



1. LES MOLÉCULES D'ADN stockent de l'information. Ce sont des enchaînements d'adénine, de thymine, de guanine et de cytosine, respectivement représentées par les lettres A, T, G et C. La molécule dessinée ici en couleurs, de séquence GCAGTCGGACTGGGC-TATGTCCGA, code la solution du problème du chemin hamiltonien représenté page suivante.

rie des nombres, par Gregory Chaitin, *Pour la Science*, septembre 1988).

Turing avait inventé un ordinateur abstrait que l'on nomme aujourd'hui «machine de Turing». Cette machine est composée de deux bandes, ainsi que d'un système nommé contrôleur à états finis, qui se déplace le long d'une bande pour y lire des données et, simultanément, se déplace le long de la deuxième bande pour y écrire d'autres données. Le contrôleur se programme à l'aide d'instructions simples, de sorte qu'il est aisé d'écrire un programme qui lit sur la bande d'entrée une chaîne composée de A, de T, de C et de G, et qui écrit la chaîne complémentaire sur la bande de sortie. La similitude avec l'ADN polymérase est patente.

La similarité est intéressante, parce que la machine de Turing est universelle, c'est-à-dire qu'elle peut être programmée pour calculer tout ce qui est calculable. Cette idée est la «thèse de Church». Elle me fit penser que l'on pourrait programmer une machine de Turing pour produire une chaîne complémentaire, pour factoriser des nombres ou pour jouer aux échecs. Je me dressai soudain dans mon

lit et fis remarquer à mon épouse : «Bon sang! Ces choses peuvent calculer.» Je ne dormis pas de la nuit, cherchant un moyen d'utiliser l'ADN pour résoudre des problèmes.

Je voulus d'abord concevoir un ordinateur à ADN analogue à une machine de Turing, où le contrôleur à états finis serait remplacé par une enzyme. Une idée identique avait déjà été proposée dix ans plus tôt par des chercheurs de la Société IBM (voir *Les limites physiques du calcul*, par Charles Bennett et Rolf Landauer, *Pour la Science*, septembre 1985). Malheureusement, si une enzyme (l'ADN polymérase) permet de créer des ADN complémentaires, où trouverait-on les molécules qui assureraient d'autres calculs importants, telle la factorisation des nombres?

Les biologistes moléculaires ont la chance inouïe de pouvoir utiliser le contenu des cellules. Nous sommes très loin de savoir créer *ex nihilo* des machines moléculaires aussi perfectionnées que l'ADN polymérase, mais trois ou quatre milliards d'années d'évolution ont engendré des cellules pleines de petites merveilles. La biotechnologie moderne emploie certaines de ces

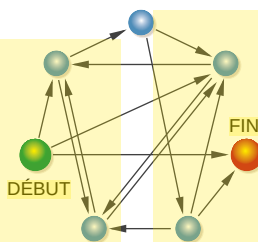
molécules, mais l'évolution n'a jamais produit de cellules qui jouent aux échecs. Aussi, pour construire des machines cellulaires aux capacités de calcul intéressantes, je devais utiliser les outils disponibles. Ces outils sont essentiellement les suivants :

1. L'appariement des bases. À chaque brin d'ADN correspond une molécule complémentaire. Dans une solution, lorsqu'un brin d'ADN rencontre l'ADN complémentaire, les deux brins s'apparient, c'est-à-dire qu'ils s'enroulent l'un autour de l'autre pour former une double hélice. Les brins ne réagissent pas chimiquement, mais se lient parce que des forces faibles, des «liaisons hydrogène», s'établissent entre les bases complémentaires. Quand un brin d'ADN rencontre un autre brin d'ADN qui ne lui est pas complémentaire ou qui ne possède pas de longues parties complémentaires, alors les deux brins ne se lient pas.

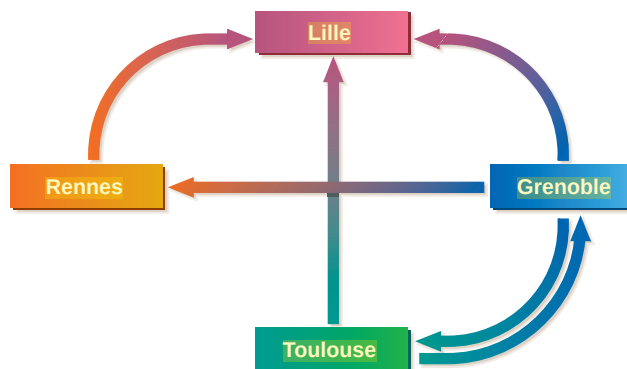
2. Des polymérases dupliquent l'information d'une molécule en créant une molécule complémentaire. Par exemple, l'ADN polymérase fabrique un brin d'ADN complémentaire à partir de n'importe quelle molé-

Le chemin hamiltonien codifié par des bases

Considérons des villes reliées par un certain nombre de vols sans escale (*en bas*). Dans l'exemple représenté, on peut aller de Grenoble à Lille, mais pas l'inverse. On cherche s'il existe un chemin qui part de Toulouse et se termine à Lille, en passant une fois seulement dans les autres villes. On attribue à chaque ville une séquence d'ADN (ACTTGCAG pour Toulouse) qui peut être considérée comme un prénom (ACTT) et d'un nom propre (GCAG). Les codes des vols peuvent alors être définis par la mise bout à bout du nom propre de la ville d'origine et du prénom de la ville de destination (*en bas, à droite*).



Les complémentaires des noms des villes s'obtiennent par le remplacement de C par G, de G par C, de A par T et de T par A. Pour ce problème particulier, un seul chemin hamiltonien relie Toulouse à Lille : il passe successivement par Grenoble et Rennes, dans cet ordre. Dans le calcul, ce chemin est représenté par GCAGTCGGACTGGGC-TATGTCCGA, une séquence d'ADN de 24 bases de longueur. Le problème représenté ci-contre, à gauche, a été résolu par un ordinateur à ADN : il fait intervenir 7 villes et 14 vols.



VILLE	ADN	COMPLÉMENTAIRE
TOULOUSE	ACTTGCAG	TGAACGTC
GRENOBLE	TGGACTG	AGCCTGAC
RENNES	GGCTATGT	CCGATACA
LILLE	CCGAGCAA	GGCTCGTT

VOL	CODE ADN DU VOL
TOULOUSE - GRENOBLE	GCAGTCGG
TOULOUSE - LILLE	GCAGCCGA
GRENOBLE - RENNES	ACTGGGCT
GRENOBLE - LILLE	ACTGCCGA
GRENOBLE - TOULOUSE	ACTGACTT
RENNES - LILLE	ATGTCCGA

Slim Films

cule d'ADN. L'ADN polymérase commence cette copie quand elle a rencontré un signal de départ qui lui indique le début de la séquence à copier. Ce signal est fourni par une amorce, une séquence d'ADN particulière et souvent courte. Dès que l'ADN polymérase trouve une amorce appariée au modèle, elle commence à ajouter des bases à l'amorce pour créer une copie complémentaire du modèle.

3. Des enzymes nommées ligases raboutent les molécules d'ADN. Lorsque deux molécules d'ADN sont proches, la ligase fait réagir une extrémité d'une molécule avec une extrémité de l'autre molécule : grâce à la liaison covalente ainsi établie, la ligase lie les deux molécules. Les cellules utilisent les ligases pour réparer les molécules d'ADN coupées par les rayonnements ultraviolets ou par d'autres rayonnements ionisants.

4. Des enzymes nommées nucléases coupent les molécules d'ADN. Par exemple, les enzymes de restriction cherchent une succession spécifique de bases dans un brin d'ADN et, quand elles l'ont trouvée, elles coupent l'ADN en deux parties. L'enzyme EcoRI, par exemple, est une enzyme de restriction qui coupe l'ADN de la bactérie *Escherichia coli* après le G de la séquence GAATTC ; elle ne coupe jamais

à un autre endroit. On a supposé que ces enzymes de restriction défendaient la bactérie contre les virus bactériophages, qui injectent leur ADN parasite dans la cellule bactérienne. La bactérie *Escherichia coli* protège son propre ADN de l'action de l'enzyme EcoRI par des méthylations (l'addition de groupes méthyle), lors de la répllication, mais tout virus qui posséderait la séquence GAATTC est inactivé, parce que son ADN est découpé en morceaux. Mon ordinateur à ADN n'utilise pas d'enzymes de restriction, mais celles-ci ont été utilisées, dans de nombreuses expériences, par d'autres équipes de recherche.

5. Des gels pour électrophorèse. Ceux-ci ne sont pas des constituants des cellules. On sépare des molécules d'ADN placées dans une solution en déposant une petite quantité de la solution sur un gel et en appliquant une différence de potentiel entre les deux extrémités du gel. Les molécules d'ADN négativement chargées migrent alors vers l'anode en se frayant un chemin à travers la structure réticulée du gel, à une vitesse qui dépend de leur longueur, les molécules les plus courtes se déplaçant plus vite que les molécules plus longues. Ainsi, les biologistes séparent les brins d'ADN selon leur longueur : l'ajout d'une substance fluorescente qui se

lie à l'ADN révèle la répartition finale des molécules d'ADN en fonction de leur longueur.

6. La synthèse de l'ADN. Aujourd'hui, on peut écrire une séquence d'ADN sur un bout de papier, l'envoyer à un laboratoire de synthèse, et recevoir quelques jours plus tard, un tube à essai qui contient approximativement 10^{18} molécules d'ADN, dont la plupart possèdent exactement la séquence demandée. On atteint facilement des longueurs de 100 bases et, pour des molécules dont la longueur est de 20 bases, le coût n'est que de 150 francs. Les molécules sont fournies déshydratées dans un tube à essai ; elles forment un petit grain, blanc et amorphe.

Comment ces divers ingrédients aident-ils à factoriser les nombres ou à jouer aux échecs ? Dans les années 1930, les mathématiciens ont démontré que deux ingrédients seulement sont indispensables pour construire un ordinateur : un moyen de stocker de l'information et un système pour effectuer quelques opérations simples. La machine de Turing stocke l'information au moyen de lettres écrites sur une bande et la manipule grâce aux instructions simples que possède le contrôleur à états finis. Un ordinateur électronique stocke l'information dans sa mémoire sous la forme de sé-