

Chapitre 7. Réseaux et systèmes biologiques

[Jacques van Helden](#) & [Andreas Zanzoni](#) (Aix-Marseille Université)

1. Théorie des graphes en mathématiques
2. Réseaux métaboliques
3. Réseaux de régulation
4. Réseaux d'interactions protéines - protéines (PPI: protein-protein interactions)
5. Réseaux gènes - maladies
6. Propriétés topologiques des réseaux biologiques
7. Conclusions

Théorie des graphes en mathématiques

Graphes mathématiques et réseaux biologiques

En mathématique, le terme **graphe** désigne une représentation formelle d'un ensemble d'entités et de relations entre elles.

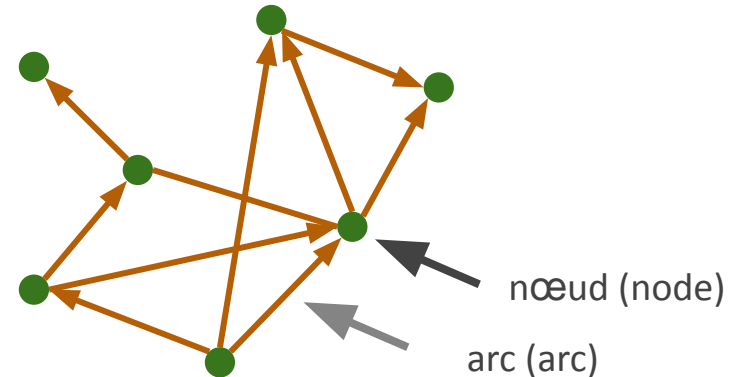
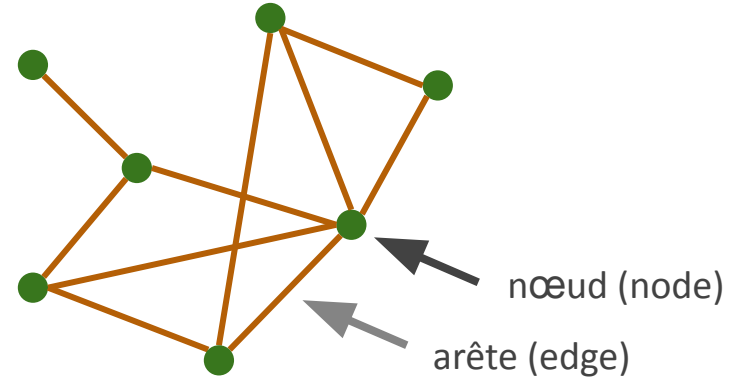
- Les entités sont dénommées **noeuds** du graphe.
- Les relations sont dénommées **arêtes** si elles sont non-orientées, et **arcs** si elles sont orientées

Les mathématiciens ont développé une **théorie des graphes**, qui traite de leurs propriétés en tant qu'objets mathématiques, et permet d'effectuer des opérations :

- Calcul de propriétés topologiques,
- Recherche de chemins
- Extraction de sous-graphes
- ...

Depuis quelques décennies, on a utilisé des graphes mathématiques pour représenter des **réseaux d'interactions** entre entités biologiques et plus particulièrement biomoléculaires :

- Réseaux métaboliques (substrats \rightarrow réactions \rightarrow produits)
- Réseaux de régulation (facteurs transcriptionnels \rightarrow gènes)
- Interactions protéine – protéine
- Co-expression de gènes à partir de données transcriptomiques

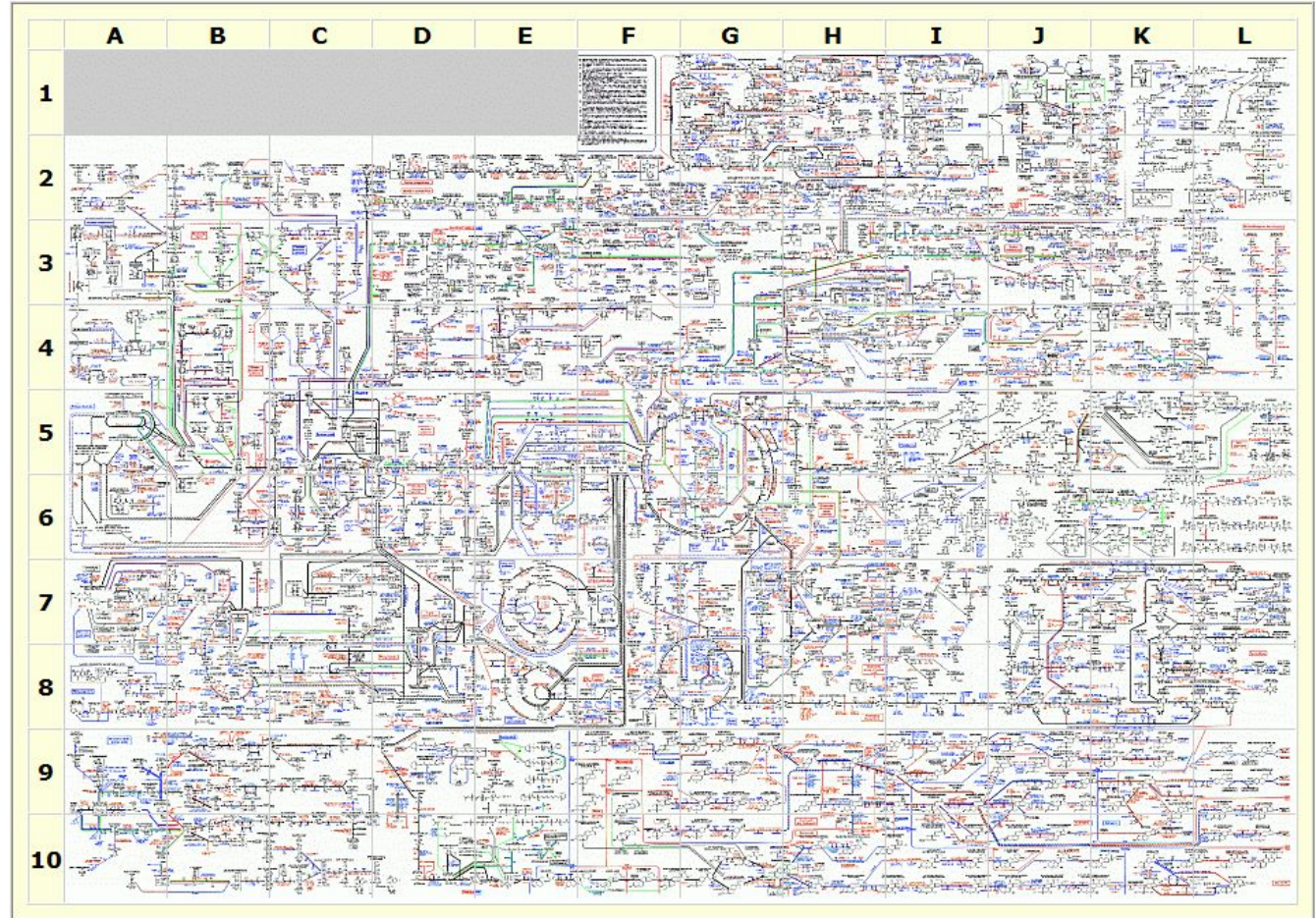


Réseaux métaboliques

Réseau métabolique : Boehringer-Mannheim wallchart

Cette figure donne la vue d'ensemble du poster métabolique (Metabolic Wall Chart), conçu par le biochimiste Gehrard Michal, et distribué depuis les années 1990 par la firme pharmaceutique Boehringer-Mannheim.

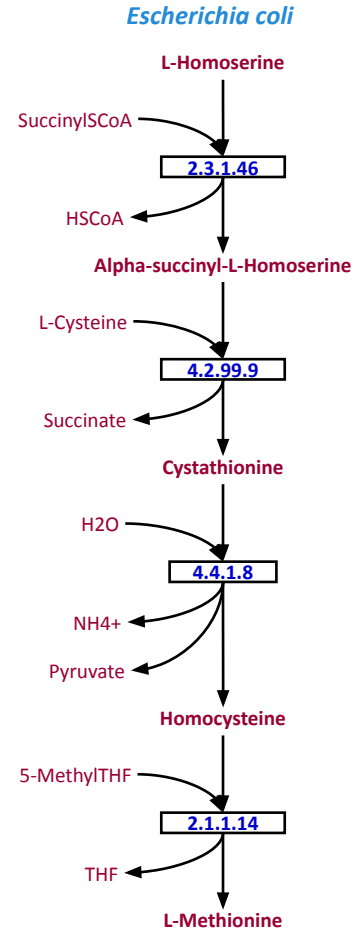
Elle figurait dans le salon ou la cuisine de tous les étudiant de médecine, qui étaient censés mémoriser la plupart de ces voies métaboliques pour passer l'examen de biochimie. Nous allons partir du détail d'une voie particulière et montrer comment on arrive à constituer un réseau métabolique de ce type.



Biosynthèse de la méthionine chez la levure *Saccharomyces cerevisiae* et la bactérie *Escherichia coli*

Biosynthèse de la L-Méthionine à partir de la L-Homosérine

- Bactérie *Escherichia coli*
 - 4 étapes
 - Les nombres associés aux réactions indiquent le type de catalyse enzymatique
 - Transsulfuration : le soufre de la L-Méthionine est transféré à partir de la L-Cystéine



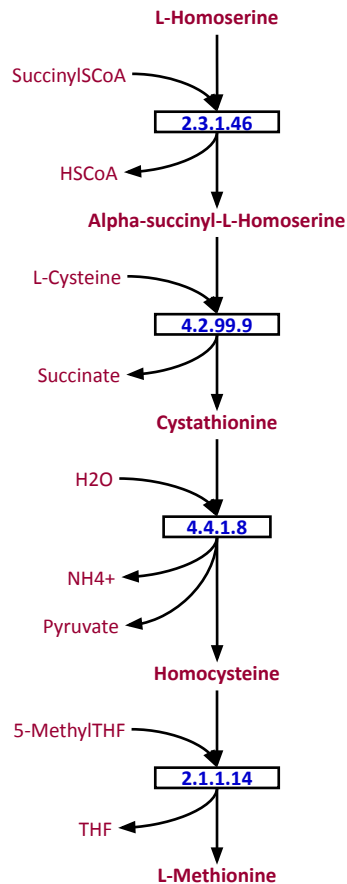
Biosynthèse de la méthionine chez la bactérie *Escherichia coli* et la levure *Saccharomyces cerevisiae*

Deux voies alternatives pour la biosynthèse de la L-méthionine à partir de la L-Homosérine.

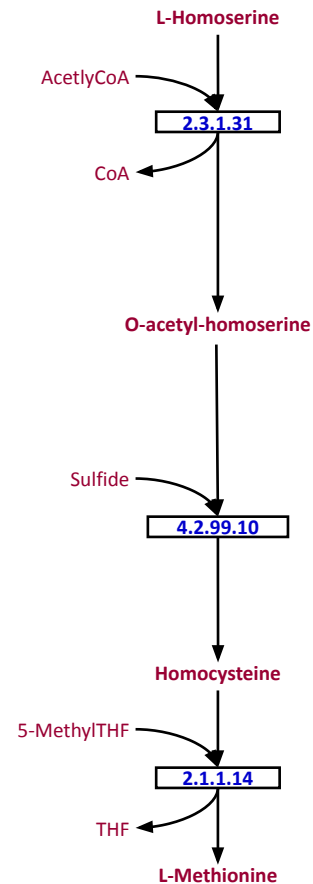
- Métabolisme de la bactérie *Escherichia coli*
 - 4 étapes
 - Les nombres associés aux réactions indiquent le type de catalyse enzymatique
 - Le soufre est initialement incorporé dans la L-Cystéine, et ensuite transféré de la L-Cystéine vers la L-Méthionine (trans-sulfuration C → M)
- Métabolisme de la levure du boulanger *Saccharomyces cerevisiae*
 - 3 étapes
 - Le soufre est initialement incorporé à partir de sulfure dans la L-Méthionine, et ensuite transféré de la L-Méthionine vers la L-Cystéine (trans-sulfuration M → C).

Dans les deux voies métaboliques, le soufre est donc incorporé directement dans un seul des deux acides aminés, et transféré ensuite à l'autre par trans-sulfuration, mais l'incorporation se fait soit dans la L-Méthionine, soit dans la L-Cystéine selon l'organisme.

Biosynthèse de la L-méthionine chez *Escherichia coli*



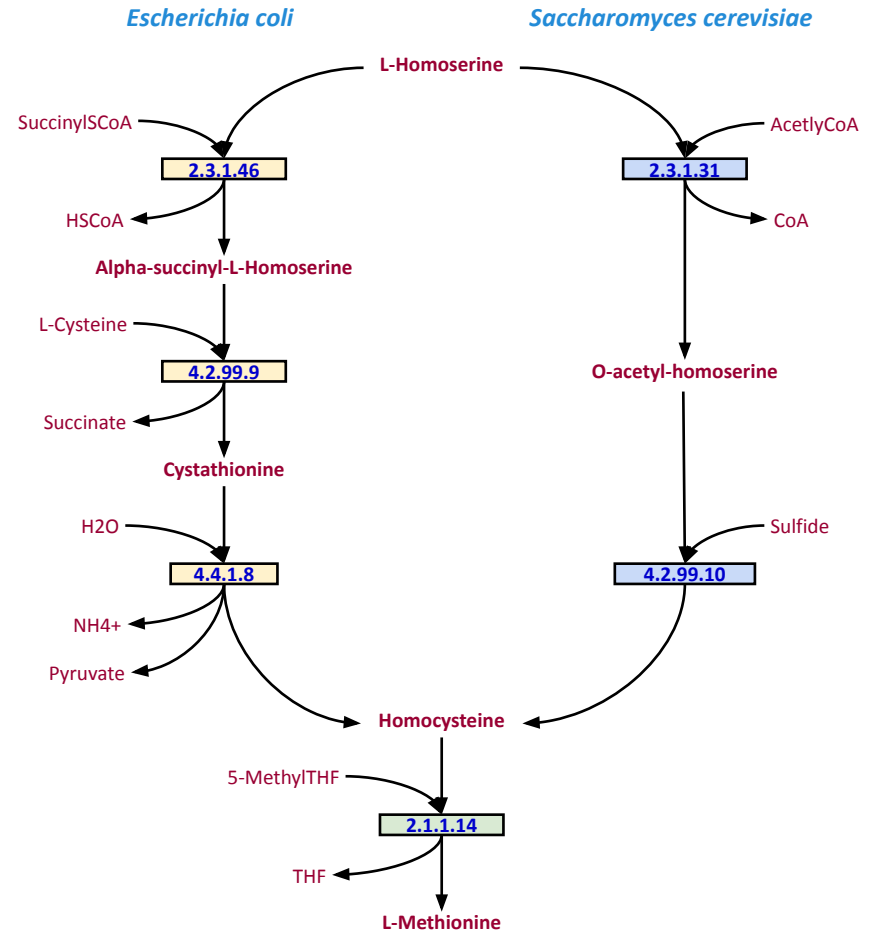
Biosynthèse de la L-méthionine chez *Saccharomyces cerevisiae*



Synthèse sous forme de graphe de la biosynthèse de la méthionine chez deux organismes

Description en termes de graphe / réseau

- Le graphe ci-joint représente une synthèse des deux voies alternatives de biosynthèse de la L-Méthionine (chaque organisme utilise un sous-ensemble des réactions).
 - Jaune** : enzymes présentes chez *E. coli*
 - Bleu** : enzymes présentes chez *S. cerevisiae*
 - Vert** : enzymes présentes chez *E.coli* et *S. cerevisiae*
- Noeuds = entités**, de 2 types pour ce réseau
 - Réactions
 - Métabolites ("compounds" en anglais)
- Arcs** (représentés par des flèches) = **relations**
 - Des substrats aux réactions (flèches entrantes des réactions)
 - Des réactions aux substrats (flèches sortantes)
- Remarque ce réseau est **biparti**
 - Deux types de noeuds (réactions et métabolites)
 - Les arcs vont toujours d'un noeud d'un type à un noeud de l'autre type
 - molécule → réaction
 - réaction → molécule



Réseau métabolique du métabolisme de la cystéine et de la méthionine (KEGG)

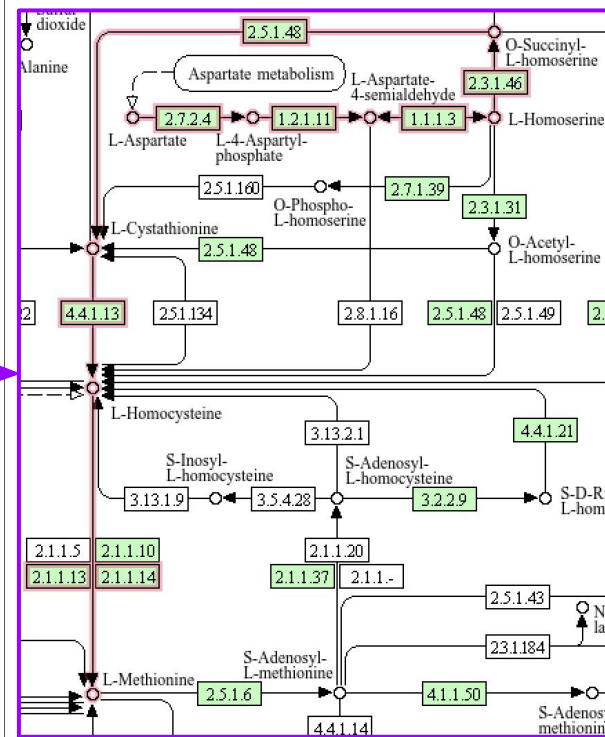
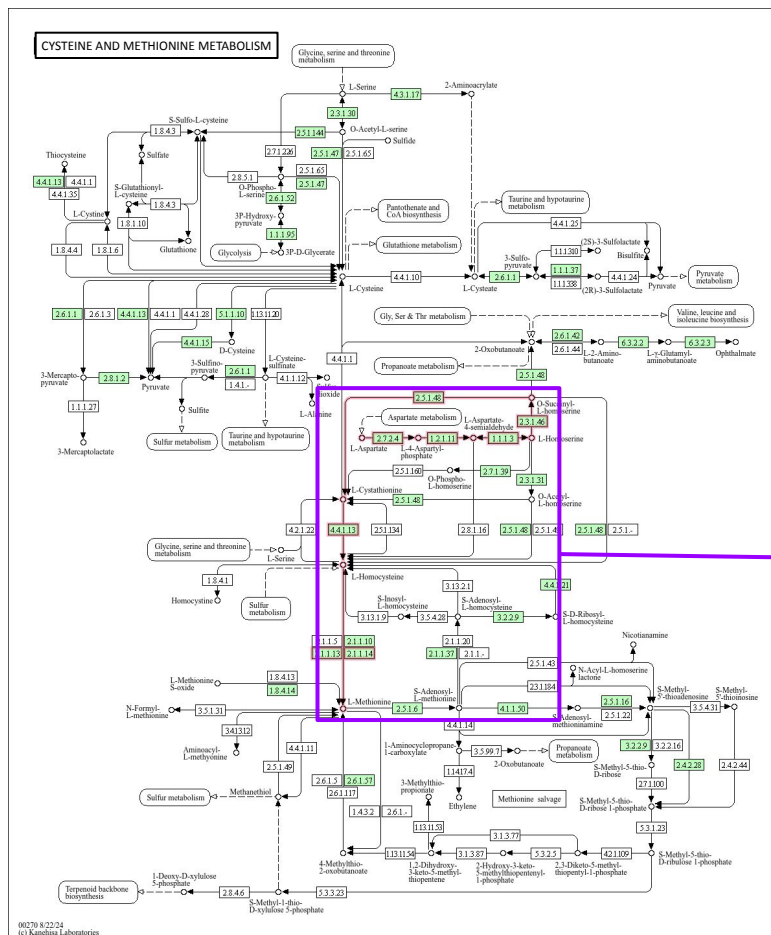
La base de données KEGG répertorie les réactions métaboliques connues chez tous les organismes.

Elle regroupe plusieurs voies en un **réseau métabolique**

Exemple: métabolisme de la cystéine + métabolisme de la méthionine en un seul réseau. "Métabolisme" inclut ici les voies de biosynthèse et de dégradation. Le réseau montre les alternatives chez tous les organismes où ces voies sont connues.

Ces réseaux sont représentés sous forme de **cartes métaboliques**, où l'on peut choisir de marquer les enzymes et voies connues chez un organisme donné.

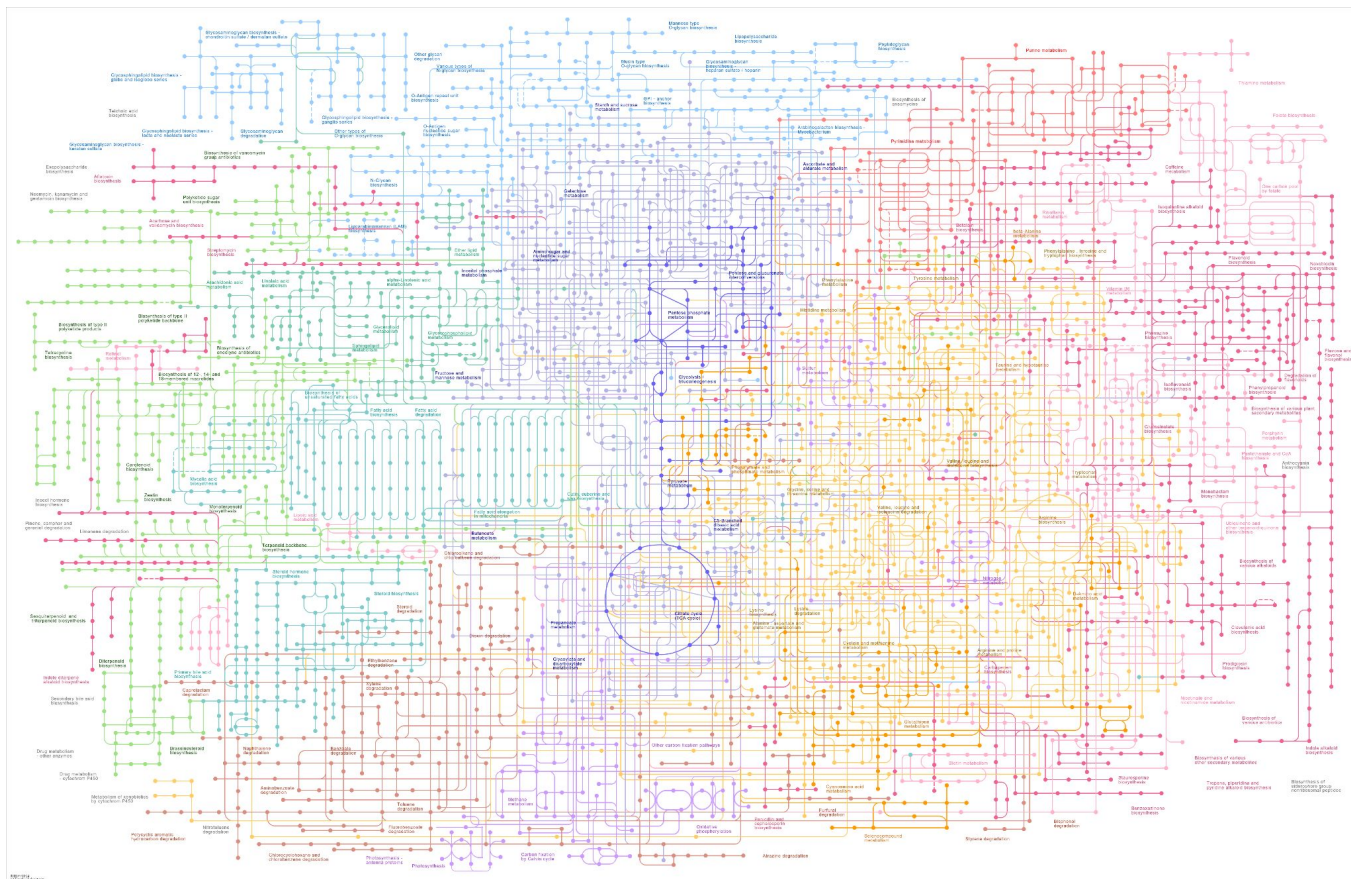
- Vert: enzymes présentes chez *Escherichia coli*.
- Rose: voie métabolique de la L-Aspartate à la L-Méthionine chez *Escherichia coli*.



Réseau métabolique global (toutes les réactions de KEGG)

Composé en agrégeant toutes les réactions connues.

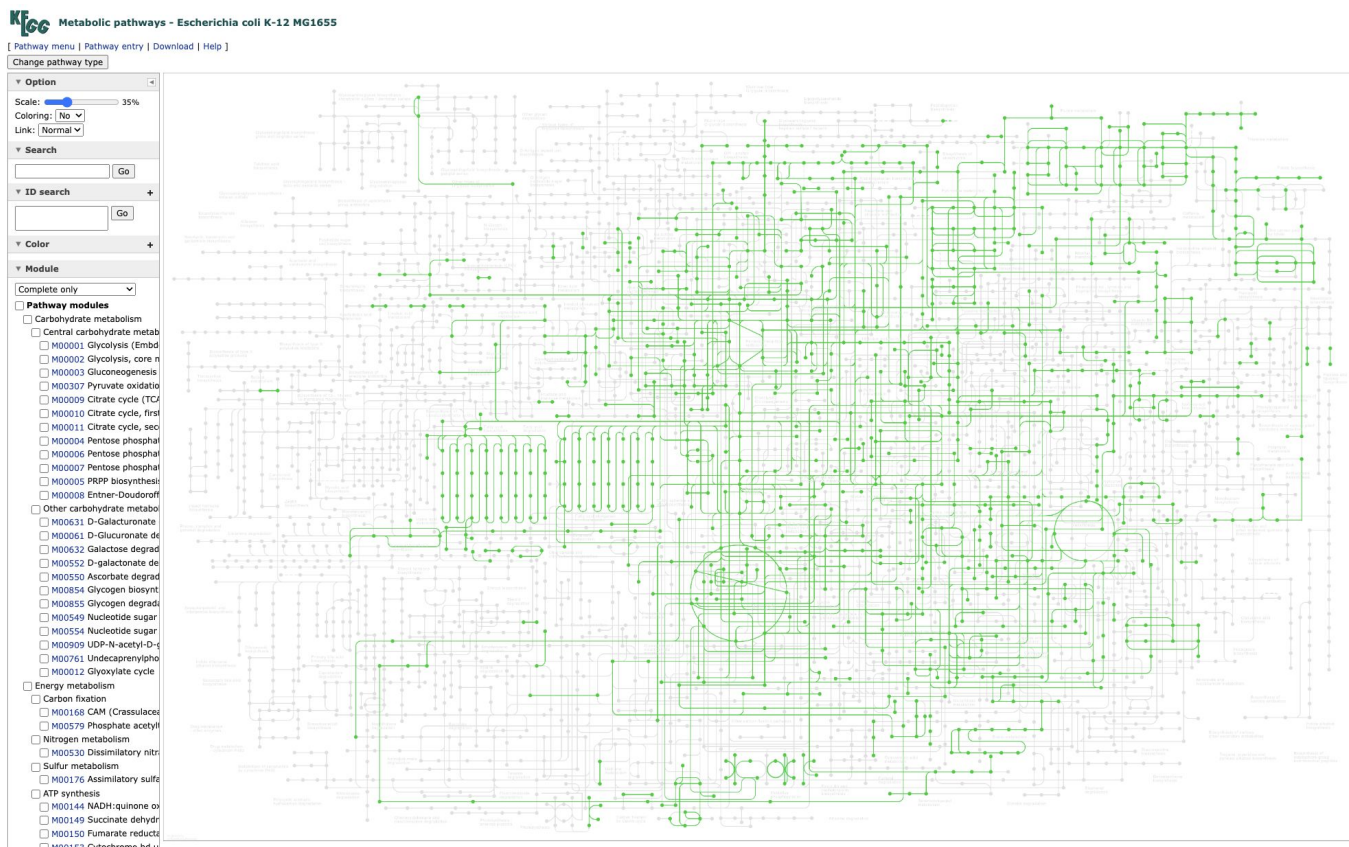
- métabolites (“compounds” en anglais)
- 12.588 réactions métaboliques
- 21.247 compounds (métabolites, petites molécules)



Réseau métabolique global d'*Escherichia coli* (base de données KEGG)

Réseau métabolique global (toutes les réactions de KEGG).

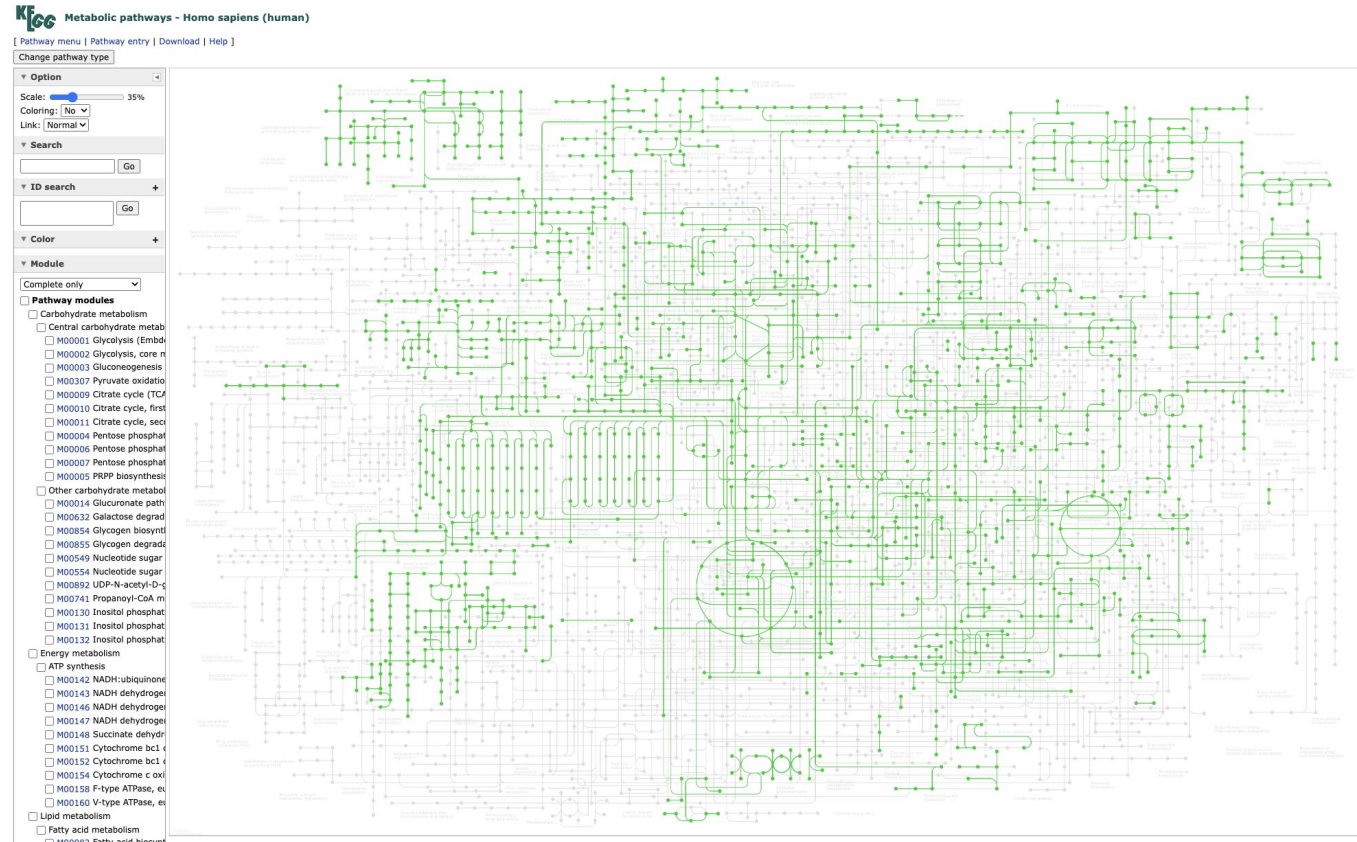
Les réactions pour lesquels on trouve des enzymes dans le génome d'*Escherichia coli* sont colorées.



Réseau métabolique global d'*Homo sapiens*(base de données KEGG)

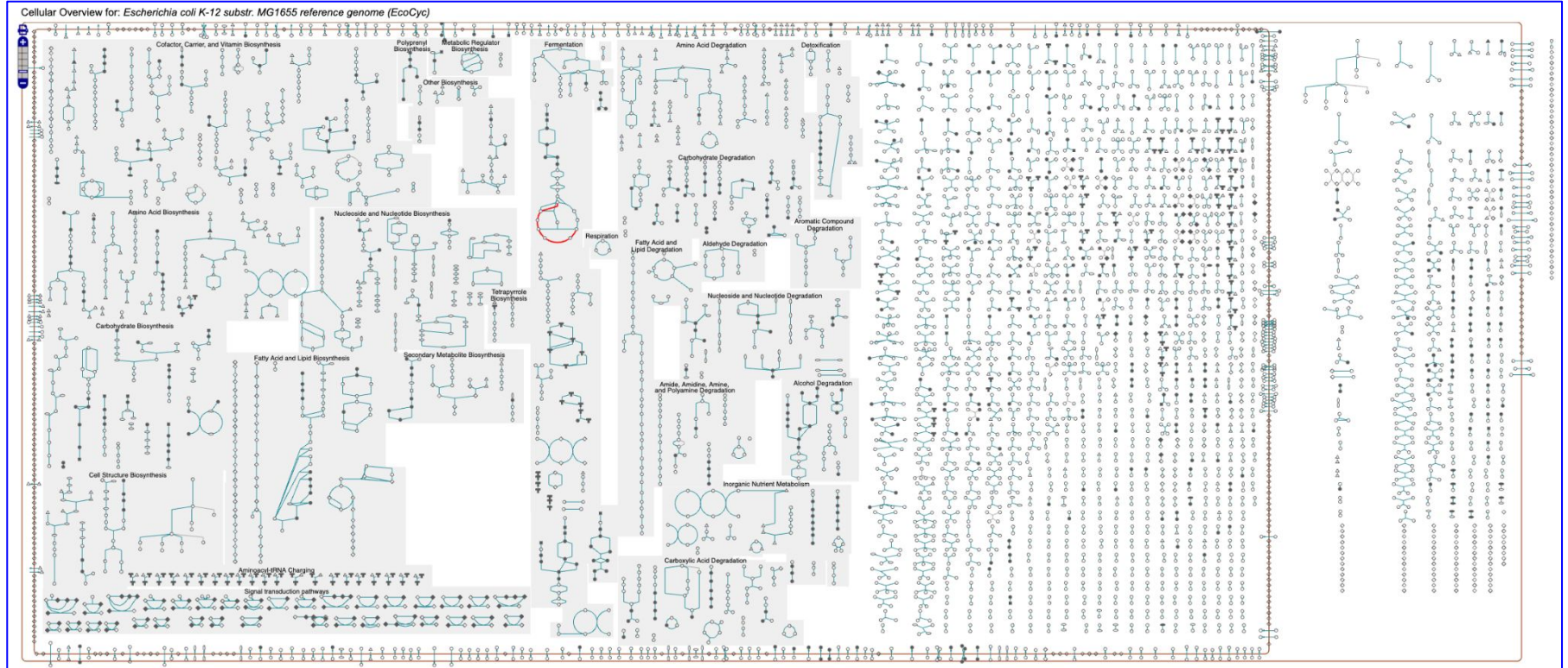
Réseau métabolique global (toutes les réactions de KEGG).

Les réactions pour lesquels on trouve des enzymes dans le génome d'*Homo sapiens* sont colorées.



EcoCyc metabolic chart

La base de données EcoCyc contient des informations détaillées sur les compounds, réactions, voies métaboliques et la régulation de la bactérie *Escherichia coli*.



Conclusion intermédiaire : réseaux métaboliques

- Tout au long du 20^è siècle, les biochimistes ont accumulé des connaissances concernant les réactions métaboliques , voies métaboliques
- Les données accumulées dans des bases de données (KEGG, EcoCyc) peuvent être représentées sous formes de réseaux composés–réactions.
- La vaste majorité de ces réactions sont catalysées par des enzymes.
- La génomique permet d'identifier dans un génome tous les gènes codant pour des enzymes, et de prédire les capacités métaboliques d'un organisme.
- Cependant, les enzymes ne sont pas toutes actives à tout moment : leur activité dépend fortement des conditions cellulaires, et de la présence de métabolites.

Une étude du réseau métabolique d'un organisme doit tenir compte de tous ces aspects.

Réseaux de régulation

Adaptation métabolique - Le shift diauxique

Durant sa thèse de doctorat, Jacques Monod étudie le phénomène de « diauxie »: quand on fait pousser des bactéries sur un mélange de deux sources de carbones (par exemple deux sucres), on observe généralement une croissance en deux phases.

La première phase correspond à la consommation, jusqu'à épuisement, d'une des deux sources (la plus favorable énergétiquement). La seconde phase correspond à la consommation de la seconde source. Le délai entre les deux phases indique une adaptation enzymatique de la cellule au changement de source de carbone.

Durant le reste de sa carrière, Jacques Monod découvrira les éléments-clés de cette adaptation métabolique :

1. Les **perméases** (Cohen & Monod, 1957)
2. La régulation transcriptionnelle et le modèle de l'**opéron** (Jacob & Monod, 1960)
3. La **régulation allostérique** (Monod, Wyman & Changeux, 1965)

ENZYMATIC ADAPTATION AND ALLOSTERIC TRANSITIONS 189

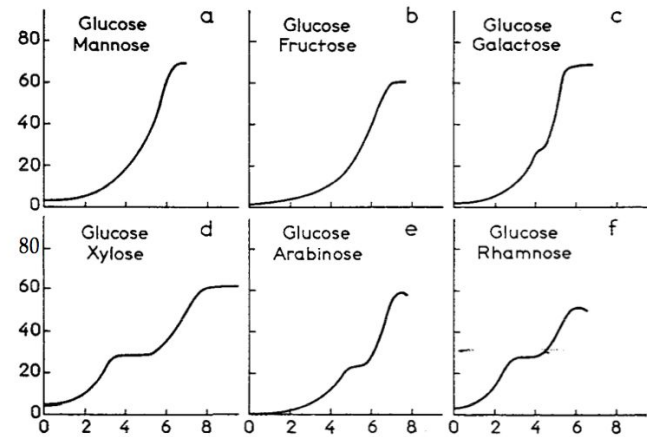


Fig.1. Growth of *E. coli* in the presence of different carbohydrate pairs serving as the only source of carbon in a synthetic medium³⁰.

Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* 12, 88–118 (1965).

[doi.org/10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6)

Cohen, G. N. & Monod, J. BACTERIAL PERMEASES. *Bacteriol Rev* 21, 169–194 (1957). doi.org/10.1128/br.21.3.169-194.1957

Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron : groupe de gènes à expression coordonnée par un opérateur [C. R. Acad. Sci. Paris 250 (1960) 1727–1729].

Republication: *Comptes Rendus Biologies* 328, 514–520 (2005). doi.org/10.1016/j.crvi.2005.04.005

Jacques Monod Nobel lecture. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1965/monod-lecture.html

Régulation de la biosynthèse de la méthionine chez *Escherichia coli*

Les voies métaboliques sont régulées à différents niveaux :

- Transcription (répresseur MetJ)
- Activité enzymatique (régulation allostérique)
- Transport des métabolites

Le réseau ci-contre, formé par les métabolites, enzymes, gènes et facteur transcriptionnel, contient plusieurs **boucles de rétroaction négative**, qui assurent un rétrocontrôle.

concentrations de méthionine et de SAM faibles

→ inactivation du répresseur MetJ

→ biosynthèse des enzymes

→ augmentation de la concentration de méthionine et de SAM

→ activation du répresseur

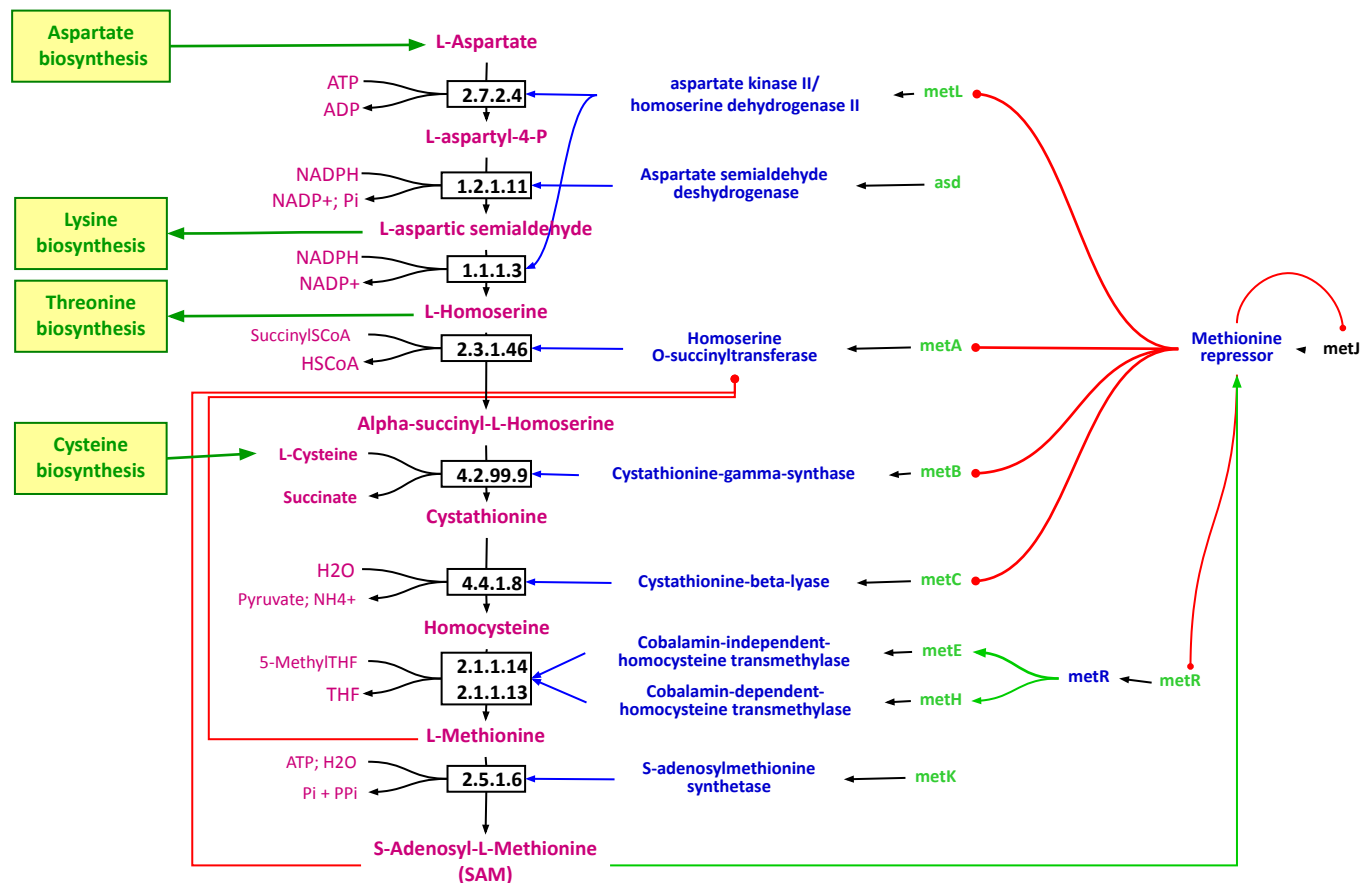
→ répression de la synthèse des enzymes

→ diminution des flux métaboliques

→ concentrations mét + SAM faibles

...

L'ensemble assure l'**homéostasie** : maintien d'une concentration constante de L-Méthionine



Réseau de régulation transcriptionnelle d'*Escherichia coli*

La base de données [RegulonDB](#) qui documente les régulations transcriptionnelles de l'organisme modèle le mieux connu concernant la régulation transcriptionnelle : *Escherichia coli*.

En s'appuyant sur cette base de connaissance, Martinez-Antonio et ses collègues construisent un réseau de régulation.

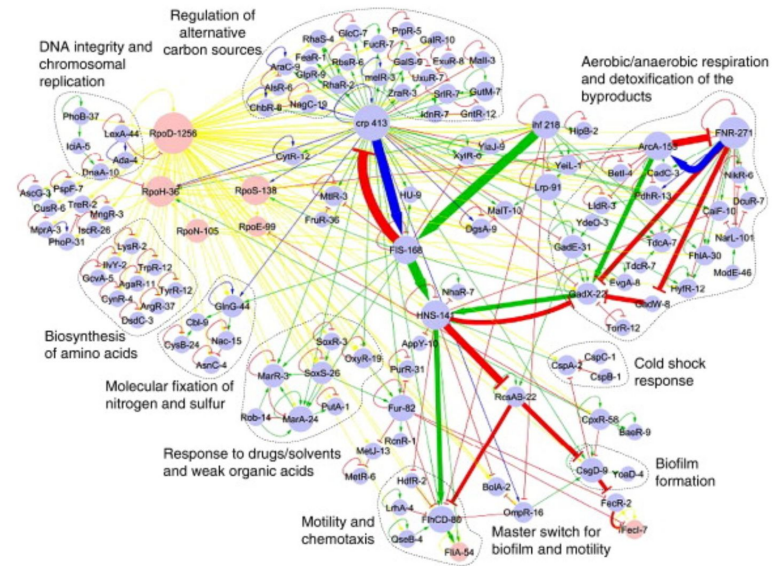
Les arcs indiquent les régulations entre un facteur transcriptionnel et le gène codant pour un autre facteur transcriptionnel.

- Vert: activation
- Rouge: répression
- Bleu: activation ou répression selon contexte

Observations

- L'énorme majorité des facteurs transcriptionnels s'auto-régulent (**rétroaction directe**, le plus souvent **négative**)
- Quelques boucles de rétroaction impliquent 2 ou plusieurs facteurs transcriptionnels (**rétroaction indirecte**)
- Groupements de facteurs transcriptionnels impliqués dans des fonctions particulières : utilisation de sources de carbone, biosynthèse des acides aminés, ...

Les **boucles de rétroaction** sont les éléments essentiels de la régulation des systèmes biologiques.



Core transcriptional regulatory network of *E. coli*. Blue and pink nodes represent genes encoding for TFs and sigma factors, respectively; each node label is accompanied with its connectivity showing the number of regulatory targets. Edges represent cross-regulatory interactions (green for activation, red for repression, blue for dual interactions and yellow for sigma transcription), whereas loops represent transcriptional autoregulations. Specific subnetworks, such as the one associated with the regulation of carbon sources, are delineated with dashed lines to distinguish different regulatory modules. This figure was generated using Cytoscape.¹⁴

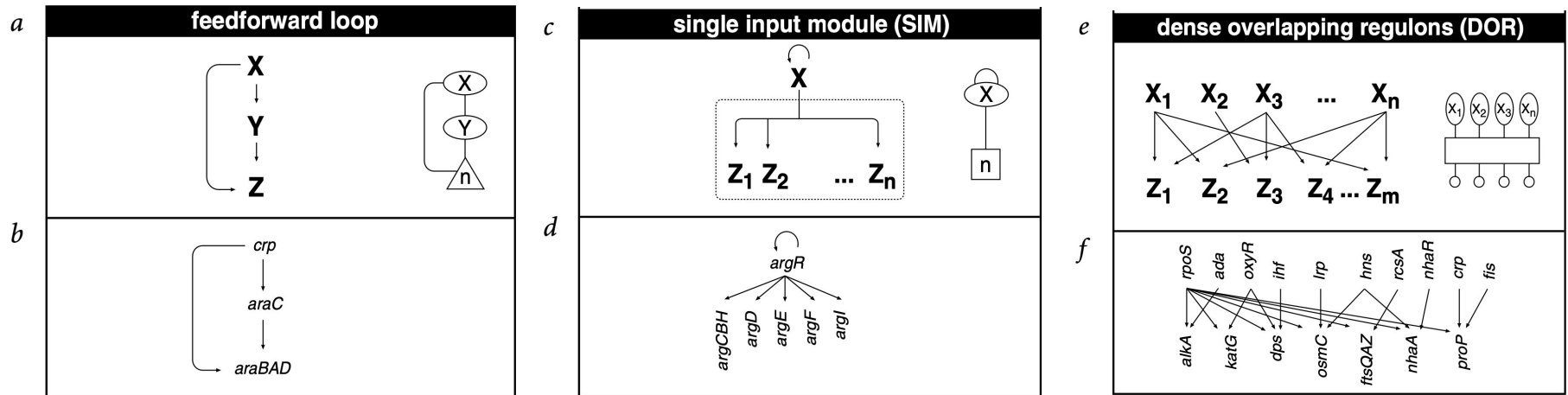
Motifs récurrents dans le réseau de régulation d'*Escherichia coli*

L'équipe de Uri Alon mène une analyse statistique des "motifs récurrents" impliquant au moins trois facteurs transcriptionnels dans le réseau de régulation d'*Escherichia coli* (sur base de RegulonDB). Ils décrivent trois types de motifs récurrents :

- Feedforward loops : X régule Z directement, mais également indirectement via Y
- Single input module (SIM) : un facteur X s'auto-régule et régule une série d'autres facteurs (Z_1, Z_2, \dots, Z_n)
- Dense overlapping regulons : entrecroisements de régulation entre une série de facteurs, et les facteurs qu'ils régulent

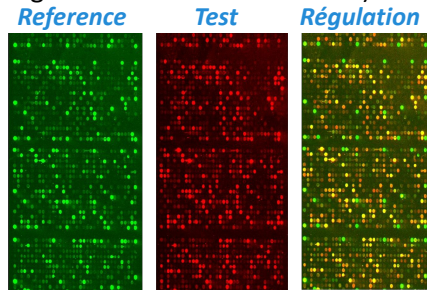
Sur base des propriétés de ce genre de circuits dans les réseaux électronique, il propose que ces motifs récurrents exercent des fonctions similaires dans les réseaux de régulation biologique.

Note : cette analyse de Uri Alon ne tient pas compte de la fréquence très élevée de boucles de rétroaction directes, présentée dans la diapo précédente.



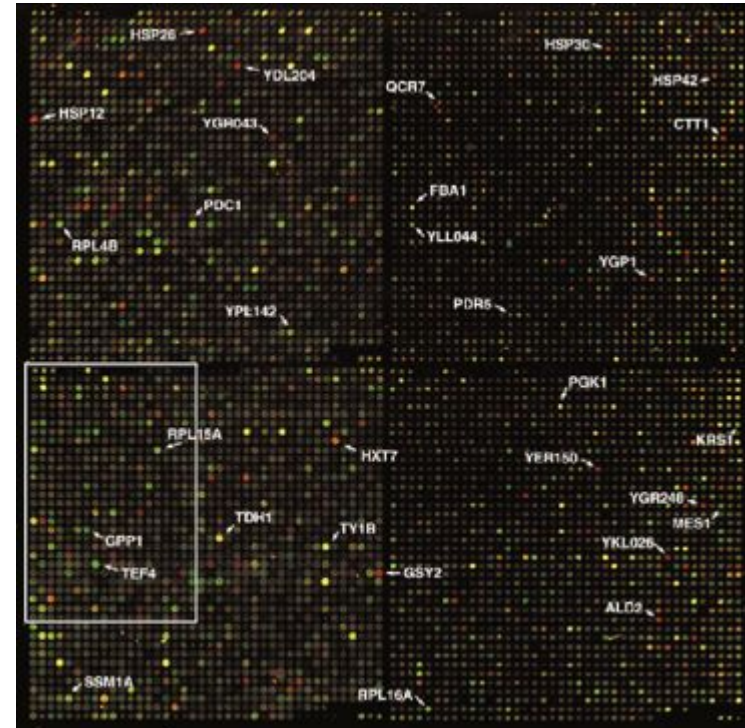
L'avènement de la transcriptomique : les biopuces de Stanford

- 1997, deRisi et collègues développent les premières biopuces transcriptomiques, qui permettent de quantifier la concentration de tous les gènes de levure dans deux conditions données, et de les comparer.
 - Canal vert : échantillon de référence
 - Canal vert : échantillon de test
- Chaque gène correspond à un transcrit
- L'intensité indique la concentration moyenne de ce transcrit dans les deux échantillons comparés.
- La couleur indique le sens de la régulation
 - Rouge: sur-exprimé dans le test par rapport à la référence
 - Vert: sous-exprimé dans le test par rapport à la référence
 - Jaune : non affecté par l'expérience (expression de niveau égal dans les deux conditions)



Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale

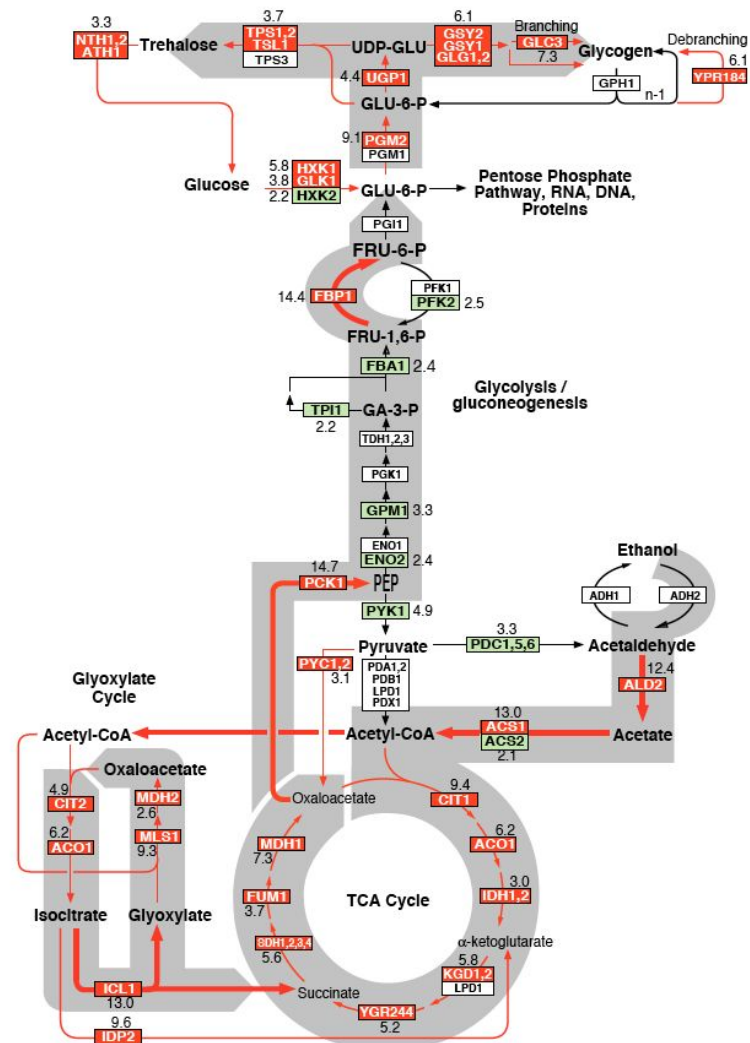
Joseph L. DeRisi, Vishwanath R. Iyer, Patrick O. Brown*



Le transcriptome du shift diauxique

Dans le même article fondateur de la transcriptomique, DeRisi et collègues (1997) mènent une expérience qu'ils appellent "diauxic shift", par référence à l'expérience de Monod. Il s'agit cependant de conditions assez différentes.

- Il s'agit d'un profil temporel du transcriptome, avec 7 points de mesure sur un total de 19h.
- Au départ, les cellules sont cultivées dans un milieu riche en glucose.
- Au fil du temps, les cellules consomment du glucose → lorsque le glucose devient limitant
 - la glycolyse s'arrête
 - la gluconéogenèse est activée pour produire du glucose
- Ce métabolisme produit également des déchets excrétés par les cellules
 - le milieu de culture devient pollué
 - activation des voies de réponse au stress
- Au total, ~1500 gènes montrent des variations d'expression d'un facteur >2 (sur-expression ou sous-expression par rapport au niveau initial).
- Cette expérience démontre donc la puissance de l'approche transcriptomique pour mesurer le niveau de transcription à l'échelle d'un génome (**transcriptome**).

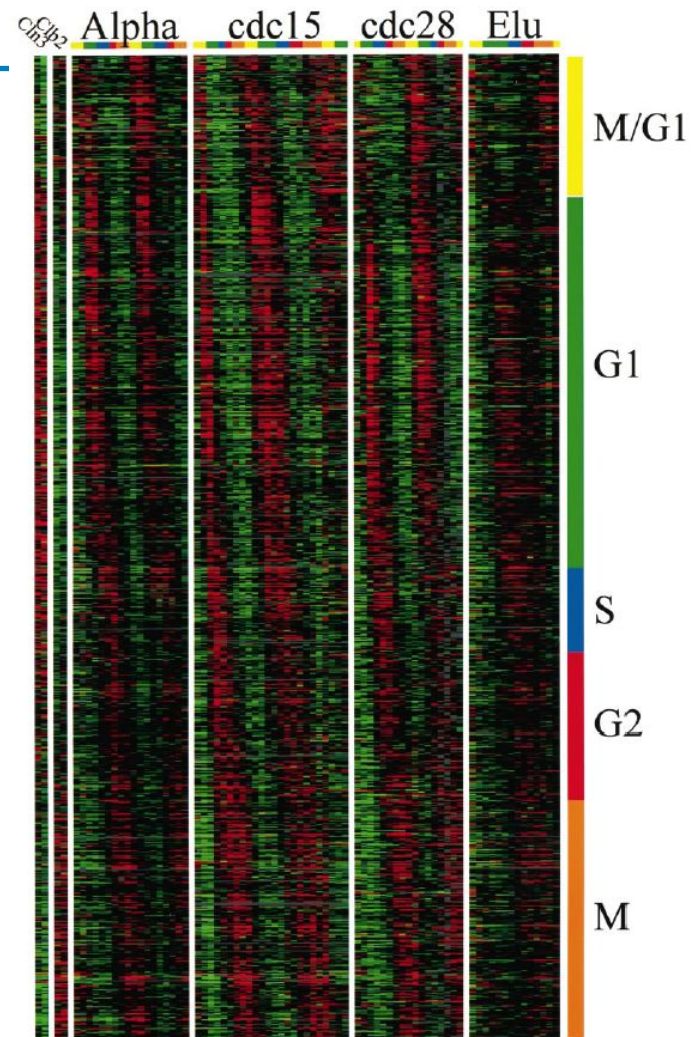


Cycle cellulaire de la levure

Une autre expérience réalisée par le même laboratoire, qui illustre bien la puissance de l'analyse transcriptomique : la détection des gènes de levure montrant des fluctuations périodiques au cours du cycle cellulaire.

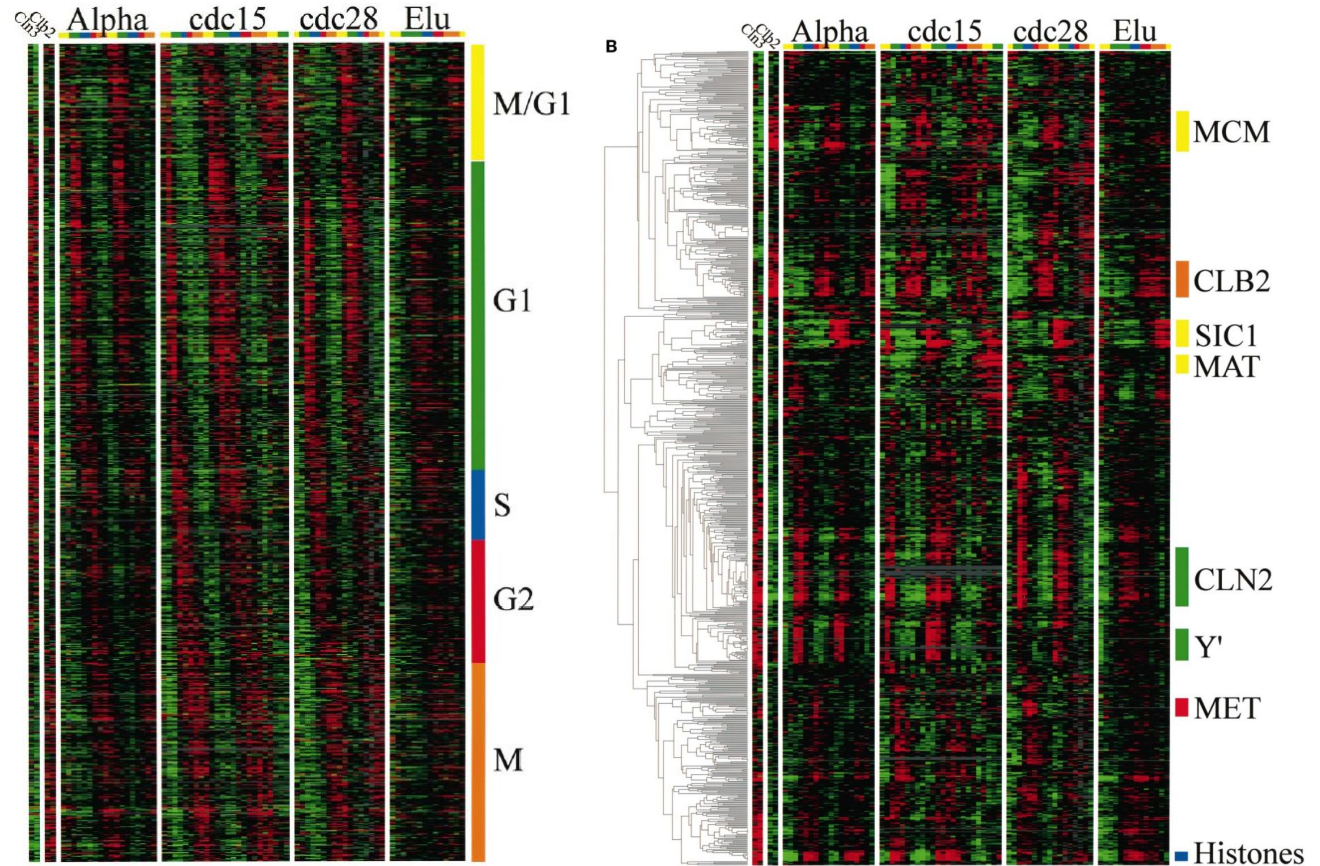
Profils transcriptomiques temporels

- Chaque ligne correspond à un gène
- Le cycle de cellules en culture est synchronisé par 4 méthodes différentes (Alpha, cdc15, cdc28 et elutriation). Au sein de chaque expérience, les colonnes correspondent à des points temporels successifs
- L'échelle de couleur indique le niveau d'expression: faible (vert) à élevé (rouge)
- Les rectangles latéraux indiquent à quelle phase du cycle (M, G1, S, G2) les gènes ont leur expression maximale



Regroupement des gènes en fonction de leurs profils d'expression

Les cartes de couleurs (gauche) peuvent être combinés avec un **algorithme de clustering hiérarchique** pour regrouper les gènes qui ont des profils transcriptomiques similaires (droite). Ceci permet d'identifier des **groupes de gènes co-régulés** par des facteurs transcriptionnels spécifiques (MCM, CLB2, SIC1, CLN2, ...).



Réponse de la levure à différentes conditions de stress

L'équipe de Stanford réalise ensuite une analyse systématique de la réponse transcriptionnelle de la levure à différents types de stress

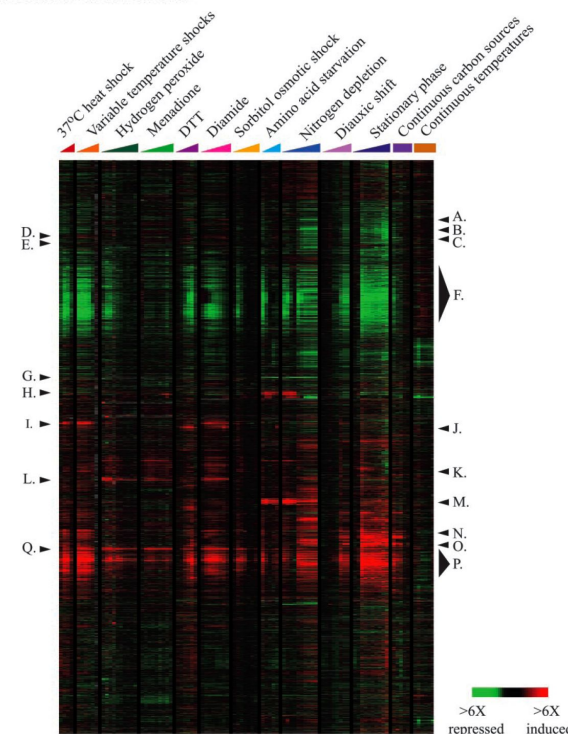
- Stress thermique
- Déplétion d'azote
- Carence en acides aminés
- ...

Ceci permet d'émettre des hypothèses concernant la fonction de gènes inconnus, et de tester ensuite ces hypothèses expérimentalement.

Molecular Biology of the Cell
Vol. 11, 4241–4257, December 2000

Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes^[1]

Audrey P. Gasch,^{*¶} Paul T. Spellman,^{*¶} Camilla M. Kao,^{*¶}
Orna Carmel-Harel,[‡] Michael B. Eisen,[§] Gisela Storz,[‡] David Botstein,[†]
and Patrick O. Brown^{*¶}



Première application médicale (1999) : signatures moléculaires de types de cancers

Golub et al. (1999) utilisent des biopuces transcriptomiques pour quantifier l'expression de 3000 gènes humains dans deux types de leucémie qui affectent des types différents de cellules (lymphoblastes ou myoblastes):

- ALL: acute lymphoblastic leukemia
- AML: acute myoblastic leukemia

Ils sélectionnent les 100 gènes les montrant les plus fortes différences d'expression entre les deux groupes (rouge versus bleu).

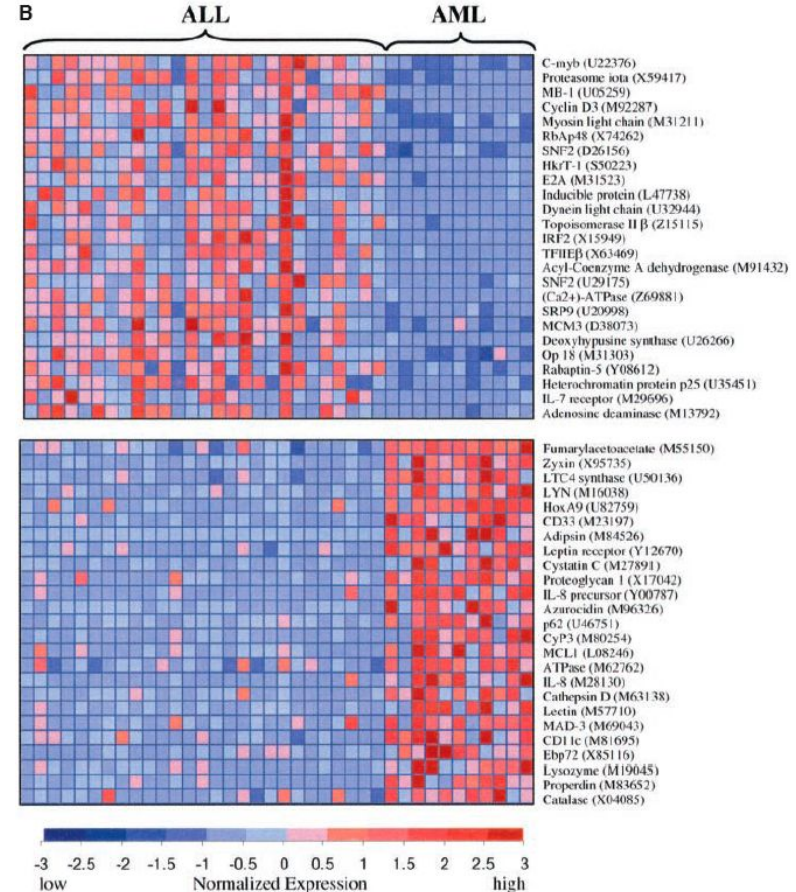
- 50 gènes sur-exprimés chez les patients ALL
- 50 gènes sur-exprimés chez les patients AML

Ils proposent d'utiliser ces gènes comme **signature moléculaire** afin de diagnostiquer les types de leucémies de futurs patients, sur base de leurs **profils transcriptomiques**.

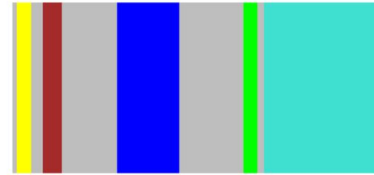
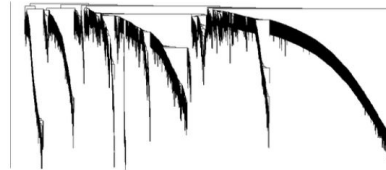
Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

T. R. Golub,^{1,2*} D. K. Slonim,^{1†} P. Tamayo,¹ C. Huard,¹
M. Gaasenbeek,¹ J. P. Mesirov,¹ H. Coller,¹ M. L. Loh,²
J. R. Downing,³ M. A. Caligiuri,⁴ C. D. Bloomfield,⁴
E. S. Lander^{1,5*}

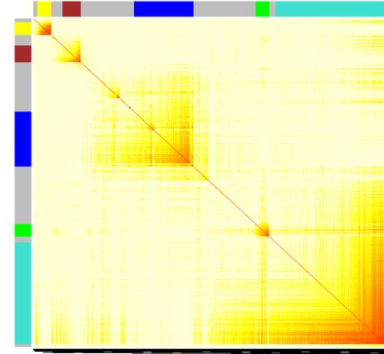
Although cancer classification has improved over the past 30 years, there has been no general approach for identifying new cancer classes (class discovery) or for assigning tumors to known classes (class prediction). Here, a generic approach to cancer classification based on gene expression monitoring by DNA microarrays is described and applied to human acute leukemias as a test case. A class discovery procedure automatically discovered the distinction between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) without previous knowledge of these classes. An automatically derived class predictor was able to determine the class of new leukemia cases. The results demonstrate the feasibility of cancer classification based solely on gene expression monitoring and suggest a general strategy for discovering and predicting cancer classes for other types of cancer, independent of previous biological knowledge.



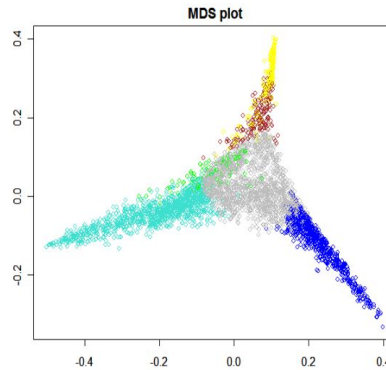
- (a) Clustering des gènes
- (b) Matrice de co-expression (on voit les groupes de gènes co-exprimés comme des rectangles colorés)
- (c) Clusters sur les composantes principales (non expliqué dans ce cours-ci)
- (d) Profils d'expression des groupes de gènes identifiés par cette méthode. Observer la cohérence, mais aussi les différences.



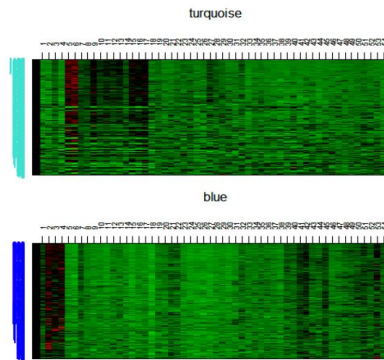
(a)



(b)



(c)



(d)

Conclusion intermédiaire : réseaux de régulation

- La régulation est omniprésente dans les processus biologiques. Elle intervient notamment dans
 - le développement embryonnaire
 - l'adaptation métabolique des cellules et des organismes, qui leur permet de maintenir un environnement interne homogène en dépit des fluctuations temporelles.
 - la réponse cellulaire à des conditions de stress
 - la réponse immunitaire
 - ...
- Depuis le début du 20^e siècle, différentes méthodes ont permis de caractériser les réseaux de régulation à large échelle
 - Transcriptome (biopuces puis séquençage massivement parallèle : RNA-seq)
 - Sites de liaison des facteurs transcriptionnels (ChIP sur chip, CHIP-seq)

Réseaux d'interactions protéines - protéines (**PPI**: protein-protein interactions)

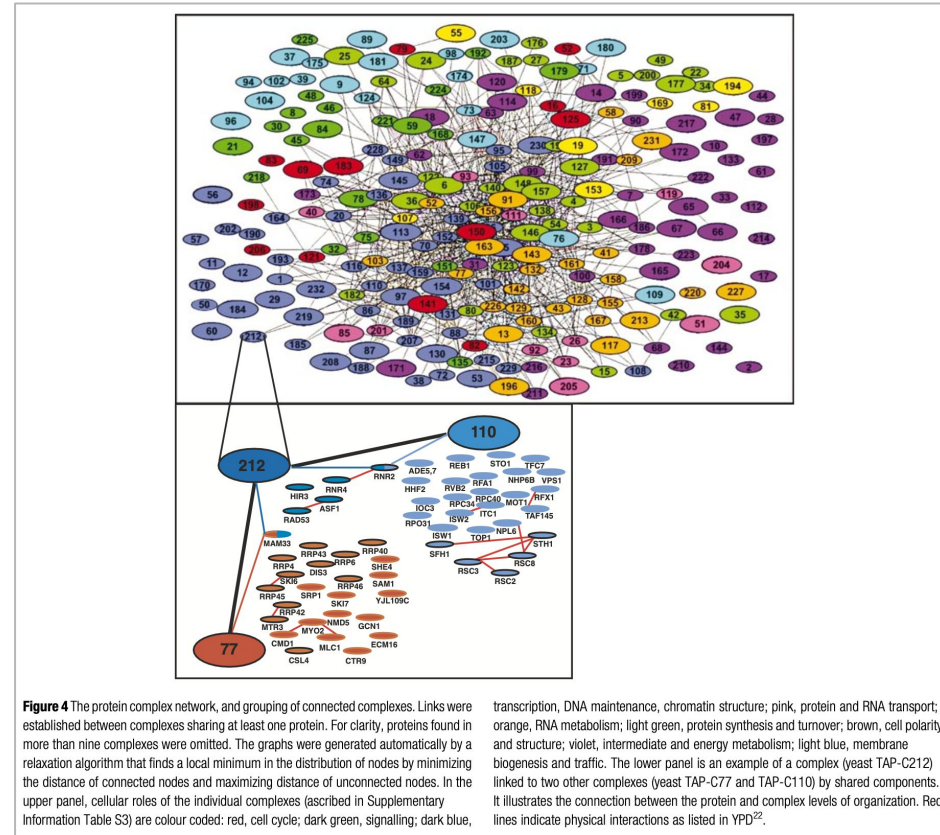
Détection d'interactions protéines-protéines

Deux méthodes développées au début des années 2000

- **Interactions binaires** : la méthode des **double hybrides** (Ito et al., 2001) permet spécifiquement de détecter des interactions entre 2 protéines.
- **Interactions de 1 à n** : approche “sonde - proie” par **spectrométrie de masse** (Gavin et al., 2002). Une protéine est utilisée comme sonde pour “attraper” les protéines qui interagissent avec elle. Cette méthode permet de détecter des interactions multiples à partir d'une même sonde, et donc de révéler les composants des complexes protéiques.

L'automatisation des procédures permet de tester des centaines voire des milliers d'interactions, on parle dès lors d'**interactome**. Cependant, le nombre potentiel d'interactions dépasse de loin la capacité de détection. Par exemple, il existe potentiellement 36.000.000 d'interactions binaires possibles entre les 6.000 protéines de la levure, et 400.000.000 entre les 20.000 protéines humaines.

Les interactomes sont donc partiels.



Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U.S.A. 98, 4569–4574 (2001).

Gavin, A.-C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147 (2002).

Différentes méthodes permettent de générer des données massives concernant les interactions protéines - protéines

Méthodes expérimentales

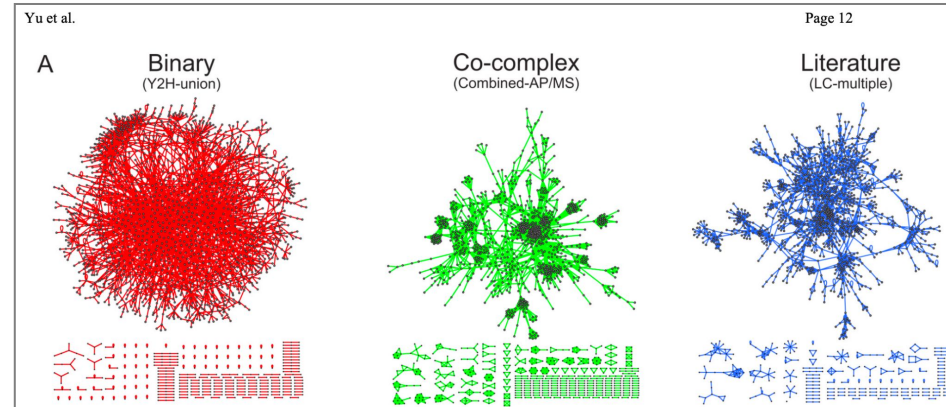
- La méthode des **doublets hybrides** détecte des interactions binaires entre deux protéines. Elle permet, indirectement, de retrouver des interactions multiples (complexes protéiques) en combinant les interactions binaires entre les différentes protéines qui appartiennent à un même complexe.
- Une autre approche consiste à “pêcher” l’ensemble des protéines liées à une protéine “hameçon”, et à les identifier par **spectrométrie de masse** (ceci suppose qu’on dispose de l’ensemble des séquences protéiques pour l’organisme considéré).

Fouille de la littérature scientifique

- Des travaux ont également été menés pour identifier les informations pertinentes concernant les interactions protéines - protéines par “**fouille de texte**” dans la littérature scientifique, en identifiant automatiquement les noms de protéines et des phrases qui indiquent une interaction.

High-Quality Binary Protein Interaction Map of the Yeast Interactome Network

Haiyuan Yu,^{1,2*} Pascal Braun,^{1,2*} Muhammed A. Yildirim,^{1,2,3*} Irma Lemmens,⁴ Kavitha Venkatesan,^{1,2} Julie Sahalie,^{1,2} Tomoko Hirozane-Kishikawa,^{1,2} Fana Gebreab,^{1,2} Na Li,^{1,2} Nicolas Simonis,^{1,2} Tong Hao,^{1,2} Jean-François Rual,^{1,2} Amélie Dricot,^{1,2} Alexei Vazquez,⁵ Ryan R. Murray,^{1,2} Christophe Simon,^{1,2} Leah Tardivo,^{1,2} Stanley Tam,^{1,2} Nenad Svrzikapa,^{1,2} Changyu Fan,^{1,2} Anne-Sophie de Smet,⁴ Adriana Motyl,⁶ Michael E. Hudson,⁶ Juyong Park,^{1,7} Xiaofeng Xin,⁸ Michael E. Cusick,^{1,2} Troy Moore,⁹ Charlie Boone,⁸ Michael Snyder,⁶ Frederick P. Roth,^{1,10} Albert-László Barabási,^{1,7} Jan Tavernier,⁴ David E. Hill,^{1,2} Marc Vidal^{1,2,†}



Réseaux d'interactions protéines-protéines

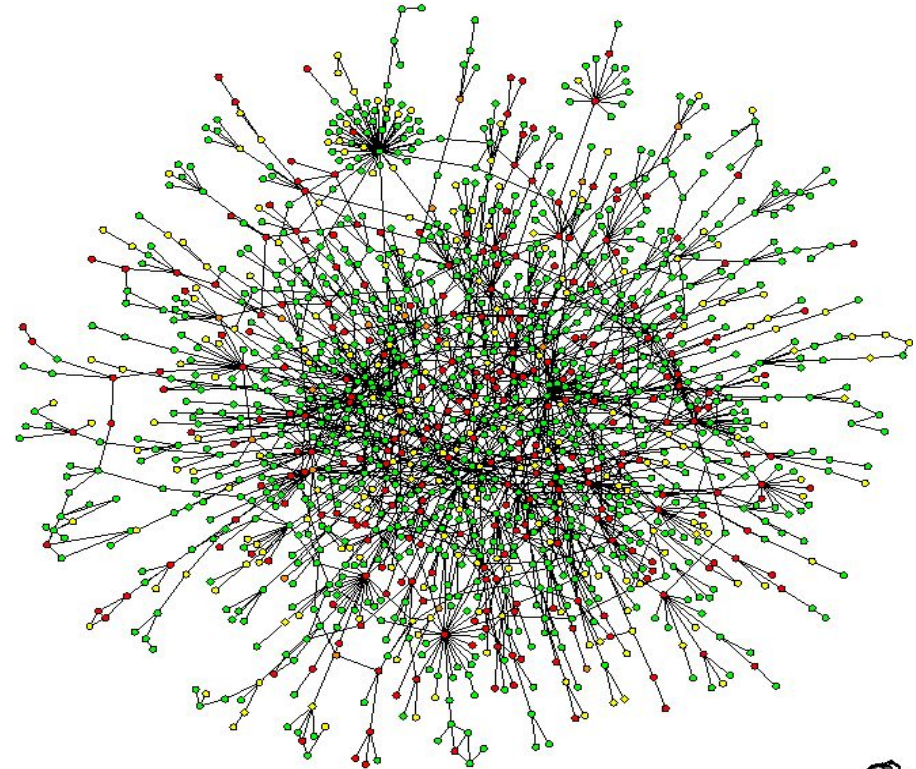
On peut représenter l'interactome sous forme d'un graphe d'interactions protéine - protéine. Les résultats donnent de gigantesques réseaux d'interactions, comprenant des centaines de noeuds (protéines) reliées par des milliers d'arêtes (interactions).

La fiabilité de ces données est cependant relative. Le taux de recouvrement entre les études indépendantes est relativement faible

- Chaque méthode retourne un certain nombre de **faux-positifs** (interactions détectées mais non-existantes)
- Un nombre d'interactions existantes échappent à la détection (**faux-négatifs**).
- Pour des raisons techniques et de moyens, ces études ne testent qu'un sous-ensemble des interactions possibles → **couverture partielle**.

Ces études apportent néanmoins des informations exploitables

- Détection de groupes de protéines interconnectées → **modules du réseau**
- Hypothèses concernant les fonctions possibles de protéines inconnues → principe d'**annotation par association** (métaphore "**coupable par association**")



Conclusion intermédiaire : réseaux d'interactions protéine-protéine

- La régulation est omniprésente dans les processus biologiques. Elle intervient notamment dans
 - le développement embryonnaire
 - l'adaptation métabolique des cellules et des organismes, qui leur permet de maintenir un environnement interne homogène en dépit des fluctuations temporelles.
 - la réponse cellulaire à des conditions de stress
 - la réponse immunitaire
 - ...
- Depuis le début du 20^e siècle, différentes méthodes ont permis de caractériser les réseaux de régulation à large échelle
 - Transcriptome (biopuces puis séquençage massivement parallèle : RNA-seq)
 - Sites de liaison des facteurs transcriptionnels (ChIP sur chip, CHIP-seq)

Réseaux gènes - maladies

Centre: Réseau maladies – gènes (“diseasome”)

- : chaque flèche indique qu’un gène est associé à une maladie (par exemple via des GWAS)
- Ce graphe est biparti : 2 types de noeuds (gène ou maladie) et les relations lient toujours un type à l’autre.

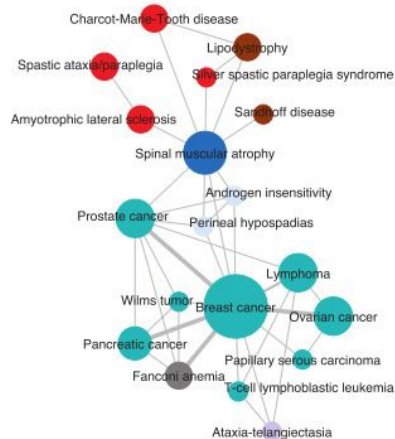
Gauche : réseau de co-morbidité

- Graphe maladie - maladie
- Dérivé du diseasome en liant 2 maladies si elles ont au moins un gène commun dans le diseasome.
- L’épaisseur des arêtes indique le nombre de gènes en commun

Droite: réseau de gènes impliqués dans les mêmes maladies.

- Deux gènes sont liés s’ils sont associés à une ou plusieurs mêmes maladies.
- L’épaisseur des arêtes indique le nombre de maladie auxquelles ils sont co-associés.

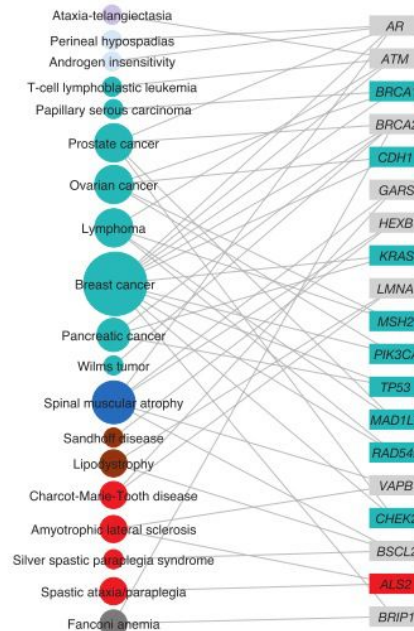
Human Disease Network (HDN)



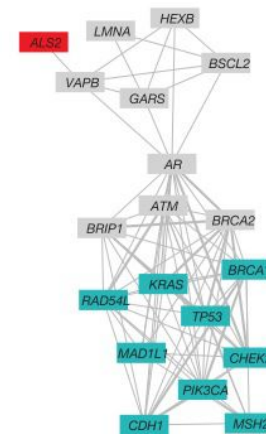
DISEASOME

disease phenotype

disease genome



Disease Gene Network (DGN)



Propriétés topologiques des réseaux biologiques

Graphes mathématiques et réseaux biologiques

En mathématique, le terme **graphe** désigne une représentation formelle d'un ensemble d'entités et de relations entre elles.

- Les entités sont dénommées **noeuds** du graphe.
- Les relations sont dénommées **arêtes** si elles sont non-orientées, et **arcs** si elles sont orientées

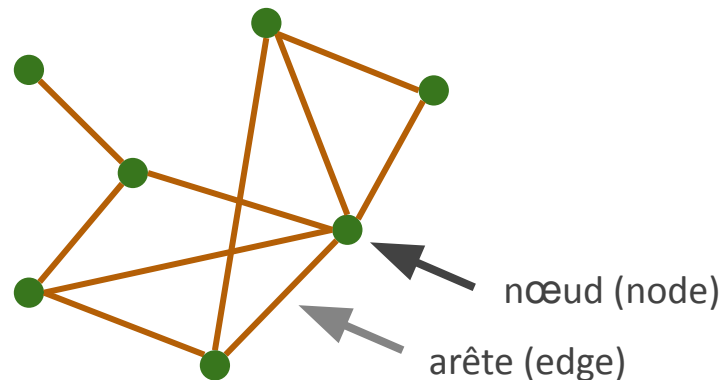
Les mathématiciens ont développé une **théorie des graphes**, qui traite de leurs propriétés en tant qu'objets mathématiques, et permet d'effectuer des opérations :

- Calcul de propriétés topologiques,
- Recherche de chemins
- Extraction de sous-graphes
- ...

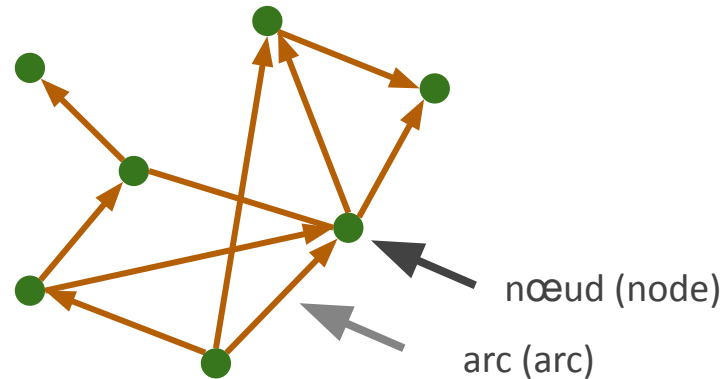
Depuis quelques décennies, on a utilisé des graphes mathématiques pour représenter des **réseaux d'interactions** entre entités biologiques et plus particulièrement biomoléculaires :

- Réseaux métaboliques (substrats \rightarrow réactions \rightarrow produits)
- Réseaux de régulation (facteurs transcriptionnels \rightarrow gènes)
- Interactions protéine – protéine
- Co-expression de gènes à partir de données transcriptomiques

Graphe non orienté



Graphe orienté



Propriétés topologiques

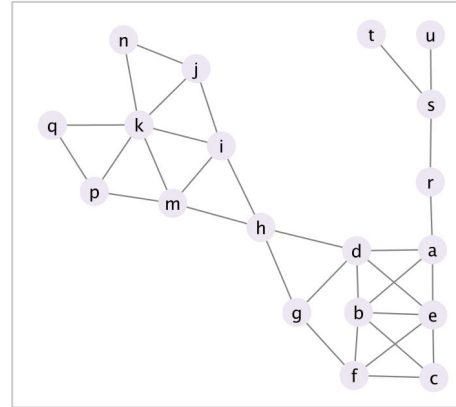
Les propriétés topologiques d'un graphe décrivent sa structure et son organisation :

Degré : le degré (*degree*) k d'un nœud est le nombre d'arêtes qui lui sont adjacentes (connectées). Il correspond donc au nombre de connexions ou nombre de premiers voisins.

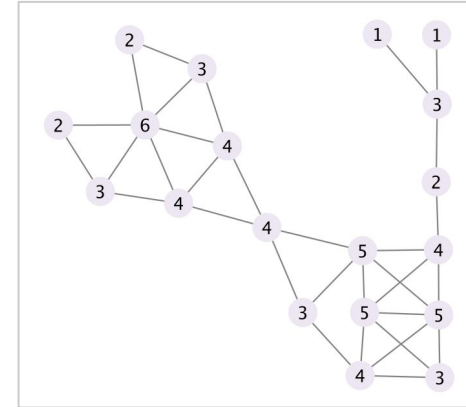
Chemin : Un chemin (*path*) dans un graphe est une suite de nœuds connectés par des arêtes, qui permet de passer d'un nœud de départ à un nœud d'arrivée en suivant les connexions existantes. Dans un chemin, chaque nœud ne peut être parcouru qu'une seule fois.

Le plus court chemin : Le plus court chemin (*shortest path*) entre deux nœuds d'un graphe est le chemin reliant ces deux nœuds en passant par le moins d'arêtes possibles. La *distance* entre deux nœuds est le nombre d'arêtes comprises dans le plus court chemin.

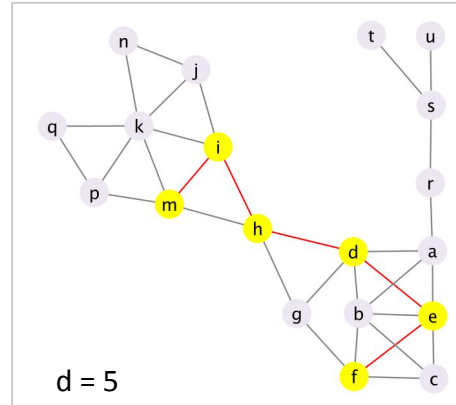
Exemple de réseau



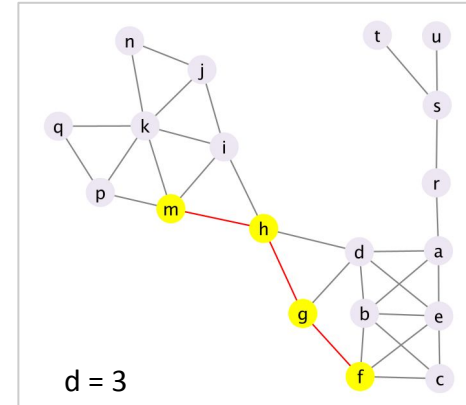
Degré de chaque nœud



Un des chemins possibles de m à f



Chemin le plus court de m à f



Propriétés topologiques

Centralité (*centrality*) : la centralité mesure l'importance d'un nœud au sein d'un réseau. Il existe plusieurs types de centralité, dont :

- **Centralité de distance** (*distance centrality*) : distance moyenne entre chaque nœud et tous les autres nœuds. Il s'agit d'une mesure globale de la centralité (dépend de l'ensemble du réseau).
- **Centralité de degré** (*degree centrality*) : la centralité de degré C_D d'un nœud v est égale à son degré. Il s'agit d'une centralité *locale* (ne dépend que du voisinage immédiat de chaque nœud).

$$C_D(v) = \text{deg}(v)$$

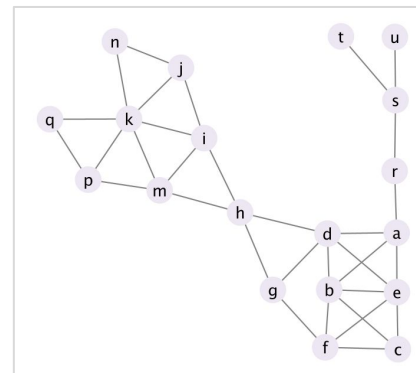
- **Centralité d'intermédiarité** (*betweenness centrality*) : "fréquence de passage" ; fréquence à laquelle un nœud v se trouve sur le plus court chemin entre deux autres nœuds (indices s et t). Il s'agit d'une mesure globale (dépend de l'ensemble du réseau)

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

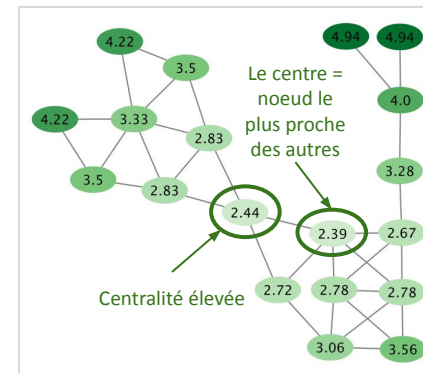
- v nœud dont on veut calculer la centralité
 - s, t indices qui énumèrent toutes les paires de nœuds distincts de v
 - σ_{st} nombre total de plus courts chemins entre les nœuds s et t
 - $\sigma_{st}(v)$ nombre de ces chemins qui passent par le nœud v
- Note : dans certains cas il peut y avoir plusieurs chemins de même taille entre deux nœuds s et t , dont certains passent par v et d'autres pas

Le centre d'un réseau dépend du choix de la mesure de centralité (exemple ci-contre).

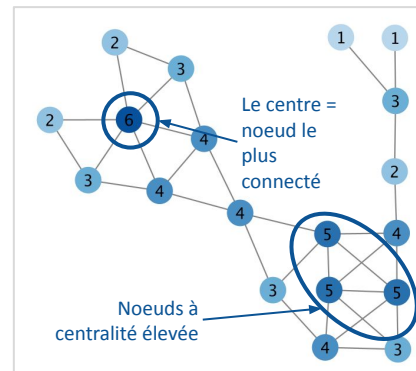
Exemple de réseau



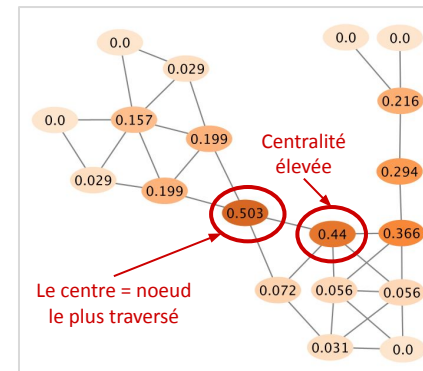
Centralité de distance



Centralité de degré



Centralité d'intermédiarité



Propriétés topologiques

Coefficient de regroupement (clustering coefficient) : pour un nœud i , le coefficient de regroupement C_i indique la proportion des paires de voisins qui sont également connectées entre elles :

$$C_i = \frac{2E_i}{k_i(k_i-1)}$$

E_i = nombre de connexions entre les voisins de i

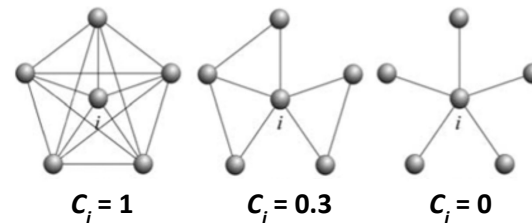
k_i = degré de i (i.e., nombre de voisins de i)

$k_i(k_i-1)/2$ = nombre d'interactions a priori possibles entre k_i nœuds

Cette mesure permet d'évaluer la densité locale (i.e., le voisinage du nœud i) des connexions dans un réseau et d'identifier des groupes de nœuds interconnectés.

Il est possible d'estimer la densité globale d'un réseau grâce à la moyenne des coefficients de regroupement de tous les nœuds (*average clustering coefficient*) :

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

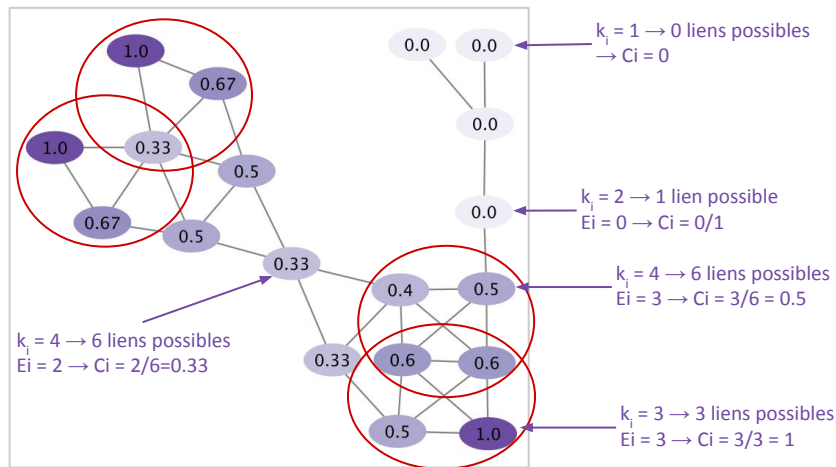


On appelle **clique**, ou **sous-graphe complet**, un sous-graphe dont les nœuds sont tous directement interconnectés.

Quand $C_i = 1$, le nœud i et tous ses voisins forment une **clique**.

Note : l'inverse n'est pas vrai – une clique peut comporter des nœuds de coefficient de clustering inférieur à 1 (exemple ci-dessous).

Coefficients de clustering et quelques cliques



Annotation par association (métaphore "coupable par association")

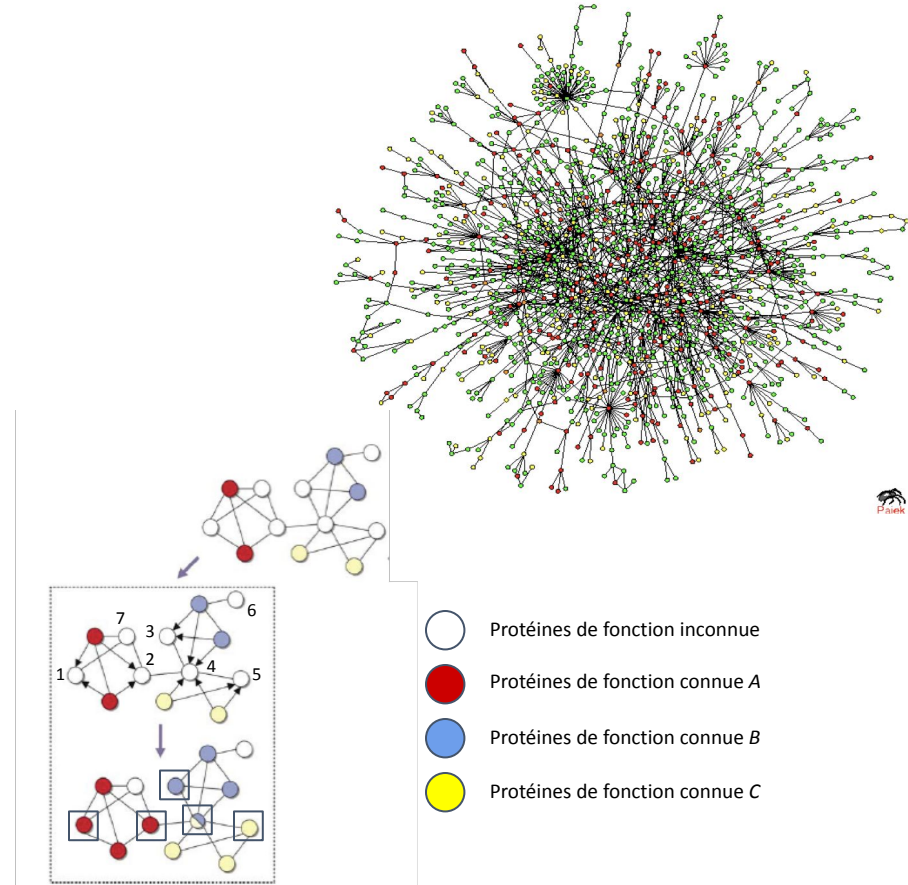
L'étude des propriétés topologiques de l'interactome nous permet d'aborder de nombreuses questions biologiques et de formuler des hypothèses concernant les fonctions possibles de protéines.

Par exemple, si une protéine de fonction inconnue est observée dans un réseau avec des protéines bien caractérisées (ayant des rôles spécifiques dans le métabolisme, la signalisation, etc.), il est probable qu'elle partage une fonction similaire ou liée.

Ce principe peut être utilisé pour identifier des nouvelles protéines potentiellement impliquées dans une maladie : si une protéine d'intérêt interagit avec plusieurs protéines connues pour leur implication dans une maladie spécifique (ex., le cancer du côlon ou maladie de Parkinson), cette protéine pourrait également jouer un rôle dans cette pathologie.

Exemple: Dans le réseau montré ici à droite on pourrait inférer la fonctions des protéines 1-5 grâce aux interactions avec des protéines de fonction connue, en considérant par exemple un seuil de 2 interactions pour assigner une fonction. Cette approche fournirait présente bien entendu un certain risque d'erreur, comme toute inférence.

Dans le cas de la protéine 4, ceci nous amènerait à lui associer une double fonction (elle interagit avec 2 protéines de fonction B, et 2 protéines de fonction C).



Sous-graphes

Un **sous-graphe** est une partie d'un graphe plus vaste, composée d'un sous-ensemble de nœuds et d'arêtes de ce graphe. Il conserve la structure et les connexions présentes dans le graphe d'origine, mais se limite aux éléments sélectionnés.

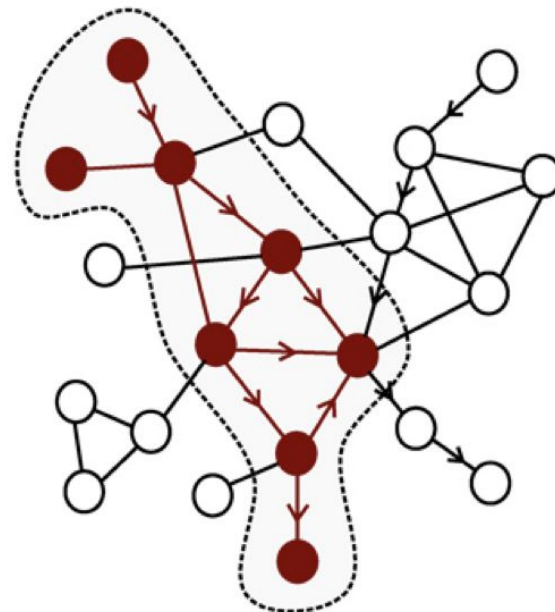
Dans les réseaux biologiques, les sous-graphes sont généralement nommés **sous-réseaux**.

Les critères pour identifier ou extraire des sous-graphes dans un réseau peuvent inclure :

- sélection de nœuds selon des **propriétés topologiques** spécifiques, tels que le coefficient de regroupement ;
- **module** ou **communauté** (ensemble de nœuds fortement interconnectés) ;
- ensemble de nœuds qui partagent des **propriétés biologiques** (facteurs transcriptionnels, enzymes, enzymes, protéines participant à un processus biologique particulier, ...), et les interactions entre ces nœuds.

Par exemple, on pourrait extraire un sous-réseau de l'interactome humain en sélectionnant

- comme nœuds, les protéines impliquées dans une pathologie donnée ;
- comme arêtes, toutes les interactions entre ce sous-ensemble de protéines.

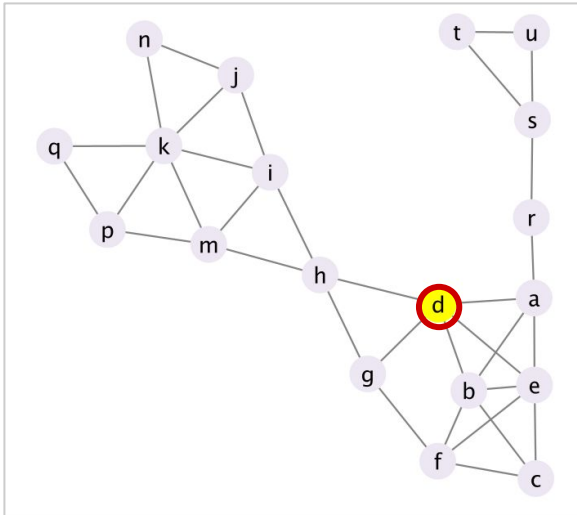


Voisinage d'un noeud

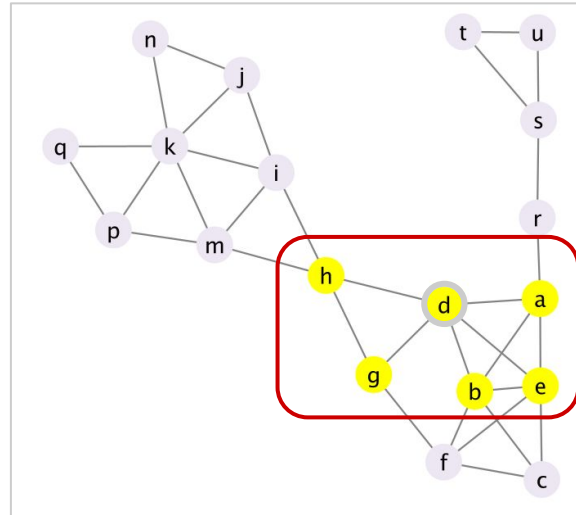
Le voisinage d'un noeud d'intérêt est l'ensemble des noeuds qui y sont connectés, en précisant une distance minimale

- Voisinage de premier ordre : ensemble des premiers voisins, c'est-à-dire les noeuds immédiatement connectés au noeud d'intérêt.
- Voisinage de deuxième ordre : ensemble des noeuds connectés directement au noeud d'intérêt, ou connectés à ses premiers voisins. Le voisinage d'ordre 2 inclut le voisinage de premier ordre.
- Voisinage d'ordre n : ensemble des noeuds connectés au noeud d'intérêt via un chemin de maximum n arêtes. Le voisinage d'ordre n inclut tous les voisinages d'ordre inférieur à n .

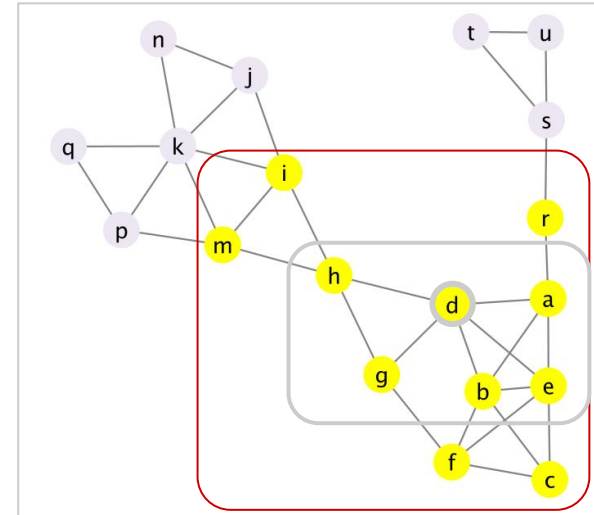
Noeud d'intérêt



Voisinage de premier ordre



Voisinage de deuxième ordre



Une équipe de chercheurs issus du domaine de la physique statistique analyse la **topologie du réseau métabolique**, en appliquant des outils précédemment utilisés pour analyser d'autres types de réseaux (internet, trajets des avions, réseaux d'acteurs ayant joué dans les mêmes films,...).

Ils mesurent des paramètres de topologie des réseaux, et constatent que le réseau métabolique

1. Suit une loi de puissance : un grand nombre de métabolites sont faiblement connectés, et quelques uns sont très fortement connectés (les "hubs").
2. Petit-monde : le nombre moyen de réactions pour convertir un métabolite à un autre est de 3. Ceci est surprenant étant donné la taille du réseau (5000 réactions, 5000 métabolites), mais résulte des hubs, qui font des transitions rapides d'un métabolite à un autre.
3. Invariance d'échelle : les propriétés telles que la taille moyenne des chemins sont robustes aux changements d'échelle → si on isole un sous-réseau aléatoire, il aura les mêmes propriétés que le réseau entier.

Par analogie avec les réseaux technologiques (internet), ils infèrent des propriétés de robustesse aux "erreurs" aléatoires (les mutations) mais de vulnérabilité aux attaques ciblées (la perte des hubs désolidarise des grands pans du réseau).

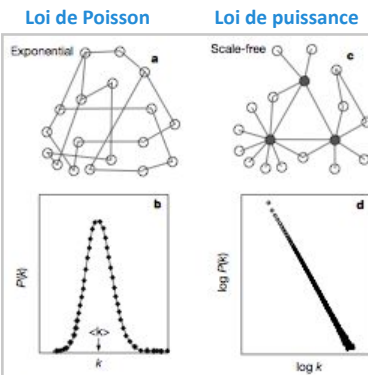
The large-scale organization of metabolic networks

H. Jeong^{*}, B. Tombor[†], R. Albert^{*}, Z. N. Oltvai[†] & A.-L. Barabási^{*}

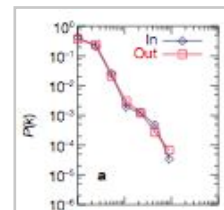
^{*} Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

[†] Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

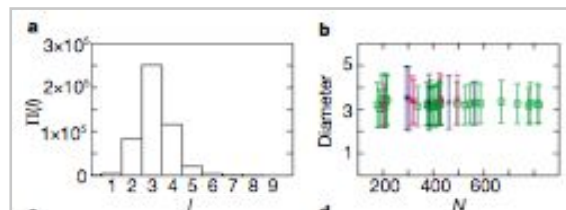
Deux modèles théoriques alternatifs



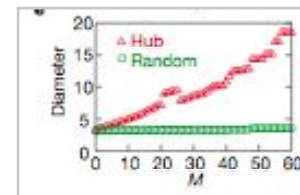
Distribution de degrés des métabolites



Petit monde



Tolérance à l'erreur Vulnérabilité aux attaques



Topologie des réseaux d'interactions protéine-protéine

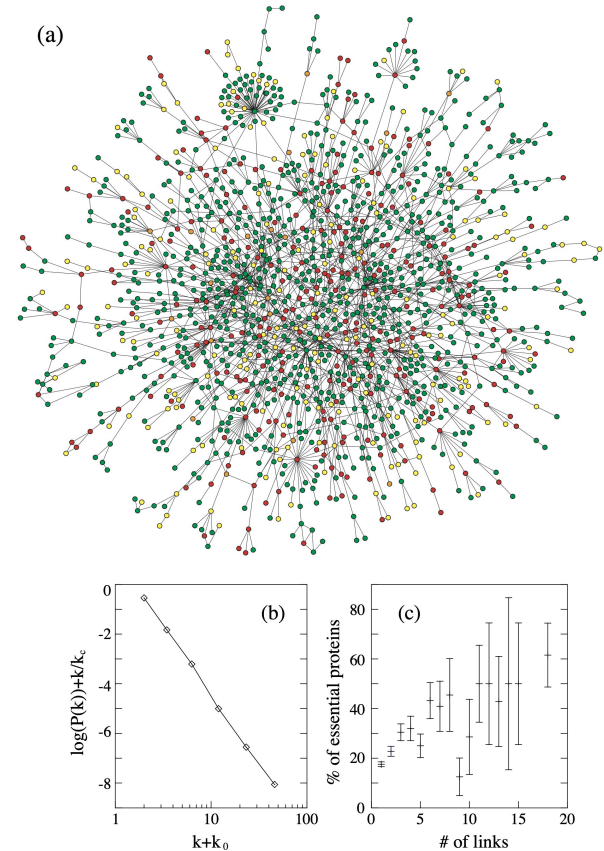
La même équipe analyse avec les mêmes méthodes la **topologie du réseau formé par l'interactome**.

Sur cette figure, les noeuds représentent des protéines et les arêtes des interactions.

Ils constatent que le réseau d'interactome comporte quelques **noeuds fortement connectés**, qu'ils appellent des **hubs**, par analogie à internet.

Ils tirent de ces observations des conclusions générales, par analogie avec leurs analyses préalables des réseaux internet : un réseau informatique présentant des "hubs" de ce type est robuste aux défaillances aléatoires de l'un ou l'autre ordinateur, mais sont particulièrement fragiles à la suppression d'un hub.

De fait, ils trouvent que les protéines les plus fortement connectées sont généralement essentielles : les mutations des gènes correspondant aux "hubs" sont généralement létales.



Sur base de ces résultats, Barabasi et Oltvai proclament que les réseaux moléculaires et cellulaires sont gouvernés par des lois universelles.

En 2009, la revue Science consacre un dossier complet à la “science des réseaux”, en mettant en valeur l’universalité des propriétés des réseaux biologiques, sociaux, économiques, techniques, ...

NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION

Albert-László Barabási* & Zoltán N. Oltvai‡

A key aim of postgenomic biomedical research is to systematically catalogue all molecules and their interactions within a living cell. There is a clear need to understand how these molecules and the interactions between them determine the function of this enormously complex machinery, both in isolation and when surrounded by other cells. Rapid advances in network biology indicate that cellular networks are governed by universal laws and offer a new conceptual framework that could potentially revolutionize our view of biology and disease pathologies in the twenty-first century.

Pushing Networks to the Limit

PERSPECTIVE

Scale-Free Networks: A Decade and Beyond

Albert-László Barabási

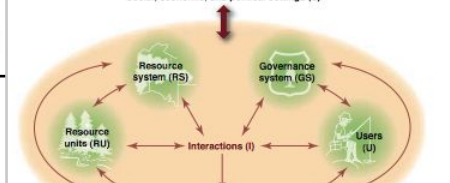
For decades, we tacitly assumed that the components of such complex systems as the cell, the society, or the Internet are randomly wired together. In the past decade, an avalanche of research has shown that many real networks, independent of their age, function, and scope, converge to similar architectures, a universality that allowed researchers from different disciplines to embrace network theory as a common paradigm. The decade-old discovery of scale-free networks was one of those events that had helped catalyze the emergence of network science, a new research field with

The web of life



Socio-ecological networks

Social, economic, and political settings (S)



Counterterrorism's New Tool: 'Metanetwork' Analysis

Researchers have created sophisticated new programs to probe beneath the surface of social interactions. How well do they work against terrorists?

PERSPECTIVE

Predicting the Behavior of Techno-Social Systems

Alessandro Vespignani

We live in an increasingly interconnected world of techno-social systems, in which infrastructures composed of different technological layers are interoperating within the social component that drives their use and development. Examples are provided by the Internet, the World Wide Web, WiFi communication technologies, and transportation and mobility infrastructures. The multiscale nature and complexity of these networks are crucial features in understanding and managing the networks. The accessibility of new data and the advances in the theory and modeling of complex networks are providing an integrated framework that brings us closer to achieving true predictive power of the behavior of techno-social systems.

PERSPECTIVE

Economic Networks: The New Challenges

Frank Schweitzer,^{1*} Giorgio Fagiolo,² Didier Sornette,^{1,3} Fernando Vega-Redondo,^{4,5} Alessandro Vespignani,^{4,6} Douglas R. White⁶

The current economic crisis illustrates a critical need for new and fundamental understanding of the structure and dynamics of economic networks. Economic systems are increasingly built on interdependencies, implemented through trans-national credit and investment networks, trade relations, or supply chains that have proven difficult to predict and control. We need, therefore, an approach that stresses the systemic complexity of economic networks and that can be used to revise and extend established paradigms in economic theory. This will facilitate the design of policies that reduce conflicts between individual interests and global efficiency, as well as reduce the risk of global failure by making economic networks more robust.

Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. Nat Rev Genet 5, 101–113 (2004). doi.org/10.1038/nrg1272

Barabási, A.-L. Scale-Free Networks: A Decade and Beyond. Science 325, 412–413 (2009).

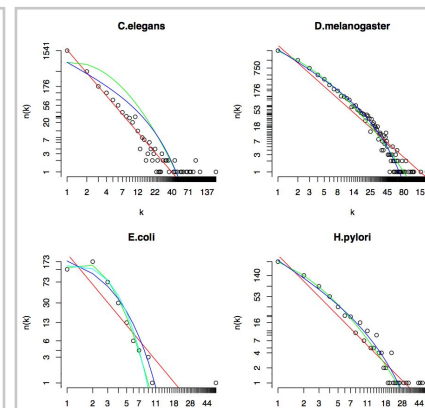
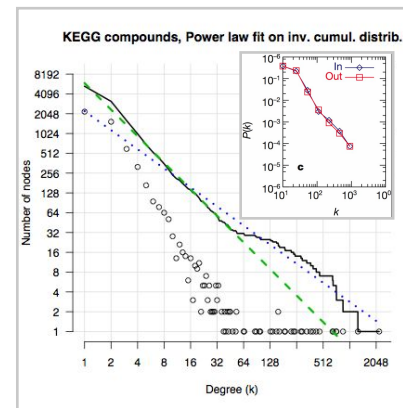
Critiques des propriétés “universelles” des réseaux biologiques

Critiques statistiques

- Des ré-analyses statistiques des mêmes données (interactome et réseaux métaboliques) ont démontré que la loi de puissance ne s'ajuste pas sur les données réelles. Les articles initiaux avaient “noyé le poisson” en regroupant ces données par classes.
- La propriété “petit monde” des réseaux métaboliques repose sur un artéfact : les “hubs” sont en fait des molécules telles que l'eau, le proton, l'oxygène etc, qui ne constituent pas des métabolites intermédiaires entre deux réactions. Les chemins “courts” passaient par des raccourcis en transformant par exemple le glucose en H₂O, puis l'H₂O en éthanol. Ceci n'a aucune pertinence biochimique.

Name	Total Degree	Metabolic function (from 71)
H ₂ O	2213	Hydrolysis, hydration
H ⁺	1269	Proton pumps
Oxygen	860	Electron acceptor
NADP ⁺	724	Coenzyme: Electron acceptor
NADPH	721	Coenzyme: Electron donor in anabolism
NAD ⁺	663	Coenzyme: Electron acceptor in catabolism
NADH	655	Coenzyme: Electron donor
ATP	466	Coenzyme: Energy donor
CO ₂	427	Last product of oxidation, precursor of photosynthesis

D-Glucose->R04094->H ₂ O->R02682->Ethanol
D-Glucose->R00300->NADH->R00754->Ethanol
D-Glucose->R00534->H ₂ O->R02359->Ethanol
D-Glucose->R02558->H ₂ O->R02682->Ethanol
D-Glucose->R00304->H ₂ O->R02359->Ethanol
D-Glucose->R02558->H ₂ O->R02359->Ethanol
D-Glucose->R05142->H ₂ O->R02682->Ethanol
D-Glucose->R00534->H ₂ O->R02682->Ethanol
D-Glucose->R01444->H ₂ O->R02682->Ethanol
D-Glucose->R04006->H ₂ O->R02359->Ethanol



M. P. H. Stumpf and P. J. Ingram (2005). Probability models for degree distributions of protein interaction networks. *Europhys. Lett.* 71:152-158.

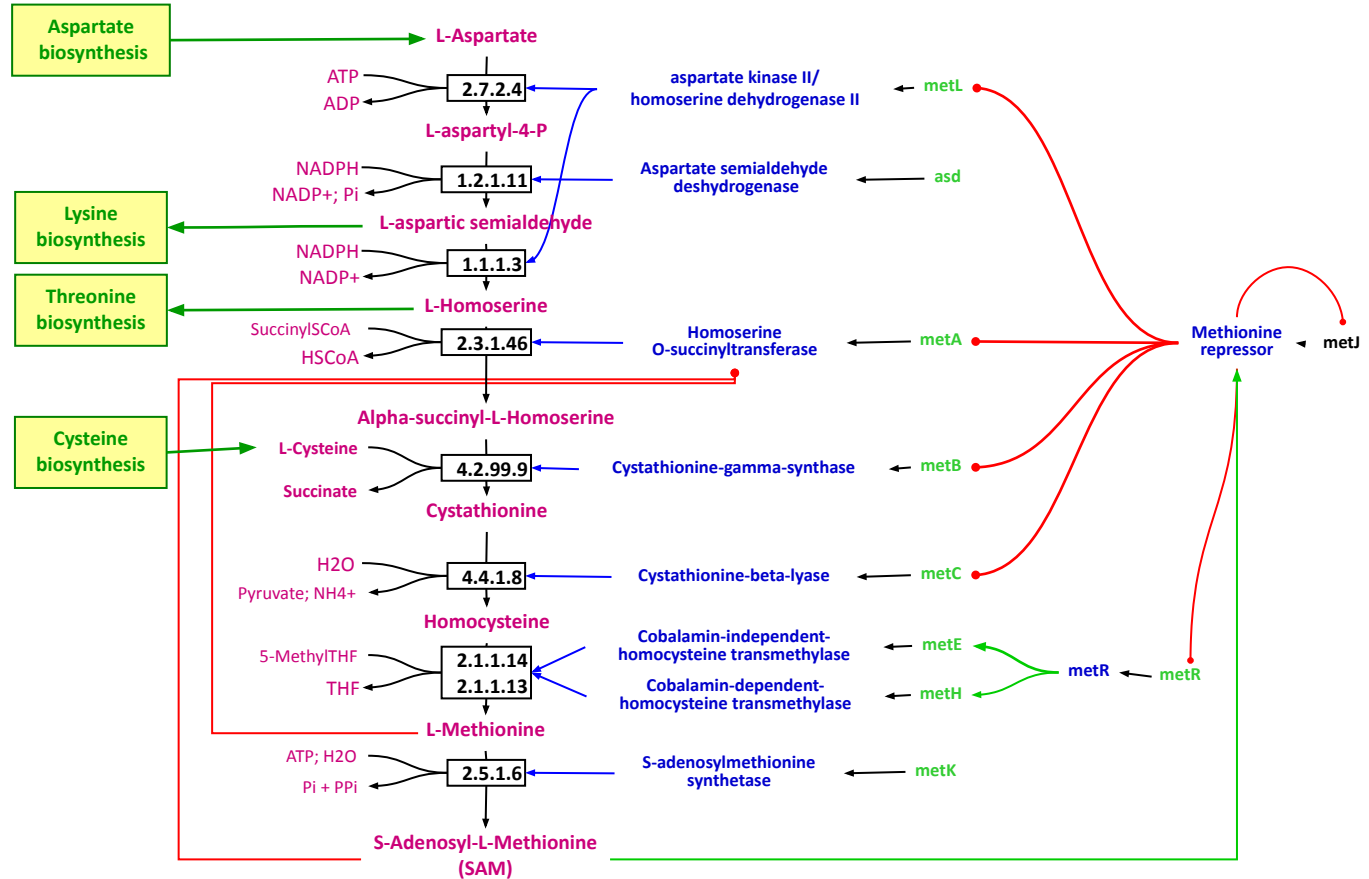
Lima-Mendez, G. & van Helden, J. The powerful law of the power law and other myths in network biology. *Mol. BioSyst.* 5, 1482 (2009).

Conclusions

- L'avènement de plusieurs technologies à haut débit et la collecte d'informations dans de grandes bases de données permettent d'extraire des informations utiles en se basant sur les interactions moléculaires de différents types (transcriptome, interactome, conservation des gènes, appartenance aux mêmes opérons, ...).
- Ces informations peuvent être représentées sous forme de réseaux, ce qui permet de leur appliquer un arsenal de méthodes informatiques et statistiques. Cependant, certaines propriétés soi-disant "universelles" imputées à ces réseaux biologiques (loi de puissance, "petit monde") reposent essentiellement sur des artefacts, et sur l'application de méthodes générales, qui se sont avérées inadéquates pour tenir compte des propriétés des réseaux biologiques.
- Pour pouvoir extraire des informations pertinentes de ces réseaux, il est essentiel de combiner une bonne connaissance des méthodes mathématiques et informatiques d'analyse des graphes avec une bonne compréhension des données biologiques représentées. Ceci permet par exemple d'identifier des modules fortement connectés, et de mieux comprendre le fonctionnement des réseaux d'interactions moléculaires.

L'intégration des données diverses

Les réseaux moléculaires ne peuvent se comprendre que si l'on intègre les différents types de données. Par exemple, pour comprendre le métabolisme d'un acide aminé, il faut intégrer les données métaboliques (molécules, réactions, catalyses enzymatiques) la régulation (transcriptionnelle, allostérique), le transport, et les autres voies métaboliques qui y sont connectées.



La segmentation de la drosophile

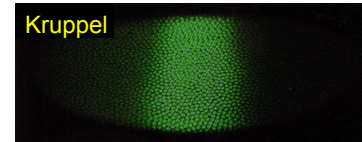
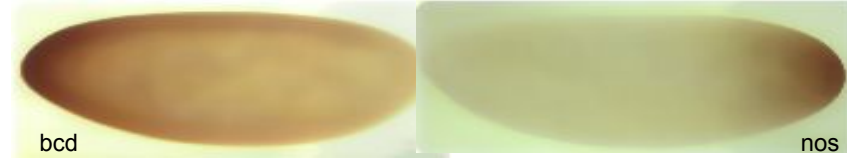
Un autre exemple est la régulation du développement embryonnaire, qui a été étudié depuis des décennies chez la drosophile.

La formation des segments repose sur un réseau de régulation transcriptionnelle remarquablement complexe.

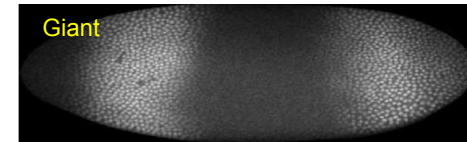
Au départ, deux ARN sont déposés aux extrémités antérieure (bicoid) et postérieure (nanos) de l'embryon. Ces ARN codent pour des facteurs transcriptionnels qui activent ou inhibent d'autres facteurs transcriptionnels (Kruppel, Giant), lesquels activent ou inhibent à leur tour d'autres facteurs (Hairy, even-skipped) etc.

On arrive ainsi progressivement à délimiter des bandes d'expression qui détermineront la formation des segments de la larve, puis de la mouche adulte.

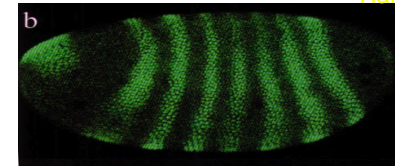
Adapted from Carroll, 2006



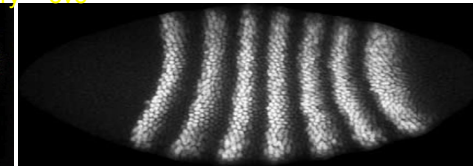
Source: Carroll, 2005.



Source: Thieffry and Sanchez (2003). <http://flyex.ams.sunysb.edu/>.



Source: Carroll, 2005.



Source: Thieffry and Sanchez (2003). <http://flyex.ams.sunysb.edu/>.



Source: Lawrence (1993). The making of a fly

La segmentation de la drosophile

On a caractérisé de façon très détaillé le réseau d'interaction entre facteurs transcriptionnels responsable de la formation progressive des segments antéro-postérieurs de la drosophile.

Des travaux de modélisation mathématique ont permis de montrer la façon dont ce réseau, extrêmement intriqué, peut expliquer les phénotypes des différents mutants, et par là de comprendre les mécanismes sous-jacents à la formation de l'embryon normal.

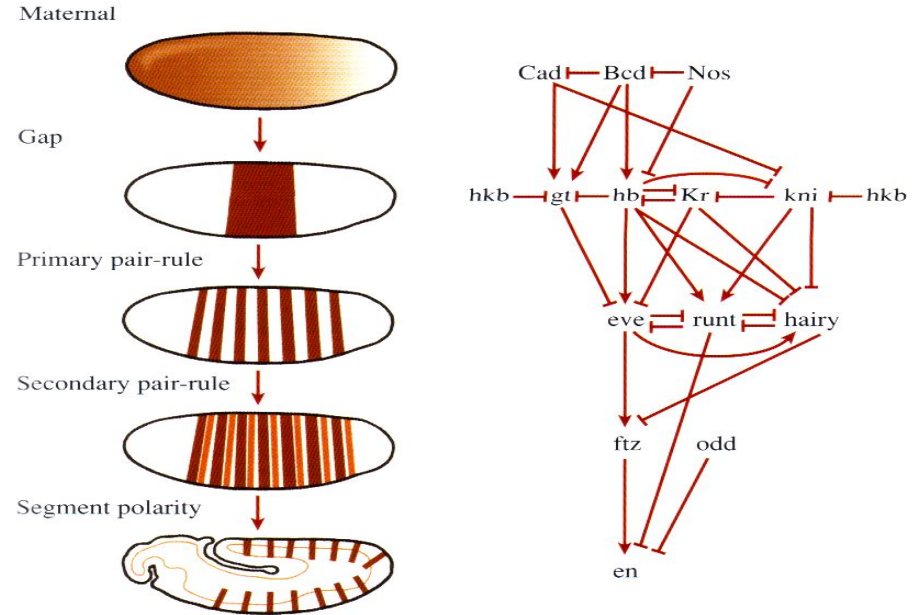


Figure 3.5

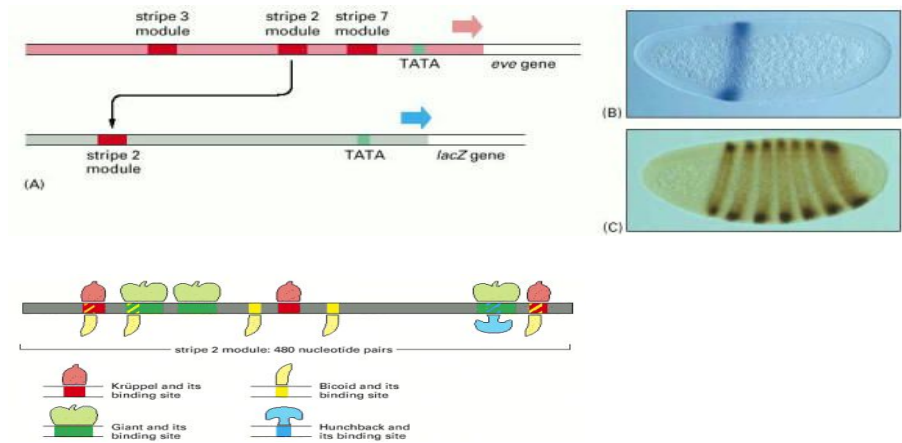
The segmentation genetic regulatory hierarchy

(left) The expression patterns of five classes of anteroposterior axis patterning genes are depicted in embryos at different stages. **(right)** Selected members of these classes are shown and the regulatory interactions between these genes are indicated. An arrow indicates a positive regulatory interaction; a line crossed at its end indicates a negative repressive regulatory relationship.

Source: Carroll, 2005. From DNA to diversity (2nd edition). Blackwell Publishing.

Réseaux de régulation du développement embryonnaire

Exemple: le gène *even-skipped* est exprimé dans 7 bandes antéro-postérieures délimitées de façon extrêmement précise. Chacune de ces bandes est contrôlée par une combinaison d'autres facteurs transcriptionnels (Krüppel, Bicoid, Giant, Hunchback, ...) qui interagissent en se liant dans des régions non-codantes en amont du gène *even-skipped*, et dans ses introns.



Quelques remarques et points de réflexion

- Les réseaux sont devenus omniprésents en biologie, pour représenter et analyser les données massives de différents types
- Ces données nous apportent un éclairage sur les composantes des systèmes biologiques
 - Génomes
 - Transcriptomes
 - Protéomes
 - Interactomes
 - Réseaux de régulation
 - Réseaux métaboliques
- Les données massives et l'énumération des composantes ne suffisent cependant pas à les rendre informatives.
- On commence à faire de la science à partir des données quand on arrive à en extraire l'information pertinente pour comprendre les mécanismes du vivant : développement des organismes, adaptation à l'environnement, évolution, ...



René Magritte – "La Clairvoyance"



René Magritte's picture "La clairvoyance" (*The foresight*), used by Alain Ghysen as front picture for a special issue on *Developmental Biology in Belgium*.