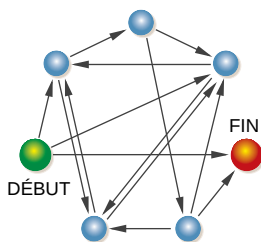
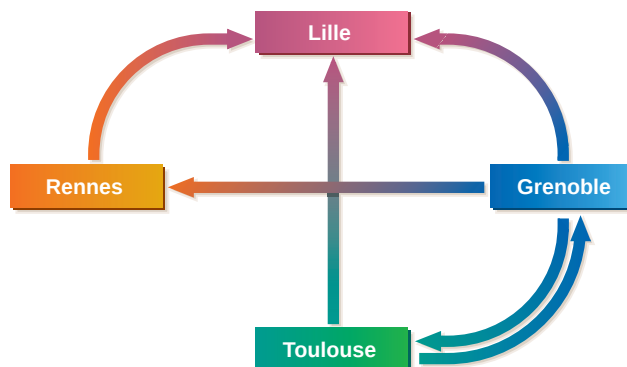


Le chemin hamiltonien codifié par des bases

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	TCGGACTG	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-

quences de 0 et de 1, et il la manipule grâce à l'unité de calcul. N'importe quelle façon de stocker l'information et n'importe quel jeu d'instruction suffisent.

Suffisent à quoi? À effectuer n'importe quel calcul. Pour synthétiser un ADN complémentaire ou pour déterminer des coups aux échecs, il suffit de stocker la bonne information et d'appliquer la bonne séquence d'opérations, c'est-à-dire d'exécuter un programme. Or, l'ADN est un excellent système de stockage de l'information : les cellules l'utilisent depuis des milliards d'années pour stocker le modèle de la vie. Les enzymes telles que les polymérases et les ligases agissent sur cette information. Était-ce suffisant pour construire une machine à calculer universelle? Grâce aux mathématiciens des années 1930, je savais que oui.

Le problème du chemin hamiltonien

Quel problème résoudre? Il ne devait pas être inventé spécialement pour la machine, et il devait démontrer les possibilités de cette nouvelle manière de calculer. J'ai choisi de résoudre le problème du chemin hamiltonien, une variante du problème du voyageur de commerce.

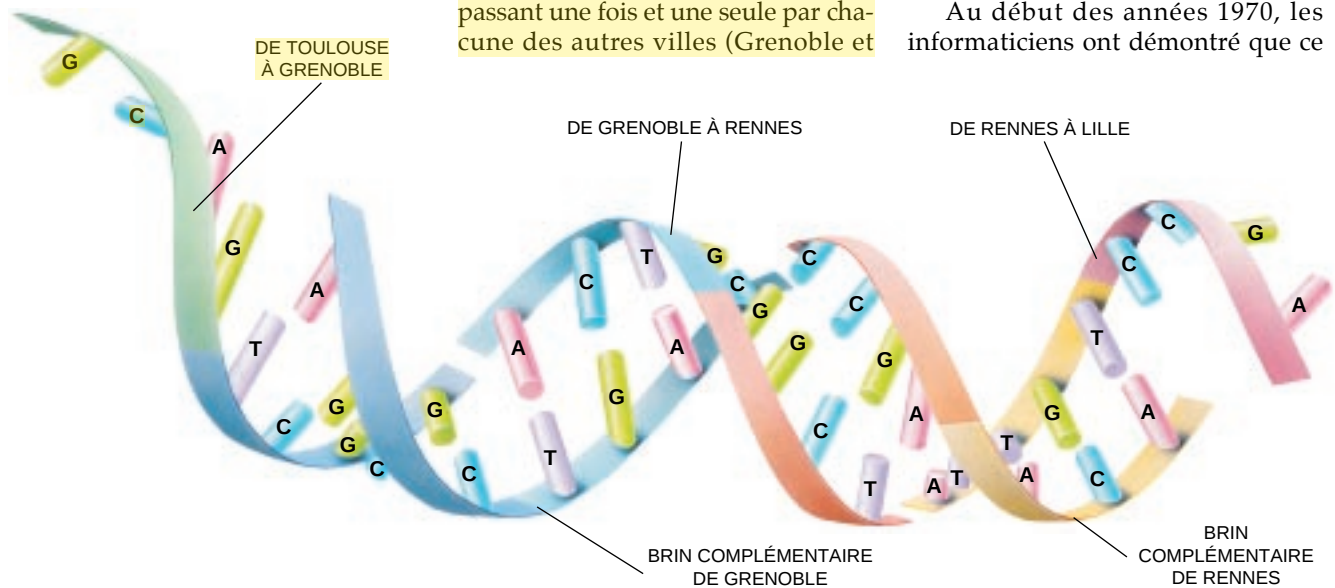
Les chemins hamiltoniens sont dus à William Rowan Hamilton, qui était astronome royal en Irlande, au milieu du XIX^e siècle. Le problème qui porte son nom est décrit dans l'encadré de la page 58 : soit une carte (graphe) où figurent des villes reliées par des vols sans escale (chemins orientés). Par exemple, on peut aller de Grenoble à Rennes, mais pas de Rennes à Grenoble. On cherche le chemin qui part de Toulouse (le sommet de départ) et se termine à Lille (le sommet d'arrivée) en passant une fois et une seule par chacune des autres villes (Grenoble et

Rennes). Un tel trajet est un chemin hamiltonien. Dans cet exemple, on voit rapidement qu'un seul chemin hamiltonien relie Toulouse à Lille ; c'est : Toulouse, Grenoble, Rennes, puis Lille. Entre Lille et Toulouse, il n'existe pas de chemin hamiltonien.

Plus généralement, étant donné un graphe orienté, un sommet de départ et un sommet d'arrivée, alors il existe un chemin hamiltonien si et seulement s'il existe un chemin qui passe par tous les sommets exactement une fois. Le problème du chemin hamiltonien est la recherche de ce chemin.

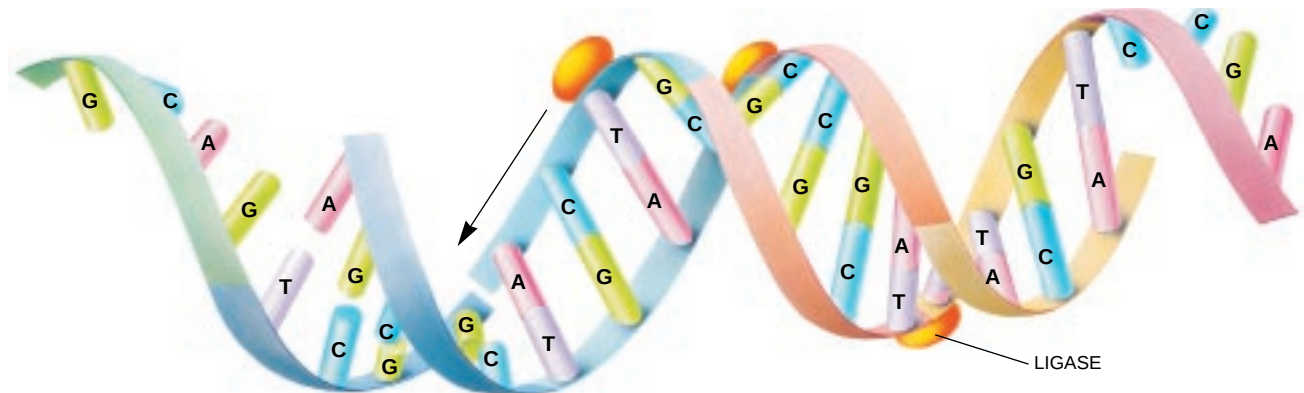
Ce problème a été très étudié par les informaticiens, mais aucun algorithme efficace (c'est-à-dire rapide) n'a été trouvé. On connaît des graphes de moins de 100 sommets pour lesquels, même en utilisant le meilleur algorithme et le meilleur ordinateur, la résolution du problème du chemin hamiltonien prendrait des centaines d'années.

Au début des années 1970, les informaticiens ont démontré que ce



2. L'ASSOCIATION DE MOLÉCULES D'ADN par appariement des C et des G, des A et des T conduit à un brin d'ADN représentant les vols (ici de Toulouse à Grenoble, de Grenoble à Rennes, et de Rennes à

Lille). Ces brins sont appariés à d'autres brins qui codent le complémentaire du nom des villes (ici Grenoble et Rennes). Des centaines de milliards d'appariements se produisent chaque seconde.



3. LES LIGASES RELIENT des molécules. Chaque fois que ces protéines trouvent deux molécules d'ADN à proximité, elles déclen-

chent une réaction des deux extrémités et relient les deux molécules en une seule par une liaison covalente.