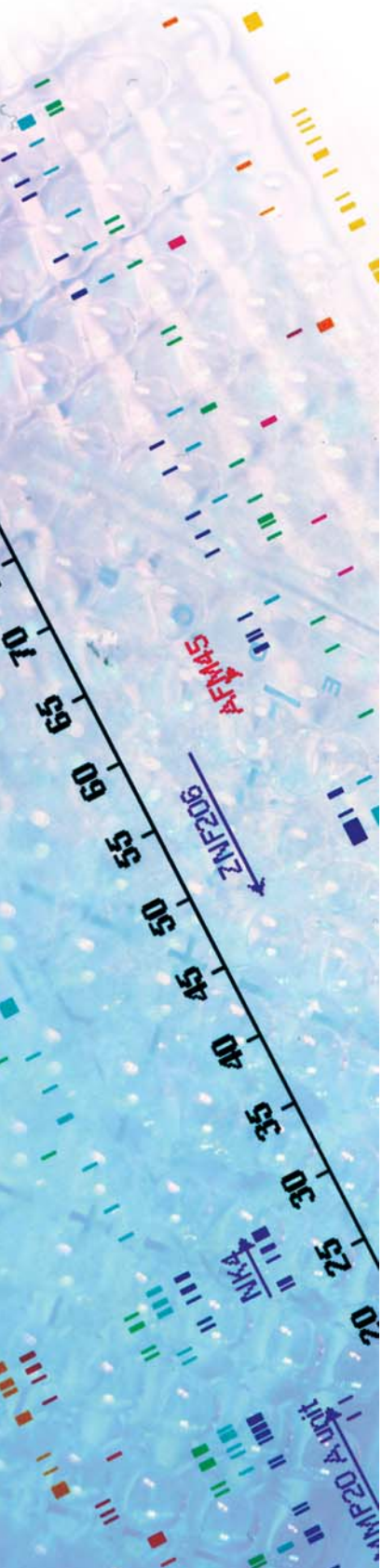
L'interprétation des séquences

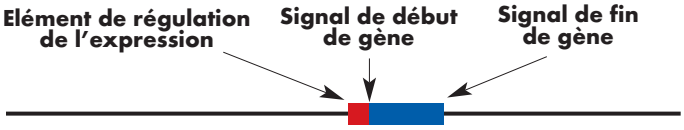
4



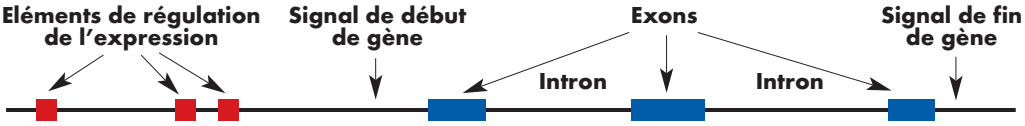
La séquence d'un fragment d'ADN doit être interprétée si l'on veut accéder à toutes les informations qu'elle renferme. Ces informations sont contenues dans différents éléments de la séquence, dont nous ne savons identifier qu'une partie ; toutefois, l'état de nos connaissances évolue rapidement dans ce domaine. Les éléments les mieux connus correspondent aux gènes. Les gènes sont délimités par des signaux de début et de fin.

Ces gènes ne s'expriment pas tous en permanence dans une cellule. Selon les besoins de la cellule, leur expression – leur copie sous forme d'une molécule d'ARN qui va diriger en général la production d'une protéine – est augmentée ou diminuée par des éléments de contrôle situés dans leur voisinage. Les gènes, les signaux de début et de fin et les éléments de contrôle peuvent être décelés, au moins en partie, à l'aide de programmes informatiques.

Gène de bactérie



Gène d'animal ou de plante



"Annoter" une séquence d'ADN consiste à essayer de reconnaître et de repérer, dans la succession des nucléotides, les éléments caractéristiques tels que nous les connaissons aujourd'hui.

La procédure d'annotation dépend aussi de l'origine de la séquence d'ADN. Les génomes des plantes et des animaux contiennent des gènes "en mosaïque" ; dans de tels gènes, la séquence biologiquement significative est morcelée en blocs (exons) qui sont réunis dans la copie ARN par l'élimination des séquences intercalaires sans signification apparente (introns). Les génomes des organismes supérieurs doivent donc être analysés différemment des génomes des bactéries, où les gènes ne sont pas interrompus.

Par l'analyse informatique, on essaie d'une part de prédire avec plus ou moins de fiabilité les parties codantes de gènes et les éléments de contrôle, et l'on procède d'autre part à des comparaisons avec les séquences des banques de données. Les comparaisons permettent de renforcer la fiabilité des prédictions et de déceler des éléments que les programmes de prédiction n'ont pas détectés.

Cependant, les résultats des interprétations effectuées à l'aide de programmes informatiques doivent ensuite être validés par des expériences biologiques. A l'inverse, ces interprétations ont l'avantage d'orienter et donc de faciliter le travail des biologistes. Le Genoscope a développé une expertise propre dans le domaine de la comparaison des génomes entiers *(voir la fiche* Comparer les génomes*)*. En particulier, la comparaison entre les génomes de l'homme et du poisson *Tetraodon nigroviridis*, ou entre ceux de la drosophile et du moustique *Anopheles gambiae*, a permis d'évaluer quantitativement le contenu en gènes de ces organismes, et d'identifier des gènes non détectés par d'autres méthodes. De telles comparaisons permettent aussi d'évaluer la qualité et le degré d'achèvement d'une annotation.