



**HAL**  
open science

# Évolution des concentrations d'enzymes dans les réseaux métaboliques

Charlotte Coton

► **To cite this version:**

Charlotte Coton. Évolution des concentrations d'enzymes dans les réseaux métaboliques. Evolution [q-bio.PE]. Université Paris-Saclay, 2021. Français. NNT : 2021UPASL109 . tel-03563233

**HAL Id: tel-03563233**

**<https://theses.hal.science/tel-03563233v1>**

Submitted on 9 Feb 2022

**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.

Évolution des concentrations d'enzymes  
dans les réseaux métaboliques  
*Evolution of enzyme concentrations  
in metabolic networks*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 577, Structure et dynamique des systèmes vivants (SDSV)  
Spécialité de doctorat : Sciences de la vie et de la santé  
Unité de recherche : Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon,  
91190, Gif-sur-Yvette, France  
Réfèrent : Faculté des sciences d'Orsay

**Thèse présentée et soutenue à Paris-Saclay,  
le 17 décembre 2021, par**

**Charlotte COTON**

**Composition du Jury**

<b>Olivier TENAILLON</b> Directeur de recherche, Inserm, Université de Paris	Président
<b>María Luz CÁRDENAS</b> Chercheuse, CNRS, Université Aix-Marseille	Rapporteuse & Examinatrice
<b>Luis-Miguel CHEVIN</b> Directeur de recherche, CNRS, Université de Montpellier	Rapporteur & Examineur
<b>Wolfram LIEBERMEISTER</b> Directeur de recherche, INRAE, Jouy-en-Josas	Examineur

**Direction de la thèse**

<b>Dominique DE VIENNE</b> Professeur émérite, Université Paris-Saclay	Directeur de thèse
<b>Christine DILLMANN</b> Professeure, Université Paris-Saclay	Co-directrice de thèse







*À mes grands-parents*



# Remerciements

Avant toute chose, je souhaite remercier Dominique DE VIENNE et Christine DILLMANN pour m’ avoir proposé ce sujet et m’ avoir guidée tout au long de ses trois années, en me faisant profiter de leur aide et leurs conseils. Je remercie en particulier Dominique, qui a su jongler avec mes inquiétudes et mon entêtement.

Merci aux membres de mon comité de thèse, Sylvain BILLIARD, Yves GIBBON, Hidde DE JONG, et plus particulièrement à Arnaud LE ROUZIC et Wolfram LIEBERMEISTER, pour leur expertise et leurs conseils pour mieux cadrer mon sujet de thèse.

Merci à Marià Luz CÀRDENAS et Luis-Miguel CHEVIN d’ avoir accepté de jouer le rôle de rapporteurs, et à Pierre CAPY, Wolfram LIEBERMEISTER et Olivier TENAILLON d’ examiner ce travail.

Merci à Pierre CAPY et Cécile FAIRHEAD pour leur super Master 2 de recherche « Génétiques, Génome et Evolution » qui m’ a permis de découvrir tant de choses sur l’ évolution.

Merci à mes professeurs de l’ école vétérinaire ONIRIS, Julie HERVÉ, Catherine MAGRAS et Nathalie RUVOEN pour m’ avoir soutenue dans cette voie de recherche.

Parmi mes collègues du Moulon, je tiens à remercier en premier lieu Rozenn LE GUYADER et Sandrine BIHAN, pour toute l’ aide qu’ elles m’ ont apportée. Rozenn, je te souhaite une excellente retraite.

Je remercie également toute l’ équipe BASE, Judith LEGRAND, Élodie MARCHADIER, Adrienne RESSAYRE, Matthieu FALQUE, Nathalie GALIC et tous les autres, et en particulier Adrienne pour les connaissances qu’ elle m’ a transmises. Les nouveaux doctorants de l’ équipe, Sacha REVILLON et Yacine DJABALI, apportent leur dynamisme et leur expertise en ping-pong.

Je remercie notamment mes collègues de bureau, Edith GAROT et Romain BENOIST pour leurs discussions et leur gaieté, malgré la crise sanitaire, avec une mention particulière pour Inoussa SANANÉ, modèle de courage et de persévérance. Inoussa, j’ espère que tu pourras réaliser les projets qui te tiennent à cœur.

Merci à Arnaud DESBIEZ-PIAT pour les nombreuses discussions sur nos sujets de thèse respectifs et pour son soutien. Je te souhaite le meilleur pour la suite.

Merci à Maéva MOLLION pour son amitié.

Merci à Baptiste ROUGER, autre doctorant de l’ année 2018, pour les moments de rigolade et pour tes conseils.



Merci à Emma FORST, Gaelle VAN FRANK et Emmanuelle BLANC pour leur dynamisme et leur joie de vivre.

Merci à Elise TOURRETTE et Xavier RAFFOUX pour leur bonne humeur et pour les discussions jouées. Je remercie également Romane LAFONTAINE, qui vient d'arriver au Moulon, et qui m'a pourtant déjà beaucoup aidée en cette fin de thèse.

Merci à Aurélien BEUGNOT, Dimitri SANCHEZ, Alizarine LORENZI, et tous les autres doctorants du Moulon, pour les moments de rigolade et de discussion lors des nombreuses pauses et goûters.

Merci aussi à tous les moulonnais que j'aurais oubliés, et qui m'ont tant apporté. J'espère un jour retrouver cette ambiance chaleureuse qui règne au Moulon.

Merci à Andreas ODORICO et Thibault CARON, doctorants EGCE et ESE, avec qui j'ai pu beaucoup discuter. J'espère que vous avez trouvé un chemin pour la suite.

Merci à Julia MARIETTE, pour son dynamisme au sein du groupe des doctorants de l'IDEEV.

Merci à Chrystelle DUHEM, Sylvie SALAMITOU, Christine VASSILIADIS, Xavier AUBRIOT, Laura PIGNON, Lola LUGAND et Victor ROCHER, avec qui j'ai travaillé dans le cadre de ma mission doctorale consacrée à la préparation de deux expositions. L'exposition « Au fil des êtres vivants : les Vertébrés, une biodiversité visible », inaugurée récemment, est un vrai succès, merci à tous !

Merci à ma famille, pour son soutien sans faille.

Merci à Julien, qui m'apporte tant tous les jours. La vie sans toi n'aurait pas la même saveur.

# Table des matières

<b>Introduction générale</b>	<b>3</b>
0.1 Modéliser l'évolution . . . . .	4
0.1.1 Les modèles en sciences . . . . .	4
0.1.2 Notions de base de génétique des populations . . . . .	5
0.1.3 Notions de base en dynamique adaptative . . . . .	7
0.1.4 Notion de paysage adaptatif . . . . .	9
0.2 La relation génotype-phénotype . . . . .	10
0.2.1 Les niveaux d'organisation biologique . . . . .	10
0.2.2 Modéliser la relation génotype-phénotype . . . . .	12
0.3 La relation enzymes-flux comme modèle de la relation génotype-phénotype . .	12
0.3.1 Les réseaux métaboliques . . . . .	12
0.3.2 La relation enzymes-flux dans le cadre de la théorie du contrôle métabolique . . . . .	13
0.3.3 Apports de la TCM en biologie évolutive . . . . .	16
0.4 Objectifs de la thèse . . . . .	18
<b>1 Article : <i>Evolution of enzyme levels in metabolic pathways : A theoretical approach</i></b>	<b>23</b>
1.1 Présentation de l'article . . . . .	23
1.2 Corps de l'article . . . . .	27
1.3 Discussion complémentaire . . . . .	65
1.3.1 Échelles de temps . . . . .	65
1.3.2 Relation entre les équilibres évolutifs des concentrations relatives et l'état stationnaire du flux . . . . .	66
1.3.3 Taille de population et neutralité sélective . . . . .	67
1.3.4 Opérons et co-régulations . . . . .	70
1.3.5 Infinité de paramètres pour un même paysage adaptatif . . . . .	70
<b>2 Article : <i>Evolution of enzyme levels in metabolic pathways : A theoretical approach.</i></b>	
<b>Part 2</b>	<b>75</b>
2.1 Présentation de l'article . . . . .	75
2.2 Corps de l'article . . . . .	79

2.3	Discussion complémentaire . . . . .	109
<b>3</b>	<b>La relation enzymes-flux dans les réseaux métaboliques : analogie avec les circuits électriques</b>	<b>113</b>
3.1	Modéliser la relation enzymes-flux dans les réseaux métaboliques . . . . .	114
3.1.1	Les modèles stœchiométriques . . . . .	114
3.1.2	Application de la TCM aux réseaux . . . . .	114
3.2	L'analogie métabolisme-électricité . . . . .	115
3.2.1	Historique de l'analogie entre réseaux métaboliques et circuits électriques	115
3.2.2	Rappels sur le fonctionnement des réactions chimiques . . . . .	117
3.2.3	Cadre de l'analogie métabolisme-électricité et équivalence des grandeurs	120
3.2.4	Les lois électriques . . . . .	123
3.2.5	Expression du flux en fonction de toutes les enzymes du réseau : recherche de l'enzyme équivalente . . . . .	125
3.2.6	Relation entre l'intensité électrique d'un circuit et la conductance d'une résistance particulière . . . . .	130
3.2.7	Discussion . . . . .	134
3.3	Perspectives . . . . .	137
	Résultats préliminaires sur l'évolution des enzymes en parallèle . . . . .	139
	Glossaire des symboles mathématiques . . . . .	141
	<b>Conclusions et perspectives</b>	<b>145</b>
	<b>Bibliographie</b>	<b>151</b>
	<b>Annexes</b>	<b>159</b>
<b>A</b>	<b>Article 1 : figures supplémentaires et démonstrations</b>	<b>161</b>
A.1	<i>Evolution of enzyme levels in metabolic pathways : A theoretical approach – Supporting Information</i> . . . . .	163
A.2	Démonstrations complémentaires . . . . .	185
<b>B</b>	<b>Article 2 : figures supplémentaires et démonstrations</b>	<b>191</b>
B.1	<i>Evolution of enzyme levels in metabolic pathways : A theoretical approach. Part 2 – Supporting Information</i> . . . . .	193
B.2	Démonstrations complémentaires . . . . .	217
<b>C</b>	<b>Chapitre 3 : démonstrations</b>	<b>219</b>
C.1	Expression de l'efficacité équivalente pour des enzymes en série . . . . .	219
C.2	Expression de l'efficacité équivalente pour des enzymes en parallèle . . . . .	222
C.3	Expression du flux pour un réseau métabolique simple . . . . .	223

C.4	Expression de l'efficacité de l'enzyme équivalente dans le cas du pont de Wheatstone . . . . .	224
C.5	Démonstration de la relation concave entre une conductance et l'intensité . . . . .	227
C.5.1	Introduction . . . . .	227
C.5.2	Wheatstone bridge like metabolic network . . . . .	227
C.5.3	Generic non-simplifiable circuit like metabolic network . . . . .	231
C.6	Analogie entre métabolisme et chaîne de production . . . . .	235
<b>D</b>	<b>Évolution des enzymes en parallèle : démonstrations</b>	<b>237</b>
D.1	Rappels . . . . .	237
D.2	Indépendance . . . . .	238
D.3	Compétition . . . . .	239



# Liste des figures

1	Évolution de la fréquence des mutations au cours du temps . . . . .	6
2	Fitness d'invasion en dynamique adaptative . . . . .	8
3	Exemple de paysage adaptatif en trois dimensions . . . . .	10
4	Les niveaux phénotypiques . . . . .	11
5	Définition du flux dans un réseau métabolique . . . . .	13
6	Relation enzymes-flux dans une chaîne linéaire . . . . .	15
7	Le coefficient de contrôle . . . . .	16
8	Modèle de la dominance inspiré de la théorie du contrôle métabolique . . . . .	17
9	Modèle de la neutralité sélective inspiré de la théorie du contrôle métabolique . . . . .	18
1.1	Variation des concentrations d'enzymes dans les trois échelles de temps . . . . .	65
1.2	Probabilité de fixation d'une mutation en fonction du coefficient de sélection $s$ pour différentes tailles de population $N$ . . . . .	68
1.3	Évolution du flux au cours du temps dans une voie métabolique linéaire à trois enzymes en cas de compétition, pour deux tailles de population . . . . .	69
1.4	Évolution du flux au cours du temps dans une voie métabolique linéaire à trois enzymes en cas d'indépendance, pour deux tailles de population . . . . .	69
3.1	Résistance et enzyme. . . . .	121
3.2	Exemple d'un réseau métabolique simple et de son équivalent électrique. . . . .	123
3.3	Exemple de circuit électrique mixte et son équivalent métabolique. . . . .	128
3.4	Pont de Wheatstone . . . . .	129
3.5	Relation entre l'efficacité d'une enzyme quelconque du réseau et l'efficacité de l'enzyme équivalente, les autres étant fixées . . . . .	132
3.6	Simulation de l'évolution des concentrations de trois enzymes catalysant la même réaction en parallèle. . . . .	140
C.1	Circuit électrique et voie métabolique en série. . . . .	219
C.2	Circuit électrique et réseau métabolique en parallèle. . . . .	222
C.3	Exemple de réseau métabolique simple à cinq enzymes avec une boucle de dérivation. Cet exemple théorique est utilisé afin d'illustrer la méthode de résolution. . . . .	223
C.4	Pont de Wheatstone : analogie métabolisme-électricité . . . . .	225

LISTE DES FIGURES

C.5 Electrical wheatstone bridge . . . . . 227

---

# **INTRODUCTION GÉNÉRALE**

---





# Introduction générale

Une propriété fondamentale du monde vivant, apparu il y a plus de trois milliards et demi d'années, est qu'il ne cesse de se transformer. L'étude de ce processus continu qu'est l'évolution<sup>1</sup> est indispensable pour comprendre l'extraordinaire diversité des êtres vivants présents actuellement sur Terre. Dans le courant des idées de son temps, Darwin chercha à expliquer scientifiquement cette diversité, sans avoir recours à une quelconque action divine. Dans *On the origin of species* [Darwin 1859], il présenta sa célèbre théorie de l'évolution par la sélection naturelle, qui repose sur trois grands principes : la variation de forme et de comportement des organismes, la sélection naturelle qui trie les organismes à chaque génération (les plus aptes à survivre et à se reproduire ont plus de descendants) et la transmission des caractères.

Inspirée par les thèses de Thomas Malthus (1766–1834), par d'innombrables observations faites lors de son voyage sur le Beagle et par l'exemple de la sélection artificielle, la théorie darwinienne de l'évolution est entièrement formulée en langage naturel. Ce n'est qu'au cours de la première moitié du XXe siècle, avec le développement de la génétique, de la génétique des populations et de la génétique quantitative que le formalisme mathématique va permettre d'étudier avec plus de précision les processus évolutifs. Enrichie de ces nouvelles approches, la théorie originelle de Darwin va devenir la théorie synthétique de l'évolution, ou néodarwinisme.

Dans la seconde moitié du XXe siècle, la découverte de niveaux de polymorphismes moléculaires très élevés dans les populations naturelles va renforcer le recours aux modèles mathématiques, dans le contexte du développement de la théorie neutraliste de l'évolution moléculaire [Kimura 1985]. À côté du rôle de la sélection, le rôle du hasard dans l'évolution est pleinement reconnu.

---

<sup>1</sup>Issu du latin *evolutio*, « action de dérouler, de parcourir », dérivé de *evolvere*, « dérouler », de *volvere*, « rouler » [Académie française 2021].

## 0.1 Modéliser l'évolution

Avant de présenter les modèles d'évolution sur lesquels je m'appuie, je vais présenter ce qu'est un modèle scientifique.

### 0.1.1 Les modèles en sciences

La notion de modèle en sciences a fait l'objet de nombreux travaux au cours des dernières décennies [Drouin 1988; Rumelhard 1988; Rosen 1993; Legay 1997; Bolker 2009; Soler 2013; Servedio *et al.* 2014]. Les modèles sont présents dans tous les domaines scientifiques, biologie, physique, économie, sciences humaines, etc. Le terme « modèle » vient de l'ancien français *modelle*, signifiant « moule » ou « représentation d'un objet » [Godefroy 1888], emprunté à l'italien *modello*, de même sens [Académie française 2021], issu lui-même du latin *modulus*, « mesure » [Legay 1997]. Une définition assez générale d'un modèle scientifique est la suivante : « "quelque chose" (objet concret, représentation imagée, système d'équations, etc.) qui se substitue au réel trop complexe, ou inaccessible à l'expérience, et qui permet de comprendre ce réel par un intermédiaire plus connu ou plus accessible à la connaissance » [Drouin 1988]. Selon Legay [1997] et Soler [2013], les modèles permettent de décrire des phénomènes, d'expliquer des mécanismes ou de prédire des comportements, sans oublier leur rôle pédagogique pour illustrer un concept ou enseigner. Pour créer un modèle, il est nécessaire de poser le problème et de choisir un point de vue, après avoir correctement analysé la situation. Un modèle est efficace s'il répond au problème posé tout en étant facile d'utilisation, et on dit qu'il est fécond s'il apporte des résultats inattendus. Pour étudier un phénomène, il peut être nécessaire d'utiliser plusieurs modèles selon les points de vue pour en saisir différents aspects [Legay 1997; Soler 2013].

Il existe plusieurs types de modèles, que l'on peut classer selon le point de vue ou l'usage. Legay [1997] distingue notamment les modèles dialectiques, basés sur un langage (verbal, mathématique, ou schématique), des modèles physiques, basés sur des objets ou phénomènes physiques pour étudier un processus plus complexe (par exemple, les modèles biologiques). Les modèles formels simples, appelés *toy models* ou *proof-of-concept models*, sont particulièrement utiles pour formaliser des idées et définir des notions dans un cadre précis, ce que les modèles verbaux et intuitifs ne permettent pas [Servedio *et al.* 2014].

Les modèles ont bien sûr leurs limites, car en tant que simplifications de la réalité, ils lui sont nécessairement infidèles [Soler 2013]. Un modèle s'applique à un problème particulier dans un cadre précis, et s'ils sont utilisés en-dehors de leur domaine de validité et du cadre de leur construction, les résultats seront probablement faux [Legay 1997; Soler 2013]. Un modèle doit également être confronté à la réalité pour vérifier sa validité, rendant l'expérimentation indispensable [Legay 1997]. Lorsque cette dernière est impossible, ce qui est souvent le cas en évolution, les simulations informatiques permettent d'« expérimenter » dans une grande diversité de situations et/ou d'échelles de temps.

## 0.1.2 Notions de base de génétique des populations

Il y a plusieurs manières d'étudier l'évolution des êtres vivants. L'une d'entre elles consiste à modéliser le changement de fréquence des gènes et/ou allèles<sup>2</sup> au cours du temps dans les populations. C'est l'objet de la génétique des populations [Wright 1937; Crow et Kimura 1970; Ohta 1992; Nielsen et Slatkin 2013]. Une population est un ensemble d'individus qui peuvent se reproduire entre eux et qui, par les caractéristiques qu'ils partagent (*ex.* un territoire commun), forment un groupe suffisamment distinct du reste de l'espèce. Chaque individu est caractérisé par son génotype, c'est-à-dire sa constitution génétique, et par sa *fitness*. La fitness d'un individu est sa valeur sélective, qui dépend de sa survie et de sa fertilité, et qui correspond au nombre de ses descendants laissés à la génération suivante [Wright 1937; Crow et Kimura 1970].

Quatre « forces évolutives »<sup>3</sup> modifient les fréquences des gènes dans les populations :

- la mutation, qui est une modification accidentelle du support de l'hérédité et donc des gènes, créant de nouveaux allèles ;
- la migration, qui apporte de nouveaux allèles à la population ;
- la sélection naturelle, qui est un tri basé sur la fitness ;
- la dérive génétique, qui est l'échantillonnage aléatoire des allèles d'une génération à l'autre dans une population.

Une cinquième « force » est parfois mentionnée, la recombinaison génétique, qui modifie la répartition des allèles le long des chromosomes, et qui, comme la mutation, apporte de la variation dans les génotypes de la population. Le régime de reproduction peut également impacter les fréquences génotypiques. À cela on peut ajouter l'influence de la démographie sur ces processus.

À chaque génération, des mutations apparaissent aléatoirement, créant de nouveaux allèles. Les mutations peuvent disparaître, ou bien se fixer dans la population, c'est-à-dire que tous les individus vont à terme porter cette mutation (figure 1). Le devenir de ces mutations dépend de deux des forces évolutives évoquées plus haut, la dérive génétique et la sélection naturelle, *via* respectivement la « taille efficace » de la population et le coefficient de sélection, mais aussi du degré de ploïdie de l'espèce [Crow et Kimura 1970].

La taille efficace d'une population (notée  $N_e$ ) est la taille d'une population « idéale » de Wright-Fisher (taille constante  $N$ , sans sélection ni migration, individus hermaphrodites, générations non chevauchantes et panmixie), où la dérive génétique serait de même ampleur que dans la population considérée [Crow et Kimura 1970]. En général, la taille efficace d'une population

<sup>2</sup>Un gène est une séquence de nucléotides codant pour une fonction, et les allèles sont les versions d'un gène.

<sup>3</sup>Le terme « force » n'est pas à prendre au sens de la physique, qui pousse dans une direction, mais plutôt au sens de mécanisme ou de processus, qui influence l'évolution.

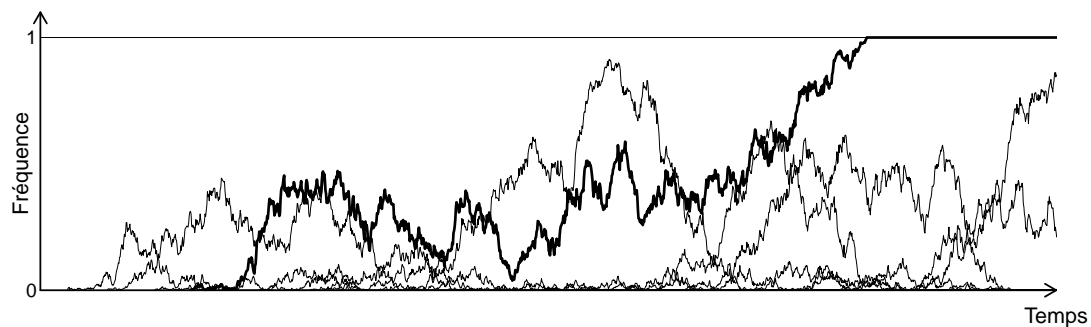


FIGURE 1 – **Évolution de la fréquence des mutations au cours du temps** (d'après Kimura [1985]). Chaque trajectoire représente l'évolution de la fréquence d'une mutation. Les trajectoires en gras indiquent une fixation de la mutation, et les autres indiquent une disparition de la mutation.

est inférieure à la taille réelle de la population, car un ou plusieurs des critères qui définissent la population idéale ne sont pas respectés dans les conditions naturelles. La dérive génétique est d'autant plus forte que la taille efficace de la population est faible. Par effet d'échantillonnage, un allèle mutant apparu dans une population de faible effectif (ex.  $N = 10$ ) a une forte probabilité d'être tiré et peut donc voir sa fréquence augmenter très rapidement aux générations suivantes et se fixer même s'il est délétère. Au contraire, si la population est très grande ( $N = 100\,000$ ), la probabilité de tirage du mutant est très faible et sa fréquence varie très peu au cours des générations.

Le coefficient de sélection, noté  $s$ , est la différence de fitness, absolue ou relative, entre deux individus. Si la population est homogène, ce que l'on considère dans ce travail, le coefficient de sélection d'une nouvelle mutation sera calculé relativement à la fitness du génotype initialement présent, appelé le génotype « résident ». Plus un génotype mutant est « performant », c'est-à-dire avec une fitness élevée par rapport au reste de la population, plus son coefficient de sélection est élevé.

Le degré de ploïdie fait référence au nombre de copies de chromosomes présentes dans les cellules. Chaque chromosome porte un ensemble de gènes, et chaque copie de chromosome possède les mêmes gènes mais avec des allèles en général différents. Une cellule haploïde possède une copie de chaque chromosome, une cellule diploïde deux copies, etc. Toutes choses égales par ailleurs, le degré de ploïdie joue sur la probabilité de fixation des mutations. Dans ma thèse j'ai considéré des populations haploïdes, pour ne pas avoir à considérer de possibles effets de dominance. En cas de dominance, la fitness de l'hétérozygote est différente de la fitness moyenne des homozygotes, ce qui implique de gérer non pas deux, mais trois valeurs de fitness différentes.

Le devenir des mutations peut être estimé à l'aide de la probabilité de fixation. Considérons une population haploïde à reproduction sexuée, de taille  $N$  ( $N$  grand et constant au fil des générations) et de taille efficace  $N_e$ . Une mutation qui affecte un individu de cette population

aura la probabilité de fixation suivante [Crow et Kimura 1970] :

$$P_f = \frac{1 - e^{-2sN_e/N}}{1 - e^{-2sN_e}}$$

où  $s$  est le coefficient de sélection de la mutation. Cette relation peut se simplifier si les croisements se font au hasard (panmixie) et que la distribution du nombre de descendants suit une loi de Poisson. Dans ce cas  $N = N_e$ , et si  $|s|$  est petit on obtient :

$$P_f = \frac{2s}{1 - e^{-2Ns}}$$

qui est la formule que nous utiliserons dans notre modèle d'évolution.

En passant aux limites, on voit que :

- pour une mutation avantageuse,  $P_f \underset{Ns \rightarrow +\infty}{\approx} 2s$ , ce qui signifie que plus une mutation est avantageuse, plus elle a de chances de se fixer
- pour une mutation neutre,  $P_f \underset{Ns \rightarrow 0}{\rightarrow} 1/N$ , et la probabilité de fixation ne dépend alors que de la taille de population
- pour une mutation délétère,  $P_f \underset{Ns \rightarrow -\infty}{\rightarrow} 0$ , ce qui signifie que plus une mutation est délétère, moins elle a de chance de se fixer.

On a coutume de distinguer trois zones différentes pour le devenir d'une nouvelle mutation [Nielsen et Slatkin 2013]. Si  $Ns > 1$ , l'allèle mutant est avantageux, et il aura d'autant plus de chances de se fixer que son avantage sélectif est important. Si  $-1 < Ns < 1$ , l'avantage sélectif du mutant est faible, la mutation a peu d'effet sur la fitness, et la mutation est dite quasi-neutre : sa probabilité de fixation ne dépend que de la taille de la population et est quasiment identique à celle d'une mutation strictement neutre. On parle alors de zone neutre. Si  $Ns < -1$ , l'allèle mutant est délétère, et il aura peu de chance de se fixer. La transition entre allèle neutre et allèle sous sélection est graduelle [Hartl *et al.* 1985]. Comme la dérive prédomine dans la zone neutre, on parle de « neutralité sélective ». Cette notion de quasi-neutralité est très importante : elle implique que la perte ou la fixation d'une mutation peut être le fait du hasard, et c'est parce que le devenir d'une mutation peut prendre de nombreuses générations que l'on observe des niveaux élevés de polymorphisme dans les populations naturelles. C'est le cœur de la théorie neutraliste de l'évolution.

### 0.1.3 Notions de base en dynamique adaptative

Une autre manière de modéliser l'évolution est de regarder l'évolution d'un caractère sur de grandes échelles de temps, en prenant en compte l'environnement dans lequel les organismes vivent et la densité de la population. C'est l'objectif de la dynamique adaptative. Dans ce cadre,

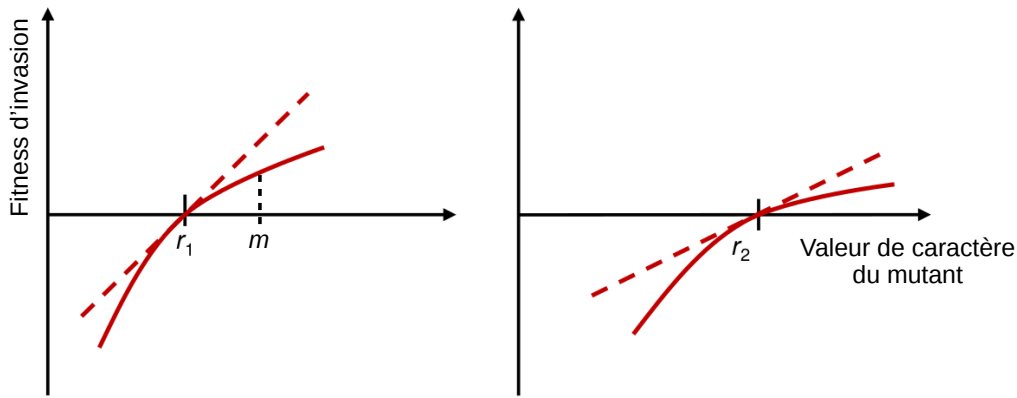


FIGURE 2 – **Fitness d’invasion en dynamique adaptative** (d’après Brännström *et al.* [2013]). La valeur du caractère chez le mutant est en abscisse, et sa fitness d’invasion, ou taux de croissance, est en ordonnée. La courbe rouge est la fitness d’invasion d’un mutant au sein d’une population résidente monomorphe de valeur  $r_1$  (à gauche) ou  $r_2$  (à droite). La fitness d’invasion est nulle pour la valeur du résident, car sa population est à l’équilibre. À gauche : La population résidente est monomorphe de valeur  $r_1$ . Un mutant avec une fitness d’invasion positive va envahir la population et remplacer le résident. C’est le cas par exemple pour un mutant de valeur  $m$ . À droite : Le mutant de valeur  $m$  a envahi la population, et est devenu le nouveau résident de valeur  $r_2$ . La fitness d’invasion a changé de forme, car elle dépend de la population résidente.

on cherche les valeurs du caractère qui optimisent la fitness. Pour cela on ignore la dérive, et on ne prend en compte que la sélection, c’est-à-dire que si un mutant a une meilleure fitness que le reste de la population, il envahira la population, sinon il disparaîtra.

On considère une population préalablement installée, dite population résidente. Dans les cas les plus simples, on considère que la population résidente est monomorphe, ou homogène, c’est-à-dire que tous les individus possèdent la même valeur  $x$  du caractère étudié. La taille de la population résidente est supposée à l’équilibre dynamique, c’est-à-dire qu’elle ne varie pas au cours du temps. Une mutation apparaît dans cette population résidente, et la valeur de l’individu mutant est  $x'$ . La mutation est de petite taille, c’est-à-dire que la valeur du caractère du mutant est proche de celle du résident ( $x' \approx x$ ). On définit la fitness d’invasion  $s_x(x')$  comme le taux de croissance, ou taux de production individuel de descendants, de ce mutant initialement rare au sein de la population résidente. Cette fitness d’invasion détermine la vitesse d’augmentation de la densité du mutant en présence du résident. Elle dépend de la valeur du caractère du mutant et de celle du résident. Il est à noter que la fitness d’invasion du résident envers lui-même est nulle, car sa densité de population est à l’équilibre. Dans un modèle simple, la vitesse d’accroissement de la population du mutant est fonction de la fitness d’invasion et du nombre d’individus mutants  $n'$  :

$$\frac{dn'}{dt} = n' s_x(x')$$

Le signe de cette fitness d’invasion indique si la mutation se fixe ou non : si elle est positive, le mutant envahit la population et devient le nouveau résident, et si elle est négative, le mutant disparaît. Une illustration de la notion de fitness d’invasion est présentée figure 2.

Dans cette modélisation, les deux échelles de temps, écologique et évolutive, sont séparées. Lorsqu'un mutant apparaît, son interaction avec la population résidente à l'échelle écologique détermine sa fitness d'invasion (*via* par exemple la compétition pour les ressources), et donc la vitesse de remplacement du résident par le mutant. À l'échelle évolutive, les mutations se succèdent sans se chevaucher car le taux de mutation est supposé faible et que chaque mutation se fixe (ou disparaît) instantanément, ce qui permet d'étudier l'évolution des caractères sur de grandes échelles de temps.

Les valeurs de caractère qui optimisent la fitness sont des équilibres évolutifs. L'équilibre le plus courant est celui qui résulte de la stratégie évolutivement stable (*Evolutionarily stable strategy*, ESS) : la population peut atteindre cet équilibre attracteur, et en ce point aucun mutant ne peut l'envahir.

#### 0.1.4 Notion de paysage adaptatif

La relation entre les génotypes ou les caractères et la fitness peut être visualisée à l'aide d'un « paysage adaptatif », ou paysage de fitness. Le concept de paysage adaptatif, initialement proposé par Wright [1932], est une métaphore utile pour visualiser la fitness à partir de l'espace des génotypes ou des caractères, et leur évolution [Gavrilets 2004]. Elle constitue un outil de modélisation pour comprendre par exemple l'adaptation ou les distributions des effets des mutations [Harmand 2017]. Le plus souvent on le représente en deux ou trois dimensions, avec les gènes ou les caractères d'intérêt sur les axes  $x$  et  $y$  et la fitness sur l'axe  $z$ . Chaque point correspond à un individu, ou éventuellement à une population monomorphe, à un temps donné. Au-delà de trois dimensions, la représentation est délicate pour le cerveau humain, et les axes  $x$  et  $y$  sont parfois la combinaison de plusieurs gènes/caractères. Ce paysage peut varier en fonction de l'environnement, et il peut prendre de multiples formes, avec des vallées et des pics : les pics sont les maximums locaux de fitness vers lesquels la population va tendre, et les vallées les minimums locaux de fitness, que la population ne peut franchir que par la dérive ou en les contournant. Une population va parcourir ce paysage au gré des mutations et de la sélection, et éventuellement de la dérive. Un exemple de paysage adaptatif en trois dimensions est présenté figure 3.

Cependant la relation entre les génotypes et les caractères n'est en aucun cas biunivoque. La manière dont les variations au niveau du génotype se traduisent en termes de différences phénotypiques est une question centrale pour l'étude des processus évolutifs, que l'on exprime en terme de « relation génotype-phénotype », ou de *GP map* [Orgogozo *et al.* 2015; de Vienne 2021].



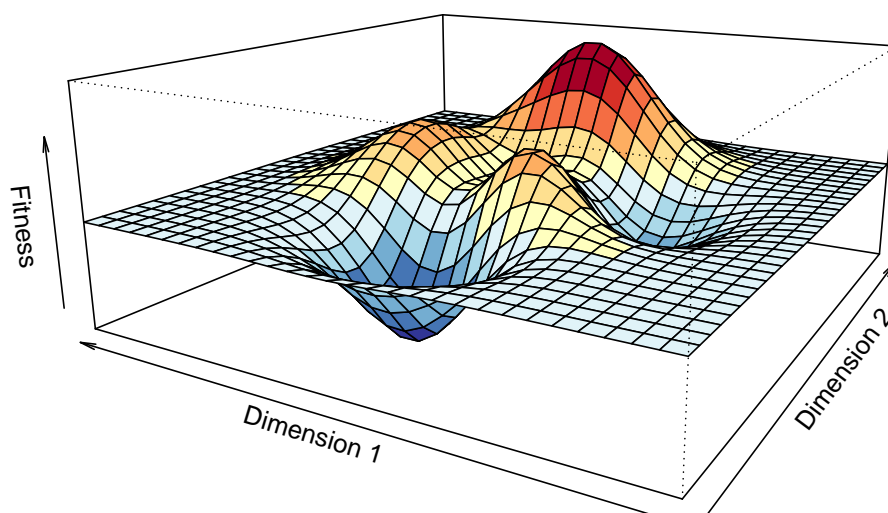


FIGURE 3 – **Exemple de paysage adaptatif en trois dimensions.** Les dimensions 1 et 2 peuvent être des gènes ou des caractères, tandis que la fitness est représentée sur l'axe  $z$ . Le paysage représenté est un exemple de relation entre les gènes (ou caractères) et la fitness. L'échelle de couleur indique le gradient de fitness, des plus faibles valeurs en bleu aux plus fortes en rouge. Les « pics » sont des maximums locaux de fitness, tandis que les « vallées » sont des minimums locaux. Les populations ont tendance à évoluer vers les maximums locaux de fitness, les vallées pouvant être franchies par dérive ou *via* des mutations à effets pléiotropes. L'épistasie est inhérente à ce type de paysage : l'effet d'une mutation d'une taille donnée selon la dimension 1 dépend de la valeur en dimension 2.

## 0.2 La relation génotype-phénotype

Le génotype est l'ensemble de l'information génétique d'un individu. Dans un sens plus restreint, c'est la composition allélique d'un ou plusieurs gènes d'intérêt chez un individu. Le phénotype<sup>4</sup> est l'expression du génotype dans un environnement donné. C'est la manière dont nous apparaît un individu, ou un caractère particulier, dans cet environnement. Dans un sens plus restreint, le phénotype est la valeur, qualitative ou quantitative, d'un caractère chez cet individu. Le phénotype dépend donc du génotype, de l'environnement (biotique et abiotique) et de l'interaction entre ce génotype et l'environnement. Depuis [Johannsen \[1911\]](#), on sait que des génotypes différents peuvent avoir le même phénotype, et que selon l'environnement un génotype donné peut présenter des phénotypes différents.

Pour se traduire à des niveaux phénotypiques intégrés, une différence génotypique doit se propager à travers une série de niveaux d'organisation de complexité croissante (figure 4).

### 0.2.1 Les niveaux d'organisation biologique

Les niveaux d'organisation – atomes, molécules, macro-molécules, cellules, tissus, organes, organismes – ne sont pas simplement des catégories commodes pour décrire et transmettre nos connaissances sur le vivant. Il y a de vraies discontinuités entre eux, en ce sens qu'il y a bien

<sup>4</sup>Du grec *phainien* « rendre visible », et de -type, tiré du grec *tupos*, « marque, empreinte » [[Académie française 2021](#)].

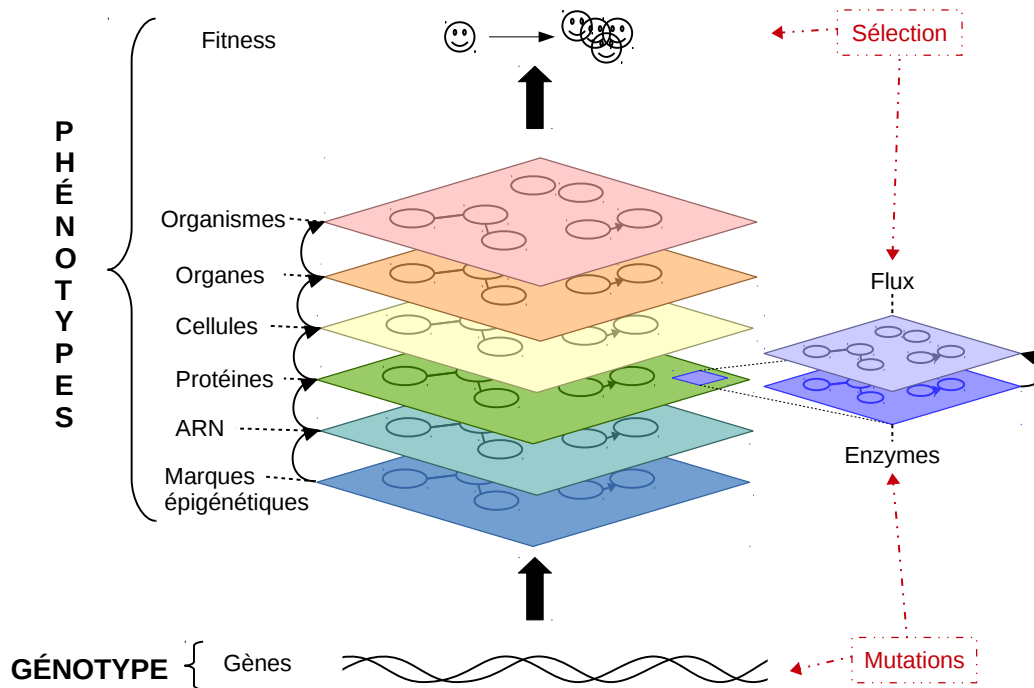


FIGURE 4 – **Les niveaux phénotypiques.** Des phénotypes peuvent être mesurés aux différents niveaux d'organisation biologique. À chaque niveau, les réseaux d'interactions permettent l'émergence de nouvelles propriétés au niveau supérieur (flèches reliant les niveaux phénotypiques adjacents), l'environnement pouvant agir sous forme de contraintes. Les effets des autres niveaux phénotypiques font partie de cet environnement. Les mutations apparaissent au niveau du génotype tandis que la sélection agit sur les phénotypes, et plus particulièrement la fitness, le phénotype « ultime ». Au niveau « protéines », les enzymes interviennent au sein de réseaux métaboliques, et de ces réseaux émergent les flux métaboliques. La relation enzymes-flux est un modèle de la relation génotype-phénotype.

plus d'interactions intra-niveaux qu'inter-niveaux [Simon 1962]. Cette hiérarchie d'organisation inhérente aux systèmes vivants a été forgée par l'évolution [Simon 1962; Vrba et Eldredge 1984; Maynard Smith et Szathmáry 1995], et c'est ce qui les rend intelligibles et accessibles à l'analyse [Bunge 1977; Mossio *et al.* 2016].

Les propriétés d'un niveau lui sont spécifiques. Selon la notion d'émergence (« Le tout est plus que la somme de ses parties »), les interactions entre éléments d'un niveau permettent l'apparition de nouvelles propriétés au niveau supérieur. Par exemple, la transparence et la viscosité de l'eau n'ont pas de sens au niveau d'une molécule d'eau, ni bien sûr au niveau des atomes d'hydrogène et d'oxygène ; la catalyse enzymatique « émerge » de l'enchaînement précis d'acides aminés dans une protéine ; des assemblages ordonnés de phospho-lipides confèrent leurs propriétés spécifiques aux membranes de la cellule ; des circuits moléculaires particuliers engendrent les rythmes biologiques, etc.

Si sur le plan structural l'organisation *bottom-up* du vivant est évidente, il en va autrement sur le plan fonctionnel car les niveaux supérieurs peuvent influencer les niveaux inférieurs (concept de *downward causation* [Campbell 1974]). Par exemple le fonctionnement d'une enzyme est influencé par la composition et l'encombrement du milieu cellulaire, ou encore l'expression de certains gènes est sous l'influence de signaux hormonaux de communication entre

cellules *via* des cascades de signalisation.

## 0.2.2 Modéliser la relation génotype-phénotype

Comprendre la relation GP est capital en sciences de l'évolution. En effet, si les mutations affectent le génotype, c'est au niveau du phénotype que la sélection naturelle agit, modifiant la structure génétique des populations. Cette compréhension de la relation GP est également au cœur d'autres disciplines, comme la biologie synthétique ou la cancérologie [Manrubia *et al.* 2021].

Cette relation est extrêmement complexe, en raison des multiples interactions intra- et inter-niveaux qui interviennent à toutes les échelles d'organisation du vivant. La démarche courante pour simplifier le problème consiste à s'intéresser aux relations entre un niveau donné, considéré comme porteur de l'information génétique, et un niveau phénotypique supérieur (pour revue, voir Kemble *et al.* [2019]; Manrubia *et al.* [2021]). Cette approche fut initiée par Wright [1934] lorsqu'il proposa le premier modèle biochimique de la dominance (voir ci-dessous), avec le flux métabolique<sup>5</sup> comme niveau phénotypique et l'activité des enzymes comme niveau génotypique.

C'est ce modèle de relation génotype-phénotype que j'ai utilisé dans mon travail de thèse, en choisissant les concentrations d'enzymes, et non pas les activités, comme niveau génotypique.

## 0.3 La relation enzymes-flux comme modèle de la relation génotype-phénotype

### 0.3.1 Les réseaux métaboliques

Les interactions entre objets biologiques peuvent être visualisées sous la forme de réseaux, présents à tous les niveaux d'organisation du vivant : réseaux d'interactions entre protéines, réseaux de signalisation, réseaux de gènes, réseaux neuronaux, etc. [Proulx *et al.* 2005]. Les réseaux peuvent être classés en fonction du rôle des objets (les noeuds) et de leurs interactions (les arêtes). Les arêtes peuvent être orientées ou non, les noeuds peuvent être homogènes ou hétérogènes. Les analogies entre réseaux autorisent des méthodes d'analyse issues de domaines parfois distants (physique, informatique, réseaux sociaux par exemple) [Proulx *et al.* 2005]. Dans le chapitre 3 j'exploite l'analogie entre réseaux métaboliques et réseaux électriques.

Les réseaux métaboliques sont parmi les réseaux biologiques les mieux connus. Les constituants chimiques de la cellule, appelés métabolites, réagissent entre eux pour former de nouvelles molécules : ce sont les réactions chimiques. Les enzymes sont des protéines qui catalysent, c'est-à-dire accélèrent, ces réactions chimiques. Les réseaux métaboliques peuvent être

---

<sup>5</sup>Le métabolisme est l'ensemble des réactions chimiques se produisant au sein d'une cellule ou d'un organisme. Le terme « métabolisme » est issu du grec *metabolê*, signifiant « changement » [Académie française 2021].

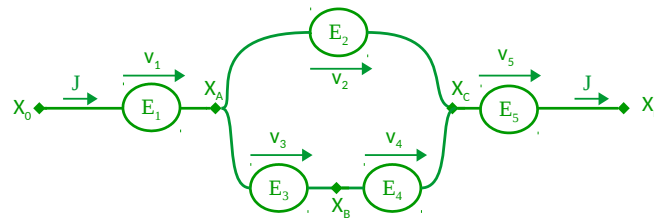


FIGURE 5 – **Définition du flux dans un réseau métabolique.** Dans un réseau métabolique, chaque réaction de vitesse  $v_i$  est catalysée par une enzyme  $E_i$ . Le réseau représenté ici possède cinq réactions.  $X_A$ ,  $X_B$  et  $X_C$  sont les métabolites intermédiaires.  $X_0$  et  $X_n$  sont respectivement le substrat et le produit du réseau. Leur vitesse de consommation et de production sont égales au flux métabolique, qui dépend de l'ensemble des vitesses des enzymes du réseau.

représentés de deux manières selon les objectifs : les enzymes sont les nœuds et les métabolites les arêtes, ou bien les enzymes sont les arêtes et les métabolites les nœuds. C'est cette dernière représentation, la plus courante, que j'utilise dans cette thèse.

Les enzymes michaeliennes sont caractérisées par leurs paramètres cinétiques – constante catalytique (ou *turnover number*)  $k_{cat}$ , constante de Michaelis-Menten  $K_M$  et éventuellement constante d'inhibition  $K_I$ . Les enzymes allostériques, qui peuvent changer de conformation en fonction de la fixation d'un ou plusieurs ligands, ont pour paramètres cinétiques la constante de dissociation  $K_d$  et l'indice de coopérativité (ou constante de Hill)  $n_H$ . Les mutations peuvent modifier l'efficacité des enzymes de deux façons : *via* des changements de leurs paramètres cinétiques ou *via* des changements de leur concentration.

Les flux métaboliques sont les variables de sortie des réseaux. Un flux métabolique est un débit, c'est-à-dire la quantité produite ou consommée d'un métabolite par unité de volume et de temps. Je fais ici la distinction entre la *vitesse d'une réaction* et le *flux métabolique* qui est le débit des métabolites d'entrée ou de sortie d'une voie métabolique et qui dépend de l'ensemble des réactions d'une voie métabolique (figure 5).

La valeur d'un flux dépend de paramètres thermodynamiques constants (constantes d'équilibres des réactions) et des paramètres génétiquement variables : les paramètres cinétiques des enzymes et leurs concentrations. Pour modéliser la relation enzymes-flux, j'ai utilisé la théorie du contrôle métabolique.

### 0.3.2 La relation enzymes-flux dans le cadre de la théorie du contrôle métabolique

Dans le but de réunir topologie (réseau métabolique) et enzymologie (étude des enzymes), [Kacser et Burns \[1973\]](#) ont développé la théorie du contrôle métabolique (TCM), dont un des axes permet de relier le flux métabolique aux propriétés des enzymes dans une voie métabolique linéaire. Indépendamment, [Heinrich et Rapoport \[1974\]](#) ont proposé un modèle similaire dans

le même objectif, mais surtout pour mieux comprendre comment sont affectés les flux par les facteurs sous-jacents que sont les métabolites et les activités des enzymes. En 1995, un article de synthèse a fait le point sur la TCM (on parle aujourd'hui plutôt de *Metabolic Control Analysis*), avec des notations mathématiques consensuelles et des définitions rigoureuses [Kacser *et al.* 1995].

La TCM « historique » considère une chaîne métabolique constituée d'une suite de réactions réversibles uni-moléculaires catalysées par des enzymes michaeliennes loin de la saturation. À l'état stationnaire, c'est-à-dire lorsque les vitesses des réactions sont constantes, toutes les vitesses sont égales et identiques au flux métabolique, qui est la vitesse de consommation du substrat de la chaîne. Le flux est alors fonction de l'ensemble des facteurs qui contrôlent les vitesses de réactions : les paramètres cinétiques des enzymes et leurs concentrations. Cette relation s'écrit mathématiquement pour une chaîne métabolique de  $n$  réactions successives :

$$J = \frac{X_0 - X_n/K_{1,n}}{\sum_{j=1}^n \frac{1}{(k_{cat,j}/K_{M_j})E_j K_{1,j-1}}} \quad (1)$$

où  $J$  est le flux métabolique de l'ensemble de la chaîne ;  $X_0$  et  $X_n$  sont respectivement les concentrations en substrat et produit de la chaîne ;  $k_{cat,j}$  et  $K_{M_j}$  sont les paramètres cinétiques, respectivement la constante catalytique et la constante de Michaelis-Menten de l'enzyme  $j$  ;  $E_j$  est la concentrations de l'enzyme  $j$  ; et  $K_{1,j}$  est le produit des constantes d'équilibres des réactions 1 à  $j$  (la constante d'équilibre constituant le « moteur » thermodynamique de la réaction, voir chapitre 3).

À partir de cette équation, on peut voir que la relation entre la concentration d'une enzyme particulière  $E_i$  et le flux est une hyperbole (figure 6) de formule générale :

$$J = \frac{X}{a + \frac{b}{E_i}} \quad (2)$$

où  $X$ ,  $a$  et  $b$  sont des constantes :

$$X = \left( X_0 - \frac{X_n}{K_{1,n}} \right) \quad a = \sum_{\substack{j=1 \\ j \neq i}}^n \frac{K_{M_j}}{k_{cat,j} E_j K_{1,j-1}} \quad b = \frac{K_{M_i}}{k_{cat,i} K_{1,i-1}}$$

L'équation 2 montre que lorsque  $E_i$  tend vers l'infini, le flux tend vers une valeur maximum  $J_{\max} = X/a$ . Cette équation permet aussi de montrer l'influence du fonds génétique sur la forme de la relation entre le flux et la concentration d'une enzyme particulière. En effet, si l'un des paramètres cinétiques de l'enzyme  $i$ , ou bien l'un des paramètres (paramètre cinétique ou concentration) d'une des enzymes  $j \neq i$ , est modifié, le degré de concavité de la courbe et le  $J_{\max}$  vont changer (figure 6). C'est un résultat important, car il fournit une base biochimique à l'épistasie : l'effet d'une mutation sur le flux dépend du fonds génétique. Autrement dit, une

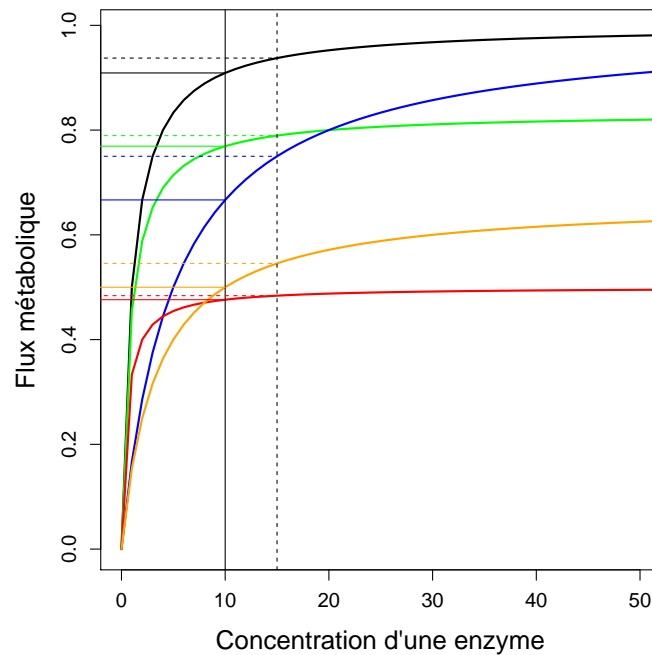


FIGURE 6 – **Relation enzymes-flux dans une chaîne linéaire.** Le flux est fonction de la concentration d'une enzyme  $i$ . La formule utilisée est  $J = \frac{E_i}{aE_i + b}$  où  $a = \sum_{j=1, j \neq i}^n \frac{1}{A_j E_j}$  et  $b = \frac{1}{A_i}$ , avec  $A_i = k_{cat,i}/K_{M_i}K_{1,i-1}$ . Courbe noire :  $a = b = 1$ ; courbe bleue :  $a = 1, b = 5$ ; courbe verte :  $a = 1.2, b = 1$ ; courbe orange :  $a = 1.5, b = 5$ ; courbe rouge :  $a = 2, b = 1$ . Le trait vertical plein indique une concentration de 10 pour l'enzyme  $i$ , et les traits horizontaux pleins indiquent le flux associé selon le fonds génétique. Le trait vertical pointillé montre une mutation de l'enzyme  $i$  de valeur +5, et les traits horizontaux pointillés sont les valeurs de flux associées. On voit clairement que selon le fonds génétique, les mutations n'ont pas le même effet sur le flux.

mutation qui modifie la valeur d'un paramètre enzymatique va modifier l'effet de toutes les mutations ultérieures, quelle que soit l'enzyme ciblée.

Pour quantifier l'effet de la variation d'un paramètre enzymatique sur le flux, on utilise le *coefficient de contrôle*, qui a donné le nom à la TCM. Le coefficient de contrôle de la concentration  $E_i$  de l'enzyme  $i$  sur le flux est :

$$C_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J}$$

C'est donc la pente normalisée de la relation entre paramètre enzymatique et flux. Un coefficient de contrôle faible signifie que la variation des paramètres de l'enzyme a peu d'effet sur le flux, et inversement, un coefficient de contrôle élevé signifie que cette variation a un fort impact sur le flux. Comme la relation entre enzymes et flux est concave et qu'elle tend vers un plateau, le coefficient de contrôle est grand lorsque que les valeurs des paramètres enzymatiques sont faibles, et petit lorsqu'elles sont élevées (figure 7).

Une propriété remarquable est que la somme des coefficients de contrôle d'une voie méta-

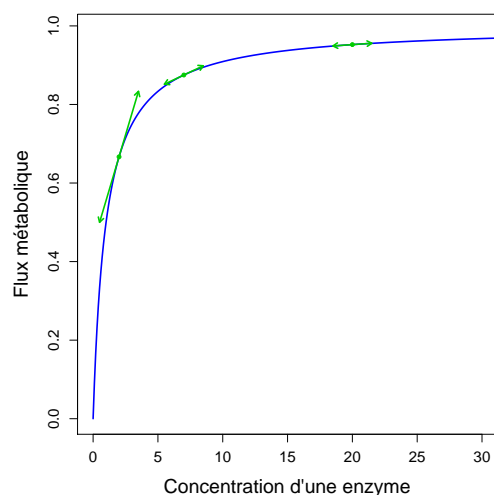


FIGURE 7 – **Illustration du coefficient de contrôle.** Le flux est fonction de la concentration d’une enzyme  $i$ , selon la même relation que dans la figure 6, avec  $a = b = 1$ . Les doubles-flèches indiquent la pente de la courbe pour différentes valeurs de  $E_i$ . Après normalisation, cette pente donne la valeur du coefficient de contrôle en ces points. Le coefficient de contrôle diminue lorsque la concentration de l’enzyme considérée augmente.

bolique vaut 1 : c’est la *propriété de sommation*, qui signifie que le contrôle du flux est partagé entre toutes les enzymes. La modification d’un coefficient de contrôle à la suite d’une mutation va modifier tous les autres – c’est une autre manière de montrer que l’épistasie est inévitable pour des caractères reliés à des flux métaboliques. En ingénierie métabolique, la propriété de sommation a permis de comprendre pourquoi l’augmentation de la concentration d’une seule enzyme était, dans la plupart des cas, insuffisante pour améliorer la productivité de la voie métabolique.

Le coefficient de contrôle suppose l’indépendance entre les paramètres enzymatiques. Or, comme nous le verrons pour les concentrations d’enzymes, il existe des contraintes cellulaires qui créent des dépendances (ressources limitées, co-régulations). Il faut dans ces cas utiliser un coefficient plus général, le coefficient de réponse combinée [de Vienne *et al.* 2001; Lion *et al.* 2004], qui mesure la sensibilité du flux vis-à-vis de la variation de la concentration d’une enzyme quand celle-ci est influencée par la variation des autres enzymes.

### 0.3.3 Apports de la TCM en biologie évolutive

La théorie du contrôle métabolique a été utilisée à maintes reprises comme modèle de la relation génotype-phénotype, en considérant les propriétés enzymatiques comme reflétant le niveau génotypique et le flux métabolique comme le phénotype d’intérêt directement lié à la fitness. La relation non-linéaire entre les propriétés des enzymes génétiquement variables et le flux a permis de rendre compte de divers effets génétiques que les approches réductionnistes étaient impuissantes à expliquer : la dominance des allèles actifs sur les moins actifs [Kacser et Burns

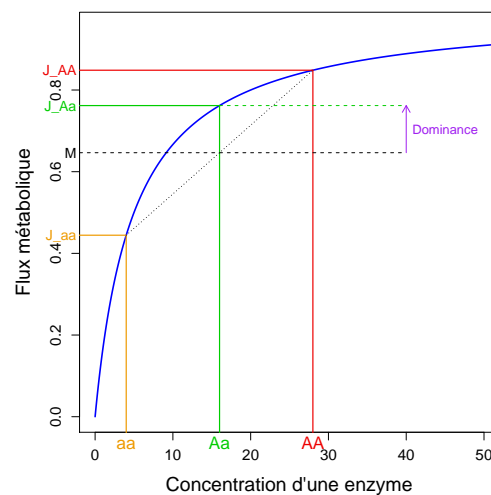


FIGURE 8 – **Modèle de la dominance inspiré de la théorie du contrôle métabolique** (schéma dérivé de Kacser et Burns [1981]). Le flux est fonction de la concentration d'une enzyme  $i$ , selon la même relation que dans la figure 6, avec  $a = 1$  et  $b = 5$ . On considère que la variation de concentration de cette enzyme est monogénique, avec deux allèles  $a$  et  $A$ , et qu'il y a additivité de la concentration chez l'hétérozygote. Le croisement du génotype  $aa$  de flux  $J_{aa}$  avec le génotype  $AA$  de flux  $J_{AA}$  donne le génotype  $Aa$  de flux  $J_{Aa}$ . Le flux du descendant hétérozygote est plus élevé que la moyenne  $M$  des flux des deux parents homozygotes : il y a dominance. Lorsque l'on prend en compte non pas un, mais plusieurs gènes, ce modèle rend compte du phénomène d'hétérosis.

1981] (figure 8), l'hétérosis<sup>6</sup> [Fiévet *et al.* 2010, 2018], la pléiotropie<sup>7</sup> [Keightley et Kacser 1987], l'épistasie [Szathmáry 1993; Keightley 1996] et la distribution en L des effets des QTL (*Quantitative Trait Loci*) [Bost *et al.* 1999].

En biologie évolutive, la TCM a permis de modéliser la dynamique de maintien/perte de gènes métaboliques après duplication complète du génome [Gout *et al.* 2009] et de déterminer mathématiquement la distribution des effets des allèles sous sélection pour un flux optimum [Clark 1991].

C'est dans ce contexte de biologie évolutive que s'inscrit cette thèse, dans la continuité des travaux de Hartl *et al.* [1985]. Pour rendre compte du fait que la plupart des polymorphismes moléculaires sont sélectivement neutres, ces auteurs ont proposé un modèle de « sélection naturelle de la neutralité sélective » qui s'appuie sur la TCM. Sous sélection pour augmenter le flux, considéré comme un proxy de la fitness, les mutations qui affectent les propriétés enzymatiques ont de moins de moins d'impact sur le flux au fur et à mesure que les valeurs des propriétés enzymatiques augmentent, jusqu'à arriver dans la zone de neutralité sélective ou zone neutre (figure 9).

En pleine controverse entre les neutralistes et les sélectionnistes, les travaux de Hartl *et al.* [1985] ont eu un certain retentissement car ils fournissaient une explication systémique à la

<sup>6</sup>Hétérosis = la valeur du descendant est différente de la valeur moyenne de ses parents (en général elle est supérieure).

<sup>7</sup>Pléiotropie = un gène a un effet sur plusieurs caractères.



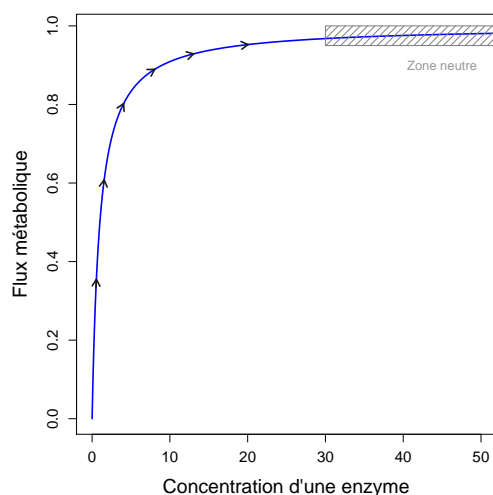


FIGURE 9 – **Modèle de la neutralité sélective inspiré de la théorie du contrôle métabolique** (d’après Hartl *et al.* [1985]). Le flux, considéré proportionnel à la fitness, est fonction de la concentration d’une enzyme  $i$  selon la même relation que dans la figure 6, avec  $a = b = 1$ . Les flèches indiquent le sens de la sélection. La zone hachurée correspond à la zone de neutralité sélective où les mutations sont quasi-neutres, avec  $2N_e s \ll 1$  pour une population haploïde,  $N_e$  étant la taille efficace de population et  $s$  le coefficient de sélection.

neutralité des polymorphismes moléculaires. Ils s’appuyaient toutefois sur deux présupposés biologiquement irréalistes :

1) Les mutations affectent toujours la même enzyme, les autres restant invariables. Comme nous l’avons vu, l’épistasie est inhérente au modèle métabolique, aussi ignorer l’évolution conjointe des enzymes d’un système revient à négliger une propriété essentielle de la relation génotype-phénotype ;

2) Aucune borne supérieure n’est imposée à la concentration de l’enzyme cible des mutations, comme si les ressources cellulaires étaient illimitées.

Par ailleurs Hartl *et al.* [1985] n’ont traité que le cas d’une chaîne métabolique linéaire. Dans des réseaux de topologie quelconque, comment le flux est-il relié aux propriétés enzymatiques, et comment évoluent-elles ?

## 0.4 Objectifs de la thèse

Dans la continuité des travaux de Hartl *et al.* [1985], l’objectif de ma thèse était d’étudier l’évolution des concentrations d’enzyme (le « génotype ») sous sélection pour augmenter le flux (le « phénotype ») en visant un niveau de généralité aussi élevé que possible. Outre la prise en compte de l’évolution conjointe de toutes les enzymes du système, la principale originalité de l’approche a été d’introduire dans les modèles deux contraintes susceptibles d’affecter les possibilités d’évolution des concentrations d’enzymes :

- **La limitation des ressources.** La quantité d'ATP, le nombre de ribosomes, certains éléments, etc., sont en quantité limitée dans la cellule. Une autre ressource moins souvent évoquée est l'espace disponible. La densité d'une cellule est telle qu'il y a du *molecular crowding* : les molécules ne peuvent circuler librement car elles se trouvent dans un espace délimité par la membrane cellulaire, ou par celle des organites, ce qui ralentit la diffusion mais diminue aussi l'activité chimique de ces molécules [Ellis 2001]. Pour tenir compte globalement de l'ensemble de ces contraintes, on considérera que la concentration totale d'enzymes affectée à une chaîne est constante – il y a donc « compétition » entre enzymes pour les ressources cellulaires.
- **Les régulations.** Ce facteur n'est en général pas considéré comme une « contrainte », mais comme un élément essentiel du fonctionnement des organismes, façonné par l'évolution, et que l'on trouve à toutes les échelles du vivant. Les pertes de régulation peuvent être très délétères à toutes les échelles. Par exemple la dérégulation de la mitose peut conduire à l'apparition de cancers ; la mobilisation importante du calcium pour la lactation au détriment de la vache en raison d'hormones hypercalcémiantes insuffisamment produites en amont lié à des erreurs d'alimentation peut conduire à la fièvre vitulaire au vêlage chez les vaches hautes productrices laitières ; chez les furets mâles, la castration chirurgicale empêche la reproduction mais dérégule la production d'hormones sexuelles conduisant à des tumeurs surrénaliennes ; certaines maladies métaboliques (diabète, hypothyroïdie, hyperthyroïdie...) sont liées à des problèmes de régulation hormonale ; la coupure des nerfs sympathique et parasympathique innervant le cœur chez les personnes avec une greffe cardiaque empêche l'adaptation de la vitesse cardiaque pendant les efforts physiques ; la prise d'antibiotique peut conduire à une dérégulation de la flore bactérienne intestinale, laissant le champ libre à des bactéries pathogènes ; etc. Les régulations permettent l'homéostasie cellulaire et rendent les organismes robustes vis-à-vis des variations environnementales. Les régulations génétiques permettent également la robustesse des organismes face aux variations génétiques [Ciliberti *et al.* 2007]. Mais la contrepartie de ces régulations est qu'elles restreignent l'espace des possibles pour des adaptations ultérieures, et à ce titre elles peuvent constituer des contraintes évolutives. Sous le terme « co-régulation entre concentrations d'enzymes », je regroupe la manifestation de tout mécanisme conduisant à la variation conjointe des concentrations de deux ou plusieurs enzymes à la suite d'une mutation.

Le sujet de thèse comportait initialement trois volets. Tout d'abord, je devais généraliser le modèle d'évolution des concentrations enzymes : 1) en considérant que des enzymes indépendantes et co-régulées peuvent co-exister au sein d'un même système ; 2) en passant des chaînes aux réseaux métaboliques ; 3) en ajoutant une fonction de coût de la synthèse protéique. Ensuite il était prévu de modéliser l'évolution des concentrations d'enzymes lorsque la sélection porte sur les teneurs en métabolites et non pas sur les flux. Enfin je devais confronter les résultats

obtenus à ceux issus de modèles plus complexes ainsi qu'à des données réelles.

Il est rapidement apparu que je ne pourrais pas traiter l'ensemble de ces questions. La confrontation aux données pouvait faire l'objet d'une thèse à part entière, de même que la sélection sur les teneurs en métabolites. J'ai également dû faire un choix entre l'extension aux réseaux et l'introduction d'une fonction de coût de la synthèse des protéines. En réalité, la contrainte « compétition pour les ressources » revient à la même question, car dans les deux cas on aboutit à un optimum pour le flux. Ainsi, il me semblait plus pertinent d'étendre en priorité le modèle aux réseaux métaboliques, qui peuvent éventuellement fournir des résultats inattendus. Je me suis donc concentrée sur la partie théorique portant sur la généralisation du modèle d'évolution.

Dans un premier temps, j'ai repris et poursuivi les travaux de Grégoire Talbot (stage de Master 1) et Maud Le Louarn (stage de fin d'études de CentraleSupélec), qui avaient commencé la recherche des équilibres évolutifs des concentrations d'enzymes dans une chaîne linéaire sous sélection pour augmenter le flux, et ce sous quatre conditions différentes (indépendance des enzymes, compétition pour les ressources, co-régulations positives et/ou négatives entre toutes les enzymes de la chaîne et compétition plus co-régulations). J'ai également travaillé sur les zones de neutralité sélective des concentrations d'enzymes sous ces différentes contraintes, ce qui faisait partie du projet originel de l'équipe, et que j'ai souhaité poursuivre. Le modèle d'évolution intègre à la fois des éléments de génétique des populations, dans la prise en compte de la dérive à travers la probabilité de fixation, et des éléments de dynamique adaptative, dans la recherche d'un optimum de fitness en décrivant l'évolution d'un caractère au cours du temps. J'ai également représenté les résultats sous forme de paysages adaptatifs, et effectué des simulations d'évolution à long terme qui ont confirmé les valeurs d'équilibres trouvées analytiquement. Cette partie constitue le chapitre 1 du manuscrit, qui correspond à un article soumis à *Journal of Theoretical Biology*.

Dans le chapitre 2, j'ai adapté le modèle d'évolution au cas plus général où les enzymes d'une même chaîne appartiennent à des groupes de co-régulations différents. Pour trouver les équilibres des concentrations relatives, il a fallu rechercher l'expression des équilibres au sein et entre les groupes de co-régulations. Dans les cas où je n'ai pas trouvé de solution analytique ou numérique, les simulations ont montré que les équilibres existaient. Ce chapitre est rédigé sous la forme d'un article qui est le *companion paper* du premier, qui sera soumis au même journal.

Dans le chapitre 3, j'ai étendu la théorie du contrôle métabolique aux réseaux métaboliques en exploitant l'analogie avec les circuits électriques. J'ai comparé précisément les formalismes, et montré comment, dans ce cadre-là, exprimer le flux en fonction de l'efficacité de toutes les enzymes d'un réseau. Je n'ai cependant pas eu le temps de chercher les équilibres évolutifs dans le cas des réseaux.

Les annexes A à C contiennent les figures supplémentaires et les démonstrations mathématiques pour les chapitres 1 à 3 respectivement.

---

# CHAPITRE 1

---



# Chapitre 1

## Article : *Evolution of enzyme levels in metabolic pathways : A theoretical approach*

Ce premier chapitre présente l'étude de l'évolution des concentrations d'enzyme sous sélection pour augmenter le flux dans une chaîne métabolique linéaire sous quatre contraintes différentes. Il est basé sur un premier article soumis à *Journal of Theoretical Biology*, et dont le manuscrit est disponible sur *BioRxiv* [Coton *et al.* 2021c]. Dans la première section, je reprends succinctement les résultats présentés dans l'article. L'article en lui-même constitue la deuxième section, avec un glossaire des symboles mathématiques utilisés. Notez que la numérotation des sections, équations, figures et tableaux est indépendante du reste du manuscrit. Enfin, dans une dernière section, je présente quelques points de discussion supplémentaires. Les *Supporting Information* de l'article, contenant les figures supplémentaires et les démonstrations, sont dans l'annexe A. Dans cette annexe figurent également quelques démonstrations complémentaires. Les programmes que j'ai écrits sont regroupés dans un package R accessible sur le CRAN (*Comprehensive R Archive Network*) sous le nom `SimEvolEnzCons`, qui permet de faire tourner des simulations d'évolution des concentrations d'enzymes et de calculer les équilibres évolutifs et les zones de neutralité sélective [Coton *et al.* 2021b].

### 1.1 Présentation de l'article

Dans cet article, je présente le modèle d'évolution des concentrations d'enzyme sous sélection pour augmenter le flux. Dans ce modèle, les mutations modifient les concentrations d'enzymes. Du fait des contraintes, une mutation ciblant la concentration d'une enzyme peut modifier la concentration des autres enzymes de la chaîne métabolique. En utilisant un modèle de dynamique adaptative, l'évolution des concentrations d'enzymes peut s'écrire sous la forme d'un système d'équations différentielles. Ce système décrit l'espérance de la variation de la concentration de chaque enzyme au cours du temps, qui est fonction du taux de mutation, de la probabilité de fixation de la mutation et de l'effet de la mutation sur l'enzyme considérée selon les

contraintes appliquées. La probabilité de fixation est directement liée au coefficient de sélection, qui dépend de l'impact de la variation de l'enzyme sur le flux.

Dans le cas d'une chaîne métabolique, nous avons étudié quatre conditions différentes, qui sont l'indépendance entre toutes les enzymes, la compétition pour les ressources, les co-régulations entre toutes les enzymes, et la combinaison de la compétition et des co-régulations.

Pour chacune de ces conditions, nous avons recherché les équilibres évolutifs, c'est-à-dire les concentrations relatives pour lesquelles le système est stable au cours du temps. En parallèle, j'ai étudié l'évolution du flux au cours du temps et les *Range of Neutral Variations* (RNV), c'est-à-dire les zones de concentrations autour de l'équilibre où les mutations sont quasiment neutres. Le recours à la notion de paysage adaptatif a permis de se représenter la relation enzymes-flux et les équilibres de manière moins abstraite.

Dans le cas où toutes les enzymes sont indépendantes, les concentrations relatives à l'équilibre sont inversement proportionnelles aux pseudo-activités, qui regroupent les paramètres cinétiques des enzymes et les constantes d'équilibre des réactions en amont de l'enzyme concernée. Les enzymes avec les plus fortes pseudo-activités ont les plus grandes RNV. Quant au flux, il peut augmenter indéfiniment.

Si les enzymes sont soumises à la compétition, modélisée par une concentration totale d'enzyme fixe, les concentrations relatives à l'équilibre sont également inversement proportionnelles aux pseudo-activités. Le flux est cependant limité, et le paysage adaptatif forme un dôme dont le sommet correspond à l'équilibre évolutif. Quant aux RNV, ce sont les enzymes avec la plus faible pseudo-activité qui ont les RNV les plus grandes à l'équilibre, en raison de la courbure du dôme.

En cas de co-régulation entre toutes les enzymes, les concentrations sont linéairement reliées et évoluent le long d'une droite dans l'espace des concentrations. Si les co-régulations sont positives, les concentrations relatives à l'équilibre sont inversement proportionnelles aux coefficients de co-régulation globaux, qui représentent l'effet de la variation de concentration d'une enzyme sur la concentration totale. Dans ce cas le flux augmente indéfiniment. S'il existe au moins une co-régulation négative, le flux est borné à cause de l'effet de balance entre les concentrations d'enzymes. Les concentrations relatives à l'équilibre sont alors dépendantes non seulement des coefficients de co-régulation mais aussi des pseudo-activités et des concentrations initiales. Quelles que soient les co-régulations, les RNV sont inversement proportionnelles aux coefficients de co-régulations globaux : plus une enzyme a un fort impact sur la concentration totale et donc sur le flux, plus sa zone neutre sera petite.

Lorsqu'il y a à la fois compétition et co-régulations entre toutes les enzymes, les concentrations sont linéairement reliées entre elles à cause des co-régulations et le flux est limité à cause de la compétition. Le paysage adaptatif est alors une section du dôme. L'équilibre évolutif correspond au flux maximum et dépend des pseudo-activités, des coefficients de co-régulations et des concentrations initiales. La RNV ne dépend alors que des coefficients de co-régulations globaux et des concentrations initiales : plus une enzyme est initialement éloignée de son équilibre,

plus sa zone neutre sera grande.

Toutes les résultats analytiques sont confirmées par des simulations d'évolution à long terme, où les mutations délétères ne sont pas négligées.

Cette étude permet de mettre en avant l'effet des contraintes sur le paysage adaptatif, conditionnant les équilibres évolutifs et la taille des zones neutres.





# Evolution of enzyme levels in metabolic pathways: A theoretical approach

Charlotte Coton<sup>†</sup>, Grégoire Talbot<sup>†</sup>, Maud Le Louarn,<sup>\*</sup>  
Christine Dillmann<sup>†</sup>, Dominique de Vienne<sup>†</sup>

May 4, 2021

5     <sup>†</sup> Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France.

## Abstract

The central role of metabolism in cell functioning and adaptation has given rise to countless studies on the evolution of enzyme-coding genes and network topology. However, very few studies have addressed the question of how enzyme concentrations change in response to positive selective pressure on the flux, considered a proxy of fitness. In particular, the way cellular constraints, such as resource limitations and co-regulation, affect the adaptive landscape of a pathway under selection has never been analyzed theoretically. To fill this gap, we developed a model of the evolution of enzyme concentrations that combines metabolic control theory and an adaptive dynamics approach, and integrates possible dependencies between enzyme concentrations. We determined the evolutionary equilibria of enzyme concentrations and their range of neutral variation, and showed that they differ with the properties of the enzymes, the constraints applied to the system and the initial enzyme concentrations. Simulations of long-term evolution confirmed all analytical and numerical predictions, even though we relaxed the simplifying assumptions used in the analytical treatment.

## Keywords

Adaptive landscape. Metabolism. Enzyme concentration. Co-regulation. Evolutionary equilibrium. Selective neutrality.

## 25 Running title

Evolution of enzyme concentrations

---

\*This article is dedicated to the memory of Maud Le Louarn, who contributed so inventively and enthusiastically to this work

# 1 Introduction

Metabolism is a property of living systems that is central to cell functioning and physiological and genetic adaptation (Cornish-Bowden and Cárdenas 2020). The metabolic networks we see today, comprising an extraordinary diversity of enzymes, cofactors and metabolites, are the product of billion years of evolution that have shaped network topology and enzyme properties.

The importance of metabolism has inspired a wealth of studies on its evolutionary dynamics over different timescales, which follow three main lines of research.

The first line of research relates to the evolution of network structure. Over evolutionary time, continuous selective pressures in diverse environments have resulted in the successive recruitment of new metabolites and enzymes that then coevolve (Noda-Garcia et al. 2018), have modified substrate specificity (Newton et al. 2018) and have determined the general features of the networks, namely functional redundancy, modularity, flux coupling and exaptation (Sambamoorthy et al. 2019). Adaptation to a new environment may involve gain/loss-of-functions and horizontal gene transfers, which rewire the metabolic networks (e.g. Pál et al. 2005; Lee and Palsson 2010; D’Souza and Kost 2016; Morrison and Badyaev 2017).

The second line of research investigates the factors that determine the degree of polymorphism and the rate of evolution of enzyme-coding genes. In addition to classical factors, such as gene dispensability, level of gene expression, protein length, protein folding stability, intron number and protein-protein interaction (see e.g. review by Larracuenté et al. 2008), various systemic constraints also influence the degree of protein polymorphism and evolution (Pál et al. 2006). Enzymes at branching points of the network or that are highly connected, and thus have pleiotropic effects, are assumed to be more subject to purifying selection; as a consequence, their coding genes have low levels of polymorphism or slow evolutionary rates (Rausher et al. 1999; Vitkup et al. 2006; Greenberg et al. 2008; Rausher 2013). Flux levels and flux control coefficients are additional factors that influence protein evolution: the genes encoding high-flux enzymes have lower evolutionary rates (Colombo et al. 2014), possibly related to enzyme abundance (Aguilar-Rodríguez and Wagner 2018); higher flux control coefficients of enzymes that are upstream in the pathway result in stronger purifying or positive selection (Flowers et al. 2007; Wright and Rausher 2010; Sellis and Longo 2015).

The third line of research is the study of the genetic variability of enzyme properties. The response of organismal fitness to selection pressures depends on genetically variable enzyme properties, namely kinetic parameters, thermodynamic stability and concentration. In the context of the historical debate between neutralism and selectionism, a number of emblematic examples of variation in enzyme properties have been studied in detail. For instance, in *Drosophila*, the fast and slow allozymes of alcohol dehydrogenase differ both in the value of their catalytic constant (the turnover number,  $k_{cat}$ ) and their abundance; also in *Drosophila* the A and B variants of G6PD have different quaternary structures, thermal stability and  $K_m$  values; in *Fundulus heteroclitus*, the allozymes of LDH-B display kinetic, abundance and stability differences, etc. (reviewed in Eanes 1999). Furthermore, quantitative proteomic studies in various species have revealed the breadth of the genetic variability, the polygenic nature and the relationship with integrated phenotypic traits of enzyme concentrations (Damerval et al. 1994; Blein-Nicolas et al. 2013; Albert and Kruglyak 2015; Chick et al. 2016; Jiang et al. 2019).

In this context, the evolution of enzyme properties in response to selective pressures on the flux, considered to be a proxy for fitness, is an important issue. This question

75 has been addressed in relation to protein folding stability, catalytic constant values and  
enzyme concentration, but to our knowledge the number of studies is quite low. Mutations  
that increase protein folding stability are favored by selection, with evolution towards near  
neutrality up to the mutation–selection balance steady state (Serohijos and Shakhnovich  
2014; Bershtein et al. 2017). Regarding enzyme activity, Kacser and Beeby (1984), using a  
80 systems biology approach, proposed a model of early enzyme evolution in which selection  
for faster growth causes a transition from few, multifunctional, poorly active enzymes to  
many, differentiated, monofunctional enzymes with high  $k_{\text{cat}}$  values. However, Heckmann  
et al. (2018) recently showed that the systemic properties of a network result in epistatic  
interactions that maintain  $k_{\text{cat}}$  values far from their theoretical maximum, preventing  
85 fitness from reaching its optimum. This result can be explained by the relationship of  
diminishing returns between turnover number and flux: the effect of a given increase in  
 $k_{\text{cat}}$  on the flux vanishes as  $k_{\text{cat}}$  increases, so that the flux eventually reaches a plateau.

Interestingly, the same type of saturation curve was invoked by Hartl et al. (1985)  
to account for the widespread selective neutrality of enzyme polymorphisms. These au-  
90 thors analyzed the evolution of enzyme quantity, equated to activity, under selection for  
higher flux, and showed that successive mutations affecting a particular enzyme drive  
the enzyme concentration towards the plateau of the concave flux-enzyme curve, where  
the mutations do not have any detectable effect, *i.e.* they are selectively nearly neutral.  
This geometrical model of the "natural selection of selective neutrality" is attractive, but  
95 surprisingly it has not been assessed with an adaptive dynamics approach. In addition,  
this model considered only a single enzyme in a given pathway and imposed no upper  
limits on enzyme quantities, even though there are well-known constraints on space and  
energy in the cell (Koehn 1991; Kurland and Dong 1996; de Vienne et al. 2001b; Ellis  
2001; Eguchi et al. 2018; Klumpp et al. 2019). Although these points were raised, the  
100 evolutionary implications were not analyzed by Hartl et al. (1985). The question of the  
optimization of enzyme kinetic parameters and concentrations in metabolic pathways was  
analyzed by Klipp and Heinrich (1994, 1999). However, their approach did not include  
any evolutionary genetics concept and did not allow them to grasp the dynamics of the  
evolutionary process.

105 Here, we develop the first adaptive dynamics model of the evolution of enzyme con-  
centrations when there is directional selection for increased flux in a pathway. Our model  
includes two biologically realistic features: (i) the concentration of each enzyme in the  
pathway can be altered by mutation, and (ii) constraints, specifically competition for  
cellular resources and/or co-regulation, can prevent enzyme concentrations from varying  
110 freely. To formalize the relationship between enzyme concentrations and flux, we used  
the metabolic control theory (MCT) framework (Kacser and Burns 1973; Heinrich and  
Rapoport 1974; Kacser et al. 1995), as generalized by Heinrich et al. (1991) and Lion  
et al. (2004) to take into account the possible interdependence of enzyme concentrations.

115 First, we present a system of ordinary differential equations that describes the evolu-  
tion of enzyme concentrations in a pathway. We then show how the flux evolves and how it  
determines the evolutionary equilibria of the relative enzyme concentrations. We demon-  
strate analytically, and validate using computer simulations, that a theoretical equilibrium  
of relative enzyme concentrations exists whatever the conditions, but that this equilibrium  
is not necessarily reached when the flux is limited by constraints on enzyme concentration  
120 variation. Thus, we differentiate the *effective* equilibrium, *i.e.* the equilibrium that is  
actually reached, from the *theoretical* equilibrium. When there is co-regulation in addi-  
tion to competition, the initial enzyme concentrations affect the evolutionary outcome.  
Finally, we show that multi-enzyme selective neutrality, *i.e.* the selective neutrality of all

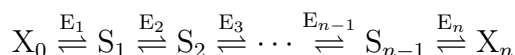
enzymes, is achieved whatever the constraints, but that the size of the range of neutral  
 125 variation largely depends on constraint-specific parameters and possibly on the initial  
 state of the system.

## 2 Material and Methods

### 2.1 Theoretical developments

#### 2.1.1 The metabolic model

130 **2.1.1.1 Flux expression** We considered a metabolic pathway of  $n$  unsaturated Michaelian  
 enzymes transforming substrate  $X_0$  into product  $X_n$  via a series of unimolecular reversible  
 reactions:



At steady state, the metabolic flux  $J$  through this pathway is (Kacser and Burns  
 1981):

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}} \quad (1)$$

135 where:

- $X = X_0 - X_n/K_{1,n}$ .  $K_{1,n}$  is the product of the successive equilibrium constants of  
 all reactions.  $X$  is assumed to be constant and symbolizes that the environment is  
 constant.
- $A_j = \frac{k_{catj}}{K_{Mj}} K_{1,j-1}$ , where  $k_{catj}$  is the turnover number of enzyme  $j$ ,  $K_{Mj}$  is the  
 140 Michaelis-Menten constant of the enzyme  $j$  for its substrate and  $K_{1,j-1}$  is the prod-  
 uct of the successive equilibrium constants up to the  $(j-1)^{th}$  reaction:  $\prod_{k=1}^{j-1} K_k$ .  
 Thus  $A_j$  is a composite parameter that includes both catalytic efficiency  $\frac{k_{catj}}{K_{Mj}}$  and a  
 product of thermodynamic constants that depends on the enzyme's position in the  
 pathway. Hereafter,  $A_j$  is referred to as the "pseudo-activity" of enzyme  $j$ .
- $E_j$  is the concentration of enzyme  $j$ . Note that  $E_j$  in italics denotes the concen-  
 145 tration of enzyme  $j$  while  $E_j$  in Roman letter denotes enzyme  $j$  (see Glossary of  
 mathematical symbols in Appendix A).

Considering the total enzyme concentration,  $E_{tot} = \sum_{j=1}^n E_j$ , and the relative concentra-  
 tion of enzyme  $j$ ,  $e_j = \frac{E_j}{E_{tot}}$ , equation 1 becomes:

$$J = E_{tot} \frac{X}{\sum_{j=1}^n \frac{1}{A_j e_j}} \quad (2)$$

150 The flux can be modified by mutations affecting  $\frac{k_{catj}}{K_{Mj}}$  or  $E_j$ . We considered only the latter,  
 because the variability of enzyme concentrations is by far the main source of variation in  
 enzyme activity in populations and species (see *Discussion*).

### 2.1.1.2 The effects of mutation when enzyme concentrations are interdependent

Let a mutation of small effect  $\nu$  targeting concentration  $E_i$  of enzyme  $i$  ( $\nu$  is positive or negative). We assume that the probability law of mutation effect  $\nu$  is identical for all enzymes, so  $\nu$  is not indexed. The concentration of enzyme  $i$  after a mutation is:

$$E_i^{(1)} = E_i + \nu \quad (3)$$

Because of constraints such as resource limitations or complex interactions between cell components, this change may not correspond to the actual effect of the mutation on the concentration of enzyme  $i$  and/or it may modify the concentration of other enzymes. Considering two causes of interdependence of enzyme concentrations, separately or jointly, we formalized the consequences of a mutation in the following way:

1. *Co-regulations.* Various direct and indirect molecular mechanisms result in co-regulation (either positive or negative) of enzyme concentrations. Thus, mutations can jointly modify the concentrations of co-regulated enzymes (*see Discussion*). To express formally the effect of a mutation targeting the concentration of enzyme  $i$  (equation 3) on the concentration of enzyme  $j$ , we write:

$$E_j^{(1)} = E_j + \beta_{ij}\nu \quad (4)$$

where  $\beta_{ij}$ , the slope of the relation between  $E_i$  and  $E_j$ , is the *co-regulation coefficient*. We assume linearity between co-regulated enzymes, so we use constant co-regulation coefficients and we pose  $\beta_{ii} = 1$ ,  $\beta_{ij} = 1/\beta_{ji}$  and, for any triplet of co-regulated enzymes  $(i, j, k)$ ,  $\beta_{ij} = \beta_{ik}\beta_{kj}$ . If enzymes  $i$  and  $j$  are not co-regulated, we have  $\beta_{ij} = \beta_{ji} = 0$ . All pairwise co-regulation coefficients can be grouped in a co-regulation matrix, noted  $\mathbf{M}_\beta$  (Supporting Information [SI.B.1.1](#)).

For  $n$  co-regulated enzymes, the total enzyme concentration after a mutation affecting enzyme  $i$  is:

$$\begin{aligned} E_{\text{tot}}^{(1)} &= \sum_{j=1}^n E_j^{(1)} = \sum_{j=1}^n (E_j + \beta_{ij}\nu) \\ &= E_{\text{tot}} + B_i\nu \end{aligned} \quad (5)$$

where  $B_i = \sum_{j=1}^n \beta_{ij}$  quantifies the impact on  $E_{\text{tot}}$  of a mutation targeting  $E_i$ . Indeed, from equations 3 and 5, we have:

$$\frac{\Delta E_{\text{tot}}}{\Delta E_i} = \frac{E_{\text{tot}}^{(1)} - E_{\text{tot}}}{E_i^{(1)} - E_i} = B_i \quad (6)$$

$B_i$  is called the *global co-regulation coefficient* (Supplementary information [SI.B.1.2](#)).

2. *Competition for resources.* Cellular resources are limited in terms of space, energy, the availability of ribosomes, etc. The total enzyme concentration a cell allocates to a given pathway is necessarily bounded, so that an increase (resp. decrease) in the concentration of an enzyme due to a mutation can result in a decrease (resp. increase) in the concentration of other enzymes (Snoep et al. 1995; Albertin et al. 2013). We modeled this trade-off by fixing  $E_{\text{tot}}$ . As a consequence, if a mutation of effect  $\nu$  targets enzyme  $i$ , the concentrations of all the other enzymes will be

modified, as will be the effect of the mutation on enzyme  $i$ . This can be viewed as a two-step process:

$$E_i \xrightarrow[\text{mutation of } E_i]{} E_i^{(1)} = E_i + \nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_i^{(2)} = E_i + \delta_i$$

185 where  $\delta_i$  is the mutation effect after applying the constraint on total enzyme concentration, which changes from  $E_{\text{tot}}^{(1)} = E_{\text{tot}} + \nu$  to  $E_{\text{tot}}$ . This constraint is assumed to respect the proportionality between concentrations, *i.e.*

$$\frac{E_i^{(1)}}{E_{\text{tot}}^{(1)}} = \frac{E_i^{(2)}}{E_{\text{tot}}} \quad (7)$$

From this relation, the expression of  $\delta_i$  can be deduced (Supporting Information [SI.B.1.3](#)):

$$\delta_i = \frac{1 - e_i}{\frac{1}{\nu} + \frac{1}{E_{\text{tot}}}} \quad (8)$$

190 where  $e_i = \frac{E_i}{E_{\text{tot}}}$ . We call  $\nu$  the *canonical effect* of the mutation and  $\delta_i$  its *actual effect*.

For enzymes  $j \neq i$ , we have

$$E_j \xrightarrow[\text{mutation of } E_i]{} E_j^{(1)} = E_j \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_j^{(2)}$$

where  $E_j^{(2)}$  is the concentration of enzyme  $j$  after applying the constraint on total enzyme concentration. To find the expression of  $E_j^{(2)}$ , we applied the proportionality rule:

$$\frac{E_j^{(1)}}{E_{\text{tot}}^{(1)}} = \frac{E_j^{(2)}}{E_{\text{tot}}} \quad (9)$$

195 and using equation 8 we get (Supporting Information [SI.B.1.3](#)):

$$E_j^{(2)} = E_j - \frac{e_j}{1 - e_i} \delta_i \quad (10)$$

3. *Co-regulation and competition.* If there is both competition and co-regulation between all enzymes, the two-step process is the same as above for enzyme  $i$  targeted by a mutation:

$$E_i \xrightarrow[\text{mutation of } E_i]{} E_i^{(1)} = E_i + \nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_i^{(2)} = E_i + \delta_i$$

but here  $E_{\text{tot}}^{(1)} = E_{\text{tot}} + B_i \nu$  (equation 5). So applying the proportionality argument (equation 7 and Supporting Information [SI.B.1.4](#)), we get:

$$\delta_i = \frac{1 - e_i B_i}{\frac{1}{\nu} + \frac{B_i}{E_{\text{tot}}}} \quad (11)$$

For enzymes  $j \neq i$ , we have

$$E_j \xrightarrow[\text{mutation of } E_i]{} E_j^{(1)} = E_j + \beta_{ij}\nu \xrightarrow[\text{constraint on } E_{\text{tot}}]{} E_j^{(2)}$$

where  $E_j^{(2)}$  is the concentration of enzyme  $j$  after co-regulation *and* reduction of the total enzyme concentration from  $E_{\text{tot}}^{(1)}$  to  $E_{\text{tot}}$ . Using equations 9 and 11, we get (Supporting Information SI.B.1.4):

200

$$E_j^{(2)} = E_j + \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i} \delta_i \quad (12)$$

In summary, non-independence of enzyme concentrations results in changes in the concentration of enzymes that are not the primary targets of the mutation. The amplitude of concentration change of non-target enzymes depends on the cause of interdependence:

Co-regulation (equation 4)	$E_j^{(1)} = E_j + \beta_{ij}\delta_i$ where $\delta_i = \nu$
Competition (equation 10)	$E_j^{(2)} = E_j - \frac{e_j}{1 - e_i} \delta_i$
Co-regulation and competition (equation 12)	$E_j^{(2)} = E_j + \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i} \delta_i$

Thus the concentration of the mutant enzyme  $E_j^m$  can be expressed as the sum of  $E_j$  and a *pseudo-mutation* effect:

$$\forall j, \quad E_j^m = E_j + \alpha_{ij}\delta_i$$

This *pseudo-mutation* effect is the product of the actual mutation effect  $\delta_i$  by a coefficient called the *redistribution coefficient*, noted  $\alpha_{ij}$  (note that  $\forall i, \alpha_{ii} = 1$ ).

205

The general expression of the redistribution coefficient is  $\alpha_{ij} = \frac{\Delta E_j}{\Delta E_i}$ , where  $\Delta E_j$  is the change of non-target enzyme  $j$  and  $\Delta E_i = \delta_i$  is the actual effect of the mutation on target enzyme  $i$ . Going to the limit when  $\delta_i$  is very small, we obtain:

$$\alpha_{ij} = \frac{\partial E_j}{\partial E_i} \quad (13)$$

which is the redistribution coefficient  $\alpha_{ij}$  defined by Lion et al. (2004).

Depending on the situations, we have:

210

1. Independence:  $\forall i, j \neq i, \alpha_{ij} = 0$
2. Co-regulation:  $\forall i, j \neq i, \alpha_{ij} = \beta_{ij} \neq 0$
3. Competition:  $\forall i, j \neq i, \alpha_{ij} = \frac{-e_j}{1 - e_i}$
4. Co-regulation and competition:  $\forall i, j \neq i, \alpha_{ij} = \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i}$



### 2.1.1.3 Quantifying the effect of enzyme concentration variation on the flux

215 To quantify the effect on the flux of variation in enzyme concentration  $E_i$ , we used the *combined response coefficient* (hereafter, the flux response coefficient) (de Vienne et al. 2001a)

$$R_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J} \quad (14)$$

The flux response coefficient is more general than the classical flux control coefficient (Kacser et al. 1995), the latter being valid only if enzymes are independent. The flux response coefficient of enzyme  $i$  includes both the flux control coefficient  $C_{E_i}^J$  of enzyme  $i$  and a term that accounts for enzyme interdependence, which contains the flux control coefficients of the other enzymes and the redistribution coefficients  $\alpha_{ij}$  (Lion et al. 2004):

$$R_{E_i}^J = C_{E_i}^J + E_i \sum_{j=1}^n C_{E_j}^J \frac{\alpha_{ij}}{E_j}$$

In Supporting Information SI.B.2, we present another expression of  $R_{E_i}^J$  that is less intuitive but more convenient for mathematical treatment.

## 2.1.2 The evolution model

220 **2.1.2.1 Adaptive dynamics** To analyze the evolution of enzyme concentrations, we used an adaptive dynamics approach (Brännström et al. 2013), assuming a constant environment.

We assumed that evolution proceeds by the successive fixation of mutations modifying enzyme concentration. We applied the mutual exclusion principle, *i.e.* two phenotypes cannot coexist because the mutations are assumed to be instantaneously fixed or eliminated (Dieckmann and Law 1996). On an evolutionary timescale, if a mutation produces a novel phenotype, either it is fixed, becoming the new resident, or it is eliminated. The vector of enzyme concentrations of the resident is noted  $\mathbf{E}^r = (E_1^r, \dots, E_n^r)$  and that of the mutant  $\mathbf{E}^m = (E_1^m, \dots, E_n^m)$ .

Mutations appear at rate  $\mu$  per time unit. They are independent and each mutation targets only one enzyme. The actual effect of a mutation on the target enzyme  $i$  is  $\delta_i$ . Target enzyme concentration after a mutation is:

$$E_i^m = E_i^r + \delta_i$$

The concentration of each of the other enzymes is:

$$E_j^m = E_j^r + \alpha_{ij}\delta_i$$

235 The selection coefficient of a mutation targeting enzyme  $i$  is:

$$s_i = \frac{w_i^m - w^r}{w^r}$$

where  $w$  is fitness. We assumed that fitness  $w$  is proportional to flux  $J$ , so we have:

$$s_i = \frac{J_i^m - J^r}{J^r} \quad (15)$$

The phenotype under selection, namely the flux, is calculated following equation 1.

Since we considered a haploid population of constant effective size  $N$  with non-overlapping generations, the fixation probability of a mutation is (Crow and Kimura 1970):

$$P_f \approx \frac{2s_i}{1 - e^{-2Ns_i}} \quad (16)$$

when  $s_i$  is small. This expression was used in our computer simulations.

In our analytical developments, we neglected deleterious mutations because their negative selection coefficients result in negligible fixation probability. Since we assumed large population sizes (for all  $i$ ,  $Ns_i \gg 1$ ), we could use the approximation:

$$P_f \approx 2s_i$$

The higher the mutant flux  $J^m$  is relative to the resident flux  $J^r$ , the higher the fixation probability of the mutation. After fixation or loss, a new mutation appears, and the process is repeated.

**2.1.2.2 The differential equation system of enzyme evolution** On a continuous timescale, the variation in concentration of enzyme  $i$  can be represented as a stochastic process. At each time  $t$ , this variation is a random variable that can be described with a complete event system:

$$\left(\frac{\partial E_i}{\partial t}\right)^{(t)} = \begin{cases} \mu 2s_i^{(t)} & \delta_i^{(t)} & \text{Fixation of mutation targeting } i \\ \mu 2s_j^{(t)} & \alpha_{ji}^{(t)} \delta_j^{(t)} \quad \forall j \neq i & \text{Fixation of mutation targeting } j \\ 1 - \sum_{k=1}^n \mu 2s_k^{(t)} & 0 & \text{No mutation or no fixation} \end{cases} \quad (17)$$

where  $\mu$  is the probability of occurrence of a mutation, assumed to be independent of the target enzyme. Because this relation is valid at each time  $t$ , we will not use the superscript  $t$  thereafter. At each time, variation in  $E_i$  can be caused by the fixation of either a mutation affecting enzyme  $i$  (probability  $2s_i$ ) or a mutation affecting any other enzyme (probability  $2s_j$ ), with mutation effect  $\delta_i$  and  $\alpha_{ji}\delta_j$ , respectively. If none of these events happen (probability  $1 - \sum_{k=1}^n \mu 2s_k$ ),  $E_i$  does not change. Therefore, the average variation in concentration of enzyme  $i$  at each time  $t$  is:

$$\left(\frac{\partial E_i}{\partial t}\right) = \mu 2s_i \delta_i + \sum_{\substack{j=1 \\ j \neq i}}^n \mu 2s_j \alpha_{ji} \delta_j \quad (18)$$

This relation is valid at all times  $t$ , and because  $\alpha_{ii} = 1$ , we can write:

$$\frac{\partial E_i}{\partial t} = 2\mu \sum_{j=1}^n s_j \alpha_{ji} \delta_j \quad (19)$$

This differential system allowed us to search for steady states – called evolutionary equilibria – of enzyme concentrations when there is selection for increased flux. When enzymes are independent,  $\partial E_i / \partial t$  is strictly positive, and thus enzyme concentrations increase indefinitely and do not reach equilibrium. Consequently, we searched for evolutionary equilibria of relative enzyme concentrations, which is given by:

$$\forall i, \quad \frac{\partial e_i}{\partial t} = 0 \quad (20)$$

265 **2.1.2.3 Selective neutrality of enzyme concentration variation** We addressed the question of the neutrality of the variations in enzyme concentrations near evolutionary equilibrium. The "neutral zone" is the area where the absolute value of the selection coefficient is close to zero – classically under a threshold of  $1/N$  for haploid populations (Kimura 1985).

270 To analytically study the way constraint-dependent factors influence the neutral zone at evolutionary equilibrium, we used an expression of  $s_i$  that explicitly contains the three relevant variables, *i.e.*  $R_{E_i}^J$ ,  $\delta_i$  and  $E_i$ . Assuming that the mutation effect  $\delta_i$  is small compared to the concentration  $E_i$  ( $\delta_i \ll E_i$ ), we can write  $\frac{\Delta J}{\Delta E_i} \approx \frac{\partial J}{\partial E_i}$ . Therefore, from equation 14, the selection coefficient is given by (Supporting Information SI.B.3.1):

$$s_i \approx R_{E_i}^J \frac{\delta_i}{E_i} \quad (21)$$

The size of the neutral zone determines what we call the *range of neutral variation* (RNV) of the concentration, the amplitude of which is:

$$RNV = |\delta_i^{\text{sup}} - \delta_i^{\text{inf}}|,$$

275 where  $\delta_i^{\text{sup}}$  (resp.  $\delta_i^{\text{inf}}$ ) is the actual effect of a mutation such that  $s_i = +\frac{1}{N}$  (resp.  $s_i = -\frac{1}{N}$ ),  $s_i$  being the superior (resp. inferior) limit of the neutral zone. To determine the  $\delta_i$ 's, we used the initial equation  $s_i = \frac{J_i^m - J^r}{J^r}$  (equation 15), where  $J^r$  is the resident phenotype and  $J_i^m$  the mutant phenotype following a mutation targeting enzyme  $i$ . From equation 1, and after transformations, we obtain (Supporting Information SI.B.3.2):

$$\sum_{j=1}^n \frac{1}{A_j(E_j + \alpha_{ij}\delta_i)} - \frac{1}{s_i + 1} \sum_{j=1}^n \frac{1}{A_j E_j} = 0 \quad (22)$$

280 The solutions of this equation for  $s_i = +1/N$  and  $s_i = -1/N$  give  $\delta_i^{\text{sup}}$  and  $\delta_i^{\text{inf}}$ , respectively. Due to the complexity of the resulting equations, we wrote a script that numerically computes the limits  $\delta_i^{\text{sup}}$  and  $\delta_i^{\text{inf}}$  of the RNVs for different scenarios of enzyme interdependence (function `uniroot` of the R Stats package [(R Development Core Team 2008)]).

## 285 2.2 Computer simulations

To assess the validity of the mathematical predictions, we performed Markov Chain Monte-Carlo simulations using the R software (R Development Core Team 2008). We considered haploid populations of size  $N = 1\,000$  evolving in a constant environment ( $X = 1$ ). At a given time step, all individuals in the population are identical, *i.e.* have the same phenotype  $J$  and the same enzyme concentrations (the "genotype")  $\mathbf{E} = (E_1, E_2, \dots, E_n)$  and the same pseudo-activities  $\mathbf{A} = (A_1, A_2, \dots, A_n)$ . Simulations were performed for a three-enzyme metabolic pathway ( $n = 3$ ), with invariable pseudo-activities  $A_1 = 1$ ,  $A_2 = 10$  and  $A_3 = 30$  and an initial total enzyme concentration  $E_{\text{tot}}^0 = 100$  in arbitrary units (see Figure 1A).

295 The vector of initial concentrations  $\mathbf{E}^0$  was changed for each simulation. Initial concentrations  $E_i^0$  were drawn in  $\mathcal{U}(0, 100)$ , and redistributed proportionally to have  $E_{\text{tot}}^0 = 100$ .  $E_{\text{tot}}$  varied over time in the case of independence and co-regulation, but was fixed at  $E_{\text{tot}} = 100$  in the case of competition.

Co-regulation coefficient values were  $\beta_{12} = 0.1$ ,  $\beta_{23} = 2$  and hence  $\beta_{13} = 0.2$  for  
 300 the positive co-regulations and  $\beta_{12} = 0.32$ ,  $\beta_{23} = -0.43$  and hence  $\beta_{13} \approx -0.138$  for  
 the negative co-regulations. The resulting co-regulation matrices  $\mathbf{M}_\beta$  were computed  
 (Supporting Information [SI.B.1.1](#)).

At each time step, a mutation randomly affected one of the enzymes. The mutation  
 effect  $\nu$  was drawn in  $\mathcal{U}(0, 1)$ ; the sign of  $\nu$  was randomly drawn (1:1 ratio). Thus the  
 305 concentration of target enzyme  $i$  changed from  $E_i^r$  to  $E_i^m = E_i^r \pm \nu$ ; the concentration  
 of all enzymes varied according to the type of dependence considered (see the two-steps  
 process described in paragraph [2.1.1.2](#)). If a concentration became negative, it was set to  
 zero. Finally, we computed the values of the flux, the selection coefficient and the fixation  
 probability of the mutant (equations [1](#), [15](#) and [16](#), respectively).

To determine whether a mutation was fixed or lost, we drew a number in  $\mathcal{U}(0, 1)$ , and  
 if this number was smaller than  $P_f$  the mutation was fixed, otherwise it was lost. If the  
 mutation was fixed, the mutant became the new resident, *i.e.* all individuals displayed the  
 same genotype and phenotype. If the mutation was lost, genotype and phenotype were  
 unchanged. We then proceeded to the next time step. Thus, a time step is the interval  
 310 between two successive mutations. Populations evolved over 125,000 time steps, and the  
 system values were recorded every 250 steps.

These simulations allowed us to relax some of the assumptions of the mathematical  
 analysis, namely small mutation effects, large population size and the absence of fixation  
 of deleterious mutations.

320 **Data and code availability** Scripts of the simulations and custom functions are writ-  
 ten in the R language ([R Development Core Team 2008](#)). Along with the simulation data,  
 they are included in a home-made package called `SimEvolEnzCons`.

### 3 Results

We have developed a theoretical framework that describes the evolution of enzyme con-  
 325 centrations in a pathway under selection for increased metabolic flux. We considered  
 that: (i) the flux can be modified by mutations targeting any enzyme in the pathway;  
 (ii) variation in enzyme concentrations can be not independent as a result of competition  
 for cellular resources and/or co-regulation. We based our developments on the principles  
 of adaptive dynamics ([Brännström et al. 2013](#)), using the enzyme-flux relationship of the  
 330 Metabolic Control Theory (MCT) ([Kacser and Burns 1973](#); [Heinrich and Rapoport 1974](#);  
[Kacser et al. 1995](#)) expanded by [Heinrich et al. \(1991\)](#) and [Lion et al. \(2004\)](#) to take into  
 account possible dependencies between enzyme concentrations.

Flux  $J$  through a linear metabolic pathway of  $n$  Michaelian enzymes was expressed as  
 a function of the parameters of all enzymes (equation [1](#) in Material & Methods):

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}}$$

where  $A_j$  is the *pseudo-activity* of enzyme  $j$ , assumed to be constant, and  $E_j$  is the  
 concentration of enzyme  $j$  that can be affected by mutation (see Glossary of mathematical  
 335 symbols in appendix [A](#)).

If there is selection for increased flux, the evolutionary rate of enzyme concentrations  
 can be expressed by a differential equation system (equation [19](#)), allowing us to: (i) search

for steady states of relative enzyme concentrations  $e_i$  ( $\forall i, \partial e_i / \partial t = 0$ ), which we call *evolutionary equilibria*; (ii) analyze flux variation during the evolutionary process.

340 To assess the validity of our analytical results, we performed simulations of long-term evolution using less stringent assumptions than in the analytical approach.

Finally, we analyzed the extent to which selection for increased metabolic flux results in the selective neutrality of mutations targeting enzyme concentrations.

### 3.1 Equilibria of relative enzyme concentrations under directional selection for increased flux

345

We searched for evolutionary equilibria of relative enzyme concentrations in four situations: independence between enzyme concentrations, fixed total enzyme concentration (due to competition), co-regulation between all enzyme concentrations and both competition and co-regulation.

#### 3.1.1 Case 1: Independence between enzymes

350

When the enzyme concentrations are independent, the evolutionary equilibrium of relative enzyme concentrations is (Supporting Information [SI.B.4](#)):

$$\forall i, \quad e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}} \quad (23)$$

Simulation results of a three-enzyme pathway (Figure [1A](#)) are consistent with this theoretical equilibrium: the evolutionary trajectories converge towards the point of the space of relative concentrations defined by equation [23](#) (Figure [1B](#)). At this point, the flux response coefficients, which quantify the sensitivity of the flux to variation in enzyme concentrations, are (Supporting Information [SI.B.4](#)):

355

$$\forall i, \quad R_{E_i}^{J*} = \frac{A_i^{-2/3}}{\sum_{j=1}^n A_j^{-2/3}} \quad (24)$$

Because the flux response coefficients are strictly positive and there is no upper limit on enzyme concentration, the flux is continuously increasing (Figure [1C](#)) while the proportions of the different enzymes remain constant at equilibrium (equation [23](#)).

360

From equations [23](#) and [24](#), we see that the relative concentrations and the flux response coefficients at equilibrium are inversely related to the pseudo-activities: enzymes with low pseudo-activity have a high concentration relative to others enzymes and exert a high control on the flux. For these enzymes, variation in abundance is less likely to be selectively neutral (see Section [3.2.2](#)).

365

#### 3.1.2 Case 2: Competition for cellular resources

In order to model the effects of limited cellular resources, the total enzyme abundance  $E_{\text{tot}}$  in the pathway was considered to be constant. When a mutation targets the concentration of an enzyme, the concentrations of all enzymes are assumed to be redistributed proportionally to maintain constant the total concentration. As a consequence, the relationship between relative enzyme concentrations and flux is a dome in the multidimensional space

370

of relative enzyme concentrations, the shape of which is determined by the vector of pseudo-activities  $\mathbf{A}$  (Figure 2A). The flux cannot exceed a maximum value  $J_{\max}$ , which is the flux value at evolutionary equilibrium.

At  $J_{\max}$  the flux response coefficients are:

$$\forall i, \quad R_{E_i}^J = 0$$

375 because at the top of the dome, the derivative of the flux with respect to  $E_i$  is null in all dimensions. From this equality, the evolutionary equilibrium of relative enzyme concentrations at  $J_{\max}$  can be found (Supporting Information SI.B.5):

$$\forall i, \quad e_i^* = \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}} \quad (25)$$

As in the case of independence, simulation results were consistent with this theoretical equilibrium (Figure 2B). As previously, the higher the pseudo-activity of the enzyme, the  
380 lower its relative concentration, but due to the difference in the value of the power of  $A_i$  ( $-1/2$  instead of  $-1/3$ ), the relative enzyme concentrations at evolutionary equilibrium were more dispersed when there was competition than when there was independence.

From equation 25, we can calculate  $J_{\max}$ , which is linearly related to  $E_{\text{tot}}$  (Supporting Information SI.B.5):

$$J_{\max} = E_{\text{tot}} \frac{X}{\left( \sum_{j=1}^n A_j^{-1/2} \right)^2} \quad (26)$$

385 Figure 2C shows the results of the simulations of flux evolution when  $E_{\text{tot}}$  is fixed. The expected  $J_{\max}$  was reached whatever the value of the initial enzyme concentrations.

### 3.1.3 Case 3: Co-regulation between enzyme concentrations

When there is co-regulation between enzyme concentrations, in the absence of competition, a mutation targeting a given enzyme results in variation in the concentration of  
390 the other enzymes. To model this phenomenon, we used the *co-regulation coefficient*  $\beta_{ij}$  – the slope of the linear relation between  $E_i$  and  $E_j$  –, which quantifies the variation in  $E_j$  when a mutation affects  $E_i$ .

When all enzyme concentrations are related to each other ( $\forall(i, j), \beta_{ij} \neq 0$ ), the evolutionary equilibrium of the relative enzyme concentrations is (Supporting Information  
395 SI.B.6.1.1):

$$\forall i, \quad e_i^* = \frac{1}{B_i} \quad (27)$$

where  $B_i = \sum_{j=1}^n \beta_{ij}$ , the *global regulation coefficient* of enzyme  $i$ , quantifies the effect of the variation of  $E_i$  on the total enzyme concentration (equation 6):  $B_i = \Delta E_{\text{tot}} / \Delta E_i$ . This equilibrium has a non-intuitive property: at this point, mutations, whatever their effect, no longer modify the *relative* enzyme concentrations (Supporting Information SI.B.6.1.2).  
400 Metaphorically, we will call this point an *evolutionary black hole*.

The enzyme concentrations are assumed to be linearly related. Consequently, in the multidimensional space of relative enzyme concentrations, the system can be described as a point moving on a straight line  $\mathcal{E}$ . The parametric equation of  $\mathcal{E}$  can be determined

from two known points, the initial relative enzyme concentrations  $\mathbf{e}^0$  and the theoretical equilibrium  $\mathbf{e}^*$  (Supporting Information [SI.B.6.2](#)):

$$\mathcal{E} : \quad \forall i, \quad e_i = \tau(e_i^* - e_i^0) + e_i^0 = \tau \left( \frac{1}{B_i} - e_i^0 \right) + e_i^0 \quad (28)$$

where  $\tau$  is the driving variable. When  $\tau = 0$ , the system is at the initial state  $\mathbf{e}^0$  and when  $\tau = 1$ , the system is at the theoretical equilibrium  $\mathbf{e}^*$ . For a given  $\mathbf{e}^0$  and  $\mathbf{M}_\beta$  (the matrix of all pairwise co-regulation coefficients), the relative concentrations  $\mathbf{e}$  are strictly determined by the value of  $\tau$ . Consistently, results of the simulations showed linear evolutionary trajectories converging towards the theoretical equilibrium (Figures [3A](#) and [3C](#)), with a clear difference between positive and negative co-regulation:

### 1. Positive co-regulation

When all  $\beta_{ij}$ 's are strictly positive, the relative enzyme concentrations converge towards the theoretical equilibrium  $\mathbf{e}^*$  located within the space of relative enzyme concentrations (Figure [3A](#)). At this point, the flux response coefficients are (Supporting Information [SI.B.6.1.3](#)):

$$\forall i, \quad R_{E_i}^{J^*} = 1 \quad (29)$$

meaning that every enzyme strictly controls the flux. This is not in contradiction with the previously mentioned evolutionary black hole, because the flux response coefficients concern the *absolute* enzyme concentrations, which increase indefinitely, as does the flux (Figure [3B](#)), while the *relative* enzyme concentrations are stuck at the theoretical equilibrium.

### 2. Negative co-regulation

If at least one co-regulation coefficient is negative, then at least one enzyme has a negative global regulation coefficient  $B_i$  (Supporting Information [SI.B.1.2](#)). As concentrations cannot be negative, the theoretical equilibrium  $\mathbf{e}^*$  is unattainable because it is outside the space of possible relative enzyme concentrations (Figure [3C](#)). There is another evolutionary equilibrium, which we called the *effective equilibrium*  $\tilde{\mathbf{e}}$ , defined by (Supporting Information [SI.B.6.3](#)):

$$\forall i, \quad R_{E_i}^J = 0 \quad (30)$$

Null response coefficients mean that in this case there is a maximum value of the flux,  $\tilde{J}_{\max}$ , for the same reason as in the case of competition: the negative co-regulation coefficients prevent the flux from exceeding a certain value because mutations that increase (resp. decrease) the concentration of an enzyme will decrease (resp. increase) the concentration of the other enzyme(s).

The *effective equilibrium*  $\tilde{\mathbf{e}}$  could not be found analytically but could be found numerically from another expression of equation [30](#) (Supporting Information [SI.B.6.3](#)):

$$\sum_{j=1}^n \frac{1}{B_j A_j (\tilde{e}_j)^2} = 0 \quad (31)$$

The black curve on Figure [3C](#) shows the set of solutions  $\tilde{\mathbf{e}}$  satisfying equation [31](#) for a given set of  $A_i$  and  $B_i$  values and a large number of initial enzyme concentrations  $\mathbf{e}^0$ . All evolutionary trajectories converge towards the theoretical equilibrium, but stop on this curve at their respective effective equilibrium. Figure [3D](#) shows the evolution of the corresponding fluxes and their respective  $\tilde{J}_{\max}$ .

### 3.1.4 Case 4: Competition *plus* co-regulation

If there is competition and co-regulation, the relationship between flux and enzyme concentrations is a dome in the multidimensional space of relative enzyme concentrations, but here the relative enzyme concentrations are constrained by the effects of co-regulation to follow line  $\mathcal{E}$  (equation 28). Consequently, the flux evolves following the curve defined by the intersection of the dome with the upright plane drawn from line  $\mathcal{E}$  in space  $(\mathbf{e}, J)$  (Figure 4). There are three remarkable points on this curve: (i) the flux value corresponding to the initial relative enzyme concentrations  $\mathbf{e}^0$  (black point); (ii) the flux value corresponding to the theoretical equilibrium of relative enzyme concentrations  $\mathbf{e}^*$  (red point); (iii) the maximum of this curve,  $\tilde{J}_{\max}$ , corresponding to the effective equilibrium of relative enzyme concentrations  $\tilde{\mathbf{e}}$  (yellow point). Note that for a given  $\mathbf{e}^*$ , the orientation of the plane is determined by  $\mathbf{e}^0$ .

The theoretical equilibrium is the same as in the case of co-regulation alone (Supporting Information SI.B.7.1.1):

$$\forall i, \quad e_i^* = \frac{1}{B_i} \quad (32)$$

and the effective equilibrium is also defined by (Supporting Information SI.B.7.2):

$$\forall i \quad R_{E_i}^J = 0 \quad (33)$$

which leads to the following equation:

$$\sum_{j=1}^n \frac{(1/B_j) - e_j^0}{A_j(\tilde{e}_j)^2} = 0, \quad (34)$$

from which  $\tilde{\mathbf{e}}$  can be found numerically. If we consider a large number of possible  $\mathbf{e}^0$ 's for given  $\mathbf{M}_\beta$  and  $\mathbf{A}$ , we can compute a set of points corresponding to  $\tilde{\mathbf{e}}$ , which defines a closed curve in the case of positive co-regulation (Figure 5A) and an open curve in the case of negative co-regulation (Figure 5C). As previously, all evolutionary trajectories end on these curves. At  $\tilde{\mathbf{e}}$  the flux reaches  $\tilde{J}_{\max}$ , which depends on  $\mathbf{e}^0$ ,  $\mathbf{M}_\beta$ ,  $\mathbf{A}$  and  $E_{\text{tot}}^0$  (Figures 5B and 5D).

When there is both competition and positive co-regulation, two possible evolutionary equilibria exist, depending on the initial relative enzyme concentrations  $\mathbf{e}^0$  (Figure 4). If the theoretical equilibrium is reached first, the evolutionary black hole, which is also valid in this situation (Supporting Information SI.B.7.1.2), will prevent any flux increase despite the selective pressure. If the effective equilibrium is reached first, selection for high flux will prevent the theoretical equilibrium from being reached because the flux value is higher at the effective equilibrium than at the theoretical equilibrium.

### 3.1.5 Summary of the results on evolutionary equilibria

In summary, the conclusions drawn from the analytical developments and the simulations are as follows:

- Independence: The flux increases continuously but the relative enzyme concentrations reach an equilibrium value that depends on the pseudo-activities of the enzymes.
- Competition: The flux increases up to a maximum value that depends on both the pseudo-activities of the enzymes and the total enzyme concentration. The relative



enzyme concentrations reach an equilibrium that depends on the pseudo-activities of the enzymes, but which differs from the one obtained in the case of independence.

- Co-regulations without competition: The relative enzyme concentrations tend towards an equilibrium that depends only on the co-regulation coefficients between enzymes. If the co-regulation coefficients are all positive, the flux increases continuously and equilibrium is reached. This equilibrium is an evolutionary black hole because, once it is reached, additional mutations cannot change the *relative* enzyme concentrations. If there is at least one negative co-regulation coefficient, the flux evolves towards a maximum and a different equilibrium is reached, called the effective equilibrium, which depends not only on the co-regulation coefficients and the pseudo-activities of the enzymes, but also on the initial enzyme concentrations.
- Competition *plus* co-regulation: This case is qualitatively similar to the case with negative co-regulation alone, with a flux maximum and a multi-factor effective equilibrium. The difference resides in the shape of the adaptive landscape.

Tables 1 and 2 summarize the evolutionary equilibria and the limit to the flux value in the different situations.

**Table 1. Theoretical  $e^*$  and effective  $\tilde{e}$  equilibrium for relative enzyme concentrations.**

	Equilibrium	Without competition	With competition
<b>Without co-regulation</b>	Theoretical	$\forall i, e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}}$	$\forall i, e_i^* = \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}}$
<b>Positive co-regulation</b>	Theoretical	$\forall i, e_i^* = 1/B_i$	$\forall i, e_i^* = 1/B_i$
	Effective	–	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$
<b>Negative co-regulation</b>	Theoretical	$\forall i, e_i^* = 1/B_i$ cannot be reached	$\forall i, e_i^* = 1/B_i$ cannot be reached
	Effective	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$	$\tilde{e} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A})$

**Table 2. Limit to the flux value.** The maximum value of the flux corresponds to the effective equilibrium.

	Without competition	With competition
<b>Without co-regulation</b>	$J \rightarrow \infty$	$J_{\max} = f(\mathbf{A}, E_{\text{tot}}^0)$
<b>Positive co-regulation</b>	$J \rightarrow \infty$	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$
<b>Negative co-regulation</b>	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$	$\tilde{J}_{\max} = f(\mathbf{e}^0, \mathbf{M}_\beta, \mathbf{A}, E_{\text{tot}}^0)$

## 3.2 The range of selective neutrality of enzyme concentrations

### 3.2.1 Selective neutrality at evolutionary equilibrium

In addition to searching for evolutionary equilibria, another key question from an evolutionary point of view is the extent to which mutations affecting enzyme concentrations become neutral under selection, *i.e.* have a negligible effect on the flux, and by extension on fitness. The classical criterion defining the neutral zone for haploid populations is a value of the selection coefficient  $s_i$  between  $-1/N$  and  $+1/N$  (Kimura 1985).

In our model, the selection coefficient is related to  $\delta_i$ ,  $R_{E_i}^J$  and  $E_i$  following the relation  $s_i \approx R_{E_i}^J \frac{\delta_i}{E_i}$  (equation 21). When  $J$  is limited (when there is competition, negative co-regulation or both competition and co-regulation),  $R_{E_i}^J$  is null at the effective equilibrium of relative enzyme concentrations, and thus mutations are nearly neutral near this point. When  $J$  is not limited (when enzymes are independent or when there is positive co-regulation), response coefficients are constant at the theoretical equilibrium and  $\frac{\delta_i}{E_i}$  tends towards zero because  $E_i$ 's increase indefinitely, and thus mutations are nearly neutral at this point. Therefore, selection for high flux leads to the near neutrality of mutations affecting any enzyme in the pathway in all situations considered.

### 3.2.2 The range of neutral variation depends on the enzymes and the constraints

Even though the mutations affecting enzyme concentrations become neutral as relative enzyme concentrations approach equilibrium, the size of the neutral zone of mutations is expected to vary depending on the enzyme properties and the constraints applied. The zone of selective neutrality determines the *range of neutral variation* (RNV) of the concentrations. The RNV size is given by  $|\delta_i^{\text{sup}} - \delta_i^{\text{inf}}|$ , where  $\delta_i^{\text{sup}}$  (resp.  $\delta_i^{\text{inf}}$ ) is the effect of a mutation such that  $s_i = +\frac{1}{N}$  (resp.  $s_i = -\frac{1}{N}$ ),  $s_i$  being the superior (resp. inferior) limit of the neutral zone. Let us examine the four situations.

**3.2.2.1 Case 1: Independence** From equations 21, 23 and 24, we deduced the selection coefficient at theoretical equilibrium:

$$s_i \approx R_{E_i}^{J*} \frac{\delta_i}{E_{\text{tot}} e_i^*} \approx \frac{A_i^{-2/3}}{\sum_{j=1}^n A_j^{-2/3}} \frac{\delta_i}{\frac{A_i^{-1/3} E_{\text{tot}}}{\sum_{j=1}^n A_j^{-1/3}}}$$

Since  $\delta_i = \nu$  when enzyme concentrations are independent, it comes:

$$s_i \approx \frac{\nu}{E_{\text{tot}}} \frac{A_i^{-1/3} \sum_{j=1}^n A_j^{-1/3}}{\sum_{j=1}^n A_j^{-2/3}} \quad (35)$$

We see that, for a given ratio  $\nu/E_{\text{tot}}$ , the pseudo-activities  $A_j$  alone determine the  $s_i$ 's. For two enzymes  $i$  and  $i'$ , if  $A_i < A_{i'}$ , then  $|s_i| > |s_{i'}|$ , *i.e.* the selection coefficients are inversely related to the pseudo-activities. Therefore, to get out of the neutral zone, a larger mutation effect is required for enzymes with large  $A_j$ 's than for those with small  $A_j$ 's (Figure S1A). This means that RNV size is positively related to the pseudo-activities

$\mathbf{A}$ , and hence is inversely related to the theoretical equilibria  $e^*$ . Numerical application was performed for a ten-enzyme pathway using equation 22 to compute the limits of the RNVs. Results were fully consistent with the theoretical predictions (Figure 6A).

**3.2.2.2 Case 2: Competition** We did not find any analytical method to study the RNV size near the equilibrium. Graphically, we observed that an enzyme with low pseudo-activity needs larger mutations to get out of the neutral zone (Figure S1B). RNVs were computed numerically from equation 22. We found that RNVs were *inversely* related to enzyme pseudo-activities, unlike in case 1, and hence were positively related to  $e^*$  (Figure 6B). Another striking observation is that, for a given total enzyme concentration, RNVs are one order of magnitude greater in the case of competition than when there is independence.

**3.2.2.3 Case 3: Co-regulation** From equation 28, the relation between *absolute* enzyme concentration and the driving variable  $\tau$  when there is co-regulation can be written as (Supporting Information SI.B.8.1):

$$E_i = E_i^0 + \frac{E_{\text{tot}}^0}{B_i} \left( -1 + \frac{1}{1 - \tau} \right) \quad (36)$$

If we consider the superior and inferior limits of the RNV for enzyme  $i$ , respectively  $E_i^{\text{sup}}$  and  $E_i^{\text{inf}}$  (*i.e.*  $E_i + \delta_i^{\text{sup}}$  and  $E_i + \delta_i^{\text{inf}}$ ), we can write:

$$|E_i^{\text{sup}} - E_i^{\text{inf}}| = \left| \frac{E_{\text{tot}}^0}{B_i} \left( \frac{1}{1 - \tau^{\text{sup}}} - \frac{1}{1 - \tau^{\text{inf}}} \right) \right| \quad (37)$$

where  $\tau^{\text{sup}}$  and  $\tau^{\text{inf}}$  are the superior and inferior limits of the driving variable.

We see that RNVs are inversely related to  $|B_i|$ 's: the enzymes that have the smallest effect on total enzyme concentration *via* co-regulation have the largest RNV (Figures 6C and 6E).

**3.2.2.4 Case 4: Competition plus co-regulation** The same method as previously was applied when we considered jointly competition and co-regulation. From equation 28, we have (Supporting Information SI.B.8.2):

$$E_i = \tau \left( \frac{E_{\text{tot}}}{B_i} - E_i^0 \right) + E_i^0 \quad (38)$$

Considering the superior and inferior limits of the RNV for enzyme  $i$  (see above), we get:

$$|E_i^{\text{sup}} - E_i^{\text{inf}}| = E_{\text{tot}} \left| \left( \frac{1}{B_i} - e_i^0 \right) (\tau^{\text{sup}} - \tau^{\text{inf}}) \right| \quad (39)$$

Thus, there is a positive relationship between RNV and  $|1/B_i - e_i^0|$ , *i.e.* the difference between the theoretical equilibrium ( $1/B_i$ ) and the initial relative concentration ( $e_i^0$ ) (Figures 6D, 6F), meaning that an enzyme whose concentration is initially far from its theoretical equilibrium will have a large RNV. A geometric representation of this relationship is shown in Figure 7 when there is positive co-regulation.

Table 3 and Figure 6 summarize the factors determining the ranking order of RNVs depending on the type of constraint.

**Table 3. Factors determining the ranking order of RNVs**

	<b>Without competition</b>	<b>With competition</b>
<b>Without regulation</b>	Same order as the $A_i$ 's	Reverse order of the $A_i$ 's
<b>With regulation</b>	Reverse order of the $ B_i $ 's	Same order as the $\left  \frac{1}{B_i} - e_i^0 \right $ 's

### 3.2.3 Measuring RNVs from simulations of long-term evolution

560 We tested these theoretical predictions by simulating the long-term evolution of a three-enzyme pathway. For each enzyme and each situation, we computed the RNV limits at each time step between 60,000 and 120,000 generations (*i.e.* once evolutionary equilibrium is reached), then we computed the mean size of the RNVs. The relationship between mean RNV value and  $A_i$ ,  $B_i$  or  $\left| \frac{1}{B_i} - e_i^0 \right|$  was fully consistent with theoretical predictions  
565 (Figure S2).

In addition, the simulations led to two observations.

- When the flux is not limited (independence and positive co-regulation), RNV size increases over the course of generations. For example, if  $\mathbf{E}^0 \approx (21.9, 30.7, 47.3)$ , mean RNV size for enzyme 1 is  $\approx 0.13$  between 60,000 and 80,000 generations, and  
570  $\approx 0.17$  between 100,000 and 120,000 generations. The reason is that, over the course of generations, a mutation of a given effect has increasingly less effect on the flux, and therefore on the selection coefficient, because enzyme concentrations increase continuously.
- When there is competition and/or negative co-regulation, the dome-shape of the flux-enzyme relationship can lead to a split of the RNVs. Near the maximum flux, an upper limit of the RNV cannot be defined because mutations are either neutral or deleterious. When the flux moves away from the maximum, advantageous mutations may occur, with the RNV having an upper and a lower limit. But interestingly,  
575 another RNV appears on the other side of the dome, due to large mutations that result in flux values that are similar to those caused by mutations of small effect.  
580 (Figure S3). The farther the flux is from the maximum, the smaller the two moieties of the RNVs and the stronger the selection to bring the concentrations towards effective equilibrium.

## 4 Discussion

585 Metabolic fluxes are commonly used as model traits in quantitative genetics and evolutionary studies, benefitting from the development of elaborate theoretical tools that enable to relate “genotype” (the genetically controlled enzyme properties) to phenotype (the flux). The multidimensional flux-enzyme relationship can be viewed as an adaptive landscape and used to analyze the dynamics of enzyme evolution. In this study, we developed a  
590 system of differential equations that describes the evolution of enzyme concentrations in a pathway under directional selection for increased metabolic flux. We considered two realistic constraints, separately or jointly, that prevent enzyme concentrations from varying independently from each other, namely competition for cellular resources and

co-regulation between enzymes. We defined two types of evolutionary equilibria of relative enzyme concentrations: a theoretical equilibrium and an effective equilibrium, the latter being reached when the flux is limited by certain constraint(s) (tables 1 and 2). In both constrained and unconstrained conditions, mutations become near neutral near the evolutionary equilibrium, however the range of neutral variation (RNVs) of enzyme concentrations differs to a large extent depending on the enzyme properties and the constraints applied (table 3). Simulation results of long-term evolution were fully consistent with the analytical predictions.

## 4.1 Modeling rationale

The effect of an enzyme  $j$  on the flux of a linear pathway depends on three factors: the ratio of kinetic parameters  $\frac{k_{\text{cat}j}}{K_{Mj}}$ , the abundance  $E_j$  and the product of the equilibrium constants from the first to the  $(j - 1)^{\text{th}}$  reaction,  $K_{1,j-1}$ . Both  $E_j$  and  $\frac{k_{\text{cat}j}}{K_{Mj}}$  can be genetically variable. Recently, Heckmann et al. (2018) simulated the evolution of turnover numbers  $k_{\text{cat}j}$  and showed that their values can hardly be maximized.

In our model, we distinguished the concentration  $E_j$  from the "pseudo-activity"  $A_j = \frac{k_{\text{cat}j}}{K_{Mj}} K_{1,j-1}$  of enzyme  $j$ , the latter being a systemic property as its value depends on upstream reactions in the pathway. We focused on enzyme concentrations, which display two features that are relevant from an evolutionary point of view:

(i) The range of natural variability is expected to be noticeably larger for enzyme concentration than for catalytic efficiency  $\frac{k_{\text{cat}j}}{K_{Mj}}$ . While genetically controlled changes in kinetic parameter values are caused by substitutions at certain sites of the enzyme-coding gene, changes in protein expression can be caused by mutations at multiple *cis* and *trans* elements affecting the gene regulatory network, the translation machinery, mechanisms of protein turnover and stability, etc. (Landry et al. 2007; Zheng et al. 2011; Gruber et al. 2012). This is consistent with the highly polygenic nature of proteomic variations observed in various species (Damerval et al. 1994; Albert and Kruglyak 2015; Chick et al. 2016). Therefore, proteome variability could play a much larger role in the evolution of pathway efficiency than the variability of kinetic parameters. Thus, in our model, we chose to consider mutations that target enzyme concentrations  $E_j$ , and let pseudo-activities  $A_j = \frac{k_{\text{cat}j}}{K_{Mj}} K_{1,j-1}$  be invariable. Note that kinetic parameters and concentration play an identical role in the expression of the flux, but the constraints applied to  $E$  do not make sense for  $k_{\text{cat}}$  and  $K_M$ .

(ii) Enzyme synthesis and degradation are energetically costly for the cell; furthermore, space and ribosome availability are limited (Koehn 1991; de Vienne et al. 2001b; Klumpp et al. 2019) Therefore, protein over-expression can result in protein burden effects (Snoep et al. 1995; Eguchi et al. 2018). Continuous selective pressure for optimizing cellular resource utilization imposes a constraint on total enzyme abundance and a trade-off between enzymes (Albertin et al. 2013).

In addition to the constraint on  $E_{\text{tot}}$ , we also considered the possibility that enzyme concentrations were co-regulated. In the cell, multiple mechanisms can result, directly or indirectly, in dependencies between enzyme quantities, *e.g.* operons in prokaryotes, metabolic gene clusters in eukaryotes (Lee and Sonnhammer 2003; Wong and Wolfe 2005)

regulation cascades (Albert and Kruglyak 2015), *trans* effects due to transcription factors (Guerin et al. 2016; Lambert et al. 2018), transposable elements (Cowley and Oakey 2013) and microRNAs (Wang et al. 2011), etc. To model the evolutionary implications of these dependencies within a single theoretical framework, we considered that a mutation targeting the concentration of an enzyme modifies the concentrations of the other enzymes. To keep things relatively simple, we chose linear and constant co-regulation coefficients, but we assumed that they could take on any value. When coordinated enzyme expression is due to a common *trans* factor or physically clustered genes, the system of differential equations has to be modified slightly because in this case the mutations target the driving variable  $\tau$  (Supporting Information SI.B.9). This changes the distribution of the mutation effects on the co-regulated enzymes, but does not modify evolutionary equilibria, which do not depend neither on the effects nor on the targets of the mutations. Further developments could rely on more complex models of gene regulatory network functioning and evolution (De Jong 2002; Spirov and Holloway 2013).

The two-step process we used to formalize the mutation effects under competition is not simply theoretically convenient, but may reflect actual events at the molecular level. First, a mutation with a canonical effect  $\nu$  affects the transcription level of an enzyme-coding gene. Second, transcripts are translated proportionally to their abundance up to the maximum total enzyme concentration allowed by the cell. This results in an actual mutation effect on the concentration of the target enzyme and in pseudo-mutation effects on the other enzymes. The process is similar if there is co-regulation in addition to competition.

As in Serohijos et al. (2012) and Heckmann et al. (2018), our evolution model is based on an adaptive dynamics approach that relies on classical population genetics concepts, such as the selection coefficient  $s$  and the fixation probability  $P_f$ , with the assumption of constant population size  $N$ . Other assumptions were made to render tractable the mathematical developments: small mutation effect relative to enzyme concentration, no fixation of deleterious mutations and large population size. We relaxed these assumptions in the simulations, and this did not affect the predictions of evolutionary equilibria of relative enzyme concentrations and maximum flux values (if applicable).

## 4.2 Constraints shape the adaptive landscape

### 4.2.1 Optimal enzyme concentrations depend on the type of constraints

The evolutionary equilibria of relative enzyme concentrations depend on the constraints applied. If enzyme concentrations vary independently, there is an inverse relationship between relative concentrations and pseudo-activities  $\mathbf{A}$  at equilibrium, which is intuitively consistent, and both enzyme concentrations and flux can increase indefinitely. This is not biologically realistic because enzyme concentrations are necessarily bounded by limited cellular resources and various regulatory mechanisms. Such constraints shape the adaptive landscape.

When enzymes compete for resources, the adaptive landscape is a dome in the multidimensional space of relative enzyme concentrations; its shape is determined by the enzyme pseudo-activities and its height is determined by the total enzyme concentration. The top of the dome corresponds to the maximum flux reached at equilibrium (Figure 2A). At that point, the relative enzyme concentrations are inversely related to the pseudo-activities  $\mathbf{A}$ , as when enzyme concentrations vary independently, but they are more dispersed. At present, we have no explanation for this result.

Does selection for high flux result in the optimal distribution of enzyme concentrations when there is competition? The theory of optimality of cost minimization has been developed in detail (Klipp and Heinrich 1999; Liebermeister et al. 2004; Noor et al. 2016). Klipp and Heinrich (1999) addressed the question of the optimal distribution of enzyme concentrations in a pathway when there is competition between enzymes. They considered a fixed flux and searched for the distribution that minimizes the total enzyme concentration, which amounts to maximizing the flux with a fixed total concentration, as we did here. Using the Lagrange multiplier method, they obtained the same equilibrium of relative enzyme concentrations as the one we obtained using an adaptive dynamics approach. Thus, the mutation-selection process for increased flux amounts to optimally minimizing the total enzyme concentration allocated to the pathway.

Co-regulation between enzymes prevents reaching this optimum. With co-regulation, the evolutionary trajectories of the relative enzyme concentrations become linear; the orientation of this line is determined by the initial enzyme concentrations and the co-regulation coefficients. Hence the flux follows the curve defined by the intersection of the dome and the upright plane drawn from the line (Figure 4). The maximal flux value on this curve,  $\tilde{J}_{\max}$ , is a local maximum corresponding to the effective equilibrium of relative enzyme concentrations.

If there is co-regulation, with or without competition, the theoretical equilibrium of relative enzyme concentrations depends on the co-regulation coefficients, according to the simple relation  $e_i^* = 1/B_i$ , where the global regulation coefficient  $B_i$  quantifies the effect of  $E_i$  variation on the total enzyme concentration. This evolutionary equilibrium implies that:

(i) The higher the effect of enzyme concentration variation on the total concentration, the lower the relative concentration of this enzyme at evolutionary equilibrium;

(ii) Pseudo-activities  $A_i$  do not appear in this relation. This does not mean that the kinetic parameters and equilibrium constants do not affect the evolutionary trajectory when there is co-regulation. Indeed,  $A_i$ 's determine the shape of the adaptive landscape, and hence the flux value at evolutionary equilibrium;

(iii) This equilibrium is an evolutionary black hole: as one gets closer, the effect of mutations on *relative* enzyme concentrations decreases, and is cancelled out at this equilibrium irrespective of the effects of these mutations. This property has an important consequence when there is competition, because it can prevent the flux from reaching its local maximum  $\tilde{J}_{\max}$ . If the initial flux value is lower than the flux value at theoretical equilibrium *and* both values are on the same side of the curve (*i.e.* the derivatives at these points have the same sign), selection will be unable to bring the flux value above the theoretical equilibrium (Figures 4B and 5A);

(iv) If there is negative co-regulation, at least one  $B_i$  is negative and theoretical equilibrium is not achievable. The flux will then reach a maximum corresponding to the effective equilibrium, whether there is competition or not.

In summary, co-regulation can be evolutionarily costly, since it can prevent the flux from reaching the maximum value given the total enzyme concentration allocated to the pathway. Therefore, increasing fitness by increasing the flux should imply having mutations that modify the co-regulation coefficients, *i.e.* that modify the orientation of line  $\mathcal{E}$  to bring  $\tilde{J}_{\max}$  closer to  $J_{\max}$ .

#### 4.2.2 The Range of Neutral Variation of enzyme concentrations depends markedly on enzyme properties, constraints and/or initial enzyme concentrations

730 From visual inspection of the simulation results, we found that relative enzyme concentrations fluctuate close to their evolutionary equilibrium, which is consistent with the mutation-selection-drift balance: selection drives the system towards equilibrium, but near the equilibrium, mutations become nearly neutral and genetic drift dominates (Figures 2B, 3C, 5A, 5C). Near the point of evolutionary equilibrium, mutations have a negligible effect on the flux, and therefore are in the neutral zone. We compared analytically the range of variation of enzyme concentrations in the neutral zone (RNVs) at evolutionary equilibrium under different conditions.

735 When enzyme concentrations vary independently, the RNVs are positively related to the pseudo-activities; the opposite is found when there is competition (Figures 6A and 6B). Thus, competition results in a positive relationship between RNVs and enzyme concentrations. In addition, competition strongly increases all RNVs for a given  $E_{\text{tot}}$ .

745 When there is only co-regulation, enzymes whose variation has the largest effect on total concentration – hence the smallest effective equilibrium of relative enzyme concentrations – have the smallest RNVs. With both competition and co-regulation, the more an enzyme is initially far from the effective equilibrium, the larger its RNV (Figure 7). Therefore, co-regulation abolishes the effect of  $A$  on RNVs, whether there is competition or not, as they did for the theoretical equilibrium values.

750 Interestingly, the existence of a flux maximum can lead to a discontinuous RNV. One part of the RNV corresponds to mutations of small effect, another part to mutations of large effect; the latter push the enzyme concentrations beyond the top of the dome, where flux values are similar to the resident flux. This optimum overshoot is inherent to Fisher’s geometric model of adaptive landscapes (Orr 1998), and is more common when phenotypes are close to the optimum (Scott and Queller 2019).

755 The magnitude of the RNV should be reflected in the level of polymorphism of the genes affecting enzyme abundance, because high-RNV enzymes can accumulate more neutral mutations in these genes than low-RNV enzymes. It would be informative to compare the levels of variability of enzyme abundances within a pathway and investigate how these are related to kinetic parameters and/or known co-regulatory interactions. 760 Quantitative proteomics databases could be exploited for this purpose.

### 4.3 Conclusion and future prospects

We combined MCT principles, an adaptive dynamics approach and simulations to analyze the evolution of enzyme concentrations in metabolic pathways under selection for increased flux. We showed that the adaptive landscape is shaped by enzyme properties and by constraints applied to the system. When there is both competition for cellular resources and co-regulation that limits the flux, which is likely to be the most common situation in a cell, selective pressures lead to equilibria of enzyme concentrations and to ranges of neutral variation that not only depend on enzyme properties and the type of constraints but also on the enzyme concentrations at the start of the selective process. 765 This evolutionary contingency comes from the effect of co-regulation between enzymes and could be mitigated by mutations modifying certain regulatory parameters. In addition, the assumption that all enzymes of the pathway are co-regulated can be relaxed by considering groups of differently co-regulated enzymes (manuscript in preparation).



## Acknowledgements

775 We warmly thank Dr. Wolfram Liebermeister and Dr. Arnaud Le Rouzic for their careful reading of the manuscript and their very helpful suggestions. We also thank Pr. Amandine Veber for her useful suggestions regarding the differential equation system of enzyme evolution and H el ene Citerne for language corrections of the manuscript.

CC was supported by a PhD thesis grant from the French Minist ere de l'Enseignement  
780 Sup erieur, de la Recherche et de l'Innovation.

The authors declare no conflict of interest.

## Author contributions

CD and DdV initiated the project. CD, MLL, GT, CC and DdV developed the model. GT, MLL and CC performed the analysis. CD and MLL wrote the simulation code.  
785 CC developed the dedicated R package. CC and DdV wrote the manuscript, with the contribution of CD.

## Literature cited

- Aguilar-Rodr iguez, J. and A. Wagner, 2018. Metabolic Determinants of Enzyme Evolution in a Genome-Scale Bacterial Metabolic Network. *Genome Biol. Evol.* 10:3076–3088.
- 790 Albert, F. W. and L. Kruglyak, 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 16:197–212.
- Albertin, W., P. Marullo, M. Bely, M. Aigle, A. Bourgeois, O. Langella, T. Balliau, D. Chevret, B. Valot, T. da Silva, C. Dillmann, D. de Vienne, and D. Sicard, 2013. Linking Post-Translational Modifications and Variation of Phenotypic Traits. *Mol Cell Proteomics* 12:720–  
795 735.
- Bershtein, S., A. W. Serohijos, and E. I. Shakhnovich, 2017. Bridging the physical scales in evolutionary biology: From protein sequence space to fitness of organisms and populations. *Curr Opin Struct Biol* 42:31–40.
- Blein-Nicolas, M., W. Albertin, B. Valot, P. Marullo, D. Sicard, C. Giraud, S. Huet, A. Bourgeois,  
800 C. Dillmann, D. de Vienne, and M. Zivy, 2013. Yeast Proteome Variations Reveal Different Adaptive Responses to Grape Must Fermentation. *Molecular Biology and Evolution* 30:1368–1383.
- Br annstr om,  .A., J. Johansson, and N. von Festenberg, 2013. The Hitchhiker's Guide to Adaptive Dynamics. *Games* 4:304–328.
- 805 Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi, 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534:500–505.
- Colombo, M., H. Laayouni, B. M. Invergo, J. Bertranpetit, and L. Montanucci, 2014. Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes. *Evolution* 68:605–  
810 613.
- Cornish-Bowden, A. and M. L. C ardenas, 2020. Contrasting theories of life: Historical context, current theories. In search of an ideal theory. *Biosystems* 188:104063.
- Cowley, M. and R. J. Oakey, 2013. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet* 9:e1003234.
- 815 Crow, J. F. and M. Kimura, 1970. An introduction to population genetics theory. Harper and Row, New York.

- Damerval, C., A. Maurice, J. M. Josse, and D. de Vienne, 1994. Quantitative Trait Loci Underlying Gene Product Variation: A Novel Perspective for Analyzing Regulation of Genome Expression. *Genetics* 137:289–301.
- 820 De Jong, H., 2002. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* 9:67–103.
- Dieckmann, U. and R. Law, 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biology* 34:579–612.
- D’Souza, G. and C. Kost, 2016. Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet* 12:e1006364.
- 825 Eanes, W., 1999. Analysis of selection on enzyme polymorphism. *Annu. Rev. Ecol.* 30:301–326.
- Eguchi, Y., K. Makanae, T. Hasunuma, Y. Ishibashi, K. Kito, and H. Moriya, 2018. Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. *eLife* 7:e34595.
- 830 Ellis, R., 2001. Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* 26:597–604.
- Flowers, J. M., E. Sezgin, S. Kumagai, D. D. Duvernell, L. M. Matzkin, P. S. Schmidt, and W. F. Eanes, 2007. Adaptive Evolution of Metabolic Pathways in *Drosophila*. *Mol Biol Evol* 24:1347–1354.
- 835 Greenberg, A. J., S. R. Stockwell, and A. G. Clark, 2008. Evolutionary Constraint and Adaptation in the Metabolic Network of *Drosophila*. *Molecular Biology and Evolution* 25:2537–2546.
- Gruber, J. D., K. Vogel, G. Kalay, and P. J. Wittkopp, 2012. Contrasting Properties of Gene-Specific Regulatory, Coding, and Copy Number Mutations in *Saccharomyces cerevisiae*: Frequency, Effects, and Dominance. *PLoS Genet* 8:e1002497.
- 840 Guerin, C., T. Joët, J. Serret, P. Lashermes, V. Vaissayre, M. D. Agbessi, T. Beule, D. Severac, P. Amblard, and J. Tregear, 2016. Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *The Plant Journal* 87:423–441.
- Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985. Limits of Adaptation: The Evolution of Selective Neutrality. *Genetics* 111:655–674.
- 845 Heckmann, D., D. C. Zielinski, and B. O. Palsson, 2018. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. *Nature Communications* 9.
- Heinrich, R. and T. Rapoport, 1974. Linear Steady-State Treatment of Enzymatic Chains - General Properties, Control and Effector Strength. *Eur. J. Biochem.* 42:89–95.
- Heinrich, R., S. Schuster, and H.-G. Holzhütter, 1991. Mathematical analysis of enzymic reaction systems using optimization principles. *European Journal of Biochemistry* 201:1–21.
- 850 Jiang, L.-G., B. Li, S.-X. Liu, H.-W. Wang, C.-P. Li, S.-H. Song, M. Beatty, G. Zastrow-Hayes, X.-H. Yang, and F. Qin, 2019. Characterization of proteome variation during modern maize breeding. *Molecular & Cellular Proteomics* 18:263–276.
- Kacser, H. and R. Beeby, 1984. Evolution of catalytic proteins: On the origin of enzyme species by means of natural selection. *J Mol Evol* 20:38–51.
- 855 Kacser, H. and J. A. Burns, 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27:65–104.
- , 1981. The Molecular Basis of Dominance. *Genetics* 97:639–666.
- Kacser, H., J. A. Burns, H. Kacser, and D. A. Fell, 1995. The control of flux : 21 years on. *Biochemical Society Transactions* 23:341–366.
- 860 Kimura, M., 1985. *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Klipp, E. and R. Heinrich, 1994. Evolutionary optimization of enzyme kinetic parameters; effect of constraints. *Journal of theoretical biology* 171:309–323.
- , 1999. Competition for enzymes in metabolic pathways:: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Biosystems* 54:1–14.
- 865 Klumpp, S., W. Bode, and P. Puri, 2019. Life in crowded conditions: Molecular crowding and beyond. *Eur. Phys. J. Spec. Top.* 227:2315–2328.
- Koehn, R., 1991. The Cost of Enzyme-Synthesis in the Genetics of Energy-Balance and Physiological Performance. *Biol. J. Linnean Soc.* 44:231–247.

- Kurland, C. G. and H. Dong, 1996. Bacterial growth inhibition by overproduction of protein. *Molecular Microbiology* 21:1–4.
- 870 Lambert, S. A., A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch, 2018. The human transcription factors. *Cell* 172:650–665.
- Landry, C. R., B. Lemos, S. A. Rifkin, W. J. Dickinson, and D. L. Hartl, 2007. Genetic Properties Influencing the Evolvability of Gene Expression. *Science* 317:118–121.
- 875 Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang, B. Oliver, and A. G. Clark, 2008. Evolution of protein-coding genes in *Drosophila*. *Trends in Genetics* 24:114–123.
- Lee, D.-H. and B. Ø. Palsson, 2010. Adaptive Evolution of *Escherichia coli* K-12 MG1655 during Growth on a Nonnative Carbon Source, l-1,2-Propanediol. *AEM* 76:4158–4168.
- 880 Lee, J. M. and E. L. Sonnhammer, 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome research* 13:875–882.
- Liebermeister, W., E. Klipp, S. Schuster, and R. Heinrich, 2004. A theory of optimal differential gene expression. *Biosystems* 76:261–278.
- Lion, S., F. Gabriel, B. Bost, J. Fiévet, C. Dillmann, and D. De Vienne, 2004. An extension to  
885 the metabolic control theory taking into account correlations between enzyme concentrations. *European Journal of Biochemistry* 271:4375–4391.
- Morrison, E. S. and A. V. Badyaev, 2017. Beyond topology: coevolution of structure and flux in metabolic networks. *J. Evol. Biol.* 30:1796–1809.
- Newton, M. S., V. L. Arcus, M. L. Gerth, and W. M. Patrick, 2018. Enzyme evolution: innovation  
890 is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* 48:110–116.
- Noda-Garcia, L., W. Liebermeister, and D. S. Tawfik, 2018. Metabolite–Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annu. Rev. Biochem.* 87:187–216.
- Noor, E., A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister, 2016. The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization.  
895 *PLOS Computational Biology* 12:e1005167.
- Orr, H. A., 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed During Adaptive Evolution. *Evolution* 52:935–949.
- Pál, C., B. Papp, and M. J. Lercher, 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* 37:1372–1375.
- 900 ———, 2006. An integrated view of protein evolution. *Nat Rev Genet* 7:337–348.
- R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Rausher, M. D., 2013. The Evolution of Genes in Branched Metabolic Pathways. *Evolution*  
905 67:34–48.
- Rausher, M. D., R. E. Miller, and P. Tiffin, 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol* 16:266–274.
- Sambamoorthy, G., H. Sinha, and K. Raman, 2019. Evolutionary design principles in metabolism. *Proc. R. Soc. B.* 286:20190098.
- 910 Scott, T. J. and D. C. Queller, 2019. Long-term evolutionary conflict, Sisyphian arms races, and power in Fisher’s geometric model. *Ecology and Evolution* 9:11243–11253.
- Sellis, D. and M. D. Longo, 2015. Patterns of variation during adaptation in functionally linked loci. *Evolution* 69:75–89.
- Serohijos, A. W. and E. I. Shakhnovich, 2014. Contribution of selection for protein folding  
915 stability in shaping the patterns of polymorphisms in coding regions. *Molecular biology and evolution* 31:165–176.
- Serohijos, A. W. R., Z. Rimas, and E. I. Shakhnovich, 2012. Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Reports* 2:249–256.
- Snoep, J. L., L. P. Yomano, H. V. Westerhoff, and L. O. Ingram, 1995. Protein burden in  
920 *Zymomonas mobilis*: negative flux and growth control due to overproduction of glycolytic

enzymes. *Microbiology* 141:2329–2337.

Spirov, A. and D. Holloway, 2013. Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. *Methods* 62:39–55.

925 de Vienne, D., B. Bost, J. Fiévet, and C. Dillmann, 2001a. Optimisation of Enzyme Concentrations for Unbranched Reaction Chains: The Concept of Combined Response Coefficient. *Acta Biotheor* 49:341–350.

de Vienne, D., B. Bost, J. Fiévet, M. Zivy, and C. Dillmann, 2001b. Genetic variability of proteome expression and metabolic control. *Plant Physiology and Biochemistry* 39:271–283.

930 Vitkup, D., P. Kharchenko, and A. Wagner, 2006. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol* 7:R39.

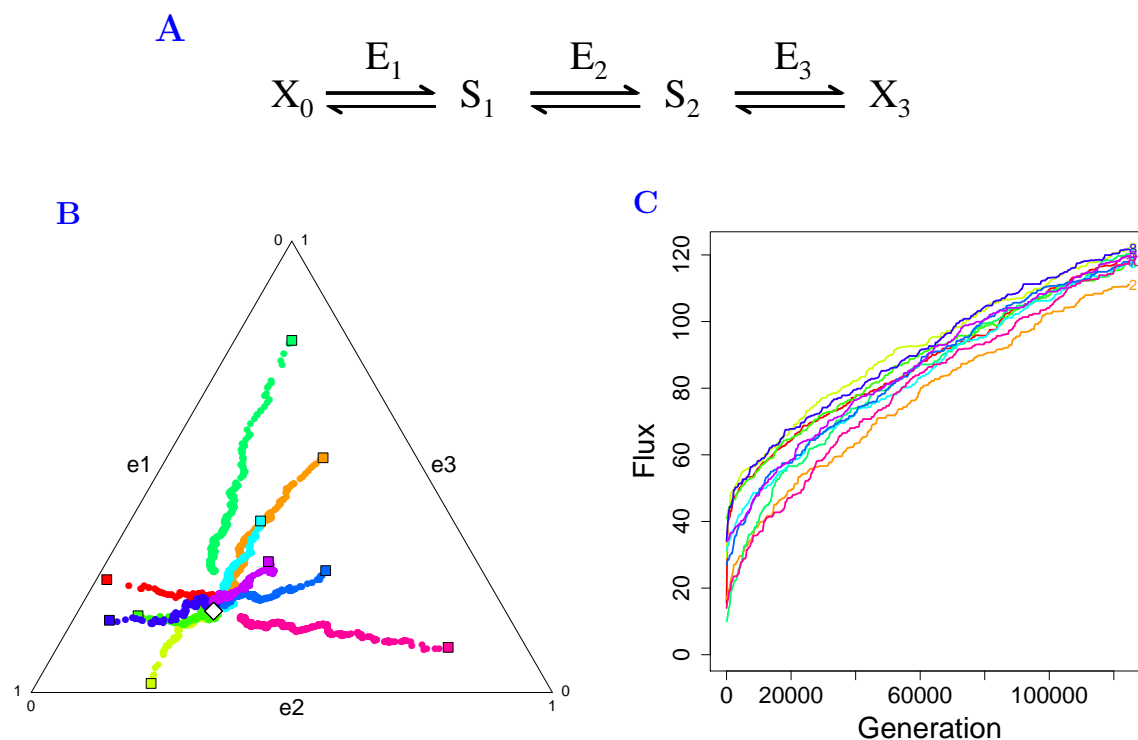
Wang, Y., X. Li, and H. Hu, 2011. Transcriptional regulation of co-expressed microRNA target genes. *Genomics* 98:445–452.

Wong, S. and K. H. Wolfe, 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nature genetics* 37:777–782.

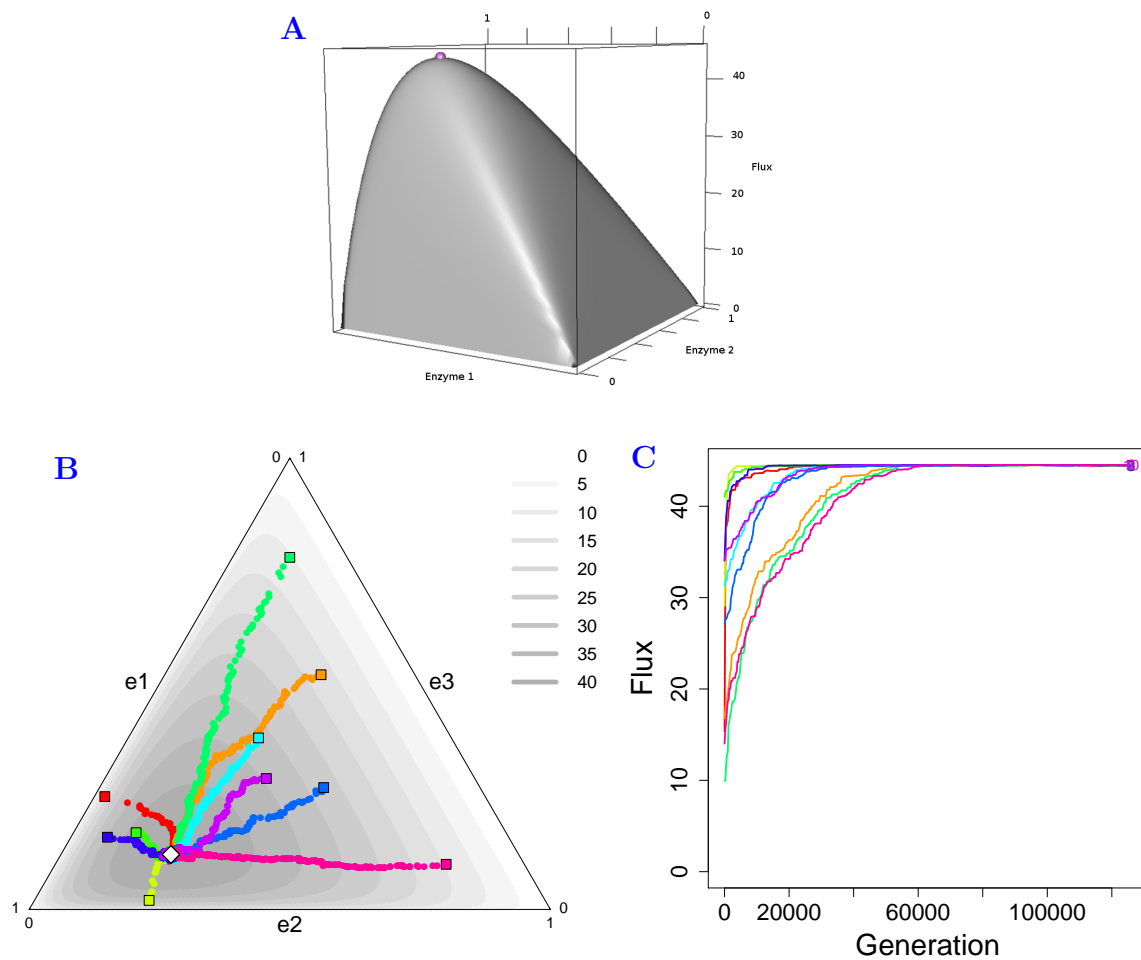
935 Wright, K. M. and M. D. Rausher, 2010. The Evolution of Control and Distribution of Adaptive Mutations in a Metabolic Pathway. *Genetics* 184:483–502.

Zheng, W., T. A. Gianoulis, K. J. Karczewski, H. Zhao, and M. Snyder, 2011. Regulatory Variation Within and Between Species. *Annu. Rev. Genom. Hum. Genet.* 12:327–346.

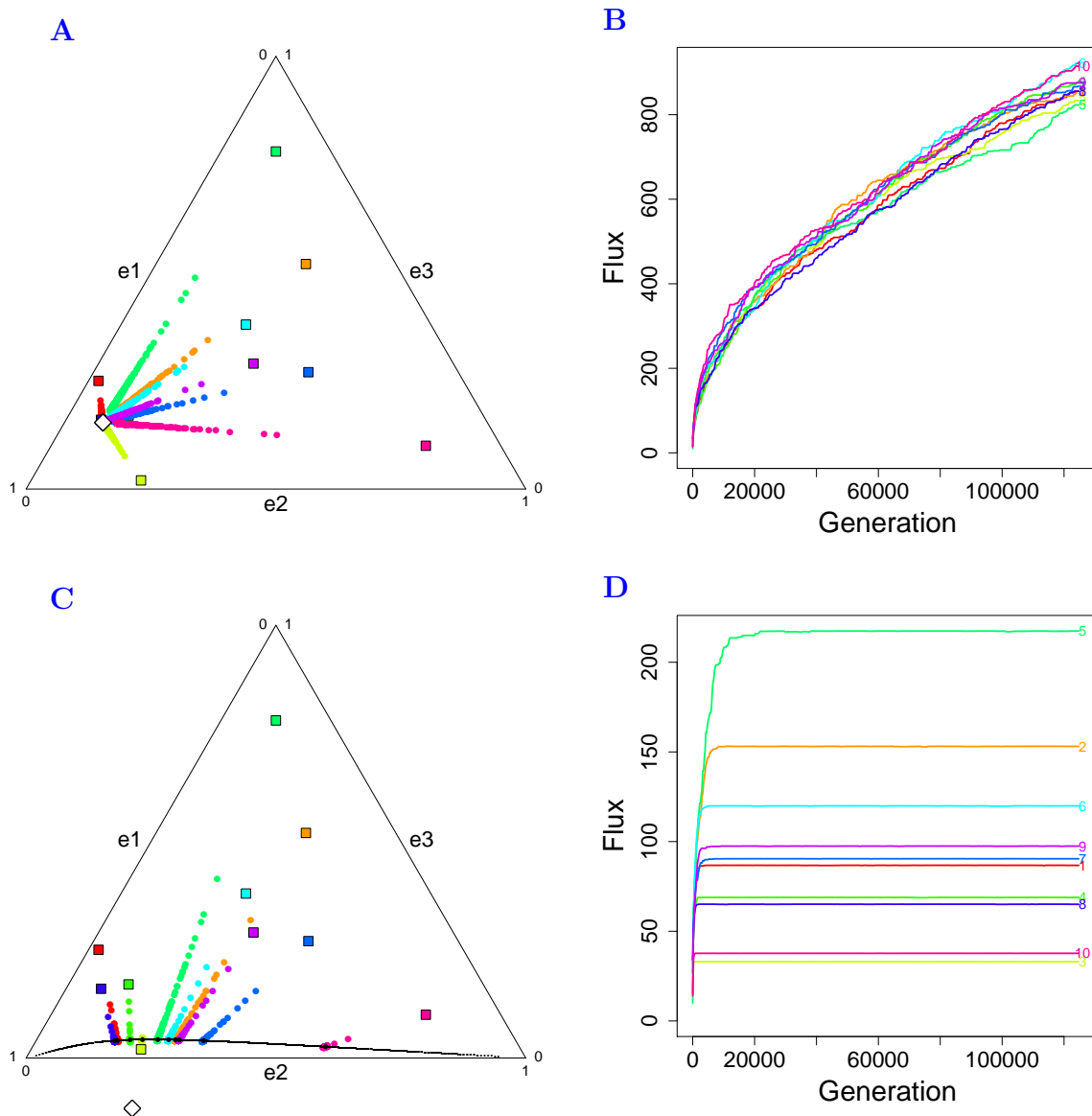
## 5 Figures



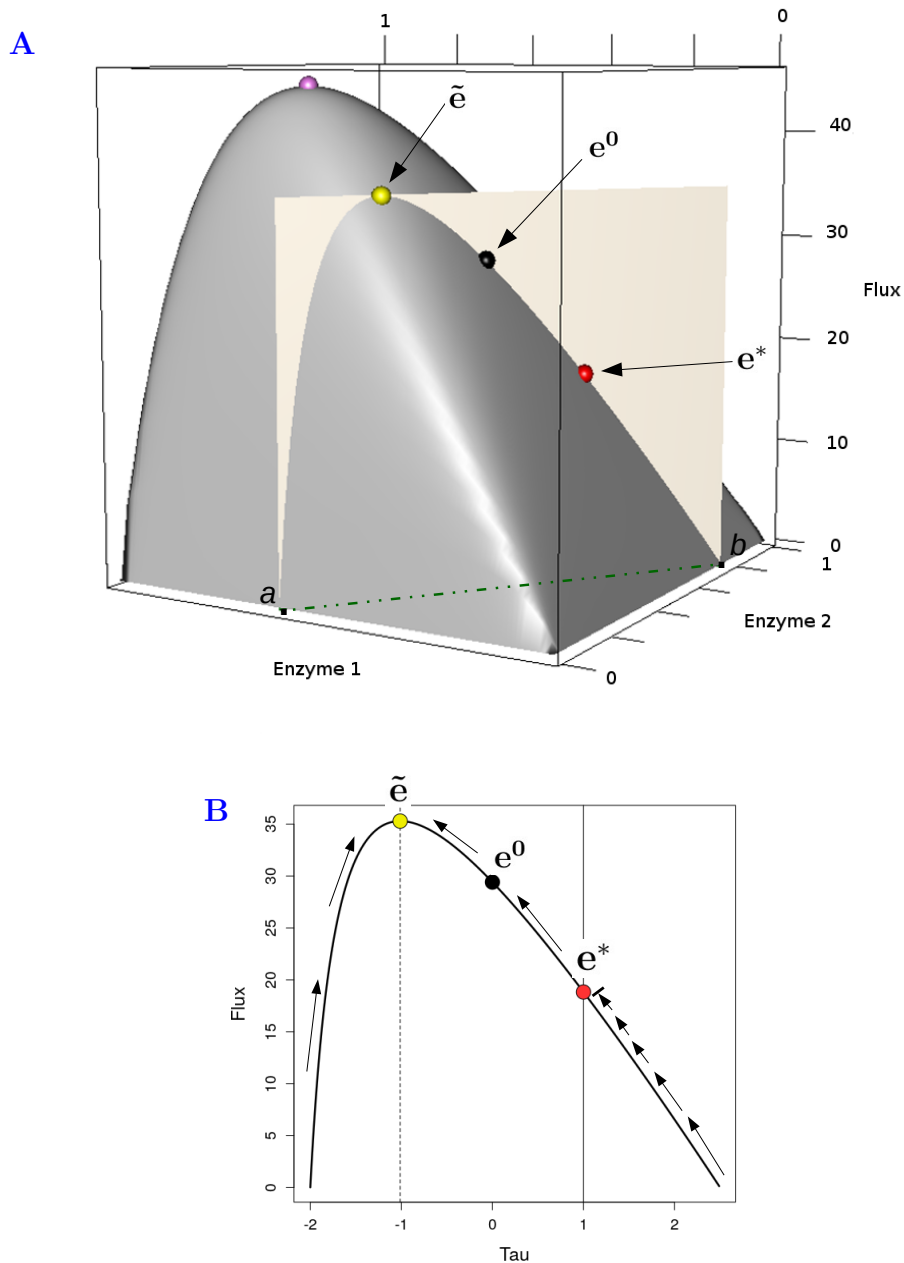
**Figure 1. Simulation of enzyme evolution in a three-enzyme pathway when enzyme concentrations vary independently.** (A) The three-enzyme pathway used in the simulations. Parameters values are:  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{tot}^0 = 100$ ,  $N = 1\,000$ . (B) Evolutionary trajectories of relative enzyme concentrations  $e_1$ ,  $e_2$  and  $e_3$  represented in a triangular diagram for ten simulations (one color per simulation) with different initial enzyme concentrations (colored squares). The relative enzyme concentrations converge towards the theoretical equilibrium  $e^*$  (white diamond). (C) Evolution of the flux over the course of generations from simulations shown in B.



**Figure 2. Simulation of enzyme evolution in a three-enzyme pathway when there is competition.** (A) The flux-enzyme relationship is dome-shaped because  $E_{\text{tot}}$  is constant. The purple point is the maximal flux  $J_{\max}$ , corresponding to the evolutionary equilibrium. (B) and (C) Representations and symbols are as in Figure 1; the grayscale in (B) corresponds to the flux values. When there is competition, the relative enzyme concentrations converge towards the equilibrium (white diamond) and the flux is limited. Parameter values are:  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{\text{tot}}^0 = 100$ ,  $N = 1\,000$ .

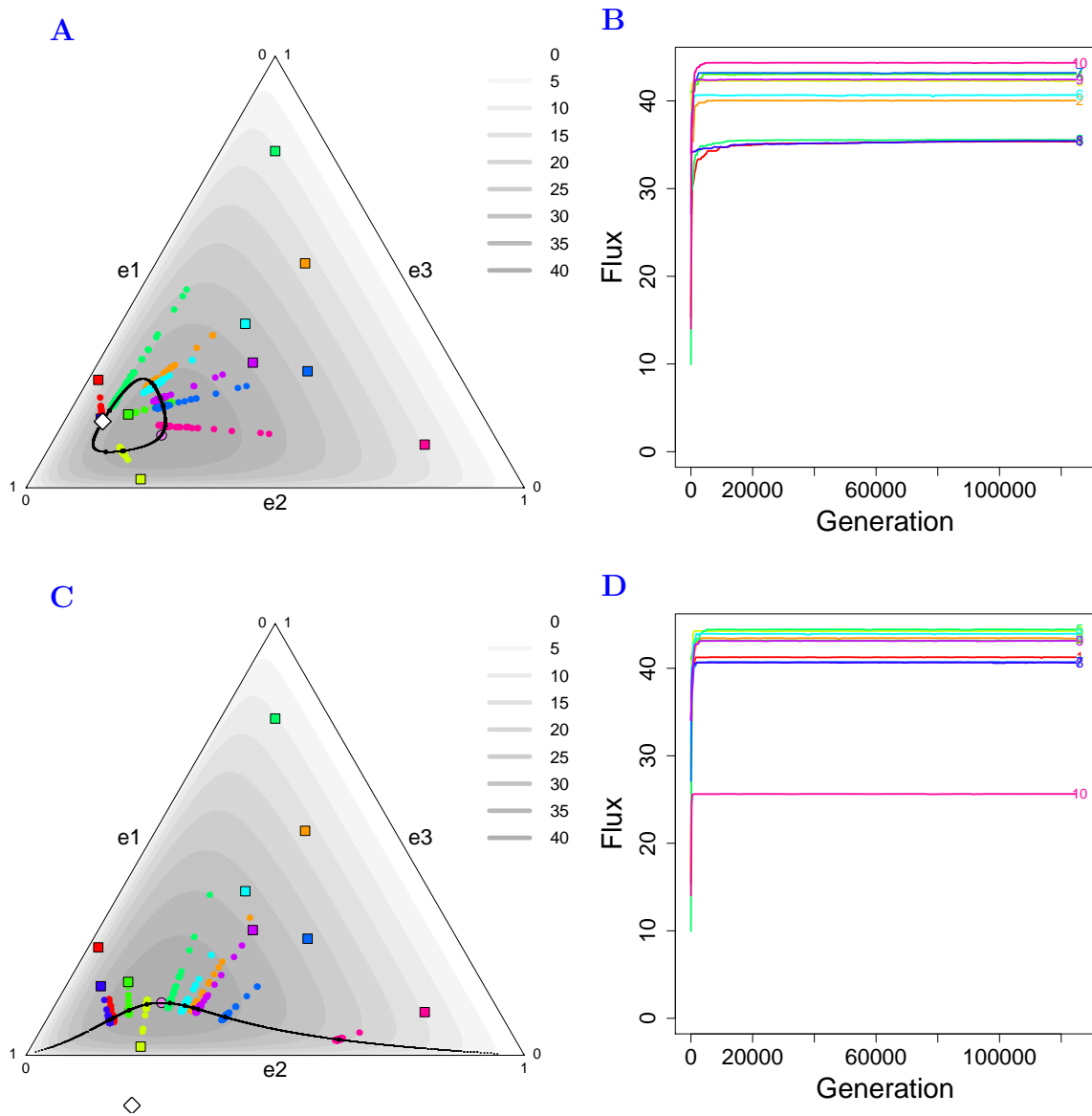


**Figure 3. Simulation of enzyme evolution in a three-enzyme pathway when there is co-regulation.** Representations and symbols are as in Figure 1. (A) and (B) Positive co-regulation. (C) and (D) Negative co-regulation. When there is co-regulation, the evolutionary trajectories in (A) and (C) are linear. In the case of negative co-regulation, the theoretical equilibrium  $\mathbf{e}^*$  (white diamond in (C)) is outside the triangle. The black curve in (C) represents the sets of points of the effective equilibrium  $\tilde{\mathbf{e}}$  determined numerically for a large number of initial concentrations  $\mathbf{e}^0$ . The relative enzyme concentrations fluctuate on both sides of this curve due to the mutation-selection-drift balance. (B) and (D) Evolution of the flux over the course of generations from simulations shown in (A) and (C), respectively. Parameter values are:  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{\text{tot}}^0 = 100$ ,  $N = 1\,000$ ,  $\beta_{12} = 0.1$ ,  $\beta_{23} = 2$  (A and B) or  $\beta_{12} = 0.32$ ,  $\beta_{23} = -0.43$  (C and D).

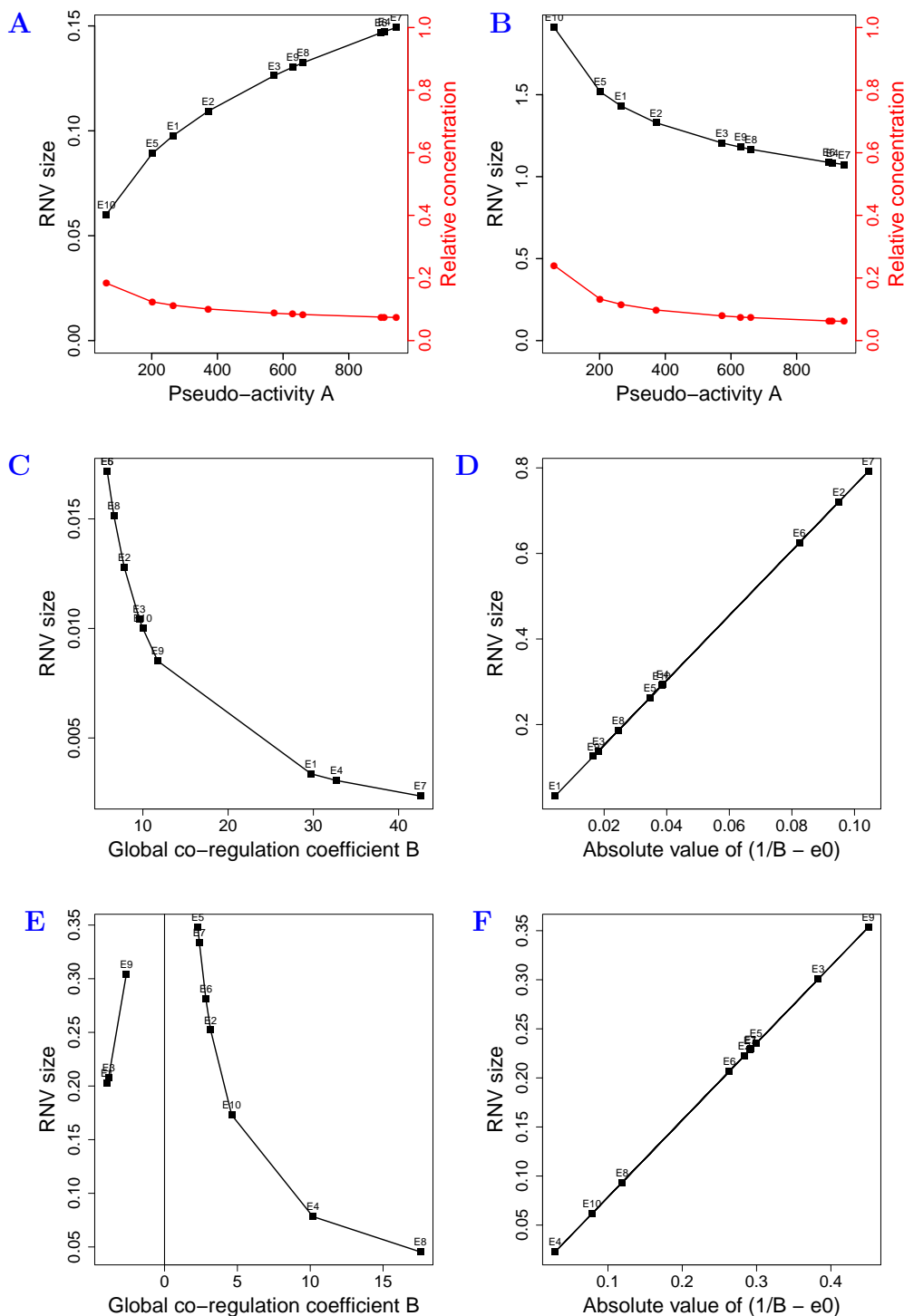


**Figure 4. Relationship between flux and relative enzyme concentrations when there is both competition and positive co-regulation in a three-enzyme pathway.** (A) When there is only competition, the relation between flux and relative concentrations of enzymes 1 and 2 is represented by a dome. The purple point indicates the maximum flux  $J_{\max}$ . The ivory section represents the upright plane drawn from line  $\mathcal{E}$  along which enzyme concentrations vary when there is positive co-regulation (green dashed line on the plane). Red point: flux value at theoretical equilibrium  $e^*$ . Black point: flux value at the initial enzyme concentrations  $e^0$ . Yellow point: maximum flux value  $\tilde{J}_{\max}$  at effective equilibrium  $\tilde{e}$ . (B) Relationship between  $J$  and  $\tau$ , the driving variable of line  $\mathcal{E}$ , corresponding to the ivory section of the dome in (A). Solid vertical line: abscissa of the theoretical equilibrium. Dashed vertical line: abscissa of the effective equilibrium corresponding to  $\tilde{J}_{\max}$ . Colored points: same as in (A). The arrows indicate the variation of flux under selection for different initial enzyme concentrations. Note that the red point (the theoretical equilibrium) is a singular point that cannot be passed if the initial flux is lower than the flux at  $e^*$  and if  $e^*$  and  $e^0$  are on the same side of the curve. Parameter values are:  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_1^0 = E_2^0 = E_3^0 = 100/3$ ,  $\beta_{12} = 2.5$ ,  $\beta_{23} = 0.6$ , corresponding to  $B_1 = 5$ ,  $B_2 = 2$  and  $B_3 = 3.33$ .

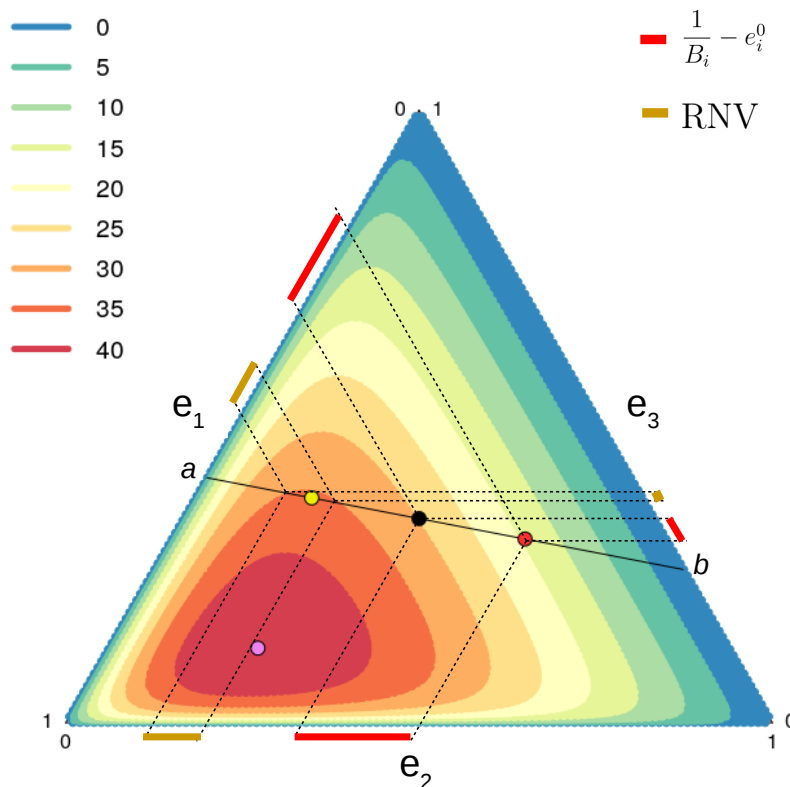




**Figure 5. Simulation of enzyme evolution in a three-enzyme pathway when there is competition and co-regulation.** Representations and symbols are as in Figure 1. (A) and (B) Competition plus positive co-regulation. (C) and (D) Competition plus negative co-regulation. In (A) and (C), the grayscales correspond to the flux values and the black curves represent the sets of points of the effective equilibrium determined numerically for a large number of initial concentrations  $e^0$ . The evolutionary trajectories converge towards, but do not reach, the theoretical equilibrium  $e^*$  (white diamonds). They fluctuate on both sides of the curves due to the mutation-selection-drift balance. (B) and (D) The evolution of the flux over the course of generations from the ten simulations shown in A and C, respectively. Parameter values are:  $X = 1$ ,  $A_1 = 10$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{\text{tot}}^0 = 100$ ,  $N = 1\,000$ ,  $\beta_{12} = 0.1$ ,  $\beta_{23} = 2$  (A and B) or  $\beta_{12} = 0.32$ ,  $\beta_{23} = -0.43$  (C and D).



**Figure 6. Constraint-dependent factors impacting the range of neutral variation (RNV) of enzyme concentrations.** Numerical applications for a ten-enzyme pathway. If enzymes vary independently (A), the RNV size at evolutionary equilibrium is positively related to the pseudo-activities – hence negatively to the relative enzyme concentrations –, whereas the opposite is found when there is competition (B). For a given total enzyme concentration, the RNV size is one order of magnitude greater when there is competition than when enzyme concentrations vary independently. If there is positive (C) or negative (E) co-regulation, RNV size is inversely related to  $|B|$ . When there is competition with positive (D) or negative (F) co-regulation, RNV size is linearly related to  $|1/B - e^0|$ . The pseudo-activities  $\mathbf{A}$  were randomly drawn between 0 and 1 000. Initial concentrations  $\mathbf{E}^0$  were randomly drawn between 0 and 100,  $\mathbf{E}^0$  was then calibrated to have  $E_{\text{tot}}^0 = 100$ . When there is co-regulation, inverses of  $B_i$ 's were randomly drawn between 0 and 100 (and randomly multiplied by +1 or -1 for negative co-regulation), then calibrated to have a sum equal to 1.



**Figure 7. Geometric representation of the RNV when there is both competition and positive co-regulation.** The triangle is a top view of the dome of Figure 4A. The color scale represents the flux levels.  $a$  and  $b$  are the ends of the line  $\mathcal{E}$ . Purple point: top of the dome. Black point: initial enzyme concentrations  $\mathbf{e}^0$ . Yellow point: effective equilibrium  $\tilde{\mathbf{e}}$ . Red point: theoretical equilibrium  $\mathbf{e}^*$ . The red bars outside the triangle represent the difference  $\left(\frac{1}{B_i} - e_i^0\right)$  for each of the three enzymes, and the gold bars represent the range of neutral variation (RNV). Note the positive relationship between  $\left(\frac{1}{B_i} - e_i^0\right)$  and RNV. The parameters values are as in Figure 4.



940 **Appendix A Glossary of mathematical symbols****Latin symbols**

$A_i$	Pseudo-activity of enzyme $i$ (the parameter includes the kinetic properties of the enzyme and thermodynamic equilibrium of upstream reactions)
$\mathbf{A}$	Vector of pseudo-activities
$B_i$	Global co-regulation coefficient, representing effect of a mutation of $E_i$ on $E_{\text{tot}}$ due to co-regulation
$E_i$	Concentration of enzyme $i$
$E_i^0$	Initial concentration of enzyme $i$
$\mathbf{E}$	Vector of enzyme concentrations, considered as the genotype
$\mathbf{E}^0$	Vector of initial enzyme concentrations
$\mathbf{E}^{(1)}$	Vector of enzyme concentrations after a mutation
$\mathbf{E}^{(2)}$	Vector of enzyme concentrations after a mutation and the application of a constraint on $E_{\text{tot}}$
$\mathbf{E}^m$	Vector of enzyme concentrations of the mutant
$\mathbf{E}^r$	Vector of enzyme concentrations of the resident
$\mathbf{E}^{(t)}$	Vector of concentrations at time $t$
$E_{\text{tot}}$	Total enzyme concentration in the metabolic pathway
$E_{\text{tot}}^0$	Initial total enzyme concentration
$e_i$	Relative concentration of enzyme $i$
$e_i^0$	Initial relative concentration of enzyme $i$
$\mathbf{e}$	Vector of relative enzyme concentrations
$\mathbf{e}^0$	Vector of initial relative enzyme concentrations
$\mathbf{e}^*$	Theoretical equilibrium of relative enzyme concentrations
$\tilde{\mathbf{e}}$	Effective equilibrium of relative enzyme concentrations, corresponding to a maximum of the flux
$\mathcal{E}$	Line in the space of relative enzyme concentrations along which the relative concentrations vary in the case of co-regulation
$J$	Metabolic flux, considered as a phenotype that is proportional to fitness
$J_i^m$	Flux resulting from a mutation targeting enzyme $i$
$J^r$	Flux of the resident
$K_i$	Equilibrium constant of reaction $i$ , catalyzed by enzyme $i$
$K_{1,i}$	Product of the equilibrium constants from the first to the $i^{\text{th}}$ reaction: $K_{1,i} = \prod_{k=1}^{k=i} K_k$
$k_{\text{cat}_i}$	Catalytic constant of enzyme $i$
$K_{M_i}$	Michaelis-Menten constant of enzyme $i$ for its substrate
$\mathbf{M}_\beta$	Matrix of co-regulation coefficients
$N$	Effective population size
$n$	Number of enzymes in the metabolic pathway
$P_f$	Fixation probability of a mutation
$R_{E_i}^J$	Flux response coefficient of enzyme $i$
$S_i$	Metabolite $i$ , product of reaction $i$ and substrate of reaction $i + 1$
$s_i$	Selection coefficient of a mutation targeting enzyme $i$
$t$	Evolutionary time
$w$	Fitness
$w^m$	Mutant fitness
$w^r$	Resident fitness
$X$	Constant representing the environment: $X = X_0 - X_n/K_{1,n}$
$X_0$	Concentration of the input substrate of the pathway
$X_n$	Concentration of the output product of the pathway

**Greek symbols**

- $\alpha_{ij}$  Redistribution coefficient, expressing the variation of  $E_j$  due to mutation affecting enzyme  $i$
- $\beta_{ij}$  Co-regulation coefficient of enzyme  $i$  on  $j$
- $\gamma_{ij}$  Competition coefficient, expression of  $\alpha_{ij}$  in the case of competition alone
- $\delta_i$  Actual effect of a mutation affecting enzyme  $i$
- $\mu$  Mutation rate at each time unit
- $\nu$  Canonical effect of a mutation
- $\tau$  Driving variable of the line  $\mathcal{E}$  in the space of relative enzyme concentrations in the case of co-regulation.



## 1.3 Discussion complémentaire

### 1.3.1 Échelles de temps

Dans notre modèle d'évolution, on peut distinguer trois échelles de temps :

1. l'échelle individuelle, correspondant à l'effet des mutations sur le phénotype de l'individu ;
2. l'échelle populationnelle ou écologique, correspondant au temps nécessaire pour la fixation des mutations ;
3. l'échelle évolutive, correspondant à la succession des mutations et à leur fixation.

La figure 1.1 montre la variation des concentrations d'enzymes dans les trois échelles de temps.

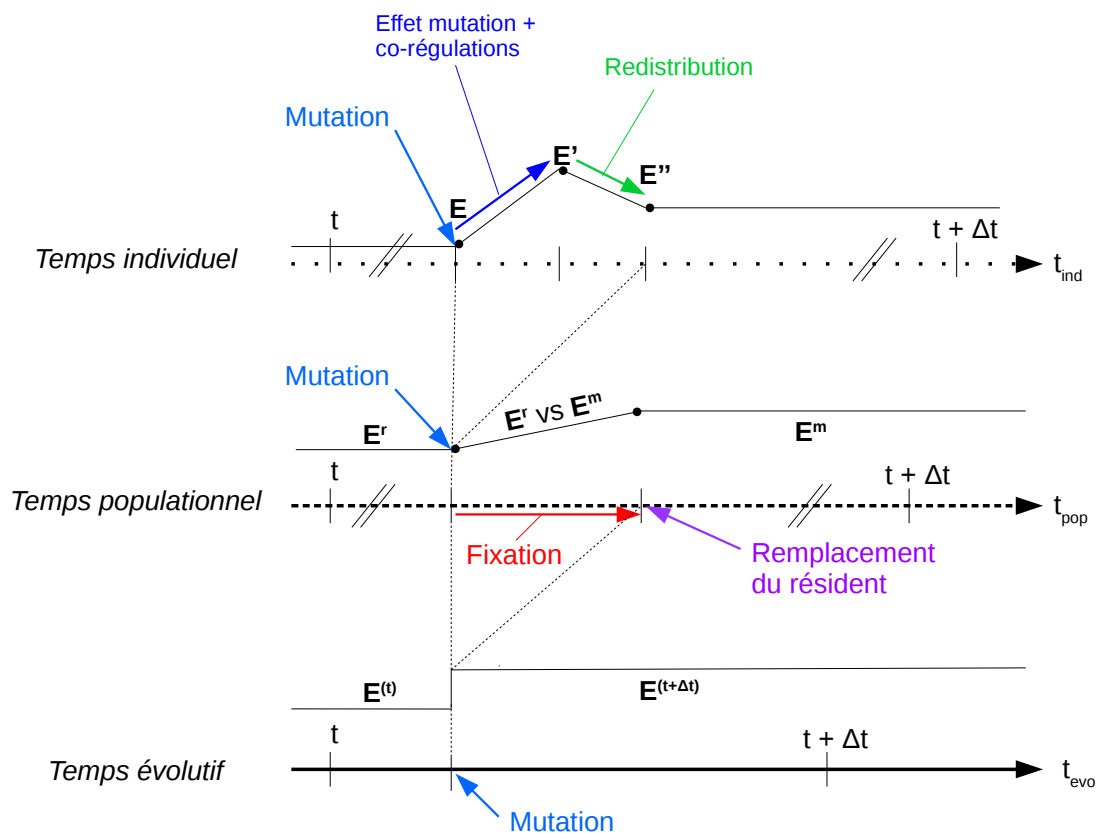


FIGURE 1.1 – **Variation des concentrations d'enzymes dans les trois échelles de temps.** L'échelle individuelle correspond au changement qui s'opère chez l'individu à cause de la mutation (ici les effets des co-régulations et de la compétition sur les concentrations *via* respectivement  $E'$  et  $E''$ ). L'échelle populationnelle correspond au temps de fixation d'une mutation (ici  $E^r$  et  $E^m$  sont respectivement les génotypes du résident et du mutant). Enfin l'échelle évolutive correspond à l'apparition et à la fixation immédiate des mutations, qui est l'échelle que nous regardons pour étudier l'évolution des concentrations.



### 1.3.2 Relation entre les équilibres évolutifs des concentrations relatives et l'état stationnaire du flux

Nous avons montré qu'en cas de co-régulations, il existait deux états stationnaires pour les concentrations relatives d'enzymes, ou équilibres évolutifs : l'équilibre théorique  $e^*$  et l'équilibre effectif  $\tilde{e}$ . L'équilibre effectif correspond à la condition  $R_{E_i}^J = 0$  pour tout  $i$ . En complément du premier article, j'ai étudié en quoi les équilibres évolutifs sont liés à l'état stationnaire du flux.

L'état stationnaire du flux est défini par :

$$\frac{\partial J}{\partial t} = 0 \quad (1.1)$$

En décomposant la dérivée en fonction des concentrations d'enzymes, on a :

$$\frac{\partial J}{\partial t} = \sum_{i=1}^n \frac{\partial J}{\partial E_i} \frac{\partial E_i}{\partial t}$$

En introduisant le coefficient de réponse  $R_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J}$  ainsi que les concentrations relatives  $e_i = E_i/E_{\text{tot}}$ , on obtient la condition suivante pour l'état stationnaire du flux :

$$\frac{\partial J}{\partial t} = \sum_{i=1}^n R_{E_i}^J \frac{J}{E_i} \left( E_{\text{tot}} \frac{\partial e_i}{\partial t} + e_i \frac{\partial E_{\text{tot}}}{\partial t} \right) = 0 \quad (1.2)$$

Les deux solutions possibles sont soit  $\forall i, R_{E_i}^J = 0$ , soit  $\forall i, \frac{\partial e_i}{\partial t} = 0$  et  $\frac{\partial E_{\text{tot}}}{\partial t} = 0$ . On ignore la solution où la somme s'annule par opposition, car cela ne correspondrait à aucun équilibre.

Par définition,  $\frac{\partial e_i}{\partial t}$  est nul quel que soit l'équilibre, théorique ou effectif.

À l'équilibre effectif, le coefficient de réponse est nul, ce qui correspond à la première solution pour l'état stationnaire du flux. En ce point, le flux est maximum. Il est logique que le flux soit à l'état stationnaire quand il est à son maximum, car on ne considère que la sélection dans notre modèle analytique.

À l'équilibre théorique, le coefficient de réponse n'est pas nul, et il faut donc regarder du côté de la deuxième solution ( $\frac{\partial E_{\text{tot}}}{\partial t} = 0$ ). En l'absence de compétition,  $E_{\text{tot}}$  peut augmenter indéfiniment à l'équilibre théorique, et le flux étant linéaire à la concentration totale, il peut également augmenter indéfiniment et n'est donc pas à l'état stationnaire. En cas de compétition et de co-régulation,  $E_{\text{tot}}$  étant constant,  $\frac{\partial E_{\text{tot}}}{\partial t}$  est toujours nul. Cependant, le coefficient de réponse n'est pas défini à l'équilibre théorique à cause de la formule du coefficient de redistribution, ainsi le flux ne peut pas être à l'état stationnaire.

Cette différence est importante, car en étudiant les états stationnaires des concentrations relatives, on pourrait en déduire que le flux est également à l'état stationnaire, mais comme on le voit, ce n'est pas toujours le cas. Avec ces éléments, on pourrait redéfinir les états stationnaires

des concentrations relatives, ou équilibres évolutifs, en tenant compte du comportement du flux :

- l'équilibre théorique  $e^*$  pour lequel le flux n'est pas à l'état stationnaire ;
- l'équilibre effectif  $\bar{e}$  correspondant au maximum du flux, et donc à l'état stationnaire du flux.

Notons que l'équilibre théorique en cas de co-régulation avec compétition est un *evolutionary black hole* : selon les concentrations initiales, le système peut être bloqué à cet équilibre car à l'approche de cet équilibre théorique, les mutations ont de moins en moins d'effet sur le système, l'empêchant d'atteindre le flux maximum et l'équilibre effectif.

### 1.3.3 Taille de population et neutralité sélective

En génétique des populations, la zone de neutralité sélective, ou zone neutre, est la zone où les mutations sont sélectivement neutres ou quasiment neutres. En termes mathématiques, c'est la zone où les mutations ont une probabilité de fixation proche de celle des mutations strictement neutres. La probabilité de fixation d'une mutation dépend du coefficient de sélection  $s$ , qui est la différence relative de fitness du mutant par rapport au génotype initialement présent, et de la taille efficace de population  $N$  :

$$P_f = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \quad (1.3)$$

À  $s$  fixé, si  $N$  est infini, la probabilité de fixation des mutations avantageuses ne dépend que du coefficient de sélection ( $P_f \underset{Ns \rightarrow +\infty}{\approx} 2s$ ), et les mutations délétères ne peuvent se fixer ( $P_f \underset{Ns \rightarrow -\infty}{\rightarrow} 0$ ). À mesure que  $N$  diminue, la probabilité de fixation augmente en raison de la dérive génétique : une mutation, même délétère, a plus de chance d'être tirée dans une petite population que dans une grande, car sa fréquence devient plus élevée. Si  $N = 1$ , la probabilité de fixation est de 1 dans le cas haploïde. Notons que si  $s$  est nul (mutation strictement neutre), la probabilité de fixation est inversement proportionnelle la taille de population. À travers  $N$  et  $s$ , les mutations sont ainsi soumises à une balance sélection-dérive.

La valeur limite de la zone neutre est classiquement définie à  $N|s| = 1$  [Nielsen et Slatkin 2013]. Initialement, on la plaçait à  $2N|s| = 1$  pour une population haploïde et  $4N|s| = 1$  pour une population diploïde [Crow et Kimura 1970], mais cette limite semblait trop restrictive [Li 1978; Kimura 1985]. La limite de la zone neutre est en réalité floue car, comme l'expliquent Hartl *et al.* [1985], la transition entre mutation sélectivement neutre et mutation sous sélection est progressive.

La figure 1.2 montre la probabilité de fixation en fonction de  $s$  pour différentes valeurs de  $N$ . La zone neutre correspond au plateau où les valeurs de probabilité de fixation sont proches de  $1/N$ . La taille de la population a une influence sur la taille de la zone neutre : plus la population est petite, plus la zone neutre est étendue et plus grande est la probabilité de fixer des mutations (délétères comme avantageuses).

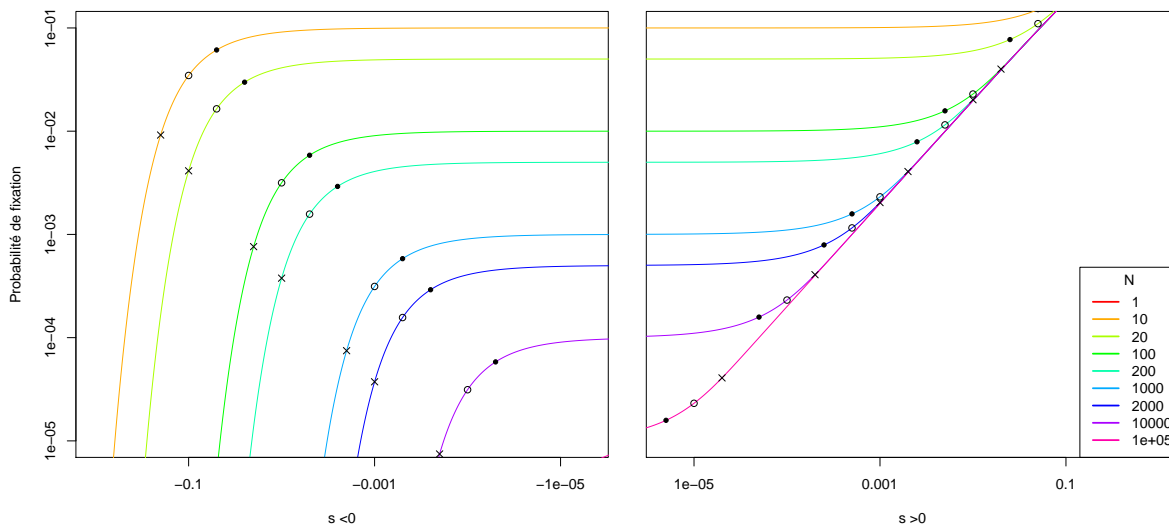


FIGURE 1.2 – **Probabilité de fixation d’une mutation en fonction du coefficient de sélection  $s$  pour différentes tailles de population  $N$ .** La probabilité de fixation est calculée selon la formule 1.3. La figure est représentée en échelles logarithmiques pour les deux axes. Le point  $s = 0$  ne pouvant être placé avec cette échelle, le graphe est divisé en deux parties : coefficient de sélection négatif à gauche et positif à droite. Les courbes peuvent se prolonger, sachant que  $P_f(s = 0) = 1/N$ . Les couleurs correspondent à différentes tailles de population  $N$  (de 10 à 100 000), indiquées en légende en bas à droite. Les croix indiquent les valeurs  $N|s| = 2$ , les cercles  $N|s| = 1$  et les points noirs  $N|s| = 1/2$ . Sur une même courbe, la zone entre les deux cercles est définie comme étant la zone neutre, car la probabilité de fixation est proche de la probabilité de sélection d’une mutation strictement neutre.

Cette différence de taille de la zone neutre se voit bien dans des simulations d’évolution que j’ai réalisées pour une voie à trois enzymes avec compétition, en comparant les résultats pour  $N = 10$  et  $N = 100\,000$  (figure 1.3). Avec les valeurs de paramètres choisies, la valeur du flux à l’équilibre est  $J_{\max} = 40.06397$ . La limite de la zone neutre se calcule grâce à la formule  $N|s| = 1$ , avec  $s = (J^m - J^r)/J^r$ . Comme le flux est maximum à l’équilibre, la zone neutre n’a qu’une limite inférieure, et chercher la limite de la zone neutre à l’équilibre revient à chercher la valeur de flux  $J_{\text{ZN}}$  telle que  $J_{\text{ZN}} = J_{\max}(1 - 1/N)$ . Lorsque  $N = 10$ , la limite est  $J_{\text{ZN}} = 36.05758$ , alors qu’avec  $N = 100\,000$  la limite est  $J_{\text{ZN}} = 40.06357$ , donc très proche de  $J_{\max}$ . La zone neutre est donc beaucoup plus large dans le premier cas que dans le deuxième. En cas d’indépendance entre les enzymes (figure 1.4), les résultats sont qualitativement les mêmes, si ce n’est que les trajectoires se dispersent car le flux n’est pas limité.

La RNV est déterminée par la forme du paysage adaptatif. C’est une projection de la zone où les mutations n’ont quasiment pas d’effet sur le flux, dites mutations quasiment neutres, dans l’espace des concentrations. J’ai analysé les RNV autour de l’équilibre.

La taille de la RNV dépend uniquement des pseudo-activités en cas d’indépendance ou de compétition seule, mais les effets sont inversés. Les enzymes avec une plus forte pseudo-activité ont une grande RNV en cas d’indépendance, mais ont une petite RNV en cas de compétition. Tout comme la différence de puissance dans les équilibres entre indépendance et compétition,

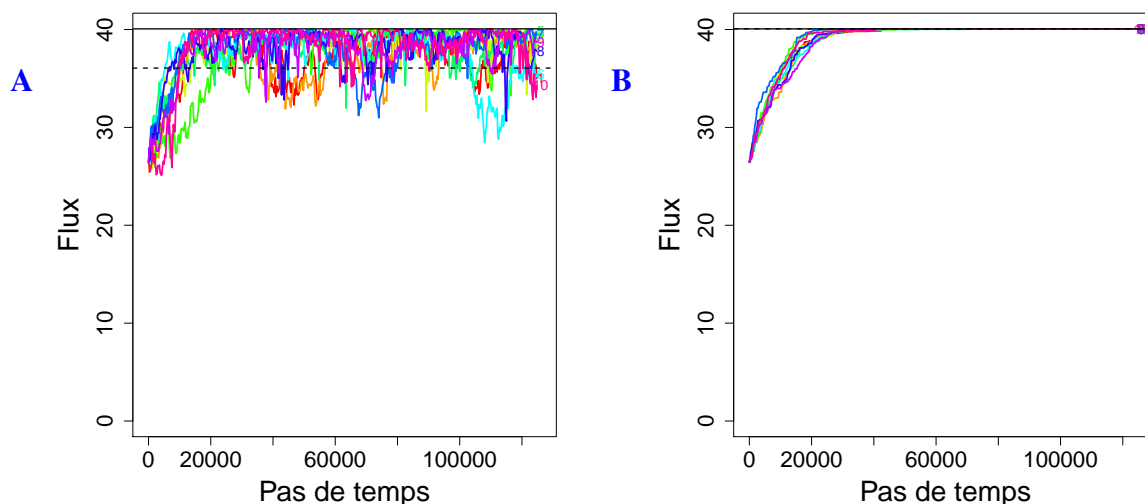


FIGURE 1.3 – **Évolution du flux au cours du temps dans une voie métabolique linéaire à trois enzymes en cas de compétition, pour deux tailles de population.** À gauche, pour  $N = 10$ ; à droite, pour  $N = 100\,000$ . Chaque couleur représente une simulation. La droite horizontale pleine correspond au flux maximum. La droite horizontale pointillé représente la limite de la zone neutre à l'équilibre. La taille des fluctuations, et donc de la zone neutre, est plus importante dans les faibles tailles de population. Paramètres :  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_1^0 = E_2^0 = E_3^0 = 30$ .

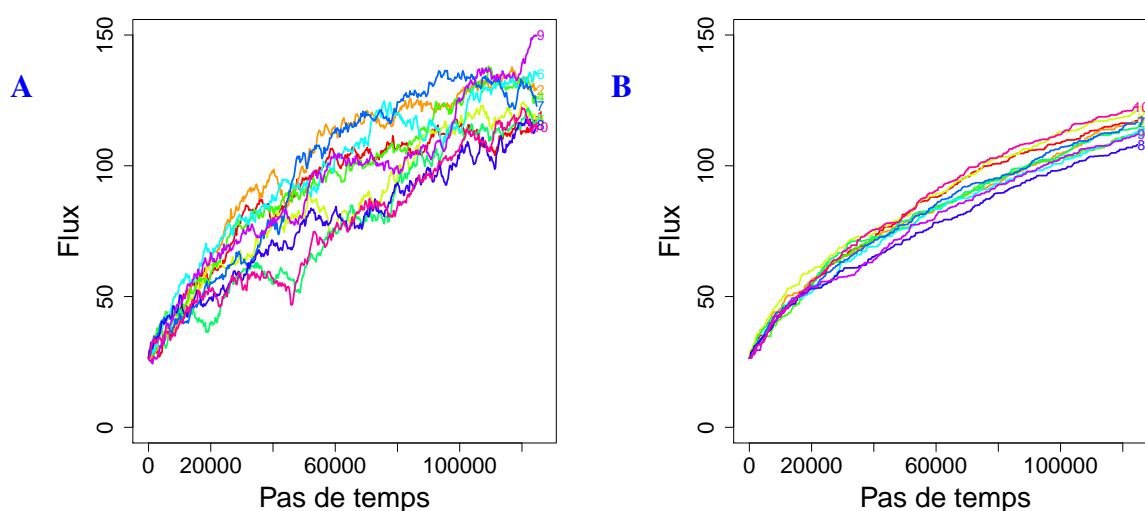


FIGURE 1.4 – **Évolution du flux au cours du temps dans une voie métabolique linéaire à trois enzymes en cas d'indépendance, pour deux tailles de population.** À gauche  $N = 10$ ; à droite  $N = 100\,000$ . Même représentation et mêmes paramètres que dans la figure 1.3. Les trajectoires évolutives sont plus dispersées dans la petite population (flux en fin de simulation compris entre 110 et 150) que dans la grande (flux entre 105 et 125).

nous n'avons trouvé aucune raison intuitive pour expliquer cette différence.

### 1.3.4 Opérons et co-régulations

Un cas particulier de co-régulation est celui des opérons, dont les gènes sont contrôlés par les mêmes promoteurs et facteurs de transcription. Pour deux enzymes, les coefficients de co-régulations  $\beta_{12}$  et  $\beta_{21}$  sont alors égaux à 1, car une modification de la transcription d'un des deux gènes modifie d'autant la transcription de l'autre, et *a fortiori*, la concentrations des deux enzymes. Les coefficients de co-régulations globaux  $B_1$  et  $B_2$ , qui sont la somme des coefficients de co-régulations sont alors égaux à 2. Et les proportions entre les concentrations des deux enzymes, quelles que soient les modifications conjointes en amont, seront toujours égales à  $1/2$ , qui est la valeur  $1/B_1 = 1/B_2$ . La situation correspondrait alors au fameux *black hole*, car à moins d'annuler les co-régulations, et donc de rendre les gènes indépendants, la proportion relative pour chaque enzyme sera toujours égale à  $1/2$ .

### 1.3.5 Infinité de paramètres pour un même paysage adaptatif

En cas de co-régulation, la droite des concentrations relatives est définie par les coefficients de co-régulations globaux  $B$  et les concentrations initiales  $E^0$ . Or une droite possède une infinité de points, et il suffit de deux points pour la décrire. Ainsi il existe une infinité de couple  $(B, E^0)$ , dès lors que leurs proportions sont conservées, permettant de décrire la même droite.

Avec compétition, le paysage adaptatif est une section du dôme par la droite des co-régulations. Ainsi un même paysage adaptatif peut être défini par une infinité de concentrations initiales, et/ou par une infinité de coefficients de co-régulation, positifs ou négatifs. Seul l'équilibre théorique sera modifié avec les coefficients de co-régulation, modifiant la position de l'*evolutionary black hole*. Selon les concentrations initiales, les enzymes peuvent se retrouver bloquées à cet équilibre théorique ou atteindre l'équilibre effectif.

Cependant, les co-régulations peuvent également être la cible des mutations, et donc être soumises à la sélection. Leurs mutations vont modifier la direction de la droite des concentrations, permettant éventuellement les concentrations d'atteindre le sommet du dôme, *i.e.* le flux maximum sans co-régulation. Dans notre modèle, chaque mutation des coefficients de co-régulation reviendrait à modifier l'état initial, qui correspondrait aux valeurs des concentrations juste avant la mutation des coefficients de co-régulation. Il existe alors une infinité de coefficients de co-régulation, positifs ou négatifs, permettant d'atteindre ce flux maximum.

Cela m'impressionne qu'une infinité de possibilités puissent donner le même résultat : cela rend l'évolution d'autant moins prévisible. De plus, ces différentes possibilités n'ont pas toutes les mêmes conséquences lors d'un changement d'environnement. Par exemple, en cas de perturbations environnementales ou de mutations des pseudo-activités, ce qui modifierait la forme du dôme, la section du dôme qui était auparavant identique entre populations sera différente selon les coefficients de co-régulation. Ainsi des populations qui sont identiques du point de vue du

flux dans un environnement donné peuvent devenir différentes et avoir des paysages adaptatifs différents selon l'environnement.



---

## **CHAPITRE 2**

---





## Chapitre 2

### **Article : *Evolution of enzyme levels in metabolic pathways : A theoretical approach. Part 2***

Ce deuxième chapitre présente une extension du modèle d'évolution des concentrations d'enzymes sous sélection pour augmenter le flux, en considérant cette fois que les enzymes d'une chaîne métabolique appartiennent à des groupes co-régulations différents, avec ou sans compétition. Il est basé sur un article compagnon au premier qui sera également soumis à *Journal of Theoretical Biology*, et dont le manuscrit est disponible sur *BioRxiv* [Coton *et al.* 2021a]. Comme pour le premier chapitre, je présente d'abord succinctement les résultats, puis le corps de l'article suivi d'un glossaire des symboles mathématiques. Je discute ensuite d'un point complémentaire qui n'avait pas sa place dans l'article. L'annexe B contient les *Supporting Information* de l'article compagnon, avec les figures supplémentaires et les démonstrations, ainsi que quelques démonstrations complémentaires n'apparaissant pas dans l'article. Les scripts d'analyse dédiés au cas des groupes de co-régulations sont disponibles dans les versions 2.0.0 et plus du package R `SimEvolEnzCons` [Coton *et al.* 2021b].

#### **2.1 Présentation de l'article**

Dans cet article, j'étends le modèle de co-régulation entre concentrations d'enzymes d'une même voie métabolique pour prendre en compte des groupes de co-régulations, avec ou sans compétition pour les ressources. Au sein de chaque groupe, les co-régulations peuvent être positives ou négatives, et les groupes ne contenant que des co-régulations positives sont appelés groupes positifs, et ceux contenant au moins une co-régulation négative sont appelés groupes négatifs. Les enzymes peuvent également être indépendantes des groupes. Comme dans le chapitre précédent, les mutations modifient les concentrations d'enzymes, et la sélection porte sur la maximisation du flux, l'évolution des concentrations des enzymes étant décrite à l'aide du

même système d'équations différentielles.

La recherche des équilibres peut se faire de façon séquentielle, en recherchant d'une part les équilibres pour les enzymes au sein des groupes de co-régulation (équilibres intra-groupes), puis la répartition des concentrations entre les groupes à l'équilibre (équilibres inter-groupes). Dans tous les cas considérés, je démontre, de façon analytique ou numérique, l'existence d'états d'équilibre stable. On peut résumer ainsi les résultats :

1. En l'absence de compétition pour les ressources (concentration totale libre), le type de co-régulation, positif ou négatif, influence fortement les équilibres, ce qui n'est pas le cas en présence de compétition pour les ressources.
2. Lorsqu'il n'y a que des groupes à co-régulations positives et/ou des enzymes indépendantes en absence de compétition, le flux augmente indéfiniment car rien ne limite l'augmentation des concentrations absolues d'enzymes. Par contre, les concentrations relatives atteignent un état d'équilibre qui dépend des pseudo-activités des enzymes et des coefficients de co-régulation. On peut définir pour chaque groupe un paramètre d'activité apparente, qui prend en compte les pseudo-activités des enzymes du groupe, pondérées par les coefficients de co-régulation globaux du groupe. Les équilibres inter-groupes sont inversement proportionnels aux activités apparentes des groupes.
3. Lorsque plusieurs groupes à co-régulation négative co-existent en absence de compétition, il existe un maximum pour le flux car les concentrations d'enzymes d'un même groupe ne peuvent toutes augmenter simultanément à cause d'un effet de balance. Ce flux maximum correspond à l'équilibre effectif, où les concentrations absolues et relatives sont optimales. Cet équilibre dépend des pseudo-activités des enzymes, des coefficients de co-régulation, mais aussi des concentrations initiales. Ainsi, chaque condition initiale, étant donné les pseudo-activités et les coefficients de co-régulation, aboutit à un unique équilibre effectif des proportions relatives et des concentrations absolues d'enzymes. Une fois l'équilibre atteint, les concentrations d'enzymes peuvent varier dans leur zone de neutralité sélective. Chacun de ces équilibres est stable, et il n'est possible de changer d'équilibre que par la modification des pseudo-activités et/ou des coefficients de co-régulation.
4. Lorsque des groupes à co-régulations positives et/ou des enzymes indépendantes co-existent avec des groupes à co-régulations négatives en l'absence de compétition, le flux est limité comme précédemment à cause des co-régulations négatives. Le flux maximal correspond à une certaine valeur optimale des concentrations absolues de ces groupes négatifs, qui dépend des concentrations initiales. Au sein des groupes, les équilibres définis précédemment sont atteints. En théorie, la concentration absolue des enzymes indépendantes ou au sein de groupes positifs peut augmenter indéfiniment, dans des proportions constantes, alors que la concentration absolue des enzymes des groupes négatifs atteint

une valeur finie optimale. En réalité, les simulations montrent que si les concentrations des enzymes des groupes négatifs oscillent bien autour de leur valeur d'équilibre, celles des enzymes des groupes positifs ou indépendantes tendent vers un plateau. La raison en est qu'au fur et à mesure que le flux s'approche du maximum, la pression de sélection sur ces enzymes diminue progressivement, ce qui conduit à un état quasi-stationnaire. Dans cet état quasi-stationnaire, la proportion relative des groupes positifs est plus importante que celle des groupes négatifs.

5. Lorsque que la concentration totale d'enzymes est fixée (compétition pour les ressources), le flux est borné par une valeur qui correspond au flux maximum en l'absence de co-régulations et qui ne dépend que des pseudo-activités (chapitre 1). En présence de co-régulations, qu'elles soient positives et/ou négatives, les concentrations relatives atteignent un équilibre effectif correspondant à un flux maximum différent, qui dépend des pseudo-activités, des coefficients de co-régulations et des concentrations initiales. Lorsque l'équilibre est atteint, les concentrations des enzymes et le flux fluctuent au sein de la zone de neutralité sélective. À noter cependant que les valeurs d'équilibre n'ont pas pu être trouvées analytiquement en cas de co-existence de plusieurs groupes de co-régulation, et que les conclusions ne reposent que sur les simulations.

Globalement, la prise en compte d'un niveau de complexité supplémentaire, en introduisant des groupes d'enzymes co-régulées, enrichit les résultats du chapitre précédent.

Tous les équilibres trouvés sont cohérents avec ceux du chapitre précédent, où toutes les enzymes sont soumises aux mêmes contraintes. Il existe toujours un état d'équilibre stable pour les concentrations relatives des enzymes. Une fois l'état d'équilibre atteint, les concentrations varient au sein d'une zone neutre (*Range of Neutral Variation*) dont l'ampleur dépend des situations et des enzymes considérées.

Les enzymes co-régulées évoluent conjointement selon une trajectoire linéaire. Une mutation affectant l'abondance d'une enzyme change l'abondance de toutes les enzymes du groupe de co-régulation dans des proportions fixées par les coefficients de co-régulation. Cette contrainte importante imposée par la co-régulation réduit fortement le nombre de degrés de liberté du système. En l'occurrence plus les groupes de co-régulation contiennent d'enzymes, c'est-à-dire que plus il y a de contraintes entre les enzymes, et plus le nombre de dimensions indépendantes du paysage adaptatif diminue, limitant le champ des possibles.

L'introduction d'un *trade-off* pour l'abondance des protéines sous la forme de co-régulations négatives et/ou d'une quantité totale fixée de ressources protéiques conduit à une valeur de flux maximale déterminée par une distribution particulière des concentrations de toutes les enzymes de la chaîne. En présence de co-régulations, ce *trade-off* aboutit à un état d'équilibre dépendant des concentrations initiales. En réalité, le simple fait d'imposer une contrainte limitant la concentration d'une enzyme dans une chaîne linéaire suffit à limiter le flux, diminuant progressivement la pression de sélection sur les autres concentrations d'enzymes. Ainsi, augmenter la

concentration des autres enzymes tant que cette contrainte ne sera pas levée limite les possibilités tant du point de vue de l'évolution qu'en ingénierie métabolique.

Si l'on considère l'ensemble des concentrations initiales possibles en cas de compétition et de groupes de co-régulation, on observe que la sélection aboutit à des patrons entre les différents équilibres effectifs des concentrations relatives. Ces patrons dépendent du système considéré (nombre d'enzymes, nombre et nature des groupes de co-régulations), et sont le plus souvent non linéaires.

# Evolution of enzyme levels in metabolic pathways: A theoretical approach. Part 2

Charlotte Coton, Christine Dillmann, Dominique de Vienne

September 30, 2021

## 5 Abstract

Metabolism is essential for cell function and adaptation. Because of their central role in metabolism, kinetic parameters and enzyme concentrations are under constant selective pressure to adapt the fluxes of the metabolic networks to the needs of the organism. In the line of various studies dealing with enzyme evolution, we recently developed a model  
10 of evolution of enzyme concentrations under selection for increased flux, considered as a proxy of fitness (Coton et al. 2021). Taking into account two realistic cellular constraints, competition for resources and co-regulations, we determined the evolutionary equilibria and the ranges of neutral variations of enzyme concentrations. In this article, we give more generality to this model, by considering that the enzymes of a pathway can be  
15 long to different groups of co-regulation. We determined the equilibria and showed that the constraints modify the adaptive landscape by limiting the number of independent dimensions. We also showed that any fixed enzyme concentration limits the flux and relaxes the selection for increasing other enzyme concentrations. Even though the model remains rather simple, the complex relationship between enzyme concentrations prevent  
20 the analysis of selective neutrality.

## Running title

Enzyme evolution in the case of co-regulation groups

# 1 Introduction

The key role of metabolism in cell functioning and adaptation accounts for the large number of studies devoted to its evolution. The main lines of research concern the evolution of metabolic network structure and properties (Noda-Garcia et al. 2018; Newton et al. 2018; Sambamoorthy et al. 2019; Morrison and Badyaev 2017), the polymorphism and rate of evolution of enzyme coding genes (Larracuenta et al. 2008; Rausher et al. 1999; Colombo et al. 2014; Flowers et al. 2007; Sellis and Longo 2015) and the evolution of enzyme properties (Eanes 1999; Kacser and Beeby 1984; Serohijos and Shakhnovich 2014; Bershtein et al. 2017; Heckmann et al. 2018). In this context, it is striking to observe that very few studies have addressed the question of the evolution of enzyme concentrations in response to selection on the metabolic flux, considered to be a proxy of fitness. Yet, quantitative proteomics in various species has revealed the extent of the genetic variability of enzyme concentrations (Damerval et al. 1994; Blein-Nicolas et al. 2013; Chick et al. 2016; Jiang et al. 2019).

In a classical paper, Hartl et al. (1985) analysed the evolution of enzyme quantity, equated to activity, under selection for higher metabolic flux. Relying on the metabolic control theory (MCT) (Kacser and Burns 1973; Heinrich and Rapoport 1974), they proposed an explanation for the selective neutrality of most of molecular polymorphisms. Under selection, the mutations that increase the flux are progressively fixed. But due to the concavity of the enzyme-flux relationship, the selection coefficients of the mutations become smaller and smaller. At the plateau of the curve, enzyme variations become nearly-neutral.

This interesting model relied on two quite restrictive assumptions: the mutations always affect the same enzyme of the pathway, and there is no limitation to enzyme concentrations, in spite of the well-documented constraints on energy and space in the cell (Kurland and Dong 1996; Ellis 2001; Klumpp et al. 2019). In addition this model has never been tested following an adaptive dynamics approach.

In a companion paper, we recently filled these gaps by developing the first adaptive dynamics model of the evolution of enzyme concentrations in linear pathways (Coton et al. 2021). We considered jointly the evolution of all enzymes, because the concentration of any enzyme of the pathway can be targeted by mutation, and we introduced in the model possible competition for cellular resources and/or co-regulations, which prevents enzyme concentrations from varying freely (Heinrich et al. 1991; Lion et al. 2004).

We wrote a system of differential equations that describes the evolution of enzyme concentrations and we analysed the enzyme properties and constraint-specific parameters that shape the adaptive landscape. We showed that relative enzyme concentrations reach evolutionary equilibria that depend on the kinetic parameters of enzymes and thermodynamic constants, and which are modified by competition and/or co-regulation. We also analysed the evolution of the flux upon selection and determined its limit value in the case of competition and/or negative co-regulation. The values of the diverse equilibria and of the flux, predicted analytically or numerically, were confirmed by computer simulations of long term evolution. Finally we showed that the selective neutrality of all enzymes is achieved at equilibrium whatever the constraints, but the range of neutral variation of concentrations vary in a large extent, depending on constraint-specific parameters. In the case of co-regulations, initial enzyme concentrations affect the evolutionary outcome.

In the previous developments, we considered as a first analysis that the enzymes of a pathway are either independent or all co-regulated to each other, with or without competition. However, enzyme concentrations of a given pathway can be differently controlled.

For example, in *Saccharomyces cerevisiae* many transcription factors (TFs) control the expression of glycolytic genes, in different ways depending on the promoters (Chambers et al. 1995). In the pathogenic yeast *Candida albicans*, two major TFs, Gal4p and Tye7p, activate the genes of the glycolytic pathway: both control diverse enzymes, but upstream enzyme genes are mainly activated by Tye7p (Askew et al. 2009). In *Escherichia coli*, only about 30 % of sequential enzymes in pathways are both regulated, including 75 % (down to 8 % for enzymes at junction) that share the same TF, *i.e.* are co-regulated (Seshasayee et al. 2009). Also in *Escherichia coli*, Goelzer et al. (2008) linked the different metabolic pathways to the regulatory networks, and shown that enzymes can be controlled by different TFs (activators or repressors), which can be in common with other enzymes, or not regulated at all.

In this work, we propose a generalization of the enzyme evolution model of Coton et al. (2021). We take into account the realistic situation where different groups of co-regulated enzymes coexist in a pathway, with or without competition between all enzymes for cellular resources. To search the equilibria we decompose the relative concentrations into intra- and inter-group relative concentrations. We found the evolutionary equilibria when there is no competition, which differ according to the type of co-regulation – positive or negative – within the groups. When there is competition, we did not find the evolutionary equilibria analytically, but the simulations show that they exist. Finally, we determine that mutations become neutral around the equilibrium, but the complex relationships between enzyme concentrations prevent us to compute the range of neutral variations. We show that increasing the number of co-regulations between enzymes diminishes the degrees of freedom of the system, that results in a smaller adaptive landscape. Particularly two negatively co-regulated enzymes are sufficient to limit the flux, which results in a progressively decrease of the selective pressure on positively co-regulated enzymes. Moreover the presence of co-regulation induces a dependence on initial concentrations.

## 2 Material & Methods

The modelling approach and simulation methods are developed in Coton et al. (2021). Hereafter we present them briefly and we complement the theoretical background to treat the general case where the enzymes of the pathway belong to different groups of co-regulation. The glossary of mathematical symbols is in appendix A.

### 2.1 Theoretical background

#### 2.1.1 Flux expression

In a linear metabolic pathway of  $n$  unsaturated Michaelian-Menten enzymes, the flux at the steady state is (Kacser and Burns 1981):

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}} \quad (1)$$

where:

- $X$  is a constant.
- $A_j$  is called *pseudo-activity* of enzyme  $j$ .  $A_j$  is a composite parameter which includes the catalytic constant  $k_{\text{cat},j}$  and the Michaelis-Menten constant  $K_{M,j}$  of enzyme  $j$ ,



and the product  $K_{1,j-1}$  of the equilibrium constants of the reactions upstream enzyme  $j$ :

$$A_j = \frac{k_{\text{cat}j}}{K_{Mj}} K_{1,j-1}$$

- $E_j$  is the concentration of enzyme  $j$ .

## 2.1.2 Formalizing the dependence between enzymes

### 110 *The redistribution coefficients*

Constraints on enzyme concentrations, such as co-regulations or competition for resources, create interdependence between enzymes. As a consequence, a mutation that changes the concentration of a target enzyme indirectly changes the concentration of the other enzymes. We model this effect by the following relations. For a target enzyme  $i$ ,  
115 we have

$$E_i \xrightarrow[\text{mutation of } E_i]{} E_i^m = E_i + \delta_i$$

where  $\delta_i$  is the actual effect of the mutation targeting enzyme  $i$ . For the other enzymes  $j \neq i$ , we have

$$E_j \xrightarrow[\text{mutation of } E_i]{} E_j^m = E_j + \alpha_{ij}\delta_i$$

where  $\alpha_{ij}$  is the *redistribution coefficient*, the value of which depends on the type of constraint applied (Lion et al. 2004; Coton et al. 2021):

- 120 1. Without any constraint,  $\alpha_{ij}$  is null.
2. In the case of co-regulation only,  $\alpha_{ij}$  is equal to  $\beta_{ij}$ , the *co-regulation coefficient*, which expresses the change in concentration of enzyme  $j$  due to co-regulation when concentration of enzyme  $i$  varies. We assumed linearity between co-regulated enzymes, so we posed:  $\beta_{ii} = 1$ ,  $\beta_{ij} = 1/\beta_{ji}$  and, for any triplet of co-regulated enzymes  
125  $(i, j, k)$ ,  $\beta_{ij} = \beta_{ik}\beta_{kj}$ . If enzymes  $i$  and  $j$  are not co-regulated, we have  $\beta_{ij} = \beta_{ji} = 0$ . All pairwise co-regulation coefficients can be grouped in a co-regulation matrix, noted  $\mathbf{M}_\beta$ . The *global co-regulation coefficient*  $B_i = \sum_{j=1}^n \beta_{ij} = \frac{\Delta E_{\text{tot}}}{\Delta E_i}$ , quantifies the impact due to co-regulation of a variation of  $E_i$  on the total enzyme concentration  $E_{\text{tot}}$ , where  $E_{\text{tot}} = \sum_{j=1}^n E_j$ .
- 130 3. In the case of competition,  $E_{\text{tot}}$  is fixed and  $\alpha_{ij} = -\frac{e_j}{1 - e_i}$ , where  $e_i$  (resp.  $e_j$ ) is the relative enzyme concentration:  $e_i = \frac{E_i}{E_{\text{tot}}}$  (resp.  $e_j = \frac{E_j}{E_{\text{tot}}}$ ).
4. In the case of co-regulation *and* competition,  $\alpha_{ij} = \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i}$ .

Note that  $\forall i, \alpha_{ii} = 1$ .

135

### The flux response coefficient

When there are dependencies between enzyme concentrations in a pathway, the classical flux control coefficient (Kacser et al. 1995) cannot be used to quantify the effect on the flux of the variation of an enzyme concentration. So we used the flux response coefficient (de Vienne et al. 2001; Coton et al. 2021) that includes not only the parameters of all enzymes, but also the redistribution coefficients  $\alpha_{ij}$ :

$$\forall i, \quad R_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J} = E_i \frac{\sum_{j=1}^n \frac{\alpha_{ij}}{A_j E_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}} \quad (2)$$

## 2.2 Co-regulation groups

In Coton et al. (2021), we assumed that in the case of co-regulation, *all* enzymes of the pathway are co-regulated, *i.e.*  $\beta_{ij} \neq 0$  for all  $i \neq j$ . Here we consider the more realistic situation where there are different groups of co-regulated enzymes.

Let  $\Phi_q$  the co-regulation group  $q$ , which includes  $m_q$  enzymes. With  $p$  co-regulation groups, we have  $\sum_{q=1}^p m_q = n$ , with  $n$  the total number of enzymes. Within a group the enzymes are co-regulated, and are independent from all enzymes outside the group. This means that for all enzyme  $i \in \Phi_q$ ,  $\beta_{ij} \neq 0$  if  $j \in \Phi_q$  and  $\beta_{ij} = 0 = \beta_{ji}$  if  $j \notin \Phi_q$ . Thus some  $\beta_{ij}$  are null in the co-regulation matrix  $\mathbf{M}_\beta$ .

We defined three possible types of co-regulation groups, denoted by the indicator variable  $\theta_q$ :

- (i) if there are only positive co-regulations,  $\theta_q = +1$  and  $\Phi_q$  is called a *positive group*;
- (ii) if there is at least one negative co-regulation in the group,  $\theta_q = -1$  and  $\Phi_q$  is called a *negative group*;
- (iii) if there is only one enzyme in  $\Phi_q$ ,  $\theta_q = 0$  and  $\Phi_q$  is called a *singleton* (the enzyme is independent from all other enzymes).

Figure 1 shows a ten-enzyme pathway with three co-regulation groups and two singletons. Note that the situations analysed by Coton et al. (2021) are particular cases of this general co-regulation model: co-regulation between *all* enzymes corresponds to  $p = 1$  and  $m_1 = n$ , and independence between all enzymes to  $p = n$  and  $m_q = 1$  for all  $q$ .

### 2.2.1 Partitioning the relative enzyme concentrations

When they are different co-regulation groups, and possibly singletons, the relative concentration of an enzyme can refer either to the sum of the enzyme concentrations of its group  $\Phi_q$ :

$$e_i^q = E_i / E^q$$

where  $E^q = \sum_{j \in \Phi_q} E_j$ , or to the total enzyme concentration in the pathway:

$$e_i = E_i / E_{\text{tot}}$$

Hereafter,  $e_i^q$  will be called the *intra-group relative concentration* and  $e_i$  will be called the *total relative concentration*.

The sum of enzyme concentrations in  $\Phi_q$  relative to total enzyme concentration is:

$$e^q = E^q / E_{\text{tot}}$$

165 which will be called the *inter-group relative concentration*.

Therefore there is the following relationship between the three relative concentrations:

$$e_i = \frac{E_i}{E_{\text{tot}}} = \frac{E_i}{E^q} \frac{E^q}{E_{\text{tot}}} = e_i^q e^q \quad (3)$$

### 2.2.2 Global co-regulation coefficient

170 Because  $\beta_{ij} \neq 0$  for enzymes of a given co-regulation group and  $\beta_{ij} = 0$  for enzymes belonging to different co-regulation groups, the co-regulation matrix  $\mathbf{M}_\beta$  can be separated into  $p$  submatrices, as shown in the example of table 3. This makes it possible to specify the expression of the global co-regulation coefficient:

$$\forall i \in \Phi_q \quad B_i = \frac{\Delta E_{\text{tot}}}{\Delta E_i} = \sum_{j \in \Phi_q} \beta_{ij} = \frac{\Delta E^q}{\Delta E_i} \quad (4)$$

175 Thus, when there are different co-regulation groups,  $B_i$  quantifies also the impact due to co-regulations of a variation of  $E_i$  on the sum of enzyme concentrations of its group, which is the same as the impact on the total concentration.

If and only if  $i$  and  $j \in \Phi_q$ , we have the relationship:

$$\beta_{ij} = \frac{B_i}{B_j} \quad (5)$$

Note that  $\sum_{i \in \Phi_q} 1/B_i = 1$  and  $\sum_{i=1}^n 1/B_i = p$ .

### 2.2.3 Parametric equation of the line of co-regulated enzyme concentrations

180 For a group  $\Phi_q$  of co-regulated enzymes, the joint variation of concentrations reduces to one the number of degrees of freedom in this group. Therefore the intra-group relative concentrations move on a straight line in the space of intra-group relative concentrations. The parametric equation of this line, noted  $\mathcal{E}^q$ , is (Coton et al. 2021):

$$\mathcal{E}^q : \quad \exists \tau^q \text{ such as } \forall i \in \Phi_q, \quad e_i^q = \tau^q \left( \frac{1}{B_i} - e_i^{q0} \right) + e_i^{q0} \quad (6)$$

where  $\tau^q$  is the driving variable and  $e_i^{q0}$  is the initial intra-group relative concentration of enzyme  $i$  in  $\Phi_q$ .

## 185 2.3 Evolution model

We studied the evolution of enzyme concentrations under selection for increased flux, assuming that the flux is proportional to fitness. We used the model developed by Coton et al. (2021), which is a differential equation system describing the variation of enzyme concentrations on a continuous timescale:

$$\frac{\partial E_i}{\partial t} = 2\mu \sum_{j=1}^n s_j \alpha_{ji} \delta_j \quad (7)$$

190 where:

- $\mu$  is the mutation rate
- $s_j = \frac{J_j^m - J^r}{J^r}$  is the selection coefficient of a mutation targeting enzyme  $j$ .  $J_j^m$  and  $J^r$  stand for the flux of the mutant and the flux of the resident, respectively.
- $\delta_j$  is the actual effect of the mutation targeting enzyme  $j$
- $\alpha_{ji}$  is the redistribution coefficient measuring the impact on  $E_i$  of a mutation targeting enzyme  $j$ .

We used this differential system to study the evolution of enzyme concentrations when there are different groups of co-regulation in the pathway. We considered two cases, with and without competition for resources between all enzymes (respectively fixed  $E_{\text{tot}}$  and free  $E_{\text{tot}}$ ).

In both cases we searched for the evolutionary equilibria of relative enzyme concentrations, defined as (Coton et al. 2021):

$$\forall i \quad \frac{\partial e_i}{\partial t} = 0 \quad (8)$$

and we analysed the evolution of the flux over the generations of selection.

## 2.4 Computer simulations

The simulation method is the same as in Coton et al. (2021), albeit with different parameters. Briefly, we performed Markov Chain Monte-Carlo simulations using the R software (R Core Team 2017). We considered haploid populations of size  $N$  evolving in a constant environment ( $X = 1$ ). At a given time step, all individuals in the population are identical, *i.e.* have the same phenotype  $J$  and the same enzyme concentrations (the "genotype")  $\mathbf{E} = (E_1, E_2, \dots, E_n)$ . Simulations were performed for pathways of different lengths ( $n = 3$  to 10), with constant pseudo-activities  $\mathbf{A} = (A_1, A_2, \dots, A_n)$  and with co-regulation matrices  $\mathbf{M}_\beta$  depending on the co-regulation structure of the pathway.

The vector of initial concentrations  $\mathbf{E}^0 = (E_1^0, E_2^0, \dots, E_n^0)$  changed at each simulation. Initial concentrations  $E_i^0$  were drawn in a uniform law  $\mathcal{U}(0, 100)$ , and redistributed proportionally to have  $E_{\text{tot}}^0 = 100$ .  $E_{\text{tot}}$  varied over time when there were no competition and was fixed at  $E_{\text{tot}} = 100$  in the case of competition.

At each time step, a mutation targets randomly one of the enzymes and affects other enzymes according to the type of dependence considered. The values of the flux, of the selection coefficient and of the fixation probability of the mutant are computed. The mutation is fixed or not according to its fixation probability. Usually, the populations evolved over 125,000 time steps, and the system values were recorded every 250 steps.

These simulations allowed us to relax some of the assumptions of the mathematical analysis, namely small mutation effects, large population size and the absence of fixation of deleterious mutations.

Simulations are performed by using function `simul.evol.enz.multiple`.

**Data and code availability** Scripts of the simulations and custom functions, written in the R language (R Core Team 2017), are available in a version 2 of the R package called `SimEvolEnzCons` (<https://CRAN.R-project.org/package=SimEvolEnzCons>).

### 3 Results

230 In a previous paper, we developed a model of joint evolution of enzyme concentrations under selection for increased flux under two cellular constraints, co-regulations between all enzyme concentrations and/or competition for resources (Coton et al. 2021).

Here, we generalize the model to the realistic case where different groups of co-regulation coexist in a pathway. Within a group, the co-regulations can be positive (positive groups), or both positive and negative (negative groups). A group that contains only one enzyme is called a *singleton*. An enzyme in a singleton is independent from all other enzymes.

We searched for the evolutionary equilibrium of relative enzyme concentrations of the pathway, such as:

$$\forall i, \frac{\partial e_i}{\partial t} = 0 \quad (9)$$

240 where  $e_i = E_i/E_{\text{tot}}$  is the total relative concentration, with  $E_i$  the concentration of enzyme  $i$  and  $E_{\text{tot}} = \sum_{j=1}^n E_j$  the total enzyme concentration.

We decomposed  $e_i$  into two parts: (i) the intra-group relative concentration  $e_i^q = E_i/E^q$ , where  $E^q$  is the sum of enzyme concentrations in the co-regulation group  $\Phi_q$ ; (ii) the inter-group relative concentration  $e^q = E^q/E_{\text{tot}}$ . These relative concentrations are linked by the relationship (see Material and Methods):

$$e_i = e_i^q e^q$$

Taking account of these relative concentrations, equation 9 becomes:

$$\forall j, \frac{\partial e_i}{\partial t} = \frac{\partial e_i^q}{\partial t} e^q + e_i^q \frac{\partial e^q}{\partial t} = 0 \quad (10)$$

Therefore, solving  $\frac{\partial e_i}{\partial t} = 0$  is equivalent to solve:

$$\forall j, \frac{\partial e_i^q}{\partial t} = 0 \quad \& \quad \frac{\partial e^q}{\partial t} = 0$$

245 We ignore the other solution when  $\partial e_i^q/\partial t$  and  $\partial e^q/\partial t$  are of opposite sign, because the equilibrium for total relative concentrations would probably be unstable.

Thus, the equilibrium of total relative enzyme concentrations implies both inter- and intra-group equilibria, that can be found separately.

250 Hereafter, we show how to find these equilibria and we analyse the flux variation over time. We considered two cases: without and with competition between all enzymes. Simulations of long-term evolution allowed us to test the validity of the analytical results.

#### 3.1 Evolutionary equilibria without competition

The evolutionary equilibria within groups depend on the type of co-regulation between enzymes. We will consider positive groups, negative groups or both kind of groups in the same pathway.

255 **3.1.1 Positive groups and singletons**

**3.1.1.1 Intra-group equilibrium** The intra-group evolutionary equilibrium is defined by

$$\forall j, \frac{\partial e_i^q}{\partial t} = 0$$

When the enzymes are all positively co-regulated, the solution of this equation is (Supporting Information B.3.1.2):

$$\forall i \in \Phi_q, e_i^{q*} = \frac{1}{B_i} \tag{11}$$

where  $B_i$  is the *global co-regulation coefficient*, which quantifies the impact due to co-regulation of a variation of  $E_i$  on the sum of enzyme concentrations of group  $\Phi_q$ .

260 This *theoretical equilibrium*  $e_i^{q*}$  is achievable since  $B_i$  is positive for all  $i$ . If the group contains only one enzyme (*singleton*), we have  $e_i^{q*} = 1 = 1/B_i$  because  $E^q = E_i$  and  $\beta_{ii} = 1$ .

**3.1.1.2 Inter-group equilibrium** The inter-group evolutionary equilibrium is defined by

$$\forall q, \frac{\partial e^q}{\partial t} = 0$$

In Supporting Information B.3.2.2, we show that when all groups are positive or singletons, the solution of this equation is:

$$\forall q, e^{q*} = \frac{(A^{q*})^{-1/3}}{\sum_{k=1}^p (A^{k*})^{-1/3}} \tag{12}$$

265 where  $A^{q*}$ , the *apparent activity of co-regulation group*  $\Phi_q$ , is a composite parameter defined such as:

$$\frac{1}{A^{q*}} = \sum_{i \in \Phi_q} \sum_{j \in \Phi_q} \frac{1}{A_j} B_j B_i^2 \tag{13}$$

We see that the apparent activity of a group depends on the pseudo-activities of every enzyme in the group, weighed by the global co-regulation coefficients of this group. At the theoretical equilibrium, the inter-group relative enzyme concentrations  $e^{q*}$  are inversely related to the  $A^{q*}$ .

270

**3.1.1.3 Total equilibrium** From equations 11 and 12, and using the relationship between relative concentrations  $e_i = e_i^q e^q$  (equation 3), we obtain the theoretical equilibrium of the total relative concentrations when there are only positive groups or singletons:

$$\forall q, \forall i \in \Phi_q, e_i^* = \frac{1}{B_i} \frac{(A^{q*})^{-1/3}}{\sum_{k=1}^p (A^{k*})^{-1/3}} \tag{14}$$

Note that if all enzymes are independent, we have  $B_i = 1$  for all  $i$ , and we find the solution  $e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}}$  for all  $i$ ; and if *all* enzymes are co-regulated, we have  $p = 1$ , and

275

we find the solution  $e_i^* = 1/B_i$  for all  $i$ . This is consistent with the results of Coton et al. (2021).

In the case of independence or positive co-regulation between enzymes, there is no limit to the increase of the flux over time.

### 3.1.2 Negative groups

**3.1.2.1 Intra-group equilibrium** If there is at least one negative co-regulation in the group, at least one  $B_i$  is negative, which prevents the intra-group theoretical equilibrium above-mentioned (equation 11) to be reached since concentrations cannot be negative. There is another solution to find the intra-group equilibrium, based on the fact that when there is negative co-regulation, the flux cannot exceed a local maximum due the trade-off between enzyme concentrations in each negative group (Coton et al. 2021). At this point, the partial derivatives of the flux with respect to the concentrations cancel, so the flux response coefficient are all null in each negative group:

$$\forall i \in \Phi_q \quad R_{E_i}^J = 0 \quad (15)$$

Using the expression of the flux response coefficient (equation 2), equation 15 becomes (Supporting Information B.3.1.3):

$$\sum_{j \in \Phi_q} \frac{1}{B_j A_j (\tilde{e}_j^q)^2} = 0 \quad (16)$$

where  $\tilde{e}_j^q$  is called the *intra-group effective equilibrium*.

To solve this equation, we used the relationship between  $e_j^q$  and  $\tau^q$ , the driving variable of the parametric equation of the line along which move the intra-group relative concentrations of co-regulated enzymes. From equation 6, equation 16 becomes:

$$\sum_{j \in \Phi_q} \frac{1}{B_j A_j (\tilde{\tau}^q (1/B_j - e_j^{q0}) + e_j^{q0})^2} = 0 \quad (17)$$

From this equation,  $\tilde{\tau}^q$ , and therefore the intra-group effective equilibrium  $\tilde{e}_j^q$ , can be found numerically.

**3.1.2.2 Inter-group equilibrium** Equation 15 does not give a solution for the inter-group effective equilibrium. So to find  $e^q$  at equilibrium, we used the relationship between the sum  $E^q$  of enzyme concentrations in group  $\Phi_q$  and the driving variable  $\tau^q$  (Supporting Information B.3.3):

$$E^q = \frac{E^{q0}}{1 - \tau^q} \quad (18)$$

where  $E^{q0}$  is the sum of initial enzyme concentrations in group  $\Phi_q$ .

If all groups are negative, the values  $\tilde{\tau}^q$  of the driving variables at the effective equilibrium (equation 17) allow to compute the values of  $\tilde{E}^q$  from equation 18. Thus we obtain the inter-group relative concentrations at effective equilibrium (Supporting Information B.3.2.3):

$$\forall q \quad \tilde{e}^q = \frac{E^{q0}}{1 - \tilde{\tau}^q} \quad (19)$$

$$\sum_{k=1}^p \frac{E^{k0}}{1 - \tilde{\tau}^k}$$

**3.1.2.3 Total equilibrium** Using equations 3, 16 and 19, the total effective equilibrium writes:

$$\forall q, \forall i \in \Phi_q \quad \tilde{e}_i = \tilde{e}_i^q \quad \tilde{e}^q = \tilde{e}_i^q \frac{E^{q0}}{1 - \tilde{\tau}^q} \quad (20)$$

$$\sum_{k=1}^p \frac{E^{k0}}{1 - \tilde{\tau}^k}$$

310 Due to the negative co-regulations between some enzyme concentrations, the flux cannot exceed a maximum. At this equilibrium, the flux reaches the following local maximum:

$$\tilde{J}_{\max} = \frac{X}{\sum_{j=1}^n \frac{1}{A_j \tilde{E}_j}} \quad (21)$$

where (Supporting Information B.3.3):

$$\forall i \in \Phi_q \quad \tilde{E}_i = E_i^0 + \frac{E^{q0}}{B_i} \left( -1 + \frac{1}{1 - \tilde{\tau}^q} \right) \quad (22)$$

### 3.1.3 Positive and negative groups

When positive (and/or singletons) and negative groups coexist in a pathway, the intra-  
315 group equilibria are reached within each group: the theoretical equilibrium  $e_i^{q*}$  in positive groups and singletons (equation 11) and the effective equilibrium  $\tilde{e}_i^q$  in negative groups (equation 16).

Regarding the *absolute* concentrations,  $E_i$  and  $E^q$  tends to infinity in the positive groups and singletons because no factor limits the increase of the  $E_i$ 's, and respectively  
320 reaches  $\tilde{E}_i$  and  $\tilde{E}^q = \frac{E^{q0}}{1 - \tilde{\tau}^q}$  in the negative groups (equations 22 and 18). Therefore, over time, every  $\tilde{E}^q$  becomes negligible *relative to* the concentrations of positive groups/singletons and  $E_{\text{tot}}$  tends to  $\sum_{\substack{q=1 \\ \theta_q \geq 0}}^p E^q$ . So,  $e^q$  (resp.  $e_j$ ) tends asymptotically to  $e^{q*}$

(resp.  $e_j^*$ ) for positive groups and singletons, and asymptotically to 0 for negative groups.

Because in the groups that reached their intra-group effective equilibrium,  $E_i$  is limited to  $\tilde{E}_i$ , the flux tends asymptotically to a local maximum that writes (Supporting  
325 Information B.3.2.4):

$$\tilde{J}_{\max} = \frac{X}{\sum_{\substack{q=1 \\ \theta_q < 0}}^p \sum_{j \in \Phi_q} \frac{1}{A_j \tilde{E}_j}} \quad (23)$$

The different equilibria are summarized in Table 1.



**Table 1. Equilibria of relative enzyme concentrations when there is no competition.**  
 $\theta_q$  is the indicator variable for the type of co-regulation in group  $\Phi_q$  (see Material & Methods).

Equilibrium	Only positive groups and singletons	Only negative groups	At least one positive and one negative group
Intra-group	$e_i^{q*} = 1/B_i$	$\tilde{e}_i^q = f(A_j, B_j, e_j^{q0}, \forall j \in \Phi_q)$	$e_i^q \rightarrow e_i^{q*}$ if $\theta_q \geq 0$ $e_i^q \rightarrow \tilde{e}_i^q$ if $\theta_q < 0$
Inter-group	$\forall q, e^{q*} = \frac{(A^{q*})^{-1/3}}{\sum_{k=1}^p (A^{k*})^{-1/3}}$ with $A^{q*} = \left( \sum_{i \in \Phi_q, j \in \Phi_q} \frac{B_i^2 B_j}{A_j} \right)^{-1}$	$\forall q, \tilde{e}^q = \frac{\frac{E^{q0}}{1-\tau^q}}{\sum_{k=1}^p \frac{E^{k0}}{1-\tau^k}}$	$e^q \rightarrow e^{q*}$ if $\theta_q \geq 0$ $e^q \rightarrow 0$ if $\theta_q < 0$
Total	$\forall q, \forall i \in \Phi_q, e_i^* = e_i^{q*} e^{q*}$	$\forall q, \forall i \in \Phi_q, \tilde{e}_i = \tilde{e}_i^q \tilde{e}^q$	$e_i \rightarrow e_i^*$ if $\theta_q \geq 0$ $e_i \rightarrow 0$ if $\theta_q < 0$
Flux	$J \rightarrow \infty$	$J \rightarrow \tilde{J}_{\max} = \frac{X}{\sum_{j=1}^n \frac{1}{A_j \tilde{E}_j}}$	$J \rightarrow \tilde{J}_{\max} = \frac{X}{\sum_{q=1}^p \sum_{j \in \Phi_q, \theta_q < 0} \frac{1}{A_j \tilde{E}_j}}$

### 3.1.4 Computer simulations

The simulations of long-term evolution for pathways of different lengths and with various types of co-regulation groups fully confirm the predictions:

- When there are only positive groups and/or singletons (figure 2A), the simulations show that the flux increases indefinitely, because there is no limit to the increase of enzyme concentrations. The total relative enzyme concentrations tend towards the predicted theoretical equilibrium, which depends on global co-regulation coefficients and enzyme pseudo-activities (figure 2B and 2C).
- When there are only negative groups (figure 3A), the *absolute* enzyme concentrations within each group are limited by the negative co-regulations. The total relative enzyme concentrations reach their effective equilibrium, which depends on enzyme pseudo-activities and on global co-regulation coefficients, but also on initial enzyme concentrations. Figures 3B and 3C show the joint evolution of total relative enzyme concentrations and flux. Whatever the initial concentrations, the flux increases until the maximum possible flux  $\tilde{J}_{\max}$  is reached, and enzyme concentrations get closer to their effective equilibrium. Then, enzyme concentrations fluctuate around the effective equilibrium while the flux stays approximately constant, which corresponds to the neutral zone.
- When positive (and/or singletons) and negative groups coexist (figure 4A), the intra-group relative concentrations are expected to reach the theoretical and effective equilibria, respectively. Due to the negative co-regulations, there is a maximum possible flux  $\tilde{J}_{\max}$ . Figures 4B and 4C show the joint evolution of total relative enzyme concentrations and flux. We see that the flux gets closer and closer to  $\tilde{J}_{\max}$ .

After some burn-in period, total *relative* concentrations decrease in negative groups and increase in positive groups. However, predicted equilibria were never reached. Figures 4D and 4E show the joint evolution of absolute enzyme concentrations and flux in the same simulations as in figures 4B and 4C. After the same burn-in  
 355 period, enzyme concentrations in positive groups increase slower and slower, while enzyme concentrations in negative groups stochastically move around the effective equilibria  $\tilde{E}_i$ . Indeed, as the flux gets closer to the maximum possible, selection intensity decreases and enzyme concentrations change within a quasi-neutral zone.

Note that even though the stability of the equilibria could not be analytically proven,  
 360 simulations show that the equilibria are stable.

## 3.2 Evolutionary equilibria with competition

Here we assume that there is competition for cellular resources between all enzymes of the pathway, so we fixed the total enzyme concentration. When there is competition, the flux cannot exceed a maximal value whatever the type of co-regulation (Coton et al.  
 365 2021). The relationship between the flux and the enzyme concentrations is a dome in the multidimensional space of the concentrations. At the top of the dome the partial derivatives of the flux with respect to the concentrations cancel for all enzymes, so we have:

$$\forall i \quad R_{E_i}^J = 0 \quad (24)$$

and at this point the enzyme concentrations are at their effective equilibrium.

In this section, we do not have to consider separately the positive and the negative  
 370 groups, because competition between enzymes create negative relationships between all enzymes.

### 3.2.1 Intra-group effective equilibrium

The intra-group relative concentration at effective equilibrium,  $\tilde{e}_i^q$ , is solution of the following equation (Supporting Information B.4.1.2):  
 375

$$\sum_{\substack{k=1 \\ \theta_k \neq 0}}^p \left( \tilde{A}^k^{-1/2} - \tilde{A}^k^{1/2} \sum_{j \in \Phi_k} \frac{1}{A_j \tilde{e}_j^k} \right) = 0 \quad (25)$$

where the apparent activity  $\tilde{A}^q$  of group  $\Phi_q$  at effective equilibrium is defined such as (provided  $\tilde{A}^q$  is positive):

$$\frac{1}{\tilde{A}^q} = \sum_{j \in \Phi_q} \frac{1}{A_j B_j (\tilde{e}_j^q)^2} \quad (26)$$

We did not find any general analytical or numerical solutions for equation 25, but the simulations show that the equilibria exist (see section 3.2.4 below). We find a numerical  
 380 solution in the particular case where there is only one co-regulation group – positive or negative – and any number of singletons. In that case, equation 25 writes after rearrangements (Supporting Information B.4.1.3):

$$\sum_{j \in \Phi_q} \left( \frac{1/B_j - e_j^{q_0}}{A_j \tilde{e}_j^{q_2}} \right) = 0 \quad (27)$$

### 3.2.2 Inter-group effective equilibrium

From equation 24, we can express  $\tilde{e}^q$ , the inter-group relative concentration at effective equilibrium (Supporting Information B.4.1.1):

$$\tilde{e}^q = \frac{\tilde{A}^q^{-1/2}}{\sum_{k=1}^p \tilde{A}^k^{-1/2}} \quad (28)$$

As previously,  $\tilde{A}^q$  (equation 26) is a weighed average of the pseudo-activities of enzymes, but its expression is different from case without competition (equation 13).

### 3.2.3 Total effective equilibrium

From equations 3, 25 and 28 we get the total relative concentration  $\tilde{e}_i$  at effective equilibrium:

$$\forall i \in \Phi_q \quad \tilde{e}_i = \tilde{e}^q \tilde{e}_i^q = \tilde{e}_i^q \frac{\tilde{A}^q^{-1/2}}{\sum_{k=1}^p \tilde{A}^k^{-1/2}} \quad (29)$$

At this effective equilibrium, the flux is at a local maximum  $\tilde{J}_{\max}$  that is lower than the maximum flux  $J_{\max}$  when there is no co-regulations, because the co-regulations impose the flux to evolve in a lower dimension space.

The equilibria are summarized in table 2.

**Table 2. Effective equilibrium of relative enzyme concentrations when there is competition.**

Equilibrium	Positive group and singleton	Negative group
Intra-group	$\tilde{e}_i^q = f(\mathbf{A}, \mathbf{M}_\beta, \mathbf{E}^0)$	
Inter-group	$\forall q, \tilde{e}^q = \frac{(\tilde{A}^q)^{-1/2}}{\sum_{k=1}^p (\tilde{A}^k)^{-1/2}}$ with $\frac{1}{\tilde{A}^q} = \sum_{j \in \Phi_q} \frac{1}{B_j A_j (\tilde{e}_j^q)^2}$	
Total	$\tilde{e}_i = \tilde{e}_i^q \tilde{e}^q$	
Flux	$J = \tilde{J}_{\max}$	

### 3.2.4 Computer simulations

Competition between enzymes creates negative relationships between all enzymes and the evolutionary issue is the same, whatever the co-regulation groups: given the initial conditions, the flux reaches a maximum  $\tilde{J}_{\max}$ , and total relative enzyme concentrations reach the effective equilibria. A numerical solution was found for the effective equilibria only in the case of a single co-regulation group and any number of singletons. Figure 5A shows an example of such situation. Figures 5B and 5C show the joint evolution of total relative enzyme concentrations and flux in the simulations. As predicted, the flux increases until a maximum  $\tilde{J}_{\max}$  that depends on the initial conditions and that is bounded by the envelop-curve of the dome corresponding to the absence of co-regulation. Once  $\tilde{J}_{\max}$

405 is reached, total relative enzyme concentrations fluctuate in a neutral zone around the effective equilibrium. Note that close to the equilibria, random fluctuations of enzyme-flux relationships delineate dome-shaped trajectories, which suggests that these trajectories evolve along a dome in lower dimension spaces as compared to the global competition dome. Another striking observation is that, before reaching the neutral zone, enzyme concentrations seem to walk erratically, sometimes decreasing before increasing again. 410 Again, this suggests evolution in a constrained solution space.

Figure 6A shows an example with one positive and one negative co-regulation group. In that case, numerical predictions for the effective equilibria are not computable. Figures 6B and 6C show the joint evolution of total relative enzyme concentrations and flux. 415 Adaptive trajectories are qualitatively the same as in the previous case: given the initial conditions, a quasi-equilibrium is reached where the flux is close to a maximum, while total relative enzyme concentrations vary into a neutral zone around a fixed value.

The dependency of the evolutionary equilibria to the initial conditions raises questions 420 about the patterns of co-variation between enzyme concentrations that could be expected after independent bouts of evolution from different initial conditions. Figure 7 shows the relationships between the equilibrium concentrations of two enzymes belonging or not to the same co-regulation group for a four-enzyme pathway, starting from different initial conditions. In all cases where effective equilibrium were computable, we found 425 a close adequacy between the predictions and the simulations (figures 7B, 7D). When all enzymes belong to the same co-regulation group (figure 7A), no relation was found (figure 7B). In all other cases (figures 7C, 7E, 7G), we found clear non-linear patterns (figures 7D, 7F, 7H). For example, in the case of a positive co-regulation group and a singleton, equilibrium concentrations of enzyme  $E_1$  and  $E_3$  belonging to the positive 430 group fall into an ellipsoid (figure 7D). Other values of co-regulation coefficients give another ellipsoidal pattern (figures 7F, 7H). Without knowing the initial conditions, this shows that the constraint added by the co-regulations makes the evolution unpredictable with regards to enzyme concentrations.

### 3.3 The range of selective neutrality of enzyme concentrations

435 The ranges of neutral variation (RNV) of enzyme concentrations close to the equilibria were compared in the different situations.

#### 3.3.1 RNV without competition

Three cases can be distinguished, as we did for the search of equilibria:

- 440 1. When there are only positive groups and singletons, the flux response coefficients  $R_{E_i}^J$  are constant at theoretical equilibrium (Supporting Information B.5). Therefore, the selection coefficient  $s_i$ , through the relationship  $s_i \approx R_{E_i}^J \frac{\delta_i}{E_i}$  (Coton et al. 2021), tends to zero because  $E_i$  increases indefinitely. The mutations have less and less effect on the flux while approaching the theoretical equilibrium.
- 445 2. When there are only negative groups, the flux response coefficients, and hence the selection coefficients, are null at effective equilibrium, so there is selective neutrality.
3. When there are both positive and negative groups, the flux response coefficients are null for negative groups at effective equilibrium and tends to zero for positive and

singleton groups (Supporting Information B.5). Therefore the selection coefficients tend to zero in every group.

450 Thus, when there is no competition, there is neutrality close to the equilibria, whatever the number and type of co-regulation groups. However, in contrast to what we did in Coton et al. (2021), we did not succeed in determining the limits of the RNV due to the complexity of the relationships between the co-regulation groups.

455 In the simulations, the neutral zone is revealed by the fluctuations of the concentrations around the effective equilibrium for negative groups (figures 3 and 4D). Figures 3B and 3C also shows that the range of neutral variation (RNV) of enzyme concentrations both depend on initial concentrations and on enzyme parameters.

### 3.3.2 RNV with competition

460 When there is competition, the response coefficients are null at effective equilibrium, and so are the selection coefficients. Therefore, there is a neutral zone around every enzyme concentration at equilibrium. However, as previously we could not untangle the factors that determine the RNV. In the simulations, the neutral zone is clearly visible (figures 5 and 6). At the end of the simulations, the concentrations fluctuate around the predicted effective equilibria, making apparent the ranges of neutral variation at the top of domes. 465 As equilibrium, RNVs probably depend on pseudo-activities, co-regulation coefficients and initial concentrations.

## 4 Discussion

The enzyme-flux relationship is commonly used to model the genotype-phenotype relationship, owing to the sound theoretical framework developed for decades that link enzymes properties (the "genotype") to the flux (the phenotype) (Wright 1934; Kacser and 470 Burns 1981; Fiévet et al. 2018). In terms of evolution, this multidimensional relationship constitutes an adaptive landscape when the flux is proportional to fitness.

In Coton et al. (2021), we developed a model of evolution of enzyme concentrations under selection for increased flux. We applied two constraints, competition for resources 475 and/or co-regulation between all enzymes, and showed the existence of two types of evolutionary equilibrium of relative concentrations, depending on the presence of constraints or not. However, considering that all enzymes are co-regulated in a pathway is a particular case. Here, we analysed the general case where there are different co-regulation groups, with or without competition, distinguishing two types of co-regulation groups: (i) the 480 positive groups when enzymes are all positively co-regulated; (ii) the negative groups when there is at least one negative co-regulation in the group. We called singletons the enzymes that are independent of all other enzymes.

We showed that selection pressure on enzyme concentrations can be decoupled within and between co-regulation groups to predict the evolutionary equilibria. In most situa- 485 tions, analytical or numerical solutions were proposed and were confronted to simulations. Combining theoretical predictions and computer simulations, we showed that a stable evolutionary equilibrium for relative enzyme concentrations always exists. In the trivial case of only positive regulation groups without competition, this equilibrium is unique whatever the initial concentrations. In all other cases, evolutionary equilibria depended 490 not only on enzyme parameters and on the gene regulatory network, but also on initial conditions. Such an evolutionary contingency emerges from the existence of constraints,

which stem either from the competition for cellular resources that limits total enzyme concentration, or from negative co-regulations, that prevents the simultaneous increase of the concentrations of the co-regulated enzymes.

#### 4.1 One negative co-regulation is sufficient to limit the flux

Even in the absence of competition, we showed that the trade-off between co-regulated enzymes in negative groups blocks the flux at a local maximum (equation 23). Such trade-offs are sufficient to limit the evolution of all enzyme concentrations, but also to limit the flux. Heckmann et al. (2018) observed a similar phenomenon. They simulated the evolution of enzyme turnovers ( $k_{\text{cat}}$ ) under selection for increased growth rate with fixed concentrations. Because enzyme concentrations and turnover numbers have concurrent roles in the expression of the flux (equation 1), constraining enzyme concentrations is similar to fixing enzyme turnovers, as we did here. Fixing some  $k_{\text{cat}}$ 's, Heckmann et al. (2018) showed that the diminishing returns in the relationship between the variable  $k_{\text{cat}}$ 's and growth rates prevented selection for high  $k_{\text{cat}}$  and maintained growth rate far from its possible maximum value. Here, we showed that any biological constraint that prevents the concentration of one enzyme from increasing despite selection for increasing the flux will result into an upper-bound for the flux at the evolutionary equilibria. Other constraints could stem from trade-offs between active sites of an enzyme (Savir et al. 2010), or from the gene regulatory network. For example, among the genes that control yeast cell cycle, the cyclins Clb1 and Clb2 deactivate different transcription factors like MBF, SBF and Swi5 that in turn control different parts of the cycle (Li et al. 2004).

Another interesting result is that upon selection, the flux increases monotonically towards the maximum, whereas enzyme concentrations decrease or increase, depending on the target of the mutations and on their effect on the flux *via* the co-regulated enzymes (figures 5 and 6). Thus, a monotonous phenotypic response to directional selection does not necessarily mean that the underlying genotypic factors evolve monotonously.

#### 4.2 Evolutionary contingency and the world of possibles

A major consequence of the existence of co-regulations is that both maximum flux and evolutionary equilibria for total relative enzyme concentrations depend on initial conditions. In this model, evolution has an historical dimension. The knowledge of the gene regulatory network and enzyme parameters is not sufficient to predict the evolutionary outcome. At most, it would be possible to predict an envelop-curve for the flux in the competition case (figures 5 and 6). The dependence to initial conditions also results in non linear patterns between enzyme concentrations at evolutionary equilibria (figure 7). This result could explain the wide range of genetic variation of enzyme concentrations observed within populations, as well as their highly polygenic nature (Damerval et al. 1994; Albert and Kruglyak 2015; Chick et al. 2016). Such dependency to initial conditions is often observed in complex biological systems. It is the case for non linear dynamical systems like the lactose operon (Laurent and Kellershohn 1999), but also for complex patterns emerging from network self-organization that can be treated using game theory (Broere et al. 2017).

Contingency comes from the fact that constraints modify the adaptive landscape by limiting the ability to move in the multidimensional space ( $\mathbf{E}, J$ ). We searched the number of degrees of freedom of the system, which is the number of independent variables necessary and sufficient to describe the system, given the constraints applied and the

fixed parameters – here the pseudo-activities  $\mathbf{A}$ , the co-regulation coefficients of matrix  $\mathbf{M}_\beta$  and the initial concentrations  $\mathbf{E}^0$  (Supporting Information C.1).

If there is no constraint on enzyme concentrations, the genotypes move freely in the multidimensional space  $(\mathbf{E}, J)$ . When there is competition between all enzymes, the adaptive landscape becomes a dome in  $(\mathbf{E}, J)$ , the shape of which depends only on the pseudo-activities (Coton et al. 2021). When there are co-regulations between *all* enzymes, the enzyme concentrations are linearly related and are fully determined by the driving variable  $\tau$ : the adaptive landscape is reduced to two dimensions,  $\tau$  and  $J$  (Coton et al. 2021). It is the same when we add competition. In that case the adaptive landscape becomes a section of the dome, oriented by the global co-regulation coefficients and the initial concentrations.

Here we considered the presence of co-regulation groups. Without competition, the enzyme concentrations are linearly related within groups, and are independent from enzyme concentrations of other groups. So the concentrations of each co-regulation group  $q$  are given by its driving variable  $\tau^q$ . The adaptive landscape is in a  $p + 1$  dimension space – the number of co-regulation groups plus the flux. With competition between all enzymes, we could not determine precisely the number of degrees of freedom. Simulations showed that the adaptive landscape is a projection in lower dimensions of the dome due to competition, and its shape is also a dome (figure 6). The effective equilibria are given by the projection of the top of the new dome in the concentration space. They depend on the number of co-regulation groups, and each additional co-regulation between two previously independent groups decreases the number of dimensions.

### 4.3 Evolutionary jumps and selective neutrality

Generalizing the results of Hartl et al. (1985) on the "natural selection of selective neutrality", we showed that in all cases there is a neutral zone of enzyme concentrations, where small changes in concentrations do not significantly change the flux that is the target of selection. The ranges of neutral variations differ according to the pseudo-activities, the co-regulation coefficients and initial conditions.

In our simulations, we found that, given the initial conditions, the system evolved quite rapidly to the equilibrium. In the same way, any mutational change in a pseudo-activity or a co-regulation coefficient would modify the adaptive landscape and change simultaneously enzyme concentrations and flux, defining new equilibria. Such evolutionary jumps, also mentioned by Heckmann et al. (2018) when mutations affect multifunctional enzymes, are reminiscent of punctuated equilibria (Gould and Eldredge 1977). Altogether, our simplified model of enzyme evolution allowed us to give support to known observations in evolution: the selective neutrality of enzymes at evolutionary equilibria, and the existence of evolutionary jumps due to rare mutations changing system's parameters.

## Acknowledgements

575 CC was supported by a PhD thesis grant from the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation.  
The authors declare no conflict of interest.

## Author contributions

580 CD and DdV initiated the project. CC developed the model, performed the analysis and the simulations and developed the dedicated R package. CC and DdV wrote the manuscript, with the contribution of CD.

## Literature cited

- Albert, F. W. and L. Kruglyak, 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 16:197–212.
- 585 Askew, C., A. Sellam, E. Epp, H. Hogues, A. Mullick, A. Nantel, and M. Whiteway, 2009. Transcriptional Regulation of Carbohydrate Metabolism in the Human Pathogen *Candida albicans*. *PLOS Pathogens* 5:e1000612.
- Bershtein, S., A. W. Serohijos, and E. I. Shakhnovich, 2017. Bridging the physical scales in evolutionary biology: From protein sequence space to fitness of organisms and populations. *Curr Opin Struct Biol* 42:31–40.
- 590 Blein-Nicolas, M., W. Albertin, B. Valot, P. Marullo, D. Sicard, C. Giraud, S. Huet, A. Bourgeois, C. Dillmann, D. de Vienne, and M. Zivy, 2013. Yeast Proteome Variations Reveal Different Adaptive Responses to Grape Must Fermentation. *Molecular Biology and Evolution* 30:1368–1383.
- 595 Broere, J., V. Buskens, J. Weesie, and H. Stoof, 2017. Network effects on coordination in asymmetric games. *Scientific Reports* 7:17016.
- Chambers, A., E. A. Packham, and I. R. Graham, 1995. Control of glycolytic gene expression in the budding yeast (*Saccharomyces cerevisiae*). *Curr Genet* 29:1–9.
- Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, 600 K. L. Svenson, G. A. Churchill, and S. P. Gygi, 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534:500–505.
- Colombo, M., H. Laayouni, B. M. Invergo, J. Bertranpetit, and L. Montanucci, 2014. Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes. *Evolution* 68:605–613.
- 605 Coton, C., G. Talbot, M. L. Louarn, C. Dillmann, and D. de Vienne, 2021. Evolution of enzyme levels in metabolic pathways: A theoretical approach. *bioRxiv* P. 2021.05.04.442631.
- Damerval, C., A. Maurice, J. M. Josse, and D. de Vienne, 1994. Quantitative Trait Loci Underlying Gene Product Variation: A Novel Perspective for Analyzing Regulation of Genome Expression. *Genetics* 137:289–301.
- 610 Eanes, W., 1999. Analysis of selection on enzyme polymorphism. *Annu. Rev. Ecol.* 30:301–326.
- Ellis, R., 2001. Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* 26:597–604.
- Fiévet, J. B., T. Nidelet, C. Dillmann, and D. de Vienne, 2018. Heterosis Is a Systemic Property Emerging From Non-linear Genotype-Phenotype Relationships: Evidence From in Vitro 615 Genetics and Computer Simulations. *Frontiers in Genetics* 9.
- Flowers, J. M., E. Sezgin, S. Kumagai, D. D. Duvernell, L. M. Matzkin, P. S. Schmidt, and W. F. Eanes, 2007. Adaptive Evolution of Metabolic Pathways in *Drosophila*. *Mol Biol Evol* 24:1347–1354.



- Goelzer, A., F. Bekkal Brikci, I. Martin-Verstraete, P. Noirot, P. Bessières, S. Aymerich, and  
620 V. Fromion, 2008. Reconstruction and analysis of the genetic and metabolic regulatory net-  
works of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol* 2:20.
- Gould, S. J. and N. Eldredge, 1977. Punctuated equilibria: the tempo and mode of evolution  
reconsidered. *Paleobiology* 3:115–151.
- Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985. Limits of Adaptation: The Evolution of  
625 Selective Neutrality. *Genetics* 111:655–674.
- Heckmann, D., D. C. Zielinski, and B. O. Palsson, 2018. Modeling genome-wide enzyme evolution  
predicts strong epistasis underlying catalytic turnover rates. *Nature Communications* 9.
- Heinrich, R. and T. Rapoport, 1974. Linear Steady-State Treatment of Enzymatic Chains -  
General Properties, Control and Effector Strength. *Eur. J. Biochem.* 42:89–95.
- 630 Heinrich, R., S. Schuster, and H.-G. Holzhütter, 1991. Mathematical analysis of enzymic reaction  
systems using optimization principles. *European Journal of Biochemistry* 201:1–21.
- Jiang, L.-G., B. Li, S.-X. Liu, H.-W. Wang, C.-P. Li, S.-H. Song, M. Beatty, G. Zastrow-Hayes,  
X.-H. Yang, and F. Qin, 2019. Characterization of proteome variation during modern maize  
breeding. *Molecular & Cellular Proteomics* 18:263–276.
- 635 Kacser, H. and R. Beeby, 1984. Evolution of catalytic proteins: On the origin of enzyme species  
by means of natural selection. *J Mol Evol* 20:38–51.
- Kacser, H. and J. A. Burns, 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27:65–104.
- , 1981. The Molecular Basis of Dominance. *Genetics* 97:639–666.
- Kacser, H., J. A. Burns, H. Kacser, and D. A. Fell, 1995. The control of flux : 21 years on.  
640 *Biochemical Society Transactions* 23:341–366.
- Klumpp, S., W. Bode, and P. Puri, 2019. Life in crowded conditions: Molecular crowding and  
beyond. *Eur. Phys. J. Spec. Top.* 227:2315–2328.
- Kurland, C. G. and H. Dong, 1996. Bacterial growth inhibition by overproduction of protein.  
*Molecular Microbiology* 21:1–4.
- 645 Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang,  
B. Oliver, and A. G. Clark, 2008. Evolution of protein-coding genes in *Drosophila*. *Trends in  
Genetics* 24:114–123.
- Laurent, M. and N. Kellershohn, 1999. Multistability: a major means of differentiation and  
evolution in biological systems. *Trends in Biochemical Sciences* 24:418–422.
- 650 Li, F., T. Long, Y. Lu, Q. Ouyang, and C. Tang, 2004. The yeast cell-cycle network is robustly  
designed. *Proceedings of the National Academy of Sciences* 101:4781–4786.
- Lion, S., F. Gabriel, B. Bost, J. Fiévet, C. Dillmann, and D. De Vienne, 2004. An extension to  
the metabolic control theory taking into account correlations between enzyme concentrations.  
*European Journal of Biochemistry* 271:4375–4391.
- 655 Morrison, E. S. and A. V. Badyaev, 2017. Beyond topology: coevolution of structure and flux  
in metabolic networks. *J. Evol. Biol.* 30:1796–1809.
- Newton, M. S., V. L. Arcus, M. L. Gerth, and W. M. Patrick, 2018. Enzyme evolution: innovation  
is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* 48:110–116.
- Noda-Garcia, L., W. Liebermeister, and D. S. Tawfik, 2018. Metabolite–Enzyme Coevolution:  
660 From Single Enzymes to Metabolic Pathways and Networks. *Annu. Rev. Biochem.* 87:187–216.
- R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation  
for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-  
900051-07-0.
- Rausher, M. D., R. E. Miller, and P. Tiffin, 1999. Patterns of evolutionary rate variation among  
665 genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol* 16:266–274.
- Sambamoorthy, G., H. Sinha, and K. Raman, 2019. Evolutionary design principles in metabolism.  
*Proc. R. Soc. B.* 286:20190098.
- Savir, Y., E. Noor, R. Milo, and T. Tlusty, 2010. Cross-species analysis traces adaptation  
of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National  
670 Academy of Sciences* 107:3475–3480.

Sellis, D. and M. D. Longo, 2015. Patterns of variation during adaptation in functionally linked loci. *Evolution* 69:75–89.

Serohijos, A. W. and E. I. Shakhnovich, 2014. Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. *Molecular biology and evolution* 31:165–176.

675

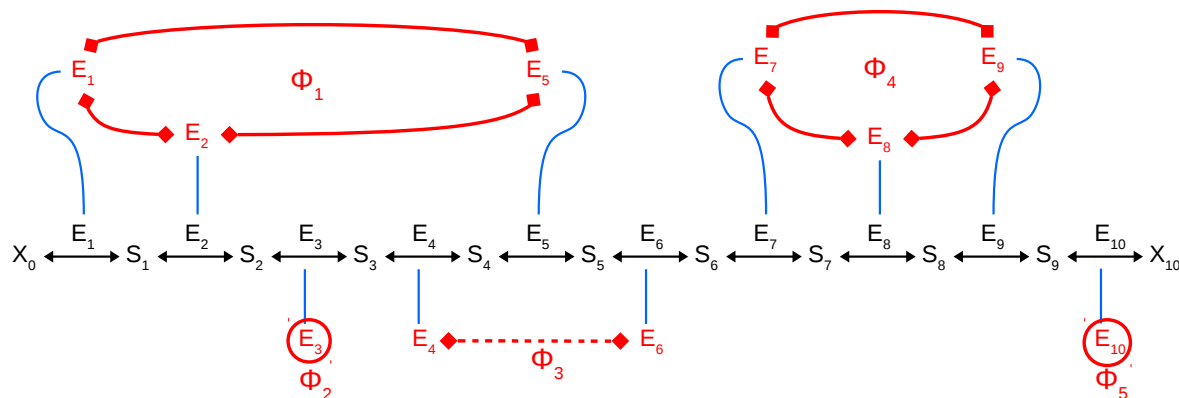
Seshasayee, A. S. N., G. M. Fraser, M. M. Babu, and N. M. Luscombe, 2009. Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*. *Genome Res.* 19:79–91.

de Vienne, D., B. Bost, J. Fiévet, and C. Dillmann, 2001. Optimisation of Enzyme Concentrations for Unbranched Reaction Chains: The Concept of Combined Response Coefficient. *Acta Biotheor* 49:341–350.

680

Wright, S., 1934. Physiological and Evolutionary Theories of Dominance. *The American Naturalist* 68:24–53.

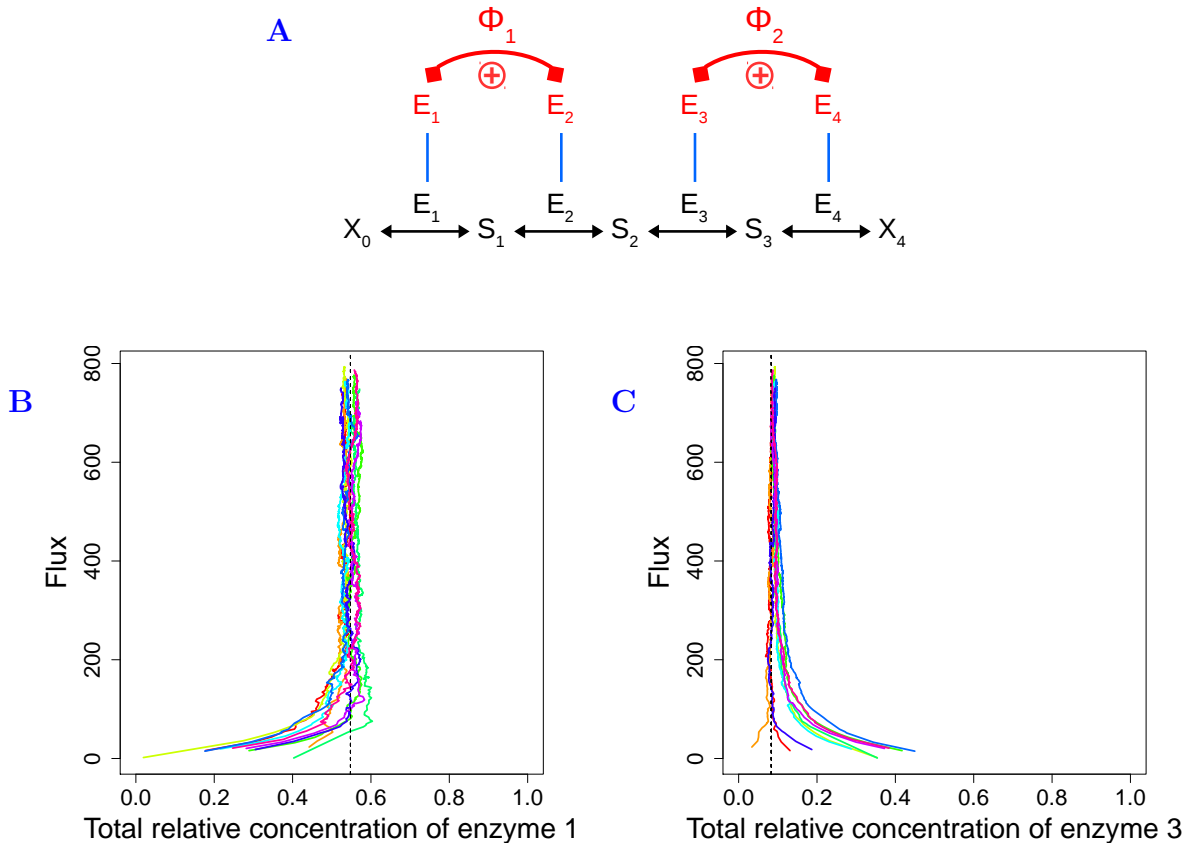
## 5 Figures



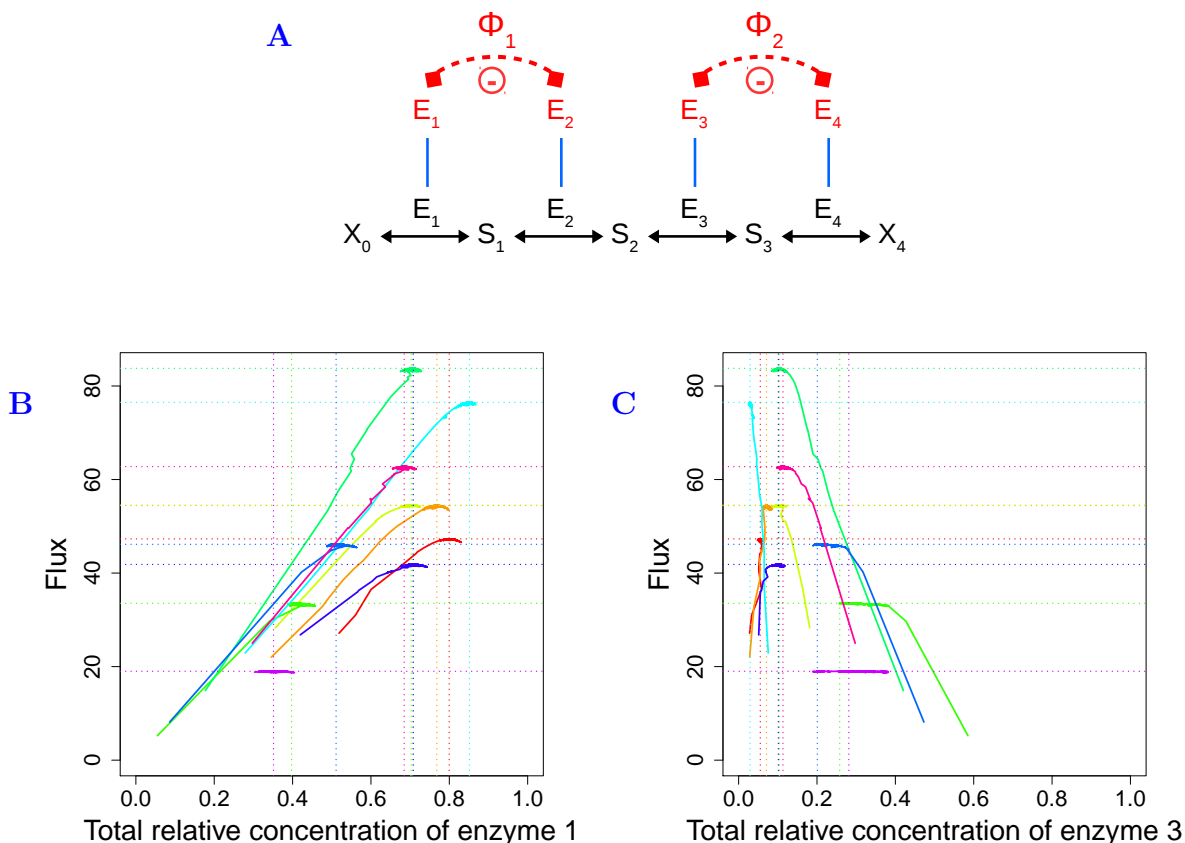
**Figure 1. Example of co-regulation groups in a linear pathway.** The pathway transforms an initial substrate  $X_0$  in a final product  $X_{10}$  through ten reversible reactions catalysed by enzymes  $E_1$  to  $E_{10}$ .  $S_i$  is the product of enzyme  $E_i$  and the substrate of enzyme  $E_{i+1}$ . There are three co-regulation groups and two singletons ( $p = 5$ ), noted  $\Phi_1$  to  $\Phi_5$ , which result from hypothetical gene regulatory processes (in red). We have  $\Phi_1 = \{1, 2, 5\}$ ;  $\Phi_2 = \{3\}$ ;  $\Phi_3 = \{4, 6\}$ ;  $\Phi_4 = \{7, 8, 9\}$ ; and  $\Phi_5 = \{10\}$ . The blue lines relate genes  $g_1$  to  $g_{10}$  to enzymes  $E_1$  to  $E_{10}$ , respectively. Red plain line and red dashed line denote positive and negative co-regulations, respectively. The singletons (circled enzymes) are not co-regulated with any other enzyme. The co-regulation matrix of this pathway is in table 3.

**Table 3. Example of co-regulation matrix  $M_\beta$**  corresponding to the pathway of figure 1. We can extract the co-regulation sub-matrix for each of the three co-regulation groups:  $\Phi_1 = \{1, 2, 5\}$ ,  $\Phi_3 = \{4, 6\}$  and  $\Phi_4 = \{7, 8, 9\}$ .

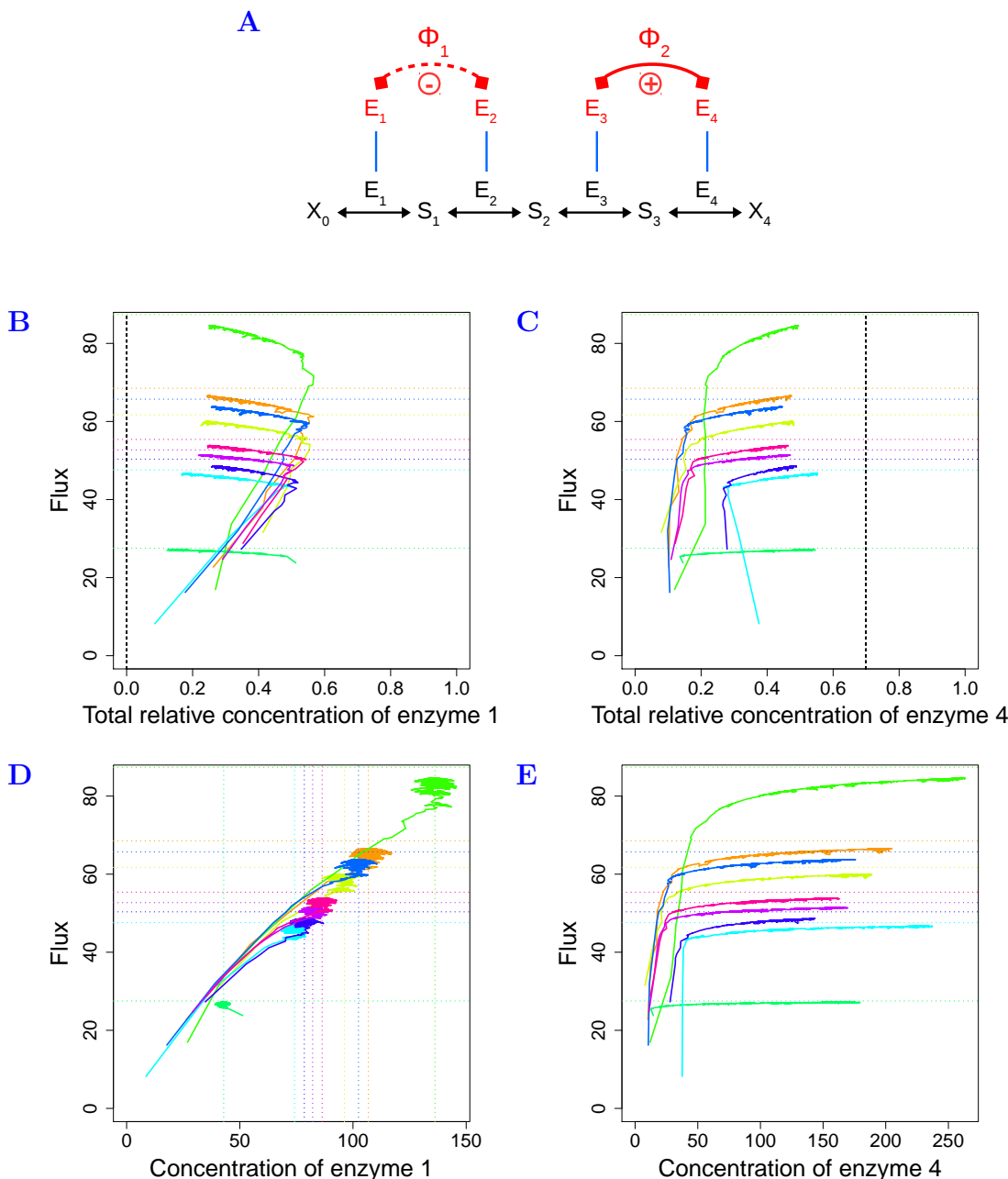
$i \backslash j$	$\beta_{-1}$	$\beta_{-2}$	$\beta_{-3}$	$\beta_{-4}$	$\beta_{-5}$	$\beta_{-6}$	$\beta_{-7}$	$\beta_{-8}$	$\beta_{-9}$	$\beta_{-10}$	Sum
$\beta_{1-}$	1	$\beta_{12}$	0	0	$\beta_{15}$	0	0	0	0	0	$B_1$
$\beta_{2-}$	$1/\beta_{12}$	1	0	0	$\beta_{21} \beta_{15}$	0	0	0	0	0	$B_2$
$\beta_{3-}$	0	0	1	0	0	0	0	0	0	0	$B_3$
$\beta_{4-}$	0	0	0	1	0	$\beta_{46}$	0	0	0	0	$B_4$
$\beta_{5-}$	$1/\beta_{15}$	$1/\beta_{25}$	0	0	1	0	0	0	0	0	$B_5$
$\beta_{6-}$	0	0	0	$1/\beta_{46}$	0	1	0	0	0	0	$B_6$
$\beta_{7-}$	0	0	0	0	0	0	1	$\beta_{78}$	$\beta_{79}$	0	$B_7$
$\beta_{8-}$	0	0	0	0	0	0	$1/\beta_{78}$	1	$\beta_{87} \beta_{79}$	0	$B_8$
$\beta_{9-}$	0	0	0	0	0	0	$1/\beta_{79}$	$1/\beta_{89}$	1	0	$B_9$
$\beta_{10-}$	0	0	0	0	0	0	0	0	0	1	$B_{10}$



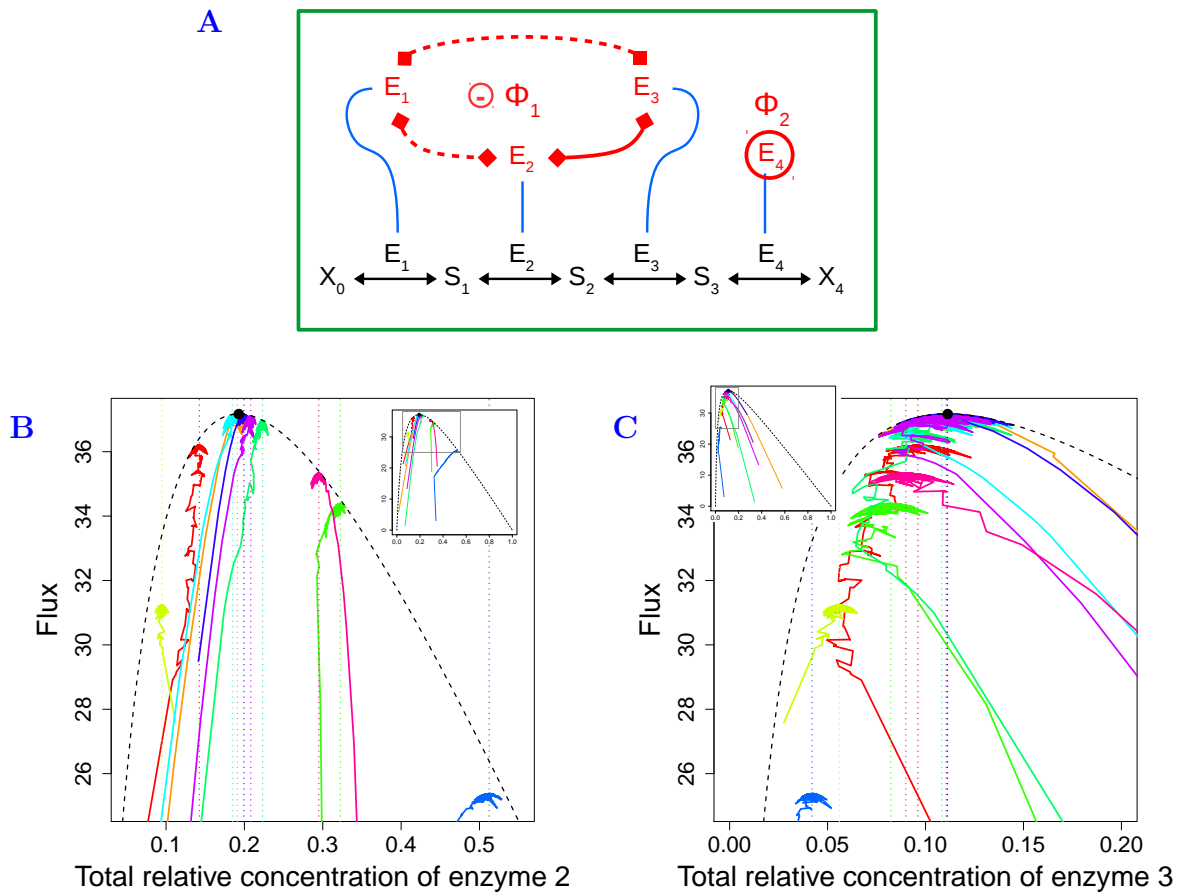
**Figure 2. Simulations of enzyme evolution in a four-enzyme pathway when there are two positive groups, without competition.** Simulations run for 1 000 000 time steps. (A) The four-enzyme pathway used in the simulations. The enzymes  $E_1$  and  $E_2$  form a positive group and  $E_3$  and  $E_4$  form another positive group (same symbols as in figure 1). Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = 0.32$ ,  $\beta_{43} = 0.43$ . (B) Relationship between flux  $J$  and total relative concentration  $e_1$  of enzyme 1. Ten simulations are represented, each with specific initial concentrations (one color per simulation). Vertical dashed lines: abscissa of the theoretical equilibria of total relative concentrations  $e_i^*$ . The total relative enzyme concentrations reach the theoretical equilibrium, within the neutral zone. (C) Same representation as in (B) for the positively-regulated enzyme 3.



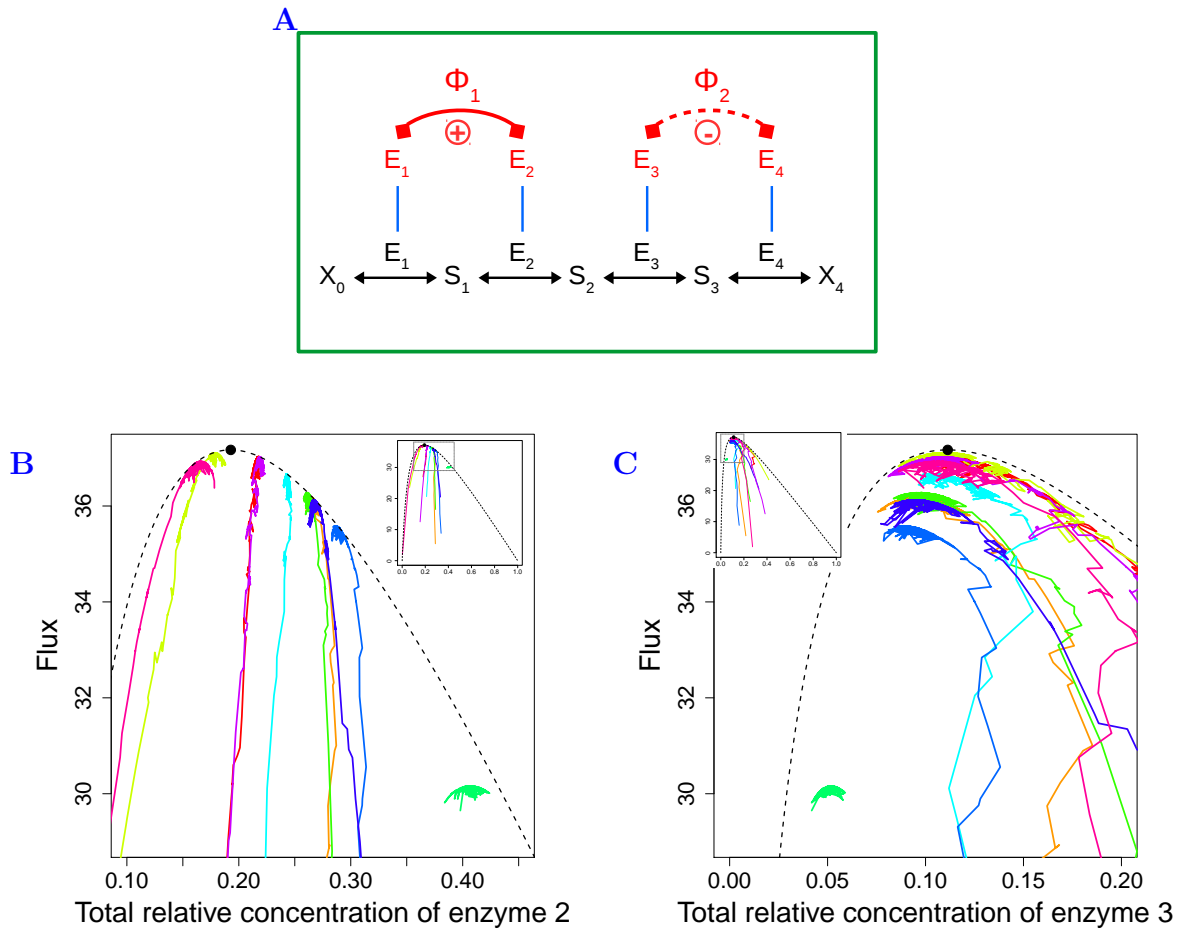
**Figure 3. Simulations of enzyme evolution in a four-enzyme pathway when there are two negative groups, without competition.** Simulations run for 125 000 time steps. (A) The four-enzyme pathway used in the simulations. The enzymes  $E_1$  and  $E_2$  form a negative group and  $E_3$  and  $E_4$  form another negative group (same symbols as in figure 1). Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = -0.1$ ,  $\beta_{43} = -0.43$ . (B) Relationship between flux  $J$  and total relative concentration  $e_1$  of enzyme 1. Ten simulations are represented, each with specific initial concentrations (one color per simulation). Colored vertical dotted lines: abscissa of the effective equilibria of total relative concentrations  $\tilde{e}_i$ . Colored horizontal dotted lines: maximal flux  $\tilde{J}_{\max}$ . The total relative enzyme concentrations reach the effective equilibrium, which correspond to the maximal flux. (C) Same representation as in (B) for the negatively-regulated enzyme 3.



**Figure 4. Simulations of enzyme evolution in a four-enzyme pathway when there are one positive and one negative groups, without competition.** Simulations run for 1 000 000 time steps. (A) The four-enzyme pathway used in the simulations. The enzymes  $E_1$  and  $E_2$  form a positive group and  $E_3$  and  $E_4$  form a negative group (same symbols as in figure 1). Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = -0.32$ ,  $\beta_{43} = 0.43$ . (B) Relationship between flux  $J$  and total relative concentration  $e_1$  of enzyme 1. Nine simulations are represented, each with specific initial concentrations (one color per simulation). Vertical dashed line: abscissa of zero, for which total relative enzyme concentrations of enzyme in negative groups are expected to tend towards. Horizontal dotted lines: maximal flux  $\tilde{J}_{\max}$ . (C) Same representation as in (B) for enzyme 4. Vertical dashed lines: abscissa of the theoretical equilibria of total relative concentrations  $e_i^*$  without consideration of negative groups. (D) Relationship between flux  $J$  and absolute concentration  $E_1$  of enzyme 1. Vertical dotted lines: abscissa of the effective equilibria of absolute concentrations  $\tilde{E}_i$ . Horizontal dotted lines: maximal flux  $\tilde{J}_{\max}$ . (E) Same representation as in (D) for positively-regulated enzyme 4.

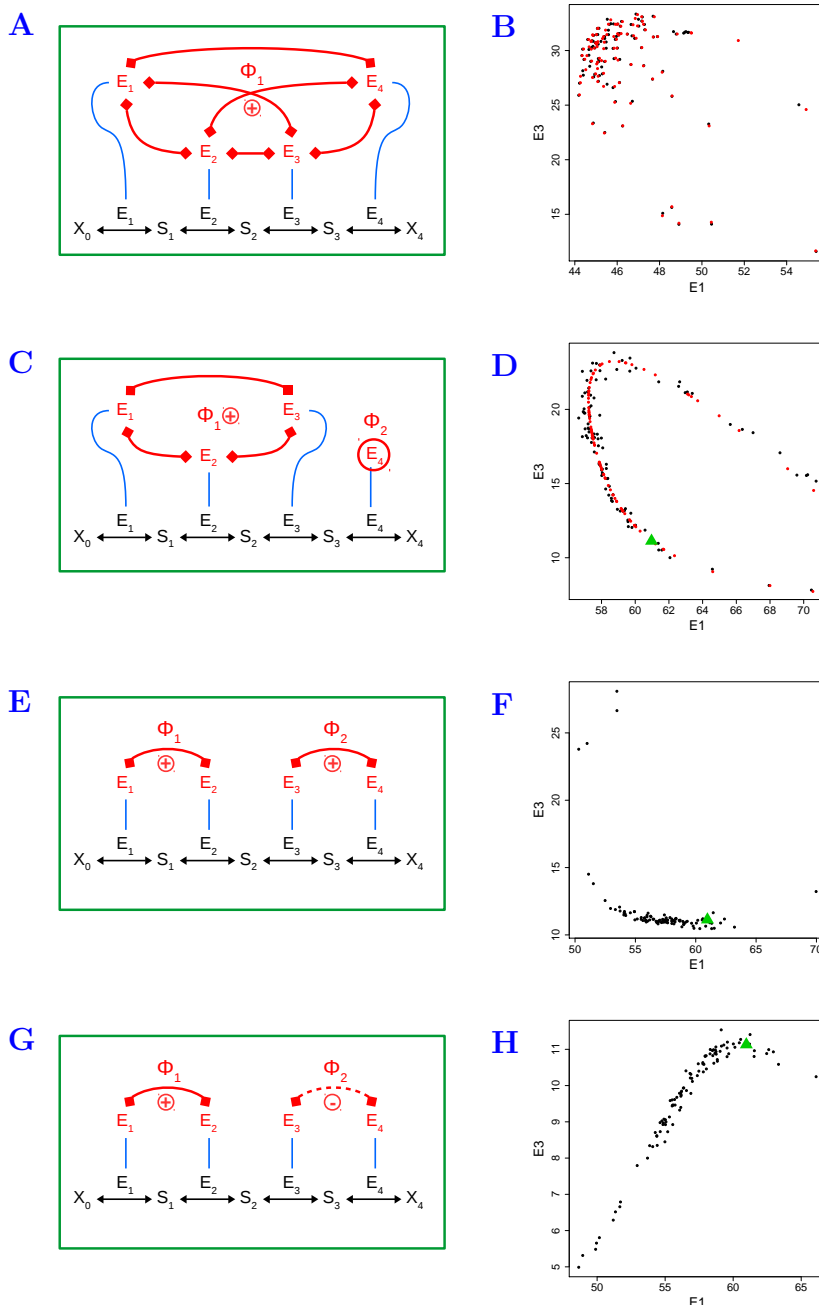


**Figure 5. Simulation of enzyme evolution in a four-enzyme pathway with one co-regulation group and one singleton when there is competition.** Simulations run for 125 000 time steps. (A) Scheme of the four-enzyme pathway (same symbols as in figure 1). Enzymes  $E_1$ ,  $E_2$  and  $E_3$  form a negative group, and enzyme  $E_4$  is independent (singleton). The total concentration is fixed. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = 0.32$ ,  $\beta_{23} = -0.43$ . (B) Relationship between flux  $J$  and total relative concentration  $e_2$  of enzyme 2. Ten simulations are represented, each with specific initial concentrations (one color per simulation). The graphic is zoomed from the grey frame in the inset, which is the entire graphic. Colored vertical dotted lines: abscissa of the effective equilibria of total relative concentrations  $\tilde{e}_i$ . Black dashed curve: envelop curve of the dome when there is no co-regulations, for which the system cannot go outside. Black point: summit of the envelop curve, which is the maximal flux  $J_{\max}$  without co-regulations. (C) Same representation as in B for enzyme 3.



**Figure 6. Simulation of enzyme evolution in a four-enzyme pathway with two co-regulation groups when there is competition.** Simulations run for 125 000 time steps. (A) Scheme of the four-enzyme pathway (same symbols as in figure 1). Enzymes  $E_1$  and  $E_2$  are positively co-regulated and enzymes  $E_3$  and  $E_4$  are negatively co-regulated. The total concentration is fixed. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = 0.32$ ,  $\beta_{43} = -0.43$ . (B and C) Same representation as in (5B) and (5C) for enzymes 2 and 3. In this case, the effective equilibrium is not computable.





**Figure 7. Simulations of evolution of enzyme concentrations for different types of co-regulation groups with competition.** *Left column:* Different co-regulation groups in a four-enzyme pathway. Same symbols as in figure 1. Parameter values: **A** = (1, 10, 30, 50),  $X = 1$ ,  $N = 10000$ . *Right column:* Relationship between the concentrations of enzymes  $E_1$  and  $E_3$  after 200 000 time steps. One hundred simulations were performed, which differed for the initial concentrations. Each black point corresponds to the end of one simulation. The red points correspond to the predicted effective equilibria, when computable. The green triangles indicate the enzyme concentrations at the maximum flux when there is no co-regulation ( $E_1 \approx 61$  and  $E_3 \approx 11.1$ ). (**A** and **B**) All enzymes are positively co-regulated ( $\beta_{12} = 0.32$ ,  $\beta_{23} = 2$ ,  $\beta_{34} = 0.1$ ). The simulated and predicted effective equilibria are close to each other, which confirms the results of our analysis. (**C** and **D**) Enzymes  $E_1$ ,  $E_2$  and  $E_3$  are positively co-regulated and enzyme  $E_4$  is independent ( $\beta_{12} = 0.1$ ,  $\beta_{23} = 2$ ). The simulated effective equilibria are close to the curve described by the predicted effective equilibria. (**E** and **F**) There are two groups of positively co-regulated enzymes,  $E_1$  and  $E_2$  on the one hand and  $E_3$  and  $E_4$  on the other hand ( $\beta_{12} = 0.32$ ,  $\beta_{34} = 0.1$ ). We hypothesize that the curve described by the black points follow the curve of effective equilibria that we could not compute. (**G** and **H**) The enzymes  $E_1$  and  $E_2$  are positively co-regulated and enzymes  $E_3$  and  $E_4$  are negatively co-regulated ( $\beta_{12} = 0.32$ ,  $\beta_{43} = -0.43$ ). The black points also describe a curve, but unlike the previous case, the relationship between  $E_1$  and  $E_3$  is positive.

685 **Appendix A Glossary of mathematical symbol****Latin symbols**

$A_i$	Pseudo-activity of enzyme $i$ , a composite parameter including the kinetic properties (catalytic constant and Michaelis-Menten constant) of the enzyme and thermodynamic equilibrium of upstream reactions)
$\mathbf{A}$	Vector of enzyme pseudo-activities
$A^{q*}$	Apparent activity of group $\Phi_q$ at theoretical equilibrium without competition
$\tilde{A}^q$	Apparent activity of group $\Phi_q$ at effective equilibrium with competition
$B_i$	Global co-regulation coefficient, representing effect of a variation of $E_i$ on $E_{\text{tot}}$ due to co-regulation, and also of $E_i$ on $E^q$
$E_i$	Concentration of enzyme $i$
$E_i^0$	Initial concentration of enzyme $i$
$E_i^m$	Concentration of a mutant enzyme $i$
$\tilde{E}_i$	Concentration of enzyme $i$ at effective equilibrium
$\mathbf{E}$	Vector of enzyme concentrations, considered as the genotype
$E^q$	Sum of enzyme concentrations in $\Phi_q$
$E^{q0}$	Initial sum of enzyme concentrations in $\Phi_q$
$E_{\text{tot}}$	Total enzyme concentration in the metabolic pathway
$E_{\text{tot}}^0$	Initial total enzyme concentration
$e_i$	Total relative concentration of enzyme $i$ ; $e_i = E_i/E_{\text{tot}}$
$e_i^0$	Initial total relative concentration of enzyme $i$
$e_i^*$	Total relative concentration of enzyme $i$ at theoretical equilibrium
$\tilde{e}_i$	Total relative concentration of enzyme $i$ at effective equilibrium
$e_i^q$	Intra-group relative concentration of enzyme $i$ in $\Phi_q$ ; $e_i^q = E_i/E^q$
$e_i^{q0}$	Initial intra-group relative concentration of enzyme $i$ in $\Phi_q$
$e_i^{q*}$	Intra-group relative concentration of enzyme $i$ in $\Phi_q$ at theoretical equilibrium
$e_i^q$	Intra-group relative concentration of enzyme $i$ in $\Phi_q$ at effective equilibrium
$e^q$	Inter-group relative concentration of enzyme of $\Phi_q$ in regard to total concentration; $e^q = E^q/E_{\text{tot}}$
$e^{q*}$	Inter-group relative concentration of enzyme $i$ in $\Phi_q$ at theoretical equilibrium
$\tilde{e}^q$	Inter-group relative concentration of enzyme $i$ in $\Phi_q$ at effective equilibrium
$\mathcal{E}^q$	Line in the space of intra-group relative enzyme concentrations along which the intra-group relative concentrations vary in the case of co-regulation for each group $q$
$J$	Metabolic flux, considered as a phenotype that is proportional to fitness
$J_i^m$	Flux resulting from a mutation targeting enzyme $i$
$J^r$	Flux of the resident
$\tilde{J}_{\text{max}}$	Maximal value of the flux, corresponding to the effective equilibrium
$\mathbf{M}_\beta$	Matrix of co-regulation coefficients
$m_q$	Number of enzymes in $\Phi_q$
$N$	Effective population size
$n$	Number of enzymes in the metabolic pathway
$p$	Total number of co-regulation groups
$R_{E_i}^J$	Flux response coefficient of enzyme $i$
$S_i$	Metabolite $i$ , product of reaction $i$ and substrate of reaction $i + 1$
$s_i$	Selection coefficient of a mutation targeting enzyme $i$
$X$	Constant representing the environment, depending on the concentrations of the input substrate $X_0$ and the output product $X_n$ of the pathway

**Greek symbols**

$\alpha_{ij}$	Redistribution coefficient, expressing the variation of $E_j$ due to a mutation targeting $E_i$
$\beta_{ij}$	Co-regulation coefficient of enzyme $i$ on $j$ , expressing the change of $E_j$ due to co-regulation when $E_i$ varies
$\delta_i$	Actual effect of a mutation affecting enzyme $i$
$\theta_q$	Indicator variable indicating the type of co-regulation in the co-regulation group $q$
$\mu$	Mutation rate at each time unit
$\nu$	Canonical effect of a mutation
$\tau^q$	Driving variable of the line $\mathcal{E}^q$ in the space of intra-group relative concentrations for enzymes in the group $\Phi_q$
$\tilde{\tau}^q$	Value of the driving variable at the effective equilibrium for the group $\Phi_q$
$\Phi_q$	Co-regulation group $q$

**Other symbols**

0	Exponent that refers to the initial state of the system
*	Exponent that refers to the theoretical equilibrium
~	Refers to the effective equilibrium

## 2.3 Discussion complémentaire

En cas de compétition, j'ai cherché les équilibres évolutifs en utilisant une autre méthode. En décomposant le flux en fonction des groupes de co-régulation, on obtient :

$$J = \frac{X}{\sum_{q=1}^p \frac{1}{E^q} \sum_{i \in \Phi_q} \frac{1}{A_i e_i^q}} \quad (2.1)$$

En supposant que les groupes sont indépendants, je me suis dit que minimiser les « sous-sommes »  $\sum_{i \in \Phi_q} \frac{1}{A_i e_i^q}$  par rapport à  $\tau^q$  reviendrait à maximiser le flux.

Cette méthode donne le même résultat que pour la recherche des équilibres intra-groupes, lorsqu'il y a un seul groupe, positif ou négatif, avec uniquement des enzymes indépendantes (annexe B.2). Toutefois, lorsque plusieurs groupes co-existent, les relations entre les  $\tau^q$  deviennent complexes, voire imprévisibles. Dans mes simulations, j'ai pu observer que selon les enzymes ciblées par les mutations, les variations des  $\tau^q$  sont différentes.



---

## **CHAPITRE 3**

---



## Chapitre 3

# La relation enzymes-flux dans les réseaux métaboliques : analogie avec les circuits électriques

Dans les chapitres précédents, nous avons étudié l'évolution des concentrations d'enzymes sous sélection pour augmenter le flux dans une chaîne métabolique linéaire. Cependant les voies métaboliques sont interconnectées, formant le réseau métabolique de la cellule, au sein duquel les métabolites peuvent être produits ou consommés par plusieurs réactions. Le formalisme de la relation enzymes-flux que nous avons utilisé, issu de la Théorie du Contrôle Métabolique (TCM), ne s'applique tel quel qu'aux chaînes linéaires. Dans un réseau de topologie quelconque, comment le flux global est-il relié aux propriétés des enzymes ? C'est la question à laquelle il faut répondre pour étudier l'évolution des concentrations d'enzymes dans le cas général et savoir si les résultats décrits dans les chapitres précédents s'appliquent, au moins qualitativement, aux réseaux métaboliques.

Il existe diverses approches pour analyser le fonctionnement des réseaux métaboliques, de la théorie des réseaux à l'analyse stoechiométrique, en passant par l'exploitation de l'information sur les propriétés des métabolites ou la relation structure/fonction des enzymes [[Hatzimanikatis et al. 2004](#)]. Je présenterai d'abord les principaux modèles connus qui relient les enzymes au flux d'un réseau et montrerai en quoi ils me semblent inadéquats pour répondre à la question posée. Dans un deuxième temps je développerai l'analogie entre les réseaux métaboliques et les circuits électriques, qui m'a permis d'exprimer le flux en fonction de toutes les enzymes du réseau et de montrer – sous réserve d'hypothèses simplificatrices – que la relation entre la concentration d'une enzyme et le flux d'un réseau était généralement concave et bornée quelle que soit la complexité du réseau.



## 3.1 Modéliser la relation enzymes-flux dans les réseaux métaboliques

### 3.1.1 Les modèles stœchiométriques

Les méthodes de *Constraint-based modeling* (CBM [Price *et al.* 2003]) ont été développées pour étudier le métabolisme à l'échelle du génome. Dans le principe, il s'agit d'estimer les valeurs de l'ensemble des flux en se basant sur la stœchiométrie. Mathématiquement, on résout le système  $Sv = 0$ , où  $S$  est la matrice des nombres stœchiométriques pour les métabolites et  $v$  est la matrice des vitesses de réaction, en considérant les réactions chimiques à l'état stationnaire. Comme il existe une infinité de solutions, il est nécessaire d'appliquer des contraintes biologiquement réalistes (pour revue, voir Petrizzelli [2019]). Par exemple dans la FBA (*Flux-Balance Analysis* [Orth *et al.* 2010]), qui est la plus ancienne méthode de CBM, on cherche les solutions de  $Sv = 0$  qui maximisent un phénotype d'intérêt, comme la biomasse ou le taux de croissance cellulaire. Pour restreindre davantage l'espace des solutions, on peut ajouter les propriétés enzymatiques connues comme contraintes supplémentaires. Par exemple, dans la *Metabolic Modeling with ENzyme kinetics*, ou MOMENT [Adadi *et al.* 2012], on implémente les constantes de catalyse  $k_{cat}$  ainsi que le poids moléculaire des enzymes, calculé à partir de la taille des gènes, puis on considère la masse totale des enzymes présentes. Les concentrations des enzymes sont alors estimées à partir des  $k_{cat}$  et des flux optimaux. Dans d'autres approches, comme la IOMA (*Integrative Omics-Metabolic Analysis* [Yizhak *et al.* 2010]), les flux sont contraints par les concentrations en métabolites et l'abondance des protéines, en supposant les flux proportionnels aux abondances d'enzymes.

Ces différentes méthodes permettent de calculer des vitesses de réactions optimales, mais ne permettent pas de relier le flux d'entrée (ou de sortie) d'un réseau métabolique quelconque avec les concentrations d'enzymes particulières. Ce sont des méthodes prédictives plutôt qu'explicatives, qui cherchent à prédire un phénotype donné à partir de l'ensemble des réactions chimiques de la cellule. Par ailleurs, les méthodes de CBM sont des alternatives aux modèles basés sur la cinétique enzymatique [Price *et al.* 2003], et ne sont donc pas conçues pour relier explicitement le flux global d'un système aux propriétés de chacune des enzymes. On peut donc difficilement les utiliser pour étudier l'évolution des propriétés enzymatiques.

### 3.1.2 Application de la TCM aux réseaux

#### a) Utilisation du paramètre de « dispensabilité »

En se basant sur la cinétique enzymatique, on peut relier le flux aux propriétés des enzymes – c'est le fondement de la TCM [Kacser *et al.* 1995]. Le formalisme de la TCM peut être étendu aux réseaux métaboliques en introduisant un paramètre de « dispensabilité » dans l'expression

du flux [Fiévet *et al.* 2006, 2010] :

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j + d_j E_{\text{tot}}}} \quad (3.1)$$

où  $X$  est une constante,  $A_j$  est la pseudo-activité de l'enzyme  $j$ ,  $E_j$  la concentration de l'enzyme  $j$ ,  $E_{\text{tot}}$  la concentration totale d'enzyme et  $d_j$  la « dispensabilité » de l'enzyme  $j$ . Lorsque  $d_j \neq 0$ , le flux ne s'annule pas si  $E_j$  tombe à zéro : l'enzyme  $j$  est « dispensable ». Lorsque  $d_j = 0$ , le flux s'annule si  $E_j$  tombe à zéro : l'enzyme est indispensable. Le paramètre  $d_j$  peut s'estimer à partir de l'ajustement hyperbolique de la courbe de titration obtenue en faisant varier la concentration de l'enzyme  $j$  de 0 à une valeur saturante pour le flux, les autres concentrations étant fixées [Fiévet *et al.* 2006].

Le paramètre de dispensabilité inclut de multiples paramètres comme variables cachées, dont les concentrations des autres enzymes du réseau. Ce n'est donc pas seulement une propriété de l'enzyme concernée mais aussi une variable systémique, que je n'ai pas vu comment intégrer dans le modèle d'évolution conjointe des enzymes d'un réseau.

## b) Utilisation de l'analogie entre les réseaux métaboliques et les circuits électriques

La méthode que j'ai choisie s'appuie également sur le formalisme de la TCM, mais en exploitant une analogie entre les réseaux métaboliques et les circuits électriques. Le flux global d'un réseau peut en effet être assimilé à l'intensité parcourant un circuit électrique, la conductance d'une résistance de ce circuit jouant un rôle similaire à celui de la concentration ou de l'activité enzymatique. Des développements théoriques récents dans le domaine des circuits électriques m'ont permis d'aborder la question de la relation entre l'efficacité des enzymes et le flux global pour un réseau de complexité quelconque.

## 3.2 L'analogie métabolisme-électricité

### 3.2.1 Historique de l'analogie entre réseaux métaboliques et circuits électriques

Le recours à des analogies est classique en recherche, tant pour leur intérêt pédagogique que scientifique, car elle permettent d'aborder l'étude de phénomènes complexes à l'aide de systèmes plus intuitifs dont le formalisme est connu [Chaumat *et al.* 1989]. En physique, on peut citer l'analogie entre les étoiles à neutrons et des fontaines à eau, qui aide à comprendre la dynamique des gaz stellaires [Foglizzo *et al.* 2012], ou encore celle de l'électrostatique et de la propagation de la chaleur [Chaumat *et al.* 1989], celle de l'électricité et de l'hydraulique [Closset 1992; Brochier et Guillemaud 2020; Oh *et al.* 2012], de l'électricité et de la mécanique [Dequidt 2015; Olson 1958], de l'électricité et de l'acoustique [Olson 1958], etc. En biologie,

on peut mentionner les analogies entre les réseaux métaboliques et les circuits économiques [Liebermeister 2014; Sugita 1954], entre les circuits électriques et les réseaux de transcription pour étudier leur modularité et leur rétroactivité [Del Vecchio *et al.* 2008], entre les circuits électriques et les réseaux de signalisation [Bhalla 2003], enfin entre les circuits électriques et les réseaux de régulation, par exemple pour expliquer la régulation de la glycolyse par les métabolites [Stavrianeas et Silverstein 2005].

À une exception près (voir ci-après), l'analogie métabolisme-électricité est évoquée dans la bibliographie en tant que *toy model* pour illustrer certains mécanismes biologiques. Par exemple Milborrow [1998] l'utilise pour expliquer son modèle d'hétérosis biochimique. Il imagine quatre gènes bi-alléliques codant des enzymes. Les enzymes sont assimilées à des résistances. Chaque enzyme forme une boucle avec deux résistances en parallèle (les produits de chaque allèle), et les quatre boucles sont en série. En calculant la résistance globale du circuit pour des parents homozygotes et leur descendance F1, il montre que la descendance F1 peut avoir une résistance plus faible que celle de la moyenne parentale, et donc une intensité électrique plus élevée, c'est-à-dire que la descendance F1 est hétérotique. Plus récemment, Yi et Dean [2019] reprennent l'analogie métabolisme-électricité pour illustrer les relations entre le fonctionnement de la voie du lactose et la fitness chez *Escherichia coli*. En assimilant l'intensité électrique au flux de lactose, les potentiels électriques aux concentrations en métabolites et la conductance électrique à l'efficacité enzymatique, ils recourent à la loi d'Ohm pour faire comprendre intuitivement la relation entre le flux métabolique, proportionnel à la fitness, et l'efficacité enzymatique. Les autres mentions de cette analogie dans la bibliographie sont encore moins poussées, et ne sont utilisées qu'en tant que métaphores à des fins pédagogiques.

À ma connaissance, l'exploitation la plus élaborée de cette analogie est celle de la *Thermodynamic-Kinetic Modeling* (TKM) [Ederer et Gilles 2007]. L'objectif de la TKM est d'inclure explicitement la thermodynamique dans l'étude des réseaux métaboliques, en assimilant les métabolites à des charges électriques, ce qui permet d'utiliser des concepts issus de l'électricité. La TKM repart de la notion de potentiel chimique afin de s'affranchir des constantes d'équilibre, en définissant un potentiel thermocinétique propre à chaque composé chimique. Ceci permet de calculer une « force thermocinétique » de la réaction analogue à la tension électrique. En connaissant la vitesse de réaction, analogue à l'intensité électrique, on peut alors déterminer une résistance thermocinétique associée à cette réaction, directement analogue à la résistance électrique. Ce formalisme, riche de notions et de notations spécifiques, m'a semblé difficile à prendre en main, notamment pour l'appliquer à l'évolution des paramètres enzymatiques.

Je vais donc présenter ici un autre formalisme pour l'analogie métabolisme-électricité, également basé sur la cinétique des réactions chimiques, mais qui s'appuie sur la TCM et qui est donc plus adapté pour répondre aux questions que nous nous posons. À terme, ce formalisme pourra être utilisé pour étudier l'évolution des enzymes.

### 3.2.2 Rappels sur le fonctionnement des réactions chimiques

Je reprends ici les notions de base sur le fonctionnement des réactions chimiques catalysées par des enzymes, nécessaires à la compréhension des notions analogues entre métabolisme et électricité, d'un point de vue thermodynamique puis cinétique, comme l'ont fait [Liebermeister et al. \[2010\]](#) et [Ederer et Gilles \[2007\]](#).

#### a) Réactions chimiques et thermodynamique

Cette partie aborde quelques notions de base de thermodynamique chimique, en laissant de côté celles qui ne sont pas nécessaires pour la compréhension de l'analogie, comme l'avancement de la réaction ou l'affinité chimique.

Une réaction chimique est une transformation de constituants chimiques, les réactifs, en d'autres composés chimiques, les produits. L'équation bilan d'une réaction s'écrit :

$$\sum_{i=1}^n \nu_i C_i = 0 \quad (3.2)$$

où  $C_i$  est le composé chimique  $i$  impliqué dans la réaction, et  $\nu_i$  le nombre stœchiométrique associé, sachant que  $\nu_i$  est positif si  $C_i$  est un produit ou négatif si  $C_i$  est un réactif. Au minimum, une réaction comprend un réactif et un produit.

Chaque composé chimique  $i$  possède un potentiel chimique  $\mu_i$ , exprimé en joule par mole (J/mol). Les potentiels chimiques permettent de définir l'enthalpie libre de réaction  $\Delta_r G$  (aussi appelée énergie libre de Gibbs) qui décrit la variation d'énergie au cours de la réaction. À pression et température constantes, on a :

$$\Delta_r G = \sum_i \nu_i \mu_i$$

Selon les premier et deuxième principes de la thermodynamique, la réaction sera spontanée des réactifs vers les produits (on dit que la réaction progresse) si  $\Delta_r G$  est négatif, c'est-à-dire si l'enthalpie de réaction  $\Delta_r H$  est négative (réaction exothermique) et/ou si l'entropie de réaction  $\Delta_r S$  est positive (augmentation du désordre). Au contraire, si  $\Delta_r G$  est positif, la réaction régresse spontanément, dans le sens des produits vers les réactifs. La réaction est dite à l'équilibre lorsque  $\Delta_r G = 0$  : l'énergie présente dans le système est alors minimale, et est en quelque sorte répartie entre les différents constituants.

Cependant, le potentiel chimique étant difficile à mesurer de façon absolue, les notions d'activité chimique  $a_i$  et de potentiel chimique standard  $\mu_i^\circ$  ont été introduites :

$$\mu_i = \mu_i^\circ + RT \ln(a_i)$$

où  $R$  la constante des gaz parfaits et  $T$  la température absolue (en Kelvin). L'activité chimique

est directement mesurable, tandis que le potentiel chimique standard est défini par rapport à un état standard conventionnel, et est tabulé. Par exemple, l'activité chimique pour un composé  $i$  dans une solution idéale est proportionnelle à sa concentration  $c_i$  :

$$a_i = \frac{c_i}{c^\circ}$$

où  $c^\circ$  est la concentration à l'état standard, définie à 1 mol/L. Dans une cellule, l'activité chimique est souvent modulée par de nombreux paramètres, par exemple le *molecular crowding* [Ellis 2001]. Pour simplifier, on se place dans le cas d'une solution idéale à pression, température et volume constants pour laquelle  $a_i \approx c_i$ .

A partir de ces notions, on peut réécrire l'enthalpie libre de réaction :

$$\Delta_r G = \Delta_r G^\circ + RT \ln(Q_r)$$

où  $\Delta_r G^\circ = \sum_i \nu_i \mu_i^\circ$  est l'enthalpie libre standard de réaction et  $Q_r = \prod_i a_i^{\nu_i}$  est le quotient de réaction (avec les produits au numérateur et réactifs au dénominateur). À l'équilibre  $\Delta_r G = 0$ , et le quotient de réaction dépend alors uniquement des potentiels chimiques standards. On en extrait la notion de constante d'équilibre  $K_{\text{eq}}$  :

$$K_{\text{eq}} = \exp\left(-\frac{\Delta_r G^\circ}{RT}\right)$$

et donc :

$$\Delta_r G = RT \ln\left(\frac{Q_r}{K_{\text{eq}}}\right)$$

Si  $Q_r < K_{\text{eq}}$ , alors  $\Delta_r G < 0$  et la réaction progresse des réactifs vers les produits, tandis que si  $Q_r > K_{\text{eq}}$ , alors  $\Delta_r G > 0$  la réaction régresse (des produits vers les réactifs). À l'équilibre, nous avons donc  $Q_r = K_{\text{eq}}$  et les solutés atteignent leur activité à l'équilibre :

$$K_{\text{eq}} = \prod_i a_{i, \text{eq}}^{\nu_i}$$

Pour résumer, la constante d'équilibre d'une réaction chimique, ou plus exactement l'écart à l'équilibre, représente le « moteur » de cette réaction : les composés chimiques évoluent vers leur concentration à l'équilibre, état qui correspond à l'énergie minimale du système.

Il existe de nombreux moyens de calculer les constantes d'équilibre. Leurs valeurs sont généralement disponibles dans la base de données *eQuilibrator* [Flamholz *et al.* 2012]. Cependant, la thermodynamique donne des informations sur le sens de la réaction, mais pas sur sa vitesse.

## b) Cinétique enzymatique

Dans une réaction chimique uni-moléculaire réversible, avec un substrat S et un produit P, de la forme



la vitesse  $v$  de la réaction, qui est à la fois la vitesse de consommation du substrat et la vitesse de production du produit, s'écrit :

$$v = \frac{d[P]}{dt} = -\frac{d[S]}{dt} = v^+ - v^- = k^+[S] - k^-[P]$$

où  $v^+$  et  $v^-$  sont les vitesses de réaction dans le sens *forward* ( $S \rightarrow P$ ) et *reverse* ( $S \leftarrow P$ ) respectivement,  $k^+$  et  $k^-$  sont les coefficients de vitesse des réactions *forward* et *reverse*, et  $[S]$  et  $[P]$  sont respectivement les concentrations du substrat et du produit.

À l'équilibre, la vitesse de la réaction est nulle et les concentrations atteignent leur valeur à l'équilibre, ce qui donne la relation dite de Haldane [Haldane 1930] :

$$K_{\text{eq}} = \frac{[P]_{\text{eq}}}{[S]_{\text{eq}}} = \frac{k^+}{k^-}$$

L'équilibre est dynamique : macroscopiquement, la réaction n'avance plus, mais au niveau microscopique les composés sont consommés et produits à la même vitesse.

La vitesse de la réaction peut se réécrire en fonction de la constante d'équilibre :

$$v = k^+ \left( [S] - \frac{[P]}{K_{\text{eq}}} \right) \quad (3.3)$$

La vitesse de la réaction peut être modifiée par différents mécanismes, mais cela n'influence pas l'état d'équilibre qui est entièrement déterminé par la thermodynamique. Notamment les enzymes, en tant que catalyseurs biologiques accélérant considérablement les réactions, vont modifier l'expression de la vitesse.

## c) Flux et état stationnaire

Dans certains articles, flux métabolique  $J$  et vitesse de réaction  $v$  sont confondus. Pour éviter toute ambiguïté je distinguerai ici les deux, comme l'ont fait Kacser et Burns [1973] dès l'origine de la TCM. La vitesse  $v$  sera réservée à *une* réaction, tandis que le flux métabolique  $J$  concernera le débit de consommation du métabolite d'entrée du réseau, égal au débit de production du métabolite de sortie de ce réseau (on supposera qu'il n'y a qu'un métabolite d'entrée et qu'un métabolite de sortie). Le flux dépend ainsi de l'ensemble des réactions du réseau.

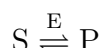
À l'état stationnaire, les concentrations et les vitesses des réactions sont constantes. Bien que les réactions soient réversibles, le système n'est pas à l'équilibre, car on considère un système ouvert, où les concentrations en substrat et produit de la voie métabolique sont maintenues

constantes, par apport de substrat et retrait de produit.

### 3.2.3 Cadre de l'analogie métabolisme-électricité et équivalence des grandeurs

Dans cette partie, je détaille le cadre de l'analogie métabolisme-électricité basée sur la cinétique enzymatique, ainsi que les grandeurs équivalentes entre les deux systèmes.

Au niveau métabolique, on considère un ensemble de réactions, toutes réversibles et uni-moléculaires, chacune étant catalysée par une enzyme michaelienne. Les réactions sont de la forme :



où S est le substrat et P le produit de la réaction, et E l'enzyme catalysant la réaction.

On se place à l'état stationnaire, c'est-à-dire que les vitesses de réactions et les concentrations en métabolites sont considérées comme constantes. La cinétique de cette réaction s'écrit [Haldane 1930] :

$$v = [E] \frac{k_{\text{cat}}}{K_M} \frac{(X_S - X_P/K_{\text{eq}})}{1 + X_S/K_M + X_P/K_M^-}$$

où  $[E]$  est la concentration de l'enzyme,  $k_{\text{cat}}$  la constante catalytique de l'enzyme,  $K_{\text{eq}}$  la constante d'équilibre,  $K_M$  la constante de Michaelis-Menten *forward* (représentant l'affinité de l'enzyme pour le substrat),  $K_M^-$  est la constante de Michaelis-Menten *reverse* (affinité pour le produit), et  $X_S$  et  $X_P$  sont respectivement la concentration en substrat et la concentration en produit de la réaction. Lorsque l'enzyme est loin de la saturation, ce qui semble être la règle car la concentration en substrat est généralement faible dans la cellule, on a  $X_S \ll K_M$  et  $X_P \ll K_M^-$ , et l'équation se simplifie en :

$$v \approx [E] \frac{k_{\text{cat}}}{K_M} \left( X_S - \frac{X_P}{K_{\text{eq}}} \right) \quad (3.4)$$

Par la suite, on notera  $F$  l'efficacité enzymatique :  $F = [E] \frac{k_{\text{cat}}}{K_M}$ .

La relation 3.4 est similaire à celle de la loi d'Ohm, qui exprime l'intensité  $I$  du courant traversant un dipôle en fonction de la tension  $U$  aux bornes du dipôle et de la résistance  $R$  de ce dipôle<sup>1</sup> :

$$I = \frac{U}{R} \quad (3.5)$$

En régime continu, les grandeurs  $U$  et  $I$  sont constantes.

Les relations 3.4 et 3.5 comprennent trois termes homologues :

<sup>1</sup>En français, le terme « résistance » fait référence à la fois au composant électrique et à la grandeur, contrairement à l'anglais, où *resistor* désigne le composant électrique et *resistance* la grandeur électrique. Pour éviter les confusions, je réserve le terme « résistance » ( $R$  italique) à la grandeur électrique, et vais désigner le composant électrique par le terme « dipôle » ( $R$  droit), un dipôle étant simplement un composant possédant deux bornes.

- la partie cinétique est le débit, de matière (vitesse  $v$ ) ou de charges électriques (intensité  $I$ )
- la partie « moteur » représente la différence d'énergie entre l'entrée et la sortie de la réaction (écart à l'équilibre chimique  $X_S - X_P/K_{eq}$ ) ou du dipôle (tension  $U$ )
- la partie modulant la vitesse de transformation de la matière (enzyme d'efficacité  $F$ ) ou le passage de l'électricité (résistance  $R$ ) permet le passage de l'énergie.

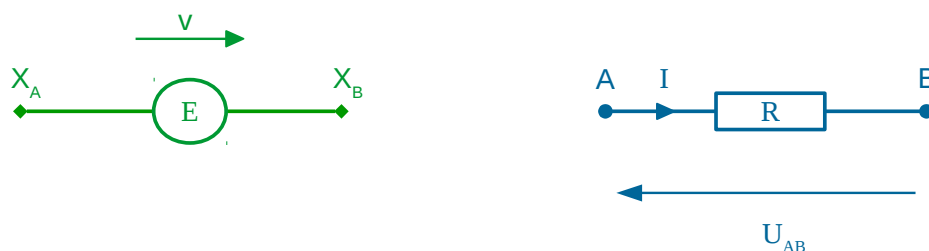


FIGURE 3.1 – **Résistance et enzyme.** Représentation graphique des éléments, métaboliques (en vert à gauche) et électriques (en bleu à droite), sur lequel se base l'analogie métabolisme-électricité. L'enzyme  $E$  d'efficacité  $F$  catalyse la réaction réversible uni-moléculaire  $X_A \rightleftharpoons X_B$  de constante d'équilibre  $K_{eq}$  avec une vitesse de réaction  $v$ ,  $X_A$  étant le substrat de l'enzyme et  $X_B$  son produit. Le dipôle  $R$  de résistance  $R$  situé entre les points  $A$  et  $B$  est traversé par un courant d'intensité  $I$ , avec une tension  $U_{AB}$  à ses bornes. Le dipôle électrique est représenté en convention récepteur : la tension et l'intensité sont représentées dans des sens opposés, afin que les deux grandeurs soient positives. En effet, les électrons circulent dans le sens opposé à l'intensité, et comme ils possèdent une charge négative, et donc une énergie potentielle électrique négative, ils circulent du potentiel électrique le plus faible (en  $B$ ) vers le plus élevé (en  $A$ ). Ainsi,  $I$  est positif, et  $U_{AB} = V_A - V_B$  l'est également. Par ailleurs, le sens de la réaction est important, car il détermine  $K_{eq}$ , mais aussi les  $k_{cat}$  et les  $K_M$  des enzymes, leur valeur différant selon que le sens *forward* de la réaction est initialement défini de  $A$  vers  $B$  ou de  $B$  vers  $A$ . Une fois le sens *forward* défini,  $v_{AB} = -v_{BA}$ , tout comme  $I_{AB} = -I_{BA}$ .

La figure 3.1 illustre cette analogie en mettant en parallèle une enzyme  $E$  catalysant la réaction  $X_A \rightleftharpoons X_B$  et un dipôle  $R$  situé entre deux bornes  $A$  et  $B$ . La vitesse de la réaction à l'état stationnaire s'écrit alors dans ce cas :

$$v = F \left( X_A - \frac{X_B}{K_{eq}} \right) \quad (3.6)$$

et l'intensité du courant qui traverse le dipôle  $R$  de résistance  $R$  est :

$$I = \frac{U_{AB}}{R} = \frac{1}{R}(V_A - V_B) \quad (3.7)$$

où  $U_{AB}$  est la tension entre  $A$  et  $B$  et  $V_A$  et  $V_B$  sont respectivement les potentiels électriques en  $A$  et  $B$ .

La conductance  $\sigma = 1/R$  est analogue à l'efficacité enzymatique  $F$ . En effet, plus une enzyme est concentrée, active et/ou affine avec son substrat, plus la réaction sera rapide. En



électricité, plus un matériau facilite le passage des électrons, plus sa conductance et donc l'intensité sont élevées.

La tension, ou différence de potentiels électriques, est analogue à la différence de concentrations en métabolites. Cependant, la constante d'équilibre  $K_{eq}$  affecte cette différence de concentrations car les composés chimiques possèdent des énergies potentielles différentes à l'équilibre, contrairement aux électrons dont l'énergie interne est uniforme à travers le circuit.

Les métabolites représentent les nœuds du réseau métabolique. La concentration en métabolite est ici analogue au potentiel électrique. Pour être plus exact, le potentiel électrique devrait être analogue au potentiel chimique, qui est directement relié à la concentration ( $\mu_i = \mu_i^\circ + RT \ln(X_i/c^\circ)$ , voir section 3.2.2), mais cela compliquerait les calculs notamment à cause du logarithme [Ederer et Gilles 2007].

Les grandeurs équivalentes entre métabolisme et électricité sont résumées dans le tableau 3.1.

TABLE 3.1 – **Grandeurs équivalentes entre électricité et métabolisme.**

<b>Électricité</b>	<b>Métabolisme</b>
Circuit électrique	Voie métabolique
Dipôle de type résistance	Enzyme michaelienne
Nœud	Métabolite X
Potentiel électrique $V$	Concentration en métabolite $X$
Intensité $I$	Vitesse de réaction $v$ et flux $J$
Tension $U_{AB}$	Écart à l'équilibre $X_A - X_B/K_{eq}$
Conductance $\sigma = 1/R$	Efficacité enzymatique $F = [E]k_{cat}/K_M$
$\emptyset$	Constante d'équilibre $K_{eq}$ (sans dimension)
Loi d'Ohm (équation 3.7)	Cinétique de réaction (équation 3.6)
Régime continu	État stationnaire

La constante d'équilibre n'a pas d'équivalence en électricité. En effet, en l'absence de générateur (ou si on le coupe) la tension est nulle, soit  $V_A = V_B$  pour deux nœuds A et B ; les électrons ne circulent plus. Pour une réaction chimique, la réaction est à l'équilibre (équilibre dynamique puisque la réaction continue à se faire dans les deux sens) si les métabolites atteignent leurs concentrations à l'équilibre, exprimées dans la constante d'équilibre  $K_{eq} = \frac{X_B^{eq}}{X_A^{eq}}$ . Cette différence vient du fait que les électrons possèdent tous la même quantité d'énergie, tandis que chaque constituant chimique porte une énergie différente (symbolisée par le potentiel chimique standard, voir section 3.2.2). L'équivalence métabolisme-électricité serait stricte si toutes les constantes d'équilibre étaient égales à 1.

### 3.2.4 Les lois électriques

Afin de savoir si les équations issues des circuits électriques peuvent s'appliquer aux réseaux métaboliques, il faut s'assurer que les lois classiquement utilisées en électricité – lois de Kirchhoff et loi d'Ohm – sont valides dans le cadre métabolique. Je m'appuierai pour cela sur la figure 3.2 qui représente un réseau métabolique simple et son équivalent électrique.

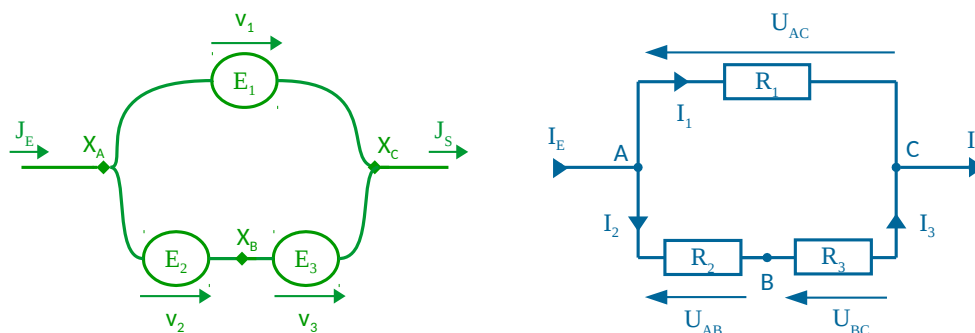


FIGURE 3.2 – **Exemple d'un réseau métabolique simple et de son équivalent électrique.** Réseau métabolique simple à gauche (en vert) et le circuit électrique équivalent à droite (en bleu). Les intensités entrante (à gauche vers A) et sortante (à droite depuis C) sont notées respectivement  $I_E$  et  $I_S$ , et les flux métaboliques correspondant sont notés  $J_E$  et  $J_S$ . Cet exemple apparaît dans le métabolisme de la sérine, avec d'un côté la transformation de la sérine en D-sérine par la sérine racémase (EC 5.1.1.18) puis en pyruvate par la D-sérine ammonia-lyase (EC 4.3.1.18), et de l'autre côté la transformation directe de la sérine en pyruvate par la L-sérine ammonia-lyase (EC 4.3.1.17).

#### a) Loi d'Ohm

*La tension aux bornes d'un dipôle est proportionnelle à l'intensité du courant qui le traverse.*

La loi d'Ohm relie l'intensité du courant  $I$  et la tension  $U$  aux bornes d'un dipôle de résistance  $R$  en régime continu. Elle est analogue à l'équation de la vitesse de réaction.

Pour la branche supérieure de la figure 3.2, on a pour la partie électrique :

$$I_1 = \frac{U_{AC}}{R_1} = \sigma_1 U_{AC}$$

où  $I_1$  est l'intensité du courant traversant le dipôle 1, et  $R_1$  et  $\sigma_1$  sont respectivement la résistance et la conductance du dipôle 1 ; et pour la partie métabolique :

$$v_1 = F_1 \left( X_A - \frac{X_C}{K_1} \right)$$

où  $v_1$ ,  $F_1$  et  $K_1$  sont respectivement la vitesse, l'efficacité enzymatique et la constante d'équilibre pour la réaction 1.

La vitesse de la réaction est proportionnelle à l'écart à l'équilibre.

**b) Loi des nœuds**

*Pour chaque nœud, la somme des intensités entrantes est égale à la somme des intensités sortantes.*

Mathématiquement, la loi des nœuds s'écrit :

$$\sum_{i \in \text{entrante}} I_i = \sum_{j \in \text{sortante}} I_j \quad (3.8)$$

Par exemple, pour le nœud A du circuit électrique de la figure 3.2, cela donne :

$$I_E = I_1 + I_2 \quad (3.9)$$

À l'état stationnaire dans un réseau métabolique, la concentration en métabolites est constante, ce qui donne par exemple pour le pool de métabolite  $X_A$  de la figure 3.2 :

$$\frac{dX_A}{dt} = J_E - v_1 - v_2 = 0 \quad (3.10)$$

où  $J_E$  est la vitesse de production et  $v_1$  et  $v_2$  sont les vitesses de consommation de  $X_A$ . On a donc :

$$J_E = v_1 + v_2 \quad (3.11)$$

Ainsi, le flux métabolique se divise de la même manière que l'intensité électrique.

Plus généralement, on peut écrire pour un pool de métabolites X :

$$\sum_{i \in \text{consommation de X}} v_i = \sum_{j \in \text{production de X}} v_j \quad (3.12)$$

ce qui est analogue à la loi des nœuds en électricité.

**c) Loi des mailles**

*La somme des tensions dans une maille d'un circuit est nulle.*

Pour le circuit électrique de la figure 3.2, cette loi s'écrit :

$$U_{AB} + U_{BC} + U_{CA} = 0 \Leftrightarrow U_{AC} = U_{AB} + U_{BC} \quad (3.13)$$

Pour la boucle métabolique équivalente, les écarts à l'équilibre analogues aux tensions  $U_{AC}$ ,

$U_{AB}$  et  $U_{BC}$  s'écrivent respectivement :

$$\begin{aligned} A \rightarrow C \text{ par } E_1 &: X_A - \frac{X_C}{K_1} \\ A \rightarrow B \text{ par } E_2 &: X_A - \frac{X_B}{K_2} \\ B \rightarrow C \text{ par } E_3 &: X_B - \frac{X_C}{K_3} \end{aligned}$$

L'application directe de la loi des mailles au métabolisme s'écrirait :  $X_A - \frac{X_C}{K_1} = X_A - \frac{X_B}{K_2} + X_B - \frac{X_C}{K_3}$ . Cette égalité est vérifiée si les constantes d'équilibre sont égales à 1, ce qui n'est pas réaliste, ou éventuellement sous d'autres conditions que je n'ai pas pu déterminer. Toutefois, on peut observer que la loi des mailles est une application à l'électricité de la loi de conservation de l'énergie, loi qui, dans le cas du métabolisme, s'exprime par la condition de Wegscheider, qui stipule que le produit des constantes d'équilibre d'un cycle fermé (réactions prises dans le même sens) sont égales à 1 [Ederer et Gilles 2007] (cela donne ici  $K_1 = K_2 K_3$ ). Malgré cette correspondance, je ne sais pas si, et sous quelles conditions, la loi des mailles peut s'appliquer au métabolisme. Comme nous le verrons, cette difficulté n'a pas empêché d'exploiter l'analogie pour répondre aux questions que nous nous posions.

### 3.2.5 Expression du flux en fonction de toutes les enzymes du réseau : recherche de l'enzyme équivalente

L'objectif est d'exprimer le flux métabolique en fonction des propriétés de l'ensemble des enzymes du réseau, afin de pouvoir analyser l'évolution des concentrations d'enzymes lorsqu'il y a sélection directionnelle sur le flux. Pour rechercher cette expression, j'ai utilisé l'analogie entre les réseaux métaboliques et les circuits électriques, en m'appuyant sur la notion de résistance (ou de conductance) équivalente.

Un circuit constitué uniquement de dipôles de type résistance peut être réduit théoriquement à un seul dipôle, dont la résistance est une fonction de l'ensemble des résistances du circuit. L'inverse de cette résistance équivalente, notée  $R_E$ , est la conductance équivalente :

$$\sigma_E = \frac{1}{R_E}$$

En vertu de la loi d'Ohm, l'intensité globale du circuit est :

$$I = \sigma_E U \tag{3.14}$$

où  $U$  est la tension aux bornes du circuit.

On voit qu'à une constante près ( $U$ ), la conductance équivalente correspond à l'intensité

dans le circuit.

De la même manière, on peut définir dans un réseau métabolique une « enzyme équivalente » d'efficacité  $F_E$  qui transformerait  $X_S$  en  $X_P$  selon une unique pseudo-réaction. Le flux dans le réseau s'écrit alors :

$$J = F_E \left( X_S - \frac{X_P}{K_E} \right) \quad (3.15)$$

où  $K_E$  la constante d'équilibre équivalente, qui ne dépend que des constantes d'équilibres individuelles.

À une constante près  $\left( X_S - \frac{X_P}{K_E} \right)$ , l'efficacité de l'enzyme équivalente correspond au flux métabolique.

Pour les circuits électriques comme pour les réseaux métaboliques, la question est donc : comment exprimer la conductance équivalente (resp. l'efficacité de l'enzyme équivalente) en fonction de la conductance de tous les dipôles (resp. l'efficacité de toutes les enzymes) du réseau ? J'aborde cette question en considérant des circuits de plus en plus complexes.

### a) Circuits en série

Dans les circuits en série, tous les dipôles sont placés les uns à la suite des autres. S'il y a  $n$  résistances, la résistance équivalente est :

$$R_E = \sum_{i=1}^n R_i \quad (3.16)$$

donc la conductance équivalente est :

$$\sigma_E = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i}} \quad (3.17)$$

Pour  $n$  réactions enzymatiques successives, l'efficacité équivalente est (annexe C.1) :

$$F_E = \frac{1}{\sum_{i=1}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}} \quad (3.18)$$

L'homologie entre ces deux équations est évidente. La différence est simplement que l'efficacité enzymatique est pondérée par le produit des constantes d'équilibre en amont de la réaction

concernée. La constante d'équilibre équivalente est  $K_E = \prod_{i=1}^n K_i$ , et donc le flux est :

$$J = \frac{X_S - \frac{X_P}{\prod_{i=1}^n K_i}}{\sum_{i=1}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}} \quad (3.19)$$

On retrouve bien ici la relation enzymes-flux (équation 1) pour une chaîne métabolique linéaire utilisée dans les chapitres 1 et 2.

### b) Circuits en dérivation

Dans les circuits en dérivation, tous les dipôles sont en parallèle et possèdent la même borne d'entrée et la même borne de sortie. La conductance équivalente pour  $n$  dipôles en dérivation est :

$$\sigma_E = \sum_{i=1}^n \sigma_i \quad (3.20)$$

De la même manière, l'efficacité enzymatique équivalente pour  $n$  enzymes en parallèle est (annexe C.2) :

$$F_E = \sum_{i=1}^n F_i \quad (3.21)$$

Les enzymes transformant un même substrat en un même produit, les réactions sont toutes thermodynamiquement équivalentes, et donc la constante d'équilibre équivalente est :

$$\forall i \quad K_E = K_i$$

et le flux s'écrit :

$$J = \sum_{i=1}^n F_i \left( X_S - \frac{X_P}{K_E} \right) \quad (3.22)$$

### c) Circuits mixtes

Les circuits mixtes sont des circuits électriques ne comportant que des dipôles en série et/ou en dérivation. Quelle que soit leur topologie, ils peuvent être progressivement réduits en regroupant de proche en proche les diverses résistances. On aboutit ainsi à une conductance équivalente exprimable en fonction des conductances de l'ensemble du circuit à partir des équations 3.17 et 3.20. Il en est de même pour l'efficacité enzymatique équivalente qui est exprimable à partir des équations 3.18 et 3.21.

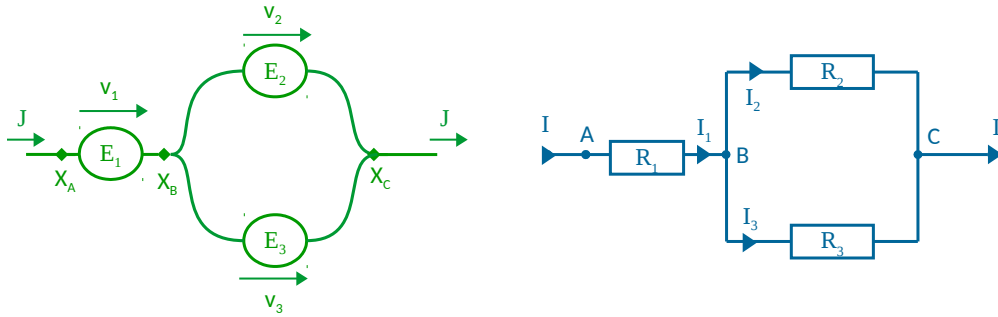


FIGURE 3.3 – **Exemple de circuit électrique mixte et son équivalent métabolique.** Réseau métabolique à gauche (en vert) et le circuit électrique mixte équivalent à droite (en bleu). Dans cet exemple, le dipôle 1 est en série avec une boucle de dérivation constituée des dipôles 2 et 3.

Considérons par exemple un circuit simple à trois dipôles possédant une boucle de dérivation, avec  $R_1$  en entrée et  $R_2$  et  $R_3$  en parallèle (figure 3.3). Dans ce cas, en appliquant la relation 3.20 à  $R_2$  et  $R_3$ , puis la relation 3.17 entre  $R_1$  et la résultante de la combinaison de  $R_2$  et  $R_3$ , on trouve facilement :

$$\sigma_E = \frac{\sigma_1(\sigma_2 + \sigma_3)}{\sigma_1 + \sigma_2 + \sigma_3} \quad (3.23)$$

Pour un réseau métabolique de même topologie, en appliquant de la même manière les relations 3.21 puis 3.18, l'efficacité enzymatique équivalente en fonction de l'efficacité des trois enzymes sera :

$$F_E = \frac{1}{\frac{1}{F_1} + \frac{1}{K_1(F_2 + F_3)}} = \frac{F_1(F_2 + F_3)}{F_1/K_1 + F_2 + F_3} \quad (3.24)$$

Là encore, la seule différence entre les deux équations porte sur la présence d'une constante d'équilibre.

Le flux est alors :

$$J = \frac{F_1(F_2 + F_3)}{F_1/K_1 + F_2 + F_3} \left( X_S - \frac{X_P}{K_E} \right) \quad (3.25)$$

où  $K_E = K_1K_2 = K_1K_3$  puisque les réactions 2 et 3 sont thermodynamiquement équivalentes.

Un autre exemple de recherche d'enzyme équivalente est présenté en annexe C.3, pour un réseau à cinq enzymes.

#### d) Circuits complexes

Dans le cas général, les circuits ne sont pas constitués que de dipôles en dérivation et en série. Pour les réduire à un dipôle équivalent, il faut utiliser des méthodes plus sophistiquées, comme la méthode des *nodal potentials* [Kagan 2015] ou encore la *Delta-Y method*, qui s'appuie sur le théorème de Kennelly [Krim et al. 2019].

Le plus simple des circuits de ce type est le pont de Wheatstone (*Wheatstone bridge*), qui présente un dipôle reliant les deux branches d'une boucle de dérivation (figure 3.4). Kagan

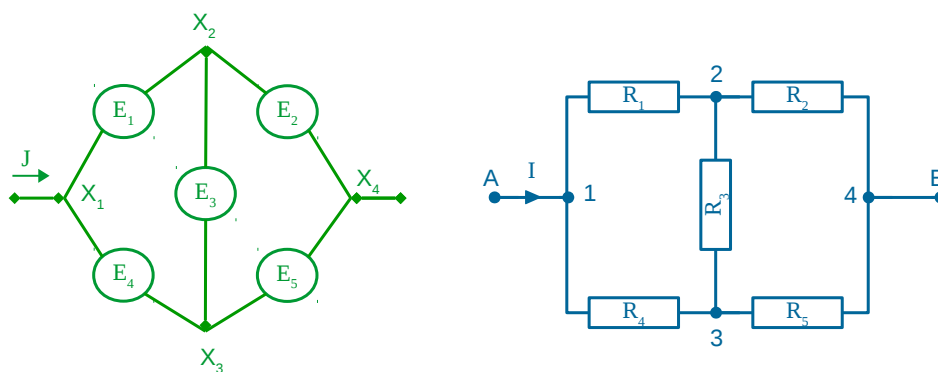


FIGURE 3.4 – **Pont de Wheatstone.** En bleu à droite le circuit électrique, et en vert à gauche son pendant métabolique. Les deux branches entre A et B sont reliées par le dipôle  $R_3$ . Les résistances ne sont alors ni en série, ni en dérivation, et le pont n'est pas réductible simplement.

[2015] démontre que la conductance équivalente s'écrit dans ce cas :

$$\sigma_E = \frac{\sigma_1\sigma_2(\sigma_3 + \sigma_4 + \sigma_5) + \sigma_2\sigma_3\sigma_4 + \sigma_1\sigma_3\sigma_5 + \sigma_4\sigma_5(\sigma_1 + \sigma_2 + \sigma_3)}{\sigma_3(\sigma_1 + \sigma_2 + \sigma_4 + \sigma_5) + (\sigma_1 + \sigma_2)(\sigma_4 + \sigma_5)} \quad (3.26)$$

En utilisant son approche, j'ai montré que l'efficacité enzymatique équivalente était (annexe C.4) :

$$F_E = \frac{F_1F_2K_3(F_3/K_3 + F_4/K_4 + F_5) + F_2F_3F_4 + F_1F_3F_5K_3 + F_4F_5K_3(F_1/K_1 + F_2 + F_3)}{F_3(F_1/K_1 + F_2 + F_4K_3/K_4 + F_5K_3) + K_3(F_1/K_1 + F_2)(F_4/K_4 + F_5)} \quad (3.27)$$

À noter qu'ici le potentiel électrique au point 4 est supposé nul. Il en est alors de même pour  $X_4$ , ce qui ne me permet pas de déterminer une constante d'équilibre équivalente.

### e) Cas général

Nous avons vu que, quelle que soit leur complexité, les circuits comme les réseaux métaboliques peuvent être progressivement réduits, ce qui permet d'obtenir une expression de la conductance équivalente (resp. de l'efficacité équivalente pour les réseaux métaboliques) en fonction de toutes les conductances (resp. de toutes les efficacités enzymatiques). À une constante près, on a donc une expression de l'intensité (resp. du flux) en fonction de tous les paramètres du système.

Mais cette approche est laborieuse et doit être adaptée à topologie. Une piste intéressante pour sa généralisation peut là encore s'appuyer sur les travaux de Kagan [2015]. Celui-ci a montré qu'un réseau de résistances de topologie quelconque possédant  $n$  nœuds ( $n > 2$ ) pouvait



être représenté à l'aide d'une matrice de conductances  $\Sigma$  :

$$\Sigma = \begin{pmatrix} c_1 & -\sigma_{1,2} & -\sigma_{1,3} & -\sigma_{1,4} & \cdots & -\sigma_{1,n} \\ -\sigma_{2,1} & c_2 & -\sigma_{2,3} & -\sigma_{2,4} & \cdots & -\sigma_{2,n} \\ -\sigma_{3,1} & -\sigma_{3,2} & c_3 & -\sigma_{3,4} & \cdots & -\sigma_{3,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{n,1} & -\sigma_{n,2} & -\sigma_{n,3} & -\sigma_{n,4} & \cdots & c_n \end{pmatrix} \quad (3.28)$$

où  $\sigma_{ij}$  est la conductance entre les nœuds  $i$  et  $j$ , et où  $c_i = \sum_{\substack{j \neq i \\ j=1}}^n \sigma_{ij}$ . Lorsque les nœuds  $i$  et  $j$  ne sont pas reliés, on a  $\sigma_{ij} = 0$ . En appliquant la loi des nœuds à chaque nœud  $i$ , et grâce à la propriété d'anti-symétrie des intensités ( $I_{ij} = -I_{ji}$ ), Kagan [2015] a montré que la conductance équivalente était :

$$\sigma_E = \frac{\det \Sigma'}{\det \Sigma''}$$

où  $\det \Sigma'$  est le déterminant de la sous-matrice supérieure gauche  $(n-1) \times (n-1)$  de  $\Sigma$  et où  $\det \Sigma''$  est la sous-matrice inférieure droite  $(n-2) \times (n-2)$  de  $\Sigma'$ . Pour appliquer cette relation aux réseaux métaboliques, il reste à introduire dans l'expression les constantes d'équilibre. Je n'ai pas eu le temps de m'y atteler, mais je crains que ce soit un travail particulièrement laborieux.

Quoi qu'il en soit, on voit que pour tout réseau métabolique, on peut exprimer l'efficacité enzymatique équivalente, donc le flux à une constante près, en fonction des efficacités de toutes les enzymes en utilisant les lois de l'électricité, même si les constantes d'équilibre compliquent les équations.

### 3.2.6 Relation entre l'intensité électrique d'un circuit et la conductance d'une résistance particulière

Comme nous l'avons vu en introduction dans le cadre de la TCM, la relation entre le flux et l'efficacité d'une enzyme donnée, et plus précisément sa concentration, est une hyperbole croissante avec asymptote horizontale (équation 2 et figure 6). Qu'en est-il dans un réseau métabolique de complexité quelconque ? Intuitivement, on peut s'attendre au même type de comportement : à mesure que l'efficacité d'une enzyme augmente, le flux va s'accroître de moins en moins vite pour finalement plafonner, car il est limité par l'efficacité de toutes les autres enzymes du système. Comme l'efficacité de l'enzyme équivalente correspond au flux métabolique à une constante près  $\left(X_S - \frac{X_P}{K_E}\right)$ , je m'appuierai sur les développements précédents pour savoir si la concavité avec la présence d'un plateau sont valides dans le cas général.

### a) Circuits en série

À partir de l'équation 3.18, la relation entre l'efficacité d'une enzyme particulière ( $F_k$ ) et l'efficacité de l'enzyme équivalente ( $F_E$ ) s'écrit :

$$F_E = \frac{1}{\frac{1}{F_k \prod_{j=1}^{k-1} K_j} + \sum_{\substack{i=1 \\ i \neq k}}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}} \quad (3.29)$$

C'est donc une hyperbole ascendante d'asymptote horizontale (figure 3.5A) :

$$\lim_{F_k \rightarrow +\infty} F_E = F_{E_{\max}} = \frac{1}{\sum_{\substack{i=1 \\ i \neq k}}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}}$$

On retrouve bien l'un des résultats majeurs de la TCM, à savoir la relation hyperbolique entre le flux et l'efficacité enzymatique. Notons que si l'une des enzymes a une efficacité nulle, le flux est également nul.

### b) Circuits en dérivation

À partir de l'équation 3.21, la relation entre l'efficacité d'une enzyme particulière ( $F_k$ ) et l'efficacité de l'enzyme équivalente ( $F_E$ ) s'écrit :

$$F_E = F_k + \sum_{\substack{i=1 \\ i \neq k}}^n F_i \quad (3.30)$$

Cette relation est donc linéaire (figure 3.5B), avec un minimum à :

$$\lim_{F_k \rightarrow 0} F_E = F_{E_{\min}} = \sum_{\substack{i=1 \\ i \neq k}}^n F_i$$

Dans ce cas, même si l'une des enzymes est inactivée, le flux n'est pas nul.

### c) Circuits mixtes

Dans un circuit mixte, la relation entre la conductance du dipôle d'intérêt et la conductance équivalente ne peut être que de deux types :

- Linéaire lorsque le dipôle d'intérêt relie directement les bornes d'entrée et de sortie du circuit (équation 3.20);
- Hyperbolique dans tous les autres cas (équation 3.17).

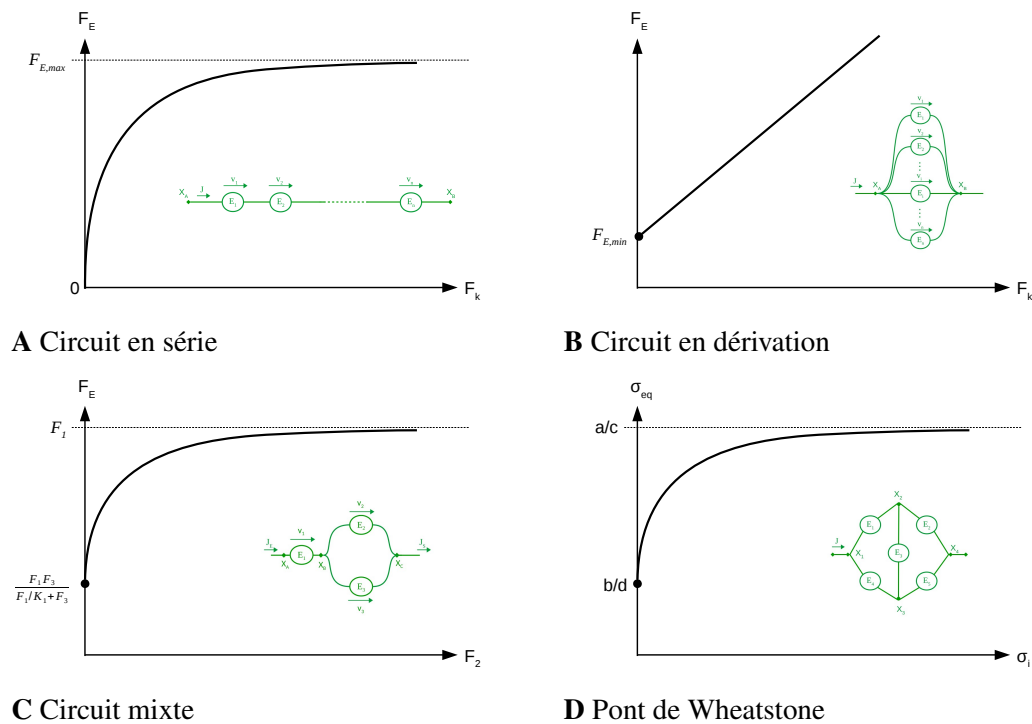


FIGURE 3.5 – **Relation entre l'efficacité d'une enzyme quelconque du réseau et l'efficacité de l'enzyme équivalente, les autres étant fixées.** (A) Dans le circuit en série, l'asymptote dépend des efficacités des autres enzymes non variables. (B) Dans le circuit en dérivation, où toutes les enzymes sont en parallèle, la relation est linéaire, et le minimum dépend des efficacités des autres enzymes. (C) Dans le circuit mixte, l'enzyme 1 est à l'entrée du réseau, et les enzymes 2 et 3 sont en parallèle. Lorsque  $F_2$  est variable, le maximum est contraint par  $F_1$  (enzyme en série) et le minimum par  $F_1$  et  $F_3$  (le flux pouvant passer par  $E_3$ ). (D) Dans le cas du pont de Wheatstone, la valeur de l'asymptote  $a/c$  et du minimum  $b/d$  dépendent aussi des efficacités des enzymes non variables.

Il en est de même pour la relation entre l'efficacité d'une enzyme et l'efficacité enzymatique équivalente (équations 3.18 et 3.21). Cela peut s'illustrer à partir du réseau métabolique de la figure 3.3 : quelle que soit l'enzyme variable on aura une relation hyperbolique avec le flux (figure 3.5C). Si l'enzyme variable est  $E_1$ , l'asymptote est  $K_1(F_2 + F_3)$ , et si c'est  $E_2$  ou  $E_3$ , l'asymptote est  $F_1$ .

#### d) Circuits complexes

Dans le cas du pont de Wheatstone, la relation entre une conductance particulière et l'intensité à travers le circuit peut être déduite directement de l'équation 3.26. En fonction d'une conductance  $\sigma_i$  quelconque, elle s'écrit :

$$\sigma_E(\sigma_i) = \frac{a\sigma_i + b}{c\sigma_i + d} \quad (3.31)$$

où  $a$ ,  $b$ ,  $c$ , et  $d$  sont des constantes qui dépendent des autres conductances du circuit, considérées comme fixes. Cette expression est le quotient de deux fonctions affines, et comme  $c$  est non nul quel que soit  $\sigma_i$ , c'est l'équation d'une hyperbole<sup>2</sup>. M. Petrizzelli a montré que cette hyperbole était monotone croissante (démonstration en annexe C.5, reproduite avec son autorisation). La courbe est comprise entre les deux limites (figure 3.5D) :

$$\lim_{\sigma_i \rightarrow 0} \sigma_E = b/d \quad \text{et} \quad \lim_{\sigma_i \rightarrow +\infty} \sigma_E = a/c$$

On peut noter que dans le cas très particulier où  $\sigma_1\sigma_5 = \sigma_2\sigma_4$ , la conductance équivalente est constante par rapport à  $\sigma_3$ , ce qui signifie que  $\sigma_3$  n'a pas d'impact sur l'intensité globale.

#### e) Cas général

Dans le cas général, l'expression qui relie la conductance équivalente  $\sigma_E$  à une conductance particulière  $\sigma_{ij}$  entre deux nœuds  $i$  et  $j$  quelconques [annexe A dans Kagan 2015] :

$$\sigma_E = \frac{A\sigma_{ij} + B}{C\sigma_{ij} + D} \quad (3.32)$$

où  $A$ ,  $B$ ,  $C$  et  $D$  sont des constantes qui dépendent des conductances de l'ensemble du circuit autres que  $\sigma_{ij}$ . Cette expression a la même forme que celle du pont de Wheatstone (équation 3.31). On a donc également une relation hyperbolique, les limites de l'hyperbole étant :

$$\lim_{\sigma_i \rightarrow 0} \sigma_E = B/D \quad \text{et} \quad \lim_{\sigma_i \rightarrow +\infty} \sigma_E = A/C$$

<sup>2</sup>Une fonction homographe  $f(x) = (ax + b)/(cx + d)$  est le quotient de deux fonctions affines. Elle est hyperbolique croissante si  $ad - cb > 0$  et  $c \neq 0$ . Si  $c = 0$ , alors la fonction est affine, et si  $ad - bc = 0$ , alors elle est constante.

Cette hyperbole est également croissante, comme M. Petrizzelli l'a montré récemment (démonstration en annexe C.5). Notons que dans les cas très particuliers où  $C$  serait nul, la relation serait linéaire, et avec  $AD - BC = 0$ , la relation serait constante. Ces cas sont très peu probables, et d'ailleurs n'ont jamais été rencontrés dans les simulations de circuits de topologies très diverses que nous avons faites avec LTSpice<sup>®3</sup>, qui ont toujours montré des hyperboles croissantes avec asymptote horizontale.

La transposition des circuits électriques aux réseaux métaboliques laisse donc penser que, quelle que soit la topologie des réseaux, et même si dans le cas général les constantes d'équilibre n'ont pas été prises en compte, la relation entre le flux et la concentration/activité d'une enzyme donnée est toujours une fonction concave croissante bornée (à quelques exceptions très particulières près).

### 3.2.7 Discussion

Dans le métabolisme comme dans l'électricité, il y a un flux d'énergie (matière ou électrons) plus ou moins rapide selon les composants (résistances ou enzymes). Dans le cas des chaînes métaboliques linéaires, on sait comment relier le flux aux propriétés de toutes les enzymes de la voie. Lorsque l'on a affaire à des réseaux, l'expression explicite de cette relation se heurte à la complexité des modèles. J'ai donc utilisé l'analogie avec les circuits électriques pour répondre à la question, en assimilant la conductance des résistances à l'efficacité enzymatique et le flux à l'intensité du courant, et en m'appuyant sur la notion de conductance équivalente. J'ai ainsi pu mettre en évidence l'homologie entre les équations.

#### a) Hypothèses et difficultés de l'analogie

Nous l'avons vu tout au long de ce travail, l'analogie métabolisme-électricité repose sur un certain nombre d'hypothèses, que je vais ici expliciter.

Tout d'abord, les réactions sont considérées comme **uni-moléculaires**, c'est-à-dire avec un seul substrat et un seul produit par réaction, pour simplifier l'écriture de l'équation de vitesse et faciliter les développements mathématiques. C'est une hypothèse forte, car la plupart des réactions cellulaires impliquent plusieurs réactifs et produits. Biologiquement, les réactions uni-moléculaires peuvent correspondre à divers cas de figure : modification d'une protéine par ajout, transformation ou suppression d'un groupe chimique, ajout séquentiel de monomères à des macromolécules, ou encore un transfert de molécules par des transporteurs. On peut également considérer des enzymes michaeliennes en ignorant les co-facteurs de type NADP souvent présents en grande quantité.

Ensuite, la cinétique enzymatique correspond à celle d'une **enzyme michaelienne loin de la saturation** catalysant une réaction réversible. Il existe divers types de cinétique enzyma-

---

<sup>3</sup>Logiciel libre téléchargeable à l'adresse : <https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html>

tique [Saa et Nielsen 2017], et en général les équations de vitesse associées se décomposent en quatre parties : la concentration de l'enzyme, une partie purement cinétique avec les propriétés de l'enzyme, une partie thermodynamique (« moteur » de la réaction) et une partie régulation de l'activité enzymatique [Liebermeister *et al.* 2010]. Ces cinétiques me semblent valides pour l'analogie tant que l'on peut isoler la partie thermodynamique de la réaction (c'est-à-dire  $X_S - X_P/K_{eq}$ ) à condition bien sûr que ces cinétiques correspondent à des réactions uni-moléculaires. Cependant, ces autres cinétiques prennent en général en compte des métabolites intermédiaires, par exemple à travers les régulations allostériques, ce qui complexifie le modèle.

Le système est supposé à l'**état stationnaire**, c'est-à-dire que les vitesses des réactions et les concentrations en métabolites sont constantes. Cet état est dynamique : il y a autant de métabolites consommés que de métabolites produits, sinon il y aurait une surcharge ou un épuisement du métabolite si les deux vitesses (entrante et sortante) étaient différentes. L'hypothèse d'état stationnaire, qui facilite énormément les analyses et qui est couramment utilisée en chimie et en biologie cellulaire, semble valable sur des durées assez longues à l'échelle de la vitesse des réactions métaboliques.

Enfin on considère un **milieu ouvert**, c'est-à-dire que le substrat du réseau métabolique est constamment ajouté et son produit constamment retirée, permettant de maintenir un état stationnaire. En effet, si on considère un milieu fermé, c'est-à-dire sans ajout ou retrait de composé chimique, les réactions chimiques iraient naturellement à l'équilibre, et la vitesse de réaction serait alors nulle quel que soit le point du réseau considéré. En électricité, ce milieu ouvert, avec un substrat et un produit constants, correspond à un générateur de tension.

Toutes ces hypothèses sont les mêmes que celles utilisées dans le développement de la relation enzymes-flux des chapitres précédents, qui a donné un modèle de la relation génotype-phénotype particulièrement fécond.

## b) Interprétation énergétique de la constante d'équilibre

Tout comme la différence d'énergie entre un point A et un point B dans un circuit électrique met en mouvement les électrons, modélisée par la différence de potentiels électriques, c'est la différence d'énergie entre les différentes espèces chimiques qui constituent le moteur de la réaction, modélisée par l'écart à l'équilibre. Quel que soit le système considéré, l'énergie tend à s'homogénéiser.

La constante d'équilibre représente la répartition des composés chimiques à l'état d'énergie minimal, qui est l'état d'équilibre. Elle exprime le fait que les composés chimiques possèdent des potentiels chimiques différents. Cette notion n'existe donc pas en électricité, car tous les électrons possèdent la même énergie.

L'analogie stricte entre métabolisme et électricité correspondrait donc à des constantes d'équilibre toutes égales à 1. Une attention particulière doit donc être portée aux constantes d'équilibre, car elles vont modifier le calcul des efficacités équivalentes.

### c) Comparaison avec la TKM

Le formalisme présenté dans ce chapitre est proche de celui de la *Thermodynamic-Kinetic Modeling* (TKM) développé par Ederer et Gilles [2007], mais il diffère sur quelques aspects.

Tout d'abord, nos objectifs sont différents. La TKM a pour objectif de mettre en valeur la thermodynamique dans les grands réseaux métaboliques, souvent oubliée lors de l'étude de leur cinétique. Mon but est plus limité : je souhaite simplement relier explicitement les propriétés enzymatiques au flux métabolique d'un réseau quelconque, afin de pouvoir étudier analytiquement leur évolution.

Notre cadre de travail est similaire d'un point de vue thermodynamique. Nos modèles se basent tous deux sur les concentrations en métabolites, et dans les deux cas c'est l'écart à l'équilibre des concentrations qui constitue le moteur de la réaction. J'utilise cependant des hypothèses plus restrictives pour la partie cinétique, notamment l'unimolécularité des réactions et la cinétique des enzymes michaeliennes.

Cependant, nos formalismes sont différents. Dans la TKM, les charges électriques  $q$  sont analogues aux concentrations en métabolites et les potentiels électriques aux potentiels chimiques. En effet, l'intensité est un débit de charges électriques ( $I = dq/dt$ ), et la vitesse de réaction est une vitesse de consommation de substrat  $S$  et de formation de produit  $P$  ( $v = dP/dt = -dS/dt$ ). Cependant, les charges électriques transportent l'énergie, et les métabolites constituent des pools d'énergie, qui est transmise *via* les réactions chimiques. De plus, les porteurs d'énergie (charges et métabolites) sont reliés différemment à leur potentiel d'énergie selon le système considéré. En électricité, le potentiel dépend des champs électriques et du placement dans le réseau, et en chimie, il est inhérent au métabolite considéré. Ederer et Gilles [2008] suggèrent que les métabolites sont des condensateurs dans la TKM : ils considèrent les métabolites comme des « stocks » d'énergie chimique, tout comme un condensateur stocke l'énergie électrique. En effet, dans la TKM, les capacités (grandeur propre aux condensateurs) sont directement reliées au potentiel chimique standard qui représente en quelque sorte l'énergie intrinsèque des composés chimiques servant habituellement à définir la constante d'équilibre. Comme les concentrations et potentiels chimiques sont liés et propres au métabolite considéré, tandis que les potentiels et les charges électriques sont indépendants, il me semblait plus pertinent de faire l'analogie entre concentrations et potentiels électriques pour simplifier les équations, tout en ignorant les charges électriques. L'intensité et la vitesse de réaction sont ainsi assimilées à des flux d'énergie, et les métabolites constituent alors les nœuds du réseau.

Dans les deux formalismes, le lien entre thermodynamique et cinétique est constitué par la « résistance ». Cependant, nos définitions mathématiques pour la résistance métabolique sont différentes, de même que sa dimension. Dans l'approche que j'utilise, on compare directement la cinétique enzymatique avec la loi d'Ohm et on fait l'analogie entre conductance et efficacité enzymatique (en  $s^{-1}$ ). De son côté, la TKM part de la définition du potentiel chimique pour écrire la partie thermodynamique et développer la notion de « force thermocinétique », analogue

à la tension électrique. En divisant cette force par la vitesse de réaction, écrite indépendamment, on obtient la résistance thermocinétique de la réaction (en  $\text{mol}^{-1} \cdot \text{m}^3 \cdot \text{s}$ ). Grâce au concept de « force thermocinétique » qui ne prend pas en compte directement les constantes d'équilibre, le formalisme de la TKM pourrait être utile, notamment pour traiter le cas général de la relation entre un flux et l'ensemble des paramètres d'un système. Cependant la recherche de la résistance thermocinétique dans le cas des enzymes m'a semblé trop complexe pour pouvoir être exploitée.

#### d) Analogies impliquant des flux d'énergie

Les analogies impliquant des flux d'énergie sont assez communes dans la bibliographie. Notamment, les analogies d'impédance [Olson 1958], parfois appelées « formalisme de la thermodynamique en réseaux » [Rumelhard 1988], se basent sur la distinction courant/force, et possèdent toujours trois parties : une partie « moteur », une partie « cinétique » et une partie « résistance » ou « vecteur d'énergie ». Le tableau 3.2 présente les équivalences entre différents systèmes (électrique, métabolique, hydraulique).

TABLE 3.2 – **Flux d'énergie : analogies entre différents systèmes.** Le tableau présente l'analogie entre métabolisme, électricité et hydraulique. On pourrait y ajouter les systèmes mécanique, acoustique et thermique, voire économique. Dans l'annexe C.6 j'utilise une métaphore avec une chaîne de production, que j'ai proposée à des fins de vulgarisation scientifique.

Système	Électrique	Métabolique	Hydraulique
Quantité	Charges électriques $q$	Matière	Volume de fluide
Cinétique	Intensité $I$	Flux métabolique $J$	Débit $Q$
Énergie	Potentiel électrique $V$	Potentiel chimique $\mu$	Charge
« Moteur »	Tension $U$	Écart à l'équilibre	Différence de charge/pression statique
Vecteur d'énergie	Résistance $R$	Réaction chimique ( $\pm$ catalyseur) liée aux constantes cinétiques $k$	Caractéristiques de perte de charge (section, longueur...)
Ajout d'énergie	Générateur électrique	Ajout substrat/retrait produit	Pompe

### 3.3 Perspectives

Grâce à la relation entre le flux et l'ensemble des efficacités enzymatiques d'un réseau, on pourrait déterminer l'effet sur le flux d'une mutation affectant une enzyme quelconque du réseau. Cela permettrait d'étudier l'évolution des enzymes sous sélection pour augmenter le flux dans le cas général. Comme l'équation de l'efficacité enzymatique équivalente dépend de l'agencement des réactions, on peut supposer que l'équilibre évolutif des concentrations d'enzymes sera directement influencé par la topologie du réseau. L'étude de l'évolution de réseaux simples (voir ci-dessous les résultats préliminaires pour un réseau ne comprenant que des enzymes en



parallèle) permettrait d'extrapoler à des réseaux plus complexes.

Par ailleurs, nous avons montré que la relation entre n'importe quelle conductance et l'intensité du courant à travers un circuit se présentait sous forme d'une courbe de saturation, quelle que soit la complexité du réseau. La cellule est un réseau de réseaux (de régulation de gènes, de signalisation, d'interactions protéiques, etc.), au sein duquel des relations concaves avec plateau semblent communes, à tous les niveaux d'organisation, de la transcription à la fitness [Fiévet *et al.* 2018; Kemble *et al.* 2019]. Au-delà des nombreux processus qui engendrent des non-linéarités locales, cette concavité pourrait être une propriété globale des réseaux, dès lors qu'il y a transfert d'énergie. L'analogie avec les circuits électriques pourrait donc, au moins qualitativement, être valable à d'autres niveaux d'organisation biologique que celui du métabolisme, ce qui permettrait d'utiliser la relation entre conductances et intensité dans un circuit constitué uniquement de résistances comme modèle de la relation génotype-phénotype.

Enfin, l'analogie métabolisme-électricité que je développe me semble avoir une portée pédagogique, par exemple pour faire comprendre le rôle des constantes d'équilibre, et notamment leur rôle dans la pondération de l'énergie par rapport à l'électricité.

## Résultats préliminaires sur l'évolution d'enzymes en parallèle

Nous avons étudié l'évolution des concentrations d'enzymes dans le cas d'une chaîne métabolique linéaire, c'est-à-dire lorsque les réactions sont successives. Un deuxième type de réseau simple est le cas où toutes les enzymes sont placées en parallèle (par exemple des isoenzymes catalysant la même réaction). En utilisant les résultats de l'analogie métabolisme-électricité (équations 3.15 et 3.21), le flux s'écrit, en dissociant ici concentrations et paramètres cinétiques des enzymes :

$$J = \sum_{i=1}^n \frac{k_{\text{cat},i}}{K_{M_i}} E_i (X_0 - X_n / K_{\text{eq}}) \quad (3.33)$$

où  $X_0$  et  $X_n$  sont respectivement les concentrations du substrat et du produit de la réaction,  $K_{\text{eq}}$  est la constante d'équilibre de la réaction,  $k_{\text{cat},i}$  est la constante de catalyse,  $K_{M_i}$  la constante de Michaelis-Menten et  $E_i$  la concentration de l'enzyme  $i$ . L'activité  $A_i$  de l'enzyme  $i$  est ici définie par  $k_{\text{cat},i}/K_{M_i}$ .

En cas d'indépendance entre toutes les enzymes, les concentrations relatives à l'équilibre sont proportionnelles aux activités (annexe D) :

$$e_j^* = \frac{A_j}{\sum_{k=1}^n A_k} \quad (3.34)$$

C'est intuitivement logique : plus une enzyme est active, plus il est avantageux qu'elle soit abondante. Les simulations sur un réseau à trois enzymes montrent que les concentrations relatives convergent bien vers l'équilibre théorique (figure 3.6A). Comme il n'y a pas de contrainte, les concentrations absolues d'enzymes et le flux augmentent indéfiniment (figure 3.6B).

En cas de compétition entre toutes les enzymes, la concentration totale d'enzymes est fixée et le flux est limité (figure 3.6D). L'équilibre des concentrations relatives ne peut pas être trouvé analytiquement (annexe D). Dans les simulations, on voit que les enzymes à faible activité vont progressivement disparaître au profit de l'enzyme dont l'activité est la plus élevée (figure 3.6C). Ce comportement est logique : par analogie avec l'électricité, la matière passe par le chemin avec le moins de résistance, et donc par l'enzyme à l'activité la plus élevée. Si les activités peuvent aussi muter, l'enzyme la plus active peut changer, si bien que la sélection ne privilégiera pas toujours la même enzyme. Ainsi plusieurs enzymes pourront continuer à co-exister.

Différentes enzymes catalysant une même réaction en parallèle correspondent en général à des isoenzymes, produits plus ou moins différents de gènes dupliqués. À terme, les isoenzymes peuvent conduire à des phénomènes de sous-fonctionnalisation.

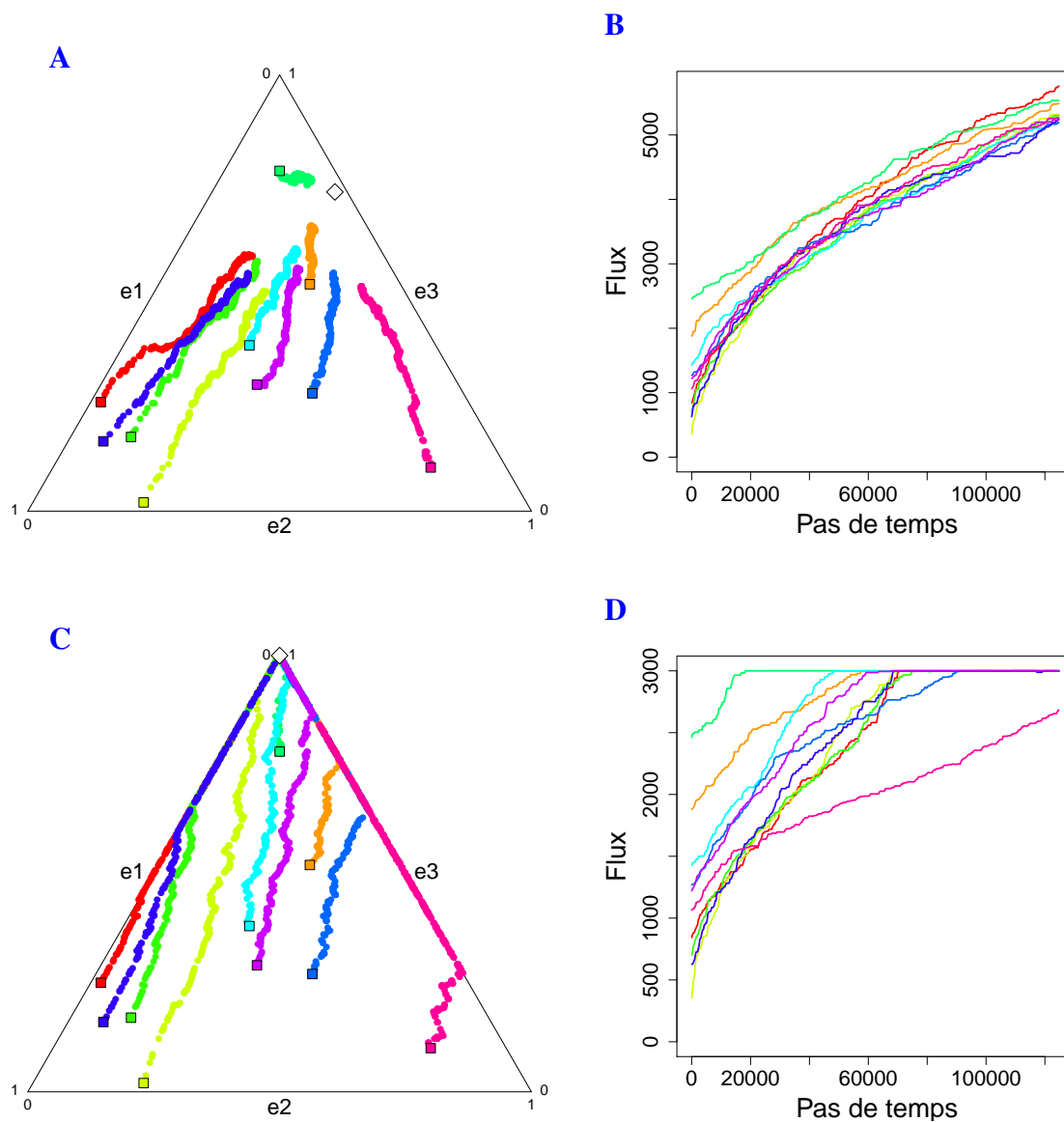


FIGURE 3.6 – **Simulation de l'évolution des concentrations de trois enzymes catalysant la même réaction en parallèle.** (A et B) Indépendance entre toutes les enzymes. (C et D) Compétition entre toutes les enzymes. (A) Trajectoires évolutives en cas d'indépendance des concentrations relatives  $e_1$ ,  $e_2$  et  $e_3$  représentées à l'aide d'un diagramme triangulaire pour dix simulations (une couleur par simulation) avec différentes concentrations initiales (carrés colorés). Les concentrations relatives d'enzyme convergent vers l'équilibre théorique (losange blanc). (C) Même représentation qu'en A pour le cas compétition entre toutes les enzymes. Les enzymes 1 et 2 disparaissent progressivement au profit de l'enzyme 3 qui a l'activité  $k_{\text{cat}}/K_M$  la plus élevée (losange blanc). (B et D) Évolution du flux au cours du temps pour les simulations présentées en figures A et C respectivement. Valeurs des paramètres :  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{\text{tot}}^0 = 100$ ,  $N = 1\,000$ .

## Glossaire des symboles mathématiques

### Symboles mathématiques utilisés dans l'analogie

$\sigma$  Conductance électrique d'un dipôle. La conductance est définie comme l'inverse de la résistance :  $\sigma = 1/R$ .

$\sigma_E$  Conductance électrique équivalente d'un circuit électrique.

$E_i$  Concentration de l'enzyme  $i$

$F_E$  Efficacité enzymatique d'une enzyme qui catalyserait une réaction chimique équivalente à un réseau métabolique quelconque, dite efficacité enzymatique équivalente.

$F_i$  Efficacité enzymatique de l'enzyme  $i$  ;  $F_i = E_i k_{\text{cat},i} / K_{M_i}$

$I_i$  Intensité du courant traversant le dipôle  $i$ .

$J$  Flux métabolique.

$K_{M_i}$  Constante de Michaelis-Menten de l'enzyme  $i$ , représentant son affinité pour son substrat.

$K_E$  Constante d'équilibre d'une réaction chimique équivalente à un réseau métabolique quelconque, dite constante d'équilibre équivalente.

$K_i$  Constante d'équilibre de la réaction chimique  $i$ . Lorsqu'on considère une réaction quelconque, peut s'écrire  $K_{\text{eq}}$ .

$R_E$  Résistance électrique équivalente d'un circuit électrique.

$R_i$  Résistance électrique du dipôle  $i$ .

$U_{AB}$  Tension entre les points A et B d'un circuit.

$V_A$  Potentiel électrique en un point A.

$X_A$  Concentration du métabolite A, réactif ou produit d'une réaction.

$X_P$  Concentration en produit d'une réaction chimique.

$X_S$  Concentration en substrat/réactif d'une réaction chimique.

$k_{\text{cat},i}$  Constante de catalyse de l'enzyme  $i$ .

$v_i$  Vitesse de la réaction chimique  $i$ .

## Thermodynamique et cinétique chimiques

Liste des symboles mathématiques utilisés dans la section 3.2.2.

$\Delta_r G$  [Thermodynamique] Enthalpie libre de réaction, ou énergie libre de Gibbs

$K_M^-$  [Cinétique] Constante de Michaelis *reverse*, représentant l'affinité de l'enzyme pour son produit.

$K_M$  [Cinétique] Constante de Michaelis *forward*, représentant l'affinité de l'enzyme pour son substrat.

$K_{eq}$  [Thermodynamique] Constante d'équilibre d'une réaction chimique.

$Q_r$  [Thermodynamique] Quotient de réaction

$R$  [Thermodynamique] Constante des gaz parfaits

$T$  [Thermodynamique] Température

$[E]$  [Cinétique] Concentration d'une enzyme.

$[P]$  [Cinétique] Concentration en produit d'une réaction chimique.

$[S]$  [Cinétique] Concentration en substrat/réactif d'une réaction chimique.

$a_i$  [Thermodynamique] Activité chimique du constituant chimique  $i$ .

$k^+$  [Cinétique] Coefficient de vitesse de la réaction *forward*.

$k^-$  [Cinétique] Coefficient de vitesse de la réaction *reverse*.

$k_{cat}$  [Cinétique] Constante de catalyse

$v$  [Cinétique] Vitesse d'une réaction chimique ;  $v = v^+ - v^-$ .

$v^+$  [Cinétique] Vitesse de réaction dans le sens *forward*, des réactifs vers les produits.

$v^-$  [Cinétique] Vitesse de réaction dans le sens *reverse*, des produits vers les réactifs.

$\nu_i$  Nombre stoechiométrique du constituant chimique  $i$ . Vaut 1 si c'est un produit et -1 si c'est un réactif.

$\mu_i$  [Thermodynamique] Potentiel chimique du constituant chimique  $i$ .

$\mu_i^\circ$  [Thermodynamique] Potentiel chimique du constituant chimique  $i$  à l'état standard.

---

## **CONCLUSIONS ET PERSPECTIVES**

---



# Conclusions et perspectives

La question de la relation génotype-phénotype est centrale en sciences de l'évolution, car si les mutations apparaissent au niveau du génotype, c'est sur les phénotypes que porte la sélection. Mais la complexité de cette relation est telle qu'elle impose le recours à des modèles. Dans cette thèse j'ai choisi comme modèle de la relation GP la relation enzymes-flux, développée notamment dans le cadre de la théorie du contrôle métabolique [Kacser *et al.* 1995], qui avait déjà permis de rendre compte de divers effets génétiques, comme la dominance, l'épistasie et l'hétérosis, ou d'éclairer des processus évolutifs comme la neutralité sélective. Dans la continuité de ces travaux, j'ai combiné la théorie du contrôle métabolique avec un modèle de dynamique adaptative pour étudier analytiquement l'évolution des concentrations d'enzymes (le « génotype ») lorsqu'il y a sélection sur le flux (le phénotype). J'ai considéré que les concentrations d'enzymes pouvaient soit varier indépendamment, soit être dépendantes en raison de deux contraintes, la compétition pour les ressources (la concentration totale d'enzymes qu'une cellule peut allouer à une voie métabolique donnée est nécessairement limitée) et/ou des co-régulations.

Le modèle d'évolution utilisé est un compromis entre réalité biologique et simplicité mathématique. Malgré les hypothèses restrictives faites (discutées dans les chapitres 1 à 3), je pense que plusieurs résultats ont une portée assez générale. J'ai d'abord montré que, quelles que soient les contraintes, les concentrations d'enzymes atteignaient un équilibre, y compris lorsque les enzymes sont réparties en groupes de co-régulation, et que les tailles des zones de neutralité sélective autour de ces équilibres différaient selon l'enzyme considérée.

J'ai aussi montré que les contraintes imposées limitaient les possibilités d'évolution, par diminution du nombre de facteurs pouvant varier indépendamment. Les *trade-off* entre concentrations limitent notamment l'augmentation des concentrations des enzymes concernées, ce qui limite le flux, et à mesure que le flux se rapproche de sa valeur maximum, la pression de sélection diminue et les concentrations des autres enzymes tendent vers un plateau. Les co-régulations négatives ont donc un coût, en ce sens qu'elles limitent le flux, mais ce coût peut être compensé par des avantages physiologiques (allocation des ressources à d'autres voies, limitation de la concentration de métabolites toxiques, etc.).

Enfin, lorsqu'il y a des co-régulations, les états d'équilibres sont sensibles aux conditions initiales, ce qui rend imprévisible le résultat de la sélection si ces conditions initiales ne sont pas connues. La trajectoire évolutive d'une population est donc en partie déterminée par son



histoire, si bien que dans des conditions naturelles, il pourrait être difficile de faire la part entre sélection et dérive pour rendre compte d'un polymorphisme naturel.

Dans le système différentiel d'évolution des concentrations d'enzymes, nous avons utilisé des espérances et nous avons négligé les mutations délétères. Les équilibres évolutifs recherchés sont donc totalement déterministes. Mais dans les simulations faites pour vérifier les résultats, où nous n'avons pas négligé les mutations délétères, on observe toujours des fluctuations autour des équilibres, d'autant plus grandes que la taille de la population est faible. Ces fluctuations révèlent le processus de mutation-sélection-dérive, où les mutations défavorables se fixent par dérive et où les mutations avantageuses sont sélectionnées, ramenant la population vers l'équilibre. Les autres sources d'aléa dans les simulations se situent dans la taille des mutations et dans l'enzyme ciblée par la mutation, ce qui affecte également la probabilité de fixation et donc la dérive. Il faut noter que si la dérive éloigne de l'optimum, elle peut aussi permettre aux populations de franchir les vallées de fitness (la *shifting-balance theory*, [Wright \[1932\]](#); [Whitlock et Phillips \[2000\]](#)). Enfin nous n'avons pas tenu compte d'une autre source d'aléa, l'environnement, qui fait qu'à un même génotype peuvent correspondre des phénotypes différents. Ainsi, les êtres vivants sont fonctionnels, mais pas forcément optimaux, car soumis au hasard des mutations, de la dérive, de l'environnement et de l'histoire évolutive.

Afin d'élargir les résultats aux réseaux métaboliques tout en gardant un formalisme proche de celui des chaînes métaboliques linéaires, j'ai choisi de creuser la piste de l'analogie entre les réseaux métaboliques et les circuits électriques. Les deux types de réseaux reposent en effet sur le même principe de flux engendré par une différence d'énergie. Sous réserve des limites de cette analogie, que j'ai analysées en comparant les deux formalismes, j'ai montré comment relier analytiquement le flux aux propriétés des enzymes du réseau, quelle que soit sa topologie, et montré que la relation entre une enzyme et le flux était, sauf cas très particuliers, une courbe de saturation. Faute de temps je n'ai malheureusement pas pu exploiter ce modèle « électrique » pour étudier l'évolution des enzymes d'un réseau.

## Perspectives

Le travail présenté dans cette thèse peut être prolongé dans différentes directions.

### a) Étudier l'effet de mutations affectant les paramètres cinétiques et les coefficients de co-régulation

Ce que nous avons appelé *pseudo-activité*  $A_j$  d'une enzyme  $j$  est un paramètre composite qui regroupe les paramètres cinétiques de l'enzyme (l'activité enzymatique  $k_{\text{cat},j}/K_{Mj}$ , que les mutations peuvent affecter) et le produit des constantes d'équilibre des réactions en amont de l'enzyme considérée ( $K_{j-1}$ , paramètre thermodynamique constant). Les pseudo-activités déterminent la forme du paysage adaptatif et les équilibres évolutifs. En cas de compétition pour les ressources, elles déterminent le flux maximum et la courbure du dôme, et donc les tailles des zones de neutralité sélective. Toutefois, les équilibres évolutifs ne dépendent pas toujours des pseudo-activités : dans le cas où toutes les enzymes sont co-régulées positivement, seuls les coefficients de co-régulation déterminent l'équilibre théorique.

Une forte pseudo-activité peut signifier une enzyme avec un fort *turnover* et/ou une grande affinité pour le substrat et/ou une contribution forte des constantes d'équilibre dans le sens de la production du substrat de l'enzyme. Dans l'expression du flux, les pseudo-activités jouent un rôle symétrique de celui des concentrations d'enzymes. Dans ce travail nous n'avons considéré que l'évolution des concentrations, car elles sont susceptibles d'être soumises à beaucoup plus de mutations que les paramètres cinétiques (voir discussion du premier article, chapitre 1). Un modèle d'évolution symétrique, dans lequel seules les paramètres cinétiques seraient susceptibles d'être modifiés par les mutations, les concentrations restant constantes, pourrait être étudié, dans l'esprit des travaux de Heckmann *et al.* [2018]. Un tel modèle n'aurait de sens qu'en cas d'indépendance entre enzymes, puisqu'on imagine mal des compétitions ou co-régulations entre paramètres cinétiques d'enzymes différentes. Par analogie avec l'expression de l'équilibre théorique des concentrations relatives, l'efficacité enzymatique relative à l'équilibre serait inversement proportionnelle au produit de deux termes : la concentration de l'enzyme et le produit des constantes d'équilibre en amont de l'enzyme considéré, soit mathématiquement  $E_j K_{j-1}$ .

Enfin, un modèle plus complet pourrait aussi impliquer des mutations sur les coefficients de co-régulation, avec les conséquences évolutives que j'ai évoquées dans la discussion complémentaire du chapitre 1.

### b) Faire évoluer les enzymes d'un réseau métabolique de topologie quelconque

Grâce à l'analogie métabolisme-électricité, il est possible de relier le flux à toutes les propriétés enzymatiques d'un réseau de complexité quelconque. Une recherche analytique des équilibres évolutifs est hors de portée, mais des simulations sur divers types de réseaux élémentaires permettraient de d'étudier l'influence de la topologie sur les équilibres, comme j'ai commencé à le

faire pour un réseau ne contenant que des enzymes en parallèle (page 139).

**c) Vérifier la validité des prédictions par une méta-analyse de la bibliographie**

Une méta-analyse de la bibliographie permettrait de vérifier la validité des prédictions. L'utilisation de bases de données comme BRENDA (enzymologie, [Schomburg et al. \[2002\]](#); [Chang et al. \[2009\]](#)), KEGG (voies métaboliques, [Kanehisa et al. \[2002\]](#)), *eQuilibrator* (constantes d'équilibre, [Flamholz et al. \[2012\]](#)), couplées à des données de protéomique ou éventuellement d'expression des gènes, permettrait de récupérer des valeurs de paramètres enzymatiques, de constantes d'équilibre de réaction et de concentrations des enzymes selon les différentes voies métaboliques. Ces données permettraient de valider les équilibres évolutifs prédits et de déterminer éventuellement quel type de contraintes s'applique à telle voie. Toutefois, la recherche des coefficients de co-régulation, et surtout des concentrations initiales, seront probablement plus complexes.

**d) Considérer d'autres cibles de sélection**

Les fortes valeurs de flux ne sont pas toujours avantageuses. Cela peut conduire à des métabolites en concentration toxique [[Wilbanks et Trinh 2017](#); [Kemble et al. 2020](#)] ou augmenter la demande énergétique pour la production de protéines [[Koehn 1991](#)]. Des caractères comme la concentrations en métabolites [[Eanes 2017](#)] ou la stabilité des protéines [[Serohijos et al. 2012](#)] sont également sous sélection, et on pourrait essayer de les introduire dans les modèles de sélection. Ajouter une sélection sur ces phénotypes modifierait la relation génotype-fitness, générant en conséquence un paysage adaptatif plus complexe. L'impact de la sélection sur les différents phénotypes pourraient alors donner des résultats potentiellement inattendus.

---

# **BIBLIOGRAPHIE**

---



# Bibliographie

- ACADÉMIE FRANÇAISE. 2021, Dictionnaire de l'Académie Française, 9e édition (de A à Savoir). [en ligne] Consulté le 30/06/2021, URL <https://academie.atilf.fr/9/>.
- ADADI R., VOLKMER B., MILO R., HEINEMANN M., SHLOMI T. 2012, Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters. *PLOS Computational Biology*, vol. 8, n° 7, p. e1002575, URL <http://dx.doi.org/10.1371/journal.pcbi.1002575>.
- BHALLA U.S. 2003, Understanding complex signaling networks through models and metaphors. *Progress in Biophysics and Molecular Biology*, vol. 81, n° 1, p. 45–65, URL [http://dx.doi.org/10.1016/s0079-6107\(02\)00046-9](http://dx.doi.org/10.1016/s0079-6107(02)00046-9).
- BOLKER J. 2009, Exemplary and Surrogate Models : Two Modes of Representation in Biology. *Perspectives in biology and medicine*, vol. 52, p. 485–99, URL <http://dx.doi.org/10.1353/pbm.0.0125>.
- BOST B., DILLMANN C., DE VIENNE D. 1999, Fluxes and metabolic pools as model traits for quantitative genetics. I. The L-shaped distribution of gene effects. *Genetics*, vol. 153, n° 4, p. 2001–2012, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460848/>.
- BROCHIER G., GUILLEMAUD M. 2020, Les composants électriques et leurs analogues en hydraulique - Résistance et condensateur. CultureSciences-Physique - Ressources scientifiques pour l'enseignement des sciences physiques. [en ligne] Consulté le 02/10/2020, URL <http://culturesciencesphysique.ens-lyon.fr/ressource/Analogie-composant-elec-hydro-RC.xml>.
- BRÄNNSTRÖM A.k., JOHANSSON J., VON FESTENBERG N. 2013, The Hitchhiker's Guide to Adaptive Dynamics. *Games*, vol. 4, n° 3, p. 304–328, URL <http://dx.doi.org/10.3390/g4030304>.
- BUNGE M. 1977, Levels and reduction. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 233, n° 3, p. R75–R82, URL <http://dx.doi.org/10.1152/ajpregu.1977.233.3.R75>.
- CAMPBELL D.T. 1974, 'Downward Causation' in Hierarchically Organised Biological Systems. Dans : F.J. Ayala, T. Dobzhansky (éditeurs), *Studies in the Philosophy of Biology : Reduction and Related Problems*, p. 179–186, London : Macmillan Education UK, URL [http://dx.doi.org/10.1007/978-1-349-01892-5\\_11](http://dx.doi.org/10.1007/978-1-349-01892-5_11).
- CASCANTE M., LLORÉNS M., MELÉNDEZ-HEVIA E., PUIGJANER J., MONTERO F., MARTÍ E. 1996, The Metabolic Productivity of the Cell Factory. *Journal of Theoretical Biology*, vol. 182, n° 3, p. 317–325, URL <http://dx.doi.org/10.1006/jtbi.1996.0170>.

- CHANG A., SCHEER M., GROTE A., SCHOMBURG I., SCHOMBURG D. 2009, BRENDA, AMENDA and FRENDA the enzyme information system : new content and tools in 2009. *Nucleic Acids Research*, vol. 37, n° suppl\_1, p. D588–D592, URL <http://dx.doi.org/10.1093/nar/gkn820>.
- CHAUMAT N., ETASSE C., HAUCHEMAILLE J., JECH B., LOBRY M., MAHEU B., THOREZ D. 1989, *L'analogie en physique*. Rapport technique 1, IREM de Rouen - Université de Haute Normandie, Mont-Saint-Aignan, URL <https://numerisation.irem.univ-mrs.fr/RO/IRO89003/IRO89003.pdf>.
- CILIBERTI S., MARTIN O.C., WAGNER A. 2007, Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. *PLOS Computational Biology*, vol. 3, n° 2, p. e15, URL <http://dx.doi.org/10.1371/journal.pcbi.0030015>.
- CLARK A.G. 1991, Mutation-Selection Balance and Metabolic Control Theory. *Genetics*, vol. 129, n° 3, p. 909–923, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1204757/>.
- CLOSSET J.L. 1992, Raisonnements en électricité et en hydrodynamique. *Aster*, vol. 14, n° 14, p. 157, URL <http://dx.doi.org/10.4267/2042/9089>.
- COTON C., DILLMANN C., DE VIENNE D. 2021a, Evolution of enzyme levels in metabolic pathways: A theoretical approach. Part 2. *bioRxiv*, p. 2021.11.18.469121, URL <http://dx.doi.org/10.1101/2021.11.18.469121>.
- COTON C., TALBOT G., LE LOUARN M., DILLMANN C., DE VIENNE D. 2021b, *SimEvolEnz-Cons : Simulation of Evolution of Enzyme Under Constraints*. R package version 2.0.0, URL <https://CRAN.R-project.org/package=SimEvolEnzCons>.
- COTON C., TALBOT G., LE LOUARN M., DILLMANN C., DE VIENNE D. 2021c, Evolution of enzyme levels in metabolic pathways: A theoretical approach. *bioRxiv*, p. 2021.05.04.442631, URL <http://dx.doi.org/10.1101/2021.05.04.442631>.
- CROW J.F., KIMURA M. 1970, *An introduction to population genetics theory*. Harper and Row, Publishers, Inc., URL <https://www.cabdirect.org/cabdirect/abstract/19710105376>.
- DARWIN C. 1859, *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London : John Murray, 1<sup>er</sup> édition.
- DEL VECCHIO D., NINFA A.J., SONTAG E.D. 2008, Modular cell biology : retroactivity and insulation. *Molecular Systems Biology*, vol. 4, n° 1, p. 161, URL <http://dx.doi.org/10.1038/msb4100204>.
- DEQUIDT C. 2015, *Étude des conditions d'efficacité d'une analogie mécanique instrumentée pour l'enseignement de l'électrocinétique en terminale scientifique*. Thèse de Doctorat en Sciences de l'Éducation, Université Blaise Pascal - Clermont-Ferrand II, Clermont-Ferrand.
- DROUIN A.M. 1988, LE MODELE EN QUESTIONS. *ASTER*, vol. 7, n° 1, p. 1–20, URL [http://documents.irevues.inist.fr/bitstream/handle/2042/9214/ASTER\\_1988\\_7\\_1.pdf?sequence=1](http://documents.irevues.inist.fr/bitstream/handle/2042/9214/ASTER_1988_7_1.pdf?sequence=1).

- EANES W.F. 2017, New views on the selection acting on genetic polymorphism in central metabolic genes. *Annals of the New York Academy of Sciences*, vol. 1389, n° 1, p. 108–123, URL <http://dx.doi.org/10.1111/nyas.13285>.
- EDERER M., GILLES E.D. 2007, Thermodynamically Feasible Kinetic Models of Reaction Networks. *Biophysical Journal*, vol. 92, n° 6, p. 1846–1857, URL <http://dx.doi.org/10.1529/biophysj.106.094094>.
- EDERER M., GILLES E.D. 2008, Thermodynamic Constraints in Kinetic Modeling: Thermodynamic-Kinetic Modeling in Comparison to Other Approaches. *Engineering in Life Sciences*, vol. 8, n° 5, p. 467–476, URL <http://dx.doi.org/10.1002/elsc.200800040>.
- ELLIS R. 2001, Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences*, vol. 26, n° 10, p. 597–604, URL [http://dx.doi.org/10.1016/S0968-0004\(01\)01938-7](http://dx.doi.org/10.1016/S0968-0004(01)01938-7).
- FIÉVET J.B., DILLMANN C., CURIEN G., DE VIENNE D. 2006, Simplified modelling of metabolic pathways for flux prediction and optimization : lessons from an in vitro reconstruction of the upper part of glycolysis. *Biochemical Journal*, vol. 396, n° Pt 2, p. 317–326, URL <http://dx.doi.org/10.1042/BJ20051520>.
- FIÉVET J.B., DILLMANN C., DE VIENNE D. 2010, Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, vol. 120, n° 2, p. 463–473, URL <http://dx.doi.org/10.1007/s00122-009-1203-2>.
- FIÉVET J.B., NIDELET T., DILLMANN C., DE VIENNE D. 2018, Heterosis Is a Systemic Property Emerging From Non-linear Genotype-Phenotype Relationships: Evidence From in Vitro Genetics and Computer Simulations. *Frontiers in Genetics*, vol. 9, URL <http://dx.doi.org/10.3389/fgene.2018.00159>.
- FLAMHOLZ A., NOOR E., BAR-EVEN A., MILO R. 2012, eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Research*, vol. 40, n° D1, p. D770–D775, URL <http://dx.doi.org/10.1093/nar/gkr874>.
- FOGLIZZO T., MASSET F., GUILLET J., DURAND G. 2012, Shallow Water Analogue of the Standing Accretion Shock Instability : Experimental Demonstration and a Two-Dimensional Model. *Physical Review Letters*, vol. 108, n° 5, p. 051 103, URL <http://dx.doi.org/10.1103/PhysRevLett.108.051103>.
- GAVRILETS S. 2004, *Fitness Landscapes and the Origin of Species*. Princeton and Oxford : Princeton University Press, google-Books-ID : duk9DwAAQBAJ.
- GODEFROY F. 1888, *Dictionnaire de l'ancienne langue française, et de tous ses dialectes du IXe au XVe siècle*, vol. 5. Paris : F. Vieweg, google-Books-ID : BJ0YAAAAIAAJ, URL <http://books.google.com/books?id=BJ0YAAAAIAAJ&oe=UTF-8>.
- GOUT J.F., DURET L., KAHN D. 2009, Differential Retention of Metabolic Genes Following Whole-Genome Duplication. *Molecular Biology and Evolution*, vol. 26, n° 5, p. 1067–1072, URL <http://dx.doi.org/10.1093/molbev/msp026>.



- HALDANE J.B.S. 1930, *Enzymes*. Longmans, Green, google-Books-ID : ECnyUM5WJQgC.
- HARMAND N. 2017, *Evolution expérimentale et spécialisation dans le paysage adaptatif d'un gradient environnemental*. Evolution [q-bio.PE], Université Montpellier, NNT : 2017MONTT062. tel-01704551.
- HARTL D.L., DYKHUIZEN D.E., DEAN A.M. 1985, Limits of Adaptation : The Evolution of Selective Neutrality. *Genetics*, vol. 111, n° 3, p. 655–674, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1202663/>.
- HATZIMANIKATIS V., LI C., IONITA J.A., BROADBELT L.J. 2004, Metabolic networks: enzyme function and metabolite structure. *Current Opinion in Structural Biology*, vol. 14, n° 3, p. 300–306, URL <http://dx.doi.org/10.1016/j.sbi.2004.04.004>.
- HECKMANN D., ZIELINSKI D.C., PALSSON B.O. 2018, Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. *Nature Communications*, vol. 9, n° 1, URL <http://dx.doi.org/10.1038/s41467-018-07649-1>.
- HEINRICH R., RAPOPORT T. 1974, Linear Steady-State Treatment of Enzymatic Chains - General Properties, Control and Effector Strength. *European Journal of Biochemistry*, vol. 42, n° 1, p. 89–95, URL <http://dx.doi.org/10.1111/j.1432-1033.1974.tb03318.x>.
- JOHANNSEN W. 1911, The Genotype Conception of Heredity. *The American Naturalist*, vol. 45, n° 531, p. 129–159, URL <http://dx.doi.org/10.1086/279202>.
- KACSER H., BURNS J.A. 1973, The control of flux. *Symposia of the Society for Experimental Biology*, vol. 27, p. 65–104.
- KACSER H., BURNS J.A. 1981, The Molecular Basis of Dominance. *Genetics*, vol. 97, n° 3-4, p. 639–666, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1214416/>.
- KACSER H., BURNS J.A., KACSER H., FELL D.A. 1995, The control of flux : 21 years on. *Biochemical Society Transactions*, vol. 23, n° 2, p. 341–366, URL <http://dx.doi.org/10.1042/bst0230341>.
- KAGAN M. 2015, On equivalent resistance of electrical circuits. *American Journal of Physics*, vol. 83, n° 1, p. 53–63, URL <http://dx.doi.org/10.1119/1.4900918>.
- KANEHISA M., GOTO S., KAWASHIMA S., NAKAYA A. 2002, The KEGG databases at GenomeNet. *Nucleic Acids Research*, vol. 30, n° 1, p. 42–46, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99091/>.
- KEIGHTLEY P.D. 1996, Metabolic models of selection response. *Journal of Theoretical Biology*, vol. 182, n° 3, p. 311–316, URL <http://dx.doi.org/10.1006/jtbi.1996.0169>.
- KEIGHTLEY P.D., KACSER H. 1987, Dominance, Pleiotropy and Metabolic Structure. *Genetics*, vol. 117, n° 2, p. 319–329, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1203207/>.

- KEMBLE H., EISENHAUER C., COUCE A., CHAPRON A., MAGNAN M., GAUTIER G., NAGARD H.L., NGHE P., TENAILLON O. 2020, Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. *Science Advances*, vol. 6, n° 23, p. eabb2236, URL <http://dx.doi.org/10.1126/sciadv.abb2236>.
- KEMBLE H., NGHE P., TENAILLON O. 2019, Recent insights into the genotype–phenotype relationship from massively parallel genetic assays. *Evolutionary Applications*, vol. 12, n° 9, p. 1721–1742, URL <http://dx.doi.org/10.1111/eva.12846>.
- KIMURA M. 1985, *The neutral theory of molecular evolution*. Cambridge : Cambridge University Press.
- KOEHN R. 1991, The Cost of Enzyme-Synthesis in the Genetics of Energy-Balance and Physiological Performance. *Biological Journal of the Linnean Society*, vol. 44, n° 3, p. 231–247, URL <http://dx.doi.org/10.1111/j.1095-8312.1991.tb00618.x>.
- KRIM A., LAKRIM A., TAHRI D. 2019, Linear Algebra Based Generalization of the Kennelly’s Theorem. *Universal Journal of Electrical and Electronic Engineering*, vol. 6, n° 3, p. 101–107, URL <http://dx.doi.org/10.13189/ujeee.2019.060303>.
- LEGAY J.M. 1997, *L’expérience et le modèle : Un discours sur la méthode*. Sciences en Questions, Paris : INRA Editions.
- LI W.H. 1978, Maintenance of genetic variability under the joint effect of mutation, selection and random drift. *Genetics*, vol. 90, p. 349–383.
- LIEBERMEISTER W. 2014, How enzyme economy shapes metabolic fluxes. *arXiv :1404.5072 [q-bio]*, arXiv : 1404.5072, URL <http://arxiv.org/abs/1404.5072>.
- LIEBERMEISTER W., UHLENDORF J., KLIPP E. 2010, Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*, vol. 26, n° 12, p. 1528–1534, URL <http://dx.doi.org/10.1093/bioinformatics/btq141>.
- LION S., GABRIEL F., BOST B., FIÉVET J., DILLMANN C., DE VIENNE D. 2004, An extension to the metabolic control theory taking into account correlations between enzyme concentrations. *European Journal of Biochemistry*, vol. 271, n° 22, p. 4375–4391, URL <http://dx.doi.org/10.1111/j.1432-1033.2004.04375.x>.
- MANRUBIA S., CUESTA J.A., AGUIRRE J., AHNERT S.E., ALTENBERG L., CANO A.V., CATALÁN P., DIAZ-URIARTE R., ELENA S.F., GARCÍA-MARTÍN J.A., HOGEWEG P., KHATRI B.S., KRUG J., LOUIS A.A., MARTIN N.S., PAYNE J.L., TARNOWSKI M.J., WEISS M. 2021, From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews*, vol. 38, p. 55–106, URL <http://dx.doi.org/10.1016/j.plrev.2021.03.004>.
- MAYNARD SMITH J., SZATHMÁRY E. 1995, *The Major Transitions in Evolution*. Oxford : Oxford University Press.
- MILBORROW B.V. 1998, A biochemical mechanism for hybrid vigour. *Journal of Experimental Botany*, vol. 49, n° 324, p. 1063–1071.

- MOSSIO M., MONTÉVIL M., LONGO G. 2016, Theoretical principles for biology: Organization. *Progress in Biophysics and Molecular Biology*, vol. 122, n° 1, p. 24–35, URL <http://dx.doi.org/10.1016/j.pbiomolbio.2016.07.005>.
- NIELSEN R., SLATKIN M. 2013, *An Introduction to Population Genetics : Theory and Applications*. Sunderland, MA, USA : Sinauer Andrew, Sinauer Associates, Inc.
- OH K.W., LEE K., AHN B., FURLANI E.P. 2012, Design of pressure-driven microfluidic networks using electric circuit analogy. *Lab on a Chip*, vol. 12, n° 3, p. 515–545, publisher : The Royal Society of Chemistry, URL <http://dx.doi.org/10.1039/C2LC20799K>.
- OHTA T. 1992, The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*, vol. 23, n° 1, p. 263–286, URL <http://dx.doi.org/10.1146/annurev.es.23.110192.001403>.
- OLSON H.F. 1958, *Dynamical analogies*. Princeton, New Jersey : Van Nostrand, 2<sup>e</sup> édition, URL <http://hdl.handle.net/2027/mdp.49015000700550>.
- ORGOGOZO V., MORIZOT B., MARTIN A. 2015, The differential view of genotype–phenotype relationships. *Frontiers in Genetics*, vol. 6, URL <http://dx.doi.org/10.3389/fgene.2015.00179>.
- ORTH J.D., THIELE I., PALSSON B.O. 2010, What is flux balance analysis? *Nature biotechnology*, vol. 28, n° 3, p. 245–248, URL <http://dx.doi.org/10.1038/nbt.1614>.
- PETRIZZELLI M. 2019, *Mathematical modelling and integration of complex biological data : analysis of the heterosis phenomenon in yeast*. Populations and Evolution [q-bio.PE], Université Paris-Saclay, NNT : 2019SACLS204.
- PRICE N.D., PAPIN J.A., SCHILLING C.H., PALSSON B.O. 2003, Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology*, vol. 21, n° 4, p. 162–169, URL [http://dx.doi.org/10.1016/S0167-7799\(03\)00030-1](http://dx.doi.org/10.1016/S0167-7799(03)00030-1).
- PROULX S.R., PROMISLOW D.E.L., PHILLIPS P.C. 2005, Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, vol. 20, n° 6, p. 345–353, URL <http://dx.doi.org/10.1016/j.tree.2005.04.004>.
- ROSEN R. 1993, On models and modeling. *Applied Mathematics and Computation*, vol. 56, n° 2-3, p. 359–372, URL [http://dx.doi.org/10.1016/0096-3003\(93\)90128-2](http://dx.doi.org/10.1016/0096-3003(93)90128-2).
- RUMELHARD G. 1988, STATUT ET ROLE DES MODELES DANS LE TRAVAIL SCIENTIFIQUE ET DANS L'ENSEIGNEMENT DE LA BIOLOGIE. *ASTER*, vol. 7, n° 2, p. 21–48, URL [http://documents.irevues.inist.fr/bitstream/handle/2042/9215/ASTER\\_1988\\_7\\_21.pdf?sequence=1](http://documents.irevues.inist.fr/bitstream/handle/2042/9215/ASTER_1988_7_21.pdf?sequence=1).
- SAA P.A., NIELSEN L.K. 2017, Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances*, vol. 35, n° 8, p. 981–1003, URL <http://dx.doi.org/10.1016/j.biotechadv.2017.09.005>.
- SCHOMBURG I., CHANG A., SCHOMBURG D. 2002, BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, vol. 30, n° 1, p. 47–49, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99121/>.

- SEROHIJOS A.W.R., RIMAS Z., SHAKHNOVICH E.I. 2012, Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Reports*, vol. 2, n° 2, p. 249–256, URL <http://dx.doi.org/10.1016/j.celrep.2012.06.022>.
- SERVEDIO M.R., BRANDVAIN Y., DHOLE S., FITZPATRICK C.L., GOLDBERG E.E., STERN C.A., CLEVE J.V., YEH D.J. 2014, Not Just a Theory—The Utility of Mathematical Models in Evolutionary Biology. *PLOS Biology*, vol. 12, n° 12, p. e1002017, URL <http://dx.doi.org/10.1371/journal.pbio.1002017>.
- SIMON H.A. 1962, The Architecture of Complexity. *Proceedings of the American Philosophical Society*, vol. 106, n° 6, p. 467–482, URL <https://www.jstor.org/stable/985254>.
- SOLER L. 2013, Qu'est-ce qu'un modèle scientifique? Des caractéristiques du modèle qui importent du point de vue de l'enseignement intégré de science et de technologie. *Spirale - Revue de recherches en éducation*, vol. 52, n° 1, p. 177–214, URL <http://dx.doi.org/10.3406/spira.2013.1067>.
- STAVRIANEAS S., SILVERSTEIN T. 2005, Teaching glycolysis regulation to undergraduates using an electrical power generation analogy. *Advances in Physiology Education*, vol. 29, n° 2, p. 128–130, publisher : American Physiological Society, URL <http://dx.doi.org/10.1152/advan.00004.2005>.
- SUGITA M. 1954, MATHEMATICAL ANALYSIS OF THE METABOLISM AND ANALOGY OF ECONOMICS. *The Annals of the Hitotsubashi Academy*, vol. 4, n° 2, p. 163–182, Publisher : Hitotsubashi University, URL <https://www.jstor.org/stable/43751296>.
- SZATHMÁRY E. 1993, Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics*, vol. 133, n° 1, p. 127–132, URL <http://www.genetics.org/content/133/1/127>.
- DE VIENNE D. 2021, What is a phenotype? History and new developments of the concept. *Genetica*, URL <http://dx.doi.org/10.1007/s10709-021-00134-6>.
- DE VIENNE D., BOST B., FIÉVET J., DILLMANN C. 2001, Optimisation of Enzyme Concentrations for Unbranched Reaction Chains: The Concept of Combined Response Coefficient. *Acta Biotheoretica*, vol. 49, n° 4, p. 341–350, URL <http://dx.doi.org/10.1023/A:1014290829329>.
- VRBA E.S., ELDREDGE N. 1984, Individuals, hierarchies and processes: towards a more complete evolutionary theory. *Paleobiology*, vol. 10, n° 2, p. 146–171, URL <http://dx.doi.org/10.1017/S0094837300008149>.
- WHITLOCK M.C., PHILLIPS P.C. 2000, The exquisite corpse: a shifting view of the shifting balance. *Trends in Ecology & Evolution*, vol. 15, n° 9, p. 347–348, URL [http://dx.doi.org/10.1016/S0169-5347\(00\)01930-3](http://dx.doi.org/10.1016/S0169-5347(00)01930-3).
- WILBANKS B., TRINH C.T. 2017, Comprehensive characterization of toxicity of fermentative metabolites on microbial growth. *Biotechnology for Biofuels*, vol. 10, URL <http://dx.doi.org/10.1186/s13068-017-0952-4>.
- WRIGHT S. 1932, The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, vol. 1, p. 356–366.

- WRIGHT S. 1934, Physiological and Evolutionary Theories of Dominance. *The American Naturalist*, vol. 68, n° 714, p. 24–53.
- WRIGHT S. 1937, The Distribution of Gene Frequencies in Populations. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 23, n° 6, p. 307–320, URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1076930/>.
- YI X., DEAN A.M. 2019, Adaptive Landscapes in the Age of Synthetic Biology. *Molecular Biology and Evolution*, vol. 36, n° 5, p. 890–907, URL <http://dx.doi.org/10.1093/molbev/msz004>.
- YIZHAK K., BENYAMINI T., LIEBERMEISTER W., RUPPIN E., SHLOMI T. 2010, Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, vol. 26, n° 12, p. i255–i260, URL <http://dx.doi.org/10.1093/bioinformatics/btq183>.

---

# ANNEXES

---



# **Annexe A**

## **Article 1 : figures supplémentaires et démonstrations**

La première section contient les *Supporting Information*, avec les figures supplémentaires et démonstrations, qui accompagnent le premier article (chapitre 1). Dans la deuxième section, je donne quelques démonstrations complémentaires, notamment pour préciser l'existence de la droite entre les concentrations en cas de co-régulation.





# A.1

## Evolution of enzyme levels in metabolic pathways: A theoretical approach

### *Supporting Information*

Charlotte Coton, Grégoire Talbot, Maud Le Louarn,  
Christine Dillmann, Dominique de Vienne

May 4, 2021

5

## SI Supporting Information

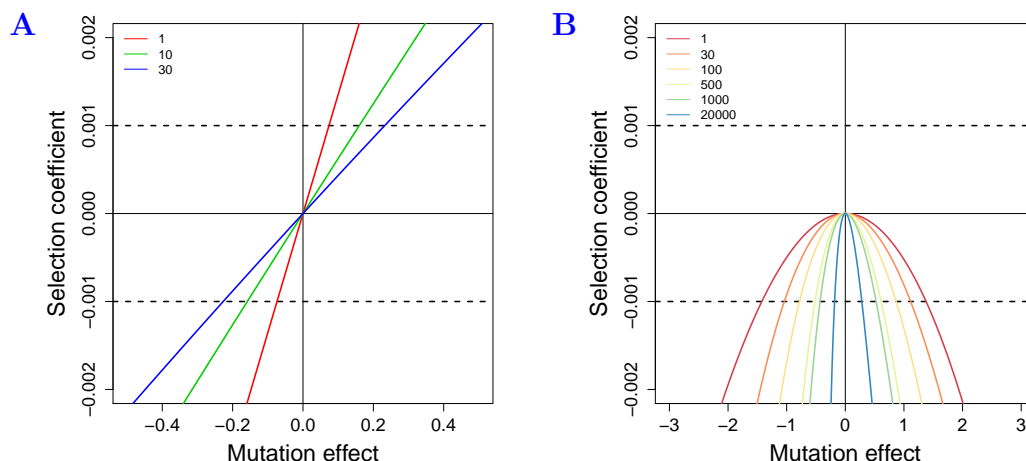
### Contents

	<b>SI Supporting Information</b>	<b>1</b>
	SI.A Supplementary Figures . . . . .	3
10	SI.B Mathematical proofs . . . . .	6
	SI.B.1 Properties of the redistribution coefficient and mutation effects . . .	6
	SI.B.1.1 The matrix of pairwise co-regulation coefficients . . . . .	6
	SI.B.1.2 Properties of the global co-regulation coefficients . . . . .	6
	SI.B.1.3 Mutation effect and redistribution coefficient in the case	
15	of competition . . . . .	6
	SI.B.1.4 Mutation effect and redistribution coefficient in the case	
	of competition plus co-regulation . . . . .	7
	SI.B.2 Explicit expression of the response coefficient . . . . .	9
	SI.B.3 Selective neutrality and range of neutral variation (RNV) . . . . .	9
20	SI.B.3.1 Expressing the selection coefficient in term of the response	
	coefficient . . . . .	9
	SI.B.3.2 Computing the limits of the RNV . . . . .	10
	SI.B.4 Evolutionary equilibrium in the case of independence (case 1) . . .	10
	SI.B.5 Evolutionary equilibrium in the case of competition (case 2) . . . .	12
25	SI.B.6 Evolutionary equilibria in the case of co-regulation (case 3) . . . .	14
	SI.B.6.1 Theoretical equilibrium . . . . .	14
	SI.B.6.1.1 Expression of the theoretical equilibrium . . . . .	14
	SI.B.6.1.2 Stability of the theoretical equilibrium . . . . .	15
	SI.B.6.1.3 Response coefficient at the theoretical equilibrium	16
30	SI.B.6.2 The parametric equation of the line relating the relative	
	enzyme concentrations. . . . .	16
	SI.B.6.3 Effective equilibrium in the case of negative co-regulation .	17
	SI.B.7 Evolutionary equilibrium in the case of competition plus co-regulation	
	(case 4) . . . . .	17
35	SI.B.7.1 Theoretical equilibrium . . . . .	17
	SI.B.7.1.1 Expression of the theoretical equilibrium . . . . .	17

---

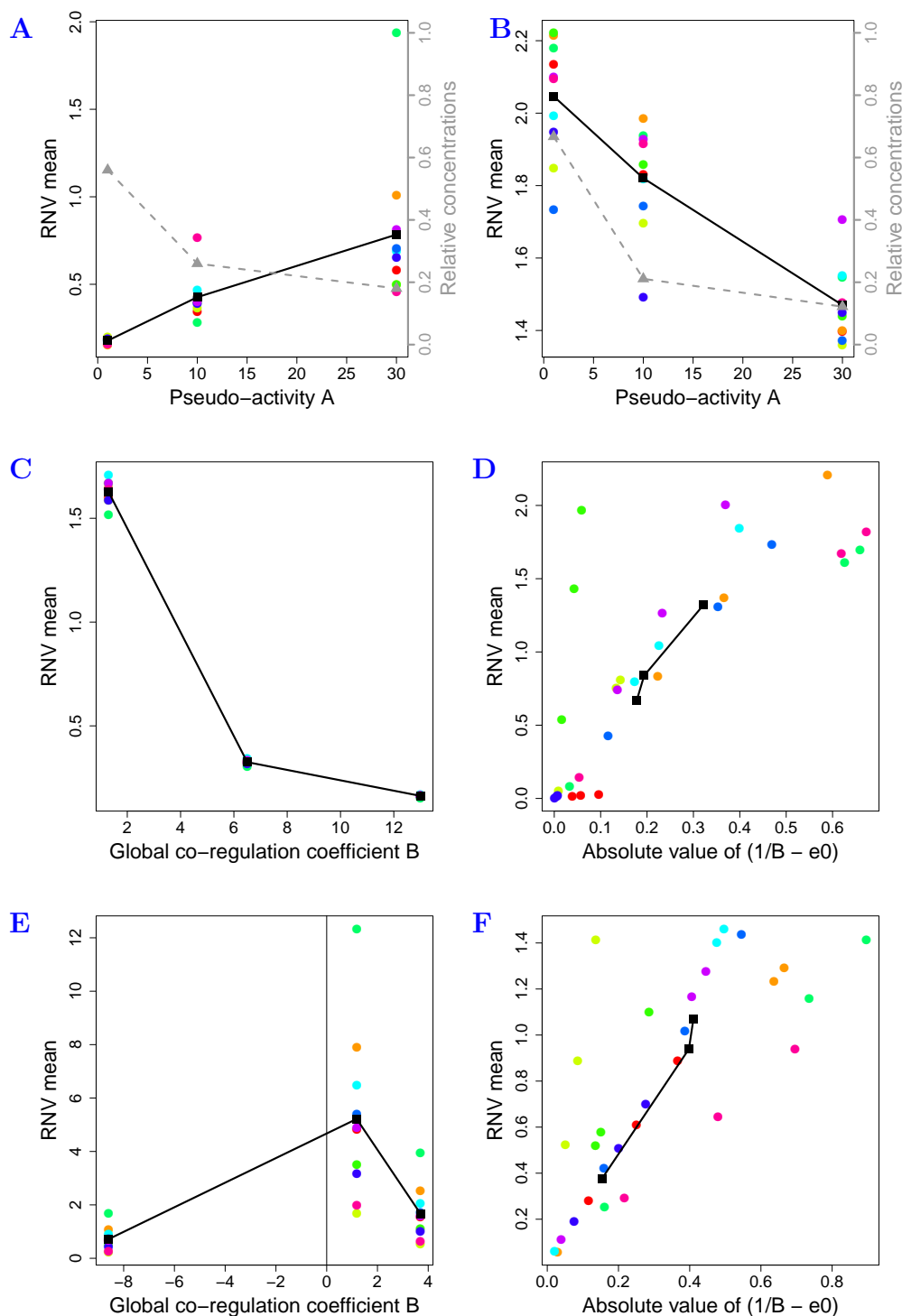
	SI.B.7.1.2 Stability of the theoretical equilibrium. . . . .	18
	SI.B.7.2 Effective equilibrium . . . . .	19
40	SI.B.8 Expressing absolute enzyme concentrations in function of the driving variable . . . . .	19
	SI.B.8.1 In the case of co-regulation . . . . .	19
	SI.B.8.2 In the case of co-regulation and competition . . . . .	20
	SI.B.9 Evolution model of the driving variable . . . . .	20
	SI.B.9.1 Redistribution coefficient and mutation effect . . . . .	20
45	SI.B.9.1.1 In the case of co-regulation . . . . .	21
	SI.B.9.1.2 In the case of competition and co-regulation . . . . .	21
	SI.B.9.2 Evolution model of the driving variable and evolutionary equilibria . . . . .	22

## SI.A Supplementary Figures

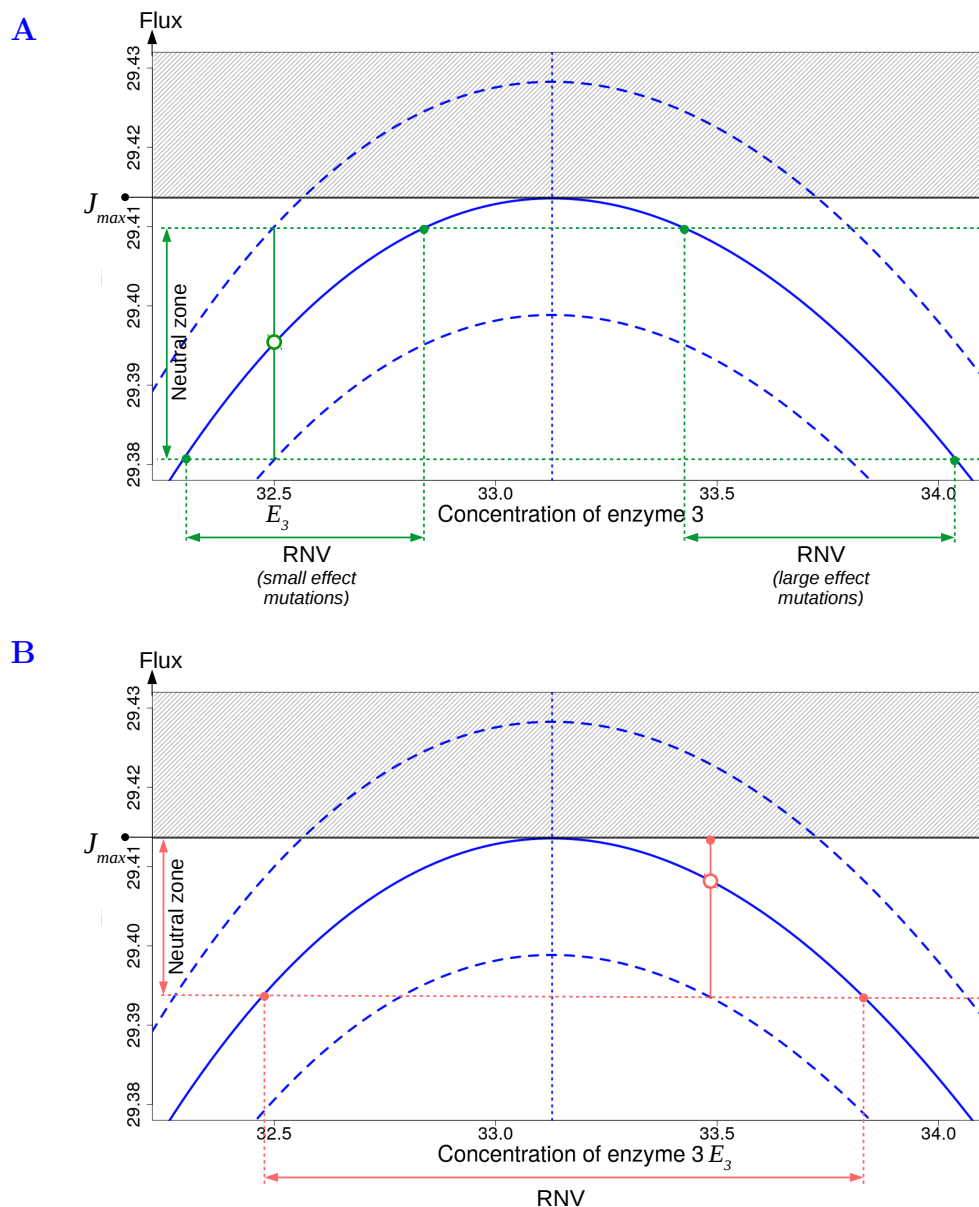


**Figure S1. Range of neutral variation in the case of independence and competition.**

(A) Relationship between the selection coefficient  $s_i = \Delta J/J$  and the actual effect of a mutation  $\delta_i$  of enzyme  $i$  at evolutionary equilibrium of relative enzyme concentrations in a pathway of three enzymes that vary independently. The pseudo-activities are  $A_1 = 1$  (red),  $A_2 = 10$  (green) and  $A_3 = 30$  (blue).  $E_i^0$ 's are respectively 55.99, 25.99 and 18.02, corresponding to the equilibrium values when  $E_{\text{tot}}^0 = 100$ . The neutral zone is located between the two horizontal dotted lines at  $s_i = \pm 0.001$ . For the most active enzyme (blue curve), larger mutation effects are required to get out the neutral zone than for the least active enzyme (red curve). The curve of enzyme with intermediate pseudo-activity (green curve) is between that of the most and least active enzyme. Therefore there is a positive relationship between pseudo-activity and RNV size. (B) Representation as in (A) of a six-enzyme pathway with competition ( $E_{\text{tot}} = 100$ ). The pseudo-activities are:  $A_1 = 10$ ,  $A_2 = 30$ ,  $A_3 = 100$ ,  $A_4 = 500$ ,  $A_5 = 1000$  and  $A_6 = 20000$  (color palette from red to blue). The selection coefficients are all negative because any deviation from the equilibrium value results in negative  $s_i$ . Unlike in (A) the most active enzymes have the narrowest range of neutral variation.



**Figure S2. Relationship between constraint-dependent factors and RNV from simulations of long-term evolution of a three-enzyme pathway.** Representation is as in Figure 6. For each of the ten simulations (one color *per* simulation), the RNV of each enzyme was computed as the mean of the RNV sizes between 60,000 and 120,000 generations, *i.e.* once equilibrium is reached. Black points are the mean values of all simulations. Gray points and dashed line in (A) and (B) indicate the relative enzyme concentrations at evolutionary equilibrium. The colored points in (D) and (F) are not vertically aligned because the different simulations do not have the same  $e^0$  values. We see that the results are qualitatively the same as in Figure 6. Note the positive relationship between RNV size and RNV variability. Parameter values are as in Figures 1 to 5.



**Figure S3. Range of neutral variation (RNV) in the case of negative co-regulation.**

The flux (blue curve) is function of the concentration of one enzyme (enzyme 3) in a three-enzyme pathway.  $\tilde{J}_{max}$  is reached at effective equilibrium  $\tilde{E}_3 = 33.13$  (vertical blue dotted line). The upper and lower dashed blue curves delimit the neutral zone of flux variation. Due to the existence of a maximum flux, the gray area is unattainable. (A) The concentration of enzyme 3 is far from the equilibrium, at  $E_3 \approx 32.5$ . The vertical green line shows the amplitude of the neutral zone (between the horizontal green dotted lines). The corresponding RNV splits into two parts (horizontal green left and right arrows), because mutations of large effect can lead to flux values that are similar to those caused by mutations of small effect. (B) The concentration of enzyme 3 is closer to the effective equilibrium, at  $E_3 \approx 33.5$ . Here the upper limit of the neutral zone cannot be achieved because it exceeds the maximum flux. The continuous RNV (horizontal red left and right arrow) is bounded by the enzyme concentrations 32.45 and 33.85 corresponding to the inferior limit of neutral zone on both sides of the curve. Parameter values are:  $X = 1$ ,  $A_1 = 1$ ,  $A_2 = 10$ ,  $A_3 = 30$ ,  $E_{tot} = 100$ ,  $\beta_{12} = -8.75$ ,  $\beta_{23} = -2.25$ .

50 **SI.B Mathematical proofs**
**SI.B.1 Properties of the redistribution coefficient and mutation effects**

**SI.B.1.1 The matrix of pairwise co-regulation coefficients** From any vector  $\boldsymbol{\beta}$  of  $\beta_{i-}$ 's, all elements of the matrix  $\mathbf{M}_\beta$  of pairwise co-regulation coefficients can be deduced from the relation  $\mathbf{M}_\beta = \boldsymbol{\beta}^{\text{inv}} \boldsymbol{\beta}^\top$ , where  $\boldsymbol{\beta}^{\text{inv}}$  is the vector of inverse elements of  $\boldsymbol{\beta}$  (Table S1). Note that  $\mathbf{M}_\beta$  is symmetric.

**Table S1. Matrix  $\mathbf{M}_\beta$  of the co-regulation coefficients.**

$i \backslash j$	$\beta_{-1}$	$\beta_{-2}$	$\cdots$	$\beta_{-i}$	$\beta_{-j}$	$\cdots$	$\beta_{-n}$	Sum
$\beta_{1-}$	1	$\beta_{i2}/\beta_{i1}$	$\cdots$	$\beta_{ii}/\beta_{i1}$	$\beta_{ij}/\beta_{i1}$	$\cdots$	$\beta_{in}/\beta_{i1}$	$B_1$
$\beta_{2-}$	$\beta_{i1}/\beta_{i2}$	1	$\cdots$	$\beta_{ii}/\beta_{i2}$	$\beta_{ij}/\beta_{i2}$	$\cdots$	$\beta_{in}/\beta_{i2}$	$B_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
$\beta_{i-}$	$\beta_{i1}/\beta_{ii}$	$\beta_{i2}/\beta_{ii}$	$\vdots$	1	$\beta_{ij}/\beta_{ii}$	$\cdots$	$\beta_{in}/\beta_{ii}$	$B_i$
$\beta_{j-}$	$\beta_{i1}/\beta_{ij}$	$\beta_{i2}/\beta_{ij}$	$\vdots$	$\beta_{ii}/\beta_{ij}$	1	$\cdots$	$\beta_{in}/\beta_{ij}$	$B_j$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$	$\vdots$
$\beta_{n-}$	$\beta_{i1}/\beta_{in}$	$\beta_{i2}/\beta_{in}$	$\vdots$	$\beta_{ii}/\beta_{in}$	$\beta_{ij}/\beta_{in}$	$\vdots$	1	$B_n$

**SI.B.1.2 Properties of the global co-regulation coefficients** Let a group of co-regulated enzymes, *i.e.*  $\beta_{ij} \neq 0$  for all  $(i, j)$ . Calculating global co-regulation coefficient leads to:

$$\forall i \quad B_i = \sum_{k=1}^n \beta_{ik} = \sum_{k=1}^n \beta_{ij} \beta_{jk} = \beta_{ij} \sum_{k=1}^n \beta_{jk} = \beta_{ij} B_j \quad (\text{S1})$$

$$\boxed{B_i = \beta_{ij} B_j} \quad (\text{S2})$$

60 Therefore:

1. If  $\forall j \neq i, \beta_{ij} > 0$ , then  $B_i = \sum_{k=1}^n \beta_{ik} > 1$  and  $B_j > 1$ .
2. If it exists at least  $(i, j)$  such as  $\beta_{ij} < 0$  and  $B_i > 0$ , then  $B_j < 0$ .
3. If it exists at least  $(i, j)$  such as  $\beta_{ij} < 0$  and  $B_i < 0$ , then  $B_j > 0$ .

Thus, if there is at least one negative co-regulation coefficient, there is at least one negative global co-regulation coefficient.

**SI.B.1.3 Mutation effect and redistribution coefficient in the case of competition** Let us consider a resident population with genotype  $\mathbf{E}$  and a mutation of canonical effect  $\nu$  targeting enzyme  $i$ . To formalize the consequences of the mutation on genotype  $\mathbf{E}$  when there is competition for cellular resources, *i.e.* constant total enzyme concentration, we considered a two-step process: first, the mutation modifies the concentration of enzyme  $i$ , second, there is reduction from  $E_{\text{tot}} + \nu$  to  $E_{\text{tot}}$ , with proportional redistribution of enzyme concentrations:

Concentration	Enzyme $i$	Enzyme $j \neq i$	Total concentration
Resident	$E_i$	$E_j$	$E_{\text{tot}}$
After mutation targeting $i$	$E_i^{(1)} = E_i + \nu$	$E_j^{(1)} = E_j$	$E_{\text{tot}}^{(1)} = E_{\text{tot}} + \nu$
After constraint on $E_{\text{tot}}$	$E_i^{(2)} = E_i + \delta_i$	$E_j^{(2)}$	$E_{\text{tot}}$

Assuming that the proportions of enzyme concentrations are conserved through the redistribution, we have for the targeted enzyme  $i$ :

$$\frac{E_i + \nu}{E_{\text{tot}} + \nu} = \frac{E_i^{(2)}}{E_{\text{tot}}} \quad (\text{S3})$$

75 The actual effect of the mutation is:

$$\delta_i = E_i^{(2)} - E_i = E_{\text{tot}} \frac{E_i + \nu}{E_{\text{tot}} + \nu} - E_i$$

As  $E_i = e_i E_{\text{tot}}$ , we get after rearrangements:

$$\delta_i = \frac{1 - e_i}{\frac{1}{\nu} + \frac{1}{E_{\text{tot}}}} \quad (\text{S4})$$

For enzymes  $j \neq i$ , we have:

$$\frac{E_j}{E_{\text{tot}} + \nu} = \frac{E_j^{(2)}}{E_{\text{tot}}} \quad (\text{S5})$$

So

$$E_j^{(2)} = \frac{E_j}{E_{\text{tot}} + \nu} E_{\text{tot}}$$

In order to express  $E_j^{(2)}$  in terms of  $E_j$  and  $\delta_i$ , we write:

$$E_j^{(2)} - E_j = E_{\text{tot}} \frac{E_j}{E_{\text{tot}} + \nu} - E_j = -\frac{E_j \nu}{E_{\text{tot}} + \nu} = -\frac{e_j}{\frac{1}{\nu} + \frac{1}{E_{\text{tot}}}}$$

80 and using equation S4, we get:

$$E_j^{(2)} = E_j - \frac{e_j}{1 - e_i} \delta_i \quad (\text{S6})$$

Thus, the term  $\frac{-e_j}{1 - e_i}$  is the redistribution coefficient  $\alpha_{ij}$  when there is competition.

**SI.B.1.4 Mutation effect and redistribution coefficient in the case of competition plus co-regulation** If there is co-regulation in addition to competition, the two-step process described above is slightly modified, because the mutation targeting enzyme  $i$  also affects indirectly the co-regulated enzymes. So we introduce the co-regulation coefficients  $\beta_{ij}$  in the first step:

Concentration	Enzyme $i$	Enzyme $j \neq i$	Total concentration
Resident	$E_i$	$E_j$	$E_{\text{tot}}$
Mutation + co-regulation	$E_i^{(1)} = E_i + \nu$	$E_j^{(1)} = E_j + \beta_{ij} \nu$	$E_{\text{tot}}^{(1)} = E_{\text{tot}} + B_i \nu$
Constraint on $E_{\text{tot}}$	$E_i^{(2)} = E_i + \delta_i$	$E_j^{(2)}$	$E_{\text{tot}}$



Assuming that the proportions of the enzyme concentrations are conserved through the redistribution, we have for the targeted enzyme  $i$ :

$$\frac{E_i + \nu}{E_{\text{tot}} + B_i \nu} = \frac{E_i^{(2)}}{E_{\text{tot}}} \quad (\text{S7})$$

The actual effect of the mutation is:

$$\delta_i = E_i^{(2)} - E_i$$

90 So, from equation S7, we have:

$$\delta_i = E_{\text{tot}} \frac{E_i + \nu}{E_{\text{tot}} + B_i \nu} - E_i$$

As  $E_i = e_i E_{\text{tot}}$ , we get after rearrangements:

$$\boxed{\delta_i = \frac{1 - B_i e_i}{\frac{1}{\nu} + \frac{B_i}{E_{\text{tot}}}}} \quad (\text{S8})$$

For enzymes  $j \neq i$ , we have:

$$\frac{E_j + \beta_{ij} \nu}{E_{\text{tot}} + B_i \nu} = \frac{E_j^{(2)}}{E_{\text{tot}}} \quad (\text{S9})$$

So:

$$E_j^{(2)} = \frac{E_j + \beta_{ij} \nu}{E_{\text{tot}} + B_i \nu} E_{\text{tot}}$$

In order to express  $E_j^{(2)}$  in terms of  $E_j$  and parameters depending on  $E_i$ , we write:

$$E_j^{(2)} - E_j = E_{\text{tot}} \frac{E_j + \beta_{ij} \nu}{E_{\text{tot}} + B_i \nu} - E_j = \frac{E_{\text{tot}} \beta_{ij} \nu - E_j B_i \nu}{E_{\text{tot}} + B_i \nu} = \frac{\beta_{ij} - B_i e_j}{\frac{1}{\nu} + \frac{B_i}{E_{\text{tot}}}}$$

95 and using equation S8, we get:

$$\boxed{E_j^{(2)} = E_j + \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i} \delta_i} \quad (\text{S10})$$

Thus, when there is competition plus co-regulation, the redistribution coefficient can be written:

$$\boxed{\alpha_{ij} = \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i}} \quad (\text{S11})$$

From this expression, we can distinguish two particular cases:

1. If there is only competition, *i.e.*  $\forall i, j \neq i, \beta_{ij} = 0$  and  $B_i = 1$ , we have:

$$\alpha_{ij} = \frac{-e_j}{1 - e_i}$$

- 100 2. If all enzymes are co-regulated, we have  $\beta_{ij} = B_i/B_j$  (equation S2), and therefore:

$$\alpha_{ij} = \frac{1/B_j - e_j}{1/B_i - e_i}$$

Note that this relation is also valid for  $\alpha_{ii} = 1$  and has the same properties as the  $\beta$ 's, *i.e.*  $\alpha_{ij} = \alpha_{ik} \alpha_{kj}$  and  $\alpha_{ij} = 1/\alpha_{ji}$ .

In the case of co-regulation without competition,  $\mathbf{E}^m$  is equal to  $\mathbf{E}^{(1)}$ , and we have  $\alpha_{ij} = \beta_{ij}$  for all  $(i, j)$  and  $\delta_i = \nu$ .

105 **SI.B.2 Explicit expression of the response coefficient**

The response coefficient is defined by:

$$\forall i \quad R_{E_i}^J = \frac{\partial J/J}{\partial E_i/E_i}$$

Introducing the expression of  $J$  (equation 1), we get:

$$\begin{aligned} \forall i \quad R_{E_i}^J &= -X \frac{E_i}{J} \frac{\frac{\partial \sum_{j=1}^n \frac{1}{A_j E_j}}{\partial E_i}}{\left( \sum_{j=1}^n \frac{1}{A_j E_j} \right)^2} \\ &= X \frac{E_i}{\frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}}} \frac{\sum_{j=1}^n \frac{\partial E_j / \partial E_i}{A_j E_j^2}}{\left( \sum_{j=1}^n \frac{1}{A_j E_j} \right)^2} \\ &= E_i \frac{\sum_{j=1}^n \frac{\partial E_j / \partial E_i}{A_j E_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}} \end{aligned}$$

As  $\alpha_{ij}$  is the derivative of  $E_j$  with respect to  $E_i$  (*Material & Methods*), we obtain:

$$\boxed{\forall i, \quad R_{E_i}^J = E_i \frac{\sum_{j=1}^n \frac{\alpha_{ij}}{A_j E_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}} = e_i \frac{\sum_{j=1}^n \frac{\alpha_{ij}}{A_j e_j^2}}{\sum_{j=1}^n \frac{1}{A_j e_j}}} \quad (\text{S12})$$

**SI.B.3 Selective neutrality and range of neutral variation (RNV)**

**SI.B.3.1 Expressing the selection coefficient in term of the response coefficient**

110 **Let us consider a mutation of actual effect  $\delta_i$  that changes the concentration of enzyme  $i$  of the resident from  $E_i^r$  to  $E_i^m$ :**

$$\delta_i = E_i^m - E_i^r = \Delta E_i$$

and the flux from  $J^r$  to  $J^m$ :

$$J^m - J^r = \Delta J$$

Because the flux is assumed to be proportional to fitness, the coefficient of selection of the mutant is:

$$s_i = \left( \frac{\Delta J}{J} \right)_i$$

115 If the mutation effect  $\delta_i$  is small compared to concentration  $E_i$ , we have:

$$\frac{s_i}{\delta_i/E_i} = \frac{\Delta J/J}{\Delta E_i/E_i} \approx \frac{\partial J/J}{\partial E_i/E_i} \quad (\text{S13})$$

where we recognize the response coefficient of enzyme  $i$ . Therefore, the coefficient of selection is:

$$\boxed{s_i \approx R_{E_i}^J \frac{\delta_i}{E_i}} \quad (\text{S14})$$

**SI.B.3.2 Computing the limits of the RNV** To determine the values  $\delta_i^{\text{sup}}$  and  $\delta_i^{\text{inf}}$  that respectively define the superior and inferior limits of the RNV, we used the equation of the selection coefficient that depends on  $J^r$  and  $J_i^m$ , the flux value before and after a mutation targeting the enzyme  $i$  (equation 15).

Writing

$$J_i^m = J^r(s_i + 1)$$

and replacing the  $J$ 's by their respective expressions, we get:

$$\frac{X}{\sum_{j=1}^n \frac{1}{A_j(E_j + \alpha_{ij}\delta_i)}} = (s_i + 1) \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}}$$

Taking the inverse, we obtain:

$$\boxed{\sum_{j=1}^n \frac{1}{A_j(E_j + \alpha_{ij}\delta_i)} = \frac{1}{s_i + 1} \sum_{j=1}^n \frac{1}{A_j E_j}} \quad (\text{S15})$$

Resolving this equation for  $s_i = 1/N$  (resp.  $s_i = -1/N$ ) gives  $\delta_i^{\text{sup}}$  (resp.  $\delta_i^{\text{inf}}$ ).

#### SI.B.4 Evolutionary equilibrium in the case of independence (case 1)

When there is independence between enzyme concentrations, we have  $\alpha_{ii} = 1$  for all  $i$  and  $\alpha_{ij} = 0$  for all  $j \neq i$  and the differential equation system describing the evolution of the absolute concentrations becomes:

$$\forall j \quad \frac{\partial E_j}{\partial t} = 2\mu s_j \delta_j$$

and the differential equation for total concentration writes:

$$\begin{aligned} \frac{\partial E_{\text{tot}}}{\partial t} &= \sum_{k=1}^n \frac{\partial E_k}{\partial t} \\ &= \sum_{k=1}^n 2\mu s_k \delta_k \\ &= 2\mu \sum_{k=1}^n s_k \delta_k \end{aligned}$$

These expressions allow us to write the differential equation system describing the evolution of the *relative* concentrations:

$$\forall j \quad \frac{\partial e_j}{\partial t} = \frac{\partial(E_j/E_{\text{tot}})}{\partial t} = \frac{\frac{\partial E_j}{\partial t} E_{\text{tot}} - E_j \frac{\partial E_{\text{tot}}}{\partial t}}{(E_{\text{tot}})^2}$$

By replacing the derivatives by their expression, we get:

$$\begin{aligned} \forall j \quad \frac{\partial e_j}{\partial t} &= \frac{E_{\text{tot}} 2\mu s_j \delta_j - e_j E_{\text{tot}} 2\mu \sum_{k=1}^n s_k \delta_k}{(E_{\text{tot}})^2} \\ &= \frac{2\mu}{E_{\text{tot}}} (s_j \delta_j - e_j \sum_{k=1}^n s_k \delta_k) \end{aligned}$$

Due to the relationship between selection and response coefficients (equation S14), we can write:

$$\begin{aligned}\forall j \quad \frac{\partial e_j}{\partial t} &= \frac{2\mu}{E_{\text{tot}}} \left( R_{E_j}^J \frac{\delta_j^2}{E_j} - e_j \sum_{k=1}^n R_{E_k}^J \frac{\delta_k^2}{E_k} \right) \\ &= \frac{2\mu}{E_{\text{tot}}^2} \left( R_{E_j}^J e_j \frac{\delta_j^2}{e_j} - e_j \sum_{k=1}^n R_{E_k}^J \frac{\delta_k^2}{e_k} \right)\end{aligned}$$

As the steady state of relative enzyme concentrations is defined by

$$\forall j, \quad \frac{\partial e_j}{\partial t} = 0$$

the relative concentrations  $e_j^*$  at evolutionary equilibrium can be found by solving:

$$\forall j, \quad R_{E_j}^{J*} \frac{\delta_j^2}{e_j^*} = e_j^* \sum_{k=1}^n R_{E_k}^{J*} \frac{\delta_k^2}{e_k^*}$$

- 130 When there is independence, the actual and canonical effects of the mutation are equal, *i.e.*  $\delta_i = \nu$  for all  $i$ . So we have:

$$\forall j, \quad R_{E_j}^{J*} \frac{\nu^2}{e_j^*} = \nu^2 e_j^* \sum_{k=1}^n \frac{R_{E_k}^{J*}}{e_k^*},$$

Leading to:

$$\forall j \quad e_j^* = \frac{\frac{R_{E_j}^{J*}}{e_j^*}}{\sum_{k=1}^n \frac{R_{E_k}^{J*}}{e_k^*}}$$

Replacing the response coefficient by its expression in the case of independence (equation S12), *i.e.* with  $\alpha_{ii} = 1$  and  $\alpha_{ij} = 0, \forall j \neq i$ , we obtain:

$$\forall j, \quad e_j^* = \frac{\frac{\frac{1}{A_j e_j^* E_{\text{tot}}}}{\sum_{l=1}^n \frac{1}{A_l E_l}} / e_j^*}{\sum_{k=1}^n \left( \frac{\frac{1}{A_k e_k^* E_{\text{tot}}}}{\sum_{l=1}^n \frac{1}{A_l E_l}} / e_k^* \right)} = \frac{\frac{1}{A_j (e_j^*)^2}}{\sum_{k=1}^n \frac{1}{A_k (e_k^*)^2}}$$

- 135 Then:

$$\sum_{k=1}^n \frac{1}{A_k (e_k^*)^2} = \frac{1}{A_j (e_j^*)^3}$$

As  $\sum_{k=1}^n \frac{1}{A_k (e_k^*)^2}$  is constant, this equality is valid for any enzyme:

$$\forall i, \forall j \quad \sum_{k=1}^n \frac{1}{A_k (e_k^*)^2} = \frac{1}{A_j (e_j^*)^3} = \frac{1}{A_i (e_i^*)^3}$$

Because the pseudo-activities are strictly positive, we get:

$$\forall j, \forall i, \quad e_j^* = \frac{A_i^{1/3} e_i^*}{A_j^{1/3}}$$

Because  $\sum_{j=1}^n e_j^* = 1$ , we have:

$$\sum_{j=1}^n e_j^* = \sum_{j=1}^n \frac{A_i^{1/3} e_i^*}{A_j^{1/3}} = A_i^{1/3} e_i^* \sum_{j=1}^n \frac{1}{A_j^{1/3}} = 1$$

This leads to the expression of the equilibrium of the relative enzyme concentrations:

$$\forall i, \quad e_i^* = \frac{A_i^{-1/3}}{\sum_{j=1}^n A_j^{-1/3}} \quad (\text{S16})$$

140 Introducing  $e_i^*$  in equation S12, we get the flux response coefficient at equilibrium of relative enzyme concentrations:

$$R_{E_i}^{J^*} = \frac{\frac{1}{A_i e_i^*}}{\sum_{k=1}^n \frac{1}{A_k e_k^*}}$$

and after rearrangements, we obtain:

$$\forall i, R_{E_i}^{J^*} = \frac{A_i^{-2/3}}{\sum_{j=1}^n A_j^{-2/3}} \quad (\text{S17})$$

### SI.B.5 Evolutionary equilibrium in the case of competition (case 2)

When there is competition,  $E_{\text{tot}}$  is constant and the redistribution coefficient is:  $\forall i, \alpha_{ii} = 1$

145 and  $\forall j \neq i, \alpha_{ij} = \frac{-e_j}{1 - e_i}$ . The response coefficient becomes:

$$\forall i, \quad R_{E_i}^J = \frac{\left( \frac{1}{A_i e_i} - \frac{e_i}{1 - e_i} \sum_{j=1, j \neq i}^n \frac{1}{A_j e_j} \right)}{\sum_{j=1}^n \frac{1}{A_j e_j}} \quad (\text{S18})$$

and the differential equation system describing the evolution of enzyme concentrations is:

$$\forall j, \quad \frac{\partial E_j}{\partial t} = 2\mu \sum_{i=1}^n s_i \alpha_{ij} \delta_i$$

As  $E_{\text{tot}}$  is constant, this system writes for the relative concentrations:

$$\forall j, \quad \frac{\partial e_j}{\partial t} = \frac{\partial(E_j/E_{\text{tot}})}{\partial t} = \frac{1}{E_{\text{tot}}} \frac{\partial E_j}{\partial t}$$

Thus:

$$\forall j, \quad \frac{\partial e_j}{\partial t} = \frac{2\mu}{E_{\text{tot}}} \sum_{i=1}^n s_i \alpha_{ij} \delta_i$$

Given the relation between selection and response coefficients (equation S14), we have:

$$\forall j, \quad \frac{\partial e_j}{\partial t} \approx \frac{2\mu}{E_{\text{tot}}} \sum_{i=1}^n R_{E_i}^J \alpha_{ij} \frac{\delta_i^2}{E_i}$$

150 In addition to the purposeless solution  $\forall i, \delta_i = 0$ , there is an obvious solution for the steady state of relative enzyme concentrations  $\frac{\partial e_i}{\partial t} = 0$ :

$$\boxed{\forall i, \quad R_{E_i}^J = 0} \quad (\text{S19})$$

From equation S18, it comes:

$$\frac{1}{A_i(e_i^*)^2} = \frac{1}{1 - e_i^*} \sum_{\substack{k=1 \\ k \neq i}}^n \frac{1}{A_k e_k^*}$$

Adding the case where  $k = i$  in the sum, we have:

$$\frac{1}{A_i(e_i^*)^2} + \frac{1}{(1 - e_i^*)A_i e_i^*} = \frac{1}{1 - e_i^*} \sum_{k=1}^n \frac{1}{A_k e_k^*}$$

After rearrangements, we get:

$$\frac{1}{A_i(e_i^*)^2} = \sum_{k=1}^n \frac{1}{A_k e_k^*}$$

155 As this equality is valid for any enzyme, we have:

$$\forall i, \forall j \quad \sum_{k=1}^n \frac{1}{A_k e_k^*} = \frac{1}{A_i(e_i^*)^2} = \frac{1}{A_j(e_j^*)^2}$$

Because the pseudo-activities are strictly positive, we get:

$$e_j^* = \frac{A_i^{1/2} e_i^*}{A_j^{1/2}}$$

and because  $\sum_{j=1}^n e_j^* = 1$ , we write:

$$\sum_{j=1}^n e_j^* = \sum_{j=1}^n \frac{A_i^{1/2} e_i^*}{A_j^{1/2}} = A_i^{1/2} e_i^* \sum_{j=1}^n \frac{1}{A_j^{1/2}} = 1$$

This leads to the equilibrium of relative concentrations when there is competition:

$$\boxed{\forall i, \quad e_i^* = \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}}} \quad (\text{S20})$$

With this expression, we can find the maximum flux value at the equilibrium:

$$J_{\text{max}} = E_{\text{tot}}^0 \frac{X}{\sum_{i=1}^n \frac{1}{A_i e_i^*}}$$

Introducing the expression of  $e_i^*$  (equation S20), we get:

$$\begin{aligned} J_{\max} &= E_{\text{tot}}^0 \frac{X}{\sum_{i=1}^n \frac{1}{A_i \left( \frac{A_i^{-1/2}}{\sum_{j=1}^n A_j^{-1/2}} \right)}} \\ &= E_{\text{tot}}^0 \frac{X}{\sum_{j=1}^n A_j^{-1/2} \sum_{i=1}^n A_i^{-1/2}} \end{aligned}$$

160 Hence:

$$\boxed{J_{\max} = E_{\text{tot}}^0 \frac{X}{\left( \sum_{j=1}^n A_j^{-1/2} \right)^2}} \quad (\text{S21})$$

## SI.B.6 Evolutionary equilibria in the case of co-regulation (case 3)

### SI.B.6.1 Theoretical equilibrium

165 **SI.B.6.1.1 Expression of the theoretical equilibrium** When there is co-regulation between all enzymes without competition ( $\forall(i, j) \alpha_{ij} = \beta_{ij} \neq 0$ ), the differential equation system becomes:

$$\frac{\partial E_j}{\partial t} = 2\mu \sum_{i=1}^n s_i \beta_{ij} \delta_i$$

In terms of *relative* concentrations, it writes:

$$\frac{\partial e_j}{\partial t} = \frac{\partial(E_j/E_{\text{tot}})}{\partial t} = \frac{\frac{\partial E_j}{\partial t} E_{\text{tot}} - E_j \frac{\partial E_{\text{tot}}}{\partial t}}{(E_{\text{tot}})^2} \quad (\text{S22})$$

Because

$$\begin{aligned} \frac{\partial E_{\text{tot}}}{\partial t} &= \sum_{j=1}^n \frac{\partial E_j}{\partial t} = \sum_{j=1}^n 2\mu \sum_{i=1}^n s_i \beta_{ij} \delta_i \\ &= 2\mu \sum_{i=1}^n s_i \delta_i \sum_{j=1}^n \beta_{ij} \\ &= 2\mu \sum_{i=1}^n s_i \delta_i B_i \end{aligned}$$

we can write:

$$\begin{aligned}
 \frac{\partial e_j}{\partial t} &= \frac{E_{tot} 2\mu \sum_{i=1}^n s_i \beta_{ij} \delta_i - E_j 2\mu \sum_{i=1}^n s_i \delta_i B_i}{(E_{tot})^2} \\
 &= \frac{2\mu}{E_{tot}} \sum_{i=1}^n s_i \beta_{ij} \delta_i - \frac{2\mu}{E_{tot}} e_j \sum_{i=1}^n s_i \delta_i B_i \\
 &= \frac{2\mu}{E_{tot}} \sum_{i=1}^n s_i \delta_i (\beta_{ij} - e_j B_i)
 \end{aligned}$$

Using the expressions of  $s_i$  (equation S14) and  $\beta_{ij}$  (equation S2), we can write:

$$\frac{\partial e_j}{\partial t} = \frac{2\mu}{E_{tot}} \sum_{i=1}^n R_{E_i}^J \frac{\delta_i^2}{E_i} \left( \frac{B_i}{B_j} - e_j B_i \right)$$

and after rearrangements, we get:

$$\frac{\partial e_j}{\partial t} = \frac{2\mu}{E_{tot}^2} (1/B_j - e_j) \sum_{i=1}^n \frac{R_{E_i}^J \delta_i^2}{e_i} B_i \quad (\text{S23})$$

An obvious solution for  $\frac{\partial e_j}{\partial t} = 0$  for all  $j$  is:

$$\boxed{\forall j, \quad e_j^* = \frac{1}{B_j}} \quad (\text{S24})$$

170 Note that if at least one co-regulation coefficient is negative, leading to at least one negative  $B_i < 0$  (equation S2), the equilibrium concentration is negative, which is not biologically possible.

### SI.B.6.1.2 Stability of the theoretical equilibrium

175 Consider a transition from a theoretical equilibrium  $\mathbf{E}^{t_1}$  to a new state  $\mathbf{E}^{t_2}$  due to a mutation of effect  $\nu$  targeting enzyme  $i$ . The relative concentration of any enzyme  $j \neq i$  after mutation is:

$$e_j^{t_2} = \frac{E_j^{t_2}}{E_{tot}^{t_2}} = \frac{E_j^{t_1} + \beta_{ij}\nu}{E_{tot}^{t_1} + B_i\nu}$$

Since the system is at the theoretical equilibrium at  $t_1$ , we have  $e_j^{t_1} = e_j^* = 1/B_j$ , so:

$$e_j^{t_2} = \frac{\frac{E_{tot}^{t_1}}{B_j} + \beta_{ij}\nu}{E_{tot}^{t_1} + B_i\nu}$$

and from  $\beta_{ij} = B_i/B_j$  (equation S2) we obtain:

$$\begin{aligned}
 e_j^{t_2} &= \frac{\frac{E_{tot}^{t_1}}{B_j} + \frac{B_i}{B_j}\nu}{E_{tot}^{t_1} + B_i\nu} \\
 &= \frac{1}{B_j}
 \end{aligned}$$



Similarly, we have for the target enzyme  $i$ :

$$e_i^{t_2} = \frac{E_i^{t_2}}{E_{tot}^{t_2}} = \frac{E_i^{t_1} + \nu}{E_{tot}^{t_1} + B_i \nu}$$

Because  $E_i = e_i E_{tot}$  and  $e_i^{t_1} = e_i^* = 1/B_i$ , we have:

$$\begin{aligned} e_i^{t_2} &= \frac{\frac{E_{tot}^{t_1}}{B_i} + \nu}{E_{tot}^{t_1} + B_i \nu} \\ &= \frac{1}{B_i} \end{aligned}$$

Thus, no mutation, whatever its effect, can modify the *relative* enzyme concentrations at the theoretical equilibrium.

**SI.B.6.1.3 Response coefficient at the theoretical equilibrium** From the general expression of the flux response coefficient (equation S12), it is possible to find the expression of the response coefficient at the equilibrium when there is co-regulation. Knowing that  $\alpha_{ij} = \beta_{ij}$  in the case of co-regulation, we have for replacing  $e_i$  by  $e_i^* = \frac{1}{B_i}$  and using the relation  $B_i = \beta_{ij} B_j$  (equation S2), we get:

$$R_{E_i}^{J^*} = e_i^* \frac{\sum_{j=1}^n \frac{\beta_{ij}}{A_j (e_j^*)^2}}{\sum_{j=1}^n \frac{1}{A_j e_j^*}} = \frac{1}{B_i} \frac{\sum_{j=1}^n \frac{B_i}{B_j} \frac{1}{A_j (1/B_j)^2}}{\sum_{j=1}^n \frac{1}{A_j / B_j}}$$

After rearrangements, we obtain:

$$\boxed{\forall i, R_{E_i}^J = 1} \quad (\text{S25})$$

Thus, the flux  $J^*$  at the theoretical equilibrium is proportional to each  $E_i$ . Its value is:

$$J^* = E_{tot} \frac{X}{\sum_{i=1}^n \frac{B_i}{A_i}}$$

**SI.B.6.2 The parametric equation of the line relating the relative enzyme concentrations.** Enzyme concentrations are all linearly related, as are the relative enzyme concentrations. So they move on a straight line of the  $n$ -dimensional space where each axis corresponds to the relative concentration of an enzyme.

This line,  $\mathcal{E}$ , has two known points: the initial point  $\mathbf{e}^0$ , named  $\mathbf{O}$ , and the theoretical equilibrium point  $\mathbf{e}^*$ , named  $\mathbf{F}$ . Thus any point  $\mathbf{M}$  of coordinates  $\mathbf{e}$  is on the line  $\mathcal{E}$  if and only if there is a real  $\tau$  such as  $\overrightarrow{OM} = \tau \times \overrightarrow{OF}$ , *i.e.* if:

$$\forall i, \quad e_i - e_i^0 = \tau(e_i^* - e_i^0)$$

We know that  $\forall i, e_i^* = 1/B_i$ , so the parametric equation of the line  $\mathcal{E}$  is:

$$\boxed{\exists \tau \in \mathbb{R} \text{ such as } \forall i, \quad e_i = \tau \left( \frac{1}{B_i} - e_i^0 \right) + e_i^0} \quad (\text{S26})$$

When  $\tau = 0$ ,  $e_i = e_i^0$  and when  $\tau = 1$ ,  $e_i = 1/B_i$ , which is the theoretical equilibrium.

**SI.B.6.3 Effective equilibrium in the case of negative co-regulation** In Supporting information [SI.B.6.1.1](#), we established the differential equation system describing the evolution of relative enzyme concentrations when there is co-regulation (equation [S23](#)):

$$\forall j, \quad \frac{\partial e_j}{\partial t} = \frac{2\mu}{E_{tot}^2} (1/B_j - e_j) \sum_{i=1}^n \frac{R_{E_i}^J \delta_i^2}{e_i} B_i \quad (\text{S27})$$

This equation is cancelled if  $\forall j, e_j = 1/B_j$ , which led to the theoretical equilibrium. Actually there is another trivial solution, which leads to the effective equilibrium:

$$\boxed{\forall i, \quad R_{E_i}^J = 0} \quad (\text{S28})$$

or, from equation [S12](#):

$$\forall i \quad \sum_{j=1}^n \frac{\beta_{ij}}{A_j(\tilde{e}_j)^2} = 0$$

This equality can only be true if at least one co-regulation coefficient is negative. Because  $\forall(i, j), \beta_{ij} = B_i/B_j$  (equation [S2](#)), we get:

$$\forall i \quad B_i \sum_{j=1}^n \frac{1}{B_j A_j (\tilde{e}_j)^2} = 0$$

or:

$$\boxed{\sum_{j=1}^n \frac{1}{B_j A_j (\tilde{e}_j)^2} = 0} \quad (\text{S29})$$

Thanks to the relationship between  $e_j$  and  $\tau$  (equation [S26](#)), searching for  $\tilde{e}_j$  for all  $j$  is equivalent to searching for  $\tilde{\tau}$  using the following equation:

$$\sum_{j=1}^n \frac{1}{B_j A_j (\tilde{\tau}(1/B_j - e_j^0) + e_j^0)^2} = 0 \quad (\text{S30})$$

Solving this equation for  $\tilde{\tau}$  allows us to calculate the relative concentrations  $\tilde{\mathbf{e}}$  at effective equilibrium.

## SI.B.7 Evolutionary equilibrium in the case of competition plus co-regulation (case 4)

### SI.B.7.1 Theoretical equilibrium

**SI.B.7.1.1 Expression of the theoretical equilibrium** When there is co-regulation and competition, we have  $\forall i \alpha_{ii} = 1$  and  $\forall(i, j \neq i) \alpha_{ij} = \frac{\beta_{ij} - B_i e_j}{1 - B_i e_i}$ . As all enzyme are co-regulated, *i.e.*  $\beta_{ij} \neq 0$  for all  $(i, j)$ , we can apply equation [S2](#) to the redistribution coefficient, leading to:

$$\alpha_{ij} = \frac{1/B_j - e_j}{1/B_i - e_i} \quad (\text{S31})$$

Note that this expression is also valid when  $j = i$ :  $\alpha_{ii} = \frac{1/B_i - e_i}{1/B_i - e_i} = 1$ . So the differential equation system becomes:

$$\forall j, \quad \frac{\partial E_j}{\partial t} = 2\mu \sum_{i=1}^n s_i \frac{1/B_j - e_j}{1/B_i - e_i} \delta_i = 2\mu (1/B_j - e_j) \sum_{i=1}^n \frac{1}{1/B_i - e_i} s_i \delta_i$$

In terms of relative enzyme concentration, we have:

$$\frac{\partial e_j}{\partial t} = \frac{\partial(E_j/E_{\text{tot}})}{\partial t}$$

or, because  $E_{\text{tot}}$  is constant:

$$\frac{\partial e_j}{\partial t} = \frac{1}{E_{\text{tot}}} \frac{\partial E_j}{\partial t}$$

Hence, the differential equation system for the relative enzyme concentrations is:

$$\forall j, \quad \frac{\partial e_j}{\partial t} = 2\mu \frac{(1/B_j - e_j)}{E_{\text{tot}}} \sum_{i=1}^n \frac{1}{1/B_i - e_i} s_i \delta_i \quad (\text{S32})$$

An obvious solution for the evolutionary equilibrium  $\forall j \frac{\partial e_j}{\partial t} = 0$  is:

$$\boxed{\forall j, \quad e_j^* = \frac{1}{B_j}} \quad (\text{S33})$$

220 At this point, the flux value is:

$$\boxed{J^* = E_{\text{tot}}^0 \frac{X}{\sum_{j=1}^n \frac{B_j}{A_j}}} \quad (\text{S34})$$

**SI.B.7.1.2 Stability of the theoretical equilibrium.** Consider a system at theoretical equilibrium  $\mathbf{E}^{t_1}$  passing to another state  $\mathbf{E}^{t_2}$  by a mutation of effect  $\nu$  affecting enzyme  $i$ . So concentration of any enzyme  $j$  at state after mutation can be written:

$$E_j^{t_2} = E_j^{t_1} + \frac{1 - \beta_{ij} e_j}{1 - B_i e_i} \delta_i$$

And for the targeted enzyme  $i$ :

$$E_i^{t_2} = E_i^{t_1} + \delta_i$$

225 where the actual effect  $\delta_i$  of the mutation is (equation S8):

$$\delta_i = \frac{1 - B_i e_i^{t_1}}{\frac{1}{\nu} + \frac{1}{E_{\text{tot}}}}$$

But because at time  $t_1$ , the system is at theoretical equilibrium, we have  $e_i^{t_1} = \frac{1}{B_i}$ , and therefore:

$$\delta_i = 0$$

So we obtain  $E_j^{t_2} = E_j^{t_1}$  and  $E_i^{t_2} = E_i^{t_1}$  whatever the canonical effect  $\nu$  of the mutation. The system is blocked at theoretical equilibrium.

230 **SI.B.7.2 Effective equilibrium** In Supporting Information [SI.B.7.1.1](#), we established the differential equation system when there is competition plus co-regulation (equation [S32](#)).

Using relationship between response and selection coefficients (equation [15](#)), equation [S32](#) becomes:

$$\forall j, \quad \frac{\partial e_j}{\partial t} = 2\mu \frac{(1/B_j - e_j)}{E_{\text{tot}}} \sum_{i=1}^n \frac{1}{1/B_i - e_i} R_{E_i}^J \frac{\delta_i^2}{E_i} \quad (\text{S35})$$

235 In addition to  $e_j = 1/B_j$ , which is the theoretical equilibrium, the system admits another obvious solution:

$$\boxed{\forall i, \quad R_{E_i}^J = 0} \quad (\text{S36})$$

Introducing the expression of  $\alpha_{ij}$  (equation [S31](#)) in the equation of the flux response coefficient (equation [S12](#)) leads to:

$$R_{E_i}^J = \frac{e_i}{1/B_i - e_i} \frac{\sum_{j=1}^n \frac{1/B_j - e_j}{A_j (e_j)^2}}{\sum_{j=1}^n \frac{1}{A_j e_j}}$$

So canceling equation [S36](#) amounts to:

$$\sum_{j=1}^n \frac{\frac{1}{B_j} - \tilde{e}_j}{A_j (\tilde{e}_j)^2} = 0$$

Thanks to the relationship between  $e_j$  and  $\tau$  (equation [28](#)), we know that searching for the  $\tilde{e}_j$ 's that cancel this equation is equivalent to searching for a particular  $\tilde{\tau}$  value. So we write:

$$\begin{aligned} \sum_{j=1}^n \frac{\frac{1}{B_j} - \tilde{e}_j}{A_j (\tilde{e}_j)^2} &= \sum_{j=1}^n \frac{1/B_j - (\tau(1/B_j - e_j^0) + e_j^0)}{A_j (\tilde{\tau}(1/B_j - e_j^0) + e_j^0)^2} \\ &= (1 - \tau) \sum_{j=1}^n \frac{(1/B_j - e_j^0)}{A_j (\tilde{\tau}(1/B_j - e_j^0) + e_j^0)^2} \end{aligned}$$

240 In addition to the trivial solution  $\tau = 1$ , which corresponds to the theoretical equilibrium, the other solution is the  $\tau$  value that cancels this equation:

$$\boxed{\sum_{j=1}^n \frac{1/B_j - e_j^0}{A_j (\tilde{\tau}(1/B_j - e_j^0) + e_j^0)^2} = 0} \quad (\text{S37})$$

Once  $\tilde{\tau}$  is found, it is easy to find the relative concentrations  $\tilde{e}$  at effective equilibrium.

## SI.B.8 Expressing absolute enzyme concentrations in function of the driving variable

245 **SI.B.8.1 In the case of co-regulation** In the parametric equation [S26](#), we can make visible the *absolute* enzyme concentrations, using the equation  $e_i = E_i/E_{\text{tot}}$ :

$$\forall i, \quad \frac{E_i}{E_{\text{tot}}} = \tau \left( \frac{1}{B_i} - \frac{E_i^0}{E_{\text{tot}}^0} \right) + \frac{E_i^0}{E_{\text{tot}}^0} \quad (\text{S38})$$

As there is not competition,  $E_{\text{tot}}$  varies with  $\tau$ . The relation between  $E_{\text{tot}}$  and  $\tau$  can be derived from the relation  $B_i = \frac{\Delta E_{\text{tot}}}{\Delta E_i}$  (equation 6). We have:

$$\begin{aligned} E_{\text{tot}} &= E_{\text{tot}}^0 + B_i(E_i - E_i^0) \\ &= E_{\text{tot}} B_i e_i + E_{\text{tot}}^0 (1 - B_i e_i^0) \\ &= E_{\text{tot}}^0 \frac{1 - B_i e_i^0}{1 - B_i e_i} \end{aligned}$$

Replacing  $e_i$  by its expression (equation S26), we get:

$$E_{\text{tot}} = E_{\text{tot}}^0 \frac{1 - B_i e_i^0}{1 - B_i(\tau(1/B_i - e_i^0) + e_i^0)}$$

which is simplified to (for  $\tau \neq 1$ ):

$$E_{\text{tot}} = \frac{E_{\text{tot}}^0}{1 - \tau} \quad (\text{S39})$$

Introducing this expression of  $E_{\text{tot}}$  in equation S38, we get:

$$\forall i, \frac{E_i(1 - \tau)}{E_{\text{tot}}^0} = \tau \left( \frac{1}{B_i} - \frac{E_i^0}{E_{\text{tot}}^0} \right) + \frac{E_i^0}{E_{\text{tot}}^0} \quad (\text{S40})$$

250 Thus:

$$E_i = E_i^0 + \frac{\tau E_{\text{tot}}^0}{(1 - \tau) B_i} \quad (\text{S41})$$

Or:

$$\boxed{E_i = E_i^0 + \frac{E_{\text{tot}}^0}{B_i} \left( -1 + \frac{1}{1 - \tau} \right)} \quad (\text{S42})$$

**SI.B.8.2 In the case of co-regulation and competition** When there is both competition and co-regulation, the difference with the previous case is that  $E_{\text{tot}}$  is constant. So, replacing  $e_i$  by  $E_i/E_{\text{tot}}$  in equation S26, we get directly:

$$\boxed{E_i = \tau \left( \frac{E_{\text{tot}}}{B_i} - E_i^0 \right) + E_i^0} \quad (\text{S43})$$

## 255 SI.B.9 Evolution model of the driving variable

We assume here that the mutations target a factor controlling the concentrations of a group of enzymes. Formally, this corresponds to mutations affecting the driving variable  $\tau$ :

$$\tau^m = \tau^r + \nu_\tau$$

where  $\nu_\tau$  is the effect of the mutation on  $\tau$ .

**SI.B.9.1 Redistribution coefficient and mutation effect** If the mutation affects  $\tau$ , the questions arise of the actual effect  $\delta_i$  of this mutation on a particular enzyme  $i$  and of the fate of the redistribution coefficients  $\alpha_{ij}$ . We examine these questions in the two possible situations.

**SI.B.9.1.1 In the case of co-regulation** For all enzyme  $i$ , we have from equation S42:

$$E_i^r = E_i^0 + \frac{E_{\text{tot}}^0}{B_i} \left( -1 + \frac{1}{1 - \tau^r} \right)$$

$$\text{and } E_i^m = E_i^0 + \frac{E_{\text{tot}}^0}{B_i} \left( -1 + \frac{1}{1 - \tau^r - \nu_\tau} \right)$$

leading to the actual effect of the mutation on enzyme  $i$ :

$$\delta_i = E_i^m - E_i^r = \frac{E_{\text{tot}}^0}{B_i} \left( \frac{1}{1 - \tau^r - \nu_\tau} - \frac{1}{1 - \tau^r} \right) \quad (\text{S44})$$

So  $\delta_i$  depends both on  $\nu_\tau$  and  $B_i$ .

265 For two enzymes  $i$  and  $j$ , we have from the previous equation:

$$(E_i^m - E_i^r) \frac{B_i}{E_{\text{tot}}^0} = (E_j^m - E_j^r) \frac{B_j}{E_{\text{tot}}^0} = \left( \frac{1}{1 - \tau^r - \nu_\tau} - \frac{1}{1 - \tau^r} \right)$$

Therefore, we obtain:

$$E_j^m = E_j^r + \frac{B_i}{B_j} (E_i^m - E_i^r)$$

and because  $B_i/B_j = \beta_{ij}$ , we have:

$$E_j^m = E_j^r + \beta_{ij} (E_i^m - E_i^r) \quad (\text{S45})$$

Thus, the equation  $\Delta E_j = \beta_{ij} \Delta E_i$  between two co-regulated enzymes is valid, whether enzyme  $i$  or the driving variable is the target of the mutation.

**SI.B.9.1.2 In the case of competition and co-regulation** For all enzyme  $i$ , we have from equation S43:

$$E_i^r = E_i^0 + \tau^r \left( \frac{E_{\text{tot}}^0}{B_i} - E_i^0 \right)$$

$$\text{and } E_i^m = E_i^0 + (\tau^r + \nu_\tau) \left( \frac{E_{\text{tot}}^0}{B_i} - E_i^0 \right)$$

270 leading to the actual effect of the mutation on enzyme  $i$ :

$$\delta_i = E_i^m - E_i^r = \nu_\tau \left( \frac{E_{\text{tot}}^0}{B_i} - E_i^0 \right) \quad (\text{S46})$$

Again  $\delta_i$  depends both on  $\nu_\tau$  and  $B_i$ .

For two enzymes  $i$  and  $j$ , we have from equation S46:

$$\frac{E_i^m - E_i^r}{\frac{E_{\text{tot}}^0}{B_i} - E_i^0} = \frac{E_j^m - E_j^r}{\frac{E_{\text{tot}}^0}{B_j} - E_j^0} = \nu_\tau$$

Therefore, we obtain:

$$E_j^m = E_j^r + \frac{E_{\text{tot}}^0/B_j - E_j^0}{E_{\text{tot}}^0/B_i - E_i^0} (E_i^m - E_i^r)$$

and because  $\frac{E_{\text{tot}}^0/B_j - E_j^0}{E_{\text{tot}}^0/B_i - E_i^0} = \alpha_{ij}$  for two co-regulated enzymes (Supporting Information 275 SI.B.1.4), we have:

$$E_j^m = E_j^r + \alpha_{ij} (E_i^m - E_i^r) \quad (\text{S47})$$

Thus, the equation  $\Delta E_j = \alpha_{ij} \Delta E_i$  between two co-regulated enzymes is valid, whether enzyme  $i$  or the driving variable is the target of the mutation.

**SI.B.9.2 Evolution model of the driving variable and evolutionary equilibria**

With the same assumptions as for the evolution model of enzyme concentrations, the variation of the driving variable  $\tau$  can be represented as a stochastic process. At each time  $t$ , its rate of variation can be described with the following complete system of events:

$$\left(\frac{\partial\tau}{\partial t}\right)^{(t)} = \begin{cases} \text{Probability} & \text{Mutation effect} & \text{Event} \\ \mu_\tau 2s_\tau^{(t)} & \nu_\tau^{(t)} & \text{Fixation of mutation targeting } \tau \\ 1 - \mu_\tau 2s_\tau^{(t)} & 0 & \text{No mutation or no fixation} \end{cases} \quad (\text{S48})$$

where  $\mu_\tau$  is the mutation rate and  $s_\tau$  the coefficient of selection ( $s_\tau = \frac{J_\tau^m - J^r}{J^r}$ ).

At each time  $t$  the *average* variation of the driving variable  $\tau$  is:

$$\frac{\partial\tau}{\partial t} = \mu_\tau 2s_\tau \nu_\tau \quad (\text{S49})$$

Assuming that mutation effects  $\nu_\tau$  are small compared to  $\tau$ , we have (Supporting Information [SI.B.3.1](#)):

$$s_\tau = \left(\frac{\Delta J}{J}\right)_\tau = \frac{\Delta J}{\Delta\tau} \frac{\Delta\tau}{J} \approx \frac{\partial J}{\partial\tau} \frac{\nu_\tau}{J}$$

So equation [S49](#) can be rewritten as:

$$\frac{\partial\tau}{\partial t} \approx 2\mu_\tau \frac{\partial J}{\partial\tau} \frac{\nu_\tau^2}{J} \quad (\text{S50})$$

To find the evolutionary equilibrium of the driving variable  $\tau$ , we have to search for  $\tau$  such as  $\frac{\partial\tau}{\partial t} = 0$ , which is equivalent, from the previous equation, to find a solution for  $\frac{\partial J}{\partial\tau} = 0$ . We can decompose this equation:

$$\begin{aligned} \frac{\partial J}{\partial\tau} &= \sum_{i=1}^n \frac{\partial J}{\partial E_i} \frac{\partial E_i}{\partial\tau} \\ &= \sum_{i=1}^n R_{E_i}^J \frac{J}{E_i} \frac{\partial E_i}{\partial\tau} \end{aligned}$$

So canceling this equation amounts to canceling the flux response coefficients  $R_{E_i}^J$  for all  $i$ , which is the definition of the effective equilibrium of relative concentrations. From that, we can find  $\tilde{\tau}$  in the cases of negative co-regulation (Supporting Information [SI.B.6.3](#)) and competition plus co-regulation (Supporting Information [SI.B.7.2](#)).

Regarding the theoretical equilibrium in the case of co-regulation plus competition, cancelling the derivative of equation [S43](#)

$$\frac{\partial E_i}{\partial\tau} = \frac{E_{\text{tot}}}{B_i} - E_i$$

leads to  $e_i^* = 1/B_i$ , which is the theoretical equilibrium. In the case of positive co-regulation alone,  $J$  increases continuously. As a consequence,  $\frac{\delta\tau}{\delta t}$  tends towards 0 (see equation [S50](#)), which corresponds to  $\tau = 1$ , the theoretical equilibrium (Supporting information [SI.B.6.2](#)).

To conclude, in the case of co-regulation, the evolution model is modified if the mutations target the driving variable  $\tau$  rather than concentrations of individual enzymes, but the evolutionary equilibria (theoretical and effective) do not change.

## A.2 Démonstrations complémentaires

### a) Co-regulation : Straight line between relative concentrations

**Step 1.** *Link between enzymes whatever mutations.* Consider a first fixed mutation targeting  $E_i$  at  $t_1$  and a second one targeting  $E_j$  at  $t_2$ . So we can write :

Initial state	First mutation	Second mutation
$E_i^0$	$E_i^{t_1} = E_i^0 + \nu^{t_1}$	$E_i^{t_2} = E_i^{t_1} + \beta_{ji}\nu^{t_2}$
$E_j^0$	$E_j^{t_1} = E_j^0 + \beta_{ij}\nu^{t_1}$	$E_j^{t_2} = E_j^{t_1} + \nu^{t_2}$
$E_k^0$	$E_k^{t_1} = E_k^0 + \beta_{ik}\nu^{t_1}$	$E_k^{t_2} = E_k^{t_1} + \beta_{jk}\nu^{t_2}$
$E_{\text{tot}}^0$	$E_{\text{tot}}^{t_1} = E_{\text{tot}}^0 + B_i\nu^{t_1}$	$E_{\text{tot}}^{t_2} = E_{\text{tot}}^{t_1} + B_j\nu^{t_2}$

where  $\nu^{t_1}$  is the size of the first mutation, and  $\nu^{t_2}$  size of the second mutation. Note that effect of mutation on mutated enzyme is equal to its size in case of regulation.

First, we will express concentration of an enzyme depending on any enzyme concentration whatever is the target of mutation. For example, we will express  $E_k$  depending on  $E_i$ . Combining third line, we have :

$$E_k^{t_2} = E_k^0 + \beta_{ik}\nu^{t_1} + \beta_{jk}\nu^{t_2}$$

By expressing size of mutation depending on  $E_i$ , we have :

$$E_k^{t_2} = E_k^0 + \beta_{ik}(E_i^{t_1} - E_i^0) + \beta_{jk} \frac{(E_i^{t_2} - E_i^{t_1})}{\beta_{ji}}$$

Thanks to co-regulation coefficients properties, *i.e.*  $\beta_{ij} = \beta_{ik}\beta_{kj}$  and  $\beta_{ij} = 1/\beta_{ji}$ , we have :

$$E_k^{t_2} = E_k^0 + \beta_{ik}(E_i^{t_1} - E_i^0) + \beta_{ik}(E_i^{t_2} - E_i^{t_1})$$

Then :

$$E_k^{t_2} = E_k^0 + \beta_{ik}(E_i^{t_2} - E_i^0)$$

By extension and with a simple demonstration by iteration, we can write concentration of any enzyme depending on initial state and any other enzyme at any time, whatever number, size or target of successive mutations :

$$E_k = E_k^0 + \beta_{ik}(E_i - E_i^0) \tag{A.1}$$

Second, we would express  $E_{\text{tot}}^{t_2}$  depending on its initial state  $E_{\text{tot}}^0$  and any concentration  $E_k$ . Combining last line, we have :

$$E_{\text{tot}}^{t_2} = E_{\text{tot}}^0 + B_i\nu^{t_1} + B_j\nu^{t_2}$$



By expressing size of mutation depending on  $E_k$ , we obtain :

$$E_{\text{tot}}^{t_2} = E_{\text{tot}}^0 + B_i \frac{(E_k^{t_1} - E_k^0)}{\beta_{ik}} + B_j \frac{(E_k^{t_2} - E_k^{t_1})}{\beta_{kj}}$$

Thanks to global co-regulation properties, textiti.e.  $B_i/\beta_{ik} = B_k$ , we have :

$$E_{\text{tot}}^{t_2} = E_{\text{tot}}^0 + B_k(E_k^{t_1} - E_k^0) + B_k(E_k^{t_2} - E_k^{t_1})$$

Then :

$$E_{\text{tot}}^{t_2} = E_{\text{tot}}^0 + B_k(E_k^{t_2} - E_k^0)$$

By extension and with a simple demonstration by iteration, we can write total concentration depending on initial state and any enzyme at any time, whatever number, size or target of successive mutations :

$$E_{\text{tot}} = E_{\text{tot}}^0 + B_k(E_k - E_k^0) \quad (\text{A.2})$$

**Step 2. Linearity between enzymes.** Because  $\beta$ 's and  $B$ 's are constant, all enzymes concentrations are linearly connected, and therefore, with total concentration. Thus relative enzyme concentrations are also linearly connected. So it exists a straight line linking relative concentrations in hyperspace of dimension  $n$ , where each axis correspond to relative concentration of one enzyme.

This line, which we call line  $\mathcal{E}$ , has two known points : initial point  $e^0$ , named  $\mathbf{O}$ , and theoretical equilibrium point  $e^*$ , named  $\mathbf{F}$ . Thus any point  $\mathbf{M}$  of coordinates  $e$  is on line  $\mathbf{E}$  if, and only if, it exists a real  $\tau$  such as  $\overrightarrow{OM} = \tau \times \overrightarrow{OF}$ , i.e. :

$$\forall i, \quad e_i - e_i^0 = \tau(e_i^* - e_i^0)$$

It leads to the parametric equation of  $\mathcal{E}$ , which is :

$$\exists \tau \in \mathbb{R} \text{ tel que } \forall i, \quad e_i = \tau(e_i^* - e_i^0) + e_i^0 \quad (\text{A.3})$$

Because  $e_i^* = 1/B_i$  for all  $i$ , we obtain this equation for  $\mathcal{E}$  :

$$\exists \tau \in \mathbb{R} \text{ tel que } \forall i, \quad e_i = \tau\left(\frac{1}{B_i} - e_i^0\right) + e_i^0 \quad (\text{A.4})$$

## b) Competition plus co-regulation : Straight line between concentrations

If all enzymes are co-regulated, we have  $B_i = \beta_{ij}B_j$ , and :

$$\forall (i, j) \quad \alpha_{ij} = \frac{1/B_j - e_j}{1/B_i - e_i}$$

Note that this relation is also valid for  $\alpha_{ii} = 1$ , and because  $\beta$ 's are constant, it gives the same properties to  $\alpha$ 's than  $\beta$ 's, *i.e.*  $\alpha_{ij} = \alpha_{ik}\alpha_{kj}$  and  $\alpha_{ij} = 1/\alpha_{ji}$ .

**Step 1. Link between enzymes whatever mutations.** Consider a first fixed mutation affecting  $E_i$  and a second one affecting  $E_j$ . If we call  $\mathbf{E}^0$  the initial state at time  $t_0$ ,  $\mathbf{E}^{t_1}$  state after mutation of  $E_i$  at time  $t_1$  and  $\mathbf{E}^{t_2}$  state after the two mutations at time  $t_2$ , we can write :

Initial state	First mutation	Second mutation
$E_i^0$	$E_i^{t_1} = E_i^0 + \delta_i^{t_1}$	$E_i^{t_2} = E_i^{t_1} + \alpha_{ji}^{t_1} \delta_j^{t_2}$
$E_j^0$	$E_j^{t_1} = E_j^0 + \alpha_{ij}^0 \delta_i^{t_1}$	$E_j^{t_2} = E_j^{t_1} + \delta_j^{t_2}$
$E_k^0$	$E_k^{t_1} = E_k^0 + \alpha_{ik}^0 \delta_i^{t_1}$	$E_k^{t_2} = E_k^{t_1} + \alpha_{jk}^{t_1} \delta_j^{t_2}$
$E_{\text{tot}}^0$	$E_{\text{tot}}^{t_1} = E_{\text{tot}}^0$	$E_{\text{tot}}^{t_2} = E_{\text{tot}}^0$

where  $\delta^{t_1}$  is the actual effect of the first mutation targeting  $i$ , and  $\delta^{t_2}$  actual effect of the second mutation targeting  $j$ .

First, we will express concentration of an enzyme depending on any enzyme concentration whatever is the target of mutation. For example, we will express  $E_k$  depending on  $E_i$ . Combining third line, we have :

$$E_k^{t_2} = E_k^0 + \alpha_{ik}^0 \delta_i^{t_1} + \alpha_{jk}^{t_1} \delta_j^{t_2}$$

By expressing effect of mutation depending on  $E_i$ , we have :

$$E_k^{t_2} = E_k^0 + \alpha_{ik}^0 (E_i^{t_1} - E_i^0) + \alpha_{jk}^{t_1} \frac{(E_i^{t_2} - E_i^{t_1})}{\alpha_{ji}^{t_1}}$$

We assumed that all enzymes are co-regulated between them, so  $\alpha_{ij}^{(t)} = \frac{1/B_j - e_j^{(t)}}{1/B_i - e_i^{(t)}}$  at any time  $t$ . Thanks to constant co-regulation coefficients, giving transitivity property to  $\alpha$ 's *i.e.*  $\alpha_{ij} = \alpha_{ik}\alpha_{kj}$  (see above), we have :

$$E_k^{t_2} = E_k^0 + \alpha_{ik}^0 (E_i^{t_1} - E_i^0) + \alpha_{ik}^{t_1} (E_i^{t_2} - E_i^{t_1})$$

If we suppose that  $\alpha$ 's are constant during time (*see demonstration in next step*), we obtain :

$$E_k^{t_2} = E_k^0 + \alpha_{ik} (E_i^{t_2} - E_i^0)$$

By extension and with a simple demonstration by iteration, we can write concentration of any enzyme depending on initial state and any other enzyme at any time, whatever number, size or target of successive mutations :

$$E_k = E_k^0 + \alpha_{ik} (E_i - E_i^0) \tag{A.5}$$

**Step 2. Constant redistribution coefficient during time.** Consider the same system as previous step, affected by same mutations.

We supposed that all enzyme enzymes are co-regulated, so we have at time  $t$  :

$$\alpha_{ij}^{(t)} = \frac{1/B_j - e_j^{(t)}}{1/B_i - e_i^{(t)}} = \frac{E_{\text{tot}}/B_j - E_j^{(t)}}{E_{\text{tot}}/B_i - E_i^{(t)}}$$

Thus we can write for two enzymes  $i$  and  $k$  :

$$\begin{aligned} \alpha_{ik}^{t_1} &= \frac{E_{\text{tot}}/B_k - E_k^{t_1}}{E_{\text{tot}}/B_i - E_i^{t_1}} \\ &= \frac{E_{\text{tot}}/B_k - E_k^0 - \alpha_{ik}^0(E_i^{t_1} - E_i^0)}{E_{\text{tot}}/B_i - E_i^{t_1}} \\ &= \frac{E_{\text{tot}}/B_k - E_k^0 - \frac{E_{\text{tot}}/B_k - E_k^0}{E_{\text{tot}}/B_i - E_i^0}(E_i^{t_1} - E_i^0)}{E_{\text{tot}}/B_i - E_i^{t_1}} \\ &= \frac{(E_{\text{tot}}/B_k - E_k^0)(E_{\text{tot}}/B_i - E_i^0 - (E_i^{t_1} - E_i^0))}{(E_{\text{tot}}/B_i - E_i^{t_1})(E_{\text{tot}}/B_i - E_i^0)} \\ &= \frac{E_{\text{tot}}/B_k - E_k^0}{E_{\text{tot}}/B_i - E_i^0} \\ &= \alpha_{ik}^0 \end{aligned}$$

By extension and with a simple demonstration by iteration, we can say that  $\alpha$ 's are constant during time whatever mutations.

**Step 3. Linearity between enzymes.** Because  $\alpha$ 's are constant, all enzymes concentrations are linearly connected. Thus relative enzyme concentrations are also linearly connected due to constant  $E_{\text{tot}}$ . So it exists a straight line linking relative concentrations in hyperspace of dimension  $n$ , where each axis correspond to relative concentration of one enzyme.

This line, which we call line  $\mathcal{E}$ , has two known points : initial point  $e^0$ , named **O**, and theoretical equilibrium point  $e^*$ , named **F**. Thus any point **M** of coordinates  $e$  is on line **E** if, and only if, it exists a real  $\tau$  such as  $\overrightarrow{OM} = \tau \times \overrightarrow{OF}$ , i.e. :

$$\forall i, \quad e_i - e_i^0 = \tau(e_i^* - e_i^0)$$

It leads to the parametric equation of  $\mathcal{E}$ , which is :

$$\exists \tau \in \mathbb{R} \text{ tel que } \forall i, \quad e_i = \tau(e_i^* - e_i^0) + e_i^0 \quad (\text{A.6})$$

Because  $e_i^* = 1/B_i$  for all  $i$ , we obtain this equation for  $\mathcal{E}$  :

$$\exists \tau \in \mathbb{R} \text{ tel que } \forall i, \quad e_i = \tau\left(\frac{1}{B_i} - e_i^0\right) + e_i^0 \quad (\text{A.7})$$

In a same way, we can write the parametric equation for the line linking absolute concentra-

tions :

$$\exists \tau \in \mathbb{R} \text{ tel que } \forall i, E_i = \tau \left( \frac{E_{\text{tot}}}{B_i} - E_i^0 \right) + E_i^0 \quad (\text{A.8})$$

which parameter  $\tau$  is the same as for line  $\mathcal{E}$ .



## **Annexe B**

### **Article 2 : figures supplémentaires et démonstrations**

Cette annexe contient les *Supporting Information* de l'article compagnon (chapitre 2), avec les figures supplémentaires et les démonstrations utiles. Je n'ai pas jugé nécessaire de mettre la démonstration rigoureuse de l'existence de la linéarité au sein des groupes de régulation, car elle est identique à celles présentées en annexe [A.2](#), en prenant en compte les enzymes au sein de groupes à la place de l'ensemble des enzymes.



## B.1

# Evolution of enzyme levels in metabolic pathways: A theoretical approach. Part 2

## Supporting Information

Charlotte Coton, Christine Dillmann, Dominique de Vienne

September 30, 2021

5

## Contents

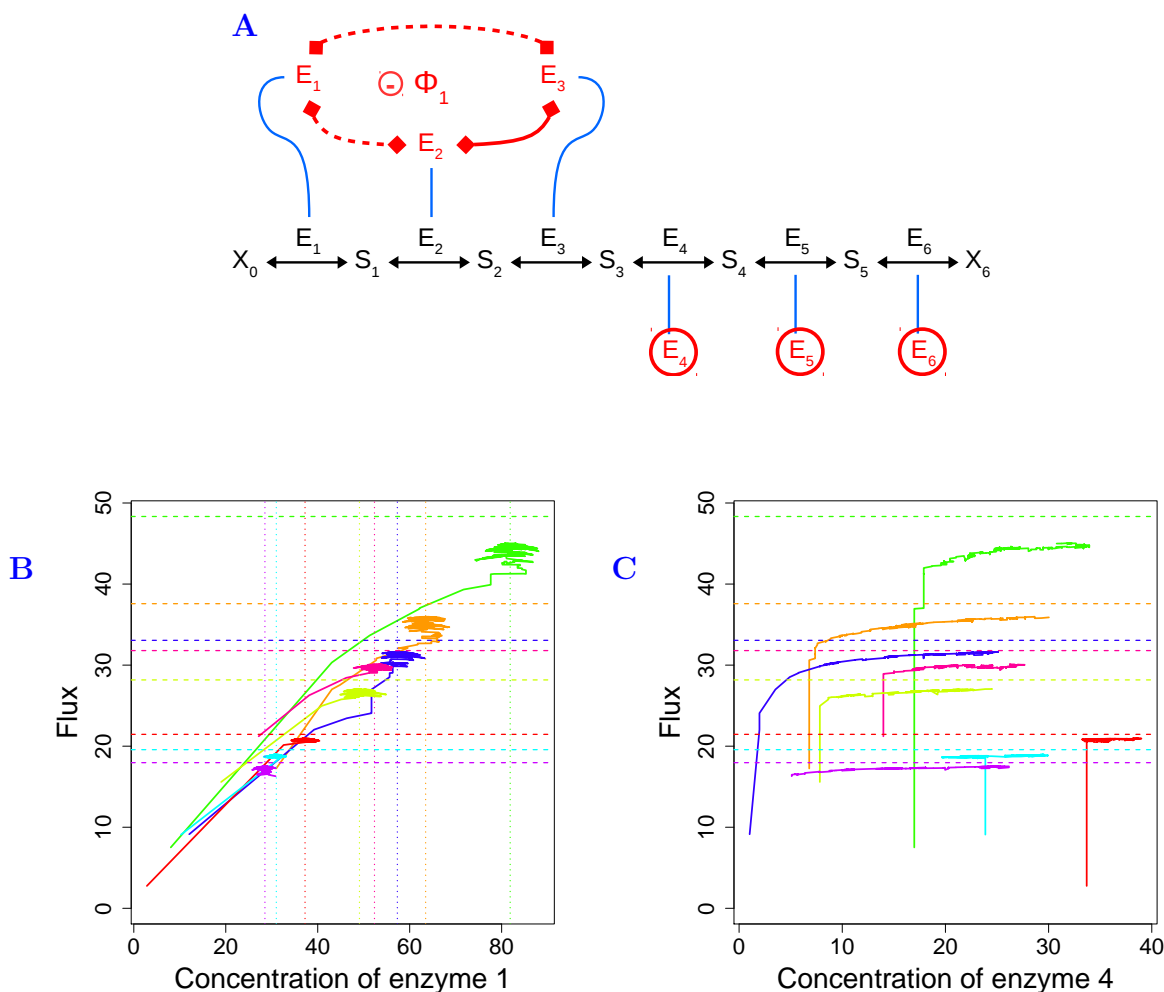
	<b>I Supplementary Figures</b>	<b>2</b>
	<b>II Mathematical Proofs</b>	<b>7</b>
	II.1 Properties of co-regulation groups . . . . .	7
10	II.2 Searching for evolutionary equilibrium . . . . .	7
	II.3 Equilibria when there is no competition . . . . .	7
	II.3.1 Equilibrium within co-regulation groups . . . . .	8
	II.3.2 Equilibrium between co-regulation groups . . . . .	10
	II.3.3 Relationship between absolute concentrations and driving variable . . . . .	13
15	II.4 Equilibria when there is competition . . . . .	14
	II.4.1 Effective equilibria . . . . .	15
	II.4.2 Intra-group theoretical equilibria . . . . .	19
	II.5 Range of neutral variation without competition . . . . .	20
	II.5.1 When there are only positive groups and singletons . . . . .	20
20	II.5.2 When there are only negative groups . . . . .	20
	II.5.3 When there are positive (and/or singletons) and negative groups . . . . .	20
	II.5.4 RNV order . . . . .	21
	<b>III Complementary analysis</b>	<b>23</b>
	III.1 Degrees of freedom depend on the applied constraint . . . . .	23

## 25 Literature cited

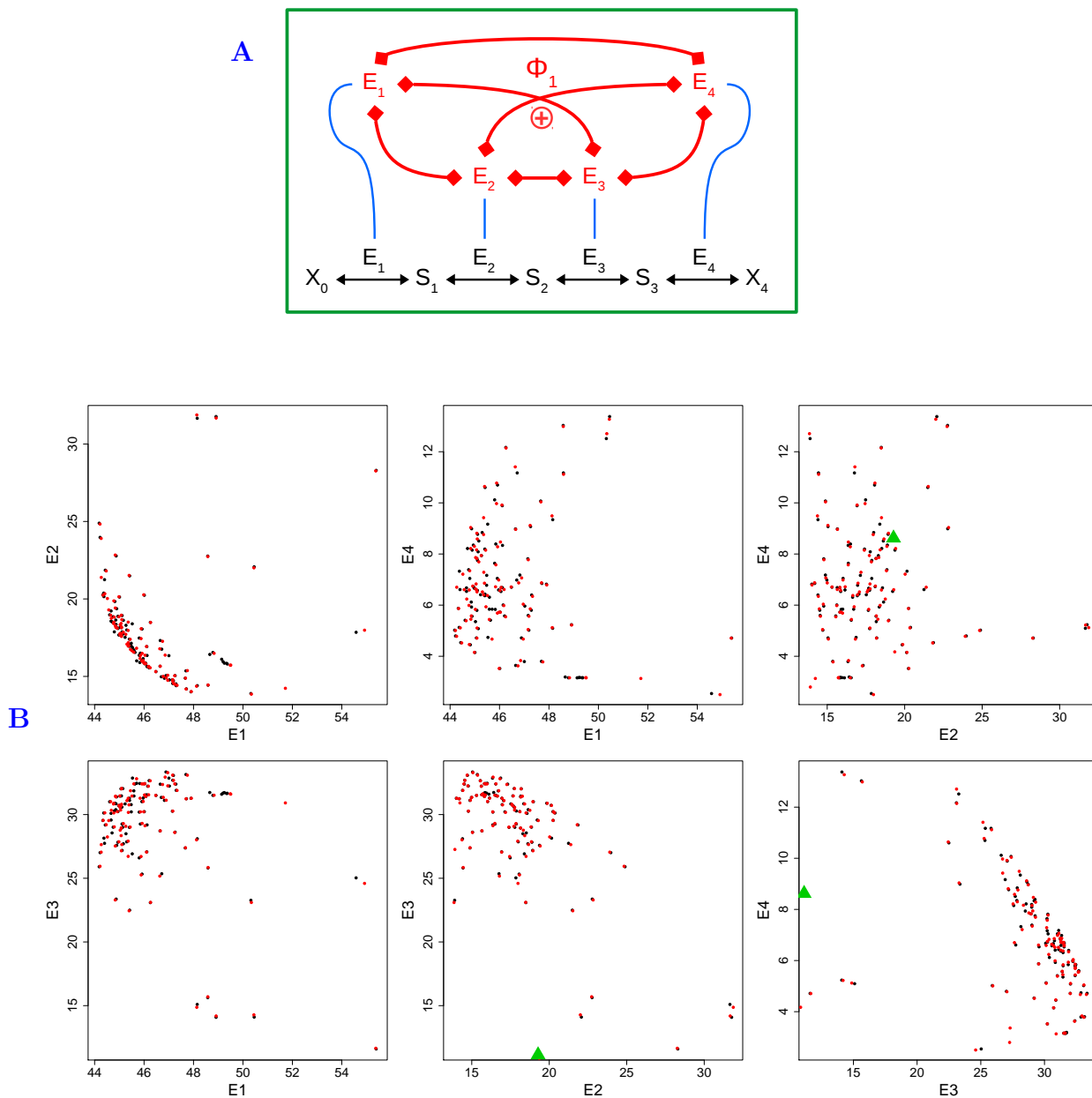
Coton, C., G. Talbot, M. L. Louarn, C. Dillmann, and D. de Vienne, 2021. Evolution of enzyme levels in metabolic pathways: A theoretical approach. bioRxiv P. 2021.05.04.442631.



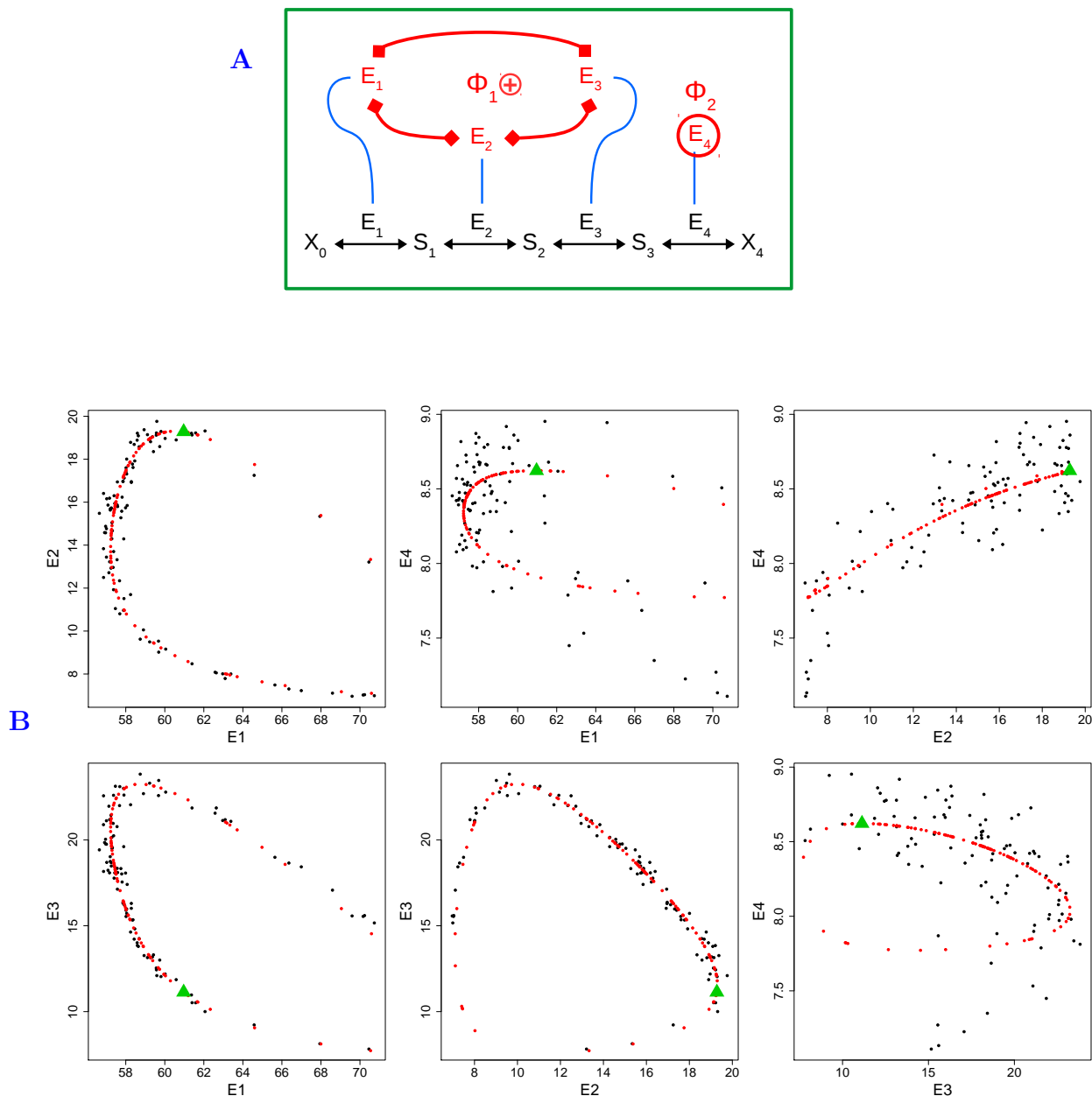
## I Supplementary Figures



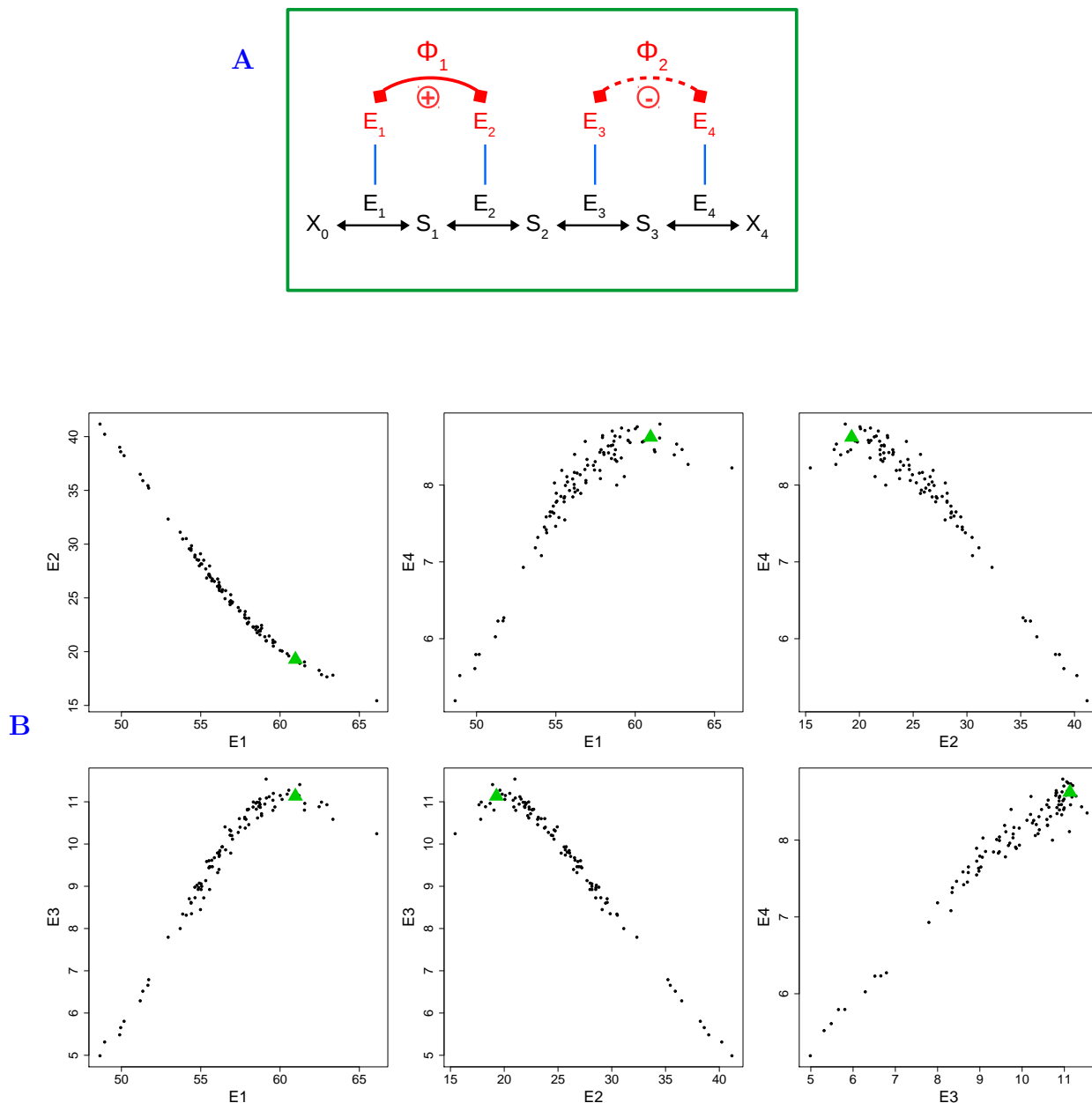
**Figure S1. Simulations of enzyme evolution in a six-enzyme pathway when there is a negative co-regulation group and three singletons, without competition.** Simulations run for 125 000 time steps. (A) The six-enzyme pathway used in the simulations. The first three enzymes  $E_1$ ,  $E_2$  and  $E_3$  form a negative group and the other enzymes are independent (same symbols as in figure 1). Parameter values:  $\mathbf{A} = (1, 10, 30, 50, 100, 200)$ ,  $X = 1$ ,  $N = 200$ ,  $\beta_{12} = -0.32$ ,  $\beta_{23} = 0.43$ . (B) Relationship between flux  $J$  and concentration of enzyme 1. Eight simulations are represented, each with specific initial concentrations (one color per simulation). Vertical dotted lines: abscissa of the effective equilibria of absolute concentrations  $\tilde{E}_i$ . Horizontal dashed lines: maximal flux  $\tilde{J}_{\max}$ . The concentrations of the negatively co-regulated enzymes first increase, then stabilize around their effective equilibrium, which prevents the flux from increasing beyond a local maximum. (C) Same representation as in (B) for independent enzyme 4. Even if there is no theoretical limit, the concentration of the independent enzyme tends towards a plateau because there is no more selective pressure when  $\tilde{J}_{\max}$  is reached.



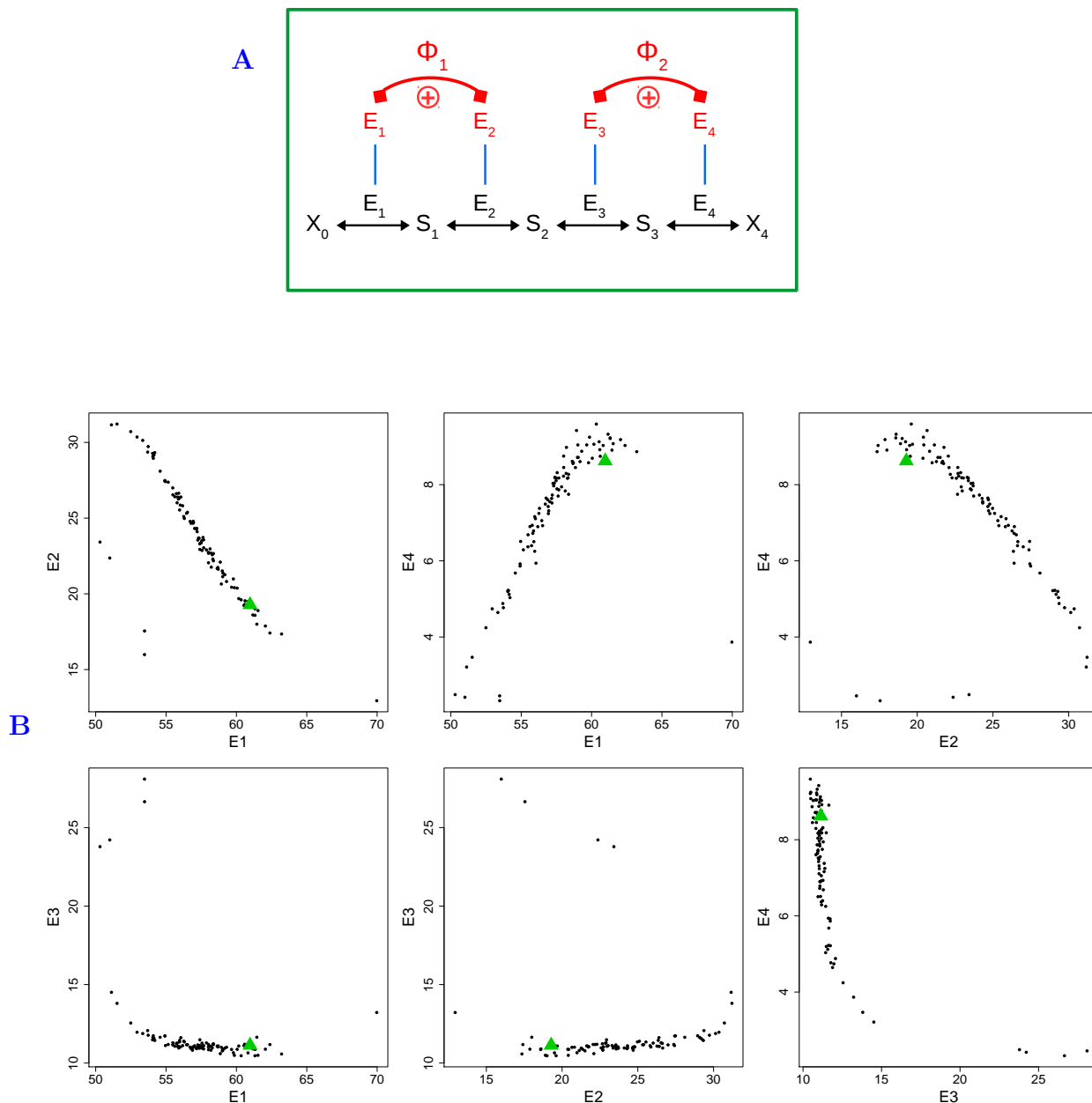
**Figure S2. End of simulation of enzyme evolution in a four-enzyme pathway when all enzymes are co-regulated with competition.** (A) The four-enzyme pathway used in these simulations. All enzymes are positively co-regulated. The green square symbolizes the competition between all enzymes. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 10000$ ,  $\beta_{12} = 0.32$ ,  $\beta_{23} = 2$   $\beta_{34} = 0.1$ . (B) Relationship between the enzymes concentrations at the end of simulation. One hundred simulations were performed, with different initial concentrations. Each black point correspond to the end of one simulation. The green triangle ( $E_1 = 61$ ,  $E_2 = 19.3$ ,  $E_3 = 11.1$ ,  $E_4 = 8.6$ ) corresponds to the maximum flux without regulation. The black points seem to be random, but actually depends on the placement of  $1/B$  and  $E^0$  is the 4D-dome.



**Figure S3. End of simulation of enzyme evolution in a four-enzyme pathway when there is one positive group and one independent enzyme with competition.** (A) The four-enzyme pathway used in these simulations. The enzymes  $E_1$ ,  $E_2$  and  $E_3$  are positively co-regulated, whereas enzyme  $E_4$  is independent. The green square symbolizes the competition between all enzymes. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 10000$ ,  $\beta_{12} = 0.1$ ,  $\beta_{23} = 2$ . (B) Relationship between the enzymes concentrations at the end of simulation. One hundred simulations were performed, with different initial concentrations. Each black point correspond to the end of one simulation. The red point correspond to the predicted effective equilibrium. The green triangle ( $E_1 = 61$ ,  $E_2 = 19.3$ ,  $E_3 = 11.1$ ,  $E_4 = 8.6$ ) corresponds to the maximum flux without regulation. The black points describe a curve close to the curve of predicted equilibria in each dimension. The average distance between simulated and predicted equilibria depends on the RNV, which is especially large ( $\approx 0.3$ ) against the value of the equilibrium for  $E_4$  (which is between 7.6 and 8.6) compared to other enzymes, causing this magnifying effect of the noise.



**Figure S4. End of simulation of enzyme evolution in a four-enzyme pathway when there are one positive and one negative groups with competition.** (A) The four-enzyme pathway used in these simulations. The enzymes  $E_1$  and  $E_2$  are positively co-regulated, whereas enzymes  $E_3$  and  $E_4$  are negatively co-regulated. The green square symbolizes the competition between all enzymes. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 10000$ ,  $\beta_{12} = 0.32$ ,  $\beta_{43} = -0.43$ . (B) Relationship between the enzymes concentrations at the end of simulation. One hundred simulations were performed, with different initial concentrations. Each black point correspond to the end of one simulation. The green triangle ( $E_1 = 61$ ,  $E_2 = 19.3$ ,  $E_3 = 11.1$ ,  $E_4 = 8.6$ ) corresponds to the maximum flux without regulation. The black points seem to describe a curve in each dimension, with more noise from  $E_1$  to  $E_4$ , probably due to different size of neutral zone. Although we cannot estimate the equilibrium, it is totally deterministic and depends on the initial concentration.



**Figure S5. End of simulation of enzyme evolution in a four-enzyme pathway when there are two positive groups with competition.** (A) The four-enzyme pathway used in these simulations. The enzymes  $E_1$  and  $E_2$  are positively co-regulated, and enzymes  $E_3$  and  $E_4$  are also positively co-regulated. The green square symbolizes the competition between all enzymes. Parameter values:  $\mathbf{A} = (1, 10, 30, 50)$ ,  $X = 1$ ,  $N = 10000$ ,  $\beta_{12} = 0.32$ ,  $\beta_{34} = 0.1$ . (B) Relationship between the enzymes concentrations at the end of simulation. One hundred simulations were performed, with different initial concentrations. Each black point correspond to the end of one simulation. The green triangle ( $E_1 = 61$ ,  $E_2 = 19.3$ ,  $E_3 = 11.1$ ,  $E_4 = 8.6$ ) corresponds to the maximum flux without regulation. The black points seem to describe a curve in each dimension, with more noise from  $E_1$  to  $E_4$ , probably due to different size of neutral zone. Although we cannot estimate the equilibrium, it is totally deterministic and depends on the initial concentration. The outsider points are probably blocked at the theoretical equilibrium for at least one group.

## II Mathematical Proofs

### 30 II.1 Properties of co-regulation groups

$\Phi_q$  is the co-regulation group  $q$ , with  $q \in \{1, p\}$ , and  $m_q$  is the number of enzymes in  $\Phi_q$ . The total number of enzymes is  $n$ , so  $\sum_{q=1}^n m_q = n$ . There is at least one enzyme by group and at most  $n - p + 1$  enzymes, *i.e.*  $\forall q, 1 \leq m_q \leq n - p + 1$ .

If there is only one enzyme in a group ( $m_q = 1$ ), this enzyme is independent from all  
35 other enzymes.

If there is only one group ( $p = 1, m_1 = n$ ), all enzymes are in this group and they are co-regulated. This is the case "Co-regulation between all enzymes".

If all enzymes are independent ( $p = n, \forall q m_q = 1$ ), there are as many groups as enzymes. This is the case "Independence between all enzymes".

### 40 II.2 Searching for evolutionary equilibrium

We search for the evolutionary equilibrium of total relative enzyme concentrations:

$$\forall j, \frac{\partial e_j}{\partial t} = 0$$

where  $e_j = E_j/E_{\text{tot}}$ . The total relative enzyme concentration can be partitioned into intra-group relative concentration  $e_j^q = E_j/E^q$ , where  $E_q = \sum_{j \in \Phi_q} E_j$ , and inter-group relative concentration  $e_q = E_q/E_{\text{tot}}$ .

Because  $e_j = e_j^q e^q$ , we have:

$$\frac{\partial e_j}{\partial t} = \frac{\partial(e_j^q e^q)}{\partial t} = \frac{\partial e_j^q}{\partial t} e^q + e_j^q \frac{\partial e^q}{\partial t} \quad (\text{S1})$$

45 Thus, solving  $\frac{\partial e_j}{\partial t} = 0$  is equivalent to solving  $\frac{\partial e_j^q}{\partial t} = 0$  and  $\frac{\partial e^q}{\partial t} = 0$ .

$$\boxed{\frac{\partial e_j}{\partial t} = 0 \iff \frac{\partial e_i^q}{\partial t} = 0 \quad \& \quad \frac{\partial e^q}{\partial t} = 0} \quad (\text{S2})$$

### II.3 Equilibria when there is no competition

Without competition for resources (free  $E_{\text{tot}}$ ), the redistribution coefficient is  $\alpha_{ij} = \beta_{ij}$  for all  $(i, j)$ , and the values of the redistribution coefficient are:

$$\forall q, \forall i \in \Phi_q, \forall j \quad \alpha_{ij} = \beta_{ij} = \begin{cases} 1/\beta_{ji} & \text{if } j \in \Phi_q \\ 1 & \text{if } j = i \\ 0 & \text{if } j \notin \Phi_q \end{cases} \quad (\text{S3})$$

50 So the differential system (equation 7) that describes the evolution of enzyme concentrations becomes:

$$\frac{\partial E_j}{\partial t} = 2\mu \sum_{i \in \Phi_q} s_i \beta_{ij} \delta_i \quad (\text{S4})$$

And the flux response coefficient writes for all  $i$  in  $\Phi_q$ :

$$R_{E_i}^J = E_i \frac{\sum_{j=1}^n \frac{\beta_{ij}}{A_j E_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}} = e_i \frac{\sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j e_j^2}}{\sum_{j=1}^n \frac{1}{A_j e_j}} \quad (\text{S5})$$

### II.3.1 Equilibrium within co-regulation groups

**II.3.1.1 Condition for equilibrium** We search  $\forall q, \forall j \in \Phi_q, e_j^{q*}$  such as  $\frac{\partial e_j^q}{\partial t} = 0$ , with  $e_j^q = E_j/E^q$ . We have:

$$\frac{\partial e_j^q}{\partial t} = \frac{\partial(E_j/E^q)}{\partial t} = \frac{\frac{\partial E_j}{\partial t} E^q - E_j \frac{\partial E^q}{\partial t}}{(E^q)^2} \quad (\text{S6})$$

From equation S4, we can write for any co-regulation group  $\Phi_q$ :

$$\begin{aligned} \frac{\partial E^q}{\partial t} &= \frac{\partial \left( \sum_{j \in \Phi_q} E_j \right)}{\partial t} = \sum_{j \in \Phi_q} \frac{\partial E_j}{\partial t} \\ &= \sum_{j \in \Phi_q} 2\mu \sum_{i \in \Phi_q} s_i \beta_{ij} \delta_i \\ &= 2\mu \sum_{i \in \Phi_q} s_i \delta_i \sum_{j \in \Phi_q} \beta_{ij} \end{aligned}$$

55 As the global co-regulation coefficient is  $B_i = \sum_{j \in \Phi_q} \beta_{ij}$  for all  $i \in \Phi_q$ , we can write:

$$\frac{\partial E^q}{\partial t} = 2\mu \sum_{i \in \Phi_q} s_i \delta_i B_i \quad (\text{S7})$$

By replacing derivatives with their expression in equation (S6), we obtain:

$$\begin{aligned} \frac{\partial e_j^q}{\partial t} &= \frac{E^q 2\mu \sum_{i \in \Phi_q} s_i \beta_{ij} \delta_i - E_j 2\mu \sum_{i \in \Phi_q} \delta_i s_i B_i}{(E^q)^2} \\ &= \frac{2\mu}{E^q} \left( \sum_{i \in \Phi_q} \beta_{ij} s_i \delta_i - e_j^q \sum_{i \in \Phi_q} s_i \delta_i B_i \right) \\ &= \frac{2\mu}{E^q} \sum_{i \in \Phi_q} s_i \delta_i (\beta_{ij} - e_j^q B_i) \end{aligned}$$

Using the relationship between the selection coefficient and flux response coefficient  $s_i = R_{E_i}^J \delta_i / E_i$  (Coton et al. 2021), and the relationship  $e_i^q = E_i / E^q$ , we get:

$$\begin{aligned} \frac{\partial e_j^q}{\partial t} &= \frac{2\mu}{E^q} \sum_{i \in \Phi_q} R_{E_i}^J \frac{\delta_i^2}{E_i} (\beta_{ij} - e_j^q B_i) \\ &= \frac{2\mu}{(E^q)^2} \sum_{i \in \Phi_q} \frac{R_{E_i}^J \delta_i^2}{e_i^q} (\beta_{ij} - e_j^q B_i) \end{aligned}$$

and from the relationship  $B_i = \beta_{ij}B_j$ , which is valid within groups, we get:

$$\begin{aligned}\frac{\partial e_j^q}{\partial t} &= \frac{2u}{(E^q)^2} \sum_{i \in \Phi_q} \frac{R_{E_i}^J \delta_i^2}{e_i^q} \left( \frac{B_i}{B_j} - e_j^q B_i \right) \\ &= \frac{2u}{(E^q)^2} \left( \frac{1}{B_j} - e_j^q \right) \sum_{i \in \Phi_q} \frac{R_{E_i}^J \delta_i^2}{e_i^q} B_i\end{aligned}$$

Therefore, the condition for evolutionary equilibrium within groups is:

$$\boxed{\frac{2u}{(E^q)^2} \left( \frac{1}{B_j} - e_j^q \right) \sum_{i \in \Phi_q} \frac{R_{E_i}^J \delta_i^2}{e_i^q} B_i = 0} \quad (\text{S8})$$

**II.3.1.2 Theoretical equilibrium within groups** An obvious solution for equation S8 is:

$$\boxed{\forall i \in \Phi_q, \quad e_i^{q*} = \frac{1}{B_i}} \quad (\text{S9})$$

Note that if at least one co-regulation coefficient is negative, leading to at least one negative  $B_i$  in the group, the equilibrium concentration is negative, which is not realistic.

**II.3.1.3 Effective equilibrium within groups** Actually equation S8 has another trivial solution:

$$\boxed{\forall i \in \Phi_q, \quad R_{E_i}^J = 0} \quad (\text{S10})$$

or, from equation S5:

$$\forall i \in \Phi_q, \quad \sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j (e_j)^2} = 0$$

As  $e_j = e_j^q e^q$ , we get:

$$\forall i \in \Phi_q, \quad \frac{1}{(e^q)^2} \sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j (e_j^q)^2} = 0$$

This equality can only be solved if at least one co-regulation coefficient is negative in group  $\Phi_q$ . Because  $\forall (i, j) \in \Phi_q, \beta_{ij} = B_i/B_j$  (equation 5), we get:

$$\forall i \in \Phi_q, \quad B_i \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j)^2} = 0$$

or

$$\boxed{\sum_{j \in \Phi_q} \frac{1}{B_j A_j (\tilde{e}_j^q)^2} = 0} \quad (\text{S11})$$

where  $\tilde{e}_j^q$  is the value at intra-group effective equilibrium that solved this equation. Thanks to the relationship between  $e_j^q$  and  $\tau^q$  (equation 6), searching for  $\tilde{e}_j^q$  for all  $j \in \Phi_q$  is equivalent to searching for  $\tilde{\tau}^q$  using the following equation:

$$\sum_{j \in \Phi_q} \frac{1}{B_j A_j (\tilde{\tau}^q (1/B_j - e_j^{q0}) + e_j^{q0})^2} = 0 \quad (\text{S12})$$



### II.3.2 Equilibrium between co-regulation groups

75 **II.3.2.1 Condition for equilibrium** We search  $\forall q$ ,  $e^{q*}$  such as  $\frac{\partial e^q}{\partial t} = 0$ , with  $e^q = E^q/E_{\text{tot}}$ . We have:

$$\frac{\partial e^q}{\partial t} = \frac{\partial(E^q/E_{\text{tot}})}{\partial t} = \frac{\frac{\partial E^q}{\partial t} E_{\text{tot}} - E^q \frac{\partial E_{\text{tot}}}{\partial t}}{(E_{\text{tot}})^2} \quad (\text{S13})$$

From equation S4, we can write for any co-regulation group  $\Phi_q$  (equation S7, Supporting Information II.3.1.1):

$$\frac{\partial E^q}{\partial t} = 2\mu \sum_{i \in \Phi_q} s_i \delta_i B_i$$

In a same way, we get for the total concentration:

$$\frac{\partial E_{\text{tot}}}{\partial t} = \sum_{j=1}^n \frac{\partial E_j}{\partial t} = \sum_{q=1}^p \frac{\partial E^q}{\partial t} = \sum_{q=1}^p 2\mu \sum_{i \in \Phi_q} s_i \delta_i B_i$$

or

$$\frac{\partial E_{\text{tot}}}{\partial t} = 2\mu \sum_{i=1}^n s_i \delta_i B_i$$

By replacing derivatives with their expression in equation (S13), we obtain:

$$\begin{aligned} \frac{\partial e^q}{\partial t} &= \frac{E_{\text{tot}} 2\mu \sum_{i \in \Phi_q} s_i \delta_i B_i - E^q 2\mu \sum_{i=1}^n s_i \delta_i B_i}{(E_{\text{tot}})^2} \\ &= \frac{2\mu}{E_{\text{tot}}} \left( \sum_{i \in \Phi_q} s_i \delta_i B_i - e^q \sum_{i=1}^n s_i \delta_i B_i \right) \end{aligned}$$

Using  $s_i = R_{E_i}^J \delta_i / E_i$  (Coton et al. 2021), we get:

$$\frac{\partial e^q}{\partial t} = \frac{2\mu}{E_{\text{tot}}} \left( \sum_{i \in \Phi_q} R_{E_i}^J \frac{\delta_i^2}{E_i} B_i - e^q \sum_{i=1}^n R_{E_i}^J \frac{\delta_i^2}{E_i} B_i \right)$$

When there is co-regulation without competition, we have  $\delta_i = \nu$  for all  $i$  (Coton et al. 2021), so:

$$\frac{\partial e^q}{\partial t} = \frac{2\mu\nu^2}{E_{\text{tot}}} \left( \sum_{i \in \Phi_q} \frac{R_{E_i}^J}{E_i} B_i - e^q \sum_{i=1}^n \frac{R_{E_i}^J}{E_i} B_i \right)$$

Thus, the condition  $\frac{\partial e^q}{\partial t} = 0$  for evolutionary equilibrium between groups is:

$$\boxed{\frac{1}{e^q} \sum_{i \in \Phi_q} \frac{R_{E_i}^J}{E_i} B_i = \sum_{i=1}^n \frac{R_{E_i}^J}{E_i} B_i} \quad (\text{S14})$$

**II.3.2.2 Positive groups and singletons** To make apparent the other co-regulation groups in equation S14, we split the sum of the right side of the equation:

$$\frac{1}{e^q} \sum_{i \in \Phi_q} \frac{R_{E_i}^J}{E_i} B_i = \sum_{k=1}^p \sum_{i \in \Phi_k} \frac{R_{E_i}^J}{E_i} B_i$$

Replacing the response coefficient by its expression (equation S5), we get:

$$\frac{1}{e^q} \sum_{i \in \Phi_q} B_i \sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j E_j^2} = \sum_{k=1}^p \sum_{i \in \Phi_k} B_i \sum_{j \in \Phi_k} \frac{\beta_{ij}}{A_j E_j^2}$$

Thus we can use the relationship  $\beta_{ij} = B_i/B_j$  (equation 5), which is valid only if  $i$  and  $j$  are in the same co-regulation group:

$$\frac{1}{e^q} \sum_{i \in \Phi_q} B_i^2 \sum_{j \in \Phi_q} \frac{1}{B_j A_j E_j^2} = \sum_{k=1}^p \sum_{i \in \Phi_k} B_i^2 \sum_{j \in \Phi_k} \frac{1}{B_j A_j E_j^2}$$

From the relationships  $e_j^q = E_j/E^q$  and  $e^q = E^q/E_{\text{tot}}$ , we get:

$$\frac{1}{e^q} \sum_{i \in \Phi_q} B_i^2 \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q e^q E_{\text{tot}})^2} = \sum_{k=1}^p \sum_{i \in \Phi_k} B_i^2 \sum_{j \in \Phi_k} \frac{1}{B_j A_j (e_j^k e^k E_{\text{tot}})^2}$$

The inter-group relative concentrations  $e^q$  and  $e^k$  can be taken out from the sums:

$$\frac{1}{(e^q)^3} \sum_{i \in \Phi_q} B_i^2 \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} = \sum_{k=1}^p \frac{1}{(e^k)^2} \sum_{i \in \Phi_k} B_i^2 \sum_{j \in \Phi_k} \frac{1}{B_j A_j (e_j^k)^2}$$

As we consider positive groups and singletons, the intra-group theoretical equilibrium is  $e_j^{q*} = 1/B_j$  (equation 11), so:

$$\frac{1}{(e^{q*})^3} \sum_{i \in \Phi_q} B_i^2 \sum_{j \in \Phi_q} \frac{B_j}{A_j} = \sum_{k=1}^p \frac{1}{(e^{k*})^2} \sum_{i \in \Phi_k} B_i^2 \sum_{j \in \Phi_k} \frac{B_j}{A_j}$$

Let's call  $A^{q*}$  the *apparent activity of co-regulation group*  $\Phi_q$  at theoretical equilibrium:

$$\boxed{A^{q*} = \left( \sum_{\substack{i \in \Phi_q \\ j \in \Phi_q}} \frac{B_i^2 B_j}{A_j} \right)^{-1}} \quad (\text{S15})$$

So the previous expression becomes:

$$\frac{1}{A^{q*} (e^{q*})^3} = \sum_{k=1}^p \frac{1}{A^{k*} (e^{k*})^2}$$

As  $\sum_{k=1}^n \frac{1}{A^{k*} (e^{k*})^2}$  is constant, this equality is valid for any co-regulation group:

$$\frac{1}{A^{q*} (e^{q*})^3} = \frac{1}{A^{q'*} (e^{q'*})^3} = \sum_{k=1}^p \frac{1}{A^{k*} (e^{k*})^2}$$

Because the apparent activities are strictly positive when there are only positive groups  
80 and singletons, we get:

$$\forall q, \forall q', \quad e^{q'*} = e^{q*} \frac{(A^q)^{1/3}}{(A^{q'})^{1/3}}$$

Since  $\sum_{q'=1}^p e^{q'} = 1$ , we have:

$$\sum_{q'=1}^p e^{q'*} = e^{q*} (A^{q*})^{1/3} \sum_{q'=1}^p \frac{1}{(A^{q'*})^{1/3}} = 1$$

This leads to the expression of the theoretical equilibrium of relative enzyme concentra-  
tions when all groups are positives or singletons:

$$\boxed{\forall q \quad e^{q*} = \frac{(1/A^{q*})^{1/3}}{\sum_{q'=1}^p (1/A^{q'*})^{1/3}}} \quad (\text{S16})$$

**II.3.2.3 Negative groups** When all groups are negative, the flux response coeffi-  
85 cients in equation S14 are all null, so we could not use the previous method to find the  
equilibria. To overcome the problem, we relied on the relationship between  $\tau^q$  and the ab-  
solute concentrations within groups,  $E^q$  (Supporting Information II.3.3. Thus, at effective  
equilibrium, the absolute concentration writes:

$$\tilde{E}^q = \frac{E^{q0}}{1 - \tilde{\tau}^q} \quad (\text{S17})$$

where  $\tilde{\tau}^q$  is the value of the driving variable at the effective equilibrium. Therefore, the  
90 inter-group relative concentrations at the effective equilibrium are:

$$\boxed{\forall q \quad \tilde{e}^q = \frac{\frac{E^{q0}}{1 - \tilde{\tau}^q}}{\sum_{k=1}^p \frac{E^{k0}}{1 - \tilde{\tau}^k}}} \quad (\text{S18})$$

**II.3.2.4 Positive and negative groups** When there are both types of groups, the  
response coefficients are null for negative groups and non-null for positive groups and  
singletons. So equation S14 is solvable only for positive groups and singletons, which  
leads to this equilibrium between these groups:

$$\forall q \text{ such as } \theta_q \geq 0 \quad \frac{E^q}{\sum_{\substack{q'=1 \\ \theta_{q'} \geq 0}}^p E^{q'}} = \frac{(1/A^{q*})^{1/3}}{\sum_{\substack{q'=1 \\ \theta_{q'} \geq 0}}^p (1/A^{q'*})^{1/3}} \quad (\text{S19})$$

95 which is similar to equation 12, by considering  $E^q / \sum_{\theta_q \geq 0} E^q$  instead of  $e^q$ .

In the negative groups the relative enzyme concentrations tend to their effective equi-  
librium.

To study the behavior of absolute concentrations, we use the relationship between absolute concentration and the driving variable (Supporting Information II.3.3):

$$\forall i \in \Phi_q \quad E_i = E_i^0 + \frac{\tau^q E^{q^0}}{(1 - \tau^q)B_i} \quad (\text{S20})$$

100 In the positive groups,  $e_i^q$  tends to the intra-group theoretical equilibrium  $1/B_i$ , so  $\tau^q$  tends to 1 (see equation 6), therefore  $E_i$  tends to infinity. The enzyme concentration in singletons also increase to infinity (Coton et al. 2021). In the negative groups,  $e_i^q$  reaches the effective equilibrium, so  $\tau^q$  tends to  $\tilde{\tau}^q$ , therefore  $E_i$  tends to  $\tilde{E}_i$ . Therefore, the absolute concentrations of enzymes in the negative groups become negligible compared  
105 to those in the positive groups and singletons. The inter-group relative concentrations tend towards the equilibrium of expression S19 in positive groups and towards zero in negative groups. Because  $\sum_{\theta_q \geq 0} E^q$  tends towards  $E_{\text{tot}}$ , we can say that for positive groups and singletons  $e^q$  tend towards  $e^{q^*}$  (equation 12) considering only positive groups and singletons in denominator instead of all groups, *i.e.*:

$$e^{q^*} = \frac{(1/A^{q^*})^{1/3}}{\sum_{\substack{q'=1 \\ \theta_{q'} \geq 0}}^p (1/A^{q'^*})^{1/3}} \quad (\text{S21})$$

110 The same applies for  $e_i$  that tend towards  $e_i^*$  (equation 14).

When there are positive and negative co-regulation groups, the flux writes:

$$J = \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}} \quad (\text{S22})$$

$$= \frac{X}{\sum_{q=1}^p \sum_{j \in \Phi_q} \frac{1}{A_j E_j}} \quad (\text{S23})$$

$$= \frac{X}{\sum_{\substack{q=1 \\ \theta_q \geq 0}}^p \sum_{j \in \Phi_q} \frac{1}{A_j E_j} + \sum_{\substack{q=1 \\ \theta_q < 0}}^p \sum_{j \in \Phi_q} \frac{1}{A_j E_j}} \quad (\text{S24})$$

As enzyme concentrations are stuck at the effective equilibrium in negative groups while they increase indefinitely in positive groups and singletons, we have:

$$J \xrightarrow{t \rightarrow \infty} \frac{X}{\sum_{\substack{q=1 \\ \theta_q < 0}}^p \sum_{j \in \Phi_q} \frac{1}{A_j \tilde{E}_j}} = \tilde{J}_{\text{max}} \quad (\text{S25})$$

### 115 II.3.3 Relationship between absolute concentrations and driving variable

The intra-group relative concentrations in a group are linearly related. The parametric equation of the line  $\mathcal{E}^q$  writes:

$$\forall i \in \Phi_q \quad e_i^q = \tau^q(1/B_i - e_i^{q^0}) - e_i^{q^0} \quad (\text{S26})$$

where  $\tau^q$  is the driving variable.

The absolute concentrations can also be expressed depending on  $\tau^q$ .

The relation between  $E^q$  and  $\tau^q$  can be derived from the relation  $B_i = \frac{\Delta E^q}{\Delta E_i}$  (equation 4). We have:

$$\begin{aligned} E^q &= E^{q^0} + B_i(E_i - E_i^0) \\ &= E^q B_i e_i^q + E^{q^0} (1 - B_i e_i^q) \\ &= E^{q^0} \frac{1 - B_i e_i^q}{1 - B_i e_i^q} \end{aligned}$$

120 Replacing  $e_i^q$  by its expression depending on  $\tau^q$  (equation 6), we get:

$$E^q = E^{q^0} \frac{1 - B_i e_i^{q^0}}{1 - B_i (\tau^q (1/B_i - e_i^{q^0}) + e_i^{q^0})}$$

which is simplified to (for  $\tau^q \neq 1$ ):

$$E^q = \frac{E^{q^0}}{1 - \tau^q} \quad (\text{S27})$$

$$\boxed{E^q = \frac{E^{q^0}}{1 - \tau^q}} \quad (\text{S28})$$

In the parametric equation 6, we can make visible the *absolute* enzyme concentrations, using the equation  $e_i^q = E_i/E^q$ :

$$\forall i, \frac{E_i}{E^q} = \tau \left( \frac{1}{B_i} - \frac{E_i^0}{E^{q^0}} \right) + \frac{E_i^0}{E^{q^0}} \quad (\text{S29})$$

Introducing in this equation the expression of  $E^q$  (equation S28), we get:

$$\forall i \in \Phi_q \frac{E_i(1 - \tau^q)}{E^{q^0}} = \tau^q \left( \frac{1}{B_i} - \frac{E_i^0}{E^{q^0}} \right) + \frac{E_i^0}{E^{q^0}}$$

Thus:

$$E_i = E_i^0 + \frac{\tau^q E^{q^0}}{(1 - \tau^q) B_i}$$

125 Or:

$$\boxed{\forall i \in \Phi_q \quad E_i = E_i^0 + \frac{E^{q^0}}{B_i} \left( -1 + \frac{1}{1 - \tau^q} \right)} \quad (\text{S30})$$

## II.4 Equilibria when there is competition

With competition for resources between all enzymes (fixed  $E_{\text{tot}}$ ), we have for the co-regulation coefficients:

$$\forall q, \forall i \in \Phi_q, \forall j \quad \beta_{ij} = \begin{cases} 1/\beta_{ji} & \text{if } j \in \Phi_q \\ 1 & \text{if } j = i \\ 0 & \text{if } j \notin \Phi_q \end{cases} \quad (\text{S31})$$

and for the redistribution coefficients:

$$\forall q, \forall i \in \Phi_q, \forall j \quad \alpha_{ij} = \frac{\beta_{ij}/B_i - e_j}{1/B_i - e_i} = \begin{cases} \frac{1/B_j - e_j}{1/B_i - e_i} & \text{if } j \in \Phi_q \\ \frac{-e_j}{1/B_i - e_i} & \text{if } j \notin \Phi_q \end{cases} \quad (\text{S32})$$

So the system of differential equations that describes the evolution of enzyme concentrations (equation 7) becomes:

$$\begin{aligned} \forall q, \forall j \in \Phi_q \quad \frac{\partial E_j}{\partial t} &= 2\mu \sum_{i=1}^n s_i \delta_i \alpha_{ij} \\ &= 2\mu \sum_{i \in \Phi_q} s_i \delta_i \frac{1/B_j - e_j}{1/B_i - e_i} + 2\mu \sum_{i \notin \Phi_q} s_i \delta_i \frac{-e_j}{1/B_i - e_i} \end{aligned}$$

130 Thus:

$$\forall q, \forall j \in \Phi_q \quad \frac{\partial E_j}{\partial t} = 2\mu \frac{1}{B_j} \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} - 2\mu e_j \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \quad (\text{S33})$$

As the total concentration is fixed with competition, we have for the total relative concentration:

$$\forall q, \forall j \in \Phi_q \quad \frac{\partial e_j}{\partial t} = \frac{1}{E_{\text{tot}}} \frac{\partial E_j}{\partial t}$$

Thanks to the relationship between the selection coefficient and the response coefficient  $s_i = R_{E_i}^J \delta_i / E_i$  (Coton et al. 2021), we get:

$$\forall q, \forall j \in \Phi_q \quad \frac{\partial e_j}{\partial t} = 2 \frac{\mu}{E_{\text{tot}}} \left( \frac{1}{B_j} \sum_{i \in \Phi_q} \frac{R_{E_i}^J \delta_i^2}{E_i (1/B_i - e_i)} - e_j \sum_{i=1}^n \frac{R_{E_i}^J \delta_i^2}{E_i (1/B_i - e_i)} \right) \quad (\text{S34})$$

#### II.4.1 Effective equilibria

A solution to the evolutionary equilibrium  $\frac{\partial e_j}{\partial t} = 0$  is:

$$\boxed{\forall i \quad R_{E_i}^J = 0} \quad (\text{S35})$$

135 which is the definition of the effective equilibrium, and corresponds to maximal flux in regard to each enzyme.

$$\begin{aligned}
\forall q, \forall i \in \Phi_q \quad R_{E_i}^J = 0 &\Leftrightarrow e_i \frac{\sum_{j=1}^n \frac{\alpha_{ij}}{A_j e_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}} = 0 \\
&\Leftrightarrow \sum_{j=1}^n \frac{\alpha_{ij}}{A_j e_j^2} = 0 \\
&\Leftrightarrow \sum_{j=1}^n \frac{\beta_{ij}/B_i - e_j}{1/B_i - e_i} \frac{1}{A_j e_j^2} = 0 \\
&\Leftrightarrow \sum_{j \in \Phi_q} \frac{1/B_j - e_j}{A_j e_j^2} + \sum_{j \notin \Phi_q} \frac{-e_j}{A_j e_j^2} = 0 \\
&\Leftrightarrow \sum_{j \in \Phi_q} \frac{1}{B_j A_j e_j^2} - \sum_{j=1}^n \frac{e_j}{A_j e_j^2} = 0
\end{aligned}$$

So the condition for the effective equilibrium is:

$$\forall q, \quad \sum_{j \in \Phi_q} \frac{1}{B_j A_j e_j^2} = \sum_{j=1}^n \frac{1}{A_j e_j} \quad (\text{S36})$$

**II.4.1.1 Between groups** Replace  $e_j$  by  $e^q e_j^q$  in the left-hand side of equation [S36](#):

$$\begin{aligned}
\sum_{j \in \Phi_q} \frac{1}{B_j A_j e_j^2} &= \sum_{j=1}^n \frac{1}{A_j e_j} \\
\Leftrightarrow \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e^q e_j^q)^2} &= \sum_{j=1}^n \frac{1}{A_j e_j} \\
\Leftrightarrow \frac{1}{(e^q)^2} \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} &= \sum_{j=1}^n \frac{1}{A_j e_j}
\end{aligned}$$

Defining  $A^q$  as the apparent activity of co-regulation group  $\Phi_q$ , such as:

$$\frac{1}{A^q} = \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} \quad (\text{S37})$$

we get:

$$\frac{1}{(e^q)^2} \frac{1}{A^q} = \sum_{j=1}^n \frac{1}{A_j e_j} \quad (\text{S38})$$

<sup>140</sup> As  $\sum_{j=1}^n \frac{1}{A_j e_j}$  is constant, this equality is valid for any group:

$$\frac{1}{(e^q)^2} \frac{1}{A^q} = \sum_{j=1}^n \frac{1}{A_j e_j} = \frac{1}{(e^k)^2} \frac{1}{A^k}$$

and:

$$(e^q)^2 = \frac{A^k (e^k)^2}{A^q}$$

Assuming that for all  $q$   $A^q$  is positive even for negative groups (which has been empirically verify with some simulations when equilibrium is computable), we have:

$$e^k = \left( e^q \frac{A^q}{A^k} \right)^{1/2}$$

As  $\sum_{k=1}^p e^k = 1$ , we can write:

$$\sum_{k=1}^p e^k = 1 = \sum_{k=1}^p e^q \left( \frac{A^q}{A^k} \right)^{1/2}$$

So at the effective equilibrium, we have:

$$\boxed{\tilde{e}^q = \frac{\tilde{A}^q^{-1/2}}{\sum_{k=1}^p \tilde{A}^k^{-1/2}}} \quad (\text{S39})$$

with

$$\frac{1}{\tilde{A}^q} = \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} \quad (\text{S40})$$

**II.4.1.2 Within groups** In expression S36, we replace  $e_j$  by  $e^q e_j^q$ :

$$\frac{1}{(e^q)^2} \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} - \sum_{k=1}^p \frac{1}{e^k} \sum_{j \in \Phi_k} \frac{1}{A_j e_j^k} = 0$$

In the first member, we recognize the expression of the apparent activity (equation S37), so we have:

$$\frac{1}{A^q (e^q)^2} - \sum_{k=1}^p \frac{1}{e^k} \sum_{j \in \Phi_k} \frac{1}{A_j e_j^k} = 0$$

145 Replacing  $e^k$  by its expression (equation S39), we get:

$$\frac{1}{A^q \left( \frac{(A^q)^{-1/2}}{\sum_{q'=1}^p (\tilde{A}^{q'})^{-1/2}} \right)^2} - \sum_{k=1}^p \frac{1}{\frac{(A^k)^{-1/2}}{\sum_{q'=1}^p (\tilde{A}^{q'})^{-1/2}}} \sum_{j \in \Phi_k} \frac{1}{A_j e_j^k} = 0$$

Leading to:

$$\sum_{q'=1}^p (\tilde{A}^{q'})^{-1/2} - \sum_{k=1}^p (A^k)^{1/2} \sum_{j \in \Phi_k} \frac{1}{A_j e_j^k} = 0$$



However, if there is only one enzyme in  $\Phi_q$ , *i.e.* if  $\theta_q = 0$ , we have  $e_i^q = 1$  and  $A^q = A_i$ . Thus  $A_i^{-1/2} - A_i^{1/2}/A_i = A_i^{-1/2} - A_i^{-1/2} = 0$ , and the elements corresponding to  $\theta_q = 0$  disappear in the expression:

$$\boxed{\sum_{\substack{k=1 \\ \theta_k \neq 0}}^p \left( (A^k)^{-1/2} - (A^k)^{1/2} \sum_{j \in \Phi_k} \frac{1}{A_j e_j^k} \right) = 0} \quad (\text{S41})$$

**II.4.1.3 Particular case: only one regulated group and singletons** When there is only one co-regulated group  $\Phi_q$  (positive or negative) and independent enzymes, we get from equation S41:

$$(A^q)^{-1/2} - (A^q)^{1/2} \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} = 0$$

or:

$$\frac{1}{A^q} = \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q}$$

Using the expression of  $A^q$  (equation S37), we get:

$$\begin{aligned} \sum_{j \in \Phi_q} \frac{1}{B_j A_j (e_j^q)^2} &= \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} \\ \Leftrightarrow \sum_{j \in \Phi_q} \left( \frac{1}{B_j A_j (e_j^q)^2} - \frac{1}{A_j e_j^q} \right) &= 0 \\ \Leftrightarrow \sum_{j \in \Phi_q} \left( \frac{1/B_j - e_j^q}{A_j (e_j^q)^2} \right) &= 0 \end{aligned}$$

Replacing  $e_j^q$  by its expression in the parametric equation of the line  $\mathcal{E}^q$  (equation 6), we get:

$$\begin{aligned} \sum_{j \in \Phi_q} \left( \frac{1/B_j - (\tau^q(1/B_j - e_j^{q0}) - e_j^{q0})}{A_j (e_j^q)^2} \right) &= 0 \\ \Leftrightarrow (1 - \tau^q) \sum_{j \in \Phi_q} \left( \frac{1/B_j - e_j^{q0}}{A_j (e_j^q)^2} \right) &= 0 \end{aligned}$$

$$\boxed{(1 - \tau^q) \sum_{j \in \Phi_q} \frac{1/B_j - e_j^{q0}}{A_j (\tau^q(1/B_j - e_j^{q0}) + e_j^{q0})^2} = 0} \quad (\text{S42})$$

Thus, there are two solutions:

- $\tau^{q*} = 1$ , which is the theoretical equilibrium (Supporting Information II.4.2), but in this case the flux response coefficient is not null
- $\tilde{\tau}^q \neq 1$ , which cancels the right part of the equation and can be found numerically. In this case,  $\tilde{\tau}^q$  nullifies the response coefficient, which corresponds to a local maximal flux. The right part of the equation is similar to the expression of the effective equilibrium when all enzymes are co-regulated.

### II.4.2 Intra-group theoretical equilibria

160 Here we search  $e_i^{q*}$  such as  $\frac{\partial e_i^q}{\partial t} = 0$ .

$$\forall q, \forall j \in \Phi_q, \quad \frac{\partial e_j^q}{\partial t} = \frac{\partial E_j / E^q}{\partial t} = \frac{E^q \frac{\partial E_j}{\partial t} - E_j \frac{\partial E^q}{\partial t}}{(E^q)^2} \quad (\text{S43})$$

From this equation S33, we can write for any co-regulation group  $\Phi_q$ :

$$\begin{aligned} \forall q, \quad \frac{\partial E^q}{\partial t} &= \frac{\partial \left( \sum_{j \in \Phi_q} E_j \right)}{\partial t} = \sum_{j \in \Phi_q} \frac{\partial E_j}{\partial t} \\ &= \sum_{j \in \Phi_q} \left( 2\mu \frac{1}{B_j} \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} - 2\mu e_j \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \right) \\ &= 2\mu \left( \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} \sum_{j \in \Phi_q} \frac{1}{B_j} - \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \sum_{j \in \Phi_q} e_j \right) \end{aligned}$$

As  $\sum_{j \in \Phi_q} e_j = \sum_{j \in \Phi_q} \frac{E_j}{E_{\text{tot}}} = \frac{E^q}{E_{\text{tot}}} = e^q$  and  $\sum_{j \in \Phi_q} \frac{1}{B_j} = 1$ , we get:

$$\forall q, \quad \frac{\partial E^q}{\partial t} = 2\mu \left( \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} - e^q \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \right) \quad (\text{S44})$$

Replacing  $\frac{\partial E_j}{\partial t}$  and  $\frac{\partial E^q}{\partial t}$  by their expressions (respectively S33 and S44) in equation S43, and with  $e_j^q = E_j / E^q$ , we have:

$$\frac{\partial e_j^q}{\partial t} = \frac{2\mu}{E^q} \left[ \left( \frac{1}{B_j} \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} - e_j \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \right) - e_j^q \left( \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} - e^q \sum_{i=1}^n \frac{s_i \delta_i}{1/B_i - e_i} \right) \right]$$

Because  $e_i = e_i^q e^q$ , we get:

$$\frac{\partial e_j^q}{\partial t} = \frac{2\mu}{E^q} \left[ \left( \frac{1}{B_j} - e_j^q \right) \sum_{i \in \Phi_q} \frac{s_i \delta_i}{1/B_i - e_i} \right]$$

165 There is an obvious solution for  $\frac{\partial e_j^q}{\partial t} = 0$ , which is the theoretical equilibrium within groups:

$$\boxed{\forall j \in \Phi_q, \quad e_j^{q*} = \frac{1}{B_j}} \quad (\text{S45})$$

This expression of  $e_i^q$  solves  $\partial e_i^q / \partial t = 0$ , but does not solve  $\partial e_i / \partial t = 0$ . However, if at least one co-regulation coefficient is negative, leading to at least one negative  $B_i < 0$  in the group, the equilibrium concentration is negative, which is impossible to reach.

## II.5 Range of neutral variation without competition

To search if the selection coefficient tends to zero at equilibrium, which corresponds to neutrality of variations of enzyme concentrations, we used the relationship between the flux response coefficients and the selection coefficients (Coton et al. 2021):

$$s_i = R_{E_i}^J \frac{\delta_i}{E_i} \quad (\text{S46})$$

### 175 II.5.1 When there are only positive groups and singletons

At the theoretical equilibrium, the response coefficients are positive (using 14 in equation S5), and  $\beta_{ij} \geq 0$  for all  $(i, j)$ , whereas  $E_i$  can increase indefinitely. Therefore the selection coefficients tends to zero.

### II.5.2 When there are only negative groups

180 At the effective equilibrium the response coefficients are null for all enzymes, therefore the selection coefficients are null.

### II.5.3 When there are positive (and/or singletons) and negative groups

When there are both types of groups, the response coefficients depend on the group in which are the enzymes, according to the relation (equation S5):

$$\forall i \in \Phi_q, \quad R_{E_i}^J = e_i \frac{\sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j e_j^2}}{\frac{n}{\sum_{j=1}^n \frac{1}{A_j e_j}}}$$

185 At equilibrium for positive groups and singletons, the numerator is strictly positive and tends toward a constant if  $\theta_q \geq 0$ . For negative groups, it is null if  $\theta_q < 0$ .

The denominator is the same for all enzymes, whatever their co-regulation groups. It can be expanded as follows:

$$\begin{aligned} \sum_{j=1}^n \frac{1}{A_j e_j} &= \sum_{q=1}^p \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q e^q} \\ &= \sum_{\substack{q=1 \\ \theta_q \geq 0}}^p \frac{1}{e^q} \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} + \sum_{\substack{q=1 \\ \theta_q < 0}}^p \frac{1}{e^q} \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} \end{aligned}$$

190 Replacing the  $e_j^q$  by their respective equilibrium (equation 11 for  $\theta_q \geq 0$  and equation 16 for  $\theta_q < 0$ ), we get:

$$\sum_{j=1}^n \frac{1}{A_j e_j} = \sum_{\substack{q=1 \\ \theta_q \geq 0}}^p \frac{1}{e^q} \sum_{\substack{j=1 \\ E_j \in \Phi_q}}^n \frac{B_j}{A_j} + \sum_{\substack{q=1 \\ \theta_q < 0}}^p \frac{1}{e^q} \sum_{\substack{j=1 \\ E_j \in \Phi_q}}^n \frac{1}{A_j e_j^q}$$

When the system tends to equilibrium within groups, (see section 3.1.3),  $e^q \xrightarrow[t \rightarrow \infty]{} 0$  for  $\theta_q < 0$  and  $e^q \xrightarrow[t \rightarrow \infty]{} e^{q^*}$  for  $\theta_q \geq 0$ . Thus, we have for the denominator:

$$\sum_{j=1}^n \frac{1}{A_j e_j} \xrightarrow[t \rightarrow \infty]{} +\infty$$

Therefore, we have for positive groups and singletons:

$$\forall i \in \Phi_q, \theta_q \geq 0, \quad R_{E_i}^J \xrightarrow[t \rightarrow \infty]{} e_i^* \frac{\sum_{j=1}^n \frac{\beta_{ij}}{A_j (e_j^*)^2}}{+\infty} = 0^+$$

$$\begin{aligned} E_i &\xrightarrow[t \rightarrow \infty]{} \infty \\ \text{and } R_{E_i}^J &\xrightarrow[t \rightarrow \infty]{} 0^+ \\ \text{so } s_i &\approx R_{E_i}^J \frac{\delta_i}{E_i} \xrightarrow[t \rightarrow \infty]{} 0 \end{aligned}$$

For negative groups, the system tends toward effective equilibrium:

$$\begin{aligned} E_i &\xrightarrow[t \rightarrow \infty]{} \tilde{E}_i \\ \text{and } R_{E_i}^J &\xrightarrow[t \rightarrow \infty]{} 0 \\ \text{so } s_i &\approx R_{E_i}^J \frac{\delta_i}{E_i} \xrightarrow[t \rightarrow \infty]{} 0 \end{aligned}$$

In conclusion, the selection coefficients tend to zero whatever the co-regulation groups.

195

## II.5.4 RNV order

Combining equations S5 and S46 and because  $\delta_i = \mu$  for all  $i$  when there is no competition, the selection coefficients writes:

$$\forall i \in \Phi_q, \quad s_i \approx \frac{\mu}{E_i} E_i \frac{\sum_{j \in \Phi_q} \frac{\beta_{ij}}{A_j E_j^2}}{\sum_{j=1}^n \frac{1}{A_j E_j}}$$

The denominator is common for all enzymes, thus we write  $Den = \sum_{j=1}^n \frac{1}{A_j E_j}$ . We use

200  $E_i = E_{\text{tot}} e_i^q e^q$ .

For positive groups and singletons, we have at theoretical equilibrium:

$$\begin{aligned}
\forall i \in \Phi_q, \quad s_i &\approx \mu \frac{\sum_{j \in \Phi_q} \beta_{ij} A_j E_j^2}{Den} \\
&\approx \frac{\mu}{E_{\text{tot}}^2 Den} \sum_{j \in \Phi_q} \frac{B_i/B_j}{A_j \left(\frac{1}{B_j}\right)^2 \left(\frac{(A^{q*})^{-1/3}}{\sum_{k=1}^p (A^{k*})^{-1/3}}\right)^2} \\
&\approx \frac{\mu \left(\sum_{k=1}^p (A^{k*})^{-1/3}\right)^2}{E_{\text{tot}}^2 Den} \sum_{j \in \Phi_q} \frac{B_i (A^{q*})^{2/3}}{A_j B_j}
\end{aligned}$$

The sum on the elements on the group is too complex to compare between groups. However, for independent enzymes, we have  $B_i = 1$  and  $A^q = A_j$ , thus the selection coefficient becomes for singletons:

$$s_i \approx \frac{\mu \left(\sum_{k=1}^p (A^{k*})^{-1/3}\right)^2}{E_{\text{tot}}^2 Den} A_j^{-1/3}$$

As RNV order is the reverse of selection coefficient order (Coton et al. 2021), we suggest that the RNV is in the same order as pseudo-activities between independent enzymes at theoretical equilibrium.

As in Coton et al. (2021), we consider the inferior and the superior limits of the RNV for enzyme  $i$  in  $\Phi_q$ . Using the relationship between  $E_i$  and  $\tau^q$  (equation S30), we have:

$$|E_i^{\text{sup}} - E_i^{\text{inf}}| = \left| \frac{E^{q0}}{B_i} \left( \frac{1}{1 - \tau^{q,\text{textsup}}} - \frac{1}{1 - \tau^{q,\text{inf}}} \right) \right|$$

where where  $\tau^{q,\text{textsup}}$  and  $\tau^{q,\text{textinf}}$  are the superior and inferior limits of the driving variable for group  $\Phi_q$ . Thus the RNVs of enzymes in the same group are inversely related to  $|B_i|$ 's.

However, we cannot compare the RNV between groups.

### III Complementary analysis

#### III.1 Degrees of freedom depend on the applied constraint

215 We showed that the evolutionary equilibrium and the flux shape depend on the constraint applied on the pathway.

In other words, each constraint modifies the adaptive landscape, by limiting the ability to move in the multidimensional space  $(\mathbf{E}, J)$ . We searched the number of degrees of freedom (df) of the system, which is the number of independent variables necessary and sufficient to describe the system, given the fixed parameters – here the pseudo-activities  $\mathbf{A}$ , 220 the co-regulation coefficients of matrix  $M_\beta$  and the initial concentrations  $\mathbf{E}^0$ . The number of df is fully determined by the relationship between enzyme concentrations, as the flux is a function of enzyme parameters (equation 1). Lower df means that the system lose dimensions on which he can move in the space  $(\mathbf{E}, J)$ , modifying the adaptive landscape. 225 The diminution of df number depends on the applied constraint:

(i) In a system without any constraint on enzymes, there are as many degrees of freedom as enzymes. The system moves freely in the multidimensional space  $(\mathbf{E}, J)$ , although there is an optimal proportions of *relative* enzyme concentrations – the evolutionary equilibrium.

230 (ii) When there is competition between all enzymes, we only have to determine  $n - 1$  enzyme concentrations to fully describe the system, because the total concentration is constant. The adaptive landscape becomes a dome in  $(\mathbf{E}, J)$ , the shape of which depending only on pseudo-activities (Coton et al. 2021).

(iii) When there is regulation between *all* enzymes, the enzyme concentrations are 235 all linearly related, and fully determined by the driving variable  $\tau$  (Coton et al. 2021): the system has one degree of freedom. Thus the adaptive landscape is reduced to two dimensions,  $\tau$  and  $J$ . The landscape shape depends on the kind of co-regulation: flux increases indefinitely if enzymes are positively co-regulated, or reaches a maximum in case or negative co-regulation.

240 (iv) When there is competition plus co-regulation between *all* enzymes, the driving variable also fully describes the system, which so has one degree of freedom, and the adaptive landscape has two dimensions. However, the flux reaches a maximum whatever the kind of co-regulations, and the adaptive landscape is a section, oriented by the global co-regulation coefficients and the initial concentrations, of the dome of competition (Coton et al. 2021). Note that if the relative concentrations are initially equal to  $1/\mathbf{B}$  ( $B_i$  245 quantifies the effect on the total concentration of a variation of  $E_i$  before applying the constraint on  $E_{\text{tot}}$ ), the system is stuck at the theoretical equilibrium – an evolutionary black-hole, because at this point, the effect of a mutation is erased (see Coton et al. 2021). In this particular case, the system has zero degree of freedom and does not go outside 250 this point.

(v) Here we worked on co-regulation groups. When there are co-regulation groups without competition, enzyme concentrations are linearly related within groups, and are independent from enzymes in other groups. So each co-regulation group can fully be described by its driving variable  $\tau_q$ . Note that for singleton groups, as the enzyme is 255 independent from the others, its concentration replaces the driving variable and so represents a degree of freedom. Thus in this case, there is as many degrees of freedom as co-regulation groups. The adaptive landscape is reduced to  $p + 1$  dimensions – the number of co-regulation groups plus the flux. As in case of regulation between all enzymes, when there is only positively co-regulated or independent enzymes, the flux can increase

260 indefinitely, whereas flux has a maximum if there is at least two negatively co-regulated enzymes (figures 2, 3, 4 and S1).

(vi) When there is co-regulation groups with competition between all enzymes, the constant total concentration complicates the relationship between enzyme concentrations. From the previous cases, we know that the number of degrees of freedom is between 1 and  
 265  $n - 1$ . We might hypothesize that the number of df is  $p - 1$  – the number of co-regulation groups minus the relationship between enzyme concentrations due to constant total concentration –, but this is not coherent with the extreme case of co-regulation between all enzyme (indeed  $p = 1$  leading to zero df). Another hypothesis for the number of df is to consider each  $E^q$  (minus one due to the constant total concentration) and each  $\tau_q$  separately, leading to formula  $p + \sum_{q=1} \mathbf{1}_{\{\theta_q \neq 0\}} - 1$ . This formula is coherent with extreme cases,  
 270

but do not take account of the complex relationship between sums of enzyme concentrations of a group  $E^q$  and the driving variables  $\tau_q$  – the same relationship that prevents us to find the effective equilibrium. This case is not very intuitive, because it occurs in a multidimensional space and because the relationship between enzyme concentrations  
 275 depends on three different factors – co-regulation within groups, independence between groups, competition between all. Note that if a co-regulation group is initially as its theoretical equilibrium (equation S45), the system loses one df as previously. Whatever the degrees of freedom, the adaptive landscape is a projection in lower dimensions of the dome of competition, and its shape is also a dome (figures 5 and 6). This projection  
 280 depends on the number of co-regulation groups, and each new co-regulation between two previously independent groups diminish the dimension number of the projection.

The equilibrium distribution in case of competition with co-regulation groups depends on the kind and number of co-regulation groups. When all enzymes are co-regulated, the distribution seems random, but actually depends on two points, initial concentrations  $\mathbf{E}^0$   
 285 and theoretical equilibrium  $1/\mathbf{B}$ , that fixed the trajectories in the  $n$ -dimensional space of enzyme concentrations (figures 7B and S2). Each effective equilibrium is the maximum of the projection of the global of competition in this specific condition. With less restriction, *i.e.* with more co-regulation groups, the concentration trajectories can move more freely in the space. Thus, the ensemble of effective equilibrium for different initial concentrations  
 290 exhibits a more visible pattern (figures 7D, S3, 7F S5, 7H and S4).

Thus the adaptive landscape arises in lower dimensions in the multidimensional space  $(\mathbf{E}, J)$ , depending on the degrees of freedom that give the different constraints. Its shape is determined by the pseudo-activities, but also the co-regulation coefficients and the initial concentrations, according to the applied constraint.

## B.2 Démonstrations complémentaires

### a) Using subsums

Effective equilibrium corresponds to maximal flux against all enzyme concentrations. As each co-regulation groups depend on its driving variable, we can write :

$$\begin{aligned}
 J &= \frac{X}{\sum_{j=1}^n \frac{1}{A_j E_j}} \\
 &= \frac{X}{\sum_{q=1}^p \sum_{j \in \Phi_q} \frac{1}{A_j E_j}} \\
 &= \frac{X}{\sum_{q=1}^p \sum_{j \in \Phi_q} \frac{1}{A_j E^q e_j^q}}
 \end{aligned}$$

Because  $e_j^q = E_j / E^q$ , we get :

$$\begin{aligned}
 J &= \frac{X}{\sum_{q=1}^p \frac{1}{E^q} \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q}} \\
 &= \frac{X}{\sum_{q=1}^p \frac{1}{E^q} \sum_{j \in \Phi_q} \frac{1}{A_j (\tau^q (1/B_j - e_j^{q0}) + e_j^{q0})}}
 \end{aligned}$$

Each element  $\sum_{j \in \Phi_q} \frac{1}{A_j e_j^q}$  are independent from other co-regulation groups, and relies on the driving variable  $\tau^q$ . So maximising the flux in regard of each enzyme concentration  $E_j$  might be equivalent to minimise each subsum in regard of its driving variable. Let us note the subsums  $S^q$  such as :

$$S^q = \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} \quad (\text{B.1})$$

Finding the minimum of  $S^q$  is the same as cancel its derivative in regard of  $\tau^q$ .

$$\begin{aligned}
 \forall i \in \Phi_q, \quad \frac{\partial S^q}{\partial \tau^q} &= \frac{\partial}{\partial \tau^q} \left( \sum_{j \in \Phi_q} \frac{1}{A_j e_j^q} \right) \\
 &= \sum_{j \in \Phi_q} \frac{1}{A_j} \frac{\partial}{\partial \tau^q} \left( \frac{1}{e_j^q} \right) \\
 &= \sum_{j \in \Phi_q} \frac{1}{A_j} \frac{-\frac{\partial e_j^q}{\partial \tau^q}}{(e_j^q)^2}
 \end{aligned}$$



Because  $e_j^q = \tau^q(1/B_j - e_j^{q0}) + e_j^{q0}$ , we get :

$$\frac{\partial S^q}{\partial \tau^q} = - \sum_{j \in \Phi_q} \frac{1/B_j - e_j^{q0}}{A_j(\tau^q(1/B_j - e_j^{q0}) + e_j^{q0})^2}$$

Thus, we can find  $\tau^q$  at effective equilibrium by solving :

$$\boxed{\sum_{j \in \Phi_q} \frac{1/B_j - e_j^{q0}}{A_j(\tilde{\tau}^q(1/B_j - e_j^{q0}) + e_j^{q0})^2} = 0} \quad (\text{B.2})$$

with :

$$\tilde{e}_j^q = \tilde{\tau}^q(1/B_j - e_j^{q0}) + e_j^{q0} \quad (\text{B.3})$$

However,  $E^q$  are not independent in the expression of flux, and their complex relationships can prevent from reaching this solution within groups.

# Annexe C

## Chapitre 3 : démonstrations

### C.1 Expression de l'efficacité équivalente pour des enzymes en série

Dans cette annexe, je montre comment exprimer l'efficacité enzymatique équivalente pour une chaîne métabolique linéaire à  $n$  enzymes catalysant la transformation d'un substrat en produit *via* une série de réactions réversibles uni-moléculaires. Cette chaîne est illustrée dans la figure C.1 avec son équivalent électrique. Comme on le verra, la démonstration est similaire à celle de la Théorie du Contrôle Métabolique [Kacser et Burns 1973; Kacser *et al.* 1995].

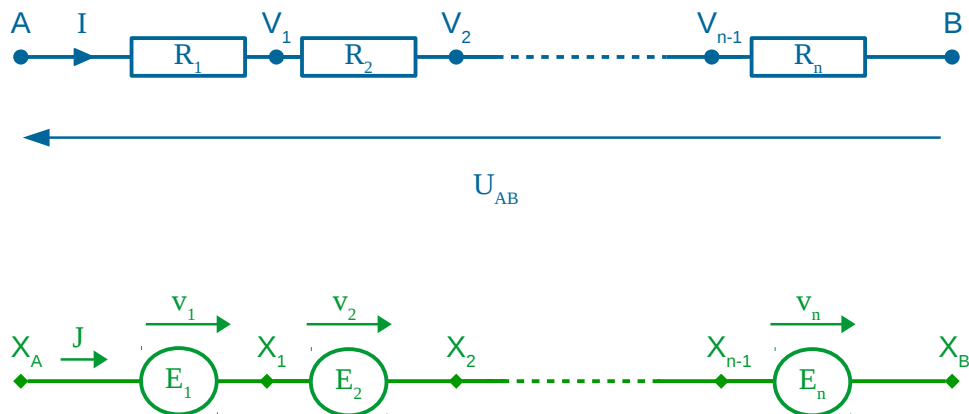


FIGURE C.1 – Circuit électrique (en haut en bleu) et voie métabolique (en bas en vert) pour  $n$  éléments en série.

L'objectif est d'obtenir la relation suivante entre le flux métabolique  $J$  à l'entrée de la voie et l'efficacité enzymatique équivalente  $F_E$  :

$$J = F_E \left( X_A - \frac{X_B}{K_E} \right) \quad (\text{C.1})$$

où  $X_A$  et  $X_B$  sont les concentrations des substrat et produit de la réaction. Pour chaque réaction  $i$ , la vitesse de réaction s'écrit :

$$v_i = F_i \left( X_{i-1} - \frac{X_i}{K_i} \right) \quad (\text{C.2})$$

où  $F_i$  est l'efficacité de l'enzyme  $i$ ,  $K_i$  la constante d'équilibre de la réaction considérée et  $X_i$  la concentration du métabolite  $X_i$ , produit de la réaction  $i$  et substrat de la réaction  $i + 1$ .

À l'état stationnaire, les vitesses de réaction sont constantes et identiques dans toute la chaîne métabolique, et sont donc égales au flux :

$$J = v_1 = v_2 = \dots = v_n \quad (\text{C.3})$$

En remplaçant les vitesses de réaction par leurs expressions, on a :

$$\begin{aligned} J &= v_1 = F_1(X_A - X_2/K_1) \\ J &= v_2 = F_2(X_2 - X_3/K_2) \\ &\vdots \\ J &= v_n = F_n(X_{n-1} - X_B/K_n) \end{aligned}$$

Pour éliminer les concentrations en métabolites inconnues, on passe les efficacités enzymatiques  $F_i$  à gauche puis on divise chaque expression par les constantes d'équilibre appropriées :

$$\begin{aligned} J/F_1 &= (X_A - X_2/K_1) \\ J/F_2 &= (X_2 - X_3/K_2) \quad \times 1/K_1 \\ &\vdots \\ J/F_n &= (X_{n-1} - X_B/K_n) \quad \times \frac{1}{K_1 K_2 \dots K_{n-1}} \end{aligned}$$

En effectuant la somme des équations du système, les concentrations en métabolites intermédiaires s'annulent et on obtient la relation suivante :

$$J \left( \frac{1}{F_1} + \frac{1}{F_2 K_1} + \dots + \frac{1}{F_n \prod_{j=1}^{n-1} K_j} \right) = X_A - \frac{X_B}{\prod_{j=1}^n K_j} \quad (\text{C.4})$$

On retrouve alors l'expression du flux en fonction des efficacités enzymatiques et des concen-

trations en substrat et produit :

$$J = \frac{X_A - \frac{X_B}{\prod_{j=1}^n K_j}}{\sum_{i=1}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}} \quad (\text{C.5})$$

En la comparant à l'équation C.8, on peut en déduire que l'efficacité d'une enzyme équivalente est :

$$F_E = \frac{1}{\sum_{i=1}^n \frac{1}{F_i \prod_{j=1}^{i-1} K_j}} \quad (\text{C.6})$$

Et que la constante d'équilibre équivalente est :

$$K_E = \prod_{j=1}^n K_j \quad (\text{C.7})$$

## C.2 Expression de l'efficacité équivalente pour des enzymes en parallèle

Dans cette annexe, je montre comment exprimer l'efficacité enzymatique équivalente pour un réseau métabolique à  $n$  enzymes en parallèle catalysant la même réaction réversible unimoléculaire. Ce réseau est illustré dans la figure C.2 avec son équivalent électrique.

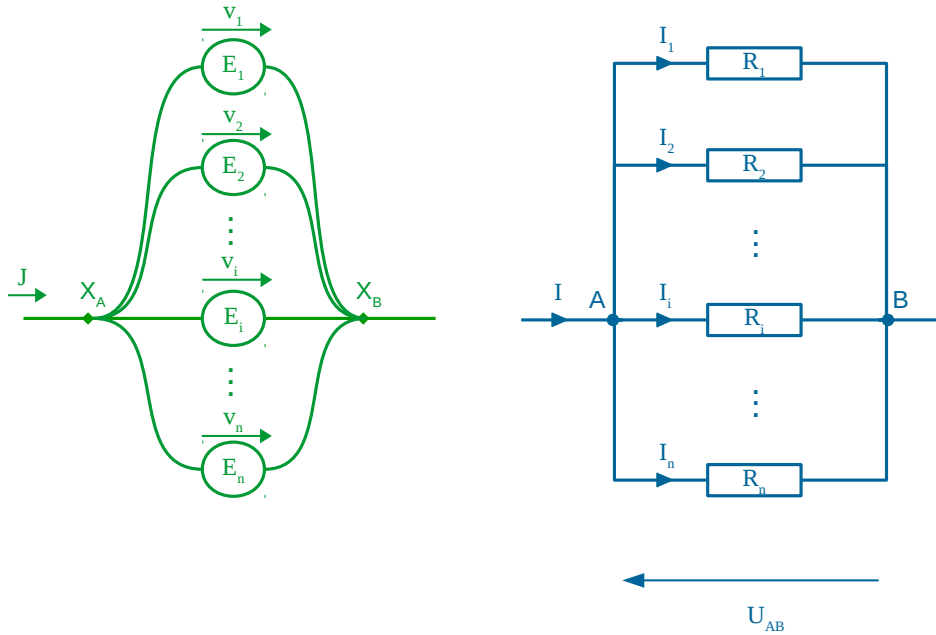


FIGURE C.2 – Réseau métabolique (à gauche en vert) et circuit électrique (à droite en bleu) pour  $n$  éléments en parallèle. Ce schéma est purement théorique et sert d'appui à la démonstration.

L'objectif est d'obtenir la relation suivante entre le flux métabolique  $J$  à l'entrée du réseau et l'efficacité enzymatique équivalente  $F_E$  :

$$J = F_E \left( X_A - \frac{X_B}{K_E} \right) \quad (\text{C.8})$$

où  $X_A$  et  $X_B$  sont les concentrations des substrat et produit de la réaction. Pour chaque réaction  $i$ , la vitesse s'écrit :

$$v_i = F_i \left( X_A - \frac{X_B}{K_i} \right) \quad (\text{C.9})$$

où  $F_i$  est l'efficacité de l'enzyme  $i$  et  $K_i$  la constante d'équilibre de la réaction considérée.

Comme toutes les réactions transforment le même substrat  $X_A$  en même produit  $X_B$ , les réactions sont thermodynamiquement équivalentes, et les constantes d'équilibre sont donc égales et on a donc :

$$\forall i \quad K_i = K_E \quad (\text{C.10})$$

Si l'on considère la vitesse de production et de consommation de  $X_A$  (équivalent de la loi

des noeuds), on a :

$$J = v_1 + v_2 + \dots + v_n = \sum_{i=1}^n v_i \quad (\text{C.11})$$

Ce qui donne :

$$J = \sum_{i=1}^n F_i \left( X_A - \frac{X_B}{K_E} \right) \quad (\text{C.12})$$

En la comparant à l'équation C.8, on obtient l'efficacité équivalente du réseau :

$$F_E = \sum_{i=1}^n F_i \quad (\text{C.13})$$

### C.3 Expression du flux pour un réseau métabolique simple

Dans cette annexe, je présente un exemple de résolution d'un circuit simple à cinq enzymes avec une boucle de dérivation (figure C.3). Cette résolution se base sur les efficacités enzymatiques équivalentes.

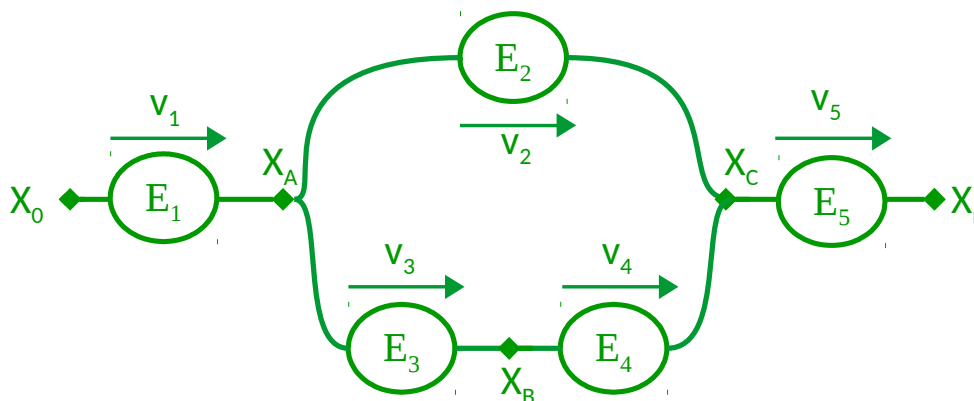


FIGURE C.3 – Exemple de réseau métabolique simple à cinq enzymes avec une boucle de dérivation. Cet exemple théorique est utilisé afin d'illustrer la méthode de résolution.

Les enzymes 3 et 4 sont en série, nous avons donc

$$\frac{1}{F_{34}} = \frac{1}{F_3} + \frac{1}{K_3 F_4}$$

$E_2$  est en parallèle avec les deux enzymes précédentes entre A et C, ce qui donne :

$$F_{AC} = F_2 + F_{34}$$

avec

$$K_{AC} = K_2 = K_3 K_4$$

Les autres enzymes sont en série, ce qui donne :

$$\frac{1}{F_E} = \frac{1}{F_1} + \frac{1}{K_1 F_{AC}} + \frac{1}{K_1 K_{AC} F_5}$$

avec comme constante d'équilibre équivalente :  $K_E = K_1 K_{AC} K_5$

Autrement dit :

$$F_E = \frac{1}{\frac{1}{F_1} + \frac{1}{K_1 \left( F_2 + \frac{1}{\frac{1}{F_3} + \frac{1}{K_3 F_4}} \right)} + \frac{1}{K_1 K_{AC} F_5}} \quad (\text{C.14})$$

Le flux s'écrira alors :

$$J = F_E (X_0 - X_n / K_E)$$

NB : si  $F_2$  disparaît (inactivité et/ou absence), alors le circuit est uniquement en série, en passant par B.  $F_2$  est alors égal à 0, et  $K_E$  et  $F_E$  (équation C.14) se simplifient en :

$$K_E = K_1 K_3 K_4 K_5$$

$$\text{et } F_E = \frac{1}{\frac{1}{F_1} + \frac{1}{K_1 \left( \frac{1}{F_3} + \frac{1}{K_3 F_4} \right)} + \frac{1}{K_1 K_3 K_4 F_5}}$$

## C.4 Expression de l'efficacité de l'enzyme équivalente dans le cas du pont de Wheatstone

Dans la cas du pont de Wheatstone, constitué par une boucle de dérivation où les deux branches sont reliées par un dipôle (figure C.4), la relation entre l'efficacité enzymatique équivalente et toutes les efficacités enzymatiques n'est pas trouvable aussi simplement que précédemment. J'ai donc adapté la démonstration de [Kagan \[2015\]](#) en introduisant les constantes d'équilibre des réactions.

La vitesse de production du métabolite de sortie  $X_4$ , qui est le flux, vaut :

$$\frac{dX_4}{dt} = J = v_2 + v_3$$

En remplaçant  $J$ ,  $v_2$  et  $v_3$  par leurs expressions, on obtient :

$$J = F_E (X_1 - X_4 / K_E) = F_2 (X_2 - X_4 / K_4) + F_5 (X_3 - X_4 / K_4)$$

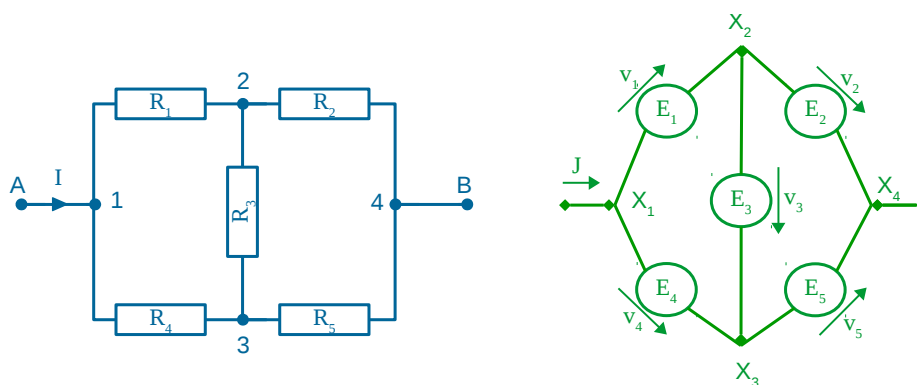


FIGURE C.4 – **Pont de Wheatstone : analogie métabolisme-électricité.** En bleu à gauche le circuit électrique, et en vert à droite son pendant métabolique. Les deux branches entre A et B sont reliées par le dipôle  $R_3$ . Les résistances ne sont alors ni en série, ni en dérivation, et le pont n'est pas réductible simplement.

En supposant que la concentration  $X_4$  du métabolite de sortie est nulle, le flux se réécrit :

$$J = X_2 F_2 + X_3 F_5 = F_E X_1 \quad (\text{C.15})$$

mais la constante d'équilibre équivalente  $K_E$  n'apparaît plus dans l'équation, nous empêchant de la déterminer.

Pour trouver les valeurs de  $X_2$  et  $X_3$ , on applique la loi des nœuds aux points 2 et 3 :

$$\begin{aligned} \frac{dX_2}{dt} = 0 &= -v_3 - v_2 + v_1 \\ &= -F_3(X_2 - X_3/K_3) - F_2(X_2 - X_4/K_2) + F_1(X_1 - X_2/K_1) \\ &= X_1 F_1 - X_2(E1/K_1 + F_2 + F_3) + X_3 F_3/K_3 + X_4 F_2/K_2 \end{aligned}$$

$$\begin{aligned} \frac{dX_3}{dt} = 0 &= v_4 + v_3 - v_5 \\ &= F_4(X_1 - X_3/K_4) + F_3(X_2 - X_3/K_3) - F_5(X_3 - X_4/K_5) \\ &= X_1 F_4 + X_2 F_3 - X_3(F_3/K_3 + F_4/K_4 + F_5) + X_4 F_5/K_5 \end{aligned}$$

Comme  $X_4$  est supposée nulle, on peut exprimer les valeurs des concentrations en métabolites intermédiaires  $X_2$  et  $X_3$  :

$$X_2 c_2 = X_1 F_1 + X_3 F_3/K_3 \quad (\text{C.16})$$

$$X_3 c_3 = X_1 F_4 + X_2 F_3 \quad (\text{C.17})$$

où  $c_2 = F_1/K_1 + F_2 + F_3$  et  $c_3 = F_3/K_3 + F_4/K_4 + F_5$ . On implémente la valeur de  $X_3$  (C.17)



dans celle de  $X_2$  (C.16), ce qui nous donne :

$$\begin{aligned} X_2 c_2 &= X_1 F_1 + \frac{X_1 F_4 + X_2 F_3}{c_3} F_3 / K_3 \\ X_2 \left( c_2 - \frac{F_3^2}{c_3 K_3} \right) &= X_1 F_1 + \frac{X_1 F_3 F_4}{c_3 K_3} \\ X_2 (c_2 c_3 K_3 - F_3^2) &= X_1 (F_1 c_3 K_3 + F_3 F_4) \\ X_2 &= X_1 \frac{F_1 c_3 K_3 + F_3 F_4}{c_2 c_3 K_3 - F_3^2} \end{aligned}$$

Procédons de la même manière pour  $X_3$ , et implémentons C.16 dans C.17 :

$$\begin{aligned} X_3 c_3 &= X_1 F_4 + X_2 F_3 \\ &= X_1 F_4 + \frac{X_1 F_1 + X_3 F_3 / K_3}{c_2} F_3 \\ X_3 \left( c_3 - \frac{F_3^2}{c_2 K_3} \right) &= X_1 \left( F_4 + \frac{F_1 F_3}{c_2} \right) \\ X_3 (c_3 c_2 K_3 - F_3^2) &= X_1 (F_4 c_2 K_3 + F_1 F_3 K_3) \\ X_3 &= X_1 \frac{F_4 c_2 K_3 + F_1 F_3 K_3}{c_3 c_2 K_3 - F_3^2} \end{aligned}$$

Écrivons  $Den = c_2 c_3 K_3 - F_3^2$  le dénominateur commun à  $X_2$  et  $X_3$  :

$$\begin{aligned} Den &= c_2 c_3 K_3 - F_3^2 \\ &= K_3 (F_1 / K_1 + F_2 + F_3) (F_3 / K_3 + F_4 / K_4 + F_5) - F_3^2 \\ &= K_3 \left( \frac{F_1 F_3}{K_1 K_3} + \frac{F_1 F_4}{K_1 K_4} + \frac{F_1 F_5}{K_1} + \frac{F_2 F_3}{K_3} + \frac{F_2 F_4}{K_4} + F_2 F_5 + \frac{F_3^2}{K_3} + \frac{F_3 F_4}{K_4} + F_3 F_5 \right) - F_3^2 \\ &= F_3 (F_1 / K_1 + F_2 + F_4 K_3 / K_4 + F_5 K_3) + K_3 (F_1 / K_1 + F_2) (F_4 / K_4 + F_5) \end{aligned}$$

En remplaçant les expressions de  $X_2$  et  $X_3$  dans l'équation du flux C.15, on obtient pour l'efficacité de l'enzyme équivalente :

$$\begin{aligned} F_E &= \frac{X_2 F_2}{X_1} + \frac{X_3 F_5}{X_1} \\ &= \frac{F_2 (F_1 c_3 K_3 + F_3 F_4) + F_5 (F_4 c_2 K_3 + F_1 F_3 K_3)}{Den} \end{aligned}$$

En remplaçant  $c_2$ ,  $c_3$  et  $Den$  par leurs expressions respectives, on obtient l'équation de l'efficacité équivalente :

$$F_E = \frac{F_1 F_2 K_3 (F_3 / K_3 + F_4 / K_4 + F_5) + F_2 F_3 F_4 + F_1 F_3 F_5 K_3 + F_4 F_5 K_3 (F_1 / K_1 + F_2 + F_3)}{F_3 (F_1 / K_1 + F_2 + F_4 K_3 / K_4 + F_5 K_3) + K_3 (F_1 / K_1 + F_2) (F_4 / K_4 + F_5)} \quad (\text{C.18})$$

## C.5 Démonstration de la relation concave entre une conductance et l'intensité

La démonstration de la concavité a été faite par Marianyela Petrizzelli, chercheuse post-doctorale à l'Institut Curie, qui m'a autorisée à la reproduire ici.

### C.5.1 Introduction

#### Similarity between metabolic networks and electric circuits.

The total flux  $J$  and the enzymatic activity  $A$  and/or the concentration of an enzyme  $[E]$  are thought to be equivalent to the current intensity ( $I \equiv J$ ) and the conductance in an electric circuit ( $1/R = \sigma \equiv A[E]$ ), respectively. Thus a circuit is like a metabolic network.

Here, we demonstrate the concave relationship between the intensity  $I$  and conductance of a resistor (all the other fixed to be constant) in resistive circuit like metabolic network.

**Definition of concavity.** A real valued function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is said to be concave if,  $\forall x, y \in \mathbb{R}$  and  $\forall \alpha \in [0, 1]$ ,

$$f((1 - \alpha)x + \alpha y) \geq (1 - \alpha)f(x) + \alpha f(y) \quad (\text{C.19})$$

### C.5.2 Wheatstone bridge like metabolic network

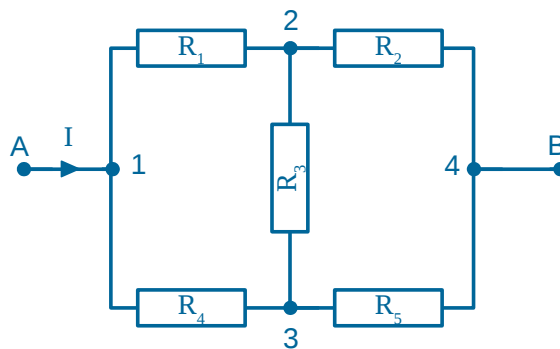


FIGURE C.5 – Electrical wheatstone bridge.

The output intensity, that is proportional to the equivalent conductance  $\sigma_{eq}$  in a circuit, is a function of the conductances in the network (eq 12 in Kagan [2015]):

$$\sigma_{eq} = \frac{\sigma_1 \sigma_2 (\sigma_3 + \sigma_4 + \sigma_5) + \sigma_2 \sigma_3 \sigma_4 + \sigma_1 \sigma_3 \sigma_5 + \sigma_4 \sigma_5 (\sigma_1 + \sigma_2 + \sigma_3)}{\sigma_3 (\sigma_1 + \sigma_2 + \sigma_4 + \sigma_5) + (\sigma_1 + \sigma_2) (\sigma_4 + \sigma_5)} \quad (\text{C.20})$$

**a) Concavity**

We want to prove that, fixed the conductance of all resistors in the network but one, let's say the  $i$ -th, the inequality

$$\sigma_{eq}((1 - \alpha)\sigma_i^x + \alpha\sigma_i^y) \geq (1 - \alpha)\sigma_{eq}(\sigma_i^x) + \alpha\sigma_{eq}(\sigma_i^y) \quad (\text{C.21})$$

is still valid.

It is easy to see that  $\forall \sigma_i, \sigma_{eq}$  can be expressed as:

$$\sigma_{eq}(\sigma_i) = \frac{a\sigma_i + b}{c\sigma_i + d} \quad (\text{C.22})$$

Substituting eq.C.22 in eq.C.21 leads to:

$$\frac{((1 - \alpha)\sigma_i^x + \alpha\sigma_i^y)a + b}{((1 - \alpha)\sigma_i^x + \alpha\sigma_i^y)c + d} \geq (1 - \alpha)\frac{\sigma_i^x a + b}{\sigma_i^x c + d} + \alpha\frac{\sigma_i^y a + b}{\sigma_i^y c + d} \quad (\text{C.23})$$

Taking the least common denominator and denoting  $\sigma^* = ((1 - \alpha)\sigma_i^x + \alpha\sigma_i^y)$ ,

$$\begin{aligned} & \frac{(\sigma^* a + b)(\sigma_i^x c + d)(\sigma_i^y c + d)}{(\sigma^* c + d)(\sigma_i^x c + d)(\sigma_i^y c + d)} \\ & \geq \frac{(1 - \alpha)(\sigma_i^x a + b)(\sigma^* c + d)(\sigma_i^y c + d) + \alpha(\sigma_i^y a + b)(\sigma^* c + d)(\sigma_i^x c + d)}{(\sigma^* c + d)(\sigma_i^x c + d)(\sigma_i^y c + d)} \end{aligned} \quad (\text{C.24})$$

Given that  $\forall_i \sigma_i \geq 0$ , and  $a, b, c, d \geq 0$ , eq. C.24 is satisfied if and only if

$$\begin{aligned} & (\sigma^* a + b)(\sigma_i^x c + d)(\sigma_i^y c + d) \\ & \geq (1 - \alpha)(\sigma_i^x a + b)(\sigma^* c + d)(\sigma_i^y c + d) + \alpha(\sigma_i^y a + b)(\sigma^* c + d)(\sigma_i^x c + d) \end{aligned} \quad (\text{C.25})$$

Thus,

$$\begin{aligned} & (\sigma^* a + b)(\sigma_i^x c + d)(\sigma_i^y c + d) \\ & \geq (\sigma^* c + d)(ad\sigma^* + \sigma_i^x \sigma_i^y ac + ((1 - \alpha)\sigma_i^y + \alpha\sigma_i^x)bc + bd) \end{aligned} \quad (\text{C.26})$$

$$\begin{aligned} & acd(\sigma^* - \sigma_i^y)(\sigma_i^x - \sigma^*) + bcd(\sigma_i^x + \sigma_i^y - \sigma^* - ((1 - \alpha)\sigma_i^y + \alpha\sigma_i^x)) \\ & \quad + bc^2(\sigma_i^x \sigma_i^y - \sigma^*((1 - \alpha)\sigma_i^y + \alpha\sigma_i^x)) \geq 0 \end{aligned} \quad (\text{C.27})$$

Noting that

$$\begin{aligned} & (\sigma^* - \sigma_i^y)(\sigma_i^x - \sigma^*) = (\sqrt{\alpha(1 - \alpha)}\sigma_i^x - \sqrt{\alpha(1 - \alpha)}\sigma_i^y)^2 \\ & \quad \sigma_i^x + \sigma_i^y - \sigma^* - ((1 - \alpha)\sigma_i^y + \alpha\sigma_i^x) = 0 \end{aligned}$$

and

$$\sigma_i^x \sigma_i^y - \sigma^* ((1 - \alpha) \sigma_i^y + \alpha \sigma_i^x) = -(\sqrt{\alpha(1 - \alpha)} \sigma_i^x - \sqrt{\alpha(1 - \alpha)} \sigma_i^y)^2$$

eq.C.27 simplifies to

$$ad - bc \geq 0 \quad (\text{C.28})$$

We need to verify this inequality  $\forall \sigma_i$ .

**i. Case 1.** For  $\sigma_i = \sigma_1$ ,

$$\sigma_{eq}^* = \frac{\sigma_1 a + b}{c \sigma_1 + d} \quad (\text{C.29})$$

with

$$a = \sigma_2(\sigma_3 + \sigma_4 + \sigma_5) + \sigma_3\sigma_5 + \sigma_4\sigma_5$$

$$b = \sigma_4\sigma_5(\sigma_2 + \sigma_3) + \sigma_2\sigma_3\sigma_4$$

$$c = \sigma_3 + \sigma_4 + \sigma_5$$

$$d = \sigma_3(\sigma_2 + \sigma_4 + \sigma_5) + \sigma_2(\sigma_4 + \sigma_5)$$

Thus, eq. C.28 becomes

$$\begin{aligned} & (\sigma_2(\sigma_3 + \sigma_4 + \sigma_5) + \sigma_3\sigma_5 + \sigma_4\sigma_5)(\sigma_3(\sigma_2 + \sigma_4 + \sigma_5) + \sigma_2(\sigma_4 + \sigma_5)) \\ & - (\sigma_4\sigma_5(\sigma_2 + \sigma_3) + \sigma_2\sigma_3\sigma_4)(\sigma_3 + \sigma_4 + \sigma_5) \geq 0 \quad (\text{C.30}) \end{aligned}$$

After simplification and rearrangements it leads to

$$(\sigma_2(\sigma_3 + \sigma_4 + \sigma_5) + \sigma_3\sigma_5)(\sigma_2\sigma_3 + \sigma_2\sigma_4 + \sigma_2\sigma_5 + \sigma_3\sigma_5) \geq 0 \quad (\text{C.31})$$

That's always verified since  $\forall i, \sigma_i \in \mathbb{R}^+$ .

**ii. Case 2.** For  $\sigma_i = \sigma_2$ ,

$$a = \sigma_1(\sigma_3 + \sigma_4 + \sigma_5) + \sigma_3\sigma_4 + \sigma_4\sigma_5$$

$$b = \sigma_4\sigma_5(\sigma_1 + \sigma_3) + \sigma_1\sigma_3\sigma_5$$

$$c = \sigma_3 + \sigma_4 + \sigma_5$$

$$d = \sigma_3(\sigma_1 + \sigma_4 + \sigma_5) + \sigma_1(\sigma_4 + \sigma_5)$$

and eq. C.28 is verified.

iii. **Case 3.** For  $\sigma_i = \sigma_3$ ,

$$a = (\sigma_1 + \sigma_4)(\sigma_2 + \sigma_5)$$

$$b = \sigma_1\sigma_2(\sigma_4 + \sigma_5) + \sigma_4\sigma_5(\sigma_1 + \sigma_2)$$

$$c = \sigma_1 + \sigma_2 + \sigma_4 + \sigma_5$$

$$d = (\sigma_1 + \sigma_2)(\sigma_4 + \sigma_5)$$

Substituting these values in eq.C.28 leads to

$$(\sigma_1\sigma_5 - \sigma_2\sigma_4)^2 \geq 0 \quad (\text{C.32})$$

that is always verified.

**Cases 4 and 5.** For  $\sigma_i = \sigma_4$  and  $\sigma_i = \sigma_5$ , by symmetry, cases 4 and 5 are equivalent to cases 1 and 2, respectively. In fact, by noting  $\sigma_4$  as  $\sigma_1$ ,  $\sigma_5$  as  $\sigma_2$ , and viceversa, it is easy to retrieve the same values for the coefficients  $a$ ,  $b$ ,  $c$  and  $d$  on the respective sections. Thus eq. C.28 is satisfy.

### b) Partial derivative of the equivalent conductance

$$\frac{\partial \sigma_{eq}}{\partial \sigma_i} = \frac{ad - bc}{(c\sigma_i + d)^2} \geq 0 \quad (\text{C.33})$$

We showed that the denominator is always greater than or equal to 0. Thus, the equivalent conductance is a monotonically increasing function of  $\sigma_i$ .

### c) Bounds

$$\lim_{\sigma_i \rightarrow 0} \sigma_{eq} = \frac{b}{d} \quad (\text{C.34})$$

To find the limit at infinity, we simply rewrite C.22:

$$\sigma_{eq}(\sigma_i) = \frac{a + b/\sigma_i}{c + d\sigma_i}$$

which gives the limit:

$$\lim_{\sigma_i \rightarrow +\infty} \sigma_{eq} = \frac{a}{c} \quad (\text{C.35})$$

Thus intensity is monotonically increasing, bounded and concave function of the conductances in a Wheatstone bridge like metabolic network.

In the particular case where  $ad - cb = 0$ , then  $\sigma_{eq}$  becomes constant against  $\sigma_i$ . It is the case for  $\sigma_3$  when  $\sigma_1\sigma_5 = \sigma_2\sigma_4$ .

### C.5.3 Generic non-simplifiable circuit like metabolic network

The junctions equations for all the node potentials in a generic circuit of  $n$  nodes ( $n > 2$ ) can be expressed as:

$$\Sigma \mathbf{V} = \mathbf{I} \quad (\text{C.36})$$

where  $\mathbf{V} = (V_1, V_2, \dots, V_n)^T$  is the vector of nodes potential,  $\mathbf{I} = (I_{out}, 0, \dots, 0, I_{in})^T$  is the vector of current flows and  $\Sigma$  the conductance matrix in the electric circuit,

$$\Sigma = \begin{pmatrix} c_1 & -\sigma_{1,2} & -\sigma_{1,3} & -\sigma_{1,4} & \dots & -\sigma_{1,n} \\ -\sigma_{2,1} & c_2 & -\sigma_{2,3} & -\sigma_{2,4} & \dots & -\sigma_{2,n} \\ -\sigma_{3,1} & -\sigma_{3,2} & c_3 & -\sigma_{3,4} & \dots & -\sigma_{3,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{n,1} & -\sigma_{n,2} & -\sigma_{n,3} & -\sigma_{n,4} & \dots & c_n \end{pmatrix} \quad (\text{C.37})$$

with  $c_i = \sum_{j \neq i} \sigma_{i,j}$ . Note that  $\sigma_{ij} = 0$  if the nodes  $i$  and  $j$  are not connected by a resistor.

Kagan [2015] derived the equivalence conductance formula for a generic non-simplifiable circuit and, by posing  $V_1 = \varepsilon$  and  $V_n = 0$ , showed that

$$\sigma_{eq} = \frac{\det \Sigma'}{\det \Sigma''} \quad (\text{C.38})$$

where  $\Sigma'$  is the upper left sub-matrix  $(n-1) \times (n-1)$  of the conductance matrix  $\Sigma$

$$\Sigma' = \begin{pmatrix} c_1 & -\sigma_{1,2} & -\sigma_{1,3} & -\sigma_{1,4} & \dots & -\sigma_{1,n-1} \\ -\sigma_{2,1} & c_2 & -\sigma_{2,3} & -\sigma_{2,4} & \dots & -\sigma_{2,n-1} \\ -\sigma_{3,1} & -\sigma_{3,2} & c_3 & -\sigma_{3,4} & \dots & -\sigma_{3,n-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{n-1,1} & -\sigma_{n-1,2} & -\sigma_{n-1,3} & -\sigma_{n-1,4} & \dots & c_{n-1} \end{pmatrix} \quad (\text{C.39})$$

and  $\Sigma''$  is the lower right sub-matrix  $(n-2) \times (n-2)$  of  $\Sigma'$

$$\Sigma'' = \begin{pmatrix} c_2 & -\sigma_{2,3} & -\sigma_{2,4} & \dots & -\sigma_{2,n-1} \\ -\sigma_{3,2} & c_3 & -\sigma_{3,4} & \dots & -\sigma_{3,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{n-1,2} & -\sigma_{n-1,3} & -\sigma_{n-1,4} & \dots & c_{n-1} \end{pmatrix} \quad (\text{C.40})$$

Kagan [2015] showed that each term in  $\det \Sigma'$  and  $\det \Sigma''$  enter with a plus. Thus, the equivalence conductance is a ratio of two polynomials of degree  $(n-1)$  and  $(n-2)$ , respectively, with only positive terms.

Again, it is possible to express  $\sigma_{eq}$  as in eq. C.22  $\forall \sigma_{i,j}, i.e$

$$\sigma_{eq} = \frac{A\sigma_{i,j} + B}{C\sigma_{i,j} + D} \quad (\text{C.41})$$

with  $A, B, C, D$  being non negative and dependent on all edge conductance other than the one between the  $i$ -th and the  $j$ -th nodes.

Thus, as eq. C.28, the inequality

$$AD - BC \geq 0 \quad (\text{C.42})$$

must be satisfied to verify the concave relationship between the conductance at a node, fixed all others, and the equivalent conductance.

Without loss of generality, we can focus on  $\sigma_{1,2}$ . Setting  $c'_1 = c_1 - \sigma_{1,2}$ ,  $c'_2 = c_2 - \sigma_{1,2}$  and  $n - 1 = k$

$$A = \begin{vmatrix} c'_1 + c'_2 & -(\sigma_{1,3} + \sigma_{2,3}) & \dots & -(\sigma_{1,k} + \sigma_{2,k}) \\ -(\sigma_{3,1} + \sigma_{3,2}) & c_3 & \dots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots \\ -(\sigma_{k,1} + \sigma_{k,2}) & -\sigma_{k,3} & \dots & c_k \end{vmatrix} \quad (\text{C.43})$$

$$B = \begin{vmatrix} c'_1 & c'_1 & -\sigma_{1,3} & \dots & -\sigma_{1,k} \\ c'_1 & c'_1 + c'_2 & -(\sigma_{1,3} + \sigma_{2,3}) & \dots & -(\sigma_{1,k} + \sigma_{2,k}) \\ -\sigma_{3,1} & -(\sigma_{3,1} + \sigma_{3,2}) & c_3 & \dots & -\sigma_{3,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{k,1} & -(\sigma_{k,1} + \sigma_{k,2}) & -\sigma_{k,3} & \dots & c_k \end{vmatrix} \quad (\text{C.44})$$

$$C = \begin{vmatrix} c_3 & \dots & -\sigma_{3,k} \\ \vdots & \ddots & \vdots \\ -\sigma_{k,3} & \dots & -c_k \end{vmatrix} \quad (\text{C.45})$$

$$D = \begin{vmatrix} c'_2 & -\sigma_{2,3} & \dots & -\sigma_{2,k} \\ -\sigma_{3,2} & c_3 & \dots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{k,2} & -\sigma_{k,3} & \dots & c_k \end{vmatrix} \quad (\text{C.46})$$

Now, we need to verify eq. C.42 to demonstrate the concave relationship between the conductance at a node, fixed all others, and the equivalent conductance. To this end let us introduce *Jacobi's* theorem first.

**Theorem** (Jacobi). Let  $M$  denote the determinant of a matrix  $M = \|m_{ij}\|_1^n$ ,  $M^c$  the determinant of the matrix of its cofactors  $M^c = \|M_{ij}\|_1^n$ ,  $1 \leq p < n$ ,  $\sigma_{(j_1 \dots j_n)}^{(i_1 \dots i_n)}$  an arbitrary

permutation of the  $n$  rows and columns of  $M$ , then

$$\begin{vmatrix} M_{i_1 j_1} & \cdots & M_{i_1 j_p} \\ M_{i_2 j_1} & \cdots & M_{i_2 j_p} \\ \vdots & \ddots & \vdots \\ M_{i_p j_1} & \cdots & M_{i_p j_p} \end{vmatrix} = (-1)^\sigma \begin{vmatrix} m_{i_{p+1} j_{p+1}} & \cdots & m_{i_{p+1} j_n} \\ m_{i_{p+2} j_{p+1}} & \cdots & m_{i_{p+2} j_n} \\ \vdots & \ddots & \vdots \\ m_{i_n j_{p+1}} & \cdots & m_{i_n j_n} \end{vmatrix} \cdot M^{p-1} \quad (\text{C.47})$$

Now, remark that  $C$  is a minor of  $B$  obtained by deleting the first two rows and columns of  $B$ . Using Jacobi's Theorem, with  $p = 2$ ,  $n = k$ ,  $i_l = l$  and  $j_l = l \forall l \in 1, \dots, k$ , we can express  $BC$  as

$$\begin{vmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{vmatrix} = (-1)^\sigma CB^{2-1} = CB \quad (\text{C.48})$$

since  $\sigma = 0$  (there are no permutations).

Using the sum property of the determinants, we can rewrite  $B$  as

$$B = \begin{vmatrix} c'_1 & 0 & -\sigma_{1,3} & \cdots & -\sigma_{1,k} \\ 0 & c'_2 & -\sigma_{2,3} & \cdots & -\sigma_{2,k} \\ -\sigma_{3,1} & -\sigma_{3,2} & c_3 & \cdots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots & \\ -\sigma_{k,1} & -\sigma_{k,2} & -\sigma_{k,3} & \cdots & c_k \end{vmatrix} \quad (\text{C.49})$$

and give an explicit expression for the co-factors in eq. C.48 :

$$B_{11} = D \quad (\text{C.50})$$

$$B_{12} = - \begin{vmatrix} 0 & -\sigma_{2,3} & \cdots & -\sigma_{2,k} \\ -\sigma_{3,1} & c_3 & \cdots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{k,1} & -\sigma_{k,3} & \cdots & c_k \end{vmatrix} = -B^{**} \quad (\text{C.51})$$

$$B_{21} = - \begin{vmatrix} 0 & -\sigma_{1,3} & \cdots & -\sigma_{1,k} \\ -\sigma_{3,2} & c_3 & \cdots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{k,2} & -\sigma_{k,3} & \cdots & c_k \end{vmatrix} = -B^* \quad (\text{C.52})$$

$$B_{22} = \begin{vmatrix} c'_1 & -\sigma_{1,3} & \cdots & -\sigma_{1,k} \\ -\sigma_{3,1} & c_3 & \cdots & -\sigma_{3,k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{k,1} & -\sigma_{k,3} & \cdots & c_k \end{vmatrix} \quad (\text{C.53})$$



Thus,

$$BC = \begin{vmatrix} D & -B^{**} \\ -B^* & B_{22} \end{vmatrix} = DB_{22} - (B^*)^2 \quad (\text{C.54})$$

where  $B^* = B^{**}$  since the respective matrices are one the transpose of the other.

On the other hand, we can re-write  $A$  as a function of  $D$ ,  $B^*$  and  $B_{22}$ . Using the properties of determinants :

$$A = 2B^* + D + B_{22} \quad (\text{C.55})$$

We can now substitute all these informations in eq. C.42 :

$$AD - BC = (2B^* + D + B_{22})D - (B_{22}D - (B^*)^2) = 2B^*D + D^2 + (B^*)^2 = (D + B^*)^2$$

that's always equal or greater than zero, thus eq. C.42 is always satisfied.

## C.6 Analogie entre métabolisme et chaîne de production

Dans la suite des analogies de flux de matière, j'ai proposé le système « chaîne de production », à but de médiation scientifique, car il est plus accessible au grand public que l'électricité ou l'hydraulique pour faire comprendre le métabolisme. J'ai pris l'exemple d'une usine de fabrication de fromages. Le système se présente comme suit : pour chaque chaîne de production, nous avons une suite de machines (par exemple, fermenteuse, mouleuse et emballeuse) qui transforme le lait en fromage emballé. Chaque machine représente une enzyme, et le nombre de machines la concentration de cette enzyme ; la performance des machines correspond à l'activité enzymatique ; le flux métabolique est la vitesse de production ; les métabolites correspondent aux différents états de la matière entre chaque étape de transformation. La force motrice est plus difficile à interpréter, car elle dépend de la façon dont la chaîne de production a été automatisée, tandis que pour les systèmes physiques, c'est uniquement une différence d'énergie qui permet le déplacement de la matière, et cette énergie n'a pas d'équivalent pour le système « chaîne de production ». Le plus simple est de considérer que l'usine est programmée pour avoir toujours un stock défini de produits finis et de matières premières, que l'on retire/ajoute régulièrement.

Cette analogie est utile pour faire comprendre notre modèle d'évolution des concentrations d'enzymes. La sélection pour maximiser le flux revient à optimiser le nombre de machines dans l'usine pour avoir la plus grande vitesse de production possible. Les contraintes appliquées aux concentrations d'enzyme peuvent aussi être expliquées dans cette analogie : l'indépendance entre enzymes correspond à un budget illimité pour acheter des machines ; la compétition pour les ressources correspond à un budget fixe ; la co-régulation correspond à un achat des machines par lots. La limite se place dans la manière dont le système est optimisé : les mutations sont totalement aléatoires, dans leur cible et leur taille, tandis que pour une usine, c'est un choix délibéré d'acheter telle ou telle machine, bien que cela puisse se faire par tâtonnement.

L'analogie entre cellule et usine de production a par ailleurs été utilisée pour expliquer l'optimisation des temps de présence des métabolites (les « stocks ») [Cascante *et al.* 1996].

TABLE C.1 – **Analogie entre métabolisme et chaîne de production de fromages.** Ce tableau est complémentaire du tableau 3.2.

Métabolisme	Chaîne de production
Métabolites	Matière transformée
Flux métabolique $J$	Vitesse de production
Potentiel chimique $\mu$	$\emptyset$
Écart à l'équilibre	Différentiel entre matières premières et produits finis
Efficacité enzymatique	Performance des machines
Concentration d'enzymes	Nombre de machines
Ressources limitées	Budget fixe
Co-régulations	Achats par lots
Évolution par sélection	Optimisation



# Annexe D

## Évolution des enzymes en parallèle : démonstrations

Cette annexe contient les démonstrations sur les résultats préliminaires de l'évolution des enzymes parallèles présentés à la fin du chapitre 3.

### D.1 Rappels

Le flux s'écrit pour  $n$  enzymes en parallèle catalysant la même réaction :

$$J = X \sum_{j=1}^n A_j E_j \quad (\text{D.1})$$

où  $X$  est une constante,  $A_j = k_{\text{cat}}/K_M$  est l'activité de l'enzyme  $j$  et  $E_j$  est la concentration de l'enzyme  $j$ .

Le système différentiel de l'évolution des concentrations d'enzymes est (voir chapitre 1) :

$$\frac{\partial E_i}{\partial t} = 2 \frac{\mu}{E_{\text{tot}}^2} \sum_{j=1}^n R_{E_j}^J \alpha_{ji} \frac{\delta_j^2}{e_j} \quad (\text{D.2})$$

où  $\mu$  est le taux de mutation,  $E_{\text{tot}}$  est la concentration totale d'enzyme,  $R_{E_j}^J$  est le coefficient de réponse de l'enzyme  $j$ ,  $\alpha_{ji}$  est le coefficient de redistribution et  $\delta_j$  est l'effet de la mutation ciblant l'enzyme  $j$ .

On cherche  $e_j^*$  tel que  $\frac{\partial e_j}{\partial t} = 0$ .

## D.2 Indépendance

En cas d'indépendance entre toutes les enzymes, l'évolution des concentrations relatives s'écrit (voir annexe A) :

$$\frac{\partial e_j}{\partial t} = 2 \frac{\mu}{E_{\text{tot}}^2} \left( R_{E_j}^J \frac{\delta_j^2}{e_j} - e_j \sum_{k=1}^n R_{E_k}^J \frac{\delta_k^2}{e_k} \right) \quad (\text{D.3})$$

Le coefficient de réponse en cas d'indépendance vaut :

$$R_{E_i}^J = \frac{\partial J}{\partial E_i} \frac{E_i}{J} = \frac{A_i E_i}{\sum_{j=1}^n A_j E_j}$$

Ainsi, en combinant ces deux équations et sachant que  $\delta_j = \nu$  pour tout  $j$ , la valeur à l'équilibre de la concentration relative de l'enzyme  $j$  est :

$$\begin{aligned} e_j^* &= \frac{R_{E_j}^J / e_j}{\sum_{k=1}^n R_{E_k}^J / e_k} \\ &= \frac{\frac{A_j E_j}{e_j \sum_{l=1}^n A_l E_l}}{\sum_{k=1}^n \frac{A_k E_k}{e_k \sum_{l=1}^n A_l E_l}} \\ &= \frac{A_j e_j E_{\text{tot}} / e_j}{\sum_{k=1}^n A_k e_k E_{\text{tot}} / e_k} \end{aligned}$$

Ainsi la concentration relative à l'équilibre est :

$$e_j^* = \frac{A_j}{\sum_{k=1}^n A_k} \quad (\text{D.4})$$

### D.3 Compétition

En cas de compétition, le coefficient de redistribution vaut :

$$\alpha_{ij} = \frac{-e_j}{1 - e_i}$$

Le coefficient de réponse devient :

$$\begin{aligned} R_{E_i}^J &= \frac{\partial J}{\partial E_i} \frac{E_i}{J} \\ &= \frac{E_i X \sum_{j=1}^n A_j \alpha_{ij}}{X \sum_{j=1}^n A_j E_j} \\ &= \frac{-E_i \sum_{j=1}^n A_j e_j}{(1 - e_i) \sum_{j=1}^n A_j E_j} \\ &= \frac{-E_i}{E_{\text{tot}}(1 - e_i)} \\ &= \frac{-e_i}{1 - e_i} \end{aligned}$$

Ainsi, l'évolution des concentrations d'enzymes devient :

$$\begin{aligned} \frac{\partial e_j}{\partial t} &= \frac{2\mu}{E_{\text{tot}}^2} \sum_{i=1}^n R_{E_i}^J \alpha_{ij} \frac{\delta_i^2}{e_i} \\ &= \frac{2\mu}{E_{\text{tot}}^2} \sum_{i=1}^n \frac{-e_i}{1 - e_i} \frac{-e_j}{1 - e_i} \frac{(1 - e_i)^2}{e_i(1/\nu + 1/E_{\text{tot}})} \\ &= \frac{2\mu}{E_{\text{tot}}^2(1/\nu + 1/E_{\text{tot}})} e_j \end{aligned}$$

Ainsi le seul moyen que la concentration d'enzyme n'évolue plus est que  $e_j^* = 0$ .











**Titre :** Évolution des concentrations d'enzymes dans les réseaux métaboliques

**Mots-clés :** Relation génotype-phénotype, Théorie du contrôle métabolique, Modélisation, Evolution, Dynamique adaptative, Analogie métabolisme-électricité

**Résumé :** La question de la relation génotype-phénotype est centrale en sciences de l'évolution. En effet, alors que les mutations concernent le génotype, c'est au niveau des phénotypes qu'opère la sélection naturelle qui va structurer le polymorphisme dans les populations. Pour analyser comment une pression de sélection à un niveau phénotypique donné se traduit au niveau sous-jacent, considéré comme porteur de l'information génotypique, j'ai choisi le modèle de la relation entre le flux à travers une voie métabolique et les concentrations des enzymes de cette voie. En m'appuyant sur la théorie du contrôle métabolique, j'ai développé un modèle mathématique d'évolution des concentrations d'enzymes lorsqu'il y a sélection directionnelle pour augmenter le flux, en envisageant deux contraintes biologiquement réalistes : la compétition pour les ressources cellulaires et/ou des co-régulations entre enzymes. J'ai montré que dans tous les cas un équilibre stable des concentrations relatives existait, mais avec des zones de neutralité sélective de tailles très variables selon les propriétés des enzymes et les contraintes appliquées. De plus j'ai montré qu'en cas de co-régulations, les équilibres dépendaient non seulement des activités des enzymes et

des co-régulations entre elles, mais aussi de leurs concentrations initiales, ce qui constitue un exemple de contingence évolutive. Enfin un *trade-off* entre les concentrations de deux enzymes suffit à limiter le flux et à empêcher les concentrations de toutes les autres enzymes d'augmenter. Toutes ces prédictions ont été confirmées par des simulations d'évolution à long terme dans lesquelles certaines des hypothèses simplificatrices faites pour les développements analytiques avaient été levées. Dans une dernière partie de la thèse, j'ai cherché à étendre le modèle d'évolution aux réseaux métaboliques en exploitant l'analogie entre réseaux métaboliques et circuits électriques, qui tous les deux impliquent des flux d'énergie dans des systèmes ramifiés. Sous réserve d'hypothèses simplificatrices, cette approche m'a permis de trouver une relation entre le flux et les efficacités de toutes les enzymes d'un réseau, et de montrer que la relation entre le flux et une enzyme particulière était toujours une courbe de saturation. On dispose donc d'un système formel pour étudier l'évolution des concentrations d'enzymes dans un réseau de topologie quelconque.

**Title:** Evolution of enzyme concentrations in metabolic networks

**Keywords:** Genotype-phenotype relationship, Metabolic control theory, Modelling, Evolution, Adaptive dynamics, Metabolism-electricity analogy

**Abstract:** The question of the genotype-phenotype relationship is essential to evolutionary science. Indeed, while the mutations concern the genotype, the natural selection acts at the phenotypic level, thus shaping the polymorphism in the populations. To analyze how selective pressure at a given phenotypic level affects the underlying level, considered as reflecting the genotypic information, I chose the model of the relationship between the flux through a metabolic pathway and the concentrations of the enzymes of this pathway. Using the metabolic control theory, I developed a mathematical model of evolution of enzyme concentrations when there is directional selection to increase flux, considering two biologically realistic constraints: competition for cellular resources and/or co-regulations between enzymes. I showed that in all cases a stable equilibrium of relative concentrations exists, but the zones of selective neutrality have widely different sizes, depending on enzyme properties and constraints applied. Furthermore, I showed that in the case of co-regulations, the equilibria depended not only on the activities of the

enzymes and the co-regulations between them, but also on their initial concentrations, which constitutes an example of evolutionary contingency. Finally, a trade-off between the concentrations of two enzymes is sufficient to limit the flux and prevent the concentrations of all other enzymes from increasing. All these predictions were confirmed by simulations of long-term evolution in which some of the simplifying assumptions made for the analytical developments were relaxed. In the last part of the thesis, I tried to extend the evolutionary model to metabolic networks by exploiting the analogy between metabolic networks and electrical circuits, both of which involve energy flows in branched systems. Subject to simplifying assumptions, this approach allowed me to find a relationship between the flux and the efficiencies of all the enzymes in a network, and to show that the relationship between the flux and a particular enzyme was always a saturation curve. This provides a formal system for studying the evolution of enzyme concentrations in a network of any topology.