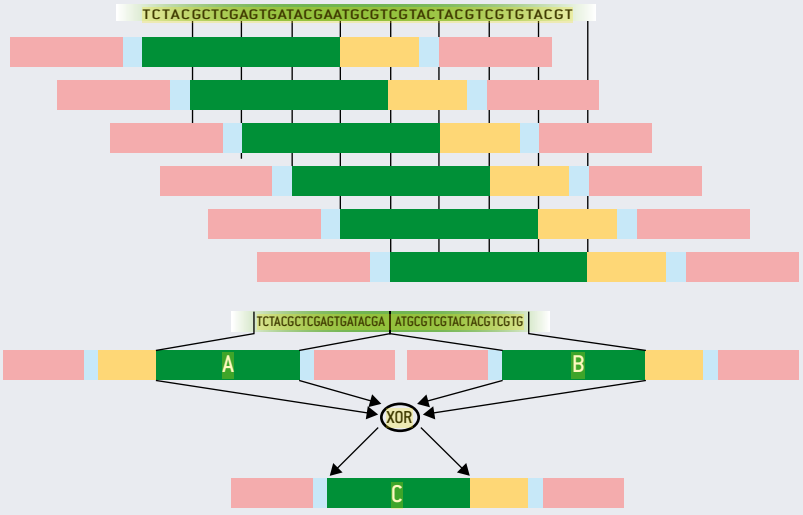
Mémorisation dans l'ADN

James Bornholt et ses collègues ont conçu et mis en application un procédé de mémorisation massive d'informations exploitant les séquences d'ADN. Le procédé se fonde sur l'inscription de couples clé-données, ce qui permet un accès direct aux informations enregistrées analogue à ce qui existe dans la mémoire d'un ordinateur utilisant un système d'adresses : pas besoin, contrairement au cas d'une bande magnétique, de tout parcourir pour retrouver une information.



Le défi relevé était particulièrement difficile pour plusieurs raisons. D'abord, les méthodes de synthèse et de lecture de l'ADN sont imparfaites, avec des taux d'erreurs de l'ordre de 1 % par nucléotide. Les séquences se dégradent aussi progressivement. Il fallait donc concevoir un système fortement redondant. Selon le type de données (image, son, texte), la présence d'une erreur dans les données est plus ou moins grave ; il fallait alors prévoir un système ajustable de redondances.

C'est ce que propose leur procédé. Le point le plus important de leur travail est cependant la mise en place d'un codage qui autorise l'accès à certaines données sans avoir à tout lire. Le système d'adressage conçu utilise la réaction en chaîne par polymérase (PCR, *polymerase chain reaction*) qui amplifie, et donc donne accès, uniquement aux données souhaitées. Bien évidemment, cette lecture choisie accélère l'accès aux données puisqu'elle évite de parcourir tout l'ADN d'une base de données moléculaire. Notons toutefois que les temps d'accès restent bien supérieurs à ceux de l'électronique et se comptent en heures.

L'information sur les séquences d'ADN doit être inscrite avec des redondances (car la transcription et la lecture amènent des erreurs), ce que plusieurs procédés permettent. On inscrit la longue séquence en s'arrangeant pour que chaque brin d'ADN (plus court que la séquence qu'on veut inscrire) contienne une portion de la séquence à mémoriser et qu'il y ait des chevauchements entre les brins *(schéma du haut)*. L'idée utilisée par l'équipe de James Bornholt consiste, pour inscrire la séquence A et la séquence B, à inscrire A sur un brin, B sur un autre et le OU exclusif (XOR) des séquences A et B (après les avoir traduites en binaire), noté C, sur un troisième brin *(schéma du bas)*. Les trois séquences A, B, C ont alors pour propriété que la connaissance de deux quelconques d'entre elles (A et B, A et C, B et C) suffit pour retrouver A et B.

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. » La mise au point de mécanismes de calcul fut cependant délicate, car ce que fait la cellule se décompose en opérations élémentaires qui ne permettent pas facilement d'effectuer des opérations complexes. En plus de la copie de brins d'ADN par l'enzyme ADN polymérase, on dispose d'enzymes nommées ligases qui raboutent des molécules d'ADN et d'« enzymes de restriction » qui coupent les molécules d'ADN à des endroits précis. L'enzyme EcoRI, par exemple, coupe l'ADN de la bactérie *Escherichia coli* après le G de la séquence de nucléotides GAATTC.

Des soupes pour mémoriser et calculer

Poursuivant son travail nocturne, dans les jours qui suivirent son eurêka, Leonard Adleman conçut et réussit à faire fonctionner un procédé de calcul à base d'ADN et d'enzymes qui résout le problème combinatoire de la recherche des chemins hamiltoniens dans un graphe. Il s'agit de trouver un chemin dans un graphe qui passe par chaque nœud du graphe une fois et une seule. Organiser une tournée visitant des villes fixées à l'avance sur un réseau routier ou aérien revient à résoudre un tel problème.

Ce problème appartient à la classe des problèmes NP-complets. Cela signifie deux choses : d'une part, qu'on ne connaît aucun algorithme rapide qui le résout dès que le nombre de nœuds du graphe devient important (« rapide » signifie que le temps d'exécution est une fonction polynomiale du nombre de nœuds du graphe) ; d'autre part, si l'on connaissait un moyen de résoudre rapidement ce problème, on pourrait aussi résoudre rapidement, c'est-à-dire en temps polynomial, tous les problèmes de la classe NP (problèmes dont on peut vérifier qu'une solution est correcte en temps polynomial).

L'algorithme proposé et mis en œuvre par Leonard Adleman en 1994 avec des soupes moléculaires réelles est le suivant :

– On engendre toutes les séquences possibles correspondant à des chemins