script. *Nature* **482**, 461, doi:10.1038/482461a (2012).
- 2 Nicolas, J. F. Elongation in chordates: a view from clonal analysis. (2015).
- 3 Wolpert, L., Tickle, C. and Martinez Arias, A. *Principles of Development*. (Oxford University Press, 2015).
- 4 Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89-93, doi:10.1038/nature16520 (2016).
- 5 Srivastava, M. *et al.* The Trichoplax genome and the nature of placozoans. *Nature* **454**, 955-960, doi:10.1038/nature07191 (2008).
- 6 Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of Xenoturbella and the position of Xenacoelomorpha. *Nature* **530**, 94-97, doi:10.1038/nature16545 (2016).
- 7 Nicolas, J. F., Dubois, P., Jakob, H., Gaillard, J. & Jacob, F. Tératocarcinome de la souris : différenciation en culture d'une lignée de cellules primitives à potentialités multiples. *Ann Microbiol* **126 A**, 3 - 22 (1975).
- 8 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000 (1987).
- 9 Bonnerot, C. & Nicolas, J.-F. in *Methods in Enzymology : Guide to techniques in mouse development* Vol. 225 (ed Academic Press) 451-469 (Wassarman, P. M. and DePamphilis, M. L., 1993).
- 10 Brinster, R. L. *et al.* Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* **27**, 223-231 (1981).
- 11 Nicolas, J. F., Bonnerot, C. & Gasca, S. Le développement embryonnaire de la souris : nouvelles approches génétiques. *Bull Inst Pasteur* **88**, 115-138 (1990).
- 12 Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**, 389-401 (1961).
- 13 Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-356 (1961).
- 14 Alon, U. *An introduction to systems biology : design principles of biological circuits*. (Chapman & Hall/CRC, 2006).
- 15 Gehring, W. J. The evolution of vision. *Wiley interdisciplinary reviews. Developmental biology* **3**, 1-40, doi:10.1002/wdev.96 (2014).
- 16 Quiring, R., Walldorf, U., Kloter, U. & Gehring, W. J. Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. *Science (New York, N.Y.)* **265**, 785-789, doi:10.1126/science.7914031 (1994).
- 17 Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (New York, N.Y.)* **267**, 1788-1792 (1995).
- 18 Czerny, T. *et al.* twin of eyeless, a second Pax-6 gene of Drosophila, acts upstream of eyeless in the control of eye development. *Molecular cell* **3**, 297-307 (1999).
- 19 Gehring, W. J. & Ikeo, K. Pax 6: mastering eye morphogenesis and eye evolution. *Trends in genetics : TIG* **15**, 371-377. (1999).
- 20 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell reports* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).
- 21 Dolgin, E. DNA's secret weapon against knots and tangles. *Nature* **544**, 284-286, doi:10.1038/544284a (2017).
- 22 Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910, doi:10.1016/j.cell.2015.07.038 (2015).

- 23 Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17507-17512, doi:10.1073/pnas.1111941109 (2012).
- 24 Sebe-Pedros, A. *et al.* The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**, 1224-1237, doi:10.1016/j.cell.2016.03.034 (2016).
- 25 Panganiban, G. *et al.* The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5162-5166 (1997).
- 26 Gebelein, B., McKay, D. J. & Mann, R. S. Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature* **431**, 653-659, doi:10.1038/nature02946 (2004).
- 27 Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).
- 28 Deschamps, J. & Duboule, D. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev* **31**, 1406-1416, doi:10.1101/gad.303123.117 (2017).
- 29 McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. A conserved DNA sequence in homeotic genes of Drosophila. Antennapedia and bithorax complexes. *Nature* **308**, 428-433 (1984).
- 30 McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283-302 (1992).
- 31 Stern, C. D. *et al.* Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? *The International journal of developmental biology* **50**, 3-15, doi:052095cs [pii] 10.1387/ijdb.052095cs (2006).
- 32 Greene, B. *L'Univers élégant.* (Robert Laffont, 2000).
- 33 Gould. *La structure de la théorie de l'évolution.* (Gallimard, 2006).
- 34 Kauffman, S. A. *The origins of order. Self-organization and selection in evolution.* (Oxford University Press, 1993).
- 35 Kauffman, S. A. *Humanity in a Creative Universe.* (Oxford University Press, 2016).
- 36 Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795-801 (1980).
- 37 Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. & Small, S. Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. *Development (Cambridge, England)* **129**, 4931-4940 (2002).
- 38 Vincent, B. J., Estrada, J. & DePace, A. H. The appeasement of Doug: a synthetic approach to enhancer biology. *Integrative biology : quantitative biosciences from nano to macro* **8**, 475-484, doi:10.1039/c5ib00321k (2016).
- 39 Vincent, B. J. *et al.* Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in *Drosophila* embryos. *PLoS genetics* **14**, e1007644, doi:10.1371/journal.pgen.1007644 (2018).
- 40 Fang, X. *et al.* Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 10286-10291, doi:10.1073/pnas.1702581114 (2017).
- 41 Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401-410 (1992).
- 42 Nüsslein-Volhard, C. *Coming to life.* (Yale University Press, 2006).
- 43 Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to diversity.* (Blackwell Science, 2001).
- 44 Mbodj, A. *et al.* Qualitative Dynamical Modelling Can Formally Explain Mesoderm Specification and Predict Novel Developmental Phenotypes. *PLoS computational biology* **12**, e1005073, doi:10.1371/journal.pcbi.1005073 (2016).

- 45 Mizutani, C. M., Meyer, N., Roelink, H. & Bier, E. Threshold-dependent BMP-mediated repression: a model for a conserved mechanism that patterns the neuroectoderm. *PLoS biology* **4**, e313, doi:10.1371/journal.pbio.0040313 (2006).
- 46 Suga, H. *et al.* The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nature communications* **4**, 2325, doi:10.1038/ncomms3325 (2013).
- 47 Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome biology* **14**, R15, doi:10.1186/gb-2013-14-2-r15 (2013).
- 48 King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783-788, doi:10.1038/nature06617 (2008).
- 49 Srivastava, M. *et al.* The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720-726, doi:10.1038/nature09201 (2010).
- 50 Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (New York, N.Y.)* **317**, 86-94, doi:10.1126/science.1139158 (2007).
- 51 Sebe-Pedros, A. *et al.* Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nature ecology & evolution* **2**, 1176-1188, doi:10.1038/s41559-018-0575-6 (2018).
- 52 Sebe-Pedros, A., Degnan, B. M. & Ruiz-Trillo, I. The origin of Metazoa: a unicellular perspective. *Nat Rev Genet* **18**, 498-512, doi:10.1038/nrg.2017.21 (2017).
- 53 Gerhart, J. 1998 Warkany lecture: signaling pathways in development. *Teratology* **60**, 226-239, doi:10.1002/(sici)1096-9926(199910)60:4<226::aid-tera7>3.0.co;2-w (1999).
- 54 Genikhovich, G. & Technau, U. On the evolution of bilaterality. *Development (Cambridge, England)* **144**, 3392-3404, doi:10.1242/dev.141507 (2017).
- 55 Adamska, M. *et al.* Wnt and TGF-beta expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. *PloS one* **2**, e1031, doi:10.1371/journal.pone.0001031 (2007).
- 56 Borisenko, I., Adamski, M., Ereskovsky, A. & Adamska, M. Surprisingly rich repertoire of Wnt genes in the demosponge *Halisarca dujardini*. *BMC evolutionary biology* **16**, 123, doi:10.1186/s12862-016-0700-6 (2016).
- 57 Hejnal, A. & Pang, K. Xenacoelomorpha's significance for understanding bilaterian evolution. *Current opinion in genetics & development* **39**, 48-54, doi:10.1016/j.gde.2016.05.019 (2016).
- 58 Bernadskaya, Y. & Christiaen, L. Transcriptional Control of Developmental Cell Behaviors. *Annual review of cell and developmental biology* **32**, 77-101, doi:10.1146/annurev-cellbio-111315-125218 (2016).
- 59 Ramesh, M. A., Malik, S. B. & Logsdon, J. M., Jr. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Current biology : CB* **15**, 185-191, doi:10.1016/j.cub.2005.01.003 (2005).
- 60 Brunet, T. & King, N. The Origin of Animal Multicellularity and Cell Differentiation. *Developmental cell* **43**, 124-140, doi:10.1016/j.devcel.2017.09.016 (2017).
- 61 DuBuc, T. Q., Stephenson, T. B., Rock, A. Q. & Martindale, M. Q. Hox and Wnt pattern the primary body axis of an anthozoan cnidarian before gastrulation. *Nature communications* **9**, 2007, doi:10.1038/s41467-018-04184-x (2018).
- 62 He, S. *et al.* An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*. *Science (New York, N.Y.)* **361**, 1377-1380, doi:10.1126/science.aar8384 (2018).
- 63 Hoffmeyer, T. T. & Burkhardt, P. Choanoflagellate models - *Monosiga brevicollis* and *Salpingoeca rosetta*. *Current opinion in genetics & development* **39**, 42-47, doi:10.1016/j.gde.2016.05.016 (2016).

- 64 Shalaeva, D. N., Galperin, M. Y. & Mulkidjanian, A. Y. Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biology direct* **10**, 63, doi:10.1186/s13062-015-0091-4 (2015).
- 65 Mulkidjanian, A. Y., Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Evolutionary primacy of sodium bioenergetics. *Biology direct* **3**, 13, doi:10.1186/1745-6150-3-13 (2008).
- 66 Lacalli, T. C. Sensory systems in amphioxus: a window on the ancestral chordate condition. *Brain, behavior and evolution* **64**, 148-162, doi:10.1159/000079744 (2004).
- 67 Vopalensky, P. *et al.* Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15383-15388, doi:10.1073/pnas.1207580109 (2012).
- 68 Spudich, J. L. The multitasking microbial sensory rhodopsins. *Trends in microbiology* **14**, 480-487, doi:10.1016/j.tim.2006.09.005 (2006).
- 69 Liegertova, M. *et al.* Cubozoan genome illuminates functional diversification of opsins and photoreceptor evolution. *Scientific reports* **5**, 11885, doi:10.1038/srep11885 (2015).
- 70 Akil, C. & Robinson, R. C. Genomes of Asgard archaea encode profilins that regulate actin. *Nature* **562**, 439-443, doi:10.1038/s41586-018-0548-6 (2018).
- 71 Nozaki, H. *et al.* Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. *Molecular phylogenetics and evolution* **53**, 872-880, doi:10.1016/j.ympev.2009.08.015 (2009).
- 72 Halder, G. *et al.* Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. *Development (Cambridge, England)* **125**, 2181-2191 (1998).
- 73 Gehring, W. J. Historical perspective on the development and evolution of eyes and photoreceptors. *The International journal of developmental biology* **48**, 707-717, doi:10.1387/ijdb.041900wg (2004).
- 74 Gehring, W. J. Chance and necessity in eye evolution. *Genome biology and evolution* **3**, 1053-1066, doi:10.1093/gbe/evr061 (2011).
- 75 Nilsson, D. E. Eye evolution and its functional basis. *Visual neuroscience* **30**, 5-20, doi:10.1017/s0952523813000035 (2013).
- 76 Hejnol, A. & Martindale, M. Q. Acoel development supports a simple planula-like urbilaterian. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **363**, 1493-1501, doi:10.1098/rstb.2007.2239 (2008).