

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

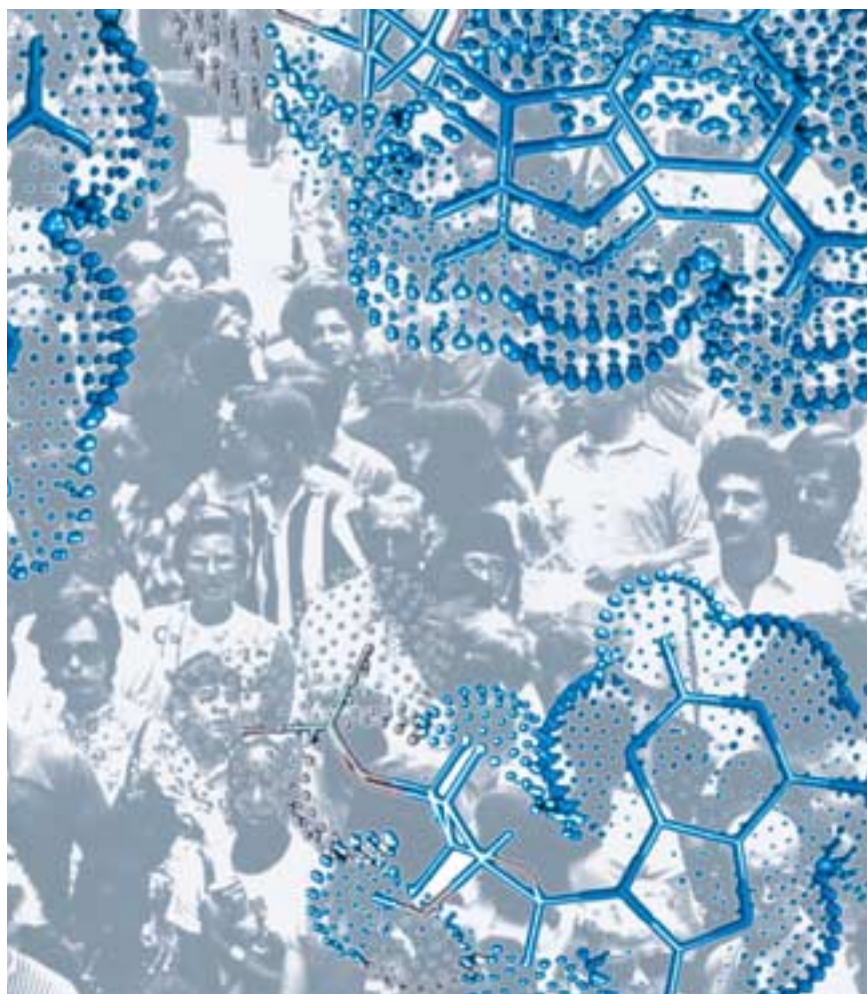
de côtés différentes ($10 \times 10 \times 10$). Toutefois, certaines combinaisons de longueurs sont géométriquement impossibles dans un triangle, et il ne reste que 500 possibilités environ. Pour coder sous forme informatique l'ensemble de ces combinaisons, c'est-à-dire 500 triangles portant les 216 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.

de ses liaisons), sa réactivité, ou encore le coût de sa synthèse ou la disponibilité commerciale des réactifs. Une analyse statistique préalable des variables permet de sélectionner les plus pertinentes. Par exemple, quand on explore une chimiothèque de peptides, molécules très flexibles, on utilise de préférence tous les descripteurs qui ne dépendent pas de la géométrie de la molécule, mais seulement de sa topologie, laquelle reste identique quelle que soit la géométrie.

L'ensemble des clés permet de définir la diversité des molécules d'une chimiothèque, de comparer deux chimiothèques, et de caractériser les «trous» de diversité, c'est-à-dire les plages de variations non décrites par certaines variables. On calcule aisément les distances séparant les molécules dans l'espace de leurs descripteurs : plus la distance qui sépare des molécules est grande, plus ces molécules sont diverses par rapport aux descripteurs utilisés.

Lors de la conception d'une chimiothèque virtuelle, on en limite la taille en évaluant la diversité apportée par chaque nouvelle molécule : soit l'enrichissement en diversité est suffisant, et la molécule est ajoutée à la banque combinatoire, soit il ne l'est pas, et la molécule est rejetée (il sera inutile de la synthétiser, puisque *a priori* elle n'apporte aucune information supplémentaire par rapport aux molécules déjà présentes dans la banque).

Une fois la chimiothèque virtuelle créée, une analyse statistique peut réduire encore la taille de la banque avec une perte limitée d'informations.

De nouveaux immunosuppresseurs

Nous avons appliqué notre méthode à la recherche d'immunosuppresseurs menée par une entreprise de biotechnologie américaine. Les immunosuppresseurs limitent, voire suppriment, les réactions du système immunitaire responsables notamment des rejets de greffe. On disposait d'une molécule qui semblait rendre des souris plus tolérantes aux greffes de cœur. Comment améliorer l'activité et la stabilité de cette molécule ?

La première molécule active était un peptide, c'est-à-dire un enchaînement d'acides aminés ; nous avons comparé sa séquence avec celle d'autres

peptides très proches dont l'activité avait été mesurée expérimentalement. Nous avons trouvé que l'on pouvait faire varier six acides aminés et que certains descripteurs permettaient de séparer les molécules actives des inactives.

Nous disposions pour cette étude de 35 types d'acides aminés différents (les 20 acides aminés naturels et 15 autres artificiels), la banque combinatoire potentielle contenait près de deux milliards de molécules, trop, même pour l'ensemble des ordinateurs de la planète ! Heureusement, les six positions variables étaient soumises à diverses contraintes physico-chimiques, et certaines combinaisons n'apportaient aucune diversité à la banque. De plus, nous savions que la synthèse de certaines combinaisons d'acides aminés ne serait pas réalisable, et nous les avons éliminées de la chimiothèque virtuelle.

Il ne restait alors plus «que» 279 936 molécules virtuelles à tester. Après avoir encore affiné les critères de sélection, il ne restait que cinq molécules présentant les caractéristiques nécessaires pour être biologiquement actives *in vivo*. Seules ces cinq molécules ont été réellement synthétisées, puis testées sur des souris greffées. Parmi ces cinq molécules, quatre ont été plus actives que la molécule testée initialement. L'une d'elles est environ 1 000 fois plus active : traitées par cet immunosuppresseur, les souris meurent de vieillesse sans avoir rejeté leur cœur étranger transplanté.

La chimie combinatoire est un outil puissant d'accès à la biodiversité. Grâce à cette stratégie, nous synthétisons seulement les molécules sélectionnées par l'ordinateur. La chimie combinatoire virtuelle optimise la sélection des molécules qui ont l'activité biologique potentielle la meilleure.

Roger LAHANA est Directeur des développements en modélisation moléculaire de la Société Synt:em.

R. LAHANA, *Les outils de la modélisation moléculaire*, in *Biofutur*, Technoscope, n° 144, pp.1-8, 1995.

R. LAHANA, *Similitude moléculaire et activité biologique*, in *Pour La Science*, pp. 54-60, juin 1995.

A. YASRI et al., *Rational Choice of Molecular Dynamics Simulation Parameters through the Use of Conformational Autocorrelation 3-D Method*, in *Protein Engineering*, vol. 9, n° 11, pp. 959-976, 1996.
