

Chimie combinatoire virtuelle

ROGER LAHANA

La modélisation de molécules obtenues par la méthode combinatoire et la prévision de leurs propriétés permettent aux chimistes de synthétiser seulement les substances qui semblent avoir l'activité biologique recherchée.

Concevoir de nouvelles molécules qui auront les propriétés thérapeutiques recherchées, tel est le Graal de tout laboratoire pharmaceutique. La chimie combinatoire (voir *La synthèse combinatoire*, par Daniel Michelet et Claude Hélène, page 50) est une puissante méthode de synthèse, offrant d'innombrables molécules originales. Toutefois, l'identification d'une molécule active parmi des centaines contenues dans le même mélange réactif reste souvent difficile. Les chimistes tentent de tourner la difficulté en modélisant les molécules obtenues par la méthode combinatoire avant de les synthétiser : ils évaluent, par des logiciels adaptés, leurs propriétés potentielles et seules les plus prometteuses de ces molécules virtuelles sont effectivement synthétisées.

Ainsi, la modélisation moléculaire facilite la sélection des molécules potentiellement intéressantes obtenues par la méthode combinatoire. L'ordinateur prévoit toutes les combinaisons que l'on pourrait obtenir en faisant réagir des molécules élémentaires judicieusement choisies, puis les chimistes-informaticiens doivent répondre à plusieurs questions : comment sélectionner les plus intéressantes ? Comment limiter le nombre de molécules étudiées sans écarter une molécule qui aurait pu, *in fine*, être intéressante ? Plusieurs de ces questions ouvertes peuvent être résumées en une seule : comment évaluer la diversité des molécules obtenues par la méthode combinatoire, c'est-à-dire déterminer en quoi elles diffèrent ?

Pendant longtemps, les algorithmes disponibles en modélisation ne visaient qu'à l'optimisation rationnelle du meilleur analogue d'une molécule déjà connue (un analogue est une molécule qui, tout en étant différente, se com-

porte comme la molécule mère). Puis, dans leur course aux nouvelles molécules, les laboratoires pharmaceutiques ont parfois abandonné cette approche rationnelle au profit d'une approche «aléatoire». On aurait, semble-t-il, davantage de chances de trouver une molécule intéressante nouvelle en testant plusieurs millions de molécules au hasard qu'en «construisant» une molécule d'après les caractéristiques idéales qu'elle devrait présenter.

En réalité, les deux approches sont complémentaires. Pour qu'une molécule nouvelle présente une activité biologique, il ne suffit pas qu'elle soit un analogue d'une molécule active connue : elle peut, par exemple, avoir une stabilité inadéquate ou interagir avec des cibles différentes de la cible visée, ce qui risque de déclencher des effets secondaires parfois désastreux.

De surcroît, les molécules des banques combinatoires – ou chimiothèques – ne sont pas conçues au hasard : elles sont fondées sur un motif moléculaire particulier, le synthon, sur lequel se feront des variations ; celles-ci sont de plus en plus souvent ciblées, et les chimistes tendent à restreindre la taille des chimiothèques (de plusieurs millions de molécules, il y a quelques années, à quelques milliers, pour faciliter l'identification des molécules actives). Toutefois, s'ils cherchent à réduire la taille des banques, les chimistes veulent conserver la biodiversité, c'est-à-dire les représentants de toutes les familles de molécules obtenues par la méthode combinatoire.

C'est alors qu'intervient la modélisation moléculaire. Fondant leur analyse sur divers paramètres, notamment sur les relations entre la structure d'une molécule et son activité, les chimistes modélisent les molécules des banques combinatoires et évaluent leur diver-

sité. Dès qu'ils ont détecté des molécules potentiellement intéressantes, ils réduisent la taille de la collection, en éliminant notamment les molécules redondantes. Ensuite, ils comparent les molécules des chimiothèques virtuelles et ils conservent uniquement les plus prometteuses. Ainsi, l'amélioration des techniques d'évaluation de la biodiversité moléculaire explique que l'on ait réussi à diminuer la taille des banques combinatoires, virtuelles, puis réelles.

L'évaluation de la diversité moléculaire

Comment évalue-t-on la diversité moléculaire ? S'il est facile d'affirmer que deux objets sont identiques, il est moins aisé de mesurer en quoi ils diffèrent. Supposons, par exemple, que l'on veuille décrire l'humanité en termes de diversité. On dira que les êtres humains appartiennent tous à la même espèce, qu'ils sont donc tous identiques, et que la diversité est nulle. Puis, si l'on différencie les êtres humains selon leur sexe, on crée deux groupes et, si l'on affine encore davantage leurs caractéristiques jusqu'à leur code génétique, on conclut que l'espèce humaine est constituée de plusieurs milliards d'individus tous différents.

Ainsi, selon les critères, ou descripteurs, une même population peut être considérée comme strictement monotone ou infiniment diverse ! Supposons maintenant que l'on classe l'ensemble des êtres vivants en fonction de leur sexe : un homme sera alors plus proche d'un cloporte mâle que d'une femme. La diversité dépend avant tout des critères que l'on choisit, elle ne correspond à aucune grandeur physique absolue.

Pour mesurer la diversité moléculaire d'une chimiothèque, on la décrit

sous forme informatique : il s'agit d'une chimiothèque virtuelle, qui n'existe que dans la mémoire d'un ordinateur, mais dont les caractéristiques sont bien définies. On peut alors optimiser les molécules dont on pense qu'elles auront une activité biologique intéressante, avant de les synthétiser.

Pour préciser la diversité moléculaire, on l'a d'abord définie comme l'opposé de la similitude : moins deux molécules se ressemblent, plus elles sont diverses. Cette quasi-lapalissade a l'avantage d'être simple, d'autant que l'on dispose d'outils informatiques de quantification de la similitude moléculaire, mais elle a plusieurs inconvénients.

Ainsi, comme dans l'exemple de la population humaine, le critère de «similitude moléculaire» est une information utile, mais partielle. En outre, pour évaluer ce critère, on doit comparer les molécules deux à deux, ce qui est rédhibitoire dès que les séries contiennent plus de quelques dizaines de molécules (le nombre de calculs nécessaires varie comme le carré du nombre de molécules) ; or, en chimie combinatoire, on doit traiter plusieurs dizaines ou centaines de milliers de molécules. Supposons que l'ordinateur mette une seconde à effectuer chaque calcul d'évaluation de l'indice de similitude : pour une chimiothèque de 100 000 molécules, il lui faudrait plus de trois siècles !

Pour pallier cette difficulté, les chimistes ont alors décrit les molécules d'une banque par des fonctions spécifiques, nommées pharmacophores. Un pharmacophore est une combinaison de groupes chimiques dont certaines caractéristiques, telles que la charge ou l'affinité pour l'eau, confèrent l'activité biologique à la molécule qui les porte. Les propriétés diffèrent selon la nature des groupes qui constituent les pharmacophores et selon la distance qui les sépare.

Supposons qu'un pharmacophore soit défini comme tous les triangles formés de trois types de groupes chimiques. Chaque triangle est décrit par ses trois sommets et par les longueurs de ses trois côtés. Si l'on considère que l'on a le choix entre six types de groupes différents, 216 combinaisons de sommets sont possibles ($6 \times 6 \times 6$) ; si la distance séparant deux sommets peut prendre dix valeurs différentes, on obtient 1 000 combinaisons de longueurs

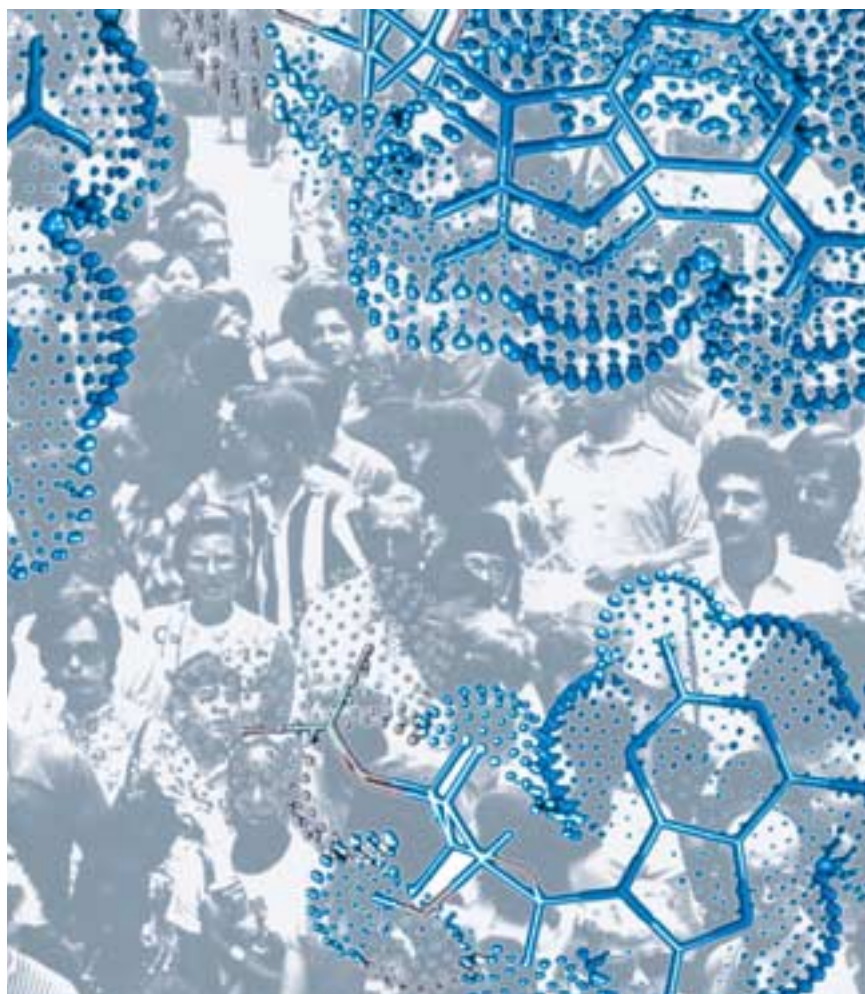
de côtés différentes ($10 \times 10 \times 10$). Toutefois, certaines combinaisons de longueurs sont géométriquement impossibles dans un triangle, et il ne reste que 500 possibilités environ. Pour coder sous forme informatique l'ensemble de ces combinaisons, c'est-à-dire 500 triangles portant les 216 pharmacophores possibles, on définit 108 000 valeurs possibles (500×216) par molécule, valeurs que l'on nomme des clés.

Cette méthode présente quelques inconvénients : un pharmacophore peut être défini par plus de trois centres, ce qui augmente exponentiellement la taille de la clé. Ainsi, si un pharmacophore est un tétraèdre défini par ses quatre sommets et par les six distances qui les séparent, la clé informatique pour une molécule est près de 10 000 fois plus lourde que celle du pharmacophore à trois sommets. Difficile de traiter des

séries de plusieurs centaines de milliers de molécules dans ces conditions (l'espace de stockage nécessaire serait égal à 1 000 gigaoctets pour 100 000 molécules, et le disque dur d'une station de travail est rarement supérieur à une dizaine de gigaoctets).

Une description complète de la diversité moléculaire

Et l'on est toujours confronté à la même question : comment décrire de la façon la plus exhaustive possible la diversité moléculaire ? Nous avons proposé une méthode qui décrit complètement une molécule par des centaines de paramètres : sa masse, son volume, le nombre de ses atomes, sa charge, sa solubilité, la présence ou l'absence de certains groupes chimiques, mais aussi sa topologie (par exemple, la ramification



DE LA DIVERSITÉ génétique à la diversité moléculaire : en combinant quelques briques de construction, les gènes façonnent des milliards d'hommes tous différents. De même, la chimie combinatoire crée d'innombrables molécules à partir d'éléments bien choisis. Avec la modélisation sur ordinateur des molécules des banques combinatoires virtuelles, on sélectionne les plus intéressantes, et l'on ne synthétise par voie chimique que les molécules potentiellement les plus intéressantes.