Chapitre 30 « L'exportation de la pensée phylogénétique en biochimie »,
par Guillaume Lecointre & Chomin Cunchillos (pages 705-730)
du livre ***Les mondes darwiniens. L'évolution de l'évolution***
(Paris, Syllepse, 2009), sous la direction de Thomas Heams,
Philippe Huneman, Guillaume Lecointre, Marc Silberstein.
www.lesmondesdarwiniens.org

CHAPITRE 30

L'exportation de la pensée phylogénétique en biochimie

Guillaume LECOINTRE[1] & Chomin CUNCHILLOS[2]

Au sein même de la communauté des systématiciens, il y eut un débat autour des années 2000 pour déterminer si la théorie de l'évolution était requise pour l'exploitation des logiciels de construction d'arbre d'inspiration cladistique, c'est-à-dire ce qu'on appelle aujourd'hui la parcimonie informatisée (Brower 2000). Rigoureusement parlant, l'algorithme cladistique, en choisissant l'« arbre » le plus parcimonieux en hypothèses de transformations des caractères le long de ses branches, maximise la contiguïté des états de caractères identiques dans l'« arbre » (appelé aussi dendrogramme, cladogramme ou, de manière encore plus éthérée, graphe connexe non cyclique). L'algorithme ne fait que montrer « qui partage quoi avec qui ». Si référence à l'évolution il y a, c'est dans la tête de celui qui a construit l'arbre. *A posteriori*, l'interprétation de ce cladogramme à la lumière de la théorie de l'évolution transforme le « qui partage quoi avec qui » en « qui est plus apparenté à qui ». *A posteriori*, tout bon biologiste s'y retrouve donc. Mais c'est concernant ce qui se passe *a priori* que le débat fit rage. Certes, on peut se demander pourquoi on construirait une matrice de caractères si ce n'est pour fonder des hypothèses d'homologie primaire (de Pinna 1991) que l'on voudrait tester ; mais qui dit « homologie primaire » dit-il évolution ? En effet, les homologies primaires n'indiquent que des ressemblances associées à un principe des connexions dont certains des pères, comme Richard Owen ou Georges Cuvier, étaient, faut-il le rappeler, fixistes. Ce rappel indique juste que le principe des connexions, sur le plan technique, ne requiert pas le transformisme, même si aujourd'hui tous ceux qui l'utilisent inscrivent l'évolution dans *leur intention*. Le débat se trouve donc reporté sur ce que les biologistes ont dans la tête lorsqu'ils codent

1. Systématicien, professeur au Muséum national d'histoire naturelle (UMR 7138 CNRS-UPMC-ENS-IRD-MNHN, « Systématique, adaptation, évolution ») et directeur du département « Systématique et évolution » du MNHN.

2. Institut Charles Darwin International.

lecointre-biochimie.indd 705 6/08/10 20:19:51

leur matrice de caractères. La plupart d'entre eux ne savent généralement pas pourquoi – en termes de relations de cause à effet *documentées* entre deux états de caractères présumés homologues – ces deux états de caractères se ressemblent. Par conséquent, l'homologie primaire figurant dans la matrice repose bien sur un pari d'où l'évolution est techniquement absente (les données moléculaires et leurs modèles traités par des approches probabilistes sont laissées de côté dans cette discussion qui ne concerne que la cladistique[3]). C'est la raison pour laquelle la plupart des matrices destinées à un traitement par parcimonie sont exemptes de pondération des caractères. Et ce n'est pas un hasard si tous les débats tournant autour de la légitimité des pondérations sont directement connectés à celui-ci. En revanche, la plupart des biologistes savent généralement pourquoi un caractère donné sera *exclu* de la matrice. La plupart des systématiciens excluront en effet de poser comme putativement homologues (homologie primaire) deux états de caractères pour lesquels on peut *présumer*, ou mieux, *documenter* une origine évolutive distincte (par l'embryologie, par l'anatomie elle-même, etc.). Ainsi, paradoxalement, c'est donc, sur le plan technique et pas seulement sur le plan des intentions, parce qu'elle ne contient pas qu'une matrice d'homologie primaire nous parle d'évolution : elle réunit des caractères qui mettent en relation des états qui pourraient éventuellement avoir été acquis par voie d'ascendance commune.

Alors, la maximisation de la contiguïté des états identiques de caractères est-elle bonne à tout faire ? Gareth Nelson & Norman Platnick (1981) pensent que oui[4]. En fait, la maximisation de la contiguïté des états de caractères identiques ne relèverait que d'un principe de cohérence maximale qui ne relève, lui-même, que de la rationalité : en sélectionnant l'arbre le plus parcimonieux, on sélectionne en même temps l'arbre le plus cohérent. En effet, plus les caractères sont cohérents entre eux dans l'arbre choisi, et moins il est besoin d'ajouter des transformations surnuméraires destinées à rendre compte des incohérences. Ce n'est pas un hasard si le nombre d'événements que requiert un cladogramme est directement proportionnel à son « indice de cohérence » (*Consistency Index*). Cohérence et parcimonie sont une seule et même mesure. C'est au nom de l'universalité du principe de maximisation de la cohérence de nos explications que l'algorithme de parcimonie se voit conféré une portée universelle. Le principe de parcimonie est partie intégrante de l'explication rationnelle en sciences…[5] et donc forcément mobilisé pour expliquer scienti-

3. Sur ce point précis, cf. Barriel (section 7), ce volume. *(NdÉ.)*

4. Au sujet de la systématique phylogénétique de Hennig (1966) : « If so, Hennig's system would be understandable not merely as the theory of "phyletic" taxonomy but as the general theory of taxonomy of whatever sort. »

5. Guillaume Lecointre développe ce point dans sa préface de Charbonnat (2007). *(NdÉ.)*

lecointre-biochimie.indd 706 6/08/10 20:19:51

fiquement toute ressemblance. Il n'y a, alors, pas besoin d'évolution, effectivement, pour justifier l'emploi de cet algorithme. Il faut juste que nous ayons de bonnes raisons de mesurer la cohérence d'un jeu de caractères entre eux, que nous voulions utiliser un graphe connexe non cyclique pour exprimer une hiérarchie (qui revient à une série d'ensembles emboîtés les uns dans les autres – cf. Hennig 1966), quelle que soit la théorie explicative sous-jacente compatible avec ces deux modes de représentation d'une hiérarchie.

1 – Les problèmes posés par l'exportation

On devine que cette question est importante concernant l'exportation des algorithmes utilisés pour l'analyse cladistique informatisée vers d'autres champs d'investigation que la systématique : linguistique[6], musicologie comparative et biochimie comparative, pour ne citer que quelques exemples. Dans ces approches comparatives d'objets qui ne sont pas des espèces (langues, musiques, voies métaboliques), l'importation d'algorithmes cladistiques explicitant « qui partage quoi avec qui » n'auraient pas besoin d'être justifiées selon Nelson & Platnick (1981). Pour d'autres auteurs, l'utilisation de tels algorithmes requiert au minimum le fameux principe du « *descent with modification* » de Darwin. En effet, ce principe donne un sens particulier au fait de maximiser la contiguïté des états de caractères identiques. Ce principe stipule que deux états de caractères trouvés chez des espèces qui ne se croisent plus aujourd'hui doivent en première instance être interprétés comme ayant été légués par voie d'ascendance commune. Ils ne se croisent plus entre eux aujourd'hui, certes, *mais* ils doivent s'être croisés entre eux hier : il fut un temps où ils avaient un ancêtre commun. Les chats et les chiens ont une truffe et des poils. Les ont-ils acquis ensemble parce qu'ils se croisent entre eux ? Non : les chats ne font pas de petits avec les chiens, tout le monde sait cela. En première instance on doit parier sur l'ascendance commune. Ce pari implique le transformisme. L'ancêtre commun était-il chat ? Si oui, il y a eu forcément transformation sur le trajet généalogique (théorique) entre les chats ancestraux et les chiens, sinon les chiens seraient chats. L'ancêtre commun était-il chien ? Si oui, il y a forcément eu transformation sur le trajet généalogique théorique entre les chiens ancestraux et les chats, sinon les chats seraient chiens. L'ancêtre commun était-il mi-chien mi-chat ? Si oui, il y a forcément eu transformation sur les deux trajets généalogiques théoriques depuis ces ancêtres, sinon les chiens et les chats ne seraient pas ce qu'ils sont. Bref, il y a une grande cohérence entre le « *descent with modification* » et le projet qui consiste à réunir les branches aux états de caractères identiques. Maximiser la contiguïté de deux (ou plus) branches portant un état de caractère donné revient à réunir ces branches de manière à leur conférer en amont un

6. Cf. Ben Hamed, ce volume. *(NdÉ.)*

lecointre-biochimie.indd 707 6/08/10 20:19:51

tronc commun sur lequel siège une hypothèse d'ancêtre commun. Notons que le « *descent with modification* » implique de postuler qu'il y a eu, dans le passé, une transformation des êtres (et de leurs attributs) génération après génération (ou « évolution » dans son sens le plus minimal). Mais la charge en postulats s'arrête là. En effet, le mécanisme par lequel cette transformation est rendue possible n'est aucunement convoqué et n'a pas besoin de l'être. Dit autrement, la vie changerait par un autre mécanisme que la sélection naturelle – dérive génétique, processus lamarckien, ou autres… –, cela ne changerait en rien la façon dont le « *descent with modification* » est postulé pour justifier de réunir sur une même branche des espèces partageant des attributs.

Nous ne nous engagerons pas ici plus avant dans le débat qui consiste à savoir si l'analyse cladistique requiert l'idée d'évolution sous sa forme « *descent with modification* », ou si cette idée est seulement souhaitable, facultative, « pas nécessaire » (Brower 2000), proscrite, selon que l'on se réfère au programme général de recherche de toute systématique *biologique* ou si l'on se restreint strictement à la perception agnostique des similitudes et au fonctionnement d'un algorithme[7]. Nous voulons rassurer les tenants des deux camps sur un point : dans les deux cas, il est justifié d'appliquer l'analyse cladistique aux voies métaboliques. Il suffit au second des deux camps, celui des « *pattern cladists* », de constater que certaines voies métaboliques se ressemblent sans être toutefois identiques, et qu'un « *pattern* » y est aisément perceptible. Pour le premier des deux camps, il nous faut convaincre que la structure du métabolisme cellulaire a quelque chose à voir avec le concept darwinien de « *descent with modification* ».

2 – Pertinence de l'application d'outils phylogénétiques aux voies du métabolisme

Les biochimistes ont beaucoup écrit au sujet de l'évolution du métabolisme, mais sans toutefois disposer d'une méthode qui leur permettrait de tester leurs hypothèses évolutives. Comme cela se faisait dans une systématique anté-hennigienne (en gros avant les années 1970), en biochimie, et jusque très récemment, des faisceaux de présomptions corroboraient (forcément) un scénario évolutif avec beaucoup de finesse et de cohérence, parfois de manière parcimonieuse[8], parfois avec un fort contenu spéculatif[9]. Ici, nous entendons expliciter les principes qui ont autorisé une véritable analyse phylogénétique des séries de réactions chimiques que l'on appelle les « voies métaboliques ». En effet, l'analyse phylogénétique a fourni les outils permettant de tester la

7. Quelques aperçus de ce débat dans Tassy, ce volume. *(NdÉ.)*

8. Jensen (1976), Cordón (1990), Cunchillos (1997).

9. Cavalier-Smith (1987), Meléndez-Hevia *et al.* (1996), Martin & Müller (1998).

lecointre-biochimie.indd 708 6/08/10 20:19:51

précédence temporelle de l'apparition de certaines réactions enzymatiques du catabolisme[10] des acides aminés par rapport à l'apparition d'autres réactions. Cet objectif revient à fournir l'ordre d'émergence dans le temps des grands types de voies métaboliques, et l'ordre d'émergence des différents types d'activité enzymatique aux origines de la vie cellulaire.

Le métabolisme cellulaire est un processus complexe constitué d'environ un millier de réactions chimiques catalysées par des enzymes, qui sont des protéines globulaires. Ces réactions sont organisées en une centaine de séquences, les voies métaboliques, qui, d'une façon générale, relient l'incorporation du nutriment, son utilisation, la synthèse des composantes du corps cellulaire, celle des réserves et de leur utilisation. Comme tout autre phénomène biologique, le métabolisme est le produit de l'évolution. Son histoire évolutive devrait pouvoir être reconstituée par l'analyse de sa complexité structurale[11]. Une telle histoire pourrait être reconstruite sur la base de la comparaison des structures des sites actifs (Knowles 1991) des enzymes elles-mêmes. Cependant, très peu d'entre eux sont connus en détail, en proportion du nombre d'enzymes répertoriées. En revanche, la réaction produite par ce site actif peut en être considérée comme un reflet extrêmement fidèle de sa structure si la réaction est hautement spécifique. De même, il est possible de reconnaître la similitude de réactions chimiques semblables réalisées par des sites actifs semblables probablement dérivés d'une protéine ancestrale ; laquelle se serait ultérieurement spécialisée dans telle ou telle variante d'une même réaction fondamentale (Jensen 1976). L'évolution des enzymes est en effet conçue comme partant d'activités généralistes vers des activités spécialisées par affirmation d'une spécificité à un substrat[12]. Par conséquent, la phylogénie des voies métaboliques peut être non seulement reconstruite à partir du seul partage d'une ou plusieurs enzyme(s) de haute spécificité, mais aussi par le partage par ces mêmes voies de types de fonctions enzymatiques (spécificité de fonction mais non spécificité des substrats), étant entendu que ceux-ci sont sous-tendus par une similitude structurale des sites actifs correspondants. Reconnaître que deux voies métaboliques partagent une réaction de haute spécificité revient à utiliser un critère d'homologie primaire strict, tandis que la reconnaissance du partage d'une même famille de réactions revient à relâcher ce critère d'homologie en le transférant dans le passé : si deux voies partagent deux enzymes semblables, c'est qu'elles ont dû partager jadis la même enzyme généraliste

10. Ensemble des réactions de dégradations moléculaires de l'organisme. L'anabolisme est l'ensemble des réactions de synthèse. Catabolisme et anabolisme constituent le métabolisme. *(NdÉ.)*

11. Schoffeniels (1984), Cordón (1990), Cunchillos (1997), Michal (1999).

12. Holden (1968), Jensen (1976), Jallon (1983), Piérard (1983), Petsko *et al.* (1993), O'Brien & Herschlag (1999).

lecointre-biochimie.indd 709 6/08/10 20:19:51

(Cunchillos & Lecointre 2002, 2003, 2005, 2007). Ce relâchement du critère d'homologie comporte un risque de convergence dont on n'a aucune raison de penser *a priori* qu'il soit supérieur à ce que l'on rencontre dans les matrices de séquences alignées ou dans les matrices de caractères morphologiques[13]. Nous avons donc ici un nouveau type de caractères : la réaction enzymatique (ou l'enzyme elle-même si sa réaction détient une très haute spécificité de substrat), un nouveau type de taxons : les voies métaboliques de chacun des acides aminés aliphatiques, et un critère d'homologie primaire : le partage par deux voies métaboliques d'une réaction spécifique traduisant la confluence de deux voies, ou le partage d'un même type de réaction ou de fonction enzymatiques, d'une même famille de fonctions, ou encore l'utilisation commune de cofacteurs. La comparaison des différences de métabolisme constatées entre espèces actuelles ne serait d'aucune aide ici, car ces différences sont surtout dues à des événements bien postérieurs aux événements inférés, et concernent des voies plus périphériques. De plus, ces différences touchent beaucoup plus la régulation des voies que leur structure (Schoffeniels 1984). Comparer les sémantides de Zuckerkandl & Pauling (1965), c'est-à-dire les séquences d'acides nucléiques ou d'acides aminés des enzymes réalisant les fonctions codées, ne serait d'aucune aide non plus : on s'exposerait à de graves problèmes d'homologie des séquences (surtout sur les parties non situées dans les sites actifs). De plus, les mutations inférées, homoplasiques[14] ou pas, seraient elles aussi bien postérieures aux événements que nous cherchons à reconstituer. Le relâchement du critère d'homologie évoqué ci-dessus fournit un type d'homologie des réactions enzymatiques quelque peu dégagé de l'homologie de séquence.Le critère d'homologie primaire une fois énoncé au niveau de l'enzyme elle-même ou de son type de réaction, comment se justifie-t-il en termes de voies métaboliques au regard des idées que se font les biochimistes de leur évolution ? Horowitz (1945) propose que le développement des voies biochimiques de synthèse a dû suivre l'ordre inverse de celui que suivent les enzymes à l'intérieur de la voie : la première enzyme à apparaître au cours de l'évolution serait la dernière de la voie, la seconde serait l'avant-dernière, et ainsi de suite (on parle de développement « rétrograde »). Le scénario explicatif est résumé dans la figure 1. Un protobionte hétérotrophe pour un substrat A finira par épuiser le milieu. A va devenir limitant. La carence en A va de fait favoriser tout protobionte capable de synthétiser A à partir de composants B et C. Une première réaction enzymatique est née, et le nouveau protobionte va se multiplier. À terme, ce sont les substrats B et C qui vont

13. Cf. Barriel, ce volume. *(NdÉ.)*

14. Dont la similitude chez différentes entités ne provient pas d'un ancêtre commun, mais est l'effet de la convergence ou de la réversion. *(NdÉ.)*

lecointre-biochimie.indd 710 6/08/10 20:19:51

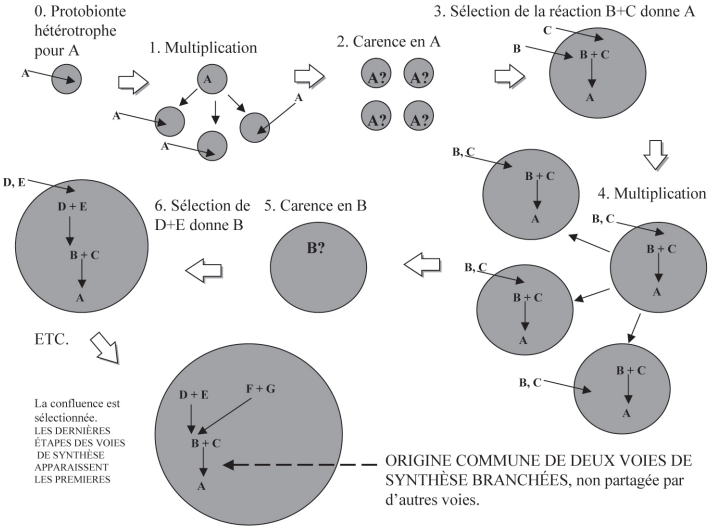

Figure 1. Évolution des voies de synthèse vues par Horowitz (1945) (cf. texte). Par l'effet de la sélection naturelle, les dernières étapes des voies de synthèse apparaissent les premières dans le temps (évolution « rétrograde »), parce que le *produit* est imposé de l'extérieur au départ. Parce que la confluence des voies est sélectionnée, l'enzyme ou la réaction enzymatique fournissant A à partir de B et C est un signe d'ascendance commune des voies DE vers A et FG vers A.

devenir limitants. Tout protobionte capable de synthétiser B et/ou C à partir d'autres substrats, disons B à partir de D et E, sera sélectionné de la même manière. Éventuellement, la possibilité d'obtenir B à partir d'autres substrats tels que F et G reste possible. Dans le schéma explicatif d'Horowitz, la confluence des voies métaboliques est sélectionnée puisqu'elle est économique. L'énergie gagnée par confluence sert à faire autre chose que ne ferait pas un protobionte compétiteur dépourvu de cette confluence et donc obligé de développer des moyens plus lourds. L'économie du métabolisme, appelée par les biochimistes « optimisation du métabolisme » est une règle fondamentale de la biochimie comparée (Schoffeniels 1984). Malgré la diffusion des idées d'Horowitz (1945), elle a été pensée soit dans un contexte non darwinien (Schoffeniels 1984)[15], soit dans un contexte résolument panglossien[16]

15. Au passage, le fait d'avoir parlé d'optimisation et non de sélection naturelle est évocateur de la culture *majoritaire* dans laquelle la biochimie et la génétique moléculaire se sont développées dans la seconde moitié du 20e siècle : largement en dehors de la pensée darwinienne. Sur cette question, cf. Morange (1994), (2005) et son chapitre dans ce volume, ainsi que Heams (2004), Kupiec (2004, 2008, 2009).

16. Cf. le passage sur le panglossisme Gouyon & Giraud, ce volume, note 6. *(NdÉ.)*

lecointre-biochimie.indd 711 6/08/10 20:19:51

(Meléndez-Hevia *et al.* 1996). Si nous considérons le résultat du raisonnement sélectif d'Horowitz, deux constats sont à faire. Premièrement, les dernières étapes des voies de synthèse apparaissent les premières. Deuxièmement, nous avons deux voies de synthèse confluentes (de DE vers A et de FG vers A) apparentées. Pourquoi apparentées ? Parce que la réaction « B + C donne A » signe une origine commune à deux voies de synthèse branchées, non partagée par d'autres voies de synthèse, si l'on prend la voie de synthèse comme taxon et la réaction comme caractère. L'acteur biochimique de cette réaction (enzyme, type de réaction, cofacteur, etc.) est le témoin qu'il s'agit de coder. S'il est codé de manière identique pour deux voies de synthèse dans une matrice, un tel codage dans la matrice en tant que « 1 » partagé est à mettre en relation avec l'ascendance commune aux deux voies du tronçon « B+C donne A » ; du moins pour ceux qui se préoccupent de justifier biologiquement leur codage. Pour certains systématiciens, cette justification n'est pas nécessaire, pour d'autres elle l'est ; et c'est la raison pour laquelle elle est développée ici. Cordón (1990) conçoit un scénario symétrique pour les voies de dégradation (fig. 2). Supposons un protobionte hétérotrophe pour A. A est dégradé en un produit (appelons-le « produit 1 »), réaction qui fournit de l'énergie dont a besoin le protobionte. Après un certain temps de multiplication, on peut supposer que A va devenir limitant dans le milieu. Seront alors sélectionnés les protobiontes capables de dégrader plus complètement le produit 1 en produit 2, et récupérer le surplus d'énergie délivrée par cette seconde réaction. Une possibilité de sélection ultérieure (non exclusive à la première) est la sélection d'un protobionte capable de dégrader un substrat B en produit 1. A et B étant devenus limitants, une étape ultérieure peut conduire à la sélection d'une dégradation encore plus poussée du produit 2 en produit 3. Deux constats symétriques au cas précédent. Premièrement, les dernières étapes des voies de dégradation apparaissent les dernières dans leur évolution (développement « antérograde »). Deuxièmement, si l'on considère que la confluence des voies est sélectionnée parce qu'elle est économique, deux longues voies de dégradation peuvent se trouver apparentées par le partage d'une enzyme ou de plusieurs enzymes distales communes. C'est donc l'allongement distal commun à plusieurs voies cataboliques qui signe à la fois leur branchement précoce et leur évolution commune. On peut bien sûr imaginer un branchement latéral tardif, et/ou très distal d'une portion de voie catabolique « opportuniste » qui n'aurait alors pas vraiment d'origine commune avec la voie sur laquelle elle se branche. Le phénomène de recrutement d'enzymes ou de voies métaboliques est une modalité assez classique de l'évolution biochimique[17]. Tout dépend de la précédence temporelle de la confluence des voies du catabolisme

17. Jensen (1976), Petsko *et al.* (1993), Copley (2000).

lecointre-biochimie.indd 712 6/08/10 20:19:52

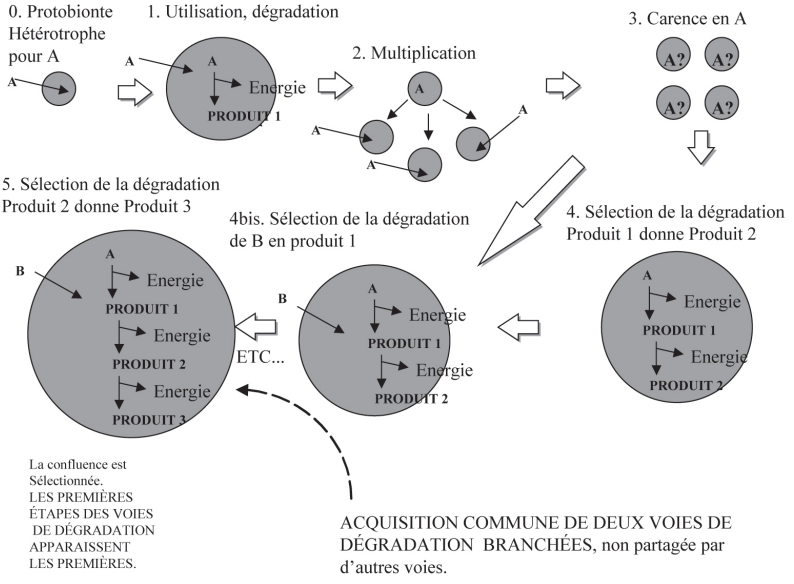

Figure 2. Évolution des voies de dégradation vues par Cordón (1990). Par l'effet de la sélection naturelle, les premières étapes des voies de dégradation apparaissent les premières dans le temps (évolution « antérograde »), parce que le *substrat* est imposé de l'extérieur au départ. Parce que la confluence des voies est sélectionnée, l'enzyme ou la réaction enzymatique fournissant le produit 3 à partir du produit 2 est un signe d'ascendance commune des voies B vers le produit 3 et A vers le produit 3, pourvu que la confluence se mette tôt en place (cf. texte).

par rapport aux étapes d'allongement de la portion commune distale. C'est un risque d'homoplasie à prendre et Cunchillos & Lecointre (2002, 2003, 2005, 2007) ont montré que ce risque n'est pas plus important dans le cas présent que pour d'autres matrices relavant de l'anatomie comparée ou de séquences alignées d'ADN.

Pour finir sur ces justifications, signalons que selon Cordón (1990), les règles d'Horowitz (1945) ne sont pas fixes et doivent être modulées en fonction des voies métaboliques. En effet, cet ordre d'apparition des réactions dans l'évolution des voies de synthèse et de dégradation est valable parce que, respectivement, le produit final ou le substrat initial a été imposé de l'extérieur, au moins initialement. Il est cependant possible d'imaginer des scénarios alternatifs pour des molécules obtenues par les cellules elles-mêmes au cours de leur évolution, par modification d'autres molécules préexistantes déjà intégrées au métabolisme. Les nouvelles séquences de synthèse se seraient développées progressivement par addition distale de nouvelles enzymes à des voies déjà développées. Il paraît clair que le type de développement suivi par

lecointre-biochimie.indd 713 6/08/10 20:19:52

une voie dépendra du rôle initial de la molécule correspondante dans l'histoire du métabolisme. C'est ainsi que Cordón (1990) a proposé un développement coïncidant avec l'ordre des enzymes dans la voie pour le catabolisme des acides aminés, l'anabolisme des acides gras et de la glycogenèse, et en sens inverse pour l'anabolisme des acides aminés, le catabolisme des acides gras et la glycolyse (dont le sens du développement a été confirmé par Fothergill-Gilmore & Michels 1993). Nous verrons ci-dessous les résultats du test opéré par Cunchillos & Lecointre (2005, 2007).

En somme, si l'on suit l'idée darwinienne que toute reconstruction phylogénétique est justifiée si il y a « filiation dans le temps des structures avec modifications », on constate que c'est précisément le cas des voies métaboliques : elles dérivent les unes des autres par modification/addition et se sont donc transformées dans le temps.

3 – Quelles sont les premières fonctions enzymatiques du vivant ? Quels furent les premiers métabolismes disponibles ? Pourquoi le catabolisme des acides aminés est-il important ? Comment se place le cycle de Krebs ?

Comparer les voies métaboliques universelles à tous les êtres vivants de manière à en inférer l'histoire nous donne les moyens d'inférer non seulement quelles sont les voies métaboliques réalisées le plus précocement à l'origine des systèmes vivants tels que nous les connaissons aujourd'hui, mais surtout les moyens d'inférer les premières fonctions enzymatiques qui ont dû être disponibles. Pour cela, la phylogénie appliquée aux voies métaboliques s'est immédiatement intéressée au catabolisme des acides aminés (Cunchillos & Lecointre 2002, 2003). En effet, il est classique de considérer que les voies du catabolisme ont dû précéder (de peu) les voies de l'anabolisme pour une raison simple : même si Horowitz (1945) conçoit un développement indépendant de l'anabolisme, de l'énergie est requise pour la réalisation des voies anaboliques. Il faut bien que cette énergie soit tirée de la rupture de liaisons chimiques de quelque molécule d'origine abiotique ou biotique présente dans le milieu. Les voies de biosynthèse se sont probablement développées en parallèle mais avec un certain délai (Cunchillos 1997). S'il faut qu'il y ait eu, aux origines de la vie, un premier catabolisme, pourquoi préférer l'hypothèse d'un catabolisme des acides aminés plutôt que celui des acides nucléiques, des acides gras ou des monosaccharides ? Plusieurs arguments importants sont à prendre en compte. On trouve naturellement des acides aminés en milieu abiotique. Il est possible de reconstituer expérimentalement et de manière abiotique des acides aminés simples (glycine, alanine, acide glutamique et acide aspartique) comme l'avaient fait Miller et Urey en 1953 à partir de molécules très

lecointre-biochimie.indd 714 6/08/10 20:19:52

simples comme l'ammoniac, l'hydrogène, le méthane et l'eau[18]. En présence de formaldéhyde, on synthétise dans certaines conditions abiotiques la sérine. D'autres schémas réactionnels furent proposés pour la synthèse abiotique des autres acides aminés aliphatiques.

D'autre part, sur le plan structural, les acides aminés sont plus riches que les autres composés cités ci-dessus. Les acides aminés ont des chaînes latérales variées, tandis que les polysaccharides, les acides gras sont de structure monotone et diffèrent entre eux, surtout par leur nombre de carbone. Parmi les candidats classiques aux premières molécules donneuses d'énergie (acides gras, monosaccharides, acides aminés), les acides aminés sont les seuls composés capables de fournir les groupes réactifs utiles à la synthèse de la plupart des autres composés biochimiques du métabolisme central. Autre argument, la complexité de leur métabolisme est la plus grande. Le catabolisme des acides aminés fait intervenir dix coenzymes, un très grand nombre d'enzymes, et les réactions successives de dégradation sont différentes. De plus, les voies de synthèse sont différentes des voies de dégradation. En comparaison, les autres catabolismes font intervenir trois coenzymes, un petit nombre d'enzymes, les étapes de dégradation sont monotones et la synthèse des composés est réalisée par fonctionnement inverse des voies de dégradation. Si l'on raisonne en termes d'économie cellulaire, on admettra que la sélection naturelle conduit nécessairement à des solutions biochimiques simples et monotones. L'hypothèse la plus parcimonieuse est donc de considérer que la complexité du métabolisme des acides aminés a été comme imposée de l'extérieur au protobionte, comme résultant d'une nécessité initiale, et qu'au sein de celui-ci, par la suite, c'est-à-dire dès que possible, la sélection naturelle n'a conduit qu'à la mise en place de réactions monotones. Il est efficace de produire beaucoup de réactions dégradatives avec peu d'enzymes. À l'inverse, on comprendrait mal, en termes de sélection naturelle, la complexité du catabolisme des acides aminés s'il s'agissait d'un métabolisme tardif. Un argument définitif, si l'on s'intéresse au métabolisme cellulaire défini comme un ensemble coordonné d'activités enzymatiques produites par les protéines, est la nécessité très précoce d'un anabolisme et d'un catabolisme des acides aminés performants pour assurer la disponibilité des acides aminés nécessaires à la synthèse de ces protéines.

Tous ces arguments expliquent que, pour qui s'intéresse aux toutes premières voies métaboliques du vivant, le catabolisme des acides aminés aliphatiques est un candidat de choix. Pour ce qui concerne les acides aminés aromatiques, leur métabolisme fait intervenir des acteurs qui n'ont pu exister qu'une fois le métabolisme des acides aminés aliphatiques mis en place. De plus, ses réactions consomment beaucoup d'oxygène ; et il fait donc figure de métabolisme plus

18. Cf. Tirard, ce volume. *(NdÉ.)*

lecointre-biochimie.indd 715 6/08/10 20:19:52

tardif (Cordón 1990) ; l'ensemble de ces faits ayant été confirmés par l'analyse phylogénétique des voies métaboliques (Cunchillos & Lecointre 2007).

Reste la question du cycle de Krebs. La comparaison du métabolisme des différents types de cellules actuelles met en évidence une grande variabilité : différentes formes d'hétérotrophie, d'autotrophisme photosynthétique et d'autotrophisme chimiosynthétique, diverses formes de respiration et de fermentation. On constate cependant sous cette diversité l'existence d'un noyau métabolique commun universel, formé de quelques cinquante voies : les voies de synthèse et de dégradation des acides aminés et des acides gras, la glycolyse, la glycogenèse, la voie des pentoses et le cycle de Krebs. Cette universalité fut détectée dès 1926 par Kluyver, qui formula le concept d'« unité de la biochimie », concept utilisé dès 1949 par Van Niel comme argument en faveur de la monophylie de toutes les cellules vivantes. Arrêtons nous un instant sur cette hypothèse. À l'encontre de cette monophylie, on peut toujours imaginer des réactions catalysées par des enzymes en milieu abiotique, c'est-à-dire en dehors de toute cellule. Il est probable qu'il a existé un métabolisme précellulaire, correspondant à l'activité enzymatique de protéines isolées, non coordonnées en un ensemble organisé. Mais ce dont nous traitons ici ne concerne pas les réactions enzymatiques isolées, mais bien un ensemble de transformations organisées en voies et fonctionnant de manière coordonnée. On ne conçoit pas facilement la mise en place d'une telle coordination sans une barrière différenciant un milieu intérieur d'un milieu extérieur, et réalisant un tri parmi les substrats. Cette coordination n'a pu se mettre en place qu'au cours d'une étape cellulaire de l'évolution, si on prend en compte le constat d'Horowitz (1945), selon lequel les réactions enzymatiques n'ont de sens (de cohérence) pour la cellule que lorsqu'on les considère dans leur ensemble. Elles en manquent totalement lorsqu'elles sont considérées individuellement ; étant donné que beaucoup de métabolites intermédiaires ne présentent aucun intérêt physiologique, et que la capacité de les produire *isolément* ne confère aucun avantage sélectif concevable. L'apparition d'une nouvelle enzyme ne sera avantageuse que lorsque son activité s'intégrera dans le métabolisme antérieur et en améliorera l'efficacité. En d'autres termes, l'intégration du métabolisme telle que la conçoit Horowitz (1945) ne prend de sens que dans un milieu clos soumis à la sélection naturelle que l'on peut appeler « protobionte » ou « cellule ».

Mais revenons au cycle de Krebs. Du fait de son universalité et surtout de sa position centrale dans le métabolisme cellulaire universel (fig. 3), point de confluence de tous les métabolismes, le cycle de Krebs (ou cycle de l'acide citrique) a été considéré comme la première étape fondamentale de l'évolution biochimique. Mais cette vision se trouve remise en cause par d'autres arguments. Les molécules qui servent de points d'entrée dans le cycle de Krebs, les céto-acides et les acyl-coenzymes A, sont des métabolites intermédiaires

lecointre-biochimie.indd 716 6/08/10 20:19:52

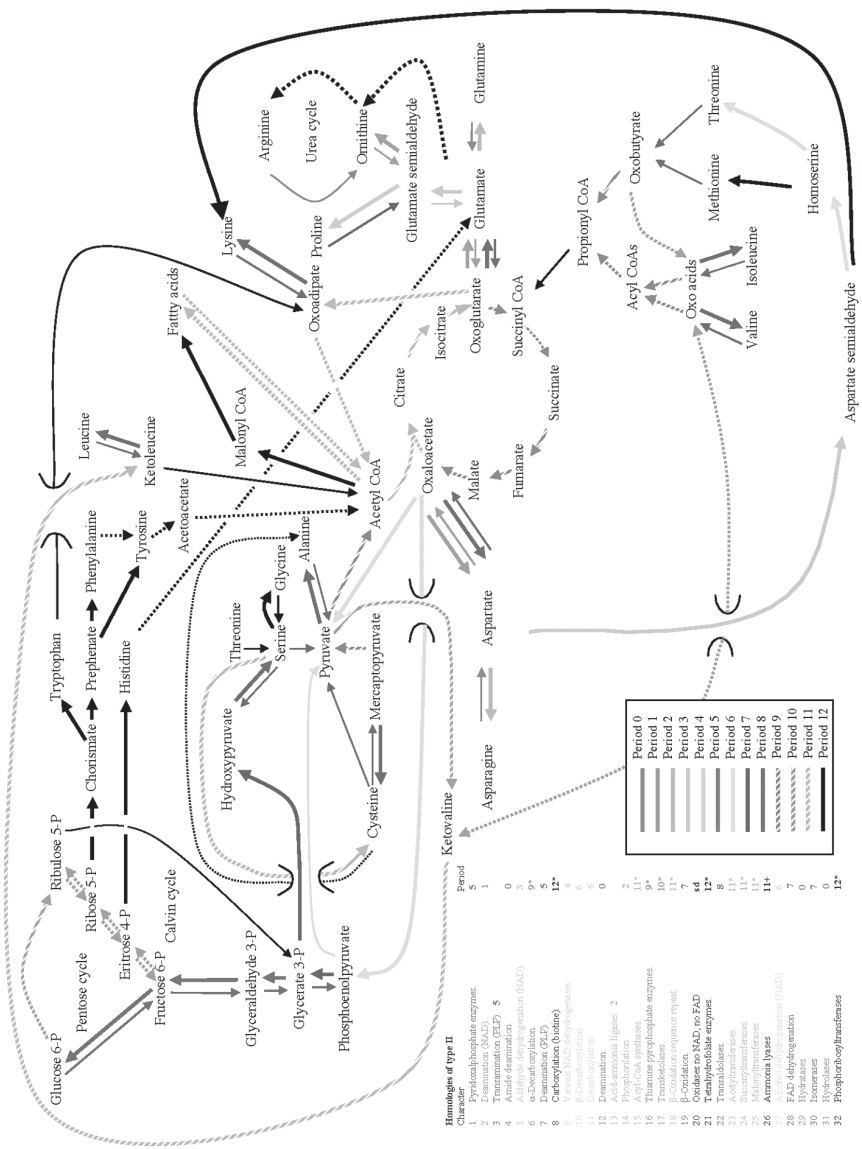

Figure 3. Résumé général des voies métaboliques universelles concernant les acides aminés, les monosaccharides et les acides gras, et de leurs points d'entrée dans le cycle de Krebs. Les niveaux de gris signalent l'ordre d'apparition des réactions tel qu'on peut le mettre en place à partir de cette analyse phylogénétique. Le noir (« période 12 ») ne correspond pas à une phase unique et homogène, mais comprend seulement toutes les phases postérieures à la dernière. Les voies en pointillés sont celles qu'il n'a pas été possible d'assigner à une période.

lecointre-biochimie.indd 717 6/08/10 20:19:53

difficilement imaginables comme disponibles en condition abiotique. Elles sont plus vraisemblablement déjà le produit d'une activité biochimique périphérique. Quant à l'universalité du cycle de Krebs, elle ne lui confère pas l'exclusivité car d'autres métabolismes cités plus haut sont tout aussi universels. L'origine du cycle de Krebs a été considérée comme secondaire et composite par Schoffeniels (1981, 1984), Gest (1981, 1987) et Meléndez-Hevia *et al.* (1996). Ces derniers auteurs considèrent que la portion oxaloacétate-alphacétoglutarate du cycle doit son origine à la biosynthèse des acides aminés (notamment la voie qui va des pyruvate et oxaloacétate au glutamate). Cependant, dans leur argumentation, le catabolisme des acides aminés comme candidat à cette origine se trouve exclu d'emblée puisqu'ils partent du postulat que la première source d'énergie est la glycolyse anaérobique. Le catabolisme des acides aminés est alors considéré comme inutile : « The selective value of a mechanism to eliminate organic material hardly built was not obvious at all » (Meléndez-Hevia *et al.* 1996). Le catabolisme des acides aminés est en fait un meilleur candidat aux premières réactions biochimiques du vivant, parce qu'il est lui-même un métabolisme périphérique qui prend son point de départ sur des molécules complexes que l'on sait déjà disponibles en conditions abiotiques (ce qui n'est pas le cas pour le glucose). De toutes façons, même si l'on accepte d'emblée le glucose comme source d'énergie première, ce n'est pas une raison suffisante pour en exclure les acides aminés qui eux, sont disponibles.

Pour tester la précédence du cycle de Krebs, il a suffit à Cunchillos & Lecointre (2002, 2003, 2005, 2007) d'intégrer à la matrice des voies métaboliques codées deux portions du cycle de Krebs, afin de constater ultérieurement dans l'arbre si les points de branchement de ces portions sont antérieurs ou postérieurs à ceux des catabolismes d'acides aminés (Cunchillos & Lecointre 2007). L'analyse phylogénétique va-t-elle permettre de répondre à ces questions ? Notre travail est la première tentative d'inférence de l'histoire du métabolisme cellulaire à l'aide de l'outil phylogénétique. Ses objectifs sont (1) l'inférence des relations de parenté de chaque voie catabolique de chacun des seize acides aminés aliphatiques ; (2) l'identification du clade le plus basal et surtout, l'ordre d'apparition dans le temps des différentes fonctions enzymatiques ; (3) de vérifier si les portions du cycle de Krebs font partie des voies les plus anciennes ou sont le produit du catabolisme des acides aminés.

4 – Hypothèses d'homologie

4.1 – Échantillonnage

Le travail réalisé a été appliqué au catabolisme et à l'anabolisme des trois grandes familles de molécules organiques utilisées par tous les êtres vivants : acides aminés, acides gras et monosaccharides. Il s'est donc agi de coder les tronçons des grandes composantes métaboliques universelles (ou quasi universelles) : cycle de Krebs, cycle de Calvin, cycle des pentose-phosphates, cycle

lecointre-biochimie.indd 718 6/08/10 20:19:53

de l'urée, glycolyse, gluconéogenèse, anabolisme et catabolisme des acides aminés et des acides gras. Pour le moment, les métabolismes plus périphériques, qui reposent pour leur réalisation de molécules produites par les grands ensembles métaboliques précédents, n'ont pas été pris en compte. Il s'agit bien sûr des polymères (synthèses et dégradations de l'amidon, du glycogène, des protéines, de l'ADN et de l'ARN) mais aussi des composés plus complexes comme les coenzymes, des triglycérides, des phospholipides, des nucléotides et des composés stéroïdiens. Nous ne traitons donc ici que du métabolisme des « briques » de base nécessaires à ces composés. Une voie est généralement définie comme l'ensemble des réactions biochimiques qui s'opèrent entre son extrémité (produit final pour les voies anaboliques) ou molécule abiotique exploitée (pour les voies cataboliques). Chaque voie métabolique est considérée comme un taxon. Lorsqu'une voie de biosynthèse (voie anabolique) ou de dégradation (voie catabolique) peut emprunter plusieurs « parcours » (par exemple cas de la dégradation de la cystéine, de l'acide aspartique, de l'asparagine, de la glutamine et de la thréonine), chacun de ces parcours est pris en compte séparément comme autant de voies – par exemple dCYS1, dCYS2, dCYS3 sont la dégradation (« d ») de la cystéine (CYS) selon les voies n° 1, n° 2, et n° 3 respectivement. Les voies de synthèses sont appelées « s » (sGLU1 pour la synthèse du glutamate selon la voie n° 1, sGLU2 pour la synthèse du glutamate selon la voie n° 2). Le cycle de l'acide citrique (« KC ») ou cycle de Krebs est incorporé en deux tronçons, chacun commençant par un céto-acide pour deux raisons. La première est que ces céto-acides constituent des points d'entrée dans le cycle, et la seconde est que la désamination d'un acide aminé fournit un céto-acide. Autrement dit, le métabolite le plus proche structuralement des acides aminés est le céto-acide. Ainsi, le tronçon « KC1 » va de l'acide oxaloacétique à l'acide alphacétoglutarique, et le tronçon « KC2 » de l'acide alphacétoglutarique à l'acide oxaloacétique. La matrice contient ainsi 75 taxons et 202 caractères (tableau 2 dans Cunchillos & Lecointre 2007), lesquels sont définis par les critères d'homologie explicités ci-dessous.

4.2 – Le codage et les critères d'homologie

Les critères d'homologie vont être exposés des plus stricts aux plus risqués, avec des exemples. L'hypothèse générale de base est que le partage par plusieurs voies métaboliques d'une même enzyme, de plusieurs enzymes successives, d'une réaction enzymatique, d'un même cofacteur, d'une même famille de fonctions sont des signes d'une ascendance commune.

Homologies de type I : le partage d'une enzyme à haute spécificité

Deux voies métaboliques utilisent la même enzyme à haute spécificité pour son substrat. C'est l'enzyme elle-même qui constitue l'hypothèse d'homologie primaire. On code 0 si l'enzyme est absente de la voie et 1 si elle est présente. Exemple : les catabolismes de la méthionine et de la thréonine emploient tous

lecointre-biochimie.indd 719 6/08/10 20:19:53

deux une enzyme qui transforme le cétobutyrate en propionyl-coenzyme A. Cette enzyme est la cétobutyrate-déshydrogénase, et sa spécificité est la même dans les deux voies. Le caractère s'appelle donc « cétobutyrate-déshydrogénase », il est codé 1 pour dTHR et dMET et 0 pour dGLY, dLYS, etc.

Homologies de type II : le partage de fonctions ou types de réactions

IIa : le partage de fonctions enzymatiques

Deux voies de dégradation utilisent la même fonction enzymatique (on ne considère pas ici la spécificité de l'enzyme pour un substrat) ; c'est-à-dire le même type de transformation. L'hypothèse sous-jacente est que deux réactions qui se ressemblent sont réalisées pas des sites actifs semblables. En d'autres termes, on fait le présupposé que ces fonctions enzymatiques sont sous-tendues par une similitude structurale des sites actifs correspondants. L'évolution supposée est qu'à son origine la fonction enzymatique est réalisée à l'aide d'un site actif « ancestral généralisé », et travaillait dans les voies où on les trouve spécialisées aujourd'hui. Par la suite, cette fonction s'est diversifiée en réactions plus spécifiques à des substrats divers mettant en œuvre des sites actifs spécialisés[19]. On code 0 si le substrat est présent mais la fonction est non remplie ; on code 1 si la fonction est remplie, et on code « ? » lorsqu'il n'y a pas le substrat requis. Exemple : les catabolismes de l'alanine et de l'acide aspartique passent par des transaminations utilisant exactement les mêmes mécanismes réactionnels, réalisés par deux enzymes similaires mais différant par leur spécificité vis-à-vis de leurs substrats respectifs : l'alanine-transaminase et l'aspartate-transaminase.

IIb : le partage de cofacteurs

Le partage de cofacteurs traduit une similarité dans les mécanismes enzymatiques. Les réactions ont même signification fonctionnelle. Si plusieurs voies emploient le même cofacteur sans similarité dans le mécanisme enzymatique associé, alors le cofacteur commun est acquis par convergence et c'est le mécanisme enzymatique lui-même qui est codé en homologie de type IIa. Exemple : en fait, ce type d'homologie ne s'est trouvé applicable qu'au pyridoxal-phosphate (PLP), dont la signification fonctionnelle est la désamination. On code 0 lorsque la désamination utilise un autre cofacteur ou lorsqu'il n'y a pas de désamination alors que la réaction est possible ; on code 1 pour une désamination-PLP (désamination directe ou transamination utilisant le PLP) ; et ? si les substrats de la voie ne permettent aucune transamination ni désamination possible (cas des deux tronçons du cycle de Krebs).

Ce critère subit des restrictions dans son application. Elles sont de trois types. Premièrement, lorsqu'un cofacteur est trop spécifiquement lié à une enzyme,

19. Holden (1968), Jallon (1983), Piérard (1983), Petsko *et al.* (1993), O'Brien & Herschlag (1999).

lecointre-biochimie.indd 720 6/08/10 20:19:53

on donne un poids de 2 à une homologie de type I. Par exemple, l'utilisation de la biotine comme cofacteur est restreinte à la seule propionyl-carboxylase. Coder le cofacteur revient à donner un poids de 2 au caractère « propionyl-carboxylase ». Ce cofacteur ne sera donc pas pris en compte car l'information qu'il apporte se superpose à celle de l'enzyme. Deuxièmement, lorsqu'un cofacteur est trop spécifique à une fonction, on donne un poids de 2 à une homologie de type IIa. Par exemple, la thiamine est spécifique des alpha-décarboxylations. Coder la thiamine revient à donner un poids de 2 aux alpha-décarboxylations. Troisièmement, il est risqué de coder des cofacteurs trop ubiquistes ; cela comporte un trop grand risque d'homoplasie. Il est nécessaire dans ce cas de coder le type de réaction associée, ce qui produit donc un retour à une homologie de type IIa. Par exemple, le NAD est utilisé dans une large gamme de réactions : NAD-désaminations, NAD-aldéhyde-acide-déshydrogénases, NAD-béta-oxydations, NAD-alpha-décarboxylations. Les recoder en tant que telles reviendrait à donner un poids de 2 à des homologies de type IIa.

IIc : communauté de famille de fonction

Ce critère d'homologie est semblable dans son principe au critère IIa, mais simplement étendu. L'idée de base est que les enzymes ont dû évoluer à partir de très faibles spécificités, vers de plus hautes spécificités ; c'est-à-dire de sites actifs ancestraux généralisés vers des sites actifs spécialisés. On code 0 lorsque la fonction n'est pas réalisée bien que possible, on code 1 lorsque la fonction est réalisée. On code « ? » lorsque la réaction est non applicable aux groupements chimiques présents dans les composés de la voie.

Par exemple, la décarboxylation (caractère 30). On code 0 lorsque la décarboxylation n'est pas réalisée bien que possible, on code 1 lorsque la décarboxylation est réalisée dans la voie. En fait, ce type d'homologie concerne seulement deux grandes familles de réactions : les décarboxylations et les désaminations (les caractères 30 et 31). On code ? pour CC1 et CC2 au caractère 31 car il n'y a pas de désamination possible au sein même du cycle de Krebs.

IId : récurrence

Des voies métaboliques peuvent se ressembler en cela qu'elles produisent de la même façon une récurrence de réactions, récurrence d'une même réaction réalisée par la même enzyme mais sur des substrats différents. Ce type de phénomène intervient dans l'allongement (ou la dégradation) de molécules monotones comme les acides gras.

4.3 – Enracinement

Si la question est d'inférer quelles furent les premières voies métaboliques et quelles furent les premières réactions enzymatiques du vivant, il se pose le problème de l'enracinement de l'arbre. En effet, en reconstruction phylogénétique, la parcimonie informatisée ne construit qu'un graphe connexe non cyclique.

lecointre-biochimie.indd 721 6/08/10 20:19:53

Pour que celui-ci soit interprétable en termes d'arbre, c'est le *biologiste* qui indique au logiciel où se place la racine, en d'autres termes le biologiste doit indiquer quel est son extra-groupe[20]. Pour le cas présent, il n'y a pas d'extra-groupe à choisir, ce qui rend le problème de l'enracinement singulier. Pour deux enracinements possibles correspondent deux problèmes. Le premier enracinement pourrait consister à enraciner l'arbre à son point moyen. Ceci suggère et même implique qu'il y a à peu près autant de quantité d'évolution de caractères le long des différentes branches. Compte tenu de la présente problématique, on ne voit pas quel « modèle d'évolution » biochimique pourrait justifier l'hypothèse d'une accumulation égale dans deux branches sœurs de l'innovation enzymatique. Cette contrainte-là n'a pas été théorisée et ne peut recevoir d'argument, ni même emporter l'intuition. Le second enracinement est celui de l'ancêtre hypothétique « tout-zéro », sachant que l'état « zéro » est donné dans les caractères pour les absences d'enzymes, les absences de fonctions réalisées, ou des absences d'utilisation d'un cofacteur. Cet enracinement-là est préférable au premier, parce qu'on peut toujours imaginer une situation ancestrale virtuelle où aucune de ces réalités biochimiques n'était présente. Cependant, il faut garder à l'esprit qu'un tel enracinement conférera automatiquement une position basale aux voies les plus simples, c'est-à-dire faisant intervenir peu d'enzymes et/ou peu de réactions, quelle que soit la nature de ces réactions.

5 – Résultats et mise en ordre des événements

Un algorithme de parcimonie standard a donc été appliqué sur la matrice en question. La figure 4 montre l'arbre obtenu (détails techniques dans Cunchillos & Lecointre 2007). Sur cet arbre, il a été défini des périodes selon une méthode que l'on résumera de la manière suivante. Les changements de caractères se produisant le long des branches sont repérés lorsque ces caractères correspondent à des homologies de type II. Ceux de ces événements d'acquisition d'une nouvelle fonction enzymatique (type II) dont l'emplacement dans l'arbre n'est pas ambigu ni multiple servent de « bornes » pour définir de nouvelles périodes. Ainsi, l'arbre peut-il être segmenté en tranches de temps (ou « périodes »), ces tranches de temps relatif étant bornées par des acquisitions d'homologies de type II. Cependant, une petite difficulté subsiste. Lorsqu'un arbre s'éloigne de la forme d'un peigne (arbre totalement asymétrique) et montre des déploiement divergents (on dit qu'il acquiert une forme de symétrie), il devient de plus en plus difficile de comparer les bornes de temps posées dans deux parties résolument divergentes de l'arbre. Ceci ne serait possible qu'en décidant une sorte d'horloge dans les acquisitions d'homologies de type II, une sorte de régularité

20. Groupe d'organismes ne faisant pas partie d'un groupe analysé phylogénétiquement et choisi afin d'orienter les transformation des caractères du primitif au dérivé. *(NdÉ.)*

lecointre-biochimie.indd 722 6/08/10 20:19:53

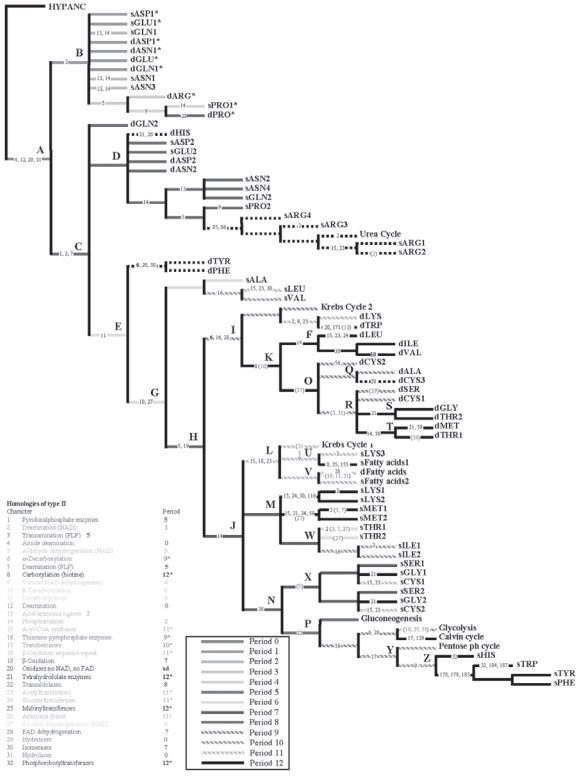

Figure 4. Consensus strict des 20 arbres équiparcimonieux fournis à partir d'une matrice de 75 taxons et 207 caractères (Cunchillos & Lecointre 2007). Les arbres sources ont 347 pas, un IC de 0,58 et un IR de 0,79. Les longueurs de branches n'ont pas de signification ici. Les chiffres renvoient à des changements concernant le caractère portant le même chiffrre dans la matrice source. Les niveaux de grisé renvoient aux périodes (cf. texte) relatives du développement métabolique tiré des acquisitions d'homologies de type II dans l'arbre. Les noms en bout de branche réfèrent au nom d'une voie métabolique (taxon) et comportent le code à trois lettres pour les acides aminés (PRO pour proline, ARG pour arginine, etc.) précédé d'un « d » (dPRO) s'il s'agit de la voie de dégradation (catabolisme) de cet acide aminé ou d'un « s » (sPRO) s'il s'agit de la voie de synthèse (anabolisme) de cet acide aminé. Les chiffres qui suivent réfèrent aux différentes façons d'aboutir à une synthèse (sPRO1, sPRO2) ou à une dégradation (dASP1, dASP2) du même acide aminé. HYPANC : extra-groupe « tout-zéro », Urea Cycle : cycle de l'urée, Krebs Cycle 2 : portion n° 2 du cycle de Krebs (cétoglutarate vers oxaloacétate), Krebs Cycle 1 : portion n° 1 du cycle de Krebs (oxaloacétate vers cétoglutarate), sFatty acids 1 : synthèse des acides gras selon la voie n° 1, sFatty acids 2 : synthèse des acides gras selon la voie n° 2, dFatty acids : dégradation des acides gras, Gluconeogenesis : glyconeogenèse, Glycolysis : Glycolyse, Calvin Cycle : cycle de Calvin.

temporelle. Pour ne pas faire cette hypothèse lourde, certaines bornes ont été mises en correspondance temporelle pour des raisons biochimiques. Ainsi, dix liens temporels joignant des parties divergentes de l'arbre ont été décrits et utilisés (bien entendu, par souci de place nous tairons ici ces aspects très techniques, ainsi que d'autres concernant la mise en période le l'arbre, pour plus d'information cf. Cunchillos & Lecointre 2007).

La figure 4 montre l'ordre relatif d'apparition des voies métaboliques du vivant au sein de douze périodes ainsi définies (de 0 à 11, la dernière – « 12 » – n'étant que la période postérieure à toutes les autres). On peut résumer les informations de la manière suivante (seules les classes de réactions sont données ici) :

Période 0 : les premières fonctions enzymatiques disponibles sont des désaminations, des hydratases et des hydrolases.

Période 1 : désaminations utilisant le cofacteur NAD.

Période 2 : phosphorylations.

Période 3 : aldéhyde déshydrogénations utilisant le NAD.

Période 4 : diverses déshydrogénases utilisant le NAD.

Période 5 : transaminations et désaminations utilisant le pyridoxal-phosphate.

Période 6 : bêta-décarboxylations, alcool-déshydrogénases utilisant le NAD.

Période 7 : bêta-oxydations, déshydrogénases utilisant le FAD, isomérases.

Période 8 : transaldolases.

Période 9 : alpha-décarboxylations, enzymes utilisant le thiamine-pyrophosphate.

Période 10 : transcétolases.

Période 11 : acétyl-coenzyme A synthétase, acétyl-transférases, séquences répétées de bêta-oxydations, succinyltransférases.

Période 12 : carboxylations utilisant la biotine, enzymes utilisant le tétra-hydrofolate, phospho-ribosyl-transférases, malonyl-transférases.

Ce qui se traduit, pour une mise en ordre de l'apparition des grandes fonctions enzymatiques :

Phase 1 : catabolisme des acides aminés.

Phase 2 : anabolisme des acides aminés.

Phase 3 : glycolyse, glycogenèse, fermeture du cycle de l'urée (peut-être disponible dès la phase 2).

Phase 4 : fermeture du cycle des pentose-phosphates.

Phase 5 : fermeture du cycle de Krebs et disponibilité du métabolisme des acides gras.

Phase 6 : fermeture du cycle de Calvin.

lecointre-biochimie.indd 724 6/08/10 20:19:53

Le cladogramme, par ses branches internes où figurent des événements, fournit les tranches de temps qui permettent d'ordonner un certain nombre d'innovations enzymatiques. On constate que cette approche fournit une confirmation de l'interprétation qu'avait proposée Cordón (1990) de l'évolution du catabolisme des acides aminés, qui est antérieur à leur anabolisme. Les premières époques, qui sont celles des désaminations et des transaminations, correspondent précisément à l'exploitation des acides aminés d'origine abiotique. Le catabolisme des acides aminés a probablement été ce que les première enzymes du monde vivant ont dû savoir faire, quelle que soit leur origine, c'est-à-dire avec ou sans ADN. Après les transaminations apparaissent les décarboxylations (phase 6). La seconde époque de Cordón était effectivement celle des décarboxylations ajoutées aux réactions précédentes. Dans nos résultats, les bêtadécarboxylations apparaissent avant les alphadécarboxylations. Il est tout à fait concevable que les bêtadécarboxylations aient été les premières puisque n'utilisant aucun cofacteur. De plus, ces décarboxylations sont spontanément possibles en milieu acide. Conformément au scénario de Cordón, après notre période 6, il est possible de dégrader et de synthétiser tous les acides aminés des groupes I et II de Cordón : asparagine, aspartate, arginine, proline, glutamine et glutamate. L'exploitation et la synthèse des acides aminés des groupes III (sérine, glycine, cystéine, alanine) et IV (thréonine, valine, isoleucine, méthionine) de Cordón se développent de la phase 5 à 11. Enfin, Cordón prévoyait une troisième phase constituée de l'acquisition tardive de réactions complexes complémentaires et de la fermeture du cycle de Krebs. Les résultats de l'analyse phylogénétique montrent une fermeture de ce cycle assez tardive.

On constate dans cet arbre une cohérence avec les vues d'Horowitz (1945) et de Cordón (1990). Les premières étapes du catabolisme des acides aminés (désaminations, transaminations) sont les premières à émerger dans l'évolution. Il faut bien garder à l'esprit qu'aucun élément de codage n'a pris en compte les idées de Cordón (1990) sur le sens de l'évolution de ces voies. Même si la méthode d'enracinement de l'arbre implique que les voies les plus simples soient à la base, il revient à la structure même des voies du catabolisme des acides aminés l'inscription des désaminations et transaminations dans les catabolismes les moins complexes (plutôt que des décarboxylations). En d'autres termes, que les voies les plus simples soient à la base du cladogramme n'a rien de surprenant compte tenu de notre méthode d'enracinement, mais que ce soient précisément des désaminations et des transaminations, premières réactions dégradatives du métabolisme, qui soient les plus simples, cela nous ne le devons pas à notre méthode d'enracinement. Le cycle de Krebs faisant intervenir des décarboxylations, il ne peut pas être, dans ce contexte, antérieur aux premières voies de dégradation des acides aminés. Il en est donc l'une des conséquences.

lecointre-biochimie.indd 725 6/08/10 20:19:53

Le catabolisme est complet avant l'anabolisme pour certains acides aminés mais pas tous. Ceci est vrai pour l'asparagine (catabolisme disponible dès la période 1), la glutamine (période 1) et l'arginine (période 3). Pour beaucoup d'autres acides aminés, la biosynthèse est rendue possible en même temps que sa dégradation, tout au moins pour un certain nombre de cas on ne peut pas dissocier l'ordre d'apparition des deux types de réactions compte tenu des outils mis en œuvre ici : il s'agit de l'aspartate (les deux sont possibles dès la période 1), le glutamate (1), la lysine (5), la cystéine (11), et plus tard encore pour la méthionine et la glycine. La biosynthèse est rendue complète avant que la dégradation ne la soit pour la proline (biosynthèse disponible dès la période 2), la thréonine (6), l'alanine (6), la sérine (7), la valine (10), l'isoleucine (10), la leucine (11), et encore plus tard pour l'histidine, la phénylalanine, le tryptophane et la tyrosine (ce sont les acides aminés aromatiques). Contrairement aux intuitions de Horowitz et Cordón, selon lesquelles le développement des voies de biosynthèse (voies anaboliques) serait rétrograde et celui des voies de dégradation (voies cataboliques) antérograde, on constate que ce schéma conceptuel est loin de s'appliquer de manière tranchée aux acides aminés. En effet, le schéma de développement est antérograde pour un jeux de voies métaboliques qui peuvent tout aussi bien être des voies de dégradation (« d ») que des voies de synthèse (« s ») : sAsn, dAsn, sGln, dGln, sPro, sArg, dAla, dSer, sCys, dCys, sMet, dMet, dThr, dIle, dVal, dLeu, dLys, sTyr, sPhe, sHis, sTrp. Le développement n'est rétrograde que pour dPro, sAla, sLys. Pour le reste des voies, le test phylogénétique est non concluant. Cette absence de corrélation provient en partie simplement des nombreuses interconnections métaboliques que les voies entretiennent entre elles, tandis que les modèles abstraits d'Horowitz et de Cordón n'en incluaient pas autant.

Il est possible de différencier dans le cycle de Krebs une succession : la portion « KC2 » (cétoglutarate vers oxaloacétate) est disponible avant la portion « KC1 » (oxaloacétate vers cétoglutarate). Dans un contexte interprétatif différent de celui-ci, c'est-à-dire en prenant la glycolyse anaérobique comme première source d'énergie (et non des acides aminés abiotiques), Gest (1981, 1987) et Meléndez-Hevia *et al.* (1996) ont décrit les processus par lesquels KC1 pourrait provenir précocement de la biosynthèse du glutamate à partir du pyruvate et de l'oxaloacétate. Les résultats de l'analyse phylogénétique ne sont pas incompatibles avec cela. La disponibilité de l'oxaloacétate par dégradation très précocement disponible de l'asparagine ou de l'aspartate (dès la période 1), celle du pyruvate (dès la période 5) rend l'hypothèse probable, la transformation du glutamate en oxaloacétate étant possible dès la période 1 ou 5 selon les processus à l'œuvre. Quoi qu'il en soit, la prise en compte d'une disponibilité d'un glutamate d'origine abiotique rend ces considérations facultatives. La glycolyse va jusqu'au pyruvate dès la période 8 et à l'acétyl-

lecointre-biochimie.indd 726 6/08/10 20:19:53

coenzyme A en période 9 (et donc complète en période 9), avant la fermeture complète du cycle de Krebs (période 11). Ceci confirme les conclusions de Meléndez-Hevia *et al.* (1996) selon lesquelles la glycolyse est complète avant la fermeture du cycle de Krebs.

6 – Conclusion

Le concept de «*descent with modification*» de Darwin s'applique aux voies métaboliques. L'exportation de la pensée phylogénétique en biochimie est fructueuse. En inventant de *nouveaux types de taxons*, les voies métaboliques, et de *nouveaux types de caractères*, les enzymes elles-mêmes, les fonctions enzymatiques, les cofacteurs utilisés, et les familles de fonctions enzymatiques utilisées par ces voies, il fut donc possible d'obtenir les premières matrices d'homologies primaires, les premières phylogénies du métabolisme universel (Cunchillos & Lecointre 2002, 2003, 2005, 2007). Leurs résultats sont compatibles avec le scénario de Cordón (1990), lequel avait été produit en dehors de toute analyse cladistique mais dans un cadre théorique plus large et bien structuré (cf. Cunchillos 2004). L'exportation de la pensée phylogénétique hennigienne en biochimie a permis d'aboutir à des conclusions dont certaines peuvent apparaître comme attendues. Cependant, c'est la transparence des procédures amenée par la formalisation des observations qu'imposent la matrice et son traitement par parcimonie informatisée qui accroît la valeur heuristique de ces conclusions. Les désaminations, les hydrolases et les hydratases furent les premières réactions enzymatiques du vivant (le vivant tel que nous le connaissons aujourd'hui). Ces réactions ont été utilisées pour exploiter les acides aminés aliphatiques d'origine abiotique (et en tout premier lieu la glutamine, l'asparagine, le glutamate, l'aspartate, puis l'arginine). Le modèle théorique d'un développement évolutif antérograde pour les voies de dégradation et d'un développement évolutif rétrograde pour les voies de biosynthèse s'applique mal aux acides aminés. Pour un certain nombre d'acides aminés aliphatiques, leur catabolisme fut rendu possible et complet avant leur synthèse. La glycolyse et la glycogenèse, ainsi que la fermeture du cycle de l'urée, sont rendus possibles avant la fermeture du cycle de Krebs et la disponibilité du métabolisme des acides gras. Le métabolisme de ces derniers, ainsi que la fermeture du cycle de Calvin, sont des événements relativement tardifs, tout comme l'est le métabolisme des acides aminés aromatiques. La méthode et les critères d'homologie développés ici rendent les hypothèses d'évolution biochimique explicites, fondées sur les patterns réactionnels, et parcimonieuses, c'est-à-dire formellement cohérentes[21].

21. ***Remerciements.*** Sont vivement remerciés Marc Silberstein, Patrick Tort, Jean-Philippe Touffut et Jean-Louis Beffa pour leur aide à la réalisation de ces travaux, dont l'interdisciplinarité fut promue par l'Institut Charles Darwin international ainsi que la compagnie Saint-Gobain et le Muséum national d'histoire naturelle.

lecointre-biochimie.indd 727 6/08/10 20:19:53

Références bibliographiques

BROWER A. (2000), "Evolution is not a necessary assumption of cladistics", *Cladistics*, 16 : 143-154.

CAVALIER-SMITH T. (1987), "The Origin of Cells : A Symbiosis between Genes, Catalysts and Membranes", *Symp. Quant. Biol.*, LII : 805-824.

COPLEY S.D. (2000), "Evolution of a metabolic pathway for degradation of a toxic xenobiotic : the patchwork approach", *T.I.B.S.,* 25 : 261-265.

CORDÓN F. (1990), *Tratado evolucionista de biologia*, Madrid, Aguilar.

CUNCHILLOS C. (1997), « Les grands axes de l'évolution du métabolisme cellulaire », *in* P. Tort (dir.), *Pour Darwin,* Paris, PUF : 425-447.

CUNCHILLOS C. (2004), « Matérialisme et théorie des unités de niveau d'intégration », *in* J. Dubessy, G. Lecointre & M. Silberstein (dir.), *Les matérialismes (et leurs détracteurs)*, Paris, Syllepse : 199-225.

CUNCHILLOS C. & LECOINTRE G. (2002), "Early steps of metabolism evolution inferred by cladistic analysis of the structure of amino acid catabolic pathways", *Comptes rendus de l'Académie des sciences,* 325 : 119-129.

CUNCHILLOS C. & LECOINTRE G. (2003), "Evolution of amino acid metabolism inferred through cladistic analysis", *Journal of Biological Chemistry*, 278 : 47960-47970.

CUNCHILLOS C. & LECOINTRE G. (2005), "Integrating the universal metabolism into a phylogenetic analysis", *Mol. Biol. Evol.*, 22(1) : 1-11.

CUNCHILLOS C. & LECOINTRE G. (2007), "Ordering events of Biochemical evolution", *Biochimie*, 89 : 555-573.

DE PINNA M.C.C. (1991), "Concepts and tests of homology in the cladistic paradigm", *Cladistics*, 7 : 367-394.

FOTHERGILL-GILMORE L.A. & MICHELS P.A.M. (1993), "Evolution of glycolysis", *Prog. Biophys. Mol. Biol.*, 59 : 105-235.

GEST H. (1981), "Evolution of the citric acid cycle and respiratory energy conversion in prokaryotes", *FEMS Microbiol. Lett.,* 12 : 209-215.

GEST H. (1987), "Evolutionary Roots of the Citric Acid Cycle in Prokaryotes", *Biochem. Soc. Symp.,* 54 : 3-16.

HEAMS T. (2004), « Biologie moléculaire : affronter la crise de la cinquantaine », *in* J. Dubessy, G. Lecointre & M. Silberstein (dir.), *Les matérialismes (et leurs détracteurs)*, Paris, Syllepse : 237-261.

HENNIG W. (1966), *Phylogenetic systematics*, Urbana and Chicago, University of Illinois Press.

HOLDEN J.T. (1968), "Evolution of Transport Systems", *J. Theoret. Biol.,* 21 : 97-102.

HOROWITZ N.H. (1945), "On the evolution of biochemical syntheses", *Proc. Natl. Acad. Sci.*, 31 : 153-157.

JALLON J.-M. (1983), « Les glutamate déshidrogénases de *Escherichia coli* a l'Homme », *in* G. Hervé (dir.), *L'évolution des protéines*, Paris, Masson : 92-103.

JENSEN R.A. (1976), "Enzyme Recruitment in Evolution of new Function", *Ann. Rev. Microbiol.,* 30 : 409-425.

KLUYVER A.J. (1926), "Die Einheit in der Biochemie", *Chemie der Zelle und Gewebe*, 13 : 134-190.

KNOWLES J.R. (1991), "Enzyme catalysis : not different, just better", *Nature*, 350 : 121-124.

KUPIEC J.-J. (2004), « Vers un darwinisme cellulaire », *in* J. Dubessy, G. Lecointre & M. Silberstein (dir.), *Les matérialismes (et leurs détracteurs)*, Paris, Syllepse : 227-236.

KUPIEC J.-J. (2008), *L'origine des individus*, Paris, Fayard.

KUPIEC J.-J., GANDRILLON O., MORANGE M. & SILBERSTEIN M. (dir.) (2009), *Le hasard au cœur de la cellule. Probabilités, déterminisme, génétique*, Paris, Syllepse.

MARTIN W. & MÜLLER M. (1998), "The hydrogen hypothesis for the first eukaryote", *Nature*, 392 : 37-41.

MELÉNDEZ-HEVIA E., WADDELL T.G. & CASCANTE M. (1996), "The Puzzle of the Krebs Citric Acid Cycle : Assembling the Pieces of Chemically Feasible Reactions, and Opportunism in the Design of Metabolic Pathways During Evolution", *J. Mol. Evol.,* 43 : 293-303.

MICHAL G. (1999), *Biochemical pathways*, New York, John Wiley and Sons.

lecointre-biochimie.indd 728 6/08/10 20:19:54

Miller S.A. (1953), "Production of Amino Acids Under Possible Primitive Earth Conditions", *Science*, 117 : 528.
Morange M. (1994), *Histoire de la biologie moléculaire*, Paris, La Découverte.
Morange M. (2005), *Les secrets du vivant,* Paris, La Découverte.
Nelson G.J. & Platnick N. (1981), *Systematics and biogeography : cladistics and vicariance*, New York, Colombia UP.
O'Brien P.J. & Herschlag D. (1999), "Catalytic promiscuity and the evolution of new enzymatic activities", *Chem. Biol.*, 6 : R91-R105.
Petsko P., Kenyon G.L., Gerlt J.A., Ringe D. & Kozarich J.W. (1993), "On the origin of enzymatic species", *T.I.B.S.*, 18 : 372-376.
Pierard A. (1983), « Évolution des systèmes de synthèse et d'utilisation du carbamoylphosphate », *in* G. Hervé (dir.), *L'évolution des protéines*, Paris, Masson : 53-66.
Schoffeniels E. (1981), *Les cahiers de biochimie,* Vaillant-Carmanne, Paris, Maloine.
Schoffeniels E. (1984), *Biochimie comparée*, Paris, Masson.
Van Niel C.B. (1949), "The 'Delft School' and the rise of general microbiology", *Bacteriological Review*, 13 : 161-174.
Zukerkandl E. & Pauling L. (1965), "Molecules as Documents of Evolutionary History", *J. Theoret. Biol.,* 8 : 357-366.

lecointre-biochimie.indd 730 6/08/10 20:19:54