

Une introduction aux algorithmes génétiques

Philippe JAUFFRET

Laboratoire des Systèmes d'Information Chimique

UMR 5076 du CNRS - Montpellier

<pj@chimie.u-strasbg.fr>



sommaire

- quelques définitions
- types de problèmes
- principe général des AG
- un exemple : minimisation de l'énergie
- utilisation des AG en chimie

Définitions

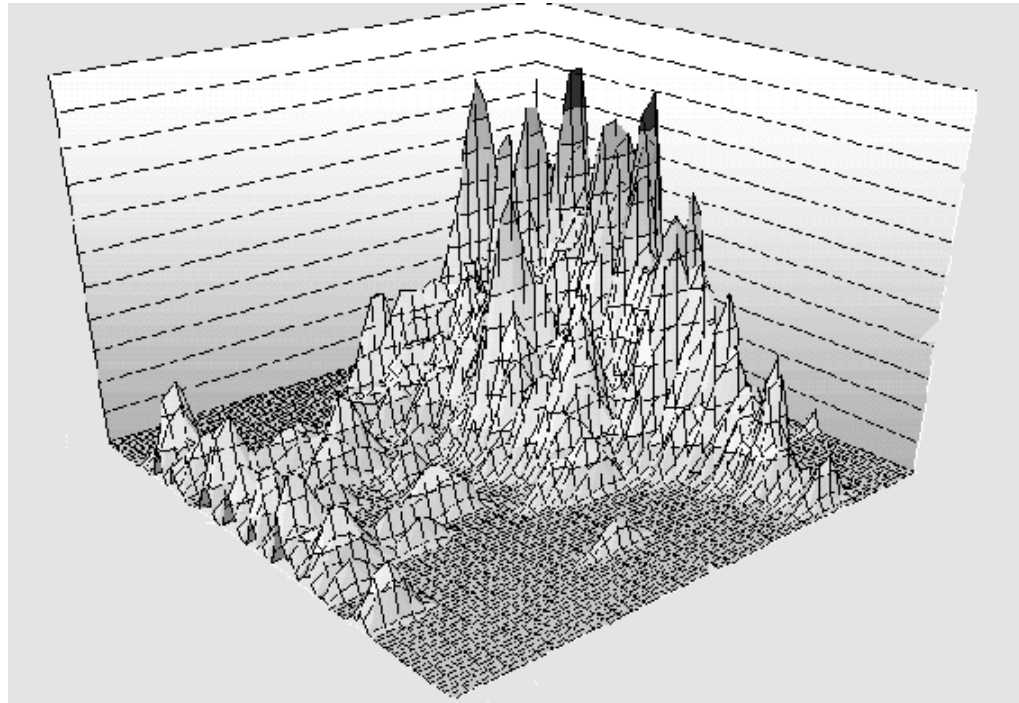
- Les algorithmes génétiques sont des techniques informatiques fondés sur les *principes de l'évolution*
- Les AG empruntent l'essentiel de leur terminologie au modèle évolutionniste : Une population évolue de génération en génération pour devenir plus adaptée aux contraintes environnementales. Ses individus, caractérisés par leurs chromosomes, s'accouplent et se reproduisent, selon le principe général du « struggle for life » (survivance du plus adapté), et l'on observe parfois des mutations.
- Bien que très réducteurs du point de vue du biologiste, les AG fournissent des mécanismes de recherche adaptative robustes et puissants.

Champs d 'application

Les algorithmes génétiques sont un mécanisme général et performant pour résoudre des problèmes pour lesquels:

- il existe un très grand nombre de solutions plus ou moins bonnes
- on ne dispose pas d 'algorithme déterministe pour calculer la ou les meilleures solutions
- l 'univers du problème est peu formalisé

Exemple typique: recherche des extréma sur une surface complexe



28/12/2002

21

5



QUITTER

L 'algorithme génétique

- $t \leftarrow 0$

initialisation au hasard
 - population initiale $P(t)$

de la population initiale
 - évaluer $P(t)$

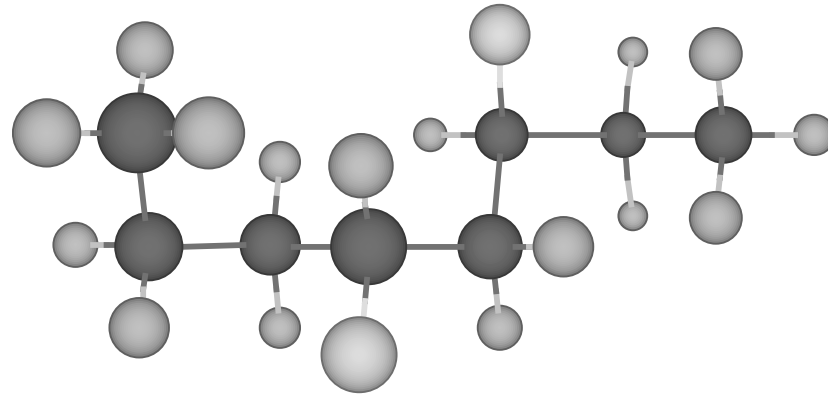
évaluer l 'adéquation (fitness) de tous les individus de la population
 - tant que (non test_de fin)

test de fin : « solution trouvée » ou nb limite d 'itérations dépassé

 - $t \leftarrow t+1$

incrémenter le compteur de temps
 - sélection des survivants
 - accoupler les reproducteurs
 - effectuer les mutations
 - $P \leftarrow$ survivants (P) + descendants
 - évaluer $P(t)$
 - fin tant que
- sous-population de parents
appliquer les opérateurs de croisement et de mutation
former la nouvelle population
et évaluer l 'adéquation de ses individus

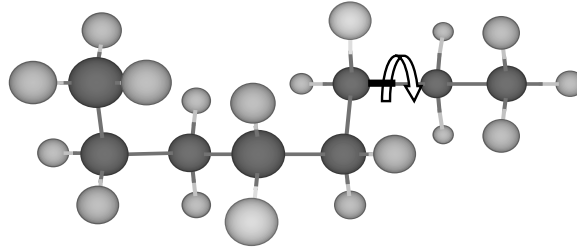
Un exemple simple: recherche de la conformation optimale d'une molécule



- On cherche la géométrie du 3,5-difluoroheptane qui minimise l'énergie de cette molécule

D'après H.M. Cartwright « Applications of Artificial Intelligence in Chemistry » Oxford University Press, 1993.

1) choix de représentation

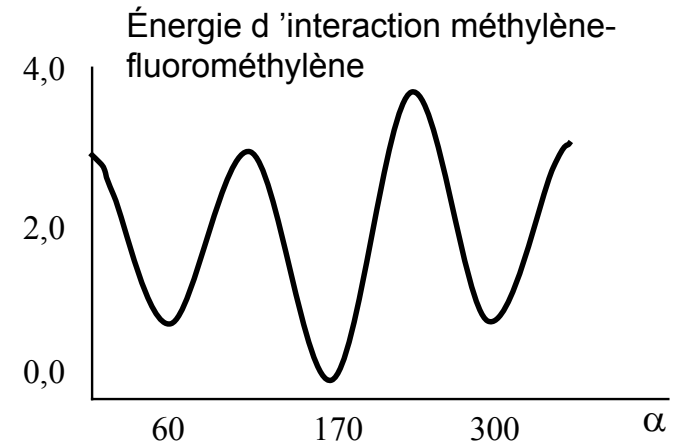
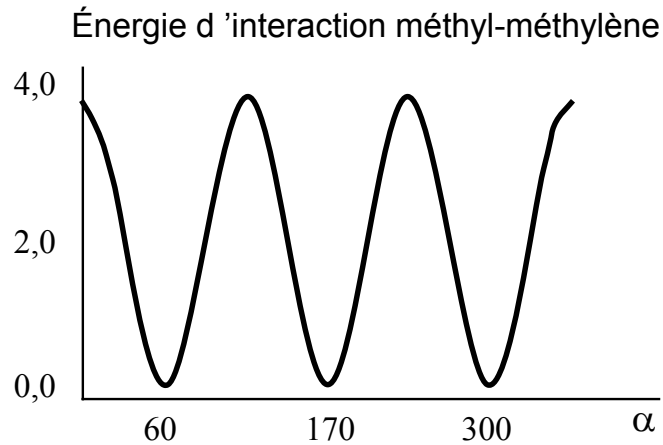


- Les différents conformères ne diffèrent que par la valeur des 6 angles dièdres formés par les liaisons C-C
- on peut donc caractériser chaque individu par un chromosome
 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$
- chaque angle dans la séquence est un gène et sa valeur un allèle.
- En représentant les angles par des entiers, il y a 360^6 individus différents ($\sim 10^{15}$)

2) population initiale

- Le choix de la taille de la population est déterminant pour l'efficacité du système :
 - plus la population est importante, plus grande est la diversité initiale, donc meilleures sont les chances de trouver une bonne solution
 - une population de grande taille augmente la charge de calcul et ralentit la convergence
 - ordres de grandeurs fréquents: 25-100
- l'auteur a choisi une population de taille 10
 - (générateur aléatoire de nombres entiers entre 0 et 359)

3) fonction d'adéquation (fitness)



$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (183, \quad 165, \quad 313, \quad 253, \quad 294, \quad 84)$$

$$E = 0,19 + 0,66 + 0,95 + 2,66 + 0,65 + 1,37$$

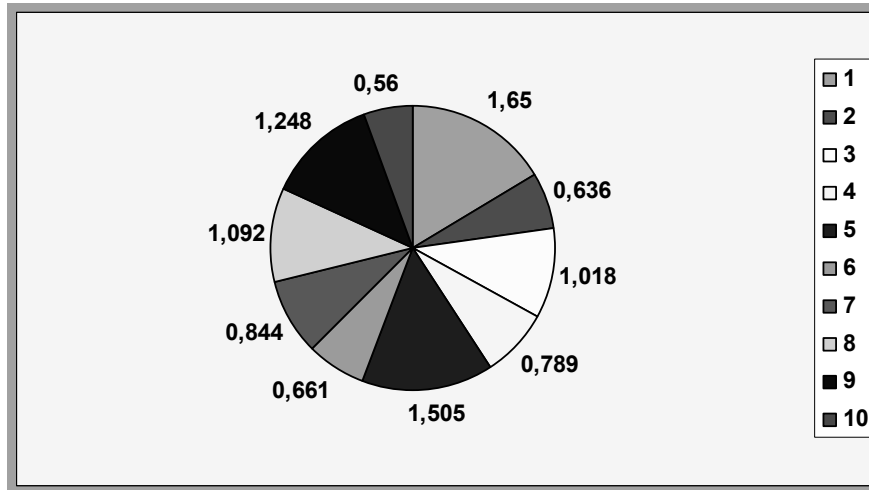
$$E = 6,50$$

$$\text{adéquation (E)} = 1,0 / (E - 0,9) = 6,502$$

fonction d'adéquation (suite)

- La fonction d'adéquation est le seul endroit dans l'algorithme où intervient la connaissance du domaine
 - généralité de la méthode et applicabilité à des domaines peu formalisés
- cette connaissance est « limitée » : on a juste besoin de savoir évaluer l'adéquation de chaque individu, mais pas comment la fonction d'adéquation évolue (comme avec la méthode du gradient)

4) sélection des « survivants » jeu de la roulette

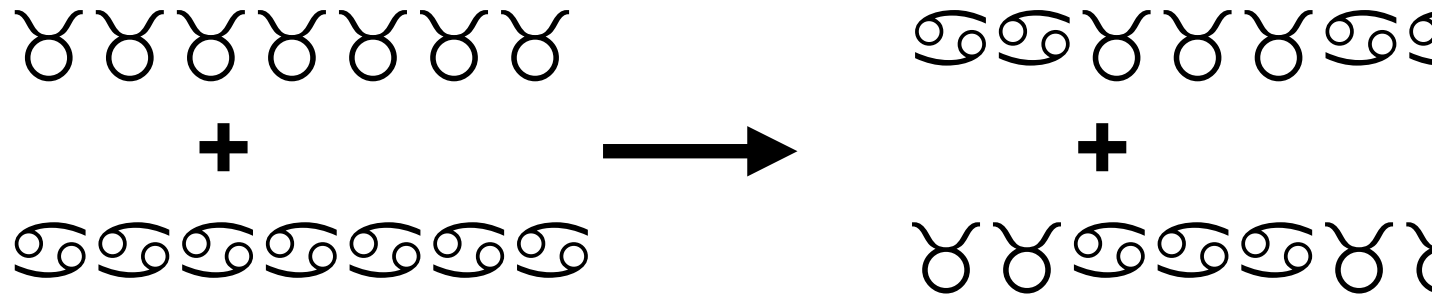


Nb survivants
2
1
1
1
1
1
1
1
1
0

- À chaque individu est associé un secteur de la roulette dont la largeur est proportionnelle à l'adéquation de cet individu.
- On sélectionne les « survivants » en tirant au hasard N individus à l'aide de cette roulette (N: taille de la population)

5) reproduction

l'opérateur de croisement

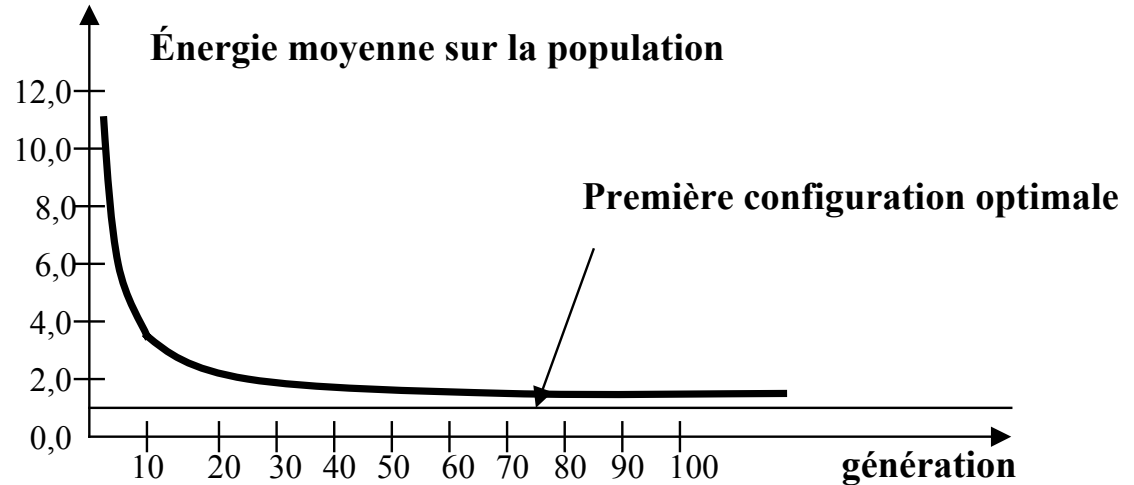


- Tout ou partie des survivants « s'accouplent » et créent des « enfants » combinant les gènes des parents.
- Les enfants s'insèrent dans la population, en remplaçant des individus moins adaptés.

6) mutation

- Objectif : introduction de « nouveau matériel génétique » pour éviter les stagnations liées à l'endogamie. Permet d'éviter les optima locaux.
- méthode : « de temps en temps », changer aléatoirement la valeur d'un allèle sur un individu.
- typiquement, la fréquence des mutations est de 1/1000
- Dans notre exemple, les auteurs ont choisi une fréquence de 1 mutation par génération.

7) résultats



- Convergence rapide de l'énergie moyenne de la population
- première configuration optimale à la 75ème génération

De nombreuses variantes

(parfois disponibles dans le même package)

Genetic Algorithms	Data Types
Generational	Binary
Steady-State	Integer
	Float
Mutation Operators	Crossover Operators
Flip Bit	One Point
Boundary	Two Point
Uniform	Uniform
Gaussian	Arithmetic
	Heuristic
Selection Operators	Termination Methods
Roulette	Generation Number
Tournament	Evolution Time
Top Percent	Fitness Convergence
Best	Population Convergence
Random	Gene Convergence

Difficulté : choix des paramètres

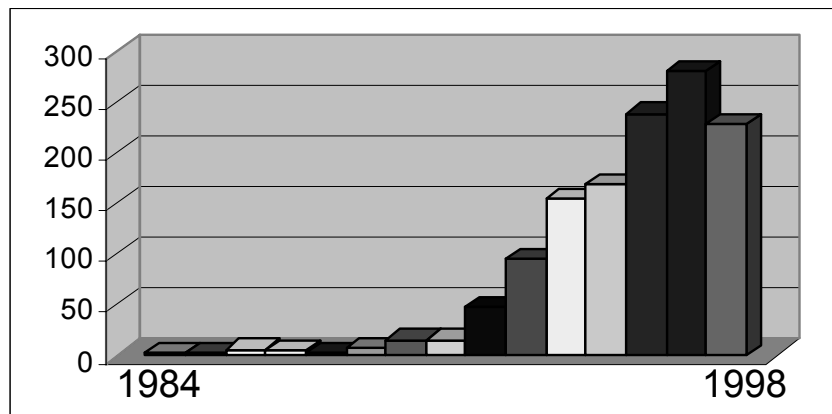
- Le choix des paramètres reste à la charge de l'utilisateur (il n'existe pas de règles autres qu'empiriques)



Quelques champs d'application en chimie

- optimisation de la géométrie de structures
- repliement des protéines
- cinétique de réaction
- élucidation de spectres
- conception de nouvelles molécules

Intérêt croissant pour les AG en chimie



Nombre de publications sur l'utilisation des algorithmes génétiques en chimie et en physique

SOURCE : J.T. Alander « An indexed bibliography of genetic algorithms in chemistry and physics », 2001, <ftp://ftp.uwasa.fi/cs/report94-1/gaCHEMPHYSbib.ps.Z>

Autres méthodes évolutionnistes

Les algorithmes génétiques font partie d'une famille de méthodes informatiques pour la résolution de problèmes appelés « algorithmes évolutionnistes »

- « programmation évolutionniste » (EP)
 - pas de contrainte sur le format de représentation
 - les opérateurs de mutation favorisent statistiquement les « petites variations » comportementales
- « stratégie d'évolution » (ES)
- « système de classification » (CFS)
- « programmation génétique » (GP)
- « algorithme symbiotique »
- etc.

Pour en savoir (un peu) plus

J. Heitkötter (ed.) « The Hitch-Hiker's Guide to Evolutionary Computation », (FAQ for comp.ai.genetic), 1998.
<ftp://rtfm.mit.edu/pub/usenet/news.answers/ai-faq/genetic/>

H. Kargupta & P. Chan (Editor), « Advances in Distributed and Parallel Knowledge Discovery », MIT Press, 2000.

E. Cantu-Paz, « Efficient and Accurate Parallel Genetic Algorithms » (Genetic Algorithms and Evolutionary Computation Volume 1), Kluwer Academic Pub, 2000.

J.T. Alander « An indexed bibliography of genetic algorithms in chemistry and physics », 2001,
<ftp://ftp.uwasa.fi/cs/report94-1/gaCHEMPHYSbib.ps.Z>

autres sites bibliographiques:

http://panizzi.shef.ac.uk/cisrg/links/ea_bib.html

<http://citeseer.nj.nec.com/317688.html>

