



HAL
open science

Identification des éléments clefs du métabolisme des lipides et de leurs régulateurs

Pierre Blavy

► **To cite this version:**

Pierre Blavy. Identification des éléments clefs du métabolisme des lipides et de leurs régulateurs. Interactions entre organismes. Université Rennes 1, 2010. Français. NNT : . tel-00541207v2

HAL Id: tel-00541207

<https://theses.hal.science/tel-00541207v2>

Submitted on 3 Dec 2010

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.



N° d'ordre 1010-9
N° de série B-204



THESE / AGROCAMPUS OUEST

Sous le sceau de l'Université Européenne de Bretagne

pour obtenir le diplôme de :

**DOCTEUR DE L'INSTITUT SUPERIEUR DES SCIENCES AGRONOMIQUES,
AGRO-ALIMENTAIRES, HORTICOLES ET DU PAYSAGE**

Ecole Doctorale : Vie-Agro-Santé

**Thèse présentée par : Pierre BLAVY
Identification des éléments clés du métabolisme des
lipides et de leurs régulateurs**

Soutenance : le 12 mars 2010 devant le jury d'examen composé de :

Président

DAVID CAUSEUR Professeur, Laboratoire de Mathématiques Appliquées,
Agrocampus-Ouest

Rapporteurs

JEAN PIERRE MAZAT Professeur, INSERM U 688 - Physiopathologie Mitochondriale,
Université de Bordeaux 2

DAVID JAMES SHERMAN Directeur de recherche, INRIA, Université Bordeaux 1

Membres du Jury

MAELA KLOAREG Maître de conférences, Laboratoire de Mathématiques Appliquées,
Agrocampus-Ouest

ANNE SIEGEL Chargée de recherche au CNRS, Symbiose, UMR 6074
CNRS-Université de Rennes 1 / INRIA IRISA (*codirectrice de thèse*)

SANDRINE LAGARRIGUE Professeur, UMR598 Génétique Animale de Rennes (GA-Ren)
Agrocampus-Ouest-INRA (*directrice de thèse*)

Équipes d'accueil :



UMR SENAH,
UMR Génétique Animale,
UR66



SYMBIOSE

Remerciements

Je remercie Jean Pierre Mazat et David James Sherman qui ont accepté le travail de rapporteur ainsi que David Causeur, Maela Kloareg, Sandrine Lagarrigue et Anne Siegel qui ont accepté d'examiner cette thèse. Merci aussi à Daniel Sauvant, Jean Sébastien Pierre et Sophie Vincent, membres du comité de thèse pour leur suivi et leurs conseils.

Je remercie aussi Anne Siegel, Jaap Van Milgen, Florence Gondret, Sandrine Lagarrigue, Ovidiu Radulescu, Pascal Martin, Hervé Guillou pour l'aide et le soutien qu'ils m'ont apporté au cours de cette thèse ainsi que pour tous les bons moments que j'ai passé avec eux. Merci aussi à François Moreews et Pauline Gloaguen avec qui ce fut un plaisir de travailler.

Je remercie aussi chaleureusement toute l'équipe Symbiose, pour son accueil et pour la bonne ambiance quotidienne dans laquelle j'ai eu la chance de passer la plupart de mes journées.

Je remercie aussi mes amis Gaspard Bertrand, Lionel Davoust, Pierre Cozanet, Bertrand Meda, Laurent Planchet, Jehane Prudhomme, Natacha Quentin, Sebastien Rochette, Charlie Theunicorn, Perrine Zeller et tous les autres pour tous les bons moments que j'ai passé avec eux, pour tout le soutien qu'ils m'ont apporté et parce que ce sont juste des gens géniaux.

Je remercie mes parents, qui m'ont toujours donné les moyens de suivre les études que je souhaitais.

Pour terminer, je remercie l'INRA qui a assuré le financement de ma thèse. Mes remerciements vont également à mes unités d'accueil : l'Equipe Symbiose (IRISA, INRIA), l' UMR Systèmes d'Elevage Nutrition Animale et Humaine (Département PHASE, INRA), l'UMR de Génétique Animale (Département Génétique Animale, INRA) et l'UR de Pharmacologie Toxicologie (Département Santé Animale, INRA).

Liste des publications et des communications

Publications dans des revues à comité de lecture

- *Transcriptome profiling of the feeding-to-fasting transition in chicken liver*, C. Désert, M.J. Duclos, P. Blavy, F. Lecerf, F. Moreews Klopp, M. Aubry, F. Herault, P. Le Roy, C. Berri, M.Douaire, C. Diot, S. Lagarrigue, 2008 [81].
- *A minimal model for hepatic fatty acid balance during fasting : Application to PPAR alpha-deficient mice*, P. Blavy , F. Gondret, H. Guillou, S. Lagarrigue, P.G.P. Martin, J. van Milgen, O. Radulescu, A. Siegel, 2009 [30].

Publications dans des actes de conférence

- *A minimal and dynamic model for fatty acid metabolism in mouse liver* P. Blavy, F. Gondret, H. Guillou, S. Lagarrigue, P. Martin, O. Radulescu, A. Siegel, J. Van Milgen. JOBIM 2008 [29].

Communications

- Colloque du programme fédérateur INRA agroBI, 2008 : présentation orale.
- Journée des doctorants INRA 2008 : présentation orale.
- Journée des doctorants INRA 2007 : poster.
- Journée INRA de génétique animale 2007 : présentation orale
- Journée INRA du département PHASE 2007 : présentation orale.

Table des matières

| | |
|--|-----------|
| Introduction Générale | 14 |
| Revue bibliographique | 17 |
| I Le métabolisme des Lipides | 19 |
| 1 Le métabolisme des lipides : une composante du métabolisme énergétique | 21 |
| 1.1 Les principales voies du métabolisme énergétique | 21 |
| 1.2 Les molécules de réserve en énergie | 23 |
| 1.2.1 Le glucose et ses polymères | 23 |
| 1.2.2 Acides gras et triglycérides | 23 |
| 1.2.3 Les protéines | 24 |
| 1.3 Les principaux organes du métabolisme énergétique | 24 |
| 1.3.1 Lieux de dégradation et de stockage des réserves énergétiques | 24 |
| 1.3.2 Coopération entre organes | 25 |
| 2 Généricité des réactions biochimiques du métabolisme des lipides | 27 |
| 2.1 Digestion des triglycérides et absorption des acides gras | 27 |
| 2.1.1 Digestion des triglycérides et absorption par l'intestin | 27 |
| 2.1.2 Estérification et synthèse de lipoprotéines | 27 |
| 2.2 Transport des lipides entre les différents organes | 27 |
| 2.2.1 Les différentes lipoprotéines | 29 |
| 2.2.2 Le transport des lipides issus de l'alimentation | 29 |
| 2.2.3 Le transport des lipides endogènes hépatiques et du cholestérol | 29 |
| 2.2.4 Le transport des acides gras libres issus du tissu adipeux | 31 |
| 2.3 Transport intracellulaire des lipides | 31 |
| 2.4 Synthèse de novo et transformation des acides gras | 31 |
| 2.4.1 La néosynthèse des acides gras | 31 |
| 2.4.2 Transformations des acides gras | 34 |
| 2.5 Dégradation des acides gras | 34 |
| 2.5.1 L'oxydation se déroule dans la mitochondrie et dans le peroxysome | 34 |
| 2.5.2 Réactions biochimiques de l'oxydation des acides gras | 34 |
| 2.5.3 Consommation d'acétyl-coenzyme A par le cycle de Krebs et la chaîne respiratoire | 38 |
| 2.5.4 Cétogenèse | 38 |

| | | |
|------------|--|-----------|
| 3 | Plasticité du métabolisme des lipides : ses principales régulations | 39 |
| 3.1 | Principales régulations hormonales | 40 |
| 3.1.1 | L'insuline | 43 |
| 3.1.2 | Le glucagon | 44 |
| 3.2 | Principales régulations de l'expression des gènes relatives au métabolisme des lipides | 45 |
| 3.2.1 | ChREBP stimule la lipogenèse en réponse au glucose | 45 |
| 3.2.2 | SREBP1 stimule la synthèse de triglycérides à l'état nourri en réponse à l'insuline | 47 |
| 3.2.3 | PPAR α stimule la β -oxydation et la cétogenèse lors du jeûne | 51 |
| 3.3 | Conclusion | 53 |
| II | Modélisation dynamique des systèmes biologiques | 55 |
| 1 | La modélisation : une approche pour comprendre les systèmes complexes | 57 |
| 1.1 | Le métabolisme des lipides est vaste et complexe | 57 |
| 1.1.1 | De nombreuses données expérimentales hétérogènes | 57 |
| 1.1.2 | De nombreuses connaissances éparpillées dans la bibliographie | 58 |
| 1.1.3 | Conclusion | 59 |
| 1.2 | Générer des connaissances dans un système complexe | 59 |
| 1.2.1 | La démarche expérimentale permet de générer des connaissances | 59 |
| 1.2.2 | Choix d'une approche systémique | 60 |
| 1.2.3 | Modéliser le système permet d'appliquer la démarche expérimentale | 61 |
| 2 | Quel(s) modèle(s) construire pour représenter le métabolisme des lipides ? | 63 |
| 2.1 | Choix des systèmes d'équations ordinaires pour modéliser les données cinétiques quantitatives. | 64 |
| 2.1.1 | Les systèmes d'équations différentielles ordinaires | 64 |
| 2.1.2 | Les formalismes dérivés des systèmes d'équations différentielles ordinaires | 65 |
| 2.1.3 | Choix du système d'équations différentielles ordinaires | 67 |
| 2.2 | Utilité et limites des modèles dynamiques | 67 |
| 2.2.1 | Les modèles dynamiques classiques pour intégrer les données qualitatives. | 68 |
| 2.2.2 | Limites des modèles dynamiques | 69 |
| 2.3 | Les graphes d'influences | 70 |
| 2.3.1 | Méthodes d'analyse des graphes d'influences | 71 |
| 2.3.2 | Méthodes de construction des graphes d'influences | 73 |
| III | Conclusion : objectifs de la thèse | 77 |
| | Résultats | 80 |
| I | Analyse d'un système complexe décrit par des données haut débit | 81 |
| 1 | Méthodes courantes d'analyse | 85 |
| 1.1 | Echantillon de publications retenues | 85 |
| 1.2 | Identification de gènes différentiellement exprimés entre les conditions expérimentales | 86 |
| 1.3 | Identification de groupes de gènes coexprimés | 86 |
| 1.4 | Interprétation biologique des groupes de gènes coexprimés | 86 |
| 1.5 | Conclusion | 87 |

| | | |
|-----------|--|------------|
| 2 | Mise en oeuvre d'une démarche d'analyse de données transcriptomiques | 89 |
| | Publication Desert <i>et al.</i> , , BMC 2008 | 89 |
| 2.1 | Contexte scientifique | 109 |
| 2.2 | Démarche | 109 |
| 2.2.1 | Identification de gènes différentiellement exprimés entre les conditions expérimentales | 110 |
| 2.2.2 | Identification de groupes de gènes coexprimés | 110 |
| 2.2.3 | Interprétation biologique des groupes de gènes coexprimés | 110 |
| 2.2.4 | Analyse manuelle d'un sous ensemble de gènes | 110 |
| 2.3 | Résultats | 110 |
| 2.3.1 | Identification des gènes différentiellement exprimés entre les conditions expérimentales | 110 |
| 2.3.2 | Identification des groupes de gènes coexprimés | 110 |
| 2.3.3 | Interprétation biologique des groupes de gènes coexprimés | 111 |
| 2.3.4 | Analyse manuelle d'un sous ensemble de gènes | 111 |
| 2.4 | Conclusion | 113 |
| 3 | Les limites des méthodes classiquement utilisées en analyse transcriptomique. | 115 |
| 3.1 | Les limites de la classification | 115 |
| 3.1.1 | Une corrélation peut avoir plusieurs explications | 115 |
| 3.1.2 | Les éléments ne sont pas tous annotés dans les ontologies | 115 |
| 3.1.3 | La qualité des annotations est cruciale | 115 |
| 3.1.4 | Une interprétation des clusters difficile | 116 |
| 3.1.5 | Exemple | 116 |
| 3.1.6 | Conclusion | 116 |
| 3.2 | Les limites de l'étude bibliographique | 117 |
| 3.2.1 | Une bibliographie partielle et fastidieuse | 117 |
| 3.2.2 | Un raisonnement difficile | 117 |
| II | Construction de bases d'interactions | 119 |
| 1 | Matériel et méthodes | 121 |
| 1.1 | Introduction | 121 |
| 1.2 | Contenu accessible des différentes bases de données de connaissances | 124 |
| 1.2.1 | Critères de choix des bases utilisées | 124 |
| 1.2.2 | Analyse des types d'informations référencées dans les bases étudiées | 125 |
| 1.2.3 | Analyse des informations exploitables automatiquement | 127 |
| 1.3 | Méthode d'évaluation du contenu des bases de connaissances | 134 |
| 1.3.1 | Choix du niveau de comparaison : les influences | 134 |
| 1.3.2 | Ingenuity [E] est-elle comparable aux autres bases? | 135 |
| 1.3.3 | Le contexte n'est pas comparable | 135 |
| 1.3.4 | Comparaison de l'origine de la bibliographie | 135 |
| 1.3.5 | Evaluation du contenu | 136 |
| 1.3.6 | Analyse des complémentarités entre bases | 137 |
| 1.3.7 | Identification d'éléments clefs | 138 |
| 2 | Résultats et discussion | 141 |
| 2.1 | Position des bases de connaissances en terme de champs disciplinaires | 141 |
| 2.1.1 | Un grand nombre de journaux et quelques journaux phares | 141 |
| 2.1.2 | Coeur de cible des bases | 143 |
| 2.1.3 | Bibliographie relative aux lipides | 145 |

| | | |
|---|--|------------|
| 2.2 | Quantité de bibliographie | 145 |
| 2.2.1 | Les bases généralistes contiennent beaucoup de bibliographie | 145 |
| 2.2.2 | Ingenuity est une base riche en bibliographie | 146 |
| 2.2.3 | Les trois bases de données ne référencent pas les mêmes publications | 146 |
| 2.2.4 | Conclusion | 146 |
| 2.3 | Quantité d'information extractible | 146 |
| 2.3.1 | En terme d'influences | 146 |
| 2.3.2 | En terme d'influences relatives au métabolisme énergétique | 148 |
| 2.3.3 | Les bases sont complémentaires | 149 |
| 2.4 | Identification d'éléments clefs | 151 |
| 2.4.1 | Des résultats globalement cohérents avec les <i>a priori</i> | 151 |
| 2.4.2 | Mise en évidence d'éléments clefs | 152 |
| 2.5 | Conclusion | 153 |
| III Analyse d'un système complexe par simplification | | 155 |
| Introduction | | 157 |
| 1 Construction d'un squelette générique des principales voies du métabolisme des acides gras | | 161 |
| 1.1 | Identification des fonctions biochimiques et des régulations clefs | 161 |
| 1.1.1 | Identification des fonctions et des régulations candidates par expertise | 161 |
| 1.1.2 | Identifications d'une liste de comportements par expertise | 162 |
| 1.1.3 | Réduction des fonctions candidates au minimum nécessaire | 163 |
| 1.2 | Mise en équation du modèle obtenu | 164 |
| 1.2.1 | Hypothèses parcimonieuses | 164 |
| 1.2.2 | Le modèle de connaissance générique et minimal obtenu | 167 |
| 1.3 | Evolution du modèle par construction ascendante et modification des équations. | 171 |
| 1.4 | Conclusion | 172 |
| 2 Modèle dynamique minimal du métabolisme hépatique des acides gras lors du jeûne | | 173 |
| 2.1 | Introduction | 173 |
| | Publication Blavy <i>et al.</i> , , JTB 2009 | 173 |
| Conclusion | | 187 |
| Conclusion et perspectives | | 191 |
| Conclusion | | 191 |
| Perspectives | | 193 |

Liste des abréviations

Liste des enzymes

| Nom | EC |
|--------------------------------|-----------|
| 3HCDH | 1.1.1.35 |
| 3KACT | 2.3.1.16 |
| glucokinase | 2.7.1.2 |
| hexokinase | 2.7.1.1 |
| Long-chain-enoyl-CoA hydratase | 4.2.1.74 |
| LCHAD | 1.1.1.211 |
| enzyme malique | 1.1.1.40 |
| PDH | 1.2.1.51 |
| PK | 2.7.1.40 |
| fructose-2,6-biphosphate | 2.7.1.105 |
| PFK | 2.7.1.11 |
| D3ECI | 5.3.3.8 |

| Nom | Identifiant KEGG |
|---------------------|------------------|
| ACSL | K01897 |
| AMPK | K07198 |
| CPT1 | K00994 |
| Delta-12-desaturase | K10255 |
| Delta-15-desaturase | K10257 |

Liste des protéines

| Nom | Identifiant Uniprot |
|------------------|---------------------|
| SREBP1-a | P36956-1 |
| SREBP1-b | P36956-2 |
| SREBP1-c | P36956-3 |
| SREBP1-isoforme4 | P36956-4 |
| ACBP | P07108 |

Liste des gènes

| Gène | Identifiant HGNC | Synonymes | Gène | Identifiant HGNC | Synonymes |
|---------|------------------|-----------|---------------|------------------|-----------|
| ACAA1 | 82 | | FBP1 | 3606 | |
| ACAA2 | 83 | | FBP2 | 3607 | |
| ACADL | 88 | | GCK | 4195 | |
| ACADM | 89 | | G6PC | 4056 | |
| ACADS | 90 | | HADH | 4799 | |
| ACADVL | 92 | | HADHA | 4801 | |
| ACAT1 | 93 | | HK2 | 4923 | |
| ACC | 84 | ACACA | HMGCL | 5005 | |
| ACLY | 115 | | HMGCS1 | 5007 | |
| ACOX1 | 119 | | HMGCS2 | 5008 | |
| ACOX2 | 120 | | INSIG1 | 6083 | |
| ACOX3 | 121 | | INSIG2 | 20452 | |
| ACOX3 | 3569 | | IRS2 | 6126 | |
| ANGPTL2 | 490 | | IDH1 | 5382 | |
| APOB | 603 | | LCAT | 6522 | |
| AWAT2 | 23251 | DGAT2L4 | LIPC | 6619 | |
| BCAT1 | 976 | | LIPE | 6621 | |
| CD36 | 1663 | | LPA | 6667 | |
| CETP | 1869 | | LPL | 6677 | |
| ChREBP | 12744 | | LSS | 6708 | |
| CLPS | 2085 | | LIPE | 6621 | |
| CPT1A | 2328 | | LXR β | 7965 | |
| CPT1B | 2329 | | ME1 | 6983 | |
| CPT2 | 2330 | | MTTP | 7467 | |
| CS | 2422 | | OGDH | 8124 | |
| CYP51A1 | 2649 | | PECI | 14601 | |
| D5D | 3574 | FADS1 | PCK1 | 8724 | PEPCK |
| FADS1 | 3574 | | PNLIP | 9155 | |
| D6D | 3575 | FADS2 | PNPLA2 | 30802 | |
| D9D | 10571 | SCD, SCD1 | PPAR α | 9232 | |
| DECR1 | 2753 | | PC | 8636 | |
| ECHS1 | 3151 | | PKLR | 9020 | |
| EHHADH | 3247 | | PFKFB1 | 8872 | |
| ETFDH | 3483 | | PFKFB1 | 8873 | |
| ELOVL1 | 14418 | | PFKL | 8876 | |
| ELOVL2 | 4416 | | PFKM | 8877 | |
| ELOVL3 | 18047 | | RXR α | 10477 | |
| ELOVL4 | 14415 | | RXR β | 10478 | |
| ELOVL5 | 21308 | | RXR γ | 10479 | |
| ELOVL6 | 15829 | | S1P | 15456 | |
| ELOVL7 | 26292 | | S2P | 15455 | |
| FABP1 | 3555 | | SAR1B | 10535 | |
| FABP2 | 3556 | | SCAP | 30634 | |
| FABP3 | 3557 | | SLC2A1 | 11005 | |
| FABP4 | 3559 | | SLC2A2 | 11006 | |
| FABP5 | 3560 | | SLC2A4 | 11009 | |
| FABP6 | 3561 | | SLC27A4 | 10998 | |
| FABP7 | 3562 | | SREBP1 | 11289 | |
| FASN | 3594 | FAS | SREBP1 | 11290 | |

Liste des ressources en ligne

Ces ressources sont directement accessibles sous forme de liens dans le texte du document électronique

| Nom | Adresse |
|---------------------|---|
| ANTLR | http://www.antlr.org/ |
| ArrayExpress | http://www.ebi.ac.uk/microarray-as/ae/ |
| <i>bayesian</i> | http://cran.r-project.org/web/views/Bayesian.html |
| bayes++ | http://bayesclasses.sourceforge.net/Bayes++.html |
| BiNoM | http://bioinfo-out.curie.fr/projects/binom |
| BioGrid | http://www.thebiogrid.org/index.php |
| Biopax | http://www.biopax.org/ |
| Logiciels | http://biopaxwiki.org/cgi-bin/moin.cgi/Biological_software_supporting_BioPAX |
| BooleanNet | http://code.google.com/p/booleannet/ |
| BNJ | http://bnj.sourceforge.net/ |
| BRENDA | http://www.brenda-enzymes.info/ |
| Cell Designer | http://www.celldesigner.org/features.html |
| cPath | http://cbio.mskcc.org/software/cpath/ |
| Cytoscape | http://www.cytoscape.org/ |
| Ensembl | http://www.ensembl.org/ |
| Ensembl-Poulet | http://www.ensembl.org/Gallus_gallus/Info/StatsTable |
| FAO | http://www.fao.org/ |
| GeneCards | http://www.genecards.org |
| GeneGo | http://www.genego.com/ |
| GINsim | http://gin.univ-mrs.fr/ |
| Gene Ontology | http://www.geneontology.org/ |
| biological process | http://www.geneontology.org/GO.doc.shtml#biological_process |
| cellular component | http://www.geneontology.org/GO.doc.shtml#cellular_component |
| molecular function | http://www.geneontology.org/GO.doc.shtml#molecular_function |
| Google Scholar | http://scholar.google.fr/ |
| GNU/GSL | http://www.gnu.org/software/gsl/ |
| HGNC | http://www.genenames.org/ |
| HMDB | http://www.hmdb.ca/ |
| Ingenuity,IPA | http://www.ingenuity.com/ |
| brochure Ingenuity | http://www.ingenuity.com/products/pathways_analysis.html |
| JavaBayes | http://www.cs.cmu.edu/~javabayes/index.html |
| Java.sql | http://java.sun.com/j2se/1.4.2/docs/api/java/sql/package-sum.html |
| KEGG | http://www.genome.jp/kegg/ |
| LipidBank | http://lipidbank.jp/ |
| Mathlab | http://www.mathworks.com/ |
| MGI | http://www.informatics.jax.org/ |
| MIAME | http://www.mged.org/Workgroups/MIAME/miame.html |
| MIPS | http://mips.helmholtz-muenchen.de/ |
| MySQL | http://www.mysql.com/ |
| MySQL API | http://dev.mysql.com/doc/ |
| mysql++ | http://tangentsoft.net/mysql++/ |
| NCBI/COG | http://www.ncbi.nlm.nih.gov/COG/ |
| ODEPACK | https://computation.llnl.gov/casc/odepack/odepack_home.html |
| Opale | http://www.nongnu.org/opale/fr/opale/presentation.html |
| OpenBayes | http://www.openbayes.org/ |
| Open Source Physics | http://www.compadre.org/osp/ |
| Pathway Commons | http://www.pathwaycommons.org/pc/ |
| paxtools | http://sourceforge.net/projects/biopax/ |
| phpMyAdmin | http://www.phpmyadmin.net/ |
| PID | http://pid.nci.nih.gov/ |

Liste des ressources en ligne

Ces ressources sont directement accessibles sous forme de liens dans le texte du document électronique

| Nom | Adresse |
|-------------------|---|
| PSI MI XML format | http://www.psidev.info/ |
| PubMatrix | http://pubmatrix.grc.nia.nih.gov/ |
| Pubmed | http://www.ncbi.nlm.nih.gov/pubmed/ |
| R | http://www.r-project.org/ |
| <i>Regulon DB</i> | http://regulondb.ccg.unam.mx/ |
| historique | http://regulondb.ccg.unam.mx/html/Database_summary.jsp |
| SABIO-RK | http://sabio.villa-bosch.de/ |
| SBML | http://sbml.org/ |
| libSBML | http://sbml.org/Software/libSBML/LibSBML_Release_Notes |
| SMD | http://smd.stanford.edu/ |
| Sundials | https://computation.llnl.gov/casc/sundials/ |
| TRANSFAC | http://www.gene-regulation.com/pub/databases.html |
| TRANSPATH | http://www.biobase-international.com/index.php?id=transpathdatabases |
| Uniprot | http://www.uniprot.org |
| Yeast Genome | http://www.yeastgenome.org/ |
| YEASTRACT | http://www.yeasttract.com/ |

Documentation de Transpath

Ces ressources sont directement accessibles sous forme de liens dans le texte du document électronique

| Nom | Adresse |
|-----------|---|
| Index | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/ |
| Pathways | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/gene.html#field |
| Reactions | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/reaction.html#field |
| Molécules | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/molecule.html#content |
| Gènes | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/gene.html |
| Effets | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/effect.html |
| Qualité | http://www.gene-regulation.com/pub/databases/transpath/6.0/doc/quality.html |
| Transfac | http://www.gene-regulation.com/pub/databases/transfac/doc |
| Gènes | http://www.gene-regulation.com/pub/databases/transfac/doc/gene1.html |

Web services

Ces ressources sont directement accessibles sous forme de liens dans le texte du document électronique

Les “#ELEMENT” sont à remplacer par l’identifiant de l’élément à requêter.

| Nom | Adresse |
|------------------|---|
| CAS | http://toolserver.org/~magnus/cas.php?cas=#ID_CAS |
| EXPASYec | http://www.expasy.org/cgi-bin/nicezyme.pl?#ID_EXPASY |
| Enzyme KEGG | http://www.genome.jp/dbget-bin/www_bget?#EC |
| Gène HGNC | http://www.genenames.org/data/hgnc_data.php?hgnc_id=#ID_HGNC |
| Metabolite KEGG | http://www.genome.jp/dbget-bin/www_bget?cpd:#ID_KEGG_COMPOUND |
| Molécule PUBCHEM | http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=#PUBCHEM_ID |
| Pathway KEGG | http://www.genome.jp/kegg/pathway/map/map#ID_PATHWAY.html |
| Protéine UNIPROT | http://www.uniprot.org/uniprot/#ID_UNIPROT |

Liste des tableaux

| | |
|---|-----|
| 0.0.1 Estimation du nombre de publications en biologie relatives à la modélisation | 16 |
| 1.3.1 Rôle des différents organes | 24 |
| 3.1.1 Régulation hormonale du métabolisme énergétique | 41 |
| 3.1.2 Régulation hormonale du métabolisme énergétique | 42 |
| 1.1.1 Estimation du nombre de publications en biologie relatives au métabolisme des lipides | 59 |
| 1.2.1 Les différentes bases de connaissances utilisées | 130 |
| 2.1.1 Nombre de journaux communs | 141 |
| 2.1.2 Nombre d'identifiants Pubmed différents par journaux | 142 |
| 2.1.3 Journaux phares des bases | 144 |
| 2.2.1 Nombre d'identifiants Pubmed communs | 145 |
| 2.3.1 Annotation des sommets par KEGG ou HGNC | 147 |
| 2.3.2 Nombre d'influences extraites des bases | 147 |
| 2.3.3 Evaluation de la redondance entre les bases | 150 |
| 2.4.1 Sommets très connectés | 153 |

Table des figures

| | |
|---|-----|
| 1.1.1 Principales réactions biochimiques du métabolisme énergétique | 22 |
| 1.3.1 Coopération entre organes du métabolisme énergétique | 25 |
| 2.1.1 Principales étapes de la synthèse des lipoprotéines | 28 |
| 2.2.1 Transport et utilisation des lipides | 30 |
| 2.4.1 Néo-synthèse des lipides | 33 |
| 2.4.2 Transformations entre acides gras | 35 |
| 2.5.1 Catabolisme des acides gras | 37 |
| 3.2.1 Mécanismes de régulation de ChREBP | 46 |
| 3.2.2 Régulation transcriptionnelle de la glycolyse et de la lipogenèse | 48 |
| 3.2.3 Activation protéolytique de SREBP1-c | 49 |
| 3.2.4 Interactions mutuelles entre PPAR α et LXR-SREBP1-c | 50 |
| 3.3.1 Les différentes échelles du métabolisme énergétique | 54 |
| 2.1.1 Principe d'un modèle d'équations différentielles linéaires par morceaux | 76 |
| 3.2.1 Erreurs et connaissance partielle | 118 |
| 1.1.1 Diverses méthodes d'exploitation de la bibliographie | 122 |
| 1.2.1 Capture d'écran d'Ingenuity | 128 |
| 1.2.2 Méthode d'évaluation du vocabulaire d'Ingenuity | 133 |
| 1.3.1 Conversions entre bases de données | 139 |
| 1.1.1 Démarche de construction d'un modèle minimal | 165 |
| 1.1.2 Simplifications d'un ensemble de fonctions biologiques régulées | 166 |
| 1.2.1 Modèle de connaissances minimal et générique du métabolisme des lipides | 168 |

Introduction Générale

Chez l'homme, les lipides constituent une proportion importante (environ 40% [72, 231]) de l'énergie obtenue par l'alimentation. Une quantité élevée de lipides alimentaires constitue un facteur de risque important de la survenue d'obésité [231], de maladies cardiovasculaires [282, 289], de dyslipidémie, de stéatose hépatique et de nombreuses autres pathologies [282, 295]. Ainsi, pour réduire l'incidence de l'obésité, la *Food and Agriculture Organisation* (FAO) préconise depuis 1994 un apport en lipides compris entre 15 et 35% de l'apport énergétique global [95]. La réduction de l'apport en lipides dans l'alimentation humaine est donc un enjeu important de santé publique.

Néanmoins, la quantité de lipides joue aussi un rôle non négligeable dans le goût, la palatabilité [340], la tendreté [108] et l'acceptabilité globale des aliments par les consommateurs. Pour ce qui concerne les produits animaux ou piscicoles, la quantité de lipides et leur répartition au sein des produits proposés sont déterminants dans la décision d'achat des consommateurs. Si ces derniers préfèrent généralement une viande pauvre en lipides visibles (lipides localisés au niveau sous cutané et intermusculaire notamment), une augmentation de la quantité de lipides stockés dans le muscle des animaux producteurs de viande améliore généralement (jusqu'au seuil de 2-3% chez le porc par exemple [100]) le goût et l'acceptabilité des produits par les consommateurs après dégustation. Enfin, la quantité et la composition en lipides dans les différents tissus sont également des facteurs déterminants de l'aptitude à la transformation industrielle des viandes ([115, 108, 230]).

Les lipides, et notamment les glycolipides et phospholipides des membranes cellulaires [345] sont principalement constitués d'acides gras. Ces derniers ont donc un rôle important dans la croissance et le développement [324] en tant qu'éléments constitutifs des structures cellulaires. De plus, certains acides gras et leurs dérivés jouent le rôle de molécules de signalisation et de régulateurs de nombreux facteurs de transcription [87, 48, 284]. Parmi eux, les dérivés des acides gras de la famille $\omega 3$ et $\omega 6$ sont par exemple les précurseurs d'hormones lipophiles telles que les leucotriènes, les prostaglandines et thromboxanes ayant un rôle dans l'inflammation, la vasomotricité ou dans l'agrégation des plaquettes.

Or l'homme est incapable de synthétiser les précurseurs des acides gras des familles $\omega 3$ et $\omega 6$. Ces derniers (l'acide linoléique noté C18 :2 $\omega 6$ et l'acide alpha-linolénique noté C18 :3 $\omega 3$) sont qualifiés d'essentiels car ils doivent être apportés par l'alimentation. Par exemple, la carence en acide alpha-linolénique ou en acides gras de la famille $\omega 6$ entraîne une altération de la structure membranaire à l'origine de problèmes dans le développement et le fonctionnement du cerveau [34]. En outre, le déséquilibre du rapport $\omega 3/\omega 6$ induit une augmentation du risque de maladies cardio-vasculaires [146, 159], d'inflammation [228] et d'autres pathologies [296, 294].

C'est pourquoi la composition en acides gras des lipides joue un rôle important dans la valeur nutritionnelle des produits alimentaires, en particulier des produits carnés et aquacoles que nous consommons largement. La composition en acides gras des lipides intervient aussi sur l'aptitude des viandes à la transformation puisqu'elle détermine le point de fusion des graisses et conditionne ainsi la fermeté des tissus [115]. De plus, les viandes riches en acides gras poly-insaturés présentent un degré de peroxydation beaucoup plus important que les viandes riches en acides gras saturés, ce qui entraîne des problèmes de rancissement [38].

La quantité totale de lipides dans les tissus animaux et humains et leur composition en acides gras est la résultante finale de leur métabolisme. Ce métabolisme est particulièrement complexe par le nombre de voies métaboliques concernées, son interaction avec le métabolisme d'autres familles biochimiques (protéines, glucides) et l'intervention de nombreux facteurs de régulation génétiques, hormonaux et métaboliques. De plus, il est extrêmement réactif aux conditions nutritionnelles (jeûne [163], alimentation [54, 144, 42], sur-alimentation [242], ...) et aux variations des besoins énergétiques liés aux états physiologiques (gestation, allaitement [17, 220] exercice physique [147, 261] ...) et pathologiques. Les voies biochimiques concernées ont fait l'objet de nombreuses publications, et plus récemment de présentations dans des bases de données en lignes telles que KEGG¹. Si les voies de régulation ont été généralement bien décrites, notamment dans les espèces animales modèles, leurs interactions et leur importance relative vis-à-vis de la quantité en lipides tissulaires, de leur répartition dans les différents tissus de l'organisme et de leur composition restent mal connues.

La modélisation est un outil qui s'impose progressivement dans la littérature (c.f. tableau 0.0.1) pour

¹<http://www.genome.jp/kegg/>

décrire, prédire et comprendre les systèmes complexes. C'est pourquoi la modélisation devrait permettre d'intégrer les données du métabolisme des lipides et de mieux comprendre son fonctionnement. Cette thèse a donc pour objet de construire deux modèles du métabolisme des lipides : a) un modèle dynamique et générique, décrivant de manière la plus simple possible la hiérarchie d'importance des voies du métabolisme des lipides grâce à quelques éléments clefs, et b) un modèle qualitatif le plus générique possible dont l'objectif est de réunir sous un même formalisme un maximum d'informations afin d'en extraire les éléments importants.

TAB. 0.0.1 – **Estimation du nombre de publications en biologie relatives à la modélisation.**

| années | nombre de publications | | rapport ($\times 10^{-3}$) |
|-----------|------------------------|---|------------------------------|
| | total | concernant la modélisation ^a | |
| 1990-1994 | 2 062 965 | 2 437 | 1.18 |
| 1995-1999 | 2 285 585 | 3 866 | 1.69 |
| 2000-2004 | 2 822 104 | 6 866 | 2.43 |
| 2005-2009 | 3 299 611 | 10 836 | 3.28 |

^aLa recherche a été effectuée dans Pubmed le 19 aout 2009. Les mots clefs *modelling* OR "*model making*" ont été recherchés pour chaque tranche d'années. Le terme *model* n'a pas été recherché car il a d'autres sens comme dans l'expression *animal model*

Une introduction bibliographique rappellera les principaux éléments du métabolisme des lipides dans le règne animal et de sa régulation, puis abordera les différentes méthodes de modélisation. Le reste de la thèse présentera la conception et la mise en œuvre des deux modèles. Enfin, une dernière partie abordera quelques éléments de discussion et les principales perspectives de ce travail.

Revue bibliographique

Première partie

Le métabolisme des Lipides

Chapitre 1

Le métabolisme des lipides : une composante du métabolisme énergétique

1.1 Les principales voies du métabolisme énergétique

L'alimentation des organismes animaux est ponctuelle, alors que ces derniers ont un besoin permanent en énergie afin d'assurer leur métabolisme basal, ainsi que leurs besoins particuliers (croissance, effort, reproduction ...). En fonction de leur balance énergétique, les animaux peuvent alors mettre en réserve l'énergie alimentaire lorsqu'elle est disponible ou utiliser ces réserves lors des phases de besoin et de restriction alimentaire. L'ensemble de ces fonctions est réalisé par les réactions biochimiques présentées figure 1.1.1. La régulation de ce métabolisme en réponse à l'alternance des phases d'alimentation et de jeûne est présentée au chapitre 3.

FIG. 1.1.1 – Principales réactions biochimiques du métabolisme énergétique.

Les molécules énergétiques issues de l'alimentation (les glucides principalement sous forme d'amidon, les protéines, et les lipides principalement sous forme de triglycérides) sont essentiellement digérées en glucose, acides aminés et acides gras. Ces nutriments sont absorbés par l'intestin. Puis, en fonction de la balance énergétique, ils sont soit catabolisés directement soit mis en réserve. Les voies métaboliques principalement actives lors du jeûne sont en rouge, celles principalement actives à l'état nourri sont en bleu, les autres sont en noir.

Production de réserves énergétiques Le glucose (et le lactate, non détaillé sur la figure) peuvent être stockés dans le foie et les muscles sous forme d'un polymère de glucose : le glycogène synthétisé lors de la glycogénogenèse. Le foie et le tissu adipeux (mais aussi de nombreux autres tissus comme les muscles) sont capables de synthétiser des triglycérides à partir d'acétyl-coenzyme A issu de la glycolyse ou du catabolisme des acides aminés, de NADPH issu du cycle des pentoses et d'ATP. La première étape est la synthèse de l'acide palmitique (C16 :0) (c.f. 2.4.1). Le C16 :0 peut éventuellement être transformé en d'autres acides gras par élongation ou désaturation (non détaillé ici, c.f. figure 2.4.2). La seconde étape est l'estérification de 3 acides gras à un glycérol (non détaillé sur la figure) afin de former des triglycérides. Ces derniers peuvent soit être stockés au sein même de ces tissus, soit exportés dans le sang pour un stockage ou une utilisation dans d'autres organes (c.f. paragraphe 2.2).

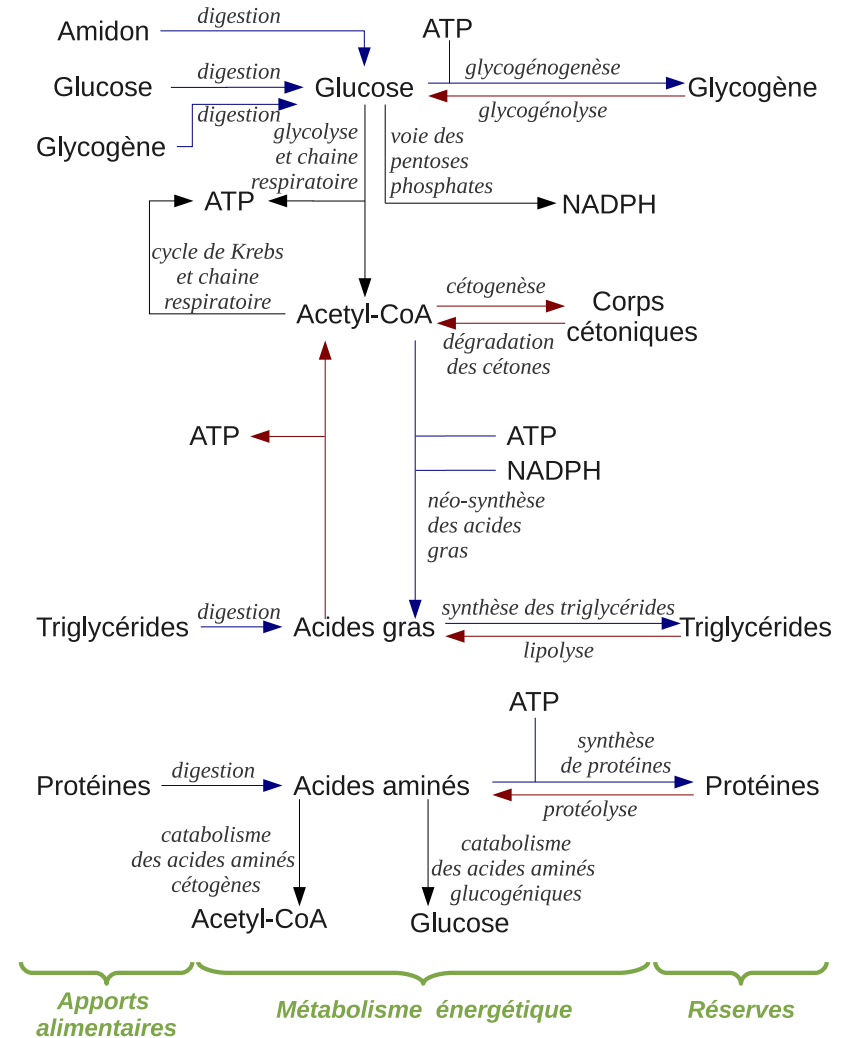
Les acides aminés issus de la digestion des protéines sont généralement utilisés pour la synthèse de protéines dont le rôle est le plus souvent structural et fonctionnel. L'excès d'acides aminés peut être converti en acétyl-coenzyme A pour les acides aminés cétogènes et en glucose pour les acides aminés glucogènes. Le glucose ainsi créé peut servir à rééquilibrer la glycémie lors du jeûne ou être stocké sous forme de glycogène. L'acétyl-coenzyme A est utilisé quant à lui pour la néo-synthèse de lipides, le cycle de Krebs qui fournit de l'ATP et le cycle pyruvate-malate (non représenté) qui fournit du NADPH.

Utilisation des réserves et production d'énergie Lors de la glycogénolyse, le foie et les muscles sont capables de cataboliser le glycogène qu'ils ont stocké afin de produire du glucose-6-phosphate (le premier intermédiaire de la glycolyse et du cycle des pentoses). Le glucose-6-phosphate peut être utilisé localement par la glycolyse afin de satisfaire les besoins de ces tissus. En anaérobie, cette dernière produit du lactate (non représenté) qui sera retransformé en glucose par le foie ; en aérobie la réaction se poursuit par le cycle de Krebs. Seul le foie est capable de transformer le glucose-6-phosphate en glucose grâce à la glucose-6-phosphatase (G6PC) puis de l'exporter dans le sang.

Pour être utilisés, les triglycérides sont hydrolysés en glycérol et en acides gras. Ces derniers peuvent être oxydés dans le même tissu (foie ou muscle) ou exportés dans le sang (par le tissu adipeux notamment) sous forme d'acides gras libres. Le glycérol est converti en glucose (non représenté) par le foie. Les acides gras circulant sont captés par les organes susceptibles de les utiliser comme le foie ou les muscles. Ils sont ensuite oxydés (c.f. 2.5) en acétyl-coenzyme A et en ATP. Dans le foie, lorsque l'acétyl-coenzyme A est produit plus rapidement qu'il est utilisé par le cycle de Krebs, il y a production de corps cétoniques par la cétogénèse. Ces derniers sont exportés dans le sang et servent de métabolite de substitution à l'acétyl-coenzyme A dans de nombreux autres organes (y compris des organes incapables d'oxyder les acides gras), ce qui court-circuite la glycolyse et permet donc une économie de glucose. Lorsque la quantité de cétones présente dans le sang est très importante, les cétones sont excrétées par les poumons et les reins.

Lors du jeûne, l'organisme est aussi capable de mobiliser ses réserves protéiques telles que les protéines "labiles" des viscères, ou en cas de jeûne important les protéines des fibres musculaires. Les protéines sont alors dégradées en acides aminés par protéolyse. Les acides aminés glucogènes sont transformés en glucose par néoglucogénèse, et les cétogènes en acétyl-coenzyme A.

Figure simplifiée réalisée à partir de Metabolism at Glance de J.G. Salway [271]



1.2 Les molécules de réserve en énergie

Les molécules de réserve appartiennent à trois familles biochimiques : les glucides, les lipides et les protéines. En plus d'être des stocks énergétiques, ces molécules ont des rôles structuraux ou fonctionnels que nous présenterons brièvement.

1.2.1 Le glucose et ses polymères

Le glucose est un métabolite crucial pour la survie de l'organisme car de nombreuses cellules comme les hématies ou les cellules nerveuses ne peuvent pas survivre sans dégrader du glucose. Le métabolisme de l'énergie passe donc par le maintien de la glycémie, c'est-à-dire de la concentration en glucose sanguin.

Cependant, le sang ne contient qu'une faible quantité de glucose, et ce dernier ne peut être stocké directement dans les cellules car son accumulation entraîne une forte augmentation de la pression osmotique intracellulaire. Le glucose est donc stocké dans les cellules sous la forme de polymères : l'amidon dans les plantes et le glycogène dans les tissus animaux. La consommation d'amidon représente une source d'énergie importante chez les animaux se nourrissant principalement de céréales comme le poulet ou le porc par exemple.

Le stockage du glucose en ses polymères a un très fort rendement énergétique¹ et son utilisation prioritaire limite donc les pertes d'énergie. Ainsi le glycogène est la première réserve synthétisée lors d'excès de glucose et la première dégradée lors d'une diminution de la glycémie. En revanche, le glycogène possède une densité énergétique² relativement faible (4.2kJ/g^{-3}) ce qui limite les quantités d'énergie stockée sous cette forme dans les muscles et le foie des animaux.

Le foie, grâce à l'activité de la glucose-6-phosphatase (G6PC) est capable de transformer le glucose-6-phosphate issu du catabolisme du glycogène en glucose, puis d'exporter ce glucose dans le sang, ce qui contribue à maintenir la glycémie. En revanche, les muscles ne sont pas capables d'opérer cette transformation en raison de la très faible activité de leur G6PC [337] et utilisent leur glucose-6-phosphate *in situ*.

1.2.2 Acides gras et triglycérides

Les triglycérides sont constitués d'une molécule de glycérol sur laquelle trois acides gras sont estérifiés. En raison de leur densité énergétique (39kJ/g [308]) beaucoup plus élevée que celle du glycogène, les triglycérides constituent la plus grande réserve énergétique des animaux.

L'ingestion de triglycérides représente une source d'énergie importante issue de l'alimentation, en particulier pour l'homme vivant dans des pays industrialisés. Lors de la digestion, les triglycérides sont hydrolysés en acides gras qui sont absorbés par l'intestin. Ce dernier les estérifie à nouveau sous forme de triglycérides qu'il libère dans la circulation sous forme de chylomicrons (c.f. 2.2.2). Après captage par les tissus, les acides gras circulants (c.f. 2.2.4) peuvent y être stockés ou immédiatement oxydés. Les acides gras peuvent aussi être néo-synthétisés par l'organisme à partir d'acétyl-coenzyme A, d'ATP et de NADPH (c.f. 2.4.1).

Les acides gras sont de longues chaînes polycarbonées se terminant par une extrémité COOH. La chaîne de carbone peut posséder une ou plusieurs insaturations⁴. Les acides gras sont généralement notés $C_n:i\omega_j$ où n représente le nombre de carbones, i le nombre d'insaturations et j la position de la dernière insaturation à partir de l'extrémité CH_3 .

Les acides gras sont apportés par l'alimentation sous forme de triglycérides ou synthétisés *de novo* par l'organisme. Les animaux sont capables de néo-synthétiser tous les acides gras des familles ω_7 et ω_9 , mais incapables de produire l'acide α -linoléique ($C_{18}:3\omega_3$) et l'acide linoléique ($C_{18}:2\omega_6$) en raison de l'ab-

¹Rendement énergétique = énergie récupérée après stockage / énergie récupérée lors d'une consommation directe

²Densité énergétique = énergie récupérée/masse

³Dans la cellule, 1g de glycogène est lié à environ 3g d'eau d'où une densité énergétique du stock de 4.2kJ/g [103] alors que la densité du glycogène pur est d'environ 17kJ/g .

⁴Une insaturation est une double liaison entre deux carbones.

sence des enzymes Delta-15-desaturase et Delta-12-desaturase (présentes chez les végétaux par exemple). L'organisme est par contre capable de transformer ces différents acides.

Les triglycérides possèdent une densité énergétique supérieure à celle du glycogène, mais un rendement énergétique beaucoup plus faible [327]. Les animaux utilisent des stocks finis de glycogènes comme réserve à court terme (première réserve synthétisée, première réserve consommée), et stockent tous le surplus de glucose sous forme de triglycérides qu'ils dégradent uniquement lorsque une bonne partie du glycogène est consommé. Ils tirent ainsi le meilleur parti des deux types de stocks.

1.2.3 Les protéines

Les protéines ont des rôles essentiellement structuraux (ex : kératine [91], myosine et actine [60,263] . . .) et fonctionnels importants (ex : enzymes, hormones [265] . . .). De plus, la synthèse de protéines (en particulier lors de la fabrication d'acides aminés non essentiels) et le renouvellement des protéines existantes constituent des mécanismes très coûteux en énergie. Hormis les protéines labiles des viscères qui sont facilement consommées à des fins énergétiques, les réserves protéiques telles que les protéines musculaires constituent généralement un stock d'énergie de secours, au cas où le glucose, le glycogène et les lipides viendraient à manquer.

Une exception notable est le cas des organismes carnivores stricts comme de nombreux poissons. Les protéines sont alors pour eux une source d'énergie importante : les acides aminés issus de la digestion des protéines sont soit retransformés en protéines dans l'organisme, soit convertis en acétyl-coenzyme A ou en glucose afin d'assurer la fourniture d'énergie ou la néo-synthèse des lipides ou du glycogène.

1.3 Les principaux organes du métabolisme énergétique

1.3.1 Lieux de dégradation et de stockage des réserves énergétiques

Toutes les cellules possèdent une activité catabolique permettant la fourniture d'énergie. Certaines cellules utilisent exclusivement du glucose comme les hématies, tandis que d'autres comme les cellules musculaires sont capables de produire de l'énergie à partir de divers substrats (glucose, corps cétoniques, acides gras).

Ces substrats proviennent soit du catabolisme des molécules alimentaires (glucose, acides gras, protéines), soit de réserves énergétiques. Les lieux de synthèse et de stockage du glycogène (la forme de réserve du glucose) sont identiques d'une espèce animale à l'autre, tandis que ceux concernant la mise en réserve et l'utilisation des lipides varient en fonction de l'espèce et de l'âge de l'animal (c.f. tableau 1.3.1).

1.3.2 Coopération entre organes

A l'exception du glycogène musculaire qui est utilisé *in situ*, les autres substrats énergétiques (acides gras, glucose) ne sont pas forcément utilisés dans leur lieu de stockage, ce qui implique à la fois un transport par voie systémique et une régulation inter-organes du métabolisme énergétique. Les principaux mécanismes de régulations seront présentés au chapitre 3. Une description synthétique des rôles des principaux organes est présentée dans la figure 1.3.1.

TAB. 1.3.1 – Rôles des différents organes dans la synthèse et la dégradation des réserves énergétiques en fonction des espèces. Nous considérons dans cette analyse les mammifères et les oiseaux.

| Organe | Foie | Tissus adipeux | Muscle |
|-------------------------------------|---|---------------------|--------|
| Fonction | | | |
| Synthèse de glycogène ^a | Toutes | Aucune | Toutes |
| Stockage de glycogène | Toutes | Aucune | Toutes |
| Exportation de glucose ^b | Toutes | Aucune | Aucune |
| Synthèse d'acides gras | | | |
| Forte chez | Homme [85], rat [109], poulet [204] | Porc [240], bovins | |
| Faible chez | Porc (sauf chez les jeunes porcs) | Homme, rat, poulet. | Tous |
| Stockage d'acides | | | |
| Forte chez | Les oiseaux migrateurs, | Toutes | |
| Faible chez | La plupart des oiseaux (sauf oie, canard ...) | Aucune | Tous |
| Cétogenèse ^c | Toutes | Aucune | Aucune |

^aLe glycogène ne circule pas tel quel dans le sang, il est toujours stocké puis utilisé par l'organe qui le synthétise, ou exporté sous forme de glucose par le foie.

^bLe foie est le seul organe possédant la glucose-6-phosphatase (G6PC), permettant de transformer le glucose-6-phosphate (issu du glycogène ou de la néoglucogenèse) en glucose et d'exporter ce dernier dans le sang. Le glucose exporté dans le sang peut être utilisé par d'autres organes.

^cLe foie est le seul organe possédant HMGCS2, l'enzyme clef de la cétogenèse

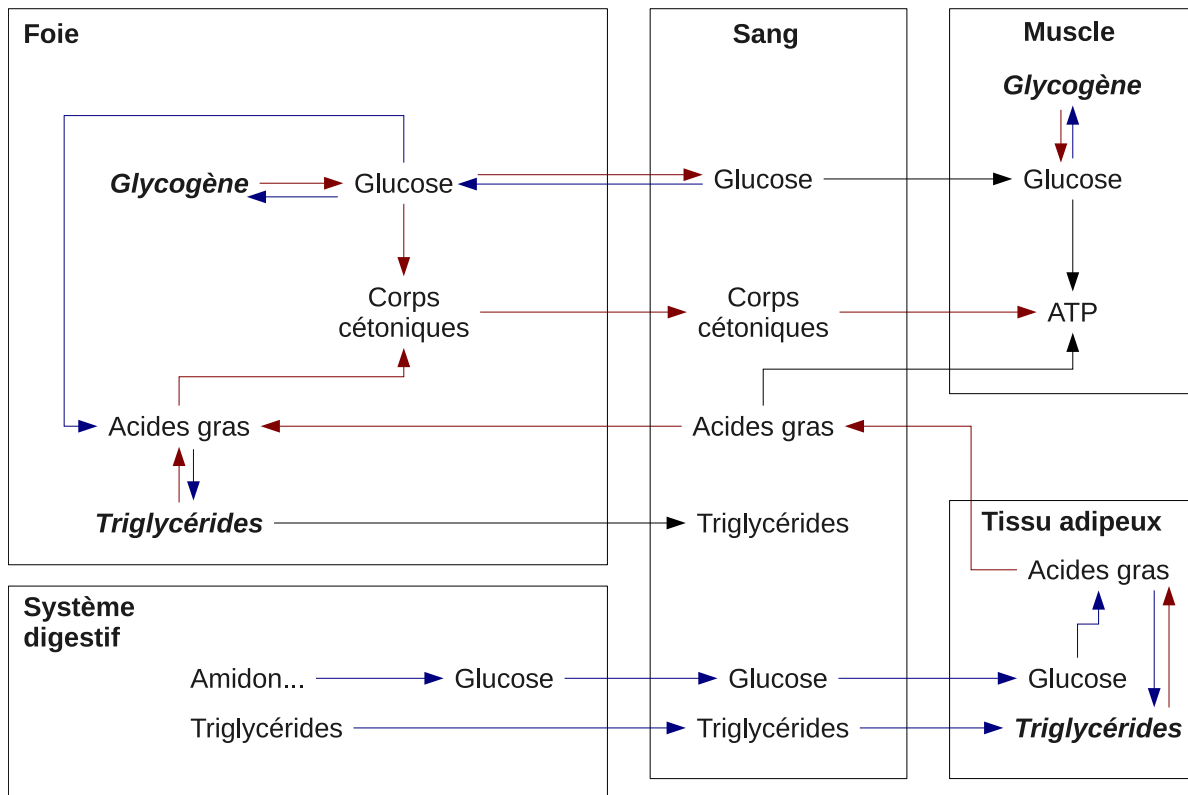


FIG. 1.3.1 – **Coopération entre organes du métabolisme énergétique : métabolisme des glucides et des lipides** Cette figure résume les principales voies du métabolisme énergétique ainsi que leur répartition entre les différents organes. Les voies actives lors du jeûne sont en **rouge**, celles actives à l'état nourri en **bleu** et les autres en **noir**. Tous les organes (non représentés) ont un catabolisme actif et dégradent du glucose, des corps cétoniques et des acides gras. La consommation de ces métabolites varie en fonction de l'organe considéré et des conditions physiologiques.

Chapitre 2

Généricité des réactions biochimiques du métabolisme des lipides

Les réactions biochimiques participant au métabolisme des lipides sont communes aux différentes espèces animales et la majorité d'entre elles sont présentes dans l'ensemble des tissus ce qui confère au métabolisme des lipides une forte généralité.

2.1 Digestion des triglycérides et absorption des acides gras

2.1.1 Digestion des triglycérides et absorption par l'intestin

Les triglycérides issus de l'alimentation ne peuvent pas franchir les membranes cellulaires et doivent donc être hydrolysés au préalable par la lipase pancréatique (PNLIP) associée à la colipase pancréatique (CLPS) en acides gras libres à longue chaîne (de plus de 18 carbones) et en 2-monoglycérides. Cette hydrolyse nécessite le passage en émulsion des gouttelettes lipidiques grâce aux sels biliaires produits par le foie.

Les acides gras libres et les 2-monoglycérides sont alors absorbés dans les entérocytes. Comme dans les autres cellules des mammifères, les acides gras franchissent la membrane par simple diffusion, grâce à un mécanisme de *flip-flop* [133] ou par diffusion facilitée grâce à un ensemble de protéines telles que la *plasma membrane fatty acid-binding protein* (FABPpm), SLC27A4 et la *fatty acid translocase murine* homologue de l'antigène CD36 humain (FAT/CD36) présentes dans tous les types cellulaires chez les mammifères [1, 251]. Dans l'intestin, le passage des acides gras à travers la membrane est aussi facilité par leur protonation qui diminue leur solubilité dans les micelles. La protonation est la conséquence de l'environnement acide créé à la surface des entérocytes par une pompe à Na^+/H^+ [24]. Une fois internalisés dans la cellule, les acides gras sont pris en charge par des protéines de transport spécifiques. Ces dernières sont décrites dans la partie 2.3. En parallèle l'intestin absorbe du cholestérol [4] et des acides gras à chaîne courte. Ces derniers diffusent au travers de l'intestin vers la veine porte et seront captés par le foie.

2.1.2 Estérification et synthèse de lipoprotéines

Les acides gras de 18 carbones et plus sont immédiatement réestérifiés en triglycérides. Ces derniers ainsi que le cholestérol sont associés à l'apoprotéine B48 (APO-B48) par le réticulum endoplasmique et l'appareil de Golgi afin de synthétiser des chylomicrons immatures comme décrit par la figure 2.1.1.

2.2 Transport des lipides entre les différents organes

Les lipides sont transportés entre les différents organes de leur lieu d'absorption et de synthèse vers leurs lieux de stockage ou d'utilisation. Comme ce sont des composés insolubles dans l'eau, ils ne peuvent pas

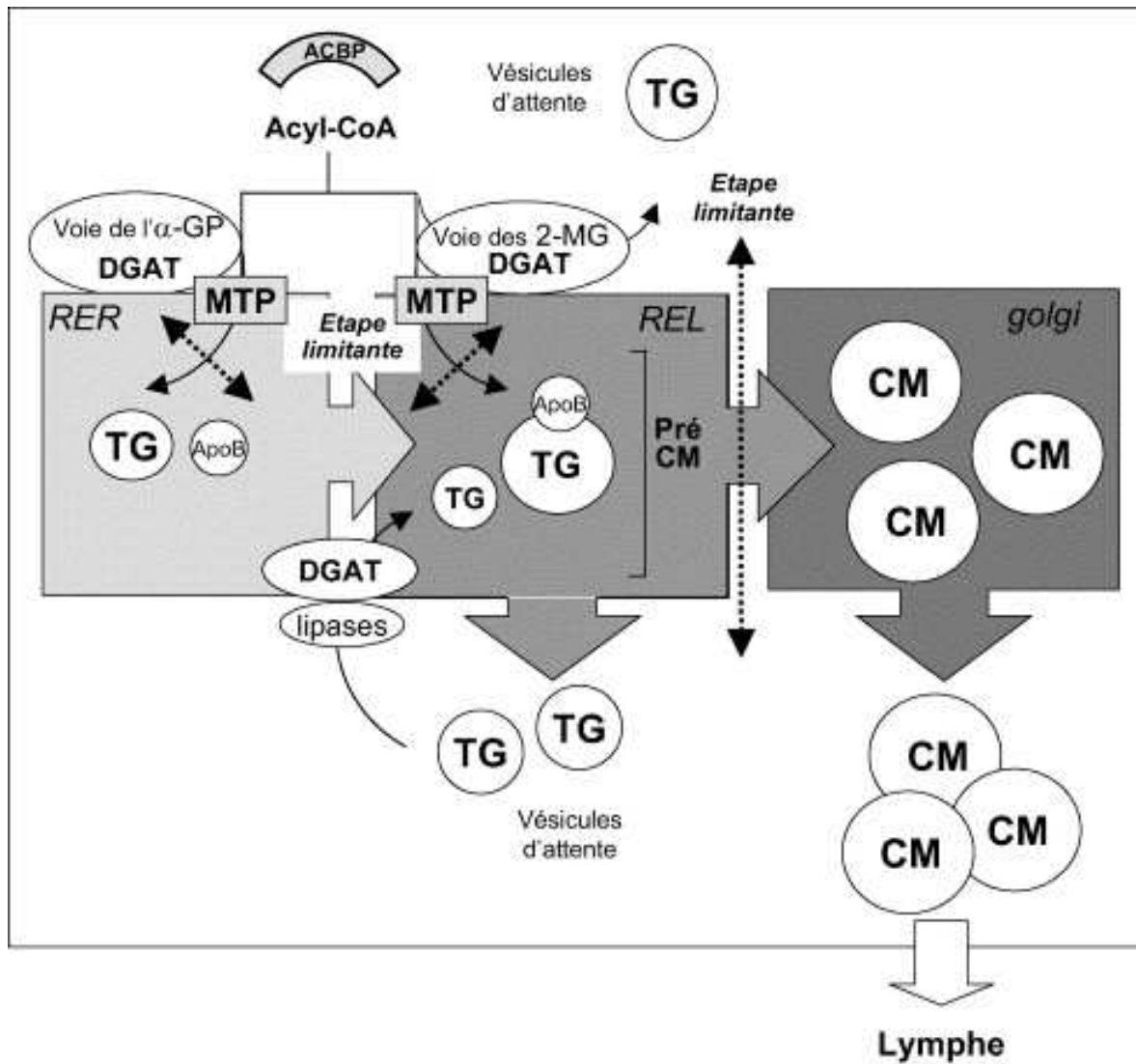


FIG. 2.1.1 – Principales étapes de la synthèse des lipoprotéines. 1) Estérification des acyl-CoA en triglycérides (TG) par les voies de l'α-glycérophosphate (α-GP) et des 2-monoglycérides (2-MG) localisées respectivement à la surface du réticulum rugueux et lisse (RER; REL); 2) transfert des TG dans le réticulum via la microsomal triglyceride transfer protein (MTP); 3) échange entre le RER et le REL. Synthèse des préchylomicrons (Pré-CM); 4) maturation des chylomicrons (CM) dans le golgi. DGAT : diacylglycérol-acyltransférase Apo B, apoprotéine B48; CM, chylomicrons; TG, triglycérides. **Figure et légende extraites de Petit *et al.*, 2007 [251].**

être transportés par simple diffusion dans le plasma. Leur transport entre les organes est facilité par les lipoprotéines [64, 231] et l'albumine [231, 352]. Il se décompose en 4 fonctions : le transport des lipides issus de l'alimentation, le transport des lipides hépatiques et le transport du cholestérol vers le foie assurés par des lipoprotéines [187] ; et le transport des acides gras libérés par les tissus adipeux assuré par l'albumine [231]. La figure 2.2.1 présente de manière synthétique les principales voies biochimiques impliquées dans le transport des lipides.

2.2.1 Les différentes lipoprotéines

A l'exception des acides gras libres, les lipides (cholestérol, triglycérides et phospholipides) sont transportés par des lipoprotéines. Il existe 5 types de lipoprotéines [64] :

- les chylomicrons sont responsables du transport des triglycérides et du cholestérol alimentaire captés par l'intestin vers les tissus périphériques ;
- les *very low density lipoproteins* (VLDL) sont majoritairement synthétisées par le foie et minoritairement par l'intestin afin d'exporter des triglycérides ;
- les *intermediate density lipoproteins* (IDL) et les *low density lipoproteins* (LDL) sont le produit terminal du catabolisme des VLDL. Elles transportent le cholestérol des HDL vers les organes périphériques ;
- les *high density lipoproteins* (HDL) transportent le cholestérol vers le foie, et servent de réserves d'apolipoprotéines.

Les lipoprotéines sont composées de triglycérides, de cholestérol, de phospholipides et d'apolipoprotéines (APO) [64]. Les apolipoprotéines sont généralement regroupées en 6 classes. Dans le cadre de ce paragraphe, nous ne présenterons que les classes ayant un rôle majeur dans le transport des lipides c'est-à-dire les classes A (A1 à A5), B (B48 et B100), C (C1 à C4) et E.

2.2.2 Le transport des lipides issus de l'alimentation

Le transport des triglycérides intestinaux et du cholestérol alimentaire vers les tissus périphériques est assuré chez les mammifères majoritairement par les chylomicrons et minoritairement par les VLDL [231]. Chez le poulet ce transport est assuré par les protomicrons. Ce paragraphe ne décrit que le cas des mammifères.

Pendant la période post-prandiale, l'intestin sécrète dans la circulation des chylomicrons immatures [231]. Ces lipoprotéines riches en triglycérides [231] récupèrent alors des APO-E et des APO-C issues des HDL pour former des chylomicrons matures. Ces derniers, sous l'action de la lipoprotéine lipase (LPL)¹ activée par les APO-C2, libèrent des acides gras et du 2-monoacylglycérol qui sont alors captés par les tissus périphériques (foie, muscle, tissu adipeux ...).

Les mécanismes de transport des acides gras à travers la membrane plasmique sont sujets à controverse. Deux mécanismes potentiels de transport des acides gras, un saturable et un insaturable ont été identifiés ([186] pour une revue). Le 2-monoacylglycérol importé dans la cellule est quant à lui hydrolysé en acide gras et en glycérol par une lipase intracellulaire.

Une fois les chylomicrons appauvris en acides gras, leurs APO-C sont à nouveau transférées vers les HDL. Ils forment alors des résidus de chylomicrons qui sont reconnus par le récepteur au couple APO-B48 - APO-E du foie [313, 352]. Ils rentrent par endocytose dans les hépatocytes puis sont hydrolysés par les lysosomes qui libèrent alors les acides aminés issus des apolipoprotéines et le cholestérol libre issu des esters de cholestérol.

2.2.3 Le transport des lipides endogènes hépatiques et du cholestérol par les VLDL, LDL et HDL

Le foie (et en moindre mesure l'intestin [231]) synthétisent des VLDL immatures à partir de triglycérides captés de l'alimentation ou néosynthétisés, d'APO-A1 [116] et d'APO-B100. Dans le sang, les VLDL immatures vont s'enrichir en APO-E et APO-C [27] au contact des HDL pour former des VLDL matures. Comme

¹Les LPL sont des isoenzymes situées à la surface endothéliale des tissus (comme le coeur, le foie et les muscles [350]).

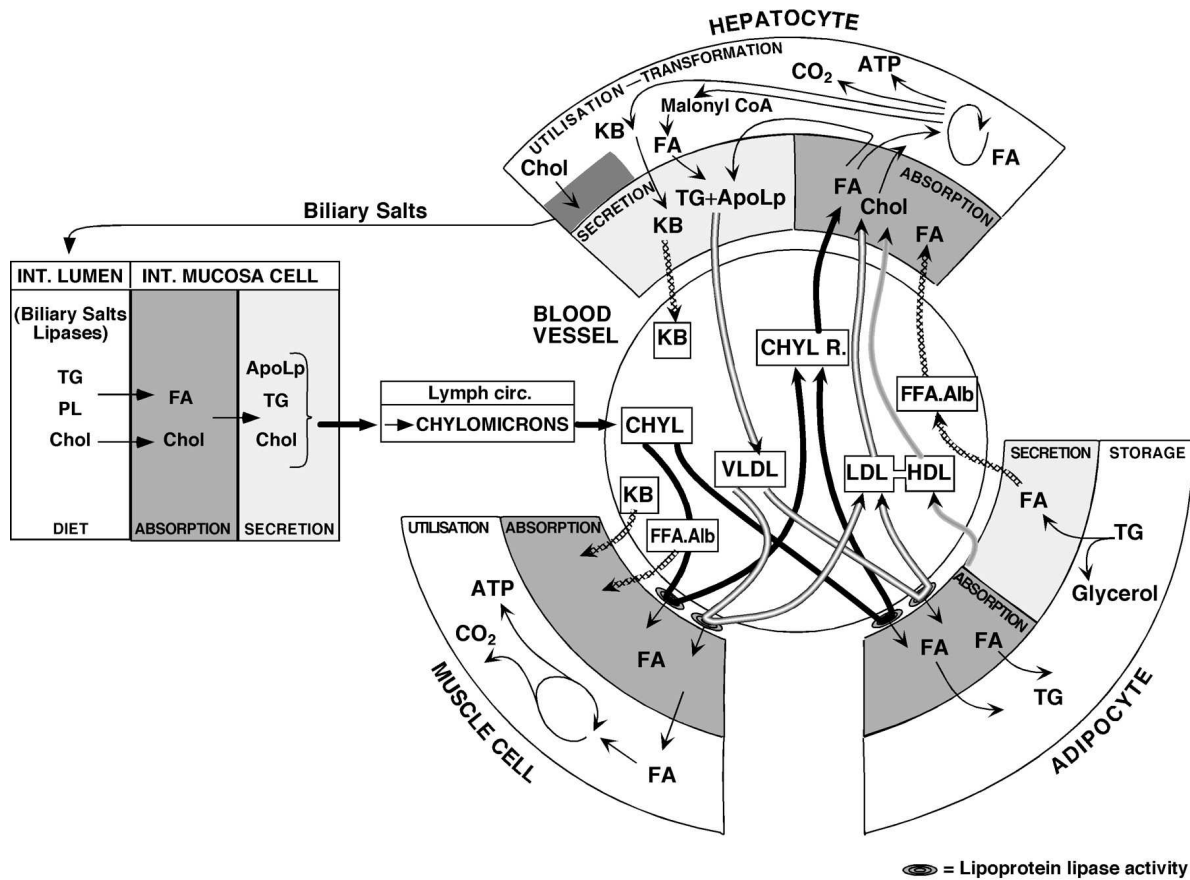


FIG. 2.2.1 – **Transport et utilisation des lipides de la lumière intestinale au foie, tissu adipeux et muscles.** Ce schéma illustre les principaux composants des différentes voies biochimiques impliquées dans le transport des lipides. L'absorption et la transformation des lipides (triglycérides (TG), cholestérol (Chol), phospholipides (PL), acides gras (FA)) par la muqueuse intestinale entraîne la synthèse de chylomicrons (CHYL), qui circulent dans la lymphe avant d'être transportés dans le sang. Les acides gras des chylomicrons, ainsi que les acides gras libres issus de l'action de la lipoprotéine lipase extracellulaire sont captés par de nombreuses cellules. Leur devenir est fonction de l'organe : stockage sous forme de triglycérides (TGs) dans les adipocytes, oxydation pour produire de l'énergie dans les muscles par exemple. Les chylomicrons ainsi appauvris donnent des résidus de chylomicrons (CHYL R.) qui sont extraits hors de la circulation sanguine par les hépatocytes. Les acides gras réestérifiés dans les hépatocytes sont sécrétés sous forme de triglycérides associés aux apolipoprotéines (ApoLp) sous forme de *very low density lipoproteins* (VLDL). Les *low density lipoproteins* (LDL) sont appauvries en acides gras par l'action de la lipoprotéine lipase présente à la surface de l'endothélium de nombreuses cellules, et enrichies en cholestérol provenant des *high density lipoproteins* (HDL); elles sont ensuite captées par les hépatocytes et participent donc au transport du cholestérol des tissus périphériques vers le foie. En cas de demande énergétique importante, la lipolyse dans les adipocytes déclenche la libération d'acides gras libres (FFA) dont la plus grande partie est transportée dans le sang complexée à l'albumine (FFA.Alb). Les acides gras libres sont captés par les hépatocytes qui les utilisent afin de produire des corps cétoniques (KB). Les corps cétoniques servent alors de carburant pour les tissus périphériques (muscle, cerveau, rein, ...). **Figure et légende d'après B. Desvergne *et al.*, 1999 [83], traduction Pierre Blavy.**

pour les chylomicrons, la présence d'APO-C2 permet l'activation de LPL entraînant la libération d'acides gras [350].

Suite à leur appauvrissement en lipides, la densité des VLDL augmente, ce qui les transforme en IDL puis en LDL. Elles transfèrent alors leurs APO-C, APO-E, triglycérides et phospholipides aux HDL en échange d'ester de cholestérol. L'échange de lipides est catalysé par l'enzyme CETP.

Les LDL ainsi formées transportent essentiellement du cholestérol qu'elles déposeront à la surface des cellules. Ces lipoprotéines peuvent aussi être reconnues par le récepteur à APO-B100. Elles sont alors endocytées et dégradées en acides aminés, cholestérol, acides gras et en d'autres résidus. L'ensemble de ces produits est utilisé par la cellule.

Le transport inverse du cholestérol vers le foie est assuré par les HDL. Les HDL sont des lipoprotéines synthétisées dans le foie à partir d'APO-A, E et C2 puis libérées dans le sang. Au fil de leur parcours, elles s'enrichissent en cholestérol à partir du cholestérol libre présent sur les membranes des cellules des tissus qu'elles traversent. Les HDL riches en cholestérol retournent dans le foie qui les absorbe par endocytose. Les esters de cholestérol sont alors hydrolysés, ce qui libère du cholestérol utilisé par le foie pour la synthèse d'hormones, d'autres lipoprotéines ou de sels biliaires par exemple. Les HDL jouent aussi le rôle de réserve d'échange d'apolipoprotéines, puisqu'elles échangent de l'APO-E et des APO-C avec les chylomicrons et les VLDL.

2.2.4 Le transport des acides gras libres issus du tissu adipeux

Lors du jeûne, le tissu adipeux riche en triglycérides hydrolyse ces derniers lors de la lipolyse en glycérol et en acides gras. Les acides gras sont alors libérés en grande quantité dans le sang [231] dans lequel ils sont transportés sous forme complexés avec l'albumine [307]. Ils sont ensuite captés très rapidement par les cellules [231] qui ont besoin d'énergie.

2.3 Transport intracellulaire des lipides

A l'intérieur du cytosol, le transport des acides gras est réalisé principalement par des *fatty acid binding protein* (FABP) [310]. Sept types de FABP ont été identifiés. Ils diffèrent par leurs affinités aux divers acides gras et par l'organe dans lequel ils sont exprimés (d'après GeneCards, FABP1 est exprimé dans le foie, l'intestin et le rein [329] FABP2 dans l'intestin, FABP3 dans le muscle et le coeur, FABP4 dans les adipocytes, FABP5 dans les cellules de la peau, FABP6 dans l'iléon et FABP7 dans le cerveau). Ceci laisse supposer un rôle de cette répartition dans les différences de réponses de ces organes au taux d'acides gras circulants et dans le devenir des acides gras au sein de ces organes.

2.4 Synthèse de novo et transformation des acides gras

2.4.1 La néosynthèse des acides gras

Les organismes animaux (autres que l'homme) consomment relativement peu de graisses alimentaires. De ce fait, ils ont développé des stratégies biochimiques leur permettant de synthétiser de novo les acides gras, afin de satisfaire leurs besoins.

La synthèse de novo d'acides gras est réalisée dans le cytosol des cellules. Elle utilise du CHO_3^- , de l'acétyl-coenzyme A issu de la glycolyse et du NADPH issu du cycle des pentoses phosphates. Ces deux dernières voies ont comme précurseur commun le glucose. La néosynthèse d'acides gras est réalisée sous le contrôle principal de deux enzymes l'acétyl-CoA carboxylase (ACC) et la fatty acid synthase (FASN). Elle nécessite la dérivation de l'acétyl-coenzyme A du cycle de Krebs mitochondrial pour une utilisation cytosolique. L'acétyl-coenzyme A ne peut pas sortir directement de la mitochondrie, il est d'abord transformé en citrate, exporté sous cette forme puis reconverti en acétyl-coenzyme A dans le cytosol. Le détail des réactions biochimiques est présenté figure 2.4.1. Ces réactions sont communes à l'ensemble des cellules capables de néosynthétiser des acides gras comme les hépatocytes et les adipocytes par exemple.

Une fois synthétisés, les acides gras peuvent être transformés en d'autres acides gras (c.f. figure 2.4.2) ou stockés sous forme de triglycérides. Si la cellule fonctionne normalement, ils ne sont pas directement dégradés ce qui constituerait un cycle futile.

FIG. 2.4.1 – Néo-synthèse des lipides.

La synthèse de novo des lipides proprement dite est réalisée par 2 enzymes : l'acétyl-coenzyme A carboxylase (ACC) et la *fatty acid synthase* (FASN). ACC est responsable de la production de malonyl-coenzyme A à partir d'acétyl-coenzyme A, de HCO_3^- et d'ATP. FASN est un complexe multienzymatique composé de 7 sous unités notées "FASN" suivi de la fonction de la sous unité, ainsi que d'un site de fixation des acyls en cours de synthèse : l'acyl-carrier protein (ACP).

La synthèse d'acides gras se déroule en 5 étapes. La première est la production de malonyl-coenzyme A par ACC. Cette étape se poursuit tout au long de la synthèse d'acides gras car chaque fois que la chaîne carbonée est allongée de 2 carbones, il y a consommation d'un malonyl-coenzyme A. Ceci confère à ACC un rôle clef dans la néo-synthèse des acides gras.

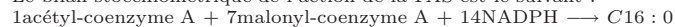
La seconde étape est l'initiation de la synthèse proprement dite par FASN : une molécule d'acétyl-coenzyme A est fixée sur l'ACP par la sous unité acyltransferase, (libérant un coenzyme ASH non représenté).

La troisième étape est l'élongation de l'acétyl-coenzyme A (une molécule à 2 carbones), en acide palmitique (C16:0) un acide gras à 16 carbones. Elle se déroule en 7 tours, chacun rajoutant 2 carbones à la chaîne carbonée fixée à l'ACP. L'initiateur de la chaîne est l'acétyl-coenzyme A-ACP. Il comporte 2 carbones. La longueur de l'acide gras au début d'un tour est noté par n , et par $n + 2$ à la fin du tour. Les sous unités responsables de l'élongation sont les mêmes à chaque tour. Tout d'abord, la sous unité acyltransferase fixe un malonyl-coenzyme A à l'ACP, (ce qui libère un coenzyme ASH non représenté). Ensuite, la sous unité β -cétolacetyl-ACP synthase catalyse l'ajout du malonyl-coenzyme A à la chaîne carbonée fixée à l'ACP pour former du β -cétolacetyl-ACP c'est à dire une molécule de $n + 2$ carbones. Cette réaction entraîne la libération d'un CO_2 et libère un emplacement de l'ACP.

La quatrième étape est la réduction de la chaîne par la sous unité β -cétolacetyl-ACP réductase ce qui consomme un NADPH (et libère un NADP non représenté). Cette molécule est alors déshydratée par la sous unité 3D-hydroxy-acyl-ACP déshydratase (ce qui libère une molécule d'eau non représentée). La dernière réaction est la réduction par l'énol-ACP reductase de l'hydroxy-acyl-ACP formant une chaîne carbonée de $n+2$ carbones lié à l'ACP. Cette chaîne va alors réaliser d'autres tours, tant que sa longueur n'est pas de 16 carbones. Le passage d'un nouveau tour est représenté par la flèche annotée $n = n + 2$.

La dernière étape a lieu lorsque la chaîne carbonée liée à l'ACP a atteint une longueur de 16 carbones : la *fatty acyl-ACP thioesterase* va briser la liaison reliant cette chaîne à l'ACP, libérant de l'acide palmitique (C16:0).

Le bilan stoechiométrique de l'action de la FAS est le suivant :

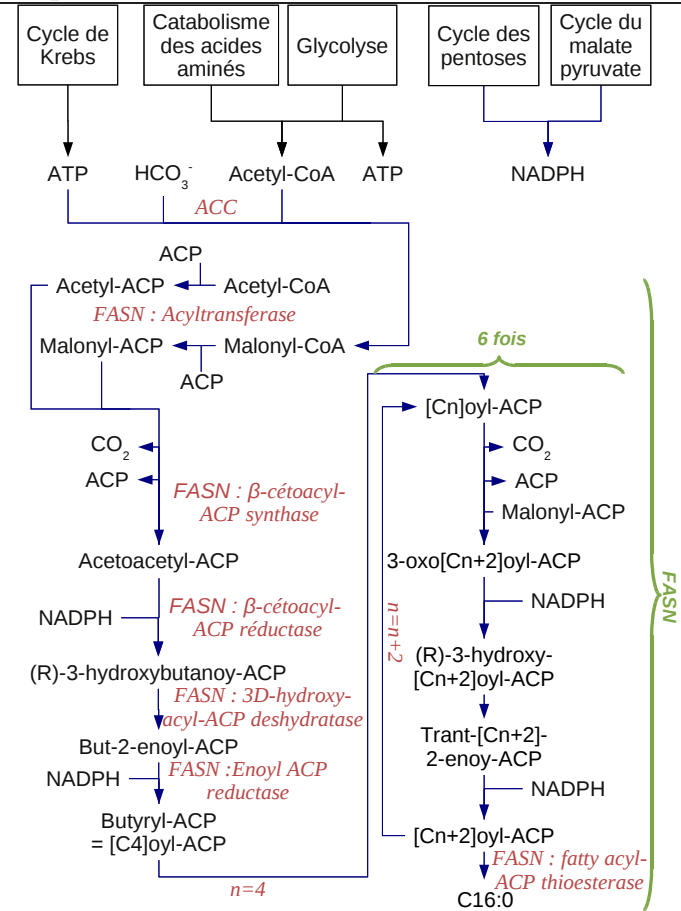


Si on intègre à ce bilan la synthèse de malonyl-coenzyme A par ACC il devient :



Le coenzyme ASH, l'eau et le NADP ne sont pas représentés dans ces bilans.

Figure simplifiée réalisée à partir de *Metabolism at Glance* de J.G. Salway et al., [271] et KEGG



2.4.2 Transformations des acides gras

Les acides gras issus de l'alimentation, ou ceux synthétisés de novo peuvent subir trois transformations majeures (c.f. figure 2.4.2) : l'élongation, réalisée par des élongases qui rajoute deux carbones sur l'extrémité COOH de l'acide gras, l'oxydation partielle qui retire deux carbones à partir de l'extrémité COOH, et la désaturation. Cette dernière réaction crée une double liaison entre deux carbones de la chaîne. En fonction de l'enzyme qui la réalise, elle ne peut avoir lieu qu'entre les carbones 5 et 6 à partir de l'extrémité COOH² lorsqu'elle est réalisée par delta-5-désaturase (D5D) entre les carbones 6 et 7 lorsqu'elle est réalisée par la delta-6-désaturase (D6D) et entre les carbones 9 et 10 lorsqu'elle est réalisée par la delta-9-désaturase (D9D).

2.5 Dégradation des acides gras

La dégradation des acides gras comporte trois réactions principales : l'oxydation, la consommation d'acétyl-coenzyme A par le cycle de Krebs et la cétogenèse.

2.5.1 L'oxydation se déroule dans la mitochondrie et dans le peroxysome

L'oxydation des acides gras est la première étape. Elle est réalisée dans la mitochondrie ou dans le peroxysome. La mitochondrie oxyde majoritairement les acides gras de 18 carbones ou moins tandis que le peroxysome prend en charge majoritairement des acides gras plus longs [192] dont il raccourcit la chaîne carbonée jusqu'à 18 carbones. Ces acides gras sont ensuite généralement pris en charge par la mitochondrie. Les mécanismes permettant de raccourcir les acides gras sont communs aux deux organites à quelques exceptions : l'oxydation mitochondriale est dépendante de la carnitine pour l'importation des acides gras, ce qui n'est pas le cas de l'oxydation dans le peroxysome, et les mécanismes de régulations diffèrent [213]. La part d'oxydation peroxysomiale peut varier entre les organes et les espèces. Elle est évaluée à 10% de l'oxydation totale dans le foie de rat [213], et 15 à 20% de l'oxydation de l'oléate (C18 :1 ω 9)³ dans le muscle de lapin [119]

2.5.2 Réactions biochimiques de l'oxydation des acides gras

L'oxydation des acides gras saturés est une répétition d'une réaction qui diminue de deux carbones la longueur de chaîne de l'acide gras à partir de l'extrémité COOH et qui produit de l'acétyl-coenzyme A et de l'ATP. Cette réaction peut se poursuivre jusqu'à dégradation complète de l'acide gras ou s'arrêter en cours de route pour produire un acide gras plus petit (on parle alors d'oxydation incomplète). L'acétyl-coenzyme A ainsi produit a deux devenir possibles : le cycle de Krebs ou la cétogenèse. Les principales réactions de l'oxydation des acides gras sont présentées figure 2.5.1.

Le cheminement des acides gras insaturés au cours de l'oxydation dépend de la position de leurs doubles liaisons. Lorsque elle se situe en position cis- Δ 3, elle est d'abord déplacée en position cis- Δ 2 par la dodecenoyl-CoA isomerase(D3ECI). Ceci forme du 2-(E)-enoyl-CoA, puis la β -oxydation se poursuit normalement. Lorsque l'acide gras contient une double liaison en position Δ 4, sa déshydrogénation donne un 2,4-dienoyl-CoA qui n'est pas un substrat de la 2-enoyl-CoA hydratase. Pour poursuivre la β -oxydation, la 2,4-dienoyl-CoA reductase (DECR1) utilise un NADPH afin de réduire le 2,4-dienoyl-CoA en trans-D3-enoyl-CoA. Ce dernier est alors converti en 2-(E)-enoyl-CoA par la D3ECI puis la β -oxydation se poursuit normalement.

Chez les animaux, l'oxydation des acides gras peut passer par des voies alternatives telles que l' ω -oxydation et l' α -oxydation. L' ω -oxydation des acides gras a lieu dans les microsomes et transforme les

²Par convention on compte toujours les carbones à partir de l'extrémité COOH. Lorsqu'on décide de les compter à partir de l'extrémité CH₃, on utilise deux notation équivalentes : n-3 ou ω 3 pour représenter la liaison entre le 3^{ème} et le 4^{ème} carbone à partir du CH₃. C'est cette notation qui est utilisée pour repérer les familles d'acides gras, ainsi le C18 :2 ω 6 possède une double liaison entre les carbones 6 et 7 à partir du CH₃, c'est-à-dire entre les carbones 12 et 13 à partir du COOH.

³L'oléate est l'acide gras majoritaire du muscle de lapin. Les acides gras plus longs, préférentiellement oxydés par le peroxysome sont fortement minoritaires dans le muscle de lapin. De ce fait, on sous-estime légèrement l'oxydation peroxysomiale totale.

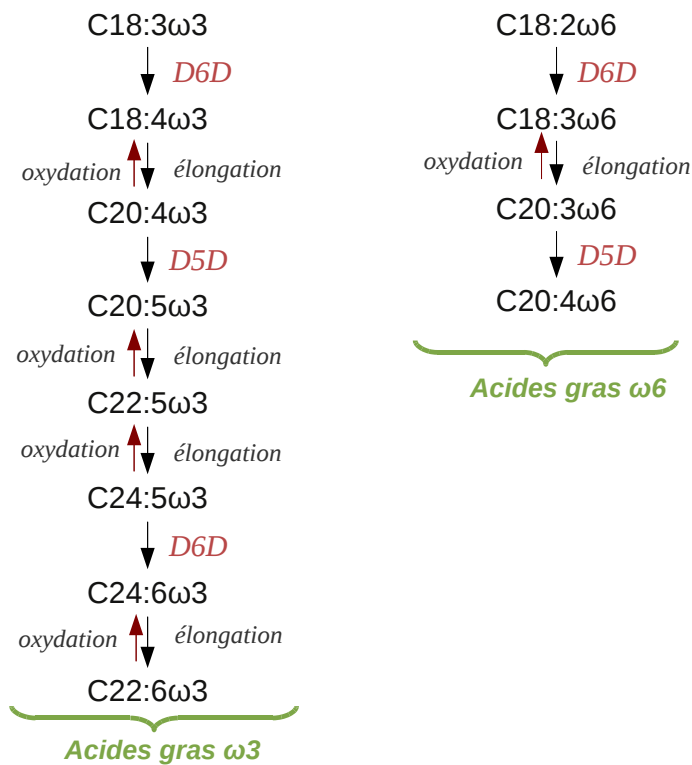
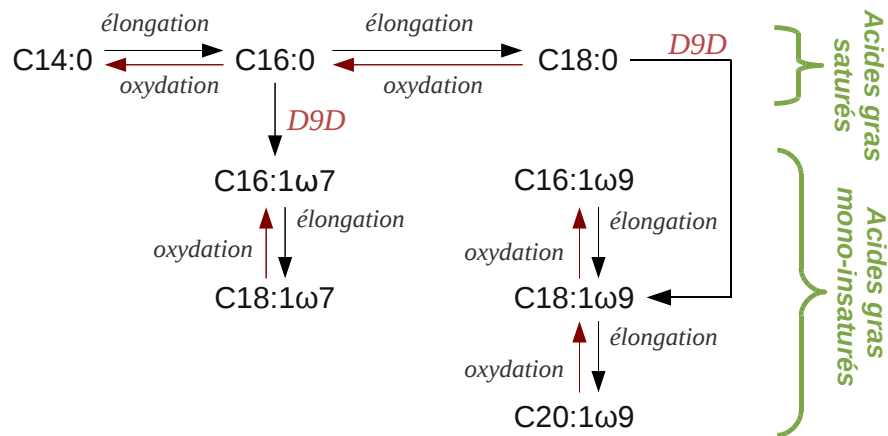


FIG. 2.4.2 – **Transformations entre acides gras.** Les acides gras sont généralement séparés en 3 classes : a) les acides gras saturés qui ne comportent aucune insaturation et monoinsaturés qui comportent une double liaison en position ω 7 ou ω 9, b) les acides gras ω 3 c) les acides gras ω 6. Les acides gras peuvent être transformés par désaturation grâce aux enzymes delta-5-désaturase (D5D) delta-6-désaturase (D6D) et delta-9-désaturase (D9D), par élongation et par oxydation partielle. Les réactions d'interconversion ne peuvent avoir lieu qu'entre acides gras d'une même classe. Ces réactions sont représentées sur la figure par une flèche. Les réactions d'importation, de néo-synthèse d'acides gras et d'oxydation complète des acides gras ne sont pas représentées.

acides gras en acides dicarboxyliques qui sont éliminés par voie urinaire [199,58,184]. Cette voie est largement minoritaire par rapport à la β -oxydation, mais peut devenir importante en cas de déficience de cette dernière [199]. L' α -oxydation est un mécanisme spécifique permettant d'oxyder les acides gras ramifiés, il ne sera pas décrit dans cette thèse (pour une revue : [332]).

FIG. 2.5.1 – Catabolisme des acides gras.

37

Hydrolyse des triglycérides En cas de besoin énergétique, les triglycérides sont hydrolysés en glycérol et en acides gras. Dans le coeur [299] et dans le tissu adipeux [191], cette réaction est catalysée notamment par LIPE. Le coeur consomme lui-même ses propres acides gras tandis que le tissu adipeux les libère dans le sang où ils circulent liés à l'albumine.

Dans le sang, les lipoprotéines lipases (LPL), des isoenzymes situés contre la paroi endothéliale, hydrolysent les triglycérides des chylomicrons et des *very low density lipoproteins* sanguins [118]. Ils libèrent des acides gras libres qui sont très rapidement captés par les cellules proches.

Dans le foie, les triglycérides intracellulaires peuvent être hydrolysés par la lipase hépatique (LIPC) en acides gras libres.

Dégradation des acides gras. Les acides pris en charge par l'*acyl carrier binding protein*(ACBP) puis importés dans la mitochondrie (grâce à l'action successive de CPT1 et CPT2 pour les acides gras longs) sont oxydés lors de la β -oxydation. Pour un acide gras saturé à n carbones, son bilan est :

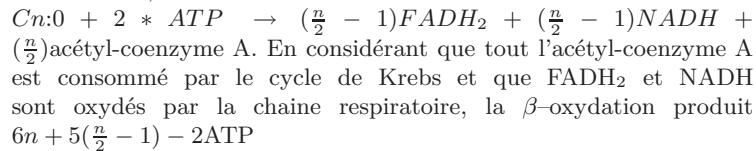
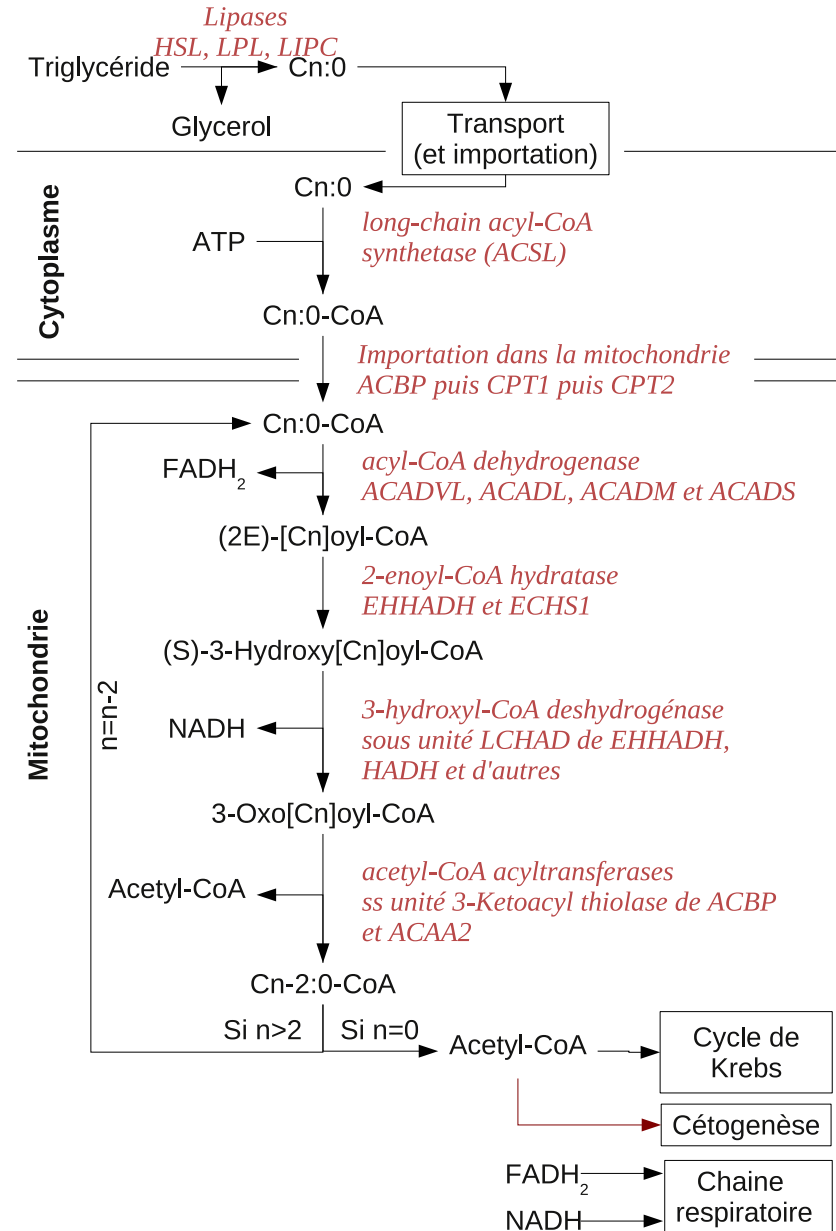


Figure simplifiée réalisée à partir de *Metabolism at Glance* de J.G. Salway [271], KEGG et la revue *Control of mitochondrial β -oxidation flux* de S. Eaton [88]. Les noms des enzymes de la revue ont été mis à jour avec HGNC.



2.5.3 Consommation d'acétyl-coenzyme A par le cycle de Krebs et la chaîne respiratoire

Dans les mitochondries, l'acétyl-coenzyme A est utilisé dans le cycle de Krebs (aussi appelé cycle des acides tricarboxyliques ou TCA) couplé à la chaîne respiratoire en présence d'oxygène pour fournir de l'ATP. Le cycle de Krebs est une succession de réactions produisant 3 NADH, 1 $FADH_2$, 1 guanosine triphosphate (GTP), 1 coenzyme A et 2 CO_2 à partir de 3 NAD, 1 FAD, 1 guanosine diphosphate (GDP), un phosphate et 1 acétyl-coenzyme A. En présence d'oxygène le NADH et l'ATP sont consommés par la chaîne respiratoire, produisant 2.5 ATP par NADH et 1.5 ATP par $FADH_2$ [271]. Pour un acétyl-coenzyme A consommé, le processus complet produit environ 10 ATP. Lorsqu'une cellule fonctionne en aérobie, le cycle de Krebs est sa principale source d'énergie.

2.5.4 Cétogenèse

Dans le foie, et seulement dans le foie, l'acétyl-coenzyme A produit lors de l'oxydation des acides gras peut être transformé en corps cétoniques par condensation. Contrairement à l'acétyl-coenzyme A qui est exclusivement intra cellulaire, les corps cétoniques sont exportés dans le sang. Ces derniers jouent alors le rôle de métabolite de substitution de l'acétyl-coenzyme A, et donc du glucose dans de nombreuses cellules. Par exemple le système nerveux de l'homme est capable, lorsqu'il est entraîné, de consommer jusqu'à 90% de corps cétoniques (et 10% de glucose). Lorsque la quantité de lipides oxydés par le foie est très importante, la quantité de cétones dans le sang peut augmenter très fortement et provoquer une crise d'acétone ; les cétones sont alors éliminées dans l'air par les poumons (ce qui provoque une haleine rance) et dans les urines.

Chapitre 3

Plasticité du métabolisme des lipides : ses principales régulations

Les variations de la balance énergétique ont des conséquences sur l'intensité des voies métaboliques à l'échelle de l'organisme comme de la cellule. Dans les conditions habituelles, les individus sont soumis à des variations quotidiennes de l'apport alimentaire. Le métabolisme animal alterne donc entre les périodes de stockage d'énergie suite à l'absorption des nutriments et les périodes de dégradation de réserves lors du jeûne ou de fortes demandes en énergie, ce qui implique de diriger les métabolites tantôt vers l'anabolisme, tantôt vers le catabolisme (pour des revues sur les effets des variations d'alimentation voir : [35, 204] chez le poulet, [268, 81] chez la truite, [303] chez la souris).

Les variations métaboliques sont coordonnées à l'échelle de l'organisme Cet aiguillage des métabolites énergétiques en fonction de la balance énergétique est fortement régulé à l'échelle de l'organisme par les niveaux circulants des différentes hormones. Les hormones sont en effet des senseurs de l'état nutritionnel global, libérées dans la circulation générale afin de délivrer un même signal pour l'ensemble de l'organisme. Dans le cadre du métabolisme énergétique, ce signal est essentiellement porté par deux hormones antagonistes régulées par la glycémie : l'insuline et le glucagon. A ces deux hormones s'ajoute l'adrénaline¹, une hormone synthétisée en cas de stress ou d'effort dont les effets sont dans la grande majorité des cas identiques à ceux du glucagon.

Des récepteurs hormonaux présents dans les différents organes permettent la transduction du message hormonal et la réponse spécifique des organes. Les hormones sont alors capables de modifier simultanément l'activité des enzymes (ce qui entraîne une variation rapide du métabolisme) et la transcription de nombreux gènes (ce qui entraîne une variation du métabolisme à plus long terme). Il en résulte une réponse coordonnée et coopérative des différents organes soit vers l'anabolisme, soit vers le catabolisme.

Les variations métaboliques sont coordonnées à l'échelle de la cellule A l'échelle de la cellule, le signal (hormonal ou nutritionnel) entraîne une cascade de réponses orientant le métabolisme cellulaire de manière exclusive soit vers l'anabolisme, soit vers le catabolisme.

Ces mécanismes, assurant l'exclusion mutuelle des différentes voies antagonistes, sont généralement très complexes. Brièvement, et uniquement pour ce qui concerne le métabolisme des lipides, nous pouvons citer :

- L'exclusion mutuelle de la voie de l'insuline et de celle du glucagon par leur effet antagoniste sur la (non) production d'AMPc, amplifiée par une cascade de kinases (ainsi que de nombreux autres mécanismes non décrits ici).
- L'effet inhibiteur du malonyl-coenzyme A sur l'oxydation des acides gras empêchant cette dernière d'avoir lieu en même temps que la néo-synthèse des acides gras.

¹L'adrénaline est aussi connue sous le nom d'épinephrine

- L’inhibition du cycle de Krebs par l’ATP régulant à la fois la consommation intracellulaire d’acétyl-coenzyme A à des fins de production d’énergie, et rendant ce dernier disponible pour la néo-synthèse des acides gras.
- L’inhibition de la glycolyse par l’ATP et l’acétyl-coenzyme A, limitant la dégradation du glucose à des fins énergétiques. Cette inhibition permet en particulier de limiter la glycolyse lorsque le catabolisme des acides gras suffit à fournir la cellule en ATP.
- L’activation d’ACC par le citrate cytoplasmique qui favorise la synthèse de C16 :0 lorsque le citrate est exporté de la mitochondrie, limitant ainsi son accumulation cytoplasmique et favorisant la néo-synthèse d’acides gras.

Globalement, la transcription des gènes des enzymes et transporteurs intervenant dans le métabolisme énergétique et lipidique obéit au même schéma de régulation orientée en deux voies : l’une active à jeûn et l’autre active à l’état nourri. Cette régulation fait intervenir une multitude de facteurs de transcription² qui modulent l’expression des gènes et constitue un domaine de recherche largement ouvert. Parmi eux, trois jouent un rôle majeur dans la régulation de l’expression des enzymes du métabolisme des lipides : SREBP1 et ChREBP actifs à l’état nourri et PPAR α actif lors du jeûne. Seuls ces trois facteurs seront détaillés dans cette thèse.

Etudier le métabolisme des lipides à l’échelle de la mécanique interne de la cellule est un vaste travail nécessitant l’analyse de la cinétique de chaque enzyme impliqué dans le métabolisme des lipides. Dans le cadre de cette thèse, nous ne développerons ce niveau de connaissance que lorsqu’il est nécessaire pour expliquer un jeu de données particulier.

3.1 Principales régulations hormonales

Globalement, l’insuline active l’anabolisme et inhibe le catabolisme des glucides et des lipides. Le glucagon a l’effet inverse. Une présentation schématique des effets de ces deux hormones est fournie par le tableau 3.1.1.

²Une recherche dans Pubmed de “*transcription factor*” AND (“*lipid*” OR “*fatty acid*” OR “*fatty acids*”)” effectuée le 28 octobre 2009 donne 2280 résultats. Cette recherche est une sous estimation de la littérature relative à ce sujet car de nombreux facteurs de transcription peuvent avoir une influence indirecte et car il faudrait aussi faire une recherche avec un mot clef associé à chaque molécule relative au métabolisme des lipides.

TAB. 3.1.1 – **Régulation hormonale du métabolisme énergétique (1/2)** Les deux principaux régulateur du métabolisme énergétique en fonction de l'alimentation sont l'insuline et le glucagon. Les éléments qu'ils activent sont notés par un +, qu'ils inhibent par un - et sur lesquels ils n'ont pas d'effet par un 0. Lorsque ces régulations font intervenir un médiateur ou sont la conséquence d'un autre mécanisme, ce dernier est mis entre parenthèses; lorsque rien n'est précisé, les régulations peuvent être directes ou indirectes.

| Fonction, élément de la fonction | Régulation par l'insuline | Régulation par le glucagon |
|----------------------------------|--|---|
| Facteurs de transcription | | |
| SREBP1 | + SREBP1-c [39, 105] muscle et tissus adipeux [347] | |
| PPAR α | | + (p38 MAPK et AMPK) [208] |
| ChREBP | | - (AMPc) [326] |
| Transport du glucose | | |
| SLC2A4 ^a | + Foie, tissu adipeux [270] | |
| G6PC ^b | - [270], transcriptionnelle [276] | + Activité G6PC [164] |
| Synthèse du glycogène | | |
| Glycogen synthase | + [45, 31, 266] + (Déphosphorylation) Activité glycogène synthase [270] | - [266, 164] - (Phosphorylation) Activité glycogène synthase [164] |
| Dégradation du glycogène | | |
| Glycogen phosphorylase kinase | - Muscle, tissu adipeux [68], foie [250] - [309] | + [164] + [164] |
| Glycogen phosphorylase | - (Glycogen phosphorylase kinase) [309] | + (Glycogen phosphorylase kinase) [164] |
| Glycolyse | | |
| Glucokinase | + (SREBP1 [104]) synthèse glucokinase [270, 160] | - [164] - [160] |
| L-pyruvate kinase | + [270] | - [164] |
| Phosphofructokinase | + [270] | |
| Pyruvate deshydrogenase | + [168] | + hépatocyte de rat [96] |
| Néoglucogénèse | | |
| PEPCK | - Expression de PEPCK [270] | + [112, 164] + Transcription (foie) [164] |
| PC | - Transcription de PC [166] | + [166] |
| FBP1 | - Expression de FBP1 [270] | + [164] |

^a SLC2A4 est aussi connu sous le nom de GLUT-4. C'est un transporteur permettant l'importation du glucose dans la cellule

^b La Glucose-6-phosphatase (G6PC) est l'enzyme hépatique responsable de la transformation du glucose-6-phosphate en glucose, ce qui permet son exportation du foie vers le sang.

TAB. 3.1.2 – Régulation hormonale du métabolisme énergétique (2/2)

| Fonction, élément de la fonction | Régulation par l'insuline | Régulation par le glucagon |
|-----------------------------------|--|---|
| Synthèse d'acides gras | + (LXR [322]) ^a . | - Foie de rat [62] |
| ACLY | + [270] | |
| ACC | + (AMPK [339,270], éventuellement SREBP1 [209,6,179]) | - (Phosphorylation d'ACC par une kinase AMPc dépendante) [145] - Foie de rat [62] |
| FASN | + Transcription de FASN [283] (SREBP1 [104]) [270] | - (AMPc) Transcription de FASN [283] |
| Elongation des acides gras | + (SREBP [227]) [334] | |
| Triglycérides | + Synthèse triglycérides [171] + Stockage dans le tissu adipeux [171] | - Synthèse triglycérides [208] - Exportation des triglycérides hors du foie [208] |
| Lipolyse | - Empêche l'induction par norepinephrine et ACTH chez le rat [229] | + [275,112]. 0 aux concentrations physiologiques chez l'homme [190] |
| LIPE | - [7,270] | + Activité de LIPE [298] dans l'adipocyte de rat 0 sur la transcription de LIPE dans l'adipocyte de rat [298]. |
| Oxydation des acides gras | | + (Acitvation céto-genèse hépatique) [249] |
| CPT1 | - Transcription CPT1 [89,291] | |
| Cétogenèse | | + [26,62,25,112] |
| Synthèse du cholestérol | + (LXR [322]) | 0 foie de rat [62] |
| Synthèse protéique | + Mécanisme [182,248], effet dans le muscle de rat [161,111] | - [11] |
| Importation acides aminés | + Hépatocyte de rat [41], mammifères [219] | |
| Protéolyse | | + Foie de rat [341,280,11] |

^aL'activation de la synthèse d'acides gras entraîne une demande en NADPH, ce qui entraîne une augmentation du cycle des pentoses [94]

3.1.1 L'insuline stimule la synthèse de réserves énergétiques et inhibe leur utilisation

L'insuline est une hormone protéique produite par les cellules β des îlots de Langerhans lorsque la glycémie est élevée, puis libérée dans le sang [141]. Cette libération fait suite aux variations du glucose sanguin. En effet, le glucose des cellules β s'équilibre avec le glucose sanguin grâce à un mécanisme de diffusion facilitée ; le catabolisme du glucose intracellulaire entraîne une augmentation du rapport ATP/ADP. L'ATP provoque alors la fermeture de canaux à potassium (K^+) ATP dépendants. Ceci cause alors une dépolarisation de la membrane cellulaire qui ouvre des canaux à calcium (Ca^{2+}). Le Ca^{2+} rentre alors dans le cytoplasme et active l'exocytose d'insuline [141] dans le sang.

L'insuline favorise le stockage de glucose et inhibe sa néo-synthèse

L'insuline stimule l'importation de glucose dans la cellule L'insuline favorise tout d'abord l'importation de glucose dans la cellule, qu'elle soit adipocytaire, musculaire ou hépatocytaire, grâce à l'activation de la synthèse et de la translocation vers la membrane cellulaire de SLC2A4³ [270]. L'insuline pourrait en revanche inhiber transitoirement l'expression des ARNm codant pour SLC2A2⁴, transporteur décrit à l'origine comme insulino-indépendant dans le foie et l'intestin notamment (pour une revue voir [203,255]). Cette inhibition est contrebalancée par l'activation de SLC2A2 par le glucose [255].

L'insuline inhibe la néoglucogenèse L'insuline inhibe transcriptionnellement la synthèse de glucose-6-phosphatase (G6PC). Cet enzyme est responsable de la transformation du glucose-6-phosphate en glucose, ce qui permet son exportation du foie vers le sang. Lorsque G6PC est inhibé, le glucose-6-phosphate s'accumule dans la cellule et inhibe la néoglucogenèse par excès de produit [270,276].

De plus, l'insuline inhibe la transcription des trois enzymes clef de la néoglucogenèse : PC [166], PEPCK [270] et FBP1 [270]⁵. Cette inhibition est cohérente avec le fonctionnement de la glycolyse à plein régime. Il en résulte une production massive d'acétyl-coenzyme A.

L'insuline stimule la synthèse de glycogène En cas d'insulinémie élevée et de réserves en glycogène faibles, le glucose-6-phosphate cellulaire est aiguillé majoritairement vers la synthèse de glycogène. En effet grâce à une cascade de seconds messagers (dont l'AMPc et la glycogène phosphorylase kinase), l'insuline active la glycogène synthase et donc la synthèse de glycogène [231,45,31,266,270]. L'insuline limite par la même occasion la dégradation de glycogène grâce à l'inhibition de la glycogène phosphorylase [231,68] et grâce à un simple "effet produit" causé par l'accumulation de glucose-6-phosphate dans la cellule [270,276].

L'insuline favorise la néo-synthèse de triglycérides et inhibe leur catabolisme

L'insuline active la glycolyse L'insuline favorise la glycolyse grâce à l'activation de la transcription de la glucokinase (GCK) [104,270,160], de la pyruvate deshydrogenase (PDH) [168] et de la pyruvate kinase (PK) [270]. Dans le foie, l'insuline active aussi la PK hépatique (PKLR) par déphosphorylation.

A l'inverse, l'insuline déphosphoryle la fructose-2,6-biphosphatase ; sous forme déphosphorylée, fructose-2,6-biphosphatase catalyse la production de fructose 2,6 bi-phosphate [153], un activateur de la phosphofructokinase (PFK) [253,328], et donc de la glycolyse. L'activation de la glycolyse, couplée à l'inhibition de la néoglucogenèse se traduit par une augmentation de la production d'acétyl-coenzyme A.

³ SLC2A4 est aussi connu sous le nom de GLUT-4. C'est un transporteur insulino-dépendant permettant l'importation du glucose dans la cellule

⁴SLC2A2 est un transporteur bidirectionnel du glucose, du fructose et du galactose aussi connu sous le nom de GLUT2

⁵ Ces 3 enzymes catalysent les réactions irréversibles de la néoglucogenèse. Toutes les autres réactions de la néoglucogenèse sont des réactions de la glycolyse qui fonctionnent dans le sens inverse.

L'insuline active la néo-synthèse d'acides gras et leur stockage sous forme de triglycérides. En présence d'insuline et de glucose, l'acétyl-coenzyme A est aiguillé principalement vers la néo-synthèse d'acides gras car l'insuline active la synthèse des enzymes clef de la lipogenèse : ACLY [270], ACC (par un mécanisme à AMPK [339, 270] et éventuellement une augmentation de sa transcription [209, 6, 179]) et FASN [283, 270, 104] Il en résulte une augmentation de la néo-synthèse de C16 :0.

L'activation de la synthèse d'acides gras entraîne une demande en NADPH, ce qui provoque une augmentation d'activité du cycle des pentoses [94]. L'utilisation d'acétyl-coenzyme A par la synthèse des acides gras empêche son accumulation cytoplasmique, il n'est ainsi pas disponible pour la cétogenèse (une voie active lorsque l'acétyl-coenzyme A est présent en grande quantité).

L'insuline active aussi l'élongation des acides gras [334, 227], ainsi que leur stockage sous forme de triglycérides dans le tissu adipeux [171].

L'insuline inhibe la dégradation des triglycérides et l'oxydation des acides gras En parallèle, l'insuline inhibe la dégradation des triglycérides en bloquant l'induction de la lipase hormonosensible (LIPE) [7, 270]. Cet enzyme, lorsqu'il est activé, participe à l'hydrolyse des triglycérides du tissu adipeux en glycérol et en acides gras. Ces derniers sont alors relargués dans le sang puis captés par les cellules. Une fois captés les acides gras induisent notamment leur catabolisme via PPAR α (c.f. 3.2.3). Ainsi le contrôle de LIPE est un élément particulièrement crucial [191] de la régulation du catabolisme des lipides.

L'insuline a aussi un effet négatif sur la transcription de CPT1 [89, 291], l'enzyme responsable de l'importation des acides gras dans la mitochondrie, lieu où ils sont oxydés.

Quelques autres rôles pour l'insuline

L'insuline stimule la synthèse de cholestérol L'insuline, grâce à son rôle sur SREBP1 et sur LXR (c.f. 3.2.2) entraîne une augmentation de la synthèse du cholestérol.

L'insuline stimule la synthèse protéique Brièvement, l'insuline entraîne aussi une augmentation de l'importation cellulaire des acides aminés [41, 219] et de la synthèse de protéines [182, 248]. Cette dernière étant particulièrement couteuse en énergie, sa stimulation à l'état nourri [161, 111] permet par exemple d'augmenter la masse musculaire lors des périodes d'abondance alimentaire.

3.1.2 Le glucagon stimule la dégradation des réserves énergétiques et inhibe l'utilisation de glucose

Le glucagon a une action généralement antagoniste à l'insuline : il inhibe la synthèse de réserves énergétiques et les mécanismes d'utilisation du glucose et stimule la dégradation de réserves et la production de glucose par glycogénolyse et néoglucogenèse.

Le glucagon stimule la néo-synthèse de glucose

Dans le foie, le glucagon favorise la dégradation du glycogène hépatique [112] et la néoglucogenèse (par activation des enzymes clefs : PEPCK,FBP1 [164] et PC [166]). Le glucose-6-phosphate ainsi produit est transformé en glucose grâce à l'activation de G6PC [164] puis exporté du foie vers le sang.

Le glucagon favorise l'utilisation d'acides gras, au dépend du glucose

Le glucagon active LIPE [275, 112]⁶ ce qui entraîne une dégradation massive des triglycérides du tissu adipeux et une libération d'acides gras libres (et de glycérol) dans le sang. Les acides gras ainsi libérés sont captés par les organes capables de les oxyder.

⁶Sauf chez l'homme aux concentrations physiologiques [190]

Le glucagon active l'oxydation des acides gras [136] ce qui entraîne une production massive d'acétyl-coenzyme A mitochondrial et d'ATP. Dans le foie, le glucagon active aussi la cétogenèse, favorisant ainsi la production de corps cétoniques à partir d'acétyl-coenzyme A mitochondrial. [249, 26, 62, 25, 112].

La β -oxydation et l'utilisation d'acétyl-coenzyme A et de corps cétoniques par le cycle de Krebs et la chaîne respiratoire permet la fourniture d'énergie. L'accumulation transitoire d'acétyl-coenzyme A mitochondrial inhibe la glycolyse par effet produit, et par inhibition de la PDH [271]. Ce mécanisme se traduit par une diminution de l'utilisation de glucose à des fins de production d'énergie.

Le glucagon inhibe la synthèse des réserves énergétiques à partir de glucose

Le glucagon inhibe la synthèse de glycogène [164], la glycolyse (inhibition de la glucokinase [160], de la L-pyruvate kinase [164], de la synthèse de fructose 2,6 bi-phosphate par phosphorylation de fructose-2,6-biphosphatase [14]), la synthèse d'acides gras (inhibition d'ACC [145, 62] et de FASN [283]) et la production et l'exportation de triglycérides par le foie [208].

3.2 Principales régulations de l'expression des gènes relatives au métabolisme des lipides

Les variations de la balance énergétique sont capables de réguler l'expression de nombreux gènes du métabolisme des lipides (voir [304] pour une revue dans le foie de rongeur, [269] pour une revue chez la truite et [82] pour une revue générale sur la régulation transcriptionnelle du métabolisme), en particulier par l'intermédiaire du glucose et des acides gras [87] notamment poly-insaturés [285, 272].

Dans cette section, nous nous proposons de décrire rapidement les principaux régulateurs transcriptionnels ayant un effet majeur sur le métabolisme des lipides [129] : ChREBP le principal médiateur du glucose [256, 78], SREBP1 le principal médiateur de l'insuline [78, 84], et PPAR α un facteur de transcription essentiellement activé par le glucagon [208, 193] et les acides gras poly-insaturés [284, 285, 175] qui est notamment responsable de l'activation du catabolisme des acides gras.

3.2.1 ChREBP stimule la lipogenèse en réponse au glucose

A l'exception de la glucokinase hépatique qui est exclusivement induite par l'insuline, la plupart des gènes de la glycolyse et de la lipogenèse sont aussi régulés par le glucose. Le facteur transcriptionnel décrit comme contrôlant l'expression de ces gènes en réponse au glucose a été nommé ChREBP. ChREBP est un facteur de la famille basic/helix-loop-helix/leucine zipper (bHLH/LZ) ; sa localisation cellulaire et son activité sont dépendantes d'un état de phosphorylation/déphosphorylation, comme expliqué ci-après (c.f. figure 3.2.1). La forme active de ChREBP est capable de se fixer à un élément CHRE (*Carbohydrate Response Element*) ce qui induit la transcription des gènes cibles. ChREBP est exprimé majoritairement dans le foie, le rein et le tissu adipeux (brun et blanc). Il est régulé positivement par le glucose et négativement par l'AMPc (un second messager régulé de manière antagoniste par le glucagon et l'insuline).

ChREBP est activé par le glucose et inhibé par le glucagon et les acides gras

Le glucose active ChREBP A l'état nourri, l'entrée du glucose dans le cycle des pentoses entraîne la production de xylulose 5-phosphate, qui active la déphosphorylation de ChREBP puis son importation dans le noyau [170, 158, 256] (c.f. figure 3.2.1). Il en résulte une activation de la transcription des gènes cibles de ChREBP en réponse au glucose.

Ces mécanismes d'activation ne sont pas les seuls suffisants pour que ChREBP ait un effet ; il faut en plus lever son inhibition [74].

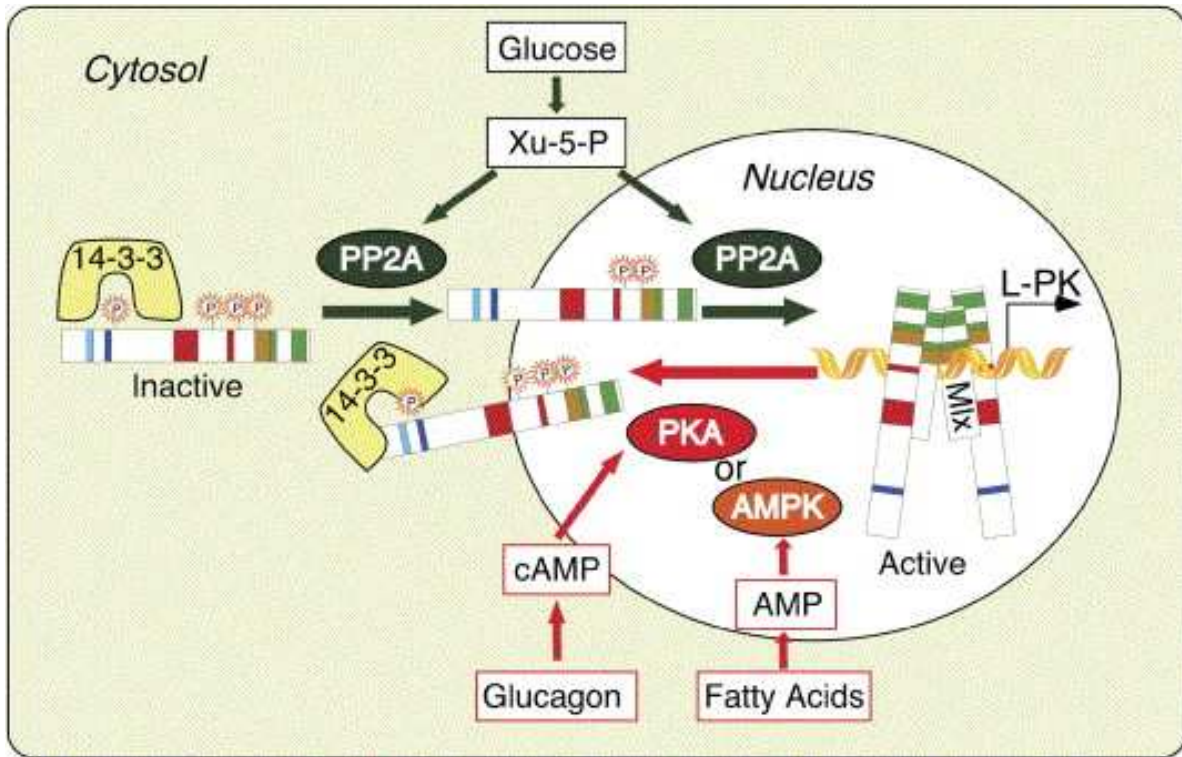


FIG. 3.2.1 – **Régulation de l'activité de ChREBP par phosphorylation.** Sous forme phosphorylée, ChREBP (représenté par un rectangle contenant des barres colorées) est activement exclu du noyau puis il est séquestré dans le cytosol par 14-3-3, formant un complexe inactif. La présence de glucose intracellulaire entraîne une augmentation du xylulose 5-phosphate (Xu-5-P) dans le cycle des pentoses entraînant l'action d'une protéine phosphatase 2A (PP2A) ce qui entraîne alors la translocation (flèches vertes) de ChREBP dans le noyau. Le ChREBP nucléaire se dimérise alors avec Mlx puis se lie aux éléments de réponse aux carbohydrates sur les promoteurs des gènes cibles tels que la pyruvate kinase hépatique (PK) ce qui augmente leur transcription. A l'inverse, le glucagon et les acides gras augmentent l'activité de kinases dépendantes à AMPc (PKA) et à l'AMP (AMPK) qui phosphorylent ChREBP, réduisant alors sa capacité de liaison avec l'ADN et provoquant l'exclusion de ChREBP hors du noyau (flèches rouges).

Figure et légende d'après Uyeda, K. *et al.*, 2006 [325].

Le glucagon et les acides gras inhibent ChREBP À l'inverse, par exemple lors de la mise à jeun, le glucagon entraîne la synthèse d'AMPc dans la cellule, ce dernier active alors une kinase AMPc dépendante qui va phosphoryler ChREBP. Une fois phosphorylé, ChREBP est alors exporté hors du noyau vers le cytosol où il est séquestré sous une forme inactive par 14-3-3 [325] (c.f. figure 3.2.1).

Les acides gras (et notamment les C18 : 2ω 6, C20 : 5ω 3, C22 : 6ω 3 [77]) sont capables d'induire la présence d'AMP cellulaire (c.f. figure 3.2.1). Les acides gras poly-insaturés C18 : 2ω 6, C20 : 5ω 3 et C22 : 6ω 3 ont également un effet négatif sur le cycle des pentoses car ils diminuent la concentration en xylulose-5-phosphate. Ceci induit une inhibition de la translocation de ChREBP, et donc une diminution de son effet transcriptionnel [77]. En parallèle, ces acides gras augmentent la dégradation de l'ARNm de ChREBP diminuant ainsi son activité [77].

ChREBP active la néo-synthèse d'acides gras, D9D et la glycolyse [156]

La forme active de ChREBP est capable d'induire la transcription des gènes de la glycolyse (PK [344,74]) et de la néo-synthèse des acides gras (ACC [158], FASN [158,344] et enzyme malique [156]⁷).

En accord avec le rôle de ChREBP sur l'expression de ses gènes cibles, la glycolyse et la néo-synthèse d'acides gras diminuent fortement (de 65% pour la synthèse d'acides gras dans le foie de souris en l'absence de ChREBP [156]).

L'absence de ChREBP inhibe aussi SCD1, une isoforme de D9D dans le foie de souris, responsable de la synthèse des acides gras des familles ω 7 et ω 9 (c.f. figure 2.4.2) [156]. L'absence de ChREBP diminue aussi la transcription de *long chain fatty-acyl elongase*, un enzyme responsable de l'élongation de C16 :0 en C18 :0 [227].

3.2.2 SREBP1 stimule la synthèse de triglycérides à l'état nourri en réponse à l'insuline

SREBP1 appartient à une famille de facteurs transcriptionnels décrits à l'origine comme régulant la transcription de gènes en relation avec la disponibilité cellulaire en cholestérol (d'où leur nom). Comme ChREBP, SREBP1 est un facteur de la famille basic/helix-loop-helix/leucine zipper (bHLH/LZ). Il possède quatre isoformes distinctes SREBP1-a, SREBP1-b, SREBP1-c et SREBP1-isoforme4 référencées dans Uniprot.

SREBP1 est connu pour activer le métabolisme des lipides en synergie avec ChREBP, en particulier lorsque l'individu est nourri (c.f. figure 3.2.2).

Régulation de SREBP1

Activation de SREBP1 par l'insuline Il a été montré que l'insuline augmente la synthèse (ARNm) [39,105], l'abondance nucléaire et la capacité de liaison de SREBP1-c avec le promoteur des gènes de la lipogenèse [84,78]. L'insuline aurait aussi un rôle direct sur le clivage protéolytique de la forme longue et immature de SREBP1-c en une forme active capable de migrer dans le noyau [140]. Le mécanisme d'action de l'insuline sur SREBP1-c fait intervenir les LXR [49], une famille de récepteurs qui activent SREBP1-c [155].

La voie de signalisation de SREBP1 et celle de PPAR α sont régulées de manière exclusive, en particulier grâce à la concurrence entre PPAR α et les *liver X receptors* (LXR) (des activateurs de SREBP) pour se complexer avec les RXR. Cette balance est détaillée figure 3.2.4.

Inhibition de SREBP1 par le glucagon Le glucagon induit des effets opposés à l'insuline sur l'expression de SREBP1-c dans le foie, via son second messager AMPc (voir [102] pour une revue). Lors du jeûne, SREBP1 décroît rapidement comme observé chez la souris [148].

⁷L'enzyme malique permet la production de NADPH à partir de malate, un intermédiaire du cycle de Krebs. Le NADPH ainsi produit est en particulier disponible pour la synthèse des acides gras.

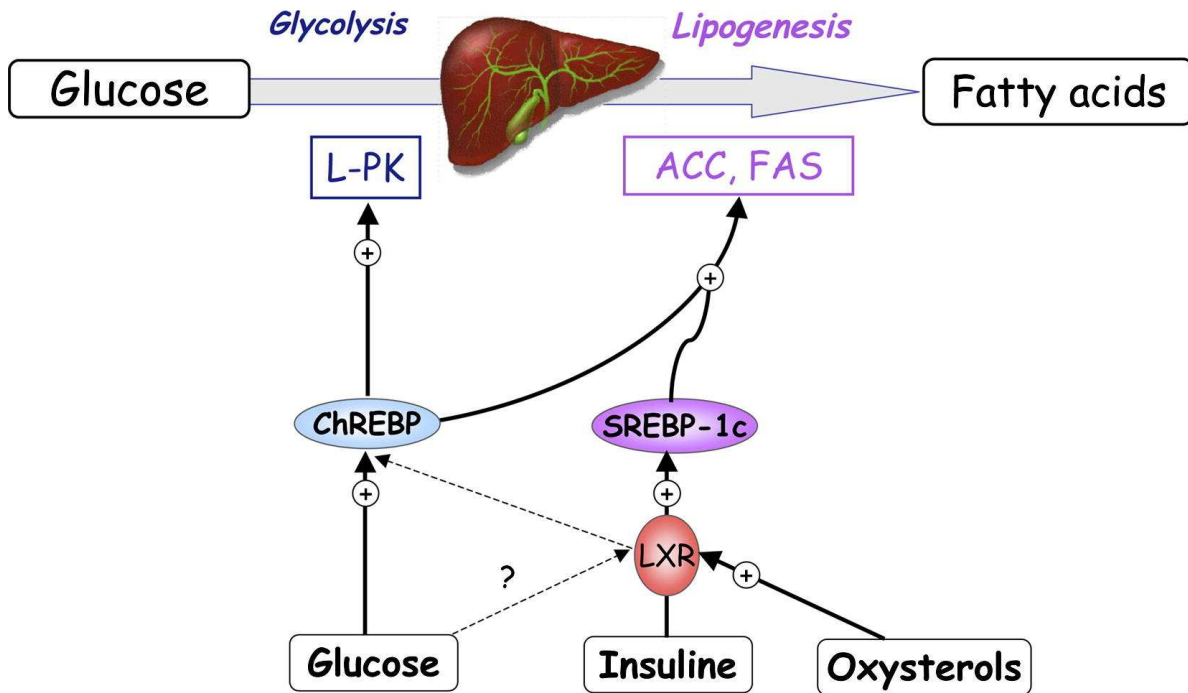


FIG. 3.2.2 – Régulation transcriptionnelle de la glycolyse et de la synthèse de lipides. Les voies de signalisation du glucose et de l'insuline sont sollicitées en réponse aux sucres alimentaires afin d'induire en synergie l'activation de la glycolyse et l'expression des gènes codants pour les enzymes de la lipogenèse. Le facteur de transcription SREBP1-c est considéré comme un médiateur majeur de l'action de l'insuline sur de nombreux gènes de la lipogenèse tels que FASN et ACC [105], via une action sur les éléments SRE des promoteurs de ces gènes. L'élément responsable de la régulation transcriptionnelle des gènes de la glycolyse comme PK par le glucose a été identifié comme étant ChREBP. Le glucose active ChREBP en stimulant son expression et en régulant son transport du cytosol vers le noyau. Une fois dans le noyau, ChREBP se lie aux *carbohydrate responsive element* (ChRE) [325] présents dans la région promotrice de ses gènes cibles. L'induction de FASN et ACC est issue de l'action combinée de ChREBP en réponse au glucose et SREBP1-c en réponse à l'insuline. ChREBP est aussi une cible directe des *liver X receptors* (LXR) [46]. Les LXR (LXR α et LXR β , pour une revue sur leurs rôles [279]) sont des récepteurs nucléaires capables d'activer directement la transcription de ACC et FASN. Ce sont aussi des médiateurs de l'activation de ces gènes par SREBP1-c [167] en réponse à l'insuline. Récemment, l'observation que le glucose se lie et active les LXR nucléaires place les LXR comme des composés centraux de la voie de signalisation du glucose [224], mais leur implication réelle dans un contexte physiologique est encore à évaluer.

Figure extraite et légende d'après de Denechaud, P.D *et al.*, 2008 [76].

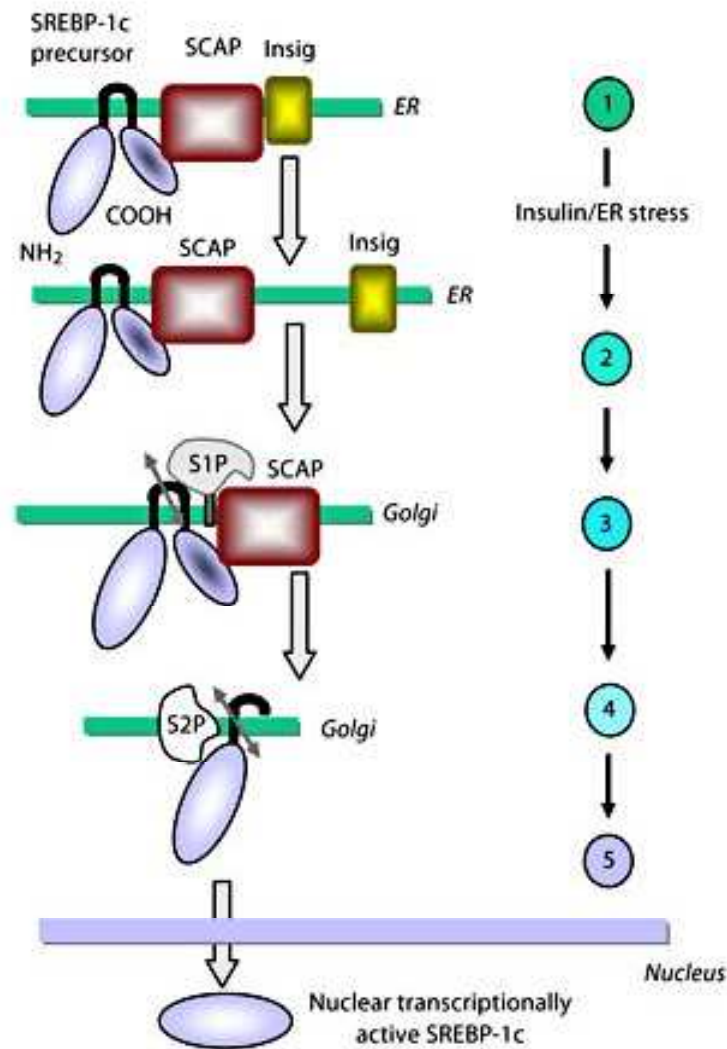


FIG. 3.2.3 – **Activation protéolytique de SREBP1-c.** L'interaction du complexe formé par SREBP1-c et la protéine SCAP avec l'*insulin induced gene* (Insig) entraîne sa séquestration dans le réticulum endoplasmique (ER). (2) En présence des signaux appropriés (insuline, stress du réticulum endoplasmique), le complexe SREBP1/SCAP migre vers l'appareil de Golgi dans lequel SREBP1-c subit deux clivages protéolytiques successifs. Ces clivages sont réalisés par *Site-1 protease* (S1P) et *Site-2 protease* (S2P) (flèches diagonales) (3) et (4), et transforment SREBP1-c en sa forme mature par clivage de la partie NH₂-terminale. La forme mature de SREBP1-c est alors transportée dans le noyau (5) où elle se lie à ses éléments de réponses spécifiques présents dans le promoteur des gènes cibles de SREBP1-c.
Figure et légende extraites de P.Ferre et F. Foufelle, 2007 [102].

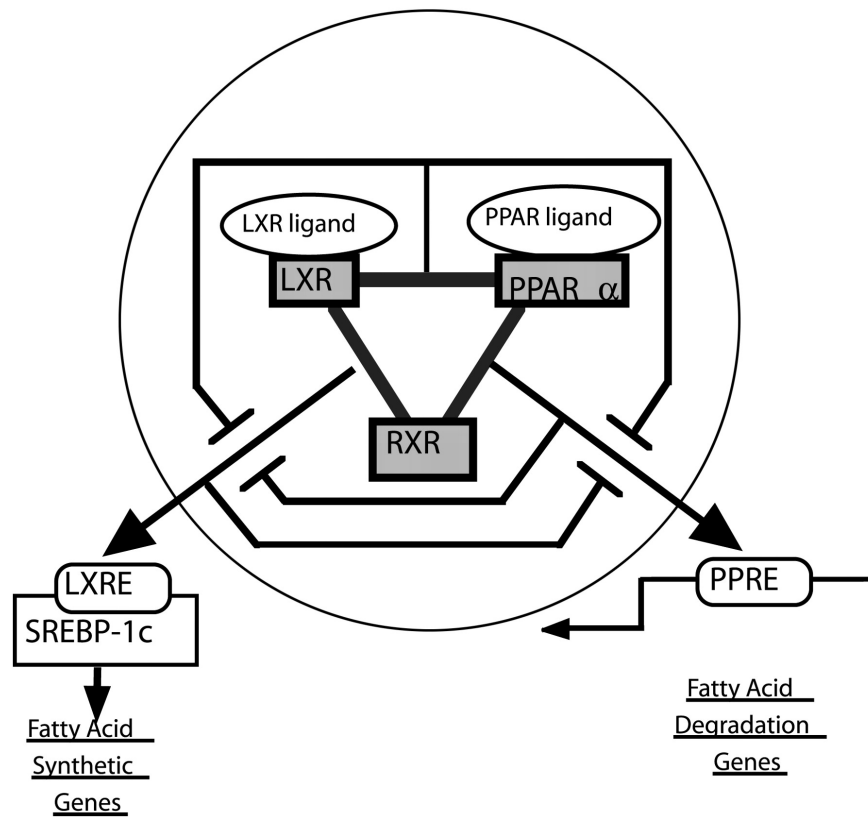


FIG. 3.2.4 – Interactions mutuelles entre PPAR α et le complexe LXR-SREBP1-c dans la régulation de la balance entre anabolisme et catabolisme des acides gras.

La complexation de RXR détermine l'orientation du métabolisme lipidique Les *retinoid X receptor* (RXR) sont une famille composée de RXR α , RXR β et RXR γ . Ils sont capables de former des complexes avec PPAR α et avec les *liver X receptor*, une famille composée de LXR α et LXR β . La balance entre PPAR α et LXR détermine donc la forme sous laquelle se retrouvera complexée RXR.

Le complexe RXR-PPAR α active le catabolisme des acides gras Lorsque RXR est complexé à PPAR α , et que ce dernier est en présence de son ligand (i.e. les acides gras poly-insaturés), il forme un complexe actif capable d'induire les gènes possédant un élément de réponse à PPAR α (PPRE) tels que les gènes du catabolisme des acides gras (*Fatty Acid Degradation Genes*).

Le complexe RXR-LXR active la synthèse des acides gras Lorsque RXR est complexé à LXR [198], il est alors capable d'induire la transcription des gènes possédant un élément de réponse à LXR (LXRE), comme SREBP1-c, un facteur de transcription capable d'induire les gènes impliqués dans la synthèse des acides gras (*Fatty Acid Synthetic Genes*). Ce complexe ne se forme pas en présence des acides gras poly-insaturés car ces derniers empêchent LXR de se lier à LXRE (non représenté) [346].

Le complexe PPAR α -LXR est inactif Lorsque PPAR α se lie à LXR, le complexe ainsi formé est inactif, il en résulte une inhibition (flèches terminées par un T) de la voie de PPAR α et de la voie de LXR.

Figure extraite de T. Ide *et al.*, 2005 [155].

Inhibition de SREBP1 par les acides gras poly-insaturés lors du jeûne Les acides gras poly-insaturés présents suite au relargage d'acides gras libres par le tissu adipeux lors du jeûne, contribuent aussi à l'inhibition de SREBP1. Les acides gras poly-insaturés ω 3 et ω 6 inhibent l'expression [244, 179] et la maturation post traductionnelle [134] de SREBP1-c, diminuant ainsi son activité [342]. De plus, les acides gras poly-insaturés bloquent la voie de signalisation de SREBP1-c par inhibition de LXR α et LXR β ⁸ [346] qui sont des activateurs de SREBP1-c.

SREBP1 active la lipogénèse et la désaturation et l'élongation des acides gras

Dans le foie, SREBP1 active la transcription de nombreux gènes de la glycolyse et de la synthèse des acides gras, [104, 19, 312, 150, 288], jouant ainsi le rôle de médiateur de l'insuline. Cette activation passe notamment par l'induction de la transcription de PKLR, d'ACC [179], de FASN [104] et des gènes responsables de l'élongation des acides gras [227] comme la *fatty acyl elongase*, ELOVL1, ELOVL2, ELOVL5, et ELOVL6 [334]. En outre, le transporteur du glucose SLC2A2 lie SREBP1-c au niveau d'une séquence SRE (voir [203] pour revue). En accord avec ces rôles transcriptionnels sur les gènes cibles, l'absence constitutionnelle de SREBP1 conduit à une baisse de la lipogénèse, comme constaté chez les souris *knockout* pour SREBP1-c [205].

Dans le foie, SREBP1 augmente aussi l'expression des gènes D5D et D6D [217, 334] qui sont les enzymes clefs de la synthèse des acides gras poly-insaturés, et augmente l'expression de D9D [178, 237, 334] qui est responsable de la désaturation des acides gras C16 :0 et C18 :0 en C16 :1 ω 7 et en C18 :1 ω 9 (c.f. figure 2.4.2).

Dans l'adipocyte, SREBP1-c aurait un effet globalement similaire à celui démontré dans le foie [180, 32]. D'autres gènes comme ceux codant pour l'hexokinase (équivalent à la glucokinase de l'hépatocyte), ou le récepteur LDL ont été caractérisés comme cibles de SREBP1-c dans l'adipocyte. De plus SREBP1 a un effet direct sur la différenciation adipocytaire chez le jeune. Dans les muscles, les gènes cibles de SREBP1 incluraient de même les gènes lipogéniques [157] et l'hexokinase II (HK2) [122].

3.2.3 PPAR α stimule la β -oxydation et la cétogénèse lors du jeûne

PPAR α est un membre des *peroxisome proliferator-activated receptors*, une famille de récepteurs nucléaires aux hormones [215]. Les PPARs sont exprimés sous différentes isoformes dont l'expression varie en fonction du type cellulaire. Chez le rat, PPAR α est très exprimé dans les hépatocytes, les cardiomyocytes, les entérocytes et les cellules des tubes proximaux du rein [36]. PPAR α initialement connu pour ses effets sur la prolifération des peroxysomes [195], est aujourd'hui considéré comme l'un des principaux médiateurs des effets du jeûne sur l'expression des gènes [201, 177].

PPAR α possède de nombreux rôles, il est en particulier capable de modifier la vitesse de transcription de nombreux gènes (voir [83] pour une revue) d'où de nombreux effets ([127] pour une liste d'effets chez la souris, [260] et [51] pour une étude des effets chez le porc). En particulier, les gènes impliqués dans le métabolisme des lipides [336, 53, 277, 107, 8, 114], des triglycérides [196] et des phospholipides [196], le métabolisme du glucose [207], le métabolisme des lipoprotéines [107], la genèse des peroxysomes [53], la transcription [53], le cycle cellulaire [53] et l'apoptose [277] sont contrôlés par PPAR α .

En accord avec sa fonction de régulateur transcriptionnel, l'absence de PPAR α entraîne diverses pathologies relatives au métabolisme des lipides (pour une revue, voir [176]) dont une dyslipidémie progressive [65], une obésité et une stéatose hépatique [65, 137].

Cette partie se propose de décrire ses principaux rôles dans le métabolisme des lipides et dans les métabolismes associés (c'est-à-dire cétogénèse et métabolisme du glucose).

PPAR α est activé lors du jeûne par le glucagon et les acides gras poly-insaturés

Lors du jeûne, le glucagon induit une augmentation de l'activité de PPAR α dans les hépatocytes grâce à un mécanisme impliquant l'AMPc et la protéine kinase A [193], et favorise sa translocation dans le noyau grâce à un mécanisme qui dépend de *Mitogen-activated protein kinases p38* (p38 MAPK) et d'AMPK [208].

⁸LXR α et LXR β sont deux isoformes de LXR ou *liver X receptor*.

De plus, lors du jeûne, les acides gras sont massivement relargués du tissu adipeux puis sont captés par différents tissus dont le foie. Parmi eux, les acides gras poly-insaturés sont capables de se lier à PPAR α [207] afin de l'activer [272]. Une fois actif, PPAR α se lie à son partenaire privilégié : le *retinoid X receptor* (RXR⁹) [174]. Le complexe RXR-PPAR α est capable d'induire la transcription des gènes portant un élément de réponse à PPAR (ou *PPAR response element* abrégé en PPRE). Ce mécanisme de liaison avec RXR permet d'exclure mutuellement les voies de PPAR α et de SREBP1 comme décrit sur figure 3.2.4.

PPAR α est un activateur de l'importation des acides gras

PPAR α favorise l'importation des acides gras et leur transport intra cellulaire par stimulation de l'expression des protéines de transport des acides gras telles que CD36 et *plasma membrane fatty acid binding protein* (FABPpm) [22] et d'acyl-CoA synthetase (ACSL) [107]. Il en résulte une importation accrue comme observé dans le coeur [336].

PPAR α est un activateur de la β -oxydation mitochondriale et de l' ω -oxydation [9]

De manière générale, PPAR α est capable d'induire la transcription de nombreux enzymes mitochondriaux. Ainsi les enzymes ACADVL, ACADL, *short chain-specific 3-ketoacyl-CoA thiolase* (i.e. très probablement ACAA2) et ACSL sont moins exprimés dans le foie des souris *knockout* PPAR α que dans les souris sauvages [8]. Chez le porc, le clofibrate (un activateur de PPAR α [210]) induit au niveau hépatique l'augmentation des gènes de la β -oxydation : CPT1A, 3-hydroxyacyl CoA dehydrogenase (3HCDH), 3-ketoacyl CoA thiolase (3KACT), HADHA, ETFDH [51].

Dans le coeur, PPAR α est capable d'induire l'expression constitutive d'au moins sept enzymes de la β -oxydation mitochondriale [336] ce qui entraîne une oxydation plus forte en présence de PPAR α [336] et une épargne de glucose [43].

Dans le muscle squelettique, le coeur et le foie, PPAR α augmente aussi l'importation mitochondriale des acides gras libres en stimulant l'expression de CPT1B¹⁰, d'où une oxydation accrue [107, 114, 51, 264].

Chez le rat, PPAR α est capable de stimuler la transcription des *cytochromes P450, family 4, subfamily A1 A2 et A3* qui sont des enzymes de ω -oxydation [185].

En revanche, PPAR α ne semble pas avoir d'effet sur la β -oxydation peroxysomiale puisque celle-ci est identique entre les souris *knockout* pour PPAR α et sauvages [336, 8]. Rappelons que la β -oxydation peroxysomiale dans le foie est sous le contrôle des LXR [152].

PPAR α est un activateur de la cétogenèse

PPAR α est capable d'induire la transcription d'HMGCS2, l'enzyme clef de la cétogenèse [139] (voir [138] pour une revue sur la régulation de la cétogenèse) chez le porc [51], la souris [138] et l'homme [151].

PPAR α est un activateur de la désaturation des acides gras [335, 128, 217]

PPAR α est un activateur de la transcription de D5D¹¹ et D6D¹², [217, 233, 315]. Sa présence permet ainsi le maintien de l'élongation et de la désaturation des acides gras polyinsaturés même en période de jeûne, situation pendant laquelle SREBP1 est fortement réprimé [217]. PPAR α est aussi connu pour activer l'expression du gène de la D9D dans le foie de souris [223]. Cependant, cet enzyme possède lors du jeûne un comportement différent de D5D et D6D [234], ce qui laisse supposer qu'il n'est pas entièrement régulé comme les deux autres désaturases.

⁹Il y a trois RXR connus : RXR α , RXR β , RXR γ

¹⁰L'enzyme responsable de l'entrée des acides gras dans la mitochondrie du muscle et du coeur, supposée limitant pour la β -oxydation

¹¹D5D est aussi connu sous le nom de FADS1

¹²D6D est aussi connu sous le nom de FADS2, cet enzyme est une étape limitante de l'élongation désaturation des acides gras poly-insaturés [55].

3.3 Conclusion

Les voies métaboliques sont communes Les voies biochimiques du métabolisme des lipides sont communes aux différents organes et aux différentes espèces. Il est donc possible d'exploiter les éléments communs afin de construire un modèle générique (i.e. valable quelque soit l'organe ou l'espèce) du métabolisme des lipides.

Les comportements sont différents Malgré un ensemble de voies communes, les comportements observés varient en fonction de l'organe et de l'espèce. Cette diversité des comportements est la conséquence de la diversité dans les voies de régulation et dans l'intensité de ces régulations.

Exploitation de la généricité Afin d'exploiter l'aspect générique des voies, nous allons proposer une méthode de construction d'un modèle valable quelque soit l'organe et l'espèce. Ce modèle sera paramétré différemment en fonction du contexte afin de faire ressortir les différences dans l'intensité de ses régulations.

Un problème complexe De nombreux éléments sont connus pour jouer directement un rôle dans le métabolisme des lipides et ces éléments interagissent beaucoup entre eux. Ces éléments sont décrits dans la bibliographie scientifique qui regroupe comme nous l'avons vu dans les chapitres précédents un grand nombre d'informations concernant ces différents éléments et leurs interactions. De plus, ce métabolisme est fortement connecté aux autres fonctions du métabolisme de l'énergie ce qui rend la définition d'une frontière d'étude difficile.

La bibliographie, bien que contenant de nombreuses informations ne propose pas de méthode afin de réunir ces dernières et d'interpréter efficacement les données pour comprendre le système dans sa globalité.

Un problème multi échelles Plusieurs échelles temporelles et spatiales interviennent dans la régulation du métabolisme des lipides (c.f. 3.3.1), et les données expérimentales ne peuvent pas toutes les couvrir simultanément, ce qui complexifie d'autant plus son étude.

L'étude simultanée et combinée de ce métabolisme à toutes ces échelles est un travail trop important pour être réalisé dans le cadre d'une thèse. De plus, en l'absence de jeux de données couvrant simultanément toutes ces échelles, il est empiriquement impossible de statuer sur la validité globale des phénomènes étudiés.

Par conséquent nous chercherons à développer une méthode d'intégration des connaissances uniquement pour les échelles qui correspondent aux principales données expérimentales disponibles.

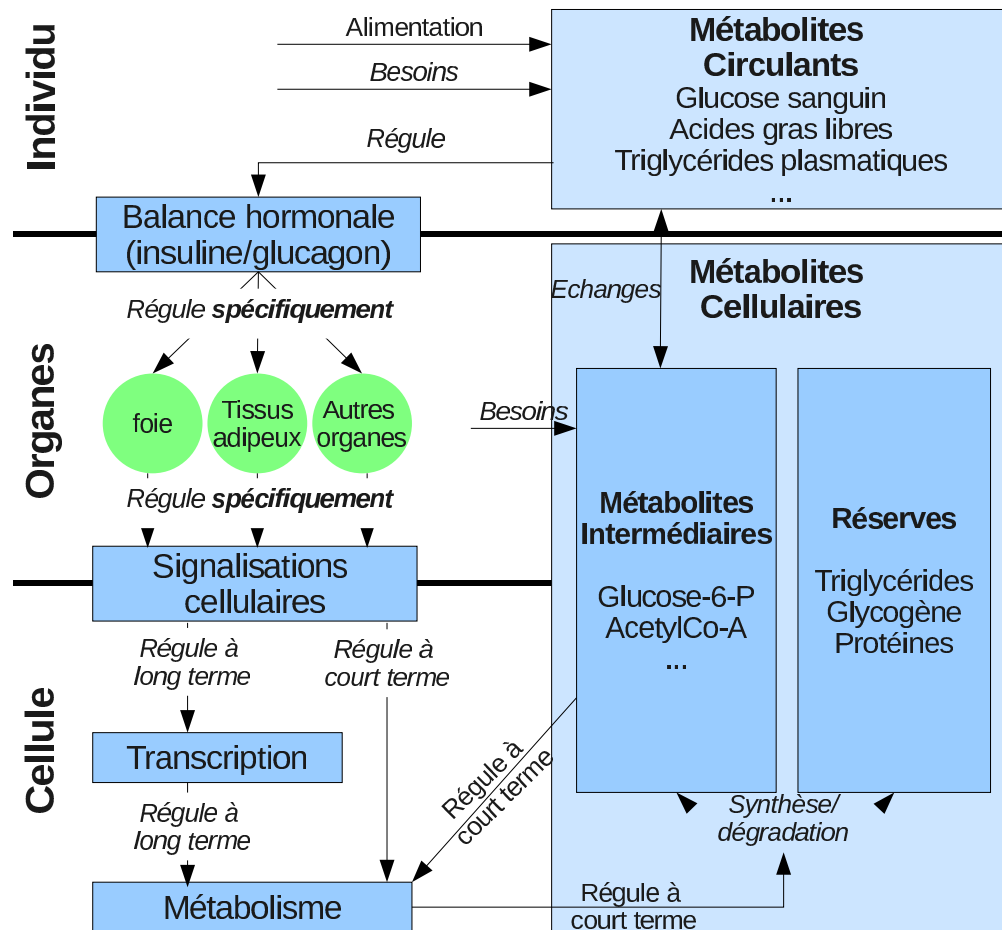


FIG. 3.3.1 – Les différentes échelles du métabolisme énergétique

Les régulations des métabolismes énergétiques et des lipides ont lieu à différentes échelles temporelles et spatiales imbriquées.

L'état physiologique global de l'individu est la conséquence du comportement des différents organes en réponse d'une part à l'apport de nutriments, et d'autre part aux besoins propres à leurs statuts (en croissance, en lactation, en exercice ...). Ces organes sont régulés de manière coordonnées par la balance hormonale qui diffuse un message unique dans l'ensemble de l'individu. Cette balance est interprétée spécifiquement dans chaque organe par la génération de seconds messagers cellulaires responsables des différences de comportement inter-organes. Au final, chaque cellule intègre alors le signal hormonal avec ses propres signaux internes afin de réguler son métabolisme. Les modifications du métabolisme cellulaire sont alors intégrés à l'échelle de l'individu, modifiant ainsi son état global.

Les différentes régulations ont lieu à différentes échelles de temps. Tout d'abord la régulation de l'activité enzymatique en réponse à un changement de l'état de la cellule (comme par exemple la régulation du cycle de Krebs en fonction du rapport ATP/ADP ^a) est extrêmement rapide. La diffusion du signal hormonal, en raison du temps nécessaire aux hormones pour atteindre les différents organes par la voie sanguine est un peu plus longue (de l'ordre de la minute). La régulation de la transcription est quant-à elle beaucoup plus longue (de l'ordre de l'heure). Ces échelles temporelles de régulation se croisent avec différents besoin de l'animal : effort ponctuel, variation quotidienne de l'alimentation, lactation, croissance... Ces différents niveaux sont à la fois connectés de manière ascendante : le niveau supérieur est défini par les comportements résultants et émergents du niveau inférieur, et descendante : le comportement du niveau inférieur est dépendant de l'état du niveau supérieur. Ces connections mutuelles rendent ainsi le système global difficile à appréhender.

^aCette régulation est de l'ordre de la seconde

Deuxième partie

**Modélisation dynamique des systèmes
biologiques**

Chapitre 1

La modélisation : une approche pour comprendre les systèmes complexes tels que le métabolisme des lipides et identifier leurs leviers de contrôle

Enjeux : contrôler les organismes La finalité de l'étude des organismes biologiques est de pouvoir contrôler ces derniers, c'est-à-dire d'influencer leurs propriétés afin qu'elles aient des valeurs intéressantes. Dans le cadre du métabolisme des lipides, on cherchera par exemple à obtenir des individus sains, ayant une croissance rapide et possédant une quantité et une composition en lipide particulière pour les tissus d'intérêts commerciaux.

Objectifs : identifier des leviers de contrôles internes Le contrôle de ces organismes passe par l'identification de leviers de contrôles, c'est-à-dire de moyens permettant d'obtenir les propriétés souhaitées. Ces leviers peuvent être des facteurs environnementaux (exemple : alimentation, température ...) ou des facteurs internes à l'individu, comme un génotype particulier par exemple. Dans le cadre de cette thèse, nous recherchons les leviers de contrôles parmi les éléments internes à l'organisme relatifs au métabolisme des lipides c'est à dire parmi les molécules du métabolisme des lipides présentes *in-vivo* et les régulateurs de ces molécules présents *in-vivo*.

Les caractères d'intérêt concernent souvent des propriétés macroscopiques comme la quantité ou la composition en acides gras d'un tissu. Nous devons donc mettre en place une stratégie qui permette à la fois d'identifier des leviers de contrôle au niveau moléculaire et de prédire l'effet de leur variation au niveau supérieur. Formulé autrement, ce problème revient à maîtriser les comportements émergents d'un système à partir du contrôle de sa mécanique interne.

Une méthode générique La méthode présentée ici s'applique au métabolisme des lipides, mais elle reste valable pour l'étude d'un système vaste et complexe dont un grand nombre d'éléments et de relations entre éléments sont connus et pour lequel de nombreuses données expérimentales sont disponibles.

1.1 Le métabolisme des lipides est vaste et complexe

1.1.1 De nombreuses données expérimentales hétérogènes

Dans cette sous section, nous tentons d'identifier les caractéristiques communes de la plupart de ces données.

Les données bas débit Ces données concernent tout d’abord les expériences “bas débit” de biochimie ayant permis d’identifier les métabolites (voir KEGG, HMDB et LipidBank) et les voies biochimiques (voir KEGG). Ces données bas débit sont essentiellement quantitatives, relativement précises, et décrivent la plupart du temps l’évolution d’un ensemble de valeurs au cours du temps.

A ces données s’ajoute la mesure (très souvent *in vitro*) des paramètres cinétiques (ex K_m , V_{max} . . .) relatifs à l’activité de nombreux enzymes (voir Brenda-enzyme, SABIO-RK). Ces données décrivent la plupart du temps la variation de la vitesse de réaction d’un enzyme en fonction d’un paramètre (ex : la quantité de substrat), les autres paramètres (ex : quantité d’enzyme, quantité d’activateur) étant constants. La connaissance de ces données permet de décrire la cinétique des voies métaboliques à condition de connaître à la fois les quantités *in-vivo* de tous les enzymes, les activités de chacun de leur substrat, produit et régulateur, ainsi que les fonctions décrivant la vitesse de réaction de chaque enzyme en fonction des activités de ses substrats de ses produits de ses régulateurs et de ces paramètres cinétiques. En pratique il nous manquera toujours les quantités d’enzymes *in-vivo*, certains paramètres cinétiques et la plupart des fonctions décrivant les vitesses de réaction (comme ACC par exemple, dont aucune fonction ne décrit à notre connaissance sa vitesse de réaction en fonction des quantités de ces régulateurs), par conséquent nous n’utiliserons pas ces données.

Les données haut débit Les données haut débit sont essentiellement obtenues grâce à des puces (ou *microarray*) qui décrivent la variation d’expression de plusieurs milliers de gènes entre quelques conditions (voir ArrayExpress, SMD pour des bases généralistes de puces). Les données haut débit issues d’autres techniques ne seront pas étudiées ici.

Conclusion : deux niveaux d’approche On peut grossièrement séparer les principales données expérimentales disponibles en deux niveaux : les données quantitatives bas débit décrivant essentiellement le métabolisme et les données transcriptomiques haut débit décrivant de manière qualitative des variations d’expressions de gènes. Bien que cette typologie soit partielle et grossière, nous nous attacherons dans cette thèse à exploiter uniquement ces deux niveaux.

Les types de données décrits ici dans le cadre du métabolisme des lipides existent pour de nombreux métabolismes. La démarche présentée dans cette thèse leur est transposable pourvu qu’il y ait suffisamment de données exploitables et de connaissances théoriques relatives aux voies biochimiques et aux régulations.

1.1.2 De nombreuses connaissances éparpillées dans la bibliographie

La bibliographie contient un ensemble de connaissances théoriques, c’est-à-dire un ensemble de propositions visant à expliquer les phénomènes observés. Basiquement la bibliographie contient essentiellement deux types d’informations : les effets et les mécanismes. Les effets sont l’interprétation théorique des résultats expérimentaux (par exemple “*Insulin stimulates the expression of genes encoding glycolytic and fatty-acid synthetic enzymes*” [270]). Les mécanismes sont une description théorique de la manière dont interagissent les molécules (par exemple “*The insulin receptor is a tyrosine kinase that undergoes autophosphorylation, and catalyses the phosphorylation of cellular proteins such as members of the IRS family, Shc and Cbl.*” voir Fig2 Saltiel *et al.*, 2001 [270] pour le mécanisme complet.)

Que ce soit de manière générale, ou relativement au métabolisme des lipides, la bibliographie est très abondante (c.f. tableau 1.1.1). La grande majorité de la bibliographie est présente sous forme de textes dans les publications. Ces publications sont dispersées, parfois dures à retrouver et nécessitent d’être lues pour en extraire leur contenu. Cette abondance ne doit pas nous faire oublier l’existence potentielle d’éléments et d’interactions inconnus ou potentiellement faux. Ainsi, même lorsque nous nous baserons sur la bibliographie, nous nous attacherons à respecter au maximum le principe de réfutabilité.

Une stratégie afin de palier à l’abondance des informations bibliographiques consiste à se tourner vers des bases de connaissances (ex : KEGG, Ingenuity, TRANSPATH, Pathway Commons). Ces bases regroupent sous une forme organisée des relations diverses entre des molécules. Elles sont issues de la lecture manuelle puis de l’annotation des publications. A notre connaissance il n’existe pas de base dédiée aux régulations du métabolisme des lipides, et nous nous sommes donc tournés vers les bases généralistes. Une autre stratégie

TAB. 1.1.1 – **Estimation du nombre de publications en biologie relatives au métabolisme des lipides.**

| années | nombre de publications | | rapport ($\times 10^{-3}$) ^b |
|-----------|------------------------|--------------------------------------|---|
| | total | métabolisme des lipides ^a | |
| 1990-1994 | 2 065 939 | 4768 | 2.3 |
| 1995-1999 | 2 286 084 | 5201 | 2.3 |
| 2000-2004 | 2 824 145 | 7995 | 2.8 |
| 2005-2009 | 3 443 583 | 9839 | 2.9 |

^aLa recherche a été effectuée dans Pubmed le 27 octobre 2009. Les mots clefs ‘‘lipid metabolism’’ OR ‘‘fatty acids metabolism’’ ont été recherchés pour chaque tranche d’années.

^bNombre de publications relatives au métabolisme des lipides/ Nombre de publications totales

consiste à se tourner vers des experts. Ces derniers ont une connaissance intégrée et précise de la bibliographie, mais cette connaissance se limite souvent à un (ou quelques) domaine(s) très pointu(s), et ne couvre pas l’intégralité du métabolisme des lipides. De plus, il est difficile de bâtir des ponts entre domaines d’expertise afin de les connecter.

1.1.3 Conclusion

Actuellement, les expériences et l’analyse de la bibliographie permettent de tester des sous ensembles de la théorie, mais ne permettent pas de l’analyser dans sa globalité. Cette approche est incompatible avec notre objectif de recherche de régulateurs parmi un vaste ensemble, car elle ne fonctionne que sur de petits sous domaines.

Il est actuellement impossible de dire s’il n’existe pas dans la bibliographie du métabolisme des lipides des informations contradictoires ni de statuer globalement sur l’accord entre les connaissances théoriques et les données expérimentales. Ceci s’explique par la taille du système étudié, par sa complexité et par la dispersion et l’absence d’intégration des informations qui le décrivent.

Ces constatations sur le métabolisme des lipides restent vraies pour un grand nombre de voies biochimiques pour lesquelles la bibliographie décrit un grand nombre de mécanismes et d’effets concernant de nombreuses molécules. La démarche développée dans cette thèse peut être transposable à ces voies.

1.2 Générer des connaissances dans un système complexe

Notre problématique de recherche de leviers de contrôle du métabolisme des lipides est un sous cas d’une problématique plus générale qui est la génération de connaissances. En sciences expérimentales, une connaissance est une proposition admise, potentiellement réfutable et corroborée par des observations. Dans cette section nous verrons appliquer la démarche expérimentale sur un système complexe afin de rechercher des leviers de contrôle du métabolisme des lipides parmi les éléments qui constituent ce métabolisme.

1.2.1 La démarche expérimentale permet de générer des connaissances

La démarche expérimentale classique permet de construire, tester, analyser et améliorer une théorie. La démarche est la suivante :

1. Récolter des observations
2. Induire une théorie à partir de ces observations. La théorie doit être cohérente avec les observations qu’elle explique.
3. Dédire des informations de la théorie.

4. Tester empiriquement la validité de la théorie (et de ses déductions). Dans le cas où la théorie est cohérente avec les observations, alors ces dernières corroborent la théorie. Dans le cas où elles ne sont pas cohérentes, l'expérience réfute la théorie, il faut alors analyser la théorie au vu des nouvelles observations afin d'en proposer une nouvelle qui respecte ces dernières. On retourne alors à la seconde étape.

Une telle démarche peut amener à proposer plusieurs théories différentes pour expliquer un même ensemble d'observations. Dans ce cas, on gardera tout d'abord les théories les plus prédictives en vertu du principe de réfutabilité puis la plus simple en vertu du principe de parcimonie.

1.2.2 Choix d'une approche systémique

Deux grandes approches sont possibles pour décrire un système comme les régulations du métabolisme des lipides : l'approche "entrée sortie" et l'approche systémique.

L'approche "entrée sortie"

L'approche "entrée sortie" consiste à évaluer comment répond un système en fonction d'un ensemble de variables d'entrée (exemple [23]). Cette approche nécessite un jeu de données comprenant un bon ensemble de valeurs des variables d'entrée couplé aux valeurs des variables de sortie ainsi qu'une fonction mathématique pour faire le lien entre ces valeurs. C'est cette fonction qui joue le rôle de théorie.

Dans le cadre de la recherche de leviers de contrôle du métabolisme des lipides, l'approche "entrée-sortie" est en pratique impossible car il y a un trop grand nombre (plusieurs milliers) de variables d'entrée pour pouvoir toutes les tester, et car la complexité du système est telle qu'il faudrait tester toutes les combinaisons possibles de valeurs de variables d'entrée pour avoir une fonction décrivant correctement sa sortie en fonction de ces entrées. En pratique il est impossible de réaliser autant d'expériences, par conséquent nous rejetons l'approche "entrée sortie" et recherchons une approche permettant de cibler les éléments à tester empiriquement.

L'approche "systémique" [330]

L'approche systémique a pour objectif de décrire la mécanique interne du système puis de déduire de cette dernière le comportement du système. Contrairement à l'approche "entrée sortie", cette dernière nécessite des données expérimentales à la fois sur les entrées et les sorties du système et sur ses éléments internes ainsi qu'une connaissance théorique de la mécanique interne du système.

Afin de cibler au mieux quels leviers de contrôle tester empiriquement, nous proposons d'utiliser le savoir théorique afin de proposer les meilleurs candidats compte tenu des connaissances disponibles. Dans notre cas, nous avons des informations théoriques sur la mécanique interne du métabolisme des lipides et nous recherchons des leviers de contrôle parmi les éléments de cette mécanique. Par conséquent, nous adoptons l'approche systémique.

Lors de cette approche, il faut garder à l'esprit que le modèle n'est pas un moyen de remplacer les expériences. C'est uniquement un moyen de choisir les expériences pertinentes à réaliser au regard des connaissances actuelles. Au final, c'est toujours la constatation empirique d'un phénomène qui a valeur de preuve, et non sa déduction théorique à partir du modèle.

Définition du système étudié

Nous considérons alors le métabolisme des lipides comme un système composé :

d'éléments : les molécules du métabolisme des lipides et ses régulateurs. Lorsque le modèle contient des compartiments, on considère un élément par molécule et par compartiment (ex : un élément *glucose sanguin*, et un élément *glucose cellulaire*).

de relation : les réactions biochimiques du métabolisme des lipides et leurs régulations.

d'une frontière : l'individu. On suppose que l'environnement influence l'individu mais pas l'inverse. Cette frontière peut être ultérieurement restreinte pour choisir un de ses éléments internes (ex l'organe), mais pas étendue.

1.2.3 Modéliser le système permet d'appliquer la démarche expérimentale

Les données expérimentales et les informations théoriques décrivant le système sont disponibles, par conséquent il est *a priori* possible d'appliquer la démarche expérimentale afin de générer des connaissances. Hélas, dans l'état actuel des choses, il est très difficile de réaliser des déductions à partir de la bibliographie à cause de son abondance et de la difficulté à intégrer un grand nombre d'effets concernant un grand nombre de molécules. Ceci bloque la démarche expérimentale à l'étape de confrontation des données à la théorie. De plus, même si une confrontation était possible, le retour sur erreur sera un véritable casse tête. Il est donc nécessaire de formaliser les connaissances théoriques dans un objet permettant de réaliser des prédictions automatiquement, puis de confronter ces prédictions aux données expérimentales. Ceci revient à construire un modèle mathématique.

Face à ce problème, nous avons fait le choix de construire des modèles mathématiques. Un modèle mathématique est défini comme une représentation des aspects essentiels d'un système existant (ou d'un système que l'on va construire) qui présente les connaissances d'un tel système sous une forme analysable. En raison de la dimension des systèmes étudiés, la construction et l'analyse de ces modèles doit pouvoir se réaliser automatiquement et dans un temps raisonnable par des algorithmes.

Les points clefs de la construction de ces modèles sont les suivants.

1. Réunir les informations théoriques.
2. Représenter ces informations dans un formalisme prédictif, analysable par des algorithmes et disposant d'une interface compréhensible par un biologiste.
3. Réaliser des prédictions afin de respecter la troisième étape de la démarche expérimentale.
4. Confronter les prédictions aux données expérimentales. Cette dernière étape implique d'avoir des données expérimentales aussi informatives que les prédictions.

Cette démarche peut amener à proposer plusieurs modèles conformes aux données observées. Une première approche consiste à réaliser de nouvelles expériences afin d'acquérir les données nécessaires pour identifier quel est le meilleur modèle. Une seconde approche est d'appliquer le principe de réfutabilité, puis le principe de parcimonie. L'application du principe de réfutabilité consiste à choisir le modèle qui fournit les prédictions les plus précises (en pratique ceci revient la plupart du temps à choisir le modèle ayant le moins de paramètres inconnus ou évalués sur les données par ajustement). L'application du principe de parcimonie consiste à conserver le modèle le plus simple parmi ceux qui restent.

Chapitre 2

Quel(s) modèle(s) construire pour représenter le métabolisme des lipides ?

Objectifs Nous souhaitons représenter le métabolisme des lipides et ses régulations comme un ensemble de molécules reliées par des interactions (c.f. 1.2.2). Cet ensemble est simplement la représentation formalisée d'un ensemble d'hypothèses. La difficulté réside dans le choix de la représentation, qui doit être suffisamment précise pour décrire les phénomènes connus et intégrer les observations, mais rester suffisamment simple pour être réfutable par l'intégration de données additionnelles.

Quelles que soient les données utilisées, nous ne nous intéresserons qu'au comportement moyen pour chaque condition et nous ne nous intéresserons pas à l'évaluation de la variabilité intra-condition : obtenir une description de la variabilité intra-condition nécessite d'avoir préalablement décrit le comportement moyen et de disposer de données obtenues sur un grand nombre de répétitions dans chaque condition, ce qui n'est pas le cas ici.

Deux modèles pour deux niveaux de données expérimentales Les données haut et bas débit décrivent deux échelles biologiques complémentaires. Les données haut débit informent sur la variation de la transcription de nombreux gènes entre quelques conditions, à partir desquelles on peut éventuellement extrapoler des variations qualitatives de quantités d'enzymes. Les données bas débit informent sur les quantités des divers métabolites au fil du temps, à partir desquelles on peut parfois déduire la vitesse des différentes voies métaboliques. Les deux niveaux (métabolique et génétique) sont en interaction : les concentrations en métabolites observées sur les données bas débit sont la résultante des voies biochimiques du métabolisme des lipides, régulées par la quantité de leurs enzymes, quantités dépendantes de leur transcription. Inversement, les métabolites sont capables d'induire des voies de signalisation ayant des effets sur la transcription des gènes. A ces régulations croisées entre les deux niveaux s'ajoutent les régulations internes comme par exemple l'influence des facteurs de transcription pour le niveau génétique, ou les inhibitions par les produits de réaction pour le niveau métabolique. La compréhension du métabolisme des lipides passe donc par la compréhension de ces deux niveaux, et donc à la fois par l'analyse des données transcriptomiques et métaboliques.

Comme les données expérimentales ont deux niveaux de détail différents, nous avons fait le choix de construire un modèle relatif à chaque type de données.

Le premier modèle a pour objectif à long terme d'analyser et d'exploiter les données haut débit en les confrontant avec les connaissances bibliographiques. Une première étape consiste à formaliser la bibliographie dans un modèle afin de pouvoir la croiser plus tard avec les données haut débit. En raison de la nature de ces données, nous avons pour objectif de construire un modèle qualitatif faisant des prédictions sur la variation de ses éléments entre conditions. En raison du nombre d'éléments dans le modèle, nous recherchons une

approche qui soit facilement automatisable et qui ne pose pas de problèmes de calcul insolubles en un temps raisonnable.

Le second modèle a pour objectif de décrire à l'échelle de l'organe les données bas débit du métabolisme des lipides de manière synthétique, dynamique et quantitative. De plus, en raison de l'importance de l'aspect temporel (comme par exemple les variations métaboliques en fonction du temps de jeûne) dans la régulation de ce métabolisme, ce modèle a pour objectif de proposer une description dynamique.

L'utilisation de deux niveaux d'approche pose le problème du lien entre ces approches. La stratégie envisagée sur le long terme est d'identifier des éléments clés potentiels dans le premier modèle, et d'évaluer leurs effets en les introduisant dans le second, suite à une bibliographie ciblée et d'éventuelles expériences supplémentaires. Cette vision à long terme ne sera pas développée dans le cadre de cette thèse.

2.1 Choix des systèmes d'équations ordinaires pour modéliser les données cinétiques quantitatives.

Cette section présente les différents formalismes utilisés pour modéliser des données cinétiques quantitatives (les systèmes d'équations différentielles ordinaires, les formalismes dérivés des équations différentielles ordinaires et les systèmes multi agents), puis discute du choix des systèmes d'équations différentielles ordinaires.

2.1.1 Les systèmes d'équations différentielles ordinaires

Formalisme [131] Les systèmes d'équations différentielles ordinaires sont un formalisme classique permettant de décrire l'évolution d'un ensemble de variables quantitatives (les éléments) au fil du temps grâce à des fonctions mathématiques (les relations).

Les variables sont des nombres réels décrivant l'état actuel du système (notés x_i pour l'élément i). Dans notre cas, ces variables représentent des quantités de molécules dans un compartiment (par exemple : $x_{glucose}$ représente la quantité de glucose dans une cellule).

Pour chaque élément, une fonction mathématique décrit sa vitesse de variation en fonction du temps et des valeurs des variables du système. Ces fonctions sont de la forme : $\frac{dx_i}{dt} = f_i(t, x_1, \dots, x_n)$, avec f_i une fonction mathématique et t le temps.

Les fonctions mathématiques peuvent éventuellement contenir des paramètres dont la valeur numérique est inconnue. Par exemple $\frac{dx_{glucose}}{dt} = -K \times x_{glucose}$ décrit une loi de dégradation d'ordre un dans laquelle K , un paramètre de valeur inconnue, représente la constante de dégradation spontanée du glucose.

Des modèles déterministes De manière générale un système dynamique est déterministe s'il permet de déterminer de manière unique le futur d'un point à partir d'un état initial. Ainsi la trajectoire d'un tel système est entièrement déterminée par l'état initial de ce système. Un état se définit par la valeur des variables et leur vitesse de variation (i.e. leur dérivée).

Les modèles d'équations différentielles ordinaires sont déterministes à condition de connaître les valeurs de leurs variables (i.e. conditions initiales) et de leurs paramètres¹. Ils produisent alors une unique trajectoire qui peut dans quelques cas simples être produite par une fonction mathématique (intégration formelle), ou dans tous les cas par une succession de valeurs approchées (intégration numérique²).

La principale difficulté lors de la construction d'un modèle d'équations différentielles ordinaires est de déterminer la valeur des paramètres du modèle et de mesurer précisément ses conditions initiales. Les valeurs des paramètres peuvent être recherchées dans la bibliographie (ex constantes cinétiques des enzymes), ou

¹Les valeurs des dérivées des variables sont alors directement déductible en calculant la valeur des fonctions mathématiques du modèle.

²Des modules d'intégration numérique existent pour les principaux logiciels de mathématiques tels que <http://www.r-project.org/> et Matlab, sous forme de programme indépendant (ex : Sundials) ou sous forme de bibliothèque pour les principaux langages de programmations (C/C++ avec GNU/GSL, Fortran avec ODEPACK et java avec Opale ...).

ajustées afin que le modèle soit cohérent avec un jeu de données. Dans ce dernier cas, il faut s'assurer que le modèle est suffisamment simple par rapport aux données disponibles pour qu'il ne soit pas surparamétré.

Lorsque le modèle comporte des paramètres indéterminés, il est parfois possible d'étudier les comportements valables quelles que soient les valeurs des paramètres indéterminés comme par exemple la présence d'attracteurs [37]. Ce sont ces comportements qui sont confrontés aux données.

Le déterminisme des modèles d'équations différentielles ordinaires doit, dans certains cas être nuancé. En effet si le modèle est chaotique ³ ou fortement sensible aux valeurs de ses paramètres ⁴, l'incertitude sur les paramètres ou sur les conditions initiales, généralement due à un ajustement à partir de données expérimentales à forte variabilité peut suffire à briser la fiabilité des prédictions. Ce cas peut se présenter lorsque le modèle différentiel est hautement non linéaire comme par exemple un phénomène de type "tout ou rien" (régulation génétique ou cascade enzymatique par exemple). Une telle situation est détectée par une étude de sensibilité [200].

Hypothèses biologiques Par définition, les variables sont des quantités (ou des concentrations) de molécules dans un compartiment, il est donc nécessaire de définir des compartiments que l'on peut considérer comme homogènes. Cette hypothèse est très fréquemment admise, généralement par parcimonie. Un contre exemple serait l'étude des milieux présentant un fort gradient de concentration.

Dans la description des fonctions différentielles, les variables représentent généralement des activités, tandis que les variables mesurées et les variables du modèle représentent des quantités de molécules (ou des concentrations). Par conséquent une modélisation rigoureuse implique de faire un lien entre activité et concentration. Dans la majorité des cas, on construit ce lien en faisant l'hypothèse que la concentration est égale à l'activité. Vérifier cette hypothèse, et construire la fonction liant activité à concentration lorsque ces dernières sont différentes n'est pas toujours évident, en particulier dans le cas où une molécule est séquestrée ou présente sous forme de petits amas (ex : gouttelette lipidique en milieu aqueux).

De plus, pour que les modèles différentiels soient valables, il faut travailler sur un système contenant dans chacun de ses compartiments un nombre de molécules suffisamment grand afin de pouvoir modéliser les variables par des nombres réels (un contre exemple est l'étude de la quantité d'ADN pour laquelle il y a soit une soit deux molécules, un bon exemple est l'étude du glucose dans le sang où de nombreuses molécules sont présentes dans un compartiment homogène).

Exemples d'utilisation L'aptitude à prédire des cinétiques quantitatives fait que ces modèles sont beaucoup utilisés pour réaliser des simulations dans de nombreux domaines (physique, biologie, dynamique des population ...). Ils sont en particulier utilisés pour la description des cinétiques de réaction [63], le comportement d'un organe [47] et la croissance des animaux d'élevage (ex : chez le boeuf [241] et le porc [254]).

2.1.2 Les formalismes dérivés des systèmes d'équations différentielles ordinaires

Les modèles stochastiques [2] Le fait d'induire des petites variations stochastiques permet de mettre en évidence un comportement chaotique par rapport à un comportement stable, d'évaluer la robustesse des modèles et de faire des prédictions sur la variabilité des variables.

Afin de prendre en compte les aspects aléatoires d'un phénomène, il est possible de rajouter du bruit aux fonctions. Le fait d'induire des petites variations stochastiques permet de mettre en évidence un comportement chaotique par rapport à un comportement stable et d'évaluer la robustesse des modèles. L'introduction de variations stochastiques permet aussi de modéliser la variabilité des variables. En pratique ces modèles sont simulés numériquement un grand nombre de fois en réalisant des tirages aléatoires afin de produire un faisceau de trajectoires.

³Un système est chaotique quand une petite variation des conditions initiales induit une forte variation des trajectoires.

⁴ Un système est sensible aux valeurs de ses paramètres si une petite variation de certains de ces paramètres induit une forte variation des trajectoires.

En simulant un vaste ⁵ ensemble de trajectoires, il est possible dans le cas où ces dernières convergent de calculer une trajectoire moyenne ⁶. Elle représente alors une trajectoire fortement probable du comportement moyen. Elle peut alors être analysée comme toute trajectoire issue d'un modèle déterministe.

Les modèles à dérivées partielles Les modèles d'équations différentielles ordinaires décrivent la dérivée du système par rapport à une unique variable indépendante : le temps. Les modèles à dérivées partielles sont une généralisation dans laquelle on décrit des dérivées partielles par rapport à un ensemble plus important de variables indépendantes. Concrètement ces modèles sont utilisés pour modéliser des phénomènes dépendant de l'espace en introduisant les dérivées partielles par rapport aux coordonnées. Ils permettent de s'abstraire de l'hypothèse des compartiments homogènes nécessaire dans les systèmes d'équations différentielles ordinaires. Dans certains cas, ces équations sont solubles analytiquement et dans la plupart des cas elles peuvent être intégrées numériquement.

Les modèles d'équations différentielles linéaires par morceaux Le principe de la construction des modèles d'équations différentielles linéaires par morceaux est présenté figure 2.1.1. Ces modèles sont particulièrement adaptés pour la description de phénomènes comportant des seuils comme les régulations génétiques (ex : [257, 123]). En effet dans ces systèmes, la vitesse de la transcription ($\frac{dARMm}{dt}$) en fonction de ces activateurs (A) suit une loi de Hills ($\frac{dARMm}{dt} = V_m \times \frac{A^n}{A^n + K}$) avec un fort coefficient ($n > 3$), ce qui la rend facilement approximable par une fonction en forme d'une unique marche d'escalier (si $A < K$, $\frac{dARMm}{dt} = 0$, sinon $\frac{dARMm}{dt} = V_m$), modélisable par deux demi-équations linéaires⁷. Ils nécessitent pour pouvoir être construits de connaître l'ordre des seuils et la position relative des attracteurs par rapport aux seuils.

La discrétisation de l'espace des variables rend ces modèles non déterministes : pour un ensemble contenant les valeurs relatives des seuils, les valeurs relatives des variables par rapport aux seuils, et les valeurs relatives des attracteurs par rapport aux seuils, plusieurs successeurs sont généralement possibles. En contre partie, comme le modèle est discret il est possible de calculer toutes les trajectoires possibles et d'analyser cet ensemble. En pratique de tels calculs sont possibles sur des systèmes relativement petits (i.e. environ 20 variables, ce qui donne un graphe des états de plusieurs dizaines de milliers de sommets).

Les systèmes multi-agents [189, 98, 245]

Les systèmes multi-agents sont composés d'un ensemble d'éléments (à petite échelle) ayant chacun un ensemble de propriétés (par exemple une molécule a une position et un type) et de règles locales décrivant l'évolution de ces éléments (par exemple : les molécules se déplacent aléatoirement, et lorsque deux molécules de type A et B se rencontrent, elles réagissent pour former C). La simulation consiste à calculer un faisceau de trajectoires (décrivant l'évolution temporelle de chaque élément) pour le système, puis à analyser ce faisceau afin de déduire des propriétés à grande échelle (comme par exemple la concentration moyenne de chaque molécule dans une région donnée au fil du temps).

En pratique ces modèles sont très utilisés pour donner une description des phénomènes émergents d'un système sachant sa mécanique interne et pour modéliser les systèmes pour lesquels la notion de concentration n'a pas de sens (par exemple en raison d'un faible nombre d'au moins un type de molécule ou d'une forte hétérogénéité spatiale).

Pour être utilisables, ces modèles nécessitent de définir toutes les propriétés des éléments. Dans le cadre du métabolisme il faudrait connaître par exemple connaître la taille de chaque molécule, la probabilité de chaque métabolite de réagir avec les enzymes sachant leurs positions relatives, la vitesse de déplacement dans le solvant de toutes les molécules . . . , ce qui est loin d'être le cas.

⁵L'ordre de grandeur est de plusieurs centaines, il dépend de la variabilité des trajectoires.

⁶Il est toujours possible de calculer une trajectoire moyenne, mais lorsque la distribution des trajectoires n'est pas monomodale, ça n'a pas vraiment de sens.

⁷Ce formalisme permet aussi de modéliser plusieurs seuils.

2.1.3 Choix du système d'équations différentielles ordinaires

Les systèmes multi-agents ne sont pas retenus car de nombreuses propriétés microscopiques des molécules sont inconnues.

Nous n'avons pas retenu le modèle d'équations linéaires par morceaux car, contrairement aux régulations génétiques, les réactions métaboliques ne sont pas facilement discrétisables. De plus les données observées sont quantitatives et ne sont donc pas complètement intégrables par ce type de modèle.

Le modèle différentiel classique comporte les garanties minimales pour satisfaire nos objectifs : il est dynamique, quantitatif et déterministe. Ses hypothèses biologiques sont *a priori* une approximation raisonnable de la réalité et ne risquent pas d'être facilement contredites par les données. Ce modèle constitue donc un choix possible.

Le modèle stochastique permet de prédire à la fois le comportement moyen et la variabilité de ce comportement. Cette description de la variabilité permet notamment d'évaluer la vraisemblance des observations ce qui facilite grandement les raisonnements. La construction de ce modèle nécessite d'introduire des lois stochastiques correctement paramétrées. Nous ne connaissons pas *a priori* la forme de ces lois et les données disponibles ne permettent pas d'ajuster précisément les paramètres de ces dernières, en particulier les paramètres responsables de la variabilité. Par conséquent, cette modélisation reste d'un niveau de détail trop élevé par rapport à nos objectifs et par rapport aux informations disponibles ; elle ne peut donc pas être réalisée de manière satisfaisante.

Le modèle à dérivées partielles offre lui aussi tous les critères requis, mais dans la mesure où nous ne disposons d'aucune donnée spatialisée intra compartiment, nous n'avons aucune information, ni aucune observation sur l'hétérogénéité de ces derniers et ne pouvons donc pas la modéliser. Nous faisons donc l'hypothèse parcimonieuse de compartiments homogènes et ne prenons pas en compte cette méthode.

Les modèles dérivés des systèmes d'équations différentielles ordinaires sont des spécifications de ces derniers dans divers cas particuliers impliquant des hypothèses supplémentaires non vérifiées (équations différentielles linéaires par morceaux) ou proposant une complexité de modélisation supérieure aux données observables afin de lever des hypothèses largement raisonnables dans notre contexte (modèle stochastique, modèle à dérivées partielles). Nous conservons donc le modèle d'équations différentielles ordinaires qui est la plus simple modélisation quantitative de la variation des quantités (ou des concentrations) d'un ensemble de molécules au cours du temps.

2.2 Les modèles dynamiques sont adaptés pour analyser les séries temporelles, mais pas pour exploiter les données de *microarray*

Objectifs Nous recherchons un modèle permettant d'intégrer les données issues de *microarray*, c'est-à-dire un modèle dans lequel ces données sont interprétables comme les valeurs des variables du modèle.

Nous recherchons un modèle capable de comparer les valeurs de ces variables avec les informations décrites dans la bibliographie, et sur lequel nous disposons d'outils d'analyse afin de rechercher des régulateurs clés du métabolisme des lipides.

Les données issues de *microarray* informent sur des variations entre conditions Les données transcriptomiques issues de puces à gènes (ou *microarray*) nous informent sur la variation d'un élément entre quelques conditions. Pour chaque couple de conditions et pour chaque gène il est possible de dire si la quantité d'ARNm a augmenté, diminué ou n'a pas eu de variation significativement différente de la variabilité expérimentale.

Dans le cadre de cette thèse, nous ne considérerons que les cas où les conditions sont comparées par paires. Les autres cas, comme par exemple les expériences où les puces décrivent la variation du transcriptome au fil du temps ne seront pas étudiés en tant que tel. Dans le cas des séries temporelles, on réalisera des comparaisons de temps deux à deux, par exemple en comparant toutes les données par rapport à un élément

de référence comme le temps 0. Cette approximation nous permet de traiter les données temporelles, mais on perd la relation d'ordre entre les différents points expérimentaux. Actuellement les jeux de données transcriptomiques contenant de nombreuses conditions, ou décrivant des séries temporelles sont rares à l'exception d'organismes modèles (ex : [188,110], [331] sur 40 gènes) et comportent peu de temps différents.

2.2.1 Les modèles dynamiques classiques pour intégrer les données qualitatives.

Les modèles dynamiques décrivent une succession d'états, ou d'ensemble d'états possibles au fil du temps. Les principaux modèles dynamiques utilisés pour décrire des données qualitatives (comme les données issues de *microarray*) sont les réseaux booléens et leur généralisation aux modèles logiques généralisés.

Les réseaux booléens

Les réseaux booléens sont des modèles très simples. Ils ont été développés afin de modéliser les systèmes répondant à des règles de type "tout ou rien" tels que les systèmes régulés génétiquement [172]. Il est possible de construire des réseaux booléens en partant d'*a priori* (bibliographiques), ou par inférence à partir de données [173]. Lorsqu'ils sont construits à partir d'*a priori*, ils permettent alors de tester partiellement la cohérence entre le réseau construits et des données issues de *microarray* (exemple chez *Escherichia coli* [130]⁸).

Elements Dans un réseau booléen un élément est une variable pouvant prendre deux valeurs 0 ou 1. Le sens du 0 et du 1 est à définir en fonction du contexte. Par exemple le 1 peut représenter une molécule présente en grande quantité ou une molécule sous forme active tandis que le 0 une molécule présente en petite quantité ou sous forme inactive.

Relations Les relations sont exprimées sous la forme de fonctions logiques décrivant l'état de chaque élément du système au temps $n+1$ en fonction de l'état d'un ensemble d'éléments au temps n (i.e. transitions). Les modèles booléens conformes à cette définition sont dynamiques, déterministes et à temps discret.

Hypothèses La réalisation d'un réseau booléen s'appuie sur les hypothèses biologiques suivantes [300] :

- Les éléments ne peuvent prendre que deux états. Ceci empêche par exemple de considérer des gradations dans les quantités ou dans les seuils de sensibilité.
- Dans les modèles booléens, les valeurs hautes et basses des éléments se situent de part et d'autre du seuil déclenchant leurs effets. Sur les données de puces les valeurs fortes et faibles se répartissent par rapport à la moyenne des valeurs observées. Il n'y a donc *a priori* aucune raison pour que les valeurs fortes et faibles de la puce coïncident avec les valeurs fortes et faibles du modèle booléen.
- Les régulations peuvent être approximées par des règles booléennes.
- Dans les modèles synchrones, les transitions ont toutes lieu en même temps. Cette hypothèse induit souvent des comportements du modèle qui ne sont pas observés expérimentalement.

Généralisation asynchrone Une généralisation de ces réseaux consiste à supposer que les transitions peuvent avoir lieu dans un ordre non déterminé [97], et une seule à la fois⁹. C'est une généralisation intéressante dans la mesure où le calcul synchrone des transitions n'est pas très cohérent avec la réalité biologique (on sait que certaines régulations sont plus rapides que d'autres par exemple). Dans ces modèles, on calcule alors l'ensemble des trajectoires possibles quel que soit l'ordre des transitions ce qui rend le modèle non déterministe.

⁸Dans cet exemple, un réseau booléen a été construit à partir des informations de *Regulon DB* puis confronté à des données expérimentales, 70 à 87% de ce réseau était cohérent avec les données expérimentales

⁹En effet la probabilité qu'un système franchisse deux seuils à la fois est *a priori* négligeable.

Généralisation stochastique Une autre généralisation consiste à supposer que les transitions ont lieu de manière stochastique [290]. La simulation de tels modèles produit un ensemble de trajectoires.

Les réseaux logiques généralisés [75]

Les réseaux logiques généralisés sont une extension des réseaux booléens. Les éléments sont des variables pouvant prendre un nombre fini de valeurs (par exemple 0=“pas du tout”, 1=“un peu”, et 2=“beaucoup”). Le choix de ces seuils a le même sens que dans les modèles booléens. Pour chaque sommet on définit une fonction mathématique qui décrit l'état vers lequel le système tend à varier en fonction de la valeur des variables. Ces fonctions peuvent contenir des paramètres comme par exemple la valeur d'un seuil. Par hypothèse, les éléments varient les uns après les autres dans un ordre inconnu. On considère alors tous les ordres possibles et on calcule un ensemble de trajectoires. Ces modèles sont donc dynamiques, à temps discret et non déterministes.

Ces modèles sont très proches des modèles d'équations linéaires par morceaux. Les intervalles des modèles linéaires par morceaux sont représentés par des valeurs discrètes dans les réseaux logiques généralisés.

Ces réseaux ont permis de décrire entre autres la dynamique du phage λ [320] et la formation de motifs ventraux-dorsaux et le contrôle des réseaux de gènes *gap* chez la drosophile [273].

Ces modèles étant non déterministes, ils prédisent un ensemble de trajectoires possibles. En pratique, la vitesse de calcul de cet ensemble dépend du nombre d'états accessibles par le système qui, dans le pire des cas, croît de manière exponentielle avec le nombre de sommets [120]. Les modèles du phage λ [320] et de la drosophile [273] contiennent 4 variables, des modèles plus récents peuvent en contenir environ une vingtaine.

Le logiciel GINsim permet de visualiser environ 500 états possibles, soit au moins l'ensemble de tous les états possibles pris par 8 variables booléennes. Des stratégies d'élagage sont mises en place, afin de ne considérer qu'un sous-ensemble des états possibles lorsque le graphe des états est trop important [120].

2.2.2 Les modèles dynamiques ne sont pas adaptés pour analyser les comparaisons entre conditions

Les modèles booléens, leurs généralisations asynchrones et aléatoires et les modèles logiques généralisés ont une interprétation facile et cohérente avec la démarche d'analyse habituelle des biologistes, ce qui facilite leur construction et leur analyse. Par contre ils analysent des données différentes des données observées sur les *microarray* et les règles logiques peuvent être délicates à écrire et à paramétrer.

Les modèles dynamiques travaillent sur des quantités alors que les comparaisons de deux conditions informent sur des variations. Les modèles booléens et les réseaux logiques généralisés modélisent des quantités (discrétisées de part et d'autre des seuils de déclenchement de leurs effets) alors que nous observons des variations. Afin de pouvoir les exploiter pour analyser des comparaisons entre deux conditions, nous devons supposer qu'une augmentation correspond à une variable au-dessus du seuil de déclenchement de son effet et qu'une diminution correspond à une variable en-dessous du seuil de déclenchement de son effet, ce dont nous n'avons *a priori* aucune idée.

Les modèles dynamiques décrivent des trajectoires alors que les comparaisons de deux conditions informent sur des variations. Ces modèles décrivent des trajectoires ou des ensembles de trajectoires alors que les comparaisons entre deux états informent sur des variations entre le début et la fin d'une expérience. Tant que nous ne disposons pas de données expérimentales portant sur des séries temporelles, nous ne pouvons pas évaluer facilement la validité de ces modèles.

De plus, dans le modèle booléen asynchrone ou modèle logique généralisé, l'espace des trajectoires possibles est *a priori* très grand, et risque fort de ne pas pouvoir être calculé sur un ensemble aussi vaste que le métabolisme des lipides et ses régulations. Introduire de l'aléatoire ne lève en rien les limites évoquées précédemment, rajoute des paramètres inconnus et complique les simulations.

La modélisation de la dynamique est difficile. La modélisation de la dynamique du système nécessite de définir des règles décrivant l'état suivant du système à partir de l'état actuel ou bien définissant un attracteur du système en fonction de son état actuel. Cette définition revient à écrire les fonctions logiques du modèle booléen, à écrire et à paramétrer les fonctions logiques du modèle logique généralisé et à définir pour chaque sous-espace euclidien les attracteurs du modèle d'équations linéaires par morceaux.

Conclusion Par conséquent, pour exploiter les données portant sur des variations entre deux conditions, nous n'utilisons ni les modèles booléens, ni leurs généralisations asynchrone ou stochastique, ni les modèles logiques généralisés, car ces derniers décrivent des trajectoires et sont difficiles à construire et à tester. Nous recherchons un formalisme décrivant des règles de causalité qui permet d'intégrer les données issues de comparaison entre conditions.

2.3 Les graphes d'influences : un formalisme adapté au croisement entre données bibliographiques et données issues de *microarray*

La description des trajectoires est superflue Lorsque nous souhaitons intégrer des données issues de *microarray* nous n'avons pas besoin de décrire la dynamique complète du système, ni de savoir comment les quantités de molécules se situent par rapport à un ensemble de seuils car les données à modéliser ne contiennent pas (ou rarement) ces informations. Nous avons simplement besoin d'un système qui mette en relation les données observées avec les connaissances et les hypothèses, sans nécessairement modéliser la dynamique.

En ne modélisant pas la dynamique, nous contournons les problèmes de construction et d'évaluation inhérents aux systèmes dynamiques. En contrepartie nous ne prédisons plus de trajectoires, ce qui n'est pas un compromis très coûteux étant donné que nous en observons très peu.

Tester la compatibilité entre données observées et un modèle est suffisant Une première approche pour analyser les données haut débit sans modéliser de dynamique est de s'intéresser à évaluer la vraisemblance de ce que l'on observe sachant un ensemble de dépendances fonctionnelles. Cette approche se réalise par la construction de modèles bayésiens [351,18] qui représentent les dépendances fonctionnelles par des probabilités conditionnelles. Ces derniers, bien que très adaptés pour la description d'une unique condition n'exploitent pas les variations, et sont donc d'un niveau de détail trop faible pour décrire la plupart des données de *microarray*.

Une autre alternative à la modélisation de la dynamique est de décrire des règles de causalité. On ne s'intéresse alors plus à la ou aux trajectoires possibles, mais aux liens entre les règles de causalité déduites de la bibliographie et les observations expérimentales. Cette dernière peut alors être exploitée pour inférer des modèles, tester la cohérence entre des données et un modèle existant ou prédire un ensemble de valeurs possibles.

Formalisme des graphes d'influences

Les graphes d'influences sont un cas particulier de graphe de dépendance fonctionnelle Les graphes d'influences, tout comme les réseaux booléens, les modèles logiques généralisés et les réseaux bayésiens sont une extension des graphes de dépendances fonctionnelles.

Un graphe de dépendance fonctionnelle est un graphe orienté (i.e. des sommets et des flèches) dans lequel les sommets représentent des éléments, et les flèches des dépendances fonctionnelles. Une dépendance fonctionnelle entre un sommet A et un sommet B (notée $A \mapsto B$) indique que la valeur (ou la variation) de A dépend de celle de B . La manière dont B dépend de A n'est pas forcément précisée dans un graphe de dépendance fonctionnelle.

Un graphe d'influences est un moyen de spécifier partiellement les dépendances fonctionnelles par des contraintes sur les valeurs possibles des sommets. Contrairement aux réseaux booléens ou aux modèles logiques généralisés, ces graphes travaillent sur des variations, et non sur des quantités.

Sommet Les sommets d'un graphe d'influences peuvent prendre 3 valeurs possibles : (+) pour un élément dont la quantité augmente entre deux conditions, (-) pour un élément dont la quantité diminue entre deux conditions et (?) pour un élément non observé ou un élément dont la variation n'est pas significativement différente de la variabilité expérimentale. On considère alors que ce sommet peut soit augmenter soit diminuer (soit (?) = (+) ou (-)).

Influences Les relations des graphes d'influences sont représentées par une flèche entre deux sommets allant d'un sommet source (A) à un sommet cible (B). Cette flèche peut être annotée par 3 types de relations : $A \xrightarrow{+} B$ pour A a un effet positif sur B (ex activation), $A \xrightarrow{-} B$ pour A a un effet négatif sur B et $A \xrightarrow{?} B$ pour A a un effet non nul sur B (soit $\xrightarrow{?} = \xrightarrow{+}$ ou $\xrightarrow{-}$).

Dans ce formalisme, un effet positif de A sur B revient à dire que la dérivée partielle de B par rapport à A est toujours positive. Ce formalisme capte la plupart des relations biologiques à de rares exceptions pour lesquelles le signe de la dérivée partielle change en fonction des valeurs des variables. On peut par exemple citer les enzymes sensibles à l'inhibition par effet de substrat, pour lesquels le substrat a un effet positif à faible quantité et négatif lorsqu'il est en fort excès et les enzymes ayant une valeur de pH optimale pour lequel H^+ a tantôt un effet positif, tantôt négatif.

2.3.1 Méthodes d'analyse des graphes d'influences

Etude de la topologie des graphes

Une fois un graphe de dépendances fonctionnelles construit, il est possible d'analyser la topologie de ce graphe [12] en recherchant par exemple les sommets très connectés au reste du réseau qui sont probablement des régulateurs clés, les composantes fortement connexes qui sont généralement interprétées comme des groupes d'éléments fonctionnant ensemble, les chemins entre deux éléments (de A vers B) qui sont des voies de signalisation possible de l'effet de A sur B ou des modules fonctionnels tels que des boucles de régulations.

Lorsqu'on dispose d'un graphe d'influences, on peut réaliser une analyse plus fine, en distinguant par exemple les boucles positives des boucles négatives qui ont des implications sur la présence d'oscillations ou de multiples attracteurs [321, 306]. La présence de certains motifs induit aussi des propriétés dynamiques du système [3].

La validité des motifs peut aussi être testée en comparant ces derniers avec des données expérimentales. Par exemple, M.J.Herrgard *et al.*, [142] ont construit un ensemble de relations à partir de *Regulon DB* dans lequel ils ont identifié cinq types de motifs de régulations : les régulons, les régulons complexes (i.e. les gènes qui reçoivent plus d'un activateur ou un inhibiteur), les interactions cibles de régulateurs, les molécules cibles et les *feed-forward loop*¹⁰. Ils ont ensuite croisé chacun de ces motifs avec des données expérimentales issues de 163 expériences chez *Escherichia coli* [287] et 904 chez la levure [126]. Pour cela ils comparent la cohérence de ce motif par rapport à un jeu de données expérimentales à la cohérence de motifs aléatoires par rapport au même jeu de données. Au final ils ont identifié que 50% des motifs de *Regulon DB* sont cohérents avec les données expérimentales utilisées, et en déduisent que de nombreuses régions du modèle sont incomplètes.

Confrontation de données et de modèles

Chez les organismes modèles *Escherichia coli* et la levure, plusieurs publications proposent de construire des graphes d'influences, de les analyser afin d'identifier des incohérences, de corriger ces dernières puis de réaliser des prédictions.

¹⁰Un exemple de *feed-forward loop* est : A active B , A et B activent C .

Chez *Escherichia coli*, R.M. Gutiérrez-Ríos *et al.*, [130] ont construit un graphe d'influences à partir de *Regulon DB*, ils ont ensuite défini des règles binaires¹¹ locales à chaque régulon puis ont évalué la cohérence de ces dernières. Ils ont alors pu conclure que de nombreuses régulations étaient redondantes.

Chez *Escherichia coli*, afin de coupler les régulations transcriptomiques et le métabolisme, M.W. Covert *et al.*, [67,66] ont couplé un réseau booléen décrivant les données transcriptomiques à un réseau métabolique dont les états quasi-stationnaires discrétisés sont calculés (à partir de l'état du modèle booléen) puis introduits dans le modèle booléen. Le premier modèle construit était incohérent avec les données expérimentales, les incohérences ont été analysées et de nouvelles règles ont été écrites afin d'améliorer la cohérence. Le modèle après correction est capable de réaliser des prédictions sur les valeurs des sommets avec une fiabilité de 90%.

Une autre méthode consiste à générer à partir du graphe d'influences un ensemble de contraintes sur les valeurs des sommets puis à tenter de résoudre ces dernières, soit en choisissant les valeurs des sommets non observés, soit en fixant le signe des influences \mapsto [292]. Cette méthode permet à partir d'un graphe d'influences et de données expérimentales de tester de manière globale si le graphe isolé est cohérent et si le graphe est cohérent avec des données. En cas d'incohérence globale, cette méthode permet d'identifier les éventuelles zones d'incohérences et dans le cas où tout est cohérent cette approche permet aussi de réaliser des prédictions sur les sommets non observés.

Les graphes d'influences sont adaptés au croisement des données *microarray* avec les données bibliographiques

L'analyse à haut débit des puces passe nécessairement par le croisement entre le plus grand ensemble possible de données observées avec les connaissances actuelles afin d'évaluer leur cohérence. De plus, l'objectif de cette thèse étant d'identifier des éléments clés parmi un vaste ensemble, l'exploitation conjointe de la bibliographie et des données haut débit constitue un outil de choix pour proposer de bons candidats.

Les graphes d'influences sont adaptés aux données haut débit car contrairement aux réseaux booléens et aux réseaux logiques généralisés qui travaillent sur des quantités par rapport à un ou des seuils, les graphes d'influences travaillent sur des variations, c'est-à-dire exactement sur le type de données observées dans les jeux de données haut débit courants.

De plus, le formalisme des graphes d'influences a l'avantage de modéliser de manière relativement simple la plupart des interactions décrites dans la bibliographie. Ils sont par exemple adaptés pour décrire les effets des transcrits de gènes sur la régulation de la transcription qui sont les interactions les plus courantes et les plus susceptibles *a priori* d'expliquer les données issues de *microarray*.

A ces avantages, l'exploitation des graphes d'influences est à la fois un bon moyen d'évaluer la fiabilité des graphes construits en vérifiant leur cohérence, et un bon moyen de rechercher des régulateurs clés par analyse de leur topologie, ou en recherchant des ensembles de sommets dont la variation induit (dans les prédictions du graphe) la variation des éléments d'intérêt.

Conclusion

Les méthodes d'analyse de topologie permettent d'identifier des éléments importants. De plus, à partir d'un graphe d'influences (même incomplet) et d'un jeu de données, il est possible de réaliser une analyse de cohérence. En cas d'incohérence, cette méthode permet d'identifier les éléments à analyser en priorité pour corriger le graphe. En cas de cohérence, cette analyse démontre la nécessité d'acquérir de nouvelles données expérimentales pour invalider le modèle et permet de réaliser des prédictions. Un graphe cohérent constitue aussi un modèle raisonnablement fiable sur lequel réaliser les analyses de topologie. Par conséquent les outils mathématiques existants offrent une piste intéressante dans notre problématique d'exploitation de données haut débit et de recherche d'éléments clés. Il est donc utile de construire un graphe d'influences du métabolisme des lipides.

Les graphes utilisés ont tous été obtenus chez des espèces modèles relativement simples et dont la régulation est *a priori* mieux décrite dans la bibliographie que celle du métabolisme des lipides. Bien que

¹¹Des règles qui sont soit respectées soit non respectées

capables de donner des résultats sur des réseaux incomplets, la transposition des méthodes autres que le test de cohérence au métabolisme des lipides nécessite de construire un graphe relativement fiable et complet du métabolisme des lipides. La faisabilité de la construction d'un tel graphe sera traitée dans la partie II.

2.3.2 Méthodes de construction des graphes d'influences

La construction de graphes d'influences se découpe en deux parties : obtenir le graphe des dépendances fonctionnelles puis définir les relations du modèle. La première partie est commune pour les modèles booléens, les réseaux logiques généralisés, les modèles bayésiens et les graphes d'influences, la seconde dépend du formalisme utilisé.

Ces deux étapes peuvent se réaliser soit à partir de la bibliographie, soit par inférence à partir de données expérimentales, soit en mélangeant ces deux approches.

Les informations nécessaires

Les informations nécessaires à la construction du graphe de dépendances fonctionnelles Tout d'abord il est nécessaire de posséder une liste d'éléments à mettre en relation. Cette liste peut être issue de la bibliographie, d'une base de connaissances ou de l'annotation d'un jeu de données expérimentales. Elle constitue les sommets du graphe.

Les deux informations minimales nécessaires pour la construction d'une dépendance fonctionnelle sont la corrélation entre les éléments (pour savoir qu'il y a une flèche qui les relie) et une information permettant de séparer la cause de la conséquence (pour savoir dans quel sens mettre la flèche).

A ces informations, il est possible d'ajouter la notion de dépendance fonctionnelle directe ou indirecte. Cette dernière permet d'identifier les dépendances fonctionnelles qui sont la résultante de dépendances fonctionnelles déjà existantes. Par exemple si $A \mapsto B$, $B \mapsto C$ et $A \mapsto C$, il est possible qu'il existe biologiquement une influence de A sur C y compris en l'absence de variations B ou bien que $A \mapsto C$ soit une pure conséquence des deux autres influences (i.e. une influence indirecte). Dans ce dernier cas, il est préférable de la retirer du modèle afin de le rendre plus informatif (C ne dépend directement que de B , il est donc possible de déduire des informations sur C d'après celles sur B même si on ne connaît rien sur A).

Les informations nécessaires à la définition des relations L'étape suivante : la construction des relations nécessite d'obtenir des informations qui décrivent comment les éléments dépendent les uns des autres, c'est-à-dire de définir les fonctions logiques dans les modèles booléens, les fonctions logiques et leurs paramètres dans les réseaux logiques généralisés, les fonctions de probabilité et leurs paramètres dans les modèles bayésiens et les signes des relations et les règles de génération de contraintes dans les graphes d'influences.

Construction des modèles par inférence

L'inférence permet de générer un modèle directement à partir de données expérimentales. Cette approche offre l'avantage de ne nécessiter aucun *a priori* et constitue par conséquent une bonne technique exploratoire, ou un bon moyen de générer de nouvelles connaissances et de valider les anciennes. L'inconvénient majeur, est qu'on ne peut pas inférer au-delà de ce que les données contiennent. Leur quantité et leur qualité sont donc cruciales.

Construction du graphe de dépendances fonctionnelles par inférence La construction du graphe de dépendances fonctionnelles peut se réaliser par inférence à partir d'un jeu de données de type *chIP-on-chip* et de connaissances sur les promoteurs des gènes. Les données *chIP-on-chip* permettent d'identifier à l'échelle d'un génome les sites sur lesquels se lie une protéine de liaison à l'ADN [10]. Lorsque on sait que ces sites jouent le rôle de promoteurs de gènes connus, on peut alors en déduire que la protéine a un effet sur ces gènes.

Cette méthode permet donc d'avoir un graphe comportant uniquement des dépendances fonctionnelles directes (car liaison protéine-gène *in-vivo*). Un exemple est la reconstruction des régulations transcriptionnelles chez *Saccharomyces cerevisiae* à partir de données *chIP-on-chip* et de réseaux d'interactions protéiques [197].

Une autre technique de construction du graphe de dépendances fonctionnelles est l'exploitation de séries temporelles [293, 305] ou de données issues de perturbations génétiques. Brièvement, les corrélations permettent de générer les flèches, leur sens est trouvé grâce à l'hypothèse d'antériorité de la cause sur la conséquence dans les séries temporelles ou en supposant que les variations sont causées par la mutation dans les mutants. Cette méthode fournit des graphes de dépendances fonctionnelles, sans préciser si ces dernières sont directes ou indirectes. Cette méthode est grandement limitée lorsque tous les éléments biologiques en interaction ne sont pas observés (ce qui est presque toujours le cas en biologie) [305].

L'inférence de réseaux est aussi réalisée par la construction de réseaux bayésien [351], ou de réseaux bayésiens dynamiques [348, 222] à partir de données expérimentales.

Construction des relations du modèle par inférence La construction des relations par inférence consiste à choisir un ensemble de relations paramétrées de manière à obtenir des prédictions cohérentes avec des données observées. En pratique on cherche soit à minimiser la distance prédictions observations (ex : modèles booléens [293], modèles booléens aléatoires [173]), soit à maximiser la vraisemblance des observations (ex : modèle bayésien, tables de vérité du modèle logique généralisé [305]).

Construction des modèles à partir de la bibliographie

Les informations disponibles dans la bibliographie Les informations présentes dans la bibliographie concernent généralement les effets des molécules et leurs mécanismes d'action. Les effets sont généralement des descriptions de ce qui varie avec ou sans une molécule particulière. Ils informent sur les dépendances fonctionnelles (sans préciser si elles sont directes ou indirectes), et donnent des informations sur les relations comme par exemple le signe des corrélations. Les effets sont des informations qui sont fondamentalement identiques à ce que contiennent les jeux de données lors de la construction par inférence.

Les mécanismes, une fois interprétés informent généralement sur les dépendances fonctionnelles et précisent la plupart du temps si ces dernières sont directes ou indirectes. Ils donnent aussi la plupart du temps des informations sur les relations.

Le plus souvent les mécanismes des régulations génétiques et des régulations de l'activité enzymatique peuvent se résumer aux trois types d'effets (\vdash^+ , \vdash^- et $\vdash^?$) décrits dans le formalisme des graphes d'influences (c.f. 2.3). A ces mécanismes il faut rajouter les réactions biochimiques pour avoir une couverture de la plupart des mécanismes décrits dans la bibliographie. L'interprétation de ces dernières n'est par contre pas du tout triviale. Par exemple si on sait que deux molécules *A* et *B* forment un complexe *C*, on obtient un graphe d'influences peu informatif comportant toutes les combinaisons de flèches possibles et l'écriture des relations n'apporte pas de connaissance sur le système.

Des descriptions plus précises existent dans la bibliographie, comme par exemple des lois cinétiques ou l'identification d'un facteur limitant d'une réaction, mais elles ne concernent généralement que de petits sous systèmes, comme un enzyme ou une partie d'une voie métabolique.

Exploiter la bibliographie nécessite un formalisme adapté Choisir d'exploiter la bibliographie nécessite d'adopter un formalisme pour lequel les informations nécessaires à l'écriture des relations sont réellement présentes dans la bibliographie. De plus, quitte à modéliser la bibliographie, autant profiter au maximum des informations qu'elle comporte afin d'interpréter les données au regard du plus vaste ensemble possible de connaissances. Lors de cette approche, on choisira donc un compromis entre un formalisme expressif pour pouvoir décrire un grand nombre d'informations présentes dans la bibliographie, et un formalisme simple pour qu'il ne nécessite pas d'informations inconnues.

Nous choisissons donc d'intégrer un maximum d'effets et de mécanismes, sans nous préoccuper des informations plus détaillées, en effet, ces informations nécessitent d'utiliser un formalisme trop riche dans lequel la

plupart des zones ne sont pas assez bien connues pour être décrites, et dont le niveau de détail des prédictions serait supérieur aux données de *microarray* rendant ainsi impossible leur exploitation.

Choix d'une méthode de construction

L'inférence est limitée par les données et la complexité des systèmes. La reconstruction du graphe de dépendances fonctionnelles nécessite des informations biologiques particulières. Si les corrélations sont faciles à obtenir, différencier les causes des conséquences à haut débit est déjà plus difficile et nécessite souvent d'exploiter des données temporelles ou des données obtenues sur des mutants ou de coupler des données de *chIP-on-chip* avec une liste de promoteurs et des gènes qui leurs sont associés. De plus, afin d'identifier les dépendances fonctionnelles indirectes, il est nécessaire d'obtenir des données biologiques qui portent sur les mécanismes comme les données de *chIP-on-chip*. Ainsi nous nous éloignons de l'objectif initial d'analyse de données issues de *microarray* comparant deux conditions. De plus, la forte complexité des systèmes eucariotes nécessite une masse de données bien plus importantes que celle qui sont disponibles.

L'inférence des relations des modèles (une fois le graphe de dépendances fonctionnelles acquis) reste souvent problématique car les modèles comportent souvent de nombreux paramètres inconnus [101].

Fondamentalement, le problème de l'inférence d'un système complexe d'un seul bloc est un problème d'ingénierie inverse difficile, et il est beaucoup plus simple de découper le système en petits morceaux facilement analysables. La bibliographie contient justement un très grand nombre de conclusions obtenues sur des sous systèmes. En pratique, l'inférence de réseaux de régulations, même après correction, est incomplète (exemple chez *Saccharomyces cerevisiae* [93]) et contient parfois des erreurs (exemple chez *Escherichia coli* [125]), même sur des organismes plus simples que les mammifères.

Choix de la construction de graphes d'influences à partir de la bibliographie La bibliographie contient toutes les informations nécessaires pour construire des graphes d'influences. De plus, ces dernier permettent de modéliser les effets et la plupart des mécanismes couramment décrits à l'exception des réactions biochimiques qui nécessitent des approximations.

Le formalisme des graphes d'influences est adapté à la description de la bibliographie à large échelle ; en effet cette dernière contient un grand nombre d'informations concernant des effets (ex : "l'injection de A dans une cellule entraîne l'augmentation de B") et des mécanismes (ex : "A est capable d'activer B par phosphorylation").

La plupart du temps, il est possible de déduire des mécanismes des informations sur leurs effets. Les seuls cas réellement problématiques sont les mécanismes ne comportant pas de distinction entre la source et la cible (on ne connaît pas le sens de la flèche), comme par exemple la capacité qu'ont deux molécules A et B à se lier entre elles (faut-il considérer cela comme la formation d'un complexe AB, d'une séquestration de A, de B ?) et réactions biochimiques sur lesquelles nous n'avons aucun *a priori*. Dans notre cas, nous commencerons déjà par l'intégration des données relatives à la régulation des gènes que nous pouvons facilement formaliser et n'aborderons pas l'intégration des réactions biochimiques.

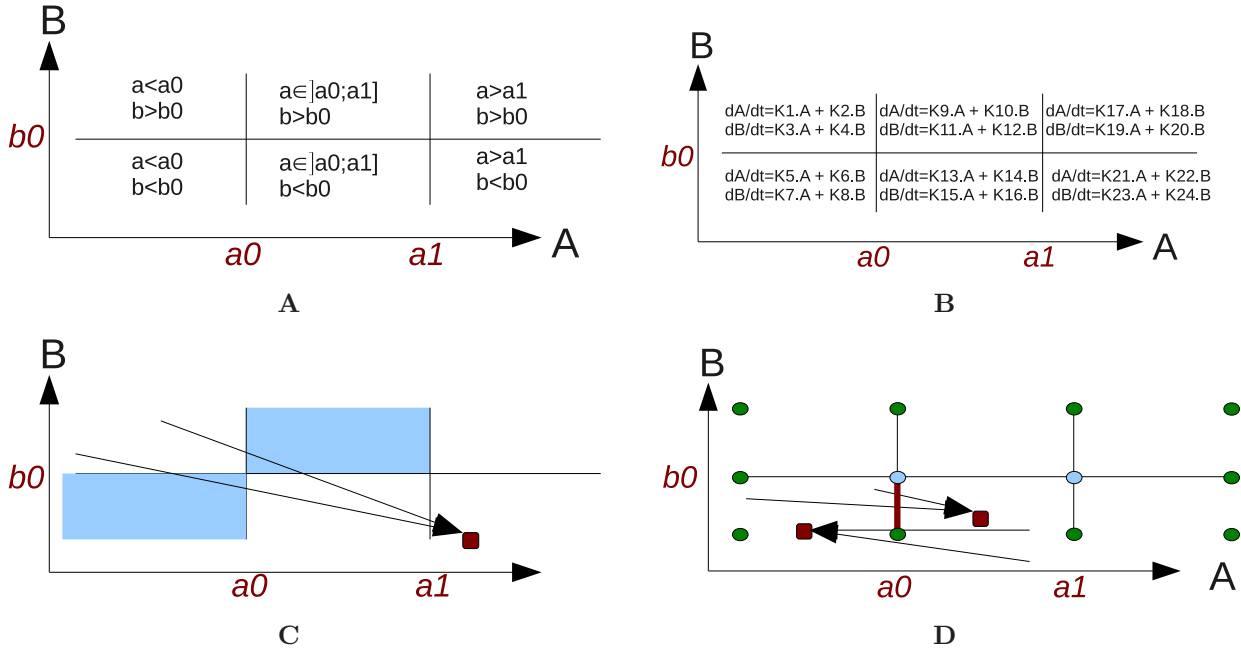


FIG. 2.1.1 – **Principe d'un modèle d'équations différentielles linéaires par morceaux** Lors d'une première étape (**A**), les variables (A et B) sont discrétisées grâce à des seuils (a_0, a_1 et b_0), formant ainsi une partition de l'espace des variables en sous-espaces euclidiens.

L'étape suivante (**B**) consiste à écrire dans chaque sous-espace euclidien un système d'équations différentielles linéaires décrivant la vitesse de variation de chaque variable (dA/dt et dB/dt) en fonction des valeurs des autres variables et de paramètres ($K1 \dots K24$).

La résolution de ces systèmes d'équations linéaires (**C**) permet pour chaque sous-espace euclidien de définir un attracteur (carré rouge) vers lequel tend (flèches noires) à évoluer le système. Le schéma représente ici uniquement l'attracteur du sous espace euclidien $a \leq a_0$, $b \geq b_0$. En connaissant le sous espace euclidien dans lequel se trouve chaque attracteur, il est alors possible de définir pour chaque sous espace euclidien, un ensemble de successeurs (cases bleus) correspondants aux futurs possibles du système.

Les successeurs peuvent être des sous espaces euclidiens (**C**), des frontières entre sous espaces euclidiens, comme par exemple la ligne rouge sur le schéma **D** due aux deux attracteurs (représentés par des carrés rouges), des intersections entre frontières (représentés par des ronds bleus), ou des limites (représentés par des ronds verts). Par exemple le rond vert en haut à droite représente un attracteur pour lequel A et B augmentent indéfiniment. Cet ensemble représente tous les états possibles du système.

Au final on décrit l'ensemble des trajectoires possibles du système par un graphe dans lequel les sommets sont les états possibles du système, et dans lequel il y a une flèche entre deux sommets (de E_1 vers E_2) si et seulement si E_2 est un successeur possible de E_1 .

Troisième partie

Conclusion : objectifs de la thèse

Problématique

Dans le cadre de cette thèse nous nous intéresserons plus particulièrement à deux questions :

- Quelle est la hiérarchie d'importance des voies métaboliques et de leur régulation vis à vis d'un contexte? La réponse à cette question permet d'identifier les éléments à étudier en priorité afin de comprendre la dynamique du métabolisme des lipides dans un contexte défini par un jeu de données expérimentales.
- Dans quelle mesure peut-on exploiter la bibliographie afin d'identifier des éléments importants?

Hiérarchie d'importance des voies métaboliques et de leur régulation

Nous considérons qu'une voie métabolique ou une régulation est importante si elle est nécessaire pour expliquer un jeu de données expérimentales. Le contexte est ici défini par le jeu de données employé. Pour cela nous construisons de manière ascendante le modèle le plus simple possible constitué de voies biochimiques et de régulations connues dans la bibliographie cohérent avec un jeu de données expérimentales.

Au cours de cette méthode, nous exploitons la généralité du métabolisme des lipides en développant une méthode ascendante de construction de modèles. Cette méthode permet de comparer les différents modèles (et donc les différents contextes) entre eux.

Exploitation haut débit de la bibliographie

L'ensemble des connaissances est trop vaste pour être exploité manuellement, nous allons donc développer une méthode la plus automatique possible afin d'exploiter les données bibliographiques. Cette méthode s'appuie sur des bases de données généralistes de bibliographie. Une première étape est donc l'évaluation de la qualité de leur contenu extractible et exploitable pour analyser des données expérimentales haut débit.

Résultats

Première partie

Analyse d'un système complexe décrit par des données transcriptomiques haut débit

Préambule : quelques caractéristiques concernant les données haut débit

Dans cette section, nous décrivons les caractéristiques des données haut débit classiquement utilisées en biologie.

De vastes jeux de données

Pour répondre aux questions biologiques, de vastes jeux de données qui décrivent l'état de nombreux éléments d'un système d'intérêt sont collectés. Ils sont ensuite analysés afin d'induire de leurs observations le comportement et la mécanique des systèmes étudiés. La collecte simultanée de nombreux éléments obtenus avec un même protocole sur un même matériel biologique est rendue possible grâce aux techniques haut débit comme les *micro-array* (puces à ADN) permettant de mesurer simultanément les niveaux d'ARN messagers de plusieurs milliers de gènes. Ces technologies ont connu des progrès importants ces dernières années en terme de coût et de fiabilité. Actuellement, les *micro-array* permettent d'obtenir des informations sur l'expression¹² de presque tous les gènes d'un génome et des techniques telles que la spectrométrie de masse permettent de mesurer les quantités de plusieurs centaines de protéines [92] ou de quelques dizaines de métabolites.

Des données essentiellement qualitatives

Bien qu'il soit possible de faire de la quantification relative avec des *micro-array*, la précision est moins bonne que des techniques à plus bas débit comme la RT-PCRq. Les quantités relatives obtenues par *micro-array* diffèrent parfois de celles obtenues par *Northern Blot* [316] ou par RT-PCR [71] en particulier pour les gènes peu ou très exprimés [259]. Les données qualitatives en terme de sens de variations ("augmente", "diminue" ou "variation non significative") restent elles beaucoup plus fiables que les données quantitatives [86]. L'exploitation de telles données nécessite donc l'utilisation de raisonnements qualitatifs.

Fiabilité des données et des conclusions

La construction de raisonnements reposant sur l'utilisation de multiples données reste un problème difficile. En effet ils reposent souvent sur l'hypothèse que toutes les données sont valides simultanément. Cette contrainte implique d'avoir une très bonne fiabilité des mesures. Par exemple, si on utilise 10 mesures, chacune étant fiable à 95%, la probabilité d'avoir toutes ces données de justes en même temps est inférieure à 60%. Pour avoir 95% de chances d'avoir simultanément 30 000¹³ éléments fiables, il faudrait une fiabilité individuelle de plus de 99.9999%.

¹²L'expression est évaluée par dosage relatif des quantités d'ARN messagers entre différents individus soumis à des conditions expérimentales différentes ou ayant des génotypes différents par exemple.

¹³Ce qui correspond environ au nombre de gènes du génome humain

En pratique, les biologistes évaluent un taux de faux positif [21]. Ce taux est souvent sous estimé car des effets potentiels comme par exemple le laboratoire dans lequel a lieu l'expérience ne sont pas pris en compte [59] dans les calculs alors qu'ils influent sur la variabilité expérimentale. Ainsi, les biologistes réalisent bien souvent une validation d'un sous ensemble des données haut débit par des expériences utilisant d'autres techniques pour mesurer l'expression des gènes (comme par exemple la RT-PCR quantitative) couplées à une expertise basée sur les comportements attendus de certains éléments.

Conclusion

L'utilisation de l'ensemble simultané de ces données, sans repasser par le bas débit, doit donc prendre en compte le risque d'avoir des données erronées et nécessite de développer soit des méthodes d'analyse robustes, soit des méthodes capables d'identifier les zones d'erreurs.

Chapitre 1

Identification à partir d'un échantillon de publications des méthodes les plus courantes pour analyser des données transcriptomiques.

L'objectif de cette partie est d'identifier les méthodes récentes couramment utilisées pour l'analyse de données transcriptomiques portant sur la comparaison entre quelques conditions. Cette analyse est réalisée par lecture d'un échantillon de publications.

1.1 Echantillon de publications retenues

Les 25 premières publications de 2008 accessibles gratuitement répondant à la recherche Pubmed "*transcriptomic conditions*" ont été analysées par lecture rapide. Parmi elles, 4 [57, 206, 221, 311] analysent des séries temporelles d'au moins 5 points, une [194] ne comportait que 13 gènes analysés par PCR, 2 [202, 333] portaient sur une méthode d'analyse théorique et une [246] concernait l'analyse de données en protéomique. Ces publications n'entrent pas dans notre problématique d'analyse de données transcriptomiques haut débit obtenues entre différentes conditions. Il reste au final 17 publications répondant à notre problématique [33, 50, 61, 79, 90, 99, 106, 143, 132, 165, 214, 236, 169, 301, 302, 304, 349].

Dans l'ensemble des 17 publications restantes K. Plaimas *et al.*, [169] croisent leurs données avec un réseau métabolique tandis que J. M. Hevel *et al.*, [143] croisent leurs données avec un réseau issu Ingenuity.

Dans le reste des publications, l'analyse des données transcriptomiques suit la démarche suivante :

- Identification de gènes différentiellement exprimés entre les conditions (15 publications [33, 50, 61, 90, 79, 99, 106, 132, 165, 214, 236, 301, 302, 304, 349])
- Regroupement des gènes coexprimés par classification non supervisée (6 publications [99, 61, 214, 90, 165, 50]) ou supervisée par une condition expérimentale [304].
- Analyse des groupes de gènes coexprimés à l'aide d'ontologies telles que Gene Ontology (2 publications [99, 165]), ou analyse directe d'un sous ensemble de gènes à l'aide de la bibliographie (ex [106]) ou d'ontologie telles que MIPS (2 publications [214, 302]) et NCBI/COG¹ [132].

Une fois cette démarche suivie, des expériences complémentaires sont éventuellement réalisées afin de vérifier empiriquement les conclusions de l'analyse.

Dans la suite de ce chapitre, nous décrirons brièvement ces méthodes d'analyse.

¹Clusters of Orthologous Groups of proteins <http://www.ncbi.nlm.nih.gov/COG/>

1.2 Identification de gènes différentiellement exprimés entre les conditions expérimentales

L'identification de gènes différentiellement exprimés se base essentiellement sur deux critères [69] : une variation d'expression suffisante (par exemple on ne sélectionnera pas les gènes, qui ne varient pas plus de deux fois entre deux conditions) et une variation significative, c'est-à-dire significativement plus grande que la variabilité expérimentale et la variabilité inter individuelle. L'identification d'un grand nombre de gènes différentiels pose des problèmes de tests multiples [286] résolus par exemple par contrôle du taux de faux positifs [21].

1.3 Identification de groupes de gènes coexprimés

Les groupes de gènes coexprimés sont mis en évidence par classification non supervisée. Les publications analysées utilisent principalement deux méthodes de classifications : la classification non supervisée hiérarchique ascendante [99, 214] et la méthode des centroïdes mobiles (ou *K-means*) [165, 90, 50].

La construction de groupes de gènes coexprimés (*clusters*) nécessite de définir deux critères : une distance entre les gènes basée sur leur profil d'expression et une méthode d'agrégation afin de regrouper les gènes proches. Dans les papiers analysés, la distance utilisée est $1 - \text{corrélation}$. Cette distance regroupe les gènes corrélés positivement entre eux et éloigne les gènes corrélés négativement. La méthode d'agrégation permet de définir à quel groupe relier un élément. Pour la classification hiérarchique ascendante, les publications analysées utilisent différents critères d'agrégation comme le lien moyen [99] ou le voisin le plus éloigné [304]. Dans *K-means*, un nombre défini *a priori* de centroïdes sont positionnés aléatoirement, les gènes sont agrégés avec le centroïde le plus proche, puis les centroïdes sont déplacés au barycentre des gènes qui leurs sont agrégés. Ce processus est répété un grand nombre de fois.

Au final que ce soit avec *K-means* ou la classification hiérarchique ascendante, on obtient des groupes de gènes coexprimés au travers des conditions expérimentales. Toute la difficulté consiste alors à interpréter ces groupes au regard de la problématique biologique de l'étude.

1.4 Interprétation biologique des groupes de gènes coexprimés

Une des méthodes couramment utilisée pour identifier des groupes de gènes d'intérêt est basée sur l'utilisation d'ontologies c'est-à-dire sur des bases de données qui associent à chaque gène un terme (ou éventuellement une hiérarchie de termes) (11 publications [61, 79, 99, 106, 132, 214, 236, 301, 302, 304, 349]). L'ontologie la plus couramment utilisée est Gene Ontology (5 publications [33, 50, 90, 99, 165, 304]). Gene Ontology regroupe 3 ontologies : la première décrit les compartiments cellulaires (cellular component) la seconde les mécanismes biologiques (biological process) et la troisième les fonctions des molécules (molecular function).

Dans 3 publications [106, 132, 302], les auteurs ont des *a priori* sur les processus biologiques à étudier (ex Respiration [106]). Ils recherchent donc le(s) terme(s) de l'ontologie d'intérêt correspondant(s) à ces processus ainsi que les gènes qui leurs sont associés. Ils interprètent ensuite ces termes en analysant la proportion de gènes sur (et sous) exprimés associés à chacun d'entre eux. Contrairement à la classification non supervisée, cette méthode nécessite de connaître *a priori* les groupes de gènes recherchés. Elle ne permet donc pas d'identifier de nouveaux groupes comme variant entre les conditions, mais seulement de décrire la variation des groupes déjà connus.

Lorsque des groupes de gènes coexprimés (ou *clusters*) sont identifiés par classification non supervisée [99, 165], leur analyse est effectuée grâce au calcul de l'enrichissement. L'enrichissement d'un terme dans un groupe de gènes est la proportion de gènes du groupe associés à ce terme par rapport à la proportion de gènes dans le jeu de données expérimentales complet associés à ce terme. Plus l'enrichissement d'un *cluster* en un terme donné est élevé et plus ce terme caractérise fonctionnellement le *cluster*. Des méthodes de calculs telles que le test de χ^2 , le test exact de Fisher ou le calcul de moyennes hypergéométriques permettent de tester si l'enrichissement est significativement supérieur à un enrichissement obtenu sous l'hypothèse où les

classes de l'ontologie sont distribuées au hasard [70]. Les résultats obtenus restent dépendants de l'ontologie choisie ainsi que de la méthode de calcul et des seuils utilisés [70].

Une fois les *clusters* caractérisés fonctionnellement par le ou les termes qui sont fortement représentés parmi leurs éléments, une étude bibliographique précise de leurs éléments est souvent réalisée (ex [99]).

L'identification d'éléments différentiellement exprimés entre conditions par une approche haut débit permet de filtrer les éléments qui varient entre les conditions sans avoir d'*a priori*, mais au final, seul un petit sous ensemble d'éléments est analysé précisément par étude manuelle de la bibliographie (détaillée en 2), ce qui nous éloigne d'une étude globale du système.

1.5 Conclusion

L'analyse des 17 articles a permis d'identifier une démarche couramment utilisée par les biologistes afin d'exploiter des données transcriptomiques. Ainsi, l'exploitation de données transcriptomiques haut débit débute par l'identification de gènes dont l'expression varie entre des conditions expérimentales, elle se poursuit par l'analyse de groupes de gènes coexprimés ou de groupes de gènes connus *a priori*. Ces groupes sont analysés à partir d'informations fonctionnelles issues de bases de données. Au final un petit sous ensemble des gènes est analysé manuellement par la reconstruction d'un sous système de relations à partir de la bibliographie.

Illustrons maintenant cette démarche au travers de l'étude approfondie d'un cas particulier que j'ai traité durant ma thèse. Cette étude nous permettra de tirer de nouveaux éléments concernant la démarche de biologiste pour analyser des données transcriptomiques.

Chapitre 2

Mise en oeuvre d'une démarche
d'analyse de données
transcriptomiques : étude de la
réponse hépatique au jeûne chez le
poulet

Research article

Open Access

Transcriptome profiling of the feeding-to-fasting transition in chicken liverColette Désert^{1,2}, Michel J Duclos³, Pierre Blavy^{1,2}, Frédéric Lecerf^{1,2}, François Moreews⁴, Christophe Klopp⁵, Marc Aubry⁶, Frédéric Herault^{1,2}, Pascale Le Roy^{1,2}, Cécile Berri³, Madeleine Douaire^{1,2}, Christian Diot^{1,2} and Sandrine Lagarrigue*^{1,2}

Address: ¹INRA, UMR 598, Génétique Animale, F-35000 Rennes, France, ²Agrocampus Ouest, UMR 598, Génétique Animale, F-35000 Rennes, France, ³INRA, UR83, Station de Recherches Avicoles, F-37380 Nouzilly, France, ⁴INRA, SIGENAE, F-35000 Rennes, France, ⁵INRA, SIGENAE, F-31000 Toulouse, France and ⁶Plateforme Transcriptome OUIST-genopole Rennes, F-35000 Rennes, France

Email: Colette Désert - colette.desert@agrocampus-rennes.fr; Michel J Duclos - michel.duclos@tours.inra.fr; Pierre Blavy - pierre.blavy@irisa.fr; Frédéric Lecerf - frederic.lecerf@agrocampus-rennes.fr; François Moreews - francois.moreews@irisa.fr; Christophe Klopp - christophe.klopp@toulouse.inra.fr; Marc Aubry - marc.aubry@univ-rennes1.fr; Frédéric Herault - frederic.herault@rennes.inra.fr; Pascale Le Roy - pascale.leroy@rennes.inra.fr; Cécile Berri - cecile.berri@tours.inra.fr; Madeleine Douaire - madeleine.douaire@rennes.inra.fr; Christian Diot - christian.diot@rennes.inra.fr; Sandrine Lagarrigue* - sandrine.lagarrigue@agrocampus-rennes.fr

* Corresponding author

Published: 17 December 2008

Received: 25 August 2008

BMC Genomics 2008, 9:611 doi:10.1186/1471-2164-9-611

Accepted: 17 December 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/611>

© 2008 Désert et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**Abstract**

Background: Starvation triggers a complex array of adaptative metabolic responses including energy-metabolic responses, a process which must imply tissue specific alterations in gene expression and in which the liver plays a central role. The present study aimed to describe the evolution of global gene expression profiles in liver of 4-week-old male chickens during a 48 h fasting period using a chicken 20 K oligoarray.

Results: A large number of genes were modulated by fasting (3532 genes with a pvalue corrected by Benjamini-Hochberg < 0.01); 2062 showed an amplitude of variation higher than +/- 40% among those, 1162 presented an human ortholog, allowing to collect functional information. Notably more genes were down-regulated than up-regulated, whatever the duration of fasting (16 h or 48 h). The number of genes differentially expressed after 48 h of fasting was 3.5-fold higher than after 16 h of fasting. Four clusters of co-expressed genes were identified by a hierarchical cluster analysis. Gene Ontology, KEGG and Ingenuity databases were then used to identify the metabolic processes associated to each cluster. After 16 h of fasting, genes involved in ketogenesis, gluconeogenesis and mitochondrial or peroxisomal fatty acid beta-oxidation, were up-regulated (cluster-1) whereas genes involved in fatty acid and cholesterol synthesis were down-regulated (cluster-2). For all genes tested, the microarray data was confirmed by quantitative RT-PCR. Most genes were altered by fasting as already reported in mammals. A notable exception was the *HMG-CoA synthase 1* gene, which was up-regulated following 16 and 48 h of fasting while the other genes involved in cholesterol metabolism were down-regulated as reported in mammalian studies. We further focused on genes not represented on the microarray and candidates for the regulation of the target

genes belonging to cluster-1 and -2 and involved in lipid metabolism. Data are provided concerning PPAR α , SREBP1, SREBP2, NR1H3 transcription factors and two desaturases (FADS1, FADS2).

Conclusion: This study evidences numerous genes altered by starvation in chickens and suggests a global repression of cellular activity in response to this stressor. The central role of lipid and acetyl-CoA metabolisms and its regulation at transcriptional level are confirmed in chicken liver in response to short-term fasting. Interesting expression modulations were observed for NR1H3, FADS1 and FADS2 genes. Further studies are needed to precise their role in the complex regulatory network controlling lipid metabolism.

Background

All animal species have evolved a metabolic response system allowing them to survive during periods of energy deprivation. The overall metabolic response to fasting operates at numerous levels and has been relatively well characterized in mammals [1-6]. In vertebrates, the liver plays a central role in this adaptive response. Deprivation of food inhibits lipogenesis and induces the release of large amounts of fatty acids from the adipose tissue, which are taken up by the liver and oxidized in the peroxisome and/or mitochondria via beta-oxidation. The majority of fatty acids are only partially oxidized to acetyl-coenzyme A (acetyl-CoA), which then condenses with itself to form ketone bodies, an important fuel for the brain. The energy released in the process of beta-oxidation is used by the liver to carry out gluconeogenesis from substrates such as glycerol, lactate, and amino acids.

Several studies showed that the inhibition of fatty acid synthesis and the induction of gluconeogenesis, ketogenesis and fatty acid beta-oxidation in response to fasting result from changes in mRNA level of genes encoding enzymes and transcription regulators involved in these metabolisms [6]. Several studies using PPAR α -null mice [7-9] have demonstrated a key role of PPAR α in this response. Although microarray-based experiments have been widely used to identify differentially expressed genes involved in numerous biological processes, only few studies have considered the effect of fasting on large-scale hepatic gene profiles. Data are available in the mouse [10,11], the pig [12] (but the study was limited to 1272 cDNA) and more recently the rainbow trout [13] and the rat [14]. In the present study, the chicken species was chosen as an important model organism that bridges the evolutionary gap between mammals and other vertebrates; a divergence that occurred about 300 million years ago. Previous studies have shown that the activities or expression of hepatic enzymes involved in lipogenesis, beta-oxidation and gluconeogenesis (ME, ACLY, ACACA [15]; CPT1A, EHHADH [16]; PEPCK [17]) and plasma metabolites and hormones levels (Glucose, lactate, pyruvate, aceto-acetate, B-hydroxybutyrate [17]; Insulin [18]) were altered during fasting in chickens. Overall the available data show that chickens share most metabolic responses

with mammals despite some distinct features. In birds lipogenesis occurs essentially in the liver [19-21] contrary to rodents or pigs in which it is regulated in both liver and adipose tissue. Regulation of gluconeogenesis differs too, essentially due to intracellular location of key enzymes [22]. Chicken plasma metabolites have different levels from those reported for mammals, especially glucose which is higher. To extend and complete previous studies conducted on candidate genes, the present study aimed to describe the evolution of global gene expression profiles in chicken liver during a 48 h fasting period by using pangenomic oligo microarrays. The second reason to choose chicken as model is that this species provides a major protein source from meat throughout the world and the excessive accumulation of lipids in birds is one of the main problems facing this industry. To our knowledge, no whole genome survey of hepatic gene expression has been reported in chickens so far [23]. The use of a recently available 20 K oligo microarray allowed us to make the first description of the alteration of hepatic gene profiles induced by fasting in this species and make some hypothesis on the regulatory mechanisms involved at the mRNA level. The genes showing a significant alteration in their expression profile during fasting were grouped in four clusters of co-expressed genes by a two-way Hierarchical Clustering Analysis. Further interpretation was based on the use of different gene annotation databases (GO, KEGG and Ingenuity databases) highlighting numerous biological processes modulated by fasting.

Results

Experimental and microarray setup

To evaluate the changes in hepatic gene expression in response to starvation, 4 week-old male chickens were submitted to 0, 16 h or 48 h of food withdrawal. Transcriptome profiling of these "Fed", "Fst16" and "Fst48" nutritional stages was carried out by using a 20 K microarray. The expression data of the 20460 gene set from 18 independent oligochips was normalized by "Lowess-fitness". Of the 20460 oligos present on the microarray, 13057 aligning with a unique coding region of the 2.1 Washington University assembly of the chicken sequence genome, were chosen for statistical analyses. The expression data were further analyzed by analysis of variance

with the eBayes procedure [24] using LIMMA package (see Methods) to identify differentially expressed genes between either of the two fasting states and the fed state. Clusters of co-expressed genes were identified by two-way Hierarchical Cluster Analysis and the degree of importance of each gene in this cluster identification was explored by Principal Component Analysis. Gene Ontology, KEGG and INGENUIITY databases were used for further interpretation, as 7419 genes (out of 13057) presented a human ortholog with a HUGO symbol allowing us to recover functional annotations from those databases.

The microarray data results were deposited in the Gene Expression Omnibus (GEO) public repository <http://www.ncbi.nlm.nih.gov/geo>. The accession number for the series is GSE11290 and the sample series can be retrieved with accession number GSM278021, GSM282110, and GSM284914 up to GSM284948. The sample series contains for each microarray the raw data (median signal) of each Cy5 and Cy3 channels as well as the normalized data (log2(ratio Cy5/Cy3))

A high number of genes differentially expressed between the two fasting states and the Fed state

A high number of gene transcripts were significantly altered by fasting, so that we considered only those presenting an amplitude of variation higher than +/- 40% between fasting and fed conditions (Figure 1A). The percentage of genes presenting such a modulation increased lower pvalues were chosen, for example for a cut-off fixed to 1.4, it increased from 58% with a pvalue of 0.01 to 95% with a pvalue of 0.0001. In the present study, the 2062 genes that satisfied the criterion of pvalue < 0.01 and absolute fold-change > 1.4 were considered for further analysis. Among these genes, the number of regulated genes was 3.5-fold higher after 48 h of fasting than after 16 h. We further performed a global analysis of the 1162 genes fulfilling those conditions and for which annotation could be retrieved through their human ortholog (Figure 1-B). The selected 0.01 pvalue that was corrected according to the false discovery rate (FDR) of Benjamini-Hochberg [25] ensures in average a number of false positive of 12 genes in this selected gene set. In this set, 190 genes were altered after 16 h of fasting, 777 after 48 h of

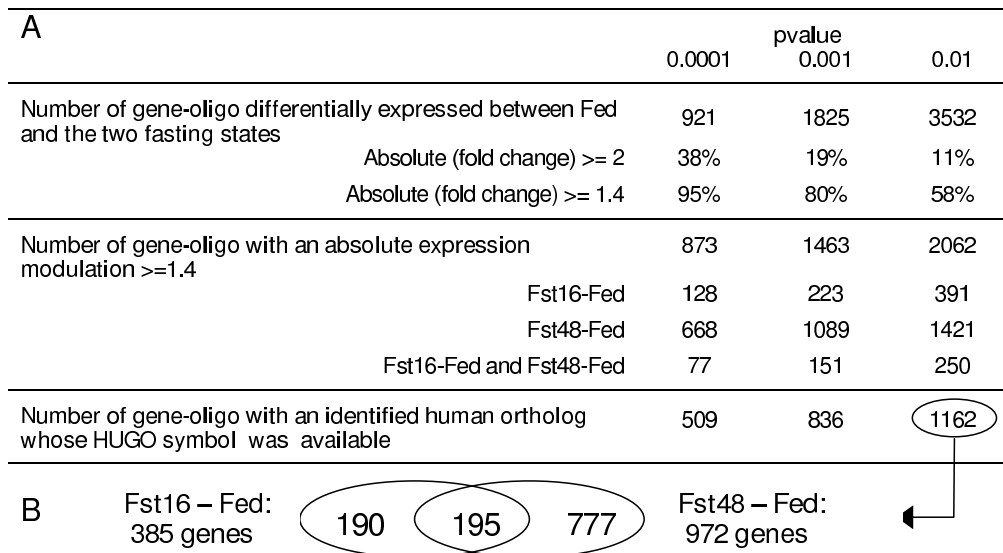


Figure 1
Genes differentially expressed between Fed and the two fasting Fst16 and Fst48 states. A: Number of genes differentially expressed according to the pvalues corrected by Benjamini-Hochberg. B: Venn diagram for the genes differentially expressed (pvalue < 0.01) in the two contrasts "16 h of fasting versus the fed state" (Fst16-Fed) and "48 h of fasting versus the fed state" (Fst48-Fed).

fasting, and 195 after 16 h and 48 h (with the concordant variations for 190 of them).

A two-way Hierarchical Cluster Analysis (HCA) was applied to the normalized dataset to identify clusters among the 18 animals and these 1162 genes (Figure 2). As expected, the animal dendrogram clearly separated the three nutritional conditions: the Fst48 state was more distant from the Fed state than the Fst16 state. The dendrogram also showed a higher homogeneity between individuals intra than inter nutritional conditions, showing that the number of chickens analyzed by condition was appropriate for such a study.

For the gene variables, four clusters were identified. The clusters 1 (277 genes) and 2 (95 genes) were respectively down- and up-regulated after 16 h of fasting. The majority of genes in cluster 1 (66%) were also significantly down-regulated after 48 h of fasting, but with an amplitude similar to that observed after 16 h. The majority of genes in cluster 2 (71%) returned to fed levels by 48 h of fasting.

The clusters 3 (517 genes) and 4 (273 genes) corresponded to the genes respectively down- and up-regulated only after 48 h of fasting. Altogether, the number of genes significantly down-regulated by fasting was higher than the number of genes up-regulated: 799 versus 360 genes. Only 3 genes (*TMEM43*, *AP3B2*, *C1orf59*) were regulated in an opposite way between the Fst16 and Fst48 states. The list of the 1162 significantly differentially expressed genes with the associated cluster is presented in an additional file 1.

In complement to the HCA, Principal Components Analysis (PCA) allowed to identify the genes contributing most to the separation of individuals. The first two components of PCA (explaining 70% of the total variability) clearly separated the three nutritional conditions (Figure 3-A) and about 80% of the genes showed a high correlation with the two principal components (> 0.7) (Figure 3-B). Further analysis of the whole dataset was then considered to identify the main biological processes regulated by fasting.

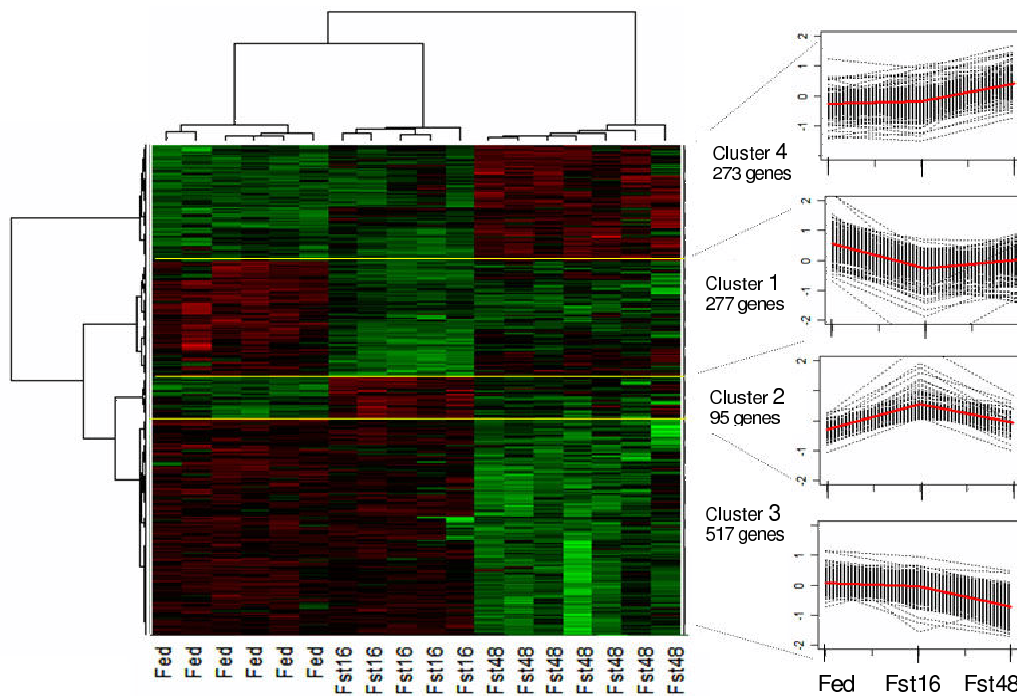


Figure 2
Two-way Hierarchical Cluster Analysis (HCA) of the gene selection. On the right, for each cluster, the curves of the expression of all the genes corresponding to the three nutritional states (the mean curve is in red).

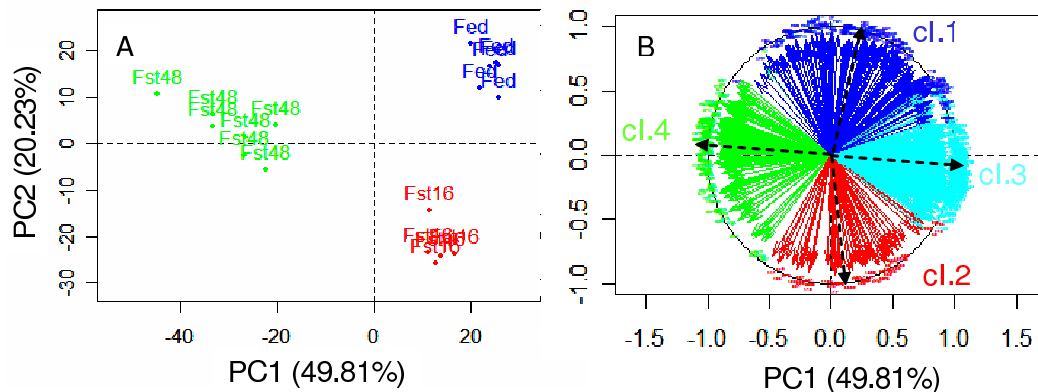


Figure 3
Principal Component Analysis (PCA) of the gene selection. The gene variables for the PCA were scaled to give them the same importance. A: individual factor map; B: gene factor map, only the genes highly correlated to the first two components are indicated ($r > 0.7$). The gene color was determined according to their HCA cluster (cl.i = cluster i).

Annotation of gene clusters using Gene Ontology (GO), KEGG and Ingenuity databases

Gene ontology (GO) terms have been widely used as a tool for a global interpretation of microarray data. Table 1 indicates the biological process GO terms enriched for each of the 4 clusters obtained by HCA (pvalue < 0.01). The most over-represented GO terms (high pvalue and at least 3 genes associated to the term) concerned "lipid biosynthesis", "fatty acid biosynthesis" corresponding to genes down-regulated after 16 h of fasting (cluster 1) and "fatty acid beta-oxidation" and "gluconeogenesis" corresponding to genes up-regulated after 16 h of fasting (cluster 2). These observations were consistent with those already reported in mammals.

Among the enriched GO terms of the cluster 3 (down-regulated after 48 h of fasting only) were "glycolysis", "regulation of protein catabolism" and surprisingly "cell cycle" with more than 30 genes associated. Of the six genes associated with "Glycolysis", *PKM2* encodes a pyruvate kinase that catalyses an irreversible reaction in glycolysis. *ATE1* and *MDM2*, two genes annotated as involved in the "regulation of protein catabolism" are part of the ubiquitin-proteasome system. No enriched GO terms was found for the cluster 4.

The biological interpretation of the gene clusters was further completed using KEGG database [26] and Ingenuity Pathway Analysis (IPA, Ingenuity Systems Inc., Redwood City, CA). The significant KEGG and IPA pathways are indicated in Table 1 with the genes associated. IPA and

KEgg pathway analyses highlighted some biological processes already revealed by GO analysis, plus other ones, showing the interest in combining different sources of annotation. Cluster 1, down-regulated at 16 h of fasting, was mainly characterized by genes involved in cholesterol metabolism and fatty acid biosynthesis, through the annotation terms "lipid metabolism", "lipid biosynthesis", "carboxylic acid biosynthesis", "fatty acid biosynthesis". Lipogenic genes (*ACACA*, *FASN*, *SCD*, *ACLY*), cholesterologenic genes (*CYP51A1*, *LSS*) and genes involved in triglyceride synthesis (*DGAT2L4*, *ANGPTL2*) were found in the list. Cluster 2, up-regulated at 16 h of fasting, was essentially characterized by genes involved in fatty acid metabolism/oxidation and acetyl-CoA metabolism through the annotation terms "fatty acid beta-oxidation", "energy derivation by oxidation of organic compounds", "gluconeogenesis", "PPAR signalling pathway", "synthesis and degradation of ketone bodies", "citrate cycle TCA cycle" and "pyruvate metabolism". The list comprised genes involved in fatty acid beta-oxidation (*PECL*, *ACAA1*, *ACOX1*, *CPT1A*, *HADHA*), ketogenesis (*HMGCL*, *ACAT1*), gluconeogenesis (*PCK1*, *FBP1*, *FBP2*) and fatty acid transport or activation (*ACSL1*, *APOB*, *FABP1*). Cluster 3, down-regulated at 48 h of fasting, was essentially characterized by genes involved in "glycolysis" (6 genes), "cell cycle" (more than 30 genes associated to this GO term) and 12 signalling pathways including "Insulin signaling pathway" and "Estrogen Receptor Signaling". Cluster 4, up-regulated at 48 h of fasting, was associated to various annotation terms, notably "cAMP-mediated Signaling" linked to glucagon/insulin balance.

Table 1: Annotation of gene clusters using Gene Ontology (GO), KEGG and Ingenuity databases

| Cluster 1 | Biological process GO terms ¹ | KEGG pathways ² | Ingenuity pathways ³ |
|-----------------------|---|--|---|
| 277 genes | <p>Lipid metabolism SLC27A4-LYPLA2-ENPP6-DGAT2L4-CYP51A1-FASN-SULT4A1-CPNE7-ANGPTL3-ACACA-LSS-ACLY-PBX1-PIGH-PLCLI-PRKAG3-PRKAA2-SCD-SBF2-PLA2G12B-MTMR3-PTPNM1 P = 5.E-04</p> <p>Lipid biosynthesis DGAT2L4-CYP51A1-FASN-ACACA-LSS-ACLY-PBX1-PIGH-PRKAG3-PRKAA2-SCD P = 2.E-03</p> <p>Fatty acid biosynthesis FASN-ACACA-PRKAG3-PRKAA2-SCD P = 6.E-03</p> <p>Regulation of action potential KCNB2-SBF2-EIF2B4 P = 7.E-03</p> | <p>Fatty acid biosynthesis ACLY-FASN-ACACA-SCD</p> | <p>Fatty Acid Biosynthesis FASN-ACACA-MCC2 P = 5.E-04</p> |
| Cluster 2 95 genes | <p>Lipid metabolism PECH-DCTN6-CPT1A-FABP1-ACSL1-ACAA1-HADHA-HMGCS1-APOB-ACOX1-SCP2-ADIPOR2-PLCZ1 P = 3.E-06</p> <p>Fatty acid metabolism PECH-CPT1A-FABP1-ACSL1-ACAA1-HADHA-ACOX1-ADIPOR2 P = 2.E-07</p> <p>Fatty acid oxidation CPT1A-HADHA-ACOX1-ADIPOR2 P = 4.E-06</p> <p>Fatty acid beta-oxidation CPT1A-HADHA-ACOX1 P = 2.E-05</p> <p>Energy derivation by oxidation of organic compounds FBP1-IDH1-PCK1-FBP2-GYG2 P = 8.E-03</p> <p>Gluconeogenesis FBP1-PCK1-FBP2 P = 4.E-04</p> | <p>Fatty acid metabolism ACAT1-PECHA-CAAI-ACOX1-HADHA-ACSL1</p> <p>PPAR signalling pathway PCK1-SCP2-FABP1-ACAA1-ACOX1-ACSL1</p> <p>Synthesis and degradation of ketone bodies ACAT1-HMGCL-ACAA1-HMGCS1</p> <p>Citrate cycle TCA cycle PCK1-ACAT1-HMGCL-IDH1-ACAA1</p> <p>Gluconeogenesis PCK1-ACAT1-HMGCL-FBP1</p> <p>Valine, leucine and isoleucine degradation ACAT1-HMGCL-ACAA1-HADHA-HMGCS1</p> <p>Lysine degradation ACAT1-HMGCL-ACAA1-HADHA-HMGCS1</p> <p>Lysine degradation ACAT1-HMGCL-ACAA1-HADHA-HMGCS1</p> | <p>Fatty Acid Metabolism CYP3A43-CYP4A22-CPT1A-ACAA1-CYP2C18-ACOX1-PECH-ACAT1-ACSL1-HADHA P = 2.E-11</p> <p>Synthesis and Degradation of Ketone Bodies I-06E01 6-42E-02 CYP3A43-CYP4A22-CPT1A-ACAA1-CYP2C18-ACOX1-PECH-CPT2-ACAT1-EHHADH-ACSL1-HADHA-ACAA1-ACAT1-HMGCL-HMGCS1-HADHA P = 2.E-11</p> <p>Pyruvate Metabolism ACAA1-ACAT1-PCK1-ACSL1-HADHA P = 2.E-05</p> <p>Valine, Leucine and Isoleucine Degradation ACAA1-ACAT1-HMGCL-HMGCS1-HADHA P = 9.E-08 6-94E00 6-54E-02 ACAA1-ACAT1-HMGCL-HMGCS2-EHHADH-HMGCS1-HADHA</p> <p>Lysine Degradation BBOX1-ACAA1-ACAT1-MLL3-HADHA P = 1.E-06</p> <p>Tryptophan Metabolism CYP3A43-CYP4A22-ACAA1-CYP2C18-ACAT1-HADHA P = 8.E-06</p> <p>Propanoate Metabolism ACAA1-ACAT1-ACSL1-HADHA P = 1.E-05 1-36E00 2-02E-02 EHHADH-HADHA 8-54E-01 1-69E-02 PCK1-HGFBP1</p> |

Table 1: Annotation of gene clusters using Gene Ontology (GO), KEGG and Ingenuity databases (Continued)

| | | | |
|---------------------------------------|---|---|--|
| <p>Cluster 3 517 genes</p> | <p>Cell cycle MPHOSPH9-TRIM13-TADA3L-YWHAQ-ESCO2-DBCI1-EJF6-ERN1-FGF6-SH3BP4-FRAPH1-GAS2-APBB2-KRAS-MAD2L1-MDM2-MLH1-MUTYH-NRAS-NDE1-PPP1CB-CDC37L1-PBK-PRK-G2-PTMS-BARD1-CLSPN-NEK4-YWHAH-MAD1L1-MCN8-CCNH-PRCI-HMTSS1-CUL7-DCLRE1A-CDC34 P = 9.E-03</p> <p>Glycolysis DLAT-ENO1-LOGDH-PGAM1-PKM2-JEVL D P = 8.E-03</p> <p>Cytokinesis CECR2-PRCI-ROCK2 P = 7.E-03</p> <p>Ceramide metabolism GALC-PRKAA1-SGPL1 P = 7.E-03</p> <p>Regulation of protein catabolism ATE1-MDM2-BARD1 P = 2.E-03</p> | <p>Insulin signalling pathway PTPN1-PRKAA1-KRAS-FRAPH1-PRKAB1-SORBS1-CRK-NRAS</p> <p>ErbB signalling pathway CAMK2A-JUN-KRAS-FRAPH1-PAK7-SRC-CRK-NRAS</p> <p>GnRH signalling pathway CAMK2A-ADCY3-JUN-KRAS-CACNA1C-SRC-NRAS</p> <p>Renal cell carcinoma JUN-SLC2A1-KRAS-PAK7-ARNT-CRK-NRAS-SELN3</p> <p>Tight junction JAM3-KRAS-CLDN16-SRC-MAGII-NRAS-CTB</p> <p>Glioma CAMK2A-KRAS-FRAPH1-MDM2-NRAS</p> | <p>Chemokine Signaling ROCK2-SRC-PLCB4-CAMK2A-NRAS-JUN-HMYL2-PPP1CB-KRAS P = 3.E-04</p> <p>Ephrin Receptor Signaling MAP3K14-SRC-NRAS-CREB3-SOS2-KRAS-CRK-ROCK2-AKT1-SORBS1-EFNBI-ACPI-ARPC4-PRKAA1-PAK7-MAK P = 6.E-04</p> <p>B Cell Receptor Signaling MAP3K14-FRAPH1-AKT1-CAMK2A-NRAS-JUN-MAP3K7-CREB3-SOS2-PI3KAP1-KRAS P = 3.E-03</p> <p>Estrogen Receptor Signaling SRC-PRKDC-CCNH-NRAS-SOS2-KRAS-GTF2A1-MED4-ESR2 P = 3.E-03</p> <p>PDGF Signaling SRC-NRAS-JUN-ACPI-SOS2-KRAS-CRK P = 4.E-03</p> <p>Wnt/fz-catenin Signaling SRC-AKT1-WF1-MAP3K7-MNT7B-EZD3-DKK2-MDM2 (includes EG-4193)-WNT15B-WNT2-SOX5 (includes EG-6660) P = 6.E-03</p> <p>Actin Cytoskeleton Signaling ABI2-TIAM1-NRAS-MYL2-ACTB-SOS2-PPP1CB-KRAS-CRK-FGF6-ROCK2-ARPC4-PRKAA1-PAK7-MAK P = 6.E-03</p> <p>JAK/Stat Signaling FRAPH1-AKT1-NRAS-SOS2-PTPN1-KRAS P = 8.E-03</p> <p>Hypoxia Signaling in the Cardiovascular System AKT1-JUN-COP55-CREB3-MDM2 (includes EG-4193)-CDC34-UBE2H-ARNT P = 6.E-04</p> |
| <p>Cluster 4 273 genes</p> | <p>TGF-beta signalling pathway ACVR2A-PPP2R2A-RBL1</p> <p>Complement and coagulation cascades F13A1-C8B-F8-PLAU</p> <p>Purine metabolism ADSL-ITPA-ATIC-POLR1B-POLR3G-POLR2C</p> <p>Gap junction CDC2-PDGFA-MAP2K2-GNAI1</p> <p>RNA polymerase POLR1B-POLR3G-POLR2C</p> <p>Histidine metabolism ADSL-GADI-ATIC</p> <p>Long term depression RARB-CASP9-STK4-MAP2K2</p> <p>Long term potentiation GRIN2B-GRIN2A-MAP2K2</p> | <p>cAMP-mediated Signaling CNGA4-MAP2K2-GNAO1-GNAI1-HTR1A-RGS12-CNGA1-P = 1.E-03 1-69°00 6-67°02 HIDAC4-TGFB3-RBL1-SIN3A</p> <p>Ephrin Receptor Signaling GRIN2B-GRIN2A-MAP2K2-PDGFA-CXCL12-GNAO1-ITGA2-GNAI1-ARPC3-CDC2-ACVR2A P = 1.E-03</p> <p>1-26°00 3-64°02 SOX9-GNAO1-RARB-TGFB3-SFRP5-ACVR2A</p> | <p>Insulin signalling pathway PTPN1-PRKAA1-KRAS-FRAPH1-PRKAB1-SORBS1-CRK-NRAS</p> <p>ErbB signalling pathway CAMK2A-JUN-KRAS-FRAPH1-PAK7-SRC-CRK-NRAS</p> <p>GnRH signalling pathway CAMK2A-ADCY3-JUN-KRAS-CACNA1C-SRC-NRAS</p> <p>Renal cell carcinoma JUN-SLC2A1-KRAS-PAK7-ARNT-CRK-NRAS-SELN3</p> <p>Tight junction JAM3-KRAS-CLDN16-SRC-MAGII-NRAS-CTB</p> <p>Glioma CAMK2A-KRAS-FRAPH1-MDM2-NRAS</p> |

The pathways sub-lined were found in at least two of the three analyses. Biological process GO terms obtained by the Gene Ontology Tree Machine software (GOTM). Are only indicated the enriched biological process GO terms with a significant level of pvalue < 0.01 (see Methods) and a minimum of 3 genes associated. 2 Kegg pathways: are only indicated those with a minimum of 3 genes associated and having a probability to be observed in the cluster 4-fold superior than the probability to obtain it by chance. 3 Ingenuity pathways: are indicated the top five Canonical pathways associated to each cluster (pvalue < 0.01). Only canonical pathways with at least 3 genes affiliated were conserved.

Following this global overview of the data, we chose to focus on genes regulated by the first 16 hours of fasting (clusters 1 and 2) that were involved in "lipid biosynthesis", "fatty acid beta-oxidation", "ketogenesis" and "gluconeogenesis". Several genes were chosen for validation of the microarray results using quantitative RT-PCR. A further group of genes encoding FADS1 and FADS2 desaturases and transcriptional factors known to be important regulators of hepatic lipid metabolism were also considered.

Genes involved in fatty acid- and cholesterol-synthesis and in fatty acid beta-oxidation, ketogenesis and gluconeogenesis

Validation by qRT-PCR

Twenty four genes involved in peroxisomal or mitochondrial beta-oxidation of fatty acids, in ketogenesis, in gluconeogenesis and in fatty acid and cholesterol synthesis were present on the microarrays, respectively 10, 3, 3, 4 and 4 genes. As indicated in Figure 4-A, after 16 h of fasting, all lipogenic genes and cholesterologenic genes were significantly down-regulated, except *HMGCR* not regulated and *HMGCS1* up-regulated. By contrast, all genes involved in ketogenesis, gluconeogenesis and mitochondrial or peroxisomal fatty acid beta-oxidation were significantly up-regulated, except *EHHADH* and *HMGCS2*. The highest amplitudes of up-regulations were observed for *CPT1A*, *PECL*, *ACAA1*, *ACOX1*, *HMGCL*, *PCK1* (~2- up to 4-fold, $p < 10^{-8}$). The genes *ACLY*, *FASN*, *ME1* and *SCD* were down-regulated by 2-fold. To confirm these microarray data, 8 of the 21 genes differentially expressed after 16 h of fasting were analyzed by qRT-PCR (Figure 4-B). The results confirmed the microarray analyses, but higher amplitudes of variation were generally measured by qRT-PCR. Notably, the up-regulation of the *HMGCS1* gene during fasting was largely confirmed (3-fold by microarray and 10-fold by quantitative RT-PCR). The 3 genes *HMGCR*, *EHHADH* and *HMGCS2* not found significantly differentially expressed by microarray were analyzed by quantitative RT-PCR as well. *EHHADH* was up-regulated during fasting (16 h or 48 h) and *HMGCR* down-regulated, as described in mammals. By contrast, *HMGCS2* was found significantly up-regulated but only after 16 h of fasting. Overall these results are consistent with those reported in mammals except for the gene *HMGCS1*, which could show species specific patterns of regulation.

Expression of genes encoding SREBP-1, PPAR α , SREBP-2 and NR1H3 transcription factors

Because Peroxisome Proliferators-Activated Receptor alpha (PPAR α), Sterol Regulatory Element Binding Protein 1 and 2 (SREBP1 and SREBP-2) and Nuclear Receptor Subfamily 1, group H, member 3 (NR1H3 also noted LXRA for Liver X receptor alpha), have been identified in mammals as critical transcription factors for the regulation of hepatic fatty acid beta-oxidation, fatty acid synthe-

sis or cholesterol metabolism [6], we measured their expression by quantitative RT-PCR (Figure 4-B). *PPAR α* : We observed an up-regulation of *PPAR α* in chicken livers after fasting (3- and 2-fold at Fst16 and Fst48 states respectively). The correlations at the mRNA level between *PPAR α* and its putative targets in chickens (*ACOX1*, *EHHADH*, *CPT1*, *HMGCS2*) were significantly high: 0.82 to 0.86 using the qRT-PCR expressions of the 18 chickens in Fed, Fst16 and Fst48 states. Using the 2.1 Washington University assembly of the *Gallus gallus* genome sequence, the upstream sequences of those genes were screened for a putative PPAR α response element (PPRE). As indicated in Table 2, potential PPRE were identified in *ACOX1*, *CPT1* and *HMGCS2*. *SREBP1*: The first 16 h of fasting caused a significant repression of the hepatic expression of *SREBP1* (2- and 4-fold for Fst16 and Fst48 states respectively). The correlations between *SREBP1* and its putative targets in chickens (*ACLY*, *ACACA*, *FASN* and *SCD*) at the mRNA level were significantly high (0.87 to 0.92 using the 18 chickens in Fed, Fst16 and Fst48 states). *SREBP2*: The transcriptional factor *SREBP2* regulates the expression of many genes involved in cholesterol synthesis. We observed by qRT-PCR a decrease of *SREBP2* gene expression after fasting, (4- and 6-fold after 16 h and 48 h of fasting) correlated to *HMGCR* gene expression. Microarray results also indicated a 2-fold decreased expression of genes encoding *CYP51A1* (Lanosterol 14 α -demethylase) and *LSS* (Lanosterol synthase), other enzymes involved in cholesterol synthesis. By contrast, *HMGCS1* was strongly up-regulated during starvation. *NR1H3*: *NR1H3* mRNA level was significantly increased (by 2-fold) after 16 h of fasting and remained unchanged after 48 h whereas genes known to be its direct targets in mammals were down-regulated by fasting. It was the case for the lipogenic genes *SREBP1*, *FASN* and *SCD* and for *Cyp7a1* (cholesterol 7 hydroxylase) encoding the rate-limiting enzyme of cholesterol degradation to bile acids.

Expression of genes involved in desaturation of fatty acids

GO term, IPA and Kegg pathway analyses highlighted *SCD* which is considered as an enzyme involved in mono-unsaturated fatty acid synthesis. However, *SCD* (also named delta9-desaturase) belongs to the fatty acid desaturase family, which also includes *FADS1* (delta5-desaturase) and *FADS2* (delta6-desaturase), two key enzymes for the production of polyunsaturated fatty acids. We observed that fasting resulted in a 14- to 18-fold inhibition of *SCD* expression (measured by qRT-PCR at 16 h and 48 h of fasting respectively, Figure 4-B). A significant repression was observed for both *FADS1* and *FADS2* genes but only after 48 h of fasting (4- or 6-fold depending on the genes, Figure 4-B). The data show that the expression profiles of *SCD* gene and the two *FADS1* and *FADS2* genes were different in response to fasting in chickens.

Discussion

To date few studies have reported global gene expression surveys in chickens. Wang et al [27] provided an analysis of the chicken adipose tissue gene expression profile using a 28 k GeneChip Chicken Genome Array (Affymetrix Inc.). Other hepatic transcriptional analyses have been reported but using only dedicated chicken microarrays: 3.2 K liver-specific microarray [28,29] or a 323 cDNA microarray [30]. Although microarray studies are more numerous in mammalian species, only few studies have considered global gene expression of the liver tissue in response to fasting. The effect of fasting has been studied in the liver of mice fasted during 24 h or 48 h using a 20 K oligo microarray [10,11] and in the liver of rats fasted during 24 h with a 30 K oligo-chip [14]. One study was conducted in the liver of pigs fasted during 24 h which used a 1272 cDNA microarray [12], and more recently one was conducted in the liver of trouts submitted to 3 weeks of food deprivation which used a 16 K microarray [13]. In the present study, we analyzed the hepatic expression of 13057 genes in 4 week-old chickens, submitted to 16 h or 48 h of starvation. Of the 7419 annotated genes on the microarray (e.g. having a human ortholog), we observed 1162 genes with an amplitude of variation higher than +/- 40% at pvalue below 0.01 (corrected for multiple test). Because our gene selection satisfied the criterion of FDR < 1% (so only 12 genes were statistically false positives) our results could be considered as reliable; and indeed all the differential genes tested were confirmed by qRT-PCR. It remains, however, that the amplitude of the differences measured by qRT-PCR were generally higher than those measured using microarrays. This may explain why *EHHADH*, *HMGCR* and *HMGCS2* were observed as differentially expressed by qRT-PCR but not by microarray; and it may suggest that the microarray analyses based on hybridization technique is more hampered by proportion of false negative compared to qRT-PCR analyses because of the lack of an amplification step. The list of the 1162 genes differentially expressed in the chicken liver after starvation is available in additional file 1.

The main observation was that a high number of genes were regulated by fasting in chicken liver and in a time dependent manner: the expression of 190 genes was altered by 16 h of fasting versus that of 777 genes was altered by 48 h. In mouse liver, Bauer et al (2004) [10] made a similar observation (131 genes and 269 genes modulated by 24 h and 48 h of fasting respectively) even if the number of differential genes was lower compared to the present study. Another important observation was that the number of genes down-regulated by fasting was higher than the number of genes up-regulated, whatever the duration of fasting (16 h or 48 h): 799 genes versus 360 genes respectively. Among the 190 genes annotated as

enzymes that represented 16% of the genes annotated, 136 genes (72%) were down-regulated. Numerous reports have shown that birds are relatively resistant to long period of starvation [31-34]. This is the case for chickens which resist to longer period of starvation than mice [33,34]. Four week old chickens survive starvation periods of more than 10 days [35], losing about 37% of body weight during this period. This suggests that the severity of a 48 h fasting period is limited. However, our observations suggest a global repression of the cellular activity in response to short-term starvation (16-48 h).

Principal components analysis clearly separated the three nutritional conditions; the most (80%) of the differentially expressed genes were correlated with the two principal components and so contributed to distinguish the three states. By a two-way hierarchical clustering analysis, 4 clusters were identified: 277 genes were down-regulated and 95 genes up-regulated after 16 h of fasting, most of them showing a similar pattern at 48 h of fasting. A prolonged starvation up to 48 h repressed 517 and induced 273 gene expressions. We then used the three databases, Gene Ontology, KEGG and INGENUITY to analyze the 4 gene clusters. These databases supply a useful tool for a global biological interpretation of microarray data. We can however notice some contradictions which could be due to the imperfection of these annotations. First, a biological process could be highlighted through genes which are not key regulators of this process but only indirectly involved in it. As an illustration, the "Valine, leucine and isoleucine degradation" process was cited from KEGG and IPA databases in cluster 2 through genes involved in ketogenesis (Table 1) even though the first key enzyme of this pathway, BCAT1 (an aminotransferase) was absent of this cluster and belonged to cluster 1. Second, the allocation of a gene to a pathway may not always be optimal. For example, Gamma-butyrobetaine dioxygenase (BBOX1, cluster 2), was associated to the "Lysine degradation pathway" in KEGG and IPA databases while this enzyme catalyses carnitine biosynthesis and should therefore be associated to "fatty acid beta-oxidation" process in which carnitine is essential for mitochondrial beta-oxidation [36]. Another limitation may arise from the fact that some particular genes may not be annotated by automatic annotation processes. For instance, the Pyruvate dehydrogenase kinase 4 gene (*PDK4* up-regulated after 16 h of fasting,) was not cited in the "gluconeogenesis" pathway associated to the cluster 2, whereas *PDK4* phosphorylates and inhibits the activity of the pyruvate dehydrogenase complex and thus promotes the conversion of pyruvate into lactate, which is used in the liver for gluconeogenesis. Other genes important in the "lipid biosynthesis" process and associated to cluster 1 were also not highlighted: Insulin receptor Signaling gene (*IRS2*) or Insulin induced gene (*INSIG1*) which mediate feedback control of cholesterol

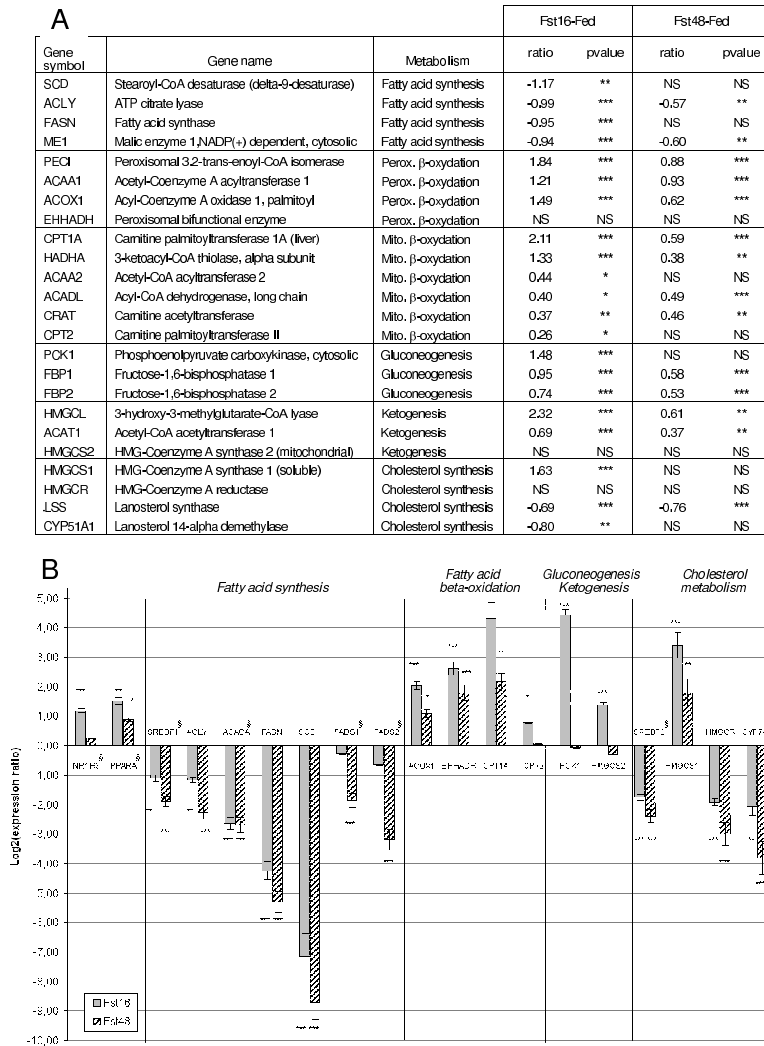


Figure 4
Identification of differentially expressed genes involved in fatty acid synthesis, beta-oxidation, ketogenesis, gluconeogenesis and cholesterol metabolism. A: Expression of genes present on microarray and analyzed by microarray procedure. Gene symbol: HGNC Hugo abbreviation of the human ortholog of the Gallus gallus gene represented by the oligo spotted on microarray. Results were expressed as a log₂ ratio of the gene expression between fasted and fed states ("Fst16-Fed" or "Fst48-Fed" contrasts). Genes in gold were analyzed by qRT-PCR (see B). P values were corrected by Benjamini-Hochberg (see Methods). * pvalue < 0.05, ** pvalue < 0.01, *** pvalue < 0.001, NS: non-significant. B: Validation of microarray results by qRT-PCR. PCRs were realized in triplicate with SyBr Green and specific primers for each gene (See Table 1). 18S ribosomal RNA was used as reference. Results were expressed as a log₂ ratio of the gene expression for the two "Fst16-Fed" and "Fst48-Fed" contrasts. Statistical significance is indicated as following: *pvalue < 0.05, **pvalue < 0.01, ***pvalue < 0.001. Genes not analyzed by microarray procedure are indicated with §

and fatty acid synthesis by controlling SREBP cleavage-activating protein (SCAP) involved in the regulation of cleavage of Sterol response element binding proteins (SREBP1 and SREBP2). Considering these imperfections, the annotation results require an expert biological knowledge to validate them and draw up precise comments and this is a limit to the full interpretation of pangenomic microarray data. However, the use of such databases, in particular when several are combined as in the present study, remains a first and useful step for a global biological interpretation as it allows highlighting the major pathways involved.

In the present study, the GO, KEGG and IPA annotation results taken together are complementary and allow us to draw up the following overview about the general biological mechanisms associated with each cluster. First, and as could be expected, they point out to an alteration of lipid and acetyl-CoA metabolisms during the first 16 h of fasting (clusters 1 and 2): lipogenic and cholesterogenic genes were down-regulated whereas genes involved in fatty acid beta-oxidation, ketogenesis, gluconeogenesis and fatty acid activation or transport were up-regulated. These results are in agreement with the previously reported alterations of glucose and lipid metabolisms in response to short-term starvation in rodents [10,14,37,38], pigs [12]

and chickens [15,19]. Second and less expected, they highlight a number of cell signalling pathways which would seem to be altered between 16 h and 48 h of fasting (clusters 3 and 4). Notably, genes in the "Insulin signalling cascade" would seem to be repressed whereas genes in the "cAMP-mediated signalling" pathway would seem to be enhanced. More than ten signalling response pathways were identified in cluster 3 (Table 1), suggesting a general repression of signalling pathways at 48 h of starvation. Many genes associated with these pathways are implicated in cellular responses, also involved in cell proliferation and differentiation. Precise comments about them would need an expert biological knowledge.

Because PPAR α , SREBP1, SREBP-2 and NR1H3 have been identified in mammals as critical transcription factors for the regulation of fatty acid beta-oxidation, fatty acid synthesis or cholesterol metabolism, we examined their expressions by qRT-PCR.

Definite proof that PPAR α plays a key role in the up-regulation of fatty acid beta-oxidation, ketogenesis and gluconeogenesis in the liver during starvation comes from several studies using PPAR α -null (KO) mice. Using a gene-by-gene approach, these studies identified PPAR α target genes [7-9,39]. Short-term fasting (12-72 h) in KO mice

Table 2: PPRE prediction in *Gallus gallus* HMGCS2, CPT1A and ACOX1 genomic DNA sequences

| Genes | Species | Ensembl ID | PPRE Pattern | Location | Strand | Methods |
|---------------|--------------------|------------------------------------|----------------------|---|--------|-------------|
| HMGCS2 | RNO | Rodriguez et al., 1994 [79] | GGGCCAaAGGTCT | Promoter region | | |
| | GGA | ENSGALG00000002960 | GGGCCAaAGGTCC | -530 | + | STAN/PATSER |
| | HSA | ENSG00000134240 | GGGTCAaAGGGCT | -118 | + | STAN |
| | MUS | ENSMUSG000000027875 | GGGCCAaAGGGAT | -104 | RC | STAN |
| CPT1A | HSA/MUS/RNO | Napal et al., 2005 [80] | AGGGGAaAGGTCA | Intronic region between exon 1 and exon 2 | | |
| | GGA | ENSGALG00000007077 | AGGGAAaAGGGCA | 4964 | + | STAN |
| | HSA | ENSG00000110090 | AGGGGAaAGGTCA | 2426 | + | STAN |
| | MUS | ENSMUSG000000024900 | AGGGGAaAGGTTA | 2180 | RC | STAN |
| ACOX1 | MUS | Tugwood et al., 1992 [81] | AGGACAaAGGTCA | Promoter region | | |
| | GGA | ENSGALG000000002159 | - | - | - | |
| | HSA | ENSG00000161533 | - | - | - | |
| | MUS | ENSMUSG000000020777 | - | - | - | |
| | RNO | Krey et al., 1995 [82] | AGGTAGaAGGTCA | Promoter region | | |
| | GGA | ENSGALG000000002159 | AGGAAGaAAGTCA | -3931 | + | STAN/PATSER |
| | HSA | ENSG00000161533 | - | - | - | STAN |
| | MUS | ENSMUSG000000020777 | AGGTAAaAGGTCA | 52 | + | STAN |

PPRE patterns used for the research is indicated in bold with the reference associated. GGA: *Gallus gallus*; HSA: Homo sapiens; MUS: Mus musculus. Nucleic acids in red indicate conserved nucleic acids with the PPREs used for the research. Location: genomic location refers to the start of the Ensembl first exon (+1). Strand: "RC" indicates that the sequence is reverse complemented compared to the genomic sequence displayed at that location, otherwise a "+" is indicated, STAN or PASTER indicates that the PPRE has been identified either by STAN or PASTER softwares respectively (see Methods). For the *ACOX1* gene, despite previous results indicating PPRE in the Mus musculus promoting DNA sequence [81], we cannot find any PPRE with the STAN procedure in any species. We used another PPRE pattern [82] identified in the *Rattus norvegicus* DNA sequence. This pattern allowed us to identify potential PPRE in *Gallus gallus* and *Mus musculus* sequences. The absence of result in the Homo sapiens sequence was consistent with the literature [78].

resulted in hepatic steatosis and hypoketonemia. Our observations also suggest a critical role of PPAR α in the up-regulation of these metabolisms in chicken liver in response to fasting. First, we observed a marked induction of PPAR α expression after 16 h and 48 h of fasting, confirming earlier results by Cogburn et al [23] in the same species. Second, we identified potential PPRE in avian genes *ACOX1*, *CPT1* and *HMGCS2* involved respectively in peroxisomal and mitochondrial beta-oxidation and ketogenesis. Third we observed a significant correlation between PPAR α and *ACOX1*, *EHHADH*, *CPT1* mRNA levels. Fourth, within the 15 genes showing the highest level of induction following 16 h of fasting (see additional file 1: cluster 2), we found 8 genes involved in fatty acid beta-oxidation (*PECL*, *ACAA1*, *ACOX1*, *CPT1A*, *HADHA*), ketogenesis (*HMGCL*, *ACAT1*), or gluconeogenesis (*PCK1*), the majority being direct targets of PPAR α .

By contrast, hepatic expression of *SREBP1* (encoding a key transcription factor controlling the expression of lipogenic enzymes in mammals) decreased after the first 16 h of fasting. We also observed a significant correlation between *SREBP1* mRNA levels and those of their putative target genes *ACLY*, *ACACA*, *FASN* and *SCD*. These results suggest a key role of *SREBP1* in the regulation of fatty acid synthesis in chickens as already suggested by others studies carried out on the same species [40,41] and demonstrated in rodents [37,38,42] using 24 h of fasting.

SREBP2 regulates the expression of many genes involved in cholesterol synthesis. Cholesterol is an essential component of animal cell membranes, and its concentration is tightly controlled by a feedback system that operates at transcriptional and posttranscriptional levels [43]. In the present study, we observed a decrease of *SREBP-2*, *HMGCR*, *LSS* and *CYP51A1* expressions during food deprivation. *HMGCR* is considered as the rate limiting enzyme of cholesterol synthesis. These results suggest a decrease of cholesterol synthesis in response to starvation in chickens, as already reported in rodents after 24 h or 48 h of fasting [10,14,44]. However, *HMGCS1* gene another gene involved in cholesterol synthesis exhibited a strikingly different pattern of expression, with a marked induction at 16 h of fasting which is contrast with the repression of *HMGCS1* expression observed during starvation in mice and pigs [10,12]. Two distinct genes condense acetyl-CoA with acetoacetyl-CoA to form HMG-CoA in the cell, the first one *HMGCS2*, located in mitochondria is involved in ketogenesis, whereas the second one, *HMGCS1* controls cholesterol synthesis in the cytosol. Like in other species, these two genes have been characterized in avian species [45], allowing us to design gene specific oligonucleotides. Four isoforms of the *HMGCS1* enzyme have been described in chicken liver [46] while only one was characterized in rat liver. The specific role of the 4 isozymes and

the mechanism to generate them are not clear and must be clarified to better understand the species difference in *HMGCS1* expression.

NR1H3 (named also *LXR α*) is a transcription factor that belongs to the nuclear hormone receptor family and was first discovered as playing a critical role in cholesterol homeostasis and bile acid metabolism through the regulation of *Cyp7a1* (cholesterol 7 hydroxylase), the rate-limiting enzyme of cholesterol degradation to bile acids [47]. It is also a master lipogenic transcription factor, directly regulating *SREBP1* [48,49], *FASN* [50], *SCD* [51] and *ChREBP* genes [52]. We therefore expected its expression profile to parallel to those of its potential avian targets during the feeding-fasting transition. Curiously, we observed that *NR1H3* mRNA level was increased at 16 h of fasting and returned to fed levels by 48 h. This result is consistent with a previous study conducted in rodents: rats fasted for 24 h increased *LXR α* mRNA level by 3-fold [53]. The role of *NR1H3* during starvation remains to be elucidated.

We then studied the effects of fasting on the hepatic expressions of the three genes encoding the desaturases *SCD*, *FADS1* and *FADS2*. *SCD* (also named delta9-desaturase) catalyzes the synthesis of monounsaturated fatty acids (MUFAs) whereas *FADS1* (delta5-desaturase) and *FADS2* (delta6-desaturase) are two key enzymes for the synthesis of highly polyunsaturated fatty acids (HUFA). All three mammalian desaturases are induced at the mRNA levels by PPAR α [54-56] and *SREBP1* [57-59]. This dual activation is a specificity of the desaturases compared to the enzymes of beta-oxidation or lipogenesis which are exclusively regulated by either PPAR α or *SREBP1* (for review, see [60]). In rodents, *SCD* gene activity is completely repressed during fasting, like the activity of other lipogenic genes (*ACLY*, *ACACA*, *FASN*) and *SREBP1* [37,38,61]. We observed the same regulation for the chicken *SCD* gene: the first 16 h of fasting resulted in a 14-fold inhibition of its expression and lasted up to 48 h of fasting. Because of their common regulation by *SREBP1* and PPAR α , similar modulations could be expected for the three genes in response to fasting. By contrast, we observed that *FADS1* and *FADS2* genes were only significantly repressed after 48 h of fasting (4- or 6-fold). In pig liver, *SCD* and *FADS1* were down-regulated after 24 h of fasting [12]. To our knowledge, no data about the *FADS1* and *FADS2* expression modulation in response to fasting are available in rodents. This difference of modulation between *SCD* gene and the two genes *FADS1* and *FADS2* in response to short-fasting suggest an additional regulatory mechanism between *SCD1* and these two latter genes. Such a hypothesis was already done by Matsuzaka et al [58], from their data obtained during fasting-refeeding treatment in which no changes of *FADS1* and *FADS2*

expression contrary to the lipogenic genes was observed. So, these two enzymes which play crucial roles in the production of HUFA, might have an important role in fasted state because HUFA were reported to inhibit SREBP-1 activity by multiple mechanism [62-65] and to be ligands of PPAR α .

Conclusion

In the present study, we successfully used a chicken 20 K oligo microarrays to analyze the alteration of hepatic gene expression profile upon starvation. We identified 1162 genes differentially expressed between the 16 h or 48 h fasting states and the fed state. We provide a valuable and publicly available resource of genes profiles altered during the first 48 h of starvation in chicken liver. After 16 h of fasting we observed an up-regulation of genes involved in fatty acid oxidation, ketogenesis, gluconeogenesis and a down-regulation of genes involved in fatty acid and cholesterol synthesis, which is consistent with earlier results obtained in mammals. After 48 h of fasting, when the number of genes showing an altered expression was much higher (about 3.5-fold higher), the annotation data suggest a repression of genes involved in numerous signalling pathways. As a whole, we observed that more genes were down-regulated than up-regulated in response to starvation. The expression profiles of candidate genes encoding key transcription factors and enzymes involved in lipid metabolism but not present on the microarray were evaluated by qRT-PCR. The results were similar to those reported in mammals except for the gene HMGCS1, which was induced at 16 h of fasting in chicken liver and repressed in mouse and pig liver. Our data also suggest that the genes *SCD*, *FADS1* and *FADS2* encoding different desaturases are regulated differently during fasting and *NR1H3* is up-regulated at 16 h of fasting while one of its target, *SREBP1* is down-regulated. Further studies should be performed to precise their role in the complex regulation array of lipid metabolism.

Methods

Animals and experimental procedures

Male broiler chicks obtained from a commercial hatchery were bred at INRA, UR083, Recherches Avicoles, F-37380 Nouzilly in accordance with European Union guidelines for animal care and under the authorization 006621 delivered to M.J. Duclos by the French Ministry of Agriculture. All birds were reared up to 1 week of age in floor pens, then raised for 2 weeks in cages under "14 h light:10 h darkness" cycles. During this period, they were fed a balanced starter diet (12.12 MJ metabolized energy/kg containing 22% crude protein). At 3 weeks of age, the chickens were weighed and assigned to seven experimental groups (n = 10 per group), equalizing body weight and variance between groups. One week later, each group was submitted to one of the following treatments: fed ad libitum (Fed), fasted for 16 h or 48 h (Fst16 and Fst48,

respectively), and four other treatments not analyzed in this present study [66]. All birds were given free access to water at all time. Following sacrifice, liver were collected, quickly frozen into liquid nitrogen and stored at -80°C until molecular analysis.

RNA isolation

Total RNA was extracted with TRIzol[®] reagent (Invitrogen, Cergy Pontoise, France) according to the manufacturer's instructions. Quality and concentration of extracted RNA were assessed using a 2100 Bioanalyzer (Agilent Technologies, Massy, France).

Array slides and annotation

The chicken 20 K array was obtained from ARK-Genomics (Roslin institute-UK: <http://www.ark-genomics.org>). The array design has been published in the ArrayExpress repository with the accession A-MEXP-820 [67] and in Gene Expression Omnibus with the platform name GPL5480 [68].

Briefly, the DNA microarray was produced from 20,460 oligonucleotides (whose the size varies from 60 to 75 nucleotides) designed using the OligoArray 2.0 software against the chicken ENSEMBL transcripts. The transcripts were selected from the chicken genome draft available in december 2004 and extensive matching of the UMIST and DT40 full length EST's with the TIGR clusters. Because the 20 K oligonucleotide set was defined in 2004–2005 from heterogeneous data sources, we checked the quality of the previously designed oligos, comparing them with the chromosomes of the 2.1 Washington University assembly of the chicken sequence genome [69]. The comparison was made using NCBI Blast with a 75% similarity threshold over 50 base pairs. Then for each high scoring pair (HSP) we retrieved the transcripts corresponding to the location using the Ensembl API (version Ensembl 43). An oligonucleotide had to be in a unique gene (even if it was spanning 2 exons) to be selected for further analyses. The corresponding annotations were then retrieved from Ensembl using the blast HSP coordinates. As results, among the 20460 gene-oligonucleotides, 13057 were identified as aligning with a unique coding region in the chicken genome sequence. As we retrieved an Ensembl gene name and/or a GO biological process term for only 32% of the 13057 oligo sub-set, we decided to rely on human orthologs (according to the "one to one" criteria of ENSEMBL annotation) which could be identified for 81% of the 13057 oligonucleotides, allowing to retrieve HGNC-HUGO gene symbol for the majority of them (70% of 81% of 13057). The annotations obtained by a bioinformatics procedure developed by SIGENAE (INRA) are available on the web site: <http://www.sigenae.org>[69]. Finally, of the 13057 oligonucleotides, 7419 presented a validated HUGO gene symbol from which we could extract more GO terms from GOA human annotation

database using the two softwares "Hugo my Genes" and "GOret" developed by the Rennes transcriptome platform: <http://ouestgenopuces.univ-rennes1.fr>.

Microarray procedures

mRNA labelling and hybridization

Five µg of each RNA sample were reverse-transcribed and Cy5 fluorescent-labelled using the ChipShot™ Direct Labeling kit (Promega, Charbonnières, France). Each Cy5-labelled RNA was hybridized to the microarray with a same Cy3-labelled reference probe according to the Transcriptome-Biochips Platform of Genopole "Toulouse Midi-Pyrénées" (France) procedure. Briefly, experiments were carried out with an automatic hybridization chamber (Discovery from Ventana Medical System, Inc). Pre-hybridization was carried out with a freshly prepared solution of 1% BSA, 2× SSC, 0.1% SDS over 30 min at 42°C. After automatic washing according to manufacturer's instructions, the slides were hybridized for 8 h in 200 µl of ChipHybe™ buffer (Ventana Medical System, Inc) containing 10 µl of labelled cDNA purified. After hybridization, the slides were washed twice for 2 min in 2× SSC/0.1% (v/v) SDS, and for 2 min in 0.1× SSC. We finally obtained 9, 7 and 7 microarrays respectively for the three nutritional conditions Fed, Fst16 and Fst48. For each gene, the fluorescence ratio reflected the relative abundance of the mRNA of interest in each experimental sample compared with the same reference mRNA. The reference allowed thus to take into account an eventual "spot × array" interaction.

Data acquisition

Detection of the fluorescence signals was made with a laser scanner (GenePix 4000A from Axon Instrument, CA) keeping a constant PMT gain for each channel. The images were then analyzed with GenePixPro 4.0 software (Axon instruments, Inc., Union City, CA). Row data file for each array containing all measured values were stored in gene-pix files compatible with the LIMMA library of R-project statistical and Bioconductor environment [70] which was used for the normalization and the analysis of the data.

Filtering for data normalization

For the normalization step, data were filtered according to 3 criterions: i) the genepix flag criterion automatically performed by GenePixPro 4.0 [71], ii) the SNR (Signal to Noise Ratio) provided also by GenePixPro and which was set to 2, iii) a asymmetry criterion of the spot which was set to 20%. For all microarrays, the mean percentages of spots discarded for the genepix flag, SNR and asymmetry criterions were 3.3%, 10.4% and 3.7% respectively. It was 16% for the three criterion taken together, showing the good quality of all technical procedures from slide production to labelling and hybridization. However, 5 arrays (3 and 2 in Fed and Fst16 conditions respectively) which

presented a percentage of spots not conformed to the SNR criterion higher than 50% were excluded from subsequent analysis. These 5 microarrays were discarded among a set of 60 microarrays corresponding to a larger experimental design with 6 treatments, confirming a global good quality of the technical procedures. The homogeneity of the background was systematically checked on each microarray by the boxplot and imageplot procedures of the LIMMA package. We finally analyzed 18 microarrays: 6, 5 and 7 microarrays for Fed, Fst16 and Fst48 conditions respectively.

Data normalization

The ratio Cy5/Cy3 used was expressed as the Log2 of the ratio of median pixel intensity of the two red and green spots. Log2 median ratio values were then normalized on each individual array (ratio centered on zero) according to the hypothesis that the majority of gene expressions do not differ between two samples. The centering was performed by "Lowess fitness" [72] to take into account the intensity dependence of the fluorescence bias. Only the array spots conform to the filtering step were used for the normalization.

Data analysis

All data analyses were performed using the R project statistical and Bioconductor environment [70]. Analysis of variance using the "eBayes" procedure [24] and analysis of contrasts were performed with LIMMA library appropriated to bi-color genepix files. Two-way Hierarchical Cluster Analysis (HCA) was performed using hclust function with "1-cor" as distance and "ward" as aggregation criterion; "heatmap" function was used to generate images. Principal Component Analysis (PCA) was performed with FactoMiner library. In this study, we were interested in comparing the two Fst16 and Fst48 conditions with the Fed state. The significance of gene expression differences between the nutritional conditions was assayed by analysis of variance followed by an analysis of the two contrasts "Fst16-Fed" and "Fst48-Fed". Because of the high test number (13057) needing an appropriated control of the false positive rate, the pvalue of each gene for each contrast was corrected according to the false discovery rate (FDR) procedure of Benjamini-Hochberg [25]; the FDR is the expected proportion of erroneously rejected null hypotheses among the rejected ones. Gene expression difference was declared significant if its corrected pvalue was $p < 0.01$. Results were further filtered by retaining only the genes exhibiting at least an amplitude of variation higher than $\pm 40\%$ between two nutritional states of interest ($(\text{absolute}(\log_2(\text{ratio}))) > 0.485$).

Gene ontology, KEGG and IPA analyses

Gene ontology (GO) constitutes a controlled vocabulary of about 20,000 terms organized in three independent

Table 3: Selected qRT-PCR primer sequences and accession numbers

| Gene symbol | Gene name | Ensembl or Gene bank Accession number | Primer sequence (forward/reverse) |
|---------------|---|---------------------------------------|---|
| ACACA | Acetyl-Coenzyme A carboxylase alpha | ENSGALG00000005439 | GAGGAGGGAAGGGAATTAGGAA CCAAGTGGCGGGACTGTT |
| ACOX1 | Acyl-Coenzyme A oxidase I | ENSGALG00000002159 | TCATCCGGTCTCTGATTGTAGGA GCACTATAGCGGATGGCAATG |
| ALCY | ATP citrate lyase | ENSGALG00000003475 | GGCGTGAATGAACTGGCTAAC TAGTCTTGGCATAGTCATAGGTCTGTTG |
| CPT1A | Carnitine palmitoyltransferase IA | ENSGALG00000007077 | CCCTGAAAATGCTGCTTTCCCTA TGGTGCCTGCAGAAAGTTTG |
| CPT2 | Carnitine palmitoyltransferase II | ENSGALG00000010681 | CCTGAACGCCCGAGAAACCT CCCTTTTCAAACCTGATGAGCAAGT |
| CYP7A1 | Cytochrome P450, family 7, subfamily A, polypeptide I | ENSGALG00000015432 | TGATGACATGAAAAAGCAAAGA CCAAAAAGTAGCAGGAATGGTGT |
| EHHADH | Peroxisomal bifunctional enzyme | ENSGALG00000006680 | TCATAGAAAGGAGCGAGAAGC AGCAGGAACCCCAACCAGT |
| FADS1 | Fatty acid desaturase 1 | ENSGALG00000007127 | CAGCACCACGCGAAACC TCTACAGAGAGCTTCTTCCCAAAG |
| FADS2 | Fatty acid desaturase 2 | ENSGALG00000007178 | CCATGATCAAGCGCAGGTT ATGTATGTGATGAAATAGCGCATGTAG |
| FASN | Fatty acid synthase | ENSGALG00000002747 | TGAAGGACCTTATCGCATTGC GCATGGGAAGCATTGTGTTGT |
| HMGCS1 | Hydroxymethylglutaryl-CoA synthase, cytoplasmic | ENSGALG00000014862 | GCTGGTGTGTTGCTATGCT TGCTGTCCCCTCTTTTGC |
| HMGCS2 | Hydroxymethylglutaryl-CoA synthase, mitochondrial | ENSGALG00000002960 | GGTGGTGTGTTGGGACAT GGTAGCACTGGATGGAGAGG |
| HMGCR | 3-hydroxy-3-methylglutaryl-Coenzyme A reductase | ENSGALG00000014948 | CTGGGTTTGGTCTTGTGTTCA ATTCCGTCTCTGCTTGTTC |
| NR1H3 | Nuclear receptor subfamily 1, group H, member 3 (LXR α) | ENSGALG00000008202 | TCCCACTCAACTCAGCACAC CAGACTTCATTCCAGCATC |
| PCK1 | Phosphoenolpyruvate carboxykinase, cytosolic | ENSGALG00000007636 | CTGCTGGTGTGCCTCTTGTA TTCCCTGGCTGTCTTTCC |
| PPARA | Peroxisome proliferative activated receptor, alpha | ENSGALG000000022985 | AGCATCCAGTCCTTCATCCA AAAAACCCCTTACAACTTCACAA |
| SCD | Stearyl-CoA desaturase (delta-9-desaturase) | ENSGALG00000005739 | TTTGCAATCGGCCGTAT TGGTAGTTGTGGAACCTTCTCCTA |
| SREBF1 | Sterol regulatory element binding protein 1 | gb:AY029224 | GTCGGCGATCCTGAGGAA CTCTTCTGCACGGCCATCTT |
| SREBF2 | Sterol regulatory element binding protein 2 | ENSGALG00000011916 | GGCTGGCTTCTCCCCAT GTTTCATCCTAACCTTTGCATCAT |

Genes in bold were present on the microarray.

hierarchies for cellular components, molecular functions, and biological processes [73]. Gene ontology analyses of clusters identified by HCA were performed using the Gene Ontology Tree Machine (GOTM) software [74]. Hypergeometric test was used as the statistical method to select enriched biological process GO terms for each cluster compared to the GO terms of the annotated genes present on the microarray (7419 genes). The biological process GO terms were considered as enriched for a level of p value < 0.01 . The biological interpretation of the gene clusters were further completed by KEGG annotation [26] and Ingenuity Pathway Analysis, the latter software using various annotation data (IPA, Ingenuity Systems Inc., Redwood City, CA). Were only conserved Kegg pathways with a minimum of 3 genes associated and having a probability to be observed in the cluster 4-fold superior than the probability to obtain it by chance. For Ingenuity pathways, we only reported in the present study the top five canonical pathways having a p value < 0.01 and at least 3 genes affiliated.

Real time quantitative RT-PCR (qRT-PCR) assay

A set of 11 genes present on the microarray was chosen for confirmation by quantitative RT-PCR, 8 of which were significantly differentially expressed between the 16 h fasting state and the Fed state. In addition, 8 candidate genes not present on the microarray were also measured by qRT-PCR (Figure 4).

Reverse transcription (RT) was carried out using the high-capacity cDNA archive kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. Briefly, 200 μ L of each reaction mixture containing 20 μ L of 10 \times RT buffer, 8 μ L of 25X dNTPs, 20 μ L of 10X random primers, 10 μ L of MultiScribe Reverse Transcriptase (50 U/ μ L), and total RNA (10 μ g) was incubated for 10 min at 25°C followed by 2 h at 37°C. A 1/10 or 1/20 dilution, dependant on the gene, of each RT reaction was further used for real time quantitative PCR (qPCR). cDNA samples were mixed with 20 μ L Absolute SYBR Green Mix (Abgene, UK) and 300, 450 or 600 nM, according to the gene, of specific reverse and forward primers (Table 3). Reaction mixtures were incubated in an iCycler iQ Multi-colour Real-Time PCR Detector (Bio-Rad, Marné la Coquette, France) programmed to conduct one cycle (95°C for 15 min) and 40 cycles (95°C for 15 s and 59°C for 45 s). A melting curve program was then performed for each gene to check the presence of a unique product with specific melting temperature. For each sample and each gene, PCR runs were performed in triplicates. For each gene, serial PCR reactions constructed with 2-fold serial dilutions from a pool of the cDNA samples were systematically added on each microplate for the calibration curve and determination of the amplification rate (R) of the Taq polymerase. For all genes including 18S, the amplification

rates were in a range of 99% to 100% and could be considered as equal to 1. So, for a same sample, the gene expression level could be normalized relative to the 18S expression level as follows: Gene normalized CT = CT_{gene} - CT_{18S} = Δ Ct. The significance of expression differences between nutritional states were analyzed by analysis of variance and analysis of contrasts on the basis of the gene normalized CT values using AOV package of R environment. For each gene, the N-fold gene expression difference between two conditions (1 versus 2) was expressed as: Fold change = $2^{\exp(-\Delta\Delta Ct)}$ with $\Delta\Delta Ct = \text{mean}(\Delta Ct)_2 - \text{mean}(\Delta Ct)_1$, where $(\Delta Ct)_i$ are the mean of the gene normalized Ct of the different samples of the condition i.

PPRE prediction in the HMGCS2, CPT1A and ACOX1 genomic DNA sequences

Several custom PERL scripts were developed to automate the PPRE detection procedure. Gallus gallus (GGA) genomic DNA sequences (including 5000 bp upstream and 3000 bp downstream sequence from the start of the first exon) were extracted from the Ensembl website by the GGA Ensembl ID. The orthologous genomic sequences of Homo sapiens and Mus musculus were automatically extracted (1-to-1 ortholog type) using the Compara API [75]. The DNA sequences were analyzed by the STAN [76] and the PATSER software [77]. The PPRE patterns used for every gene were designed from previous results found in the literature [78-82]. For STAN PPRE detection, one degree of freedom were applied on each part of the DR1 (e.g., for HMGCS2, the PPRE pattern syntax was: "AGACCT":1, 1...1, "TGGCCC":1, see Nicolas et al. [76] for details on the syntax used in STAN). For PATSER detection, Positional Weight Matrix (PWM) were generated using CONSENSUS software [77] using PPRE sequences described by previous studies. Raw results obtained by both softwares were post-analyzed by another custom PERL script.

Abbreviations

PPAR α : Peroxisome Proliferators-Activated Receptor alpha; FADS1: Fatty Acid Desaturase 1 (delta5-desaturase); FADS2: Fatty Acid Desaturase 2 (delta6-desaturase); SCD: Stearoyl CoA Desaturase 1 (delta9-desaturase); SREBP1 and 2: Sterol Regulatory Element Binding Protein 1 and 2; NR1H3: Nuclear Receptor Subfamily 1, group H, member 3 (noted LXR α for Liver X receptor alpha); HMGCS1: 3-Hydroxy-3-MethylGlutaryl-CoA Synthase 1; ME: Malic Enzyme; ACLY: ATP Citrate Lyase; ACACA: Acetyl CoA Carboxylase; CPT1A: Carnitine Palmitoyl-Transferase 1A; EHHADH: Enoyl-CoA Hydratase/3-HydroxyAcyl CoA DeHydrogenase; PEPCK: Phosphoenolpyruvate CarboxyKinase; PPRE: PPAR Element; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; qRT-PCR: quantitative Reverse Transcription-

Polymerase Chain Reaction; HGNC: HUGO Gene Nomenclature Committee; FDR: False Discovery Rate.

Authors' contributions

CD and SL were responsible for generating the gene expression data and for interpretation and analysis of the results. CD and CB carried out the qRT-PCR experiments. PB, FH and SL performed the statistical analyses with R. FL performed the PPRE prediction. MJD designed and provided the animal experimental design, contributed advice on the analysis and critically reviewed the manuscript. FM, CK, MA and SL contributed to the annotations of the 20 K microarray. PLR and SL defined the microarray experimental design. CD and SL drafted the manuscript. MD, CDi, and CB reviewed the manuscript. SL proposed and supervised the research. All authors read and approved the final manuscript.

Additional material

Additional file 1

List of the 1162 genes differentially expressed in the chicken liver after 16 h or 48 h of starvation ordered according to their HCA cluster. These 1162 genes i) were significantly differentially expressed between fasting and fed states (p value < 0.01), ii) exhibited at least an expression difference between two conditions exceeding an absolute 1.4 fold ratio, iii) were annotated through their human ortholog. The selected 0.01 p value that was corrected according to the false discovery rate (FDR) of Benjamini-Hochberg [25] ensures in average a number of false positive of 12 genes in the selected gene set. For each cluster, genes were ordered by decreasing \log_2 (ratio) of the "Fst16-Fed" contrast for the clusters 1 and 2 and then of the "Fst48-Fed" contrast for the clusters 3 and 4. Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2164-9-611-S1.xls]

Acknowledgements

This research was supported by the French Ministry for Research through an ACI IMPBio program based on interdisciplinarity, by the Institut National de la Recherche Agronomique (INRA) and by the Direction Générale de l'Enseignement et de la Recherche of the Ministry of Agriculture. PB is a Ph.D. fellow supported by INRA. We acknowledge Lidwine Trouille and Véronique Leberre from the transcriptomic platform of Toulouse for performing the hybridizations, Richard Talbot and Alison Downing from Ark-Genomics for providing the microarrays and Cecile Duby and Estelle Godet for participating to RNA extraction or qRT-PCR.

References

- Seitz HJ, Muller MJ, Krone W, Tarnowski W: **Coordinate control of intermediary metabolism in rat liver by the insulin/glucagon ratio during starvation and after glucose refeeding. Regulatory significance of long-chain acyl-CoA and cyclic AMP.** *Arch Biochem Biophys* 1977, **183**(2):647-663.
- Christiansen RZ: **Regulation of palmitate metabolism by carnitine and glucagon in hepatocytes isolated from fasted and carbohydrate refeed rats.** *Biochim Biophys Acta* 1977, **488**(2):249-262.
- Parvin R, Pande SV: **Enhancement of mitochondrial carnitine and carnitine acylcarnitine translocase-mediated transport of fatty acids into liver mitochondria under ketogenic conditions.** *J Biol Chem* 1979, **254**(12):5423-5429.
- Bergseth S, Lund H, Poisson JP, Bremer J, Davis-Van Thienen W, Davis EJ: **Carnitine palmitoyltransferase: activation and inactivation in liver mitochondria from fed, fasted, hypo- and hyperthyroid rats.** *Biochim Biophys Acta* 1986, **876**(3):551-558.
- Berge G van den: **The role of the liver in metabolic homeostasis: implications for inborn errors of metabolism.** *J Inher Metab Dis* 1991, **14**(4):407-420.
- Desvergne B, Michalik L, Wahli W: **Transcriptional regulation of metabolism.** *Physiol Rev* 2006, **86**(2):465-514.
- Kroetz DL, Yook P, Costet P, Bianchi P, Pineau T: **Peroxisome proliferator-activated receptor alpha controls the hepatic CYP4A induction adaptive response to starvation and diabetes.** *J Biol Chem* 1998, **273**(47):31581-31589.
- Kersten S, Seydoux J, Peters JM, Gonzalez FJ, Desvergne B, Wahli W: **Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting.** *J Clin Invest* 1999, **103**(11):1489-1498.
- Leone TC, Weinheimer CJ, Kelly DP: **A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders.** *Proc Natl Acad Sci USA* 1999, **96**(13):7473-7478.
- Bauer M, Hamm AC, Bonaus M, Jacob A, Jaekel J, Schorle H, Pankratz MJ, Katzenberger JD: **Starvation response in mouse liver shows strong correlation with life-span-prolonging processes.** *Physiol Genomics* 2004, **17**(2):230-244.
- Rakhshandehroo M, Sanderson LM, Matilainen M, Stienstra R, Carlberg C, de Groot PJ, Muller M, Kersten S: **Comprehensive Analysis of PPARalpha-Dependent Regulation of Hepatic Lipid Metabolism by Expression Profiling.** *PPAR Res* 2007, **2007**:26839.
- Cheon Y, Nara TY, Band MR, Beever JE, Wallig MA, Nakamura MT: **Induction of overlapping genes by fasting and a peroxisome proliferator in pigs: evidence of functional PPARalpha in nonproliferating species.** *Am J Physiol Regul Integr Comp Physiol* 2005, **288**(6):R1525-1535.
- Salem M, Silverstein J, Rexroad CE 3rd, Yao J: **Effect of starvation on global gene expression and proteolysis in rainbow trout (*Oncorhynchus mykiss*).** *BMC Genomics* 2007, **8**:328.
- Nakai Y, Hashida H, Kadota K, Minami M, Shimizu K, Matsumoto I, Kato H, Abe K: **Up-regulation of genes related to the ubiquitin-proteasome system in the brown adipose tissue of 24-h fasted rats.** *Biosci Biotechnol Biochem* 2008, **72**(1):139-148.
- Calabotta DF, Cherry JA, Siegel PB, Jones DE: **Lipogenesis and lipolysis in fed and fasted chicks from high and low body weight lines.** *Poult Sci* 1985, **64**(4):700-704.
- Skiba-Cassy S, Collin A, Chartrin P, Medale F, Simon J, Duclos MJ, Tesseraud S: **Chicken liver and muscle carnitine palmitoyltransferase I: nutritional regulation of messengers.** *Comp Biochem Physiol B Biochem Mol Biol* 2007, **147**(2):278-287.
- Brady LJ, Romsos DR, Brady FS, Bergen VWG, Leveille GA: **The effects of fasting on body composition, glucose turnover, enzymes and metabolites in the chicken.** *J Nutr* 1978, **108**(4):648-657.
- Rosebrough RVW, McMurtry JP, Steele NC: **Protein and energy relations in the broiler chicken. Chronic or acute effects of alternating protein or intermittent feeding regimens on broiler lipid metabolism.** *Br J Nutr* 1989, **61**(2):223-233.
- Leveille GA, Romsos DR, Yeh Y, O'Hea EK: **Lipid biosynthesis in the chick. A consideration of site of synthesis, influence of diet and possible regulatory mechanisms.** *Poult Sci* 1975, **54**(4):1075-1093.
- Griffin HD, Guo K, Windsor D, Butterwith SC: **Adipose tissue lipogenesis and fat deposition in leaner broiler chickens.** *J Nutr* 1992, **122**(2):363-368.
- O'Hea EK, Leveille GA: **Lipogenesis in isolated adipose tissue of the domestic chick (*Gallus domesticus*).** *Comp Biochem Physiol* 1968, **26**(1):111-120.
- Hanson RVW, Reshef L: **Regulation of phosphoenolpyruvate carboxylkinase (GTP) gene expression.** *Annu Rev Biochem* 1997, **66**:581-611.
- Cogburn LA, Porter TE, Duclos MJ, Simon J, Burgess SC, Zhu JJ, Cheng HH, Dodgson JB, Burnside J: **Functional genomics of the chicken—a model organism.** *Poult Sci* 2007, **86**(10):2059-2094.

24. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
25. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Stat Soc B* 1995, **57**:289-300.
26. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006:D354-357.
27. Wang HB, Li H, Wang QG, Zhang XY, Wang SZ, Wang YX, Wang XP: **Profiling of chicken adipose tissue gene expression by genome array.** *BMC Genomics* 2007, **8**:193.
28. Cogburn LA, Wang X, Carre W, Reijo L, Aggrey SE, Duclos MJ, Simon J, Porter TE: **Functional genomics in chickens: development of integrated-systems microarrays for transcriptional profiling and discovery of regulatory pathways.** *Comp Funct Genom* 2004, **5**:253-261.
29. Duclos MJ, Wang X, Carre W, Reijo L, Simon J, Cogburn LA: **Nutritional regulation of global gene expression in chicken liver during fasting and re-feeding.** *Plant Animal Genom XII Conf* 2004; San Diego, CA 2004.
30. Bourneuf E, Herault F, Chicault C, Carre W, Assaf S, Monnier A, Motier S, Lagarrigue S, Douaire M, Mosser J, et al.: **Microarray analysis of differential gene expression in the liver of lean and fat chickens.** *Gene* 2006, **372**:162-170.
31. Hazelwood RL: **The intermediary metabolism of birds.** Volume 2. New York: Academic Press; 1972.
32. Webster AB: **Physiology and behavior of the hen during induced molt.** *Poult Sci* 2003, **82**(6):992-1002.
33. Bierer BV, Eleazer TH, Roebuck DE: **Effect of Feed and Water Deprivation on Chickens, Turkeys, and Laboratory Mammals.** *Poult Sci* 1965, **44**:768-773.
34. Cuendet GS, Loten EG, Cameron DP, Renold AE, Marliss EB: **Hormone-substrate responses to total fasting in lean and obese mice.** *Am J Physiol* 1975, **228**(1):276-283.
35. Bierer BV, Eleazer TH, Barnett BD: **The effect of feed and water deprivation on water and feed consumption, body weight and mortality in broiler chickens of various ages.** *Poult Sci* 1966, **45**:1045-1050.
36. van Vlies N, Ferdinandusse S, Turkenburg M, Wanders RJ, Vaz FM: **PPAR alpha-activation results in enhanced carnitine biosynthesis and OCTN2-mediated hepatic carnitine accumulation.** *Biochim Biophys Acta* 2007, **1767**(9):1134-1142.
37. Shimano H, Yahagi N, Amemiya-Kudo M, Hasty AH, Osuga J, Tamura Y, Shionoin F, Iizuka Y, Ohashi K, Harada K, et al.: **Sterol regulatory element-binding protein-1 as a key transcription factor for nutritional induction of lipogenic enzyme genes.** *J Biol Chem* 1999, **274**(50):35832-35839.
38. Horton JD, Bashmakov Y, Shimomura I, Shimano H: **Regulation of sterol regulatory element binding proteins in livers of fasted and refed mice.** *Proc Natl Acad Sci USA* 1998, **95**(11):5987-5992.
39. Hashimoto T, Cook VWS, Qi C, Yeldandi AV, Reddy JK, Rao MS: **Defect in peroxisome proliferator-activated receptor alpha-inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting.** *J Biol Chem* 2000, **275**(37):28918-28928.
40. Gondret F, Ferre P, Dugail I: **ADD-1/SREBP-1 is a major determinant of tissue differential lipogenic capacity in mammalian and avian species.** *J Lipid Res* 2001, **42**(1):106-113.
41. Assaf S, Hazard D, Pitel F, Morisson M, Alizadeh M, Gondret F, Diot C, Vignal A, Douaire M, Lagarrigue S: **Cloning of cDNA encoding the nuclear form of chicken sterol response element binding protein-2 (SREBP-2), chromosomal localization, and tissue expression of chicken SREBP-1 and -2 genes.** *Poult Sci* 2003, **82**(1):54-61.
42. Kim JB, Sanraf P, Wright M, Yao KM, Mueller E, Solanes G, Lowell BB, Spiegelman BM: **Nutritional and insulin regulation of fatty acid synthetase and leptin gene expression through ADD1/SREBP1.** *J Clin Invest* 1998, **101**(1):1-9.
43. Goldstein JL, DeBose-Boyd RA, Brown MS: **Protein sensors for membrane sterols.** *Cell* 2006, **124**(1):35-46.
44. Tomkins GM, Chaikoff IL: **Cholesterol synthesis by liver. I. Influence of fasting and of diet.** *J Biol Chem* 1952, **196**(2):569-573.
45. Kattar-Cooley PA, Wang HH, Mende-Mueller LM, Miziorko HM: **Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes encode the cholesterogenic and ketogenic isozymes.** *Arch Biochem Biophys* 1990, **283**(2):523-529.
46. Clinkenbeard KD, Sugiyama T, Reed WD, Lane MD: **Cytoplasmic 3-Hydroxy-3-methylglutaryl Coenzyme A Synthase from liver.** *J Biol Chem* 1975, **250**(8):3124-3135.
47. Chiang JY, Kimmel R, Stroup D: **Regulation of cholesterol 7alpha-hydroxylase gene (CYP7A1) transcription by the liver orphan receptor (LXRalpha).** *Gene* 2001, **262**(1-2):257-265.
48. Repa JJ, Liang G, Ou J, Bashmakov Y, Lobaccaro JM, Shimomura I, Shan B, Brown MS, Goldstein JL, Mangelsdorf DJ: **Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta.** *Genes Dev* 2000, **14**(22):2819-2830.
49. Yoshikawa T, Shimano H, Amemiya-Kudo M, Yahagi N, Hasty AH, Matsuzaka T, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, et al.: **Identification of liver X receptor-retinoid X receptor as an activator of the sterol regulatory element-binding protein 1c gene promoter.** *Mol Cell Biol* 2001, **21**(9):2991-3000.
50. Joseph SB, Laffitte BA, Patel PH, Watson MA, Matsukuma KE, Walczak R, Collins JL, Osborne TF, Tontonoz P: **Direct and indirect mechanisms for regulation of fatty acid synthase gene expression by liver X receptors.** *J Biol Chem* 2002, **277**(13):11019-11025.
51. Chu K, Miyazaki M, Man WC, Ntambi JM: **Stearoyl-coenzyme A desaturase 1 deficiency protects against hypertriglyceridemia and increases plasma high-density lipoprotein cholesterol induced by liver X receptor activation.** *Mol Cell Biol* 2006, **26**(18):6786-6798.
52. Cha JY, Repa JJ: **The liver X receptor (LXR) and hepatic lipogenesis. The carbohydrate-response element-binding protein is a target gene of LXR.** *J Biol Chem* 2007, **282**(1):743-751.
53. Tobin KA, Steineger HH, Alberti S, Spydevoold O, Auwerx J, Gustafsson JA, Nebb HI: **Cross-talk between fatty acid and cholesterol metabolism mediated by liver X receptor-alpha.** *Mol Endocrinol* 2000, **14**(5):741-752.
54. Tang C, Cho HP, Nakamura MT, Clarke SD: **Regulation of human delta-6 desaturase gene transcription: identification of a functional direct repeat-1 element.** *J Lipid Res* 2003, **44**(4):686-695.
55. Guillou H, Martin P, Jan S, D'Andrea S, Roulet A, Catheline D, Rioux V, Pineau T, LeGrand P: **Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice.** *Lipids* 2002, **37**(10):981-989.
56. Miller CW, Ntambi JM: **Peroxisome proliferators induce mouse liver stearoyl-CoA desaturase 1 gene expression.** *Proc Natl Acad Sci USA* 1996, **93**(18):9443-9448.
57. Nara TY, He VWS, Tang C, Clarke SD, Nakamura MT: **The E-box like sterol regulatory element mediates the suppression of human Delta-6 desaturase gene by highly unsaturated fatty acids.** *Biochem Biophys Res Commun* 2002, **296**(1):111-117.
58. Matsuzaka T, Shimano H, Yahagi N, Amemiya-Kudo M, Yoshikawa T, Hasty AH, Tamura Y, Osuga J, Okazaki H, Iizuka Y, et al.: **Dual regulation of mouse Delta(5)- and Delta(6)-desaturase gene expression by SREBP-1 and PPARalpha.** *J Lipid Res* 2002, **43**(1):107-114.
59. Tabor DE, Kim JB, Spiegelman BM, Edwards PA: **Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2.** *J Biol Chem* 1999, **274**(29):20603-20610.
60. Nakamura MT, Nara TY: **Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases.** *Annu Rev Nutr* 2004, **24**:345-376.
61. Liang G, Yang J, Horton JD, Hammer RE, Goldstein JL, Brown MS: **Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c.** *J Biol Chem* 2002, **277**(11):9520-9528.
62. Xu J, Teran-Garcia M, Park JH, Nakamura MT, Clarke SD: **Polyunsaturated fatty acids suppress hepatic sterol regulatory element-binding protein-1 expression by accelerating transcript decay.** *J Biol Chem* 2001, **276**(13):9800-9807.
63. Xu J, Nakamura MT, Cho HP, Clarke SD: **Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A mechanism for the coordinate**

- suppression of lipogenic genes by polyunsaturated fats. *J Biol Chem* 1999, **274**(33):23577-23583.**
64. Yahagi N, Shimano H, Hasty AH, Amemiya-Kudo M, Okazaki H, Tamura Y, Iizuka Y, Shionoiri F, Ohashi K, Osuga J, et al: **A crucial role of sterol regulatory element-binding protein-1 in the regulation of lipogenic gene expression by polyunsaturated fatty acids.** *J Biol Chem* 1999, **274**(50):35840-35844.
 65. Kim HJ, Takahashi M, Ezaki O: **Fish oil feeding decreases mature sterol regulatory element-binding protein 1 (SREBP-1) by down-regulation of SREBP-1c mRNA in mouse liver. A possible mechanism for down-regulation of lipogenic enzyme mRNAs.** *J Biol Chem* 1999, **274**(36):25892-25898.
 66. Guernec A, Chevalier B, Duclos MJ: **Nutrient supply enhances both IGF-I and MSTN mRNA levels in chicken skeletal muscle.** *Domest Anim Endocrinol* 2004, **26**(2):143-154.
 67. **ArrayExpress** [<http://www.ebi.ac.uk/microarray-as/aer/>]
 68. **Gene expression omnibus** [<http://www.ncbi.nlm.nih.gov/projects/geo/>]
 69. Klopp C, Moreews F, Aubry M, Lagarrigue S: **Contribution To The Annotations Of The Chicken 20 K Oligo Microarray Of ARK-Genomics.** *Plant & Animal Genomes XVI Conference: January 12-16 2008; San Diego, CA 2008.*
 70. **Bioconductor** [<http://www.bioconductor.org/>]
 71. **GenePix User's guide and tutorial** [http://microarray.okstate.edu/files/GenePix_Pro.pdf]
 72. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30**(4):e15.
 73. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**(1):25-29.
 74. Zhang B, Schmoyer D, Kirov S, Snoddy J: **GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies.** *BMC Bioinformatics* 2004, **5**:16.
 75. **Ensembl Compara API** [<http://www.ensembl.org/info/using/api/compara/index.html>]
 76. Nicolas J, Durand P, Ranchy G, Tempel S, Valin AS: **Suffix-tree analyser (STAN): looking for nucleotidic and peptidic patterns in chromosomes.** *Bioinformatics* 2005, **21**(24):4408-4410.
 77. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.** *Bioinformatics* 1999, **15**(7-8):563-577.
 78. Lemay DG, Hwang DH: **Genome-wide identification of peroxisome proliferator response elements using integrated computational genomics.** *J Lipid Res* 2006, **47**(7):1583-1587.
 79. Rodriguez JC, Gil-Gomez G, Hegardt FG, Haro D: **Peroxisome proliferator-activated receptor mediates induction of the mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase gene by fatty acids.** *J Biol Chem* 1994, **269**(29):18767-18772.
 80. Napal L, Marrero PF, Haro D: **An intronic peroxisome proliferator-activated receptor-binding sequence mediates fatty acid induction of the human carnitine palmitoyltransferase 1A.** *J Mol Biol* 2005, **354**(4):751-759.
 81. Tugwood JD, Issemann I, Anderson RG, Bundell KR, McPheat WL, Green S: **The mouse peroxisome proliferator activated receptor recognizes a response element in the 5' flanking sequence of the rat acyl CoA oxidase gene.** *Embo J* 1992, **11**(2):433-439.
 82. Krey G, Mahfoudi A, Wahli W: **Functional interactions of peroxisome proliferator-activated receptor, retinoid-X receptor, and Sp1 in the transcriptional regulation of the acyl-coenzyme-A oxidase promoter.** *Mol Endocrinol* 1995, **9**(2):219-231.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Parallèlement à l'analyse d'articles (c.f. paragraphe précédent), j'ai pu mettre en oeuvre, en collaboration avec des biologistes, une démarche courante d'analyse de données transcriptomiques (c.f. paragraphe précédent). Cette démarche s'appuie sur un jeu de données créé dans l'objectif de décrire l'évolution globale des profils d'expression des gènes dans le foie de poulets mâles de 4 semaines en réponse à deux temps de jeûne (16 et 48 heures). Ces travaux ont été publiés (c.f. 2).

2.1 Contexte scientifique

De nombreuses études [82] ont montré que l'inhibition de la synthèse des lipides et l'induction de la néoglucogenèse, de la synthèse de cétones et de la β -oxydation des acides gras en réponse au jeûne sont des conséquences des variations des quantités d'ARN messagers des gènes codants des enzymes clés de ces fonctions ainsi que des facteurs de transcription connus pour réguler ces fonctions.

Les expériences de puces sont couramment utilisées pour identifier des gènes (et donc des processus biologiques) dont l'expression varie entre des conditions mais peu d'entre elles ont considéré l'effet du jeûne sur un grand nombre de gènes du foie. Des données sont disponibles chez la souris [15, 260], le porc [51], la truite arc-en-ciel [268] et le rat [232].

Dans notre étude, le poulet a été choisi car cette espèce est un organisme modèle important qui fait le lien dans l'évolution entre les mammifères et les autres vertébrés (une divergence qui a eu lieu il y a 300 millions d'années). De plus, le poulet est une espèce d'intérêt agronomique en raison de la consommation de sa viande et de la production d'oeufs. L'accumulation excessive de lipides dans les carcasses de poulets (environ 15% du poids total d'un poulet) est un des problèmes principaux de la filière avicole. Le gras, majoritairement localisé dans la cavité abdominale (5%), n'est pas valorisé que ce soit lors de la commercialisation de poulets entiers ou de produits élaborés.

Chez le poulet, des expériences précédentes ont montré que l'activité ou l'expression d'enzymes hépatiques impliqués dans la lipogenèse (ME1, ACLY, ACC) [40], la β -oxydation (CPT1A, EHHADH) [297] et la néoglucogenèse (PEPCK [35]) ainsi que les quantités de métabolites du plasma (glucose, lactate, pyruvate, aceto-acetate, B-hydroxybutyrate [35]) et d'insuline [267] varient au cours du jeûne.

Au final, les données disponibles montrent que le poulet possède la plupart des réponses métaboliques des mammifères à quelques exceptions près. Chez les oiseaux la lipogenèse a lieu dans le foie [204, 124, 239] contrairement aux rongeurs et au porc chez qui la lipogenèse est régulée à la fois dans le foie et dans le tissu adipeux. La régulation de la néoglucogenèse est elle aussi différente chez le poulet, essentiellement en raison de la localisation intracellulaire de ses enzymes clés [135]. La quantité de glucose plasmatique est plus élevée chez le poulet que chez les autres mammifères, et les quantités de métabolites plasmatiques sont différentes.

Afin de compléter les études précédentes réalisées sur des gènes candidats, notre étude utilise des puces pangénomiques afin de décrire l'évolution globale de l'expression des gènes dans le foie de poulets durant 48 heures de jeûne. À notre connaissance, aucune étude concernant l'expression des gènes de foie de poulet n'a été publiée auparavant. Ces travaux décrivent pour la première fois la variation de l'expression des gènes du foie induite par le jeûne chez des poulets grâce à une puce contenant 20 461 gènes. Elle propose ensuite quelques hypothèses sur la régulation (au niveau des ARN messagers) des mécanismes impliqués dans la réponse au jeûne.

2.2 Démarche

La démarche utilisée est similaire à celle décrite en 1, pour plus de détail voir le matériel et méthodes décrit dans la publication de Désert *et al.*, présentée page 102.

2.2.1 Identification de gènes différentiellement exprimés entre les conditions expérimentales

La première étape consiste à sélectionner les gènes différentiellement exprimés entre l'état nourri et 16 heures ou 48 heures de jeûne. Le choix a été fait de ne retenir que les gènes différentiellement exprimés dans au moins un de ces deux contrastes dont la variation d'expression est supérieure à 1.4 et dont le taux de faux positif (contrôlé par la méthode de Benjamini et Hochberg [21]) est inférieur à 1%.

2.2.2 Identification de groupes de gènes coexprimés

La seconde étape est la réalisation de groupes de gènes coexprimés par classification hiérarchique ascendante, une méthode non supervisée. La distance utilisée ($1 - correlation$) regroupe les gènes corrélés positivement, tout en éloignant ceux qui sont corrélés négativement. Le critère d'agrégation est le critère de Ward.

2.2.3 Interprétation biologique des groupes de gènes coexprimés

La troisième étape est le croisement de ces groupes de gènes coexprimés avec les termes des ontologies KEGG, Gene Ontology et Ingenuity. L'orthologue¹ humain des gènes est recherché afin d'obtenir des annotations. Ensuite l'enrichissement des groupes de gènes coexprimés en termes des ontologies est calculé et sa significativité est évaluée grâce au test hypergéométrique.

2.2.4 Analyse manuelle d'un sous ensemble de gènes

La dernière étape est l'étude approfondie de gènes impliqués dans la synthèse des acides gras et du cholestérol et dans la dégradation des acides gras. Ces processus ont été identifiés comme régulés par le jeûne grâce à l'analyse haut débit précédente. Pour cela, l'expression de 24 gènes a été mesurée par RT-PCRq et une analyse de la bibliographie a été effectuée pour chacun d'entre eux.

2.3 Résultats

2.3.1 Identification des gènes différentiellement exprimés entre les conditions expérimentales

La recherche de gènes différentiellement exprimés (c.f. 2.2.1) a permis de mettre en évidence 2062 gènes sur les 20 461 présents sur la puce. Une majorité d'entre eux sont régulés négativement (799 sont régulés négativement, 360 positivement). Le nombre de gènes différentiellement exprimés au bout de 48 heures est 3.5 fois plus important qu'au bout de 16 heures.

2.3.2 Identification des groupes de gènes coexprimés

Parmi les 2062 gènes différentiellement exprimés, 1162 gènes possèdent un orthologue annoté. Ces 1162 gènes ont été regroupés en 4 groupes de gènes coexprimés. Ces groupes sont décrits dans la figure 2 de la publication (c.f. page 93). Brièvement le **cluster 1 (277 gènes)** contient les gènes dont l'expression a diminué après 16 heures de jeûne, diminution qui reste stable à 48 heures. Le **cluster 2 (95 gènes)** contient les gènes dont l'expression a augmenté à 16 heures de jeûne, avant de retourner à leur niveau basal à 48 heures de jeûne. Le **cluster 3 (517 gènes)** contient les gènes dont l'expression reste stable à 16 heures et a diminué à 48 heures. Le **cluster 4 (517 gènes)** contient les gènes dont l'expression reste stable après 16 heures de jeûne et a augmenté à 48 heures.

¹Deux séquences homologues de deux espèces différentes sont orthologues si elles descendent d'une séquence unique présente dans le dernier ancêtre commun aux deux espèces. On suppose ici que deux gènes orthologues ont la même fonction.

2.3.3 Interprétation biologique des groupes de gènes coexprimés

Les termes sur-représentés dans les clusters sont présentés dans le tableau 1 de la publication pages 95 et 96.

Le **cluster 1 (277 gènes)** est caractérisé par des gènes impliqués dans la synthèse d'acides gras et de cholestérol comme le montre son enrichissement en termes des bases de données KEGG, Gene Ontology et Ingenuity : “*lipid metabolism*”, “*lipid biosynthesis*”, “*carboxylic acid biosynthesis*”, “*fatty acid biosynthesis*”. Les différents termes de ce cluster sont associés aux gènes impliqués dans la lipogénèse (ACACA, FASN, SCD, ACLY), la synthèse de cholestérol (CYP51A1, LSS) et la synthèse de triglycérides (DGAT2L4, ANGPTL2).

Ces fonctions sont connues comme étant régulées négativement lors du jeûne [16, 52], ce qui est cohérent avec le comportement de ce *cluster* en réponse au jeûne.

Le **cluster 2 (95 gènes)** est caractérisé par des gènes impliqués dans l'oxydation, l'oxydation des acides gras et dans le métabolisme de l'acétyl-coenzyme A comme le montrent les termes : “*fatty acid beta-oxidation*”, “*energy derivation by oxidation of organic compounds*”, “*gluconeogenesis*”, “*PPAR signalling pathway*”, “*synthesis and degradation of ketone bodies*”, “*citrate cycle TCA cycle*” et “*pyruvate metabolism*”. Ce cluster est aussi enrichi en termes concernant le métabolisme de l'acide propionique et le catabolisme des acides aminés. Ces différents termes sont associés aux gènes impliqués dans la β -oxydation (PECI, ACAA1, ACOX1, CPT1A, HADHA), la cétogénèse (HMGCL, ACAT1), la néoglucogénèse (PCK1, FBP1, FBP2) et dans le transport des acides gras (ACOX3, APOB, FABP1). Ces fonctions sont liées à la consommation d'acides gras et à l'économie de glucose et ont dans cette expérience le même comportement que celui habituellement observé chez les mammifères au cours du jeûne [16, 52].

Le **cluster 3 (517 gènes)** est essentiellement caractérisé par des gènes impliqués dans la glycolyse (6 gènes), le cycle cellulaire (plus de 30 gènes sont associés au terme Gene Ontology “*Cell cycle*”) et dans 12 voies de signalisation dont celle de l'insuline et du récepteur aux oestrogènes.

Le **cluster 4 (517 gènes)** est enrichi en termes divers, dont la voie de signalisation de l'AMP cyclique, un second messager intervenant dans la régulation du métabolisme par l'insuline et le glucagon (deux hormones dont la quantité varie au cours du jeûne).

2.3.4 Analyse manuelle d'un sous ensemble de gènes

Analyse fine des désaturases et de leur régulation

Les désaturations sont des réactions biochimiques permettant de réaliser des transformations d'acides gras (figure 2.4.2). Brièvement la $\Delta 9$ désaturase (SCD ou D9D) catalyse la synthèse d'acides gras monoinsaturés, tandis que la $\Delta 5$ désaturase (FADS1 ou D5D) et la $\Delta 6$ désaturase (FADS2 ou D6D) sont des enzymes clés de la transformation des acides gras essentiels en acides gras fortement insaturés (comme le C22:6 ω 3 et le C20:4 ω 6)

Chez les mammifères, PPAR α [315, 127, 223] et SREBP1 [235, 217, 314] induisent la synthèse d'ARN messagers des trois désaturases. Cette double induction est spécifique aux désaturases [234] ; en effet, les enzymes de la β -oxydation sont régulés par PPAR α et pas par SREBP1 tandis que ceux de la synthèse des acides gras sont régulés par SREBP1 mais pas par PPAR α . Chez les rongeurs l'activité de SCD est complètement réprimée lors du jeûne comme d'autres enzymes de la synthèse d'acides gras (ACLY, ACACA, FASN) et leur régulateur : SREBP1 [288, 149, 205]. Dans le foie de porc, SCD et FADS1 sont régulées négativement au bout de 24 heures de jeûne [51].

La même régulation est observée pour le gène SCD chez le poulet : lors des 16 premières heures de jeûne, l'expression de ce gène est inhibée drastiquement ($\log_2(\frac{\text{jeun16h}}{\text{nouri}}) = 5$ en *microarray* et -7.1 en RT-PCRq) et cette inhibition se poursuit jusqu'à 48 heures de jeûne ($\log_2(\frac{\text{jeun48h}}{\text{nouri}})$ non significativement différent de 0 en *microarray* et égal à -8.7 en RT-PCRq).

Les trois désaturases sont régulées par les deux mêmes régulateurs : SREBP1 et PPAR α . Sous l'hypothèse que des éléments régulés par les mêmes régulateurs varient de la même manière, on s'attendrait à observer le même comportement pour FADS1 et FADS2 que pour SCD, soit une forte diminution lors des 16 premières

heures. Pourtant FADS1 et FADS2 sont significativement inhibées seulement au bout de 48 heures de jeûne et d'un facteur plus faible ($\log_2(\frac{\text{jeun48h}}{\text{nourri}}$ de -1.9 pour FADS1 et -3.1 pour FADS2 en RT-PCRq).

Cette différence de modulation entre SCD et les gènes FADS1 et FADS2 en réponse à un jeûne de moins de 16 heures suggère qu'il existe un mécanisme de régulation supplémentaire qui n'est pas le même pour SCD que pour FADS1 et FADS2. Une telle hypothèse a déjà été avancée par Matsuzaka *et al.*, [217] à partir de leur jeu de données obtenues lors d'un jeûne suivi d'une renutrition. Dans cette expérience, FADS1 et FADS2 restaient stables tandis que les gènes de la lipogenèse variaient. Dans la partie 2, la modélisation du métabolisme hépatique des acides lors du jeûne chez des souris sauvages et *knockout* pour PPAR α nous a conduit à doser les ARNm de FADS2 [30]. Ces derniers sont présents en plus grande quantité chez le sauvage que chez le mutant, ce qui confirme le rôle de PPAR α dans l'activation de FADS2, de plus ils diminuent fortement à 48 heures de jeûne chez le mutant (et à 72 heures chez le sauvage) ce qui laisse supposer qu'un activateur autre que PPAR α et SREBP1 maintient leur expression en début de jeûne.

De plus, l'expression de NR1H3 est observée comme augmentant après 16 heures de jeûne, or NR1H3 est un activateur de SREBP1 dont l'expression décroît. A ces régulations, il faut rajouter le rôles des acides gras fortement insaturés dont la production nécessite FADS1 et FADS2. Ces acides gras inhibent SREBP1 par divers mécanismes [342,179] et sont des ligands de PPAR α . Ces résultats suggèrent d'étudier plus précisément les rôles de ces gènes dans le métabolisme des lipides ainsi que leurs régulations en réponse au jeûne comme indiqué page 12 de la discussion de Desert *et al.*, (c.f. 2).

L'expression de HMGCS1 est différente de celle des autres gènes codants pour les enzymes de la synthèse du cholestérol

L'analyse manuelle des gènes a permis de mettre en évidence que l'HMG-CoA synthase 1 (HMGCS1), un gène responsable de la synthèse de cholestérol, ne se comporte pas comme les autres gènes de cette fonction. L'expression des gènes responsables de la synthèse de cholestérol diminue au cours du jeûne, alors que l'expression de HMGCS1 augmente à 16 et à 48 heures de jeûne, contrairement à ce qui a été observé chez la souris [16] et le porc [52] chez qui l'expression de ce gène diminue. Le rôle de HMGCS1 chez le poulet reste donc à préciser.

Discussion sur l'utilisation de termes fonctionnels pour l'interprétation biologique de clusters

Un cluster peut être composé de termes indirectement liés à sa fonction, sans que les régulateurs clefs soient présents Lors de l'analyse manuelle du cluster 2, on se rend compte qu'il est annoté par le terme "*Valine, leucine and isoleucine degradation*" alors que l'enzyme clef de cette voie : BCAT1 (une aminotransférase) est absent de ce *cluster* et appartient au *cluster* 1. Ainsi, un cluster peut être enrichi en termes décrivant un processus biologique au travers de gènes qui interviennent indirectement dans ce processus alors qu'ils ne constituent pas des régulateurs clefs de ce processus. Un tel biais nécessite une analyse fine et manuelle pour être détecté et ne pas conclure trop hâtivement.

Certains gènes clefs n'apparaissent pas dans les clusters pour lesquels ils interviennent L'*insulin receptor substrate 2* (IRS2) et l'*insulin induced gene* (INSIG1) sont des gènes importants dans la synthèse des lipides. INSIG1 joue un rôle dans le rétro contrôle de la synthèse du cholestérol et des acides gras en contrôlant SCAP, une protéine impliquée dans la régulation par clivage de SREBP1 et SREBP1. Bien que présent dans le *cluster* 1, ces gènes ne sont pas ressortis lors du calcul d'enrichissement des *clusters* alors qu'ils auraient dû apparaître dans "*lipid synthesis*" probablement car dans les bases de données KEGG et Gene Ontology ils ne sont pas encore associés à ce terme malgré une bibliographie abondante les concernant. Par conséquent, l'utilisation d'une ontologie permet de mettre en évidence des voies biologiques importantes, mais certaines de ces voies peuvent passer inaperçues dans cette méthode du fait d'une annotation partielle.

Par ailleurs, nous montrons que l'utilisation de plusieurs ontologies(ici KEGG, IPA et Gene Ontology) permet de mettre en évidence des voies biologiques différentes et constitue une première étape utile pour l'interprétation globale des données transcriptomiques.

2.4 Conclusion

La démarche suivie dans cet article est dans ses grandes lignes identique à celle dégagée lors de l'analyse des publications. Elle consiste à identifier des gènes variant entre les conditions expérimentales, à les regrouper sans *a priori* en fonction de leurs corrélations, puis à interpréter les groupes ainsi formés en s'appuyant sur des ontologies. La dernière étape est une analyse bas débit de certains gènes ou groupes de gènes.

L'analyse des clusters permet de dégager sans *a priori* les principales fonctions biologiques dont la régulation transcriptionnelle varie en fonction des conditions expérimentales. C'est une amélioration par rapport aux méthodes supervisées, car il est possible d'obtenir des *clusters* inattendus tout à fait pertinents. Dans cette expérience, les observations chez le poulet sont dans leur grande majorité cohérentes avec celles observées dans d'autres espèces, ce qui corrobore la validité de la méthode.

Lors de cette étape, l'utilisation de plusieurs ontologies (IPA, KEGG et Gene Ontology) a permis d'interpréter fonctionnellement chaque *cluster*, chacune proposant ses propres termes, ce qui illustre la diversité des informations présentes dans les différentes sources de données et l'intérêt d'en utiliser plusieurs.

L'analyse manuelle permet d'introduire dans le raisonnement un ensemble d'informations issues de la bibliographie qui n'étaient pas présentes lors de l'analyse des clusters. Dans cet exemple, ces informations portaient sur le rôle clef d'un élément dans une fonction (ici BCAT1 dans le catabolisme de la valine, la leucine et isoleucine) ou dans un métabolisme (ici IRS2 dans le métabolisme des lipides) ainsi que sur les mécanismes de régulations connus (ici le rôle de PPAR α et SREBP1, NR1H3 et des acides gras fortement insaturés dans la régulation des désaturases). De plus, l'analyse manuelle a permis d'identifier de nouveaux éléments (dosés par RT-PCRq) comme importants pour la compréhension de la régulation du métabolisme des lipides lors du jeûne. L'introduction de ces informations a permis une meilleure interprétation des clusters et de leurs éléments.

De plus, l'analyse manuelle a permis de mettre en évidence que certains gènes (ici HMGCS1) avaient un comportement différent des autres gènes du *cluster* auxquels ils appartiennent. Ainsi la granularité du cluster s'avère trop grossière pour décrire certains comportements.

Chapitre 3

Les limites des méthodes classiquement utilisées en analyse transcriptomique.

Les méthodes courantes pour analyser les données haut débit de transcriptomiques sont l'identification et l'interprétation de groupes de gènes coexprimés (*clusters*) et l'analyse approfondie de quelques gènes. Nous allons étudier les limites de ces approches puis proposer dans le chapitre II une démarche afin de faciliter l'automatisation du raisonnement afin de s'affranchir de certaines d'entre elles.

3.1 Les limites de la classification

3.1.1 Une corrélation peut avoir plusieurs explications

Lors de l'analyse des corrélations, des phénomènes différents peuvent conduire au même ensemble d'observable. Si deux éléments A et B sont corrélés : soit A régule B, soit B régule A, soit C (un élément qui n'est peut être pas mesuré) régule A et B, soit la corrélation est due au grand nombre de variables mesurées. Les *clusters* contiennent donc des éléments coexprimés regroupés pour diverses raisons. Il est ainsi possible d'observer dans un même cluster des éléments appartenant à des processus biologiquement dépendants ou biologiquement indépendants.

3.1.2 Les éléments ne sont pas tous annotés dans les ontologies

Une fois les clusters réalisés, l'utilisation d'enrichissement de termes issus d'ontologie est un premier pas afin de les analyser. Dans ce cas l'efficacité de l'analyse est directement dépendante du contenu des ontologies.

Tout d'abord les ontologies actuelles ne couvrent généralement pas l'ensemble des éléments présents sur la puce, mais elles sont en constante évolution. Ceci s'explique car on ne connaît pas la fonction de tous les gènes. De plus il n'existe pas toujours d'ontologies décrivant les espèces d'intérêt, ce qui nécessite d'utiliser des orthologues afin d'extrapoler une espèce proche à l'espèce d'intérêt.

Par exemple, dans une expérience de mise à jeun du poulet [81] seulement 1162 gènes sur 2062 gènes d'intérêt présentent un orthologue annoté dans HGNC. Le problème est d'autant plus important lorsque l'enrichissement d'un cluster en termes de l'ontologie est le fruit d'un faible nombre de ses éléments.

3.1.3 La qualité des annotations est cruciale

La qualité des annotations des gènes dans les ontologies est un élément déterminant de l'interprétation des *clusters*. Un procédé biologique peut apparaître comme représenté dans un *cluster* car il contient des

régulateurs indirects du procédé, comme présenté en 2.3.4.

Deuxièmement, certains gènes ne sont pas toujours classés dans le meilleur pathway possible, par exemple le gène BBOX1 du cluster 2 est annoté comme intervenant dans la dégradation de la lysine dans KEGG et IPA, alors que cet enzyme catalyse la synthèse de carnitine ¹ et devrait être associé à la bêta oxydation [81].

La dépendance vis-à-vis d'une source de données de qualité inconnue pose le problème de l'évaluation de la qualité de cette source, point que nous tenterons d'évaluer dans la partie suivante (c.f. II).

3.1.4 Une interprétation des clusters difficile

Les termes sont parfois généralistes

Lorsque l'ontologie est généraliste, les *clusters* peuvent être identifiés par des termes très vagues comme “*lipid metabolism*” [81], ce qui rend leur interprétation difficile. Dans des ontologies hiérarchisées telles que Gene Ontology ces termes représentent les éléments de l'arbre des termes proches de la racine et décrivent donc une multitude de phénomènes.

Le comportement d'un terme de l'ontologie n'est pas directement déductible du comportement de ces éléments

Les ontologies font généralement l'hypothèse qu'un gène annoté par un terme, a un effet positif sur ce terme. Pourtant extrapoler la variation des termes à celles du *cluster*, n'est pas directement possible sans faire des hypothèses supplémentaires sur la mécanique interne du *cluster* et sur sa dépendance vis-à-vis du reste.

3.1.5 Exemple

Lors d'une expérience de mise à jeun chez le poulet (c.f. 2.3), le terme “*lipid synthesis*” a été identifié comme bien représenté dans le *cluster* contenant les gènes dont l'expression au cours du jeûne est décroissante. La conclusion d'une synthèse décroissante dans le métabolisme des lipides nécessite de faire l'hypothèse que les gènes mesurés sont des activateurs de la synthèse de lipides (hypothèse garantie par Gene Ontology qui ne considère que les activations), et que cette voie n'est pas activée par un autre mécanisme (une hypothèse parcimonieuse raisonnable dans un contexte de jeûne). Sans ces hypothèses, on aurait très bien pu supposer avoir à faire à des inhibiteurs, ou à un mélange activateurs-inhibiteurs, ou bien à l'existence d'un régulateur majeur non mesuré.

3.1.6 Conclusion

Au final, le calcul de l'enrichissement des clusters en termes fonctionnels donne une information sur le domaine d'étude nécessaire pour les interpréter et sur la variation globale de l'expression des éléments des grandes fonctions biologiques. Bien que cette méthode fournisse un outil descriptif synthétisant la variation globale de la variation transcriptomique des fonctions biologiques, elle ne permet pas une interprétation fine des éléments présents dans les *clusters*.

Pour interpréter finement un *cluster*, et comprendre les implications des variations de ces éléments, une étude manuelle de la bibliographie est réalisée par le biologiste, éventuellement suivie par une nouvelle expérience afin de tester ses conclusions.

¹Une enzyme de la bêta oxydation

3.2 Les limites de l'étude bibliographique

3.2.1 Une bibliographie partielle et fastidieuse

La première limite de l'approche est due à l'acquisition manuelle de la bibliographie. En effet cette étape est longue, fastidieuse et ne peut être réalisée que sur un petit nombre d'éléments. Une autre approche consiste à se tourner vers des experts qui ont lu la totalité ou la quasi totalité de la bibliographie sur un sujet précis. En raison de la masse de connaissances disponibles, cette expertise est difficile, et se limite à des sujets extrêmement précis.

La connaissance partielle d'un système peut amener à produire des erreurs d'interprétation, ou à limiter la capacité de prédiction, comme illustré figure 3.2.1. ².

Bien qu'une connaissance totale du système soit hors de portée, plus la connaissance est exhaustive, meilleurs sont les résultats. Pour faire au mieux il faudrait donc exploiter un maximum de bibliographie, ce qui en raison de la quantité d'informations, ne peut se faire que par une approche automatique.

Un des points clefs de la partie suivante (c.f. II) va être de proposer une méthode basée sur l'exploitation haut débit des bases de données d'interactions dans le but d'automatiser la tâche d'extraction de la bibliographie et d'interprétation des données au vue de la bibliographie extraite.

3.2.2 Un raisonnement difficile

Un système vaste et complexe

La seconde limite est due aux capacités d'analyse de l'être humain. Les capacités mémorielles permettent de retenir qu'une faible quantité d'informations face à la masse disponible dans la littérature. De plus, il est extrêmement difficile de résoudre des systèmes complexes, c'est-à-dire les systèmes comportant un grand nombre d'interactions.

Un raisonnement simplifié

Cette limite justifie la démarche de simplification et d'étude partielle appliquée par le biologiste. Tout d'abord cette démarche a pour but de proposer une expérience, et n'a pas la prétention de servir de preuve. Dans ce cadre, étant donné que les ressources allouées à l'exploration sont limitées, mieux vaut un raisonnement pragmatique partiel ou comportant des *a priori* issus d'expertise qui peut tomber juste qu'un raisonnement exact qui nécessite des moyens colossaux ou qui n'aboutit jamais. Ces simplifications découlent généralement de l'application du principe de parcimonie et d'hypothèses de travail issues du contexte expérimental.

Un raisonnement implicite

Le raisonnement du biologiste est basé sur de nombreuses hypothèses qui ne sont pas toujours rédigées dans les publications [80], soit car elles sont triviales, soit car elles sont couramment admises dans le domaine d'étude, soit car il faut pour y accéder remonter une longue chaîne de publications.

L'absence de rédaction explicite de ces hypothèses n'est pas néfaste pour une analyse bas débit par des experts car si les observations viennent à infirmer une hypothèse, l'expert saura l'identifier et la remettre en cause. Lorsque l'hypothèse fautive appartient à un autre domaine d'étude que celui sur lequel on possède de l'expertise, sa recherche peut s'avérer être très fastidieuse. Lorsqu'on souhaite appliquer ce raisonnement sur des jeux de données haut débit, il faut poser les hypothèses de manière systématique pour pouvoir automatiser, et de manière explicite pour pouvoir remonter à la ou aux hypothèses en cause en cas d'incohérence entre prédictions et données observées.

²Ces erreurs sont dues à l'hypothèse implicite qu'il n'existe pas d'éléments, ni d'interactions ayant de l'influence en dehors des éléments représentés dans le système. C'est une hypothèse de travail parcimonieuse tout à fait censée dans le cadre d'une démarche empirique. Cette hypothèse est amenée à être contredite lors de la confrontation entre le système connu et les données.

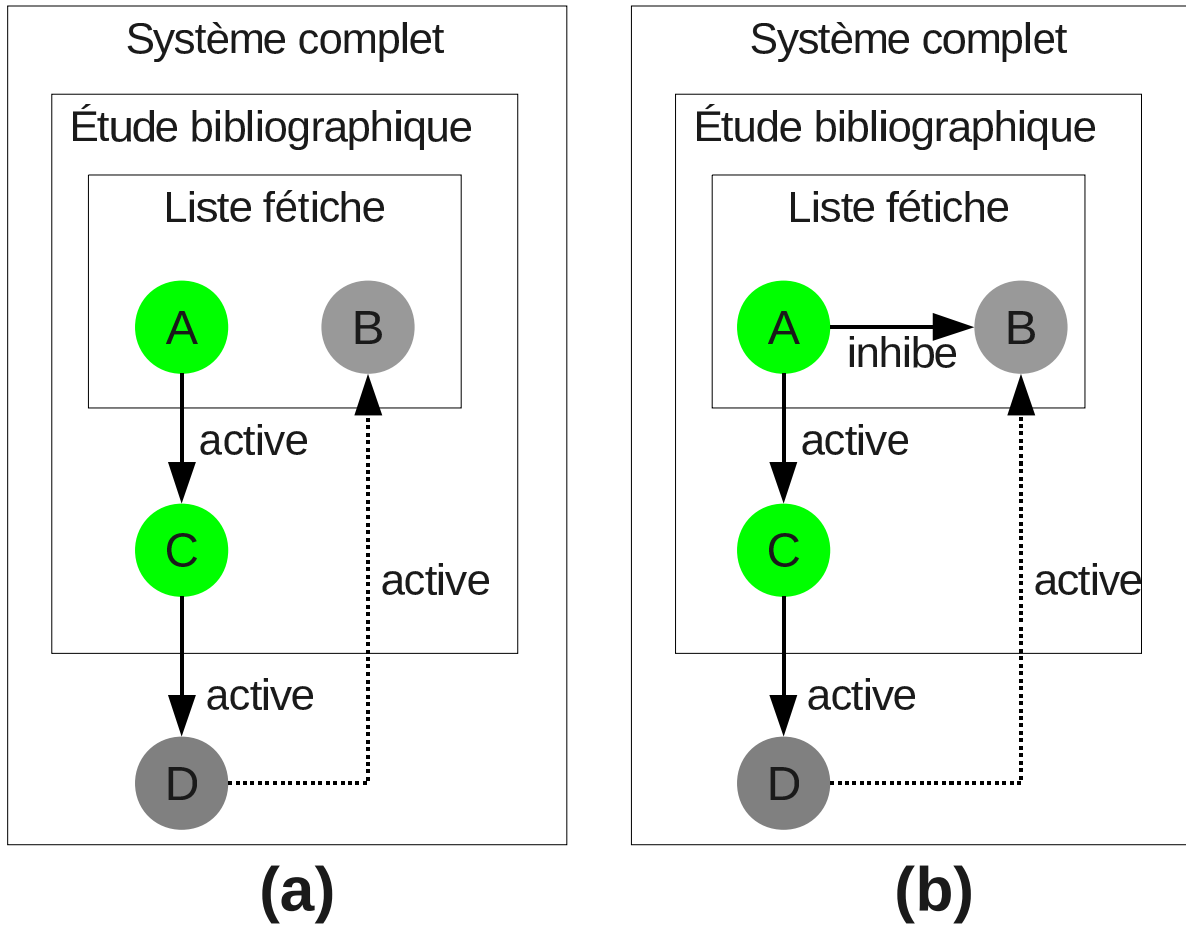


FIG. 3.2.1 – Une connaissance partielle des informations diminue la capacité de prédiction et peut entraîner des conclusions erronées. Les éléments dont la variation est connue sont notés par des cercles verts, ceux dont la variation est inconnue par des cercles gris. Les interactions connues entre éléments sont notées par des flèches continues, les inconnues par des flèches en pointillés. Dans l'exemple (a), la connaissance d'un sous système n'est pas suffisante pour déduire la variation de B alors qu'elle peut être déduite du système complet. Dans l'exemple (b), la connaissance du sous système indique que B varie dans le sens contraire de A, or le sens de variation de B ne peut pas être déduit du système complet.

Deuxième partie

Construction de bases d'interactions dans un formalisme permettant de tester la cohérence entre données observées et connaissances bibliographiques

Chapitre 1

Matériel et méthodes

1.1 Introduction

L'analyse manuelle des données permet d'exploiter uniquement un petit sous-ensemble des connaissances disponibles.

Les données expérimentales sont disponibles pour une exploitation automatique Le choix des données à exploiter est entièrement manuel car il nécessite une expertise propre au contexte expérimental de leur obtention. Par contre, une fois les données à exploiter obtenues, ces dernières sont analysables automatiquement. En effet, elles sont disponibles sous forme d'objets numériques annotés (par exemple grâce aux informations MIAME) dont le format est décrit (ou trivial) et peuvent donc être utilisées directement sans formalisme supplémentaire.

Une fois les données obtenues, nous avons cherché à savoir s'il est possible d'analyser automatiquement une liste de gènes dont on connaît le sens de variation entre deux (ou quelques) conditions.

Extraction des publications grâce à des moteurs de recherche Que ce soit pour une analyse manuelle ou automatique, les diverses publications rapportant des données expérimentales ou des informations sur des mécanismes et des régulations peuvent être sélectionnées par l'intermédiaire de moteurs de recherche lexicaux, comme Pubmed ou Google Scholar (pour une liste voir [5]) ou directement choisies par un expert (c.f. figure 1.1.1). L'utilisation de moteurs de recherche textuels est basée sur la présence de mots clés dans le texte des publications et non sur le sens de ce que l'on recherche. Il est ainsi impossible de formuler des requêtes portant sur les relations entre éléments, comme par exemple : "quels sont les activateurs de PPAR alpha?", ou d'obtenir des résultats pertinents quand on recherche un mot possédant plusieurs sens¹. Un autre problème de la recherche par mot clef est la présence de synonymes pour décrire un même concept [13], ce qui complique grandement les recherches lorsqu'une personne souhaite exploiter la bibliographie d'un domaine pour lequel elle n'est pas experte². Une solution pour résoudre ce problème est de s'appuyer sur des bases de synonymes [13], il faut alors rechercher toutes les combinaisons de synonymes possibles, ce qui peut s'avérer fastidieux. Les moteurs de recherche sont donc un bon outil pour extraire des publications relatives à un ensemble de mots clés connus *a priori*, mais ils ne sont pas prévus pour extraire de l'information des publications.

¹Par exemple le mot "<modèle>" peut avoir le sens d'un animal dont le comportement est représentatif de ce qui se passe chez l'homme (ex : [113]) ou avoir le sens de modèle mathématique (ex : [252]).

²Par exemple, Pubmed renvoie 1567 publications pour " "peroxisome proliferator activated receptor alpha" " et 2236 publications pour " "PPAR alpha" "

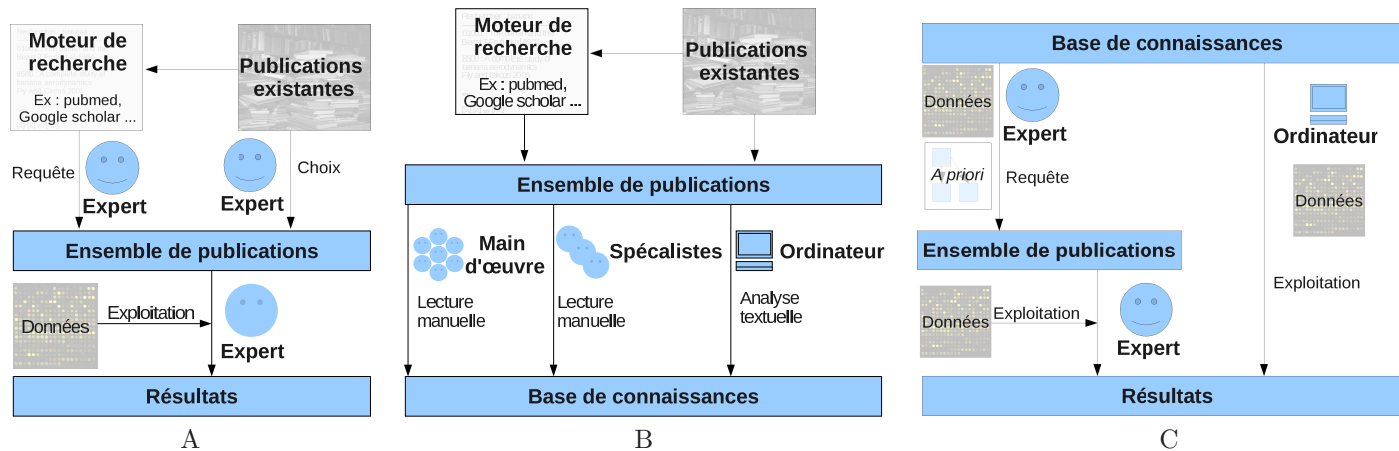


FIG. 1.1.1 – Diverses méthodes d'exploitation de la bibliographie

A) Les informations bibliographiques sont sélectionnées par les experts, éventuellement par l'intermédiaire d'un moteur de recherche. Ces informations sont ensuite utilisées par l'expert afin d'exploiter des données expérimentales.

B) Les informations bibliographiques peuvent aussi être réunies sous un même formalisme dans une base de données de connaissances. Cette base peut être construite par la lecture manuelle de la bibliographie par des spécialistes d'un domaine particulier ou par une main d'oeuvre plus large. Une autre alternative est d'extraire l'information automatiquement grâce à une machine.

C) Les informations présentes dans une base de données peuvent soit être analysées manuellement par des experts qui dans ce cas sélectionnent un ensemble d'éléments de la base (grâce à des *a priori*, comme par exemple une voie métabolique, ou grâce à leurs données, comme par exemple un *cluster* d'éléments coexprimés), ils vont alors lire la bibliographie relative à ces éléments puis les analyser. Une autre approche consiste à utiliser l'ensemble (ou un vaste sous-ensemble) de la base, et à analyser ce dernier grâce à un ordinateur.

Photographies :

books in a stack (a stack of books),

publiée sous Creative Commons attribution 2.0 par austinevan <http://www.flickr.com/photos/austinevan/1225274637/>.

microarray, publiée sous Creative Commons attribution 2.5 par www.sgn.cornell.edu.

ordinateur, publiée sous Creative Commons attribution 2.0 par conanil <http://www.flickr.com/photos/24559254@N00/1186276193>

Extraction automatique des informations présentes dans les publications L'extraction d'informations à partir d'un ensemble de publications peut soit se réaliser en lisant les publications, soit automatiquement. L'analyse automatisée de textes repose sur deux étapes principales [162] : l'identification des molécules mises en jeu et l'identification des interactions. L'identification des molécules est un problème important car les noms utilisés dans les publications ne sont pas normés, malgré l'effort réalisé depuis quelques années pour construire des ontologies [162]. Cette étape nécessite donc de construire un dictionnaire de synonymes spécifique des noms de molécules attendus, ce qui est particulièrement compliqué pour les noms des métabolites qui n'ont pas fait l'objet d'une normalisation unique adoptée pour toutes les publications ³.

Lors de la seconde étape, il est simple d'identifier des co-occurrences de termes dans les publications ⁴ mais l'analyse de la signification des relations nécessite une approche plus complexe : un *tokenizer* permet d'identifier le type grammatical des mots (ex : nom, verbe ...) puis un arbre syntaxique est construit et analysé grâce à des règles de grammaires. Le *tokenizer* nécessite pour fonctionner un dictionnaire associant à chaque mot sa fonction, ce qui nécessite de construire des listes de vocabulaires relatives aux termes courants et au vocabulaire spécifique du domaine d'étude. L'analyse de l'arbre syntaxique repose sur un ensemble de règles qui peuvent soit être écrites à la main, soit induites à partir d'un corpus de phrases déjà interprétées (à la main). De tels outils ont été développés par exemple pour extraire spécifiquement les interactions entre protéines [318] ⁵.

Il n'y a aucun algorithme générique réellement au point capable d'extraire de l'information formalisée à partir de textes écrits en anglais [5] et les principales recherches réalisées actuellement se concentrent sur la recherche d'interaction entre protéines ([183, 225]). De plus, les algorithmes existants sont à la fois dépendants de listes de termes (que l'on peut obtenir dans des ontologies) et de règles d'interprétation des arbres syntaxiques dont l'apprentissage n'est pas évident. A ce jour, il paraît donc plus efficace de se tourner vers la lecture des publications.

Exploitation de bases de données de connaissances Les informations bibliographiques peuvent aussi être extraites et formalisées par la lecture manuelle de la bibliographie par un petit nombre de spécialistes d'un domaine particulier. Dans cette approche, les informations relatives au sujet d'étude sont extraites, interprétées et formalisées par des personnes compétentes sur le sujet, ce qui constitue un important gage de qualité. En contrepartie, cette approche ne peut se réaliser que sur un petit domaine du fait du champ de compétences particulier des experts. Ces bases expertes ne sont souvent pas publiées ou disponibles pour la communauté scientifique. C'est le cas par exemple de Gardon, base experte du métabolisme des lipides, que nous utiliserons dans la suite de ce travail. Cette base contient à la fois un ensemble d'éléments considérés comme importants dans le métabolisme des lipides par les experts et un ensemble de référence *a priori* relativement fiable.

A plus grande échelle, la lecture des publications peut être aussi réalisée par un vaste ensemble de personnes afin de couvrir un champ disciplinaire plus large qu'une base experte. C'est ce que proposent les bases de données de connaissances généralistes telles qu'Ingenuity ou TRANSPATH qui affirment couvrir un domaine d'étude beaucoup plus large que les bases expertes ⁶.

³Par exemple, le C18 : 3ω 3, le C18 : $3n$ -3, le *linoléic acid*, l' *α -linolenic acid*, le *all-cis-9,12,15-octadecatrienoic acid*, le *cis,cis,cis-9,12,15-Octadecatrienoic acid*, le *(9Z,12Z,15Z)-9,12,15-Octadecatrienoic acid* sont tous des synonymes faisant référence à l'acide alpha-linoléique. Il faut rajouter à ces termes les abréviations courantes comme *LNA* dont la signification dépend de leur contexte.

⁴Les co-occurrences sont par exemple exploitées dans PubMatrix pour faciliter l'annotation de listes de gènes [20]

⁵Par exemple en 2003, l'extraction à partir de textes bruts d'interactions de type protéines-protéines identifie 83.5% des noms de protéines avec une spécificité de 93.1% et 63.9% des interactions avec une fiabilité de 70.2%. En 2008, la précision sur les interactions entre protéines est de 80 à 90% [225]

⁶Ingenuity affirme dans sa FAQ couvrir entre autre les interactions moléculaires et les régulations (" *The Ingenuity Knowledge Base is a repository of molecular interactions, regulatory events, gene-to-phenotype associations, and chemical knowledge that provides the building blocks for pathway construction. It is the largest knowledge base of its kind, with millions of findings pulled from the full text of the life sciences literature by Ph.D.-level life scientists, and features the most descriptive and detailed structure, and the highest degree of accuracy.*")

TRANSPATH affirme couvrir la régulation des gènes et des facteurs de transcriptions (" *TRANSPATH is an information system on gene-regulatory pathways. It focuses on pathways involved in the regulation of transcription factors. Elements of the relevant signal transduction pathways such as hormones, enzymes, complexes and transcription factors are stored together*")

Problématique L'utilisation de ces bases de données pose la question de l'évaluation de leur qualité et de leur contenu. En effet, pour un sujet tel que le métabolisme des lipides et ses régulations qui est plus spécifique que le domaine d'étude des bases généralistes et plus vaste que ce qui peut être couvert par une base experte, il est difficile de savoir quelle est la meilleure approche (base généraliste, base experte ou combinaison de plusieurs bases), et quelle est l'information réellement exploitable dans les bases disponibles.

Une fois l'information réunie et formalisée dans des bases de données, la démarche actuelle d'exploitation consiste à sélectionner un ensemble d'éléments de la base à partir d'un jeu de données ou d'un ensemble d'*a priori* afin d'obtenir un schéma mettant en relation des éléments. Ce schéma sert de base à l'expert qui va lire la bibliographie associée à chacun de ces éléments et de ces relations. Il va ensuite interpréter le schéma afin d'obtenir des résultats (c.f. figure 1.1.1).

1.2 Contenu accessible des différentes bases de données de connaissances

1.2.1 Critères de choix des bases utilisées

Le domaine des bases de données d'interactions est en constante évolution, et de nombreuses ressources sont disponibles (pour une liste voir <http://www.pathguide.org/> et BioGrid). En 2006 les deux organismes les mieux décrits étaient *Escherichia coli* avec *Regulon DB*⁷ qui contient essentiellement des mécanismes de régulations transcriptomiques et *Saccharomyces cerevisiae* avec Yeast Genome⁸ qui décrit essentiellement des relations entre protéines, YEASTRACT [317, 226] qui décrit des relations génétiques et N. Guelzim *et al.*, [126] qui proposent une analyse du réseau de régulations génétiques. Chez l'Homme, de nombreuses bases référencent des interactions entre protéines (pour une revue, voir [216]).

Dans le cadre de notre problématique, nous ne souhaitons pas exploiter les bases spécialisées pour les espèces modèles *Escherichia coli* et *Saccharomyces cerevisiae*, car ces dernières sont *a priori* phylogénétiquement trop éloignées de l'homme et des espèces d'intérêt agronomique. De plus, la connaissance d'interactions entre protéines est difficilement exploitable pour réaliser des prédictions sur le comportement du métabolisme, un de nos objectifs. Par conséquent, nous nous sommes tournés vers des bases plus proches des espèces d'intérêts qui comportent des interactions relatives à la régulation de la transcription et du métabolisme.

Nous avons donc étudié trois bases de données : Ingenuity, TRANSPATH et Gardon qui représentent un panel de ce qui était accessible dans le cadre de cette thèse. Ce choix a été réalisé à l'initiation de ce travail et néglige donc de nouvelles bases de données comme Pathway Commons (une base publique agrégeant dans un même formalisme des données référencées dans d'autres bases) qui n'avait que peu de contenu au début de cette thèse et qui connaît actuellement une croissance rapide⁹. Cette dernière base est très prometteuse, en particulier par ses efforts de formalisme, de définitions des termes et d'accessibilité pour des travaux haut débit.

Ingenuity est une base généraliste commerciale développée par *Ingenuity Systems Inc.* largement utilisée¹⁰ par les biologistes. Elle contient essentiellement des interactions entre entités biologiques et des données relatives aux effets des médicaments¹¹. D'après la documentation, ces informations sont soit extraites d'autres bases de données¹², soit extraites de la littérature¹³; la documentation ne précise pas sur quels critères

with information about their interaction.”)

⁷L'histoire de *Regulon DB* référence : 2 025 *regulatory interactions* en 2006 et 2 396 en 2009.

⁸Yeast Genome est une base d'interactions maintenant incluse dans BioGrid.

⁹D'après le site de Pathway Commons, cette base contient le 12 janvier 2010 : 421 395 interactions, 88 509 éléments et 441 organismes différents.

¹⁰La recherche du 23 juin 2009 dans Google Scholar d' " *Ingenuity Pathway Analysis* " fournit 1 550 résultats, celle de " *Ingenuity Systems* " 1 250

¹¹La documentation indique : " *[Ingenuity Pathways Analysis]'s expert-extracted content includes modeled relationships between chemicals, proteins, genes, complexes, cells, cellular components, tissues, drugs, cellular processes, diseases, and clinical phenotypes* ".

¹²Pour la liste voir http://www.ingenuity.com/products/pathways_analysis.html

¹³La documentation précise : " *The vast majority of Findings in IPA are manually curated and modeled by a team of Ph.D. scientists from primary literature sources, including peer-reviewed journal articles, review articles, and textbooks* ".

sont choisies les publications incluses dans cette base.

TRANSPATH est une seconde base commerciale généraliste développée par *BIOBASE Biological Databases* contenant des interactions entre entités biologiques (signalisation et mécanismes) ¹⁴. Cette base est elle aussi très utilisée ¹⁵. Elle est annotée manuellement par des experts. La documentation n'indique pas sur quels critères sont choisies les publications incluses dans TRANSPATH.

Gardon est une base experte réalisée en interne par des spécialistes du métabolisme des lipides qui ont sélectionné manuellement des relations importantes dans le métabolisme des lipides à partir de la bibliographie. Elle contient des interactions entre molécules à deux niveaux : le mécanisme (ex : phosphoryle) et l'effet (ex : active). Elle n'est pas accessible à la communauté scientifique mais son utilisation dans le cadre de cette thèse permet à la fois d'avoir un référent fiable concernant le métabolisme des lipides et d'évaluer s'il y a un gain lors de l'utilisation d'une base experte du métabolisme des lipides par rapport à l'utilisation des bases généralistes existantes.

1.2.2 Analyse des types d'informations référencées dans les bases étudiées

Nous décrivons d'abord les différentes informations présentes dans les trois bases de données choisies, nous évaluerons parmi elles celles qui sont extractibles automatiquement, puis une fois extraites, nous les comparerons.

Ingenuity

Une base bibliographique d'interactions biologiques Ingenuity est une solution intégrée réunissant à la fois une base de données bibliographiques et un ensemble de logiciels pour exploiter ces dernières. Nous avons analysé manuellement le contenu de cette base afin d'identifier les types d'informations réellement présents.

Des relations entre éléments Ingenuity contient un ensemble de termes (ex : un nom de molécule, une pathologie . . .) qui sont mis en relation au sein d'une phrase en vocabulaire contrôlé. Le vocabulaire contrôlé est un ensemble fini de mots qui permet de décrire les relations entre les termes.

Chaque terme et chaque phrase sont annotés par :

- Une ou plusieurs publications, repérées par un identifiant Pubmed
- Parfois, une citation du texte de la publication à partir de laquelle le terme ou la phrase en vocabulaire contrôlé sont extraits.

Les phrases d'Ingenuity décrivent des mécanismes de régulation, des réactions biochimiques, des effets, des fragments de molécules (ex : La protéine *A* a une sous unité *B*), des résultats expérimentaux et des rôles dans une pathologie.

Les termes d'Ingenuity décrivent des molécules, des fragments de molécules, des séquences, des pathologies, d'autres relations (ce qui permet de combiner plusieurs relations dans les phrases en vocabulaire contrôlé). Les termes d'Ingenuity ne sont pas normalisés. Ils sont parfois suivis d'un identifiant entre crochets mais la base à laquelle cet identifiant fait référence n'est pas précisée. Parfois il y a seulement un nom sans identifiant.

Les informations de contexte d'Ingenuity décrivent de manière optionnelle et dans un vocabulaire non normalisé, le niveau (gène ARNm ou protéine), l'espèce ou la lignée cellulaire, le lieu (organe, tissu, cellule), si l'expérience a eu lieu dans un mutant ou un *knockout*, l'âge des individus et parfois d'autres informations.

¹⁴La documentation de transpath indique "TRANSPATH provides information on signaling molecules, metabolic enzymes, second messengers, endogenous metabolites, miRNAs and the reactions they are involved in, resulting in a complex network of interconnected chains of reactions and pathways."

¹⁵3 800 résultats pour la recherche de " "Transpath" " dans Google Scholar le 23 juin 2009

Un moteur de recherche de termes relatifs à une saisie Ingenuity intègre aussi un moteur de recherche de termes qui permet de renvoyer un ensemble de termes proches de l'élément recherché. Aucune documentation n'a permis de comprendre exactement quelle était la sortie de ce moteur de recherche en fonction de son entrée, ni de savoir sur quels critères des éléments étaient considérés comme proches.

Les éléments obtenus suite à une requête ne sont pas des synonymes de l'élément recherché, en effet les résultats des recherches ne forment pas systématiquement une classe d'objets (exemple : en recherchant A on obtient B, puis en recherchant B on obtient C. Si le moteur renvoyait vraiment des synonymes, en recherchant A, B ou C on aurait obtenu l'ensemble A, B et C). Ce moteur est un outil pratique pour réaliser rapidement une recherche, mais il ne peut pas être exploité en tant que base de synonymes.

Une ontologie des molécules Ingenuity contient une ontologie de molécules qui annote ces dernières avec les voies métaboliques auxquelles elles appartiennent (nommées "*Canonical pathways*")¹⁶.

TRANSPATH

Transpath est une base de données dont le rôle est de décrire les mécanismes de transduction des signaux. Cette base contient des informations sur des molécules et des mécanismes de transduction de signaux essentiellement extraits manuellement de la littérature scientifique [56]. En plus des données, cette base propose une interface de visualisation et un outil d'analyse de données issues de *microarray*. Ces outils ne seront pas développés ici (pour un descriptif voir : [56]).

Molécules Les molécules sont annotées par des identifiants pointant vers des bases de données¹⁷. De plus, TRANSPATH propose un système d'alias permettant de réunir sous une même molécule plusieurs synonymes. TRANSPATH différencie aussi (avec une nomenclature normalisée) les différentes formes (ex : phosphorylée, ubiquitinée . . .) d'une même molécule. Il considère ces différentes formes comme des molécules différentes.

Relations Dans TRANSPATH, les mécanismes de transduction du signal sont décrits par des relations. Une relation contient :

- une liste de participants : ex la molécule *A* et la molécule *B*
- un type : par exemple activation
- le rôle de chacun des participants : ex *A* est l'activateur, *B* est l'activé.

Les relations entre molécules sont de deux types : l'effet (*semantic reactions*) et le mécanisme (*mechanistic reactions*). Les types de relations sont organisés dans une ontologie ; dans cette dernière il est possible de déduire l'effet à partir du mécanisme. Les différents types de relations sont nommés par un ensemble fini de termes dont la définition est clairement explicitée dans la documentation. Les relations sont annotées par la ou les publications sources à partir desquelles elles ont été extraites.

Contexte Transpath contient aussi pour chaque relation des informations sur le matériel et la méthode qui ont permis de la mettre en évidence. A partir de ces informations TRANSPATH construit un indicateur de qualité (la règle de construction est décrite dans la documentation) pour évaluer la fiabilité de cette dernière.

Gardon

Gardon est une base experte du métabolisme des lipides développée à l'IRISA et à Agrocampus-Ouest. Elle contient des relations entre molécules annotées par de la bibliographie.

¹⁶Cette ontologie a par exemple été utilisée pour analyser des données issues de *microarray*, c.f. 2.3.3.

¹⁷La documentation répertorie une liste non exhaustive de bases de référence : SwissProt, EMBL, InterPro, Entrez Gene, UniGene, GO, DIP, BIND, and HyperCLDB, PDB, PROSITE, Flybase, MGD

Molécules Les molécules sont annotées avec un identifiant pointant vers une base de données de molécules (HGNC, CAS, MGI, KEGG compound et MGD [28]) ou par leur nom courant. Elles sont annotées par leur type (produit de gène, substance ou complexe) et éventuellement par l'identifiant d'une autre molécule dont elles sont la forme active.

Relations Les relations sont annotées par une petite liste de termes décrivant des mécanismes. Ces termes sont, lorsque cela est possible, associés à leur effet. A chaque relation est associé un ou plusieurs identifiants Pubmed.

Contexte Gardon contient des informations de contexte : modèle physiologique ou aphysiologique, localisation cellulaire, niveau (ADN, ARN, protéine, activité ou non applicable), si l'expérience a eu lieu chez un hybride et un champ libre de commentaire.

Conclusion

Gardon et TRANSPATH, bien que construites séparément, ont toutes les deux une structure et une interface très similaires : un ensemble de molécules mises en relation. Elles diffèrent essentiellement par la manière dont est décrit le contexte pour lequel les relations ont été mises en évidence.

Ingenuity a un contenu beaucoup plus vaste que ces deux bases. Elle ne se limite pas à décrire des relations entre molécules annotées par leur contexte. Elle référence aussi d'autres informations comme l'implication de molécules ou de mécanismes dans des pathologies ou l'influence d'un médicament sur un phénomène.

Au final, l'information commune se résume à un contexte dont le contenu dépend de la base d'origine, et des relations entre molécules.

1.2.3 Analyse des informations exploitables automatiquement

Dans le cadre de notre problématique d'exploitation à haut débit des bases de données de connaissances, seule l'information extractible automatiquement peut être exploitée. Afin de l'identifier, nous analysons l'interface logicielle (ou "Application programming interface" (API)) mise à disposition par les bases analysées.

Interface des bases de données

Ingenuity L'interface d'Ingenuity est essentiellement composée d'outils graphiques permettant de construire un ensemble de sommets mis en relation. Une capture d'écran de cette interface est présentée figure 1.2.1.

L'interface utilisateur ne permet pas de réaliser des requêtes de manière automatisée afin de questionner la base. De plus, la sortie d'Ingenuity n'est pas exportable dans un format utilisable par un programme et l'affichage des objets sur l'écran est limité à 255 sommets¹⁸. La sortie la plus interprétable est la page web relative à chaque sommet. Cette dernière contient en effet une liste d'adresses url vers les pages html contenant la description de toutes les relations dans lesquelles est impliqué le sommet. Contacté sur ce sujet Ingenuity nous a indiqué que la licence ne permet pas d'accéder à une interface de requête utilisable par une machine (comme une base de données SQL, un *web service* ou un fichier plat), que ce soit pour la base complète ou pour un sous ensemble de cette base relatif au métabolisme des lipides.

TRANSPATH Transpath est une base de données entièrement disponible sous forme de fichiers texte dont le format est intégralement décrit dans la documentation. Ces fichiers suffisent à fournir une interface triviale dont l'accès au contenu nécessite uniquement l'écriture d'un *parser*¹⁹.

¹⁸Le dernier test de cette limitation effectué au cours de la Thèse date de 2008.

¹⁹Un *parser* est un programme informatique dont le rôle est de lire un fichier dont la structure est connue afin de stocker son contenu sous une forme exploitable par un autre programme. Par exemple, un *parser* peut lire un fichier de nombres au format texte séparés par des espaces, afin de stocker ces derniers en mémoire vive sous forme binaire dans un tableau d'entiers.

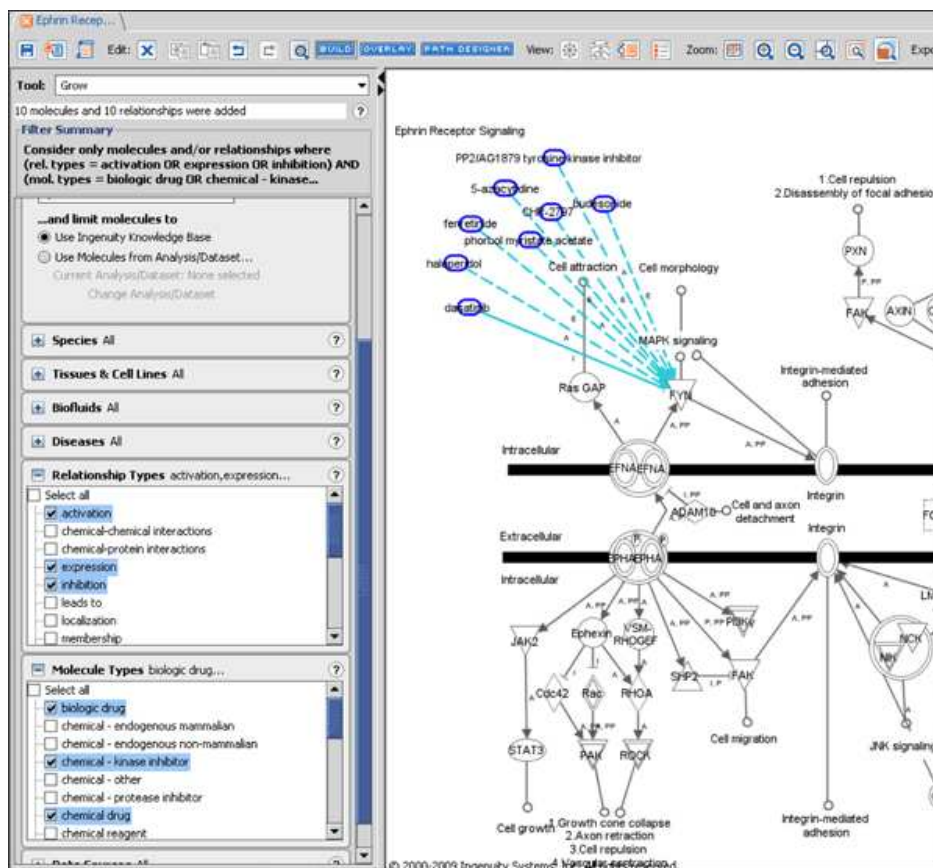


FIG. 1.2.1 – Capture d’écran de l’interface utilisateur d’Ingenuity. Ingenuity propose un ensemble d’outils graphiques permettant de construire rapidement un réseau d’éléments. Il est par exemple possible de rechercher les éléments correspondant à une chaîne de caractères, d’étendre le graphe aux voisins d’un sommet, ou de sélectionner un sous-graphe en fonction du type de relation ou de sommets. L’exemple ci-dessus montre à quoi peut ressembler un réseau obtenu grâce à cette base. L’image est extraite de la documentation officielle d’Ingenuity disponible sur http://www.ingenuity.com/products/pathways_analysis.html.

Gardon Gardon est une base de donnée relationnelle implémentée avec MySQL. Ce système de gestion de base de données dispose d’une interface standardisée utilisable facilement par de nombreux programmes informatiques (Java avec Java.sql, C++ avec mysql++, C avec MySQL API, ou bien dans un navigateur avec phpMyAdmin).

Conclusion Gardon propose une API directement exploitable sans développement en s’appuyant sur une bibliothèque logicielle déjà existante. Cette API permet d’accéder facilement à l’intégralité de l’information et de sélectionner facilement les éléments d’intérêt grâce au langage SQL.

TRANSPATH propose un format plus bas niveau dans lequel toute l’information est facilement disponible mais qui ne permet pas de réaliser rapidement des requêtes.

Ingenuity ne propose pas d’interface adaptée au haut débit. Les requêtes ne peuvent que se réaliser manuellement (sommet par sommet) et l’information obtenue suite à la requête n’est pas disponible dans un format directement interprétable par une machine. Nous avons donc dû développer une méthodologie d’extraction spécifique.

Méthode d'exploitation automatique des bases de données

Les informations communes des bases de données sont des relations et des influences entre molécules annotées par de la bibliographie. Après avoir défini à quel niveau nous souhaitons extraire ces informations, nous définissons une méthodologie pour les obtenir. Les bases extraites et les bases utilisées ultérieurement dans cette partie sont présentées tableau 1.2.1.

Niveau de détail des informations extraites Les informations bibliographiques sont facilement accessibles dans toutes les bases et sont toujours représentées sous la forme d'identifiants Pubmed. Nous utiliserons donc cette information telle quelle pour décrire l'origine de la bibliographie.

La documentation de TRANSPATH et la disponibilité des experts pour Gardon nous ont permis d'interpréter de manière claire leur relations (ex : A est un régulateur de B), nous les avons donc extraites. Nous les avons ensuite interprétées sous formes d'influences ($A \overset{?}{\mapsto} B$) en utilisant les règles de conversions définies par ces bases.

Pour Ingenuity, la seule information accessible est un ensemble de phrases écrites en vocabulaire contrôlé. Le sens des mots du vocabulaire contrôlé n'est pas documenté, et leur interprétation (même manuelle) n'est pas évidente. Par conséquent, en l'absence de sens à donner aux mots décrivant les relations, il n'est pas possible d'exploiter ces relations directement. Par contre, il est possible d'interpréter certains mots de vocabulaire contrôlé en terme d'influences²⁰. Les informations que nous avons extraites dans Ingenuity concernent donc le niveau de détail des influences.

²⁰Une influence de A vers B indique que la variation de A a un effet sur B . Cette influence peut être positive ($A \overset{+}{\mapsto} B$) : la dérivée partielle de B par rapport à A (dB/dA) est positive, négative ($A \overset{-}{\mapsto} B$) : dB/dA est négative ou de sens inconnu ($A \overset{?}{\mapsto} B$) : dB/dA est soit positive soit négative mais ne change pas de signe.

TAB. 1.2.1 – Les différentes bases de connaissances utilisées

| Bases de données d'origines (Gardon, TRANSPATH) et extraites (Ingenuity) | | | |
|--|--|---------------------|--|
| Bases d'origine | Nombre de sommets | Nombre de relations | |
| Gardon | 4 361 | 1 874 | Les bases Gardon et TRANSPATH contiennent des relations interprétables en influences. Ingenuity contient des phrases en vocabulaire contrôlé contenant une ou plusieurs relations. Le sens de ces dernières n'est ni évident ni documenté mais reste interprétable en terme d'influences. Par conséquent, nous n'avons pas extrait les relations, mais leur interprétation en terme d'influences (base Ingenuity [E]). De plus, l'intégralité d'Ingenuity n'est pas accessible, seul un sous ensemble est extrait à partir d'une liste de 1 738 gènes ^a . |
| TRANSPATH | 408 617 | 166 151 | |
| Partie extraite Ingenuity | Phrases en vocabulaire contrôlé 57 555 | | |
| Graphes d'influences | | | |
| Graphes d'influences | Nombre de sommets | Nombre d'influences | |
| Gardon [E] | 513 | 935 | Ces graphes d'influences sont construits en interprétant les relations ou les phrases des bases d'origine (ou de la partie extraite) en terme d'influences. |
| TRANSPATH [E] | 62 235 | 131 295 | |
| Ingenuity [E] | 3 459 | 11 673 | |
| Gardon [S] | 315 | 654 | Ces graphes contiennent les influences des graphes [E] dont les deux sommets sont annotés par un identifiant HGNC ou KEGG. Cette sélection nous garantit qu'une même molécule porte le même nom dans tous les graphes (c.f.1.3.6). |
| TRANSPATH [S] | 8 025 | 22 775 | |
| Ingenuity [S] | 2 274 | 9 116 | |
| Gardon [R] | 247 | 411 | Les bases notées R sont issues de la sélection à partir des bases extraites de toutes les influences dont au moins un des deux sommets est présent dans la liste de 1 738 qui a servi de base à l'extraction d'Ingenuity (c.f.1.3.2). |
| TRANSPATH [R] | 4 101 | 6 426 | |
| Ingenuity [R] | 2 427 | 6 033 | |
| Gardon [SR] | 177 | 323 | Ces graphes contiennent les influences des graphes [E] dont les deux sommets sont identifiés par un identifiant HGNC ou KEGG et dont toutes les influences ont au moins un de leur sommet présent dans la liste de 1 738 qui a servi de base à l'extraction d'Ingenuity. |
| TRANSPATH [SR] | 2 734 | 4 491 | |
| Ingenuity [SR] | 1 742 | 5 244 | |

^aCette liste contient un ensemble de gènes identifiés par expertise comme importants dans le métabolisme des lipides ainsi que les gènes dont l'expression varie lors de la mise à jeûn chez le poulet c.f.(2)

Méthode d'extraction de Gardon et TRANSPATH Au début de ma thèse, plusieurs formats permettaient de représenter des relations entre molécules : SBML et Biopax deux formats d'échange de données couramment utilisés en bioinformatique ; et les bases de données relationnelles²¹ : un outil générique pour stocker des objets en relation.

Le format Biopax est beaucoup plus intuitif pour des biologistes qu'une base de données relationnelle et SBML et Biopax sont facilement interopérables. L'interopérabilité est un point clef de l'automatisation des traitements et est donc un élément à considérer pour toute analyse à large échelle. En effet, l'utilisation d'un format interopérable permet de récupérer des informations de multiples sources et de s'assurer de la pérennité et de la diffusion des bases construites. De plus, SBML et Biopax permettent de représenter facilement des relations entre plusieurs éléments, comme par exemple une réaction catalysée par un enzyme. La contrepartie à leur utilisation est l'acquisition de compétences pour exploiter ces outils, ainsi que l'analyse en profondeur du sens des relations que ce format définit afin de s'assurer qu'elles sont convertibles en influences.

Actuellement, Biopax et SBML disposent d'un support logiciel important. Pour Biopax : les plugins BiNoM et cPath permettent son exploitation dans Cytoscape, et paxtools permet d'exploiter ce format en Java. Le format SBML peut être utilisé à l'aide de sa bibliothèque consacrée : libSBML. Ces formats constituent actuellement une alternative sérieuse à l'utilisation de bases de données relationnelles.

Lorsque ma thèse a débuté, les outils pour exploiter ces formats n'étaient pas suffisamment stables : pour SBML, le *changelog* de libSBML référençait encore de fréquentes corrections de *bugs* ; pour Biopax, un plugin fonctionnel pour Cytoscape était disponible mais il n'y avait pas de documentation pour utiliser l'API Cytoscape et le fonctionnement des plugins dans ce dernier restait assez aléatoire. De plus l'API Cytoscape n'est pas normalisée et change au gré des versions ce qui rend risqué l'investissement en développement logiciel dans un tel outil.

D'un point de vue pratique, les bases de données relationnelles permettent de stocker des objets en relation et fournissent une interface qui permet d'écrire des requêtes relativement complexes avec peu de code et de manière assez simple, ce qui évite le développement de petits utilitaires.

Nous avons fait le choix de conserver la structure de Gardon et de convertir les relations de TRANSPATH dans le même formalisme que Gardon (un travail réalisé par Pauline Gloaguen lors de son stage de Master [117]). Ce choix nous évite le développement d'une nouvelle structure de base de données et permet d'utiliser des outils communs pour exploiter ces deux bases.

Méthode d'extraction des influences d'Ingenuity La méthode d'extraction proposée est semi-automatique : nous entrons manuellement un élément à rechercher dans l'interface de requête puis toujours à la main, nous affichons la page html décrivant les informations relatives à cet élément. Cette page contient des liens vers les différentes pages html décrivant les relations dans lesquelles l'élément requêté est impliqué. Nous analysons alors chacune des pages décrivant ces relations.

L'aspect manuel de cette méthode de requête nous oblige à nous limiter à un sous-ensemble de ce que contient Ingenuity. Pour cela nous réunissons un ensemble d'éléments *a priori* importants pour le métabolisme des lipides identifiés par expertise avec un ensemble d'éléments impliqués dans le métabolisme énergétique identifiés expérimentalement par la variation de leur expression lors de la mise à jeûn chez le poulet (c.f. 2). Une fois réunis, nous obtenons 1 738 noms de gènes. Cette requête nous a permis d'extraire 44 683 pages web.

A partir de ces pages, un important travail d'ingénierie a été réalisé avec François Moreews, Ingénieur d'étude INRIA, Sigenae. Brièvement, les identifiants Pubmed et les phrases en vocabulaire contrôlé ont été extraits des pages, nous avons obtenu 57 555 phrases uniques et 7 867 identifiants Pubmed uniques. Nous avons ensuite reconstruit par ingénierie inverse la grammaire utilisée par Ingenuity en s'appuyant sur les expressions régulières identifiées lors de l'analyse de vocabulaire (c.f. figure 1.2.2). Cette grammaire a alors été codée dans ANTLR²² pour les relations dont nous pouvons interpréter le sens. A l'aide de ce logiciel,

²¹Plusieurs systèmes de gestion de bases de données existent (Oracle, MySQL, postgre SQL ...) Nous avons choisi d'utiliser MySQL car il dispose d'un support logiciel important et nous avons les compétences pour l'exploiter.

²²ANTLR est un outil de reconnaissance de langage qui permet d'analyser un texte à partir de la définition de sa grammaire. Pour plus d'informations voir : <http://www.antlr.org/>

chaque phrase en vocabulaire contrôlé a été convertie sous forme d'un arbre syntaxique que nous avons analysé grâce à un ensemble de règles afin d'en déduire des influences.

L'analyse de l'arbre syntaxique couplé à la recherche de noms de molécules dans les chaînes de caractères nous a permis d'extraire 11 673 influences.

Méthode d'extraction du nom des molécules d'Ingenuity Contrairement à Gardon et TRANSPATH, l'interface d'Ingenuity ne propose pas de référence explicite vers une base de données de molécules. Pire, ses termes sont parfois des molécules, parfois autre chose (une maladie par exemple) et la notion même de molécule n'est pas définie (par exemple est-ce que A est une molécule différente de A phosphorylé, ou bien une classe regroupant de manière indifférenciée A non phosphorylé et A phosphorylé?). En première approche nous avons fait le choix de considérer comme une molécule dans Ingenuity toute chaîne de caractères contenant une et une seule sous chaîne parmi l'ensemble des identifiants HGNC et des noms de métabolites KEGG.

Parmi les termes du vocabulaire contrôlé d'Ingenuity, nous avons pu repérer un ensemble de mots servant à définir des complexes. Ces derniers sont traités par notre méthode d'extraction et les différents termes formant les morceaux du complexe sont extraits. Ils sont alors considérés comme des molécules et identifiés comme précédemment. Si tous les éléments d'un complexe possèdent un identifiant HGNC ou KEGG le complexe est conservé. On crée alors un nouveau nom de molécule à partir de ces identifiants.

Cette méthode introduit une importante perte d'information et regroupe parfois sous un même identifiant plusieurs formes de la même molécule, ou plusieurs molécules contenant la même chaîne de caractères.

Discussion : des pertes d'informations à tous les niveaux lors de l'extraction d'Ingenuity Cette méthode d'extraction est une méthode qui comporte des pertes de sens et de données à chacune de ces étapes : l'extraction à partir d'une liste de gènes nous limite dans notre analyse à un sous-ensemble d'Ingenuity et l'étape de formalisation en graphes d'influences ignore de nombreuses relations et il est difficile d'évaluer sa qualité.

Par conséquent, cette thèse évalue uniquement ce que nous avons pu extraire après d'importants efforts de modélisation sous forme de graphes d'influences, ce qui limite la portée des informations contenues dans Ingenuity.

Méthode d'extraction du contexte Les bases proposent des informations de contexte (espèce, *in-vivo* vs *in-vitro* ...). Ces informations sont importantes, en particulier lorsque l'on souhaite corriger une erreur. En effet, nous ne savons pas si les données issues de différentes espèces ou de différents contextes biologiques sont extrapolables aux jeux de données que nous souhaitons analyser.

Par conséquent, nous conservons dans la mesure du possible les informations relatives au contexte que nous pouvons extraire. L'exploitation directe de ces informations par un programme informatique (comme un filtre ne sélectionnant par exemple que les expériences *in-vivo* obtenues chez la souris) est assez difficile car les termes décrivant les contextes ne sont pas normalisés.

Sans cette normalisation, nous avons fait le choix de conserver en langage naturel les informations de contexte et de considérer par défaut que les relations sont extrapolables aux cas analysés. Ces dernières restent tout de même exploitables à bas débit, comme par exemple dans le cas où une petite zone d'incohérence ou un petit sous-ensemble d'éléments clefs a été identifié.

Dans l'objectif d'analyser le métabolisme des lipides dans un cas non pathologique, le choix d'extrapoler par défaut est *a priori* raisonnable pour Gardon (une base experte) et pour TRANSPATH. Il est plus discutable pour Ingenuity : cette base contient (d'après une lecture manuelle de quelques pages) de nombreuses relations valables chez des animaux mutants ou dans des cas pathologiques.

Traçabilité Afin de toujours pouvoir remonter à la source, nous conservons toujours un lien indiquant à partir de quoi sont construits les objets que nous utilisons. Nous créons donc un lien entre les influences et les relations (Gardon et TRANSPATH) ou les phrases en vocabulaire contrôlé (Ingenuity) dont elles sont

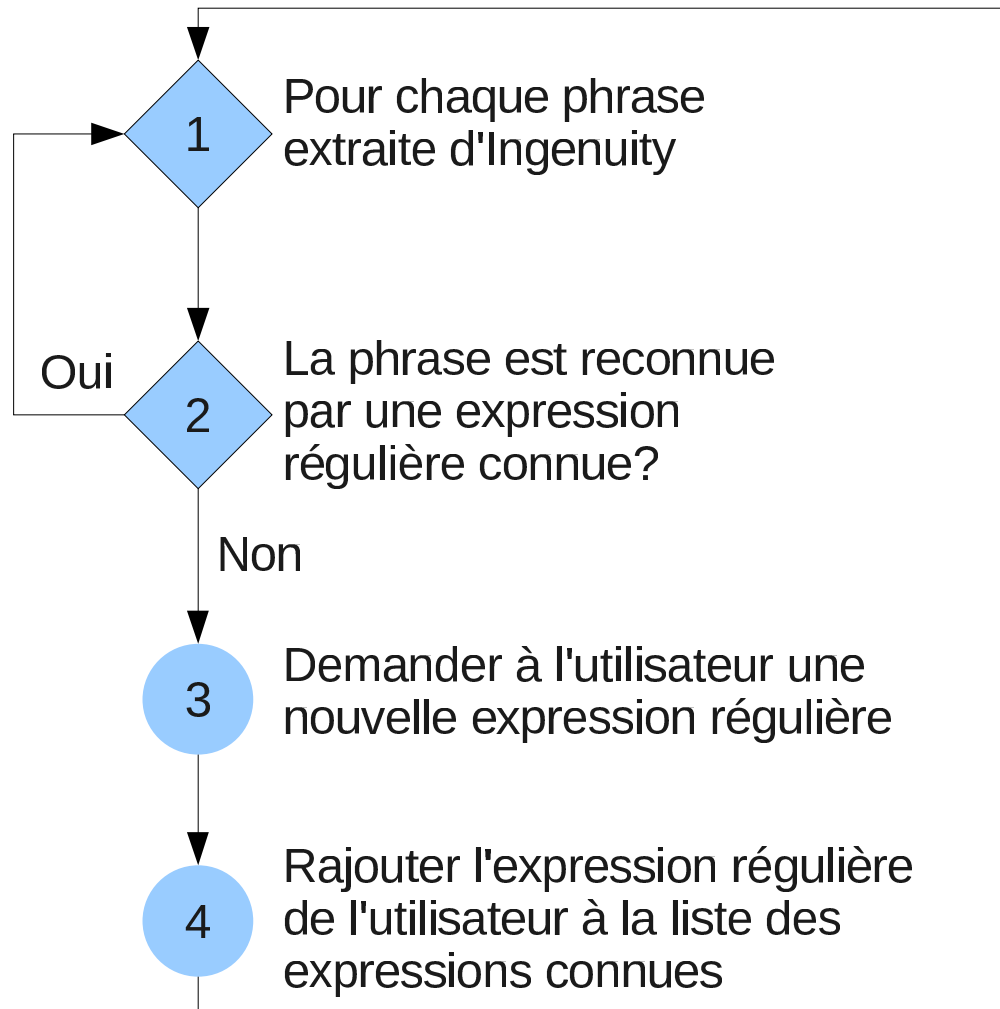


FIG. 1.2.2 – Méthode d'évaluation du vocabulaire d'Ingenuity

Tout d'abord, nous initialisons le programme avec une liste d'expressions régulières vide.

1) Pour les 10 000 premières phrases extraites d'Ingenuity nous réalisons le programme suivant :

2) Nous testons si la phrase est reconnue par une expression régulière. Si c'est le cas, alors l'analyse de cette phrase ne contient pas de nouveaux mots du vocabulaire contrôlé d'Ingenuity, et nous passons à la phrase suivante.

3) Si la phrase actuelle n'est reconnue par aucune expression régulière, nous entrons manuellement l'expression régulière correspondante, et annotons cette dernière par un type afin de pouvoir analyser les différents variants d'une même expression (par exemple les expressions régulières contenant "activate" et "activation of" seront notées de la même manière).

4) Nous rajoutons l'expression régulière entrée par l'utilisateur à la liste d'expressions régulières.

issues. Ces relations ou ces phrases sont dans toutes les bases associées à des identifiants bibliographiques que nous conservons lors du passage aux influences.

D'un point de vue ingénierie, ces liens nous permettent de déboguer les chaînes de programmes construits, et d'un point de vue exploitation, ils permettent d'appliquer la démarche du biologiste : une fois une incohérence ou un ensemble d'éléments clefs identifiés (automatiquement) par l'exploitation de la base de données (et éventuellement de données supplémentaires), le biologiste peut remonter aux publications sources afin de proposer une analyse plus fine, d'identifier des erreurs ou de proposer une nouvelle théorie.

1.3 Méthode d'évaluation du contenu des bases de connaissances

Lors de cette étude nous n'avons pas de référence absolue à laquelle comparer les bases, par conséquent nous comparons les bases entre elles afin de donner un positionnement relatif de ces dernières. Dans ce positionnement, la comparaison avec Gardon permet d'évaluer la pertinence des bases par rapport à ce qui est considéré comme important par les experts du métabolisme des lipides, tandis que les comparaisons entre les deux bases généralistes nous permettent d'évaluer laquelle est la plus adaptée pour la réalisation d'analyses automatiques de son graphe d'influence. Pour terminer, la comparaison des éléments communs entre les bases nous informe sur une stratégie d'exploitation : faut-il utiliser la meilleure des bases ou en combiner plusieurs, et si oui lesquelles ?

Bien que les bases de données disponibles affirment toutes décrire des interactions entre des molécules, elles ne contiennent pas toutes les mêmes informations. De plus, ces informations ne sont pas représentées de la même manière d'une base à l'autre. Afin de pouvoir réaliser des comparaisons, nous identifions donc l'information commune (indépendamment de sa représentation) puis nous représentons cette information dans un formalisme commun à toutes les bases de données.

1.3.1 Choix du niveau de comparaison : les influences

Transpath et Gardon contiennent des relations et des influences

Transpath et Gardon définissent suffisamment précisément leurs relations pour établir des tables d'équivalence entre les relations de chaque base. Il est donc possible de comparer entre elles la plupart des relations de Gardon et TRANSPATH. De plus, il est possible de déduire de ces relations un ensemble d'influences, comme présenté en 1.2.3.

La partie extraite d'Ingenuity contient des influences.

Les phrases Ingenuity empiètent parfois de nombreuses informations dont le découpage en relations n'est pas facile ; il est donc difficile d'extraire proprement un ensemble de relations.

De plus, Ingenuity ne propose pas de documentation suffisamment claire pour savoir ce que représentent biologiquement les termes utilisés pour mettre les éléments en relation. Par exemple, il n'est pas possible de savoir si "*A is an activator of B*" dans Ingenuity a le même sens que *A active B* dans Gardon et TRANSPATH. Par conséquent, même en extrayant correctement les relations, nous ne pouvons pas comprendre précisément leur signification afin de les interpréter, ce qui rend leur utilisation directe impossible.

Le seul niveau réellement exploitable d'Ingenuity est l'influence car nous pouvons à la fois extraire certaines influences et une fois extraites, nous savons leur donner une signification.

Conclusion : seules les influences sont comparables

La comparaison entre les relations n'a de sens que si ces dernières ont la même signification (ou des significations comparables) entre les différentes bases. En contrepartie, la signification des influences est indépendante de la base dont elles sont issues.

Si on souhaite rester au niveau de la relation, il n'est pas possible de réaliser des comparaisons entre les relations de TRANSPATH ou Gardon et les phrases en vocabulaire contrôlé d'Ingenuity. Nous avons

donc fait le choix de conserver un niveau de détail moins précis : le graphe d'influences. En nous limitant à ce niveau, nous perdons des informations, mais nous nous plaçons à un niveau pour lequel nous pouvons comparer beaucoup plus facilement les bases (c.f. figure 1.3.1).

1.3.2 Ingenuity [E] est-elle comparable aux autres bases ?

Gardon [E] et TRANSPATH [E] sont issues de l'extraction complète des bases de données sources alors qu'Ingenuity [E] n'est qu'un sous ensemble de ce qui est réellement disponible dans Ingenuity. Comparer ces trois bases sur un pied d'égalité est donc "injuste" envers Ingenuity.

La perte d'information entre Ingenuity et Ingenuity [E] s'opère à deux niveaux : la requête puisque nous avons extrait d'Ingenuity uniquement les pages web relatives à une liste de 1 738 gènes et non toutes les relations ; et la méthodologie d'extraction des influences qui n'exploite pas toutes les relations présentes dans les pages web.

Une première option est de ne pas prendre en compte cette différence ; on évalue alors ce qui a été extrait d'Ingenuity. Ce choix est une évaluation pertinente de cette base, en effet il reflète ce qui est exploitable dans le cadre de cette thèse.

Si on souhaite évaluer le positionnement relatif de cette base vis-à-vis de ce qu'on peut en extraire sans développer de nouvelle méthodologie, il faut corriger le biais dû à la méthode de requête. Pour cela, nous mettons toutes les bases sur un pied d'égalité en sélectionnant uniquement les influences dont au moins une des deux molécules est incluse dans la liste de 1 738 gènes qui a servi de base pour l'extraction d'Ingenuity. Les bases ainsi construites sont les bases Gardon [R], TRANSPATH [R] et Ingenuity [R]²³.

Cette méthode nous permet de positionner Ingenuity en terme de quantité de contenu relatif au métabolisme énergétique vis-à-vis des autres bases. L'extrapolation de ce classement à un autre cas est *a priori* fortement dépendante du domaine disciplinaire du cas en question et les données produites doivent donc être considérées comme un exemple.

Le biais dû à la mauvaise extraction des influences à partir des pages web reste pour l'instant insoluble. Il n'est donc pas possible d'évaluer le contenu réel d'Ingenuity en terme d'influences, mais seulement ce que nous avons pu analyser automatiquement en utilisant l'interface disponible.

1.3.3 Le contexte n'est pas comparable

Les trois bases définissent des informations relatives au contexte d'étude. Ces informations sont très difficilement comparables d'une base à l'autre car elles concernent des types variés (espèce, organe, lignée cellulaire, âge des patients), elles ne sont pas toujours présentes et leur contenu n'est pas normalisé.

En l'absence de normalisation des informations de contexte, nous ne pouvons pas exploiter ce dernier automatiquement, et donc nous ne pouvons pas le comparer. La définition d'une norme étant un travail beaucoup trop vaste pour être entrepris au cours de cette thèse, nous ne comparons pas les contextes.

1.3.4 Comparaison de l'origine de la bibliographie

Les trois bases associent un identifiant Pubmed à une relation (Gardon et TRANSPATH) ou à une phrase en vocabulaire contrôlé (Ingenuity). Cet identifiant est une information fiable et directement interprétable sur la source de données dont est extraite la relation ou la phrase. Une fois convertis en influences, il est alors possible d'associer ces mêmes identifiants Pubmed aux influences.

Il est donc possible pour les trois bases de comparer l'origine de la bibliographie ayant servi à obtenir les influences que nous avons extraites en se basant sur les identifiants Pubmed.

La couverture de la bibliographie se décline en deux volets : sa position en terme de champs disciplinaires et sa quantité. Nous évaluerons ces deux volets pour les bases entières (Gardon et TRANSPATH), ainsi que pour la sous-partie de ces bases relative au métabolisme énergétique (Gardon [R], TRANSPATH [R]

²³Cette sélection est aussi appliquée à Ingenuity car les phrases extraites contiennent plusieurs relations (et plusieurs influences), et car Ingenuity ne précise pas le critère de lien entre l'élément recherché et les phrases obtenues.

et Ingenuity [R]) obtenue par sélection sur une liste de 1 738 gènes identifiés comme importants pour ce métabolisme.

Position en terme de champs disciplinaires

D'un coté, nous souhaitons exploiter des bases qui couvrent un vaste domaine de la bibliographie, afin d'identifier de nouveaux éléments comme potentiellement intéressants, y compris parmi les éléments qui ne sont pas déjà connus pour avoir un rôle clef dans notre problématique. D'un autre côté, nous souhaitons aussi avoir une expertise importante dans le domaine du métabolisme des lipides, car nous étudions particulièrement ce domaine.

Afin d'obtenir une information sur les champs disciplinaires couverts par les différentes bases de données, nous recherchons automatiquement dans quel journal a été publié chaque identifiant Pubmed grâce au *webservice* <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/> (c.f. tableau 2.1.1). Nous comptabilisons alors pour chaque journal le nombre d'identifiants Pubmed différents qu'il référence (c.f. tableau 2.1.2). L'analyse directe de cette répartition nous permet de savoir si les bases de données utilisent de nombreuses sources ou concentrent leur analyse sur un faible nombre de journaux phares.

Nous trions ensuite cette liste par nombre d'identifiants Pubmed différents et analysons manuellement le titre des journaux phares (i.e. les premiers de la liste) de chaque base. Cette méthode donne un aperçu du positionnement relatif des bases, mais n'informe pas de manière absolue sur les thématiques choisies. En effet un journal qui contient beaucoup de publications a mathématiquement plus de chance d'être en tête de liste, y compris si la base choisit aléatoirement ses publications.

Quantité de sources bibliographiques différentes

Nous évaluons deux points : la quantité totale de bibliographie lue pour la réalisation des bases en comptant le nombre d'identifiants Pubmed²⁴, et la redondance entre les bases en calculant les éléments communs.

Lors de cette comparaison, nous comparons tout d'abord la bibliographie relative aux influences extraites afin de donner un positionnement de ce qui a été extrait des différentes bases de données.

Cette sélection nous permet aussi de positionner Ingenuity vis-à-vis des autres bases en terme de quantité de contenu relatif au métabolisme énergétique car elle limite la pénalité d'Ingenuity dans le classement dûe à la requête d'un sous-ensemble de ces éléments, et non de l'intégralité de la base.

1.3.5 Evaluation du contenu

Quantité d'information extractible en terme d'influences

Ce point est évalué en comparant le nombre de sommets et le nombre d'influences extraites des bases Gardon [E], TRANSPATH [E] et Ingenuity [E]. Les chiffres obtenus pour Gardon [E] et TRANSPATH [E] sont représentatifs du contenu réel de la base, ceux obtenus pour Ingenuity [E] sont représentatifs de ce qui a été extrait.

Quantité d'information extractible en terme d'influences relatives au métabolisme énergétique

La liste de 1 738 gènes construite pour requêter Ingenuity contient les noms de gènes qui sont soit considérés comme importants dans le métabolisme des lipides par les experts, soit expérimentalement mis en évidence comme importants dans le métabolisme énergétique. L'analyse de la couverture des bases vis-à-vis de cette liste nous informe donc sur leur aptitude à répondre à l'analyse d'un jeu de données particulier sachant qu'un ensemble de molécules clefs a déjà été identifié par expertise.

²⁴Ceci nous informe sur la diversité des sources bibliographiques, en effet rien ne nous permet de savoir si la publication a été lue dans sa totalité, ou si seulement un extrait a été analysé.

Ce point est évalué en comparant le nombre de sommets et le nombre d'influences extraites des bases obtenues après sélection sur la liste des 1 738 gènes relatifs au métabolisme énergétique (bases Gardon [R], TRANSPATH [R] et Ingenuity [R]).

Cette sélection nous permet aussi de positionner Ingenuity vis-à-vis des autres bases en terme de quantité de contenu relatif au métabolisme énergétique en limitant sa pénalité dans le classement due à la requête d'un sous ensemble de ces éléments, et non de l'intégralité de la base.

1.3.6 Analyse des complémentarités entre bases

Synonymes entre bases

Existence de synonymes entre les bases Les bases TRANSPATH et Gardon garantissent qu'une même molécule est toujours représentée par un unique identifiant au sein de ces bases. Ingenuity n'offre pas cette garantie directement mais notre méthode d'extraction traite ce problème.

Par contre, rien ne nous assure qu'une même molécule porte le même nom dans deux bases différentes. En effet, certaines molécules de Gardon ne sont pas identifiées par une référence à une base de molécules mais par leur nom commun. De plus les trois bases font référence à des bases de molécules différentes (par exemple TRANSPATH utilise beaucoup les termes CAS, tandis que la partie extraite d'Ingenuity utilise des identifiants KEGG).

L'existence de synonymes pose problème pour évaluer la redondance entre les bases puisque deux molécules portant des identifiants différents sont considérées à tort comme différentes.

Traitement des synonymes entre les bases Tout d'abord, nous avons construit un dictionnaire de synonymes qui associe à chaque synonyme d'une même molécule un identifiant unique. Pour les molécules présentes dans le dictionnaire, nous remplaçons leur nom par leur identifiant unique. Pour les molécules qui ne sont pas dans le dictionnaire, nous retirons ces dernières ainsi que les relations dans les quelles elles sont impliquées.

La construction d'un dictionnaire des synonymes oblige à identifier un ensemble de bases de données de molécules qui soit possèdent des liens entre elles pour identifier les synonymes d'une base à l'autre, soit contiennent des ensembles de molécules qui ne sont pas synonymes.

Dans le cadre de cette thèse nous nous sommes limités aux bases de références HGNC (noms de gènes chez l'homme) et KEGG (noms d'enzymes et métabolites). Nous passons donc à côté d'autres molécules comme celles identifiées par CAS par exemple. Une amélioration future consisterait à mieux gérer ces synonymes et à introduire en tant que synonymes les molécules issues de gènes orthologues afin de pouvoir obtenir une base commune aux différentes espèces.

La sélection des influences des bases de données Gardon [E], TRANSPATH [E] et Ingenuity [E] dont toutes les influences contiennent deux sommets annotés par un identifiant HGNC ou KEGG nous a permis d'obtenir les bases Gardon [S], TRANSPATH [S] et Ingenuity [S].

Complémentarité entre les bases

La complémentarité entre les bases est évaluée en calculant le taux de redondance c'est-à-dire le rapport : nombre d'éléments communs / nombre d'éléments dans la plus petite des bases comparées, les éléments étant soit des molécules, soit des influences, soit des identifiants Pubmed, soit des journaux.

Lorsque ce rapport concerne des influences, il est calculé sur les bases obtenues après sélection des influences dont tous les sommets sont annotés par un identifiant KEGG ou HGNC afin de ne pas avoir de synonymes entre les bases. Lorsqu'il concerne de la bibliographie, il est calculé sans réaliser cette sélection car il n'y a pas de problèmes de synonymes.

Le taux de redondance est calculé pour chaque couple de base, afin d'évaluer leurs éléments communs, et pour Gardon et l'union de TRANSPATH et Ingenuity ($Gardon \cap (TRANSPATH + Ingenuity)$) afin d'évaluer la pertinence de l'utilisation d'une base experte par rapport à ce qui est disponible dans les bases généralistes.

Le calcul de la redondance entre les bases Gardon [*S*], TRANSPATH [*S*] et Ingenuity [*S*] nous informe sur l'information commune des bases extraites tandis que son calcul sur Gardon [*SR*], TRANSPATH [*SR*] et Ingenuity [*SR*] nous informe sur la partie relative au métabolisme énergétique des bases.

1.3.7 Identification d'éléments clefs

En première approche, nous ordonnons les sommets du graphe d'influences extrait à partir de la liste de 1 738 gènes par le nombre d'influences entrantes ou sortantes de ce sommet. Un nombre élevé d'influences entrantes indique que le sommet est régulé par une multitude de molécules, tandis qu'un nombre élevé d'influences sortantes indique que le sommet régule de nombreuses molécules. Nous obtenons alors un ensemble ordonné d'éléments comportant les principaux carrefours des mécanismes de transduction du signal ayant un lien avec le métabolisme énergétique.

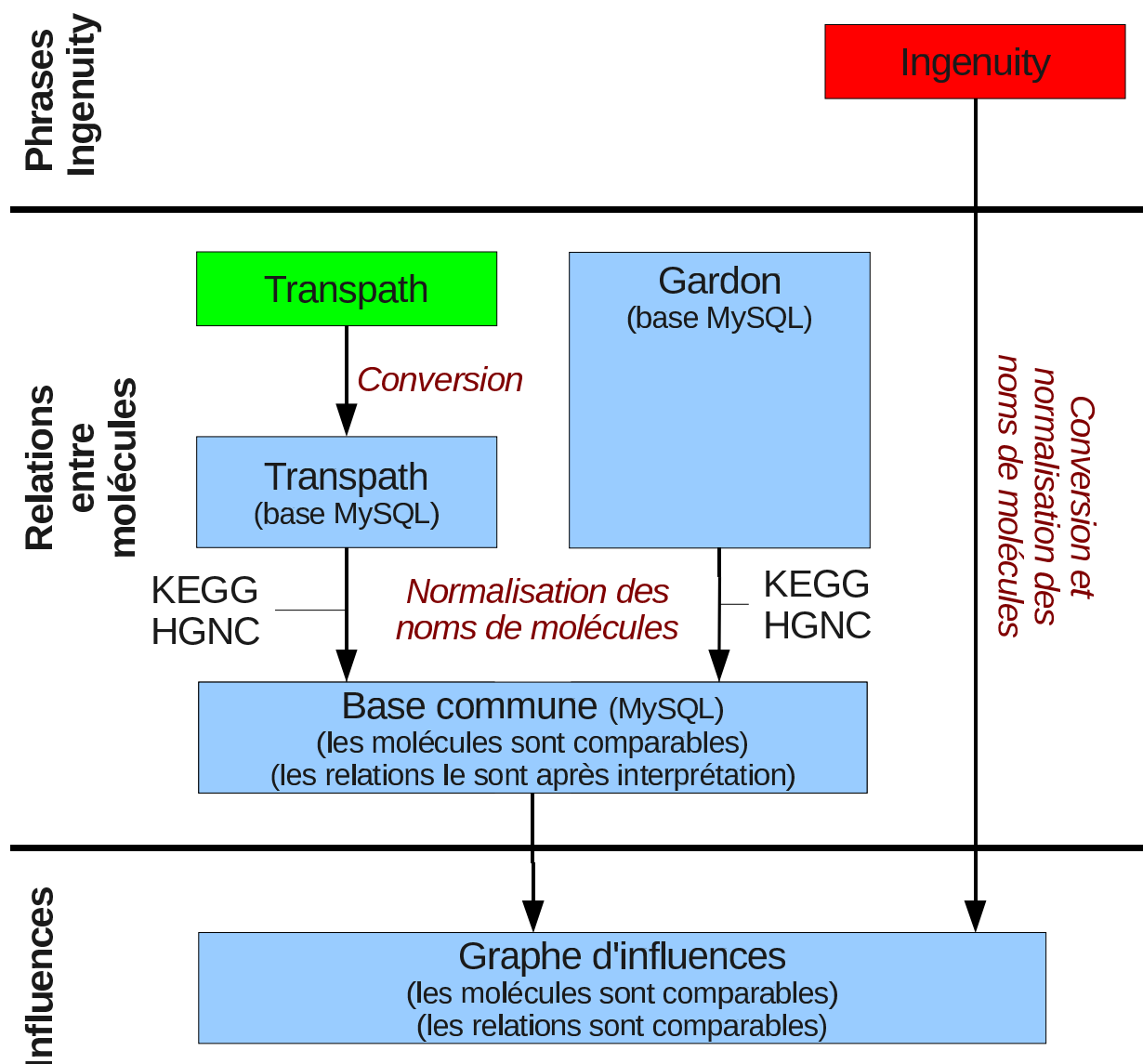


FIG. 1.3.1 – **Conversions entre bases de données** Transpath est convertie dans un formalisme proche de celui de Gardon (i.e. une base de données relationnelle MySQL), puis les noms de molécules sont normalisés. Il en résulte une base de données commune. En interprétant les différents types de relations de Gardon ou de TRANSPATH, il est possible de faire des comparaisons entre ces deux bases à ce niveau de détail. En raison de l'absence de documentation et de la sortie d'Ingenuity dont le format n'est pas exploitable, nous ne pouvons pas représenter cette base dans ce formalisme. Afin de réaliser des comparaisons, nous sommes forcés de descendre d'un niveau de détail et de construire un graphe d'influences. Notre objectif étant de comparer ces trois bases, nous n'avons pas exploité le niveau des relations pour comparer seulement TRANSPATH et Gardon et avons concentré nos efforts sur la comparaison des trois bases au niveau des influences.

Chapitre 2

Résultats et discussion

2.1 Position des bases de connaissances en terme de champs disciplinaires

2.1.1 Un grand nombre de journaux et quelques journaux phares

La majorité du contenu des bases est extraite d'un faible nombre de journaux phares. En effet les 10 premiers journaux des bases (en terme de nombre d'identifiants Pubmed différents) contiennent la plupart des identifiants Pubmed (les dix journaux phares de Gardon contiennent 78% des identifiants Pubmed différents, TRANSPATH : 52%, Ingenuity : 66%, Gardon [R] : 77%, TRANSPATH [R] 62% et Ingenuity [R] 75%). En revanche, les identifiants Pubmed restants sont dispersés entre de nombreux journaux qui apparaissent de 1 à 6 fois chacun.

TAB. 2.1.1 – Nombre de journaux communs

| Bases | Journaux (communs) | Couverture (%) ^a |
|---|--------------------|-----------------------------|
| Gardon [E] | 37 | |
| TRANSPATH [E] | 345 | |
| Ingenuity [E] | 505 | |
| Gardon [E] \cap TRANSPATH [E] | 33 | 89 |
| Gardon [E] \cap Ingenuity [E] | 35 | 94 |
| TRANSPATH [E] \cap Ingenuity [E] | 237 | 69 |
| Gardon [E] \cap (TRANSPATH [E] + Ingenuity [E]) | 36 | 97 |
| Gardon [R] | 33 | |
| TRANSPATH [R] | 163 | |
| Ingenuity [R] | 171 | |
| Gardon [R] \cap TRANSPATH [R] | 24 | 73 |
| Gardon [R] \cap Ingenuity [R] | 28 | 85 |
| TRANSPATH [R] \cap Ingenuity [R] | 88 | 54 |
| Gardon [R] \cap (TRANSPATH [R] + Ingenuity [R]) | 30 | 91 |

^aNombre de journaux communs / Nombre de journaux de la plus petite base

TAB. 2.1.2 – Nombre d’identifiants Pubmed différents par journaux

| Base Nombre d’identifiants Pubmed différents par journaux | Nombre de journaux (proportion) | | | | | |
|---|---------------------------------|------------------------|--|---------------------|------------------------|--|
| | Gardon [<i>E</i>] | TRANSPATH [<i>E</i>] | Ingenuity [<i>E</i>] ^{<i>a</i>} | Gardon [<i>R</i>] | TRANSPATH [<i>R</i>] | Ingenuity [<i>R</i>] ^{<i>b</i>} |
| = 1 | 18 (49%) | 140 (28%) | 147(43%) | 15 (45%) | 71 (44%) | 83 (49%) |
| [2; 5] | 11 (30%) | 138 (27%) | 110(32%) | 12 (36%) | 50 (31%) | 43 (25%) |
| [6; 10] | 4 (11%) | 68 (13%) | 20(6%) | 2 (6%) | 13 (8%) | 13 (8%) |
| [11; 15] | 1 (3%) | 25 (5%) | 15(4%) | 3 (9%) | 6 (4%) | 4 (2%) |
| [16; 20] | 1 (3%) | 11 (2%) | 7(2%) | 0 (0%) | 2 (1%) | 8 (5%) |
| [21; 25] | 1 (3%) | 7 (1%) | 3(1%) | 0 (0%) | 4 (2%) | 5 (3%) |
| [26; 50] | 0 (0%) | 32 (6%) | 16(5%) | 1 (3%) | 12 (7%) | 4 (2%) |
| ≥ 50 | 1 (3%) | 84 (17%) | 26(7%) | 0 (0%) | 5 (3%) | 11 (6%) |

^{*a*}Ingenuity contient en plus 171 phrases pour lesquelles nous n’avons pas extrait de Pubmed.

^{*b*}Ingenuity [*R*] contient en plus 69 phrases pour lesquelles nous n’avons pas extrait de Pubmed.

2.1.2 Coeur de cible des bases

Les dix premiers journaux phares des bases et les *ex aequo* sont présentés tableau 2.1.3. Ils donnent un aperçu de la politique des bases relativement fiable puisqu'ils contiennent à eux seuls plus de la moitié des publications.

On remarque tout d'abord que J Biol Chem est le journal dont est extraite la majorité des interactions des bases, ce qui est cohérent à la fois avec l'*impact factor* de ce journal¹, sa ligne éditoriale et l'objectif des bases de décrire des connaissances de biochimie. Ce classement contient aussi d'autres grands journaux de biochimie (EMBO J, Biochem Biophys Res Commun, Biochem J, Biochim Biophys Acta) et des journaux généralistes (Proc Natl Acad Sci (i.e. PNAS)).

De plus TRANSPATH et Ingenuity font aussi référence à plusieurs journaux spécialisés en santé humaine comme Oncogène (spécialisé cancer, dans TRANSPATH et Ingenuity), J Immunol (spécialisé immunologie, dans TRANSPATH et Ingenuity) et J Clin Invest (spécialisé dans les mécanismes biologiques relatifs aux pathologies, dans Ingenuity).

L'analyse de cette liste fait aussi ressortir la spécialisation de Gardon. En effet cette base contient deux journaux spécialisés lipides dans sa liste de journaux phares (J Lipid Res et Lipids) ainsi qu'un journal de nutrition : J Nutr. Tout comme les bases expertes, Gardon contient un grand nombre de publications extraites de journaux généralistes pour lesquels il est impossible de se prononcer sur le champ disciplinaire à la seule lecture du titre du journal.

¹D'après l'éditeur, ce journal reste le plus cité, voir <http://www.jbc.org/site/misc/journalranking.xhtml>

TAB. 2.1.3 – Journaux phares des bases

| Gardon [E] | | TRANSPATH [E] | | Ingenuity [E] | |
|----------------------------|---------------------|----------------------------|--------|--------------------------|--------|
| Journal | Pubmed ^a | Journal | Pubmed | Journal | Pubmed |
| J Biol Chem | 62 | J Biol Chem | 7468 | J Biol Chem | 1964 |
| Proc Natl Acad Sci U S A | 21 | Mol Cell Biol | 1509 | Endocrinology | 871 |
| Biochem Biophys Res Commun | 17 | Proc Natl Acad Sci U S A | 1174 | Proc Natl Acad Sci U S A | 564 |
| J Lipid Res | 15 | J Immunol | 1037 | Cancer Res | 329 |
| Mol Cell Biol | 8 | Biochem Biophys Res Commun | 957 | J Lipid Res | 328 |
| Biochem J | 8 | Oncogene | 862 | Mol Cell Biol | 258 |
| Lipids | 6 | EMBO J | 733 | Oncogene | 243 |
| Biochim Biophys Acta | 6 | Blood | 574 | J Immunol | 225 |
| J Nutr | 5 | Biochem J | 548 | Mol Endocrinol | 201 |
| Genes Dev | 4 | FEBS Lett | 517 | J Clin Invest | 189 |
| Diabetes | 4 | | | | |

| Gardon [R] | | TRANSPATH [R] | | Ingenuity [R] | |
|----------------------------|--------|----------------------------|--------|--------------------------|--------|
| Journal | Pubmed | Journal | Pubmed | Journal | Pubmed |
| J Biol Chem | 47 | J Biol Chem | 599 | J Biol Chem | 912 |
| Biochem Biophys Res Commun | 14 | Mol Cell Biol | 160 | Endocrinology | 419 |
| J Lipid Res | 13 | EMBO J | 90 | Proc Natl Acad Sci U S A | 224 |
| Proc Natl Acad Sci U S A | 12 | Proc Natl Acad Sci U S A | 89 | J Lipid Res | 155 |
| Mol Cell Biol | 8 | Biochem Biophys Res Commun | 66 | Cancer Res | 140 |
| Biochem J | 7 | Cancer Res | 48 | Oncogene | 123 |
| Lipids | 5 | Mol Cell | 48 | Mol Cell Biol | 122 |
| Biochim Biophys Acta | 4 | FEBS Lett | 43 | Mol Endocrinol | 107 |
| J Nutr | 4 | Mol Endocrinol | 42 | J Immunol | 88 |
| Biosci Biotechnol Biochem | 3 | Oncogene | 40 | J Clin Invest | 80 |
| Cell | 3 | | | | |
| Genes Dev | 3 | | | | |
| Mol Endocrinol | 3 | | | | |

^aNombre de Pubmed différents dans le journal

2.1.3 Bibliographie relative aux lipides

Pour la majorité des journaux présents dans la liste phare de Gardon dont la thématique est relative au métabolisme des lipides, les bases généralistes les citent avec un nombre d'identifiants Pubmed différents bien plus élevé que Gardon².

Les seuls cas où la base experte contient plus de bibliographie extraite de journaux relatifs au métabolisme des lipides que les bases généralistes concernent : Lipids (en dernière position dans TRANSPATH avec un identifiant Pubmed et 78 ème position dans Ingenuity avec six identifiants Pubmed) et J Nutr qui n'est pas présent dans TRANSPATH.

Par conséquent, à quelques exceptions près, les bases généralistes contiennent plus d'information relative au métabolisme des lipides que la base experte, bien que leur coeur de cible soit plus généraliste. Ceci confirme, sous réserve de pertinence des informations référencées, leur intérêt pour une problématique trop vaste pour une expertise et dont la frontière est incertaine sur des thématiques telle que le métabolisme des lipides.

2.2 Quantité de bibliographie

TAB. 2.2.1 – Nombre d'identifiants Pubmed communs

| Bases | Pubmed (communs) | Couverture (%) ^a |
|--|---------------------|-----------------------------|
| Gardon | 196 | |
| TRANSPATH | 29 240 ^b | |
| Ingenuity | 7 867 | |
| Gardon [E] ∩ TRANSPATH [E] | 40 | 20 |
| Gardon [E] ∩ Ingenuity [E] | 103 | 53 |
| TRANSPATH [E] ∩ Ingenuity [E] | 969 | 12 |
| Gardon [E] ∩ (TRANSPATH [E] + Ingenuity [E]) | 115 | 59 |
| Gardon [R] | 151 | |
| TRANSPATH [R] | 1 979 | |
| Ingenuity [R] | 3 156 | |
| Gardon [R] ∩ TRANSPATH [R] | 7 | 5.6 |
| Gardon [R] ∩ Ingenuity [R] | 50 | 33 |
| TRANSPATH [R] ∩ Ingenuity [R] | 67 | 3.4 |
| Gardon [R] ∩ (TRANSPATH [R] + Ingenuity [R]) | 54 | 35 |

^aNombre d'identifiants Pubmed communs / Nombre d'identifiants Pubmed de la plus petite base

^bUn identifiant Pubmed parmi ceux référencés dans TRANSPATH était erroné et ne pointait pas vers une publication. Il a été manuellement retiré de l'analyse

2.2.1 Les bases généralistes contiennent beaucoup de bibliographie

Lorsqu'on analyse la quantité totale de bibliographie, les bases généralistes arrivent en tête de classement suivies de loin par Gardon [E], ce qui est bien évidemment cohérent avec les objectifs de ces bases (c.f. tableau 2.2.1).

Même après sélection sur les 1 738 gènes identifiés comme importants dans le métabolisme énergétique, Gardon reste toujours loin derrière les bases généralistes. Par conséquent Gardon traite beaucoup moins de publications que les bases expertes, y compris sur un sujet incluant son domaine d'expertise.

² J Nutr est en 54 ème position dans Ingenuity avec 13 identifiants Pubmed, J Lipid Res apparaît en 5ème position dans Ingenuity (sans doute en raison du biais de requête) et en 68ème position dans TRANSPATH (avec 64 identifiants Pubmed), Diabetes arrive en 21 ème position dans Ingenuity avec 58 identifiants Pubmed et en 73ème position dans TRANSPATH avec 62 identifiants Pubmed.

2.2.2 Ingenuity est une base riche en bibliographie

Après sélection sur les 1 738 gènes identifiés comme importants dans le métabolisme énergétique, Ingenuity [R] contient beaucoup plus d'identifiants Pubmed que TRANSPATH [R] ce qui rend cette base intéressante pour identifier des publications relatives à un ensemble de gènes.

L'exploitation de la bibliographie à partir d'Ingenuity est tout de même plus difficile qu'à partir de TRANSPATH ; en effet l'absence d'API nécessite d'extraire à la main une page web par élément de la liste de requêtes³. Une fois les pages web récupérées, l'extraction automatique des identifiants Pubmed est facile.

Par conséquent, lorsque l'objectif est d'identifier un grand nombre de publications relatives à une liste de quelques milliers de molécules, Ingenuity est - sous réserve que la bibliographie extraite soit pertinente - la source à exploiter en priorité, malgré l'aspect manuel de son utilisation.

2.2.3 Les trois bases de données ne référencent pas les mêmes publications

L'intersection entre TRANSPATH et Ingenuity est faible avant et après sélection sur la liste des 1 738 gènes. Par conséquent il est intéressant d'exploiter de manière conjointe ces deux bases de données.

Gardon est une base très mal couverte, que ce soit par TRANSPATH ou par Ingenuity. Par conséquent, si l'objectif est d'identifier des publications relatives au métabolisme énergétique, on ne peut pas se dispenser de réaliser une base experte.

Inversement, Gardon ne contient pas toute l'information présente dans les bases généralistes, y compris lorsqu'on se restreint à la liste des 1 738 gènes considérés comme importants.

2.2.4 Conclusion

Au final les différentes bases n'exploitent pas les mêmes sources bibliographiques, et il y a donc un gain important à en exploiter plusieurs afin de couvrir un maximum de bibliographie. De plus, étant donné que nous ne savons pas à quel point l'information est redondante d'une publication à l'autre, nous ne pouvons pas extrapoler ce résultat au contenu des bases et devons donc analyser ce dernier.

2.3 Quantité d'information extractible

2.3.1 Quantité d'information extractible en terme d'influences

Nombre de sommets et d'influences Lorsque nous analysons le nombre de sommets (c.f. tableau 2.3.1) et d'influences (c.f. tableau 2.3.2) présents dans les bases entières, TRANSPATH [E] et Ingenuity [E] devancent de loin Gardon [E]. L'importance de l'écart confirme qu'il est impossible d'espérer construire une base de l'envergure de TRANSPATH à l'échelle d'un ou de quelques laboratoires. Ceci est appuyé par le nombre de relations de Gardon⁴ (1 874) qui est bien plus petit que celui de TRANSPATH (166 151).

Nombre de sommets annotés Dans ces trois bases Gardon [E] contient un pourcentage important d'éléments annotés dans HGNC et KEGG ; ceci s'explique par la méthode de construction de la base qui a consisté à rechercher systématiquement les références HGNC de tous ces éléments. Ingenuity [E] a aussi un fort pourcentage de sommets annotés en raison de la méthode d'extraction. Les 100% ne sont pas atteints en raison de la présence de complexes (c.f. 1.2.3). TRANSPATH [E] comporte seulement 13% d'éléments annotés par ces bases, le reste est annoté par d'autres bases non exploitées ici telles que CAS, ou contient des complexes formés à partir de molécules. Ces 13% représentent tout de même une quantité importante d'informations qu'il est possible d'étendre ultérieurement en exploitant de nouvelles bases d'identifiants de molécules.

³Un ordre de grandeur est 5 secondes par page et par personne

⁴Dans la base d'origine, avant conversion en influences

TAB. 2.3.1 – Annotation des sommets par KEGG ou HGNC

| Base | Nombre de sommets | Nombre de sommets normalisés ^a | Rapport (%) ^b |
|------------------------|-------------------|---|--------------------------|
| Gardon | 513 | 336 | 65 |
| TRANSPATH | 62 235 | 8 129 | 13 |
| Ingenuity ^c | 3 459 | 2 402 | 69 |
| Gardon [R] | 247 | 182 | 44 |
| TRANSPATH [R] | 4 101 | 2 742 | 66 |
| Ingenuity [R] | 2 427 | 1 758 | 72 ^d |

^aSommets annotés par un identifiant HGNC ou KEGG

^bNombre de sommets normalisés / nombre de sommets total

^cCes chiffres ne doivent pas être comparés avec les autres bases car KEGG et TRANSPATH sont des bases complètes tandis qu’Ingenuity représente ce que nous en avons extrait. Par conséquent toute comparaison sous-estime le contenu réel d’Ingenuity.

^dLe chiffre de 100% n’est pas obtenu car la méthode d’extraction des relations détecte les relations de types “complexes”, il y a donc 28% du contenu de la partie extraite d’Ingenuity qui est composé de complexes d’éléments référencés dans KEGG ou HGNC.

TAB. 2.3.2 – Nombre d’influences extraites des bases de données bibliographiques

| Bases | Nombre d’influences | Pertes lors de la sélection (%) ^a |
|---------------|---------------------|--|
| Gardon [E] | 935 | |
| TRANSPATH [E] | 131 295 | |
| Ingenuity [E] | 11 673 | |
| Gardon [R] | 411 | 56 |
| TRANSPATH [R] | 6 426 | 95 |
| Ingenuity [R] | 6 033 | 48 |

^aLes bases notées [R] ont été obtenues en sélectionnant toutes les influences dont au moins un des sommets est présent dans une liste de 1 738 gènes. La procédure est détaillé en 1.3.2.

2.3.2 Quantité d'information extractible en terme d'influences relatives au métabolisme énergétique

Effet de la sélection sur la liste des 1 738 gènes relatifs au métabolisme énergétique sur le nombre de sommets

La sélection des bases qui ne conserve que les interactions dont au moins un élément est dans la liste des 1 738 gènes fait perdre un grand nombre de sommets de TRANSPATH [E] (on perd 93% des sommets) ce qui est normal car on réduit une vaste base entière à un petit sous-ensemble. Les pertes d'Ingenuity [E] sont faibles (29%) en raison de la méthode d'extraction de cette base. Celles de Gardon [E] (19%) sont encore moins élevées, ce qui veut dire que les experts n'ont identifié comme importants que peu de sommets qui n'appartiennent pas à la liste des 1 738 ayant servi à la sélection.

L'étape de sélection n'influe pas sur la relation d'ordre en terme de quantité de sommets entre les bases : que ce soit lorsque nous considérons tous les sommets ou lorsque nous considérons les sommets annotés dans HGNC ou KEGG, TRANSPATH reste le premier. Dans le cas de notre liste de 1 738 gènes, il est donc plus productif de réaliser une sélection de TRANSPATH [E] pour obtenir des sommets ayant un lien avec ces éléments que de tenter de les extraire d'Ingenuity [E] ou de les faire identifier par des experts. Le temps de travail (quelques heures pour TRANSPATH [E], plusieurs semaines pour Ingenuity et Gardon) plaide d'autant plus en faveur de l'utilisation d'une licence TRANSPATH [E] pour identifier un grand nombre de molécules.

Ce résultat doit, pour le moment, être nuancé ; en effet rien ne dit que les bases contiennent toutes les mêmes sommets, ni que le grand nombre de sommets de TRANSPATH [E] ou Ingenuity [E] ne s'explique pas par un contenu abondant, mais peu fiable.

Effet sur le nombre de sommets annotés de la sélection des sommets inclus dans la liste des 1 738 gènes relatifs au métabolisme énergétique

Une fois les bases sélectionnées sur la liste des 1 738 gènes, ces dernières contiennent, comme attendu, un pourcentage d'éléments annotés dans KEGG ou HGNC bien plus élevé qu'avant sélection puisque chaque influence contient au moins un identifiant HGNC. On n'obtient pas 100% car le second élément de l'influence n'est pas forcément un élément de la liste des 1 738 gènes.

Les pourcentages d'éléments annotés dans HGNC ou KEGG obtenus restent faibles (sauf pour Ingenuity [R] en raison de la méthode d'extraction), ce qui signifie que des problèmes de synonymes entre bases peuvent encore exister sur ces bases, même après sélection sur la liste des 1 738 gènes.

Effet de la sélection sur la liste des 1 738 gènes relatifs au métabolisme énergétique sur le nombre d'influences

La sélection diminue de manière importante le nombre d'influences des bases pour les mêmes raisons que celles exposées lors de l'analyse des sommets.

Après sélection, la base issue de TRANSPATH comporte un peu plus d'influences que celle extraite d'Ingenuity (6 426 contre 6 033). De plus l'extraction à partir de TRANSPATH demande un travail beaucoup plus simple que l'extraction à partir d'Ingenuity. Cette base contient de plus des informations qui sont *a priori* plus fiables car la signification des termes décrivant les relations est détaillée dans la documentation et car la méthode d'exploitation est plus directe. Par conséquent s'il faut exploiter une unique base afin de récupérer un maximum d'influences pour analyser une liste d'éléments importants dans le métabolisme énergétique TRANSPATH est, sur le plan quantitatif le meilleur candidat.

Les autres bases peuvent néanmoins apporter de l'information complémentaire à TRANSPATH : Ingenuity [S] contient presque autant d'influences que TRANSPATH [S] et Gardon contient *a priori* des informations importantes dans notre cadre d'étude car elles ont été sélectionnées par des experts.

Conclusion

En terme de quantité d'information directement exploitable, ou en terme de quantité d'information relative au métabolisme énergétique, le meilleur choix consiste à se tourner vers TRANSPATH, qui en plus de proposer le contenu le plus vaste nécessite le temps d'exploitation le plus faible. Il est suivi de près par Ingenuity mais l'aspect laborieux de l'exploitation de cette base, et l'incertitude sur le sens de son contenu diminuent grandement son intérêt. Gardon arrive sans surprise à la troisième position, ce qui est normal pour une base experte.

2.3.3 Les bases sont complémentaires

L'étude des sources bibliographiques employées suggère que les différentes bases contiennent des informations potentiellement complémentaires. Nous allons donc évaluer l'utilité de l'exploitation des différentes bases en calculant leurs éléments communs. Les résultats du calcul des éléments communs sont présentés tableau 2.3.3.

TAB. 2.3.3 – Evaluation de la redondance entre les bases

| Base | Sommets communs | Taux de redondance (%) ^b | Influences communes ^a | Taux de redondance (%) ^c |
|--|-----------------|-------------------------------------|----------------------------------|-------------------------------------|
| Gardon [S] | 315 | | 654 | |
| TRANSPATH [S] | 8 025 | | 22 775 | |
| Ingenuity [S] | 2 274 | | 9 116 | |
| Gardon [S] \cap TRANSPATH [S] | 225 | 71 | 10 | 1.5 |
| Gardon [S] \cap Ingenuity [S] | 163 | 52 | 41 | 6.2 |
| TRANSPATH [S] \cap Ingenuity [S] | 1 387 | 61 | 177 | 1.2 |
| Gardon [S] \cap (TRANSPATH [S] + Ingenuity [S]) ^d | 240 | 76 | 47 | 7.2 |
| Gardon [SR] | 177 | | 323 | |
| TRANSPATH [SR] | 2 734 | | 4 491 | |
| Ingenuity [SR] | 1 724 | | 5 244 | |
| Gardon [SR] \cap TRANSPATH [SR] | 106 | 60 | 8 | 2.4 |
| Gardon [SR] \cap Ingenuity [SR] | 104 | 59 | 32 | 9.9 |
| TRANSPATH [SR] \cap Ingenuity [SR] | 611 | 35 | 68 | 1.5 |
| Gardon [SR] \cap (TRANSPATH [SR] + Ingenuity [SR]) | 126 | 71 | 37 | 11 |

^aSeules les influences contenant deux sommets possédant un identifiant HGNC ou KEGG sont utilisées dans cette analyse afin de corriger le biais de synonymes (c.f. ??). Deux influences sont considérées comme identiques si elles ont les mêmes sommes, le même sens, et le même signe

^bNombre de sommets communs / nombre de sommets de la plus petite base

^cNombre d'influences communes / nombre d'influences de la plus petite base

^dElements communs entre Gardon [S] et la réunion de TRANSPATH [S] et Ingenuity [S].

Les bases généralistes ne contiennent pas les informations de la base experte

La base Gardon est une base réalisée par expertise, une tâche qui demande une quantité de travail importante. Une alternative éventuelle serait de n'exploiter que les bases généralistes. Afin d'évaluer la pertinence de cette alternative, nous calculons la proportion de Gardon contenue dans les bases généralistes.

Le calcul du nombre d'éléments communs entre Gardon [S] et TRANSPATH [S] ou Ingenuity [S] montre tout d'abord que la majorité des sommets Gardon [S] se retrouve dans chacune de ces deux bases. Par contre, lorsque l'on s'intéresse aux influences, les bases ne se recouvrent pas du tout.

Ce point démontre les lacunes des bases généralistes pour décrire des interactions sur le métabolisme des lipides, un sujet plus précis que leur objectif. Par conséquent, dans l'état actuel des choses, l'expertise ne peut pas être substituée par la simple exploitation de bases généralistes. Il est donc tout à fait utile de combiner une base experte avec une (ou plusieurs) bases généralistes.

TRANSPATH ne contient pas la partie extraite d'Ingenuity

L'exploitation de plusieurs bases généralistes multiplie les coûts de licences et les efforts d'exploitation. Pour évaluer le gain à utiliser une source supplémentaire, nous évaluons la quantité d'information commune entre ce que nous avons extrait d'Ingenuity, et ce qui est disponible dans TRANSPATH.

Là encore, l'intersection en terme d'influences est très faible. Par conséquent, sous réserve que les informations extraites d'Ingenuity soient pertinentes, il est judicieux d'exploiter les deux bases.

L'analyse après sélection sur la liste des 1 738 gènes n'améliore pas beaucoup le pourcentage de recouvrement en terme d'influence. On peut donc raisonnablement supposer que le faible recouvrement en terme d'influences ne s'explique pas par une extraction d'un sous-ensemble trop petit d'Ingenuity, mais bien par un contenu différent.

2.4 Identification d'éléments clefs

Les dix premiers éléments en terme d'influences entrantes, sortantes et totales sont présentés tableau 2.4.1. Ces éléments sont des carrefours des mécanismes de transduction du signal ayant un lien avec le métabolisme énergétique.

2.4.1 Des résultats globalement cohérents avec les *a priori*.

Des éléments clefs cohérents avec les résultats issus du *clustering* Parmi les éléments clefs, on retrouve des gènes connus pour réguler (entre autre) les principales fonctions mises en évidence lors de l'étape de *clustering* des données de *microarray* issues de la mise à jeûn du poulet (c.f. 2.3). Ces éléments sont les gènes relatifs à la voie de l'insuline (INS⁵ IRS1⁶ et IRS2⁷), au catabolisme des acides gras (ACOX1⁸), à la synthèse des acides gras (NR1H3⁹ ACACA¹⁰ et FASN¹¹) et au métabolisme du cholestérol HMGCR¹².

Des éléments clefs cohérents avec la bibliographie Deux éléments relatifs au transport des lipides : ABCA1¹³ et FABP4¹⁴ ont aussi été mis en évidence. Comme attendu après l'étude bibliographique (c.f.2.3).

⁵Le gène de l'insuline

⁶*Insulin receptor substrate 1*

⁷*Insulin receptor substrate 2*

⁸*Peroxisomal acyl-coenzyme A oxidase 1*; un enzyme de la β -oxydation.

⁹LXR α , un ligand de SREBP1.

¹⁰*Acetyl-Coenzyme A carboxylase alpha* (aussi connu sous le nom d'ACC)

¹¹*Fatty acid synthase*

¹²*3-hydroxy-3-methylglutaryl-Coenzyme A reductase*, une étape limitante fortement régulée de la synthèse du cholestérol (d'après <http://www.ncbi.nlm.nih.gov/>).

¹³*ATP-binding cassette, sub-family A (ABC1), member 1*. La protéine codée par ce gène est un des principaux régulateur de l'homéostasie du cholestérol et des phospholipides.

¹⁴*Fatty acid binding protein 4*; un transporteur des acides gras.

Le transport des lipides est un carrefour important dans la régulation du métabolisme des lipides, il n'est donc pas surprenant de retrouver des éléments de cette fonction parmi les éléments clef.

La mise en évidence de ces deux éléments en particulier parmi l'ensemble des éléments du transport des lipides est par contre, une information qui n'était pas connue suite à l'étude bibliographique.

Une méthode dépendante de la qualité des sources Parmi les éléments identifiés, FAS (i.e., *Fas* (*TNF receptor superfamily, member 6*))¹⁵ est le seul qui n'est pas décrit dans la bibliographie comme un élément relatif au métabolisme énergétique et sa présence est inattendue. Ce terme est en effet identifié comme important à tort en raison de son homonymie dans Ingenuity avec "FAS" pour *fatty acid synthase*¹⁶. Ce point fait ressortir de manière concrète toute la difficulté à analyser des données dont les éléments ne sont pas normalisés.

2.4.2 Mise en évidence d'éléments clefs

Identification d'éléments dont l'importance n'était pas évidente *a priori* En plus des éléments dont l'importance était évidente *a priori*, l'analyse des sommets les plus connectés a permis d'identifier comme importants PPARC1A, MAPK1, UBC IGF1 et IGF2. Ces éléments sont particulièrement bien décrits dans la bibliographie, ce qui est une explication à leur présence dans le haut du classement en terme d'influences qui les concernent.

Les deux facteurs de croissances IGF1¹⁷ et IGF2¹⁸ interviennent notamment dans le métabolisme, la croissance cellulaire et la prolifération des cellules et dans de nombreuses autres voies de signalisation [343].

PPARGC1A¹⁹ est un gène dont la protéine est un coactivateur de la transcription des gènes du métabolisme énergétique. Cette protéine permet l'interaction de PPAR γ (un régulateur du stockage des acides gras et du métabolisme du glucose) avec divers facteurs de transcription. Elle intervient aussi dans la régulation de la genèse des mitochondries et elle possède divers autres rôles non détaillés ici²⁰.

MAPK1²¹ a aussi été identifiée comme un élément fortement régulé (en raison du nombre élevé d'influences entrantes). La protéine codée par ce gène est impliquée dans de nombreux mécanismes de régulation. Elle joue notamment un rôle dans la régulation de la synthèse des acides gras [238]. Cette protéine fait partie de la famille des MAP kinases. Les MAP kinases interviennent dans la prolifération et la différenciation des cellules (voir notamment MAPK14 ou p38MAPK, et les protéines ERK [243, 323]), dans la régulation de la transcription et dans le contrôle du développement [281, 44]. Cette protéine est un élément d'un immense carrefour de nombreuses voies de signalisation dont le rôle est bien plus vaste que celui du métabolisme énergétique.

L'UBC²² est une protéine capable de se fixer à d'autres protéines pour adresser ces dernières vers la dégradation dans le protéasome. Cette protéine joue aussi un rôle dans la maintenance de la structure chromatique, dans la régulation de l'expression des gènes, dans la réponse au stress, dans la genèse des ribosomes, dans la réparation de l'ADN²³ et dans le métabolisme de l'énergie [338].

Ces quatre molécules sont décrites dans la bibliographie pour avoir un rôle dans le métabolisme énergétique ou le métabolisme des lipides, ce qui est cohérent avec leur présence dans la liste. Elles participent aussi à la régulation de nombreuses autres fonctions biologiques. Cette analyse a donc fait ressortir leur importance en tant que carrefour de diverses voies de signalisation. Ces molécules sont de bons candidats potentiels²⁴ en terme de régulateurs clefs du métabolisme en réponse au jeûne.

¹⁵La protéine codée par ce gène est impliquée dans de nombreux mécanismes de régulations. Elle joue notamment un rôle dans l'apoptose.

¹⁶Le nom HGNC officiel pour la fatty acid synthase n'est pas FAS mais FASN.

¹⁷Insulin-like growth factor 1 (somatomedin C)

¹⁸Insulin-like growth factor 2 (somatomedin A)

¹⁹PPAR γ , coactivator 1 alpha

²⁰D'après www.ncbi.nlm.nih.gov.

²¹Mitogen-activated protein kinase 1

²²*Ubiquitin C*

²³D'après www.genecards.org

²⁴Leur rôle dans ce métabolisme nécessite tout de même d'être validé expérimentalement.

Un métabolisme très connecté aux autres fonctions D'un côté PPARGC1A et MAPK1 sont directement reliées par au moins une influence à un élément de la liste des 1 738 gènes et d'un autre côté ces molécules sont décrites comme intervenant dans la régulation de nombreuses fonctions biologiques. Il existe donc des liens directs (en terme d'influences) entre le métabolisme énergétique et d'autres fonctions de la cellule. Parmi ces fonctions, certaines influent directement sur la consommation d'énergie (comme par exemple la prolifération cellulaire). La présence de ces liens illustre la complexité de ce métabolisme et rend difficile la construction d'une frontière pour l'étudier comme une entité isolée.

TAB. 2.4.1 – Liste des 10 sommets les plus connectés au reste du graphe

| Nombre d'influences totales ^a | | Nombre d'influences entrantes ^b | | Nombre d'influences sortantes ^c | |
|--|--------|--|--------|--|--------|
| Identifiant | nombre | Identifiant | nombre | Identifiant | nombre |
| [HGNC]-IGF1 | 922 | [HGNC]-FASN | 357 | [HGNC]-IGF1 | 707 |
| [HGNC]-FASN | 605 | [HGNC]-IRS1 | 352 | [HGNC]-INS | 529 |
| [HGNC]-IRS1 | 601 | [HGNC]-IGF1 | 215 | [HGNC]-IRS1 | 249 |
| [HGNC]-INS | 561 | [HGNC]-IRS2 | 171 | [HGNC]-FASN | 248 |
| [HGNC]-FAS ^d | 266 | [HGNC]-ACACA | 131 | [HGNC]-FAS | 208 |
| [HGNC]-IRS2 | 259 | [HGNC]-ABCA1 | 120 | [HGNC]-UBC | 172 |
| [HGNC]-PPARGC1A | 247 | [HGNC]-HMGCR | 108 | [HGNC]-PPARGC1A | 148 |
| [HGNC]-IGF2 | 231 | [HGNC]-FABP4 | 101 | [HGNC]-NR1H3 | 147 |
| [HGNC]-NR1H3 | 204 | [HGNC]-PPARGC1A | 99 | [HGNC]-IGF2 | 139 |
| [HGNC]-ABCA1 | 184 | [HGNC]-MAPK1 | 98 | [HGNC]-ACOX1 | 114 |
| [HGNC]-UBC | 184 | | | | |

^aReprésente le nombre d'influences dans lesquelles le sommet est soit la source soit la cible

^bReprésente le nombre d'influences dans lesquelles le sommet est la cible

^cReprésente le nombre d'influences dans lesquelles le sommet est la source

^dFAS est (*TNF receptor superfamily, member 6*), ce n'est pas la fatty acid synthase qui est notée FASN.

2.5 Conclusion

L'extraction d'influences à partir des bases de données bibliographiques est possible En se limitant aux éléments relatifs au métabolisme énergétique ²⁵ et aux sommets annotés par un terme HGNC ou KEGG (bases [SR]), on obtient un graphe de 9 667 interactions et 3 865 sommets. En ne se limitant pas aux éléments annotés dans KEGG ou HGNC (bases [R]), le graphe construit contient 12 391 interactions et 5 903 sommets.

Tous les éléments d'intérêt ne sont pas décrits TRANSPATH [R] contient 47% des gènes de la liste des 1 738 identifiés comme importants dans le métabolisme énergétique, Ingenuity [R] 14% et Gardon [R] 4%. Une fois ces trois bases réunies, le graphe d'influences construit contient 49% des gènes de la liste.

Au final, malgré une quantité d'informations importante, seulement la moitié des gènes d'intérêts est impliquée dans au moins une influence. De plus, l'exploitation d'Ingenuity et de Gardon augmente la couverture seulement de 2% par rapport à l'utilisation de TRANSPATH. En effet, les bases de données contiennent une forte proportion de sommets communs (c.f. tableau 2.3.3), et cet ensemble ne suffit pas à traiter tous les éléments de notre liste d'intérêt.

Cette structure reflète probablement la structure de la bibliographie : les bases traitent un ensemble commun de sommets bien décrits dans les publications (i.e. les sommets de la liste présents dans les bases et les sommets des influences récupérées) tandis que les sommets peu décrits ne sont présents dans aucune des bases.

²⁵En sélectionnant sur la liste de 1 738, c.f.1.3.5.

Les influences des bases sont complémentaires Bien que les bases analysées aient de nombreux sommets communs, elles contiennent des influences différentes. Il est donc utile d'exploiter l'ensemble de ces bases afin d'avoir une meilleure couverture des influences relatives au métabolisme énergétique.

Les bases généralistes ne couvrent pas la base experte L'ensemble des influences extraites des deux bases généralistes TRANSPATH et Ingenuity ne recouvre pas les influences extraites de Gardon. Par conséquent sur un sujet tel que le métabolisme des lipides (le coeur de cible de Gardon), les bases généralistes ont d'importantes lacunes et ne peuvent pas se substituer entièrement à la réalisation d'une base experte.

L'extraction d'influences à partir de TRANSPATH est plus aisée qu'à partir d'Ingenuity Parmi les bases analysées, TRANSPATH est la meilleure source pour l'extraction d'influences, que ce soit de manière générale ou relativement au métabolisme énergétique. En effet, en plus d'un contenu important, TRANSPATH met à disposition une structure de données facilement exploitable. Cette dernière contient des relations dont le sens est clairement défini dans la documentation et des molécules identifiées par une référence vers des bases classiques de molécules.

Identification d'éléments clefs Au final, la construction de graphes d'influences permet de mettre en évidence des éléments clefs en relation avec un ensemble de gènes. La plupart de ces éléments sont cohérents avec les résultats du *clustering* et correspondent aux éléments connus par l'expertise et couramment décrits dans la bibliographie, ce qui corrobore cette méthode.

Cette méthode identifie aussi des éléments dont le rôle dans le métabolisme énergétique est connu, mais dont l'importance n'est pas évidente. La construction de graphes d'influences est donc utile pour repérer les principaux carrefours de transduction du signal auxquels sont connectés les éléments d'une liste.

La simple analyse des sommets les plus connectés est néanmoins une méthodologie brutale. Elle est fortement biaisée par l'abondance de bibliographie sur certains éléments : elle peut mettre en évidence un sommet clef très décrit dans un contexte différent du métabolisme énergétique et qui possède une seule interaction en lien avec les 1 738 de la liste. Par conséquent elle n'exploite pas toute la structure du réseau et de meilleurs résultats restent envisageables suite au développement de méthodes d'analyses plus fines de ce graphe, par exemple en exploitant sa topologie de manière plus globale et/ou en intégrant dans l'analyse la variation de certains sommets mesurée expérimentalement.

Troisième partie

Analyse d'un système complexe par simplification : application au métabolisme des acides gras lors du jeûne chez la souris

Introduction

Le métabolisme des acides gras est fortement régulé, puisque son comportement varie en fonction de l'espèce, du génotype, de l'organe et des conditions expérimentales mises en jeu. Les voies métaboliques sont généralement communes à tous ces contextes, mais leur importance relative varie. Or, si les réactions biochimiques sous-jacentes des voies qui participent au métabolisme des lipides, leurs enzymes et leurs principaux régulateurs sont relativement bien connues et bien référencées dans de nombreux manuels (ex : [231]) et bases de connaissances (ex : KEGG), les lois de réponse cinétique de ces fonctions, ainsi que leur hiérarchie d'importance vis-à-vis de la quantité et de la composition en lipides cellulaires restent mal connues. En effet, s'il existe de nombreux jeux de données décrivant les variations d'un ou plusieurs éléments du métabolisme des lipides, il reste difficile d'en déterminer la quantité d'informations déductibles et d'en identifier les éléments clefs. La modélisation apparaît alors comme un outil permettant de décrire de manière formalisée et analysable mathématiquement le métabolisme des lipides afin de simuler son comportement et d'identifier les éléments responsables des phénotypes observés.

Actuellement, de nombreux modèles dynamiques se sont attachés à décrire une voie particulière du métabolisme énergétique, et notamment la glycolyse [154, 319]. Ces modèles mettent en œuvre un niveau de détail important (24 réactions dans [154], 18 réactions dans [319]). Le niveau de détail de ces réactions implique l'introduction dans le modèle de nombreux paramètres cinétiques (38 chez [154], 58 chez [319]) dont l'évaluation des valeurs n'est pas évidente. Lorsque l'ensemble des paramètres cinétiques n'est pas facilement évaluable, des méthodes mathématiques proposent de réduire la complexité des modèles en agrégeant certains paramètres grâce à des contraintes, en supprimant des variables [73] ou en exploitant les corrélations entre métabolites [211] tout en respectant certaines propriétés cinétiques. D'autres modèles se sont plutôt concentrés sur la régulation de quelques éléments (par exemple, la régulation génétique de l'opéron lactose chez *Escherichia coli* [274]). A notre connaissance, aucun modèle dynamique n'avait tenté de décrire la hiérarchie d'importance des voies du métabolisme des lipides dans la variation de la quantité et de la composition des tissus en acides gras.

Choix d'une modélisation dynamique Rendre compte de l'aspect dynamique du métabolisme des lipides et des acides gras est particulièrement important d'un point de vue biologique puisque les priorités dans les synthèses ou dans l'utilisation des différentes réserves énergétiques (i.e. glucides, lipides et protéines) au niveau de l'organisme vont varier en fonction de la balance énergétique. Cette balance est par exemple dépendante de l'alimentation (quantité, temporalité, types de nutriments...), de l'intensité d'exercice physique, etc, et donc aussi du temps. En outre, les voies de régulation du métabolisme des lipides, par leur nature diverse (modulation de l'expression de gènes, de l'activité des enzymes, des cinétiques de réactions, des translocations des transporteurs de substrat...) interviennent à différentes échelles de temps (de la minute à la journée). Pour être validé empiriquement, un modèle dynamique doit être confronté à des données dynamiques. En ce qui concerne le métabolisme des lipides et plus généralement le métabolisme de l'énergie, des jeux de données décrivant la variation de quantités de métabolites au cours du temps dans diverses conditions expérimentales, dans différents organes et dans différentes espèces existent aujourd'hui dans la littérature. Ainsi l'évaluation de modèles dynamiques relatifs au métabolisme des lipides devrait être *a priori* réalisable.

Généricité A première vue, la diversité dans les comportements observés, les contextes expérimentaux (ex : organe, espèce) et dans les conditions mises en œuvre inciterait à réaliser un modèle spécifique à chaque cas. Ainsi on profiterait des hypothèses propres à chaque cas sans subir de contraintes extérieures, ce qui permettrait un développement relativement rapide de chacun des modèles. Cependant, une telle méthode ne permettra pas d'exploiter un ensemble de cas proches. A l'inverse, penser un modèle en vue de sa réutilisation ultérieure dans différents cas (organe, espèce, condition expérimentale ...) permettra de recycler à moindre frais les réalisations des travaux précédents, et surtout de réaliser des comparaisons entre ces différents cas.

En contrepartie, le développement d'un modèle générique implique de fixer une partie constante entre modèles (donc indépendante des hypothèses spécifiques à chaque cas) et de prévoir une stratégie d'évolution du modèle lors de l'intégration d'informations qui n'étaient pas connues (ou pas exploitées) lors de la conception de la partie constante.

Simplicité La modélisation sous forme d'équations différentielles ordinaires de tous les mécanismes biochimiques décrits dans la littérature implique l'introduction dans le modèle d'un grand nombre de paramètres cinétiques, comme par exemple la constante de Michaelis-Menten (Km) et la vitesse maximum d'une réaction enzymatique (Vm), la quantité d'enzymes, les dépendances des fonctions vis-à-vis de leurs régulateurs ... Bien que certains d'entre eux soit connus *in-vitro* (Km et Vm des enzymes par exemple), une grande partie des informations nécessaires reste inconnue ou dépendante de chaque contexte expérimental (quantité d'enzymes par exemple). Les données présentes dans la littérature ne permettent pas de contraindre suffisamment un modèle ainsi construit pour obtenir un ajustement unique de tous ses paramètres ; le nombre de degré de liberté introduit par les paramètres permettrait alors d'ajuster à peu près n'importe quelles données expérimentales, ce qui remet en cause la pertinence de la validation empirique du modèle. De plus, certaines lois de réponses d'un mécanisme ne sont pas connues, comme par exemple la dépendance d'une vitesse de réaction en fonction de plusieurs régulateurs ou des régulations mettant en œuvre plusieurs mécanismes (ex : régulation d'ACC [181]).

Le niveau de détail des mécanismes n'est donc pas adapté à la modélisation des données cinétiques du métabolisme des lipides et reste incompatible avec l'objectif de compréhension que nous nous sommes fixés. De plus, les fortes interconnections entre les éléments du métabolisme de l'énergie (c.f. figure 1.1.1) rendent difficile la définition *a priori* d'une frontière d'étude des réactions biochimiques (par exemple, où s'arrête le métabolisme des lipides et où commence celui des protéines?).

La recherche d'un modèle le plus simple possible et cohérent avec les observations expérimentales permet de limiter grandement le nombre de paramètres introduits en se limitant à décrire les éléments les plus importants pour le contexte étudié. Cette démarche résout ainsi les problèmes de sur-paramétrisation tout en remplissant l'objectif de compréhension.

Objectif et stratégie Afin d'identifier la hiérarchie d'importance des voies métaboliques dans la variation de la quantité d'acides gras tissulaires, et de pouvoir ultérieurement réaliser des comparaisons entre conditions, notre travail a consisté à élaborer une stratégie de modélisation simple et générique. Afin de satisfaire les objectifs de simplicité, nous avons proposé des modèles ne décrivant que le nécessaire pour rester cohérent avec les données observées, tout en ignorant parcimonieusement toute complexité superflue. Afin de respecter la généralité, nous avons développé une méthode ascendante de construction de modèles dans laquelle chaque ancien modèle est un cas particulier du nouveau modèle. Cette méthode permet notamment de comparer entre eux les paramètres des différents modèles construits. En pratique un nouveau modèle est obtenu à partir d'un ancien modèle par simplification ou par ajout d'un élément (séparation d'un paramètre en deux, rajout d'un flux ou utilisation d'un cas limite).

L'application de cette démarche a débouché sur la construction d'un squelette de modèle valide *a priori* quelque soit l'espèce et l'organe étudiés. Ce squelette sert de base à la création de tout nouveau modèle du métabolisme des lipides. Il a alors été dérivé afin de réaliser un modèle dynamique simple décrivant la composition hépatique en acides gras chez les souris de deux génotypes différents. Grâce à cette démarche, nous avons pu mettre en évidence les voies biochimiques suffisantes et nécessaires pour décrire les variations de quantité et de composition des acides gras dans ce contexte. Conformément à notre méthode de construc-

tion ascendante, les modèles obtenus dans les deux géotypes partageaient des paramètres ayant la même signification biologique. Leur comparaison nous a permis d'identifier les principales différences de comportements entre ces deux géotypes et notamment de proposer des pistes de travaux futurs sur la régulation des désaturases.

Chapitre 1

Construction d'un squelette générique des principales voies du métabolisme des acides gras

Le schéma général de construction du modèle et son squelette générique ont été présentés aux Journées Ouvertes de Biologie, Informatique et Mathématique (JOBIM, Lille, 30 juin - 2 juillet 2008). Nous en présentons ici les tenants et aboutissants.

Pour réaliser un squelette de modèle métabolique valide quels que soient l'organe, l'espèce et les conditions expérimentales, nous avons fait appel à un panel d'experts afin de proposer un ensemble de comportements biologiques généraux représentatifs du métabolisme des lipides. Ensuite nous avons recherché un modèle le plus simple possible basé sur les voies et les régulations biologiques existantes, et qui respecte cet ensemble de comportements.

1.1 Identification des fonctions biochimiques et des régulations nécessaires pour respecter une liste de comportements

Identifier *a priori* des éléments par expertise pose la question de la définition du niveau de détail à retenir (peut-on représenter la glycolyse par une seule fonction, ou faut-il détailler tous ces enzymes?) et de la quantité d'informations à intégrer (faut-il inclure le métabolisme des acides aminés dans celui des lipides?). Il est donc difficile de définir *a priori* ce que l'on entend par "élément important".

Pour résoudre ce problème, un ensemble de fonctions candidates a d'abord été identifié par des experts à partir de la bibliographie. Cet ensemble de fonctions a été ensuite réduit ou enrichi jusqu'à obtenir un ensemble minimal qui respecte une liste de comportements biologiques bien connus. On définit ainsi comme important l'ensemble des éléments nécessaires et suffisants pour satisfaire les comportements.

1.1.1 Identification des fonctions et des régulations candidates par expertise

Parmi les voies biochimiques décrites dans KEGG, [231] et dans les manuels de biochimie [271, 231], les experts ont retenu les voies généralement admises pour influencer de façon majeure les quantités d'acides gras.

A priori, et comme présenté dans notre introduction bibliographique (c.f. I), les fonctions importantes pour synthétiser et dégrader des acides gras peuvent être catégorisées en :

la glycolyse qui transforme en anaérobie le glucose en acétyl-coenzyme A, produit du NADH et utilise et produit de l'ATP ;

le **cycle de Krebs** qui utilise l'acétyl-coenzyme A pour produire de l'ATP, du FADH₂ et du NADH en aérobie ;

la **chaîne respiratoire** qui utilise en aérobie le NADH et le FADH₂ pour produire de l'ATP ;

la **cétogenèse** qui transforme l'excès d'acétyl-coenzyme A en corps cétoniques ;

le **cycle des pentoses** (ou *shunt* des hexoses monophosphates) qui convertit le glucose-6-phosphate en dérivés à 5 carbones, l'énergie dégagée étant stockée sous forme de NADPH¹ ;

l'**oxydation des acides gras** qui dégrade les acides gras, afin de produire de l'ATP et de l'acétyl-coenzyme A ;

la **néosynthèse d'acides gras** qui consomme du NADPH et de l'acétyl-coenzyme A pour produire des acides gras ;

l'**importation/exportation des acides gras dans/hors de la cellule** .

Lors de cette catégorisation, le choix a été fait de séparer les fonctions autour de pivots métaboliques, c'est-à-dire des éléments pouvant avoir plusieurs devenir possibles dans le métabolisme. Ainsi dans notre cas, la glycolyse se poursuit jusqu'à l'acétyl-coenzyme A et non jusqu'au pyruvate comme classiquement décrit, car l'acétyl-coenzyme A a plusieurs devenir : la cétogenèse, la synthèse d'acides gras ou le cycle de Krebs. De la même manière, la synthèse d'acides gras commence avec l'exportation de l'acétyl-coenzyme A mitochondrial vers le cytosol tandis que le pyruvate ne peut (d'après nos hypothèses) que se transformer en acétyl-coenzyme A.

Les régulations transcriptionnelles importantes pour ces fonctions sont notamment l'activation de la transcription des enzymes clefs de la synthèse d'acides gras (FASN et ACC) par SREBP1 [121] et l'activation de nombreux enzymes de l'oxydation des acides gras par PPAR α [201]. Lors de cette étape, il faut garder à l'esprit que le modèle construit ne comportera que deux variables régulatrices : l'une se comportant comme SREBP1, et l'autre comme PPAR α . Dans la réalité ces variables peuvent en fait représenter tout un ensemble de régulateurs corrélés agissant en synergie comme par exemple SREBP1, ChREBP et l'insuline.

Les aspects relatifs au métabolisme anaérobie strict (lactate), au stockage des lipides (estérification des acides gras en triglycérides), au transport intracellulaire des métabolites (glucose, acides gras, citrate ...), ainsi que les connections des voies biochimiques décrites dans le modèle avec le reste du métabolisme énergétique n'ont pas été retenus. Ils pourront néanmoins être rajoutés *a posteriori* s'ils s'avèrent nécessaires pour ajuster un jeu de données particulier.

1.1.2 Identifications d'une liste de comportements par expertise

Nous nous sommes focalisés sur les comportements relatifs à la variation du métabolisme des lipides en fonction de la disponibilité en glucose et qui correspondent à des réponses observées dans la majorité des cellules animales. De plus, des comportements témoignant d'une régulation transcriptionnelle de ces voies ont été rajoutés afin de pouvoir prendre en compte dans le modèle des effets et régulations à plus long terme.

Les comportements retenus sont les suivants :

- Les cellules ont besoin de consommer de l'ATP afin d'assurer *a minima* leur métabolisme basal.
- L'ATP ne s'accumule pas.
- Lorsque la glycémie est élevée, la plupart des cellules oxydent le glucose plutôt que les acides gras.
- La glycolyse n'est pas active lorsque la cellule n'a ni besoin d'acétyl-coenzyme A ni d'ATP. ²
- La synthèse de novo d'acides gras est active uniquement lorsque les besoins énergétiques du métabolisme basal ont été satisfaits.

¹Le NADPH est un élément important de la synthèse d'acides gras et de stéroïdes

²En effet, l'insuline, une hormone dont la quantité est fortement corrélée avec la glycémie, favorise les transporteurs au glucose SLC2A4 en particulier dans les adipocytes et les fibres musculaires. De plus l'hyperglycémie seule majore le captage du glucose par SLC2A1 (GLUT1) ou SLC2A2 (GLUT2) qui présentent une constante de Michaelis-Menten élevée pour le glucose, puis son oxydation, sans doute par l'activation de la PDH [212]. Dans de futurs modèles, une oxydation préférentielle des acides gras au dépend du glucose pourra être introduite dans des cellules telles que les fibres oxydatives à vitesse de contraction lentes lors d'un exercice prolongé (théorie du *cross-over point* décrivant la compétition entre dégradation de glucose et oxydation acides gras dans la production d'ATP [262]).

- La synthèse de novo d'acides gras est active dans les organes capables de la réaliser (foie, muscle, tissu adipeux) uniquement si l'acétyl-coenzyme A produit par la dégradation des glucides est présent en quantité suffisante. Ceci implique généralement que le glucose soit présent en quantité suffisante. ³
- La synthèse des enzymes impliquées dans la néo-synthèse des acides gras (ACC et FASN) a lieu uniquement lorsque la glycémie est élevée [180, 179] ⁴.
- L'oxydation des acides gras n'a lieu que si la glycémie est faible. ⁵.
- La cétogenèse a lieu uniquement s'il y a oxydation des acides gras. Ceci implique donc (voir comportement précédent) que la cétogenèse n'a lieu que si la glycémie est faible. ⁶.
- La synthèse et l'oxydation des acides gras n'ont pas lieu simultanément. ⁷
- Les enzymes de l'oxydation des acides gras et de la cétogenèse sont synthétisés lorsque la glycémie est faible ⁸.
- PPAR α est synthétisé lors du jeûne tandis que SREBP1 est synthétisé à l'état nourri.

1.1.3 Réduction des fonctions candidates au minimum nécessaire pour respecter les comportements

Méthode d'obtention d'un modèle minimal respectant les comportements

Afin de garantir un modèle respectant les comportements, nous écrivons un premier modèle qualitatif de connaissances comprenant toutes les fonctions biologiques et les régulations identifiées par les experts. Nous vérifions alors que les comportements sont bien respectés dans tous les scénarios envisagés. Si ce n'est pas le cas, nous retournons à l'étape d'expertise et rajoutons une fonction, un intermédiaire dans une fonction existante ou une régulation. Nous vérifions alors ce nouveau modèle et réitérons ainsi l'étape de complexification jusqu'à respecter tous les comportements. Cette démarche est schématisée figure 1.1.1.A. Au final nous obtenons un modèle suffisant pour respecter les comportements.

A partir de ce modèle, nous recherchons si cet ensemble peut être simplifié : pour chacun de ses éléments (métabolite, régulation, fonction biologique), nous écrivons un sous modèle dans lequel cet élément n'est pas présent. La technique de simplification est présentée figure 1.1.2.

Cette technique garantit d'avoir un modèle dans lequel tous les éléments sont nécessaires et suffisants, mais ne garantit pas l'unicité du modèle obtenu. Afin de maximiser nos chances d'avoir un modèle minimal proche de la réalité, les simplifications ont d'abord été faites à partir d'hypothèses biologiques raisonnables, puis à défaut à partir d'hypothèses parcimonieuses.

³Nous supposons pour ce comportement que la régulation du métabolisme est telle que la voie la plus économe en énergie est toujours choisie par rapport à d'autres voies moins efficaces. Par exemple, il est plus coûteux en énergie de synthétiser des acides gras à partir du glucose puis de les oxyder pour produire de l'ATP que d'oxyder directement le glucose dans le cycle de Krebs.

⁴La synthèse de ces enzymes est fortement stimulée par SREBP1, un facteur de transcription dont l'expression est corrélée avec la glycémie. SREBP1 est à la fois activé par le glucose et l'insuline et inhibé par les acides gras poly-insaturés.

⁵Certains organes ne peuvent pas survivre sans dégrader du glucose même en cas de raréfaction de ce dernier (jeûne, régime pauvre en glucides). Dans d'autres organes à forte capacité oxydative, on observe alors une réorientation du métabolisme vers l'oxydation d'autres substrats énergétiques tels que les acides gras. A l'inverse en cas de forte glycémie, tous les organes oxydent préférentiellement le glucose dont le rendement énergétique est bien meilleurs en cas d'oxydation directe plutôt qu'après sa transformation en acides gras

⁶Lors d'un jeûne prolongé, l'oxydation des acides gras produit une grande quantité d'acétyl-coenzyme A. Dans le foie, cet acétyl-coenzyme A n'est que très peu dégradé par le cycle de Krebs car ce dernier est bloqué ou ralenti par la mise en place de la néoglucogenèse et car la β -oxydation des acides gras apporte déjà au foie une grande quantité d'ATP. L'acétyl-coenzyme A est alors transformé en corps cétoniques qui sont exportés dans le sang puis captés par de nombreux autres organes ; ils sont alors oxydés dans le cycle de Krebs, remplaçant ainsi l'acétyl-coenzyme A issu de la glycolyse.

⁷Cette hypothèse a été posée *a priori*, en supposant que la cellule ne réalise pas de cycle futile. Garantir ce comportement dans le modèle nous a obligé par la suite à rajouter à notre squelette initial l'inhibition allostérique de CPT1 (une enzyme responsable de l'entrée des acides gras longs dans la mitochondrie et qui constitue une étape limitante de la β -oxydation) par le malonyl-coenzyme A, un intermédiaire réactionnel de la néosynthèse d'acides gras.

⁸Cette régulation est assurée par PPAR α (c.f. 3.2.3)

Hypothèses biologiques

Nous avons tout d'abord supposé que le métabolisme était aérobie. Cette hypothèse est sans doute vraie pour le foie et les tissus adipeux, mais ne rend pas compte de la réalité dans le muscle. En effet, même si les muscles représentent globalement avec le foie les principaux tissus oxydant les acides gras, les fibres glycolytiques présentes en majorité dans les muscles à vitesse de contraction rapide ont un métabolisme basé principalement sur la transformation anaérobie du glucose en lactate. Cette hypothèse sera donc une des premières à lever si le modèle doit rendre compte de données sur le métabolisme musculaire, en particulier dans le cadre d'un effort ponctuel et intensif. Dans les autres cas, cette hypothèse nous a permis d'éliminer le métabolisme du lactate, de retirer les intermédiaires NADH, et FADH₂ et de fusionner la chaîne respiratoire avec la glycolyse et le cycle de Krebs.

Nous avons aussi fait l'hypothèse que la variation de l'exportation des sucres à 5 carbones par le cycle des pentoses n'avait pas une influence importante sur le métabolisme des lipides. Nous avons ainsi pu réduire le cycle des pentoses à une fonction unique produisant du NADPH à partir du glucose. De plus en considérant que tout le NADPH est produit d'une manière ou d'une autre (cycle des pentoses ou navette à malate) à partir du glucose, nous avons pu considérer la synthèse des acides gras comme une fonction à 3 paramètres avec le glucose en entrée, les acides gras en sortie et le malonyl-coenzyme A comme seul intermédiaire à conserver pour respecter la non simultanéité de la synthèse et de l'oxydation des acides gras.

1.2 Mise en équation du modèle obtenu

Une fois le modèle simplifié, nous écrivons un ensemble d'équations le plus simple possible et garantissant la liste de comportements observés afin de rendre compte des flux métaboliques. Lors de cette étape, un maximum d'hypothèses parcimonieuses a été appliqué afin de garantir la simplicité. En plus de devoir respecter ces critères, le modèle construit doit pouvoir évoluer afin de s'adapter à de nouvelles données. Cela implique de développer une stratégie d'écriture spécifique afin de respecter la règle "tout ancien modèle doit être un cas particulier du nouveau modèle".

1.2.1 Hypothèses parcimonieuses

Les hypothèses parcimonieuses sont les suivantes :

- Nous considérons que les fonctions biologiques se comportent comme une unique réaction. Si cette fonction possède une étape limitante unique et clairement identifiée, cette hypothèse est validée ; dans tous les autres cas, cette hypothèse est une approximation plus ou moins grossière de la réalité. Lever cette hypothèse nécessite de disposer de données expérimentalement fines rendant compte de la réponse de la fonction biologique à ses paramètres d'entrée. En pratique, ceci ne peut se faire qu'avec des jeux de données cinétiques très précis qui couvrent des plages de valeurs permettant d'observer plus d'une étape limitante.
- Une réaction (ou une fonction biologique), lorsqu'elle a lieu, est soit considérée comme linéaire en fonction de la quantité de substrat (cas de l'enzyme en excès), soit comme constante (cas de la présence de substrat en excès avec une activité enzymatique constante). Le choix entre ces deux hypothèses a été guidé par la bibliographie ou les données à expliquer.
- Si et seulement si cette modélisation n'est pas suffisante pour rendre compte des comportements biologiques ou des données expérimentales à expliquer, les réactions (ou les fonctions biologiques) sont modélisées par une cinétique de Michaelis-Menten. Cette modélisation constitue généralement une bonne approximation des données observées, y compris lorsque la réaction sous jacente est plus complexe (fonction de Hills, réaction non unique ...). Réfuter cette hypothèse nécessite là aussi soit des données concernant une plage de valeurs mettant en évidence plus d'une étape limitante, soit des données mettant en évidence un phénomène de régulation, soit des données concernant des intermédiaires réactionnels.
- Si un élément a une influence sur la vitesse d'une réaction, nous considérons que son effet est constant pour une condition expérimentale donnée.

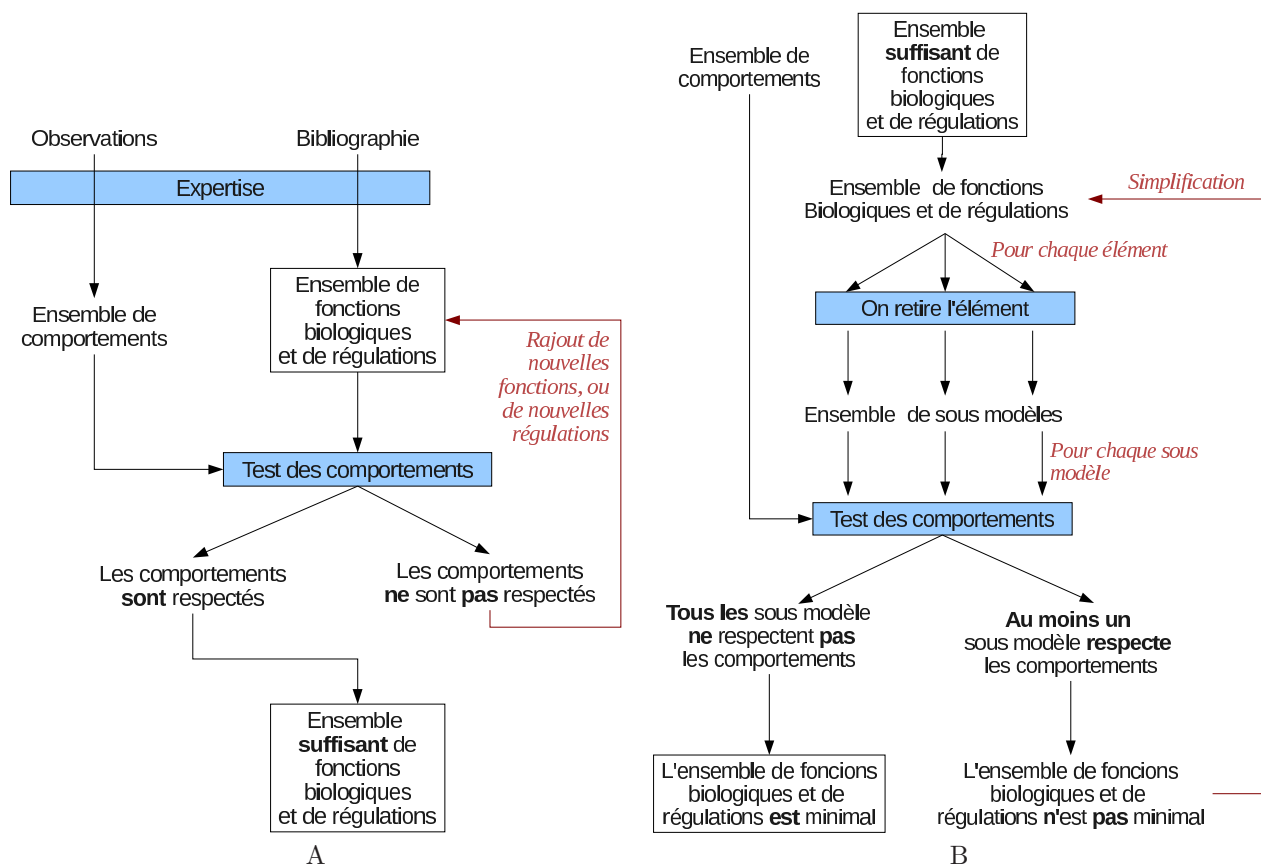


FIG. 1.1.1 – Démarche de construction d'un modèle minimal. Dans une première étape (A), on construit un modèle dont le contenu garantit le respect des comportements. Une fois cette étape réalisée, ce modèle est simplifié (B). On obtient alors un modèle minimal dans lequel la perte de n'importe quel élément (simplification ou retrait d'une fonction biologique, retrait d'une régulation) entraîne le non respect d'au moins un comportement.

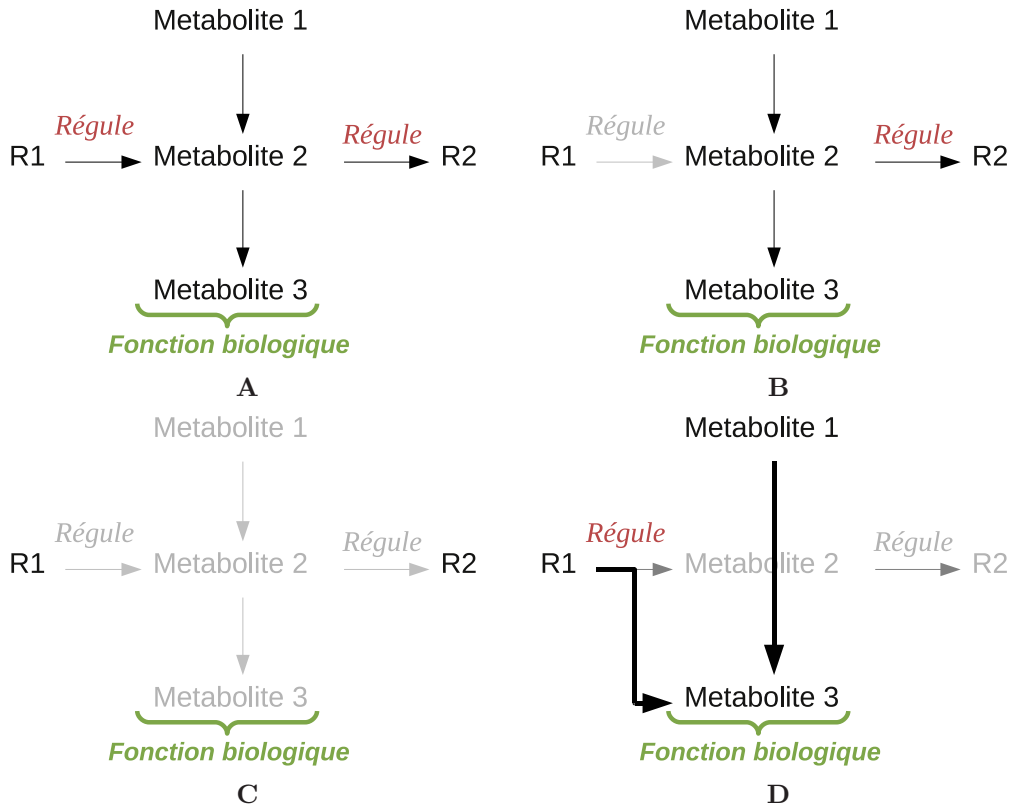


FIG. 1.1.2 – Méthode de simplification d'un ensemble de fonctions biologiques régulées. Considérons une fonction biologique (**A**) composée de 3 métabolites dont le métabolite intermédiaire (métabolite 2) est régulé par R1 et régule R2. R1, R2, le métabolite 1 et le métabolite 3 peuvent être éventuellement connectés au reste du modèle (non représenté). Plusieurs simplifications de cette fonction sont possibles :

- 1) retirer une régulation (en gris, (**B**)) en considérant que la variation de sa cible est indépendante de celle de son régulateur (soit car il est constant, soit car il n'a pas d'influence). Il suffit alors d'effacer la flèche correspondante.
- 2) retirer une fonction biologique complète (**C**), en considérant que la variation des éléments de la fonction biologique n'a aucune influence sur le reste du système. Il suffit d'effacer cette dernière, ainsi que les flèches de régulations pointant à partir ou vers rien.
- 3) retirer un intermédiaire dans une fonction (**D**) en considérant que l'intermédiaire ne régule aucun élément en dehors de la fonction. Cela revient à considérer que la connaissance du métabolite intermédiaire n'est pas nécessaire pour décrire l'état des autres éléments du système. On relie alors ses précurseurs à ses produits (flèches en gras) et on supprime cet intermédiaire du graphe (en gris).

Ces hypothèses parcimonieuses sont considérées comme de bonnes approximations tant qu'aucun élément ne nous incite à les rejeter.

1.2.2 Le modèle de connaissance générique et minimal obtenu

Le modèle de connaissances obtenu est présenté figure 1.2.1. Ce modèle présente une liste minimale de fonctions biologiques et de leurs régulations afin de respecter les comportements mis en évidence par les experts.

De manière triviale, le modèle doit au moins comporter l'acétyl-coenzyme A (i.e. le pivot métabolique central), le glucose, les acides gras, les cétones, l'ATP, SREBP1, PPAR α et les voies de la glycolyse, de la synthèse et d'oxydation des acides gras, la cétogenèse et les voies de synthèse des enzymes de l'oxydation et de la synthèse des acides gras car ce sont les éléments décrits dans les comportements.

FIG. 1.2.1 – **Modèle de connaissances minimal et générique du métabolisme des lipides cohérent avec les comportements identifiés par les experts.**

Le modèle décrit un ensemble de voies biochimiques nécessaires et suffisantes pour respecter les comportements identifiés par les experts (c.f. sous section 1.1.2 pour la liste).

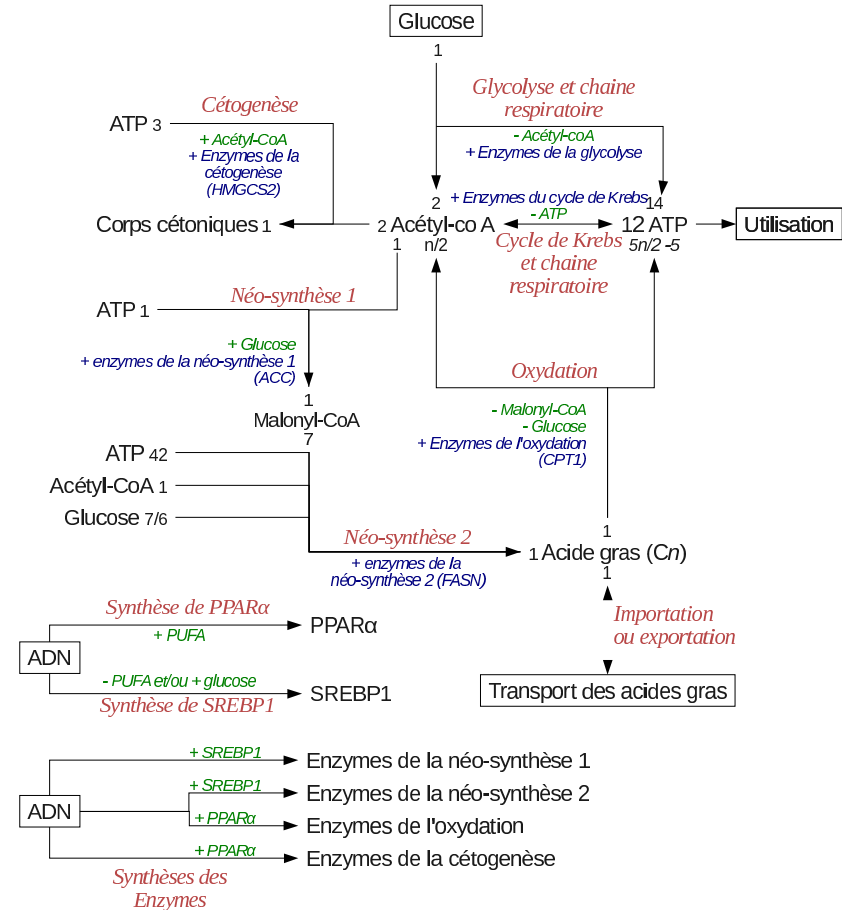
Les voies biochimiques sont représentées par des flèches **noires**. A ces voies s'ajoutent la dégradation des facteurs de transcription et des enzymes (non représentés) qui sont des réactions chimiques d'ordre 1 en fonction de l'élément dégradé. Les activateurs (+) et les inhibiteurs (-) des voies métaboliques sont représentés en **vert**. Lorsque les substrats ne sont pas représentés parmi les activateurs des voies, nous faisons l'hypothèse parcimonieuse que les substrats ne sont pas limitants tant qu'ils sont présents en quantité non nulle.

Les coefficients stoechiométriques des réactions sont représentés par des nombres, ou par n qui représente le nombre de carbones de l'acide gras qui subit l'oxydation. Le coefficient de l'ATP (consommé sous forme de NADH) de la cétogenèse est approximatif et concerne le cas où l'acétoacétate est réduit en β -hydroxy butyrate. Ce dernier peut aussi se transformer en acétone + CO_2 , ce qui ne consomme pas de NADH. La connaissance de ces coefficients n'est pas importante pour respecter les comportements définis par les experts (car ils sont qualitatifs), mais est utile en aval, afin d'écrire les équations.

Les cadres représentent les variables libres du modèle. Leur comportement doit être décrit par une fonction mathématique dépendante du temps.

SREBP1 est à la fois activé par l'insuline (corrélée au glucose) et inhibé par les acides gras poly-insaturés (PUFA). Or les quantités de glucose et de PUFA varient généralement de manière opposée, ce qui empêche de distinguer leurs effets relatifs. Nous choisissons donc de modéliser cette régulation conjointe par une unique variable que nous choisissons d'indexer sur la quantité de acides gras poly-insaturés.

Afin d'assurer la généralité, nous introduisons pour chaque fonction du modèle une dépendance vis-à-vis de sa quantité d'enzymes (en **bleu**). Lorsque l'enzyme limitant est connu, il est noté entre parenthèses.



Frontières du modèle L'importation et l'exportation des acides gras sont des fonctions fortement dépendantes de la situation extra cellulaire (i.e. des acides gras libres circulants), et de l'état intra cellulaire (le type et la quantité des différents transporteurs ou les quantités d'acides gras cytoplasmiques par exemple). Dans un premier temps, nous considérerons parcimonieusement que l'importation et l'exportation sont uniquement régulés par le milieu extérieur du modèle et nous les décrirons donc par des fonctions mathématiques uniquement dépendantes du temps. De la même manière, nous considérerons que le glucose est une fonction directement dépendante de la glycémie.

Nous considérons la consommation d'ATP comme une fonction qui ne dépend pas de l'état de la cellule. Nous supposons par conséquent que cette dernière satisfait son métabolisme basal (supposé constant) ou répond à un stimulus extérieur. Le métabolisme basal est modélisé par une consommation constante d'ATP. Si la quantité d'ATP disponible est insuffisante pour satisfaire cette demande, nous considérons alors que la cellule est morte. La réponse à un stimulus extérieur, comme par exemple un effort physique variable ou l'avancement de la croissance chez un jeune animal, est modélisée par une consommation d'ATP dépendante du temps.

Glycolyse et chaîne respiratoire L'inhibition de la glycolyse par l'acétyl-coenzyme A suffit à respecter les comportements dans les cellules en aérobie. En effet, lorsque l'acétyl-coenzyme A est produit plus vite par la glycolyse qu'il n'est consommé (notamment par le cycle de Krebs pour produire de l'ATP), il s'accumule et finit par inhiber la glycolyse. Si dans le futur nous souhaitons par exemple introduire le métabolisme anaérobie, pour lequel la glycolyse est capable de fournir de l'ATP sans produire d'acétyl-coenzyme A, il sera nécessaire de réguler la glycolyse aussi par l'ATP comme largement observé *in-vivo*.

Cycle de Krebs et chaîne respiratoire Dans un modèle aérobie, l'ATP peut être produit par la glycolyse ou par le cycle de Krebs. La régulation du cycle de Krebs par l'ATP suffit à respecter le comportement de la non accumulation d'ATP.

En effet, dans ce modèle, lorsque le cycle de Krebs est inhibé, la glycolyse ne peut à elle seule fournir de grandes quantités d'ATP, car il faut évacuer l'acétyl-coenzyme A produit. Ce dernier ne peut pas être utilisé pour la cétogenèse car cette dernière n'a pas lieu à partir de l'acétyl-coenzyme A issu de la glycolyse, et s'il est utilisé pour produire des lipides, le bilan en ATP est négatif.

Cette inhibition du cycle de Krebs par l'ATP est cohérente avec les connaissances biologiques puisque la citrate synthase (CS) et l' α cétoglutarate deshydrogénase (OGDH) sont inhibées par l'ATP. De plus le NADH inhibe fortement la PDH, l'isocitrate deshydrogénase (IDH1), OGDH et CS. Or le NADH s'accumule généralement lorsque la chaîne respiratoire tourne au ralenti, c'est-à-dire lorsque cette dernière est inhibée par l'ATP (ou par le manque de dioxygène).

Dans ce modèle il aurait été possible d'introduire d'autres régulations pour garantir le comportement de la non accumulation de l'ATP, comme par exemple une quantité fixée d'ADP. Un tel choix n'a pas été retenu car il introduit une molécule supplémentaire ce qui complique un peu le modèle et car il induit une inhibition de l'oxydation des acides gras et de la glycolyse lorsque la concentration d'ADP est nulle.

Oxydation des acides gras Lorsque l'individu est nourri, il oxyde préférentiellement le glucose ce qui implique d'inhiber l'oxydation quand l'individu est nourri. De plus, l'oxydation et la néo-synthèse des acides gras doivent s'exclure mutuellement.

Nous savons tout d'abord que les enzymes de l'oxydation sont peu synthétisés lorsque l'individu est nourri, en particulier car PPAR α est présent à son niveau basal, mais cette régulation nous semble insuffisante car elle implique une demi-vie très courte à la fois pour les enzymes de l'oxydation et pour les éléments de la voie de signalisation de PPAR α (hypothèse dont nous ignorons la véracité).

Nous savons de plus, que lorsque l'individu est nourri, de nombreux organes synthétisent des acides gras et que le malonyl-coenzyme A (un intermédiaire réactionnel de cette synthèse) est un très bon inhibiteur de CPT1 [218], l'étape limitante de la β -oxydation. Introduire cette régulation suffit alors pour exclure synthèse et dégradation de lipides, et pour inhiber l'oxydation dans les cellules capables de synthétiser des lipides.

Dans les cellules qui ne synthétisent pas de lipides, cette inhibition n'est pas suffisante pour inhiber l'oxydation des acides gras à l'état nourri, il faudrait en plus rajouter une inhibition par le glucose (ou un élément qui lui est corrélé). Les cellules étudiées (foie, muscle, adipocyte) sont capables à des degrés divers de synthétiser des lipides en fonction de l'espèce étudiée. Pour rester dans un cadre générique, il est donc nécessaire de conserver l'inhibition de l'oxydation des acides gras par le glucose. Cette dernière est très probablement réalisée *in-vivo* par l'insuline qui inhibe l'activité et la transcription CPT1 et augmente sa sensibilité au malonyl-coenzyme A [247].

La régulation de CPT1 par le glucose, redondante avec celle par PPAR α , implique que nous pouvons faire l'hypothèse parcimonieuse que la quantité d'enzymes de l'oxydation est constante jusqu'à preuve du contraire.

Cétogenèse La cétogenèse a lieu uniquement s'il y a oxydation des acides gras. Un mécanisme simple consiste à n'activer cette dernière que lorsque les quantités d'acétyl-coenzyme A sont élevées, ce qui est justement le cas lorsque l'oxydation des acides gras a lieu.

Néo-synthèse des acides gras L'introduction de la régulation de l'oxydation des acides gras par le malonyl-coenzyme A nous oblige à découper la synthèse d'acides gras en deux : synthèse1 (synthèse de malonyl-coenzyme A par ACC) et synthèse2 (synthèse d'acides gras par FASN).

La synthèse de novo d'acides gras est active uniquement lorsque les besoins énergétiques du métabolisme basal ont été satisfaits. Une première idée afin de satisfaire cette étape est de considérer la régulation de la synthèse de FASN et d'ACC par SREBP1. Cette régulation est certes capable d'induire la synthèse lors de la transition à jeun/nourri, mais son inhibition lors du jeûne implique aussi, comme pour l'oxydation, d'attribuer une demi-vie faible pour FASN ou ACC et pour les intermédiaires de la voie de signalisation de SREBP1 ce qui est faux (par exemple FASN a une demi-vie de 32 heures).

Dans la mesure où les variations de la quantité d'enzymes induites par les variations SREBP1 sont insuffisantes pour expliquer la diminution de la néo-synthèse des acides gras lors du jeûne, nous recherchons une régulation supplémentaire. Nous savons que l'insuline régule fortement ACC [339], et ce à des échelles de temps bien plus courtes que l'effet de SREBP1. Par conséquent nous rajoutons dans le modèle la régulation d'ACC par le glucose, une variable très corrélée à l'insuline.

Cette nouvelle régulation à court terme suffit, indépendamment de l'effet de SREBP1 sur les quantités d'enzymes, à satisfaire l'arrêt de la néo-synthèse des acides gras lors du jeûne. Par conséquent nous pouvons considérer *a priori* les quantités d'enzymes FASN et ACC comme constantes. Dans le cas où des données expérimentales nous permettent d'accéder à ces deux enzymes ou si elles concernent des durées suffisamment longues pour que l'effet des variations de quantités d'enzymes soit perceptible, nous lèverons cette hypothèse.

Synthèse et dégradation des facteurs de transcription PPAR α est connu pour être activé par les acides gras poly-insaturés, ce qui suffit à expliquer son comportement à jeun.

SREBP1 est à la fois activé par l'insuline et inhibé par les acides gras poly-insaturés. Or les quantités d'insuline et de PUFA varient généralement de manière opposée, ce qui empêche de distinguer leurs effets relatifs. Nous choisissons donc de modéliser cette régulation conjointe par une unique variable que nous choisissons d'indexer sur la quantité d'acides gras poly-insaturés.

Pour que SREBP1 et PPAR α ne s'accumulent pas indéfiniment et soient dégradés, nous faisons l'hypothèse que la production des facteurs de transcription est compensée par leur dégradation. Pour modéliser cette hypothèse nous faisons l'hypothèse que les quantités de SREBP1 et de PPAR α sont proportionnelles aux acides gras poly-insaturés.

Synthèse d'enzymes La synthèse des enzymes de la néo-synthèse des acides gras est connue pour être fortement activée par SREBP1 (et d'autres facteurs corrélés à ce dernier), tandis que la synthèse des enzymes de l'oxydation et de la cétogenèse est connue pour être activée par PPAR α . Ces deux régulations sont suffisantes pour expliquer les comportements.

Dans les modèles que nous construisons, nous faisons l'hypothèse que les quantités d'enzymes sont *a priori* constantes, et nous ne considérons leurs variations que si les données nous permettent de les prendre en compte. Dans ce cas, soit nous supposons qu'ils sont proportionnels à leurs facteurs de transcription, soit nous introduisons leur synthèse et leur dégradation.

Généricité du modèle construit Les éléments précédemment introduits sont suffisants pour respecter les différents comportements, mais ne permettent pas *a priori* au modèle d'avoir la souplesse nécessaire pour s'adapter aux divers organes et diverses espèces. Il est par exemple difficile de trouver une dépendance de la cétogénèse vis-à-vis de l'acétyl-coenzyme A qui fasse que cette dernière est active dans le foie et pas dans les autres organes. Par conséquent (et conformément aux connaissances biologiques), nous rajoutons pour chaque fonction un paramètre décrivant la quantité et l'efficacité des enzymes de cette fonction. Nous pourrions par la suite décider par parcimonie de choisir des quantités d'enzymes égales jusqu'à preuve du contraire entre les divers organes et les diverses espèces.

1.3 Evolution du modèle par construction ascendante et modification des équations.

Lors de la construction d'un modèle, il est impossible d'écrire dès le départ un modèle qui intègre toute la connaissance existante. De plus l'acquisition de nouvelles données et de nouvelles connaissances ou la remise en cause des hypothèses ayant permis de construire le modèle peuvent amener ce dernier à changer. Il est donc nécessaire de proposer dès le début une stratégie de modification du modèle. Cette stratégie doit respecter au maximum la compatibilité ascendante des modèles, c'est-à-dire que chaque ancien modèle doit être un cas particulier du nouveau modèle. Dans le cas d'un modèle dynamique, cela revient à dire que l'ancien modèle est un nouveau modèle avec des valeurs particulières de ses paramètres ou de ses variables. Assurer la compatibilité ascendante permet de se baser sur l'ancien modèle lors de la construction du nouveau et d'avoir des paramètres comparables d'un modèle à l'autre. Ce dernier point est celui qui sera à la base de l'approche générique : en effet lorsqu'un modèle de base sera modifié pour prendre en compte un nouveau contexte, l'ancien et le nouveau modèle resteront comparables.

Afin de conserver la généralité, la première stratégie est d'autoriser les valeurs des paramètres à varier en fonction du modèle. Cette stratégie simple garantit la compatibilité ascendante des modèles, mais ne permet pas d'obtenir une marge de manœuvre suffisante dans les cas où la forme des équations doit être modifiée. Dans ce cas, nous utilisons une méthode permettant de découper un paramètre, de rajouter de nouveaux flux et de complexifier un flux existant à condition de généraliser un cas limite.

Séparer un paramètre en deux Si on souhaite complexifier un modèle, $dx/dt = A \times y$ avec A un paramètre et y une variable suite à l'acquisition de nouvelles données comme la quantité moyenne d'enzyme (E), il est alors possible de couper le paramètre A en $V.E$ afin d'obtenir le modèle $dx/dt = E.V \times y$ avec V et E comme paramètres. L'ancien et le nouveau modèle restent comparables car il est possible de déduire du nouveau modèle la valeur de A .⁹

Rajouter un nouveau flux Si l'on souhaite étendre le modèle initial $dx/dt = A \times y$ avec un flux $f(z)$, suite par exemple à un nouveau dosage concernant une nouvelle variable z , il est possible d'écrire un nouveau modèle : $dx/dt = A \times y + B \times f(z)$ avec B un nouveau paramètre et $f(z)$ une fonction dépendante de la variable mesurée. L'ancien modèle reste bien un cas particulier du nouveau avec $B = 0$.

Complexifier un flux Un premier jeu d'expériences est modélisé correctement par un ensemble de flux (comme par exemple $dx/dt = A \times y$). Plus tard de nouvelles données expérimentales sont disponibles (comme

⁹Il est par contre impossible de déduire de l'ancien modèle les valeurs de E et V .

par exemple de données concernant une plage de valeurs plus large pour y que dans le jeu initial). Ces nouvelles données infirment le modèle initial.

Il est alors possible de complexifier les flux initiaux, en rajoutant une fonction (quasi) nulle dans l'ancien modèle, et non nulle dans le nouveau. Par exemple on écrira dans le nouveau modèle $dx/dt = A \times y / (y/K + 1)$ avec K un nouveau paramètre. L'ancien modèle reste un cas particulier du nouveau modèle sous l'hypothèse $y/K \ll 1$ (ce qui revient à dire par exemple que l'enzyme est loin de la saturation pour les valeurs de $y/K \ll 1$).

Ce cas implique de rajouter *a posteriori* une hypothèse à l'ancien modèle qu'il faut vérifier (ici $y/K \ll 1$). Dans le cas où cette hypothèse n'est pas valide, elle remet en cause l'ancien modèle qu'il convient de réévaluer au regard des nouvelles connaissances, ce qui brise la démarche ascendante.

Les limites de l'approche Cette démarche relativement souple garantit la compatibilité ascendante des modèles mais ne garantit en rien l'unicité des paramètres. Lors d'un ajustement réussi il est donc très important de considérer chaque jeu de paramètres comme "un jeu de paramètres ajustant les données parmi d'autres jeux de paramètres possibles". De plus si l'unicité des paramètres a été prouvée pour un modèle, elle ne s'extrapole pas aux modèles ascendants et il faut donc la tester à nouveau. Ceci peut par exemple poser problème lors du rajout d'un flux : l'ancien modèle est $dx/dt = A \times y$ avec une valeur unique de A égale à A_1 ajustant les anciennes données, le nouveau modèle est : $dx/dt = A \times y + B \times f(x)$, l'ancien modèle est bien le nouveau modèle avec $A = A_1$ et $B = 0$, mais il est possible que d'autres couples (A, B) ajustent aussi bien (ou mieux) les anciennes données.

De plus lorsqu'on se rend compte que l'équation d'un ancien modèle était fautive, ou qu'une hypothèse permettant de la considérer comme un cas limite n'est pas valide, on perd la compatibilité ascendante, ce qui oblige à reconstruire et à tester à nouveau l'ensemble des modèles existants.

1.4 Conclusion

Suite à l'étude bibliographique, la démarche empirique a permis de mettre en évidence un squelette de voies biochimiques nécessaires pour respecter des comportements biologiques. Une stratégie d'écriture d'équations a été développée afin de pouvoir construire les modèles de manière ascendante au fur et à mesure de l'intégration de nouvelles données. L'étape suivante est le croisement de ce squelette avec des données expérimentales de cinétique.

Chapitre 2

Modèle dynamique minimal du métabolisme hépatique des acides gras lors du jeûne : application aux souris sauvages et déficientes en PPAR α

2.1 Introduction

Le squelette identifié lors du chapitre précédent a été confronté à des données biologiques décrivant la variation des quantités d'acides gras dans le foie et le tissu adipeux de souris sauvages et mutantes pour PPAR α pendant 72 heures de jeûne.

Ce squelette est trop complexe pour rendre compte de la situation particulière de cette étude. Tout d'abord, le modèle générique comporte des fonctions biologiques sur lesquelles nous n'avons pas de données expérimentales. Intégrer dans un même modèle toutes ces fonctions créerait un modèle sur-paramétré qui serait capable d'ajuster à tort les données. De plus de nombreuses fonctions comme la synthèse des lipides peuvent être considérées ou approximées comme étant inactives lors du jeûne, en particulier en raison de la forte décroissance de SREBP1.

Nous allons donc construire le modèle sous forme d'équations différentielles ordinaires par une démarche ascendante en intégrant pas à pas les éléments identifiés lors de l'étape précédente (et éventuellement de nouveaux éléments, si les éléments existants ne sont pas suffisants pour rendre compte des données observées) jusqu'à ajuster les données. Nous obtiendrons au final un ensemble d'éléments nécessaires pour expliquer les observations, et apporterons la preuve que tous les sous-ensembles qui ne permettent pas d'ajuster les données ne sont pas suffisants.

La construction de ce modèle a fait l'objet d'une publication (Blavy *et al.*, 2009 [30]) présentée ci dessous.



A minimal model for hepatic fatty acid balance during fasting: Application to PPAR alpha-deficient mice[☆]

P. Blavy^{a,g}, F. Gondret^a, H. Guillou^b, S. Lagarrigue^{c,d,e}, P.G.P. Martin^b, J. van Milgen^a, O. Radulescu^{g,h,*}, A. Siegel^{f,g}

^a INRA, UMR 1079 Systèmes d'Élevage, Nutrition Animale et Humaine (SENAH), F-35590 Saint Gilles, France

^b Laboratoire de Pharmacologie et Toxicologie, Institut National de la Recherche Agronomique, INRA UR66, 180 chemin de Tournefeuille, BP93173, F-31027 Toulouse 3, France

^c Agrocampus Ouest, UMR 598 Génétique Animale, F-35000 Rennes, France

^d INRA, UMR 598 Génétique Animale, F-35000 Rennes, France

^e IFR 140, CEAS, F-35000 Rennes, France

^f CNRS, UMR 6074 IRISA, Campus de Beaulieu, F-35042 Rennes, France

^g INRIA Rennes Bretagne Atlantique, Symbiose-Project, Campus de Beaulieu, F-35042 Rennes, France

^h Université de Rennes 1, UMR 6625 IRMAR, Campus de Beaulieu, F-35042 Rennes, France

ARTICLE INFO

Article history:

Received 21 March 2009

Received in revised form

25 June 2009

Accepted 16 July 2009

Available online 25 July 2009

Keywords:

Fatty acid metabolism

Systems biology

Fasting

Knockout mice

Modeling

ABSTRACT

The purpose of this study is to identify the hierarchy of importance amongst pathways involved in fatty acid (FA) metabolism and their regulators in the control of hepatic FA composition. A modeling approach was applied to experimental data obtained during fasting in PPAR α knockout (KO) mice and wild-type mice. A step-by-step procedure was used in which a very simple model was completed by additional pathways until the model fitted correctly the measured quantities of FA in the liver. The resulting model included FA uptake by the liver, FA oxidation, elongation and desaturation of FA, which were found active in both genotypes during fasting.

From the model analysis we concluded that PPAR α had a strong effect on FA oxidation. There were no indications that this effect changes during the fasting period, and it was thus considered to be constant.

In PPAR α KO mice, FA uptake was identified as the main pathway responsible for FA variation in the liver. The models showed that FA were oxidized at a constant and small rate, whereas desaturation of FA also occurred during fasting.

The latter observation was rather unexpected, but was confirmed experimentally by the measurement of delta-6-desaturase mRNA using real-time quantitative PCR (QPCR). These results confirm that mathematical models can be a useful tool in identifying new biological hypotheses and nutritional routes in metabolism.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Fatty acids (FA) are the main constituents of lipids in the body and are the building blocks for glyco- and phospholipids of cell membranes. FA also play an important role in energy metabolism,

allowing the storage of energy in a very dense form as triglycerides, which can be oxidized later when energy is needed. The FA can also act as signaling molecules and behave as regulators of several transcription factors (Duplus et al., 2000). Therefore, they play crucial roles in normal growth and development (e.g., Uauy et al., 2007) but also in coronary artery disease (Seidelin, 1995; Shirai, 2004), dyslipidemia, hepatic steatosis and other pathologies (Seidelin, 1995; Simopoulos, 1991).

The balance between synthesis and degradation of FA is regulated by nutrient supply and the energy needs of the organism. In humans and mice, the liver plays a central role in the endogenous synthesis of FA (Muiruri and Leveille, 1970). During fasting, FA are released from adipose tissue (AT) by lipolysis, and serve as sources of energy in other organs. Circulating free FA are extensively taken up by the liver in fasting

[☆] Results were partially presented at the JOBIM-2008 meeting held in Lille, France.

* Corresponding author at: INRIA Rennes Bretagne Atlantique, Symbiose-Project, Campus de Beaulieu, F-35042 Rennes, France. Tel.: +33 299847306; fax: +33 299847171.

E-mail addresses: pierre.blavy@irisa.fr (P. Blavy), florence.gondret@rennes.inra.fr (F. Gondret), Herve.Guillou@toulouse.inra.fr (H. Guillou), Sandrine.Lagarrigue@agrocampus-ouest.fr (S. Lagarrigue), Pascal.Martin@toulouse.inra.fr (P.G.P. Martin), jaap.vanmilgen@rennes.inra.fr (J. van Milgen), ovidiu.radulescu@inria.fr (O. Radulescu), anne.siegel@irisa.fr (A. Siegel).

rodents (Remesy and Demigne, 1983). In the liver, FA can be a source of substrates for the synthesis of ketone bodies, which can be used as fuel by extrahepatic tissues.

The synthesis, degradation, and transformation of FA in hepatic cells are catalyzed by over 300 enzymatic reactions (Kanehisa et al., 2008) involved in distinct pathways (e.g., FA oxidation and elongation). These reactions are regulated at the metabolic and genetic levels by various hormones (e.g., insulin (Campbell et al., 1992), leptin (Unger et al., 1999)) and nutrients (e.g., polyunsaturated FA (Sessler and Ntambi, 1998)). However, the simple aggregation of abundant literature data cannot account for all underlying interactions responsible for both FA metabolism and lipid phenotype. A better and more comprehensive understanding of FA metabolism is needed to identify routes that will allow for the nutritional modulation of lipid deposition that may help preventing or curing lipid-related disorders. Therefore, it is critical to identify the main pathways and regulators involved in the control of FA metabolism. To achieve this goal, we used a modeling approach.

Mathematical models are powerful tools to combine information using a common formalism. Models are frequently used to describe, predict and test hypotheses. Depending on the objective, models can have different levels of detail, ranging from very basic molecular mechanisms (e.g., Fattal and Ben-Shaul, 1993) to an empirical “black box” approach (e.g., Forns et al., 2002). Models can be useful to explore and better understand FA metabolism and its regulation within a cell, as well as between organs involved in lipid metabolism. To provide a full description of the dynamics of the system considered as a homogeneous organ (e.g., Calvetti et al., 2008) or including the heterogeneous nature of the organ (e.g., spatial models Chalhoub et al., 2007b), metabolic reactions are modeled as differential equations or analyzed by convex optimization techniques such as flux balance analysis. Spatial models typically use partial differential equations.

A contrasted situation is necessary to identify the key mechanisms involved in the regulation of FA metabolism. In the present study, we considered the kinetics of FA metabolism during fasting in both wild-type and peroxisome proliferator-activated receptor alpha (PPAR α) knockout (KO) mice. Fasting triggers complex adaptive metabolic responses, including a switch to rely on FA and ketone bodies for ATP synthesis (Leone et al., 1999) and an increased capacity for mitochondrial FA oxidation in tissues with high energy demands (Hashimoto et al., 2000; Leone et al., 1999). The PPAR α is considered as the master regulator of FA homeostasis (Desvergne and Wahli, 1999; Leone et al., 1999). A genome-wide transcriptomic approach in mice has recently pointed out the role of PPAR α in the liver in the regulation of FA oxidation and ketone body production during fasting (Sokolović et al., 2008). Animals lacking PPAR α appear to be unable to increase the capacity for cellular FA utilization (Leone et al., 1999). Monitoring the variation in FA composition in tissues of wild-type and PPAR α KO mice during fasting provides a useful experimental data set to understand the regulation of FA metabolism and to develop computational models describing this metabolism.

To our knowledge, there is only one model (Chalhoub et al., 2007a) focused on lipid metabolism in the liver during fasting conditions. This detailed mathematical model was based on differential equations and simulated gluconeogenesis during a 24-h fasting period in the perfused rat liver. This model included key reactions for FA metabolism such as FA uptake, synthesis of triglycerides, and FA oxidation. Because of the large number of reactions involved, many of these were aggregated in series. Moreover, this model did not intend to predict the variation in FA composition in the liver and did not include the genetic regulation of FA metabolism. It intended to predict concentrations and fluxes of intermediate metabolites involved in FA metabolism and

gluconeogenesis in response to changes in various substrate concentrations in the perfused liver.

The aim of the present study was to identify the most relevant pathways and their regulators involved in hepatic FA metabolism. Based on experimental data and information from the literature, a model (based on differential equations) was developed that allows to explain the variation in FA composition of the mouse liver during a fed-to-fasting transition. Additional experimental measurements (mRNA expression of delta-6-desaturase, a key enzyme of polyunsaturated FA synthesis) were carried out to evaluate hypotheses that were formulated following model analyses.

2. Materials and methods

2.1. Biological experiments

2.1.1. Experimental data

Experimental data were obtained in a 8-week-old male wild-type C57BL/6J (WT) and PPAR α KO mice (Lee et al., 1995; Costet et al., 1998) over a 72-h fasting period. Three to six mice for each genotype were sacrificed at different time points (0, 3, 6, 9, 12, 18, 24, 36, 48, 60, and 72 h) after the last meal. Before fasting, all animals were fed ad libitum a rodent diet 2018 from Harlan Teklad (Gannat, France). At each time point, animal body weight was recorded and liver and epididymal white adipose tissue (AT) were dissected, weighed, frozen in liquid nitrogen immediately after dissection, and stored at -80°C until analysis. Total lipid content was determined in liver and AT as described previously (Martin et al., 2007). Proportions of individual FA were determined by gas chromatography analysis of FA methyl esters. The quantities of each FA (C14:0, C16:0, C16:1 ω 9, C16:1 ω 7, C18:0, C18:1 ω 9, C18:1 ω 7, C18:2 ω 6, C18:3 ω 3, C20:1 ω 9, C20:3 ω 6, C20:4 ω 6, C20:5 ω 3, C22:6 ω 3) in the liver and AT were calculated considering the relative proportions of FA in total lipids and the mass of liver and AT. Time-related variations in FA quantity were analyzed using the $lm()$ linear regression implemented in R 2.8.1 (www.r-project.org).

2.1.2. Gene expression

Hepatic total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instruction. Complementary cDNA was synthesized from 2 μg of total RNA using random primers and superscript II (Invitrogen) reverse transcriptase according to the manufacturer's instructions. The mRNA levels of delta-6-desaturase (D6D) were measured by quantitative real-time PCR (QPCR) after 0, 24, 48, and 72 h of fasting using 3'-TCCAGTACCAGATCATCATGACAA-5' as forward primer and 3'-GGTGTAGAAGAAACGCATATAGTAGCTG-5' as reverse primer. Amplifications were performed on an ABI Prism 7000 sequence detection system (Applied Biosystems, Courtaboeuf, France). The QPCR data were normalized by TATA-box binding protein (TBP) expression levels (TBP-F: ACTTCGTGCAAGAAATGCTGAA, TBP-R: GCAGTTGTCCGTGGCTCTCT) and analyzed using the DART-PCR software (Peirson et al., 2003).

2.2. Principles of model development

A step-by-step procedure was applied to include different metabolic pathways and regulators until the model fitted the data well and no reasonable improvement could be obtained.

Due to the absence of at least one complex regulatory system, the PPAR α KO mice can be considered as a relatively simpler model compared to wild-type mice. Therefore we first fitted the

Table 1

Summary of models obtained with the step-by-step procedure. Parameters s_i representing proportions of complete oxidation occur in Eqs. (3), (4) giving fluxes ϕ_i^{pd} .

| Model | Differential equations |
|---|---|
| Model 1 $\forall i \in \{\text{all FA}\}$ | $\frac{dX_i^c}{dt} = \phi_i^{pn}$ Differential equations |
| Model 2 $\forall i \in \{C14 : 0, C16 : 0, C16 : 1w7, C18 : 0, C18 : 1w7, C18 : 3w3, C18 : 2w6, C20 : 5w3, C22 : 6w3, C20 : 3w6, C20 : 4w6\}$ $i = C18 : 1w9$ | $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd}$ with $s_i = 1$ $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd}$ with $s_i = 0$ in PPAR α KO mice and $s_i = 1$ in wild-type mice |
| $i = C16 : 1w9$ | $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd} + \phi_{C18:1w9}^{pcompleteoxi}$ with $s_i = 1$ |
| Model 3 $\forall i \in \{C14 : 0, C16 : 0, C16 : 1w7, C18 : 0, C18 : 1w7\}$ $\forall i \in \{C18 : 3w3, C18 : 2w6\}$ $\forall i \in \{5w3, 5w6\}$ $i = C18 : 1w9$ | Differential equations $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd}$ with $s_i = 1$ $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd} - \phi_i^{Desat}$ with $s_i = 1$ $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd} + \phi_i^{Desat}$ with $s_i = 1$ $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd}$ with $s_i = 0$ in PPAR α KO mice and $s_i = 1$ in wild-type mice |
| $i = C16 : 1w9$ | $\frac{dX_i^c}{dt} = \phi_i^{pn} - \phi_i^{pd} + \phi_{C18:1w9}^{pIncompleteOxi}$ with $s_i = 1$ |

The models predict variations in hepatic fatty acid (FA) composition during fasting.

successive models on data obtained in PPAR α KO mice. In order to allow comparisons between genotypes within the same model (or modeling framework), we then extrapolated to wild-type mice the values of the parameters that should be common to both genotypes. We consider that the KO mice is just a simpler and easier to study submodel of the wild-type model. This resulted in a minimal model of a set of ordinary differential equations describing FA metabolism in the liver during fasting (Table 1).

The main pathways used in our model of FA metabolism are consistent with common biochemistry knowledge (Kanehisa et al., 2008; Murray et al., 2006), and include (Fig. 1) FA uptake, FA oxidation in mitochondria and peroxisomes, ketogenesis, the tricarboxylic acid cycle (where ATP is produced from acetyl-coA), synthesis of malonyl-CoA from acetyl-coA, synthesis of palmitic acid (C16:0) from malonyl-CoA and acetyl-coA, glycolysis to produce acetyl-coA from glucose, and elongation and desaturation of polyunsaturated FA.

Several key regulators of FA metabolism have been described in the literature. They include PPAR α , which is a transcription factor that activates FA oxidation (Leone et al., 1999; Desvergne and Wahli, 1999; Lee et al., 1995) and FA elongation and desaturation (Wang et al., 2006; Guillou et al., 2002). Other regulators are SREBP1, which is a transcription factor that activates FA synthesis (Horton et al., 2002; Jakobsson et al., 2006) as well as FA elongation and desaturation (Matsuzaka et al., 2002) and malonyl-CoA, which inhibits FA oxidation.

Because of the experimental conditions and the resulting data in the present study, a decision was made as to which variables should be included in the model. Metabolic pathways and regulators that were supposed to be inactive or constant during fasting were not considered (Fig. 1). During the early hours of fasting (3–6 h), the level of SREBP1 markedly decreases in the liver (Horton et al., 1998), which results in a fasting-mediated transcriptional down-regulation of the gene encoding FA synthase (i.e., the key enzyme of de novo FA synthesis (Shimano et al., 1999; Horton et al., 2002; Kim et al., 1998)). Therefore, SREBP1 was assumed to be negligible in our experimental conditions, and de novo FA synthesis was supposed not to occur in our experimental conditions. Hence, the influence of SREBP1 on elongation and

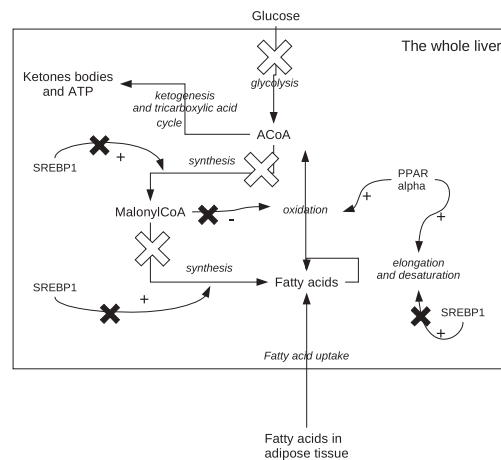


Fig. 1. Pathways that influence fatty acid composition in the liver. The main pathways influencing hepatic fatty acid quantities are indicated by straight arrows. Acetyl-coenzyme A (AcoA) and malonyl-coenzyme A (malonyl-CoA) are pivots in hepatic intermediary metabolism. Transcriptional regulators of fatty acid synthesis, oxidation or desaturation are the sterol regulatory element binding transcription factor 1 (SREBP1) and peroxisome proliferator-activated receptor alpha (PPAR α). (+) signs for an activation, whereas (–) signs for an inhibition of the pathway. During fasting, the pathways with a white cross are assumed to be inactive, and regulations marked with a black cross are assumed to have no direct influence on the variations of fatty acid composition in the liver.

desaturation was not considered. Moreover, since malonyl-CoA is an intermediate in de novo FA synthesis, the malonyl-CoA-dependent inhibition of mitochondrial FA uptake was not considered. Acetyl-coA is a pivot in intermediary metabolism between catabolic and anabolic processes (van Milgen, 2002).

Table 2
Fatty acid contents (μ moles) in the liver of wild type or PPAR α knockout mice during fasting.

| Genotype | PPAR α knockout | | | | Wild type | | | | |
|------------------|------------------------|--------|-------|--------------------|----------------------|--------|-------|--------|----------------------|
| | Fasting time (h) | 0 | 72 | 72/0h ^a | P value ^b | 0 | 72 | 72/0h | P value ^a |
| Fatty acid | | | | | | | | | |
| C14:0 | 0.43 | 10.17 | 23.45 | <0.001 | 0.26 | 3.07 | 11.81 | <0.001 | |
| C16:0 | 31.77 | 129.76 | 4.08 | <0.001 | 23.26 | 53.49 | 2.30 | <0.001 | |
| C16:1 ω 9 | 0.51 | 16.53 | 32.41 | <0.001 | 0.38 | 5.31 | 13.97 | <0.001 | |
| C16:1 ω 7 | 1.34 | 21.06 | 15.72 | <0.001 | 1.91 | 6.73 | 3.52 | 0.001 | |
| C18:0 | 15.79 | 34.54 | 2.19 | <0.001 | 8.69 | 15.92 | 1.83 | <0.001 | |
| C18:1 ω 9 | 14.09 | 257.35 | 18.26 | <0.001 | 11.99 | 85.43 | 7.13 | <0.001 | |
| C18:1 ω 7 | 1.74 | 19.12 | 10.99 | <0.001 | 1.94 | 6.24 | 3.22 | <0.001 | |
| C18:2 ω 6 | 38.57 | 455.88 | 11.81 | <0.001 | 15.11 | 114.84 | 7.60 | <0.001 | |
| C18:3 ω 3 | 1.21 | 9.99 | 8.26 | <0.001 | 0.41 | 2.50 | 6.10 | <0.001 | |
| C20:1 ω 9 | 0.61 | 6.17 | 10.11 | <0.001 | 0.29 | 2.87 | 9.90 | <0.001 | |
| C20:3 ω 6 | 1.06 | 1.99 | 1.88 | 0.017 | 1.22 | 1.04 | 0.85 | 0.037 | |
| C20:4 ω 6 | 12.45 | 15.54 | 1.25 | 0.825 | 9.16 | 11.1 | 1.18 | 0.196 | |
| C20:5 ω 3 | 0.26 | 0.36 | 1.38 | 0.115 | 0.20 | 0.97 | 4.85 | <0.001 | |
| C22:6 ω 3 | 6.30 | 10.97 | 1.74 | 0.021 | 4.09 | 14.31 | 3.50 | <0.001 | |

^a Ratios between the 72 and 0h of fasting are presented although these ratios may be largely influenced by some very low concentrations observed under the fed condition ($t = 0$ h).

^b P value associated to the slope of the linear regression: $Fatty\ acid\ quantity = slope \times time + intercept$ in each genotype. This value was considered to be significant for P value <0.05.

However, because the quantity of acetyl-coA has no direct effect on the FA composition in the liver, there was no need to consider the three pathways affecting acetyl-coA (i.e., glycolysis, ketogenesis and the tricarboxylic acid cycle).

2.3. Model description

2.3.1. Choice of variables

In the various models we considered, we included all measured metabolites, except those involved in intermediate metabolism (i.e., molecules participating in a considered reaction that are neither a product nor a substrate). Therefore, when the elongation and desaturation pathways were added, we included the sum of $S\omega 3 = C20:5\omega 3 + C22:6\omega 3$ and $S\omega 6 = C20:3\omega 6 + C20:4\omega 6$ and not the individual quantities of C20:5 ω 3 and C20:3 ω 6, which were considered as intermediates. Individual quantities of other FA were included in the model.

The quantity of PPAR α was not explicitly included as a variable in the model. However, the difference in numerical values of model parameters between wild-type and PPAR α mice represents the effect of PPAR α .

2.3.2. Fluxes

Liver uptake: The uptake of FA in the liver from the blood was difficult to assess in the experiments. During fasting, free FA are released from AT by lipolysis and serve as energy sources for other organs including the liver. We considered that the observed variation in FA composition in AT would reflect the FA uptake by the liver. Input fluxes of individual FA in liver (Φ_i^{AT}) were thus considered to be proportional to FA released from AT by lipolysis during fasting. The parameter K_m represents the proportionality factor relating FA release from AT to FA uptake by the liver. The K_m was supposed to be the same for all FA. During fasting, FA are released at a time-constant rate from AT, with time-constant rates that differ between FA. To evaluate this rate, a linear regression for each FA in AT (X_i^{AT}) was performed as $X_i^{AT} = a_i \times time + intercept_i$. The slope (a_i) of this relation was then used to estimate FA uptake by the liver:

$$\Phi_i^{AT} = K_m * a_i. \tag{1}$$

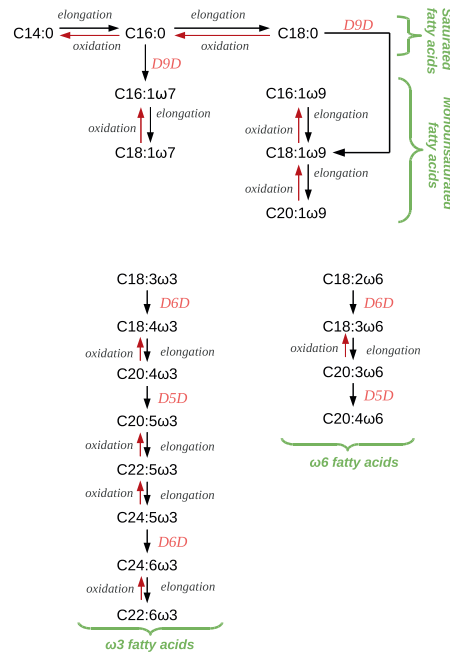


Fig. 2. Transformations between fatty acids. Fatty acids are separated in three classes: saturated and monounsaturated, ω 6, and ω 3. The enzymes regulating desaturation and elongation are shown as delta-5-desaturase (D5D), delta-6-desaturase (D6D) and delta-9-desaturase (D9D). Reactions of interconversion that occur only between fatty acids of a same class are represented by a solid line. The reactions allowing complete oxidation, importation, or synthesis of fatty acids from glucose are not shown.

Oxidation: The reaction flux for total oxidation (Φ_i^{Oxi}) of the i th FA in the liver (X_i^L) during fed-to-fast transition was expressed as a Michaelis–Menten equation in simplified form, using a constant

parameter representing the oxidation rate (k_{oxi}) and a relative affinity parameter for each FA (b_i):

$$\phi_i^{Oxi} = k_{oxi} \frac{X_i^j b_i}{\sum_j X_j^j b_j} \quad \text{with} \quad \sum_i b_i = 1. \quad (2)$$

We hypothesized that the quantities of FA were not limiting for the reaction. This assumption was supported by the fact that the sum of FA in the liver increased from 0 to 72 h since the onset of fasting (Table 2).

The FA do not always undergo complete oxidation resulting in acetyl-coA and ATP. Consequently, shorter chain FA intermediates may be generated, which can accumulate in the liver. We considered incomplete oxidation of a FA as the reaction producing a FA that is two carbons shorter than the original FA. Therefore, the total oxidation flux ϕ_i^{Oxi} was separated into two fluxes representing either complete ($\phi_i^{CompleteOxi}$) or incomplete ($\phi_i^{IncompleteOxi}$) oxidation fluxes:

$$\phi_i^{CompleteOxi} = s_i * \phi_i^{Oxi}, \quad (3)$$

$$\phi_i^{IncompleteOxi} = (1 - s_i) * \phi_i^{Oxi}, \quad (4)$$

where s_i represents the proportion of complete oxidation of the total oxidation.

In the models considered (Table 1), we had to choose between three hypotheses: the i th FA is only oxidized completely ($s_i = 1$), the i th FA is only partially oxidized leading to the generation of a two carbons shorter FA ($s_i = 0$), or the i th FA is both partially and totally oxidized ($s_i \in]0; 1[$). These hypotheses will be tested going from the simplest ($s_i = 1$) to the most complicated hypothesis ($s_i \in]0; 1[$) in a step-by-step procedure (Section 2.3.3). As will be shown later, the third hypothesis was not necessary to obtain a correct fit of the model to the data.

Elongation and desaturation: Unsaturated FA consist of mono-unsaturated and polyunsaturated FA. Based on the location of the last double bond, polyunsaturated FA are classified as $\omega 3$ or $\omega 6$. Within each class, FA can be transformed through elongation and desaturation reactions (Fig. 2). The enzyme D6D has been considered a rate-limiting enzyme (Cho et al., 1999) of the FA elongation/desaturation pathway, which coincide with the observation that the hepatic quantities of C18:2 $\omega 6$ and C18:3 $\omega 3$ increased from 0 to 72 h of fasting (Table 2). Because D6D is common to $\omega 3$ and $\omega 6$ desaturation pathways, C18:2 $\omega 6$ and C18:3 $\omega 3$ are in competition for this enzyme. Elongation and desaturation fluxes ($\phi_{C18:2\omega 6}^{Desat}$ and $\phi_{C18:3\omega 3}^{Desat}$) of C18:2 $\omega 6$ and C18:3 $\omega 3$ were included in the rate of elongation and desaturation (k_{desat}), and the relative affinity of D6D for the i th FA (d_i) can be given as

$$\phi_{C18:3\omega 3}^{Desat} = k_{desat} \frac{C18:3\omega 3 * d_{C18:3\omega 3}}{C18:3\omega 3 * d_{C18:3\omega 3} + C18:2\omega 6 * d_{C18:2\omega 6}}, \quad (5)$$

$$\phi_{C18:2\omega 6}^{Desat} = k_{desat} \frac{C18:2\omega 6 * d_{C18:2\omega 6}}{C18:3\omega 3 * d_{C18:3\omega 3} + C18:2\omega 6 * d_{C18:2\omega 6}}, \quad (6)$$

with $\sum_i d_i = 1$.

2.3.3. Choosing the optimal level of detail by a step-by-step procedure

A step-by-step procedure was applied until the model included just enough detail to fit the change in FA quantities over time.

- Step 1: We started with a model including only one flux (i.e., uptake of FA from AT by the liver).

- Step 2: We estimated model parameters from the experimental data (see the parameter estimation method below).
- Step 3: If a parameter set was found allowing a close fit of model predictions to the observed data, the model (and its parameters) was accepted and the procedure ended (see the fitting criterion below). If not, model output was analyzed, and additional pathways and regulators were included. The selection of pathways and regulations was based on model behavior and biological knowledge. We then went back to step 2.

Parameters estimation: An initial guess of parameter estimates was found by a linear approximation of the model under the assumption that $\sum_{j \in FA} X_j^j b_j$ is constant. This sum was evaluated by replacing X_j^j by the mean of observed value for the j th FA. Parameters were then estimated using the Nelder and Mead optimization (Nelder and Mead, 1965) implemented in R 2.8.0 *optim* procedure, by minimizing the sum of squared deviations between observed data and predicted values.

Goodness of fit: Fitting quality was estimated using the coefficient of variation of the mean squared prediction error (cvMSPE) (Tedeschi, 2006; Bibby, 1977). The mean squared prediction error measures the distance between observed and predicted values, and can be decomposed in central error (CE), regression error (RE), and disturbance error (DE). The CE describes the contribution of distance between mean values of observed and predicted data. The RE describes how the slope of the linear regression between predicted and observed data differs from one. The DE is the remaining error. The cvMSPE was used to standardize the results, so that different FA could be compared. Models resulting in a low cvMSPE combined with an important contribution of DE are preferred and we considered results acceptable if $DE > 25\%$ (an arbitrarily chosen value).

3. Results and discussion

In both mouse genotypes, FA quantities in the liver were generally greater after a 72 h fasting period compared with the fed state (Table 2). For only three FA, there was no difference in the quantity of FA between 0 and 72 h. This concerned C20:4 $\omega 6$ in wild-type mice and C20:4 $\omega 6$ and C20:5 $\omega 3$ in PPAR α KO mice. There are indications in the literature showing that FA can accumulate in the liver during fasting in both PPAR α KO and wild-type mice using Sudan Black staining of lipids (Lee et al., 2004) or Oil Red O staining of neutral lipids (Hashimoto et al., 2000). The amplitudes of fasting-induced FA accumulations observed in the present study are consistent with a previous study of wild-type and PPAR α knockout mice under the fed and starved (72 h) conditions (see Lee et al., 2004).

This likely reflects the major role played by the liver during fasting to cope with the metabolism of FA coming from adipose tissue lipolysis. The quantities of most FA increased more markedly in PPAR α KO livers compared to wild-type controls (Table 2). As shown below, this likely results from the impaired hepatic FA beta-oxidation in PPAR α KO mice. Saturated and especially monounsaturated FA were most strongly increased than polyunsaturated FA in both genotypes. Only C20:5 $\omega 3$ and C22:6 $\omega 3$ increased more strongly in wild-type than in PPAR α KO livers during fasting (Table 2).

3.1. Model structure

According to the step-by-step procedure, three successive models were built, differing in the degree of complexity. Model 1 included only FA uptake, whereas model 2 included both FA uptake

and FA oxidation. Model 3 was similar to model 2 but also included elongation and desaturation of FA. The equations used for these models are given in Table 1. These models are shown in Fig. 3.

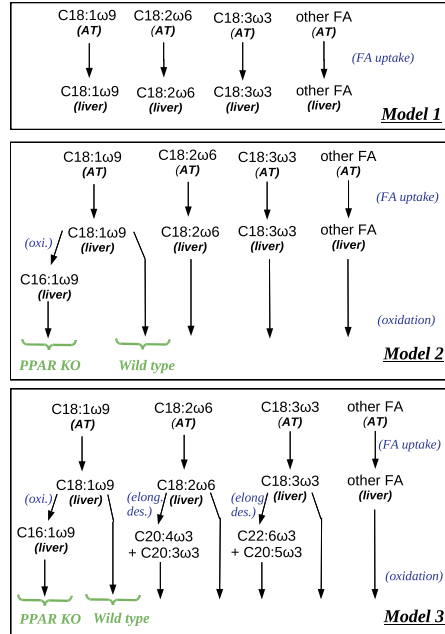


Fig. 3. Models built for wild-type and PPAR α knockout (KO) mice according to a step-by-step procedure detailed in Section 2.3. Model 1 included only fatty acids (FA) uptake from adipose tissues (AT) to liver. Model 2 included both FA uptake and FA oxidation (oxi.) in liver. Model 3 was similar to model 2 but also included elongation and desaturation (elong. des.) of FA.

For each model, the values of the parameters were first estimated in the simpler PPAR α KO mice model. The parameter values were then extrapolated to wild-type mice. When the model with extrapolated parameter values did not fit to the data for wild-type mice, then new parameter values were estimated from wild-type data (Table 3).

3.2. Model 1: accumulation of FA by the liver

This model included uptake of FA by the liver proportional to AT efflux. The proportionality constant was assumed to be the same for all FA. This constant was evaluated in PPAR α KO mice and then extrapolated to wild-type mice.

3.2.1. During fasting, FA release from AT differs between genotypes

Fatty acid composition in mouse liver (Table 2) and AT (Table 4) in the fed state was different in PPAR α KO mice than in wild-type mice. The FA quantities (except for C20:5 ω 3 and C22:6 ω 3) were generally greater in the liver and lower (except for C18:2 ω 6 and C18:3 ω 3) in AT of KO mice compared with the wild-type mice. The variation in FA quantities in AT during fasting could be described by linear functions of time, as shown by the high correlation coefficients in Table 4. The regression slopes (Table 4) were lower in PPAR α KO mice than in WT animals, suggesting that lipolysis from AT was less active during fasting in the absence of PPAR α (Fig. 4).

A lesser reduction in epididymal fat pad weights in KO mice compared with wild-type has been observed by others (Lee et al., 2004), suggesting that the mobilization of fat depots is delayed during starvation in mice that lack PPAR α .

This potential effect of PPAR α on adipose tissue lipolysis during fasting contrasts with its low expression in this tissue (Bookout et al., 2006). However, it has been shown that despite a low expression level PPAR α regulates glycerol kinase (GyK) in white adipose tissue (Mazzucotelli et al., 2007). Interestingly, PPAR α KO mice exhibit a higher basal AT GyK expression compared to wild-type mice (Mazzucotelli et al., 2007). Since GyK is involved in FA recycling, its activity may contribute to counteract the effect of

Table 3
Calculated parameters of the different models built to depict fatty acid composition in the liver of wild-type or PPAR α knockout mice during fasting.

| Genotype | PPAR α knockout mice | | | Wild-type mice | | |
|---------------------------------|-----------------------------|----------|----------|----------------|----------|----------|
| | Model 1 | Model 2 | Model 3 | Model 1 | Model 2 | Model 3 |
| K_{in} | 2.9 | 2.9 | 2.9 | 2.9 | 2.9 | 2.9 |
| K_{oxi} (mmolh $^{-1}$) | 0 | 1.05e-04 | 1.05e-04 | 0 | 9.92e-03 | 9.92e-03 |
| s_i ($i \neq C18:1\omega9$) | 1 | 1 | 1 | 1 | 1 | 1 |
| $S_{C18:1\omega9}$ | 0 | 0 | 0 | 1 | 1 | 1 |
| $\hat{b}_{C14:0}$ | 0 | 0 | 0 | 1.43e-01 | 1.43e-01 | 1.43e-01 |
| $\hat{b}_{C16:0}$ | 0 | 0 | 0 | 6.96e-02 | 6.96e-02 | 6.96e-02 |
| $S_{C16:1\omega9}$ | 0 | 0 | 0 | 2.51e-02 | 2.51e-02 | 2.51e-02 |
| $S_{C16:1\omega7}$ | 0 | 0 | 0 | 2.22e-01 | 2.22e-01 | 2.22e-01 |
| $S_{C18:0}$ | 0 | 0 | 0 | 1.67e-02 | 1.67e-02 | 1.67e-02 |
| $S_{C18:1\omega9}$ | 1 | 1 | 1 | 9.51e-02 | 9.51e-02 | 9.51e-02 |
| $S_{C18:1\omega7}$ | 0 | 0 | 0 | 1.03e-01 | 1.03e-01 | 1.03e-01 |
| $S_{C18:2\omega6}$ | 0 | 0 | 0 | 6.39e-02 | 6.39e-02 | 6.39e-02 |
| $S_{C18:3\omega3}$ | 0 | 0 | 0 | 1.54e-01 | 1.54e-01 | 1.54e-01 |
| $S_{C20:1\omega9}$ | 0 | 0 | 0 | 7.18e-02 | 7.18e-02 | 7.18e-02 |
| $S_{C20:3\omega6}$ | 0 | 0 | 0 | 3.44e-02 | 3.44e-02 | 3.44e-02 |
| $S_{C20:4\omega6}$ | 0 | 0 | 0 | 8.58e-04 | 8.58e-04 | 8.58e-04 |
| $S_{C20:5\omega3}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $S_{C22:6\omega3}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $S_{C22:6\omega3}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| K_{desat} (molh $^{-1}$) | 0 | 0 | 4.92e-05 | 0 | 0 | 9.2e-05 |
| $d_{C18:3\omega3}$ | 1 | 1 | 1 | 0.99956 | 0.99956 | 0.99956 |
| $d_{C18:6}$ | 0 | 0 | 0 | 4.4e-04 | 4.4e-04 | 4.4e-04 |

Table 4
Estimates of linear regressions for fatty acid contents during a 72 h of fast in adipose tissue of wild type or PPAR α knockout mice.

| Genotype | PPAR α knockout | | | Wild-type | | |
|------------------|--------------------------|-------------------------|--------|--------------------------|-------------------------|-------|
| | Intercept (μ moles) | Slope (nmol h $^{-1}$) | R 2 | Intercept (μ moles) | Slope (nmol h $^{-1}$) | R |
| Fatty acids | | | | | | |
| C14:0 | 3.91 | -42.68 | -0.81 | 4.64 | -59.58 | -0.98 |
| C16:0 | 57.22 | -623.63 | -0.88 | 71.28 | -962.29 | -0.99 |
| C16:1 ω 9 | 2.34 | -23.38 | -0.87 | 2.63 | -31.80 | -0.98 |
| C16:1 ω 7 | 15.50 | -183.58 | -0.87 | 21.13 | -311.27 | -0.96 |
| C18:0 | 5.76 | -53.23 | -0.85 | 6.49 | -75.27 | -0.99 |
| C18:1 ω 9 | 87.23 | -903.41 | -0.87 | 105.20 | -1352.02 | -0.99 |
| C18:1 ω 7 | 6.57 | -67.63 | -0.86 | 8.72 | -113.66 | -0.98 |
| C18:2 ω 6 | 137.19 | -1448.85 | -0.87 | 114.91 | -1485.97 | -1.00 |
| C18:3 ω 3 | 6.17 | -72.85 | -0.90 | 5.83 | -83.98 | -0.98 |
| C20:1 ω 9 | 1.70 | -14.58 | -0.81 | 2.85 | -28.57 | -0.95 |
| C20:3 ω 6 | 0.27 | -2.37 | -0.80 | 0.58 | -6.74 | -0.96 |
| C20:4 ω 6 | 0.44 | -4.00 | -0.86 | 0.89 | -11.36 | -0.95 |
| C20:5 ω 3 | 0.14 | -1.14 | -0.69 | 0.33 | -3.61 | -0.94 |
| C22:6 ω 3 | 0.14 | -0.76 | -0.47 | 0.66 | -7.51 | -0.94 |
| Average | 23.18 | -245.86 | -0.81 | 24.72 | -323.83 | -0.97 |

^a Pearson correlation coefficient.

lipolysis in PPAR α KO AT. On the other hand, PPAR α -dependent regulations in other tissues expressing PPAR α could have indirect effects on AT through interorgan communications. For example, it has been proposed that PPAR α may regulate specific genes in the brain which could result in increased Glut4 expression in the AT (Knauf et al., 2006). In this study, high Glut4 expression was proposed to increase glucose clearance in the adipose tissue of PPAR α KO mice thus contributing to their hypoglycemia during fasting. Increased glucose input in adipose tissue of fasted PPAR α KO mice may also contribute to counteract the effects of lipolysis on AT lipid content by stimulating inappropriately FA synthesis.

3.2.2. During fasting, FA hepatic input flux is assumed to be proportional to the FA release from AT

During fasting, FA released by lipolysis in AT serve as an energy source for other organs, including the liver. The free FA uptake by the liver has been described by others as a linear function of the total free FA concentration in blood (Berk and Stump, 1999), until saturation of the uptake system (Sorrentino and Berk, 1993).

In the absence of information in our experiments on the free FA flux in blood (i.e., FA concentrations and blood flow), we assumed that hepatic input flux of FA was proportional to the FA release from AT during fasting, using a same K_{in} proportionality constant for all FA and both genotypes.

3.2.3. FA uptake is the major phenomenon involved in the variation of FA in PPAR α KO mice, but not in wild-type mice where oxidation is important too

The cvMSPE estimates (Table 5) showed that the sum of all FA and all individual FA quantities except C16:1 ω 7, C16:1 ω 9, C22:6 ω 3, and C18:3 ω 3 were correctly fitted in PPAR α KO mice. By contrast, in wild-type mice neither the sum nor the individual FA are correctly fitted. This result was anticipated for wild-type mice, as during fasting, liver FA oxidation is known to be more active in wild-type mice than in PPAR α ones (Le May et al., 2000). Therefore, the uptake of FA was the major phenomenon involved in the variation of hepatic FA during fasting in PPAR α KO mice, but not in wild-type mice.

3.2.4. Hepatic FA input can be modeled as a simple proportional function of FA variation in AT

The good fitting of most FA in PPAR α KO mice confirmed that hepatic FA input can be modeled as a simple proportional function

of FA variation in AT. This also suggests that FA uptake is the main pathway responsible for variation in hepatic FA composition in PPAR α KO mice during the fed-to-fasting transition. It also suggests that selective importation of FA was negligible.

3.2.5. C16:1 ω 7 has a specific two times dynamic

The quantity of C16:1 ω 7 was poorly fitted in PPAR α KO mice and was overestimated from 50 to 72 h (Fig. 4). The kinetics of C16:1 ω 7 hepatic accumulation seems to follow a specific time pattern with accumulation from 0 to 50h, to remain constant thereafter. This pattern cannot be explained by importation or specific oxidation, which should have occurred at a constant rate during the fed-to-fasting transition. Recently, it was shown that C16:1 ω 7 can act as an AT-derived signal and has, unlike other FA, a systemic metabolic effect (Cao et al., 2008). It is possible that C16:1 ω 7 has a very specific metabolic fate and not only targets the liver but also other tissues such as muscle. This hypothesis remains to be explored further experimentally, and the lack-of-fit for C16:1 ω 7 was therefore not further addressed in our model.

3.2.6. The bad fitting of C16:1 ω 9 and C22:6 ω 3 in PPAR α KO mice is unlikely to be explained by a specific FA uptake

The poor fitting of the quantities of C16:1 ω 9 and C22:6 ω 3 in the liver of PPAR α KO mice can be explained by specific higher rates of uptake or by reactions that produce these FA from other FA. Cellular uptake of FA as well as intracellular transport of FA is mediated through simple diffusion, facilitated diffusion, or carrier-mediated transport (Berk and Stump, 1999). It has been shown (Oikari et al., 2008) that cytosolic acyl-carrier binding protein (ACBP) has a higher affinity for C14–C22 FA than for medium chain FA (C8–C12). To our knowledge, selective regulation of long-chain FA uptake has not yet been described. When we calculate the selective uptake for C16:1 ω 9, and C22:6 ω 3 required to fit the data, it appears that these values should have been 2–12 times greater than the common rate determined for the other FA. Therefore, we assumed that transformations between FA leading to the production of C16:1 ω 9 and C22:6 ω 3 would be a more likely scenario to explain the accumulation of these FA in the liver, rather than the selective uptake.

3.2.7. The bad fitting of C16:1 ω 9 in PPAR α KO mice is likely explained by incomplete oxidation of C18:1 ω 9

C16:1 ω 9 can be produced by oxidation from C18:1 ω 9 (Fig. 2). In PPAR α KO mice, the proportion of C16:1 ω 9 in the liver is very

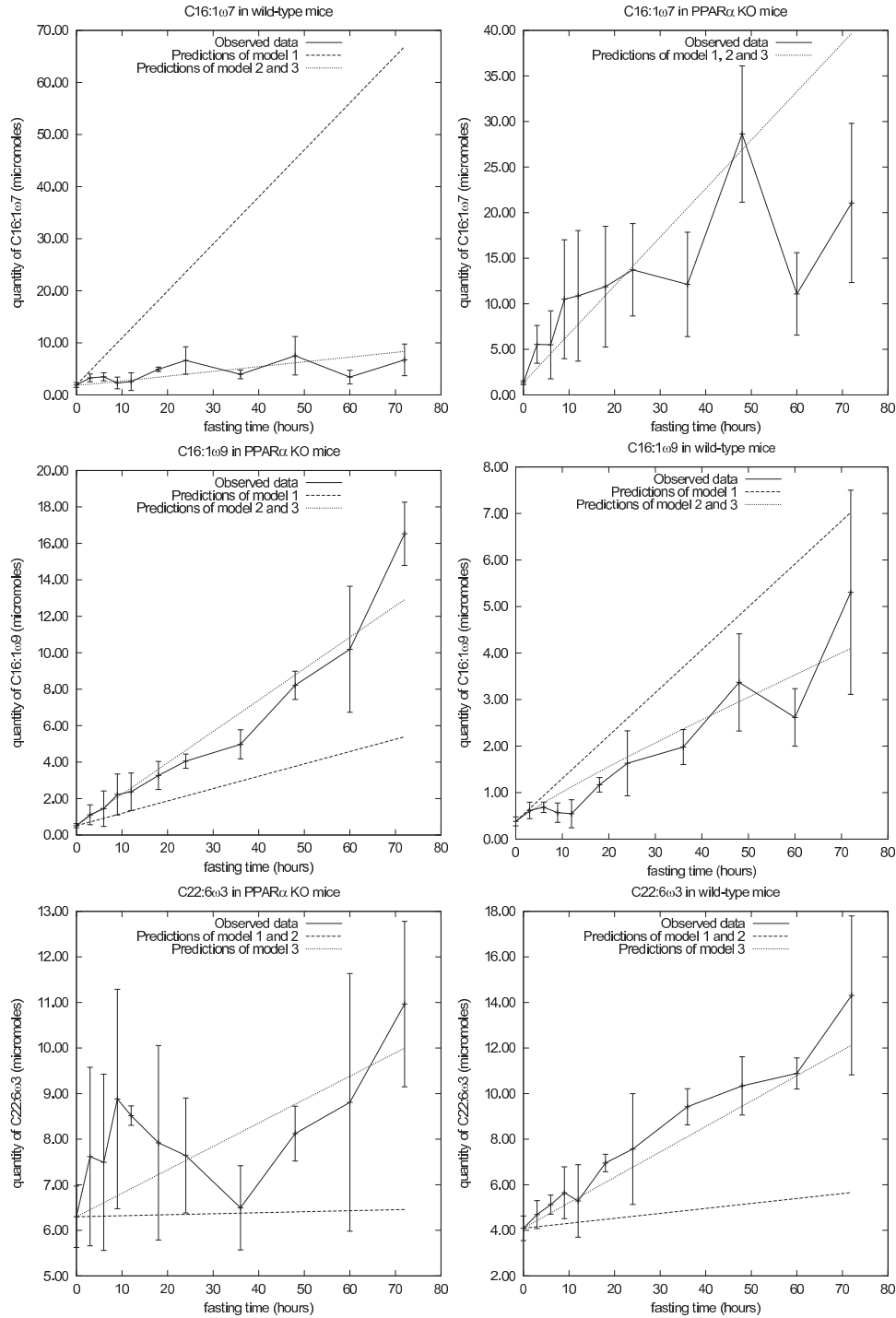


Fig. 4. Predicted and observed quantities of fatty acids in the liver of wild type or PPAR α knockout (KO) mice during fasting.

Table 5
MSPE coefficient of variation and its decomposition in the model 1 predicting fatty acid composition in the liver of wild type or PPAR α knockout mice during fasting.

| Genotype | PPAR α knockout mice | | | | Wild-type mice | | | |
|-------------------------------------|-----------------------------|--------|--------|-----------|-----------------|--------|--------|-----------|
| | cvMSPE | CE (%) | RE (%) | DE (%) | cvMSPE | CE (%) | RE (%) | DE (%) |
| C14:0 | 2.77e-01 | 0 | 24 | 76 | 4.42e+00 | 55 | 44 | 0 |
| C16:0 | 3.16e-01 | 6 | 57 | 37 | 2.15e+00 | 52 | 48 | 1 |
| C16:1 ω 9 | 8.29e-01 | 43 | 49 | 8 | 8.23e-01 | 59 | 30 | 11 |
| C16:1 ω 7 | 7.73e-01 | 12 | 60 | 28 | 6.89e+00 | 53 | 47 | 0 |
| C18:0 | 1.35e-01 | 34 | 5 | 60 | 5.28e-01 | 62 | 35 | 3 |
| C18:1 ω 9 | 2.59e-01 | 49 | 7 | 44 | 2.94e+00 | 54 | 45 | 1 |
| C18:1 ω 7 | 2.22e-01 | 43 | 4 | 53 | 3.19e+00 | 56 | 44 | 0 |
| C18:2 ω 6 | 2.91e-01 | 55 | 7 | 38 | 1.98e+00 | 51 | 48 | 1 |
| C18:3 ω 3 | 6.62e-01 | 13 | 67 | 19 | 4.76e+00 | 51 | 49 | 1 |
| C20:1 ω 9 | 3.95e-01 | 37 | 36 | 27 | 2.23e+00 | 60 | 38 | 2 |
| C20:3 ω 6 | 1.88e-01 | 1 | 1 | 98 | 1.10e+00 | 67 | 30 | 3 |
| C20:4 ω 6 | 9.94e-02 | 3 | 2 | 95 | 7.64e-02 | 0 | 44 | 56 |
| C20:5 ω 3 | 4.7e-01 | 0 | 54 | 46 | 1.5e-01 | 13 | 4 | 83 |
| C22:6 ω 3 | 2.58e-01 | 68 | 11 | 21 | 5.15e-01 | 58 | 40 | 2 |
| Total FA | 2.12e-01 | 37 | 0 | 63 | 2.07e+00 | 53 | 47 | 1 |
| Total ω 9 ^b | 2.79e-01 | 51 | 12 | 37 | 2.83e+00 | 55 | 45 | 1 |
| Total ω 3 ^c | 2.37e-01 | 2 | 52 | 47 | 3.68e-01 | 38 | 55 | 7 |
| Total ω 6 ^d | 2.75e-01 | 54 | 6 | 39 | 1.66e+00 | 50 | 48 | 1 |
| C20:5 ω 3 + C22:6 ω 3 | 2.48e-01 | 68 | 8 | 25 | 4.86e-01 | 58 | 40 | 2 |
| C20:3 ω 6 + C20:4 ω 6 | 1.04e-01 | 2 | 1 | 97 | 1.34e-01 | 32 | 52 | 17 |
| Total sat + mono ^e | 1.94e-01 | 12 | 10 | 78 | 2.52e+00 | 54 | 46 | 0 |

This model included only fatty acid uptake from adipose tissue.

^a Mean squared prediction error (MSPE) expressed as a coefficient of variation (cv). This value measures the distance between observed and predicted value. MSPE can be decomposed in central error (CE) + regression error (RE) + disturbance error (DE). CE describes the contribution of distance between mean values of observed and predicted data. The RE describes how the slope of the linear regression between predicted and observed data differs from one. DE is the remaining error. Fatty acid quantity is considered badly fitted if $DE \leq 25\%$; these values are written in bold font.

^b C16:1 ω 9 + C18:1 ω 9 + C20:1 ω 9.

^c Sum of C18:3 ω 3, C20:5 ω 3, C22:6 ω 3.

^d Sum of C18:2 ω 6, C20:3 ω 6, C20:4 ω 6.

^e Sum of saturated (C14:0 + C16:0 + C18:0) and monounsaturated (C16:1 ω 7, C16:1 ω 9, C18:1 ω 7, C18:1 ω 9) fatty acids.

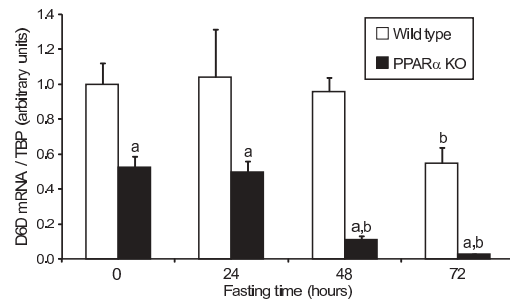


Fig. 5. Measured mRNA of delta-6-desaturase in wild-type and PPAR α knockout mice liver during a 72h fasting. The QPCR data were normalized by TATA-box binding protein (TBP) expression levels. (a) indicates a significant difference (student test, $p < 0.05$) between the two genotypes at each time point and (b) indicates a significant difference (student test, $p < 0.05$) since the onset of fasting. Error bars indicate the standard error of the mean.

small (1.7% of total liver FA at 72 h of fasting), compared with that of C18:1 ω 9 (26% of total hepatic FA at 72 h). Therefore, inclusion of the incomplete oxidation of C18:1 ω 9 into C16:1 ω 9 might fit the latter FA, without reducing the quality of fit for C18:1 ω 9 (this reaction will be included in model 2).

3.2.8. The bad fitting of C22:6 ω 3 in PPAR α KO mice is likely explained by elongation/desaturation of C18:3 ω 3

C22:6 ω 3 was underestimated in PPAR α KO mice and can only be produced by elongation and desaturation (Fig. 2) from C18:3 ω 3, which was overestimated. Therefore, it seems plausible

that the absence of the elongation/desaturation pathway in model 1 is the cause of the observed lack-of-fit. These reactions will be added in model 3 (Section 3.4).

3.2.9. Building the next model

Model 2 will include incomplete oxidation of C18:1 ω 9 to C16:1 ω 9 in PPAR α KO mice, and complete oxidation of all FA in wild-type mice.

3.3. Model 2: accumulation and (complete or partial) oxidation of FA by the liver

This model included both FA uptake (see model 1 in Section 3.2) and FA oxidation in the liver of both genotypes.

3.3.1. Both PPAR α KO and wild-type mice have an active oxidation during fasting but with much higher rates in wild-type mice

In PPAR α KO mice, only the partial oxidation of C18:1 ω 9 was considered ($b_{C18:1\omega9} = 1$ whereas $b_i = 0$ for the other FA). This oxidation was considered to be incomplete, and produced C16:1 ω 9, which was modeled as $s_{C18:1\omega9} = 0$. As anticipated, the disturbance error (DE) of C16:1 ω 9 was greater for model 2 (87%, Table 6) than from model 1 (8%, Table 5). Comparatively, the DE for C18:1 ω 9 did not decrease notably (from 44% to 36%, Tables 5 and 6). In wild-type mice, the complete oxidation of all FA was required to obtain a better fit of the model, as shown by the increase of the DE for most of the FA (Tables 5 and 6). The estimated rate of oxidation of C18:1 ω 9 in PPAR α KO mice was only 1% of that estimated in wild-type mice during a 72 h fasting period, confirming that wild-type mice had a much higher hepatic FA oxidation compared to PPAR α KO mice.

Table 6
MSPE coefficient of variation and its decomposition in the model 2 predicting fatty acids composition in mouse liver during fasting.

| Genotype | PPAR α knockout mice | | | | Wild-type mice | | | |
|-------------------------------------|-----------------------------|---------------------|--------|--------|----------------|--------|--------|--------|
| | Statistics | cvMSPE ^a | CE (%) | RE (%) | DE (%) | cvMSPE | CE (%) | RE (%) |
| C14:0 | 2.77e-01 | 0 | 24 | 76 | 3.18e-01 | 0 | 10 | 90 |
| C16:0 | 3.16e-01 | 6 | 57 | 37 | 2.52e-01 | 3 | 46 | 51 |
| C16:1 ω 7 | 2.52e-01 | 0 | 13 | 87 | 3.15e-01 | 4 | 11 | 85 |
| C16:1 ω 9 | 7.73e-01 | 12 | 60 | 28 | 4.07e-01 | 0 | 31 | 69 |
| C18:0 | 1.35e-01 | 34 | 5 | 60 | 7.75e-02 | 9 | 1 | 90 |
| C18:1 ω 7 | 2.86e-01 | 53 | 11 | 36 | 2.3e-01 | 0 | 0 | 100 |
| C18:1 ω 9 | 2.22e-01 | 43 | 4 | 53 | 1.97e-01 | 0 | 15 | 84 |
| C18:2 ω 6 | 2.91e-01 | 55 | 7 | 38 | 2.11e-01 | 0 | 5 | 95 |
| C18:3 ω 3 | 6.62e-01 | 13 | 67 | 19 | 3.74e-01 | 0 | 23 | 76 |
| C20:1 ω 9 | 3.95e-01 | 37 | 36 | 27 | 3.51e-01 | 1 | 23 | 76 |
| C20:3 ω 6 | 1.88e-01 | 1 | 98 | 1 | 1.28e-01 | 23 | 4 | 73 |
| C20:4 ω 6 | 9.94e-02 | 3 | 2 | 95 | 7.16e-02 | 11 | 24 | 65 |
| C20:5 ω 3 | 4.7e-01 | 0 | 54 | 46 | 1.5e-01 | 13 | 4 | 83 |
| C22:6 ω 3 | 2.58e-01 | 68 | 11 | 21 | 5.15e-01 | 58 | 40 | 2 |
| Total FA | 2.12e-01 | 37 | 0 | 63 | 1.80e-01 | 3 | 5 | 91 |
| Total ω 9 ^b | 2.79e-01 | 51 | 12 | 37 | 2.31e-01 | 0 | 1 | 99 |
| Total ω 3 ^c | 2.37e-01 | 2 | 52 | 47 | 4.01e-01 | 62 | 32 | 5 |
| Total ω 6 ^d | 2.75e-01 | 54 | 6 | 39 | 1.81e-01 | 0 | 6 | 93 |
| C20:5 ω 3 + C22:6 ω 3 | 2.48e-01 | 68 | 8 | 25 | 4.86e-01 | 58 | 40 | 2 |
| C20:3 ω 6 + C20:4 ω 6 | 1.04e-01 | 2 | 1 | 97 | 7.28e-02 | 13 | 28 | 58 |
| Total sat + monof ^e | 1.94e-01 | 12 | 10 | 78 | 1.91e-01 | 1 | 12 | 87 |

This model includes fatty acid influx from adipose tissue and oxidation of fatty acids.

^a MSPE, CE, RE and DE are defined in footnote a of Table 5.

^b C16:1 ω 9 + C18:1 ω 9 + C20:1 ω 9.

^c Sum of C18:3 ω 3, C20:5 ω 3, C22:6 ω 3.

^d Sum of C18:2 ω 6, C20:3 ω 6, C20:4 ω 6.

^e Sum of saturated (C14:0 + C16:0 + C18:0) and monounsaturated (C16:1 ω 7, C16:1 ω 9, C18:1 ω 7, C18:1 ω 9) fatty acids.

Fasting is thought to induce a rise of hepatic lipids sensed by PPAR α that, in turn, stimulates the expression of oxidative genes (Kersten et al., 1999). Studies using PPAR α KO mice have provided evidence that the absence of functional PPAR α decreases basal levels of β -oxidation of C16:0 in PPAR α KO mice, but induces no difference in the metabolism of C24:0 compared with wild-type mice (Aoyama et al., 1998). The hepatic expression of most oxidative genes was not induced by fasting in PPAR α KO mice (Leone et al., 1999). Finally, beta-hydroxybutyrate (i.e., an important ketone product of liver FA oxidation) in PPAR α KO mice is 14% of that in wild-type mice fasted for 24 h (Kersten et al., 1999).

Considered together, the introduction of a small oxidation rate for PPAR α KO mice during fasting is not inconsistent with the literature results indicated above. It should be kept in mind that this option will likely underestimate the oxidation rate, because a model with a higher FA uptake rate and a higher oxidation rate would produce similar results. Furthermore, it is possible that the oxidation of FA other than C18:1 ω 9 also occurred in PPAR α KO mice during fasting. However, this was not considered because results from model 1 suggested that these oxidation rates are small. Introducing additional oxidation reactions was not necessary to obtain reasonable model predictions (Table 2).

In wild-type mice, the simplest hypothesis of complete oxidation ($s_i = 1$) of FA was sufficient to model the FA content in the liver. Therefore, a step-by-step procedure did not include the use of more refined hypotheses by combining complete and incomplete oxidation. The calculated affinity coefficients for the various FA regarding oxidation are given in Table 3 and the cvMSPE values are in Table 6. The accuracy of the model was generally good, except for C20:5 ω 3 and C22:6 ω 3, which have DE proportions smaller than 25%. It is known that FA are oxidized in peroxisomes or mitochondria at different rates according to chain length and degree of unsaturation (Mahler et al., 1953; Fritz, 1959; Shindo and Hashimoto, 1978; Mannaerts et al., 1979; Hiltunen

et al., 1986). To our knowledge, the relative rates of oxidation of FA are unknown. Even though FA may be oxidized at different rates, the model fitted the experimental data well without the need for a time-dependent regulation of oxidation during the fed-to-fasting transition. This suggests that the PPAR α -dependent induction of FA oxidation during fasting is either small compared to its constitutive effect, or that it saturates relatively quickly during fasting. The saturation hypothesis may be tested in a specific study focusing on the first 10–20 h of fasting by evaluating the expression of genes and enzymes involved in the oxidation of FA.

3.3.2. Building the next model

Model 2 only included FA uptake and oxidation in the liver, but many of the short-comings of model 1 could be resolved by including the oxidation of FA. As discussed for model 1, the poor quality of fit for C22:6 ω 3 cannot be resolved by oxidation and the results for this FA were not better in model 2 than they were in model 1 (Table 6). As indicated above, C22:6 ω 3 can only be produced by elongation and desaturation (Fig. 2) from C18:3 ω 3. Thus, to build model 3 we added the elongation and desaturation pathway to model 2. In wild-type mice, the oxidation rate had to be estimated again to account for the use of C18:3 ω 3 by both oxidation and elongation and desaturation.

3.4. Model 3: accumulation, oxidation, and desaturation and elongation of FA by the liver

Essential FA metabolism involves the desaturation and elongation to synthesize very long-chain polyunsaturated FA from C18:3 ω 3 and C18:2 ω 6. Because delta-6-desaturase (D6D) is common to ω 3 and ω 6 desaturation pathways, C18:2 ω 6 and C18:3 ω 3 are in competition for this rate-limiting enzyme (Fig. 2).

Table 7
MSPE coefficient of variation and its decomposition in the model 3 predicting fatty acids composition in mouse liver during fasting.

| Genotype | PPAR α knockout mice | | | | Wild-type mice | | | |
|-------------------------------------|-----------------------------|---------------------|--------|--------|----------------|--------|--------|--------|
| | Statistics | cvMSPE ^a | CE (%) | RE (%) | DE (%) | cvMSPE | CE (%) | RE (%) |
| C14:0 | 2.77e-01 | 0 | 24 | 76 | 3.23e-01 | 0 | 12 | 88 |
| C16:0 | 3.16e-01 | 6 | 57 | 37 | 2.47e-01 | 5 | 42 | 53 |
| C16:1 ω 9 | 2.52e-01 | 0 | 13 | 87 | 3.17e-01 | 3 | 12 | 85 |
| C16:1 ω 7 | 7.73e-01 | 12 | 60 | 28 | 4e-01 | 1 | 28 | 71 |
| C18:0 | 1.35e-01 | 34 | 5 | 60 | 7.96e-02 | 13 | 3 | 84 |
| C18:1 ω 9 | 2.86e-01 | 53 | 11 | 36 | 2.31e-01 | 0 | 1 | 99 |
| C18:1 ω 7 | 2.22e-01 | 43 | 4 | 53 | 1.93e-01 | 2 | 11 | 87 |
| C18:2 ω 6 | 2.91e-01 | 55 | 7 | 38 | 2.09e-01 | 1 | 3 | 96 |
| C18:3 ω 3 | 4.41e-01 | 0 | 56 | 44 | 5.08e-01 | 56 | 1 | 44 |
| C20:1 ω 9 | 3.95e-01 | 37 | 36 | 27 | 3.57e-01 | 1 | 26 | 74 |
| C20:3 ω 6 | 1.88e-01 | 1 | 1 | 98 | 1.30e-01 | 28 | 3 | 69 |
| C20:4 ω 6 | 9.94e-02 | 3 | 2 | 95 | 7.25e-02 | 6 | 30 | 64 |
| C20:5 ω 3 | 4.7e-01 | 0 | 54 | 46 | 1.5e-01 | 13 | 4 | 83 |
| C22:6 ω 3 | 1.40e-01 | 14 | 16 | 70 | 91.19e-01 | 51 | 20 | 29 |
| Total FA | 2.12e-01 | 37 | 0 | 63 | 1.80e-01 | 3 | 5 | 91 |
| Total ω 9 ^b | 2.79e-01 | 51 | 12 | 37 | 2.33e-01 | 0 | 2 | 98 |
| Total ω 3 ^c | 2.37e-01 | 2 | 52 | 47 | 1.68e-01 | 60 | 8 | 33 |
| Total ω 6 ^d | 2.75e-01 | 54 | 6 | 39 | 1.8e-01 | 1 | 4 | 95 |
| C20:5 ω 3 + C22:6 ω 3 | 1.50e-01 | 12 | 20 | 68 | 1.17e-01 | 51 | 17 | 32 |
| C20:3 ω 6 + C20:4 ω 6 | 1.04e-01 | 2 | 1 | 97 | 7.34e-02 | 9 | 33 | 57 |
| Total sat + mono ^e | 1.94e-01 | 12 | 10 | 78 | 1.89e-01 | 3 | 9 | 89 |

This model includes fatty acid influx from adipose tissue, oxidation of fatty acids and biosynthesis of polyunsaturated fatty acids.

^a MSPE, CE, RE and DE are defined in footnote a of Table 5.

^b C16:1 ω 9 + C18:1 ω 9 + C20:1 ω 9.

^c Sum of C18:3 ω 3, C20:5 ω 3, C22:6 ω 3.

^d Sum of C18:2 ω 6, C20:3 ω 6, C20:4 ω 6.

^e Sum of saturated (C14:0 + C16:0 + C18:0) and monounsaturated (C16:1 ω 7, C16:1 ω 9, C18:1 ω 7, C18:1 ω 9) fatty acids.

3.4.1. Elongation and desaturation process is active during fasting in both genotypes

For both genotypes model 3 (FA uptake, oxidation and elongation/desaturation) fitted the observed accumulation of C22:6 ω 3 better (Table 7) than model 2 (Table 6). This suggests that elongation and desaturation of C18:3 ω 3 to C22:6 ω 3 are active in both genotypes during fasting. The need to include an active elongation and desaturation pathway in the liver of mice under fasting conditions in PPAR α KO mice is surprising.

The rate of desaturation and elongation is generally considered to be limited by D6D (Cho et al., 1999), which is transcriptionally activated by SREBP1 (Matsuzaka et al., 2002). The nuclear form of SREBP-1 in the liver of mice has a very low, barely detectable level after 6 h of fasting (Horton et al., 1998). A second transcriptional activator of D6D is PPAR α (Matsuzaka et al., 2002). Considering that PPAR α was invalidated in PPAR α KO mice, and that SREBP-1 is inhibited by fasting (Horton et al., 1998), transcription and activity of D6D should have rapidly decreased during the first hours of fasting in PPAR α KO livers.

3.4.2. Experimental validation: mRNA expression of D6D is stable during the first 24 h of fasting

To support at least in part the results of our model, mRNA levels of D6D were monitored by real-time quantitative PCR in the liver of mice after 0, 24, 48, and 72 h of fasting. As shown in Fig. 5, D6D mRNA levels remained constant during the first 24 h of fasting in the two genotypes.

D6D mRNA levels significantly decreased at 72 and 48 h in wild-type and PPAR α KO livers respectively. Therefore, D6D expression seems to be regulated in a time-dependent manner during fasting and its regulation partially depends on PPAR α expression only after 24 h. This result is consistent with the need to incorporate the desaturation/elongation pathway in our model for both genotypes to correctly fit the FA data. Additionally, these data suggest that the mechanisms by which D6D mRNA expres-

sion is regulated during fasting involve other molecular players than the known major regulators PPAR α and SREBP1. These unknown regulators remain to be identified but could respond to several hormonal or metabolic signals that are modulated during fasting, possibly differentially between wild-type and PPAR α KO mice. Several hormones have been shown to reduce D6D expression or activity in the liver (for a review see Brenner, 2003) including glucagon and glucocorticoids which are increased during fasting. However, to our knowledge, the accurate kinetics of these hormonal changes have not yet been described in wild-type and PPAR α knockout mice. Hence, it is difficult to predict whether such hormonal signal may differentially influence D6D expression between these two genotypes. On the other hand, several metabolic parameters differ between fasted wild-type and PPAR α KO mice. For instance, the latter exhibit hypoglycemia and reduced metabolic rates during fasting (Kersten et al., 1999) in addition to impaired FA oxidation (Le May et al., 2000). In the liver, glucose and FA (Guillou et al., 2008) may influence the transcription of enzymes involved in FA metabolism. Changes in glucose, FA or other metabolites that may influence the differential expression of D6D reported here remain to be investigated.

4. Conclusion

A simple model including fatty acids uptake, oxidation and elongation/desaturation was able to predict correctly the variation of most fatty acids in the liver of both PPAR α KO and wild-type mice. This model included parameter estimates in adipose tissue and liver to explain the change in fatty acid content in the liver during fasting.

Experimental measurements in different organs obtained in the same animals and under the same experimental conditions are strongly needed in the future to predict the dynamics of fatty acids in a given organ. The presence of a basal oxidation in both PPAR α KO

and wild-type mice with a rate that depended on the genotype but not on the time of fasting shows that PPAR α has a constitutive effect on oxidation but little or no time-dependent effects.

The presence of an active elongation and desaturation in both genotypes was surprising in the liver of fasting mice. It confirms and strengthens earlier results (Nakamura and Nara, 2003) suggesting that the unsaturated fatty acids content in tissues is maintained within physiological ranges by feedback regulation of synthetic pathways.

The regulation of desaturases by PPAR α is different from the main role of PPAR α in inducing oxidation (Nakamura and Nara, 2003) and cannot be explained by the differential behavior of various desaturases such as delta-6-desaturase and stearoyl-coenzyme A desaturase (Radulescu et al., 2006). Our results suggest that mechanisms other than PPAR α activation are likely to contribute to the regulation of delta-6-desaturase activity during fasting.

In the future, the model could be used to design specific experiments aiming at a better understanding of lipid metabolism as regulated by the nutritional status. The same set of equations could be used in other tissues such as muscle, in other animal species, to predict inter-species variability in lipid metabolism or in animals fed diets with different FA compositions.

References

- Aoyama, T., Peters, J., Iritani, N., Nakajima, T., Furihata, K., Hashimoto, T., Gonzalez, F., 1998. Altered constitutive expression of fatty acid-metabolizing enzymes in mice lacking the peroxisome proliferator-activated receptor α (PPAR α). *Journal of Biological Chemistry* 273 (10), 5678–5684.
- Berk, P., Stump, D., 1999. Mechanisms of cellular uptake of long chain free fatty acids. *Molecular and Cellular Biochemistry* 192 (1), 17–31.
- Bibby, J., 1977. *Predictions and Improved Estimation in Linear Models*. Wiley, Chichester.
- Bookout, A., Jeong, Y., Downes, M., Yu, R., Evans, R., Mangelsdorf, D., 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. *Cell* 126 (4), 789–799.
- Brenner, R., 2003. Hormonal modulation of delta6 and delta5 desaturases: case of diabetes. *Prostaglandins Leukotrienes & Essential Fatty Acids* 68 (2), 151–162.
- Calvetti, D., Kucevski, A., Somersalo, E., 2008. A mathematical model of liver metabolism: from steady state to dynamic. *Journal of Physics: Conference Series* 124(1), 012012.
- Campbell, P., Carlson, M., Hill, J., Nurjhan, N., 1992. Regulation of free fatty acid metabolism by insulin in humans: role of lipolysis and reesterification. *American Journal of Physiology—Endocrinology and Metabolism* 263 (6), 1063–1069.
- Cao, H., Gerhold, K., Mayers, J., Wieser, M., Watkins, S., Hotamisligil, G., 2008. Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* 134 (6), 933–944.
- Chalhoub, E., Hanson, R., Belovich, J., 2007a. A computer model of gluconeogenesis and lipid metabolism in the perfused liver. *American Journal of Physiology—Endocrinology and Metabolism*, 00161–2007.
- Chalhoub, E., Xie, L., Balasubramanian, V., Kim, J., Belovich, J., 2007b. A distributed model of carbohydrate transport and metabolism in the liver during rest and high-intensity exercise. *Annals of Biomedical Engineering* 35 (3), 474–491.
- Cho, H., Nakamura, M., Clarke, S., 1999. Cloning expression and nutritional regulation of the mammalian Δ -6 desaturase. *Journal of Biological Chemistry* 274 (1), 471–477.
- Costet, P., Legendre, C., More, J., Edgar, A., Galtier, P., Pineau, T., 1998. Peroxisome proliferator-activated receptor α -isoform deficiency leads to progressive dyslipidemia with sexually dimorphic obesity and steatosis. *Journal of Biological Chemistry* 273 (45), 29577–29585.
- Desvergne, B., Wahli, W., 1999. Peroxisome proliferator-activated receptors: nuclear control of metabolism I. *Endocrine Reviews* 20 (5), 649–688.
- Dupuis, E., Glorian, M., Forest, C., 2000. Fatty acid regulation of gene transcription. *Journal of Biological Chemistry* 275 (40), 30749–30752.
- Fattal, D., Ben-Shaul, A., 1993. A molecular model for lipid-protein interaction in membranes: the role of hydrophobic mismatch. *Biophysical Journal* 65 (5), 1795–1809.
- Forns, X., Ampurdanes, S., Llovet, J., Aponte, J., Quinto, L., Martinez-Bauer, E., Bruguera, M., Sanchez-Tapias, J., Rodes, J., 2002. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 36 (4), 986–992.
- Fritz, I., 1959. Action of carnitine on long chain fatty acid oxidation by liver. *American Journal of Physiology* 197 (2), 297–304.
- Guillou, H., Martin, P., Jan, S., D'Andrea, S., Roulet, A., Catheline, D., Rioux, V., Pineau, T., Legrand, P., 2002. Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice. *Lipids* 37 (10), 981–989.
- Guillou, H., Martin, P., Pineau, T., 2008. Transcriptional regulation of hepatic fatty acid metabolism. *Subcellular Biochemistry* 49, 3–47.
- Hashimoto, T., Cook, W., Qi, C., Yeldandi, A., Reddy, J., Rao, M., 2000. Defect in peroxisome proliferator-activated receptor α -inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting. *Journal of Biological Chemistry* 275 (37), 28918–28928.
- Hiltunen, J., Karki, T., Hassinen, I., Osmundsen, H., 1986. Beta oxidation of polyunsaturated fatty acids by rat liver peroxisomes: a role for 2, 4-dienoyl-coenzyme A reductase in peroxisomal beta-oxidation. *Journal of Biological Chemistry* 261 (35), 16484–16493.
- Horton, J., Bashmakov, Y., Shimomura, I., Shimano, H., 1998. Regulation of sterol regulatory element binding proteins in livers of fasted and re-fed mice. *Proceedings of the National Academy of Sciences* 95 (11), 5987–5992.
- Horton, J., Goldstein, J., Brown, M., 2002. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation* 109 (9), 1125.
- Jakobsson, A., Westerberg, R., Jacobsson, A., 2006. Fatty acid elongases in mammals: their regulation and roles in metabolism. *Progress in Lipid Research* 45 (3), 237–249.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., Yamaniishi, Y., 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Research* 36, D480–D484.
- Kersten, S., Seydoux, J., Peters, J., Gonzalez, F., Desvergne, B., Wahli, W., 1999. Peroxisome proliferator-activated receptor α mediates the adaptive response to fasting. *Journal of Clinical Investigation* 103 (11), 1489.
- Kim, J., Sarraf, P., Wright, M., Yao, K., Mueller, E., Solanes, G., Lowell, B., Spiegelman, B., 1998. Nutritional and insulin regulation of fatty acid synthetase and leptin gene expression through ADD1/SREBP1. *Journal of Clinical Investigation* 101 (1), 1.
- Knauf, C., Rieusset, J., Foretz, M., Cani, P., Udry, M., Hosokawa, M., Martinez, E., Bringart, M., Waget, A., Kersten, S., Desvergne, B., Gremlich, S., Wahli, W., Seydoux, J., Delzenne, N., Thorens, B., Burcelin, R., 2006. Peroxisome proliferator-activated receptor-alpha-null mice have increased white adipose tissue glucose utilization, glut4, and fat mass: role in liver and brain. *Endocrinology* 147 (9), 4067–4078.
- Le May, C., Pineau, T., Bigot, K., Kohl, C., Girard, J., Pégurier, J., 2000. Reduced hepatic fatty acid oxidation in fasting PPAR α null mice is due to impaired mitochondrial hydroxymethylglutaryl-coA synthase gene expression. *FEBS Letters* 475 (3), 163–166.
- Lee, S., Chan, W., Lo, C., Wan, D., Tsang, D., Cheung, W., 2004. Requirement of PPAR α in maintaining phospholipid and triglyceride homeostasis during energy deprivation. *The Journal of Lipid Research*, M400078.
- Lee, S., Pineau, T., Drago, J., Lee, E., Owens, J., Kroetz, D., Fernandez-Salguero, P., Westphal, H., Gonzalez, F., 1995. Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Molecular and Cellular Biology* 15 (6), 3012–3022.
- Leone, T., Weinheimer, C., Kelly, D., 1999. A critical role for the peroxisome proliferator-activated receptor α (PPAR α) in the cellular fasting response: The PPAR α -null mouse as a model of fatty acid oxidation disorders. *Proceedings of the National Academy of Sciences* 96 (13), 7473–7478.
- Mahler, H., Wakil, S., Bock, R., 1953. Studies on fatty acid oxidation I: enzymatic activation of fatty acids. *Journal of Biological Chemistry* 204 (1), 453–468.
- Mannaerts, G., Debeer, L., Thomas, J., De Schepper, P., 1979. Mitochondrial and peroxisomal fatty acid oxidation in liver homogenates and isolated hepatocytes from control and clofibrate-treated rats. *Journal of Biological Chemistry* 254 (11), 4585–4595.
- Martin, P., Guillou, H., Lasserre, F., Déjean, S., Lan, A., Pascucci, J., San Cristobal, M., Legrand, P., Besse, P., Pineau, T., 2007. Novel aspects of PPAR α -mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology* 45 (3), 767–777.
- Matsuzaka, T., Shimano, H., Yahagi, N., Amemiya-Kudo, M., Yoshikawa, T., Hasty, A., Tamura, Y., Osuga, J., Okazaki, H., Iizuka, Y., et al., 2002. Dual regulation of mouse Δ -5- and Δ -6-desaturase gene expression by SREBP-1 and PPAR α . *The Journal of Lipid Research* 43 (1), 107–114.
- Mazzucotelli, A., Viguerie, N., Tiraby, C., Annicotte, J., Mairal, A., Klimcakova, E., Lepin, E., Delmar, P., Dejean, S., Tavernier, G., Lefort, C., Hidalgo, J., Pineau, T., Fajas, L., Clément, K., Langin, D., 2007. The transcriptional coactivator peroxisome proliferator-activated receptor (PPAR)gamma coactivator-1 alpha and the nuclear receptor PPAR alpha control the expression of glycerol kinase and metabolism genes independently of PPAR gamma activation in human white adipocytes. *Diabetes* 56 (10), 2467–2475.
- Muiruri, K., Leveille, G., 1970. In vitro fatty acid synthesis and enzyme activity in liver and adipose tissue of the mouse. *International Journal of Biochemistry* 1, 254–256.
- Murray, R., Granner, D., Rodwell, V., 2006. *Harper's Illustrated Biochemistry*, 27th ed The McGraw-Hill Companies.
- Nakamura, M., Nara, T., 2003. Essential fatty acid synthesis and its regulation in mammals. *Prostaglandins Leukotrienes & Essential Fatty Acids* 68 (2), 145–150.
- Nelder, J., Mead, R., 1965. A simplex method for function minimization. *The Computer Journal* 7 (4), 308.
- Oikari S., Ahtialansaari, T., Huotari, A., Kiehne, K., Folsch, U., Wolfram, S., Janne, J., Alhonen, L., Herzog, K., 2008. Effect of medium and long chain fatty acid diet on PPARs and SREBP-1 expression and glucose homeostasis in ACBP over-expressing transgenic rats. *Acta Physiol (Oxf)* 194 (1), 57–65.

- Peirson, S., Butler, J., Foster, R., 2003. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research* 31 (14), e73.
- Radulescu, O., Lagarrigue, S., Siegel, A., Veber, P., Le Borgne, M., 2006. Topology and static response of interaction networks in molecular biology. *Journal of the Royal Society Interface* 3 (6), 185–196.
- Remesy, C., Denigne, C., 1983. Changes in availability of glucogenic and ketogenic substrates and liver metabolism in fed or starved rats. *Annals of Nutrition and Metabolism* 27 (1), 57–70.
- Seidelin, K., 1995. Fatty acid composition of adipose tissue in humans: implications for the dietary fat-serum cholesterol-CHD issue. *Progress in Lipid Research* 34 (3), 199–217.
- Sessler, A., Ntambi, J., 1998. Polyunsaturated fatty acid regulation of gene expression. *Journal of Nutrition* 128 (6), 923–926.
- Shimano, H., Yahagi, N., Amemiya-Kudo, M., Hasty, A., Osuga, J., Tamura, Y., Shionoiri, F., Iizuka, Y., Ohashi, K., Harada, K., et al., 1999. Sterol regulatory element-binding protein-1 as a key transcription factor for nutritional induction of lipogenic enzyme genes. *Journal of Biological Chemistry* 274 (50), 35832–35839.
- Shindo, Y., Hashimoto, T., 1978. Acyl-coenzyme a synthetase and fatty acid oxidation in rat liver peroxisomes. *Journal of Biochemistry* 84 (5), 1177.
- Shirai, K., 2004. Obesity as the core of the metabolic syndrome and the management of coronary heart disease. *Current Medical Research and Opinion* 20 (3), 295–304.
- Simopoulos, A., 1991. Omega-3 fatty acids in health and disease and in growth and development. *American Journal of Clinical Nutrition* 54, 438 (see comments).
- Sokolović, M., Sokolović, A., Wehkamp, D., van Themaat, E., de Waart, D., Gilhuijs-Pederson, L., Nikolsky, Y., van Kampen, A., Hakvoort, T., Lamers, W., 2008. The transcriptomic signature of fasting murine liver. *BMC Genomics* 9, 528.
- Sorrentino, D., Berk, P., 1993. Free fatty acids and the sinusoidal plasma membrane: concepts, trends and controversies. In: Tavoloni, N., Berk, P.D. (Eds.), *Hepatic Transport and Bile Secretion: Physiology and Pathophysiology*, Raven Press, New York, USA.
- Tedeschi, L., 2006. Assessment of the adequacy of mathematical models. *Agricultural Systems* 89 (2–3), 225–247.
- Uauy, R., Mena, P., Rojas, C., 2007. Essential fatty acids in early life: structural and functional role. *Proceedings of the Nutrition Society* 59 (01), 3–15.
- Unger, R., Zhou, Y., Orci, L., 1999. Regulation of fatty acid homeostasis in cells: novel role of leptin. *Proceedings of the National Academy of Sciences* 96 (5), 2327–2332.
- van Milgen, J., 2002. Modeling biochemical aspects of energy metabolism in mammals I. *Journal of Nutrition* 132 (10), 3195–3202.
- Wang, Y., Botolin, D., Xu, J., Christian, B., Mitchell, E., Jayaprakasam, B., Nair, M., Peters, J., Busik, J., Olson, L., et al., 2006. Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity. *The Journal of Lipid Research* 47 (9), 2028.

Conclusion

Une démarche de construction de modèle générique a été développée afin d'établir et de faire évoluer de manière ascendante des modèles comparables entre eux. Cette méthode a été appliquée pour décrire et comprendre la hiérarchie d'importance des voies métaboliques vis-à-vis de la composition en acides gras dans le foie de souris en relation avec leur génotype.

Elle a permis de construire un modèle très simple, de mettre en évidence la hiérarchie d'importance des voies métaboliques dans ce contexte et d'identifier l'existence d'une régulation particulière des désaturases lors du jeûne, y compris en l'absence constitutionnelle de PPAR α . Ces résultats ont permis d'améliorer le squelette d'équation du modèle de base par l'ajout de l'élongation-désaturation, une voie bien décrite dans la bibliographie, et par une nouvelle régulation hypothétique de cette voie.

Conclusion et perspectives

Conclusion

La recherche d'éléments clefs du métabolisme des lipides est une problématique fondamentalement dépendante des données expérimentales et bibliographiques disponibles. En raison de l'hétérogénéité des données, deux approches complémentaires ont été développées pour exploiter différents types de données disponibles.

Afin d'exploiter les données dynamiques et quantitatives obtenues sur quelques métabolites, nous avons développé un modèle dynamique abstrait du métabolisme des lipides. Ce modèle a permis d'intégrer des données métaboliques obtenues dans le foie de souris et d'identifier la hiérarchie d'importance des voies métaboliques vis-à-vis de la composition hépatique en acides gras au cours du jeûne. De plus, afin d'exploiter la généralité du métabolisme des lipides, une méthodologie de construction ascendante de modèles a été développée. Cette méthodologie permet de comparer des modèles obtenus dans des contextes différents. En pratique, elle a permis de comparer des souris sauvages avec des souris dans lesquelles PPAR α est invalidé constitutionnellement (souris *knockout* PPAR α ^{-/-}), et de déduire de cette comparaison l'existence d'un régulateur des désaturases de la voie des ω 3 et ω 6 autre que PPAR α et SREBP1. Ce résultat était inattendu, même après l'analyse statistique des données initiales.

L'exploitation des données transcriptomiques par les méthodes d'analyse statistiques classiques a permis de mettre en évidence des ensembles de gènes coexprimés. Ils ont alors été annotés en les croisant avec des ontologies afin de les caractériser fonctionnellement. Ceci a permis d'identifier les principales fonctions régulées transcriptomiquement au cours des conditions expérimentales. L'analyse manuelle de certains éléments a permis une interprétation plus fine des groupes de gènes coexprimés. Que ce soit lors de l'étude détaillée de la régulation de la transcription du foie de poulet au cours du jeûne, ou lors de l'étude bibliographique de quelques publications relatives à l'analyse de données transcriptionnelles, l'exploitation manuelle des informations bibliographiques a été identifiée comme une étape limitante de l'exploitation des données haut débit.

Afin d'évaluer la quantité, la qualité et l'accessibilité des informations bibliographiques disponibles pour une exploitation à haut débit, nous avons expertisé trois bases de données bibliographiques disponibles au démarrage de cette thèse. Cette analyse a permis de quantifier la couverture des bases de données analysées vis-à-vis de l'analyse d'une liste de gènes relatifs au métabolisme énergétique. Elle a permis de faire ressortir la complémentarité de ces bases en terme d'influences et de contenu bibliographique et la nécessité d'exploiter les bases de données généralistes en parallèle de la réalisation d'une base experte.

A partir des bases analysées, nous avons extrait un graphe d'influences. En première approche, l'analyse topologique du graphe d'influences a permis d'identifier les éléments les plus connectés ayant un lien avec le métabolisme énergétique. Elle a révélé des éléments bien connus du métabolisme énergétique mais d'autres *a priori* plus inattendus car essentiellement décrits pour avoir un rôle dans d'autres fonctions (exemple : les régulateurs de la croissance cellulaire). De plus, le graphe d'influence est un formalisme adapté pour exploiter des comparaisons entre conditions comme par exemple celles décrites par les données issues de *microarray*. Nous espérons donc utiliser ce dernier pour interpréter les données haut débit disponibles.

Perspectives

Perspectives à court terme

Améliorer le graphe d'influences extrait

Le graphe d'influences construit n'exploite pas toutes les données présentes dans les bases sources, en particulier à cause des problèmes de synonymes. Il serait donc intéressant d'ajouter les identifiants CAS¹ à la liste des bases de référence afin de pouvoir mieux exploiter l'information présente dans TRANSPATH. Introduire CAS nécessite de modéliser une relation entre CAS et HGNC pour prendre en compte le lien entre chaque gène et ses protéines, et de traiter les synonymes pour les métabolites identifiés par KEGG.

TRANSPATH et Ingenuity contiennent des complexes que nous n'exploitons pas, il serait donc intéressant de modéliser proprement la description des complexes. Pour les complexes dont le nombre de molécules est connu, il suffira de trier par ordre alphabétique les identifiants des molécules formant le complexe et de préciser leur nombre. Par contre lorsque ce nombre est inconnu, il n'est pas évident de statuer sur ce qui doit être considéré comme synonyme.

La base de données Pathway Commons contient actuellement beaucoup plus d'informations qu'au début de cette thèse. De plus, l'information présente dans cette base est organisée de manière à être exploitée à haut débit. Par conséquent il serait utile d'évaluer la pertinence de cette base vis-à-vis du métabolisme des lipides, puis éventuellement d'intégrer cette dernière au graphe d'influences comme l'ont été la base experte Gardon et les bases généralistes TRANSPATH et Ingenuity.

Une autre option plus compliquée est d'améliorer l'extraction d'Ingenuity en essayant d'obtenir de la documentation ou en utilisant les technologies d'analyse de langage naturel pour analyser de manière plus approfondie les phrases présentes dans l'interface utilisateur d'Ingenuity. En effet, l'analyse de ces phrases est un cas simple d'analyse de langage naturel dans lequel les règles de grammaire sont relativement peu nombreuses et pour lequel une même phrase a toujours le même sens [183].

Exploiter le graphe d'influences

Le graphe d'influences construit au cours de cette thèse n'a pas été exploité dans sa totalité. La méthode utilisée pour identifier des éléments clés fait ressortir les éléments connus pour avoir de nombreuses interactions dont l'une d'entre elles met en jeu un élément du métabolisme de l'énergie. Afin d'obtenir directement des régulateurs clés de ce métabolisme, il serait intéressant d'identifier le plus petit ensemble d'éléments dont la variation suffit à expliquer les variations observées des éléments connus du métabolisme énergétique. Néanmoins, cette approche nécessite le développement d'algorithmes dont le temps de calcul n'est pas connu *a priori* et repose sur la qualité du graphe d'influences construit.

Il serait aussi intéressant d'analyser plus en détail la topologie globale du réseau construit et de croiser cette dernière avec les informations fonctionnelles présentes dans les ontologies et les informations expérimentales. On obtiendrait alors un croisement entre trois types de *clusters* : topologiques (composantes fortement connexes par exemples), fonctionnels et expérimentaux (éléments qui varient ensemble). La visualisation de

¹CAS est une base d'annotation de molécules organiques contenant de nombreuses protéines et métabolites.

ces informations nous permettrait en particulier d'avoir une idée générale de l'organisation du métabolisme des lipides.

Exploiter la généricité du modèle dynamique pour réaliser des comparaisons

Confronter le modèle générique du métabolisme lipides avec des données obtenues dans différents organes, espèces et conditions permettrait de réaliser des comparaisons afin d'identifier quelles sont les voies génériques ou spécifiques responsables des différences comportementales observées. Cette comparaison permettra par exemple de savoir si la lipogenèse dans l'hépatocyte de poulet met en jeu les mêmes fonctions biologiques et les mêmes voies de transduction du signal que dans l'adipocyte de porc.

Cette démarche nécessite néanmoins de s'appuyer sur des données obtenues dans des organes où au moins l'entrée des acides gras et la vitesse de leur oxydation est connue, sans quoi ce modèle ne peut pas être déterminé. De plus, au cours de la modélisation il est possible que la mesure d'autres éléments s'avère nécessaire pour décrire correctement le modèle. Il est donc primordial de coopérer dans cette approche avec les biologistes dans la construction des futurs plans d'expériences afin d'obtenir des jeux de données informatifs sur le modèle construit.

Visualiser les connaissances haut débit

Une fois des graphes de plusieurs milliers de sommets construits (tels que les graphes d'influences par exemple), leur exploration reste problématique car peu d'outils permettent de les visualiser. Cytoscape, l'outil le plus couramment utilisé consomme trop de mémoire pour charger un graphe de plusieurs milliers de sommets et ne permet pas de disposer ces derniers de manière organisée sur un plan. Les autres logiciels testés (GraphViz et Yed) ne sont pas vraiment meilleurs.

Il est donc utile d'expertiser de manière plus approfondie l'offre logicielle existante. Dans le cas où cette dernière n'est pas satisfaisante, il serait éventuellement utile à plus long terme d'implémenter quelques algorithmes existants (ex :F. Schreiber *et al.*, [278] ont développé un algorithme de visualisation organisée des réseaux biologiques; Quigley *et al.*, [258] ont développé un algorithme de représentation des graphes relativement rapide) afin de proposer une solution d'exploration et de visualisation.

Perspectives à plus long terme

Intégration verticale

Au cours de cette thèse deux modèles disjoints ont été développés : un vaste modèle qualitatif (i.e. le graphe d'influences) et un petit modèle quantitatif dynamique. Ces deux formalismes sont incompatibles car le modèle qualitatif ne contient pas les informations nécessaires pour déduire une dynamique et car la conversion en qualitatif du modèle dynamique aboutit à une sous exploitation de ses données pour apporter une quantité d'informations minime par rapport à une expérience haut débit.

Pour réunir ces deux modèles nous proposons donc d'exploiter le modèle qualitatif afin d'identifier un ensemble d'éléments potentiellement importants dans le métabolisme des lipides, puis une fois ces éléments identifiés, de les mesurer expérimentalement afin d'avoir leur variation au fil du temps puis de les injecter dans le modèle dynamique afin de tester leur nécessité pour expliquer un comportement et l'effet de leur dynamique sur le système.

Une difficulté importante de cette approche, consiste à définir une frontière autour des éléments identifiés comme importants pour savoir quelles relations il faut inclure dans le modèle lorsqu'ils sont présents. Ceci pose aussi problème pour construire une expérience permettant de tester leur importance ou les mécanismes de leur régulation. Ce point peut s'avérer complexe en raison de l'existence de voies de contournement, de corrélations fortuites, de nombreuses régulations dont l'importance n'est pas connue *a priori*, d'une forte variabilité dans les échelles de temps mises en jeu et de difficultés expérimentales pour obtenir des mesures précises. La construction de telles expériences s'appuie donc à la fois sur le modèle et sur l'expertise des spécialistes pour identifier ce qui est susceptible d'influencer les éléments clefs.

Intégration transversale

De manière générale, l'approche systémique nécessite de pouvoir considérer l'objet étudié comme un tout. Or, les données expérimentales et les connaissances détenues par les experts ne concernent que de petits sous systèmes. De plus il est beaucoup plus simple d'expérimenter, d'analyser, de modéliser et de comprendre des petits sous systèmes qu'un gigantesque système. En contrepartie, une fois le système découpé en sous systèmes, ces derniers doivent pouvoir s'assembler pour fonctionner ensemble.

Pour que les divers sous systèmes puissent fonctionner ensemble, il est nécessaire de définir leur frontière (variables d'entrée et de sortie), les hypothèses nécessaires pour que le modèle soit valide ², et les garanties fournies par le modèle ³. Une fois ces éléments décrits, il est possible de combiner des modèles en reliant la sortie d'un modèle à l'entrée d'un autre et en vérifiant que les garanties du premier modèle suffisent à respecter les hypothèses du second.

Au cours de cette thèse, seul le sous système du modèle générique du métabolisme des lipides correspondant aux fonctions nécessaires pour expliquer la composition en acides gras au cours du jeûne et dans le foie a été évalué. Il serait donc intéressant de tester ce modèle dans d'autres conditions pour évaluer les fonctions actives à l'état nourri et les mécanismes de régulations lors des changements d'alimentation. On obtiendrait alors un modèle générique plus polyvalent, capable de simuler différents contextes.

L'étape suivante est de modéliser le fonctionnement d'un individu en autant de sous systèmes que d'organes responsables du métabolisme des lipides, puis de définir une interface à chaque organe décrivant par exemple les principaux régulateurs externes (ex : hormones) et les molécules qu'il échange (ex : glucose et acides gras). Une version spécifique du modèle générique pourra alors être paramétrée dans chaque organe (en la confrontant avec les données expérimentales adéquates). La dernière étape serait de réunir ensemble les différents sous-systèmes afin de simuler le métabolisme des lipides à l'échelle de l'individu.

La représentation de la connaissance est un problème ouvert

Afin de pouvoir exploiter automatiquement les données disponibles, une description formalisée des connaissances est nécessaire. Elle a pour rôle de fournir un contenu exploitable sur lequel utiliser les algorithmes d'analyse. De plus, il est utile de décrire les hypothèses de raisonnement afin de pouvoir les remettre en cause en cas d'incohérence entre un modèle et une observation et mener ainsi à bien la démarche expérimentale.

Cette thèse a illustré ce problème sur le métabolisme des lipides et a proposé une solution en construisant des graphes d'influences à partir de bases de données de connaissances. Cette solution permet d'obtenir des résultats et constitue une première approche afin d'utiliser les outils actuellement disponibles.

Néanmoins une partie très importante de la bibliographie n'est pas exploitée. Ceci s'explique par la faible couverture des bases de données vis-à-vis de l'ensemble des connaissances comme l'a démontré la comparaison avec la base experte Gardon, par la difficulté à formaliser les informations existantes dans une base de données et par la difficulté à exploiter ces informations une fois organisées dans les bases.

La connaissance est difficile à formaliser Tout d'abord, malgré les efforts réalisés par des consortiums (ex : Biopax, SBML) pour représenter la connaissance biologique, il est souvent difficile d'exprimer dans un langage formel un ensemble de mécanismes entre éléments, et les noms des éléments ne sont pas tous normalisés ⁴. Le niveau de détail des connaissances est aussi très variable. Par exemple certains éléments représentent une molécule (ex : C18 :0) tandis que d'autres représentent une classe dont le contenu est plus

²Par exemple, dans un modèle où l'oxydation des acides gras est constante, une hypothèse nécessaire est que les quantités d'acides gras sont toujours positives.

³Par exemple, un modèle dans lequel l'oxydation des acides gras est proportionnelle à leur quantité et leur entrée/production constante garantit que les acides gras ne s'accumuleront pas indéfiniment.

⁴Un effort récent a été réalisé pour les noms de gènes humains avec HGNC ou pour les protéines avec CAS. Pour les métabolismes et les complexes il n'y a pas de norme couramment admise.

ou moins connu (ex : les acides gras). Les interactions décrivent des mécanismes ⁵ ou des effets ⁶ plus ou moins précis.

La fiabilité des connaissances décrites dans la bibliographie est très variable, allant du phénomène largement observé à l'hypothèse de travail et les connaissances sont souvent extrapolées hors du cadre dans lequel elles ont été validées expérimentalement. La fiabilité d'une extrapolation est elle aussi dépendante de la "distance" entre le contexte expérimental où elles ont été mise en évidence, et celui où elles sont utilisées.

Une autre problématique consiste donc à décrire de manière normalisée le contexte expérimental dans lequel les données et les connaissances ont été obtenues et celui dans lequel elles sont raisonnablement extrapolables.

Conclusion

L'exploitation des bases de connaissances est totalement dépendante de leur contenu et de leur formalisme. Ce point bien que présent en amont de la thèse reste un élément crucial de l'intégration à haut débit des connaissances, et constitue un important défi dans la compréhension globale des systèmes biologiques.

De même, pour la construction d'un modèle dynamique, un petit nombre de données est suffisant. Ces données doivent au minimum décrire des cinétiques aux bonnes échelles de temps et contenir la mesure des éléments présents à la frontière du modèle. Dans le cadre où la mécanique interne du modèle est étudiée, ces données doivent en plus contenir la mesure d'un ensemble d'éléments internes au modèle. Il n'est par contre pas toujours évident de définir *a priori* le bon ensemble d'éléments à mesurer ni la ou les bonnes échelles de temps ce qui nécessite un ensemble d'aller-retour entre le modèle (pour construire l'expérience) et l'expérience pour améliorer le modèle.

Par conséquent, la démarche de modélisation doit être pensée en amont de la production de connaissances ou de l'organisation des connaissances existantes afin de définir un cadre général dans lequel les connaissances sont exploitables. La construction de ce cadre est un important défi pour la compréhension des système qui nécessite d'unir des compétences dispersées dans de nombreux champs disciplinaires et dans de nombreuses communautés scientifiques.

⁵Ex : phosphoryle, est impliqué dans une réaction. . .

⁶Ex : active, inhibe, augmente la quantité de X moles/heure . . .

Bibliographie

- [1] N. Abumrad, C. Harmon, and A. Ibrahimi. Membrane transport of long-chain fatty acids evidence for a facilitated process. *The Journal of Lipid Research*, 39(12) :2309–2318, 1998.
- [2] G. Adomian. *Nonlinear stochastic operator equations*. Academic Press, 1986.
- [3] U. Alon. *An introduction to systems biology : design principles of biological circuits*. CRC Press, 2006.
- [4] S.W. Altmann, H.R. Davis, L. Zhu, X. Yao, L.M. Hoos, G. Tetzloff, S.P.N. Iyer, M. Maguire, A. Golovko, M. Zeng, et al. Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption, 2004.
- [5] S. Ananiadou, D.B. Kell, and J. Tsujii. Text mining and its potential applications in systems biology. *Trends in biotechnology*, 24(12) :571–579, 2006.
- [6] C. Andreolas, G. da Silva Xavier, F. Diraison, C. Zhao, A. Varadi, F. Lopez-Casillas, P. Ferré, F. Foufelle, and G.A. Rutter. Stimulation of acetyl-CoA carboxylase gene expression by glucose requires insulin release and sterol regulatory element binding protein 1c in pancreatic MIN6 β -cells. *Diabetes*, 51(8) :2536, 2002.
- [7] M.W. Anthonsen, L. Ronnstrand, C. Wernstedt, E. Degerman, and C. Holm. Identification of novel phosphorylation sites in hormone-sensitive lipase that are phosphorylated in response to isoproterenol and govern activation properties in vitro. *Journal of Biological Chemistry*, 273(1) :215–221, 1998.
- [8] T. Aoyama, J.M. Peters, N. Iritani, T. Nakajima, K. Furihata, T. Hashimoto, and F.J. Gonzalez. Altered Constitutive Expression of Fatty Acid-metabolizing Enzymes in Mice Lacking the Peroxisome Proliferator-activated Receptor α (PPAR α). *Journal of Biological Chemistry*, 273(10) :5678–5684, 1998.
- [9] T. Aoyama, J.M. Peters, N. Iritani, T. Nakajima, K. Furihata, T. Hashimoto, and F.J. Gonzalez. Altered constitutive expression of fatty acid-metabolizing enzymes in mice lacking the peroxisome proliferator-activated receptor α (PPAR α). *Journal of Biological Chemistry*, 273(10) :5678–5684, 1998.
- [10] O. Aparicio, JV Geisberg, and K. Struhl. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Current protocols in cell biology/editorial board, Juan S. Bonifacino...[et al.]*, 2004.
- [11] MS Ayuso-Parrilla, A. Martin-Requero, J. Pérez-Días, and R. Parrilla. Role of glucagon on the control of hepatic protein synthesis and degradation in the rat in vivo. *Journal of Biological Chemistry*, 251(24) :7785–7790, 1976.
- [12] A.L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439) :509, 1999.
- [13] J.C. Barnes and J. Robertson. Conceptual biology : a semantic issue and more. *Nature*, 417(6889) :587–588, 2002.
- [14] R. Bartrons, L. Hue, E. Van Schaftingen, and HG Hers. Hormonal control of fructose 2, 6-bisphosphate concentration in isolated rat hepatocytes. *Biochemical Journal*, 214(3) :829, 1983.
- [15] M. Bauer, A.C. Hamm, M. Bonaus, A. Jacob, J. Jaekel, H. Schorle, M.J. Pankratz, and J.D. Katzenberger. Starvation response in mouse liver shows strong correlation with life-span-prolonging processes. *Physiological Genomics*, 17(2) :230, 2004.

- [16] M. Bauer, A.C. Hamm, M. Bonaus, A. Jacob, J. Jaekel, H. Schorle, M.J. Pankratz, and J.D. Katzenberger. Starvation response in mouse liver shows strong correlation with life-span-prolonging processes. *Physiological Genomics*, 17(2) :230, 2004.
- [17] D.E. Bauman and W.B. Currie. Partitioning of nutrients during pregnancy and lactation : a review of mechanisms involving homeostasis and homeorhesis. *Journal of Dairy Science*, 63(9) :1514, 1980.
- [18] T. Bayes and T. Bayes. An essay towards solving a problem in the doctrine of chances. 1763. *MD Comput*, 8(3) :157–71, 1991.
- [19] D. Bécard, I. Hainault, D. Azzout-Marniche, L. Bertry-Coussot, P. Ferré, and F. Fougère. Adenovirus-mediated overexpression of sterol regulatory element binding protein-1c mimics insulin effects on hepatic gene expression and glucose homeostasis in diabetic mice. *Diabetes*, 50(11) :2425, 2001.
- [20] K.G. Becker, D.A. Hosack, G. Dennis, R.A. Lempicki, T.J. Bright, C. Cheadle, and J. Engel. PubMatrix : a tool for multiplex literature mining. *BMC bioinformatics*, 4(1) :61, 2003.
- [21] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
- [22] C.R. Benton, D.P.Y. Koonen, J. Calles-Escandon, N.N. Tandon, J.F.C. Glatz, J.J.F.P. Luiken, J.J. Heikkilä, and A. Bonen. Differential effects of contraction and PPAR agonists on the expression of fatty acid transporters in rat skeletal muscle. *The Journal of Physiology*, 573(1) :199, 2006.
- [23] RN Bergman, YZ Ider, CR Bowden, and C. Cobelli. Quantitative estimation of insulin sensitivity. *American Journal of Physiology- Gastrointestinal and Liver Physiology*, 236(6) :667, 1979.
- [24] P. Besnard and I. Niot. Role of lipid binding protein in intestinal absorption of long chain fatty acids. *Fat Digestion and Absorption*, pages 96–118, 2000.
- [25] PD Bewsher and J. Ashmore. Ketogenic and lipolytic effects of glucagon on liver. *Biochemical and biophysical research communications*, 24(3) :431, 1966.
- [26] M. Beylot. Regulation of in vivo ketogenesis : role of free fatty acids and control by epinephrine, thyroid hormones, insulin and glucagon. *Diabetes & metabolism*, 22(5) :299, 1996.
- [27] J. Björkegren, A. Hamsten, RW Milne, and F. Karpe. Alterations of VLDL composition during alimentary lipemia. *The Journal of Lipid Research*, 38(2) :301–314, 1997.
- [28] J.A. Blake, J.E. Richardson, C.J. Bult, J.A. Kadin, and J.T. Eppig. MGD : the mouse genome database. *Nucleic Acids Research*, 31(1) :193, 2003.
- [29] P. Blavy, F. Gondret, H. Guillou, S. Lagarrigue, P. Martin, O. Radulescu, A. Siegel, and J. Van Milgen. A minimal and dynamic model for fatty acid metabolism in mouse liver. *Actes des Journées Ouvertes de Biologie, Informatique et Mathématiques-JOBIM'08*, pages 83–88.
- [30] P. Blavy, F. Gondret, H. Guillou, S. Lagarrigue, P.G.P. Martin, J. van Milgen, O. Radulescu, and A. Siegel. A minimal model for hepatic fatty acid balance during fasting : Application to ppar alpha-deficient mice. *Journal of Theoretical Biology*, In Press, Uncorrected Proof :–, 2009.
- [31] C. Bogardus, S. Lillioja, K. Stone, and D. Mott. Correlation between muscle glycogen synthase activity and in vivo insulin action in man. *J Clin Invest*, 73(4) :1185–1190, 1984.
- [32] M. Boizard, X. Le Liepvre, P. Lemarchand, F. Fougère, P. Ferre, and I. Dugail. Obesity-related overexpression of fatty-acid synthase gene in adipose tissue involves sterol regulatory element-binding protein transcription factors. *Journal of Biological Chemistry*, 273(44) :29164, 1998.
- [33] O. Bouchabke-Coussa, M.L. Quashie, J. Seoane-Redondo, M.N. Fortabat, C. Gery, A. Yu, D. Linderme, J. Trouverie, F. Granier, E. Téoulé, et al. ESKIMO 1 is a key gene involved in water economy as well as cold acclimation and salt tolerance. *BMC Plant Biology*, 8(1) :125, 2008.
- [34] JM Bourre. Roles of unsaturated fatty acids (especially omega-3 fatty acids) in the brain at various ages and during ageing. *Journal of Nutrition, Health & Aging*, 8(3) :163–174, 2004.

- [35] L.J. Brady, D.R. Romsos, P.S. Brady, W.G. Bergen, and G.A. Leveille. The effects of fasting on body composition, glucose turnover, enzymes and metabolites in the chicken. *Journal of Nutrition*, 108(4) :648, 1978.
- [36] O. Braissant, F. Fougère, C. Scotto, M. Dauca, and W. Wahli. Differential expression of peroxisome proliferator-activated receptors (PPARs) : tissue distribution of PPAR- α , - β , and - γ in the adult rat. *Endocrinology*, 137(1) :354–366, 1996.
- [37] K.S. Brown and J.P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E*, 68(2) :21904, 2003.
- [38] EA Bryhni, NP Kjos, R. Ofstad, and M. Hunt. Polyunsaturated fat and fish oil in diets for growing-finishing pigs : effects on fatty acid composition and meat, fat, and sausage quality. *Meat Science*, 62(1) :1–8, 2002.
- [39] L.M. Cagen, X. Deng, H.G. Wilcox, E.A. Park, R. Raghov, and M.B. Elam. Insulin activates the rat sterol-regulatory-element-binding protein 1c (SREBP-1c) promoter through the combinatorial actions of SREBP, LXR, Sp-1 and NF-Y cis-acting elements. *Biochemical Journal*, 385(1) :207–216, 2005.
- [40] DF Calabotta, JA Cherry, PB Siegel, and DE Jones. Lipogenesis and lipolysis in fed and fasted chicks from high and low body weight lines. *Poultry science (USA)*, 1985.
- [41] A. Cam and P. Freychet. Effect of insulin on amino acid transport in isolated rat hepatocytes. *Diabetologia*, 15(2) :117–123, 1978.
- [42] D. Cameron-Smith, L.M. Burke, D.J. Angus, R.J. Tunstall, G.R. Cox, A. Bonen, J.A. Hawley, and M. Hargreaves. A short-term, high-fat diet up-regulates lipid metabolism and gene expression in human skeletal muscle. *American Journal of Clinical Nutrition*, 77(2) :313, 2003.
- [43] F.M. Campbell, R. Kozak, A. Wagner, J.Y. Altarejos, J.R.B. Dyck, D.D. Belke, D.L. Severson, D.P. Kelly, and G.D. Lopaschuk. METABOLISM AND BIOENERGETICS-A role for peroxisome proliferator-activated receptor α (PPAR α) in the control of cardiac malonyl-CoA levels. Reduced fatty acid oxidation rates and increased glucose. *Journal of Biological Chemistry*, 277(6) :4098–4103, 2002.
- [44] E. Cano and L.C. Mahadevan. Parallel signal processing among mammalian MAPKs. *Trends in Biochemical Sciences*, 20(3) :117–122, 1995.
- [45] RB Ceddia, WN William, and R. Curi. Comparing effects of leptin and insulin on glucose metabolism in skeletal muscle : evidence for an effect of leptin on glucose uptake and decarboxylation. *International journal of obesity*, 23(1) :75–82, 1999.
- [46] J.Y. Cha and J.J. Repa. The liver X receptor (LXR) and hepatic lipogenesis : the carbohydrate-response element-binding protein is a target gene of LXR. *Journal of Biological Chemistry*, 282(1) :743, 2007.
- [47] E. Chalhoub, R. Hanson, and J.M. Belovich. A computer model of gluconeogenesis and lipid metabolism in the perfused liver. *American Journal of Physiology- Endocrinology And Metabolism*, pages 00161–2007, 2007.
- [48] A. Chawla, J.J. Repa, R.M. Evans, and D.J. Mangelsdorf. Nuclear receptors and lipid physiology : opening the X-files, 2001.
- [49] G. Chen, G. Liang, J. Ou, J.L. Goldstein, and M.S. Brown. Central role for liver X receptor in insulin-mediated activation of Srebp-1c transcription and stimulation of fatty acid synthesis in liver. *Proceedings of the National Academy of Sciences*, 101(31) :11245, 2004.
- [50] Z. Chen, C.H.C. Cheng, J. Zhang, L. Cao, L. Chen, L. Zhou, Y. Jin, H. Ye, C. Deng, Z. Dai, et al. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences*, 105(35) :12944, 2008.
- [51] Y. Cheon, T.Y. Nara, M.R. Band, J.E. Beever, M.A. Wallig, and M.T. Nakamura. Induction of overlapping genes by fasting and a peroxisome proliferator in pigs : evidence of functional PPAR α in nonproliferating species. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 288(6) :1525–1535, 2005.

- [52] Y. Cheon, T.Y. Nara, M.R. Band, J.E. Beever, M.A. Wallig, and M.T. Nakamura. Induction of overlapping genes by fasting and a peroxisome proliferator in pigs : evidence of functional PPAR α in nonproliferating species. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 288(6) :1525–1535, 2005.
- [53] M. Cherkaoui-Malki, K. Meyer, WQ Cao, N. Latruffe, AV Yeldandi, MS Rao, CA Bradfield, and JK Reddy. Identification of novel peroxisome proliferator-activated receptor alpha (PPARalpha) target genes in mouse liver using cDNA microarray analysis. *Gene expression*, 9(6) :291, 2001.
- [54] Y. Chilliard. Dietary fat and adipose tissue metabolism in ruminants, pigs, and rodents : a review. *Journal of Dairy Science*, 76(12) :3897–3931, 1993.
- [55] H.P. Cho, M.T. Nakamura, and S.D. Clarke. Cloning, expression, and nutritional regulation of the mammalian Δ -6 desaturase. *Journal of Biological Chemistry*, 274(1) :471–477, 1999.
- [56] C. Choi, M. Krull, A. Kel, O. Kel-Margoulis, S. Pistor, A. Potapov, N. Voss, and E. Wingender. TRANSPATH® a high quality database focused on signal transduction. *Comparative and functional genomics*, 5(2) :163–168, 2004.
- [57] K. Chourey, W. Wei, X.F. Wan, and D.K. Thompson. Transcriptome analysis reveals response regulator SO 2426-mediated gene expression in *Shewanella oneidensis* MR-1 under chromate challenge. *BMC genomics*, 9(1) :395, 2008.
- [58] E. Christensen, M. Gronn, TA Hagve, and BO Christophersen. Omega-oxidation of fatty acids studied in isolated liver cells. *Biochim Biophys Acta*, 1081(2) :167–73, 1991.
- [59] G.A. Churchill. Fundamentals of experimental design for cDNA microarrays. *nature genetics*, 32(supp) :490–495, 2002.
- [60] M. Clarke and JA Spudich. Nonmuscle contractile proteins : the role of actin and myosin in cell motility and shape determination. *Annual Review of Biochemistry*, 46(1) :797–822, 1977.
- [61] D. Clémentine, R. Emma, M. Helene, B. Philippe, L. Pascal, and C.B. Muriel. Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics*, 9, 2008.
- [62] GA Cook, RC Nielsen, RA Hawkins, MA Mehlman, MR Lakshmanan, and RL Veech. Effect of glucagon on hepatic malonyl coenzyme A concentration and on lipid synthesis. *Journal of Biological Chemistry*, 252(12) :4421–4424, 1977.
- [63] A. Cornish-Bowden and A. Cornish-Bowden. *Fundamentals of enzyme kinetics*. Portland Press London, 1995.
- [64] B. Corvilain. Lipoprotein metabolism. *Revue médicale de Bruxelles*, 18(1) :3, 1997.
- [65] P. Costet, C. Legendre, J. More, A. Edgar, P. Galtier, and T. Pineau. Peroxisome Proliferator-activated Receptor α -Isoform Deficiency Leads to Progressive Dyslipidemia with Sexually Dimorphic Obesity and Steatosis. *Journal of Biological Chemistry*, 273(45) :29577–29585, 1998.
- [66] M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, and B.O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987) :92–96, 2004.
- [67] M.W. Covert, C.H. Schilling, and B. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of theoretical biology*, 213(1) :73–88, 2001.
- [68] D.A.E. Cross, P.W. Watt, M. Shaw, J. van der Kaay, C.P. Downes, J.C. Holder, and P. Cohen. Insulin activates protein kinaseB, inhibits glycogen synthase kinase-3 and activates glycogen synthase by rapamycin-insensitive pathways in skeletal muscle and adipose tissue. *FEBS letters*, 406(1-2) :211–215, 1997.
- [69] X. Cui and G.A. Churchill. Statistical tests for differential expression in cDNA microarray experiments. *Genome biology*, 4(4) :210, 2003.
- [70] R.K. Curtis, M. Orešič, and A. Vidal-Puig. Pathways to the analysis of microarray data. *TRENDS in Biotechnology*, 23(8) :429–435, 2005.

- [71] P.B. Dallas, N.G. Gottardo, M.J. Firth, A.H. Beesley, K. Hoffmann, P.A. Terry, J.R. Freitas, J.M. Boag, A.J. Cummings, and U.R. Kees. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? *BMC genomics*, 6(1) :59, 2005.
- [72] E. Danforth Jr. Diet and obesity. *American Journal of Clinical Nutrition*, 41(5) :1132, 1985.
- [73] S. Dano, M.F. Madsen, H. Schmidt, and G. Cedersund. Reduction of a biochemical model with preservation of its basic dynamic properties. *FEBS Journal*, 273(21) :4862–4877, 2006.
- [74] M.N. Davies, B.L. O’Callaghan, and H.C. Towle. Glucose activates ChREBP by increasing its rate of nuclear entry and relieving repression of its transcriptional activity. *Journal of Biological Chemistry*, 283(35) :24029, 2008.
- [75] H. De Jong. Modeling and simulation of genetic regulatory systems : a literature review. *Journal of computational biology*, 9(1) :67–103, 2002.
- [76] P.D. Denechaud, R. Dentin, J. Girard, and C. Postic. Role of ChREBP in hepatic steatosis and insulin resistance. *FEBS letters*, 582(1) :68–73, 2008.
- [77] R. Dentin, F. Benhamed, J.P. Pégrier, F. Fougelle, B. Viollet, S. Vaulont, J. Girard, and C. Postic. Polyunsaturated fatty acids suppress glycolytic and lipogenic genes through the inhibition of ChREBP nuclear protein translocation. *Journal of Clinical Investigation*, 115(10) :2843–2854, 2005.
- [78] R. Dentin, J. Girard, and C. Postic. Carbohydrate responsive element binding protein (ChREBP) and sterol regulatory element binding protein-1c (SREBP-1c) : two key regulators of glucose metabolism and lipid synthesis in liver. *Biochimie*, 87(1) :81–86, 2005.
- [79] N. Derome, B. Bougas, SM Rogers, AR Whiteley, A. Labbe, J. Laroche, and L. Bernatchez. Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (*Coregonus* sp., Salmonidae). *Genetics*, 179(4) :1903, 2008.
- [80] N.A. Desbiens. The presence of hypotheses in the scientific literature. *J Philos Sci Law*, 6 :50–7, 2006.
- [81] C. Desert, M.J. Duclos, P. Blavy, F. Lecerf, F. Moreews, C. Klopp, M. Aubry, F. Herault, P. Le Roy, C. Berri, et al. Transcriptome profiling of the feeding-to-fasting transition in chicken liver. *BMC genomics*, 9(1) :611, 2008.
- [82] B. Desvergne, L. Michalik, and W. Wahli. Transcriptional regulation of metabolism. *Physiological reviews*, 86(2) :465–514, 2006.
- [83] B. Desvergne and W. Wahli. Peroxisome Proliferator-Activated Receptors : Nuclear Control of Metabolism. *Endocrine Reviews*, 20(5) :649–688, 1999.
- [84] N. Dif, V. Euthine, E. Gonnet, M. Laville, H. Vidal, and E. Lefai. Insulin activates human sterol-regulatory-element-binding protein-1c (SREBP-1c) promoter through SRE motifs. *Biochemical Journal*, 400(Pt 1) :179, 2006.
- [85] F. Diraison, V. Yankah, D. Letexier, E. Dusserre, P. Jones, and M. Beylot. Differences in the regulation of adipose tissue and liver lipogenesis by carbohydrates in humans. *The Journal of Lipid Research*, 44(4) :846, 2003.
- [86] S. Draghici, P. Khatri, A.C. Eklund, and Z. Szallasi. Reliability and reproducibility issues in DNA microarray measurements. *TRENDS in Genetics*, 22(2) :101–109, 2006.
- [87] E. Duplus, M. Glorian, and C. Forest. Fatty Acid Regulation of Gene Transcription. *Journal of Biological Chemistry*, 275(40) :30749–30752, 2000.
- [88] S. Eaton. Control of mitochondrial β -oxidation flux. *Progress in Lipid Research*, 41(3) :197–239, 2002.
- [89] S. Eaton, K. Bartlett, and M. Pourfarzam. Mammalian mitochondrial beta-oxidation. *Biochem J*, 320(Pt 2) :345–357, 1996.
- [90] J. Ehltling, S.G. Chowrira, N. Mattheus, D.S. Aeschliman, G.I. Arimura, and J. Bohlmann. Comparative transcriptome analysis of *Arabidopsis thaliana* infested by diamond back moth(*Plutella xylostella*) larvae reveals signatures of stress response, secondary metabolism, and signalling. *BMC genomics*, 9(1) :154, 2008.

- [91] R. Eichner, TT Sun, and U. Aebi. The role of keratin subfamilies and keratin pairs in the formation of human epidermal intermediate filaments. *Journal of Cell Biology*, 102(5) :1767–1777, 1986.
- [92] J.E. Elias, W. Haas, B.K. Faherty, and S.P. Gygi. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods*, 2 :667–675, 2005.
- [93] J. Ernst, Q.K. Beg, K.A. Kay, G. Balázsi, Z.N. Oltvai, and Z. Bar-Joseph. A semi-supervised method for predicting transcription factor–gene interactions in *Escherichia coli*. *PLoS Computational Biology*, 4(3), 2008.
- [94] I. Fabregat, E. Revilla, and A. Machado. Short-term control of the pentose phosphate cycle by insulin could be modulated by the NADPH/NADP ratio in rat adipocytes and hepatocytes. *Biochemical and biophysical research communications*, 146(2) :920, 1987.
- [95] FAO. Experts’ recommendations on fats and oils in human nutrition. *Food and Nutrition Paper No. 57*, 1994.
- [96] HR Fatania, TC Vary, and PJ Randle. Modulation of pyruvate dehydrogenase kinase activity in cultured hepatocytes by glucagon and n-octanoate. *Biochemical Journal*, 234(1) :233, 1986.
- [97] A. Faure, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14), 2006.
- [98] J. Ferber. *Multi-agent systems : an introduction to distributed artificial intelligence*. Addison-Wesley, 1998.
- [99] P. Fernandez, J. Di Rienzo, L. Fernandez, H.E. Hopp, N. Paniego, and R.A. Heinz. Transcriptomic identification of candidate genes involved in sunflower responses to chilling and salt stresses based on cDNA microarray analysis. *BMC Plant Biology*, 8(1) :11, 2008.
- [100] X. Fernandez, G. Monin, A. Talmant, J. Mourot, and B. Lebret. Influence of intramuscular fat content on the quality of pig meat. Consumer acceptability of *m. longissimus lumborum*. *Meat Science*, 53(1) :67–72, 1999.
- [101] F. Ferrazzi, P. Magni, L. Sacchi, A. Nuzzo, U. Petrovič, and R. Bellazzi. Inferring gene regulatory networks by integrating static and dynamic data. *International Journal of Medical Informatics*, 76 :462–475, 2007.
- [102] P. Ferre and F. Fougelle. SREBP-1c transcription factor and lipid homeostasis : clinical perspective. *Horm Res*, 68(2) :72–82, 2007.
- [103] JP Flatt. Use and storage of carbohydrate and fat. *American Journal of Clinical Nutrition*, 61(4) :952–959, 1995.
- [104] M. Foretz, C. Guichard, P. Ferré, and F. Fougelle. Sterol regulatory element binding protein-1c is a major mediator of insulin action on the hepatic expression of glucokinase and lipogenesis-related genes, 1999.
- [105] F. Fougelle and P. Ferré. New perspectives in the regulation of hepatic glycolytic and lipogenic genes by insulin and glucose : a role for the transcription factor sterol regulatory element binding protein-1c. *BIOCHEMICAL JOURNAL-LONDON-*, 366(2) :377–391, 2002.
- [106] W.L. Franck, W.S. Chang, J. Qiu, M. Sugawara, M.J. Sadowsky, S.A. Smith, and G. Stacey. Whole-Genome Transcriptional Profiling of *Bradyrhizobium japonicum* during Chemoautotrophic Growth? *Journal of Bacteriology*, 190(20) :6697–6705, 2008.
- [107] J.C. Fruchart, P. Duriez, and B. Staels. Peroxisome proliferator-activated receptor- α activators regulate genes governing lipoprotein metabolism, vascular inflammation and atherosclerosis. *Current opinion in lipidology*, 10 :245–258, 1999.
- [108] G. Gandemer. Lipids in muscles and adipose tissues, changes during processing and sensory properties of meat products. *Meat Science*, 62(3) :309–321, 2002.
- [109] G. Gandemer, G. Pascal, and G. Durand. Lipogenic capacity and relative contribution of the different tissues and organs to lipid synthesis in male rat. *Reproduction Nutrition Développement*, 23(3) :575–586, 1983.

- [110] R. Garcia, C. Bermejo, C. Grau, R. Perez, J.M. Rodriguez-Pena, J. Francois, C. Nombela, and J. Arroyo. The global transcriptional response to transient cell wall damage in *Saccharomyces cerevisiae* and its regulation by the cell integrity signaling pathway. *Journal of Biological Chemistry*, 279(15) :15183, 2004.
- [111] P.J. Garlick, F. Margaret, V.R. Preedy, and C.N. and Metabolism Unit. The effect of insulin infusion and food intake on muscle protein synthesis in postabsorptive rats. *Biochem. J*, 210 :669–676, 1983.
- [112] JE Gerich, M. Lorenzi, DM Bier, E. Tsalikian, V. Schneider, JH Karam, and PH Forsham. Effects of physiologic levels of glucagon and growth hormone on human carbohydrate and lipid metabolism. Studies involving administration of exogenous hormone during suppression of endogenous hormone secretion with somatostatin. *Journal of Clinical Investigation*, 57(4) :875, 1976.
- [113] R. Giavazzi, JM Jessup, DE Campbell, SM Walker, and IJ Fidler. Experimental nude mouse model of human colorectal cancer liver metastases. *Journal of the National Cancer Institute*, 77(6) :1303–1308, 1986.
- [114] A.J. Gilde, K.A.J.M. van der Lee, P.H.M. Willemsen, G. Chinetti, F.R. van der Leij, G.J. van der Vusse, B. Staels, and M. van Bilsen. Peroxisome proliferator-activated receptor (PPAR){alpha} and PPAR {beta}/{delta}, but not PPAR {gamma}, modulate the expression of genes involved in cardiac lipid metabolism. *Circulation Research*, 92(5) :518, 2003.
- [115] JP Girard, J. Bout, and D. Salort. Lipides et qualités des tissus adipeux et musculaires de porc ; facteurs de variation. *Journées de la Recherche Porcine en France*, 20 :255–278, 1988.
- [116] RM Glickman and PHR Green. The intestine as a source of apolipoprotein A1. *Proceedings of the National Academy of Sciences*, 74(6) :2569–2573, 1977.
- [117] Pauline Gloaguen. Détermination de la meilleure technique d’acquisition d’un réseau de régulation à partir des connaissances avec pour application le métabolisme des lipides. Mémoire de M2, université de Rennes 1.
- [118] IJ Goldberg. Lipoprotein lipase and lipolysis : central roles in lipoprotein metabolism and atherogenesis. *The Journal of Lipid Research*, 37(4) :693–707, 1996.
- [119] F. Gondret, J. Sanjay, D. Marie, H. Patrick, V. Céline, H. Louis-Marie, and H. Jean-François. Unusual metabolic characteristics in skeletal muscles of transgenic rabbits for human lipoprotein lipase. *Lipids in Health and Disease*, 2004.
- [120] A.G. Gonzalez, A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya. GINsim : A software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, 84(2) :91–100, 2006.
- [121] Y. Gosmain, N. Dif, V. Berbe, E. Loizon, J. Rieusset, H. Vidal, and E. Lefai. Regulation of SREBP-1 expression and transcriptional action on HKII and FAS genes during fasting and refeeding in rat tissues. *The Journal of Lipid Research*, 46(4) :697–705, 2005.
- [122] Y. Gosmain, E. Lefai, S. Ryser, M. Roques, and H. Vidal. Sterol regulatory element-binding protein-1 mediates the effect of insulin on hexokinase II gene expression in human muscle cells. *Diabetes*, 53(2) :321, 2004.
- [123] J.L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical systems*, 17(4) :299–316, 2002.
- [124] H.D. Griffin, K. Guo, D. Windsor, and S.C. Butterwith. Adipose tissue lipogenesis and fat deposition in leaner broiler chickens. *Journal of Nutrition*, 122(2) :363, 1992.
- [125] B. Gr
"unenfelder and E.A. Winzeler. Treasures and traps in genome-wide data sets : case examples from yeast. *Nature Reviews Genetics*, 3(9) :653–661, 2002.
- [126] N. Guelzim, S. Bottani, P. Bourguine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nature genetics*, 31(1) :60–63, 2002.

- [127] H. Guillou, P. Martin, S. Jan, S. D'Andrea, A. Roulet, D. Catheline, V. Rioux, T. Pineau, and P. Legrand. Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor α -deficient mice. *Lipids*, 37(10) :981–989, 2002.
- [128] H Guillou, P Martin, S Jan, S D'Andrea, A Roulet, D Catheline, V Rioux, T Pineau, and P Legrand. Comparative effect of fenofibrate on hepatic desaturases in wild-type and peroxisome proliferator-activated receptor alpha-deficient mice. *Lipids*, 37(10) :981–9, 2002.
- [129] H. Guillou, PG Martin, and T. Pineau. Transcriptional regulation of hepatic fatty acid metabolism. *Subcell Biochem*, 49 :3–47, 2008.
- [130] R.M. Gutiérrez-Ríos, D.A. Rosenblueth, J.A. Loza, A.M. Huerta, J.D. Glasner, F.R. Blattner, and J. Collado-Vides. Regulatory network of *Escherichia coli* : consistency between literature knowledge and microarray profiles. *Genome research*, 13(11) :2435, 2003.
- [131] R. Haberman. *Mathematical models*. Prentice-Hall, 1977.
- [132] T. Hain, H. Hossain, S.S. Chatterjee, S. Machata, U. Volk, S. Wagner, B. Brors, S. Haas, C.T. Kuenne, A. Billion, et al. Temporal transcriptomic analysis of the *Listeria monocytogenes* EGD-e σ B regulon. *BMC microbiology*, 8(1) :20, 2008.
- [133] J.A. Hamilton, R.A. Johnson, B. Corkey, and F. Kamp. Fatty acid transport. *Journal of molecular neuroscience*, 16(2) :99–108, 2001.
- [134] V.C. Hannah, J. Ou, A. Luong, J.L. Goldstein, and M.S. Brown. Unsaturated fatty acids down-regulate SREBP isoforms 1a and 1c by two mechanisms in HEK-293 cells. *Journal of Biological Chemistry*, 276(6) :4365–4372, 2001.
- [135] R.W. Hanson and L. Reshef. Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. *Annual review of biochemistry*, 66(1) :581–611, 1997.
- [136] Y. Harano, A. Kashiwagi, H. Kojima, M. Suzuki, T. Hashimoto, and Y. Shigeta. Phosphorylation of carnitine palmitoyltransferase and activation by glucagon in isolated rat hepatocytes. *FEBS letters*, 188(2) :267, 1985.
- [137] T. Hashimoto, W.S. Cook, C. Qi, A.V. Yeldandi, J.K. Reddy, and M.S. Rao. Defect in peroxisome proliferator-activated receptor α -inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting. *Journal of Biological Chemistry*, 275(37) :28918–28928, 2000.
- [138] F.G. Hegardt. Transcriptional regulation of mitochondrial HMG-CoA synthase in the control of ketogenesis. *Biochimie*, 80(10) :803–806, 1998.
- [139] FG Hegardt. Mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase : a control enzyme in ketogenesis. *Biochemical Journal*, 338(Pt 3) :569, 1999.
- [140] B.D. Hegarty, A. Bobard, I. Hainault, P. Ferré, P. Bossard, and F. Foufelle. Distinct roles of insulin and liver X receptor in the induction and cleavage of sterol regulatory element-binding protein-1c. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3) :791, 2005.
- [141] JC Henquin. Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes*, 49(11) :1751–1760, 2000.
- [142] M.J. Herrgård, B.S. Lee, V. Portnoy, and B.Ø. Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome research*, 16(5) :627, 2006.
- [143] J.M. Hevel, L.C. Olson-Buelow, B. Ganesan, J.R. Stevens, J.P. Hardman, and A.E. Aust. Novel functional view of the crocidolite asbestos-treated A 549 human lung epithelial transcriptome reveals an intricate network of pathways with opposing functions. *BMC genomics*, 9(1) :376, 2008.
- [144] R.E. HODGES and WA Krehl. The role of carbohydrates in lipid metabolism. *American Journal of Clinical Nutrition*, 17(5) :334–346, 1965.
- [145] R. HOLLAND, L.A. WITTERS, and D.G. HARDIE. Glucagon inhibits fatty acid synthesis in isolated hepatocytes via phosphorylation of acetyl-CoA carboxylase by cyclic-AMP-dependent protein kinase. *European Journal of Biochemistry*, 140(2) :325–333, 1984.

- [146] L. Hooper, R.L. Thompson, R.A. Harrison, C.D. Summerbell, A.R. Ness, H.J. Moore, H.V. Worthington, P.N. Durrington, J.P.T. Higgins, N.E. Capps, et al. Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer : systematic review. *British Medical Journal*, 332(7544) :752–760, 2006.
- [147] JF Horowitz and S. Klein. Lipid metabolism during endurance exercise. *American Journal of Clinical Nutrition*, 72(2 ; SUPP/1) :558–563, 2000.
- [148] J.D. Horton, Y. Bashmakov, I. Shimomura, and H. Shimano. Regulation of sterol regulatory element binding proteins in livers of fasted and refed mice. *Proceedings of the National Academy of Sciences*, 95(11) :5987–5992, 1998.
- [149] J.D. Horton, Y. Bashmakov, I. Shimomura, and H. Shimano. Regulation of sterol regulatory element binding proteins in livers of fasted and refed mice, 1998.
- [150] J.D. Horton, J.L. Goldstein, and M.S. Brown. SREBPs : activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation*, 109(9) :1125, 2002.
- [151] M.H. Hsu,
"U. Savas, K.J. Griffin, and E.F. Johnson. Identification of Peroxisome Proliferator-responsive Human Genes by Elevated Expression of the Peroxisome Proliferator-activated Receptor alpha in HepG 2 Cells. *Journal of Biological Chemistry*, 276(30) :27950–27958, 2001.
- [152] T. Hu, P. Foxworthy, A. Siesky, J.V. Ficorilli, H. Gao, S. Li, M. Christe, T. Ryan, G. Cao, P. Eacho, et al. Hepatic peroxisomal fatty acid β -oxidation is regulated by liver X receptor α . *Endocrinology*, 146(12) :5380–5387, 2005.
- [153] L. Hue, C. Beauloye, A.S. Marsin, L. Bertrand, S. Horman, and M.H. Rider. Insulin and ischemia stimulate glycolysis by acting on the same targets through different and opposing signaling pathways. *Journal of molecular and cellular cardiology*, 34(9) :1091–1097, 2002.
- [154] F. Hynne, S. Danø, and PG Sørensen. Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophysical chemistry*, 94(1-2) :121–163, 2001.
- [155] T. Ide, H. Shimano, T. Yoshikawa, N. Yahagi, M. Amemiya-Kudo, T. Matsuzaka, M. Nakakuki, S. Yatoh, Y. Iizuka, S. Tomita, et al. Cross-talk between peroxisome proliferator-activated receptor (PPAR){alpha} and liver X receptor (LXR) in nutritional regulation of fatty acid metabolism. II. LXRs suppress lipid degradation gene promoters through inhibition of PPAR signaling. *Molecular Endocrinology*, 17(7) :1255, 2003.
- [156] K. Iizuka, R.K. Bruick, G. Liang, J.D. Horton, and K. Uyeda. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proceedings of the National Academy of Sciences*, 101(19) :7281, 2004.
- [157] S. Ikeda, H. Miyazaki, T. Nakatani, Y. Kai, Y. Kamei, S. Miura, N. Tsuboyama-Kasaoka, and O. Ezaki. Up-regulation of SREBP-1c and lipogenic genes in skeletal muscles after exercise training. *Biochemical and Biophysical Research Communications*, 296(2) :395–400, 2002.
- [158] S. Ishii, K. Iizuka, B.C. Miller, and K. Uyeda. Carbohydrate response element binding protein directly promotes lipogenic enzyme gene transcription. *Proceedings of the National Academy of Sciences*, 101(44) :15597–15602, 2004.
- [159] H.M. Ismail. The role of omega-3 fatty acids in cardiac protection : an overview. *Front Biosci*, 10 :1079–88, 2005.
- [160] PB Iynedjian, D. Jotterand, T. Noursipikel, M. Asfari, and PR Pilot. Transcriptional induction of glucokinase gene by insulin in cultured liver cells and its repression by the glucagon-cAMP system. *Journal of Biological Chemistry*, 264(36) :21824–21829, 1989.
- [161] LS Jefferson, JO Koehler, and HE Morgan. Effect of insulin on protein synthesis in skeletal muscle of an isolated perfused preparation of rat hemi-corpus. *Proceedings of the National Academy of Sciences*, 69(4) :816–820, 1972.

- [162] L.J. Jensen, J. Saric, P. Bork, et al. Literature mining for the biologist : from information retrieval to biological discovery. *Nature Reviews Genetics*, 7(2) :119–129, 2006.
- [163] M.D. Jensen, K. Ekberg, and B.R. Landau. Lipid metabolism during fasting. *American Journal of Physiology- Endocrinology And Metabolism*, 281(4) :789–793, 2001.
- [164] G. Jiang and B.B. Zhang. Glucagon and regulation of glucose metabolism, 2003.
- [165] Y.H. Jin, P.E. Dunlap, S.J. McBride, H. Al-Refai, P.R. Bushel, and J.H. Freedman. Global transcriptome and deletome profiles of yeast exposed to transition metals. *PLoS Genetics*, 4(4), 2008.
- [166] S. Jitrapakdee and JC Wallace. Structure, function and regulation of pyruvate carboxylase. *Biochemical Journal*, 340(Pt 1) :1, 1999.
- [167] S.B. Joseph, B.A. Laffitte, P.H. Patel, M.A. Watson, K.E. Matsukuma, R. Walczak, J.L. Collins, T.F. Osborne, and P. Tontonoz. Direct and indirect mechanisms for regulation of fatty acid synthase gene expression by liver X receptors. *Journal of Biological Chemistry*, 277(13) :11019–11025, 2002.
- [168] RL Jungas. Hormonal regulation of pyruvate dehydrogenase. *Metabolism : clinical and experimental*, 20(1) :43, 1971.
- [169] K., J. P. Plaimas, M. Mallm, F. Oswald, V. Svara, R. Sourjik, R. Eils, and König. Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Systems Biology*, 2 :67, 2008.
- [170] T. Kabashima, T. Kawaguchi, B.E. Wadzinski, and K. Uyeda. Xylulose 5-phosphate mediates glucose-induced lipogenesis by xylulose 5-phosphate-activated protein phosphatase in rat liver. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9) :5107, 2003.
- [171] B.B. Kahn and J.S. Flier. Obesity and insulin resistance, 2000.
- [172] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10) :770–780, 2008.
- [173] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein. Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences*, 100(25) :14796, 2003.
- [174] H. Keller, C. Dreyer, J. Medin, A. Mahfoudi, K. Ozato, and W. Wahli. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proceedings of the National Academy of Sciences*, 90(6) :2160–2164, 1993.
- [175] S. Kersten. Effects of fatty acids on gene expression : role of peroxisome proliferator-activated receptor α , liver X receptor α and sterol regulatory element-binding protein-1c. *Proceedings of the Nutrition Society*, 61(03) :371–374, 2008.
- [176] S. Kersten, B. Desvergne, and W. Wahli. Roles of PPARs in health and disease. *Nature*, 405 :421–424, 2000.
- [177] S. Kersten, J. Seydoux, J.M. Peters, F.J. Gonzalez, B. Desvergne, and W. Wahli. Peroxisome proliferator-activated receptor α mediates the adaptive response to fasting. *Journal of Clinical Investigation*, 103(11) :1489–1498, 1999.
- [178] H.J. Kim, M. Miyazaki, and J.M. Ntambi. Dietary cholesterol opposes PUFA-mediated repression of the stearoyl-CoA desaturase-1 gene by SREBP-1 independent mechanism. *The Journal of Lipid Research*, 43(10) :1750, 2002.
- [179] H.J. Kim, M. Takahashi, and O. Ezaki. Fish Oil Feeding Decreases Mature Sterol Regulatory Element-binding Protein 1 (SREBP-1) by Down-regulation of SREBP-1c mRNA in Mouse Liver A POSSIBLE MECHANISM FOR DOWN-REGULATION OF LIPOGENIC ENZYME mRNAs. *Journal of Biological Chemistry*, 274(36) :25892–25898, 1999.
- [180] JB Kim, P. Sarraf, M. Wright, KM Yao, E. Mueller, G. Solanes, BB Lowell, and BM Spiegelman. Nutritional and insulin regulation of fatty acid synthetase and leptin gene expression through ADD1/SREBP1. *Journal of Clinical Investigation*, 101(1) :1, 1998.

- [181] K.H. Kim. Regulation of mammalian acetyl-coenzyme A carboxylase. *Annual review of nutrition*, 17(1) :77–99, 1997.
- [182] SR Kimball, CV Jurasinski, JC Lawrence, and LS Jefferson. Insulin stimulates protein synthesis in skeletal muscle by enhancing the association of eIF-4E and eIF-4G. *American Journal of Physiology-Cell Physiology*, 272(2) :754–759, 1997.
- [183] M. Krallinger, A. Morgan, L. Smith, F. Leitner, L. Tanabe, J. Wilbur, L. Hirschman, and A. Valencia. Evaluation of text-mining systems for biology : overview of the Second BioCreative community challenge. *Genome biology*, 9(Suppl 2) :S1, 2008.
- [184] HA Krebs and R. Hems. Fatty acid metabolism in the perfused rat liver. *The Biochemical journal*, 119(3) :525, 1970.
- [185] D.L. Kroetz, P. Yook, P. Costet, P. Bianchi, and T. Pineau. Peroxisome proliferator-activated receptor α controls the hepatic CYP4A induction adaptive response to starvation and diabetes. *Journal of Biological Chemistry*, 273(47) :31581–31589, 1998.
- [186] Wolf-H. Kunau, V. Dommes, and H. Schultz. β -oxidation of fatty acids in mitochondria, peroxisomes, and bacteria : a century of continued progress. *Lipid Research*, 34 :267–342, 1995.
- [187] P.O. Kwiterovich Jr. The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides : a current review. *American Journal of Cardiology*, 86 :5–10, 2000.
- [188] L.C. Lai, M.T. Kissinger, P.V. Burke, and K.E. Kwast. Comparison of the transcriptomic" stress response" evoked by antimycin A and oxygen deprivation in *saccharomyces cerevisiae*. *BMC genomics*, 9(1) :627, 2008.
- [189] C. Lales1&2, N. Parisey, J.P. Mazat, and M. Beurton-Aimar1&2. Simulation of mitochondrial metabolism using multi-agents system. In *First International Workshop on Multi-Agent Systems for Medicine, Computational Biology, and Bioinformatics*, page 137, 2005.
- [190] D. Langin. Control of fatty acid and glycerol release in adipose tissue lipolysis. *Comptes rendus-Biologies*, 329(8) :598–607, 2006.
- [191] D. Langin, C. Holm, and M. Lafontan. Adipocyte hormone-sensitive lipase : a major regulator of lipid metabolism. *Proceedings of the Nutrition Society*, 55(1B) :93–109, 2007.
- [192] PB Lazarow. Rat liver peroxisomes catalyze the beta oxidation of fatty acids. *Journal of Biological Chemistry*, 253(5) :1522–1528, 1978.
- [193] G. Lazennec, L. Canaple, D. Saugy, and W. Wahli. Activation of peroxisome proliferator-activated receptors (PPARs) by their ligands and protein kinase A activators. *Molecular Endocrinology*, 14(12) :1962, 2000.
- [194] A. Le Bail, S.M. Dittami, P.O. de Franco, S. Rousvoal, M.J. Cock, T. Tonon, and B. Charrier. Normalisation genes for expression analyses in the brown alga model *Ectocarpus siliculosus*. *BMC Molecular Biology*, 9(1) :75, 2008.
- [195] SS Lee, T. Pineau, J. Drago, EJ Lee, JW Owens, DL Kroetz, PM Fernandez-Salguero, H. Westphal, and FJ Gonzalez. Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Molecular and Cellular Biology*, 15(6) :3012–3022, 1995.
- [196] S.S.T. Lee, W.Y. Chan, C.K.C. Lo, D.C.C. Wan, D.S.C. Tsang, and W.T. Cheung. Requirement of PPARalpha in maintaining phospholipid and triglyceride homeostasis during energy deprivation. *The Journal of Lipid Research*, page M400078, 2004.
- [197] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594) :799, 2002.
- [198] J.M. Lehmann, S.A. Kliewer, L.B. Moore, T.A. Smith-Oliver, B.B. Oliver, J.L. Su, S.S. Sundseth, D.A. Winegar, D.E. Blanchard, T.A. Spencer, et al. Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. *Journal of Biological Chemistry*, 272(6) :3137, 1997.

- [199] A.L. Lehninger, D.L. Nelson, and M.M. Cox. *Lehninger principles of biochemistry*. WH Freeman, 2004.
- [200] J.R. Leis and M.A. Kramer. The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Transactions on Mathematical Software (TOMS)*, 14(1) :45–60, 1988.
- [201] T.C. Leone, C.J. Weinheimer, and D.P. Kelly. A critical role for the peroxisome proliferator-activated receptor α (PPAR α) in the cellular fasting response : The PPAR α -null mouse as a model of fatty acid oxidation disorders. *Proceedings of the National Academy of Sciences*, 96(13) :7473–7478, 1999.
- [202] A. Lesne and A. Benecke. Feature context-dependency and complexity-reduction in probability landscapes for integrative genomics. *Theoretical Biology and Medical Modelling*, 5(1) :21, 2008.
- [203] A. Leturque, E. Brot-Laroche, and M. Le Gall. GLUT2 mutations, translocation, and receptor function in diet sugar managing. *American Journal of Physiology- Endocrinology And Metabolism*, 296(5) :E985, 2009.
- [204] GA Leveille, DR Romsos, Y. Yeh, and EK O’Hea. Lipid biosynthesis in the chick. A consideration of site of synthesis, influence of diet and possible regulatory mechanisms. *Poultry science*, 54(4) :1075, 1975.
- [205] G. Liang, J. Yang, J.D. Horton, R.E. Hammer, J.L. Goldstein, and M.S. Brown. Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *Journal of Biological Chemistry*, 277(11) :9520–9528, 2002.
- [206] L. Liang-Chuan, K. Matthew, B. Patricia, and K. Kurt. Comparison of the transcriptomic" stress response" evoked by antimycin A and oxygen deprivation in saccharomyces cerevisiae. *BMC Genomics*, 9, 2008.
- [207] Q. Lin, S.E. Ruuska, N.S. Shaw, D. Dong, and N. Noy. Ligand Selectivity of the Peroxisome Proliferator-Activated Receptor [alpha]. *Biochemistry*, 38(1) :185–190, 1999.
- [208] C. Longuet, E.M. Sinclair, A. Maida, L.L. Baggio, M. Maziarz, M.J. Charron, and D.J. Drucker. The Glucagon receptor is required for the adaptive metabolic response to fasting. *Cell Metabolism*, 8(5) :359–371, 2008.
- [209] J.M. Lopez, M.K. Bennett, H.B. Sanchez, J.M. Rosenfeld, and T.F. Osborne. Sterol regulation of acetyl coenzyme A carboxylase : a mechanism for coordinate control of cellular lipid, 1996.
- [210] S. Luci, B. Giemsa, H. Kluge, and K. Eder. Clofibrate causes an upregulation of PPAR- α target genes but does not alter expression of SREBP target genes in liver and adipose tissue of pigs. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 293(1) :R70, 2007.
- [211] J. Maertens, B.M.R. Donckels, G. Lequeux, and P.A. Vanrolleghem. Metabolic model reduction by metabolite pooling on the basis of dynamic phase planes and metabolite correlation analysis. *Proceedings of the Conference on Modeling and Simulation in Biology, Medicine and Biomedical Engineering. Linkping, Sweden,*, 2005.
- [212] LJ Mandarino, A. Consoli, A. Jain, and DE Kelley. Differential regulation of intracellular glucose metabolism by glucose and insulin in human muscle. *American Journal of Physiology- Endocrinology And Metabolism*, 265(6) :898–905, 1993.
- [213] GP Mannaerts, LJ Debeer, J. Thomas, and PJ De Schepper. Mitochondrial and peroxisomal fatty acid oxidation in liver homogenates and isolated hepatocytes from control and clofibrate-treated rats. *Journal of Biological Chemistry*, 254(11) :4585–4595, 1979.
- [214] J. Mao, T. Habib, M. Shenwu, B. Kang, W. Allen, L.S. Robertson, J. Yang, and Y. Deng. Transcriptome profiling of Saccharomyces cerevisiae mutants lacking C2H2 zinc finger proteins. *BMC genomics*, 9(Suppl 1) :S14, 2008.
- [215] P.G.P. Martin. Etude du potentiel de régulation génique exercé par l’isoforme alpha du récepteur activé par les proliférateurs de peroxysomes (PPARalpha), 2007.

- [216] S. Mathivanan, B. Periaswamy, TKB Gandhi, K. Kandasamy, S. Suresh, R. Mohmood, YL Ramachandra, and A. Pandey. An evaluation of human protein-protein interaction data in the public domain. *BMC bioinformatics*, 7(Suppl 5) :S19, 2006.
- [217] T. Matsuzaka, H. Shimano, N. Yahagi, M. Amemiya-Kudo, T. Yoshikawa, A.H. Hasty, Y. Tamura, J. Osuga, H. Okazaki, Y. Iizuka, et al. Dual regulation of mouse $\Delta 5$ - and $\Delta 6$ -desaturase gene expression by SREBP-1 and PPAR α . *The Journal of Lipid Research*, 43(1) :107–114, 2002.
- [218] JD McGarry, SE Mills, CS Long, and DW Foster. Observations on the affinity for carnitine, and malonyl-CoA sensitivity, of carnitine palmitoyltransferase I in animal and human tissues. Demonstration of the presence of malonyl-CoA in non-hepatic tissues of the rat. *Biochemical Journal*, 214(1) :21, 1983.
- [219] JD McGivan and M. Pastor-Anglada. Regulatory and molecular aspects of mammalian amino acid transport. *Biochemical Journal*, 299(Pt 2) :321, 1994.
- [220] J.P. McNamara. Lipid metabolism in adipose tissue during lactation : a model of a metabolic control system. *Journal of Nutrition*, 124(8 Suppl) :1383S, 1994.
- [221] T.P. Michael, T.C. Mockler, G. Breton, C. McEntee, A. Byer, J.D. Trout, S.P. Hazen, R. Shen, H.D. Priest, C.M. Sullivan, et al. Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet*, 4(2) :e14, 2008.
- [222] V. Mihajlovic and dr. M. Petkovic. Dynamic bayesian networks : A state of the art. *CTIT technical reports series*, TR-CTIT-34, 2001. DMW-project.
- [223] C.W. Miller and J.M. Ntambi. Peroxisome proliferators induce mouse liver stearyl-CoA desaturase 1 gene expression, 1996.
- [224] N. Mitro, P.A. Mak, L. Vargas, C. Godio, E. Hampton, V. Molteni, A. Kreuzsch, and E. Saez. The nuclear receptor LXR is a glucose sensor. *Nature*, 445(7124) :219–223, 2006.
- [225] Y. Miyao, K. Sagae, R. Saetre, T. Matsuzaki, and J. Tsujii. Evaluating contributions of natural language parsers to protein-protein interaction extraction. *Bioinformatics*, 25(3) :394, 2009.
- [226] P.T. Monteiro, N.D. Mendes, M.C. Teixeira, S. d’Orey, S. Tenreiro, N.P. Mira, H. Pais, A.P. Francisco, A.M. Carvalho, A.B. Lourenco, et al. YEASTRACT-DISCOVERER : new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucleic acids research*, 2007.
- [227] Y.A. Moon, N.A. Shah, S. Mohapatra, J.A. Warrington, and J.D. Horton. Identification of a mammalian long chain fatty acyl elongase regulated by sterol regulatory element-binding proteins. *Journal of Biological Chemistry*, 276(48) :45358, 2001.
- [228] T.A. Mori and L.J. Beilin. Omega-3 fatty acids and inflammation. *Current Atherosclerosis Reports*, 6(6) :461–467, 2004.
- [229] C. Morimoto, T. Tsujita, and H. Okuda. Antilipolytic actions of insulin on basal and hormone-induced lipolysis in rat adipocytes. *The Journal of Lipid Research*, 39(5) :957–962, 1998.
- [230] J. Mourot and D. Hermier. Lipids in monogastric animal meat. *Reproduction Nutrition Development*, 41(2) :109–118, 2001.
- [231] Robert K. Murray, Daryl K. Granner, and Victor W. Rodwell. *Harper’s Illustrated Biochemistry, 27th Edition*. The McGraw-Hill Companies, 2006.
- [232] Y. Nakai, H. Hashida, K. Kadota, M. Minami, K. Shimizu, I. Matsumoto, H. Kato, and K. Abe. Up-regulation of genes related to the ubiquitin-proteasome system in the brown adipose tissue of 24-h-fasted rats. *Bioscience, Biotechnology, and Biochemistry*, page 711300675, 2008.
- [233] MT Nakamura and TY Nara. Essential fatty acid synthesis and its regulation in mammals. *Prostaglandins, Leukotrienes & Essential Fatty Acids*, 68(2) :145–150, 2003.
- [234] MT Nakamura and TY Nara. Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annual review of nutrition*, 24 :345, 2004.

- [235] T.Y. Nara, W.S. He, C. Tang, S.D. Clarke, and M.T. Nakamura. The E-box like sterol regulatory element mediates the suppression of human Δ -6 desaturase gene by highly unsaturated fatty acids. *Biochemical and Biophysical Research Communications*, 296(1) :111–117, 2002.
- [236] C. Neal, Z. Magdalena, F.M. Ali, P. Diane, M.M. Anne-Marie, and T. Patricia. Influence of monolayer, spheroid, and tumor growth conditions on chromosome 3 gene expression in tumorigenic epithelial ovarian cancer cell lines. *BMC Medical Genomics*, 1, 2008.
- [237] J.M. Ntambi. Regulation of stearoyl-CoA desaturase by polyunsaturated fatty acids and cholesterol. *The Journal of Lipid Research*, 40(9) :1549, 1999.
- [238] M.J. O’Connell and J.O. McInerney. Adaptive evolution of the human fatty acid synthase gene : Support for the cancer selection and fat utilization hypotheses? *Gene*, 360(2) :151–159, 2005.
- [239] EK O’Hea and GA Leveille. Lipogenesis in isolated adipose tissue of the domestic chick (*Gallus domesticus*). *Comparative biochemistry and physiology*, 26(1) :111, 1968.
- [240] E.K. O’Hea and G.A. Leveille. Significance of adipose tissue and liver as sites of fatty acid synthesis in the pig and the efficiency of utilization of various substrates for lipogenesis. *Journal of Nutrition*, 99(3) :338, 1969.
- [241] JW Oltjen, AC Bywater, RL Baldwin, and WN Garrett. Development of a dynamic model of beef cattle growth and composition. *Journal of Animal Science*, 62(1) :86, 1986.
- [242] JM Oppert. Adaptation to overfeeding in humans : metabolic aspects ; Adaptation à la suralimentation chez l’homme : aspects métaboliques. *Cahiers de Nutrition et de Diététique (France)*, 2001.
- [243] R.J. Orton, O.E. Sturm, V. Vyshemirsky, M. Calder, D.R. Gilbert, and W. Kolch. Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. *Biochemical Journal*, 392(Pt 2) :249, 2005.
- [244] J. Ou, H. Tu, B. Shan, A. Luk, R.A. DeBose-Boyd, Y. Bashmakov, J.L. Goldstein, and M.S. Brown. Unsaturated fatty acids inhibit transcription of the sterol regulatory element-binding protein-1c (SREBP-1c) gene by antagonizing ligand-dependent activation of the LXR. *Proceedings of the National Academy of Sciences*, 98(11) :6027, 2001.
- [245] L. Panait and S. Luke. Cooperative multi-agent learning : The state of the art. *Autonomous Agents and Multi-Agent Systems*, 11(3) :387–434, 2005.
- [246] J. Pandhal, P.C. Wright, and C.A. Biggs. Proteomics with a pinch of salt : A cyanobacterial perspective. *Saline Systems*, 4(1) :1, 2008.
- [247] EA Park, RL Mynatt, GA Cook, and K. Kashfi. Insulin regulates enzyme activity, malonyl-CoA sensitivity and mRNA abundance of hepatic carnitine palmitoyltransferase-I. *Biochemical Journal*, 310(Pt 3) :853, 1995.
- [248] A. Pause, G.J. Belsham, A.C. Gingras, O. Donzé, T.A. Lin, J.C. Lawrence, and N. Sonenberg. Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5’-cap function. *Nature*, 1994.
- [249] JP Pegorier, MV Garcia-Garcia, C. Prip-Buus, PH Duee, C. Kohl, and J. Girard. Induction of ketogenesis and fatty acid oxidation by glucagon and cyclic AMP in cultured hepatocytes from rabbit fetuses. Evidence for a decreased sensitivity of carnitine palmitoyltransferase I to malonyl-CoA inhibition after glucagon or cyclic AMP treatment. *The Biochemical journal*, 264(1) :93, 1989.
- [250] K.F. Petersen, D. Laurent, D.L. Rothman, G.W. Cline, and G.I. Shulman. Mechanism by which glucose and insulin inhibit net hepatic glycogenolysis in humans, 1998.
- [251] V. Petit, I. Niot, H. Poirier, and P. Besnard. Absorption intestinale des acides gras : faits et incertitudes Fatty acids intestinal absorption : facts and uncertainties. *Nutrition Clinique et Métabolisme*, 21(1) :38–45, 2007.
- [252] M.W. Pfaffl. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, 29(9) :e45, 2001.

- [253] SJ Pilkis, MR El-Maghrabi, J. Pilkis, TH Claus, and DA Cumming. Fructose 2, 6-bisphosphate. A new activator of phosphofructokinase. *Journal of Biological Chemistry*, 256(7) :3171, 1981.
- [254] C. Pomar, DL Harris, and F. Minvielle. Computer simulation model of swine production systems : I. Modeling the growth of young pigs, 1991.
- [255] C. Postic, R. Burcelin, F. Rencurel, JP Pegorier, M. Loizeau, J. Girard, and A. Leturque. Evidence for a transient inhibitory effect of insulin on GLUT2 expression in the liver : studies in vivo and in vitro. *Biochemical Journal*, 293(Pt 1) :119, 1993.
- [256] C. Postic and J. Girard. Contribution of de novo fatty acid synthesis to hepatic steatosis and insulin resistance : lessons from genetically engineered mice. *The Journal of clinical investigation*, 118(3) :829, 2008.
- [257] EI Prokudina, RY Valeev, and RN Tchuraev. A new method for the analysis of the dynamics of the molecular genetic control systems. II : application of the method of generalized threshold models in the investigation of concrete genetic systems. *Journal of theoretical biology*, 151(1) :89–110, 1991.
- [258] A. Quigley and P. Eades. FADE : Graph drawing, clustering, and visual abstraction. *Lecture notes in computer science*, pages 197–210, 2001.
- [259] M.S. Rajeevan, S.D. Vernon, N. Taysavang, and E.R. Unger. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR, 2001.
- [260] M. Rakhshandehroo, L.M. Sanderson, M. Matilainen, R. Stienstra, C. Carlberg, P.J. de Groot, M. Muller, and S. Kersten. Comprehensive analysis of PPARalpha-dependent regulation of hepatic lipid metabolism by expression profiling. *PPAR Res*, 26839, 2007.
- [261] R.F. Ranallo and E.C. Rhodes. Lipid metabolism during exercise. *Sports Medicine*, 26(1) :29, 1998.
- [262] PJ Randle, PB Garland, CN Hales, and EA Newsholme. The glucose fatty-acid cycle. Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet*, 1(7285) :785, 1963.
- [263] I. Rayment, HM Holden, M. Whittaker, CB Yohn, M. Lorenz, KC Holmes, and RA Milligan. Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, 261(5117) :58, 1993.
- [264] J.K. Reddy, S. Rao, et al. Lipid metabolism and liver inflammation. II. Fatty liver disease and fatty acid oxidation. *American Journal of Physiology- Gastrointestinal and Liver Physiology*, 290(5) :G852, 2006.
- [265] J.S. RICHARDS and A.R. MIDGLEY. Protein hormone action : a key to understanding ovarian follicular and luteal cell development. *Biology of Reproduction*, 14(1) :82–94, 1976.
- [266] M. Roden, G. Perseghin, K.F. Petersen, J.H. Hwang, G.W. Cline, K. Gerow, D.L. Rothman, and G.I. Shulman. The roles of insulin and glucagon in the regulation of hepatic glycogen synthesis and turnover in humans. *Journal of Clinical Investigation*, 97(3) :642–648, 1996.
- [267] RW Rosebrough, JP McMurtry, and NC Steele. Protein and energy relations in the broiler chicken. *British Journal of Nutrition*, 61(02) :223–233, 2007.
- [268] M. Salem, J. Silverstein, C.E. Rexroad III, and J. Yao. Effect of starvation on global gene expression and proteolysis in rainbow trout (*Oncorhynchus mykiss*). *BMC genomics*, 8(328) :1471–2164, 2007.
- [269] M. Salem, J. Silverstein, C.E. Rexroad III, and J. Yao. Effect of starvation on global gene expression and proteolysis in rainbow trout (*Oncorhynchus mykiss*). *BMC genomics*, 8(328) :1471–2164, 2007.
- [270] A.R. Saltiel and C.R. Kahn. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414(6865) :799–806, 2001.
- [271] JG Salway and DK Granner. *Metabolism at a Glance*. Blackwell Pub, 2004.
- [272] H. Sampath and J.M. Ntambi. Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annual Review of Nutrition*, 2005.
- [273] L. Sanchez and D. Thieffry. A logical analysis of the Drosophila gap-gene system. *Journal of theoretical Biology*, 211(2) :115–141, 2001.

- [274] M. Santillán. Bistable Behavior in a Model of the lac Operon in Escherichia coli with Variable Growth Rate. *Biophysical Journal*, 94(6) :2065–2081, 2008.
- [275] DS Schade, W. Woodside, and RP Eaton. The role of glucagon in the regulation of plasma lipids. *Metabolism : clinical and experimental*, 28(8) :874, 1979.
- [276] D. Schmoll, K.S. Walker, D.R. Alessi, R. Grempler, A. Burchell, S. Guo, R. Walther, and T.G. Unterman. Regulation of glucose-6-phosphatase gene expression by protein kinase Balpha and the forkhead transcription factor FKHR. Evidence for insulin response unit-dependent and-independent effects of insulin on promoter activity. *Journal of Biological Chemistry*, 275(46) :36324–36333, 2000.
- [277] K. Schoonjans, B. Staels, and J. Auwerx. Role of the peroxisome proliferator-activated receptor (PPAR) in mediating the effects of fibrates and fatty acids on gene expression. *The Journal of Lipid Research*, 37(5) :907–925, 1996.
- [278] F. Schreiber, T. Dwyer, K. Marriott, and M. Wybrow. A generic algorithm for layout of biological networks. *BMC bioinformatics*, 10(1) :375, 2009.
- [279] J.R. Schultz, H. Tu, A. Luk, J.J. Repa, J.C. Medina, L. Li, S. Schwendner, S. Wang, M. Thoolen, D.J. Mangelsdorf, et al. Role of LXRs in control of lipogenesis. *Genes & Development*, 14(22) :2831–2838, 2000.
- [280] C.M. Schworer and G.E. Mortimore. Glucagon-induced autophagy and proteolysis in rat liver : mediation by selective deprivation of intracellular amino acids. *Proceedings of the National Academy of Sciences*, 76(7) :3169–3173, 1979.
- [281] R. Seger and EG Krebs. The MAPK signaling cascade. *The FASEB Journal*, 9(9) :726, 1995.
- [282] K.N. Seidelin. Fatty acid composition of adipose tissue in humans. Implications for the dietary fat-serum cholesterol-CHD issue. *Progress in Lipid Research*, 34(3) :199–217, 1995.
- [283] C.F. Semenkovich. Regulation of fatty acid synthase (FAS). *Progress in lipid research*, 36(1) :43–53, 1997.
- [284] A.M. Sessler and J.M. Ntambi. Polyunsaturated Fatty Acid Regulation of Gene Expression. *Journal of Nutrition*, 128(6) :923–926, 1998.
- [285] A.M. Sessler and J.M. Ntambi. Polyunsaturated fatty acid regulation of gene expression. *Journal of Nutrition*, 128(6) :923–926, 1998.
- [286] JP Shaffer. Multiple hypothesis testing. *Annual Review of Psychology*, 46(1) :561–584, 1995.
- [287] S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1) :64–68, 2002.
- [288] H. Shimano, N. Yahagi, M. Amemiya-Kudo, A.H. Hasty, J. Osuga, Y. Tamura, F. Shionoiri, Y. Iizuka, K. Ohashi, K. Harada, et al. Sterol regulatory element-binding protein-1 as a key transcription factor for nutritional induction of lipogenic enzyme genes. *Journal of Biological Chemistry*, 274(50) :35832–35839, 1999.
- [289] K. Shirai. Obesity as the core of the metabolic syndrome and the management of coronary heart disease. *Current Medical Research and Opinion*, 20(3) :295–304, 2004.
- [290] I. SHMULEVICH, E.R. DOUGHERTY, and WEI ZHANG. From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks. *PROCEEDINGS OF THE IEEE*, 90(11), 2002.
- [291] LS Sidossis, CA Stuart, GI Shulman, GD Lopaschuk, and RR Wolfe. Glucose plus insulin regulate fat oxidation by controlling the rate of fatty acid entry into the mitochondria. *Journal of Clinical Investigation*, 98(10) :2244, 1996.
- [292] A. Siegel, O. Radulescu, M. Le Borgne, P. Veber, J. Ouy, and S. Lagarrigue. Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. *Biosystems*, 84(2) :153–174, 2006.
- [293] A. Silvescu and V. Honavar. Temporal boolean network models of genetic networks and their inference from gene expression time series. *Complex Systems*, 13(1) :61–78, 2001.

- [294] AP Simopoulos. Omega-3 fatty acids in health and disease and in growth and development. *American Journal of Clinical Nutrition*, 54(3) :438–463, 1991.
- [295] AP Simopoulos. Omega-3 fatty acids in health and disease and in growth and development (see comments). *Am. J. Clin. Nutr.*, 54 :438, 1991.
- [296] AP Simopoulos. The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomedicine & Pharmacotherapy*, 56(8) :365–379, 2002.
- [297] S. Skiba-Cassy, A. Collin, P. Chartrin, F. Médale, J. Simon, M.J. Duclos, and S. Tesseraud. Chicken liver and muscle carnitine palmitoyltransferase 1 : Nutritional regulation of messengers. *Comparative Biochemistry and Physiology, Part B*, 147(2) :278–287, 2007.
- [298] BG Slavin, JM Ong, and PA Kern. Hormonal regulation of hormone-sensitive lipase activity and mRNA levels in isolated rat adipocytes. *The Journal of Lipid Research*, 35(9) :1535–1541, 1994.
- [299] C.A. Small, A.J. Garton, and S.J. Yeaman. The presence and role of hormone-sensitive lipase in heart muscle. *Biochem J*, 258(1) :67–72, 1989.
- [300] P. Smolen, D.A. Baxter, and J.H. Byrne. Modeling transcriptional control in gene networks-methods, recent results, and future directions. *Bulletin of Mathematical Biology*, 62(2) :247–292, 2000.
- [301] J. Soini, C. Falschlehner, C. Liedert, J. Bernhardt, J. Vuoristo, and P. Neubauer. Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W 3110. *Microbial Cell Factories*, 7(1) :30, 2008.
- [302] A.J. Soitamo, M. Piippo, Y. Allahverdiyeva, N. Battchikova, and E.M. Aro. Light has a specific role in modulating Arabidopsis gene expression at low temperature. *BMC Plant Biology*, 8(1) :13, 2008.
- [303] M. Sokolović, A. Sokolović, D. Wehkamp, E.V.L. van Themaat, D.R. de Waart, L.A. Gilhuijs-Pederson, Y. Nikolsky, A.H.C. van Kampen, T. Hakvoort, and W.H. Lamers. The transcriptomic signature of fasting murine liver. *BMC genomics*, 9(1) :528, 2008.
- [304] M. Sokolović, A. Sokolović, D. Wehkamp, E.V.L. van Themaat, D.R. de Waart, L.A. Gilhuijs-Pederson, Y. Nikolsky, A.H.C. van Kampen, T.B.M. Hakvoort, and W.H. Lamers. The transcriptomic signature of fasting murine liver. *BMC Genomics*, 9 :528, 2008.
- [305] M.J. Song, C.K. Lewis, E.R. Lance, E.J. Chesler, R.K. Yordanova, M.A. Langston, K.H. Lodowski, and S.E. Bergeson. Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. *EURASIP Journal on Bioinformatics and Systems Biology*, 2009, 2009.
- [306] C. Soulé. Mathematical approaches to differentiation and gene regulation. *Comptes rendus-Biologies*, 329(1) :13–20, 2006.
- [307] AA Spector. Fatty acid binding to plasma albumin. *The Journal of Lipid Research*, 16(3) :165–179, 1975.
- [308] P. Sprumont and C.M. Thiebault. *Le sport après 50 ans*. De Boeck, 2005.
- [309] A.K. Srivastava and S.K. Pandey. Potential mechanism (s) involved in the regulation of glycogen synthesis by insulin. *Molecular and cellular biochemistry*, 182(1) :135–141, 1998.
- [310] A. Stahl. A current review of fatty acid transport proteins (SLC27). *Pflügers Archiv European Journal of Physiology*, 447(5) :722–727, 2004.
- [311] J. Stöckel, E.A. Welsh, M. Liberton, R. Kunnvakkam, R. Aurora, and H.B. Pakrasi. Global transcriptomic analysis of Cyanospora 51142 reveals robust diurnal oscillation of central metabolic processes. *Proceedings of the National Academy of Sciences*, 105(16) :6156, 2008.
- [312] A.K. Stoekman and H.C. Towle. The Role of SREBP-1 c in Nutritional Regulation of Lipogenic Enzyme Gene Expression. *Journal of Biological Chemistry*, 277(30) :27029–27035, 2002.
- [313] F. Sultan, D. Lagrange, and S. Griglio. Métabolisme intravasculaire des chylomicrons et des remnants de chylomicrons. *Nutrition clinique et métabolisme(Paris)*, 9(1) :43–51, 1995.

- [314] D.E. Tabor, J.B. Kim, B.M. Spiegelman, and P.A. Edwards. Identification of conserved cis-elements and transcription factors required for sterol-regulated transcription of stearoyl-CoA desaturase 1 and 2. *Journal of Biological Chemistry*, 274(29) :20603–20610, 1999.
- [315] C. Tang, H.P. Cho, M.T. Nakamura, and S.D. Clarke. Regulation of human Δ^6 desaturase gene transcription. *Journal of lipid research*, 44 :686–695, 2003.
- [316] M. Taniguchi, K. Miura, H. Iwao, and S. Yamanaka. Quantitative assessment of DNA microarrays comparison with Northern blot analyses. *Genomics*, 71(1) :34–39, 2001.
- [317] M.C. Teixeira, P. Monteiro, P. Jain, S. Tenreiro, A.R. Fernandes, N.P. Mira, M. Alenquer, A.T. Freitas, A.L. Oliveira, and I. Sá-Correia. The YEASTRACT database : a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic acids research*, 34(Database Issue) :D446, 2006.
- [318] J.M. Temkin and M.R. Gilder. Extraction of protein interaction information from unstructured text using a context-free grammar. *Bioinformatics*, 19(16) :2046, 2003.
- [319] B. Teusink, J. Passarge, C.A. Reijenga, E. Esgalhado, C.C. van der Weijden, M. Schepper, M.C. Walsh, B.M. Bakker, K. van Dam, H.V. Westerhoff, et al. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry, 2000.
- [320] D. Thieffry and R. Thomas. Dynamical behaviour of biological regulatory networks-II. Immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57(2) :277–297, 1995.
- [321] R. Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In *Numerical methods in the study of critical phenomena : proceedings of a colloquium, Carry-le-Rouet, France, June 2-4, 1980*, page 180. Springer, 1981.
- [322] K.A.R. Tobin, S.M. Ulven, G.U. Schuster, H.H. Steineger, S.M. Andresen, J.A. Gustafsson, and H.I. Nebb. Liver X receptors as insulin-mediating factors in fatty acid and cholesterol biosynthesis. *Journal of Biological Chemistry*, 277(12) :10691–10697, 2002.
- [323] R. Treisman. Regulation of transcription by MAP kinase cascades. *Current opinion in cell biology*, 8(2) :205–215, 1996.
- [324] R. Uauy, P. Mena, and C. Rojas. Essential fatty acids in early life : structural and functional role. *Proceedings of the Nutrition Society*, 59(01) :3–15, 2007.
- [325] K. Uyeda and J.J. Repa. Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metabolism*, 4(2) :107–110, 2006.
- [326] K. Uyeda, H. Yamashita, and T. Kawaguchi. Carbohydrate responsive element-binding protein (ChREBP) : a key regulator of glucose metabolism and fat storage. *Biochemical pharmacology*, 63(12) :2075–2080, 2002.
- [327] J. van Milgen. Modeling Biochemical Aspects of Energy Metabolism in Mammals 1. *Journal of Nutrition*, 132(10) :3195–3202, 2002.
- [328] E. Van Schaftingen, MF Jett, L. Hue, and HG Hers. Control of liver 6-phosphofructokinase by fructose 2, 6-bisphosphate and other effectors. *Proceedings of the National Academy of Sciences*, 78(6) :3483, 1981.
- [329] J.H. Veerkamp. Fatty acid transport and fatty acid-binding proteins. *Proceedings of the Nutrition Society*, 54(01) :23–37, 2007.
- [330] L. Von Bertalanffy. General system theory-a critical review. *Educational Administration and the behavioral Sciences : A Systems Perspective*, page 5, 1973.
- [331] T.T. Vu and J. Vohradsky. Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*. *Nucleic acids research*, 35(1) :279, 2007.
- [332] R.J.A. Wanders, G.A. Jansen, and M.D. Lloyd. Phytanic acid alpha-oxidation, new insights into an old problem : a review. *BBA-Molecular and Cell Biology of Lipids*, 1631(2) :119–135, 2003.

- [333] L. Wang, M. Montano, M. Rarick, and P. Sebastiani. Conditional clustering of temporal expression profiles. *BMC bioinformatics*, 9(1) :147, 2008.
- [334] Y. Wang, D. Botolin, J. Xu, B. Christian, E. Mitchell, B. Jayaprakasam, M. Nair, J.M. Peters, J. Busik, L.K. Olson, et al. Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity. *The Journal of Lipid Research*, 47(9) :2028, 2006.
- [335] Y. Wang, D. Botolin, J. Xu, B. Christian, E. Mitchell, B. Jayaprakasam, M. Nair, J.M. Peters, J. Busik, L.K. Olson, et al. Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity. *The Journal of Lipid Research*, 47(9) :2028, 2006.
- [336] K. Watanabe, H. Fujii, T. Takahashi, M. Kodama, Y. Aizawa, Y. Ohta, T. Ono, G. Hasegawa, M. Naito, T. Nakajima, et al. Constitutive Regulation of Cardiac Fatty Acid Metabolism through Peroxisome Proliferator-activated Receptor α Associated with Age-dependent Cardiac Toxicity. *Journal of Biological Chemistry*, 275(29) :22293–22299, 2000.
- [337] G. Weber and A. Cantero. Glucose-6-phosphatase activity in normal, precancerous, and neoplastic tissues. *Cancer research*, 15(2) :105, 1955.
- [338] R.L. Welchman, C. Gordon, and R.J. Mayer. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nature Reviews Molecular Cell Biology*, 6(8) :599–609, 2005.
- [339] LA Witters and BE Kemp. Insulin activation of acetyl-CoA carboxylase accompanied by inhibition of the 5'-AMP-activated protein kinase. *Journal of Biological Chemistry*, 267(5) :2864–2867, 1992.
- [340] JD Wood, RI Richardson, GR Nute, AV Fisher, MM Campo, E. Kasapidou, PR Sheard, and M. Enser. Effects of fatty acids on meat quality : a review. *Meat Science*, 66(1) :21–32, 2004.
- [341] K.H. Woodside, W.F. Ward, and G.E. Mortimore. Effects of glucagon on general protein degradation and synthesis in perfused rat liver. *Journal of Biological Chemistry*, 249(17) :5458–5463, 1974.
- [342] J. Xu, M. Teran-Garcia, J.H.Y. Park, M.T. Nakamura, and S.D. Clarke. Polyunsaturated fatty acids suppress hepatic sterol regulatory element-binding protein-1 expression by accelerating transcript decay. *Journal of Biological Chemistry*, 276(13) :9800–9807, 2001.
- [343] S. Yakar, J.L. Liu, A.M. Fernandez, Y. Wu, A.V. Schally, J. Frystyk, S.D. Chernausk, W. Mejia, and D. Le Roith. Liver-specific igf-1 gene deletion leads to muscle insulin insensitivity. *Diabetes*, 50(5) :1110, 2001.
- [344] H. Yamashita, M. Takenoshita, M. Sakurai, R.K. Bruick, W.J. Henzel, W. Shillinglaw, D. Arnot, and K. Uyeda. A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proceedings of the National Academy of Sciences of the United States of America*, 98(16) :9116, 2001.
- [345] P. Yeagle. *The membranes of cells*. Academic Press, 1993.
- [346] T. Yoshikawa, H. Shimano, N. Yahagi, T. Ide, M. Amemiya-Kudo, T. Matsuzaka, M. Nakakuki, S. Tomita, H. Okazaki, Y. Tamura, et al. Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements. *Journal of Biological Chemistry*, 277(3) :1705, 2002.
- [347] A. Young. Inhibition of glucagon secretion. *Advances in pharmacology (San Diego, Calif.)*, 52 :151, 2005.
- [348] J. Yu, V.A. Smith, P.P. Wang, A.J. Hartemink, and E.D. Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data, 2004.
- [349] J. Yun, B. Jeon, Y.W. Barton, P. Plummer, Q. Zhang, and S. Ryu. Role of the DksA-Like Protein in the Pathogenesis and Diverse Metabolic Activity of *Campylobacter jejuni*. *Journal of bacteriology*, 190(13) :4512–4520, 2008.
- [350] R. Zechner. The tissue-specific expression of lipoprotein lipase : implications for energy and lipoprotein metabolism. *Current opinion in lipidology*, 8(2) :77, 1997.

- [351] NL Zhang and D. Poole. Exploiting causal independence in Bayesian network inference. *Arxiv preprint cs/9612101*, 1996.
- [352] Claude Zinsou. Cours de Biochimie Métabolique. [http ://pagesperso-orange.fr/CBZINSOU/CH-19.htm](http://pagesperso-orange.fr/CBZINSOU/CH-19.htm).