Chapitre 8

Conclusion

And ever, as the story drained
The wells of fancy dry
And faintly strove that weary one
To put the subject by,
"The rest next time -" "It is *next time!"*
The happy voices cry.

Alice in Wonderland. Lewis Carroll.

Comme cela est probablement le cas de beaucoup de travaux, celui-ci s'achève sur une note à la fois pessimiste et optimiste. Le pessimisme se trouve justifié par les diverses limitations qui ont été détaillées dans le chapitre précédent.

Ainsi que nous l'avons alors évoqué, la plupart de ces limitations ne sont pas dues à notre choix initial de ne regarder qu'à l'intérieur d'une fenêtre nous obligeant à ne considérer que l'aspect linéaire des macromolécules biologiques. Ces limitations découlent plutôt du fait que nous n'avons exploité qu'une toute petite partie de l'espace sur lequel s'ouvre cette fenêtre.

Plus que la contrainte d'ordre elle-même, c'est une vision trop étroite de la notion de ressemblance qui nous paraît diminuer la portée de ce que nos algorithmes permettent d'identifier. D'une façon générale, la plupart des méthodes d'analyse des macromolécules (ADN/ARN ou protéines) par des moyens informatiques ne considèrent ces objets que sous un seul point de vue à la fois. Il peut s'agir, soit du point de vue de la structure primaire (la séquence de nucléotides ou d'acides aminés), soit de celui de la structure secondaire ou tertiaire (les chaînes dans l'espace avec ou sans contrainte d'ordre). De plus, on ignore souvent, ou on simplifie dans le meilleur des cas, le fait que les macromolécules ont une histoire évolutive et ne peuvent être traitées isolément et indépendamment les unes des autres.

L'argument que nous voulons avancer ici est que nous devons dorénavant essayer de maintenir plusieurs points de vue simultanément. Nous pouvons imaginer au moins deux voies possibles vers ce but. L'une part d'une analyse des macromolécules en tant qu'objets tridimensionnels et introduit le point de vue de la séquence, de l'évolution ou de la fonction comme autant de contraintes qui vont lui permettre de limiter l'espace de recherche. L'autre voie part de la représentation linéaire des macromolécules (les séquences mais également les structures avec une contrainte d'ordre) et introduit d'autres points de vue (contextuels, structuraux ou évolutifs) afin d'élargir la notion de ressemblance. La première voie présente l'avantage de partir de la représentation la plus complète qu'il est possible d'avoir des macromolécules considérées en dehors de leur milieu naturel (la cellule), la seconde bénéficie du pouvoir et de la flexibilité que lui confère la simplicité de sa représentation linéaire.

C'est cette seconde approche que nous souhaitons poursuivre pour plusieurs raisons.

La première est nous pensons que c'est l'approche la plus adaptée à la compréhension de la relation entre séquence et structure. Cette relation a été beaucoup étudiée, soit dans le cadre de la prédiction de l'une à partir de l'autre, soit dans celui de la modélisation d'une structure en particulier. Les méthodes actuelles de prédiction aussi bien que de modélisation ont un domaine d'application relativement limité. Nous souhaitons comprendre pourquoi et explorer plus précisément la nature de cette relation. La seconde raison pour laquelle nous désirons poursuivre la voie élargissant les définitions de ressemblance à partir d'une représentation linéaire provient de notre sentiment que cette représentation est en fait beaucoup plus riche qu'on ne l'imagine. Les résultats initiaux que nous avons présentés dans le chapitre consacré aux illustrations nous semblent suffisamment encourageants pour nous renforcer dans cette impression.

Cette voie doit bien sûr nous conduire à des approches dont l'aspect lexical ne résidera plus que dans l'existence d'un ordre linéaire sur les éléments composants la macromolécule. Il est clair qu'il faudra étendre la définition d'un mot bien au-delà de ce qu'il représente actuellement à notre esprit. Nous avons déjà commencé à travailler avec des modèles qui sont des objets plus complexes, établis à partir d'une couverture qui condense plusieurs relations de similarité en même temps. Nous avons également mentionné une première façon de modifier de manière plus radicale le sens du terme 'mot' lorsque nous avons introduit les 'faux mots' structuraux composés de symboles qui sont non contigus le long de la chaîne polypeptidique. Il sera sans doute nécessaire d'aller vers des représentations beaucoup plus élaborées encore.

L'aspect qui nous semble cependant le plus passionnant dans ce travail est que non seulement il nous conduit à réfléchir sur des procédés algorithmiques ou des mécanismes biologiques, mais qu'il nous oblige également à nous poser des questions plus fondamentales sur notre représentation des objets. Finalement, notre optimisme se fonde sur le fait qu'arrivés au bout d'une première exploration de la notion de ressemblance, dans ce cas des macromolécules biologiques, nous n'avons pas le sentiment d'être parvenus à la fin d'un parcours.