



HAL
open science

L'alphabet, la Machine et l'ADN : l'incomplétude causale de la théorie de la programmation en biologie moléculaire

Giuseppe Longo, Pierre-Emmanuel Tendero

► To cite this version:

Giuseppe Longo, Pierre-Emmanuel Tendero. L'alphabet, la Machine et l'ADN : l'incomplétude causale de la théorie de la programmation en biologie moléculaire. Miquel, Paul-Antoine. Biologie du XXIe siècle : évolution des concepts fondateurs, DeBoeck, 2008, 978-2804155742. hal-03319651

HAL Id: hal-03319651

<https://hal-ens.archives-ouvertes.fr/hal-03319651>

Submitted on 12 Aug 2021

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.

L'alphabet, la Machine et l'ADN : l'incomplétude causale de la théorie de la programmation en biologie moléculaire¹

Giuseppe Longo

Laboratoire d'Informatique
CNRS et École Normale Supérieure, Paris
et CREA, École Polytechnique
<http://www.di.ens.fr/users/longo>

Pierre-Emmanuel Tendero

CREA
CNRS et École Polytechnique, Paris

Introduction

La difficulté immense d'analyse des phénomènes du vivant pousse souvent les scientifiques à recourir à la métaphore, avant de proposer (voire à la place de) un modèle, dans le sens physico-mathématique du terme. Cette attitude, bien rare en physique, se justifie par la moindre mathématisation des sciences du vivant, où la variabilité est conjuguée à l'invariance, la non-stationnarité ou généricité des évolutions possibles prévaut sur la spécificité des géodésiques, ces dernières étant des propriétés essentielles aux approches mathématiques telles qu'on les connaît en physique (voir [Bailly, Longo, 2006]). Or, « le modèle simplifie, la métaphore complique » [Nouvel, 2002]. Elle ajoute de l'information, elle fait référence à un (autre) cadre conceptuel prégnant, un univers de méthodes et de connaissances que l'on transfère sur celui de départ. Elle peut être utile, voire nécessaire, mais il faut essayer de comprendre ses conséquences, en particulier les méthodes, les concepts et l'histoire qu'on transfère ainsi sur l'objet d'étude. De plus, lorsqu'on fait référence à une métaphore explicative, celle-ci « devient l'objet d'une séduction pour la pensée. Si on l'utilise alors comme suggestion pour la solution d'une question philosophique, on parviendra, à la faveur de cette confusion, à faire apparaître cette métaphore comme une "conséquence philosophique" » de l'analyse scientifique [Nouvel, 2002].

On discutera ici certains aspects de deux métaphores communes pour le rôle et l'action du génome, tout en proposant d'abord des cadres historiques et conceptuels propres aux notions utilisées métaphoriquement, l'alphabet et l'ordinateur digital avec ses programmes. On essaiera aussi d'esquisser une analyse critique de certains des acquis empiriques sur la structure de l'ADN, tout en étant conscients des limites de nos choix, peut-être arbitraires, dans l'immense richesse de la littérature sur le sujet. En particulier, on mettra en évidence les insuffisances conceptuelles d'une structuration causale de l'ontogenèse, qui serait complètement décrite dans l'alphabet moléculaire, en termes de théorie de la programmation, avec ses notions intrinsèques de détermination/déduction. On s'interrogera également sur la possibilité de proposer une analyse de l'ontogenèse à partir des méthodes usuelles de la physique, dont l'analyse différentielle des paramètres et des observables.

1. L'alphabet, en quelques lignes.

Dans [Havelock, 1976 ; Herrenschildt, 1996], on raconte l'histoire fascinante de la naissance de

¹ In "Evolution des concepts fondateurs de la biologie du XXI^e siècle", (Miquel ed.) DeBoeck, Paris, 2008.

l'écriture, à partir des symbolismes magiques et de l'idéogramme, jusqu'aux notations syllabiques et à l'alphabet. Ce dernier, un grand tournant qui va marquer notre culture, est un passage extraordinaire vers l'abstraction. Le flux du langage, un chant continu, est découpé dans des fragments sans signification, des lettres, des consonnes en fait, qui ne doivent pas avoir de signification, ni d'histoire. Contrairement au hiéroglyphe, à l'idéogramme, la lettre, composante élémentaire (indécomposable) de l'alphabet, est très simple et insignifiante ; par contre, un idéogramme, même élémentaire, peut être très complexe, dans sa structure ou dans son rapport à l'espace de la signification évoquée. De plus, l'alphabet force un dualisme nouveau : on départage la pensée, organisée dans la langue (écrite), de l'objet de pensée, du monde ; en fait, le sujet connaissant de l'objet à connaître.

Voilà donc une pratique historique des plus anciennes, d'abstraction de plus en plus poussée, en tant que parfaitement détachée du monde : des signes qui obligent à prendre du recul par rapport à la signification, à réfléchir à ce que l'on dit en tant qu'on signifie et qui poseront à la culture grecque, la première avec une notation complète pour les voyelles, le problème de la signification (*semeion*). Que veut dire ce jeu de signes insignifiants ? Le fait que l'alpha soit, peut-être, une tête de taureau penchée, n'a aucune importance : on *doit* oublier toute origine évocatrice des lettres, pour lire et comprendre. On les juxtapose et un signifié émerge. On arrive même à composer une phrase qui dit « cette phrase est fausse », paradoxe ahurissant - les chinois n'ont pas produit ce genre de paradoxes : leurs jeux de mots sont plutôt des contrapositions de significations (un même vendeur qui vend une épée qui perfore n'importe quel bouclier et un bouclier qui résiste à toute épée). Et Aristote, en développant de façon analytique les idées de Platon, démarrera la sémantique moderne : « la neige est blanche » si la neige est blanche, expliquera Tarski sans beaucoup d'originalité - les signes insignifiants, juxtaposés, font référence à la vérité, qui est dans le monde. “Adequatio intellectus et rei” disaient déjà, suivant Aristote, ses Commentateurs, car, pour lui, l'intellect, la pensée, est l'impression de lettre de l'alphabet sur notre corps, tout comme elles sont imprimées sur une tablette de cire [Cappuccio, 2003], réflexion audacieuse, à l'aube de la philosophie. Il y a donc *ici* la pensée, alphabétique, voire “succession de 0 et 1” (disait-on en Intelligence Artificielle), parfaitement détachée de toute signification (une pensée formelle, un Turing-calcul programmable disent encore certains logiciens-philosophes), *là* le monde des choses, bien séparés : l'adéquation de la première au deuxième constitue le critère de (la) vérité. Cette pensée-calcul sans signification aura son analogue dans le paradigme du programme génétique.

Mais comment émerge-t-elle cette signification à partir d'un jeu de signes alphabétiques sans signification ? C'est le *phonème* qui produit le sens : c'est-à-dire, il faut prononcer, chanter, pour qu'il y ait émergence du sens. Il faut revenir au chant originaire, ce continu de la communication humaine, où la distance mesurée entre les mots n'est pas plus grande que celle qui est entre les syllabes. Ce chant découpé, atomisé par la construction alphabétique, est l'invention qui a marqué notre civilisation [Ong, 1988 ; Sini, 1994,] : c'est à partir de l'alphabet que l'intelligibilité est donnée par le découpage du monde en atomes individuellement insignifiants (Démocrite) ; même le raisonnement est intelligible et certain, seulement si découpé en anneaux atomiques, élémentaires et simples, sorte d'alphabet de la pensée (Descartes, mais aussi, leur façon, Frege et Hilbert) ; et l'on associera ainsi des lettres aux bases dans l'ADN. Mais, nous insistons, ces atomes, dont les consonnes sont des “pitch” linguistiques sans sens, produisent le sens *par la voix*, le phonème. Un chant que l'on doit produire à haute voix, nécessairement pour apprendre, comme pratique tout enfant, mais même pour comprendre. La lecture est faite à haute voix, pendant des siècles. Sant'Ambrogio (IV^e siècle, en Lombardie), un des rares lecteurs en silence

cité, a frôlé l'hérésie, car il suscitait merveille et inquiétude : il faut produire des sons dans le dialogue lu avec Dieu. En fait, la lecture était presque exclusivement sacrée et Rome s'inquiète. Ambrogio et sa secte de lecteurs silencieux sera mieux maîtrisée, en le sanctifiant (on connaît l'importance du dialogue intime avec Dieu chez les anti-papistes, quelque siècle plus tard : le Pape de l'époque avait bien compris l'enjeu). Mais même dans le silence de la voix, il faut produire des sons, mentalement au moins, pour lire l'alphabet : *le phonème est nécessaire à la transition vers le sens*. Tout alphabet est phonétique. Par contre, un idéogramme évoque une signification sans médiation sonore : si vous avez besoin des toilettes, dans une gare du monde, et vous voyez l'idéogramme correspondant, vous agissez, sans aucun besoin de prononcer le mot, dans aucune langue, pour y aller. Nous tous, alphabétisés, nous produisons en lisant le miracle du musicien expérimenté qui entend la musique dans sa tête, quand il lit une partition - une autre formidable écriture alphabétique-phonétique, découpage atomique du continu musical, aux liens de continuité souvent explicites. On cherchera l'analogie du phonème dans la lecture alphabétique du génome.

2. La machine

La force de l'alphabet, ces rails pour la pensée, d'Aristote à Descartes et aux systèmes arithmétisant des formalismes logiques modernes, nous a donné la plus formidable des sciences au monde : l'intelligibilité du complexe par son découpage en composantes élémentaires et simples, atomiques, insécables et certaines, dépourvues des ambiguïtés du sens. Et la Machine de Turing (MdT), l'ordinateur digital, sera l'apothéose moderne de la culture alphabétique, synthèse ultime de la science de Démocrite, Galilée, Descartes et Laplace. Mais il ne faut pas se tromper : la MdT, l'ordinateur moderne, est sûrement alphabétique, par codage du monde dans un alphabet de 0 et de 1, *mais elle s'est spécifiée d'une façon bien particulière*. Elle a son origine dans une histoire très riche, qui mélange la culture des horloges mécaniques, la logique booléenne et, surtout, *le tournant arithmétique dans les fondements des mathématiques*.

C'est le « délire » (Frege, 1884) proposé par les géométries non-euclidiennes (la géométrie riemannienne, en particulier), la catastrophe de leur signification dans l'espace sensible, qui pousse les pères de la logique moderne vers un fondement des mathématiques exclusivement logico-formel, sans rapport à l'espace et au temps, un rapport si fortement mis à mal par la perte de l'unité des espaces de l'intuition sensible et physique, l'espace d'Euclide et de Descartes, l'absolu de Newton. L'Arithmétique de Dedekind-Peano, devra être alors, pour Frege et Hilbert (quoique de façon très différente), la "bottom line" de toute certitude. Le nombre entier, concept absolu (Frege) ou notion formelle dans le cadre d'une théorie démontrablement non contradictoire *par leurs propres moyens* (Hilbert), en dehors de l'espace et du temps, en dehors du monde, doit garantir les fondements certains, élémentaires et simples, de tout l'édifice des mathématiques, une fois reconstruits ou encodés en Arithmétique, le continu de l'Analyse (Cantor-Dedekind, 1880) et toutes les géométries (Hilbert, 1899). Et la *règle*, absolue, arithmétique, logique ou formelle, *engendre la conséquence logique*, reconstruit, voire produit, la pensée mathématique ; règle, pour Hilbert, formelle, car à appliquer d'une façon « potentiellement mécanisable », un pas élémentaire et simple, atomique, après l'autre, sans les ambiguïtés possibles du sens. Et voilà la récursivité à la Herbrand-Gödel (1930-31), et plus encore le lambda-calcul de Church (1932) et la Machine de Turing (1935) qui spécifient enfin ces mathématiques du formel-potentiellement-mécanisable dont rêvaient Peano et Hilbert : le calcul, suite finie de suites finies de 0 et 1, procède par "sequence-matching" et "sequence-replacement",

et rien d'autre. Et la machine arithmétique développe ses calculs comme des preuves formelles : elles sont en fait identiques (isomorphisme de Curry-Howard).

À partir des règles bien formalisées, en amont de tout calcul, le processus effectif déroule sa puissance expressive d'une façon *déterministe et (effectivement) prédictible*. Puisque voilà l'enjeu : en passant à côté, programmatiquement, de la géométrisation de la physique, de Riemann-Einstein (Relativité) et Poincaré (Systèmes Dynamiques), les fondements arithmétiques des mathématiques gardent une culture laplacienne, voire aristotélicienne (Tarski), chaque fois qu'ils sont confrontés au monde (la détermination formelle-mathématique implique la prédictibilité, voire la décidabilité conjecturée par Hilbert, comme par Laplace en physique). Turing le dit très bien dans son article de 1950, en répétant plusieurs fois que *sa machine est laplacienne* (« la prédiction est possible » dit-il, même si cela peut-être difficile *en pratique*), lui qui depuis 1948 a commencé à s'intéresser aux équations non-linéaires pour une dynamique déterministe de la morphogenèse, dont l'imprédictibilité est *théorique* ([Turing, 1952] ; voir [Longo, 2002], pour une discussion)². Expliquons nous.

Certainement, donc, la MdT est une machine alphabétique, mais son contexte d'expressivité est bien spécifique, riche d'une histoire logico-mathématique qu'on ne doit pas oublier, surtout quand on y fait référence comme métaphore. Et ce n'est sûrement pas l'évocation du son, le phonème, qui en ouvre le parcours vers la signification. Sa sémantique (opérationnelle) est donnée par l'application de la règle de calcul, cet absolu de l'écriture hors du monde, sur une base de données exacte et absolue, sans espace ni temps. Car la MdT *n'est pas* dans l'espace physique : des théorèmes garantissent que la dimension (cartésienne) de sa structure "physique" (matérielle) n'a pas de rôle pour l'expressivité du calcul, *contrairement à tout phénomène physique*. Son temps, sécrété par l'horloge interne, *n'est pas* un temps, tout comme il n'y a pas de distance dans un univers vide, où il n'y a qu'un mètre : un univers tout juste avec un mètre, voire une horloge, isolé dans le vide relationnel. Fort justement, Turing définit sa Machine une *Logical Computing Machine* - un homme dans l'acte minimal (élémentaire et simple) de la pensée alphabétique, des 0 et des 1, manipulés par l'application ordonnée de règles logico-formelles ; son implantation physique, spatio-temporelle, n'a aucune importance. Et *sa nature prédictible est au coeur du calcul séquentiel* : elle peut se comprendre par le rôle essentiel de l'itération à l'identique, toujours possible ; une symétrie par translation temporelle qui caractérise l'application des règles de Frege-Hilbert-Turing (voir [Longo, 2002 ; Bailly, Longo, 2006]).

L'enrichissement pratique de l'informatique par la distribution des calculs et des machines dans l'espace-temps physique, en réseaux, voire les systèmes distribués et concurrents, rendent le calcul *non-absolu en principe* : différents parcours spatio-temporels peuvent produire des résultats différents, à partir des mêmes programmes. Cela n'empêche que tout soit fait en informatique pour éviter cet inconvénient et le succès pratique est remarquable (une approche intéressante à l'analyse et la gestion géométrique de ces enjeux se trouve dans [Goubault, 2000], où l'on utilise la théorie de l'homotopie – une théorie des parcours connexes spatio-temporels ; dans [Aceto *et al.*, 2002] on développe des analyses mathématiques et l'on explique la perte de l'absolu - ainsi que la préservation de l'exactitude - dans les réseaux de calculs distribués et

² "In a Discrete State Machine,... it is always possible to predict all future states". "This is reminiscent of Laplace's view ... The prediction which we are considering is, however, rather nearer to practicability than that considered by Laplace" [Turing, 1950; p. 47 and ff.]. Et cette fois la prédiction est praticable, puisque, nous verrons, une structure discrète a un accès exact aux données. Pour Shroedinger aussi, si les chromosomes, avec leur structure discrète et finie, contiennent le "code" dans son entièreté, alors la connaissance complète de ce code est possible et elle nous mettrait à la place "of the all-penetrating mind, once conceived by Laplace", quant à prévision phénotypique [Shroedinger, 1951 ; pp. 22-23].

concurrents, pour lesquels, justement, *la MdT ne constitue plus un modèle mathématique adéquat*). Toutefois, *les bases de données et les programmes restent exactes* : le niveau d'observabilité est donné par la discrétisation et la dynamique est "seulement" spatio-temporelle, *elle n'est pas due à la structure de détermination*, en tant que départage entre détermination globale et causalité locale, comme dans les dynamiques non-linéaires en physique (voir [Bailly, Longo, 2006]). Expliquons nous.

La MdT est laplacienne, comme l'on disait plus haut, car la symétrie par translation temporelle de la structure causale (l'itérabilité à l'identique) est synonyme de prédictibilité de son cadre dynamique déterministe. Son univers mathématique exact (digital) garantit cette propriété cruciale, car c'est la causalité locale, en dessous de la mesure physique possible (voire de la discrétisation), qui produit l'imprédictibilité mathématique des *systèmes continus* (comme les appelle Turing, 1950), c'est-à-dire des dynamiques déterministes non-linéaires, dans notre langage (soumis à l'"exponential drift", précise Turing en 1952 ; sensibles aux conditions initiales ou au contour, disons-nous aujourd'hui, dont le rôle des variations même inobservables est intégré dans la théorie – voilà le sens du modèle turingien de la morphogenèse). Cette causalité locale du non observable n'a mathématiquement pas de sens dans des structures de données topologiquement discrètes, dont le niveau d'observabilité est donné par la structure digitale et rien ne se passe en dessous, par engagement théorique (et pratique ; certains intelligents artificiels classiques nous disent souvent ici que les machines aussi se trompent, par des défauts du matériel ou des bogues des programmes par exemple – ce qui serait synonyme d'intelligence, voire de créativité ; mais la faillite logicielle ou matérielle, rarissime, a toute autre mesure de probabilité par rapport à la certitude de la non itération propre à n'importe quel système dynamique, quelque peu sensible aux conditions aux contours).

En plus, la MdT est surtout une machine logico-arithmétique : elle suit la règle formelle et rien d'autre, une règle donnée en amont, indépendante de tout matériel, dans un langage de type arithmétique - voilà aussi pourquoi, dans sa structure discrète, elle agit dans un cadre, artificiel, à la Newton-Laplace, où la loi est normative. Dans l'univers de la physique moderne, on ne propose plus des lois formelles qui, en amont, mouleraient le monde, mais on analyse des phénomènes par la construction de cadres métriques (de mesure), faits de pics et vallées, des géométries adéquates où des géodésiques (qui minimisent les longueurs dans la métrique donnée), voire des symétries et des invariants, nous donnent une intelligibilité mathématique possible ([van Frassen, 1994 ; Bailly, Longo, 2006]). C'est dans ce cadre qu'on peut faire enfin une distinction entre stabilité structurelle globale et causalité locale (une distinction qu'on pourrait attribuer au débat R. Thom vs. I. Prigogine) : une perturbation/variation en dessous de l'observable *cause/est à l'origine de la trajectoire spécifique*. *Cette distinction n'a pas de sens dans des structures dont la topologie naturelle est discrète*, comme les bases de données digitales et les systèmes informatiques. Voilà, très rapidement donc, la structure causale implicite, de type laplacien, à la notion de "programme" (voir [Bailly, Longo, 2006] ; on y reviendra dans le § 4.1).

3. L'ADN.

Passons maintenant des quelques certitudes qu'on peut avoir au sujet de la théorie de la calculabilité, si bien ancrée sur l'élémentaire et le simple (tout en atteignant le très compliqué, par composition du simple), à une analyse critique du recours à la notion de programme génétique en tant que cadre d'interprétation des liens qui unissent cet élémentaire très complexe qu'est la cellule vivante (élémentaire, du point de vue du vivant, car si on la découpe elle meurt, mais complexe, peut-être infiniment complexe, par rapport à toute mesure physique) à cette

composante discrète si importante des systèmes biologiques : l'ADN. Dans la suite de ce paragraphe, on s'intéressera plus particulièrement à la question du rôle physiologique des gènes et on s'interrogera, dans un premier temps, sur les définitions successives qui en ont été proposées, ainsi que sur la compatibilité de ces définitions entre elles (§3.1). En vue d'expliciter ce qui, à notre sens, fait ici problème, on essaiera (§3.2) de mettre en évidence le fait que, malgré les assertions historiques réitérées de la génétique, l'existence d'un lien de causalité direct entre gènes et caractères ne peut être solidement établie sur la seule base des données expérimentales dans la mesure où ces dernières se présentent *a priori* comme le résultat d'expériences *différentielles*, en un sens qu'on essaiera d'expliciter grâce à des références empruntées à la physique et à l'informatique. La question principale qui nous occupera alors sera celle de savoir s'il existe un cadre théorique adéquat permettant de concilier les assertions de la génétique (classique mais aussi moléculaire) et les faits expérimentaux. On soulignera, en particulier que, de ce point de vue, la théorie de la programmation est incapable *par principe* de fournir un tel cadre (§3.2.2). On se demandera ensuite (§§3.2.3 à 3.5) si, passant outre cette première restriction (*qui reste néanmoins dirimante pour la théorie du programme génétique*) et en admettant, malgré tout, l'existence d'un lien de causalité direct entre génotype et phénotype, *au moins dans le cas de la biologie moléculaire* le recours à la métaphore des machines de Turing et à certaines notions apparentées de la théorie de l'information permet de construire un cadre conceptuel robuste capable d'unifier sous le concept de gène les propriétés structurelles et fonctionnelles de l'ADN.

3.1 Le concept de gène du point de vue de l'histoire de la génétique moderne.

Du point de vue de l'histoire de la génétique moderne, il nous semble qu'on peut repérer trois grandes caractérisations successives de la notion de gène qui se sont superposées les unes aux autres pour finalement aboutir à une définition opératoire, certes, très précise, mais qui est devenue *problématique dans le cas des eucaryotes supérieurs*. En effet, le gène a d'abord été défini comme une *hypothétique unité fonctionnelle de recombinaison* (c'est-à-dire d'assortiment et ségrégation), objectivable dans le cadre d'expériences d'hybridation par le fait qu'elle permettait de formuler des prédictions, *en probabilité*, relativement à la fréquence des caractères dans la descendance d'une génération parentale dont le génotype était supposé connu [Mendel, 1907]. Par la suite, cette définition a été complétée, à partir des travaux de Morgan et de son collaborateur Hermann J. Muller [Morgan, 1926], par l'hypothèse supplémentaire selon laquelle le gène était le lieu d'une *modification structurelle au niveau des chromosomes*, modification qui était censée intervenir dans la *transmission différentielle des caractères*. En d'autres termes, le gène est alors devenu *une unité structurelle de mutation*, unité qui, du reste, était encore seulement hypothétique puisque la localisation des gènes au niveau des chromosomes ne reposait que *sur une simple analogie formelle* entre le constat empirique de la recombinaison des caractères dans la descendance des individus et certains comportements remarquables des chromosomes au moment de la méiose (c'est-à-dire les *crossing over*, "mécanismes" par lesquels les chromosomes appariés, se séparant pour former des cellules haploïdes, échangent "au hasard" certaines portions homologues de chromatine). Enfin, avec la découverte de la structure tridimensionnelle de l'ADN [Watson et Crick, 1953, Crick, 1957], le concept de gène a reçu, pour la première fois, une *caractérisation à la fois structurelle et fonctionnelle* à travers la formulation, dans les termes de la biologie moléculaire, de l'hypothèse - ou plutôt du "dogme central" - de l'existence d'une relation entre gènes et protéines que la découverte du code génétique devait bientôt confirmer. Or, ce qui, à notre avis, pose très clairement problème aujourd'hui de ce point de vue, c'est qu'il semble de plus en plus difficile de remonter le fil de cette histoire, à tel point, d'ailleurs, qu'il paraît même envisageable que ces trois déterminations

successives de la notion de gène ne se rapportent pas à des entités et/ou processus identiques. La question qu'on peut se poser, du reste, face à ce problème, est évidemment celle de l'apport que représente, en termes d'explication théorique, l'introduction des métaphores génétiques du programme et du langage/alphabet.

3.2 Structure et fonction dans la définition du concept de gène.

D'un point de vue très général, on peut distinguer, dans ces caractérisations successives de la notion de gène, deux aspects remarquables, indéfectiblement attachés l'un à l'autre bien que leur conciliation se révèle nettement problématique. Le premier de ces aspects, même s'il n'est apparu que secondairement d'un point de vue historique, est celui de la *caractérisation structurelle des gènes*, tandis que le second concerne *leur définition fonctionnelle*.

3.2.1 Quelques précisions quant à l'orientation des remarques suivantes.

Il est, certes, clair que les travaux de cartographie morganiens suggéraient assez nettement l'existence d'une caractérisation structurelle des gènes. Néanmoins, il nous semble fondamentalement problématique d'affirmer de manière péremptoire que cette caractérisation structurelle est absolument superposable à celle que propose la biologie moléculaire aujourd'hui.

En premier lieu, en effet, les *loci* repérés par le biais des protocoles d'hybridation morganiens n'étaient associés à des caractères que sur la base d'une interprétation *causale, très peu argumentée sur le plan physiologique*³, d'un ensemble de *corrélations empiriques* établies entre, d'une part, la distribution observable des recombinaisons de paires de caractères au sein de la descendance d'une génération parentale donnée et, d'autre part, la fréquence des *crossing over* au niveau des cellules germinales des individus. Or, ce type de méthodologie statistique laisse fondamentalement ouvertes les trois questions suivantes, qui sont cruciales, aujourd'hui, dès lors que se pose la question de la réinterprétation des énoncés de la génétique classique dans le vocabulaire de la biologie moléculaire.

1) Sur la base de quel principe physico-chimique *a priori* et/ou d'évidence empirique indiscutable peut-on affirmer que les localisations en question *concernaient la seule molécule d'ADN* ?

2) Si les corrélations observées doivent être interprétées dans un vocabulaire causal - ce qui ne va pas aussi clairement de soi qu'on pourrait le croire (*cf.*, par exemple, §. 3.2.2 et 5.2) -, la relation causale ainsi mise en évidence est-elle nécessairement une relation de type *bottom/up* telle qu'elle existe dans les mécanismes de transcription et de traduction de l'information génétique⁴ ?

3) Quel rôle causal peut-on attribuer, sur la base de ces analyses, à la structure tridimensionnelle de la chromatine, laquelle a indubitablement une influence non-négligeable sur la cinématique des réactions chimiques dont le noyau des cellules est le siège ?

³ Et il ne saurait en être autrement étant donnés les moyens d'investigation moléculaire disponibles à l'époque.

⁴ Lesquels sont d'ailleurs, à cet égard, les seuls pour lesquels la biologie moléculaire dispose d'une explication théorique systématique au sens le plus strict du terme.

En second lieu, on peut également remarquer que, depuis les années 60, avec la définition structurelle des gènes en rapport avec la synthèse de protéines, la question de la relation entre génotype et phénotype ne se pose plus, désormais, au niveau des caractères observables en tant que tels mais à celui du métabolisme cellulaire, ce qui indique, de la part de la biologie moléculaire, un très net recul, en termes de domaine de validité explicative, par rapport à la génétique classique. On verra d'ailleurs plus loin que, à strictement parler, on n'arrive même pas au métabolisme cellulaire proprement dit mais, encore que d'une manière problématique (cf. §3.4.2), *au rôle enzymatique potentiel des protéines dans les cycles biochimiques spécifiques du vivant*⁵.

Toutefois, il existe encore aujourd'hui, en marge de la biologie moléculaire au sens strict du terme, des travaux très importants qui mettent en évidence le rôle de certaines *modifications* de la structure du génome des espèces dans la modification d'un certain nombre d'éléments phénotypiques associés. Par exemple, on sait très bien que des modifications dans la disposition de gènes homéotiques induisent des effets *tératogènes* sur le développement des embryons de drosophile. Mais il y a là, précisément, quelque chose de problématique à nos yeux de logiciens et d'informaticiens : cet aspect tératologique, ou simplement *différentiel*, propre à la méthodologie de la génétique ne semble pas se prêter aussi simplement qu'il y paraît à une analyse rétrospective en *termes causaux*, voire déductifs comme en programmation. Discutons de cet enjeu, pour nous fondamental.

3.2.2 Méthodes différentielles et structure causale.

La méthode empirique dominante en génétique moléculaire consiste dans l'induction de modifications génotypiques et l'observation d'éventuelles variations phénotypiques. De quelle manière est-il possible d'interpréter des corrélations empiriques de ce type dans un vocabulaire causal et comment une telle théorisation se rapporterait-elle aux pratiques habituelles de la physique, pour laquelle les méthodes différentielles constituent un protocole expérimental courant, ou de l'informatique, où la notion de "programme" a une origine logico-déductive et ses applications pratiques ?

En physique, on appelle "méthode différentielle", la pratique expérimentale qui consiste à dériver certaines relations causales en fixant tous les paramètres d'un système, sauf un que l'on fait varier dans un certain domaine restreint de valeurs. Et cela peut conduire à proposer une relation entre paramètres et observables (en fait, une équation) et, ainsi, *mettre en évidence des relations causales* en tant que *brisures formelles de symétrie*. Considérons, par exemple le cas de l'équation du gaz parfait à l'équilibre, $pV = KT$; une fois cette équation assertée, une brisure de sa symétrie formelle, correspondant à la brisure d'un équilibre, s'interprétera comme l'expression d'une relation causale *possible* entre les différents paramètres (par exemple, une variation de p , à gauche, cause une variation de T à droite...).

Par application de la méthode différentielle en physique, une telle équation peut donc être reconstruite *empiriquement* en faisant varier la pression p seulement, par exemple, dans une certaine fenêtre de valeurs. Mais, comme cela a précisément été le cas pour l'équation des gaz parfaits 50 ans après sa formulation historique, une relation de ce type peut aussi (et l'on serait tenté de dire "doit", en physique) être *déduite mathématiquement* et de manière *a priori* dans un

⁵ Ces cycles, d'ailleurs, ne se limitent pas à de simples réactions chimiques entre protéines mais font intervenir, parfois de manière cruciale, diverses espèces chimiques, moléculaires ou non, au sujet desquelles il est bien difficile (pensons au cas des ions...) d'affirmer que leurs propriétés physico-chimiques sont placées sous la dépendance de gènes.

cadre théorique plus général, celui de la physique statistique au moins pour la question qui nous occupe. Ce qui nous semble important, de ce point de vue, c'est qu'alors la causalité mise en évidence dans les deux cas n'est pas la même, puisque, dans le second, on l'aura dérivée d'une équation différentielle, qui propose un cadre de *détermination théorique* systématique et qui justifie (de manière *a priori* d'un point de vue déductif, même si elle lui succède généralement d'un point de vue historique) l'analyse différentielle empirique. Et cette équation est basée sur des principes très généraux, les principes géodésiques (ou de moindre action) qui régissent l'analyse des trajectoires d'un gaz, transférée ensuite à la limite thermodynamique (une limite infinie, une intégrale). Or, tant que ce cadre manque, *une équation du type $pV = KT$ est seulement une détermination formelle possible d'une corrélation empirique* et ne se prête pas, en tant que telle, à une analyse proprement physique en termes de causalité objective, [Bailly, Longo, 2006]. En particulier, il y a *toujours* une infinité de lois - d'équations - de ce genre (c'est-à-dire, polynomiales) permettant de réaliser une interpolation entre les mêmes données empiriques, dans la fenêtre des valeurs admissibles des paramètres ; tandis que la dérivation formelle dans le cadre d'une théorie bien explicitée ne produit *qu'une seule* solution et le recours à l'expérience se limite alors à la détermination de la constante K - ou à l'éventuelle falsification de *toute* la théorie, sans permettre l'astuce de changer juste l'équation.

Pour en revenir au problème particulier qui nous préoccupe dans ce paragraphe, c'est-à-dire, celui de l'existence d'un lien de causalité éventuel entre génotype et phénotype, il nous faut donc souligner le fait que, bien que la méthode différentielle en génétique prétende atteindre la rigueur démonstrative des théories physiques, l'absence d'un cadre théorique plus général ne permet pas de donner valeur de preuve scientifique aux seules expériences de tératogénie actuellement disponibles. Ce qui rend "causalement incomplet" le discours de la génétique de ce point de vue consiste simplement dans le fait que cette dernière ne se trouve pas en position d'établir l'existence formelle d'un lien de *causation* direct entre génotype et phénotype. On peut en fait soupçonner ici la persistance implicite d'une certaine croyance causale pré-théorique (voire pré-scientifique), un peu comme si les généticiens partageaient du principe qu'ils pouvaient, en droit, produire une *détermination théorique* systématique par l'évocation de la métaphore du programme informatique, sans nécessairement ressentir le besoin de l'expliciter dans un cadre théorique très précis (physique, informatique, ou mieux : *proprement biologique*). Ce que l'on veut dire, et que l'on exprimera encore d'une autre façon par la suite (§§. 4.3 et 5.2), c'est que d'une étude simplement différentielle du lien entre génotype et phénotype, on ne saurait dériver immédiatement une théorie causale, voire déductive comme en programmation (car un programme est une "déduction"), au sein de laquelle un génotype donné produirait un phénotype spécifique, voire une gamme de phénotypes en fonction de certains paramètres de développement (c'est-à-dire, ce que les biologistes moléculaires appellent "norme de réaction")⁶.

⁶ On peut illustrer le caractère problématique de tels liens causaux de deux manières (en physique et en informatique). En premier lieu, en effet, les théoriciens de la mécanique quantique savent très bien, étant donnée la place centrale qu'a occupé le débat autour de ce cas dès les premières formulations de la théorie, que, dans le cas d'un état intriqué, une mesure sur une particule du système - mesure qui *est* une modification de l'état de cette particule - induit instantanément une modification identique de l'état de la particule associée. Or, cela ne démontre pas qu'il y a une relation de causalité entre les deux événements ; et d'ailleurs il ne saurait être question de relation de causalité dans ce cas, du fait de la finitude de la vitesse de la lumière. En second lieu, et de manière symétrique, en informatique, une modification de l'ordre des instructions d'un programme ne modifie pas nécessairement le résultat d'un calcul alors qu'il existe des raisons fortes, notamment dans le cadre paradigmatique des machines de Turing, de parler d'une relation de causalité (laplacienne) à propos du lien entre programmes et calculs (voir ci-dessus, § 2). Et, par suite d'ailleurs, il est clair qu'en logique et en informatique, l'analyse de la sémantique de la déduction ne saurait se contenter d'une simple méthode différentielle, voire perturbative, pour établir un lien de causalité entre

La singularité physique du vivant est telle de ce point de vue qu'une théorie physicaliste de sa détermination mériterait une analyse aussi fine qu'en physique quantique ou en relativité, où les espaces de Minkowski nous ont donné un autre aperçu original des relations causales. Une analyse, toutefois, tout à fait différente de celles propres aux théories physiques modernes (qui sont loin d'être unifiées et diffèrent fondamentalement des cadres classiques), par suite, certainement très lointaine de la prédictibilité laplacienne que Turing lui-même attribue à ses machines à états discrets, [Turing, 1950]. Une des raisons fortes en faveur de cette singularité tient à l'enchevêtrement et au bouclage entre les niveaux d'organisation propres au vivant. L'intégration et la régulation, concepts qui s'expriment assez mal dans les théories physiques courantes, - et qui sont, en particulier, bien différents des feed-back de la cybernétique - sont une des composantes de cet enchevêtrement et bouclage causal⁷. A cela, il faut encore ajouter que, ces aspects singuliers ayant été intégrés, il faudra encore rendre compte d'un point de vue strictement physicaliste de cette caractéristique remarquable des systèmes biologiques qu'est l'unité d'un organisme, que l'on essaye d'analyser ailleurs comme un système de type critique aux *longueurs de corrélations*, entre paramètres, de la taille de l'objet lui-même (voir Conclusion).

Il est clair que ces différents éléments, en particulier la complexité des corrélations, ne manqueront pas de rendre d'autant plus difficile l'application rigoureuse de la méthode différentielle consistant à fixer tous les paramètres (et les observables) sauf un, faire varier ce dernier et en observer seulement les conséquences directes. Mais peut-être pourra-t-on le faire grâce à une théorie variationnelle (*d'un type sans doute inédit en physique*) des phénomènes *loin de l'équilibre* et près de (voire dans (!)) la criticalité, où les longueurs des corrélations physiques divergent (on y reviendra).

On ajoutera, d'ailleurs, que, dans un cas tel que celui de la phénocopie, connu depuis [Goldschmidt, 1938], les biologistes du développement sont capables de créer des "clones phénotypiques", simplement en modifiant, de diverses manières, les conditions expérimentales du développement (voir [Stewart, 2004] pour des références récentes). D'un point de vue

programmes et fonctions calculées. Il existe néanmoins un type d'analyse, en théorie de la programmation, qui peut ressembler à une méthode différentielle : la technique du "böhm-out" [Barendregt, 1984]. Le théorème qui en est à l'origine permet de démontrer qu'une différence entre deux programmes, en tant que termes du lambda-calcul (dans leurs arbres de Böhm, à un niveau fini) induit une différence dans leur sémantique, *étant donnés* une sémantique opérationnelle bien précise ou certains domaines mathématiques d'interprétation, [Barendregt, Longo, 1980]. Mais ce résultat, qui peut être considéré de type "différentiel", ne donne point la sémantique d'un programme (voire d'un lambda-terme), son "phénotype" pour ainsi dire ! Il caractérise syntaxiquement les termes avec *la même sémantique*, dans une structure sémantique déjà donnée (ou, au moins et plus techniquement, il démontre que deux termes en forme normale syntaxiquement différents ne calculent pas la même fonction). Encore plus qu'en physique, donc, on donne *d'abord* un cadre conceptuel interprétatif cohérent – sound, en anglais (avec sa métrique ou topologie, son interprétation mathématique parfaitement claire... sorte de cadre de la signification-détermination comparable à celui de la physique statistique, pour continuer l'analogie), *ensuite* on affine l'analyse de l'égalité entre termes par une méthode différentielle. Encore une fois, l'analyse des variations contribue à mieux expliciter des significations, voire les processus/fonctions/"phénotype" engendrés, *une fois donnée* une interprétation directe et cohérente, bref ce que signifie/calcule un programme tout à fait en général. Pour cette raison *la notion de programme génétique ne permet pas de rendre compte du lien entre gènes et caractères seulement à partir d'une analyse différentielle*, car il lui manque (une proposition théorique autonome et scientifiquement solide de) une détermination de ce lien dans le sens direct.

⁷ Un exemple très simple mais qui suffit à souligner les difficultés d'une analyse causale directe est le suivant : une modification de la concentration en certaines catégories ioniques dans le sang, par exemple, peut causer une baisse de la pression, contrairement à l'effet physique attendu, suite à des mécanismes de régulation qui "over-react" dans le sens inverse (effets ago-antagonistes, [Bernard-Weil, 2002]).

différentiel, donc, la pression, la composition chimique du milieu ou l'intensité locale du champ électromagnétique, en tant que paramètres de l'écosystème, et, plus généralement, tout le contexte extragénomique, *sont* causes du développement de l'embryon tout autant que le génome, si on reste à un niveau *purement empirique*. On peut, par exemple, induire des effets tératogéniques similaires à ceux induits par le déplacement d'un gène homéotique (voir § 3.2.1), simplement en modifiant la pression à un certain moment de l'embryogenèse. Et dès lors, l'assignation de causalité, ou plutôt, de primauté causale au génome dans la production d'un phénotype donné *ne paraît pas jouir d'une assise empirique indiscutable et manque d'un encadrement théorique rigoureux*. Pour revenir à l'exemple qu'on vient de citer, on ne voit pas de raison de principe pour exclure d'emblée la possibilité que certaines modifications de la disposition des gènes homéotiques *causent* en fait un changement de la réactivité de l'embryon à la pression....

Pour nous résumer, on vient d'essayer de mettre en évidence, en l'illustrant de diverses manières, le fait que l'existence de corrélations empiriques entre les modifications de deux structures physiques ou simplement logiques ne démontre pas l'existence d'une relation causale directe entre elles ; et que, réciproquement, l'inexistence de telles corrélations ne démontre pas l'absence d'un tel lien de causalité. Dans le cas de la génétique développementale, on reste donc simplement, à notre avis, et malgré une remarquable richesse expérimentale, au niveau de la constatation de l'existence de corrélations empiriques entre les modifications du génotype et du phénotype.

A propos de la question du rôle du contexte dans la méthode différentielle et de son interprétation dans le cadre de la métaphore du programme génétique, il y a encore deux remarques à faire. Premièrement, s'il est vrai que le fonctionnement *réel* d'un programme d'ordinateur suppose également un contexte (un système d'exploitation, du matériel d'un certain type), la théorie de la programmation est cependant née *in abstracto*, hors du monde, en tant que théorie de la calculabilité, à partir des Machines mathématiques de Turing (qui n'ont pas besoin d'un contexte *physique*, comme on le rappelait plus haut – c'est l'homme dans l'acte de pensée minimal) ; et cette indépendance en droit à l'égard du monde physique demeure essentielle aux développements pratiques de l'informatique - même de la concurrence et des réseaux - en tant que science du logiciel *portable*, c'est à dire du logiciel qui, dans des contextes opérationnels et physique adéquats, *s'itère à l'identique* (si jamais on fait appel aux aspects physiques des ordinateurs, ce n'est précisément que pour en expliquer le dysfonctionnement et non le fonctionnement réputé "normal" ; en un certain sens, même les problèmes posés par la concurrence dans les réseaux distribués sont justement des "problèmes" auxquels on cherche à remédier par une bonne théorie). Par suite, il semble douteux que l'éventuelle information codée dans l'ADN puisse être interprétée dans des termes empruntés à la théorie de la programmation en tant que telle : le contexte d'expression des génomes, n'est, en effet, certainement pas un lieu passif d'itération à l'identique, mais se laisse, bien plutôt, décrire comme un espace actif de la co-constitution ontogénétique (cf. §3.5), dont *la variabilité est aussi importante que la stabilité*, ce qui, à strictement parler, *n'a pas son équivalent en informatique*. Deuxièmement, la méthode différentielle telle qu'elle est appliquée en physique suppose, nous l'avons vu, de faibles - voire une absence de - corrélations (finement analysées) entre les paramètres contextuels et les variables dont on étudie le comportement en fonction de certains paramètres jugés *a priori* plus pertinents, hypothèse qu'il ne semble pas licite de proposer dans l'analyse des contextes extragénomiques ou épigénétiques de l'ontogenèse. Car l'aspect des systèmes biologiques qui, à

notre avis, constitue le plus grand défi pour les théories physiques courantes consiste donc dans l'existence de ce "champ" particulier, propre à la cellule vivante, où (presque) tout est corrélé à tout et où les effets de "résonance" paraissent infiniment (et l'on utilise ce mot dans un sens mathématique que l'on espère mieux préciser ailleurs) plus complexes que dans n'importe quelle dynamique physique, même non-linéaire.

Plus modestement, cet article se propose en dernière analyse de montrer qu'une éventuelle primauté du génome, même si elle était avérée par delà même ces méthodes différentielles, n'aurait absolument rien de comparable à la *structure de détermination logique propre à la notion formelle de programme*⁸.

Quoi qu'il en soit, pour l'instant de la réponse à ces questions, il nous semble, pour revenir au problème de la succession historique des trois paradigmes de la théorie génétique d'où nous étions partis, que le discours de la théorie de l'hérédité en biologie moléculaire (lois de Mendel y compris) s'est aujourd'hui déplacé pour n'être plus restreint qu'à l'explication des mécanismes moléculaires de réplication et d'expression de l'information génétique dans les massifs cellulaires [Pentris *et al.*, 1983]. Cette remarque et celles qui la précèdent n'ont, en ce sens, pas d'autre prétention que de vouloir indiquer l'existence d'un éventuel biais d'interprétation qui pourrait être introduit dans une relecture *a posteriori* des propositions de la génétique classique avant l'avènement du paradigme moléculaire et, en particulier, de la métaphore du programme génétique. La question que l'on se propose de traiter par la suite est, en fait, celle de la pertinence de cette dernière métaphore et des notions qui lui sont connexes, du point de vue de la compréhension de la manière dont les aspects structurels et fonctionnels caractéristiques des gènes peuvent être synthétisés, avec les seuls outils conceptuels de la biologie moléculaire, en un tout cohérent capable de donner une explication plausible de la manière dont ces derniers *participent à la mise en forme du vivant*.

3.2.3 Un gène - des protéines vs. une protéine - aucun gène.

Il nous semble, d'après la lecture d'un ensemble sélectif - qu'on imagine bien volontiers trop restreint face à la densité des données empiriques - de travaux de biologistes moléculaires contemporains sur quelques dizaines d'années, que les gènes de structure sont aujourd'hui assimilés à des portions d'ADN potentiellement associées, *via* le code génétique, à des protéines dont les acides aminés successifs sont déterminés par la série des triplets de nucléotides qui les constituent. Néanmoins, on remarque déjà que cette définition n'est pas, à strictement parler, une caractérisation purement structurelle des gènes puisqu'elle est fondamentalement indissociable de leur définition fonctionnelle, qui est celle de servir de matrice à la synthèse d'"une" protéine, du moins dans le cas des organismes les plus simples. En effet, dans le cas des eucaryotes supérieurs, la chose semble nettement plus complexe, dans la mesure où il apparaît que, chez ces derniers, les gènes ne peuvent plus être simplement conçus comme des segments ininterrompus d'ADN en correspondance linéaire avec des protéines spécifiques. À cet égard, on ne peut s'empêcher de remarquer ici que le terme de "détrit" généralement employé à propos des

⁸ Cette restriction n'exclut pas, bien évidemment, qu'il existe des cas très locaux (comme ceux, par exemple, de certaines maladies génétiques rares) pour lesquels le lien entre génotype et phénotype relève précisément de ce type de causalité. Toutefois, d'une part il s'agit encore une fois de différences qui induisent des différences, de l'autre il ne va pas de soi que ces cas particuliers soient, en fait, le modèle général sur la base duquel tous les cas possibles pourront être analysés. Il se pourrait même qu'il s'agisse de l'exact opposé : ainsi, en physique, on peut se souvenir que l'évidence aristotélicienne selon laquelle tout mobile particulier s'arrête dès qu'on cesse de le pousser s'oppose précisément au principe d'inertie galiléen, qui seul est théoriquement pertinent.

portions non-codantes des gènes morcelés au sein des génomes des espèces eucaryotes, semble des plus inappropriés : il est clair, en effet, comme tendent à le montrer nombre de publications, que la présence de ces portions non-associées à des protéines *n'est pas indifférente à la capacité d'adaptation des cellules à leur environnement* dans la mesure où, en aval de la transcription des gènes, le phénomène d'épissage alternatif⁹ procure à ces dernières une plasticité certaine du point de vue de la régulation de leurs activités métaboliques, en rapport avec leur fonction organique¹⁰. Or, c'est précisément là que surgit une difficulté fondamentale pour la définition des gènes en termes structurels dès lors que l'on voudra l'interpréter dans un vocabulaire fonctionnel puisque, dans un très grand nombre de cas, il n'existe plus de relation linéaire entre gènes et produits de l'expression de l'information qu'ils contiennent. Du reste, cette relation dégénérée (voir § 4.2.1) est parfois portée à son acmé par l'existence de mécanismes de modification extragénomique de la séquence des transcrits primaires qui deviennent alors susceptibles de conduire à la synthèse de protéines *n'ayant aucun antécédent à l'intérieur du génome de la cellule elle-même*. On pourrait également ajouter à cela que la distinction entre intron et exon n'est même plus pertinente pour la définition du rôle physiologique des gènes aujourd'hui dans la mesure où il est apparu que la synthèse de certaines protéines faisait précisément intervenir des portions d'ADN intronique.

Il existe, néanmoins, une solution face à ce problème de "dégénérescence empirique" de la notion de gène, solution qui consiste à rapporter la définition des gènes, non plus à certaines portions d'ADN, mais à leurs équivalents en termes d'ARN. Pourtant, cette dernière alternative ne saurait être satisfaisante dans la mesure où elle remet sérieusement en question l'interprétation traditionnelle du rôle des gènes dans la transmission des caractères, en particulier du fait qu'il faudrait alors intégrer l'idée problématique que ces derniers puissent avoir, à l'image des segments d'ARN conduisant à la synthèse de protéines, une existence discontinue *dans l'espace et dans le temps*, ce qui invaliderait, de ce fait-même, la primauté absolue accordée à l'ADN dans le processus de développement.

3.3 La notion d'information dans la définition du concept de gène.

Il nous semble que le recours au vocabulaire des théories de l'information au sein de la biologie moléculaire reste généralement très informel. Ainsi, la notion d'"information" génétique elle-même, utilisée en référence à l'existence d'un code - du reste, dégénéré ou plutôt, à strictement parler, redondant (voir ci-dessous) - reliant certaines séries de nucléotides à des acides aminés, semble aujourd'hui incapable de rendre compte du lien entre gènes et protéines, du fait, notamment, qu'elle perpétue, moyennant un jeu de mot aristotélicien¹¹, le dogme central de la

⁹ c'est-à-dire l'ensemble des "mécanismes" de régulation post-transcriptionnels destinés à amener le transcrit primaire à maturation en excisant les segments d'ARN messenger associés aux séquences introniques du gène puis en réarrangeant les portions codantes de diverses manières.

¹⁰ Le nombre des variants protéiques qui peuvent être synthétisés à partir d'un même gène peut ainsi atteindre plusieurs centaines, comme c'est, par exemple, le cas du gène cSlo des cellules ciliées de l'oreille interne du poulet pour lequel ce nombre atteint les 576 (voir [Black, 1998]). Les protéines en question, du reste, ont cette propriété remarquable qu'elles interviennent dans la cellule de sorte à modifier ses différentes fréquences de résonance, ce qui semble bien aller dans le sens d'un gain de plasticité et, donc, d'adaptabilité des cellules de l'oreille interne du poulet.

¹¹ Le jeu de mot en question *est* aristotélicien, au sens où il renvoie à l'idée que les gènes "informeront" les protéines, c'est-à-dire, les "mettront en forme", de la même façon que, pour le Philosophe, le devenir, chez les êtres de nature, se traduit par l'imposition d'une forme sensible <μορφη> - dont le modèle idéal renvoie à une forme intelligible <ειδος> - à travers la mise en place d'une certaine organisation <λογος> qui les détermine en tant qu'actualisation d'une certaine fonction <εν-τελεχεια> au sens le plus large du terme.

biologie moléculaire d'après lequel la fonction d'un gène pourrait être déduite de la seule séquence de ses nucléotides, ce qui, étant donné l'existence des épissages alternatifs, est, désormais, *largement sujet à caution d'un point de vue strictement physiologique*.

De ce point de vue, l'un des meilleurs exemples qu'on peut donner du caractère très informel du recours à ce vocabulaire est certainement celui de l'interprétation des expériences d'inactivation de certains gènes <knockouts>. Il a, en effet, été constaté à de multiples reprises que ce type d'expériences ne conduisait pas nécessairement à la suppression pure et simple de la fonction à laquelle les gènes étudiés avaient été traditionnellement attachés, avec parfois cette conséquence pour le moins inattendue que cette dernière s'en trouvait même améliorée. Et ce fait est précisément interprété comme mettant en évidence la grande redondance des voies métaboliques dans le vivant [Tautz, 1992 ; Thomas, 1993]. Or, si l'on prend au sérieux l'idée d'introduire la rigueur scientifique des théories de l'information en génétique, alors on doit bien convenir que le terme-même de redondance est ici inadéquat dans la mesure où il ne permet absolument pas de rendre compte de *l'originalité de ce phénomène biologique* par rapport aux systèmes cybernétiques. Le concept de redondance s'est, en effet, imposé, en cybernétique, d'un point de vue strictement pragmatique, en vue de réguler l'introduction d'erreurs dans le traitement automatique des messages. La redondance ainsi définie est donc, au premier chef, *une propriété structurelle de l'information* qui n'affecte aucunement sa fonction/signification. Or, dans le cas des génomes, le qualificatif de redondant n'est pas seulement appliqué aux gènes dont la structure est répétée, avec quelques variantes dépourvues de rôle fonctionnel, en vue de la traduction conforme d'au moins l'un d'entre eux, mais aussi à ceux dont la ou les protéine(s) associée(s) *ont une fonction suffisamment proche pour se substituer l'une à l'autre*. Il serait donc plus conforme à l'usage consacré du terme de redondance de parler à leur sujet de redondance fonctionnelle ou plutôt, pour reprendre un terme qui semble très prégnant dans les analyses de Gerald Edelman [Edelman et Gally, 2001], de *dégénérescence*, un concept dont on parlera et qui, par nombre de ses aspects, semble renvoyer au concept de multiréalisabilité en philosophie de l'esprit¹². Mais alors, tout en remarquant l'énorme importance d'une analyse de cet apparent "surplus" informationnel dans le génome, on doit également souligner que l'analogie cybernétique est ici des plus lâches et *ne permet pas d'introduire toute la rigueur du concept de redondance en théorie de l'information*¹³.

¹² On pense qu'il est important de souligner, à cet égard, le fait qu'il n'existe, à notre connaissance, aucun argument probant en faveur de l'existence d'une éventuelle relation *générique* originelle entre structure et fonction, perdue au cours de l'évolution, comme paraissent le sous-entendre certains biologistes moléculaires, et ce à la différence explicite d'Edelman.

¹³ On voudrait dès maintenant souligner la spécificité du concept de dégénérescence en biologie : elle signifie, contrairement à ce que pourrait laisser entendre le paragraphe précédent en première lecture, beaucoup plus qu'un simple "surplus" ou "déficit" informationnel. Ainsi, comme on l'a vu, les expériences de "knock-out" montrent que *parfois* la suppression de pans entiers d'ADN peut n'avoir absolument aucune "conséquence" observable sur le développement des individus, du moins, dans certaines limites raisonnables. Par contraste, il existe aussi certains cas où la modification d'un seul nucléotide "conduit" à des changements phénotypiques importants. Et, à n'en pas douter, il doit exister des cas où ces deux phénomènes sont observables *pour une même portion d'ADN*. Ce qui nous semble fondamental ici, c'est que, précisément, la possibilité de co-occurrence de ces deux phénomènes (*possibilité que rien ne permet en principe d'écarter*) suggère qu'on ne peut pas *a priori* privilégier l'explication d'un seul de ces aspects remarquables du rôle physiologique des gènes au détriment de l'autre, en remettant à plus tard l'explication, par composition du connu, de ce qui reste encore inconnu. Il nous semble, à cet égard, qu'une distinction conceptuelle rigoureuse, non seulement entre redondance et dégénérescence, mais encore, à l'intérieur de cette dernière notion, entre dégénérescence structurelle et dégénérescence fonctionnelle, pourrait nous faire saisir plus clairement ce qui

3.4 La question du statut du concept de gènes dans la métaphore du programme génétique.

Il nous semble également que la notion de programme génétique pose de sérieux problèmes théoriques d'interprétation dans la mesure où, en pratique, elle semble n'être employée que de manière très locale, de sorte qu'*elle ne permet pas de rendre intelligible le rôle physiologique des gènes dans sa toute sa globalité*. Il y a sans doute plusieurs raisons à cela, dont les suivantes.

On se souvient déjà qu'à l'origine, la notion de programme génétique a été développée dans l'horizon des travaux de Monod, Lwoff et Jacob concernant la régulation de l'expression de l'opéron lactose chez *Escherichia Coli* [Jacob et Monod, 1959]. Néanmoins, il est clair que cette dernière notion était plus ou moins implicitement présente dès les années '30 dans certaines réflexions de [Morgan, 1934] et [Goldschmidt, 1938], par exemple, à propos du rôle des gènes dans le processus de développement embryonnaire, à cette différence près, toutefois, que ces derniers pensaient précisément que ce rôle *n'était pas prépondérant*, mais dépendait étroitement de régulations cytoplasmiques. Ce que l'on veut dire en précisant cela, c'est que les travaux de Monod, Lwoff et Jacob *ont contribué à la formulation de la notion de programme génétique en biologie moléculaire*, au sens, par exemple, où ils lui auraient fourni l'assise expérimentale qui lui avait fait défaut auparavant. Ainsi, on peut se souvenir qu'à la même époque, Ernst Mayr avait également introduit cette notion de manière totalement indépendante, à l'occasion d'un discours théorique consacré au problème de la causalité en biologie [Mayr, 1959]. Néanmoins, quelle que soit la nature des fondements théoriques et/ou expérimentaux du concept en tant que tel, il se trouve que ces mécanismes de régulation, isolés dans le cas d'un génome bactérien, sont en fait également légion dans celui des eucaryotes, ce qui laisse à penser que les processus métaboliques au sein des cellules sont soumis à des contrôles du même type tout au long de la vie des individus. Le rôle central de ces mécanismes dans le métabolisme cellulaire est, de fait, absolument indéniable, mais on peut se demander si le recours à la métaphore du programme génétique elle-même était bien nécessaire pour rendre intelligibles ces phénomènes propres au vivant. En fait, on peut faire le constat du fait que, par certains de ses aspects, la notion de programme introduit un biais théorique important dans la lecture des données expérimentales. En voici quelques exemples.

3.4.1 Gènes de structure vs. gènes de régulation.

Tout d'abord, le modèle initial de l'opéron lactose introduit une distinction très ambivalente entre les notions de gènes de structure, lesquels sont supposés n'avoir pour fonction que de permettre la synthèse de molécules nécessaires à la poursuite du métabolisme cellulaire, et de gènes régulateurs, qui peuvent, quant à eux, être de différents types selon leur mode d'intervention sur la régulation de l'expression des gènes de structure. Or, c'est précisément de ce point de vue qu'ils se prêtent assez mal à une interprétation en termes d'information génétique dans la mesure où les portions d'ADN auxquelles ils correspondent ne sont parfois associées à la synthèse d'aucune protéine en particulier et ne servent fondamentalement qu'à favoriser la liaison de complexes moléculaires induisant des modifications épigénétiques de la structure du génome permettant l'ouverture ou la fermeture de certains cadres de lecture adjacents.

Le cas des gènes de régulation fait donc intervenir une nouvelle difficulté dans la définition générale des gènes en rapport avec une activité enzymatique des protéines puisqu'il existe

distingue ces deux aspects remarquables chez les organismes eucaryotes au sein desquels un déficit fonctionnel quantifiable (cf., par exemple, les estimations successives données au cours du séquençage du génome humain) est conjugué à un apparent surplus structurel (les gènes sont morcelés, répétés, peuvent se substituer l'un à l'autre...).

visiblement *des liens entre ces deux entités qui ne font aucunement intervenir la médiation d'un code*. Ce problème n'est d'ailleurs pas exclusivement un problème théorique relatif à la seule définition des gènes en termes moléculaires puisqu'il conduit également à formuler des questions très précises sur le plan pratique dès lors qu'il s'agit de comptabiliser les portions d'ADN susceptibles de se comporter comme des gènes dans le génome des espèces "décryptées" [Bernet, 2001; Boffelli, Nobrega et Rubin, 2004]. En outre, le fait qu'une partie des mécanismes de régulation de l'expression des génomes ne repose pas, à strictement parler, sur des relations analysables en termes de code, au sens formel de la théorie de l'information, mais plutôt sur des interactions physico-chimiques dynamiques, telles que celles étudiées par la chimie statistique, laisse ouverte la possibilité d'interprétations alternatives du rôle physiologique des gènes *à partir du même corpus de faits expérimentaux*. En premier lieu, en effet, rien ne permet *a priori* d'exclure que l'ADN puisse n'être, vis-à-vis du cytoplasme, qu'une simple "banque de données analogique" soumise à un tri par un programme situé en aval de la traduction, à savoir dans les organites qui participent au métabolisme cellulaire, du moins si l'on tient à conserver la métaphore informatique [Atlan et Koppel, 1990]. De fait, on peut également contester, en un second temps, l'aspect déterministe, que nous avons dit être de type laplacien, que la notion de programme *impose a priori* aux relations entre gènes et protéines (qu'il s'agisse de relations de traduction ou de régulation), et ce du fait que *ce type de déterminisme est fondamentalement incompatible avec le caractère probabiliste des prédictions de la chimie statistique*¹⁴ [Creager et Gaudillière, 1996 ; Kupiec, 1996]. Et on ne voit pas de raison physiquement recevable au nom de laquelle on pourrait supposer que les molécules intervenant dans les réactions propres aux phénomènes vitaux échappent à ce type de restriction de l'analyse chimique.

3.4.2 La notion de stéréospécificité.

Il a été longtemps question de l'existence de molécules *spécifiques* capables de réguler la synthèse des protéines nécessaires au métabolisme cellulaire. Il semble toutefois que la très grande majorité des facteurs moléculaires isolés à ce jour n'est jamais, à proprement parler, spécifique mais, bien plutôt, *ubiquitaire et totipotente* de sorte qu'il est fréquent qu'il soit fait appel à certains cofacteurs encore non-isolés pour expliquer *l'absence constatée de spécificité des protéines étudiées*. De fait, le concept de stéréospécificité est-il hautement problématique dans l'interprétation des phénomènes physicochimiques propres au vivant puisqu'il signifie que les molécules qui interviennent dans la régulation de l'expression génétique jouissent d'une *propriété de liaison spécifique* sur des substrats particuliers, *à l'exclusion de tous les autres*. Or, la plasticité des macromolécules est un phénomène bien connu de la cinétique chimique puisqu'il n'est pas rare que ces dernières possèdent plusieurs isomères enthalpiques entre lesquels elles oscillent en permanence sous l'effet de l'agitation thermique *tant qu'aucune contrainte ne leur est appliquée*. Qui plus est, lorsque les contraintes en question consistent en la réalisation de liaisons *non-covalentes* avec certains substrats, la stabilité du complexe tridimensionnel qui résulte de cette interaction *n'est que temporaire* dans une réaction enzymatique. C'est là, du reste, un fait bien connu de la cinétique chimique puisque la plupart des réactions de ce type *sont partiellement réversibles* de sorte que les équations différentielles qui les caractérisent *ne concernent que l'évolution des concentrations et non le comportement des molécules individuelles elles-mêmes*

¹⁴ Ces prédictions sont, en effet, probabilistes dans la mesure où elles concernent le comportement *global* de populations de molécules et non le comportement individuel de chacune de ces molécules, lequel reste *soumis à l'influence perturbatrice de l'agitation thermique*.

(voir la note précédente).

Il n'y a donc, *a priori*, aucune raison fondamentale justifiant le fait que la non-spécificité apparente des protéines du vivant soit un sujet d'étonnement et donne lieu à d'interminables enquêtes empiriques et/ou théoriques puisque l'hypothèse de la stéréospécificité n'est pas une condition *sine qua non* de la validité des analyses de la chimie statistique.

3.4.3 La question du rôle développemental des gènes et le clonage.

On ne saurait, finalement, donner une meilleure illustration du caractère hautement problématique de la notion de programme génétique qu'à travers l'exemple du clonage, auquel il est pourtant bien souvent fait appel pour attester de la maîtrise *technique* du génome des espèces par la biologie moléculaire.

Il est bien clair que, d'une manière tout à fait générale, le problème du développement embryonnaire est sans doute celui qui a tenu le plus longtemps en échec toutes les théories de l'hérédité successives. De fait, l'aspect du développement qui pose les plus nettes difficultés à la théorie du programme génétique est-il ce fait fondamental que les cellules en division sont capables de se différencier sélectivement avant de se répliquer (plus ou moins) à l'identique. Si l'on tient à privilégier le rôle physiologique des gènes par rapport aux activités métaboliques de l'ooplaste au cours du développement, ce processus peut aussi être présenté comme celui de l'*activation différentielle des gènes*, dans la formation de tissus et organes. Le problème que cela pose n'est cependant pas insurmontable *a priori*, ne serait-ce que parce qu'il n'y a pas de raison de principe pour exclure l'existence de certains gènes promoteurs du processus de différenciation lui-même¹⁵. Toutefois, on doit reconnaître aujourd'hui que *de tels gènes n'ont toujours pas été mis en évidence, sauf de manière très locale*, dans le cas de la morphogenèse de certains organes [Goodwin, 1985]. Mais même si de tels gènes avaient un rôle plus général, la question de *leur primauté dans le processus de développement* en tant que tel n'en demeurerait pas moins problématique puisqu'elle ne permettrait pas d'expliquer que le processus de différenciation soit *irréversible*, indiquant, par là, la possibilité de l'existence de modifications épigénétiques du matériel génétique lui-même, lesquelles se révéleraient "somatiquement héréditaires", si une telle notion a bien un sens. Mais c'est fondamentalement le fait que le clonage nécessite le transfert du noyau d'une cellule différenciée dans un ooplaste *adéquat* qui suggère le plus nettement que le déroulement du fameux programme génétique n'est pas indépendant des conditions extragénomiques et, en particulier, de certaines contraintes cytoplasmiques.

La première chose qui semble nécessaire, en pratique, pour "reprogrammer" le noyau d'une cellule différenciée en vue de son implantation dans un ooplaste est, en effet, d'induire une restructuration complète de sa chromatine. Ainsi, le succès du clonage de Dolly par Ian Wilmut n'apparaît-il pas véritablement comme la conséquence d'un enrichissement notable des connaissances en matière de rôle développemental des gènes. Et cela puisqu'en dernière analyse, on ne fait que découvrir aujourd'hui la très grande complexité des relations entre la réceptivité du matériel génétique à certaines contraintes cytoplasmiques et les modifications épigénétiques de l'expression des gènes - et on pourrait presque dire que le clonage ne se justifie que pour élucider

¹⁵ La difficulté est, en fait, ici que *de ce seul point de vue*, c'est-à-dire si l'on se contente de postuler, sans évidence empirique et cadre théorique pertinent, que chaque étape du processus de différenciation est placée sous la dépendance de certaines combinaisons de gènes, la théorie du programme reste infalsifiable, au sens où les explications de la différenciation cellulaire seront au mieux circulaires, et, dans le pire des cas, indéfiniment régressives (des gènes régulateurs qui régulent l'expression des gènes régulateurs...).

précisément ces aspects. La naissance de Dolly fut, d'ailleurs, une surprise pour tout le monde, Wilmut et son équipe y compris, qui lancèrent, à l'époque, l'hypothèse que leur succès reposait sur une méthode particulière de culture des cellules donneuses en l'absence d'albumine, ce qui était censé les avoir placées dans un état de latence favorable à de telles modifications [Wilmut *et al.*, 1997, Lewontin, 1997, Ahouse et Keller, 1997]. Mais, comme l'on sait, il est apparu par la suite que ce régime n'avait strictement rien à voir avec ce succès puisque des résultats similaires ont été obtenus sans le recours à la privation d'albumine.

4. La machine et l'ADN.

Mais alors, peut-on encore croire que l'ADN est un programme dans le sens de l'informatique ? Est-il alphabétique ? Nous reviendrons d'abord sur la notion de programme génétique. Comme le dit fort justement Danchin, il s'agit ici *d'une métaphore*, à saisir comme suggestion pour l'intelligibilité, car l'ADN n'est pas une construction humaine, pratique ou conceptuelle, comme l'alphabet ou la Machine de Turing (MdT). Mais quand on projette en arrière, sur les phénomènes naturels, une construction humaine (l'alphabet, les horloges, les machines...), il faut bien réfléchir, comme nous avons essayé de le proposer, à l'histoire constitutive de ces constructions, très riches en pratiques humaines, voire en logiques internes, qui en marquent le sens et le rôle possible pour l'intelligibilité.

4.1 La règle, le calcul et leur contexte.

Dans les deux premières sections, on a mis en évidence les liens de continuité et les différences entre alphabet et MdT. Machine alphabétique, cette dernière est aussi cartésienne et laplacienne : revenons-y brièvement. Elle est cartésienne aussi à cause du dualisme féroce qui la caractérise : la grande idée de Turing, en fait, consiste (par rapport aux autres formalismes mathématiques pour la calculabilité) dans le départage mathématique très net entre logiciel et matériel. C'est là ce qui démarre la notion moderne de programme et la théorie logico-mathématique de la programmation. Ce départage, purement conceptuel en '35, a produit l'informatique moderne, où la portabilité du logiciel est au cœur de toute application possible, tout comme sa parfaite itérabilité théorique (mais aussi pratique).

L'enjeu central d'une comparaison, voire d'une intelligibilité par métaphores, consiste dans l'analyse des différents contextes d'expression, de l'alphabet, de la machine à états discrets (la MdT), du génome. Car au fond, avant toute expression, il y a un terrain, très important, commun à ces trois structures de l'intellect et du monde : la nature essentiellement discrète du support notationnel, conceptuel ou chimique. Discrète, on y revient, dans le sens mathématique suivant : la topologie discrète organise *naturellement* ces structures (dans le sens informel mais rigoureux de la géométrie, voire de la Théorie des Catégories). Mais *c'est le contexte opératoire, qui engendre le sens, voire le calcul, voire encore les structures du vivant* (le phénotype), suivant les trois cas que l'on compare ici.

On a déjà résumé brièvement la thèse au cœur des analyses comparatives, entre physique et calculabilité, dans [Longo, 2002 ; Bailly, Longo, 2006] : la calculabilité est une théorie de l'itération, à partir de la recursion primitive (Herbrand, Gödel), jusqu'à la portabilité du logiciel (forme des plus remarquables d'itérabilité). Même dans les réseaux concurrents, dont les mathématiques (difficiles) sont en train de se faire, tout est fait pour forcer la certitude de l'itération. L'espace et le temps physiques, même de type relativiste ([Matherat, Jaekel, 2003]), posent des problèmes difficiles à la synchronisation de processus concurrents asynchrones, mais

nous savons tous que si l'on ouvre mille fois une page web lointaine, on obtient, sauf dans des exceptions rares, toujours la même page, à l'identique, indépendamment de l'accès spatio-temporel (et l'on devient furieux s'il n'en ait pas ainsi : l'exactitude et l'indépendance de la physique, matérielle et spatio-temporelle, restent l'objectif de cette extraordinaire construction mathématique qu'est l'informatique). La programmation, séquentielle ou concurrente est faite pour cela : la règle *doit* dominer, sans exceptions. Et s'il y en a, si l'interaction dans l'espace et le temps d'un réseau, imprévue par Turing, peut, bien rarement, gêner, une science des processus concurrents est mise en place pour l'éviter, la contenir. Et on y arrive, avec la force mathématique des règles (les instructions d'un programme), qui s'empilent sur les règles.

Exactitude et itération : la règle dirige le calcul. L'ordinateur n'engendre aucun calcul sans règle, le programme lui-même est un ensemble de règles (*Regulae ad directionem ... calculi*). *L'aléatoire n'existe pas en calculabilité* : les générateurs pseudo-aléatoires sont de petits programmes qui engendrent des suites de 0 et 1, *parfaitement itérables dans les mêmes conditions initiales*, mais qui donnent des bonnes distributions de probabilité dans les espaces de suites (elles *semblent* aléatoires, elles "imitent" l'aléatoire). Tout comme pour Laplace, grand mathématicien aussi des probabilités, le hasard est parfaitement étranger à la détermination par la règle. Or, depuis Poincaré, ou, mieux encore, depuis Kolmogorov, on a compris qu'un processus déterministe (*classique*) est aléatoire quand, *itéré dans les mêmes conditions aux contours, il ne suit pas la même évolution*. Cela n'existe pas en théorie de la programmation séquentielle (de la calculabilité) ; dans les systèmes concurrents, il est rare, il est dû au contexte physique (ou humain) et on l'évite à tout prix (et l'on y arrive).

En fait, la pratique de la programmation concurrente nous donne essentiellement le même type d'évolution que les processus séquentiels du point de vue de la détermination et de la prédictibilité (non pas le même ensemble de processus !), voire tout est fait (sémaphores, "interleaving", voir [Aceto et al., 2003]) pour qu'il en soit ainsi, malgré le manque d'un cadre théorique (unique et) convaincant. Le but, largement atteint en pratique, est d'implémenter le processus imaginé, de façon essentiellement insensible aux variations aux contours, y compris celles qui présentent un problème théorique dans la conception du logiciel, à savoir celles qui sont dues à la concurrence-synchronisation de processus distribués dans l'espace et le temps physique. Bref, les règles sont *en amont* de tout processus digital : voilà la conception classique, newtonienne (voire fréguenne) de *loi physique* (ou logique), absolu de la nature (voire de la pensée) enfin réalisé dans la machine car elle dicte son parcours. Et l'exactitude de l'univers digital garantit la prédictibilité, au moins en termes d'itération : la prévision laplacienne est possible dans la machine à états discrets, en rappelant encore [Turing, 1950]. Et cela puisque les dynamiques sensibles aux variations en dessous de l'observable *sont exclues par principe* dans une structure où il n'y a que du discret, bien observable. Lancez deux fois le programme qui engendre le plus étrange des attracteurs de Lorentz et vous obtiendrez exactement la même image digitale, *ce qui est absolument impossible pour une turbulence physique* - le fait est que *la structure causale a changé en passant à l'implantation sur la machine* [Bailly, Longo, 2006] (et si une fois sur un million il n'en est pas ainsi, mesurez la probabilité de ces exceptions par rapport à la non-itérabilité des turbulences naturelles ! l'ajout d'un générateur aléatoire dans le contexte de cette discussion, où l'on vise l'intelligibilité, frôle moins l'analyse scientifique que l'arnaque, au vu des considérations faites ci-dessus).

Qu'en est-il de cette *insensibilité programmée au contexte* propre à la machine digitale, dans le cas du génome ? On retrouve sans doute des éléments d'une structure discrète dans les séquences des nucléotides, ces lettres qui font penser à l'alphabet, dont on parlera. Mais *le rôle*

que tient le contexte d'expression dans l'analyse des données et l'évolution des processus change radicalement.

Tout d'abord, s'agissant d'une phénoménalité naturelle, il faudrait laisser tomber, on le disait ci-dessus, la conception classique de lois de la Nature : conception normative, dérivée, depuis Aristote jusqu'à Newton, de pratiques sociales et religieuses (la loi civile et pénale, la loi divine). C'est bien cette normativité qui fait la force de la programmation ; mais *ce n'est pas ainsi que procède la physique lorsqu'elle cherche aujourd'hui à rendre intelligibles les phénomènes qui nous entourent*. La co-constitution du sens et l'intelligibilité scientifique sont le résultat d'une pratique de la connaissance qui essaye de proposer des espaces conceptuels (des espaces de phases, des métriques adéquates), qui intègrent le sujet connaissant, dans une polarité active sujet-objet. La relativité et la mécanique quantique ont proposé ce type de regard, comme l'on disait plus haut, et leurs approches révolutionnaires *ont changé la démarche scientifique* : la trajectoire d'une planète *n'est pas* un parcours qui obéit aux lois/équations de Newton, *mais une géodésique dans une variété riemannienne adéquate* ; l'évolution d'un système quantique se comprend en terme d'hamiltonien et de symétries. Que la programmation de la machine digitale suive de vieux paradigmes, c'est d'une part *intrinsèque à sa nature essentiellement artificielle* (nous sommes Dieu pour la machine, nous dictons la loi - avec quelque difficulté à se faire obéir quand on les distribue dans l'espace-temps physique) ; de l'autre, ceci est dû à l'*exclusion historique des nouveaux paradigmes scientifiques dans les analyses post-fregéennes des fondements des mathématiques*, ces fondements logico-formels dont l'invention de la machine est largement dérivée depuis les années '30.

Peut-on, malgré tout cela, transférer ce paradigme scientifique laplacien (la détermination équationnelle-formelle implique la prédictibilité de l'évolution) dans le cadre du vivant, de la génomique en particulier ? Encore une fois, le charme du discret est fortement justifié et on y reviendra, mais le programme (séquentiel, concurrent) est *bien plus qu'une écriture alphabétique* : il dérive, on espère l'avoir assez dit, et contient/impose une organisation du monde et une logique fortes, qui s'expriment dans un contexte de règles formelles, dont *la certitude itérative, voire la portabilité, indépendamment du contexte*, est la propriété principale. Et ce discret génétique, qui est sans doute là, dans la structure de l'ADN, *dans quel contexte s'exprime-t-il ?*

4.2 L'ADN et son contexte.

Au niveau nanométrique, le niveau de son expression primaire, l'ADN est soumis à un tourbillon violent et rapide (avec des variations d'activité, dans les limites de l'intervalle thermodynamique de viabilité) : des milliers de particules bombardent ses nucléotides. De temps en temps, il arrive qu'un nucléotide heurte une base ou une polymérase compatibles, après des milliers de frottements qui n'ont pas aboutis. Mais cette compatibilité, voire complémentarité, n'est pas un oui/non net : il existe des *degrés* de stéréocomplémentarité et des correspondances géométriques grossières peuvent parfois permettre, par adaptation induite, un couplage entre biomères dont la compatibilité peut dépendre de la moindre variation du contexte, par exemple une oscillation thermique. Déjà à ce niveau, la "règle" est floue, adaptive, *fortement contextuelle*, basée sur un processus dont la totalité de l'évolution importe autant que la rencontre locale. Il n'existe pas de programmation informatique, même concurrente, qui démarre de cette façon tourbillonnante (fort heureusement, en vue des buts pour lesquels nous avons inventé les ordinateurs : la programmation est *téléologiquement orientée* et, par exemple, même la notion *physique* "d'attracteur d'un programme" n'aurait pas de sens).

Continuons. Parfois, le système de duplication associe, par exemple, une G à une C ; toutefois, il se peut aussi que l'on ait le couplage GT. Ce duplex moins fréquent peut être gardé, mais il peut aussi être muté. Il n'est pas convainquant d'encadrer ces phénomènes en disant qu'il y a des règles universelles avec des exceptions, même fréquentes, car une petite mutation de ce genre, ou une autre jamais vue dans les laboratoires, peut être à l'origine d'un arbre phylogénétique... comme celui auquel nous appartenons. Il vaut mieux dire qu'à certaines situations correspond un éventail de possibilités, un peu à la manière des généticiens qui utilisent le concept de "norme de réaction" (un génotype peut donner naissance à une gamme de phénotypes en fonction du contexte) mais tout en précisant, de notre point de vue, que la liste des phénotypes possibles associés à une norme de réaction, n'est pas donnée (programmée) *a priori*, dès le début de l'évolution des espèces (!). Ils seraient plutôt des possibles dans un cadre non-stationnaire, dirait-on en physique, presque tous, mais pas tous, équivalents : la sélection, locale (à l'intérieur de la cellule), globale (ontogénétique, voire phylogénétique) fera le tri des ceux qui sont viables à un moment précis de l'ontogenèse ou de la phylogenèse (quelle catastrophe si l'on programmait ainsi !). Et si l'on inventait un jour, en informatique, une méthode de "sélection darwinienne" des programmes dans un cadre concurrent, il faudrait simuler la variabilité du vivant, qui précède la sélection, avec tout son cadre d'individuation et dégénérescence, dont on parlera. Car la variabilité, dans les phénomènes du vivant, *n'est pas un défaut* ; au contraire, ce qui est très important, chez le vivant, *sa singularité physique*, c'est que jamais une cellule n'est identique à la cellule mère. Et cette variabilité est la *condition de possibilité* même de la sélection somatique aussi bien que phylogénétique et, donc, du développement des individus au même titre que de l'évolution des espèces.

Bref, même en physique classique, la règle, la loi, est un instrument de l'intelligibilité, elle n'est pas inscrite dans le monde ni dans un logiciel pré-établi (sauf si on est encore galiléonewtonien ou on fait de la programmation) : *elle est de nature épistémique*. En tout cas, en physique, *une loi avec des exceptions est fautive et doit donc être abandonnée*¹⁶. Il serait absurde de se donner, comme instrument d'intelligibilité, en biologie, une notion de règle qui ne permet pas de comprendre une grande partie de ce qui compte. En fait, en biologie, il n'y a presque jamais de règle parfaitement rigide, à la Newton-Frege-Turing, avec des exceptions, mais il y a plutôt *des possibles dans un cadre globalement (relativement) stable, mais pas trop*. Les marges mobiles d'un attracteur, dont les composantes suivraient des trajectoires viables à l'intérieur de l'attracteur, donnent peut-être une meilleure image des processus ontogénétiques et phylogénétiques, tout comme les marges d'une criticité étendue dont on parlera. Mais il faut aussi ajouter à ces métaphores physiques, *la non-stationnarité de l'espace des phases*, qui se forment au cours du processus évolutif lui-même - et cela en contraste avec *toutes* les théories physiques courantes (voir [Bailly, Longo, 2006]).

Plus encore : en informatique, un résultat différent de la détermination prédictible par rapport à la règle programmée, est tout simplement une erreur. Pour cette raison, par exemple, on ne fait pas confiance aux ordinateurs pour qu'ils évoluent tous seuls et l'on préfère payer pour acheter du nouveau logiciel¹⁷.

¹⁶ Quant aux "exceptions" en informatique, il s'agit d'une notion présente en programmation, qui fait référence à des bifurcations déterministes, mais sans criticalité physique (avec ses fluctuations associées), donc encore de type laplacien – des lois avec des alternatives prévues.

¹⁷ Les fantaisies de l'intelligence artificielle classique à cet égard ont été un des plus grands gaspillages d'argent de l'histoire de la science : on pourrait citer de nombreux projets, à partir de 1956 aux USA jusqu'à la "cinquième génération" japonaise ; la biologie moléculaire serait-elle sur la même voie, encore une fois ancrée sur le mythe de la

Comment rendre intelligible ce mélange singulier de stabilité globale (l'individu vivant en formation ou formé) et instabilité locale, couplée parfois à leur dual : une extrême rigidité locale (certaines complémentarités stériques sont très rigides) et une plasticité globale (celle d'un organe comme le cerveau, même au cours de la vie adulte), ainsi qu'une stabilité dynamique globale et des variations locales en cascade ? Même la physique des plus dynamiques des systèmes paraît insuffisante : l'invariance mathématique en est toujours le cœur et la variation, si bien saisie dans son importance (les coefficients de Lyapounov de certains systèmes décrivent très bien le rôle de la perturbation/variation minime, dans le temps), est bien différente de la variabilité biologique. Celle-ci se base sur l'*individuation*, notion étrangère à la physique, tout comme la notion de pathologie.

Pour conclure : individuation, variabilité, cadre de stabilité changeante, instabilité maintenue dans des limites plastiques (le pathologique, le déviant peut devenir un nouveau parcours évolutif), non-stationnarité de l'espace des évolutions possibles et ce à partir de l'action du génome, suite discrète *immergée dans une cellule*. Tout cela reste donc à saisir, peut-être avec un jeu complexe et à inventer de discret et de continu mathématiques, avec des invariances propres et des stabilités où le fonctionnement utilise le dysfonctionnement, bien loin d'une théorie pure du discret arithmétique comme la calculabilité.

4.2.1. La dégénérescence et le programme.

Mais même ce discret, si important dans le phénomène du vivant, a une propriété bien singulière, pour la calculabilité, voire pour la physique : *la dégénérescence*. Cette notion a été introduit par Edelman et Tononi, reprise par d'autres auteurs [Edelman, Tononi, 2000 ; Edelman, Gally, 2001] relativement au fonctionnement cérébral ; elle a son origine dans les travaux d'Edelman au sujet du système immunitaire. Bref, une structure est dégénérescente, si des sous-structures non isomorphes peuvent donner lieu à une même fonctionnalité et une structure donnée peut s'exprimer dans plusieurs fonctionnalités. Ce concept reprend et généralise celui de *redondance*, bien connu en théorie de l'information, qui est une théorie de la communication [Shannon, Waever, 1975], et en informatique, mais il en diffère très nettement. Dans cette perspective, on peut distinguer entre *dégénérescence "fonctionnelle"* (des systèmes non isomorphes participant à une seule et même fonction) et *dégénérescence "systémique"* (un même système participant à des fonctions distinctes), voir [Bailly, Longo, 2006]. Dans [Edelman, Gally, 2001], on observe que la dégénérescence est *ubiquitaire dans les systèmes du vivant*, à partir du génome. Même donc sur une structure discrète, elle nous propose une causalité bien particulière. Il faut considérer, en plus, que, en biologie, quand on dit « même » (fonction, par exemple), cela ne veut absolument pas dire « parfaitement identique », comme on peut le prétendre pour un objet mathématique, une fonction calculée par un programme, un programme, une suite de nombres... Pour la dégénérescence biologique, des sous-structures non isomorphes (bien qu'"analogues", dans un sens sans doute à préciser selon les contextes) ou une structure donnée engendrent *presque* la même fonction, dans un contexte similaire. Une petite fluctuation dans le contexte, même d'origine physique, peut alors engendrer dans le temps une expression, une fonction par exemple, très différente. Cette analyse causale de la variabilité du vivant est compatible, mais enrichit nettement celle basée sur la notion de mutation (les deux n'ont pas de sens en théorie de la programmation et elles sont très peu utilisées en pratique, que l'on sache... fort heureusement).

En conclusion, loin donc d'apporter une vision claire du vivant, la notion de programme

règle, du programme, du calcul laplacien qui tout engendre ?

génétique, en s'appuyant sur l'hypothèse selon laquelle il existerait des molécules spécifiques (plutôt que dégénérées) capables de réguler la synthèse des protéines nécessaires au métabolisme cellulaire, n'a pas manqué de se heurter, en pratique, à de nombreuses difficultés, dont quelques unes ont été rappelées ici et dans le § 3.

En particulier, d'un point de vue qu'on qualifierait volontiers de nominaliste, il n'est pas très convaincant, au sein d'une discipline scientifique, d'avoir toujours recours à des explications qui tiennent généralement de la simple description de phénomènes et sont interprétées (voire théorisées) en des termes empruntés, de manière problématique, à d'autres disciplines sans en avoir importé, en même temps, la rigueur méthodologique. Car ces explications présentent alors le défaut majeur de multiplier le nombre des cofacteurs cachés (des variables cachées ?) dès que l'expérience ne semble plus se prêter docilement aux prescriptions du modèle.

On a, en ce sens, déjà eu l'occasion de souligner que, compte tenu de ces éléments, la notion de programme génétique, bien loin d'avoir renforcé le rôle des gènes dans les processus biologiques, a sérieusement érodée la possibilité-même d'en donner une définition univoque en termes d'information génétique, ne serait-ce que parce que, non seulement, les gènes ne peuvent plus être sans ambiguïté associés à une succession de nucléotides codant une protéine, mais encore, parce que *même si tel était le cas*, la succession des acides aminés au sein de la protéine en question ne permettrait pas de déterminer la fonction qui lui est associée, à *moins de faire référence aux contraintes qui seront appliquées au niveau du cytoplasme*. Et, de fait, il semble bien aujourd'hui qu'il soit impossible de faire l'économie d'une référence aux contextes épigénétique et extragénomique dans lesquels s'exprime le genome, avec toute leur immense complexité biologique, pour en donner une interprétation minimalement cohérente [Allis et Junuwein, 2001, Hieter et Boguski, 1997, Turner, 2002].

4.2.2. Un dernier point de méthode à propos de la notion de programme : la question du finalisme contingent.

Dans la critique que nous avons esquissée de la méthode différentielle en biologie moléculaire (§ 3.2.2), il a été question d'une incomplétude théorique du point de vue de l'inférence de "lois causales". En particulier, nous avons mis en évidence le fait important, à nos yeux, d'un point de vue purement logique que de simples corrélations empiriques entre certaines modifications de l'ADN et des différences observables dans le phénotype des individus ne donnent pas de façon immédiate et/ou évidente une "loi" qui corrèle causalement les diverses structures de l'ADN à leurs "conséquences" réputées normales. Il faut souligner, ici encore, la spécificité de la méthode différentielle en physique, dès lors qu'il est question d'établir formellement une dépendance causale quelconque entre variables. En effet, outre les analyses de corrélation/décorrélation de variables dont nous avons déjà parlé, les physiciens ont généralement besoin de se doter d'un cadre d'interprétation théorique essentiel à la construction de l'objectivité scientifique.

En premier lieu, le nombre fini d'expériences et, par suite, le caractère discret des données ne permet de proposer une *unique* équation que moyennant un certain critère d'"optimalité conceptuelle" : autrement dit, on cherche l'équation (le polynôme, la courbe, ...) ayant la forme mathématique "la plus simple". Il est évident que la validité de ce critère *n'est pas un absolu*, mais *son explicitation* est au cœur de la force méthodologique de la physique, depuis Copernic et Kepler. En second lieu, la méthode différentielle en physique trouve, dans une large mesure, sa justification mathématique dans les méthodes variationnelles du calcul différentiel : de petites perturbations (infinitésimales) permettent de reconstruire des géodésiques. Dans les deux cas, on le voit, l'analyse mathématique présuppose un cadre continu - des "surfaces" spatio-temporelles (voire conceptuelles) lisses (smooths) - où ces critères d'optimalité ont un sens et une cohérence

strictement physico-mathématiques.

Or, ce cadre fait précisément défaut dans les analyses différentielles en biologie, sans doute à cause de l'immense difficulté à démêler, du fait de leur singularité physique, les relations causales "intriquées" caractéristiques du vivant, qui nous paraissent sortir même des plus dynamiques des cadres théoriques de la physique. Et, en aucune façon le "modèle" (la métaphore) du programme ne fournit un tel cadre du fait de ses racines computationnelles, cuasalement laplaciennes. On a déjà rappelé, en fait, que ces dernières trouvent leur source dans une théorie de la calculabilité (séquentielle ou, pour le peu qui existe, de la concurrence) dont les conséquences au niveau des régimes de causalité ont été analysées ailleurs [Bailly, Longo, 2006] et sont très éloignées de la smoothness des théories physico-mathématiques auxquelles on fait implicitement référence en biologie, en proposant une analyse différentielle (dans une note on mentionnait que la "sémantique différentielle", en informatique, n'existe pas et pour des bonnes raisons). En somme, le "défaut" en matière d'explication qu'on a relevé en génétique ne tient pas à l'analyse différentielle elle-même, une technique qui nous a donné, au cours du XX^e siècle des informations et résultats d'un immense intérêt, mais il doit, à notre avis, être imputé aux tentatives de l'encadrer dans des notions plus ou moins naïves de "programmation génétique".

Mais alors, on peut se demander ce qui pousse tant de biologistes, du plus haut niveau scientifique, vers ce mythe du programme génétique. Il est clair que la structure discrète de l'ADN, si bien décrite par des suites de lettres qui ne manquent pas d'évoquer l'encodage d'un langage formel, suggère une telle analogie par le jeu métaphorique d'une représentation commode. Toutefois cela ne suffit pas à justifier en soi la référence à la notion de programme en tant que telle. Sans doute le besoin, si fort en biologie, de trouver une justification (voire une explication prétendument réductionniste) au *finalisme* des processus du vivant transparait-il ici bien plus nettement. Par le biais d'un jeu de langage autour de la notion formelle de programme, on glisse subrepticement, en effet, vers une conception du vivant plus ou moins explicitement imprégnée de finalisme (cette fameuse maîtresse que l'on ne montre pas, mais dont on ne peut se passer...). Et voilà, salvatrice, la notion de programme, notion laïque qui, sans besoin de Dieu, contient *une fin* dans ses lignes de code si bien matérialisées dans nos machines modernes. Le programme informatique est fait pour... , il a un but, mais il est, ou peut être, encodé matériellement. Toutefois, nous ne connaissons pas de programme informatique fonctionnant qui n'ait été auparavant tapé sur un clavier : est-ce alors Dieu qui tape sur le clavier moléculaire ? Certes non : c'est l'Evolution nous dira-t-on... Deus ex-machina, l'évolution vient ici combler les lacunes explicatives de la métaphore du programme elle-même, en rendant apparemment intelligible l'inscription dans la matière d'un projet pour le futur... Du moins si l'on exclut la question hautement problématique d'un point de vue strictement physique de l'origine du matériel génétique en tant qu'il constitue la condition de possibilité de cette inscription elle-même.

On nous dira que nous sommes peut-être en train de finasser, car il s'agissait ici seulement d'une question de langage. Le problème, néanmoins, est que ce langage tend à s'imposer comme *le* paradigme dominant dans la communauté scientifique, y compris pour ce qui est de la formulation de projets de recherche qui, semble-t-il, devraient pourtant rester raisonnablement ouverts à la possibilité d'une faillite des paradigmes (on se demande, en effet, comment la science pourrait bien progresser s'il en était autrement...) : on cherche, par exemple, majoritairement et depuis très longtemps, la cause des cancers dans des mutations d'un ADN soudainement mal programmé et c'est seulement depuis quelques années que l'attention de certains audacieux s'est déplacée *aussi* vers le contexte d'expression des génomes.

Mais revenons-en au problème de la finalité, en essayant de l'aborder d'un point de vue physicaliste possible. Considérons un rocher qui forme une pente avec des irrégularités données

et sur lequel la pluie, par l'action répétée au cours du temps de frottements et de petites turbulences, produit graduellement un creux. Il est clair alors que, si les conditions auxquelles le rocher est exposé sont suffisamment constantes au cours du temps, ce premier creux va, à son tour, engendrer une nouvelle turbulence qui, sans doute, fera elle aussi apparaître un deuxième creux, en quelque sorte "fils" du premier. Mais alors y aurait-il vraiment du sens à dire que la première encoche est un *programme*, écrit par l'évolution physique, pour engendrer la deuxième ? Ou encore, qu'elle contient de *l'information* sur la forme, des encoches futures avec leur apparente spécificité structurelle ? Sur cet exemple on voit bien qu'on peut *toujours*, d'un point de vue simplement descriptif, "fabriquer" une finalité apparente en jouant le jeu rhétorique de la programmation ou en utilisant la notion, si chargée dans la culture aristotélicienne, d'information ; mais cela rend-il les phénomènes ainsi décrits *physiquement* intelligibles ? Il n'est pas question ici de soulever la question hautement problématique du rôle éventuel de la finalité dans les sciences, mais l'exemple de l'attitude historique de la physique face à cette question pourra peut-être nous montrer qu'il existe des alternatives, même *physicalistes*, à la notion de programme pour rendre compte d'une apparente finalité dans les systèmes étudiés par la biologie.

Jusqu'à la première moitié du XIXe siècle, en effet, nombres d'auteurs voyaient dans le critère d'optimalité des trajectoires physiques une forme de finalité en acte au sein des sciences de la Nature : le grave ou la lumière était ainsi réputé "choisir" la trajectoire optimale pour "aller vers...". C'est seulement avec l'avènement des méthodes variationnelles que ce type de raisonnement a finalement été abandonné au profit d'explications au sein desquelles les géodésiques étaient induites (causées) par la structure de l'espace elle-même (espace physique ou, plus généralement, espace des phases) . Exit, dès lors, le finalisme en physique... Toutefois, entendons-nous bien : il ne s'agit ici que de discréditer un certain type de discours sur le vivant qui prétend répondre à la question du finalisme en l'intégrant, par un simple jeu de langage, à la description de ses objets, tout en se prétendant physicaliste, d'une façon, pour nous, bien paradoxale. Mais la question de la finalité éventuelle propre au vivant va, à notre avis, certainement bien au delà de celle d'une relation possible entre gènes et caractères, même simplement au niveau moléculaire.

Il ne nous paraît pas extravagant, de ce point de vue, de se demander si une étude des propriétés des systèmes biologiques à partir des méthodes de la physique ne doit pas accepter, malgré qu'on en ait, une certaine forme minimale de finalité. Toutefois, il sera sans doute plus sage de réserver cette inévitable et toujours renouvelée référence au finalisme en biologie au passage peut-être le moins accessible à la connaissance, à savoir celui de la formation origininaire du vivant ; et cela précisément, pour employer une expression chère aux informaticiens, par *défaut* de connaissance¹⁸. Mais pour le reste, il faudrait abandonner, surtout si l'on prétend garder une perspective authentiquement physicaliste, cette référence obsessionnelle à la finalité implicitement associée au terme de programme et, par suite, à celui de gène, identifié à ces

¹⁸ Dans le cadre de l'immense débat sur le finalisme en biologie, et dans cette recherche de minimalité, nous osons faire référence ici à celle dont on parle dans [Bailly, Longo, 2006]. Nous pensons à une finalité "contingente" car matérielle, non programmée et qui pourrait être autre qu'elle n'est, voire ne pas être ; une finalité qui constitue, en outre, le fondement implicite de tout discours sur le vivant, à savoir : *la survie* (de l'individu, de l'espèce). Dès la constitution de la première structure vivante, en effet, passage évolutif sans doute le moins connu, ce qui a compté dans cette formation physique inédite et singulière et l'a définitivement différenciée en tant que telle du reste de la matière inerte - même si, par ailleurs, cela n'était pas nécessairement prévu à l'origine ou inscrit dans un projet - n'est pas autre chose que sa capacité à survivre. Sinon, cette matière ne saurait précisément pas être considérée comme vivante, ou elle ne serait plus là. Bien évidemment, il s'agit maintenant de cerner cette finalité contingente dans un cadre plus précis, par exemple celui de la criticité étendue auquel on travaille (voir [Bailly, Longo, 2006]).

“encoches” de l’ADN, sa structure chimique ! Le système vasculaire est, de ce point de vue, tout aussi finalisé que la canalisation qui se forme peu à peu par accumulation d’encoches sur un rocher, à ceci près, toutefois, que, à la différence du canal géodésique creusé par l’eau dans une structure physique (un espace des phases) qui lui a pré-existé, le système vasculaire se forme *en même temps* que le reste du corps... Et c’est là, sans doute, que vient se greffer cette mémoire matérielle qu’est l’ADN avec son contexte cellulaire, mémoire qui semble constituer, par le jeu de frictions répétées tout au long de la phylogenèse et de l’ontogenèse, une caractéristique spécifique aux processus biologiques, tout à fait étrangère aux dynamiques physiques usuelles.

En conclusion, remplaçons la pente physique du rocher (voire la *cause* de cette pente, parfaitement inconnue en biologie), par cette finalité contingente, une seule, la survie, et essayons de construire des espaces conceptuels, autant que possible de type physico-mathématiques, voire proprement biologique, pour une intelligibilité qui évite d’insérer une fin, de type informatique, dans l’ADN. Comprendre cette spécificité est sans doute l’un des plus grands enjeux à venir de la biologie et amènera très certainement l’émergence de théories en marge des théories physiques existantes, comme nous le suggérerons dans la conclusion.

5. L’alphabet et l’ADN

Quel sens donner, maintenant, à la métaphore “l’ADN, structure alphabétique du vivant” ? Certainement cette métaphore engage-t-elle moins que celle de Machine logico-formelle, dans le sens d’Aristote-Descartes-Newton-Laplace-Turing (!) ; un sens trop fort, trop riche d’histoire et de déterminations le plus souvent incompatibles entre elles. Ainsi, dans une certaine mesure, on peut défendre l’idée selon laquelle les lettres sont bien là, dans l’ADN, que leur structure discrète et “sans signification” est bien visible, et qu’en un certain sens elles composent même des mots (les gènes ?). Mais que et comment signifient ces mots ? L’alphabet, on l’a dit, *s’exprime par le phonème* : la parole, le chant fait le lien avec un contexte de signification. Et il y a, en ce sens, des efforts remarquables des grammairiens pour nous donner des “règles” pour passer des suites de lettres au phonème - et de là au sens - et réciproquement. Mais ces règles, dans leur modeste généralité, sont *des co-constituées, des géodésiques de l’histoire* : elles ne préexistent pas à la langue. L’écriture alphabétique a figé la parole, le flux continu de la communication, tant que faire se peut avec les langues vivantes. Dans la fixation pratique, longue construction historique, qui est passée par le symbole magique, communicatif, signifiant, l’idéogramme, la syllabe et, enfin, le pitch vocal isolé par la consonne écrite, on a extrait les invariants, les règles, un “*a posteriori*” d’une pratique de la communauté humaine communicante. Des règles plus au moins stables, sûrement pas parfaitement stables et forcément incomplètes. Les règles de lecture capturent un fragment des sons possibles dans une langue : le fragment qui deviendra la langue savante ou standard. Un dialecte est souvent impossible à écrire : l’usage phonétique dominant exclut les autres, qui n’ont pas de règles d’écriture ou qui violent, pour s’exprimer, les règles. On y verrait presque l’éventail des possibles de l’expressivité génomique. En fait, les langues vivantes sont ... vivantes, tout comme la cellule. Un flux dynamique similaire, globalement (relativement) stable, mais capable, dans le temps, de changements radicaux ; localement intégré à des exceptions, qui ne sont pas nécessairement des monstres, car elles pourront engendrer une expressivité nouvelle qui peut devenir la nouvelle règle, voire une nouvelle langue. Voilà ce paysage de trajectoires possibles, voire de parcours génériques : tous les possibles sont peut-être explorés, dans l’histoire de l’expression phonétique-signifiante d’une langue écrite. Les règles, les lois, seraient donc des co-constituées de la friction de structures, à la fois rigides et plastiques,

le résultat de pratiques historiques.

Mais alors, oui, gardons pour le moment la métaphore "l'ADN est un alphabet", dont les règles, très peu stables, ont été constituées à partir de certaines exceptions essentielles à la dynamique globale, où la différence entre règles et éventails d'exceptions possibles serait essentiellement statistique, selon des distributions changeantes au cours de l'évolution. Si les langues naturelles ont des ressemblances avec ces évolutions historiques, rien de comparable n'est décrit en physique ; même des cascades de phénomènes critique, qui dépendent d'une histoire, ne permettent pas de rendre compte de notions aussi fondamentales pour la biologie que celles de fonction ou de pathologie, au coeur des évolutions possibles.

Un alphabet donc immergé par l'interaction de "biomères" dans un contexte évolutif, dans le sens riche que l'on vient de décrire, la cellule, l'organisme vivant - tout comme le phonème immerge l'alphabet dans *des formes de vie, avec leur historicité*.

Tout comme l'alphabet, l'ADN est incomplet : il est loin de contenir toute l'information nécessaire à l'embryogenèse. En fait, avec l'alphabet phonétique, on ne capture pas, disait-on plus haut, les *variantes* dialectales d'une langue, parfois essentielles à l'*expressivité* d'un discours ; de même, une lecture mécanique, selon des règles strictes, ne transmet pas les émotions, les sensations, bref tout ce qui compte le plus dans la communication humaine - et aux intonations, aux nuances des accents, nous ajoutons toujours le sourire (qu'on saisit même à la radio), le geste qui peut radicalement *changer la signification* d'un discours. Parfois, la lecture mécanique devient incompréhensible, bien avant même qu'incomplète. L'alphabet est donc essentiellement *incomplet par rapport à l'interaction signifiante, le but de toute langue*, tout comme les formalismes logiques sont démontrablement incomplets par rapport à la preuve (et au sens structurel en) mathématique. Dans les deux cas, de l'ADN et de l'alphabet, l'information pertinente se trouve aussi (voire surtout) *dans le contexte*: la cellule, voire l'organisme avec ses cascades de régulation (hormonales, nerveuses), dans le premier cas, les formes de vie de la communauté linguistique dans le second, tout cela étant situé *en deçà et au delà des règles de la structuration discrète*. En fait, *l'information se constitue dans la friction entre ces alphabets et leurs contextes très dynamiques*. Et l'alphabet aussi, à sa façon, est dégénéréscence : une même phrase écrite dans des contextes d'expression, écrits, oraux, différents, peut avoir des significations différentes. Un même sens peut être convoyé par des phrases différentes.

La métaphore alphabétique peut donc être suggestive pour l'ADN : sa structure discrète est bien saisie par l'analogie, pourvue que l'on développe une théorie non-naïve des langues, avec leur richesse, leur historicité, leur polysémie, leurs branchements dans des espaces de signification très vastes. Une géométrie du sens, que certains linguistes envisagent aujourd'hui, peut en donner une idée. Mais, dès que l'on propose une théorie à la fragrance fréguenne des langues (à la Chomsky, par exemple) et du génome, la règle aristotélo-cartésienne de la Machine, en amont de tout processus, avec sa rigidité prédictible, laplacienne, *empêche le regard scientifique moderne sur deux dynamiques des plus riches au monde* ; jusqu'à trouver le gène de l'infidélité conjugale (Young et al., *Nature*, 400, 766-788, 1999) mélange ordinaire de plaquage informatique et de prédestination protestante au Bien ou au Mal...

5.1 ptqqpdr, wtpss, zwtsss...

Toutefois, toutefois, il y a un petit problème avec cette analogie, même en la poussant dans le sens d'une plus grande attention à la richesse et à l'historicité "naturelle" des langues et de leur écriture, vers une sémiotique des gestes qui intègre toute l'expressivité de l'interaction humaine. Nous n'insérons jamais (presque jamais ?) des ppqqrr, wtwss, zwzsss... au beau milieu d'une

phrase écrite, et très rarement dans un mot. Par contre, d'après l'état des connaissances en génomique, dans près d'un tiers des "mots" du génome des eucariotes (à moins que par "mots" on entende les seules parties codantes des gènes, mais alors on perdrait tout espoir de pouvoir en donner une définition strictement structurelle...), un pourcentage important de l'écriture - parfois jusqu'à 95% (!) des séquences - paraît être strictement non-codant.

De plus, contrairement aux inscriptions sur le marbre des anciens (illisibles pour le profane), nous laissons, dans nos langues, des espaces bien marqués entre les mots. Et l'on utilise des points bien clairs, des virgules et maints autres signes perceptibles, pour marquer le début et la fin d'une phrase signifiante (à notre époque ; ça n'a pas été toujours ainsi). Quoique le sens soit toujours à repérer dans le contexte richissime dont on parlait, un marquage des limites des mots et des phrases caractérise toute expression écrite moderne. Malgré maintes tentatives, il nous semble que les biologistes moléculaires rencontrent des difficultés autant pratiques que théoriques pour repérer le début et la fin d'un gène *en tant qu'unité expressive autonome*. En pratique, ces deux difficultés deviennent d'ailleurs colossales dans le cas du patrimoine génétique des plantes : il apparaît, semble-t-il, que les redondances, le mélange des parties codantes et non-codantes, les formes étendues de dégénérescence, servent ici à compenser, par le biais d'une grande richesse génomique, la modeste mobilité des plantes, incapables à s'adapter par la fuite à un environnement changeant. On ne connaît pas d'alphabet historique qui ait remédié d'une façon similaire au sédentarisme de son peuple créateur (laissons tomber la programmation, car vraiment personne n'écrit ptqqpdr, wtpss, zwtsss... au beau milieu d'un programme, quoique Microsoft ait parfois utilisé cette ruse pour compliquer les interfaces et gêner la concurrence).

Si l'on exclut donc d'improbables astuces de Dieu, pour empêcher les martiens de copier ou d'interférer avec notre génome, on voit mal comment on pourrait rendre compte, dans les termes d'une construction humaine en partie "artificielle" telle que l'alphabet, de nombreux aspects de l'ADN, remarquables de par leur singularité physique. Et tel est bien le nœud du problème : la structure de nos langues, qui est au fondement de notre culture, est un constitué historique, continuellement révisé, en tant que tel, par l'invention de l'alphabet, entre autres, mais aussi par l'action consciente des académiciens et des grammairiens, qui effacent en partie cette histoire et peuvent conduire à croire que l'alphabet et la grammaire, même simplifiés, ont précédé la langue, comme leur détermination formelle. Il s'agit bien évidemment d'une illusion rétrospective : la mise en place d'un alphabet et d'une grammaire, répond à un aspect particulier, propre aux structures (partiellement) artificielles de nos langues, qui se base sur et propose un paradigme culturel *de simplicité* (voire de minimalité) au niveau de ses composantes élémentaires.

Toutefois, ce n'est pas au niveau de ces briques élémentaires du langage que l'on va pouvoir situer, que ce soit d'un point de vue simplement logique ou d'un point de vue dynamique au sens physique du terme, le niveau phénoménal pertinent pour une analyse de la façon dont les mots véhiculent un ou plusieurs sens, mais, bien plutôt, *dans l'épaisseur historique des langues*. Celle-ci interdit d'emblée tout nettoyage simplifiant et parfait, un debugging formel dirait un programmeur (pensez à l'absurdité apparente de la grammaire et de l'orthographe français, riches d'histoire, de redondances, d'ambiguïtés et d'inutilités, de détrit - que les collègues français attribuent plutôt à l'anglais). Comme disent les computationnalistes : les langues sont des langages logico-formels mal ficelés (quel incompetent ce Bon Dieu !). L'évolution biologique, tout comme celle des langues, se nourrit de son histoire et, en même temps, *efface son origine* : voilà une des raisons pour lesquelles la biologie ne peut pas être simplement une physique avec une histoire. Et cela laisse entrevoir quelques éléments d'une analogie métaphorique possible entre langue et génome.

Toutefois, disait-on, on sera encore bien loin d'obtenir, dans le même temps, un apport explicatif véritablement opérationnel, ne serait-ce que sur le plan mathématique : les ambiguïtés structurelles/fonctionnelles, la polysémie, la pléiotropie, la redondance, la dégénérescence du génome et, bien sûr, l'efficacité évolutive qui leur est associée sont certes quantitativement et qualitativement bien différentes de celles des langues. Elles réclament, de la même façon, bien plus que l'importation de paradigmes mathématiques ayant déjà constitué par ailleurs leur domaine de validité, *l'invention d'un cadre théorique d'analyse et d'un vocabulaire autonomes*, en rupture avec, mais sans nécessairement l'invalider, tout ce qui a déjà été fait dans d'autres disciplines scientifiques. Qui niera, par exemple, que tel a été le cas des théories de l'information dans les années '50 ?

5.2 Les mutations.

Nous avons discuté plus haut (§3.2.2) de certains problèmes soulevés par l'interprétation en termes causaux des expériences de génétique développementale qui permettent de modifier le phénotype des individus à partir de certaines permutations effectuées dans l'organisation de leur génome. Dans des cas de ce type, donc, on déplace ou on rend silencieuses des portions plus ou moins vastes d'ADN et l'on observe des modifications associées de certains caractères observables sans pouvoir véritablement rendre compte, *dans un vocabulaire strictement physicaliste*, des *liens de causalité* permettant de passer du génotype au phénotype. Si l'on tient à forcer, ici encore, la métaphore linguistique, on pourra donc dire qu'on réussit à modifier la signification d'un discours par le simple jeu d'une permutation de mots ou de phrases entières. Mais qu'en est-il, pour revenir à la question des méthodes différentielles à un niveau plus élémentaire, de la pertinence de la métaphore alphabétique elle-même du point de vue des corrélations observées entre gènes, en tant que "mots" de l'ADN, et caractères ? On s'intéressera en fait ici aux problèmes d'interprétation soulevés par la notion de mutation, entendue, au sens strict de la biologie moléculaire, comme le changement ou la délétion d'une base, dans les contextes respectifs de l'ontogenèse et de la phylogenèse.

On sait que, d'un point de vue historique, l'une des questions les plus délicates laissées ouvertes par Darwin a sans doute été celle de *l'origine des variations spontanées dans les caractères des individus*. Il va sans dire qu'il en allait de la complétude de sa propre théorie de l'évolution puisque c'étaient précisément ces variations qui étaient censées permettre à la sélection naturelle d'opérer comme un facteur de pression dans la lutte pour l'existence à l'intérieur des populations. *Grosso modo*, ce n'est véritablement qu'avec les travaux de Morgan, Sturtevant, Muller et Bridges sur la drosophile, dans la première partie du XXème siècle, que va commencer de s'élaborer une première ébauche de réponse à cette question sur la base d'une réinterprétation "matérialiste" de l'hérédité sur la base de la théorie des mutations issue des travaux de de Vries. Or, la notion centrale de mutation dans la génétique classique se prête à de nombreuses objections critiques qui ne sont pas étrangères à celles que Morgan lui-même avait précisément adressées à la méthodologie de Mendel. Il lui reprochait essentiellement, en effet, d'avoir fait du découpage en caractères une simple affaire de langage, en particulier en s'autorisant à décomposer en sous-caractères inobservés tout caractère qui ne se distribuait pas selon les fameuses lois qu'il avait édictées dans le cas des expériences d'hybridation. En fait, ce que Morgan dénonçait dans cette "astuce" de la méthode d'analyse mendélienne, c'était précisément son caractère de *clef universelle pour répondre à tous les problèmes de la théorie de l'hérédité*. Avec cette dernière remarque, il est clair que Morgan remettait en question la

scientificité-même de la théorie de Mendel, avec une critique très fine sur le plan épistémologique puisqu'en un certain sens, elle revient à souligner *l'inadéquation du seul critère de validation expérimentale des théories*. Pourtant, la solution que lui-même a proposé ne semble pas satisfaisante, et ce pour une raison qui rejoint la critique que nous avons déjà proposée des méthodes différentielles en génétique. Il existe, en effet, un certain flou dans l'usage que fait Morgan de la notion de mutation, devenue indissociable de celle de gène depuis [de Vries, 1909].

Un premier aspect problématique de cette notion est ainsi qu'elle lui sert, tour à tour, à désigner tout changement observable dans le phénotype d'un individu, puis, là où la théorie des mutations issue des travaux de de Vries avait pourtant introduit le concept intermédiaire de *prémuation*, une modification associée, *de nature inconnue*, sur une portion de chromosome. Et cette ambiguïté sur le plan du vocabulaire n'a, à cet égard, servi qu'à déplacer la méthode mendélienne de décomposition en sous-caractères pour la réintroduire, au niveau des mutations elles-mêmes, en postulant, notamment, que toute variation continue ou par degrés d'un caractère, comme la couleur de certaines fleurs, peut *a priori* être ramenée à une combinaison de mutations discrètes. Mais, ce qui pose le plus nettement problème à nos yeux, c'est, encore une fois, que la méthodologie "morgano-mendélienne" repose *exclusivement* sur une étude des corrélations empiriques observables entre la fréquence des *crossing over* dans la lignée germinale et celle des recombinaison de mutations phénotypiques au sein de la descendance d'un individu donné. Morgan souligne d'ailleurs à plusieurs reprises que la question du lien physiologique entre gènes et caractères *concerne l'embryologie seule et non la génétique*. Or, on peut fondamentalement contester à ce type d'hypothèse théorique qu'elle puisse mettre la génétique en position de se prononcer sur la nature d'un éventuel lien de causalité entre gènes et caractères. En particulier, la génétique classique n'a jamais montré que ses facteurs-unités *engendraient* les caractères auxquels ils étaient associés (ainsi que le suggère le terme de gène lui-même). Et ce, pour deux raisons : d'une part, une simple analyse empirique de corrélations ne permet pas de conclure à l'existence d'un lien de causalité effective entre les différents niveaux d'organisation observés ; et, d'autre part, nul ne saurait inférer l'existence nécessaire d'un gène du type sauvage sur la seule base de l'existence probable d'un gène muté : encore une fois, en reprenant notre discussion du § 3.2.2, une analyse différentielle ne donne pas immédiatement une *loi* générale explicative du cas réputé sauvage (et l'on peut se demander si la notion de gène sauvage a véritablement un sens en dehors d'expériences d'hybridation strictement délimitées étant donné le rôle que sont censées jouer les mutations dans la théorie néo-darwinienne de l'évolution, comme on en discutera ci-après).

Pour conclure la réflexion sur cet aspect, observons que, une fois décodé le génome, on s'aperçoit que la notion de « gène », qui a dominé le XXème siècle, est un concept aussi flou que celui de phlogistique en vigueur au XIXème, car dépourvu de toute définition structurelle, physico-chimique, voire de toute délimitation spatio-temporelle. En général, un gène est ... « ce qui engendre » : des protéines ou des enzymes, parfois plusieurs, disions-nous – on ne parle presque plus de caractères. Et, qui plus est et sauf des cas particuliers qui ne donnent pas une loi dans le sens de la physique, il est situé dans un espace et un moment *indéterminés* qui vont de l'ADN aux activités cytoplasmiques de l'ARN : comme nous observons dans 3.2.3 pour les eucaryotes, même la structure primaire des protéines engendrées peut dépendre du contexte cytoplasmique de transfert de l'ARN

Un second aspect de la méthodologie différentielle ne manque d'ailleurs pas de mettre en doute la simple possibilité d'une réinterprétation des travaux de cartographie génétique à la lumière des données empiriques que la biologie moléculaire ne cesse de confirmer aujourd'hui.

Dans *The Mechanism of mendelian heredity* (pp. 172 et *passim.*), Morgan est, en effet, le premier à relever systématiquement les éléments factuels qui interdisent toute tentative d'associer biunivoquement un gène à un caractère. Ces observations sont d'ailleurs devenues, depuis quelques années, des arguments importants dans le discours d'une partie de la littérature en biologie théorique cherchant à placer le concept de "dégénérescence" au cœur des discussions contemporaines, notamment afin de montrer que le réductionnisme physicaliste *n'est pas une approche pertinente en biologie*. Parmi ces éléments, on peut, par exemple, citer le fait qu'*une même mutation génotypique affecte toujours plusieurs caractères observables en même temps*, parfois sans rapport direct les uns avec les autres¹⁹, et que, réciproquement, *un même caractère peut être modifié sous l'effet de diverses mutations situées en des loci bien distincts*²⁰. Il semble, de ce point de vue, qu'il faudrait vraiment forcer la métaphore alphabétique pour qu'elle nous aide à comprendre des faits aussi remarquables. Sans doute une bonne notion de tri sur une certaine base de données nous éclairerait-elle beaucoup plus, du moins si on cherche à la comprendre dans le cadre de contextes d'expression différents, chose qui est encore inhabituelle en informatique aujourd'hui.

Les remarques précédentes ne concernaient que la question du lien entre gènes et caractères au sein d'un individu, *modulo une certaine ontogenèse, inanalysée dans le cas de la génétique classique*. On a déjà mis en doute plus haut (§ 3.2.2) le fait que ce lien pouvait être analysé d'un point de vue causal à partir des seules expériences tératogènes de la génétique développementale. Mais un autre aspect de la notion de mutation nous semble problématique, du point de vue de la théorie de l'évolution elle-même cette fois.

Les théories néo-darwiniennes de l'évolution ont, en effet, fait de la mutation l'un des concepts centraux des analyses, autant théoriques que pratiques, de la phylogénèse. L'une des hypothèses avancées est que le changement d'un seul nucléotide, induit, au besoin, par un rayon cosmique ou un effet quantique, *peut* démarrer une chaîne causale qui aboutit à une modification évolutive importante. Mais, comment saisir cela dans les termes d'un changement de signification d'un discours au cours de l'histoire, à partir du simple changement/déplacement d'une lettre ou d'un mot dans une phrase, voire d'une phrase dans un discours, sans tomber dans les apories du fameux singe dactylographe de Thomas Huxley ? Si la référence à l'alphabet/base de données peut rendre partiellement compte du caractère discret de la structure de l'ADN en tant que mémoire génétique, il reste, encore et toujours, que la plus grande difficulté propre à ces analyses par analogie est qu'elles ne permettent pas de saisir de manière intelligible le passage de l'alphabet lui-même au *sens* du discours. Et cette difficulté s'accroît d'autant plus que l'histoire des systèmes biologiques tend précisément à accumuler et compliquer les systèmes élémentaires en les constituant en organismes, avec tous les raffinements de causalité intriquée (au sens large) que suggère ce terme, causalité que des méthodes simplement différentielles laisseront toujours inanalysée dans la mesure où, par définition elles sont systématiquement insensibles à tout ce qui demeure invariant.

Il est clair, de ce point de vue, que, d'une manière très générale, si l'on retire des explications évolutionnaires les arguments de la transmission directe des caractères acquis et de la traduction

¹⁹ Ainsi, chez la drosophile, la mutation "yeux blancs" semble empiriquement être associée à la durée de vie des individus *sans qu'il soit possible d'établir un lien physiologique clair entre ces deux caractères*.

²⁰ Pour reprendre l'exemple précédent, Morgan a ainsi recensé plus de vingt-cinq mutations susceptibles de modifier le caractère "couleur de l'oeil" chez la drosophile.

univoque des génotypes individuels en termes de portions d'ADN, le problème de l'adaptation *évidente* des organismes à leur environnement devient très vite d'une difficulté telle que le recours à une simple théorie des mutations n'est guère éclairant, voire devient tout bonnement douteux. On a ainsi montré qu'avec un taux de mutations de l'ordre de 1 nucléotide pour 100 000 au cours d'une réplication, ce qui correspond à peu près aux fréquences moyennes observées chez les organismes non-bactériens, et si l'on suppose que l'occurrence d'une mutation double la probabilité d'apparition d'une nouvelle mutation, il faudrait encore attendre 274 milliards d'années pour qu'une mutation simultanée de 5 nucléotides touche l'ensemble d'une population de 100 millions d'éphémères. En soi, un tel argument n'a absolument rien de dirimant pour la théorie de l'évolution entendue *comme une théorie descriptive*. D'ailleurs, le zoologue George G. Simpson [Simpson, 1953] notait déjà combien l'idée-même d'une évolution des espèces fondée sur la simple succession de mutations ponctuelles est *absurde eu égard aux données de la paléontologie*, puisque les différentes étapes de l'orthogénie animale montrent comment *de brusques ruptures sont introduites régulièrement dans la succession des individus*²¹.

Le problème de l'adaptation des espèces à leur environnement pose donc de sérieux problèmes à la théorie de l'évolution si l'on se contente d'analyser l'ADN et les modifications qui l'affectent au cours de l'histoire des individus dans le seul cadre paradigmatique d'une métaphore alphabétique. Car, quoique cette métaphore engage, apparemment, très peu et ouvre la voie vers l'univers richissime des analyses des langues vivantes, le problème de l'expression de cet alphabet, et particulièrement la question de ce qui pourrait jouer, en génétique, le rôle du phonème et des formes de vie historiques de la communauté langagière, de son rôle et son histoire donc, par rapport à la signification, reste entièrement ouvert. On pourrait même dire que ce problème pose avec une grande netteté la question de la possible incomplétude des théories physiques dans le cas de l'explication de l'origine et de l'évolution des formes vivantes, c'est-à-dire *des systèmes biologiques en tant qu'ils ont une histoire*, et une histoire qui, de surcroît disions-nous, efface une partie de son origine.

Conclusion

La complexité des composantes élémentaires des phénomènes naturels reste aujourd'hui le grand défi scientifique. En mécanique quantique tout d'abord, avec la non-localité et la non-séparabilité des particules élémentaires, avec les cordes et les théories à dix dimensions, dont six compactifiées, pour saisir l'élémentaire, mais également en biologie, où la complexité de la cellule, contexte nécessaire à l'expression de l'ADN, est probablement mathématiquement infinie par rapport à toute mesure physique. Les passages historiques par une théorie mécaniciste, qui essaie de saisir le complexe par la réduction à l'élémentaire et *simple*, ont fait leur temps. Tous nos engins mécaniques, toutes nos constructions artificielles, étant conçus "bottom up", sont donnés par composition de l'élémentaire simple (les horloges, les ordinateurs, les langages de programmation... les programmes). Il n'en est vraisemblablement pas ainsi pour des structures dont l'histoire naturelle est riche en aller-retours, back-tracks, interactions entre différents niveaux d'organisation, où l'élémentaire est un co-constitué dynamique, le résultat d'unités globales très complexes, donc, en principe, lui-même très complexe. L'embryogenèse,

²¹ Dans le cas du cheval, dont il est l'un des grands spécialistes, Simpson a estimé que l'évolution orthogénique a pris seulement 65 millions d'années, le cheval moderne ou *equus* étant le résultat de la succession, entre l'éocène et le quaternaire, d'une quinzaine de genres.

typiquement, *est* un processus “top-down” : tout commence par une cellule, unité organique complexe, qui se différencie (et d'ailleurs, en un certain sens, cette complexité de l'élémentaire biologique est *la* condition de possibilité du processus de différenciation lui-même).

Mais même le plus déterministe des systèmes physiques, notre système planétaire, *n'est pas une grande horloge*, comme on l'a très longtemps cru (malgré les doutes de Newton et Laplace, qui étaient des grands). Sa nature *essentiellement chaotique* ([Laskar, 1992 ; 1994]) démontre que la détermination équationnelle globale, qui correspond aux géodésiques de l'espace-temps, ne suffit pas à en saisir l'évolution en termes causaux ; il est en fait *causalement* sensible à des variations/perturbations en dessous de toute mesure théoriquement possible. C'est donc l'élémentaire des phénomènes locaux - qui peuvent être très complexes - qui interfère avec l'évolution globale. Et dans ce cas précisément, une théorie de géodésiques perturbées permet de comprendre quelque chose ; tandis qu'une vision en termes de lois ou de programmes, si elle peut donner lieu à de bonnes *imitations* computationnelles, avec leur propre apport d'intelligibilité, doit, néanmoins, être fondamentalement distinguée d'une modélisation mathématique explicative et/ou prédictive au sens le plus strict du terme (voir [Longo, 2002] pour la distinction entre imitation informatique et modélisation mathématique, d'ailleurs implicite chez [Turing, 1950 ; 1952]).

Les biologistes doivent avoir l'audace de hurler *la spécificité de leur besoin théorique*. La science procède tout d'abord *par différenciation des champs phénoménaux*. Darwin n'a pas pris des idées des théories physiques de l'époque, encore moins de la mécanique si poussée de son temps, mais il a proposé (suivant aussi d'autres : les idées ont toujours une histoire) une théorie absolument originale - tout au plus, avec des contaminations sociologiques, diront certains. Ce qui importe ici, c'est que, par son travail de réflexion théorique, *il a distingué la théorie du vivant de toute théorie physique*. On voit là une analogie avec un autre grand moment théorique, à l'intérieur de la physique même, qui a brisé l'unité prétendue de la microphysique et de l'astrophysique à la fin du XIX siècle, soutenue par la proposition d'un modèle planétaire de l'atome : l'invention de la physique quantique. Certains physiciens eurent l'audace de dire : non, la structure de la détermination que nous proposons est *radicalement, voire irréductiblement, différente*. Et des pivots centraux de l'intelligibilité physique ont, par là-même, été renversés : pas de localité, pas de séparabilité, pas de trajectoires (!), le champ est linéaire.

Les biologistes devraient en faire autant, un peu comme Darwin du reste (et on en voit des signes, ici et là). Une séparation conceptuelle nette aide à saisir une phénoménalité très difficile ; *ensuite*, on visera à l'unification, aux ponts de langage et à la dérivation logique des théories, mais on ne se laissera pas séduire, entre temps, par des analogies métaphoriques trompeuses, riches d'une histoire forte et d'une structure logique autonome. En physique, la séparation théorique n'empêche pas que, près de cent ans après, on entrevoit certains éléments d'une unification possible entre champs quantique et relativiste (à vrai dire *deux* unifications possibles, apparemment incompatibles...) ; mais, dans l'intervalle, la physique quantique, avec une intelligibilité toute particulière et de nombreuses interprétations, n'aura pas manqué de changer le monde par son efficacité et la culture scientifique par *ses méthodes et son autonomie explicative*.

La compréhension de la structure spatiale-discrète de l'ADN est une des grandes avancées scientifiques du XXe siècle, mais son activité dans un contexte fort complexe et dynamique *reste à comprendre*. Notre impression première persiste : en l'absence d'une théorie forte et autonome de la cellule vivante en tant qu'organisme, avec tous ses niveaux d'organisation, une spécificité théorique qui doit être pensée comme à la fois *dynamique et physiquement singulière*, inconnue aux théories physiques actuelles, il sera difficile de se donner un cadre d'intelligibilité comparable à ceux de la physique. Encore une fois, nous ne doutons pas qu'il n'y ait que du

“matériel physique” là, dans le monde, mais nous pensons que les théories physiques courantes, avec leurs structures de la détermination, (en particulier la théorie causalement laplacienne de la programmation), ne sont pas adéquates pour rendre intelligible la matière organisée qui intéresse la biologie, tout comme les dynamiques classiques et relativistes nous disent très peu sur la microphysique. Sans un changement de point de vue radical, on aura toujours des difficultés insurmontables pour saisir la place occupée, que ce soit au niveau du métabolisme cellulaire, du développement individuel ou de l'évolution des espèces, par cette trace discrétisée de l'histoire des êtres vivants (et d'aucun autre des systèmes traités par les théories physiques) qu'est l'ADN avec son contexte d'expression.

Au sujet de la détermination et de la causalité, deux pistes parmi maintes autres possibles nous paraissent pertinentes, voire nous fascinent. L'enjeu de la longueur de corrélation entre variables (et leurs interactions avec les phénomènes), qui dans un organisme vivant paraît atteindre les dimensions de l'organisme lui-même, pourrait peut-être être analysé à la limite des phénomènes proprement physiques, comme on essaie de faire dans [Bailly, Longo, 2006]. Il s'agirait d'analyser la dynamique du vivant comme *situation critique étendue*. En fait, la criticalité physique permet de saisir le passage du local au global, par exemple dans les transitions de phases. On y décrit le changement par des passages ponctuels, instantanés si le paramètre temps est en jeu, où des corrélations s'établissent entre toutes les composantes du phénomène et produisent une nouvelle unité ; celle-ci, par rapport à “l'échelle” précédente, donne lieu à des mesures mathématiquement infinie. En physique, alors, on dit que certaines valeurs divergent, comme dans une singularité mathématique, voire qu'elles sortent de l'observable physique. Ces divergences physiques, peuvent acquérir un sens biologique, si elles nous aident à saisir la formation d'une unité critique, mais *étendue*, dans le temps et l'espace, *loin de l'équilibre*, stable et instable, voire préservée dans des rails de stabilité par cette unité de corrélation (d'auto-organisation) physiquement implausible, car mathématiquement infinie. Une des difficultés techniques résiderait dans l'invention de bonnes techniques de renormalisation, à plusieurs (une infinité de?) paramètres, bien au delà des théorisations physiques courantes. L'avantage immédiat, par rapport à ce dont on a discuté, consisterait dans une intelligibilité possible du rôle des moindres variations de cette composante discrète de l'hérédité, l'ADN, dans l'induction d'énormes changements au niveau phénotypique : les effets de résonance et d'intrication propres aux longueurs de corrélation de la criticité étendue justifieraient ces changements d'échelle causale, sans donner un rôle de cause, dans le sens de la détermination laplacienne, voire “directe et entière” à l'ADN, en tant que programme²². On pourrait peut-être aussi mieux comprendre en ces termes les notions génétiques de mutations, de pléiotropie, de polygenie ou d'épistasie : c'est l'activité de l'organisme (cellule, métazoaire) dans son unité, son *action-réaction* organisée, qui leur donnerait du sens.

Une autre piste pourrait être probablement basée sur une logique de l'interaction, qui sache aussi utiliser les mathématiques du continu, grâce auxquelles on comprend mieux les dynamiques physiques dont nul ne doute de la présence chez le vivant. Par ce biais, on devrait viser à un changement intrinsèque du cadre logique – tout comme dans certaines propositions récentes pour la logique du quantique -, qui pourrait être mieux adapté à rendre compte de l'intrication des niveaux d'organisation et des interactions propres aux phénomènes du vivant.

²² Et cela nous reconforterait au sujet de ce 1% seulement de différence, difficilement interprétable dans les paradigmes dominants en génétique, qui paraît exister entre l'ADN de certains grands singes et celui du lecteur de cette note. En fait, nous faisons plutôt confiance à la sensibilité-au-contexte/stabilité propre à la criticité étendue de notre lecteur, ainsi qu'à sa richesse extragénomique et son expérience épigénétique.

Remerciements Des discussions très enrichissantes avec Antoine Danchin ont motivé cet article. F. Bailly, J.-J. Kupiec, M. Mossio, B. Saulnier et J. Stewart ont été des interlocuteurs critiques essentiels à son écriture.

Références

(Des versions préliminaires ou revues des articles de Longo sont "downloadable" de "<http://www.di.ens.fr/users/longo>" ou Google: search : Giuseppe Longo).

Aceto L., Longo G., Victor B. (eds.) **The difference between Sequential and Concurrent Computations. Special issue, Mathematical Structures in Computer Science**, Cambridge U. Press, n. 4-5, 2003.

Ahouse J. C. et Keller E. F., "Writing and Reading about 'Dolly'", **BioEssays**, Tome 19, 741-742, 1997.

Allis C. et Jenuwein T., "Translating the Histone Code", **Science**, Tome 293, 1074-1080, 2001.

Apter M. J., **Cybernetics and Development**, Pergamon, 1966.

Aristote, **Physique**.

Aristote, **Traité de l'Âme**.

Atlan H., **L'Organisation biologique et la théorie de l'information**, Hermann, 1972.

Atlan H. et Koppel M., "The Cellular Computer DNA ; program or data", **Bulletin of Mathematical Biology**, Tome 52, 335-348, 1990.

Bailly F., Longo G. **Mathématiques et sciences de la nature. La singularité physique du vivant**. Hermann, Paris, 2006.

Barendregt H. **The lambda-calculus: its syntax, its semantics**, North-Holland, rev. edit., 1984.

Barendregt H., Longo G. "Equality of lambda terms in the model T^{\square} ," *in* **To H.B. Curry: Essays in Combinatory Logic, Lambda-calculus and Formalism**, (Seldin, Hindley eds.) Academic Press, London, pp. 303-337, 1980.

Bartel D., "MicroRNAs : Genomics, Biogenesis, Mechanism and Function", **Cell**, Tome 116, 281-297, 2004.

Beadle G. W. et E. L. Tatum E. L., "Genetic Control of Developmental Reactions", **American Naturalist**, Tome 75, 107-116, 1941.

Bernard-Weil E. "Ago-Antagonistic systèmes" *In* **Quantum Mechanics, Mathematics, Cognition and Action**. (M. Mugur-Schachter ed.), Kluwer, pp. 433-463, 2002.

Bernot A., **Analyse de génomes, transcriptomes et protéines**, Dunod, Paris, 2001.

- Beurton P.-J. et al, **The Concept of the Gene in Development and evolution : Historical and Epistemological Perspectives**, Cambridge University Press, 2000.
- Bjedov I., Denamu E., Gerard B., Matic I., Tenailon O., Rardman M., Souza V. et Taddei F., "Stress-induced Mutagenesis in bacteria", **Science**, Tome 300, 1404-1409, 2003.
- Black D. L., "Splicing in the Inner Ear : a familiar tune, but what are the instruments ?", **Neuron**, Tome 20, 165-168, 1998.
- Boffelli D., Nobrega M. et Rubin E., "Comparative genomics at the vertebrates extremes", **Nature Reviews Genetics**, Tome 5, 456-465, 2004.
- Boguski M. et Hieter P., "Functional Genomics : it's all about how you read it", **Science**, Tome 278, 601-602, 1997.
- Bonner J., **The Molecular biology of development**, Oxford University Press, 1965.
- Brenner S., Dove W., Herskowitz I. et Thomas R., "Genes and Development : molecular and logical themes", **Genetics**, Tome 126, 479-486, 1990.
- Burnet F. M., **Le Programme et l'Erreur** (1978), Albin Michel, 1982.
- Campbell K. H. S., Kind A. J., Schnieke A. E., McWhir J. et Wilmut I., "Viable Offspring Derived from Fetal and Adult Mammalian Cells", **Nature**, Tome 385, 810-813, 1997.
- Cappuccio M., "Traces of computational mind : From wax tablets to Turing machine", **Géométrie et cognition**, (Longo ed.), Editions rue d'Ulm, Tome 124, s.5, 43- 60, 2003.
- Chapouthier G. et Matras J.-J., "La Néguentropie : un artefact ?", **Fundamenta Scientiæ**, Tome 2, 141-151, 1984.
- Collectif, **Biologie moléculaire de la cellule**, Flammarion, 1995.
- Collectif, **Le Concept d'information dans la science contemporaine**, Cahiers de Royaumont, Minuit-Gauthier-Villars, 1965.
- Creager A. N. et Gaudillière J.-P., "Meanings in Search of Experiments and *Vice-versa* : the invention of allosteric regulation in Paris and Berkeley, 1889-1968", **Historical Studies in the Physical and Biological Sciences**, Tome 27, 1-89, 1996.
- Crick F. H. C., "On Protein Synthesis", **Symposia of the Society for Experimental Biology**, Tome 12, 138-163, 1957.
- Crick F. H. C. et Watson J. D., "Molecular Structure of Nucleic Acids", **Nature**, Tome 171, 737-738, 1953.
- Crick F. H. C. et Watson J. D., "Genetical Implications of the Structure of Deoxyribonucleic Acid", **Nature**, Tome 171, 964, 1953.
- Eddy S., "Non-coding RNA Genes and the Modern RNA World", **Nature Reviews Genetics**, Tome 2, 919-929, 2001.

- Edelman G. M. et Gally J. A., "Degeneracy and Complexity in biological systems", **Proceedings of the National Academy of Science**, Tome 24, 13763-13768, 2001.
- Edelman G., Tononi G. **A Universe of Consciousness. How Matter Becomes Imagination**, Basic Books, 2000.
- Enard W. *et al.*, "Intra- and Inter-specific Variation in Primate Gene Expression Patterns", **Science**, Tome 296, 340-343, 2002.
- Feingold J. et Serre J.-L., **Génétique humaine, De la transmission des caractères à l'analyse de l'ADN**, INSERM-Nathan, 1993.
- van Frassen B. **Lois et symétries**, Vrin, Paris, 1994.
- Gilbert W., "Why Genes in Pieces ?", **Nature**, Tome 271, 501, 1978.
- Godell B., Gouyon P. H., Maynard-Smith J., Radman M., Taddei F. et Toupance B., "Role of mutator alleles in adaptive evolution", **Nature**, Tome 387, 700-702, 1997.
- Goldschmidt R., **Physiological Genetics**, McGraw-Hill, 1938.
- Goodwin B., "What are the Causes of Morphogenesis ?", **BioEssays**, Tome 3, 32-36, 1985.
- Goubault E. (ed.) **Geometry in Concurrency, Special issue, Mathematical Structures in Computer Science**, Cambridge U.P., vol.10, n.4, 2000.
- Gros F., **Les Secrets du gène**, O. Jacob/Points-Seuil, 1991.
- Havelock, E. A. **Origins of Western Literacy**. Toronto: Ontario Institute for Studies in Education, 1976.
- Herrenschmidt C. et al. (eds), **L'Orient et nous**, Albin-Michel, 1996.
- Jablonka E. et Lamb M., **Epigenetic Inheritance and Evolution**, Oxford University Press, 1995.
- Jacob F. et Monod J., "Gènes de structure et gènes de régulation dans la biosynthèse des protéines", **Comptes rendus de l'Académie des Sciences de Paris**, Tome 349, 1282-1284, 1959.
- Kupiec J.-J., "A Chance-Selection Model for Cellular Differentiation", **Cells, Death & Differentiation**, Tome 3, 385-390, 1996.
- Kupiec J.-J. et Sonigo P., **Ni Dieu, ni gène**, Seuil, 2000.
- Laskar J., "La stabilité du système solaire" in **Chaos et déterminisme** (Dahan et al. eds), Seuil, 1992.
- Laskar J., "Large scale chaos in the Solar System", **Astron. Astrophysics**, 287, L9 L12, 1994.
- Lewontine R., "The Confusion over Cloning", **New York Review of Book**, octobre 1997.
- Longo G. "Laplace, Turing et la géométrie impossible du "jeu de l'imitation" : aléas, déterminisme et programmes dans le test de Turing", **Intellectica**, 35, 2002/2 (version anglaise 2007, téléchargeable).

- Mayr E., "Cause and Effect in Biology", **Science**, Tome 24, 1-14, 1959.
- Matherat P., Jaekel M.-T. "Concurrent computing machines and physical space-time", *special issue*, **Mathematical Structures in Computer Science**, Cambridge U.P., vol.13, n. 5, 2003.
- Mendel J., "Recherches sur des hybrides végétaux", traduction de A. Chappellier, **Bulletin scientifique de la France et de la Belgique**, Tome 41, 371-419, 1907.
- Morange M., **Histoire de la biologie moléculaire**, La Découverte, 1994.
- Morange M., **La Part des gènes**, Odile Jacon, 1998.
- Morgan T. H., **The Theory of the Gene** (1926), Yale University Press, 1928.
- Morgan T. H., **Embryologie et Génétique** (1934), traduction de J. Rostand, Gallimard, 1936.
- Nijhout F., "The Nature of Robustness in Development", **BioEssays**, Tome 24, 553-563, 2002.
- Nouvel P. "Modèles et métaphores" *dans* **Enquête sur le concept de modèle**, Nouvel P. (ed.), Presses Univ. de France, 2002.
- Ong W. J. **Orality and Literacy: The Technologizing of the Word**. New Accents. Ed. Terence Hawkes. New York: Methuen, 1988.
- Pentris S., Tooze J., Hunt T. (eds.) **DNA makes RNAmakes Protein**, Elsevier, 1983.
- Pichot A. **Histoire de la notion de gène**, Flammarion, 1999 .
- Portin P., "The Concept of the Gene : short history and present status", **Quarterly Review of Biology**, Tome 68, 173-223, 1993.
- Shannon C. E. et Weaver W., **Théorie mathématique de la communication**, traduction de J. Cosnier, G. Dahan et S. Economidès, Retz-C. E. P. L., 1975.
- Shroedinger E., **What is life?** , 1951.
- Simpson G. G., **The Major features of Evolution**, Harvard University Press, 1953.
- Stewart J. **La vie existe-t-elle ?** Vuibert, Paris, 2004.
- Sturtevant A. H., **A History of Genetics**, McGraw-Hill, 1965.
- Tautz D., "Redundancies, Development and the Flow of Information", **BioEssays**, Tome 14, 263-266, 1992.
- Thomas J. H., "Thinking about Genetic Redundancy", **Trends in Genetics**, Tome 9, 395-399, 1993.
- Turing A. "Computing Machines and Intelligence", **Mind**, LIX, 1950 (page references to its reprinted version *in* M. Boden ed., Oxford Univ. Press, 1990; traduction française et introduction *dans* A. Turing, J.-Y. Girard, **La machine de Turing**, Seuil, 1991).

Turing A. M. "The Chemical Basis of Morphogenesis" **Philo. Trans. Royal Soc.**, B237, 37-72, 1952.

Turner B., "Cellular Memory and the Histone Code", **Cell**, Tome 111, 285-291, 2002.

de Vries H., **Espèces et variétés, leur naissance par mutations**, traduction de L. Blaringhem, Alcan, 1909.