

BRUCE
ALBERTS

ALEXANDER
JOHNSON

JULIAN
LEWIS

DAVID
MORGAN

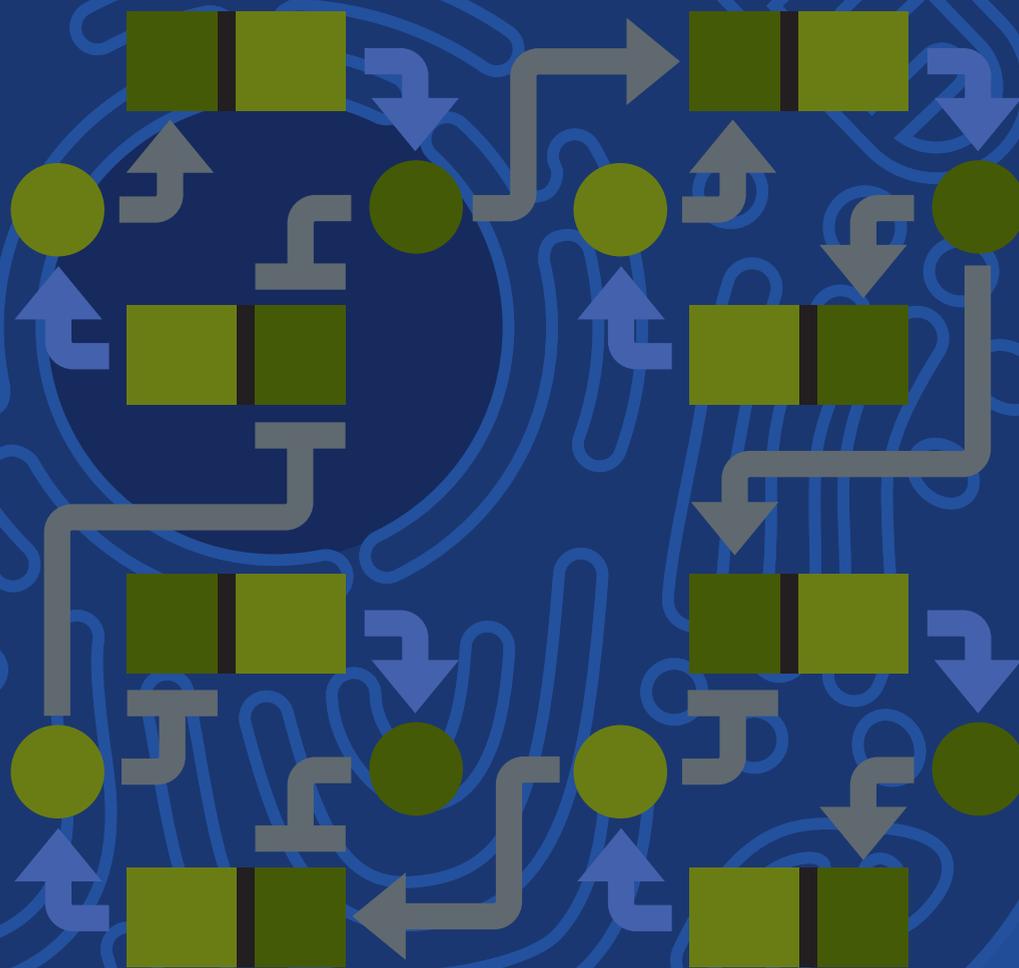
MARTIN
RAFF

KEITH
ROBERTS

PETER
WALTER

Biologie moléculaire de LA CELLULE

Sixième édition



Lavoisier
Médecine
SCIENCES

INTRODUCTION À LA CELLULE

CHAPITRE

1

Cellules et génomes

La surface de notre planète est peuplée de choses vivantes – usines chimiques bizarres, organisées de façon compliquée, qui prélèvent de la matière dans leur environnement et utilisent ce matériel brut pour produire des copies d’elles-mêmes. Ces organismes vivants sont extraordinairement divers. Que pourrait-il y avoir de plus différent qu’un tigre et un fragment d’algue, ou une bactérie et un arbre ? Pourtant nos ancêtres, sans rien savoir des cellules ou de l’ADN, savaient que tous ces organismes avaient quelque chose en commun. Ils ont appelé ce quelque chose « la vie » et s’en sont émerveillés. Ils ont essayé de le définir, et désespéraient de ne pouvoir expliquer ce que c’était, et comment cela fonctionnait par rapport aux choses non vivantes.

Les découvertes du siècle dernier n’ont pas diminué l’émerveillement, bien au contraire. Mais elles ont élucidé le mystère central de la nature de la vie. Nous pouvons maintenant constater que tous les êtres vivants sont constitués de cellules : de petites unités enveloppées de membranes, remplies d’une solution aqueuse concentrée de produits chimiques, et dotées de l’extraordinaire capacité de créer des copies d’elles-mêmes en croissant et en se divisant ensuite en deux.

Puisque les cellules sont les unités fondamentales de la vie, il est logique de s’adresser à la *biologie cellulaire* – l’étude de la structure, de la fonction et du comportement des cellules – pour chercher des réponses aux questions de ce qu’est la vie et comment elle fonctionne. Avec une meilleure compréhension des cellules et de leur évolution, nous pouvons commencer à aborder les grands problèmes historiques de la vie sur Terre : ses origines mystérieuses, sa diversité étonnante, et sa dissémination dans tous les habitats imaginables. En effet, comme l’a souligné il y a longtemps le pionnier de la biologie cellulaire E. B. Wilson, « la clé de tous les problèmes biologiques doit en fin de compte être recherchée dans la cellule ; car chaque organisme vivant est, ou a été à un moment donné, une cellule ».

Malgré leur apparente diversité, les êtres vivants, vus de l’intérieur, sont fondamentalement similaires. Ainsi toute la biologie n’est qu’un contrepoint entre deux thèmes : l’étonnante variété des caractères individuels ; l’étonnante constance des mécanismes fondamentaux. Dans ce premier chapitre, nous commencerons par définir les caractères universels communs à tous les organismes vivants sur notre planète. Nous examinerons ensuite rapidement la diversité des cellules. Et nous verrons comment, grâce au code moléculaire commun dans lequel sont inscrites toutes les spécifications des organismes vivants, il nous est possible de lire, mesurer, et déchiffrer ces spécifications pour comprendre de façon assez satisfaisante toutes les formes de vie, de la plus petite à la plus grande.

DANS CE CHAPITRE

LES CARACTÉRISTIQUES
UNIVERSELLES DES CELLULES
SUR LA TERRE

LA DIVERSITÉ DES GÉNOMES
ET L’ARBRE DE LA VIE

L’INFORMATION GÉNÉTIQUE
CHEZ LES EUCARYOTES

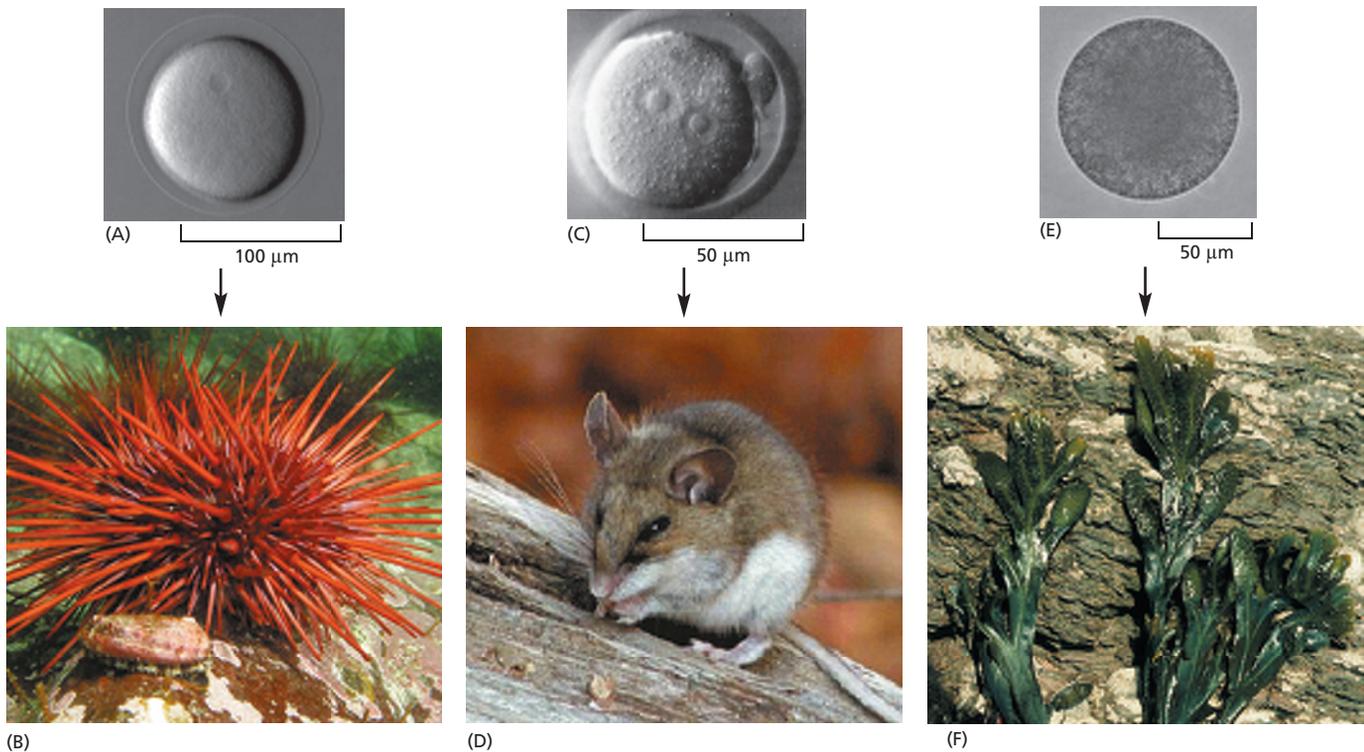


Figure 1-1 Les informations héréditaires contenues dans l'œuf fécondé déterminent toute la structure de l'organisme multicellulaire. Bien que les cellules de départ semblent superficiellement similaires, comme indiqué : un œuf d'oursin de mer donne naissance à un oursin de mer (A et B). Un œuf de souris donne naissance à une souris (C et D). Un œuf d'algue *Fucus* donne naissance à une algue *Fucus* (E et F). (A, dû à l'obligeance de David McClay; B, dû à l'obligeance de M. Gibbs, Oxford Scientific Films; C, dû à l'obligeance de Patricia Calarco, d'après G. Martin, *Science* 209 : 768-776, 1980. Avec l'autorisation d'AAAS; D, dû à l'obligeance de O. Newman, Oxford Scientific Films; E et F, dû à l'obligeance de Colin Brownlee.)

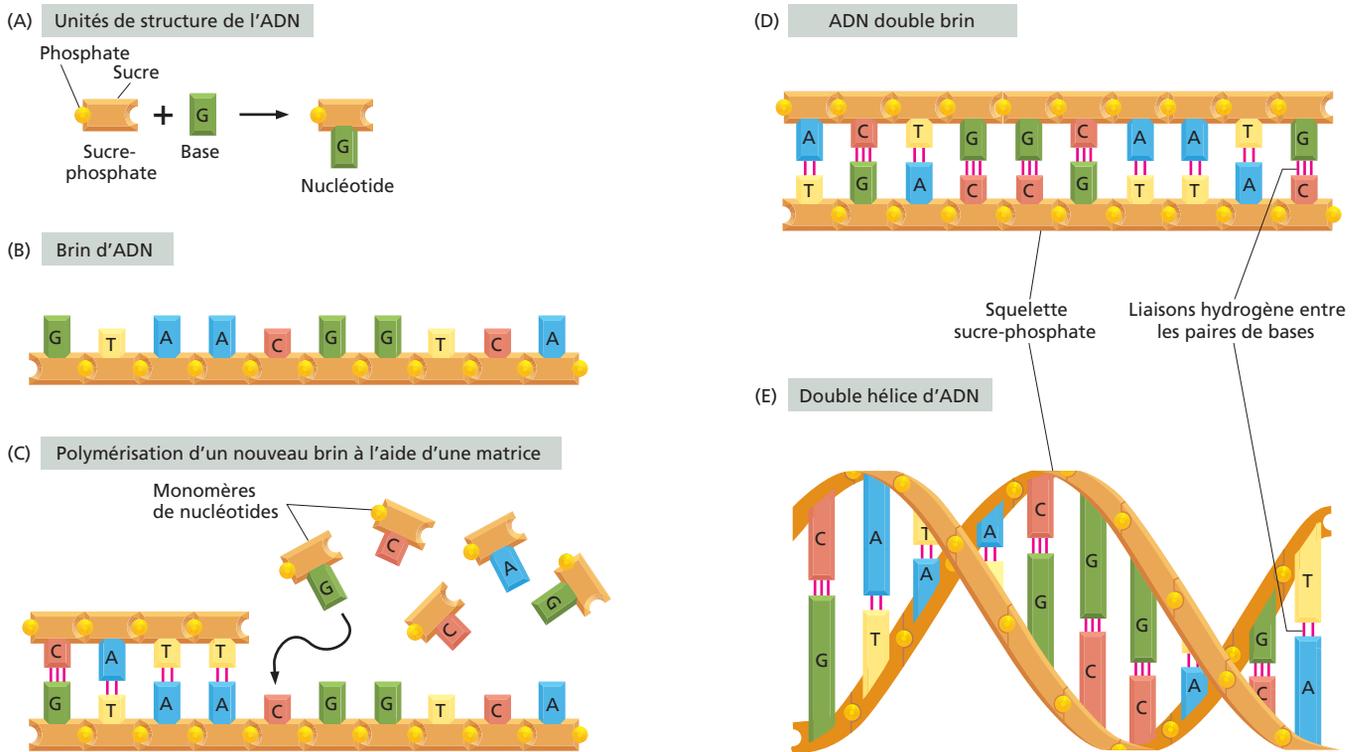
LES CARACTÉRISTIQUES UNIVERSELLES DES CELLULES SUR LA TERRE

On estime qu'il y a aujourd'hui plus de 10 millions – peut-être 100 millions – d'espèces vivantes sur la Terre. Chaque espèce est différente, et chacune se reproduit fidèlement, produisant une descendance qui appartient à la même espèce : les parents transmettent, avec un détail extraordinaire, l'information qui va déterminer les caractéristiques de leur progéniture. Ce phénomène d'*hérédité* est au centre de la définition de la vie : il fait la différence entre la vie et d'autres processus, comme la croissance d'un cristal, la combustion d'une bougie ou la formation de vagues sur l'eau, processus au cours desquels des phénomènes réguliers se produisent mais sans le même type de lien entre les particularités des parents et celles des descendants. Comme la flamme d'une bougie, l'organisme vivant doit consommer de l'énergie libre pour créer et maintenir son organisation. Mais la vie utilise cette énergie libre pour déclencher l'ensemble des processus chimiques extrêmement complexes qui sont spécifiés par l'information héréditaire.

La plupart des organismes vivants sont des cellules isolées. D'autres, comme nous-mêmes, sont de grandes « cités multicellulaires » dans lesquelles des groupes de cellules assurent des fonctions spécialisées et sont reliés par des systèmes de communication complexes. Mais même pour l'agrégat de plus de 10^{13} cellules formant un corps humain, la totalité de l'organisme est le résultat de divisions cellulaires à partir d'une cellule unique. Cette cellule unique est donc porteuse de toute l'information héréditaire qui définit chaque espèce (Figure 1-1). Cette cellule contient la machinerie nécessaire à la collecte de matériaux bruts dans l'environnement, et à la construction, à partir de ces matériaux, d'une nouvelle cellule à son image, complète, et contenant une nouvelle copie de son information héréditaire. Chaque cellule, quelle qu'elle soit, est vraiment stupéfiante.

Toutes les cellules stockent leur information génétique à l'aide du même code chimique linéaire : l'ADN

Les ordinateurs nous ont habitués à concevoir l'information comme une quantité mesurable – un million d'octets (pour enregistrer quelques centaines de pages ou une image à partir d'une caméra digitale), 600 millions d'octets pour la musique d'un CD, etc. Les ordinateurs nous ont aussi permis de prendre conscience que la même information pouvait être enregistrée sous différentes formes physiques : les disques et les bandes magnétiques que nous utilisons il y a vingt ans pour nos



archives électroniques sont devenus illisibles sur les machines d'aujourd'hui. Les cellules vivantes, comme les ordinateurs, stockent l'information, et on estime qu'elles ont évolué et se sont diversifiées depuis plus de 3,5 milliards d'années. On pourrait s'attendre à ce que toutes ne stockent pas l'information sous la même forme, ou que les archives d'un type cellulaire ne soient pas lisibles par la machinerie d'un autre type. Et pourtant, c'est bien le cas. Toutes les cellules vivantes sur la Terre stockent leur information génétique sous forme de molécules d'ADN double brin – longues chaînes *polymériques* sans embranchements, appariées, toujours formées des quatre mêmes *monomères*. Ces monomères, des composés chimiques appelés nucléotides, ont des surnoms tirés d'un alphabet à quatre lettres – A, T, C, G – et ils sont attachés les uns aux autres en une longue séquence linéaire qui code l'information génétique exactement comme la séquence de 1 et de 0 code l'information dans un fichier d'ordinateur. Nous pouvons prendre un fragment d'ADN d'une cellule humaine et l'insérer dans une bactérie, ou un fragment d'ADN bactérien pour l'insérer dans une cellule humaine, et l'information sera lue, interprétée et copiée avec succès. À l'aide de méthodes chimiques, les scientifiques ont appris comment lire la séquence complète des monomères dans n'importe quelle molécule d'ADN – sur une longueur de plusieurs millions de nucléotides – et donc comment déchiffrer l'information génétique contenue dans chaque organisme.

Toutes les cellules reproduisent leur information génétique par polymérisation à partir d'une matrice

Les mécanismes qui rendent la vie possible sont basés sur la structure de la molécule d'ADN double brin. Chaque monomère d'un simple brin d'ADN – c'est-à-dire chaque **nucléotide** – est constitué de deux parties : un sucre (désoxyribose) auquel est attaché un groupement phosphate, et une *base* qui peut être soit l'adénine (A), soit la guanine (G), soit la cytosine (C), soit la thymine (T) (**Figure 1-2**). Chaque sucre est relié au suivant par le groupement phosphate, créant ainsi une chaîne de polymères dont le squelette est formé de répétitions sucre-phosphate, avec une série de bases qui en dépassent. Le polymère d'ADN s'allonge par addition des monomères à une extrémité. S'il s'agit d'un brin unique isolé, ces monomères peuvent, en principe, s'ajouter dans n'importe quel ordre parce que chacun s'attache au suivant de la même manière, par la partie de la molécule qui leur est commune. Mais dans une cellule vivante, l'ADN n'est pas synthétisé sous forme d'un brin unique isolé, mais sur une matrice formée d'un brin d'ADN préexistant. Les bases faisant saillie sur le

Figure 1-2 L'ADN et ses unités de construction. (A) L'ADN est composé de sous-unités simples, appelées nucléotides, chacune composée d'une molécule de sucre-phosphate sur laquelle est lié un groupement latéral azoté, ou base. Les bases sont de quatre types (adénine, guanine, cytosine et thymine) et correspondent à quatre nucléotides distincts désignés par A, G, C et T. (B) Un brin unique d'ADN est composé de nucléotides reliés par des liaisons sucre-phosphate. Remarquez que chaque unité sucre-phosphate est asymétrique, ce qui donne au squelette du brin d'ADN une polarité définie ou sens. Cette polarité guide le processus moléculaire qui interprète et copie les informations de l'ADN dans la cellule; l'information est toujours « lue » dans le même sens, comme un texte écrit en français est toujours lu de gauche à droite. (C) Par une polymérisation à l'aide d'une matrice, la séquence des nucléotides d'un brin d'ADN existant dicte la séquence dans laquelle les nucléotides sont reliés dans le nouveau brin d'ADN; T d'un brin s'apparie avec A de l'autre et G d'un brin s'apparie avec C de l'autre. Le nouveau brin possède une séquence de nucléotides complémentaire à celle de l'ancien brin et son squelette a un sens opposé : correspondant à GTAA... du brin d'origine, il a...TTAC. (D) Une molécule d'ADN normale est composée de deux brins complémentaires. Les nucléotides à l'intérieur de chaque brin sont reliés par une liaison chimique forte (ou covalente); les nucléotides complémentaires sur les brins opposés sont maintenus ensemble par des liaisons hydrogène, plus faibles. (E) Les deux brins s'enroulent l'un sur l'autre pour former une double hélice – structure robuste qui peut recevoir n'importe quelle séquence de nucléotides sans que sa structure de base soit altérée.

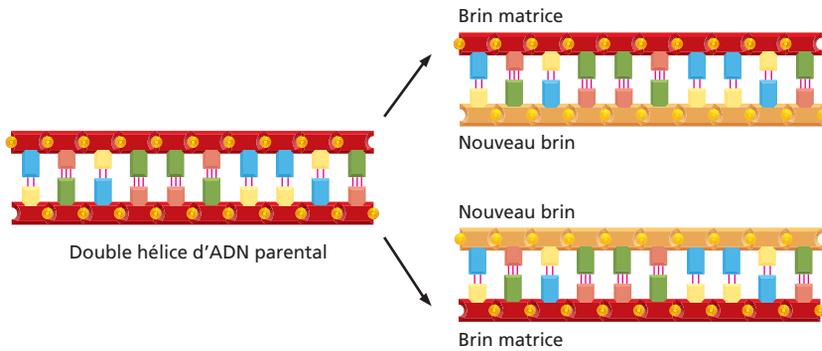


Figure 1-3 Duplication de l'information génétique par réplication de l'ADN. Au cours de ce processus, les deux brins de la double hélice d'ADN sont séparés et chacun sert de matrice à la synthèse d'un nouveau brin complémentaire.

brin existant se fixent aux bases du brin en cours de synthèse, selon une règle stricte définie par les structures complémentaires des bases : A se fixe à T, et C se fixe à G. Cet appariement des bases maintient les nouveaux monomères en place et contrôle le choix du monomère suivant à fixer sur le brin en croissance. De cette manière se crée une structure double brin formée de deux séquences exactement complémentaires de A, C, T et G. Les deux brins s'enroulent l'un autour de l'autre, formant une double hélice d'ADN (Figure 1-2E).

Les liaisons entre les bases appariées sont faibles en comparaison des liaisons sucre-phosphate, et cela permet aux deux brins d'ADN d'être séparés sans rupture de leurs squelettes. Chaque brin peut alors, de la même manière, servir de matrice pour la synthèse d'un nouveau brin d'ADN qui lui est complémentaire - c'est-à-dire une nouvelle copie de l'information génétique (Figure 1-3). Dans les différents types cellulaires, ce processus de **réplication de l'ADN** se produit à des vitesses différentes, avec des contrôles différents pour le démarrer ou l'arrêter, et des molécules auxiliaires différentes pour le faciliter. Mais le phénomène de base est universel : l'ADN est le lieu de stockage de l'information héréditaire, et la *polymérisation à partir d'une matrice* est le mode de copie de cette information dans tout le monde vivant.

Toutes les cellules transcrivent des portions de leur information génétique en une même forme intermédiaire : l'ARN

Pour remplir sa fonction de porteur d'information, l'ADN doit faire plus que juste se recopier. Il doit aussi *exprimer* cette information, c'est-à-dire permettre à cette information de guider la synthèse d'autres molécules dans la cellule. Cette expression se produit dans tous les organismes vivants grâce au même mécanisme, conduisant d'abord et avant tout à la synthèse de deux autres classes de polymères clés : les ARN et les protéines. Ce processus (qui sera étudié en détail dans les Chapitres 6 et 7) commence par une polymérisation à partir d'une matrice, étape appelée **transcription**, dans laquelle des segments de la séquence d'ADN sont utilisés comme matrice pour la synthèse de molécules plus courtes d'un polymère très proche, l'**acide ribonucléique** ou ARN. Ensuite, au cours de l'étape plus complexe de la **traduction**, beaucoup de ces molécules d'ARN dirigent la synthèse d'une classe de polymères totalement différents sur le plan chimique - les *protéines* (Figure 1-4).

Dans l'ARN, le squelette est formé d'un sucre légèrement différent de celui de l'ADN - le ribose au lieu du désoxyribose - et une des quatre bases est légèrement différente - l'uracile (U) au lieu de la thymine (T). Mais les trois autres bases - A, C et G - sont les mêmes et les quatre bases s'apparient avec leurs équivalents complémentaires de l'ADN - A, U, C et G de l'ARN avec T, A, G et C de l'ADN. Pendant la transcription, les monomères d'ARN sont alignés sur un brin matrice d'ADN et choisis en vue de la polymérisation, exactement comme les monomères d'ADN sont choisis au cours de la réplication. Le résultat est un polymère dont la séquence en nucléotides reproduit fidèlement une partie de l'information génétique de la cellule, bien qu'elle soit écrite dans un alphabet légèrement différent, fait de monomères d'ARN plutôt que de monomères d'ADN.

Le même segment d'ADN peut être utilisé de façon répétée pour guider la synthèse de nombreuses molécules d'ARN identiques. Ainsi, alors que les archives cellulaires de l'information génétique sous forme d'ADN sont fixées et sacro-saintes, ces *transcrits*

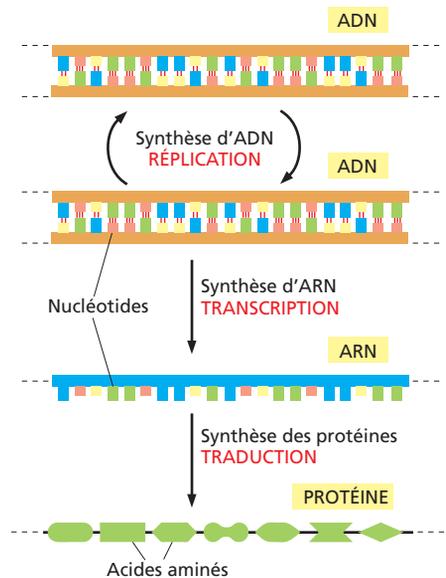


Figure 1-4 De l'ADN aux protéines. L'information génétique est lue et mise en service grâce à un processus en deux étapes. Tout d'abord, lors de la *transcription*, des segments de la séquence d'ADN sont utilisés pour guider la synthèse de molécules d'ARN. Puis, lors de la *traduction*, les molécules d'ARN sont utilisées pour guider la synthèse de molécules de protéines.

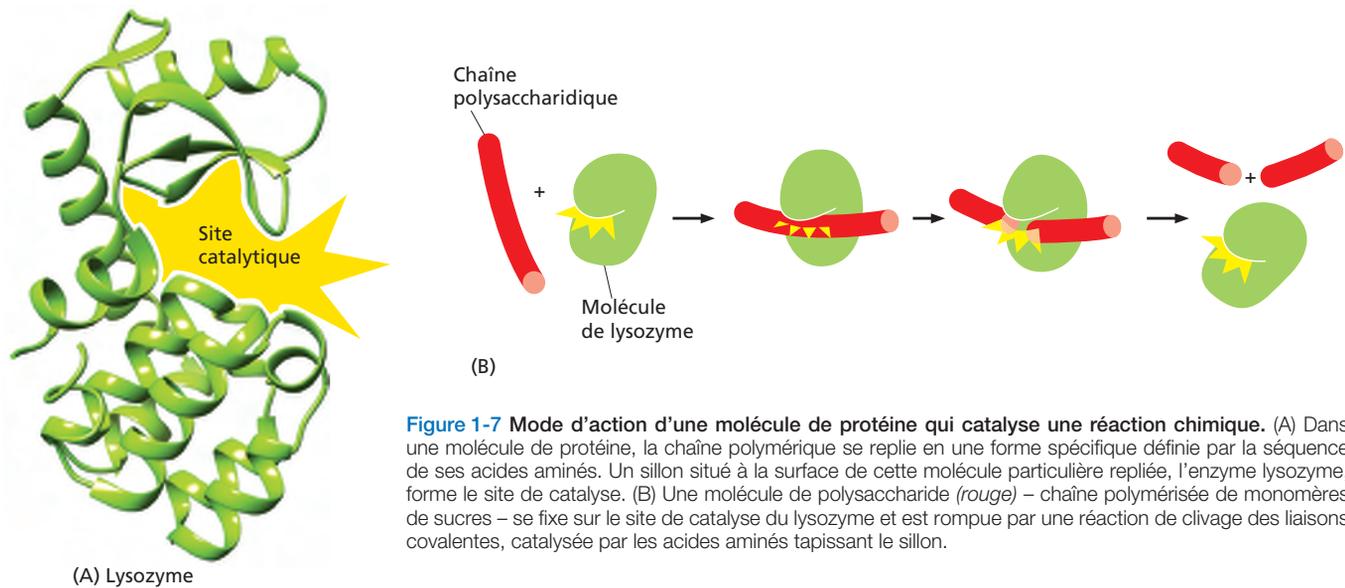


Figure 1-7 Mode d'action d'une molécule de protéine qui catalyse une réaction chimique. (A) Dans une molécule de protéine, la chaîne polymérique se replie en une forme spécifique définie par la séquence de ses acides aminés. Un sillon situé à la surface de cette molécule particulière repliée, l'enzyme lysozyme, forme le site de catalyse. (B) Une molécule de polysaccharide (rouge) – chaîne polymérisée de monomères de sucres – se fixe sur le site de catalyse du lysozyme et est rompue par une réaction de clivage des liaisons covalentes, catalysée par les acides aminés tapissant le sillon.

Les monomères des protéines, les **acides aminés**, sont très différents de ceux de l'ADN ou de l'ARN, et il y en a 20 au lieu de 4. Chaque acide aminé est construit autour de la même structure centrale, par l'intermédiaire de laquelle il peut se lier d'une façon standard à n'importe quel autre acide aminé du répertoire; des groupements latéraux attachés à cette structure centrale donnent à chaque acide aminé des propriétés chimiques particulières. Chacune des molécules de protéine est un **polypeptide**, créé en joignant ses acides aminés dans une séquence particulière. Grâce à des milliards d'années d'évolution, cette séquence a été sélectionnée pour donner une fonction utile à la protéine. Ainsi, par pliage en une forme tridimensionnelle précise possédant des sites réactifs à sa surface (Figure 1-7A), ces polymères d'acides aminés peuvent se fixer avec une spécificité élevée à d'autres molécules et peuvent agir en tant qu'**enzymes** pour catalyser des réactions qui établissent, ou rompent, des liaisons covalentes. De cette manière, ils dirigent la grande majorité des processus cellulaires (Figure 1-7B).

Les protéines ont aussi beaucoup d'autres fonctions – maintenir des structures, générer des mouvements, percevoir des signaux et bien d'autres – chaque molécule de protéine assurant une fonction spécifique selon sa propre séquence en acides aminés, qui est définie génétiquement. Les protéines sont donc avant tout les molécules principales qui mettent en action l'information génétique de la cellule.

Ainsi, les polynucléotides spécifient la séquence en acides aminés des protéines. Les protéines, à leur tour, catalysent de nombreuses réactions chimiques, entre autres celles qui vont conduire à la synthèse de nouvelles molécules d'ADN. D'un point de vue plus fondamental, une cellule vivante est une collection de catalyseurs doués d'auto-réplication qui consomme des aliments et les transforme pour obtenir les blocs constitutifs et l'énergie nécessaires à la fabrication d'autres catalyseurs, et qui élimine les matériaux inutilisés sous forme de déchets (Figure 1-8A). Une boucle de rétroaction située entre les protéines et les polynucléotides est à la base du comportement autocatalytique et autoreproductif des organismes vivants (Figure 1-8 B).

Toutes les cellules traduisent l'ARN en protéines de la même manière

La façon dont l'information contenue dans l'ADN spécifie la production de protéines était un mystère complet dans les années 1950 lorsque la structure double brin de l'ADN s'était révélée pour la première fois être la base de l'hérédité. Mais dans les années qui suivirent, les scientifiques élucidèrent les élégants mécanismes à la base de la transmission de l'information génétique. La traduction de l'information génétique de l'alphabet à 4 lettres des polynucléotides vers celui à 20 lettres des protéines est un processus complexe. Par certains aspects, les règles de cette traduction semblent claires et rationnelles, mais par d'autres elles semblent étrangement arbitraires, du fait qu'elles sont (à quelques exceptions mineures près) identiques dans tous les organismes vivants. On suppose que ces caractères arbitraires sont les reflets gelés d'accidents survenus dans les premiers temps de l'histoire de la vie. Ils ont pour origine des propriétés acquises par hasard par les tout premiers organismes, propriétés qui ont été transmises héréditairement et se sont enfouies si profondément dans la constitution de toutes les cellules vivantes, qu'elles n'ont pas pu être modifiées sans provoquer des effets désastreux.

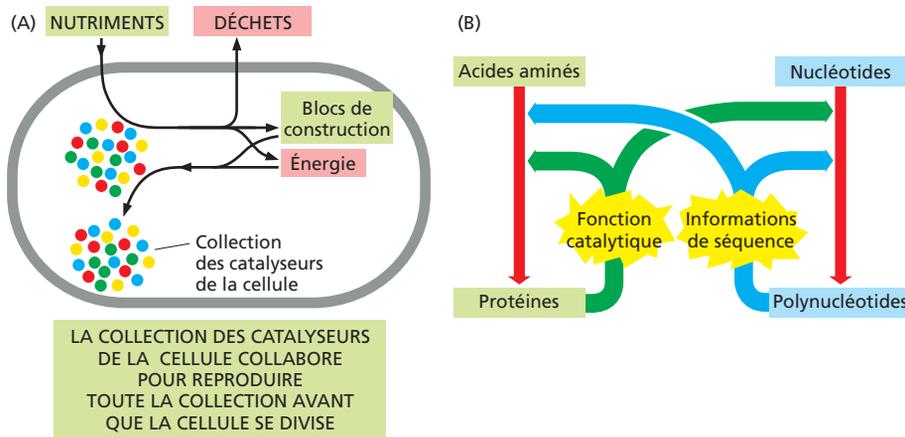


Figure 1-8 La vie est un processus auto-catalytique. (A) La cellule vue comme une collection de catalyseurs capables d'auto-réplication. (B) Les polynucléotides (les acides nucléiques ADN et ARN, qui sont des polymères de nucléotides) fournissent l'information de leurs séquences, tandis que les protéines (polymères d'acides aminés) assurent la plupart des fonctions catalytiques qui dirigent – grâce à un ensemble complexe de réactions chimiques – la synthèse d'autres polynucléotides et protéines de mêmes types.

Il se trouve que l'information contenue dans la séquence d'une molécule d'ARN messager est lue par groupes de trois nucléotides : chaque triplet de nucléotides, ou *codon*, code de façon spécifique un unique acide aminé de la protéine correspondante. Étant donné que le nombre de triplets distincts qui peuvent être formés à partir de quatre nucléotides est 4^3 , il existe 64 codons possibles, tous présents dans la nature. Cependant, il n'y a que 20 acides aminés naturels. Cela veut dire qu'il y a nécessairement de nombreux cas dans lesquels plusieurs codons correspondent au même aminoacide. Ce *code génétique* est lu par une classe spéciale de petites molécules d'ARN, les *ARN de transfert (ARNt)*. Chaque type d'ARNt se fixe par une de ses extrémités à un acide aminé spécifique, et présente à l'autre extrémité une séquence spécifique de trois nucléotides – un *anticodon* – qui lui permet de reconnaître, par appariement de bases, un codon particulier ou un ensemble de codons dans l'ARNm. La chimie complexe qui permet à ces ARNt de traduire une séquence spécifique de nucléotides A, C, G et U, présente dans une molécule d'ARNm, en une séquence spécifique d'acides aminés formant une molécule de protéine, a lieu sur le *ribosome*, une grande machine multimoléculaire composée de protéines et d'*ARN ribosomique*. Tous ces processus sont décrits en détail dans le Chapitre 6.

Chaque protéine est codée par un gène spécifique

Les molécules d'ADN sont en général très grandes et contiennent les spécifications de milliers de protéines. Des séquences spéciales dans l'ADN servent de ponctuation, définissant où l'information codant chaque protéine commence et se termine. Des segments individuels d'une longue séquence d'ADN sont transcrits en molécules d'ARNm distinctes, codant des protéines différentes. Chacun de ces segments d'ADN représente un *gène*. Une complication vient du fait que les molécules d'ARNm, transcrites à partir d'un même segment d'ADN, peuvent subir des processus de maturation différents, pour donner naissance à différentes versions d'une protéine, en particulier dans les cellules les plus complexes, comme celles des végétaux ou des animaux. En outre, certains segments d'ADN – un plus petit nombre – sont transcrits en molécules d'ARN qui ne sont pas traduites, mais qui ont des fonctions catalytiques, régulatrices ou structurales ; ces segments d'ADN sont également considérés comme des gènes. On définit donc un gène comme le segment d'une séquence d'ADN correspondant à une seule protéine ou à un groupe de versions de cette protéine, ou à une seule molécule d'ARN catalytique, régulatrice ou structurale.

Dans toutes les cellules, l'*expression* de chaque gène est régulée : au lieu de produire tout son répertoire de protéines possibles, en permanence, et à haut débit, la cellule ajuste la vitesse de transcription et de traduction des différents gènes de façon indépendante, en fonction de ses besoins. Des segments d'ADN *régulateurs* sont dispersés parmi les segments qui codent la protéine, et ces régions non codantes se lient à des molécules de protéines particulières qui contrôlent la vitesse de transcription locale. La quantité et l'organisation de ces ADN régulateurs varient énormément d'une classe d'organismes à l'autre, mais la stratégie de base est universelle. De cette manière, le *génom*e de la cellule – c'est-à-dire la totalité de son information génétique représentée par sa séquence ADN complète – dicte non seulement la nature des protéines de la cellule, mais aussi où et quand ces protéines doivent être produites.

La vie a besoin d'énergie libre

Une cellule vivante est un système chimique dynamique, qui fonctionne très loin de l'équilibre chimique. Pour qu'une cellule puisse grandir ou produire une nouvelle cellule à son image, elle doit prélever dans son environnement de l'énergie libre et des matériaux bruts pour activer les réactions de synthèse nécessaires. Cette consommation d'énergie libre est fondamentale pour la vie. Quand elle s'arrête, la cellule décline vers l'équilibre chimique et meurt rapidement.

L'information génétique est également fondamentale pour la vie, et de l'énergie libre est nécessaire pour la propagation de cette information. Par exemple, le choix d'un bit d'information – c'est-à-dire un choix « oui » ou « non » entre deux alternatives également probables – coûte une quantité définie d'énergie libre qui peut être calculée. Pour établir cette relation quantitative, un raisonnement approfondi est nécessaire, ainsi qu'une définition précise du terme « énergie libre » que nous expliquerons dans le Chapitre 2. Mais l'idée peut facilement être comprise intuitivement.

Imaginez les molécules dans une cellule comme un essaim d'objets, possédant chacun une énergie thermique, se déplaçant violemment au hasard, et se cognant les uns contre les autres. Pour fournir une information génétique précise – par exemple sous la forme d'une séquence ADN – les molécules de cette foule sauvage doivent être capturées, organisées en un ordre particulier défini par une matrice préexistante, et reliées de façon définie. Les liaisons qui maintiennent les molécules à une place appropriée sur la matrice et qui les relient entre elles doivent être assez fortes pour résister aux effets désorganisateur de l'agitation thermique. Le processus est rendu possible par la consommation d'énergie libre, nécessaire pour assurer la formation de liaisons adéquates et robustes. Dans le cas le plus simple, les molécules peuvent être comparées à des pièges à ressort, prêts à se refermer sur un état plus stable, d'énergie inférieure, en présence de la proie appropriée; quand elles s'accrochent ensemble en créant des liaisons, leur énergie disponible, c'est-à-dire leur énergie libre, comme l'énergie du ressort dans un piège, est libérée et dissipée sous forme de chaleur. Dans une cellule, les processus chimiques qui sous-tendent le transfert d'information sont plus complexes, mais le principe de base est le même : il faut dépenser de l'énergie libre pour créer de l'ordre.

Pour reproduire fidèlement son information génétique, et pour produire effectivement toutes les molécules complexes spécifiées par cette information, la cellule a donc besoin d'énergie libre et doit l'importer, d'une manière ou d'une autre, à partir de son environnement. Comme nous le verrons dans le Chapitre 2, l'énergie libre requise par les cellules animales provient des liaisons chimiques présentes dans les molécules des aliments que les animaux mangent, tandis que les végétaux obtiennent leur énergie libre de la lumière du soleil.

Toutes les cellules sont des usines biochimiques qui utilisent les mêmes unités de construction moléculaires

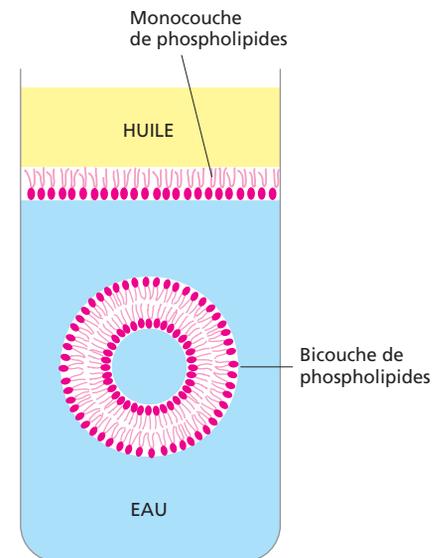
Parce que toutes les cellules produisent de l'ADN, de l'ARN et des protéines, toutes les cellules doivent contenir et manipuler des collections semblables de petites molécules, comme des sucres simples, des nucléotides et des acides aminés, et d'autres substances universellement requises. Toutes les cellules, par exemple, requièrent le nucléotide phosphorylé *ATP* (*adénosine triphosphate*), non seulement en tant que bloc de construction pour la synthèse de l'ARN et de l'ADN, mais aussi en tant que support et transporteur de l'énergie libre qui est nécessaire pour entraîner un grand nombre de réactions chimiques dans la cellule.

Bien que toutes les cellules fonctionnent comme des usines biochimiques d'un type globalement similaire, de nombreux détails de leurs transactions entre petites molécules sont différents. Certains organismes, comme les végétaux, n'ont besoin que des nutriments les plus simples et captent l'énergie de la lumière du soleil pour produire toutes leurs petites molécules organiques. D'autres organismes, comme les animaux, se nourrissent d'organismes vivants et doivent trouver beaucoup de leurs petites molécules organiques toutes faites. Nous reviendrons sur ce point plus tard.

Toutes les cellules sont entourées d'une membrane plasmique à travers laquelle doivent passer les nutriments et les déchets

Une autre caractéristique universelle est que chaque cellule est entourée par une membrane – la **membrane plasmique**. Cette enveloppe agit comme une barrière sélective permettant à la cellule de concentrer les nutriments collectés dans

Figure 1-9 Formation d'une membrane par des molécules de phospholipides amphipathiques. Les phospholipides ont une tête hydrophile (aimant l'eau, phosphate) et une queue hydrophobe (évitant l'eau, chaîne hydrocarbonée). À l'interface entre l'huile et l'eau, ils s'organisent en un seul feuillet, les groupes de tête face à l'eau et les groupes de queue face à l'huile. Mais lorsqu'ils sont immergés dans l'eau, ils s'agrègent pour former une double couche enfermant un compartiment aqueux, comme indiqué.



l'environnement et de retenir les produits qu'elle synthétise pour son usage propre, et aussi d'excréter les déchets. Sans membrane plasmique, la cellule ne pourrait pas maintenir son intégrité en tant que système chimique coordonné.

Les molécules qui forment la membrane ont une propriété physicochimique simple : elles sont *amphipathiques* – c'est-à-dire qu'elles comportent une partie hydrophobe (insoluble dans l'eau) et une partie hydrophile (soluble dans l'eau). Ces molécules forment spontanément des agrégats quand elles sont mises dans l'eau, plaçant leurs parties hydrophobes le plus possible au contact les unes des autres pour les protéger de l'eau, tandis que leurs parties hydrophiles restent exposées. Les molécules amphipathiques qui ont une forme appropriée, telles que les molécules de phospholipides qui forment la plus grande partie de la membrane plasmique, s'agrègent spontanément dans l'eau en créant une *bicouche* qui forme de petites vésicules fermées (**Figure 1-9**). Le phénomène peut être reproduit dans un tube à essai simplement en mélangeant des phospholipides et de l'eau ; dans des conditions appropriées, des gouttelettes se forment, dont le contenu aqueux est isolé du milieu extérieur.

Bien que les détails chimiques varient, les queues hydrophobes des principales molécules des membranes de toutes les cellules sont des polymères hydrocarbonés ($-\text{CH}_2-\text{CH}_2-\text{CH}_2-$), et leur assemblage spontané pour former une vésicule à double couche n'est qu'un des nombreux exemples d'un important principe général : les cellules produisent des molécules dont les propriétés chimiques les poussent à *s'assembler* spontanément pour former les structures dont la cellule a besoin.

La limite de la cellule ne peut pas être totalement imperméable. Pour grandir et se reproduire, une cellule doit être capable d'importer des matériaux bruts et d'exporter ses déchets à travers la membrane plasmique. Toutes les cellules présentent donc des protéines spécialisées, enchâssées dans leur membrane, qui transportent des molécules spécifiques d'un côté à l'autre de cette membrane. Certaines de ces *protéines de transport membranaire*, comme certaines des protéines qui catalysent les principales réactions des petites molécules dans la cellule, ont été si bien conservées au cours de l'évolution que nous pouvons reconnaître un air de famille entre elles, en comparant différents organismes vivants même s'ils appartiennent aux groupes les plus distants.

Les protéines de transport de la membrane déterminent en grande partie quelles molécules vont entrer dans la cellule, et les protéines catalytiques à l'intérieur de la cellule déterminent les réactions chimiques que ces molécules vont subir. Ainsi, en définissant les protéines que les cellules doivent produire, l'information génétique enregistrée dans la séquence de l'ADN dicte toute la chimie de la cellule ; et non seulement sa chimie, mais aussi sa forme et son comportement qui sont essentiellement construits et contrôlés par les protéines de la cellule.

Une cellule vivante peut exister avec moins de 500 gènes

Les principes de base du transfert de l'information biologique sont assez simples, mais quelle est la complexité des vraies cellules vivantes ? Quels sont, en particulier, ses besoins minimaux ? Nous pouvons obtenir une indication grossière en examinant une espèce dont le génome est le plus petit des génomes cellulaires connus – la bactérie *Mycoplasma genitalium* (**Figure 1-10**). Cet organisme est un parasite des mammifères et il trouve, dans son environnement, de nombreuses petites molécules prêtes à l'emploi. Il doit néanmoins synthétiser les grosses molécules – ADN, ARN et protéines – nécessaires au processus de base de l'hérédité. Il a environ 530 gènes, dont 400 sont essentiels. Son génome de 580 070 paires de bases représente 145 018 octets d'information, à peu près ce qu'il faudrait pour enregistrer le texte d'un chapitre de ce livre. La biologie cellulaire est peut-être compliquée, mais pas d'une manière inabordable.

Il faut probablement au moins 300 gènes pour qu'une cellule soit viable dans l'environnement d'aujourd'hui, bien que l'ensemble commun à toutes les espèces vivantes ne contienne que 60 gènes.

Résumé

La cellule individuelle est l'unité de matière vivante minimale capable d'autoreproduction, et composée d'une collection de catalyseurs se répliquant de façon autonome. Le principe de base permettant cette reproduction est la transmission de l'information génétique aux cellules de la descendance. Toutes les cellules, sur notre planète, stockent leur information génétique sous la même forme chimique – l'ADN double brin. La cellule reproduit cette information en séparant les brins d'ADN appariés et en les utilisant comme matrice pour la polymérisation d'un nouveau brin d'ADN dont la séquence nucléotidique lui sera complémentaire. La même stratégie de polymérisation sur une matrice est utilisée pour transcrire des portions de l'information depuis l'ADN vers les molécules d'un polymère très proche, l'ARN. Ces molécules d'ARN vont ensuite diriger la synthèse des molécules de protéines par un procédé plus complexe appelé traduction, qui fait intervenir une grande machine multimoléculaire, le ribosome. Les protéines sont les principaux catalyseurs de presque toutes les réactions chimiques de la cellule; parmi leurs autres fonctions, elles assurent l'importation ou l'exportation sélective de petites molécules à travers la membrane plasmique qui forme les limites de la cellule. La fonction spécifique de chaque protéine dépend de sa séquence en acides aminés, laquelle est définie par la séquence en nucléotides d'un segment correspondant de l'ADN – le gène codant cette protéine. Ainsi, le génome de la cellule détermine la chimie de celle-ci; et cette chimie est fondamentalement la même dans toutes les cellules vivantes, car elle doit assurer la synthèse d'ADN, d'ARN et de protéines. Les cellules les plus simples connues peuvent survivre avec environ 400 gènes.

LA DIVERSITÉ DES GÉNOMES ET L'ARBRE DE LA VIE

Le succès des organismes vivants, basé sur l'ADN, l'ARN et les protéines, a été spectaculaire. La vie a peuplé les océans, couvert la terre ferme, infiltré la croûte terrestre et modelé la surface de notre planète. Notre atmosphère riche en oxygène, les dépôts de charbon et de pétrole, les couches de mines de fer, les falaises de craie, de calcaire ou de marbre, sont tous des produits directs ou indirects de l'activité biologique passée sur la Terre.

Les organismes vivants ne sont pas confinés au royaume familier et tempéré des terres, de l'eau et de la lumière solaire, habité par les végétaux et les animaux qui les mangent. On les trouve aussi dans les fonds obscurs des océans, dans la boue chaude des volcans, dans des mares situées sous la surface gelée de l'Antarctique, et enterrés à des kilomètres de profondeur dans la croûte terrestre. Les créatures qui vivent dans ces environnements extrêmes ne nous sont en général pas familières, non seulement parce qu'elles sont inaccessibles, mais aussi parce qu'elles sont la plupart du temps microscopiques. Dans des habitats plus accueillants aussi, la plupart des organismes sont trop petits pour que nous puissions les voir sans équipement spécial : ils passent plutôt inaperçus, sauf s'ils sont la cause de maladie ou... pourrissent la charpente de nos maisons. Pourtant, les microorganismes constituent la plus grande partie de la matière vivante de notre planète. Ce n'est que récemment, grâce aux nouvelles méthodes d'analyse moléculaire et en particulier l'analyse des séquences d'ADN, que nous avons commencé à nous faire une image de la vie sur la Terre qui ne soit pas considérablement déformée par notre perspective de gros animaux vivant sur la terre sèche.

Dans cette partie, nous allons examiner la diversité des organismes et les relations qu'ils ont entre eux. Comme l'information génétique de tous les organismes est écrite dans la langue universelle des séquences d'ADN et que ces séquences, pour n'importe quel organisme, peuvent être obtenues facilement par des techniques biochimiques standard, il est maintenant possible de caractériser, cataloguer et comparer n'importe quel ensemble d'organismes vivants en se référant à ces séquences. À partir de ces comparaisons, nous pouvons estimer la place de chaque organisme dans l'arbre généalogique des espèces vivantes – l'« arbre de la vie ». Mais avant de décrire ce que cette approche va nous révéler, nous devons d'abord examiner comment les cellules se procurent, dans des environnements différents, la matière et l'énergie dont elles ont besoin pour survivre et proliférer, et de quelle manière certains organismes dépendent des autres pour leurs besoins chimiques de base.

Les cellules peuvent utiliser une grande variété de sources d'énergie libre

Les organismes vivants obtiennent leur énergie libre de différentes manières. Les uns, comme les animaux, les champignons et les nombreuses différentes bactéries qui vivent dans l'intestin humain, l'obtiennent en se nourrissant d'autres êtres vivants ou des

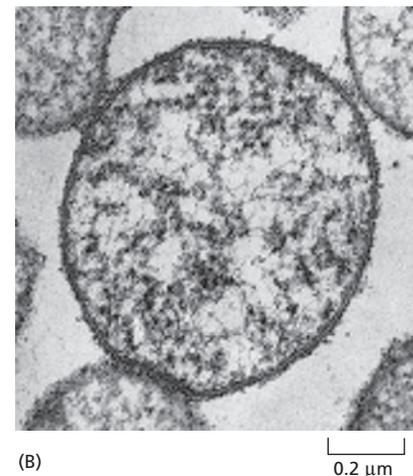
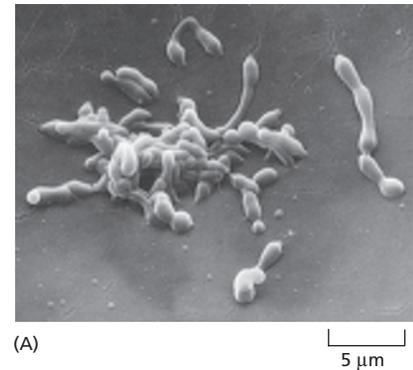


Figure 1-10 *Mycoplasma genitalium*. (A) Photographie en microscopie électronique à balayage qui montre la forme irrégulière de cette petite bactérie, due à l'absence de toute paroi rigide. (B) Coupe transversale (photographie en microscopie électronique à transmission) d'une cellule de *Mycoplasma*. Sur les 530 gènes de *Mycoplasma genitalium*, 43 codent les ARN de transfert, les ARN ribosomiques et d'autres ARN non messagers. Les fonctions des 339 gènes codant les protéines sont connues, ou peuvent être devinées; parmi celles-ci, 154 sont impliquées dans la réplication, la transcription et la traduction de l'ADN et à des processus voisins impliquant ADN, ARN et protéines; 98 dans les structures de la membrane ou de la surface de la cellule; 46 dans le transport des nutriments et d'autres molécules à travers la membrane; 71 dans la conversion énergétique, la synthèse et la dégradation des petites molécules; et 12 dans la régulation de la division cellulaire et d'autres processus. Remarquez que ces catégories se chevauchent partiellement, de sorte que certains gènes sont cités deux fois. (A, avec l'autorisation de S. Razin et al., *Infect. Immun.* 30 : 538-546, 1980. Avec l'autorisation de the American Society for Microbiology; B, dû à l'obligeance de Roger Cole, in *Medical Microbiology*, 4th ed. [S. Baron ed.]. Galveston : University of Texas Medical Branch, 1996.)

produits organiques que ceux-ci produisent; ces organismes sont appelés *organotrophes* (du grec *trophe* qui signifie «nourriture»). D'autres tirent directement leur énergie du monde non vivant. Ces convertisseurs d'énergie primaire se répartissent en deux catégories : ceux qui tirent leur énergie de la lumière solaire, et ceux qui la tirent des systèmes riches en énergie de composants inorganiques de l'environnement (des systèmes chimiques qui sont très loin de l'équilibre). Les premiers sont appelés *phototrophes* (se nourrissant de la lumière du soleil); les derniers sont appelés *lithotrophes* (se nourrissant des pierres). Les organismes organotrophes ne pourraient pas survivre sans ces convertisseurs d'énergie primaires, qui représentent la forme de vie la plus abondante.

Parmi les organismes phototrophes, on trouve de nombreuses bactéries, des algues et des végétaux dont nous – et pratiquement tous les autres êtres vivants que nous voyons autour de nous – dépendons. Les organismes phototrophes ont totalement modifié notre environnement chimique : l'oxygène de l'atmosphère terrestre est un sous-produit de leurs activités de biosynthèse.

Les organismes lithotrophes ne sont pas aussi évidents dans notre entourage parce qu'ils sont microscopiques et vivent pour la plupart dans des habitats peu fréquentés par l'homme : les profondeurs des océans, l'épaisseur de la croûte terrestre ou d'autres environnements inhospitaliers. Mais ils sont un constituant majeur du monde vivant et ils doivent être pris en compte pour toute considération sur l'histoire de la vie sur la Terre.

Certains lithotrophes tirent leur énergie de réactions *aérobies*, qui utilisent l'oxygène moléculaire de l'environnement; puisque l'O₂ atmosphérique est produit, en fin de compte, par les organismes vivants, ces lithotrophes aérobies se nourrissent en un sens de produits de la vie passée. Il y a, cependant, d'autres lithotrophes qui vivent dans des conditions anaérobies, dans des endroits où l'oxygène moléculaire est peu ou pas présent. Ce sont des circonstances similaires à celles qui existaient dans les premiers jours de la vie sur Terre, avant que l'oxygène ne se soit accumulé.

Les plus spectaculaires de ces sites sont les très chaudes *cheminées hydrothermiques* du plancher des océans Pacifique et Atlantique. Ils sont situés là où le plancher océanique s'étale pendant que de nouvelles parties de la croûte terrestre se forment par remontée progressive de matériel de l'intérieur de la Terre (Figure 1-11). L'eau de mer qui coule vers les fonds est chauffée et renvoyée vers le haut comme un geyser sous-marin, transportant avec elle un courant de produits chimiques provenant de la roche chaude du fond. Un mélange typique peut contenir H₂S, H₂, CO, Mn²⁺, Fe²⁺, Ni²⁺, CH₄, NH₄⁺ et des composés contenant du phosphore. Une population dense de microbes vit dans le voisinage de la cheminée, prospérant malgré ce régime austère

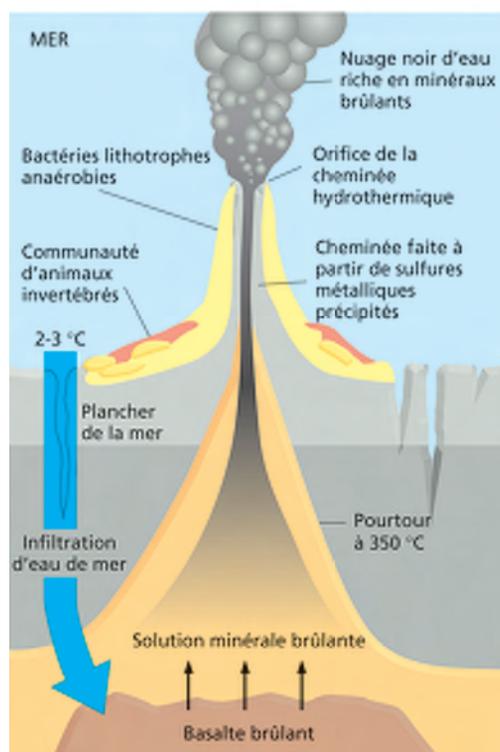


Figure 1-11 Géologie d'une cheminée hydrothermique chaude sur le plancher de l'océan. Comme indiqué, l'eau s'infiltré à travers la roche en fusion jaillissant de l'intérieur de la Terre. Elle est chauffée et entraînée vers la surface, transportant des minéraux arrachés à la roche brûlante. Un gradient de température se met en place, entre plus de 350 °C près du centre de la cheminée et 2-3 °C dans l'océan alentour. Les minéraux se condensent par perte d'eau lorsqu'ils se refroidissent, et forment une cheminée. Différentes classes d'organismes se multiplient aux différentes températures et vivent dans les différentes régions de la cheminée. Une cheminée typique peut mesurer quelques mètres de haut, crachant de l'eau chaude chargée de minéraux avec une vitesse de débit de 1-2 m/s.

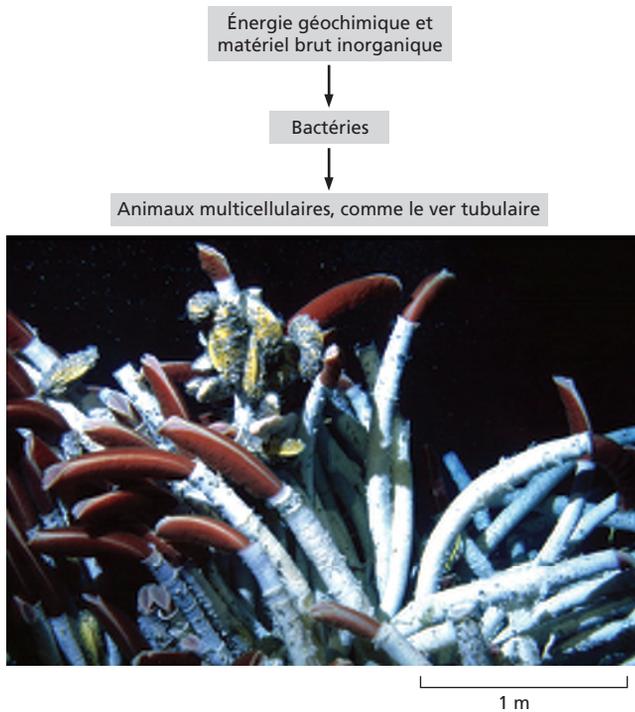


Figure 1-12 Organismes vivant à une profondeur de 2500 mètres à proximité d'un événement dans le fond de l'océan. Près de l'orifice de la cheminée, à des températures proches de 120 °C, vivent diverses espèces lithotrophes de bactéries et d'archaea (archéobactéries), directement approvisionnées par l'énergie géochimique. Un peu plus loin, là où la température est plus faible, vivent divers animaux invertébrés qui se nourrissent de ces microorganismes. Les plus remarquables sont ces vers tubulaires géants (2 mètres), *Riftia pachyptila*, qui, au lieu de se nourrir des cellules lithotrophes, vivent en symbiose avec elles; des organes spécialisés de ces vers abritent un très grand nombre de bactéries symbiotiques qui oxydent le soufre. Ces dernières exploitent l'énergie géothermique et fournissent de la nourriture à leur hôte, qui ne possède ni bouche, ni appareil digestif, ni anus. On pense que ce ver tubulaire a évolué à partir d'animaux plus conventionnels et s'est secondairement adapté à la vie au niveau des cheminées hydrothermiques. (Dû à l'obligeance de Monika Bright, University of Vienna, Austria.)

en recueillant l'énergie libre créée par les réactions chimiques se produisant entre les composés disponibles. D'autres organismes – praires, moules et vers marins géants – vivent de ces microbes présents à l'orifice de la cheminée, formant un véritable écosystème comparable à celui des végétaux et des animaux auquel nous appartenons, mais utilisant l'énergie géochimique plutôt que celle de la lumière (Figure 1-12).

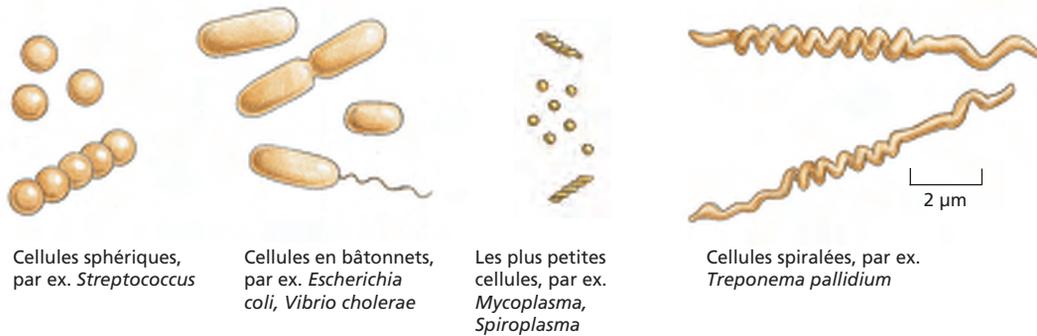
Certaines cellules fixent l'azote et le gaz carbonique pour les autres

Pour qu'une cellule vivante soit faite, il faut non seulement des matériaux, mais aussi de l'énergie libre. L'ADN, l'ARN et les protéines ne contiennent que six éléments : l'hydrogène, le carbone, l'azote, l'oxygène, le soufre et le phosphore. Ils sont tous présents en abondance dans l'environnement non vivant de la Terre, dans les roches, l'eau et l'atmosphère. Mais ils ne sont pas présents sous des formes chimiques qui permettent une incorporation facile dans des molécules biologiques. Le N_2 et le CO_2 atmosphériques, en particulier, sont extrêmement peu réactifs. Une grande quantité d'énergie libre est requise pour activer les réactions qui permettent à ces molécules inorganiques de se transformer en composés organiques nécessaires aux biosynthèses ultérieures – c'est-à-dire pour fixer l'azote et le gaz carbonique et rendre les atomes N et C utilisables par les organismes vivants. De nombreux types cellulaires vivants sont dépourvus de la machinerie biochimique permettant de réaliser cette fixation; ils comptent plutôt sur d'autres classes de cellules pour faire le travail à leur place. Nous, qui sommes des animaux, nous dépendons des végétaux pour nos besoins en carbone organique et en composés azotés. Quant aux végétaux, ils peuvent fixer le dioxyde de carbone de l'atmosphère, mais ils n'ont pas la capacité de fixer l'azote atmosphérique; ils dépendent en partie des bactéries fixatrices d'azote pour subvenir à leur besoin de composés azotés. Les végétaux de la famille des pois, par exemple, abritent dans les nodules de leurs racines des bactéries symbiotiques capables de fixer l'azote.

Les cellules vivantes sont donc très largement différentes, si on considère certains aspects parmi les plus élémentaires de leur biochimie. Il n'est donc pas surprenant que certaines cellules, dont les besoins et les capacités sont complémentaires, aient développé des associations étroites. Certaines de ces associations, comme nous le verrons plus loin, ont évolué au point que les partenaires ont perdu leur identité propre : ils se sont unis pour ne former qu'une seule cellule composite.

La plus grande diversité biochimique est trouvée parmi les cellules procaryotes

Grâce à un simple examen microscopique, on sait depuis longtemps classer les organismes vivants en deux groupes, sur la base de leur structure : les eucaryotes et les



Cellules sphériques, par ex. *Streptococcus*
 Cellules en bâtonnets, par ex. *Escherichia coli*, *Vibrio cholerae*
 Les plus petites cellules, par ex. *Mycoplasma*, *Spiroplasma*
 Cellules spiralées, par ex. *Treponema pallidum*

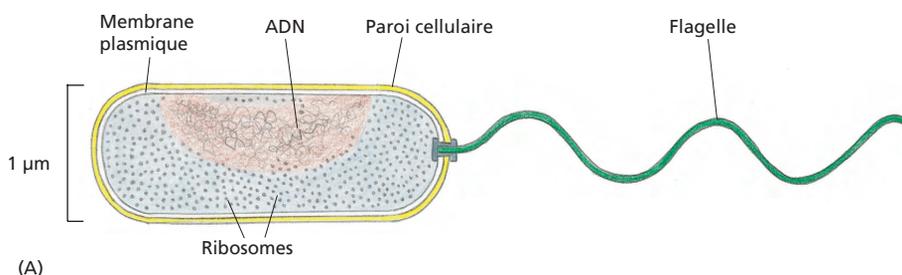
procaryotes. Les eucaryotes conservent leur ADN dans un compartiment intracellulaire distinct et entouré d'une membrane, le noyau. (Le terme vient des mots grecs *eu*, bien ou vrai, et *karyon*, noyau). Les procaryotes n'ont pas de compartiment nucléaire distinct pour loger leur ADN. Les végétaux, les champignons et les animaux sont des eucaryotes; les bactéries sont des procaryotes – de même que les archéobactéries, une classe particulière de procaryotes dont nous parlerons plus loin.

La plupart des cellules procaryotes sont petites et d'apparence simple (Figure 1-13), et vivent le plus souvent comme des individus indépendants, ou en communautés peu organisées, plutôt que sous forme d'organismes multicellulaires. Elles sont généralement rondes ou en forme de bâtonnets et leur dimension linéaire est de quelques micromètres. Elles ont souvent une épaisse enveloppe protectrice, appelée *paroi cellulaire*, sous laquelle une membrane plasmique renferme un unique compartiment cytoplasmique contenant l'ADN, l'ARN, les protéines et de nombreuses petites molécules nécessaires à leur vie. Au microscope électronique, l'intérieur de la cellule apparaît comme une matrice de texture variable sans aucune structure interne organisée (Figure 1-14).

Les cellules procaryotes vivent dans une très grande variété de niches écologiques et elles ont des propriétés chimiques étonnamment variables – bien plus que les cellules eucaryotes. Les espèces organotrophes peuvent se nourrir de pratiquement n'importe quel type de molécule organique, des sucres et des acides aminés aux hydrocarbures et au gaz méthane. Les espèces phototrophes (Figure 1-15) récoltent l'énergie de la lumière par différents procédés, certaines produisant de l'oxygène et d'autres non. Les espèces lithotrophes peuvent vivre avec un régime simple de nutriments inorganiques, prenant leur carbone directement du CO₂ et utilisant H₂S pour produire leur énergie (Figure 1-16) – ou même en utilisant H₂, ou Fe²⁺ ou du soufre ou n'importe quel composé chimique qu'elles trouvent dans leur environnement.

Figure 1-13 Formes et tailles de quelques bactéries. La plupart des bactéries sont petites, comme on le voit ici, mais il existe aussi des espèces géantes. Un exemple extrême (non représenté) est la bactérie en forme de cigare *Epulopiscium fishelsoni*, qui vit dans l'intestin d'un poisson chirurgien et peut mesurer jusqu'à 600 µm.

Figure 1-14 Structure d'une bactérie. (A) La bactérie *Vibrio cholerae*, montrant son organisation interne simple. Comme de nombreuses autres espèces, *Vibrio* présente un appendice hélicoïdal à une extrémité – un flagelle – qui tourne comme une hélice pour propulser la cellule. Il peut infecter l'intestin grêle de l'homme et provoquer le choléra; la diarrhée sévère qui accompagne cette maladie tue plus de 100 000 personnes par an. (B) Une photographie en microscopie électronique d'une coupe longitudinale d'*Escherichia coli* (*E. coli*), bactérie très largement étudiée. L'ADN de la cellule est concentré dans la région légèrement colorée. Faisant partie de notre flore intestinale normale, *E. coli* est apparenté à *Vibrio*, et il a de nombreux flagelles répartis sur sa surface qui ne sont pas visibles sur ce cliché. (B, dû à l'obligeance de E. Kellenberger.)



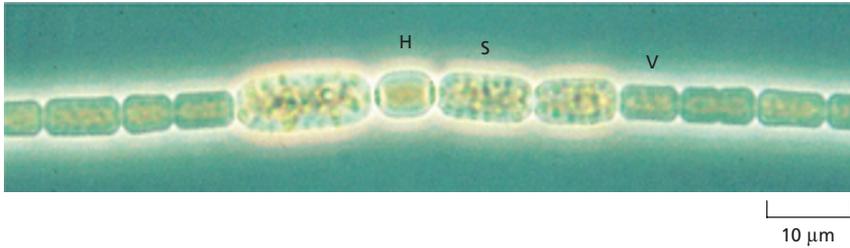


Figure 1-15 *Anabaena cylindrica*, bactérie phototrophe, vue au microscope optique. Les cellules de cette espèce forment de longs filaments multicellulaires. La plupart des cellules (désignées par V), effectuent la photosynthèse, alors que d'autres se sont spécialisées dans la fixation de l'azote (désignées par H), ou se développent en spores résistantes (désignées par S). (Dû à l'obligeance de Dave G. Adams.)

Une grande partie du monde des organismes microscopiques est encore inexplorée. Les méthodes traditionnelles de la bactériologie nous ont appris à connaître les espèces qui peuvent être isolées et cultivées en laboratoire. Mais l'analyse des séquences d'ADN de populations bactériennes et d'archéobactéries dans des échantillons provenant de leur habitat naturel – comme le sol, ou l'eau des océans, ou même la bouche humaine – nous a montré que la plupart de ces espèces ne pouvaient pas être cultivées par les techniques standard de laboratoire. D'après une estimation, 99 p. 100 des espèces de procaryotes restent encore à caractériser. On les a détectées seulement par leur ADN, et il n'a pas encore été possible de cultiver la grande majorité d'entre elles au laboratoire.

L'arbre de la vie comporte trois branches primitives : les bactéries, les archéobactéries (archaea) et les eucaryotes

La classification des êtres vivants était traditionnellement effectuée en fonction de leur aspect extérieur : nous voyons aisément qu'un poisson a des yeux, des mâchoires, une colonne vertébrale, un cerveau, etc., tout comme nous, à la différence d'un ver ; qu'un rosier est plus proche d'un pommier que d'une herbe. Comme Darwin l'a montré, nous pouvons facilement interpréter ces ressemblances en termes d'évolution à partir d'un ancêtre commun, et nous pouvons trouver les restes de beaucoup de ces ancêtres conservés sous forme de fossiles. Il a ainsi été possible de commencer à établir un arbre généalogique des êtres vivants, montrant les différentes lignées de descendance et les points de branchements qui correspondent au moment, au cours de l'histoire, où les ancêtres d'un groupe d'espèces sont devenus différents les uns des autres.

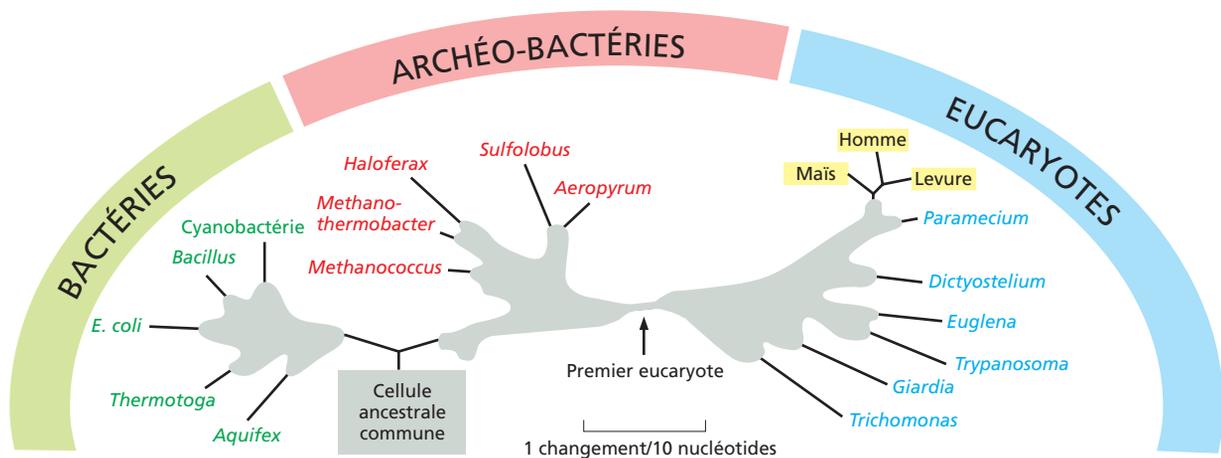
Toutefois, quand les différences entre les organismes deviennent très importantes, ces méthodes ne sont plus efficaces. Comment pouvons-nous décider si un champignon est plus proche d'une plante ou d'un animal ? Et quand on arrive aux procaryotes, la tâche est encore plus difficile : un bâtonnet ou une sphère microscopiques se ressemblent beaucoup. Les microbiologistes ont donc tenté de classer les procaryotes en fonction de leur biochimie et de leurs besoins nutritionnels. Mais cette approche présente aussi des pièges. Parmi l'extraordinaire variété des comportements biochimiques, il est difficile de savoir quelles sont les différences qui reflètent vraiment l'évolution.

Maintenant, l'analyse des génomes nous a apporté un outil plus simple, plus direct et beaucoup plus puissant pour étudier les relations entre les espèces au cours de l'évolution. La séquence complète d'ADN d'un organisme définit sa nature avec une précision presque parfaite et de manière exhaustive. De plus, cette définition est sous forme digitale – une suite de lettres – qui peut être entrée directement dans un ordinateur et comparée à l'information correspondante de n'importe quel autre organisme vivant. Comme l'ADN est sujet à des changements aléatoires qui s'accumulent sur de longues périodes de temps (comme nous allons bientôt le voir), le nombre de différences entre les séquences d'ADN de deux organismes donne une indication directe, objective et quantitative de la distance évolutive qui existe entre eux.

Cette approche a montré que les organismes qui étaient traditionnellement regroupés sous le terme de « bactéries » pouvaient être aussi divergents en termes d'évolution qu'un procaryote d'un eucaryote. On sait maintenant que les procaryotes comportent deux groupes distincts qui ont divergé très tôt dans l'histoire de la vie sur la Terre, avant même que les eucaryotes ne divergent en un groupe séparé. Ces deux groupes de procaryotes sont appelés **bactéries** (ou eubactéries) et **archéobactéries**. Des analyses génomiques détaillées ont récemment révélé que la première cellule eucaryote a été formée après qu'un type particulier d'une ancienne cellule Archaea



Figure 1-16 Une bactérie lithotrophe. *Beggiatoa*, qui vit dans un environnement soufré, obtient son énergie de l'oxydation d' H_2S et peut fixer du carbone même dans le noir. Remarquez les dépôts jaunes de soufre à l'intérieur de la cellule. (Dû à l'obligeance de Ralph W. Wolfe.)



a englouti une ancienne bactérie (voir Figure 12-3). On considère ainsi que le monde vivant d'aujourd'hui se compose de trois catégories ou *domaines* majeurs : les bactéries, les archéobactéries et les eucaryotes (Figure 1-17).

Les archéobactéries habitent souvent dans des environnements que nous, humains, évitons, comme les marécages, les égouts, les profondeurs de l'océan, les eaux salées et les sources acides chaudes, mais elles sont aussi largement répandues dans des environnements moins extrêmes et plus proches de nous, depuis la terre et les lacs jusqu'à l'estomac du bétail. Par leur apparence extérieure, elles ne sont pas faciles à distinguer des bactéries. Sur le plan moléculaire, les archéobactéries semblent plus proches des eucaryotes en ce qui concerne la transmission de l'information génétique (réplication, transcription, traduction), mais plus proches des bactéries en ce qui concerne le métabolisme et la conversion de l'énergie. Nous verrons plus loin comment cela peut s'expliquer.

Certains gènes évoluent rapidement ; d'autres sont très conservés

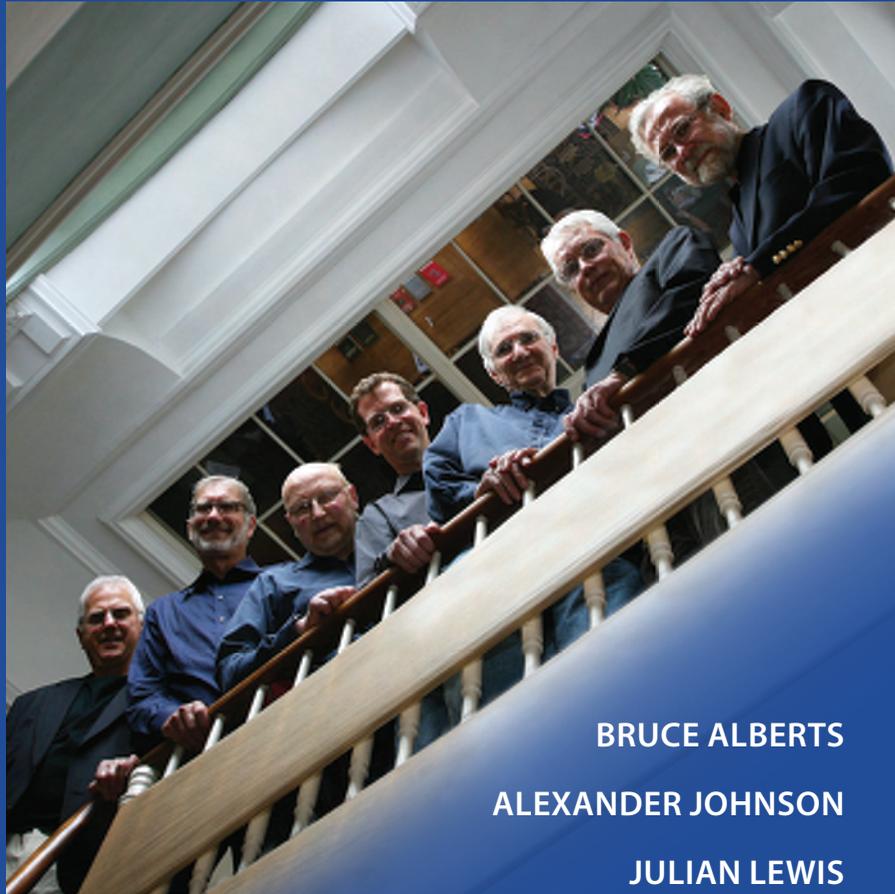
Au cours du stockage comme de la copie de l'information génétique, des accidents et des erreurs aléatoires surviennent, qui modifient la séquence nucléotidique – c'est-à-dire qui créent des **mutations**. Ainsi, quand une cellule se divise, il arrive souvent que les deux cellules filles ne soient pas identiques l'une à l'autre ou à la cellule parentale. Rarement, l'erreur peut provoquer une amélioration ; le plus souvent elle n'entraînera aucune différence pour l'avenir de la cellule. Mais dans de nombreux cas, l'erreur entraînera des dommages sérieux, par exemple en provoquant une rupture de la séquence codant une protéine clé. Les changements dus au premier type d'erreur vont tendre à être perpétués parce que la cellule modifiée a plus de chance de se reproduire. Les changements dus aux erreurs du deuxième type – les changements *sélectivement neutres* – seront perpétués ou non : dans la compétition pour des ressources limitées, seul le hasard décidera laquelle de la cellule initiale ou de sa cousine modifiée survivra. Mais les changements provoquant des dommages sérieux ne mènent nulle part : la cellule qui les a subis meurt sans laisser de descendance. C'est par la répétition de ce cycle sans fin d'erreurs et d'essais – de *mutation* et de *sélection naturelle* – que les organismes évoluent : leur définition génétique change, leur donnant de nouveaux moyens d'exploiter plus efficacement l'environnement pour survivre dans la compétition avec les autres et pour se reproduire avec succès.

Certaines parties du génome ont changé plus facilement que d'autres au cours de l'évolution. Un segment d'ADN qui ne code pas une protéine et qui ne joue pas un rôle régulateur particulier est libre de changer à une vitesse qui n'est limitée que par la fréquence des erreurs dues au hasard. Au contraire, un gène qui code une protéine essentielle et hautement spécialisée ou une molécule d'ARN, ne peut pas être modifié si facilement : si une erreur survient, les cellules défectueuses sont presque toujours éliminées. Ce genre de gène est donc *hautement conservé*. Grâce aux 3,5 milliards d'années ou plus d'histoire de l'évolution, de nombreuses caractéristiques du génome ont changé au-delà de toute possibilité d'identification, mais les gènes les mieux conservés sont restés parfaitement reconnaissables dans toutes les espèces vivantes.

Figure 1-17 Les trois divisions majeures (domaines) du monde vivant. Remarquez qu'à l'origine le mot *bactérie* était utilisé en référence aux procaryotes en général, mais plus récemment il a été redéfini pour se référer spécifiquement aux eubactéries. L'arbre montré sur cette figure est basé sur la comparaison, dans les différentes espèces, de la séquence de nucléotides d'une sous-unité d'ARN ribosomique (ARNr). La longueur des lignes du diagramme représente le nombre estimé de modifications qui se sont produites, dans chaque lignée, au cours de l'évolution (voir Figure 1-18). Les parties de l'arbre en grisé représentent les incertitudes des scientifiques sur le processus réel de divergence de ces espèces au cours de l'évolution ; les comparaisons des séquences de nucléotides ou d'acides aminés de molécules autres que l'ARNr, ainsi que d'autres arguments, peuvent conduire à des arbres un peu différents. Comme indiqué, le noyau de la cellule eucaryote est maintenant supposé avoir émergé d'une sous-branche des archaea, de sorte que, au début, l'arbre de la vie avait seulement deux branches – bactéries et archaea.

Biologie moléculaire de
LA CELLULE

Sixième édition



BRUCE ALBERTS

ALEXANDER JOHNSON

JULIAN LEWIS

DAVID MORGAN

MARTIN RAFF

KEITH ROBERTS

PETER WALTER

