الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

N° Réf :……

Centre Universitaire
Abd elhafid Boussouf Mila

Institut des sciences et de la technologie
Département de Mathématiques et Informatiques

Mémoire préparé en vue de l'obtention du diplôme de Master

En: **Informatique**

Spécialité : Sciences et Technologies de l'Information et de la Communication(STIC)

Approche pour la conception de médicaments assistée par ordinateur

Préparé par :

Bouanane Hanane
Benloucif Ahlam

Soutenue devant le jury

Encadré par Bouchekouf Asma……. M.A.B
Président : Bouchemal Nardjes …….M.A.B
Examinateur : Zekiouk Mounira……..M.A.A

Année universitaire : 2015/2016

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

N° Réf :……

Centre Universitaire
Abd elhafid Boussouf Mila

Institut des sciences et de la technologie Département de Mathématiques et Informatiques

Mémoire préparé en vue de l'obtention du diplôme de Master

En: **Informatique**

Spécialité : Sciences et Technologies de l'Information et de la Communication(STIC)

Multi objectif genitic algorithme pour de novo drug design

Préparé par :

Bouanane Hanane
Benloucif Ahlam

Soutenue devant le jury

Encadré par Bouchekouf Asma……. M.A.B
Président : Bouchemal Nardjes …….M.A.B
Examinateur : Zekiouk Mounira……..M.A.A

Année universitaire : 2015/2016

Remerciements

Après la louange et grâce à Dieu qui nous a donné La force et la patience d'accomplir ce modeste travail.

Je remercie vivement Mlle Bouchekouf Asma notre encadreur pour sa présence, son aide

Notre vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à ce travail en acceptant de l'examiner et de l'enrichir par leurs propositions.

Nous tenons à exprimer mes sincères remerciements à tout le personnel de l'institut de sciences et de la technologie du Centre Universitaire de Mila surtout les enseignants qui nous ont enseigné durant ces cinq années d'étude.

Nous remercions également toutes les personnes qui nous ont aidés, de près ou de loin pour la réalisation de ce travail en particulier à nos parents, et nos frères, sœurs, amis qui nous avons encouragé surtout monsieur F.Bensaadi,et Mahdadi Abla et les développeurs de Library CDK monsieur Egon Wiligagen et monsieur John M, qui nous ont soutenu durant tout notre cursus.

Nous remercions tous les étudiants de la promotion 2015/2016 pour avoir été liés et unis tout au long de cette année et tous ceux qui ont collaboré de près ou de loin à l'élaboration de ce travail.

Dédicaces

Je dédie ce travail à:

Ma très cher mère « Seloua » et à mon très cher père « kadour »

A mon bien aimé petit frère: « Mouatez Bi Allah »

A mes sœurs : « Manar » et « Abir » et « Malak »

A le souvenir de mon frére décédé «Yasser »

A tous les membres de la famille « Benloucif » surtout : mes tentes « Rabiaa » et « Fatima », et de la famille « Zendaoui » surtout : mes tentes « Samia » « Yamina » et « zahia » et ma grande mère « Zoulikha »et

A le souvenir de mes grands pères décédés «Aissa »et «Sherif » et ma grande mère « Zahia »

Et touts mes oncles, cousins et cousines.

A tous les amies : ma très cher amie et dans le même temps un membre de ce travail« Hanan» et tous mes amies : « Soumia » et « Amel »et « Iman »

A tous mes enseignants

À toute autre personne qui m'a encouragé ou aidé au long de mes études

Ahlam

Dédicaces

Je dédie ce travail à:
Ma très cher mère « Mounira » et à mon très cher père « Abd Alkarim »
A mon frère: « Samir »
A mes soeurs : «Ahlam » et a mes petites aimé sœurs « Aya » et « Ritaj ».
A tous les membres de la famille « Bouanane » surtout : ma tentes « Nora » et a tous mes tentes et mes oncles ma très cher grande mère « Saada ».
Et a tous mes cousins et cousines surtout a mes cousines «Linda », « Nassima », «Mina », « Asia»
« Rima », « Asma », « Sonia », «Hasna» à mon cousin et mon grand frère « Salah » et tous mes cousins, A le souvenir de mes grands pères décédés «Ammar »et «Ahsan » et surtout mon grand père «Ammar » et ma grande mère «Zineb »
A tous les amies et surtout ma très cher amie et dans le même temps un membre de ce travail « Ahlam».
A tous mes enseignants de primaire jusqu'à l'université a tous les collègues de 2éme année Master STIC
À toute autre personne qui m'a encouragé ou aidé au long de mes études.

Hanane

Table Des Matières

REMERCIMENTS I

Dédicace..... II

Table Des Matières..... IV

LISTE DES FIGURES..... VII

LISTE DES TABLES..... X

LISTE DES ACRONYMES XI

Résume.....XIII

Introduction générale..... 1

Chapitre 1 : Introduction à la chemoinformatique

Introduction.....5

1. la chemoinformatique

1.1. Historique de la chemoinformatique 5

1.2. Définition de la chemoinformatique 6

1.3. Objectifs de la chemoinformatique 7

1.4. Les domaines d'application de la chemoinformatique 7

2. les concepts de base 10

2.1. L'espace chimique 10

2.2. Les grand domains de la chemoinformatique 12

2.3. La théorie des graphes et les graphes moléculaires.....15

2.4. La notion « drug like » et regle de 5.....23

3. Grands Défis pour Chemoinformatics.....24

4. La propriété ADME/TOX 25

CONCLUSION……...28

Chapitre 2 : Le Drug design

Introduction.....29

1. Définition.....29

2. Le processus de découvert de médicament (Drug discovery process….30

3. Les types de drug design.....33

4. La conception de médicament assisté par ordinateur (computer_aided drug design) 34

5. les approches de drug design 36

5.1. Strucure_based drug design 37

5.2. Ligand_based drug design.....41

5.3. La différence entre ligand_based et structre_based drug design ……..46

6. De novo drug design.....47

Conclusion……...52

Chapitre 3 : de novo drug design et l'optimisation multi_objectif

Introduction ……...53

Partie 1 : l'optimisation multi objectif 53

1.1. Définition d'un problème d'optimisation.....53

1.2. Quelque notion sur l'optimisation 53

1.3. Les problèmes d'optimisation mono objective.....54

1.4. Les problèmes d'optimisation multi objectif 54

1.5. Résolution d'un problème d'optimisation 55

1.6. Le front de Pareto les solutions dominé et les solutions non dominé..57

Partie 2 : de novo Drug design et l optimisation multi objectif……58

2.1. Pour quoi de novo drug design est un problème d'optimisation multi objectif ? 58

2.2. Les méthodes d'optimisation pour de novo Drug design.....58

2.3. Le filtrage.....65

3. Les algorithmes proposés 66

3.1. L'algorithme MEGA (The Multi-objective Evolutionary Graph Algorithm) ……66

3.2. L'algorithme MOEA/D (multi objective evolutionary algorithm based decomposition) 68

3.3. L'algorithme NSGA (Non dominated Soring Genetic Algorithm).....71

3.4. Comparaison des algorithmes étudiés 75

Conclusion ……..…76

Chapitre 4 : les outils et les détaille de travaille

Introduction……....77

1. les outils de la chemoinformatique……..……....77

2. Les bases de donné de la chemoinformatique.....85

3. Appliqué L'algorithme MOGA (Multi Objective genetic algorithm) sur problème DND avec les détailles de notre travaille.....*87*

3.1.Définition l'algorithme MOGA.....87

3.2.Principe de l'algorithme MOGA.....88

3.3. Notre Proposition *88*

3.4.Le fonctionnement de notre outil 96

Conclusion……...…….……...98

Chapitre 5 : la validation et les résultats obtenus

Introduction ……..……..99

1. Les données ……...……..99

2. La configuration des paramètres ……100

3. Les testes et les résultats obtenu ……..104

Conclusion ……...110

Conclusion générale ……111

Les références bibliographiques…….…..113

Liste Des Figures

Chapitre1

Figure 1.1 : la chemoinformatique et la bioinformatique ……9
Figure 1.2: la différence entre chemoinformatique et la bioinformatique …....10
Figure 1.3. Définitions des quatre grands espaces chimiques……..…11
Figure 1.4 : Criblage à haut débit……..…14
Figure 1.5 : processus Docking/Scoring…….15
Figure 1.6 :v1…v6=les nœuds, e1…e6=les liens ……..…16
Figure 1.7. Exemple de graphe moléculaire……...16
Figure 1.8. Représentation 2D de la molécule "acétaminophène"…….……19
Figure 1.9. Tableau de correspondance de "l'acétaminophène" ……...20
Figure 1.10 : Tableau de connexion non redondant ……..…20
Figure 1. 11 : Les raisons de rejet des médicaments…… *25*
Figure 1.12 : La propriété ADME/TOX……..26

Chapitre 2

Figure 2.1 : le processus de conception de nouveau médicament temps et cout……29
Figure 2.2 : les étapes de processus de découvert de nouveau médicamen……30
Figure 2.3 : le mode de fonctionnement de computer aided drug desig…….…35
Figure 2.4 : le CADD et le drug descovry……..…35
Figure 2.5 : les approches de drug design…….……36
Figure 2.6: le diagramme de processus structred_based Drug design …….... *37*
Figure 2.7 : le docking et le scoring .Le R représente la structure réceptrice A, B, et C représente les petites molécules qui se lient avec le récepteur …….38
Figure 2.8 : les mécanismes de docking …….39
Figure 2.9: reconstruction de ligand dans le site actif……..……40
Figure 2.10 : phormacophore ……42

Figure 2.11: le QSAR ……42
Figure 2.12 : classification en descripteur……43
Figure 2.13 : les déférents descripteurs……45
Figure 2.14 : QSR 3D et les déférents descripteurs ……46
Figure 2.15 : les types de complémentaire……..…47
Figure 2.16: placement de ligand dans le site actif ……..…….47
Figure 2.17 : le principe de novo Drug design ……48
Figure 2.18: le processus de novo drug design ……..48
Figure 2.19: le growing ……..…49
Figure 2.20: le linking ……50
Figure 2.21: la stratégie link/grow…….……50
Figure 2.22: la stratégie lattice……...…51

Chapitre 3

Figure 3.1 : classification des méthodes de résolution……56
Figure 3.2: le front de Pareto les solutions dominé et les solutions non dominé……..…57
Figure 3.3 : le processus e général de l'optimisation multi objectif……..……61
Figure 3.4 : Illustration de la notion d'optimum de Pareto pour les composés représentés par des points dans une parcelle de la propriété 1 par rapport à la propriété……..…64
Figure 3.5 : le diagramme de fonctionnement de l'algorithme MEGA …..…...67
Figure 3.6: le fonctionnement de l'algorithme PMEGA ……...…68
Figure 3.7 : fonctionnement de l'algorithme NSGA……..…71
Figure 3.8 : principe de fonctionnement de l'algorithme NSGAII……..…72
Figure 3.9: fonctionnement de l'algorithme NSGAIII……..…..73

Chapitre 4 :

Figure 4.1 : Un exemple courant d'analyse à l'intérieur KNIME……..…81
Figure 4.2. Diagramme UML expliquant la hiérarchie d'héritage et les

dépendances entre les classes fondamentales de la CDK……..….82
Figure 4.3 : Le principe général de l'algorithme MOGA……..…….....88
Figure 4.4 : La représentation des chromosomes en nombre entier……...…91
Figure 4.5: Exemple de sélection par roulette……..……..93
Figure 4.6 : Le mode de fonctionnement de notre outil MOGA-DND……97

Chapitre 5 :

Figure 5.1 : Visualisation 2D d'un fragment par guhautil…….……99
Figure 5.2 : La structure 2D des 2 molécules de référence……..……...100
Figure 5.3 : L'interface pour la configuration des paramètres pour l'algorithm…….……101
Figure 5.4 : Interface pour ajouter l'emplacement de la Library acide……..……101
Figure 5.5 : Evaluation de la fonction de fitness pour chaque taille de population……..……102
Figure 5.6 : Valeur de la fonction de fitness pour chaque modification dans le nombre d'itération…….……103
Figure 5.7 : Interface qui affiche la meilleure solution……105
Figure 5.8 : Un graphe qui représente les résultats obtenue……106
Figure 5.9 : Les structures 2D des meilleures molécules obtenues……..……107
Figure 5.10 : Graphe qui représente les résultats obtenu ……108
Figure 5.11 : L'interface qui visualise le meilleur résultat et les valeurs de la fonction de fitness……..109
Figure 5.12 : Les structures 2D des meilleurs résultats……..……110

Liste Des Tables

Chapitre 1

Tableau 1.1 : Exemple de descripteurs en fonction de la dimensionnalité de la structure de départ ……19

Tableau 1.2: Les formats utilisés en chemoinformatique avec des exemples…23

Chapitre 2

Tableau 2.1: le tableau qui illustre la différence entre structre_based et ligand_based drug design……46

Chapitre 3

Tableau 3. 1 : les approches utilisées par les méthodes DND pour adresser la présence de multiple objectif…….……61

Tableau 3. 2 : comparaison entre les algorithmes étudiés……..75

Chapitre 4

Tableau 4.1: les outils de la chemoinformatique……..……..78

Tableau 4.2. Liste de base de données connue en chemoinformatique……...86

Chapitre 5

Tableau 5.1 : les valeurs de la fonction de finesse obtenue on modifiant la taille de la population sur 4 exécutions……..*102*

Tableau 5.2 : les valeurs de la fonction d'évaluation pour chaque nombre d'itération……..*103*

Tableau 5.3 : les résultats obtenus et les valeurs de la fonction objectif pour le cas d'etude 1(lidociane) ……...…*105*

Tableau 5.4 : tableau des résultats obtenu pour le cas d'étude 2 (furano_pyrimidine) ……*108*

Liste Des Acronymes

DD	*Drug Design*
DND	*De Novo Drug Design*
QSAR	*Quantitative structure-activity relationship*
QSPR	*Quantitative structure-property relationship*
HTS	*High-throughput screening*
MDL	*Molecular Design Limited*
MOL	*Molecular*
SDF	*Structure Data Format*
SMILES	*Simplified Molecular Input Line Entry Specification*
ADME/TOX	*Absorption, /Distribution/ Métabolisme /Excrétion/Toxicité*
SBDD	*Structure-Based Drug Design*
PK	*PharmacoKinetics*
PD	*PharmacoDynamics*
IND	*Investigational New Drug*
NMR	*Nuclear magnetic resonance*
EA	*Evolutionary algorithm*
GA	*Genetic Algorithm*
MOGA	*Multi-Objectif Genetic Algorithm*
MOEA/D	*Multi Objective Evolutionary Algorithm Based Decomposition*
NSGA	*Non dominated Soring Genetic Algorithm*
CADD	*Computer_Aided Drug Design*

MOOP	*Multi-Objective Optimization Problem*
SOOP	*Single-Objective Optimization Problem*
TCoef	*Tanimoto Coefficient*
OBA	*Oral Bio-Availability*
GP	*Genetic Programming*
EP	*Evolutionary programming*
CDK	*Chemistry Development Kit*

Résumé :

Le processus de découvert de médicament traditionnelle connue comme le nom de Drug design rationnelle est un processus très Long et couteux, il prend de 12 a 15 ans. La conception de médicament assisté par ordinateur (computer aided drug design CADD) est une nouvelle technique qui apparait ces dernieres années, cette technique permet de développer des nouveaux médicaments en utilisant l'ordinateur, l'utilisation de l'ordinateur dans ce domaine permet de produire des médicaments efficaces en réduisant le temps et le cout de développement. Il existe plusieurs approches dans cette technique, parmi ces approches nous avons l'approche de novo drug design. C'est une approche très puissante qui permet de générer des médicaments qui satisfirent certaines propriétés comme la propriété favorable ADME/TOX. A cause de développement continue dans le domaine de pharmacie, et de la nouveauté de la conception de médicament assisté par ordinateur, nous proposons de réaliser un outil qui permet de générer des nouveaux molécules (médicaments) qui satisfaire certains propriétés, ici nous choisissons de satisfaire deux propriétés : la bio disponibilité orale OBA et la propriété de similarité a un médicament connu en utilisant le coefficient de Tanimoto. Notre objectif ce n'est pas seulement de satisfaire ces deux propriétés mais aussi de définir ce nouveau domaine, de donner une base de recherche et de développement, et d'aider les pharmacies en Algérie de développer des médicaments de qualité en réduisant le temps le cout et les efforts. Le problème DND est un problème NP_complét pour cela dans notre outil MOGA_DND nous essayons d'appliquer les algorithmes génétique multi objectifs, en utilisant l'approche d'optimisation de somme pondéré par agrégation des deux objectifs (OBA et le coefficient de Tanimoto). Une étude expérimentale en utilisant un ensemble de Library de fragment construire a partir d'une base de donné de la chemoinformatique , et la Library open source java CDK (Chimistry Developement Kit) est réalisé. Cette étude donne des résultats prometteurs.

Mots clés : la chemoinformatique, l'optimisation multi objectif, computer aided Drug design, les algorithmes génétiques, le Drug design, Chimistry Developement Kit.

الملخص :

إن العملية التقليدية لاكتشاف الأدوية الجديدة هي عملية معقدة تتطلب الكثير من الجهد و المال إذ تأخذ ما يقارب إل 12 أو 15 عاما. تعتبر صناعة الأدوية المدعمة بالحاسوب تقنية جديدة ظهرت في السنوات الأخيرة هذه التقنية تسمح بإنتاج أدوية جديدة في وقت معقول و بتكلفة مقبولة. هناك عدة مناهج تحتويها هذه التقنية من بينها التصميم دو نوفو و هي تقنية فعالة تسمح بصناعة أدوية ذات نوعية و تلبي بعض الأهداف .ونظرا للتطور الحاصل في ميدان صناعة الأدوية في الجزائر بالإضافة إلى أن ميدان صناعة الأدوية المدعمة بالحاسوب تعتبر مجالا جديدا في الجزائر اقترحنا إنشاء أداة تساعد على صناعة الأدوية باستعمال الحاسوب و التي تلبي هدفين الأول هو مشابهة الجزيئة المنتجة لجزيئة معروفة باستعمال معامل تانيموتو و الثانية هي خاصية التوافر الحيوي عن طريق الفم باستعمال القاعدة الخماسية لليبنسكي. هدفنا من خلال هذه الأداة ليس فقط توفير هاتين الخاصيتين و إنما التعريف بهذا المجال الجديد و إعطاء قاعدة للبحث و التطوير و مساعدة الصيدليين في الجزائر على إنتاج أدوية ذات نوعية بتكلفة و بزمن قصيرين و توفير الجهد.يعتبر التصميم دو نوفو مشكلة معقدة جدا و متعددة الأهداف . حاولنا في أداتنا هذه تطبيق الخوارزمية الجينية متعددة الأهداف. و هذا باستعمال منهج تجميع الأهداف في هدف واحد معقد .دراسة تجريبية باستعمال مجموعة من المكتبات لجزيئات مستخرجة من قاعدة معطيات كيميائية و المكتبة سي دي كا للغة البرمجة جافا تم تطبيقها و التي أعطت نتائج واعدة .

Abstract:

The traditional process of drug discovery known as the name of rational Drug design is a very Lang and expensive process, it takes 12 to 15 years, the computer-aided drug design (CADD) is a new technique that appear in recent years, this technique allows the development of new drugs using the computer, use the computer in this area can produce effective drugs by reducing the time and cost of development, there are several approaches in this technique, among these approaches we have the approach of de novo drug design. This is a very powerful approach to generate drugs which satisfy some properties as the property favorable ADME / TOX. Because of continuous development in the pharmacy field, and the novelty of the drug design computer assisted, we propose to create a tool to generate new molecules (drugs) that satisfy certain properties, here we choose to meet two properties, bio oral availability OBA and drug like property (the similarity to a known drug) using the Tanimoto coefficient, our goal is not just to satisfy both properties but also to define this new field, giving a base of research and development, and to assist pharmacies in Algeria to develop quality drugs by reducing the time the cost and efforts. DND is an NP_hard and multi objective problem, in our tool MOGA_DND we try to apply genetic algorithms multi objectives, using the approach of weighted sum optimization by aggregating two goals (OBA and Tanimoto coefficient), a experimental study using a set of Library of fragments constructing from a chemoinformatic data base, and the open source java library CDK (Chimistry Developement Kit) is done, this study gives promising results.

Key words: chemoinformatique, the multi-objective optimization, Computer Aided Drug Design, genetic algorithms, the Drug design, Chimistry Developement Kit.

Introduction générale :

1. Contexte du travail et motivations

Ces dernières années, on assiste à une situation problématique pour la recherche pharmaceutique.Les investissements sont en augmentation constante, tandis que le nombre de nouvelles entités chimiques introduites sur le marché n'augmente pas. Cela signifie que le coût de mise au point d'un médicament est : tout comme les investissements, en augmentation constante. L'une des raisons de cette augmentation est le durcissement des critères d'acceptation des médicaments par les organismes gouvernementaux.

L'informatique joue un rôle croissant dans la recherche en Chimie. Des secteurs très variés de la recherche fondamentale ou appliquée nécessitent des spécialités du traitement informatique, de l'information chimique, de la modélisation moléculaire ou de la chimie théorique.

La chimie se prête à un traitement informatique car elle est complexe et nécessite des capacités d'acquisition, de traitement et d'archivage considérables. Des bases de données importantes se constituent à travers le monde pour permettre aux chercheurs de suivre quasiment en temps réel l'avancement de la chimie. Le but est de montrer l'implication de l'informatique dans différentes applications de la chimie. Ce domaine est appelé "*Chemoinformatique*".

Le terme "*chemoinformatique* " est apparu il y a quelques années et a rapidement gagné l'utilisation répandue. La définition la plus large de la chemoinformatique « est l'application des méthodes d'informatique pour résoudre des problèmes chimiques ».

La chimoiformatique à des applications différentes dans des domaines variés et l'application la plus communément admise est dans le domaine de la découverte de médicaments «Drug Discovery» qui est aussi connue sous le nom de conception de médicaments « Drug Design, (DD) » qui existe depuis plus d'une décennie.

Le domaine de la «conception de médicaments» peut être défini par un ensemble de processus inventifs capables de trouver de nouveaux médicaments. Un médicament est une petite molécule biologiquement actif qui actif ou inhibite le fonctionnement d'une cible thérapeutique.

Traditionnellement, les médicaments ont été découverts grâce à la recherche de produits naturels et les librairies chimiques de molécules synthétiques biologiquement actives. Par conséquent, le processus de la découverte de médicaments est un processus long et couteux pour l'industrie. De nos jour et avec l'utilisation des méthodes informatique dans ce domaine le processus de découverte de médicament réduit temps et cout.

La conception de médicament a l'aide d'un outil informatique connu sous le nom de la conception de médicament assisté par ordinateur ou en anglais computer aided drug design ou la conception « in silico », joue un rôle très important de nos jours dans la conception des médicaments mise sur le marché et surtout dans les payes développé. Cette approche de conception rationnelle de médicaments est considérée comme l'un des domaines de recherche de la chemoinformatique qui couvre, à ce jour une large gamme d'applications.

2. Problématique

La technique de conception de novo de médicament est l'une des techniques les plus récent et efficace dans le processus de Drug Design assisté par ordinateur, cette technique se décompose en deux catégorie principales, la conception de médicament basé sur la structure cette approche se base sur la connaissance de la structure de cible biologique, le médicament est conçue molécule par molécule ce type de conception prend un temps d'exécution long , la deuxième approche c'est la conception de médicament basé ligand c'est une approche qui concevoir le médicament par assemblage des fragments moléculaires il prend un temps d'exécution raisonnable par rapport au premier approche.

La conception de novo est un problème d'optimisation combinatoire multi objectif, c'est un problème NP complet et sa taille augment en fonction de nombre de Library de fragment utilisé et en fonction de nombre d'objectif considéré, l'optimisation combinatoire multi objectif est un processus qui consiste à optimiser simultanément deux ou plusieurs objectifs, soumis à certaines contraintes, pour cela une méthode de recherche exhaustif ne peuvent pas être raisonnable pour ce type de problème, il existe deux types des méthodes d'optimisation, les méthodes exacte et les méthodes approché, une méthode d'optimisation exacte sachant qu'une telle complexité est accentuée par le coût de l'évaluation des fonctions objectifs qui est souvent prohibitif. Une méthode d'optimisation approché est plus efficace dans ce type de problème pour gérer la recherche dans l'espace issu de la combinaison d'un nombre important de fragments et pour choisir une molécule vérifiant au mieux les propriétés requises en un

temps raisonnable. L'algorithme génétique est une méthode d'optimisation stochastique qui se base sur la méthode d'optimisation approché et qui montre leur efficacité de résoudre plusieurs problèmes d'optimisation combinatoire, pour cela nous choisissons d'appliquer ce type d'algorithme dans notre problème en appliquant différent méthodes de sélection.

Plusieurs objectifs sont prise en compte simultanément dans ce type de problème y compris la propriété drug like(la similarité a une molécule connue), la propriété ADME (Absorption, Distribution, Métabolisme, Excrétion) et plusieurs autre propriété qui sont très importants dans la recherche pharmaceutique. , il est noté qu'à chaque étape de processus de découverte de médicaments, un certain nombre de molécules ne doivent pas passer au stade suivant de développement. L'une des causes les plus fréquents d'échec de composés tête de série "lead" dans les derniers stades de ce processus est le manque de considération de ces contraintes de conception (la satisfaction de ces objectifs considérés).

A cause de complexité d'optimisation de plusieurs objectifs simultanément la plupart des méthodes d'optimisation dans le domaine de DND préfère d'ignorer la nature multi objectif et de regrouper les objectifs dans un seul objectif composite en donnant un poids a chaque objectif.

3. Objectifs et schéma de solution

L'objectif de notre outil développé MOGA_DND (Multi Objectif Genetic Algorithme) est de concevoir des nouveaux médicaments à partir d'un ensemble de Library de fragment, ce médicament satisfaire deux propriétés, la similarité à une molécule de référence calculé par le coefficient de Tanimoto et la bio disponibilité orale calculé par les règles de 5 de lipinski.

Les composants principaux de notre outil sont : un ensemble de Library de fragments construit a partir d'une base de donné de la chemoinformatique, la Library open source CDK (Chimistry Developement Kit) une Library open source java développé pour la bio et la chemoinformatique.

L'idée principale de notre outil consiste donc à concevoir à l'aide de CDK une molécule composite «drug-like» à partir de plusieurs fragments qui doit satisfaire simultanément deux objectifs d'une importance pharmaceutique qui sont: drug-likness de notre molécule conçue et sa similitude à une molécule de référence. Le premier objectif est calculé à l'aide de l'évaluateur de la biodisponibilité qui se base sur les "règles des 5" de Lipinski cet objectif est

nécessaires pour la sélection de composés «drug-like». Tandis que, le second objectif est calculé en utilisant l'évaluateur de similarité qui se base le "coefficient de Tanimoto" pour mesurer la similarité entre une molécule conçue et une molécule de référence.

4. L'organisation de mémoire

Notre mémoire est organisé auteur de 5 chapitres :

Chapitre1 : Dans le premier chapitre nous présentons une introduction mettant en exergue les aspects généraux relatifs au domaine de la chemoinformatique et ses notions fondamentales.

Chapitre2 : Dans le deuxième chapitre nous présentons la définition et les concepts fondamentaux de notre domaine de recherche le Drug design et de novo Drug design avec les différentes approches qu'existe dans la littérature.

Chapitre3 : dans le troisième chapitre nous représentons l'optimisation combinatoire avec les différentes approches qui existent dans la première partie, dans la deuxième partie nous présentons les méthodes d'optimisation multi objectif dans notre domaine de novo drug design, puis les trois algorithme étudié avec les variantes de chaque algorithme et leur avantage et inconvénient.

Chapitre4 : dans le quatrième chapitre nous présentons les outils de la chemoinformatique, nous choisissons 3 outils pour chaque outil nous présentons la définition et les fonctionnalités principale, puis nous citons les base de donné de la chemoinformatique, et nous terminons ce chapitre par l'algorithme choisie ; définition, principe avec les détaille de travaille faite.

Chapitre5 : le cinquième chapitre représente les études faites en appliquant l'algorithme MOGA_DND et les résultats obtenus.

Chapitre 1

Introduction à la chemoinformatique

Introduction

Le terme «chemoinformatique» a été inventé il y a seulement quelques années, mais il a gagné rapidement une utilisation généralisée. Des ateliers et des colloques consacrés exclusivement à chemoinformatique ont été organisés, et les revues sont pleines de publicités de positions pour chemoinformatique spécialistes. Parce que le nom chemoinformatique est nouveau, il y a encore des vues différentes de la portée du domaine et même sur l'orthographe du nom. Deux orthographes, "chemoinformatique" et "chimionformatique" sont actuellement utilisés.

Il est clair que la chimie est une discipline scientifique qui est en grande partie construite sur des observations expérimentales et des données. La quantité de données et d'informations accumulées est toutefois énorme, et la taille de cette montagne augmente à une vitesse croissante. Le problème est donc d'extraire des connaissances à partir de ces données et ces informations, et utiliser ces connaissances pour faire des prédictions. Ceci est où chemoinformatique est utile.

La chemoinformatique représente le domaine de notre étude. Aussi, nous présentons dans ce chapitre une introduction mettant en exergue les aspects généraux relatifs à ce domaine. Nous commençons d'abord, par une brève histoire de cette nouvelle discipline, ses principaux objectifs, ainsi que ses différentes applications dans les différents domaines. Nous présenterons ensuite, les concepts de base de la chemoinformatique nécessaires à la compréhension des méthodes informatiques et des applications développées dans cette mémoire.

1. la chemoinformatique

1.1. Historique de la chemoinformatique :

La chemoinformatique est une nouvelle discipline apparue il y a environ 40 ans. Au début des années soixante, dans le but d'élucider la structure de composés chimiques inconnus, les données provenant des méthodes existantes (spectroscopie) ont été mises en commun sur informatique, C'était la naissance de la chemoinformatique.[1]

- ➔ en 1964 : Le projet DENDRAL initié à l'université de Stanford a été le premier à développer des générateurs de structures chimiques à partir de spectres de masses.

- ➔ à la fin des années 60 : Sasaki à l'université de technologie de Toyohashi et Munk à l'université d'Arizona ont utilisé plusieurs méthodes de spectroscopie afin d'élucider la structure chimique de leurs composés.
- ➔ En 1969 : Corey et Wipke ont présenté un travail similaire concernant les systèmes de représentation des molécules.
- ➔ Peu après d'autres groupes comme Ugi , Hendrickson et Gelernter ont développés des systèmes pour représenter des molécules organiques. Dès lors, beaucoup d'articles concernant cette discipline ont commencé à apparaitre dans les journaux scientifiques.
- ➔ C'est seulement en 1998 que F.K Brown a défini pour la première fois cette discipline comme étant la chemoinformatique. [1]

1.2. Définition de la chemoinformatique :

La chemoinformatique (parfois aussi appelé chéminformatique,chémioinformatique ou chemie informatique)est une discipline scientifique relativement récente qui consiste à étudier et à résoudre les problèmes relatifs à la chimie en appliquant des méthodes informatiques[2]

La notion de la chemoinformatique a émergé pour la première fois lors du développement d'un médicament en 1998.

- **La definition de Franck BROWN :**

  "Chemoinformatics is the mixing of those resources information (information technology and management) to transform data into information and information into knowledge for intended purpose of making better decisions faster in the area of drug lead identification and optimization"F. K. Brown. [3, 4]

- **G.Paris(1999) :**

  " La chemoinformatique est un terme générique qui regroupe la conception, la création, l'organisation, la gestion, l'extraction, l'analyse, la valorisation, la visualisation et l'utilisation de l'information chimique." [5]

- **La définition de J. Gasteiger:**

  J. Gasteiger a définit la chemoinformatique en 2003 comme suite :

  "L'application de méthodes de l'informatique pour résoudre des problèmes chimiques ". JGasteiger**.** [6]

➢ **Varnek.A et Baskin.I** ont dit que:

"Chemoinformatics is a field based on the representation of molecules as objects (graphs or vectors) in a chemical space".[7]

Depuis, l'acception du terme chemoinformatique s'est étendue pour désigner le traitement informatique de l'information chimique en général.

1.3. Objectifs de la chemoinformatique :

La chemoinformatique est là pour répondre et aider les chimistes à résoudre certains des problèmes fondamentaux suivants:

1. Concevoir des molécules avec des propriétés désirées: L'objectif majeur d'un chercheur en chimie est de produire des composés possédant les propriétés souhaitées et d'étudier la relation quantitative structure-activité qui peut être employée pour la prédiction de la propriété de nouvelles molécules.

2. Concevoir la réaction chimique et synthétiser des nouveaux composés: la conception de la réaction comprend les réactifs utilisés pour synthétiser les composés (les produits) issus de cette opération chimique.

3. Analyser et élucider les structures obtenues lors de la réaction : Il est nécessaire d'établir la structure du produit due à la réaction en utilisant les différents outils pour déterminer la structure.

4. Transformer les données en informations et celles-ci en connaissances, afin de prendre la meilleure décision possible. [8]

1.4. Les domaines d'application de la chemoinformatique :

La chemoinformatique est un nouvelle discipline qui est utilisé dans différents domaine telle que : l'apprentissage automatique, la chimie computationnelle, pharmacologie, la visualisation des données …etc. Nous citerons les certaines de ces domaines :

1.4.1. La chemoinformatique et l'apprentissage automatique

La chemoinformatique utilise des méthodes issues de l'informatique, plus particulièrement la théorie des graphes et l'apprentissage automatique (Machine Learning), afin de classifier ou prédire les propriétés de bases de molécules. certaines méthodes d'apprentissage de la machine sont très populaires dans la chemoinformatique et particulièrement dans le modèle de relation quantitative

structure-activité (QSAR) qui se base sur la corrélation entre un ensemble de descripteurs associés à la molécule et une propriété à prédire [10]. Ce principe nécessite l'application de la théorie des graphes, pour développer des nouveaux descripteurs et des noyaux sur graphes, et d'appliquer des méthodes d'apprentissage capables de traiter des données structurées [9]. Cependant, la première représentation in silico de la molécule a été développée en basant sur une liste de descripteurs qui peut être calculée à partir de la structure, des propriétés physiques ou bien encore de l'activité biologique de la molécule. L'ensemble de ces descripteurs est regroupé au sein d'un vecteur de taille fixe. Cette dernière représentation permet d'appliquer à la chemoinformatique un vaste ensemble de méthodes numériques définies dans le cadre de l'analyse de données et des méthodes d'apprentissage.

1.4.2. La chemoinformatique et la chimie computationnelle

Depuis nombre d'années, la bioinformatique, la chemoinformatique et la chimie computationnelle (ou numérique) sont utilisées en conjonction avec la recherche expérimentale pour faciliter l'analyse de séquences et de génomes, la prédiction de la structure ou de la fonction de protéines, la modélisation de systèmes biologiques au niveau moléculaire, la conception de nouveaux médicaments, l'étude des propriétés moléculaires, la prédiction de réactions chimiques, etc. Étant donné la quantité toujours croissante de données biologiques, pharmacologiques et chimiques qui ne cessent de s'accumuler dans de multiples banques de données publiques et privées, il n'y a aucun doute que des méthodes informatiques permettant une analyse efficace et l'extraction d'informations pertinentes de ces banques de données continueront de jouer un rôle crucial dans les décennies à venir. Ces méthodes, combinées avec des approches de modélisation moléculaire à la fine pointe de la technologie, devraient notamment continuer de contribuer à l'acquisition d'une meilleure compréhension des mécanismes impliqués dans diverses maladies, et permettre la découverte et le développement de nouvelles approches thérapeutiques. [11]

1.4.3. La chemoinformatique et pharmacologie :

La chemoinformatique est notamment utilisée en pharmacologie pour la découverte de nouvelles molécules actives et la prédiction de propriétés à partir de structures moléculaires. [12]

1.4.4. La chemoinformatique et la visualisation des données:

Comme il est connu la visualisation d'information (données) est un domaine d'informatique dont l'objet d'étude est la représentation visuelle de données, principalement abstraites, sur une Interface graphique. la chemoinformatique utilise ce domaine informatique pour la visualisation abstrait des informations ou des données chimique sur des graphes comme: les molécules ,les fragments, les liaisons chimique ...etc. , pour facilitée les traitements chimique comme les prédiction des nouveaux structures ,création des liaison entre des molécules, réaction, similarité ....etc. [13]

1.4.5. La chemoinformatique et la bioinformatique :

La chemoinformatique puisse être moins connue que la bioinformatique, elle a une histoire considérable. La différence entre ces deux disciplines, c'est que la première se concentre sur l'étude de la structure des petites molécules, tandis que la deuxième est appliquée à l'analyse des séquences biologiques.

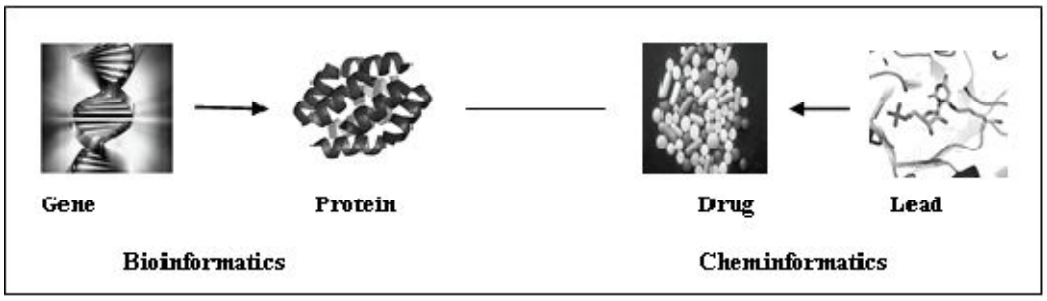


Figure 1.1 : la chemoinformatique et la bioinformatique [38]

Les différences entre ces deux disciplines sont citées dans la figure suivant :

1978	Bioinformatique	Chemoinformatique	1998
1977	Assemblage de séquences	Elucidation de structures	1960s
1970	Homologie	Similarité	1970
1995	Prédiction de fonctions	Prédiction d'activités	1947
2000s	Inférence de réseaux	Générateur de réseaux	1970s
1960s	Dynamique des réseaux	Cinétique	1960s
1960s	Pliage des protéines	Analyse conformationelle Docking	1960s
1960s	Modélisation moléculaire	Modélisation moléculaire	1960s

Figure 1.2: la différence entre chemoinformatique et la bioinformatique [39]

2. les concepts de base

Cette section est consacrée à la définition et la présentation des concepts de base de la chemoinformatique nécessaires à la compréhension des méthodes informatiques et des applications développées dans cette mémoire.

2.1. L'espace chimique :

2.1.1. Définition :

L'espace chimique est l'espace qui contient toutes les molécules théoriquement possibles et est donc théoriquement infini**.**

2.1.2. Types de l'espace chimique

Hann et Oprea ont défini quatre types d'espaces chimiques (Figure) [14]

- **Virtuel :** l'espace chimique virtuel regroupe tous les composés qu'il serait possible de synthétiser. Ce nombre est grossièrement estimé à 10^{60}. [15]

- **Tangible :** l'espace tangible contient toutes les molécules qui peuvent être facilement synthétisées. Une publication récente estime que ce nombre de composés est compris entre 10^{20} et 10^{24} [16].

- **Global :** l'espace global regroupe tous les composés synthétisés. Il est impossible de connaître exactement le nombre de molécules synthétisées dans le

monde, car beaucoup de molécules sont synthétisées pour la recherche industrielle et donc confidentielles. On estime très approximativement ce nombre à 80 millions.

- **Réel :** l'espace réel correspond à tous les composés possédés par un organisme. Le nombre de composés maximum que possède une société est estimé à 10 millions. [17]

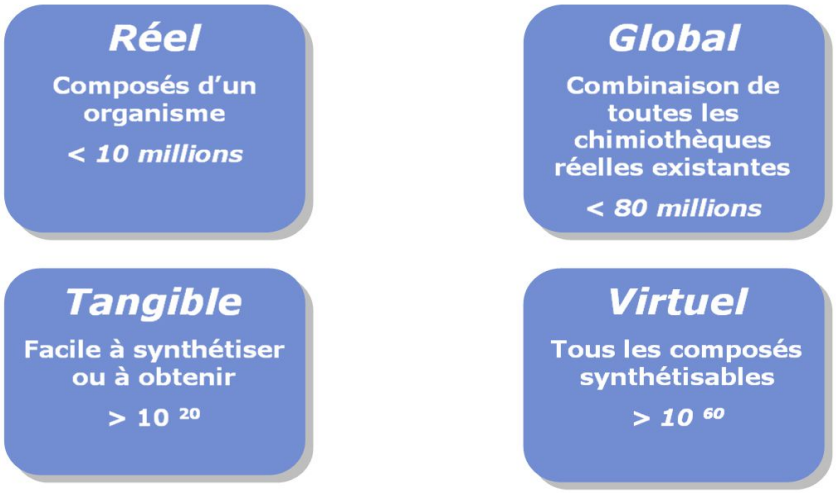


Figure 1.3 : Définitions des quatre grands espaces chimiques. [40]

2.1.3. Bases de données moléculaire :

-Définition :

Une base de données chimique (moléculaire) est une base de données (éventuellement bibliographique) spécifiquement dédiée à l'information chimique. La plupart des bases de données chimiques stockent des informations sur des molécules stables. Les grandes bases de données chimiques devraient être capables d'assurer le stockage et la recherche d'informations sur des millions de molécules (ou autres objets chimiques). [18]

-Exemple sur base de donné chimique:

- Chemical Abstracts Service (1975)
- Cambridge Structure Database (1984)
- Beilstein (1990)
- Gmelin (1990)

- ChemInformRX (1991)
- SpecInfo (1991)
- PubChem (2004)
- ChemSpider(2009)
- etc.

2.1.4. Les bibliothèques virtuelles :

Bibliothèque virtuelle est une bibliothèque qui n'a pas d'existence physique, construit uniquement sous forme électronique ou sur papier. Les blocs de construction nécessaires à une telle bibliothèque ne peuvent pas exister, et les étapes chimiques pour une telle bibliothèque ne peuvent pas avoir été testées. Ces bibliothèques sont utilisées dans la conception et l'évaluation de la possibilité libraires. [19]

Les données chimiques peuvent se rapporter à des molécules réelles ou virtuelles. Les bibliothèques virtuelles de composés peuvent être générées de différentes manières d'explorer l'espace chimique et formuler des hypothèses de nouveaux composés ayant des propriétés souhaitées. Les bibliothèques virtuelles des classes de composés (médicaments, produits naturels, des produits de synthèse orientée vers la diversité) ont été générées récemment. [20]

2.2. Les grand domaines de la chemoinformatique:

2.2.1. Le criblage:

On distingue deux types de criblages :

- Le criblage réel à haut débit
- Le criblage virtuel

A. Criblage réel à haut debit (High-throughput screening):

Le criblage réel permet quand à lui de tester rapidement *in-vitro* l'activité de composés biologiques. On est cependant limité par le nombre de composés qu'il est possible de tester en un temps raisonnable et par le coût des tests.

L'expression de criblage ou criblage à haut débit (high-throughput screening, HTS) désigne dans le domaine de la pharmacologie ,les techniques visant à étudier et à identifier dans les chimiothèques et ciblothèques, des molécules aux propriétés nouvelles, biologiquement actives.

L'expression haut débit évoque ici l'utilisation de la robotique, de l'informatique et de la bio-informatique pour accélérer la phase de test des molécules, protéines, catalyseurs, etc. en vue de processus de production de médicaments, …etc. [21]

B. Criblage virtuel (Virtual screening) (ou essai in silico)

Le criblage virtuel est réalisé *in-silico*. Il permet de réaliser de manière rapide et à moindre coût des prédictions de l'activité des molécules.

A la différence de criblage à haut débit, le criblage virtuel de calcul comprend le dépistage (screening) dans les bibliothèques de composés silico, au moyen de diverses méthodes telles que l'amarrage (doking), pour identifier les membres susceptibles de posséder des propriétés désirées telles que l'activité biologique contre une cible donnée. Dans certains cas, la chimie combinatoire est utilisé dans le développement de la bibliothèque d'augmenter l'efficacité dans l'exploitation minière de l'espace chimique. Plus communément, une bibliothèque diversifiée de petites molécules ou des produits naturels est projeté. [21]

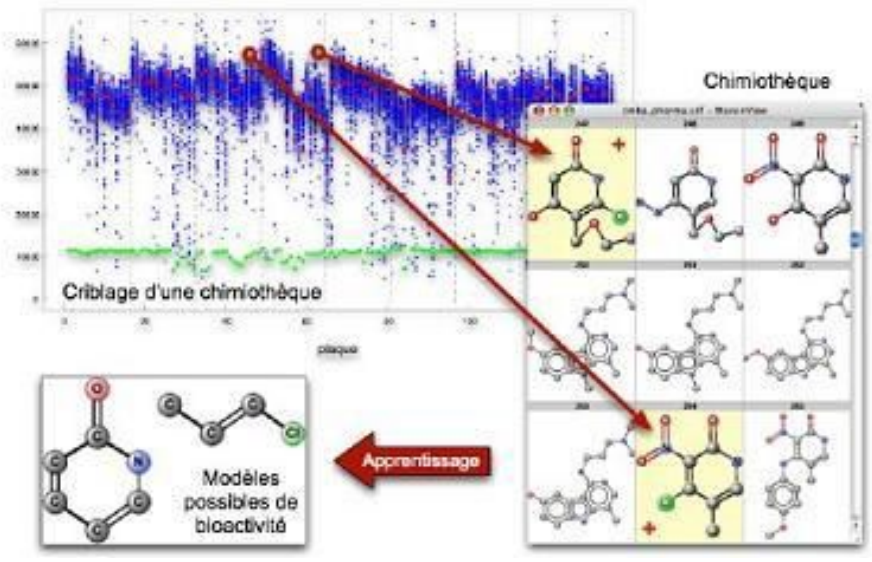


Figure 1.4 : Criblage à haut débit.[41]

2.2.2. Amarrage/ striage (doking/ Scoring):

Face à la volonté de découvrir toujours plus de médicaments, il est indispensable de sélectionner rationnellement, pour une cible donnée, des composés issus de la chimie organique.

Des méthodes de modélisation moléculaire, dont le Docking, permettent d'isoler, à partir de larges chimiothèques, les quelques molécules les plus potentiellement actives. Ce protocole de Docking nécessite la structure du récepteur étudié, obtenue par cristallographie X ou à partir de données RMN, duquel sera extrait le site actif. L'algorithme va alors y placer chaque molécule et évaluer le complexe ligand/récepteur nouvellement formé. Le positionnement de la molécule ainsi que l'aspect interactionnel sera calculé par une fonction de scoring qui classera chaque composé par une valeur numérique. [22]

Ce processus Docking-Scoring appliqué aux librairies de molécules est appelé criblage virtuel qui nous citons précédemment.

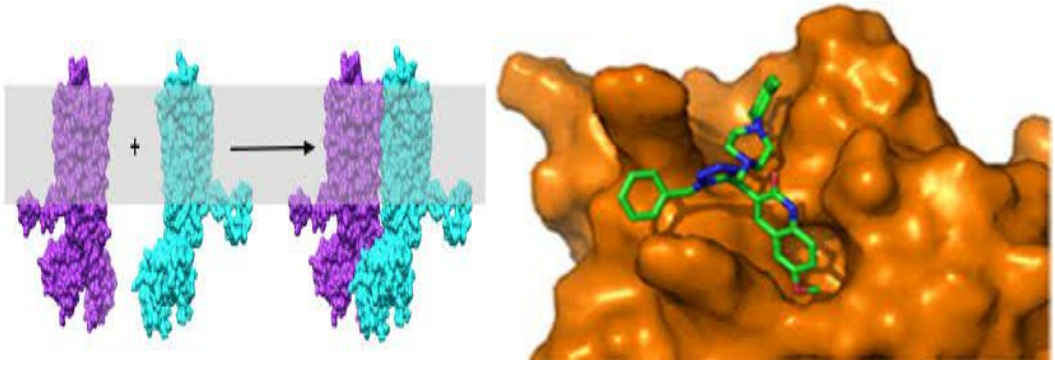


Figure 1.5 : processus Docking/Scoring[22]

2.2.4. La conception des médicaments (drug design):

Drug design, parfois appelé la conception de médicaments comme rationnel ou la conception simplement rationnelle, est le procédé de l'invention de trouver de nouveaux médicaments basés sur la connaissance d'une cible biologique. [23]

Le médicament est le plus souvent une petite molécule organique qui active ou inhibe la fonction d'une molécule telle qu'une protéine, qui à son tour entraîne un bénéfice thérapeutique au patient. Dans le sens le plus fondamental, la conception de médicaments implique la conception de molécules qui sont complémentaires dans la forme et la charge à la cible biomoléculaire avec lequel

ils interagissent et donc vont se lier à elle. Drug design repose souvent, mais pas nécessairement sur les techniques de modélisation informatique. [24] Ce type de modélisation est souvent désigné comme la conception de médicaments assistée par ordinateur. Enfin, la conception de médicaments qui repose sur la connaissance de la structure tridimensionnelle de la cible biomoléculaire est connue comme la conception de médicaments basée sur la structure. [24] En plus de petites molécules, les produits biopharmaceutiques et en particulier des anticorps thérapeutiques sont une classe de plus en plus importante de médicaments et les méthodes de calcul pour améliorer l'affinité, la sélectivité et la stabilité de ces protéines thérapeutiques ont également été développées. [25] et notre travail est consterné sur cette domaine de la chemoinformatique pour créer un nouveau médicament.

2.3. La théorie des graphes et les graphes moléculaires :

La chemoinformatique peut s'appuyer sur une représentation des molécules sous forme de graphes. Nous citons dans quelques définitions relatives à la théorie des graphes nécessaires pour comprendre facilement le principe du graphe moléculaire utilisé fréquemment dans la chemoinformatique pour représenter graphiquement les molécules.

2.3.1. La théorie des graphes :

La technique de la théorie des graphes est largement appliquée en informatique pour représenter une donnée structurée;

Cette théorie est fondée à partir de quelques lois simples: Un ensemble de points, dont certaines paires sont directement reliées par un ou plusieurs liens. Ces liens peuvent être orientés; dans ce cas, le graphe est dit orienté. Sinon, les liens sont symétriques, et le graphe est non orienté. Dans la littérature récente de la théorie des graphes, les points sont appelés les sommets ou les nœuds; les liens sont appelés arêtes dans les graphes non orientés (figure 1.7) et arcs dans un graphe orienté.

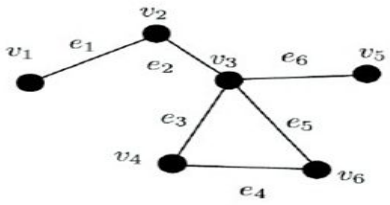


Figure 1.6 :v1…v6=les nœuds, e1…e6=les liens [42]

2.3.2. Les graphes moléculaires :

Une représentation usuelle des molécules est définie par le graphe moléculaire (figure1.8). Le graphe moléculaire est un graphe simple étiqueté et non orienté représentant la structure d'une molécule. L'ensemble des sommets encode les atomes et l'ensemble des arêtes représente les liaisons covalentes entre les atomes [26]. Chaque sommet est étiqueté par l'élément chimique de l'atome correspondant (carbone (C), oxygène (O), azote (N), chlore (Cl), etc.) et les arêtes par le type de liaison covalente (simple, double, triple ou aromatique). Dans ce graphe non orienté les atomes d'hydrogène sont exclus du graphe pour simplifier les calculs.

Les graphes moléculaires sont représentés graphiquement en utilisant plusieurs conventions

- L'élément chimique carbone (C) n'est pas explicitement représenté dans la représentation graphique du graphe moléculaire ;
- les atomes d'hydrogène ne sont pas explicitement représentés dans le graphe moléculaire. Leur présence est implicitement encodée par le degré des autres atomes
- L'étiquette d'une arête d'un graphe moléculaire est graphiquement représentée par une liaison composée par un nombre de traits encodant le type de liaison (1 trait pour une liaison simple, 2 et 3 traits pour les liaisons doubles et triples et une alternance de traits simples et doubles pour les liaisons aromatiques d'un cycle).

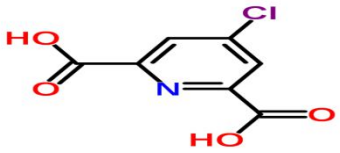


Figure 1.7 : Exemple de graphe moléculaire.

2.3.3. Descripteurs moléculaires

1) Définition :

Une molécule peut être représentée par un ensemble de valeurs numériques appelées descripteurs moléculaires (vecteur de description). Ces valeurs peuvent être obtenues expérimentalement mais le plus souvent, elles sont calculées à partir de la structure de la molécule.

Les descripteurs moléculaires jouent un rôle important pour calculer la similarité entre molécules chimiques. Cependant ils peuvent être également utilisés en complément dans des méthodes de fouille de données ou d'apprentissage automatique. Plusieurs logiciels sont disponibles pour calculer ces différents descripteurs, comme: QSARIS, VolSurf, Dragon.

Enfin des librairies informatiques écrites en Java telles que le CDK (Chemistry Development Kit) [27] ont été développées pour la chemoinformatique et la bioinformatique. Elles permettent à des utilisateurs avertis d'intégrer le calcul de leurs descripteurs à une plateforme informatique déjà existante

2) Regroupements les descripteurs :

Il est possible de classer ces différents descripteurs existants (5000 types) de différentes façons. En effet, on peut organiser les descripteurs selon leur type (quantitatif, qualitatif, structuré), selon leur dimension, ou selon la nature des propriétés décrites.

- **Type numérique/structuré du descripteur**

On distingue quatre types principaux de formes numériques que peuvent prendre les descripteurs:

a. Les variables simples

Ils peuvent prendre une valeur qualitative ou une valeur numérique dite quantitative:

- Les quantitatives sont variables ne prenant que des valeurs numériques. on distincte 2 types de variables quantitatives; les variables quantitatives continues (par exemple le poids

moléculaire) et les variables quantitatives discontinues (par exemple le nombre d'atomes d'une molécule).

- Les variables qualitatives ou symboliques sont des variables ne prenant que des valeurs non numériques appelées caractéristiques ou modalités; par exemple la variable taille peut prendre les modalités (grand, moyen, petit); la couleur peut prendre les modalités (jaune, vert, rouge).

b. Les fingerprints

La représentation du fingerprint de la structure moléculaire est une forme particulièrement complexe de descripteurs. Les fingerprints sont des vecteurs de bits (ou valeurs binaires). Chaque bit prend la valeur 0 ou 1. Cette valeur code l'absence ou la présence de certains fragments structuraux ou d'autres caractéristiques (propriété physico-chimique, etc.). Les fingerprints restent l'un des types de descripteurs les plus utilisés en chemoinformatique pour décrire les molécules et les comparer entre elles

c. Les Vecteurs et les matrices

Les descripteurs peuvent être de forme plus complexe comme des vecteurs ou des matrices. Il peut s'agir de matrices de connectivité indiquant pour chaque couple d'atomes la présence ou non d'une liaison.

d. Les graphes

Ce type de descripteurs, tels que décrits précédemment permet de représenter chaque molécule par un ensemble de nœuds et d'arêtes. Cette représentation permet d'utiliser des algorithmes de comparaison de graphes pour calculer la similarité entre les molécules chimiques.

➢ **Classement selon la dimensionnalité**

Les descripteurs moléculaires sont classés en trois catégories en fonction de leur dimension [28]:

- **1D :** descripteurs représentant diverses propriétés calculées à partir de la formule brute (e.g C6H6O pour le phénol) de la molécule (e. g. nombre d'atomes, poids moléculaire).
- **2D :** descripteurs présentant l'information structurelle pouvant être calculée à partir de la structure en deux dimensions d'une molécule (e.g. nombre de cycle de benzène).

- **3D :** descripteurs représentant l'information dérivée de la représentation en trois dimensions des molécules (e.g. surface et volume moléculaire)

Structure de la molécule	Informations	Exemple de descripteurs
1D	Formule brute : atomes présents	Masse moléculaire Présence / nombre d'un atome donné
2D	Enchaînement des atomes Type des atomes et des liaisons	Méthodes fragmentales (log P, réfractivité molaire…) Fingerprints
3D	Structure minimisée / Conformations	Surfaces Volumes Pharmacophores

Tableau 1.1 : Exemple de descripteurs en fonction de la dimensionnalité de la structure de départ [43]

2.3.4. Représentation de structures 2D

La représentation 2D (figure 1.9) décrit les atomes présentés dans une molécule, et les liaisons entre eux. En effet, les informations relatives à cette structure 2D peuvent être exprimées sous forme tabulaire appelé "table de connexion", pour pouvoir être facilement manipulée par l'ordinateur. Lorsqu'une molécule a besoin d'être représentée dans un fichier, il est possible de compresser l'information contenue dans la table de connexion en une forme plus facile à lire. Ceci, exige plusieurs formats y compris : la notation SMILES, etc.

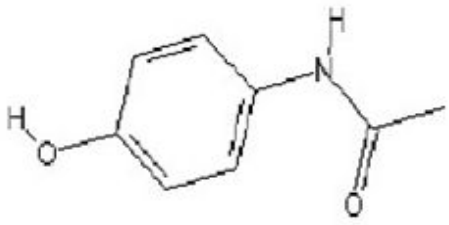


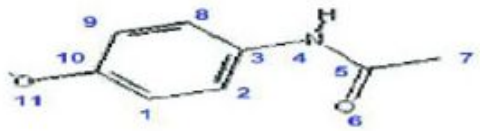
Figure 1.8 : Représentation 2D de la molécule "acétaminophène".

2.3.4.1. Les tableaux de connexion

Le tableau de connexion enregistre toutes les informations se trouvant dans la structure 2D, c.-à-d. les atomes constituant une molécule et les liens existant entre eux. Avant qu'un tableau de connexion se produise, on doit commencer

par la numérotation des atomes de la molécule pour produire un tableau de correspondance. Puis, on construit le tableau de connexion tel que, le numéro de ligne et de la colonne représente le nombre donné à l'atome. Par exemple, si une liaison existe entre les atomes 5 et 8, alors un "1" est placé à l'intersection de la ligne 5 et de la colonne 8 (ainsi, la ligne 8 et la colonne 5), sinon, un "0" est placé. En outre, nous pouvons utiliser 2 pour représenter un double liaison, 3 pour représenter un triple liaison, etc.

Cependant, puisque le tableau de connexion est symétrique par rapport à la diagonale, nous devrons stocker seulement la moitié de ce tableau dans un autre tableau appelé "tableau de connexion non redondant». Le tableau de correspondance et le tableau connexion non redondant de "l'acétaminophène" sont respectivement illustrés sur les figures 1.10 et 1.11 suivantes.



Num	Atom Type
1	C
2	C
3	C
4	N
5	C
6	O
7	C
8	C
9	C
10	C
11	O

Figure 1.9 : Tableau de correspondance de "l'acétaminophène" non redondant

	1	2	3	4	5	6	7	9	10	11	12
1											
2	1										
3	0	2									
4	0	0	1								
5	0	0	0	1							
6	0	0	0	0	2						
7	0	0	0	0	1	0					
9	0	0	1	0	0	0	0				
10	0	0	0	0	0	0	0	2			
11	2	0	0	0	0	0	0	0	1		
12	0	0	0	0	0	0	0	0	0	1	

Figure 1.10 : Tableau de connexion

2.3.4.2. Les formats de fichiers :

Le traitement des informations chimiques à donné lieu à de très nombreux formats de représentation de molécules. Du fait que ces différents travaux ont été conduits sans prédéfinir une norme standard, plusieurs formats coexistent. Parmi eux, les formats les plus populaires sont: SDF et MOL proposé par MDL [29], SMILES [30], etc. Les formats MDL (sdf ou mol) contiennent les coordonnées cartésiennes de chaque atome, et éventuellement leur charge partielle (mol2), ainsi que toutes les liaisons formant la molécule.

Le format SMILES permet la représentation d'une molécule par une chaine de caractères, symbolisée par un enchainement d'atomes et de liaisons.

A. Les formats MDL

L'information chimique est classiquement stockée dans différents types de format. Une des normes de l'industrie les plus largement utilisées sont les formats de fichiers de table chimiques, Ils sont communément appelés MDL MOL (parfois molfile) pour les molécules simples, et MDL SDF (parfois fichier SDfile) pour les molécules et les données multiples.

Les fichiers de données MDL (Molecular Design Limited) sont des fichiers texte qui adhèrent à un format strict pour représenter des enregistrements multiples de la structure chimique et les champs de données associés. Ceux-ci sont acceptés par de nombreux outils, y compris JChemPaint,Jmol, etc.[31]

i. Le format de fichier MDL MOL

Pour l'acquisition des données chimiques, les molécules et les fragments devront être en format ".mol". Les fichiers MDL MOL sont généralement classés comme des fichiers de données qui contiennent des informations sur les données moléculaires, atome, les coordonnées et informations de connectivité au format texte brut. En termes plus techniques, le molfile se compose de trois informations d'en-tête de ligne, info atome dans le tableau de connexion, les connexions des obligations et des types et des sections.

ii. Le format de fichier MDL SDF

Le format MDL SDF est une extension de fichier MDL MOL. L'une des caractéristiques principales du format SDF est sa capacité à inclure les données associées.

B. Le format SMILES

Les notations de lignes convertissent des tableaux de connexion de la structure chimique en une chaîne de caractères, et une séquence de lettres, en utilisant un ensemble de règles. Cependant, la plus élégante notation de la ligne moléculaire est SMILES (Simplified Molecular Input Line Entry Specification SMILES) [32]. Le code SMILES permet de mémoriser les molécules et d'illustrer les différentes données en sortie des méthodes de calcul.

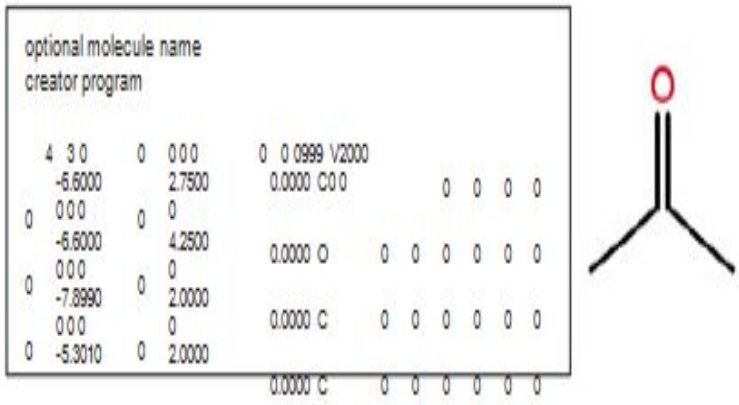
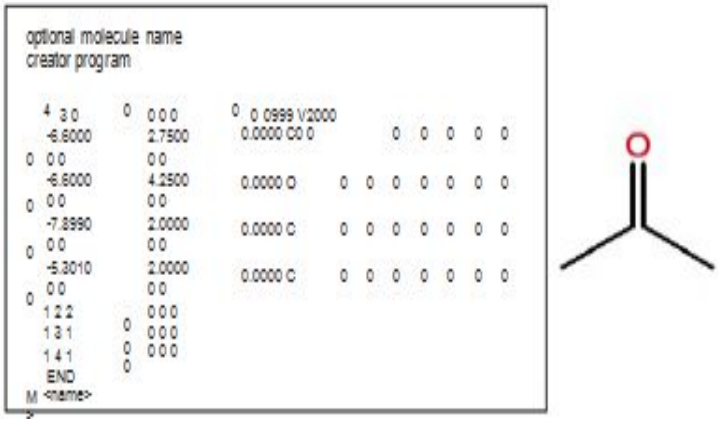
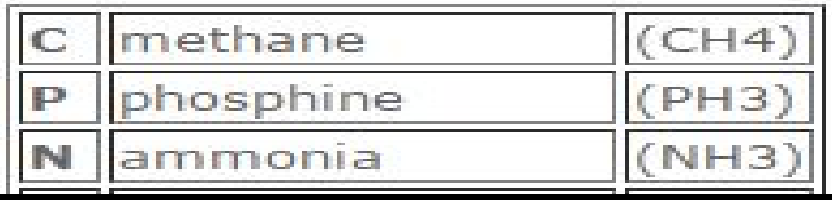
L'avantage principal de ce format est qu'il est léger, permettant le stockage de bases de molécules de grande taille dans un espace mémoire restreint.

- **Règles de représentation SMILE :**
  - Les atomes d'hydrogène ne sont pas représentés. Autres atomes sont

représentés par leurs symboles atomiques, généralement en majuscules B, C, N, O, F, P, S, Cl, Br et I.

- Les liaisons sont représentées par les '=','#' et ':' pour, respectivement doubles, triples, et aromatiques liaisons.
- les cycles sont représentés par un nombre commun suivant les atomes qui relient pour former le cycle.
- Les branches sont codées par des parenthèses.
- Enfin, il est nécessaire pour une déconnexion à coder explicitement avec le point final ou le caractère ('.').

Dans le tableau suivant certains exemples sur chaque type de format

Format	Examples	
MDL format	**mol (molecular)**	
	Sdf (Structure-Data-Format)	
SMILES format	**Atomes**	

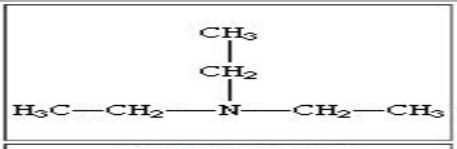
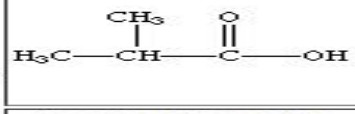
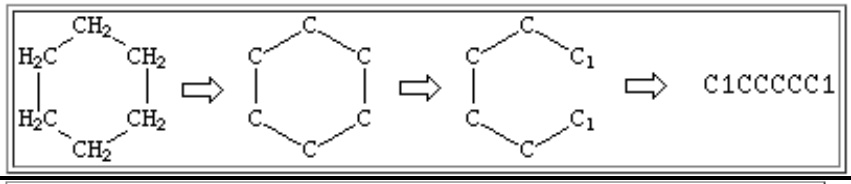
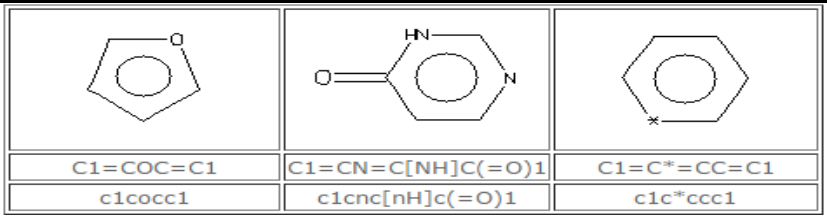
	Liaisons	CC ethane (CH3CH3) C=O formaldehyde (CH2O) C=C ethene (CH2=CH2)
	Ramifications	 CCN(CC)CC **Triethylamine**  CC(C)C(=O)O **Isobutyric acid**
	Cycles	 C1CCCCC1
	Aromaticité	 C1=COC=C1 c1cocc1 C1=CN=C[NH]C(=O)1 c1cnc[nH]c(=O)1 C1=C*=CC=C1 c1c*ccc1

Tableau 1.2 : Les formats utilisés en chemoinformatique avec des exemples

2.4. La notion « drug like » et regle de 5 :

- **« Drug like » :**

Il est devenu indispensable d'éliminer les mauvais composés ayant les propriétés physico-chimiques les moins similaires avec les médicaments disponibles sur le marché afin de réduire les coûts de développement de médicaments, pour cela de nombreuses études ont été menées sur les propriétés physico-chimiques des médicaments ,Un contributeur majeur dans le domaine de la caractérisation de composés « drug-like » est Lipinski avec la « règle des 5 » [33]. Cette règle est la plus utilisée pour l'identification des composés « drug-like » [34]. D'après cette règle, les composés ne validant pas au moins deux des critères suivants ont de très fortes chances d'avoir des problèmes d'absorption ou de perméabilité :

- masse moléculaire ≤ 500 Da
- log P ≤ 5
- accepteurs de liaisons H ≤ 10
- donneurs de liaisons H ≤ 5

La « règle des 5 » a été mise au point à partir de composés administrables par voie orale. Ce n'est donc pas une méthode pour distinguer les composés étant potentiellement des médicaments de ceux n'en étant pas, mais plutôt une méthode pour identifier les composés ayant une faible absorption ou une faible perméabilité.

3. Grands Défis pour Chemoinformatique :

Il y a trois domaines "grand défi". Ils devraient être un objectif important pour chemoinformatique.

3.1. Surmonter étals la découverte de médicaments

Après les succès impressionnants dans la découverte de médicaments vers la fin du siècle dernier, la productivité dans l'industrie pharmaceutique a diminué en charges ont augmenté. Chemoinformatique peut aider en permettant, des expériences virtuelles bon marché rapide de prioriser les expériences réelles. En plus de la recherche de découverte de médicaments est effectuée dans les universités, les instituts et les petites entreprises, et les solutions, il faudra des pièces de chemoinformatique, la bioinformatique et d'autres disciplines, les connaissances de la chemoinformatique et les outils devraient être aussi largement que possible.

3.2. La chimie verte et le réchauffement climatique

Le réchauffement climatique et la préservation de l'environnement sera l'un des plus grands défis pour l'humanité ce siècle. Fondamentale à ce sera de trouver des produits chimiques qui sont moins polluantes ou moins toxiques pour l'environnement, ou l'amélioration de l'utilisation chimique pour minimiser l'impact sur l'environnement (par exemple dans la pétrochimie). Chemoinformatique a déjà beaucoup à offrir grâce à la toxicologie computationnelle et la modélisation prédictive.

3.3. Compréhension la vie du point de vue chimique

Produits chimiques sont jugés de plus en plus important dans les fonctions cellulaires, par exemple par le biais de petits modulateurs de molécules et épigénétique. Cela a conduit à des domaines tels que la biologie chimique, et plus

récemment les systèmes de chimie (Ludlow et Otto, 2008) et de la biologie des systèmes chimiques (Oprea et al. 2007), qui cherchent à comprendre les systèmes biologiques à partir d'un point de vue de la chimie. Intégration des méthodes de la Cheminformatique et la bioinformatique sera clé. [35]

4. La propriété ADME/TOX :

C'est la propriété favorable du médicament, la majorité des médicaments a rejeté a cause de ce propriété ce figure la représente les causes de rejet des médicaments :

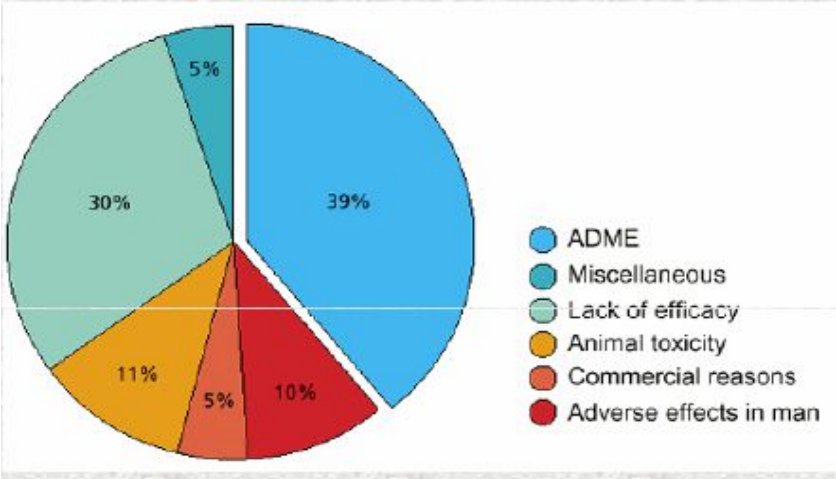


Figure 1. 11 : les raisons de rejet des médicaments. [44]

4.1. Définition :

ADME est une abréviation de la pharmacocinétique et la pharmacologie pour "absorption, distribution, métabolisme et l'excrétion" et décrit la disposition d'un composé pharmaceutique dans un organisme. Les quatre critères ont tous une influence aux niveaux de la drogue et de la cinétique de l'exposition au médicament dans les tissus et donc influencer la performance et de l'activité pharmacologique du composé comme un médicament. [36]

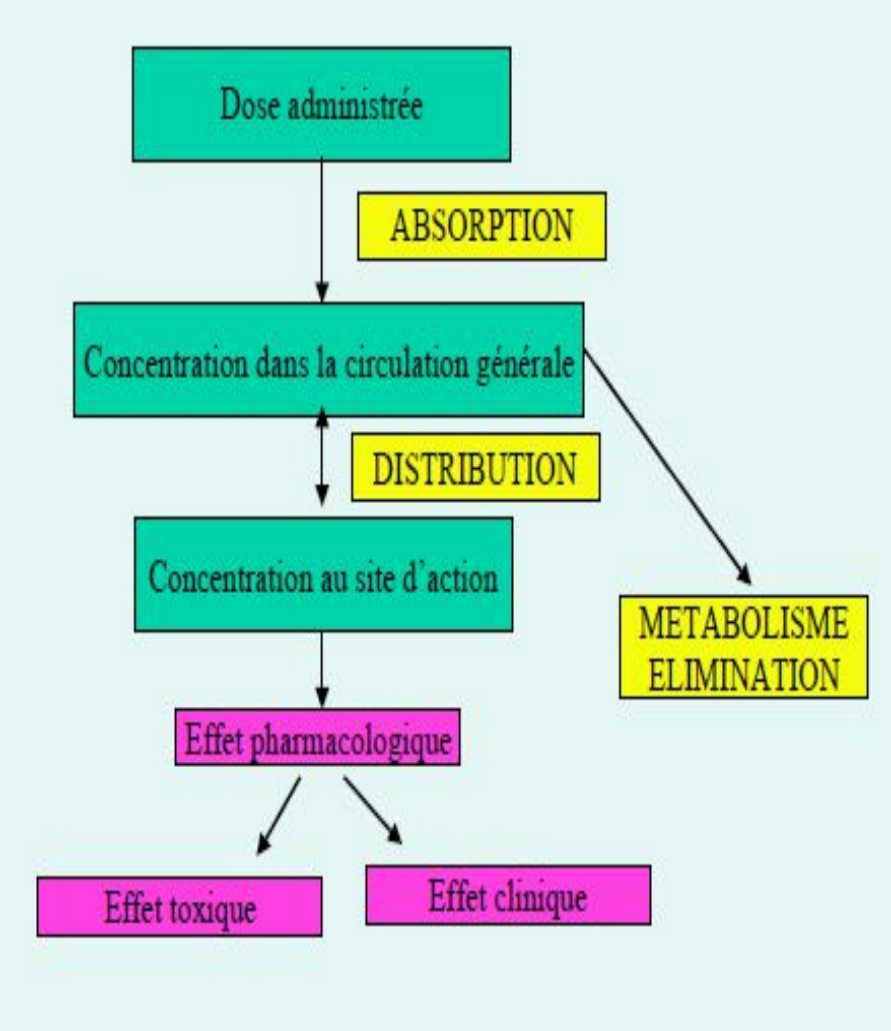


Figure 1.12: la propriété ADME/TOX. [44]

4.2. Les propriétés :

A. Absorption:

Détermine la biodisponibilité d'un compose. Un compose doit généralement franchir un certain nombre de barrières pour atteindre le sang principal vecteur de diffusion. Ce phénomène fait référence au concept de la biodisponibilité orale (Oral Bio-Availability ,OBA) [37]

B. Distribution:

Pour être véhiculée dans l'organisme, une substance peut se fixer à différentes protéines plasmatiques. Elle peut également interagir avec différentes molécules et s'y accumuler. Il y a une compétition entre diffusion et accumulation. [37]

C. Métabolisme :

Le métabolisme est une biotransformation (modification de la structure chimique) de la molécule par des réactions enzymatiques. Après l'absorption et la distribution dans l'organisme, les médicaments peuvent être :

– éliminés totalement ou partiellement sous forme inchangée, (ex : aminosides)

– ou subir des transformations enzymatiques

Le métabolisme peut se fait surtout au niveau **du foie** +++ (poumon, rein, intestin...)

L'objectif de métabolisme est de rendre les molécules plus hydrosolubles donc plus facilement éliminables par voie urinaire,La vois d'élimination peut être :

– **Rénale** : principale voie d'élimination

– **Biliaire : « cycle entéro-hépatique »**

– Réabsorption dans l'intestin du médicament excrété par la bile

– Autres voies : respiratoire, mammaire, salive, sperme, cheveux…etc [37]

D. Excrétion :

Les composés et de leurs métabolites ont besoin d'être éliminé de l'organisme par excrétion, habituellement par les reins (urine) ou dans les fèces. Sauf l'excrétion est terminée, l'accumulation de substances étrangères peut affecter le métabolisme normal. Il existe trois principaux sites où se produit excrétion des médicaments. Le rein est le site le plus important et il est où les produits sont excrétés dans l'urine. L'excrétion biliaire ou l'excrétion fécale est le processus qui initie dans le foie et passe au travers de l'intestin jusqu'à ce que les produits soient finalement excrétés avec les matières fécales des produits ou des déchets.

La dernière méthode principale d'excrétion est par les poumons (par exemple, les gaz anesthésiques). L'excrétion des médicaments par le rein implique 3 mécanismes principaux: la filtration glomérulaire de médicament non lié. Sécrétion active de (libre et liée aux protéines) médicaments par les transporteurs (par exemple des anions tels que l'urate, de la pénicilline, glucuronide, conjugués sulfate) ou des cations tels que la choline, l'histamine. Le filtrat concentré 100 fois dans les tubules pour un gradient de concentration favorable de sorte qu'il peut être sécrété par diffusion passive et sortit par l'urine. [37]

E. Toxicité :

Parfois, la toxicité potentielle ou réelle du composé est prise en compte (ADME-Tox ou ADMET). Lorsque la Libération de la substance (de revêtement de protection, ou d'autres excipients) est considérée, nous parlons de LADME. Computationnel chimistes tentent de prédire les qualités ADME-Tox de composés par des méthodes comme QSPR ou QSAR. La voie d'administration influe critique ADME. [37]

CONCLUSION :

Dans ce chapitre nous présentons un nouvelle discipline qui apparus dans ces dernière décennies c'est la chemoinformatique, nous avons présenté un bref survol historique de cette nouvelle discipline, ses principaux objectifs, ainsi que ses différentes applications dans les différents domaines. Nous avons aussi présenté dans ce chapitre, les notions de base sur lesquelles repose la chemoinformatique (espace chemique ,le graphe moléculaire, les descripteurs,…etc.). Différents classements des descripteurs ont été discutés, notamment ceux en rapport avec les descripteurs sur les propriétés physico-chimiques et les descripteurs numériques (fingerprints) que nous les considérons comme des méthodes viables pour répondre de manière précise à notre problématique (sélectionner les composés "drug-like"). En revanche, lors de notre recherche de médicaments, un filtrage doit être utilisé pour éliminer les mauvais candidats-médicaments ayant les propriétés physico-chimiques les moins similaires avec les médicaments disponibles sur le marché. Ce filtrage se fait grâce aux règles de Lipinski, nécessaire pour la sélection de composés "drug-like" . Ceci exige l'utilisation d'un descripteur numérique (fingerprints) pour représenter les caractéristiques des molécules, nous nous intéressons maintenant au traitement du problème de conception de médicaments « Drug design» par des approches computationnelles qui est l'objectif de notre travail. Cette approche de conception rationnelle de médicaments est l'un des domaines de recherche de la chemoinformatique qui couvre, à ce jour une large gamme d'applications.

Chapitre 2

Le drug design

Introduction :

Le processus de découverte de médicaments est fondamentalement un patient à vocation scientifique, où les chercheurs cherchent à améliorer les médicaments existants ou d'inventer une toute nouvelle entité chimique, qui devrait être idéalement plus puissant que n'importe quel médicament existant d'une catégorie similaire. Si non, alors au moins il devrait être plus sûr que ceux existant. Ce processus prend un temps très long et il est très coûteux la figure ci-dessus représente les étapes de ce processus :

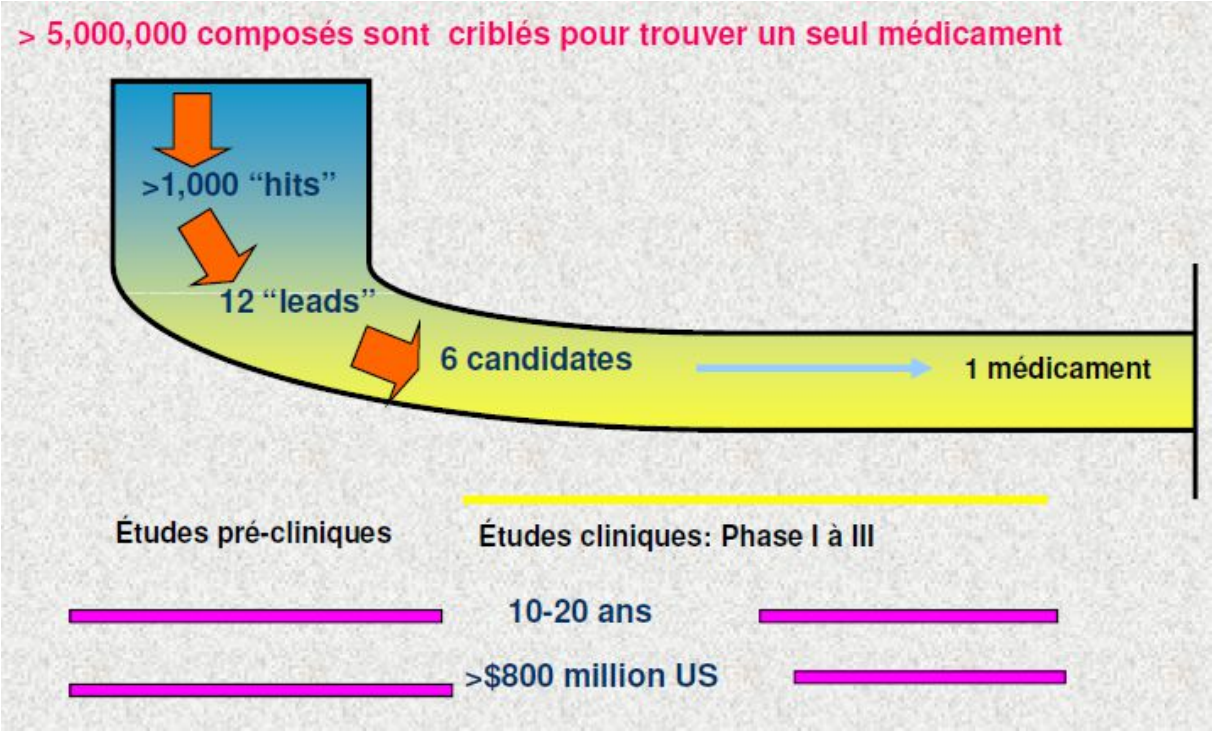


Figure 2.1: le processus de conception de nouveau médicament (temps et cout) [3]

1. Définition drug design:

Drug design, parfois appelé la conception de médicaments comme rationnel (rational drug design) ou conception simplement rationnel, est le procédé de l'invention pour trouver de nouveaux médicaments. [1]

Dans le sens le plus fondamental, la conception de médicaments implique la conception des molécules qui sont complémentaires dans la forme et la charge à la cible biomoléculaire avec laquel ils interagissent, et donc vont se lier à elle. Drug design repose souvent, mais pas nécessairement sur les techniques de modélisation informatique. [2] Ce type de modélisation est souvent désigné comme la conception de médicaments assistée par ordinateur (Computer-aided drug design). Enfin, la conception de médicaments qui repose sur la connaissance de la

structure tridimensionnelle de la cible biomoléculaire est connue comme la conception de médicaments basée sur la structure (structure-based drug design).

Le médicament : est le plus souvent d'une petite molécule organique qui active ou inhibe la fonction d'une molécule telle qu'une protéine, qui à son tour entraîne un bénéfice thérapeutique au patient.

2. Le processus de découvert de médicament (Drug discovery process) :

La figure suivante illustre les étapes de découvert d'un médicament :

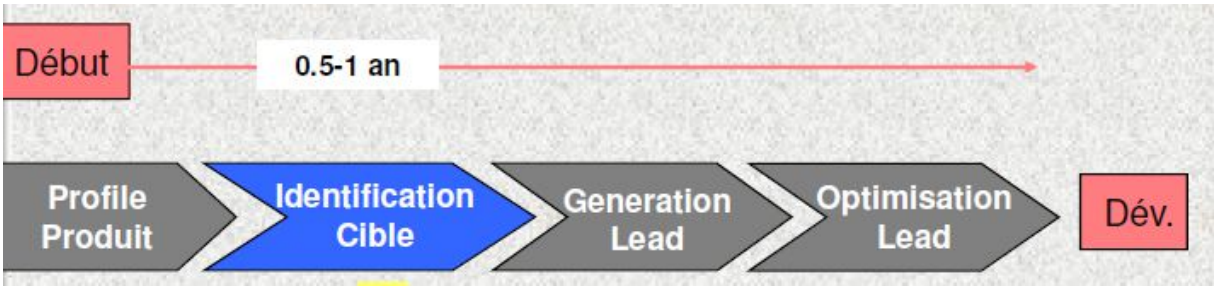


Figure 2.2 : les étapes du processus de découverte de nouveau médicament [3]

A. L'identification de cible biologique :

Il faut définir une entité biologique associé avec la maladie ou le virus, cette dernière a une grande potentielle pour le développement d'un médicament. Un cible peut provenir a partir :

- D'un mécanisme d'action de médicaments déjà connue ou de ligands naturels.
- D'une compréhension des processus cellulaires et physiologiques et/ou du mécanisme de la maladie.
- D'une mutation dans les gènes impliqués spécifiquement pour la maladie.
- Ou bien d'une façon aléatoire.

L'ensemble de médicament cible est environ 500 molécule biologique Certaines molécules sont plus aisées a cibler (plus facile pour une petite molécule de perturber une liaison d'un ligand endogène que des interactions protéines-protéines).

On va choisir un cible ou :

a) le site de liaison est bien défini.

b) un succès historique contre d'autre cible cellulaire. [3]

B. L'étape d'identification des hits:

Identifier des molécules ayant la capacité d'interagir avec la cible choisie et susceptible de moduler ses effets sur la pathologie. La sélection se fait à partir de librairies naturels, de chimiot_hèques propriétaires, ou à l'aide de technique de criblage haut débit (HTS pour Hight Throughput Screening). [3]

C. Recherche de la structure de lead:

- **qu'est ce qu'un lead :** Un composé lead dans la découverte de médicaments est un composé chimique qui a une activité pharmacologique ou biologique susceptible d'être thérapeutiquement utile (une molécule candidate pour être médicament), mais peut encore avoir une structure optimale qui doit être modifié pour répondre mieux à la cible. Sa structure chimique est utilisée comme point de départ pour les modifications chimiques afin d'améliorer l'activité, la sélectivité ou les paramètres pharmacocinétique. [5]

- **les sources de lead :** un lead peut provenir :

✓ De la nature (Plantes (fleurs, arbres, arbustes), Micro-organismes (bactérie, fungi), Animaux (grenouille, serpent, scorpion), Marin (coraux, bactérie, poisson, etc.), Biochimiques (Neurotransmetteurs, hormones)).

✓ De synthèse chimique ou combinatoire.

✓ De virtuelle rationnelle (conception assisté par ordinateur (Docking, vHTS, etc)).

D. Optimisation de lead :

Une fois la structure de la molécule est identifié, la molécule candidate passe par une série de tests afin de fournir une évaluation précoce de leurs caractéristiques ADME / Tox et de la pharmacocinétique. Ici, les chimistes en étroite collaboration avec les pharmacologues vont étudier attentivement l'activité de la structure relation et se synthétiser ces autres dérivés, de manière à obtenir un composé qui a la meilleure activité souhaitée possible. Des approches sont utilisé pour l'optimisation de lead comme Structure-Based Drug Design (SBDD), quantitative relation structure-activité (QSAR) et Drug Design assistée par ordinateur (CDAO). Toutes ces approches génèrent une énorme quantité de données, de manière à aider le chimiste dans l'optimisation pour prendre la meilleure structure possible, avec la meilleure action souhaitée possible.

E. Les études précliniques

L'objectif principal des études précliniques est de vérifier la sécurité de nouveau médicament développé. On peut jamais tester le nouveaux médicament produit sur le corps humain il faut à l' avance faire des études sur les animaux afin de tester les propriétés pharmacocinétiques et les effets de ce nouveaux médicament sur le corps. Cette phase, traite généralement avec élucider le mode d'action de la molécule et de se faire une idée de la pharmacocinétique (PK) et la pharmacodynamique (PD) de la molécule. Cependant, le plus important est les données toxicologiques obtenues à partir de l'étude fait sur les animaux, ce qui donne l'estimation approximative des effets indésirables possibles qui peuvent être susceptibles d'être vu au cours de la thérapie.

Celles-ci sont réalisées en deux étapes, des études in vitro et des études in vivo. Les études in vitro faire usage de différentes lignées cellulaires et des préparations de tissu. Les études in-vivo sont effectuées sur les animaux vivants et on observe les changements dans le comportement de l'animal.

F. Les essais cliniques

La prochaine étape après les études précliniques est l'étape des études cliniques, l'essai proprement dit de la molécule chez les volontaires humains. Cette phase permet d'évaluer l'innocuité et l'efficacité de la nouvelle molécule. Cette phase permet également de recueillir des informations sur les effets toxicologiques dans le corps humain, qui ne sont pas détecté dans l'étape précédente. Avant le début de cette étape, l'innovateur doit déposer une demande, à savoir, «Investigational New Drug (IND)", l'innovateur peuvent procéder à des études cliniques.

4 étapes pour cette étude la quatrième étape ce fait après le lacement de médicament sur le marché :

Études de phase 1 sont généralement effectuées sur des volontaires humains en bonne santé et sur un petit groupe de personnes.

Études de phase 2 : sont généralement réalisées sur une petite population de la maladie cible. Dans cette phase, l'efficacité la sécurité, le métabolisme et la pharmacocinétique du médicament sont évaluées sur le corps humain.

Études de phase 3 : sont des études approfondies et multiples sites. Cette phase, couvre un grand groupe de personnes atteintes de maladie cible. Cette phase est essentiellement une phase de confirmation thérapeutique, comme tous les paramètres

étudiés dans la phase 2 de l'étude sont confirmés dans cette phase. **Études de phase 4** :(de survillienne post-commercialisation) sont effectuées, après que le médicament a été lancé sur le marché. La société poursuit sa surveillance du médicament. Le raisonnement derrière cette phase est de vérifier toute nouvelle réaction indésirable grave ou qui n'était pas détectée dans les phases antérieures et qui peut être observé dans cette phase. S'il se trouve qu'une réaction indésirable grave est observée, la société peut retirer le médicament du marché.

3. Les types de drug design :

3.1. Cibles médicamenteuses (Drug targets) :

Une cible biomoléculaire (le plus souvent une protéine ou un acide nucléique) est une molécule clé impliquée dans un métabolisme particulier qui est associée à un état pathologique ou d'une pathologie spécifique ou à la survie de l'ineffectivité d'un agent pathogène microbien. Les cibles potentielles de médicaments ne sont pas nécessairement des pathogènes, mais doivent par définition être modificateur de la maladie. [5] Dans certains cas, les petites molécules seront conçues pour améliorer ou inhiber la fonction de cible dans la maladie spécifique. Les petites molécules conçues (par exemples les agonistes des récepteurs, des antagonistes, des agonistes inverses, des modulateurs, des activateurs d'enzymes, des inhibiteurs, des ouvertures des canaux ioniques ou des inhibiteurs) [6] sont complémentaires au site de liaison de la cible [7], ces derniers peuvent être conçues de manière à ne pas affecter d'autres molécules importantes "hors cible" (souvent désignés comme anti cibles) depuis les interactions médicamenteuses avec les molécules hors-cible peut conduire à des effets secondaires indésirables.

3.2. Drug design rationnelle (rational drug design) :

Contrairement aux méthodes traditionnelles de la découverte de médicaments, qui reposent sur des tests d'essai-erreur de substances chimiques sur des cellules ou des animaux d'élevage, et correspondant à des effets apparents aux traitements, conception rationnelle de médicaments (également appelé la pharmacologie inverse) commence par une hypothèse que la modulation d'une cible biologique spécifique peut avoir une valeur thérapeutique.

4. La conception de médicament assisté par ordinateur (computer_aided drug design) :

La conception de médicaments assistée par ordinateur utilise la chimie computationnelle pour découvrir, améliorer ou étudier les médicaments biologiquement connexes et les molécules actives.

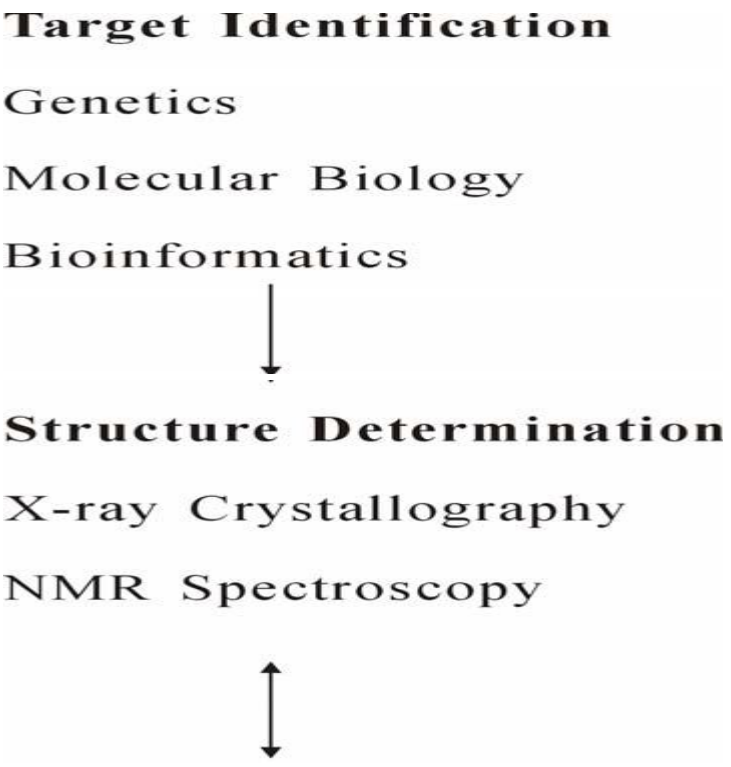
4.1. Objectif :

L'objectif le plus fondamental est de prédire si une molécule donnée va se lier à une cible et si oui, comment fortement.

La mécanique moléculaire ou la dynamique moléculaire sont les plus souvent utilisés pour prédire la conformation de la petite molécule conçue avec une cible biologique.

Cette offre de prédiction semi-quantitative de l'affinité de liaison. En outre, la fonction de notation basée sur la connaissance peut être utilisée pour fournir des estimations de l'affinité de liaison. [12]

4.2. Comment le CADD travaille:



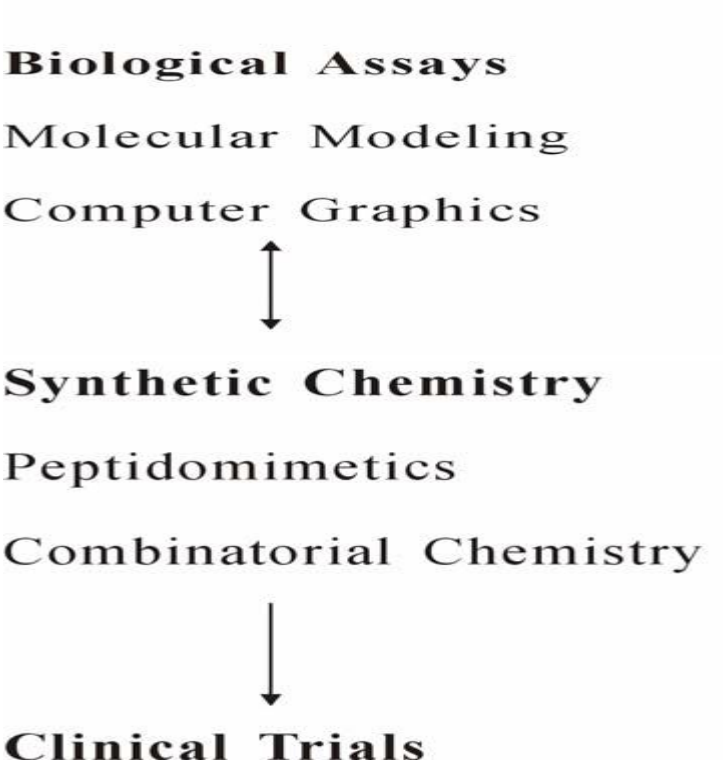


Figure 2.3: le mode de fonctionnement de computer aided drug design [13]

4.3. La conception de médicament assisté par ordinateur et le processus drug descovry :

La figure suivant représente les étapes de drug descovry qui utilise le CADD

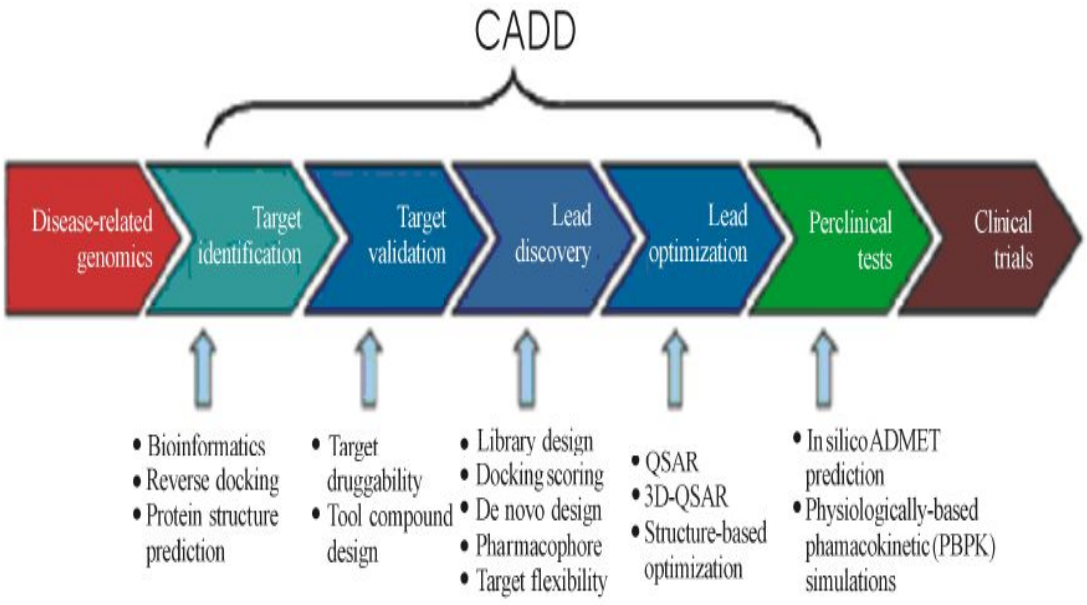


Figure 2.4: le CADD et le drug descovry[15]

La conception de médicament assisté par ordinateurs peut être utilisée à l'une des étapes suivantes de la découverte de médicament:

1. Identification de succès en utilisant le criblage virtuel (la conception structure-based ou ligand-based)

2. L'optimisation hit-to-lead de l'affinité et la sélectivité (conception basée sur la structure, QSAR, etc.)

3. l'optimisation de lead: l'optimisation d'autres propriétés pharmaceutiques tout en maintenant l'affinité afin de remédier à l'insuffisance de la prédiction calculée par la fonction de score de liaison, l'interaction protéine-ligand et composé et les informations sur la structure 3D de composé sont utilisées pour analyse. [10]

5. les approches de drug design :

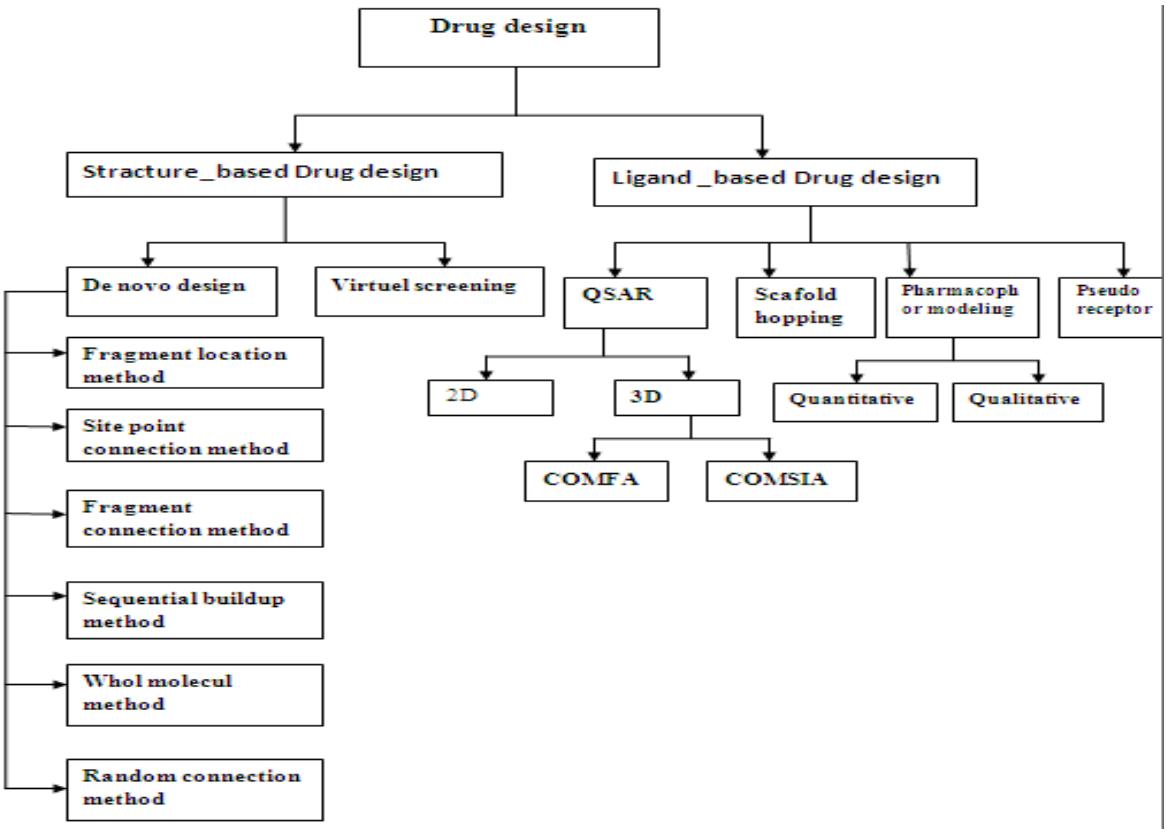


Figure 2.5: les approches de Drug design [23]

Il existe deux grands types de conception de médicaments. La première est appelée conception de médicaments basée ligand (legand_based) et la second, la conception de médicaments basée structure (structure-based).

5.1.Structure_based drug design :

5.1.1. Les méthodes de conception de médicament basé sur la structure :

On peut deviser les méthodes de conception des médicaments en trois catégorie :

1) **Le criblage virtuel :** c'est l'identification de nouveau ligand pour un récepteur donné en recherchant dans des grandes bases de données de structures des petites molécules pour trouver le site de liaison de récepteur en utilisant des programmes d'accueils approximatifs rapides.
2) **la conception de novo de nouveau médicament :** Dans ce procédé, des molécules ligands sont construits en respectant les contraintes de site de liaison par assemblage de petits morceaux d'une manière pas à pas. Ces pièces peuvent être soit des atomes individuels ou des fragments moléculaires. Le principal avantage d'un tel procédé est que des nouvelles structures, non contenus dans une base de données, peuvent être suggérées. [12] [13] [14]
3) l'optimisation de ligand connu en évaluant analogues proposées au sein de la cavité de liaison.

5.1.2. Le processus de structure_based drug design :

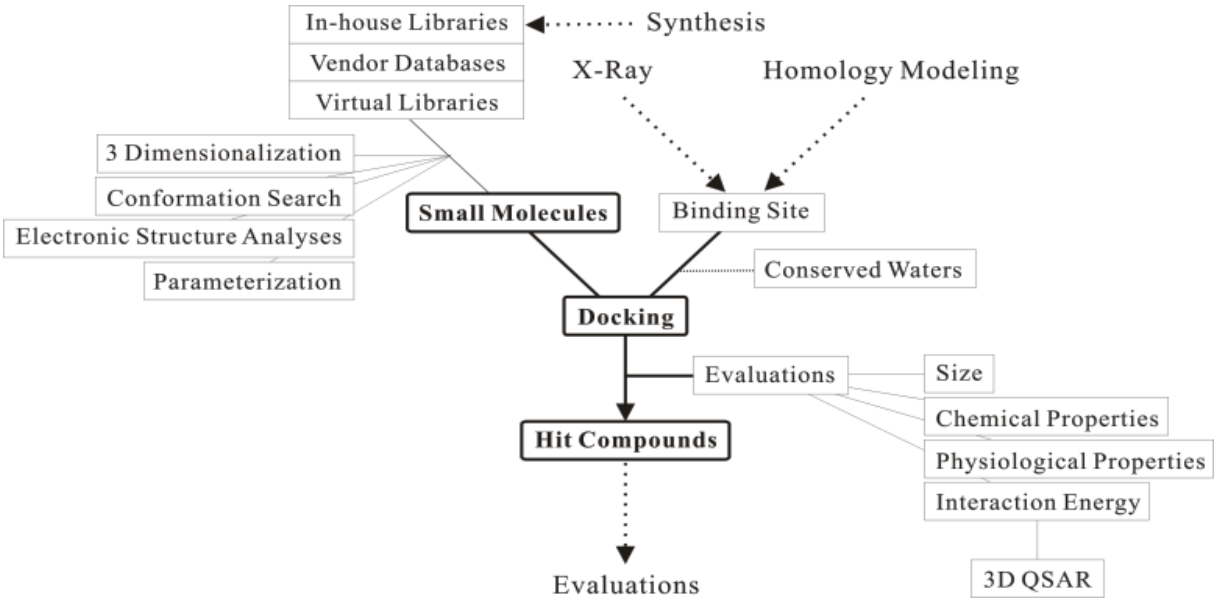


Figure 2.6: le diagramme de processus structred_based Drug design [25]

a. Choix d'un objectif de drogue

Dans la conception de médicaments basée sur la structure, il faut choisir a partir d'une base de donné biologique ou chimique une molécule cible idéale qui se lie fortement a la maladie afin d'exécuter une fonction.

b. L'évaluation d'une structure pour structure-Based Drug Design :

Une fois que la cible a été identifiée, il est nécessaire d'obtenir des informations structurales précises. Il existe trois principales méthodes pour la détermination de la structure qui sont utiles pour la conception de médicaments: cristallographie aux rayons X, NMR.

c. Déterminer la structure à partir de la cristallographie ou NMR :

Le NMR est une technique capable de fournir des structures à très haute résolution qui sont nécessaires pour déterminer le niveau atomique précise une description des sites de liaison d'un ligand.

* Les choses se cristallisent souvent mieux en présence du ligand à la suite de la stabilité accrue (moins les régions de disquettes).

* Une fois que des techniques de cristallisation ont été élaborés pour un complexe, ils devraient être assez similaires complexes subséquents.

* Bien adapté pour l'étude des petits échantillons des molécules qui ont été présélectionnés par une méthode précédente mais pas vraiment adapté pour des projections de la bibliothèque. **[21]**

d. Le Docking et le scoring :

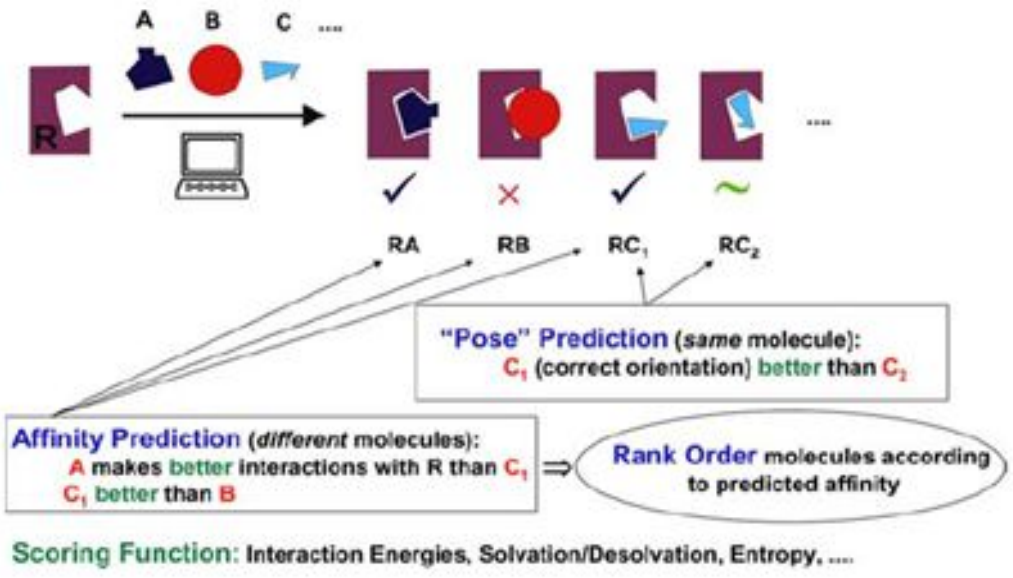


Figure 2.7 : le docking et le scoring .Le R représente la structure réceptrice A, B, et C représente les petites molécules qui se lient avec le récepteur. [21]

1) Définition de docking :

Le docking est une méthode qui prédit l'orientation d'une molécule par rapport à une autre pour avoir le complexe le plus stable. Il est fréquemment utilisé sur l'étude de la cible moléculaire des médicaments et réduire les essais expérimentaux.

Il existe deux approches de docking moléculaire, l'approche Basée sur la complémentarité des surfaces, et l'approche basée sur le calcul de l'énergie du complexe.

Le docking moléculaire est l'une des méthodes les plus fréquemment utilisées dans la conception de médicaments basée sur la structure, en raison de sa capacité à prédire la conformation de liaison des petites molécules ligands au site de liaison ciblent appropriée.

La caractérisation du comportement de liaison joue un rôle important dans la conception rationnelle des médicaments, ainsi que d'élucider les processus biochimiques fondamentaux. [48]

2) Les mécaniques de docking :

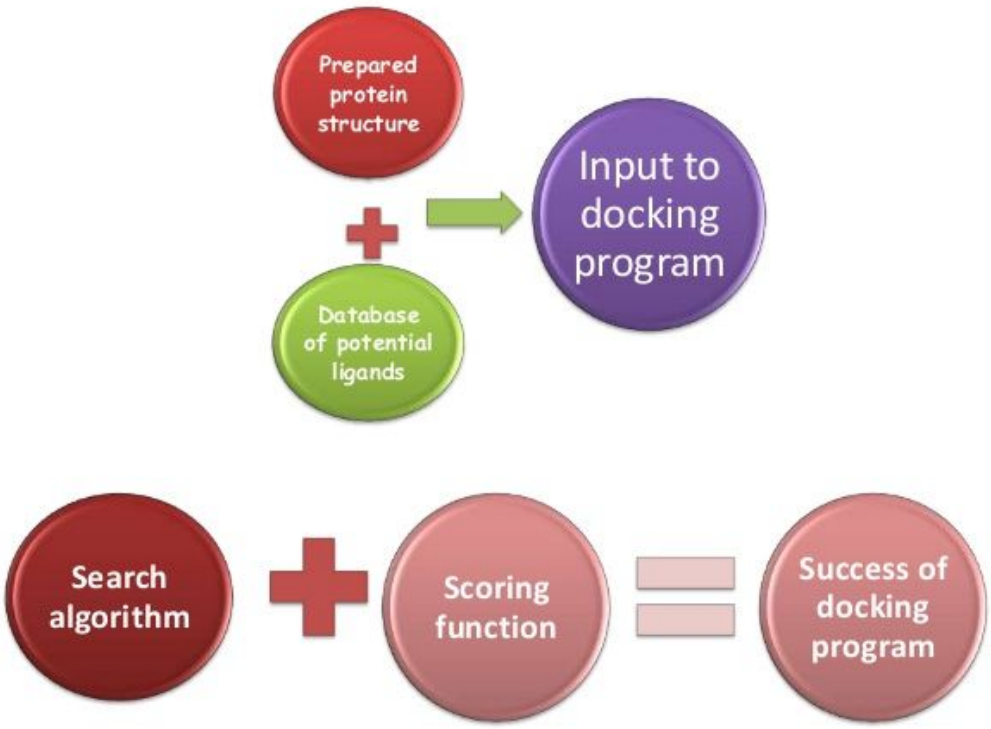


Figure 2.8 : les mécanismes de docking [42]

3) Les étapes de Docking :

Le docking se dévisse en 3 étapes principales, la caractérisation du site actif, le positionnement du ligand dans le site actif, et l'évaluation des interactions entre le ligand et la protéine réceptrice (définition d'un score).

I. caractérisation de site actif :

Il existe 2 méthodes de Docking, la première consiste à utiliser les cavités de la protéine pour définir une image négative du site actif, puis un jeu de sphères se superposant (DOCK). La deuxième c'est la définition des descripteurs ensuite recherchés à la surface de la protéine, il existe deux types de descripteurs les descripteurs chimiques (comme l'aromatique) ou physico-chimiques (hydrophobicité, potentiel électrostatique).

II. positionnement du ligand dans le site actif :

Le ligand peut être définit comme rigide ou flexible, dans le cas générale le récepteur est toujours définit comme rigide, les premiers approches considère toujours que le récepteur est rigide.

III. Reconstruction de ligand dans le site actif :

Il faut à l' avance identifier un ou plusieurs fragments significatifs dans le ligand, puis de les Placer dans le site actif en essayant de maximiser les contacts favorables, chaque orientation de ces fragments est le point de départ d'une étude conformationelle du ligand entier. **[5]**

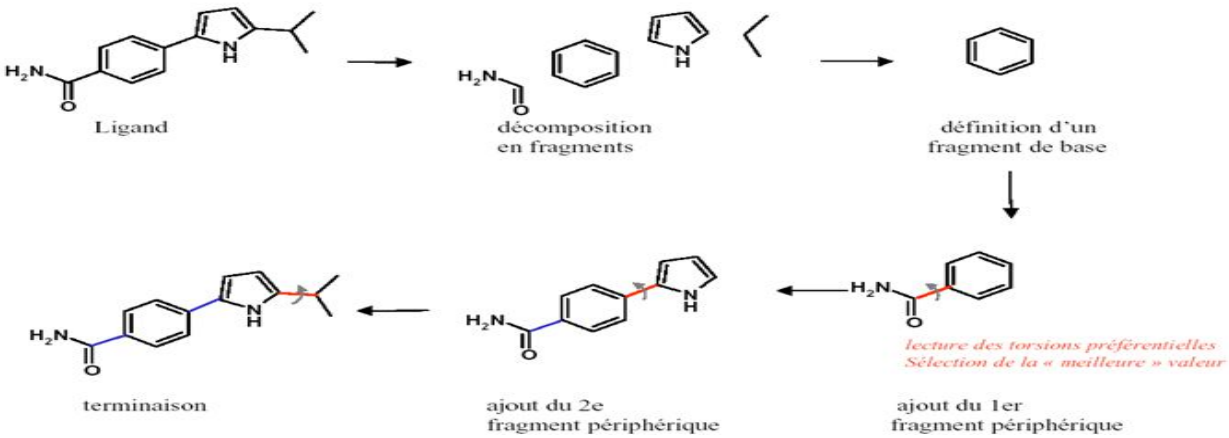


Figure 2.9: reconstruction de ligand dans le site actif [7].

IV. Le scoring ou la prédiction d'affinité :

Les fonctions de score d'affinité sont ensuite appliquées à la pose ou plus énergiquement N meilleures poses trouvées pour chaque molécule, et la comparaison

des scores d'affinité pour des molécules différentes donnent leur rang-ordre relatif. L'hypothèse implicite est que, pour une molécule donnée la meilleure pose selon le score d'affinité est parmi les n sauvé poses identifiés avec le score de dock.

Et voila la formule générale pour le scoring :

dGbind = dGint + dGsolv + dGconf + dGmotion

(dGint) : interactions ligand-récepteur spécifique

(dGsolv) :les interactions du ligand et le récepteur avec un solvant , la conformation

(dGconf) : les changements dans le ligand et le récepteur

(dGmotion) : mouvements dans la protéine et le ligand complexe au cours de la formation

5.2. ligand_based drug design :

La conception de médicaments basée Ligand ou le drug design indirect est une approche utilisée en cas d'absence des informations 3D du récepteur et elle repose sur la connaissance des molécules qui se lient à la cible biologique d'intérêt.la relation quantitative structure d'activité 3D (3D QSAR) et la modélisation de pharmacophore sont les outils les plus importants et les plus utilisés dans la conception de médicament à base de ligand. Ils peuvent fournir des modèles prédictifs appropriés pour l'identification et l'optimisation de lead [32].

5.2.1. La modélisation pharmacophore :

1) Définition :

Un pharmacophore est l'ensemble des caractéristiques stériques et électroniques qui sont nécessaires pour assurer les interactions optimales d'une molécule avec une structure cible biologique spécifique et déclencher (ou bloquer) sa réponse biologique. Un pharmacophore ne représente pas une vraie ou d'une molécule réelle association de groupes fonctionnels, mais un concept purement abstrait qui tient compte des capacités d'interactions moléculaires communs d'un groupe de composés à l'égard de leur structure cible.

2) Le mode de fonctionnement de pharmacophore :

Voila un schéma qui représente le fonctionnement de pharmacophore :

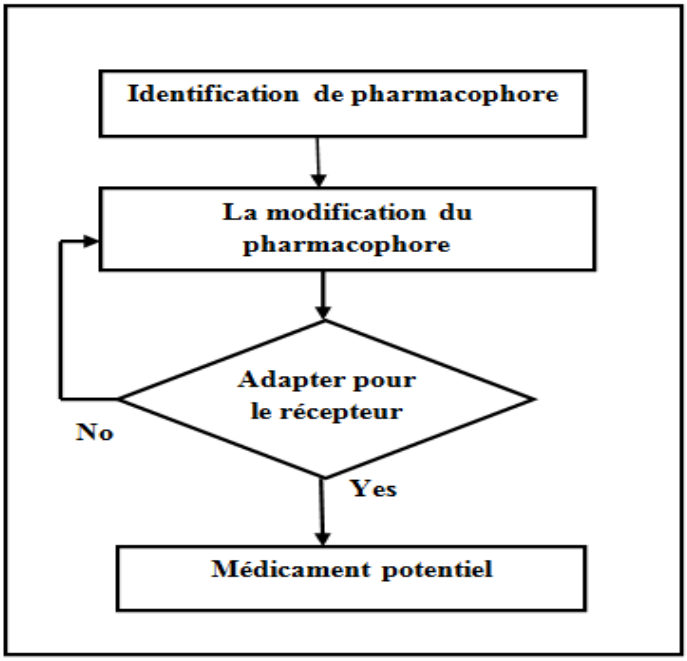


Figure 2.10 : pharmacophore [43]

5.2.2. Le modèle QSAR :

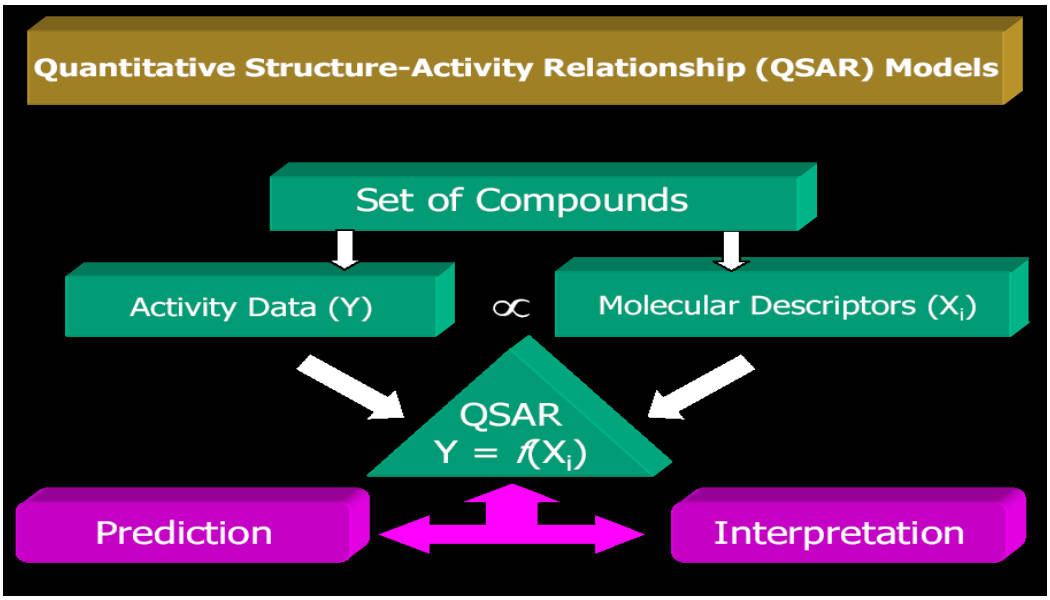


Figure 2.11: le QSAR [7]

5.2.2.1. Définition :

QSAR, qui signifie "relations quantitatives structure-activité», est une méthode qui relie la structure chimique à l'activité biologique ou chimique à l'aide de modèles mathématiques [46]. Si l'activité d'un ensemble de ligands peut être déterminée, un modèle peut être construit pour décrire cette relation. Contrairement à un modèle de pharmacophore, qui code seulement les caractéristiques essentielles d'un ligand actif, le modèle QSAR permet de déterminer l'effet d'une certaine propriété sur

l'activité d'une molécule. Par exemple, le modèle QSAR peut révéler une propriété négatif, ou encore un effet positif faible sur l'activité du ligand. Ces informations ne sont pas disponibles en utilisant un modèle de pharmacophore [47], la forme générale de ces modèles est la suivante:

Activité ou Propriété = f(Descripteurs) [4]

5.2.2.3. Les descripteurs:

1) Définition :

Un descripteur est un paramètre unique ou un ensemble de paramètre combiné dans une valeur réelle, binaire ou un vecteur. Un descripteur peut être une masse moléculaire, une masse molaire, le nombre d'atome de chlore d'iode d'oxygène ou d'autre atome, le nombre de cycles a 6, le nombre d'atomes aromatiques, le nombre de liaisons..., pKa, logP, tous ces descripteurs sont faciles à calculer et leurs valeurs sont précises et avec peu d'ambigüité (sauf logP). [4]

Figure 2.12: classification en descripteur [4]

5.2.2.4. Les paramètres de QSAR:

I. Hydrophobie(LogP) :

la reconnaissance moléculaire dépend fortement à des interactions hydrophobes entre ligands et récepteurs. Hydrophobie des solutés peut être facilement déterminée par la mesure de coefficient de partage désigné comme P. Les coefficients de partage sont additif constitutives, les propriétés liées à l'énergie libre. Log P représente le caractère hydrophobe global d'une molécule, qui comprend la somme des contributions hydrophobes de la molécule "parent" et son substituant.

II. Électronique(Ka) :

les attributs électroniques des molécules sont intimement liés à leurs réactivités chimiques et des activités biologiques. La mesure dans laquelle une réaction donnée

répond à la perturbation électronique constitue une mesure des demandes électroniques de cette réaction, qui est déterminée par un mécanisme est calculé.

III. Les effets stériques :

La quantification des effets stériques est complexe et difficile dans toutes les autres situations, en particulier au niveau moléculaire. Les stériques sont d'une importance capitale dans les interactions ligand-récepteur, ainsi que dans les phénomènes de transport dans les systèmes biologiques.

5.2.2.5. Les types de QSAR :

A. QSAR 1D :

- **QSAR 1D modèle Free-Wilson:**

Activité de dérivés du N, N diméthyle-a-Boromo phenylethylamine sur les récepteurs adrénergiques.

L'activité biologique (Act) est exprimée comme le logarithme de l'inverse d'une concentration critique C (en mol/kg chez le rat), équivalent a –log Ki ou -log IC50. [4]

$$Act = \log\frac{1}{C} = \sum a_i x_i + \mu$$

- **QSAR 1D modèle Hansch 1964 :**

Le premier à établir une relation entre l'activité biologique d'une série de composes et leurs propriétés physico-chimiques Notion de lipophilicite (LogP) et de paramètres électroniques (constante de Hammet) pour la première fois dans le modèle. Modèle non-linéaire (parabolique) [4]

$$Act = \log\frac{1}{C} = -a_i\,(\log p)^2 + a_2(\log p) + a_3\sigma + \cdots + b$$

B. QSAR 2D :

Le QSAR 2D se base sur le Principe de similarité chimique (similarité de propriétés biologiques) il existe plusieurs type de descripteurs pour ce type de QSAR.

1. **Descripteur topologiques:**

On représente usuellement les connectivités sous forme de matrice. Des lors on peut faire de nombreuse opérations pour comparer celle si.

2. **Descripteurs de charges:**

- Charge globale (avec les ions)
- Charge partielle d'un atome donné sur une structure

3. **Descripteurs fragmentaux:**

On va prendre en compte des propriétés physicochimiques de partie de la molécule. Cela revient à analyser les différences de groupes fonctionnels2 .[4]

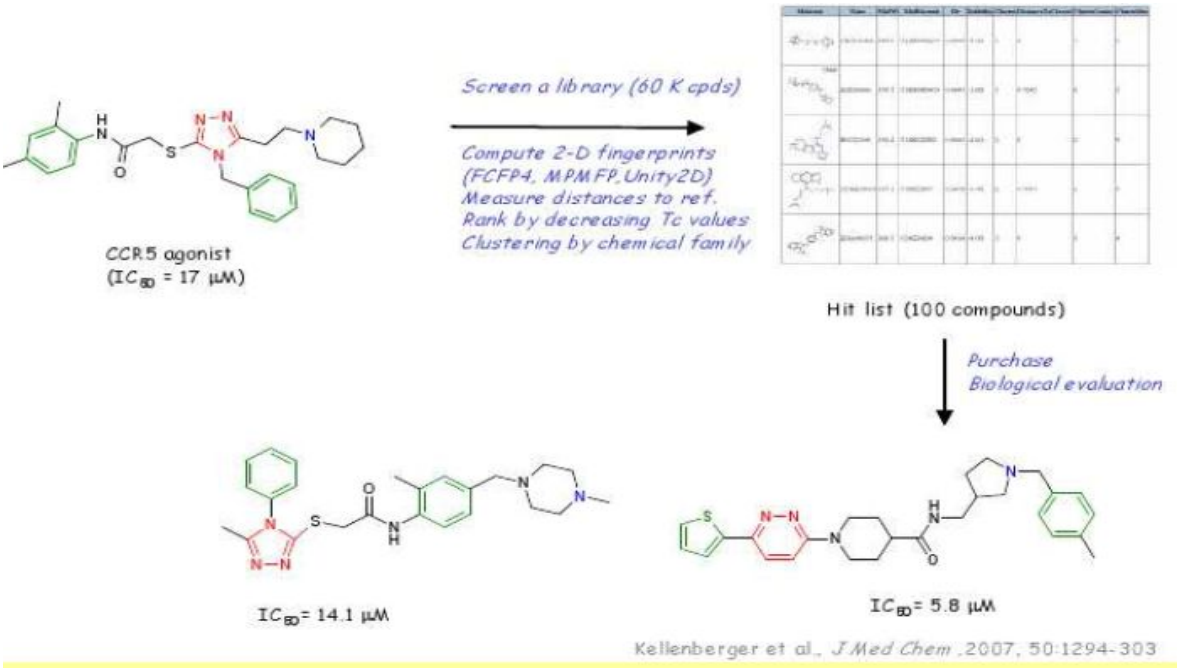


Figure 2.13: les déférents descripteurs [3]

C. QSAR 3D :

L'expression de QSAR tridimensionnelle (3D-QSAR) réfère à l'application de calculs de champs de forces nécessitant des structures tridimensionnelles, comme la cristallographie protéique. Elle utilise des potentiels calculés, et elle traite des champs stériques (forme de l'objet) et électrostatiques en fonction énergie appliquée. [44]

L'espace de données ainsi créé est ensuite habituellement réduit par une extraction de caractéristique (voir aussi réduction dimensionnelle).

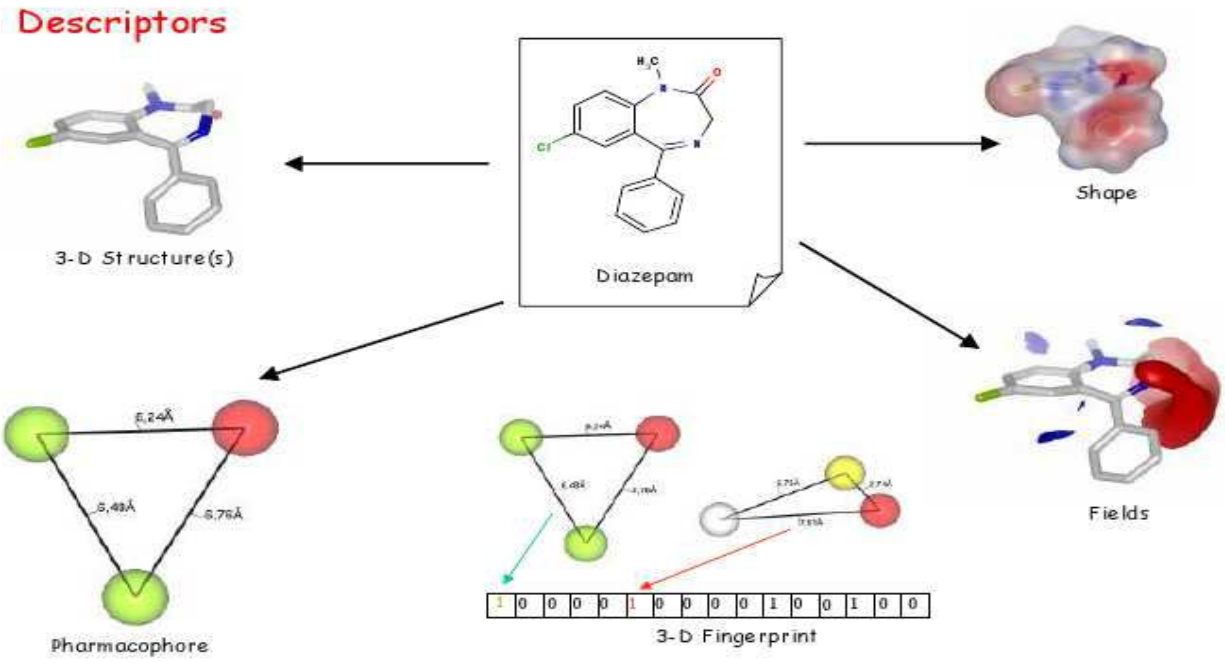


Figure 2.14 : QSR 3D et les déférents descripteurs [4]

5.3. La différence entre ligand_based et structre_based drug design :

La conception de médicament basé structure (SBDD) et la conception basé ligand (LBDD) sont des zones actives de recherche dans les deux domaines universitaires et commerciaux. Les différences entre le SBDD et le LBDD sont illustré dans le tableau si dessus :

Ligand_based drug design	Structure_based drug design
• Les informations structurelles 3D du médicament cible est pas nécessaire, mais repose sur la connaissance des petites molécules et leur puissance de liaison au médicament cible d'intérêt. • L'activité de la structure quantitative 3D relations (3D QSAR) et modélisation pharmacophore sont couramment utilisés. • Fournit des modèles prédictifs appropriés pour l'identification et l'optimisation de lead	•Les informations structurelles 3D de la cible de la drogue est une condition préalable pour le développement de son inhibiteur. •La structure de la cible est déterminée par des techniques expérimentales telles que la cristallographie aux rayons X par rapport RMN. • les méthodes de calcul comme le filetage et la modélisation d'homologie sont utilisé pour prédire la structure des protéines. •HTS et les méthodes de docking sont utilisés pour trouver des petits coups de molécules.

Tableau 2.1: le tableau qui illustre la différence entre structre_based et ligand_based drug design [37]

6.De novo drug design :

6.1. Définition :

La conception de novo est une technique utilisée pour la conception de nouveau médicament, le médicament doit être non peptidique, petite molécule ligands pour les sites de liaison macromoléculaires. [1]

« Is the design of bioactive compounds by the incremental, construction of a ligand model within the receptor or enzyme active site, the structure of which is known from X-ray or nuclear magnetic resonance (NMR) » [2]

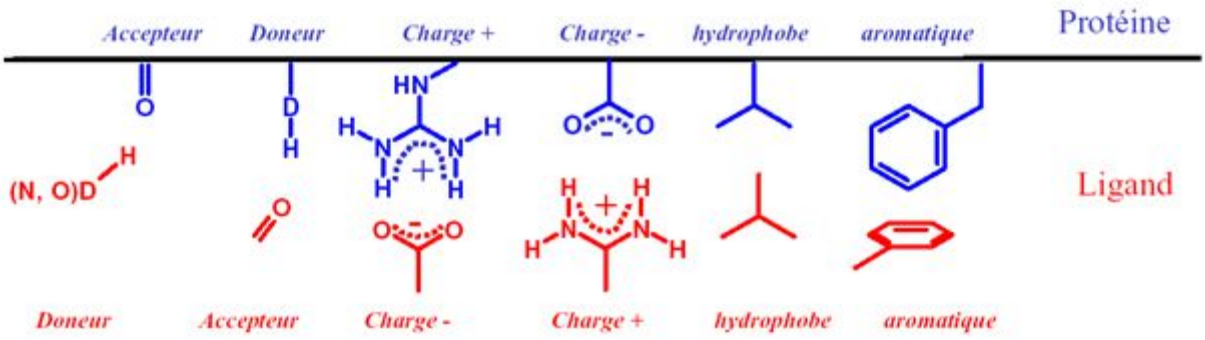


Figure 2.15: les types de complémentaire [41]

6.2. Principe:

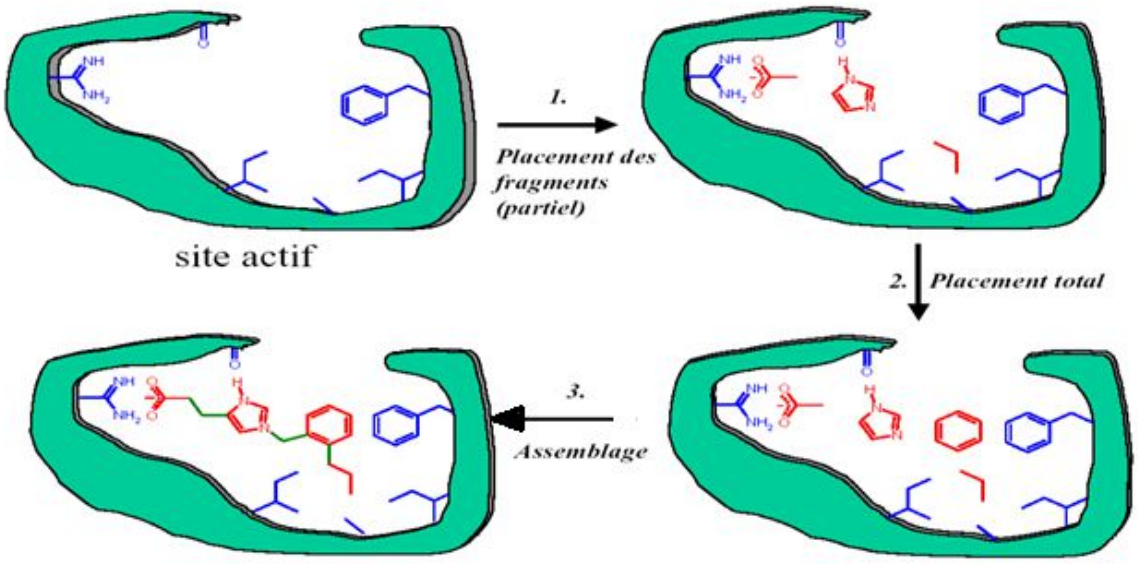


Figure 2.16: placement de ligand dans le site actif [41]

- L'approche vise la complémentarité à 3 dimensions a un site de liaison de la protéine cible. Ces propriétés peuvent inclure l'électrostatique, la forme, la taille. Lipophile, l'aromaticité, etc.
- il faut utiliser un logicielle pour produire des composés et évaluer les interactions intermoléculaires. Actuellement, la faisabilité chimique n'est pas considérée dans ce processus.
- Des fragments moléculaires sont joints individuellement à un point d'un ligand partiel dans la cavité de liaison de docking. Une évaluation de la "forme" avec la protéine est déterminée et le scoring peut être marqué. Cette amarrage et la notation des ligands potentiels peut être effectuée manuellement, cependant la génération automatisée des ligands potentiels fournit la vitesse nécessaires pour évaluer de nombreuses structures possibles, et donne une vue impartiale de l'ajustement des fragments. Cette technique peut également être utilisée pour concevoir des liens entre des fragments de molécules plus grosses de fragment accueil.

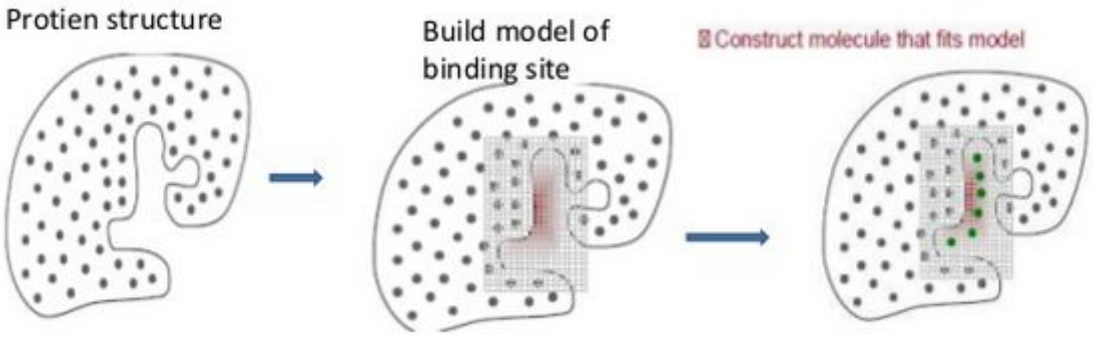


Figure 2.17 : le principe de novo Drug design [34]

6.3. Le processus de novo Drug design :

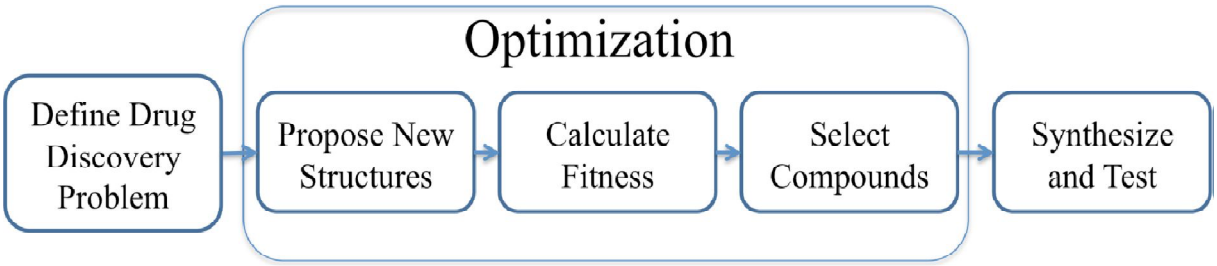


Figure 2.18: le processus de novo drug design [35]

Le processus De Novo Drug Design ce fait sur 6 étapes chaque étape sera détailler par la suite:

1) Production de contraintes primaires potentiels
2) Calcul des sites d'interaction
3) Construire une méthode
4) Essaye ou notation
5) les stratégies de recherche
6) les contraintes cibles secondaires

6.3.1. Contraintes cibles primaires

les contraintes cibles primaires sont les molécules capable de faire une interaction avec le récepteur de ligand, il existe 2 type de contraintes cibles primaires

1) Receptor base: l'interaction de la forme du récepteur de base pour la conception de médicaments

2) à base de ligand: la fonction ligand de cible en tant que clé. Dans la conception novo de la structure de la cible devrait être connue d'une haute résolution et la liaison au site doit être bien définie. Cela devrait définit non seulement une contrainte de forme, mais les sites d'interaction hypothétiques généralement constitué de liaisons hydrogène, interactions non covalentes électrostatiques et autres. Ceux-ci peuvent réduire considérablement l'espace de l'échantillon, comme des liaisons hydrogène et d'autres interactions anisotropes peuvent définir des orientations spécifiques. [34]

6.3.2. Les stratégies de construction :

Il existe 4 types de stratégie pour la construction de la nouvelle molécule :

1) Growing
2) Linking
3) Lattice based sampling
4) Molecular dynamics based methods

I. Le growing:

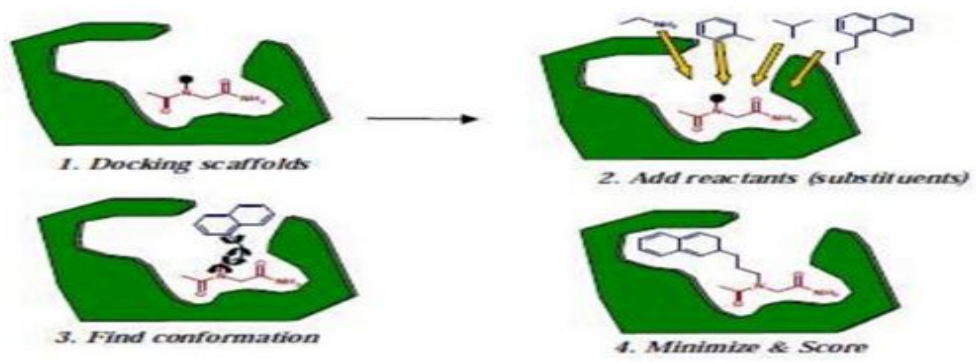


Figure 2.19: le growing [34]

Les étapes:

• Un bloc des fragments est le point de départ ou de la graine

• Les fragments sont ajoutés pour fournir des interactions de block des fragments avec le site de liaison

• Ceux-ci comprennent des chaînes simples d'hydrocarbures, amines, alcools, et même anneau simple

• Dans le cas de multiples graines, la croissance est généralement simultanée et continue jusqu'à ce que toutes les pièces ont été intégrées dans une seule molécule. [34]

II. Le linking :

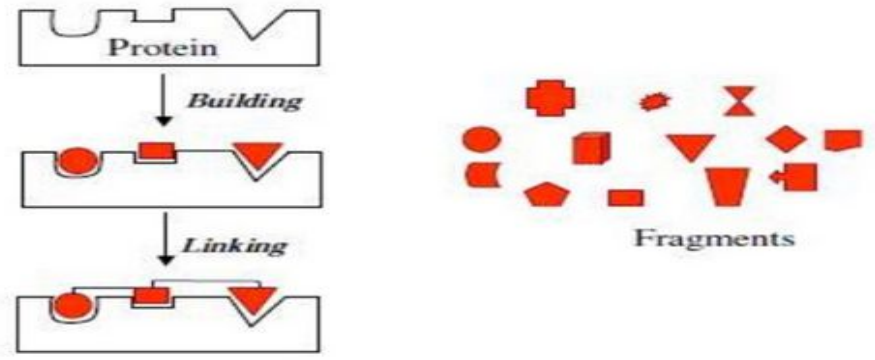


Figure 2.20: le linking [34]

Les étapes :

• Les atomes de fragments ou des blocs de construction sont soit placés dans des sites d'interaction clés ou pre_docked utilisant un autre programme

• Ils sont reliés entre eux par des règles prédéfinies pour donner une molécule complète

• Les groupes de liaison ou linkers peuvent être prédéfinis ou générés pour satisfaire toutes les conditions requises. [34]

III. La stratégie Link/Grow :

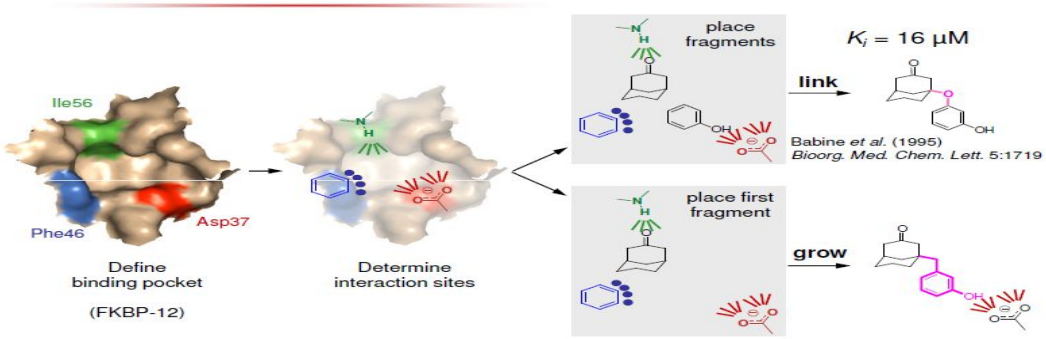


Figure 2.21: la stratégie link/grow [34]

IV. Méthode basée Lattice :

- Le treillis est placé dans le site de liaison et les atomes autour des sites clés sont joints en utilisant le chemin le plus court.
- Ensuite, diverses versions dont chacune comprend translation, rotation ou mutation d'atomes sont guidées par la fonction d'énergie potentielle pour aboutir finalement à une molécule cible.
- Les blocs de construction sont initialement placés au hasard et ensuite par simulation MD (Dynamique moléculaire) autorisés à réorganiser.
- Après chaque réarrangement certaines obligations ont été brisées et le processus est répété

Au cours de cette structure haute procédure de notation ont été stockés pour une évaluation. [34]

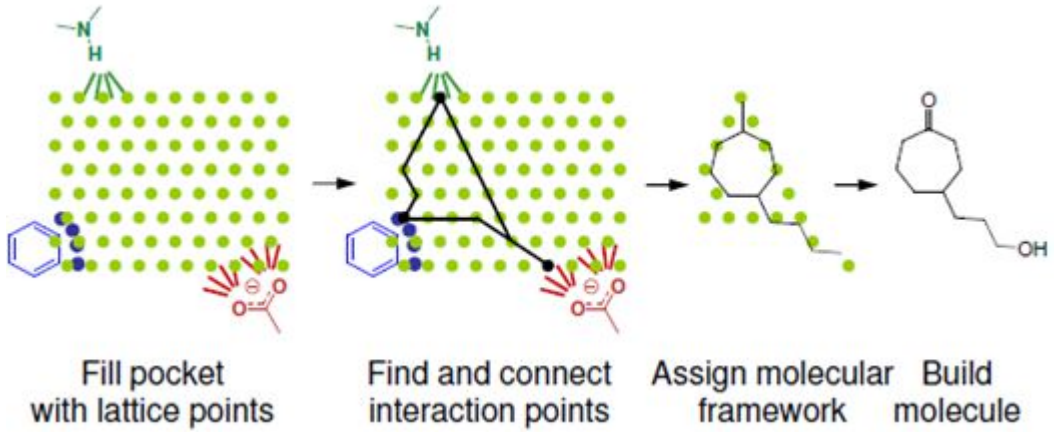


Figure 2.22: la stratégie lattice [34].

6.3.3. Contraintes cibles secondaires

L'affinité de liaison à lui seul ne suffit pas pour concevoir des médicaments efficaces il existe d'autre Propriétés essentielles comprennent l'absorption efficace, la distribution, le métabolisme, l'excrétion et la toxicité pour évaluer la performance de la molécule conçu.

6.4. Problème et les défis de novo Drug design :

- le défi principal c'est la représentation correcte du problème
- trouver des solutions candidat valable et l'évaluation de la qualité de solution proposée
- La satisfaction de ces besoins dans le cadre de DND (De Novo Drug Design) exige, entre autres, le codage des méthodes d'évaluation de la structure chimique

pertinentes sur le plan pharmaceutique, la mise en œuvre d'un moteur de génération de structure chimique virtuel, et l'utilisation d'une méthode d'optimisation pour explorer l'espace de recherche chimique

- la recherche exhaustive est difficile et des fois impossible pour ce type de problème

- Le problème d'échantillonnage de la structure : comment assembler les composés candidat, par exemple, sur la base-fragment à base atome.

- Le problème de notation : comment évaluer la qualité molécule, par exemple le docking et le scoring récepteur-ligand 3D (nécessite la structure de récepteur), ou d'une mesure de similarité basée ligand (nécessite un ligand de référence).

- Le problème d'optimisation : comment naviguer systématiquement l'espace de recherche, par exemple, par la recherche en largeur d'abord en profondeur d'abord, échantillonnage de Monte Carlo avec le critère Metropolis, les algorithmes évolutionnaires, ou une structure exhaustive énumération. [35]

Conclusion :

Dans ce chapitre la nous essayons d'illustrer tous les notions fondamentaux concernant notre thème de recherche le Drug design avec les différentes approches et les méthodes qui existe. Le Drug design c'est la conception de médicament en générale il existe plusieurs type de Drug design, le Drug design rationnelle(rational Drug design) c'est la conception de médicament traditionnelle basé sur les essayes expérimentaux, le Drug design assisté par ordinateur(computer aided Drug design) ce qui est nous intéresse, il existe deux types ou approches de Drug design assisté par ordinateur, la conception de médicament basé structure(structre_based Drug design), et la conception de médicament basé ligand(ligand_based Drug design), la première approche se réalise soit en utilisant le criblage virtuelle ou la conception de novo(la conception de nouveau médicament) ,la deuxième approche(la conception de médicament basé ligand) se réalise en utilisant soit le modèle QSAR ou le modèle phormacophore. On peut conclure que la conception de médicament assisté par ordinateur résume le processus de Drug descovery temps et cout.

Chapitre 3

De novo drug design et l'optimisation multi objectif

Introduction :

Un médicament de qualité est un médicament qui satisfaire certain propriété surtout la propriété ADME/TOX, les méthodes d'optimisation multi objectif guider l'optimisation simultanément de plusieurs paramètre dans la conception de médicament d'une façon rapide et efficace.

L'optimisation multi objectif peut être combinée avec de novo drug design pour automatiser la génération et l'optimisation d'un très grand nombre de nouveaux composés.

Partie 1 : l'optimisation multi objectif :

1.1. Définition d'un problème d'optimisation:

L'optimisation multi objectif : est une branche de l'optimisation combinatoire dont la particularité est de chercher à optimiser simultanément plusieurs objectifs d'un même problème (contre un seul objectif pour l'optimisation combinatoire classique). Elle se distingue de l'optimisation multidisciplinaire par le fait que les objectifs à optimiser portent ici sur un seul problème. [1]

- **L'optimisation combinatoire :** Dans sa forme la plus générale, un problème d'optimisation combinatoire (on dit aussi d'**optimisation discrète**) consiste à trouver dans un ensemble discret un parmi les meilleurs sous-ensembles (ou solutions) réalisables, la notion de meilleure solution étant définie par une fonction objectif.

- **L'optimisation multidisciplinaire :** L'optimisation multidisciplinaire (OMD ou MDO en anglais) est un domaine d'ingénierie qui utilise des méthodes d'optimisation afin de résoudre des problèmes de conception mettant en œuvre plusieurs disciplines.

1.2. Quelque notion sur l'optimisation :

- **Le domaine des variables de décision:** soit continu et on parle alors de problème continu, soit discret et on parle donc de problème combinatoire.
- **la nature de la fonction objectif à optimiser**: soit linéaire et on parle de fonction linéaire, soit non linéaire et on parle de fonction non linéaire.
- **sa taille:** problème de petite ou de grande taille.

- **le nombre de fonctions objectifs à optimiser:** soit une fonction scalaire et on parle alors de problème mono-objectif, soit une fonction vectorielle et on parle donc de problème multi objectif.
- **la présence ou non des contraintes:** on parle de problème sans contrainte ou avec contrainte.
- **l'environnement:** problème dynamique (la fonction objectif change avec le temps).
- un problème d'optimisation peut être un problème de minimisation ou un problème de maximisation.

1.3. Les problèmes d'optimisation mono objective:

Les techniques d'optimisation Objectif simples sont des techniques visant à résoudre les problèmes où les solutions doivent satisfaire une seule fonction objective. Étant donné que ces problèmes essaient de trouver l'optimum d'une fonction d'un objectif qu'ils ont en général une seule solution. Dans ces problèmes les techniques d'optimisation Objectif simples sont les plus couramment utilisées. Un problème d'optimisation Objectif Simple est défini comme suite:

$$\begin{cases} minimize\ f(x) \\ gj(x) \geq 0\ (j = 0, \dots, m) \\ x \in X \subset R^n \end{cases} \quad (1)$$

f(x) : est la fonction objectif qui peut être maximisé ou minimisé

$g_j(x)$: le nombre de contrainte

x : Un vecteur de n variable de décision $x = (x1, x2, \dots, xn)^T$

X : représente la région faisable

1.4.Les problèmes d'optimisation multi objectif :

L'optimisation multi objectif, aussi connue sous le nom d'optimisation multicritère est un processus qui consiste à optimiser simultanément deux ou plusieurs objectifs, soumis à certaines contraintes. Un problème multi objectif est caractérisé, non pas par une solution unique, mais par un nombre de solutions qui réalisent des compromis sur les différents objectifs [3]. Il procède donc par la

d'optimisation va résoudre le problème en cherchant les meilleures solutions en fonction de ces critères.

1.4.1. Formulation d'un problème d'optimisation multi objectif :

Formuler un problème d'optimisation revient à d'identifier clairement trois ingrédients: les variables d'optimisation ou de décision, les objectifs du problème et les contraintes. D'une façon générale, le problème d'optimisation multi objectif est exprimé par l'équation suivant [4,5]:

$$\min(ou\ max)\ f(x) = [f1(x), f2(x), \dots, fm(x)], x \in E = R^m\ et\ m > 2$$

$$\begin{cases} gi(x) \geq 0, & i = 1, \dots, p \\ hj(x) = 0, & j = 1, \dots, q \\ x_{kMin} \leq x_k \leq x_{kMax}, & k = 1, \dots, n \end{cases} \qquad (2)$$

m>=2 : le nombre de fonction objectif

f(x)=f$_1$(x),…..,f$_m$(x) : le vecteur des fonctions a optimiser

x=(x$_1$,….,x$_m$) : le vecteur des variables de décision

g$_i$(x) : les contraintes d'inégalité

h$_j$(x) : les contraintes d'égalité et les contraintes de domaine

x$_{kMin}$ et ***x$_{kMax}$*** sont les bornes supérieure et inférieure des variables de décisions ***x$_k$***

R^m c'est l'ensemble des solutions réalisables.

1.5. Résolution d'un problème d'optimisation :

La classification des problèmes d'optimisation Variable suivant le point de vue considéré :

- Algorithmes déterministes / algorithmes stochastiques
- Algorithmes de recherche locale / algo de recherche globale
- Algorithmes d'optimisation locale/globale
  - **Algorithmes d'optimisation locale:** Tout algorithme piégé par le premier optimum rencontré. Ne permettant pas d'obtenir une solution proche de l'optimum global en raison de la trop grande cardinalité de l'espace de recherche
  - **Algorithmes d'optimisation globale:** Tout algorithme qui n'est pas sensible aux minima locaux. Algorithme permettant d'obtenir une solution proche de

l'optimum global

- **L'algorithme stochastique:** elles ont un comportement non-déterministe d'un problème à un autre, d'une instance à une autre, et même d'une exécution à une autre. [6]
- **Les méthodes exactes:** permettent de calculer la (ou les) solution(s) optimale(s), on est sûr qu'il n'existe aucune solution de meilleure qualité. Cependant, si cette classe de méthodes permet de parcourir l'espace de recherche de façon intelligente en réduisant le nombre de solutions visitées, elle n'est efficace qu'avec des problèmes spécifiques et/ou de petites tailles. [6]
- **L'heuristique:** Cette classe ne permet pas de donner une garantie d'optimalité ; cependant, elles sont bien adaptées aux problèmes complexes de grande taille. En effet, une heuristique permet de calculer une solution approchée en un temps raisonnable comparativement à une méthode exacte. [6]
- **Le méta heuristique :** ce sont des heuristiques dédiées à un problème donné et celles plus génériques, ces dernières peuvent être appliquées à un large panel de problèmes. [6]

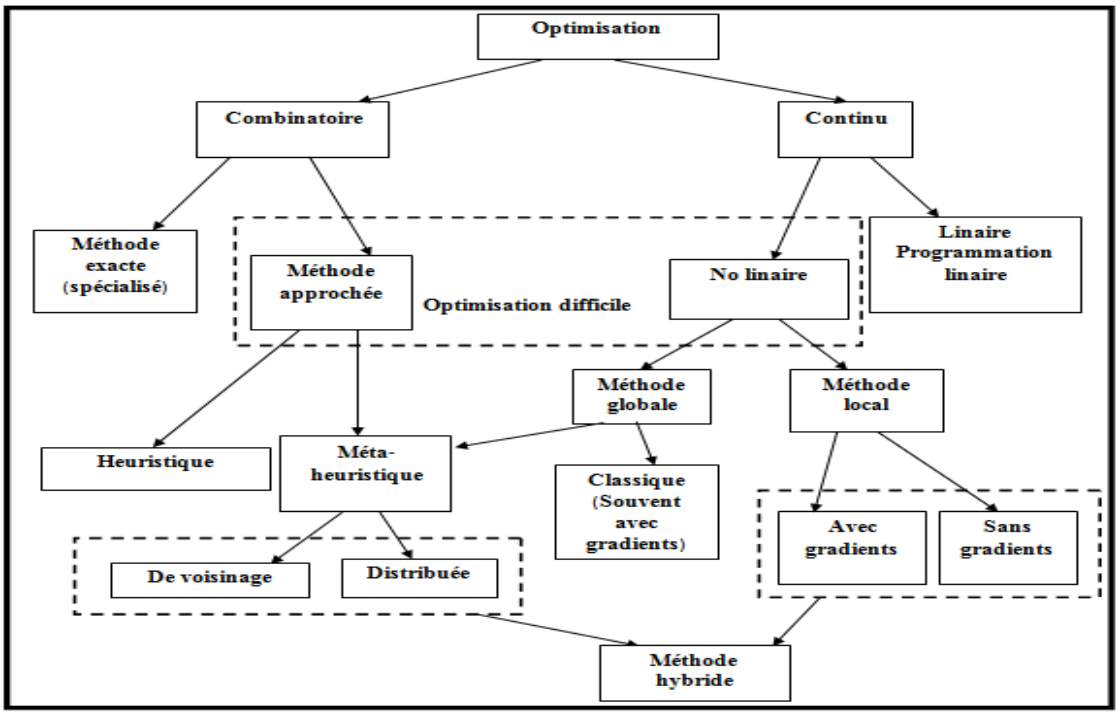


Figure 3.1 : classification des méthodes de résolution

1.6. Le front de Pareto les solutions dominé et les solutions non dominé :

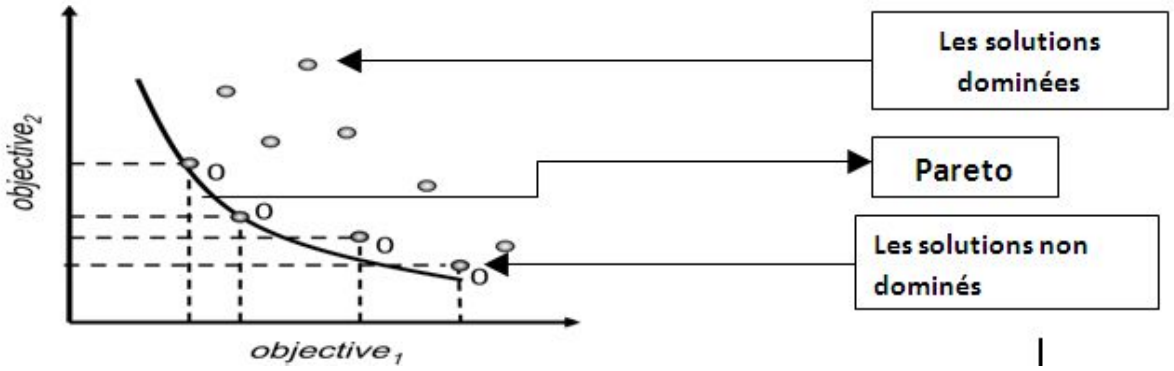


Figure 3.2: le front de Pareto les solutions dominé et les solutions non dominé [7]

A. **Solution non dominée:** lorsque l'amélioration des performances sur un objectif tend à aggraver la performance dans un autre. Ces solutions multiples «meilleurs», connus sous le nom non dominée, n'ont pas d'autres solutions qui sont mieux qu'eux dans tous les objectifs considérés. Connu comme la surface de compromis ou le front de Pareto. [7]

B. **Solution dominée :** les solutions sont dits dominées s'il existe une ou plusieurs solutions dans l'ensemble qui présentent de meilleures performances dans tous les objectifs. [7]

1.6.1. Les algorithmes d'optimisation multi objectif :

Les phénomènes physiques ou biologiques ont été à la source de nombreux algorithmes s'en inspirant plus ou moins librement. Ainsi les réseaux de neurones artificiels s'inspirent du fonctionnement du cerveau humain, l'algorithme de recuit simulé de la thermodynamique, et les algorithmes évolutionnaires (AEs) (dont les plus connus sont les algorithmes génétiques) de l'´evolution darwinienne des populations biologiques.

Les algorithmes les plus célèbres et les plus utilisé dans notre domaine de recherche sont les algorithmes évolutionnaires.

Partie 2 : de novo Drug design et l optimisation multi objectif

2.1. Pour quoi de novo drug design est un problème d'optimisation multi objectif ?

Par nature le processus de découverte de médicament est un processus multi objectif :

- Il faut estimer plusieurs caractéristiques pour les composés produits : la puissance, la propriété ADME, la toxicité, le cout…etc.
- Le résultat obtenue par les déverses expérience de criblage en parallèle

De coté information il faut exploiter les informations disponible de ressource et de nature déverse puis appliquer une méthode d'optimisation multi objectif. [8]

2.2. Les méthodes d'optimisation pour de novo Drug design :

Il existe 2 catégories de méthode d'optimisation multi objectif :

- ✓ **Des méthodes qui ignore la nature multi objectif** afin de simplifier le problème et qui concentre sur la conception de molécules en satisfaisant un seul objectif, soit prédit affinité de liaison à une protéine cible connue ou, similitude avec un ligand.
- ✓ **Des méthodes reconnaissent l'existence de multiples objectifs** dans la découverte de médicaments

- Parmi les méthodes d'optimisation les plus couramment utilisés dans le DND est les algorithmes évolutionnaires (EA).
- Au début le DND fondées sur la recherche combinatoire qui explorent l'espace des solutions utilisant des techniques comme en largeur ou en profondeur d'abord une recherche pour générer des conceptions de produits chimiques répondant aux contraintes imposées.

2.2.1. Les méthodes DND a objectif simple (Single objective DND methods) :

Les méthodes suivantes de cette approche se répartissent en deux catégories selon que l'objectif poursuivi est ligand ou cibles basées :

I. Les approches fondé ligand :

Il utilise généralement une méthode appelé TOPAS (en français Système d' Affectation de topologie) en anglais (TOPology Assigning System) [10]

a) Définition de TOPAS :

TOPAS (TOPology Assising System) des nouveaux composés moléculaires sont suggérées dans un processus cyclique entièrement automatisé, en prenant une structure donnée comme un point de référence (de la graine ou de la structure du modèle) de plus au lieu de générer des architectures moléculaires contenant des caractéristiques structurelles indésirable ou des composés de synthèse insolubles(Un problème rencontré par beaucoup de novo procédure de conception) ,TOPAS était équipé d'un ensemble limité de médicaments dérivés blocs de construction qui ont été obtenus à partir de la fragmentation avant rétro synthétique de l'indice de drogue dans le monde. [11]

b) L'algorithme de conception TOPAS :

TOPAS est basé sur un algorithme évolutionnaire simples un (1, n) stratégie d'évolution.

- Le choix d'une EA spécifique est presque arbitraire à condition que le processus d'optimisation est guidé par la stratégie d'adaptation des paramètres du capital sous-jacent est fondé sur un processus de recherche stochastique à partir d'un point arbitraire de l'espace de recherche des parents dite initiale
- Un ensemble de n variante sont générés
- Une distribution en forme de cloche de la variante est générée pour chaque génération (cycle optimisé), à centrer les parents .ce signifie que la plupart sont très similaires à la structure de leur parent, et avec l'augmentation de la dissemblance (distance) le nombre de variante diminue.
- Une valeur de remise en forme a été calculé pour chaque variante, et les plus forts a été choisi comme le parent de la génération suivante dans l'étude de parent ce processus cyclique a été contraint de mettre fin après un nombre n de générations sorcière avéré suffisant pour la convergence sur un optimum local de conditionnement physique.

Le processus a été supposé avoir atteint un optimum de remise en forme en l'absence de modification structurelle réussie a eu lieu au cours des dix dernières générations, et le paramètre de stratégie étape taille se rapproche d'une valeur de 0. [11]

II. Les approches fondées sur des objectifs :

Reposent sur la disponibilité d'une description détaillée du récepteur cible d'intérêt pour concevoir des structures chimiques prédit de bien lier.

Cette approche utilise le docking/scoring pour L'évaluation des produits de conception qui donnent une indication de l'affinité probable de chaque composé virtuelle au récepteur. [12]

- **Principe :**

Les méthodes d'optimisation Search-based sont les plus utilisées dans les travaux récents en combinaison avec un moteur de synthèse de composé virtuel, pour concevoir des composés qui remplissent une fonction objective basée sur le docking. [13, 14,15]

- Alternativement, certaines méthodes utilisent la connaissance disponible sur le site récepteur pour identifier et caractériser les régions qui peuvent être impliqués dans des interactions chimiques
- À la suite, une approche de construction supplémentaire peut être utilisée pour générer des composés virtuels à travers le couplage des régions clés de récepteurs avec des fragments moléculaires qui peuvent théoriquement interagir et former des liaisons chimiques
- Les Structures chimiques spéciaux, appelés linkers, sont utilisés pour relier les fragments moléculaires prenant en compte des considérations de géométrie
- La connaissance de l'emplacement du récepteur peut également être utilisée pour dériver un modèle de composés ayant une affinité de liaison prédite dans ce cas, le but du processus de conception de novo est de générer des composés correspondant à ce modèle

2.2.2. Les méthodes DND multi-objectif (multi objective DND methods) :

La procédure d'optimisation multi objective générale est illustrée dans la figure ci-dessus :

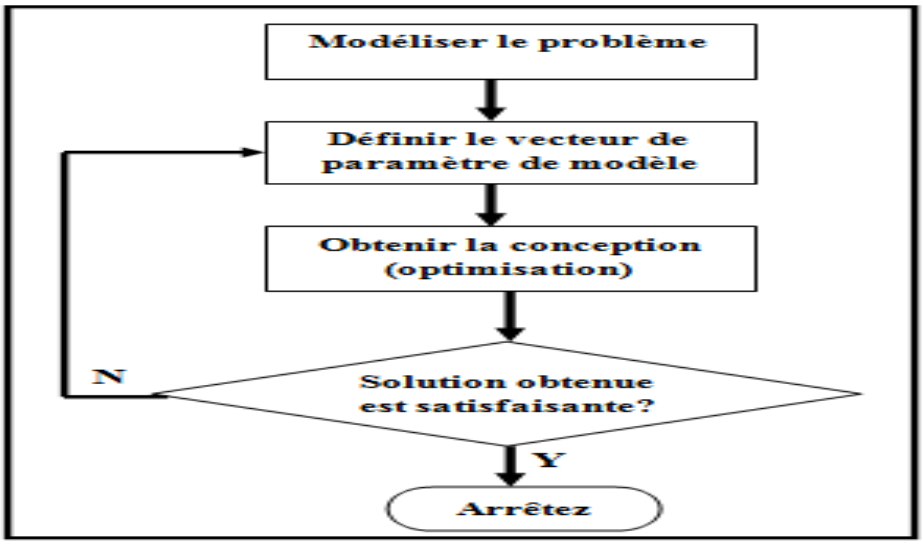


Figure 3.3 : le processus e général de l'optimisation multi objectif. [16]

1) Classification des méthodes multi objectif pour de novo drug design :

Les approches multi objective			
Type de méthode	**interaction de l'utilisateur**	**méthodes**	**exemples d'applications**
A priori	**avant l'optimisation**	**Moyenne-pondérée:** spécification des connaissances d'expert en fonction de poids aux objectifs, les objectifs globaux en une seule grâce à la combinaison linéaire des moyennes-pondérée pesées de tous les objectifs	**Chemical Genesis ,LEA, GANDI , FOG , PHDD**
		Opportunité **(desirability):** spécification des connaissances d'experts de la cession de l'opportunité d'objectifs, les objectifs globaux en une seule grâce à des fonctions de désirabilité	**MOOP-DESIRE**
A posteriori	**après l'optimisation**	**Pareto-fondé:** processus d'optimisation a lieu sans utilisation de la connaissance a priori; produire un ensemble de solutions optimales, des connaissances spécialisées utilisées pour sélectionner un ensemble de solutions souhaitées après optimisation	**COG., MEGA , PLD**
progressive	**lors de l'optimisation**	**interactive:** permettre à l'utilisateur d'interagir avec le processus d'optimisation pour guider la recherche, l'utilisateur agit comme la fonction de score de fitness	**MoleculeEvoluator ,Mobius**

Tableau 3. 1 : les approches utilisées par les méthodes DND pour adresser la présence de multiple objectif

A. Les méthodes A priori :

Les méthodes a priori exigent que des informations de préférence suffisante est exprimé avant le processus de solution. [17] Des exemples bien connus d'une des méthodes a priori comprennent la méthode de la fonction d'utilité, la méthode lexicographique, et la programmation de but.

fonction d'utilité :

Dans le procédé de la, il est supposé que la fonction d'utilité du décideur est disponible. Une cartographie u :$Y \to R$ est une fonction d'utilité si pour tout y^1 a y^2 et u (y^2)=u (y^2) si le décideur est indifférent pour y^1 et y^2 la fonction d'utilité spécifier un ordremax $u(f(x))subject\ to\ x \in X$, mais en pratique, il est très difficile de construire une fonction d'utilité qui représente avec précision les préférences du décideur [18] en particulier depuis le front de Pareto est inconnu avant le début de l'optimisation.

i. Méthode lexicographique

Suppose que les objectifs peuvent être classés dans l'ordre d'importance. On peut supposer, sans perte de généralité, que les fonctions objectives sont dans l'ordre d'importance de sorte que f_l est le plus important et f_k le moins important pour le décideur. La méthode lexicographique consiste à résoudre une série de problèmes d'optimisation mono-objective de la forme

$$\min f_l(x)$$
$$s.t\ f_j \leq y_j^*, j = 1, \dots, l-1, x \in X \qquad (3)$$

Ou y_j^* est la valeur optimale du problème ci-dessus avec l=j. Ainsi y_j^* :=min $\{f_1(x) \backslash x \in X\}$ et chaque nouveau problème de la forme dans le problème ci-dessus dans la séquence ajoute une nouvelle contrainte que l va de 1 à k.

ii. la méthode composite :

La majorité des méthodes multi-objectifs combinent les nombreux objectifs en un seul avant l'application d'une méthode d'optimisation Ces méthodes décidées a priori et de manière efficace l'importance relative de chaque objectif existant, souvent par l'association d'un poids pour chacun d'entre eux, pour générer un nouveau, c'est l'objectif composite [7].

- **Avantage :**

  Les mêmes algorithmes utilisés pour résoudre les problèmes à simple objectif peuvent être utilisés pour des problèmes multi-objectifs

- **Inconvénient :**
  - Inconvénients de la méthode comprennent la nécessité de choisir une pondération appropriée pour les différents objectifs, même si la relation entre eux n'est pas claire,
  - Pour la génération de solutions 'meilleurs' aucune information associée a leur placement sur le front de Pareto dans l'ensemble des solutions non dominé.

iii. Méthode de désirabilité:

Une approche pour éviter la dureté artificielle de filtres simples est une méthode qui concerne la valeur d'une propriété à la «désirabilité» de ce résultat, en utilisant une «fonction désirabilité» [10]. Ceci est une fonction mathématique qui traduit la valeur d'une propriété dans un nombre entre 0 et 1, ce qui représente la façon souhaitable que résultat serait; une opportunité de 1 indique que la valeur de la propriété est idéale, tandis que 0 correspond à un résultat tout à fait inacceptable. [2]

- **L'avantage**

  Fonctions désirabilité offrent une plus grande flexibilité que les filtres dans la définition des exigences en matière de propriété pour un composé de succès, l'importance de chaque propriété individuelle à l'objectif global d'un projet et les compromis acceptables si un composé idéal ne peut être identifié.

- **Inconvénient :**

  Toutefois, afin de définir les fonctions de désirabilité et leurs poids pour un objectif spécifique du projet, nécessite une connaissance a priori des valeurs idéales de propriété composés et des compromis acceptables. La complexité des données maintenant générés dans la découverte de médicaments signifie que ce ne soit pas toujours clair, même pour un scientifique expérimenté.

B. Les méthodes a posteriori :

Les méthodes a posteriori visent à produire toutes les solutions optimales de Pareto ou un sous-ensemble représentatif des solutions Pareto optimales. La plupart des méthodes a posteriori tombent dans l'une des deux catégories suivantes: la programmation mathématique à base a posteriori méthodes, où un algorithme est

répété et chaque exécution de l'algorithme produit une solution optimale au sens de Pareto, et algorithmes évolutionnaires où une exécution de l'algorithme produit un ensemble de solutions Pareto optimales.

- **Approches basées sur l'optimisation de Pareto:**

L'optimisation de Pareto est basée sur le principe ou il avoir plusieurs résultats possible qui représentent différents soldes, «optimal» de biens et non pas une seul. Une solution optimale au sens de Pareto (un composé dans le cadre de la découverte de médicaments) est celui pour lequel il n'y a pas une autre solution qui est meilleure dans toutes les autres propriétés ces solution sont dites non dominé et ils sont inclus dans le front de Pareto. [23] Une illustration de ce concept est représentée sur la figure 2.

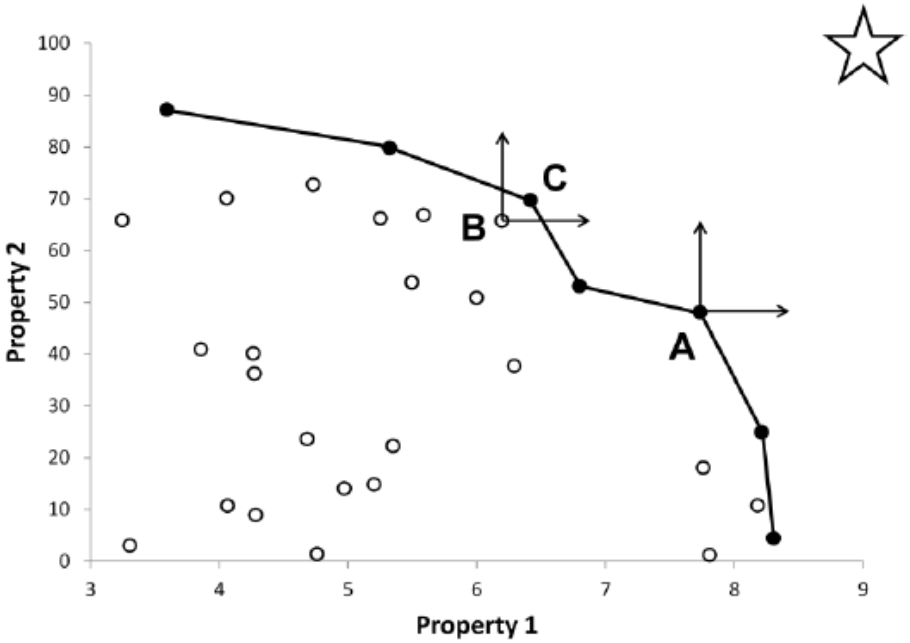


Figure 3.4 : Illustration de la notion d'optimum de Pareto pour les composés représentés par des points dans une parcelle de la propriété 1 par rapport à la propriété 2. [22]

L'objectif idéal correspond à l'angle supérieur droit de la parcelle, comme indiqué par l'étoile. Des points solides sont optimum de Pareto ou «non-dominé»; dans le cas du point A, il n'y a pas de points avec une valeur plus élevée pour les deux paramètres. Cependant, les milieux ouverts ne sont pas optimum de Pareto; par exemple le point B est «dominé» par point C.

C. Les méthodes progressives ou les méthodes interactives :

Dans les méthodes interactives, le processus de solution est itératif et le décideur interagit en permanence avec la méthode pour rechercher la solution la plus préférée .En d'autres termes, on prévoit que le décideur exprime des préférences à chaque itération afin d'obtenir des solutions Pareto optimales qui sont d'intérêt pour lui / elle et apprendre ce genre de solutions sont réalisables.

Les méthodes interactives travaillées généralement comme suite :

1) initialiser (par exemple le calcul idéal et approchée nadir vecteurs objectifs et les montrer au décideur)

2) générer un point de départ optimal au sens de Pareto (en utilisant par exemple une méthode sans préférence ou solution donnée par le décideur)

3) demander des informations sur les préférences du décideur (par exemple les niveaux d'aspiration ou le nombre de nouvelles solutions pour être générés)

4) générer une nouvelle solution optimale au sens de Pareto en fonction des préférences et montrer / eux et éventuellement d'autres informations sur le problème pour le décideur

5) si plusieurs solutions ont été générés, demander au décideur de choisir la meilleure solution à ce jour

6) arrêter, si le décideur veut; Sinon, passez à l'étape 3).

Au lieu de convergence mathématique qui est souvent utilisée comme critère d'arrêt dans les procédés d'optimisation mathématiques, une convergence psychologique est soulignée dans les méthodes interactives. D'une manière générale, une méthode est terminée lorsque le décideur est convaincu qu'il a trouvé la solution la plus préférée disponible.

2.3. Le filtrage :

La méthode la plus courante pour MPO est d'appliquer des filtres de propriété multiples pour rejeter les composés qui ne répondent pas à tous les critères de propriété. Bien que la simplicité de filtrage est très attrayante.

- **Inconvénient :**
  - Établissent des distinctions artificiellement difficiles entre les composés ayant des propriétés similaires Cette distinction dure est encore aggravée par l'incertitude dans les données à laquelle un filtre peut être appliqué.

- Ces incertitudes accumulent lorsque nous appliquons plusieurs filtres dans l'ordre.
- Pour cette raison, les filtres doivent être traités avec prudence. Dans les situations où de bonnes possibilités sont abondantes, il peut être approprié d'appliquer plusieurs filtres. Toutefois, lorsque le coût d'une occasion manquée est élevé, ce qui est fréquemment le cas dans la découverte de médicaments, le risque de rejet à tort de bons composés peut être trop grand.

3. Les algorithmes proposés :

Il existe plusieurs algorithme d'optimisation multi objectif dans la littérature pour notre travaille nous étudions trois entre ils les plus récent et les plus utilisé pour les MPOs puis nous choisissons un parmi ces algorithmes pour l'appliquer dans notre domaine de novo drug design :

3.1. L'algorithme MEGA (The Multi-objective Evolutionary Graph Algorithm) :

3.1.1. Définition :

Nicolaou et al. Décrivent MEGA (Multi-objective Evolutionary Graph Algorithm), une méthode qui combine les algorithmes évolutionnaires avec les techniques de recherche locale afin de permettre l'utilisation des connaissances de problèmes spécifiques lors de la recherche et d'amélioration des performances et d'évolutivité [30]. Ils proposent un cadre algorithmique général pour la conception de structures chimiques (graphes moléculaires) qui répondent à plusieurs exigences d'importance pharmaceutique.

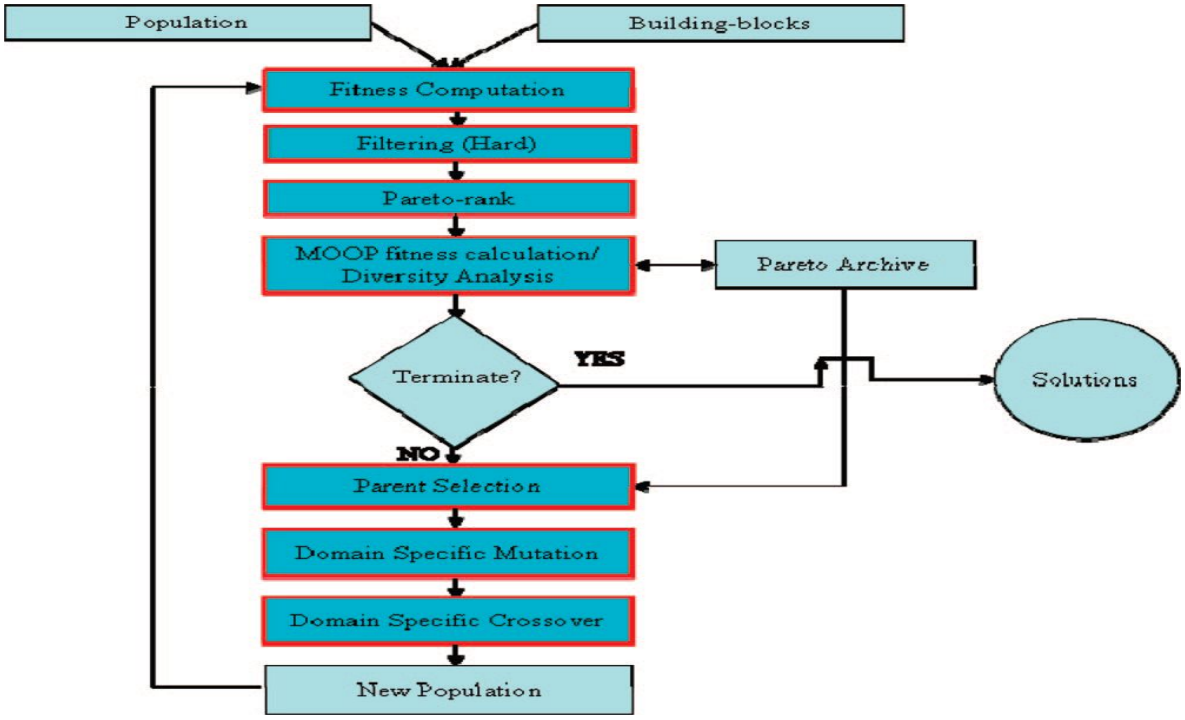


Figure 3.5 : le diagramme de fonctionnement de l'algorithme MEGA [29]

3.1.2. L'algorithme PMEGA(Le MEGA parallèle):

PMEGA est une extension de MEGA qui exploite le parallélisme afin de réduire le temps d'exécution. MEGA et PMEGA ont été développés principalement pour la conception de graphiques optimales répondant à des objectifs multiples. Un accent particulier a été mis sur le problème de la conception de petits graphiques moléculaires aux propriétés thérapeutiques communément appelées conception de novo de drogue. [32]

C'est un algorithme créé par Christos Kannas, dans leur travaille il applique l'algorithme avec des changements dans le nombre de sous population et la capacité de processeur et il compare les résultats obtenue avec les résultats de MEGA. Les résultats obtenus sont comme suit :

- Des solutions de qualité égale à MEGA
- Un temps d'exécution réduit par rapport à MEGA

Points (a) à (c) peut être réalisé en utilisant une méthode de parallélisassions, le point (d) doit être un sujet de recherches ultérieures

I. Le fonctionnement de PMEGA :

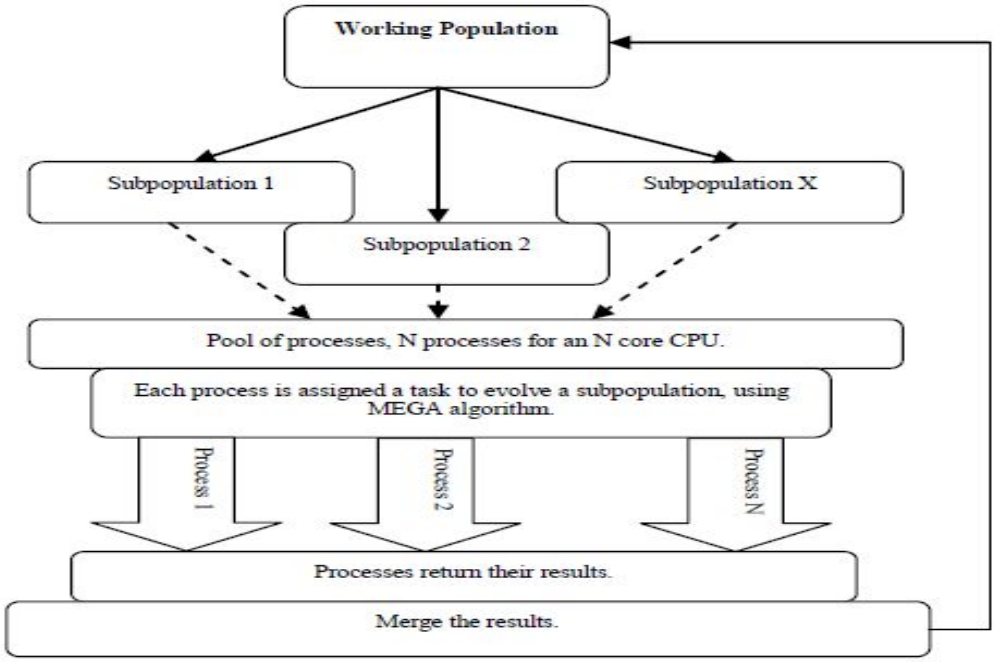


Figure 3.6: le fonctionnement de l'algorithme PMEGA [32]

3.2. L'algorithme MOEA/D (multi objective evolutionary algorithm based decomposition) :

3.2.1. Définition:

C'est un algorithme d'optimisation base sur la décomposition qui consiste a:

1) Décomposer explicitement le problème d'optimisation multi objectif en N sous problème d'optimisations scalaires.

2) Il résout ces sous-problèmes simultanément par l'évolution d'une population de solutions.

3) A chaque génération, la population est composée de la meilleure solution trouvée jusqu'à présent (à savoir depuis le début de l'exécution de l'algorithme) pour chaque sous-problème.

4) Les relations de voisinage entre ces sous-problèmes sont définies en fonction des distances entre leurs vecteurs de coefficients d'agrégation les solutions optimales aux deux sous-voisins devraient être très similaires

5) Chaque sous-problème (par exemple, la fonction d'agrégation scalaire) est optimisée dans MOEA / D en utilisant des informations uniquement de ses sous-problèmes voisins. [33]

3.2.2. Les variantes de l'algorithme MOEA/D :

- **L'algorithme MOGA (multi objectif genetic algorithm) :**

C'est une variante de MOEA/D qui utilise les algorithmes génétiques, la différence principale c'est le traitement efficace de différant forme de front de Pareto

- **L'algorithme MOGLS:**

MOGLS a été proposée par Ishibuchi et Murata , et encore améliorées par Jaszkiewicz . L'idée de base est de reformuler la MOP (1) que l'optimisation simultanée de toutes les fonctions de TCHEBYCHEFF pondérés ou toutes les fonctions de la somme pondérée. [35]

Input:
• MOP (1);
• A stopping criterion;
•K: the size of temporary elite population;
•S: the size of initial population.
Output:
Step 1) Initialization:
Step 1.1) Generate S initial solutions $x^1,\ldots,x^s$ randomly or by a problem-specific method. Then, CS is initialized to be $\{x_1,\ldots,x^s\}$.
Step 1.2) Initialize $z= (z_1,\ldots,z_m)^T$ by a problem-specific method.
Step 1.3) EP is initialized to be the set of the -values of All the non dominated solutions in CS.
Step 2) Update:
Step 2.1) Reproduction:
Uniformly randomly generate a weight vector.
From CS select K the best solutions, with regard to the
Tchebycheff aggregation function ^{gte} with the weight vector λ , to form a temporary elite population (TEP).
Draw at random two solutions from, and then generate a new solution from these two solutions by using genetic operators.
Step 2.2) Improvement: Apply a problem-specific repair/Improvement heuristic on y to generate y'.
Step 2.3) Update of z: For each i=1,…, m, if $z_j< f_j(y')$, then set $z_j=f_j(y')$.
Step 2.4) Update of Solutions in TEP:
If y' is better than the worst solution in TEP with regard to g^{te}
With the weight vector λ and different from any solutions
In with TEP regard to -values, then add it to the set CS.
If y' the size of CS is larger than K× *S*, delete the oldest solution in CS.
Step 2.5) Update of EP:
Remove from EP all the vectors λ dominated by F(y') .
Add F (y') to EP if no vectors in EP dominates F(y') .
Step 3) Stopping Criteria: If stopping criteria is satisfied,
Then stop and output. Otherwise, go to **Step 2**.

- **ALGORITHME PAλ-MOEA / D:**

L'algorithme PAλ-MOEA / D est basée sur l'approche Tchebycheff, Certains paramètres sont définis comme:

- ✓ N : est la taille de la population;
- ✓ T : est le nombre de la région avoisinante de vecteurs de pondération;
- ✓ Archive : est la population externe qui stocke les solutions non dominées.

Il y a quatre étapes dans le paλ-MOEA / D:

Étape 1 : est d'initialiser la première population, vecteurs de poids, le quartier et la population externe (Archive);

Étape 2 : est de générer les prochaines descendants, la population de mise à jour et Archives;

Étape 3 : est la méthode de paλ

Étape 4 : est d'arrêter l'algorithme lorsque le nombre d'évaluations de la fonction de remise en forme est plus grand que max_evals. [36]

- **L'algorithme EASS:**

C'est un algorithme proposé par Cai Dai et Yuping Wang il se compose de trois parties : la classification des solutions, la stratégie de mise à jour, et de la stratégie de sélection

Etape 1 (initialisation). Donné N vecteurs de direction$(\gamma^1, \gamma^2, \dots, \gamma^N)$, Générer aléatoirement une population initiale $POP(k)$et sa taille est N, soit K=0, Ensemble $Z_i = min\{f_i(x)|x \in POP(k)\}, 1 \leq i \leq m$

Étape 2 (remise en forme). Solutions de POP(k) sont d'abord divisées en N classes par la formule (5) et la valeur de condition physique de chaque solution en POP(k) est calculée par la distance de l'encombrement. Puis, de meilleures solutions sont choisies parmi la population POP(k) et mis dans la population POP. Dans ce monde, la sélection du tournoi binaire est utilisée.

Étape 3 (nouvelles solutions). Appliquer les opérateurs génétiques pour la population mère pour générer progéniture. L'ensemble de tous ces enfants est notée O.

Étape 4 (mise à jour).Z est d'abord mis à jour. Pour chaque $j = 1, \dots, m$ if $z_j > \min\{f_j(x)|x \in O\}$

$z_j = min\{f_j(x)|x \in O\}$Les solutions $POP(k) \cup C$ sont tout d'abord classée par la formule (5); puis meilleures solutions sont sélectionnés par la stratégie de mise à jour de la section 3.2 et mis en $POP(k+1)$ soit k=k+1.

Étape 5 (résiliation). Si la condition d'arrêt est satisfaite, arrêtez; Sinon, passez à l'étape 2. [37]

3.3. L'algorithme NSGA (Non dominated Soring Genetic Algorithm) :

3.3.1. Définition :

L'algorithme génétique tri non dominé est un algorithme d'optimisation Multiple Objectif (MOO) et est une instance d'un algorithme évolutionnaire du champ de calcul évolutif.

NSGA est une extension de l'algorithme génétique pour plusieurs d'optimisation de la fonction objectif. Elle est liée à d'autres Evolutionary Multiple Objective Optimization Algorithms (de Emoo) (ou Multiple Objective Evolutionary Algorithms MOEA), tels que Vector-Evaluated Genetic Algorithm (VEGA), Strength Pareto Evolutionary Algorithm (SPEA), et Pareto Archived Evolution Strategy (PAES).

3.3.2. Principe de NSGA:

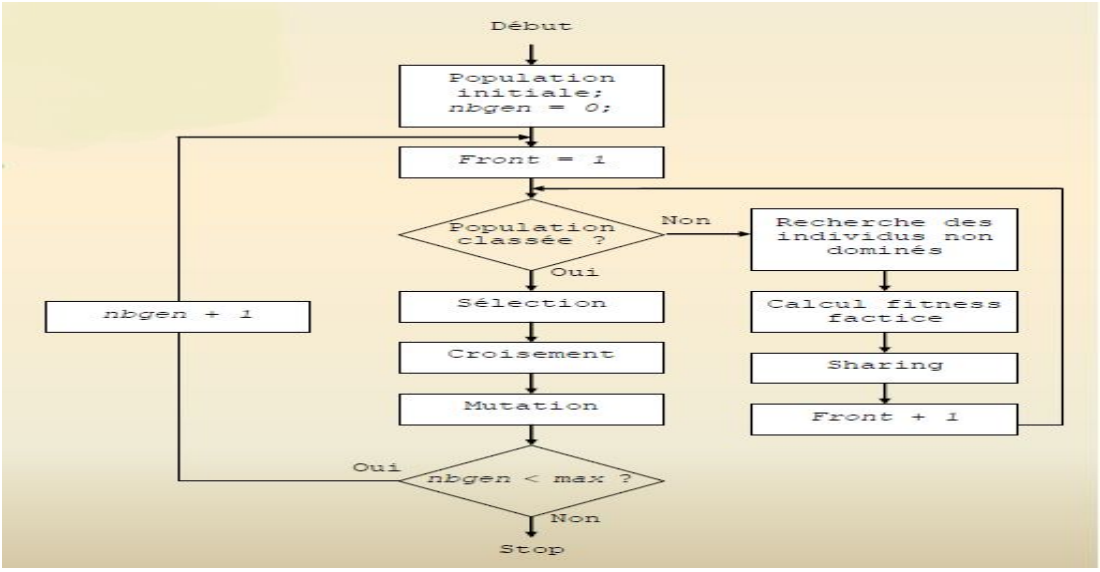


Figure 3.7 : fonctionnement de l'algorithme NSGA

3.3.3. Les variantes de l'algorithme NSGA :

Il existe quatre versions de l'algorithme, le NSGA classique et le formulaire mis à jour et le moment canonique NSGA-II et NSGA-III et la dernière version l'U- NSGA-III.

A. NSGAII :

❖ **Définition :**

NSGA-II est la deuxième version de la célèbre "non dominé par algorithme génétique tri" sur la base des travaux du Pr Kalyanmoy Deb pour résoudre des problèmes simples et multiples non convexes et non lisses optimisation objectives.

❖ **Principe de fonctionnement :**

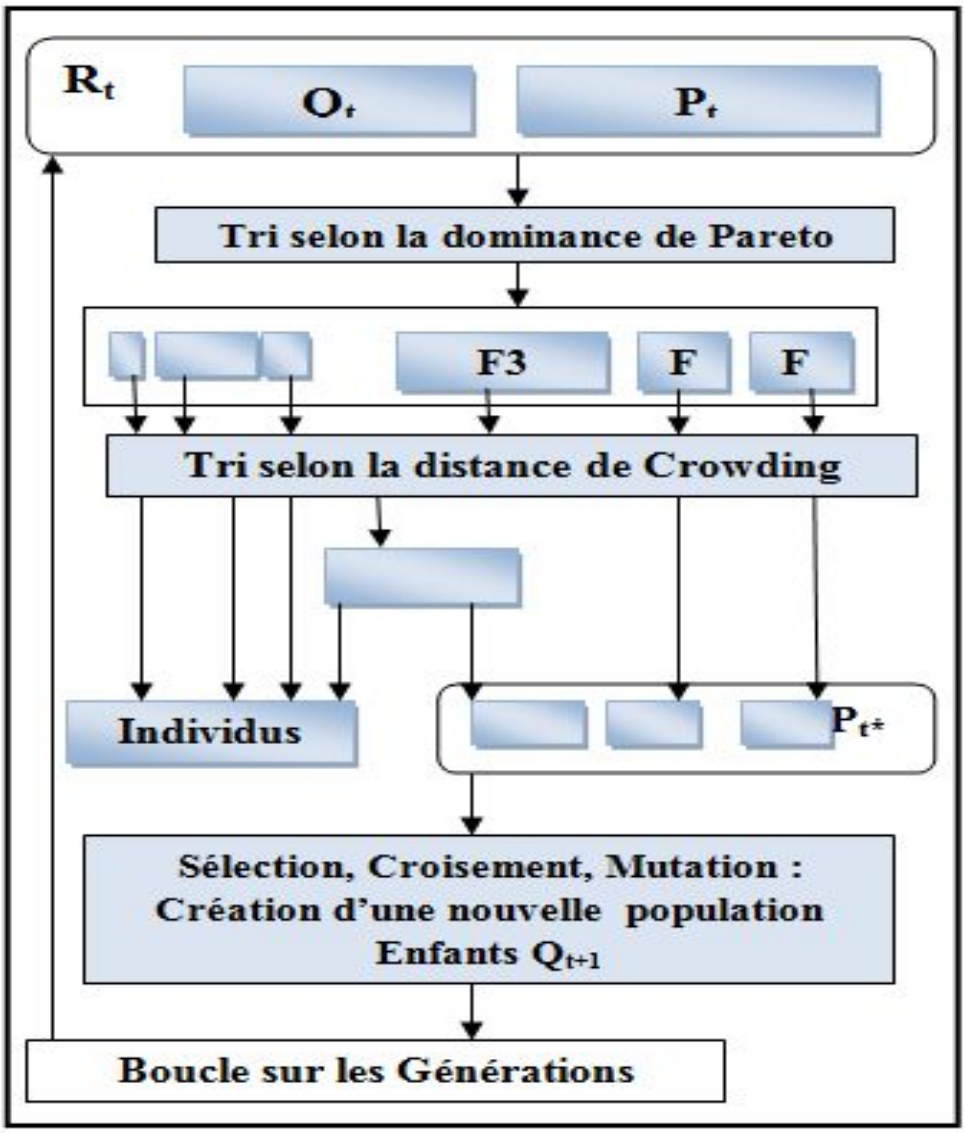


Figure 3.8 : principe de fonctionnement de l'algorithme NSGAII

B. NSGAIII

La structure de base reste similaire à NSGA-II avec des changements significatifs dans le mécanisme de sélection élitiste et la création de la population de la progéniture.

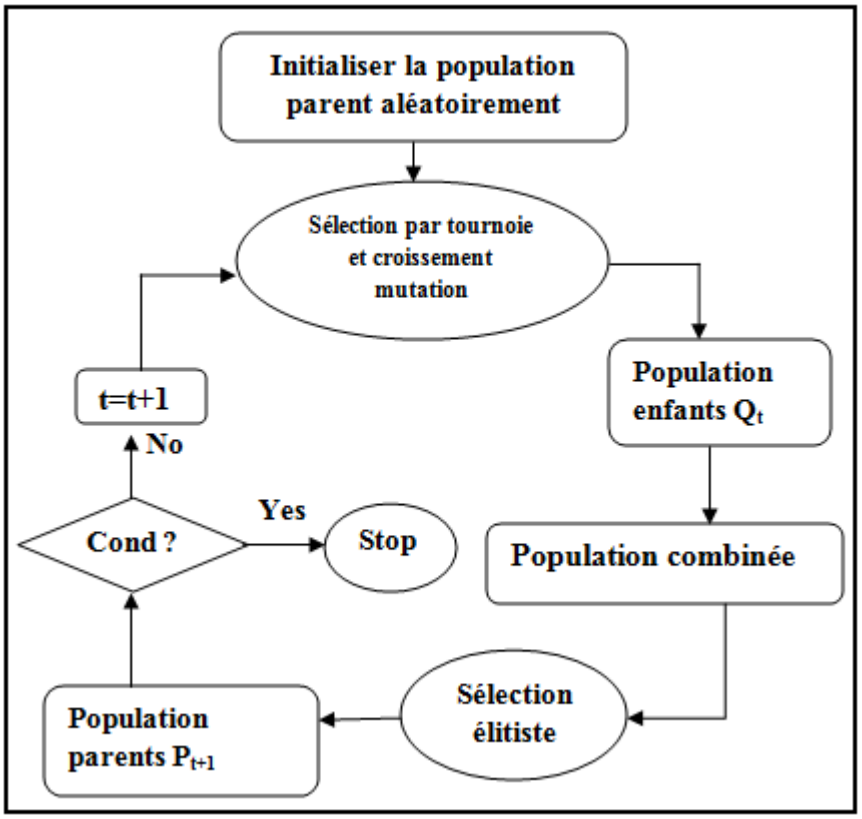


Figure 3.9: fonctionnement de l'algorithme NSGAIII

❖ **Avantage de NSGA-III**

1. L'étude NSGA-III original, Il a été démontré de bien travailler de trois à 15 objective et d'autres problèmes.

2. Un aspect clé de NSGA-III est qu'il ne nécessite aucun paramètre supplémentaire.

3. La méthode a également été étendue pour gérer les contraintes sans introduire de nouveau paramètre.

4. Cette étude a également mis en place une approche de calcul rapide par lequel l'ensemble de points de référence est mis à jour de manière adaptative à la volée sur la base du statut d'association de chaque point de référence sur un certain nombre de générations.

C. U_NSGAIII :

La méthode U-NSGA-III proposée peut conserver les caractéristiques de l'algorithme NSGA-III d'origine, comme l'a montré NSGA-III de bien travaillé sur trois ou plus objectives.

3.4. Comparaison des algorithmes étudiés :

Algorithme	MEGA		MOEA/D					NSGA		
variante	MEGA	PMEGA	MOEA/D classique	MOGA	MOGLS	PAλ	EASS	NSGAII	NSGAIII	U-NSGAIII
Avantage	-Applicable au problème mono- et multi- objectifs -pas de perte d'information associé au codage de structure Les Chromosomes Graph-Based - les résultats de MEGA pour les tests mono-objectif démontrer la capacité de l'algorithme pour explorer l'espace chimique donné - MEGA produire des solutions de qualité .	- PMEGA produire des solutions de qualité égale à MEGA. - Réduction de temps d'exécution à cause de parallélisassions et en fonction de la puissance de processeur utilisé - Exploiter les processeurs multi-core -PMEGA augmente accélération maximale	-la normalisation peuvent être incorporés dans MOEA / D pour traiter les objectives disparates échelles. -Il est très naturel d'utiliser des méthodes d'optimisation scalaires dans MOEA / D - L'efficacité de calcul - Évolutivité à de nombreux problèmes et de haute capacité de recherche pour les problèmes d'optimisation combinatoire.	-la capacité de traitement des PF non convexe	- MOGLS stocke l'ensemble des solutions actuelles CS et sa population externe EP (pas de besoin de mise à jour). - réduction de la complexité	- capable de générer un nombre arbitraire de vecteurs de poids, même si le nombre d'objectifs est supérieur à deux - Il peut ajuster automatiquement vecteurs de pondération pour disperser concave PF, alors assembler des vecteurs de pondération pour la PF convexe - une meilleure convergence et des solutions plus uniformément réparties que classique MOEA / D et NSGA-II	-une meilleur convergence des solutions obtenues à la vraie PF par rapport MOEA/D et NSGAII; -une meilleur couverture des solutions obtenues à la vraie PF; -la diversité des solutions obtenues.	-pas d'utilisation de stratégies d'élitisme, empêchent la perte de bonnes solutions	-NSGA-III , bien travailler de trois à 15 objective. - NSGA-III ne nécessite aucun paramètre supplémentaire. -La méthode a été étendue pour gérer les contraintes sans introduire de nouveau paramètre - NSGA-III utilise un ensemble de directions de référence pour maintenir la diversité	- U-NSGA-III de bien travaillé et efficacité sur problèmes mono- ou multi- ou beaucoup-objectives
Inconvénient	-problème de perte des solutions non dominé lors d'utilisation de mécanisme de sélection élitisme -problème de lenteur de temps d'exécution -dans les problèmes mono-objectifs focalisation sur l'affinité de liaison - la prudence lors de choix d'objectif pour éviter le surentraînement et génération d'individus sur-ajustement	-le grande nombre de sous population similaire causé des solutions similaire qui ne change pas la diversité -problème dans le mécanisme de sélection -le nombre d'itération nécessaire pour la convergence est élevé	- MOEA / D est incapable de produire un nombre arbitraire de vecteurs de poids lorsque le nombre d'objectifs est supérieur à deux; -les vecteurs de pondération classique dans MOEA / D restent inchangées pour les différentes formes des PF. -incapable de traiter des fronts de Pareto non convexe - la complexité	- pas de mécanisme pour garder la meilleure solution trouvée	- le cout de calcule de certain étape est augmenter par rapport MOEA/D	-besoin d'amélioration pour les problèmes multi objectif dure		-Des paramètres supplément aires doivent être définis pour garantir la diversité -moin de dévirsité par rapport MOEAD et NSGAIII moin de densité des solution et faible convergence vers le PF.	-difficulté de résoudre des problèmes mono et bi objectif -l'aléatoire de processus de sélection.	

Tableau 3. 2 : comparaison entre les algorithmes étudiés

Nous essayons ici d'étudié ces trois algorithme afin de choisir un entre ils pour l'appliquer sur notre problème de novo drug design a la suite nous détaillons l'algorithme choisie pour notre étude.

Conclusion

dans ce chapitre nous présentons l'optimisation multi et mono objectif généralement, puis les différentes méthode et approche d'optimisation dans notre domaine (de novo drug design), ce chapitre est divisé en deux partie , dans la première partie nous présentons l'optimisation combinatoire mono et multi objectif et les algorithme qui existe mais nous concentrons sur les algorithmes évolutionnaire, dans la deuxième partie les méthodes d'optimisation pour de novo drug design dans le cas multi et mono objectif avec les travaux de chaque méthode , cette partie est terminé par une étude de trois algorithmes afin de choisir un algorithme a appliquer en vue des avantages et des inconvénients de chaque algorithme, le chapitre suivant présente les détaille de l'algorithme choisis et les détaille d'implémentation

Chapitre 4

Les outils et les détaille de travaille

Introduction :

Afin de réduire le temps et le cout de conception d'un nouveau médicament la conception de médicament assisté par ordinateur est utilisée, cette technique permet la virtualité des étapes de conception et la vérification de certaine propriété de médicament conçue avant de le synthétiser.

La conception de novo de nouveau médicament est un processus qui nécessite l'optimisation de plusieurs objectif, pour cela un algorithme d'optimisation multi objectif est utilisé, l'algorithme le plus utilisé dans ce type de problème c'est l'algorithme évolutionnaire et ces variantes a cause de leur efficacité.

Nous essayons ici d'utiliser les algorithmes génétiques afin de réaliser un outil pour la conception de novo de nouveau médicament en utilisant un outil et une base de donné de la chemoinformatique, cet outil permet de concevoir une nouvelle molécule similaire a une molécule de référence (une molécule connue) et de vérifie certain propriété.

1. Les outils de la chemoinfomatique :

Les outils de la Chemoinformatique sont un ensemble de bibliothèques comprenant des codes sources pour différents algorithmes / fonctions qui permettent aux chemoinformaticians de développer leurs propres applications logicielles pour une utilisation possible dans la recherche de similarité structurelle, le criblage virtuel, l'exploitation minière de base de données, analyse des relations structure-activité… etc. Le développement d'open-source des outils de chimio a commencé il y a plus d'une décennie, et jusqu'à présent, de nombreux outils très fonctionnels ont été développés [1]. Certains outils ont été développés à partir de zéro, par exemple, kit de développement Chemistry (CDK) et Open Babel, tandis que d'autres, tels que RDKit [2] et Indigo [3] boîtes à outils, ont été faites open source en faisant don du code source en interne sous des licences libérales.

Voila un tableau qui représente les différentes outils de la chemoinformatique dans ce qui suite nous détaillerons 3 outils :

Sr. No.	Name	Link	Programming language (wrapper, if any)	Operating system(s)	License
1	Chemistry Development Kit (CDK)	http://sourceforge.net/projects/cdk/	Java	Platform independent	LGPL
2	RDKit	http://www.rdkit.org/	C++ (Python, Java and C# wrapper)	Mac, Windows, and Linux	BSD
3	Open Babel	http://openbabel.org/	C++ (Java, .NET platform, Perl, Python, and Ruby wrapper)	Windows, Mac OS X, Linux	GPL
4	Cinfony	https://code.google.com/p/cinfony/ https://github.com/cinfony/cinfony	Python, Jython	Platform independent	BSD
5	Small Molecule Subgraph Detector (SMSD)	http://www.ebi.ac.uk/thornton-srv/software/SMSD/	Java	Platform independent	Creative Common (CC)
6	Biochemical Algorithms Library (BALL)	http://www.ball-project.org	C++	Windows, Linux, and Mac OS X	LPGL and GPL
7	Indigo	http://www.ggasoftware.com/opensource/indigo	C++ (Python, Java, and C# wrappers available)	Windows, Linux, and Mac OS X	GPL
8	jCompoundMapper	http://jcompoundmapper.sourceforge.net/	Java	Platform independent	LPGL
9	chem^f	https://github.com/stefan-hoeck/chemf http://www.scala-lang.org/	Scala (runs on Java platform)	Platform independent	GPL
10	Cheminformatics in Python (ChemoPy)	https://code.google.com/p/pychem/downloads/list	Python	Linux and Windows	–
11	ChemmineR	http://www.bioconductor.org/packages/release/bioc/html/ChemmineR.html	Statistical programming environment "R"	Windows and Mac OS X	Artistic 2.0
12	Compound-Protein Interaction with R (Rcpi)	http://bioconductor.org/packages/release/bioc/html/Rcpi.html	Statistical programming environment "R"	Windows and Mac OS X	Artistic 2.0
13	Chemkit	http://wiki.chemkit.org/Main_Page	C++	Windows, Mac, and Linux	BSD

Tableau 4.1: les outils de la chemoinformatique. [1]

1. 1. Le CDK (Chimistry Development Kit):

1.1.1. Définition:

Le Chemistry Development Kit(CDK) est une bibliothèque open source java pour la Chemo et la bio informatique. Le CDK a été créé le 27-29 Septembre 2000 à l'Université de Notre Dame par Christoph Steinbeck, Egon Willighagen et Dan Gezelter, les développeurs de Jmol et JChemPaint à l'époque de fournir une base de code commune. La première publication du code source a été faite le 11 mai 2011. [2] Depuis plus de 75 personnes ont contribué au projet, [3] conduisant

à un ensemble riche de fonctionnalités, comme indiqué ci-dessous. La nouvelles CDK produit entre le 2004 et 2007 était la lettre d'information du projet dont tous les articles sont disponibles à partir d'un archive publique. [4]

C'est la bibliothèque la plus utilisé jusqu'a aujourd'hui par la majorité des projets comme (JChemPaint , SENECA , NMRShiftDB , Padel descripteur , Jmol , JOELib, Nomen, Coffre-Base).

1.1.2 . Les fonctionnalités :

- **Pour la Chemoinformatique** :

- la modification et la génération du diagramme 2D.
- génération de géométrie 3D.
- requêtes de recherche de sous-structure en utilisant des structures exactes et la similarité SMARTES.
- calcul des descripteurs QSAR. [5]
- calcul d'empreintes digitales. [6]
- calculs de champ de force.
- de nombreux formats d'entrée / sortie chimiques, y compris SMILES, CML et format MDL.
- générateurs de structure. [7]

- **Pour la bioinformatique :**

- La détection de site actif de protéine
- la détection de ligand idoine [8]
- l'identification des métabolites [9]
- bases de données de la voie
- 2D et 3D protéines descripteurs [10]

1.1. 3. Les travaux connexes :

Le paradigme de workflow permet aux scientifiques de créer des workflows flexible génériques en utilisant une interface graphique dans laquelle différents types de nœuds ou de composants de logiciel sont reliés par des arcs ou des tubes [11], qui peuvent ensuite être adaptés à l'évolution des besoins.

Les Workflows sont de plus en plus utilisés dans les recherche sur la chemoinformatique , Ils sont largement utilisées pour de multiples applications de la chemoinformatique telles que le criblage virtuel, le docking moléculaire, la conception de novo de médicaments, l'étude QSAR, etc.

Les plus connues des systèmes de workflow open-source dans le domaine de la chemoinformatique sont Taverna, KNIME, et Cancer Grid (une extension de Taverna).

a) Taverna :

Taverna est une open source et le domaine du système de gestion de flux de travail indépendant, c'est une suite d'outils utilisés pour concevoir et exécuter des workflows scientifiques et aider dans les expérimentations in silico. Taverna a été créé par l'équipe de myGrid. Les outils de Taverna comprennent le Workbench (application client de bureau), l'outil de ligne de commande (pour une exécution rapide des flux de travail à partir d'un terminal), le serveur (pour l'exécution à distance des flux de travail) et le plugin Player (interface Web pour la présentation des flux de travail pour l'exécution à distance). [12]

Pour étendre les fonctionnalités de Taverna dans le domaine de la chemoinformatique, l'outil CDK a été combiné au système de workflow Taverna.

Ce plug-in CDK-Taverna [13] fournit 164 différents nœuds appelés "travailleurs". CDK-Taverna a le potentiel de devenir une alternative libre aux outils existants de workflow propriétaires.

b) KNIME (Konstanz Information Miner) :

Le Konstanz information Miner est un environnement modulaire qui permet un montage facile et visuel avec une exécution interactive d'un pipeline de données. Il est conçu comme un enseignement, la recherche et la collaboration plate-forme, ce qui permet une intégration facile de nouveaux algorithmes, la manipulation des données ou des méthodes de visualisation en tant que nouveaux modules ou nœuds. [14] Figure 1.3 montre une capture d'écran d'un flux d'analyse.

KNIME a été développé par une équipe d'ingénieurs à l'Université de Constance (Allemagne); sa combinaison avec le CDK (KNIME-CDK [97]) permet d'offrir une large gamme de fonctionnalités et d'applications dans le domaine de la chemoinformatique (convertir des formats moléculaires, générer des fingerprints, etc.). [15]

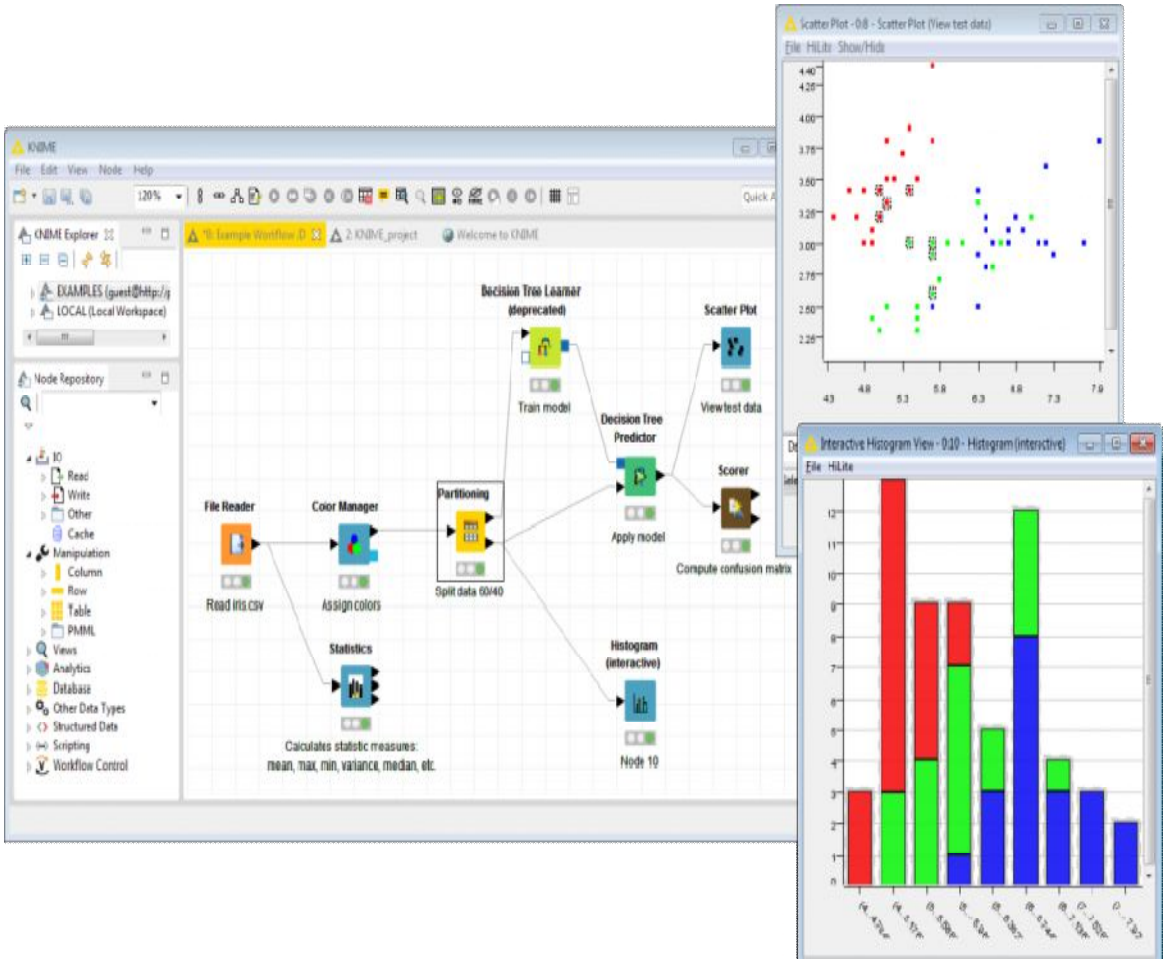


Figure 4.1 : Un exemple courant d'analyse à l'intérieur KNIME. [1]

1.1.4. Comment la Library CDK est organisé :

Les classes contenues dans la section racine du package de la hiérarchie de la CDK "org.openscience.cdk" sont toutes les représentations formalisées des concepts chimiques de base tels que des atomes, des liaisons, des molécules, etc. La figure 3 montre un diagramme UML expliquant la hiérarchie d'héritage et les dépendances entre les classes fondamentales de la CDK. Ils montrent le rôle central de la classe ChemObject, qui est la superclasse de toutes les autres classes et fournit des méthodes pour stocker les propriétés les plus complexes pour tout objet CDK dérivé.

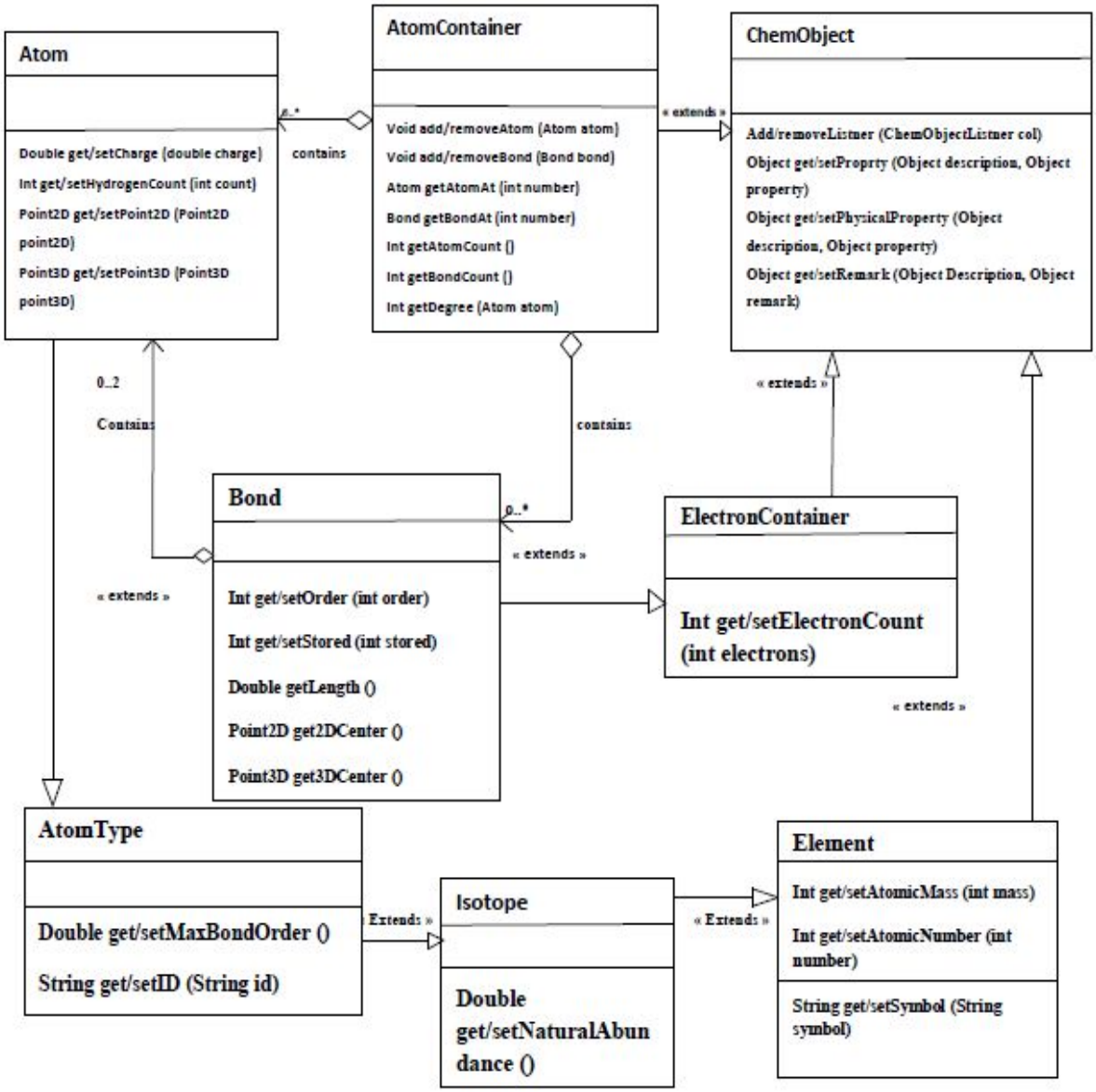


Figure 4.2. Diagramme UML expliquant la hiérarchie d'héritage et les dépendances entre les classes fondamentales de la CDK. [14]

La première et probablement la plus évidente chaîne d'héritage à mentionner dans les classes de base il qu'Atom étendant AtomeType étendant Isotope étendant Elément. Ceci est non seulement logique d'un point de vue chimique, mais fournit également la base d'un mécanisme simple pour la création d'Atoms, AtomTypes, Isotopes, et des éléments basés sur des sous-classes d'une même classe d'outil IsotopeFactory.

Placer l'Atom dans une longue chaîne d'héritage fournit des points d'accès au centre vers les différents niveaux d'information. Un autre niveau d'abstraction est incorporé par l'AtomContainer et l'ElectronContainer. L'ElectronContainer forme la base pour les

constructions telles que les liaisons et les Orbitals, alors que l'AtomContainer est le stockage envisagé pour Atomes ainsi que leurs liaisons c'est la superclasse pour Rings, Molécules et substructions.

1.1.5. Quelque exemple sur l'utilisation de CDK : [15]

- **Lire un fichier « .mol »**

```
//read file
File sdfF = new File("C:/acid/5558.mol");
   IteratingMDLReader read = new IteratingMDLReader(new FileInputStream(sdfF), DefaultChemObjectBuilder.getInstance());
    Molecule mol = null;
    while (read.hasNext()) {
       mol = (Molecule) read.next();
```

- **Créer un atome de carbone**

```
IAtom atom = DefaultChemObjectBuilder.getInstance().newAtom("C");
```

- **Créer le format SMILES. Pour cela, utiliser le Parser Smiles:**

```
SmilesParser sp = new SmilesParser(DefaultChemObjectBuilder.getInstance());
IMolecule molecule = sp.parseSmiles("CCCC(CC)CC=CC=CC");
```

- **Générer des fingerprints :**

```
BitSet fingerprint = Fingerprinter.getFingerprint(molecule)
```

1.2. RDKit :

1.2.1. Définition :

Le RDKit a été développé et utilisé pour la découverte Rational pendant la période 2000-2006, pour construire des modèles prédictifs pour l'absorption, distribution, métabolisme, l'élimination, la toxicité et l'activité biologique. [1]

1.2.2. Les fonctionnalités :

RDKit offre diverses fonctionnalités telles que :

- Permet de lire des différents formats chimique I / O, y compris : SMILES / SMARTS et le format de données de structure (SDF).
- arbre de données Thor (TDT)
- Corina mol2 et Protein Data Bank (PDB)
- la recherche de la sous-structure
- SMILES canoniques
- transformations chimiques (par exemple, supprimer substructures correspondant)
- réactions chimiques
- sérialisation moléculaire
- sélection / diversité de similitude
- pharmacophores 2D
- analyse hiérarchique sous-graphe / fragment
- détermination d'échafaudage Bemis et Murcko
- procédure rétro synthétique combinatoire d'analyse (RECAP)
- forme à base de similitude
- forme basée sur l'alignement
- filtrage de groupe fonctionnel
- Bibliothèque de descripteur moléculaire

1.3. Open Babel :

1.3.1. Définition :

C'est une bibliothèque open source C++, elle est construite pour soutenir inter conversion entre les différents formats de fichiers utilisés dans chimio, modélisation moléculaire, et des domaines connexes. [16]

1.3.2. Fonctionnalités :

- effectuer la lecture, l'écriture et inter conversion de plus de 111 formats de fichiers chimique.
- Il prend en charge les fichiers de molécules de filtrage
- calcule des descripteurs de groupes de contribution tels que LogP, la surface polaire (PSA).et réfractivité molaire (MR).
- Il fournit des empreintes moléculaires extensible et fonctions de la mécanique moléculaire.

- Il extraire des informations supplémentaires en plus de la structure moléculaire. Par exemple :
  - les champs de propriété peuvent être lus à partir de fichier SDF.
  - des informations de cellule unité peuvent être extraites à partir de fichiers CIF.
  - les fréquences de vibration peuvent être extraites des fichiers journaux de chimie computationnelle.
- Il offre une solution pour faire face à l'augmentation du nombre de formats de fichiers
- la recherche de conformer
- représentation 2D
- le filtrage
- la conversion par lots
- l'infrastructure et la recherche par similarité. Pour les développeurs de logiciels

2. Les bases de donné de la chemoinformatique :

La base de données est un outil permettant de stocker et de retrouver l'intégralité de données brutes ou d'informations en rapport avec un thème ou une activité, Les bases de données qui sont largement utilisées en chemoinformatique avec leurs liens officiels et une brève description pour chacune d'elles sont illustrés dans le tableau 4 ci-dessous:

N°	Bases de données	Liens officiels	Brève description
1	QSAR DataBank	http://qsardb.org/repository/	QSAR DataBank
2	VAMMPIRE	http://vammpire.pharmchem.uni-frankfurt. de/vammpire/	Conception et optimisation de médicaments basés sur la structure
3	PubChem3D	https://pubchem.ncbi.nlm.nih.gov/	Référentiel ouvert pour les petites molécules et leur activité biologique
4	MMsINC	http://mms.dsfarm.unipd.it/MMsINC	Base de données des structures chimiques
5	CREDO	http://marid.bioc.cam.ac.uk/credo	Base de données des interactions protéine ligand
6	ChemBank	http://chembank.broadinstitute.org/	Base de données de petites molécules
7	DrugBank	http://www.drugbank.ca/	Informations sur la cible biologique et les médicaments
8	ChemDB	http://cdb.ics.uci.edu	Base de données des petites molécules
9	ChemMine	http://chemminedb.ucr.edu/	Base de données d'extraction de composés pour chimie du génome
10	National Cancer Institute (NCI) 3D database	http://www.cancer.gov/cancertopics/pdq/cancerdatabase	Base de données des médicaments Anticancéreux
11	ZINC	http://zinc.docking.org	Base de données gratuite des petites molécules disponibles dans le commerce
12	ChEMBL	https://www.ebi.ac.uk/chembl/	Molécules bioactives avec des propriétés "drug-like"
13	Therapeutic Target Database (TTD)	http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp	Base de données des médicaments
14	PharmGKB	http://www.pharmgkb.org/	Ressources de connaissances pharmacogénomique
15	STITCH	http://stitch.embl.de/	Ressources pour explorer les interactions connues et prévues des produits chimiques et des protéines
16	SuperTarget	http://bioinf-apache.charite.de/supertarget_v2/	Base de données des médicaments et des protéines
17	ChemSpider	http://www.chemspider.com/About.aspx	Base de données des structures chimiques

Tableau 4.2. Liste de base de données connue en chemoinformatique.

Dans notre travaille nous avons utilisé la base de données « ChemSpider » pour obtenir les fragments acide et amine nécessaires qui sont téléchargé en format «.mol».

- **ChemSpider :**

ChemSpider est un moteur de recherche de chimie. C'est une base de données gratuite. Elle offre l'accès à des millions de structures chimiques et intègre une multitude d'autres prestations de service en ligne. ChemSpider est la plus riche source d'information unique de la chimie.

Il a été construit avec l'intention de l'agrégation et de l'indexation chimique des structures et leurs informations associées dans un seul référentiel consultable et rend disponible à tout le monde, sans frais. [17]

3. Appliquer L'algorithme MOGA (Multi Objective genetic algorithm) sur problème DND avec les détailles de notre travaille:

3.1. algorithme génétique (GA) :

Algorithme génétique (GA) est une technique stochastique aléatoire qui est approprié pour un certain nombre de problèmes d'optimisation [7], qui imite l'idée de Darwin du processus d'évolution naturelle [7]. GA est la technique la plus populaire de l'algorithme évolutionnaire.

Lorsque le GA utilise une fonction unique objective pour arriver à la solution optimale, c'est l'algorithme génétique mono objectif. Lorsque le GA utilise plusieurs objectifs parfois contradictoires pour arriver à la meilleure solution, c'est le GA multi-objectifs.

L'algorithme génétique multi objectif MOGA est classé dans le cadre des algorithmes MOEA/D (multi objectif evolutionary algorithme based decomposition), l'approche de décomposition utilisé dans ce type d'algorithme c'est l'approche somme pondérée ou l'approche de Pareto rang.

Dans l'approche de somme pondéré un poids est donné pour chaque fonction objectif et le problème multi objectif est convertit vers un problème mono objectif afin de prendre des meilleures solutions (la méthode composite).

Dans l'approche de Pareto rang, toute la population est classée selon la règle de dominance et une valeur de remise en forme est attribuée à chaque solution en fonction de son rang dans la population, au lieu de sa valeur de fonction objectif réel.

3.2. principe de MOGA :

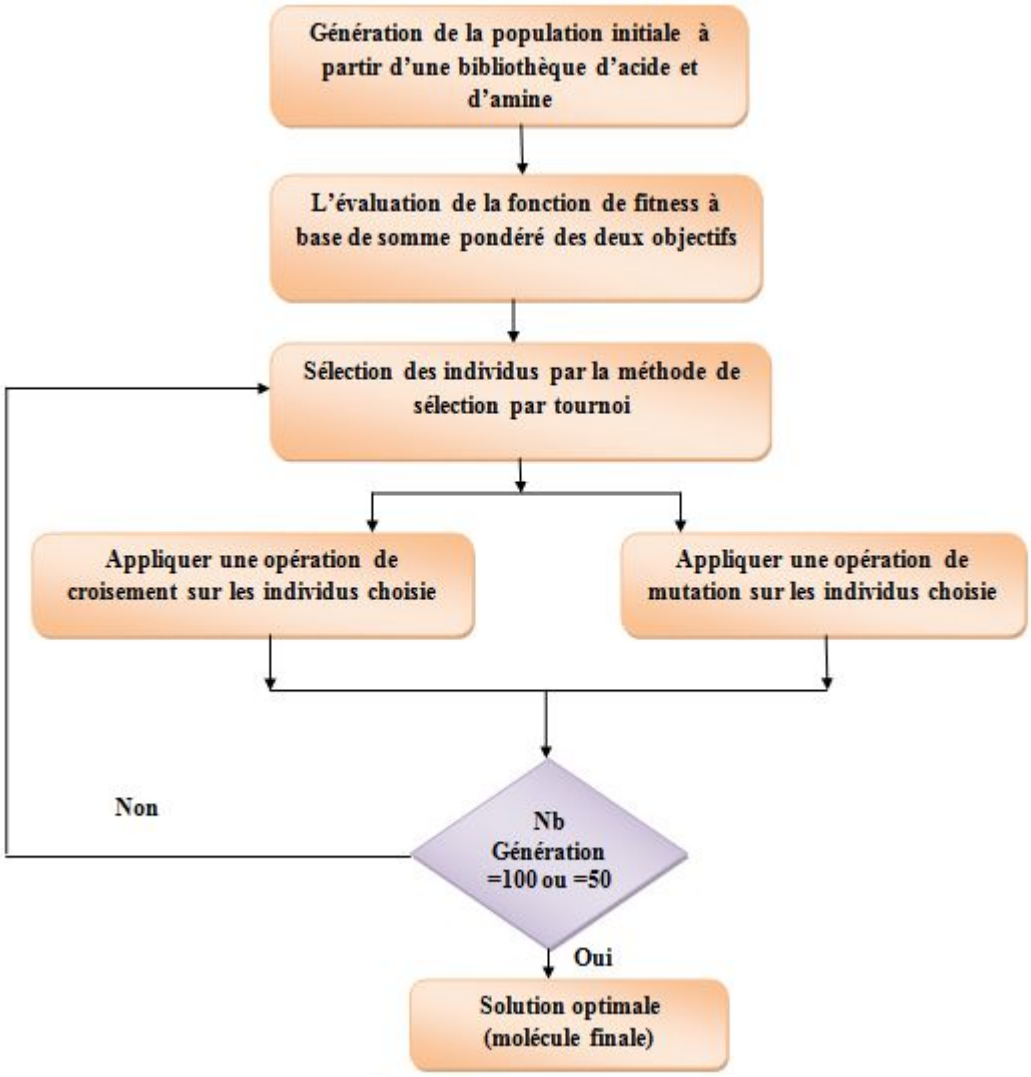


Figure 4.3 : le principe générale de l'algorithme MOGA

3.3. Notre Proposition :

3.3.1. L'idée de base :

L'idée est d'utiliser l'algorithme génétique multi objectif afin de réaliser la conception de novo des nouveaux médicaments en utilisant l'approche de somme pondéré. La solution (molécule) n'est pas forcement la meilleure solution qui existe. Au lieu de cela, la molécule doit être la meilleure solution optimale qui répond aux objectifs de conceptions:

Notre étude est basée sur 2 objectifs, Premièrement il faut vérifier la biodisponibilité de la molécule conçue défini par la règle de Lipinski 5, puis la

similarité de cette molécule a une molécule de référence connue (la propriété de drug_likness) calculé par le coefficient de Tanimoto.

3.3.2. Formulation de problème de DND :

Nous avons un nombre k de Library de fragment :

$L_i = \{f_{i1},\ldots, f_{il}\}$

Telle que $L_{i=1,\ldots,k}$ l'ensemble de Library de fragment , les f_i c'est l'ensemble des fragments moléculaire dans la Library des fragments i, l c'est la taille d'une Library des fragments .

Et nous avons :

$F= \{F_1 ,F_2 ,\ldots,F_m\}$ un ensemble de m fonction objectif .

Le but principale dans le problème de conception de nouveau médicament est de créer des nouveaux molécules M_{new} en prenants des molécules a partir des k Library, cette nouvelle molécule maximise la valeur de l'ensemble des m fonctions objectifs. Mathématiquement notre problème consiste de trouver les n combinaisons optimales $(f^*_{1*1},\ldots,f^*_{l*l})$ des fragments à partir des k Library de fragment$\{L1 \times L2 \times \ldots \times LK\}$, qui optimise les m fonction objectifs M_{new}= Max $\{F_1,F_2 ,\ldots,F_l\}$.

3.3.3. Discussion :

Dans l'outil proposé, l'algorithme génétique multi-objectif est appliquée pour trouver des molécules de médicaments comme optimales, qui sont similaires à une molécule de référence connue à l'aide de novo conception de médicaments.

L'application du système proposé est représentée par la conception de novo des molécules de médicament analogue à partir d'une banque de fragments d'acides et d'amines extraits de médicaments connus. La conception des molécules ont été guidés en utilisant deux fonctions objectives, un score de similarité (Tanimoto similarité) à une molécule de référence connue (lidocaïne et Furano-pyrimidine) et un score de biodisponibilité orale (la règle 5 de Lipinski). Le système multi-objectifs de novo drogue conception proposée pourrait être utilisée pour concevoir des molécules drug-like pour diverses maladies.

3.4.4. Les étapes de l'algorithme proposé :

1) génération de la population initiale :

La population initiale est constituée d'un ensemble de 50 chromosomes, chaque chromosome est représenté sous la forme d'un vecteur de nombres entiers, la valeur représentative de leur identité dans les banques de fragments d'acide et amine. A la fin de la conception actuelle, seuls deux chromosomes de gènes sont utilisés. Chaque chromosome représente une molécule de médicament possible, comme cela peut être synthétisé à partir des deux fragments (un acide et une amine) qui composent les différents gènes. Ici, le premier gène représente le fragment acide et le second gène représentent le fragment d'amine. La banque de fragments utilisés dans la conception des chromosomes se compose de 31 acides et 157 amines extraits de médicaments connus [13].ces fragments sont téléchargé a partir de site officielle de la base de donné moléculaire Chemspider (http://www.chemspider.com/).

La génération de la population initiale se fait d'une façon aléatoire comme suite

```
indiceFragment1 : entier ;
indiceFragment2 : entier ;
Chromosome : vecteur d'entier ;
Tant que la taille des k Library des fragments  n'est pas attient faire
Pour i allant de 1 à taille de Library acide faire
Pour j allant de 1 à taille de Library amine  faire
indiceFragment1= choisir aléatoirement l'indice d'acide par la formule
    (math.random ()*100) mode 2 ;
indiceFragment2= choisir aléatoirement l'indice d'amine par la
    formule (math.random ()*100) mode 2 ;
Chromosome= Chromosome [indiceFragment1] [indiceFragment2] ;
```

Voila un schéma qui représente la représentation des chromosomes sous forme d'un vecteur d'entier :

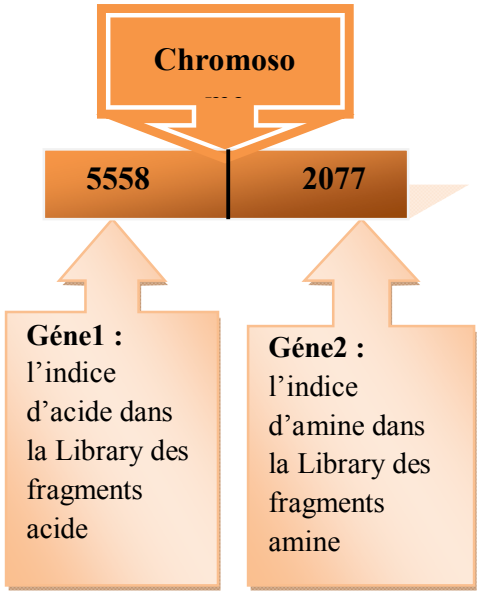


Figure 4.4 : la représentation des chromosomes en nombre entier

2) Evaluation initiale des individus (la fonction objective) :

Notre fonction objectif est de concevoir des nouveaux molécules d'une très bon valeur de biodisponibilité orale (OBA), et qui est également similaire a une molécule de référence connue(Tcoef) , donc notre problème est un problème d'optimisation multi objectif qui consiste a maximiser deux objectif simultanément, et comme nous l'avons indiquer précédemment nous utilisons une méthode d'optimisation composite qui consiste a combiner les objectifs en un seul et de donner un poids pour chaque objectif donc la formule générale de notre fonction objectif est comme suite :

$f(y) = \sum_{i=1}^{m} a_i * f_i(y)$

Ou les a_i sont les poids associé a chaque objectif telle que $a_i \geq 0$ et $\sum_{i=0}^{N} a_i = 1$, y est le vecteur de paramètre de la fonction f_i la molécule synthétisé, m le nombre d'objectif (m=2), pour notre travaille la fonction objectif est comme suite :

$$F(y) = a_1 * OBA(y) + a_2 * TCoeff(y, Mref)$$

Ou y c'est la solution, M_{ref} c'est la molécule de référence

a) La biodisponibilité orale (Oral Bio Avialibility OBA) :

La **biodisponibilité** se définit comme étant la **fraction** de la dose de médicament administré qui atteint la circulation générale **et** la **vitesse** à laquelle elle l'atteint [19]. Dans notre travaille l'évaluation de cette propriété ce fait par les règles de 5 de Lipinski. La règle de Lipinski de cinq également connu comme la règle de Pfizer de cinq ou tout simplement la règle de cinq (RO5) est une règle de base pour évaluer drug likeness. La règle a été formulée par Christopher A. Lipinski en 1997, basée sur l'observation que les médicaments administrés par voie orale plus sont relativement petites et modérément lipophiles molécules. [20] [21]. La règle décrit les propriétés moléculaires importantes pour la pharmacocinétique d'un médicament dans le corps humain, y compris leur absorption, la distribution, le métabolisme et l'excrétion («ADME»). Toutefois, la règle ne signifie pas si un composé est pharmacologiquement actif.

La règle de Lipinski stipule que, en général, un médicament actif par voie orale n'a pas plus d'une violation des critères suivants:

- Le nombre d'hydrogène donneur <=5
- Le nombre d'hydrogène accepteur <=10
- La masse moléculaire<=500
- Le coefficient de partage entre l'eau et un solvant organique logP<=5

Notez que tous les nombres sont multiples de cinq, qui est l'origine du nom de la règle.

Pour notre travaille les contraintes sont comme suite :

Si tous les paramètres sont vitrifiés alors :
Score OBA=1
Si un des paramètres n'est pas vérifié alors :
Scor OBA=0.75
Si deux paramètres non pas vitrifiés alors :
Score OBA=0.5
Si trois paramètres ne sont pas vérifiés alors
Score OBA=0.25
Si aucun paramètre n'est vitrifié alors :
Score OBA=0

b) Le coefficient de Tanimoto :

Tanimoto c'est le coefficient de similarité (la similitude 2D avec une molécule de référence). C'est une mesure d'évaluation de similarité chimique structurelle entre deux molécules [22], [23]. Il est utile de concevoir des molécules avec un ensemble prédéfini de propriétés, dans ce cas, la similitude chimique à une molécule de référence connue, de sorte que les molécules nouvellement conçus auront similaire à celle de la molécule de référence fonction / activité. Le score Tanimoto coefficient de similarité est compris entre 0 et 1. Un score de similarité élevée signifie une bonne similarité à la molécule de référence. Un score nul signifie qu'il n'existe pas de similarité entre 2 molécules.

c) La méthode de sélection :

Généralement dans les algorithmes génétiques il existe plusieurs méthodes de sélection :

I. Sélection par roulette (*wheel*)

Les parents sont sélectionnés en fonction de leur performance. Meilleur est le résultat codé par un chromosome, plus grandes sont ses chances d'être sélectionné. Il faut imaginer une sorte de roulette de casino sur laquelle sont placés tous les chromosomes de la population, la place accordée à chacun des chromosomes étant en relation avec sa valeur d'adaptation [24]. Cette roulette est représentée par la figure 1

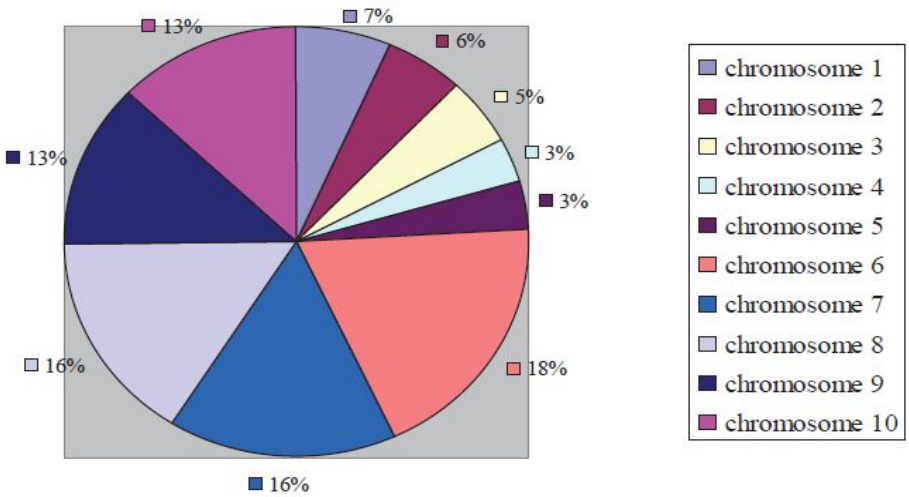


Figure 4.5: **Exemple de sélection par roulette. [25]**

Ensuite, la bille est lancée et s'arrête sur un chromosome. Les meilleurs chromosomes peuvent ainsi être tirés plusieurs fois et les plus mauvais ne jamais être sélectionnés. Cela peut être simulé par l'algorithme suivant :

1. On calcule la somme $S1$ de toutes les fonctions d'évaluation d'une population.
2. On génère un nombre r entre 0 et $S1$.
3. On calcule ensuite une somme $S2$ des évaluations en s'arrêtant dès que r est dépassé.
4. Le dernier chromosome dont la fonction d'évaluation vient d'être ajoutée est sélectionné.

II. Sélection par tournoi

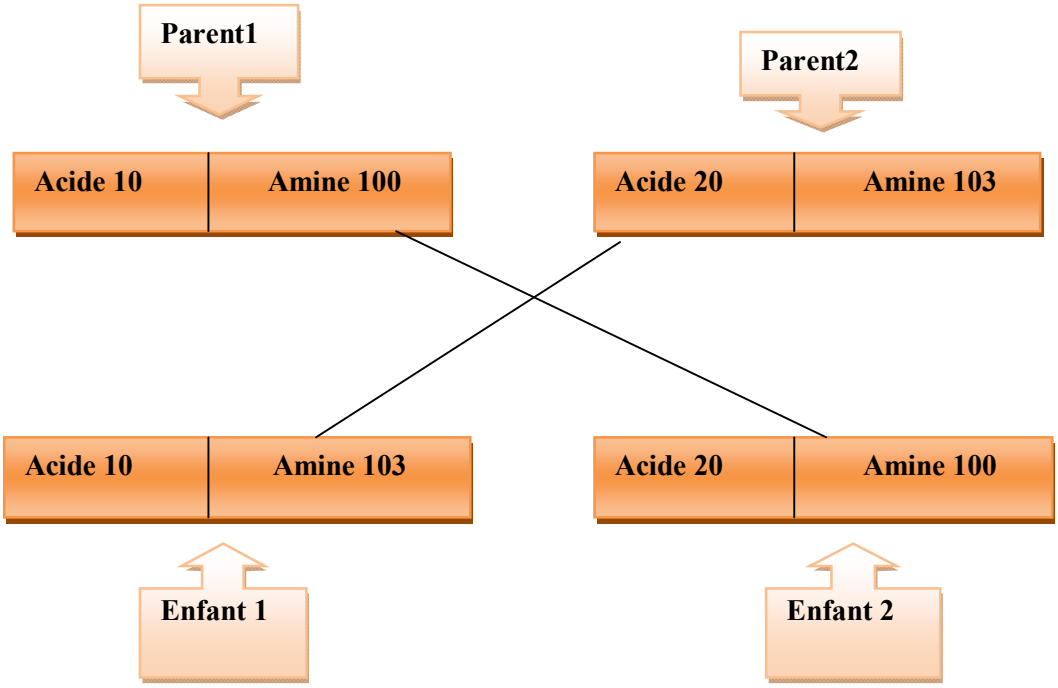
Sur une population de m chromosomes, on forme m paires de chromosomes. Dans les paramètres de l'AG, on détermine une probabilité de victoire du plus fort. Cette probabilité représente la chance qu'a le meilleur chromosome de chaque paire d'être sélectionné. Cette probabilité doit être grande (entre 70% et 100%). A partir des m paires, on détermine ainsi m individus pour la reproduction [24].

III. Elitisme

A la création d'une nouvelle population, il y a de grandes chances que les meilleurs chromosomes soient perdus après les opérations d'hybridation et de mutation. Pour éviter cela, on utilise la méthode d'élitisme. Elle consiste à copier un ou plusieurs des meilleurs chromosomes dans la nouvelle génération. Ensuite, on génère le reste de la population selon l'algorithme de reproduction usuel. Cette méthode améliore considérablement les algorithmes génétiques, car elle permet de ne pas perdre les meilleures solutions [24].

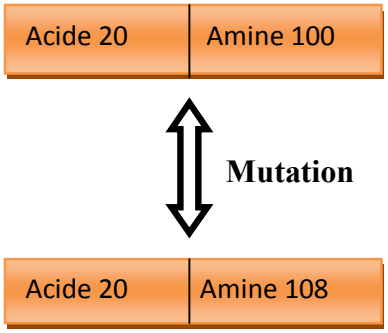
d) Le croisement :

Le croisement a pour but d'enrichir la diversité de la population en manipulant la structure des chromosomes. Classiquement, les croisements sont envisagés avec deux parents et génèrent deux enfants. Dans notre cas le croisement se fait comme illustre le schéma ci-dessus :



e) L'opérateur de mutation :

L'opérateur de mutation apporte aux algorithmes génétiques la propriété de périodicité de parcours d'espace. Cette propriété indique que l'algorithme génétique sera susceptible d'atteindre tous les points de l'espace d'état, sans pour autant les parcourir tous dans le processus de résolution. Ainsi en toute rigueur, l'algorithme génétique peut converger sans croisement, et certaines implantations fonctionnent de cette manière. Les propriétés de convergence des algorithmes génétiques sont donc fortement dépendantes de cet opérateur sur le plan théorique. Dans notre cas l'opérateur de mutation se fait comme illustre le schéma ci-dessus :



3.4.Le fonctionnement de notre outil :

Nous avons trois Library de fragments, la première Library contient 28 acides, la deuxième Library contient 108 amines, les acides et les amines sont des fragments extrait à partir de médicament connus, la troisième Library contient deux molécules ce sont les molécules de références : furano-pyrimidine et lidocaine pour calculer la similarité dans les deux cas mono et multi objectif.

3.4.1. Combinaison acide/amine :

La formule générale d'un acide est comme suite **R-COOH**

La formule générale d'une amine est comme suite **R'-NH2**

Notre idée est de trouver l'emplacement des atomes OH dans l'acide correspondent et l'emplacement de l'atome H dans le fragment amine on utilisons le package SMARTS de CDK ,puis de trouver les bonds (les liaisons)entre l'atome C et les atomes O,H dans l'acide et la liaisons entre N et H dans l'amine, puis d'enlever les atomes O et H à partir d'acide et H à partir d'amine avec les liaisons de chaque atome ,puis de sauvegarder l'emplacement ou nous enlevons , et enfin de créer une liaison entre l'atome C d'acide et l'atome N d'amine :

R-COOH + R'-NH2 ⟶ R-OCNH - R'+H2O

3.4.2. Les évaluateurs :

Nous avons deux évaluateurs le premier correspond la règle de 5 de Lipinski qui se compose de 4 propriétés calculé par le package CDK "org.openscience.cdk.qsar", le deuxième correspond le coefficient de Tanimoto ce dernier est calculé aussi par le même package par rapport a une molécule de référence.

3.4.3. Application de l'algorithme :

Appliquer les opérateurs de l'algorithme génétique pour générer d'autre molécule et trouver l'optimum.

3.4.4. Affichage de la solution optimale :

L'affichage ce fait par les packages de jar guhautils, la structure optimale est affiché dans un panel.

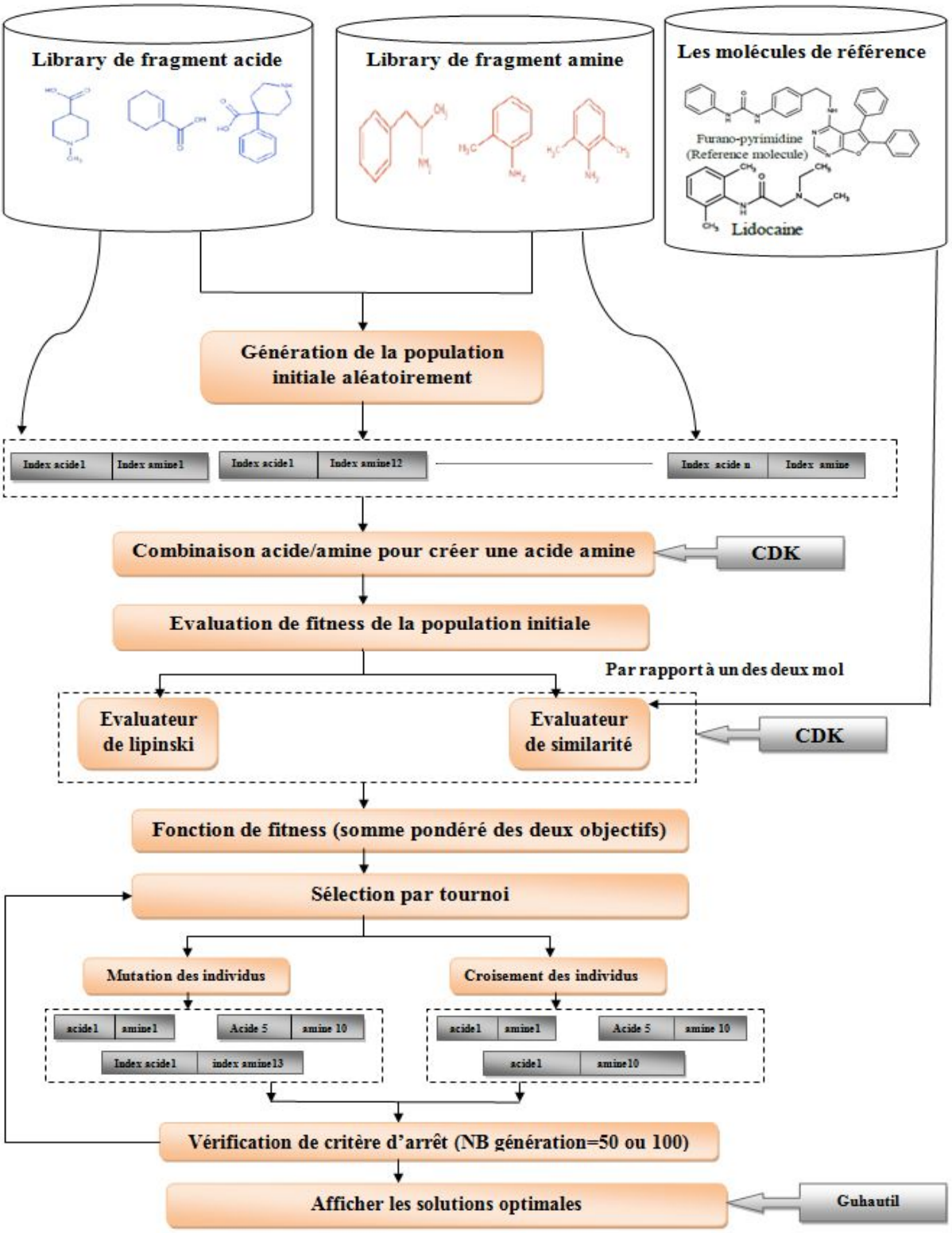


Figure 4.6 : le mode de fonctionnement de notre outil MOGA-DND

Conclusion :

Dans ce chapitre nous avons fait le tour des différents outils de la chemoinformatique, en spécifiant leurs définitions et leurs fonctionnalités et en détaillant l'outil CDK choisie pour notre travaille. Puis, nous avons entamé les bases de données de la chemoiformatique avec une définition de la base Chemspider car c'est la base de données utilisé pour extraire les fragments. Ensuite, nous avons présenté l'algorithme MOGA (définition et principe) .A la suite, nous avons formulé mathématiquement notre problème.. Enfin, les détaille de notre travaille sont présenter ; l'idée de base avec une discussion générale les étapes de l'algorithme proposé (MOGA_DND) et le mode de fonctionnement d'outil réalisé. Dans le prochain chapitre nous présentons les résultats obtenus en appliquant l'algorithme proposé.

Chapitre 5

La validation et les résultats obtenus

Introduction :

Dans ce chapitre nous présentons les résultats obtenus par l'application de l'algorithme proposé, deux cas d'étude sont effectué afin d'évaluer la performance de l'outil proposé par application mono et multi objectif, les résultats sont représenté sous forme des tableaux et des graphes, avec une visualisation 2D de la structure des meilleures molécules et les interfaces finaux.

1. Les données :

Dans notre étude expérimentale nous utilisons 3 Library de fragment une Library de fragment acide constitué de 31 acides, une Library de fragment amine constitué de 157 amines, et une autre Library constitué de 2 molécules de référence tous les fragments sont téléchargé en format .mol :

1.1. Les Library de fragments :

Les nouvelle molécules qui satisfaire les propriété étudiés la similarité et la biodisponibilité orale sont généré par des réactions chimique entre les fragments acide et les fragment amine, une concaténation entre les acides carboxylique (R-COOH) et les amines primaire ou secondaire (R-NH2 ,R-NH2-R') se fait au cour d'exécution de notre outil, une évaluation de la fonction objectif est fait par l'outil CDK , puis les molécules qui satisfaire les propriétés sont sauvegardé dans un fichier.mol et affichés dans un panel (la visualisation 2D ce fait a l'aide de l'outil guhautil associé au CDK) , ci-dessus une figure qui représente la visualisation 2D d'une molécule par guhautil :

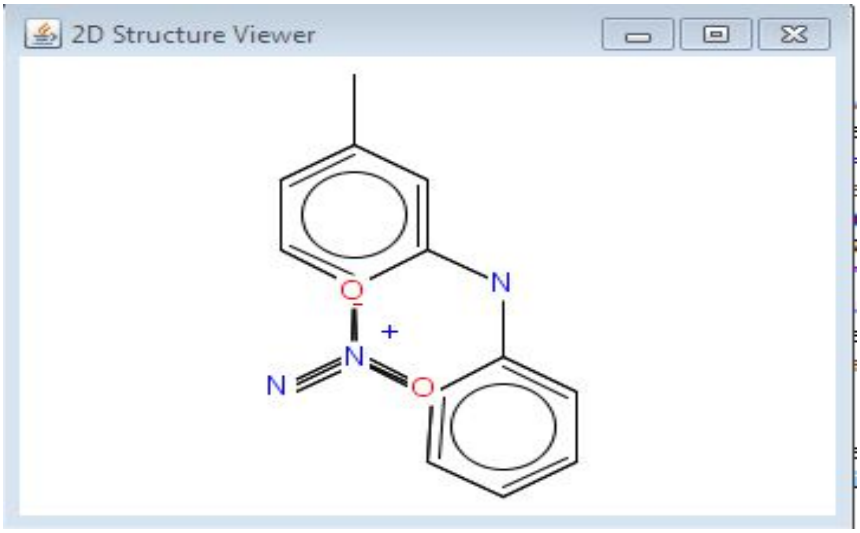


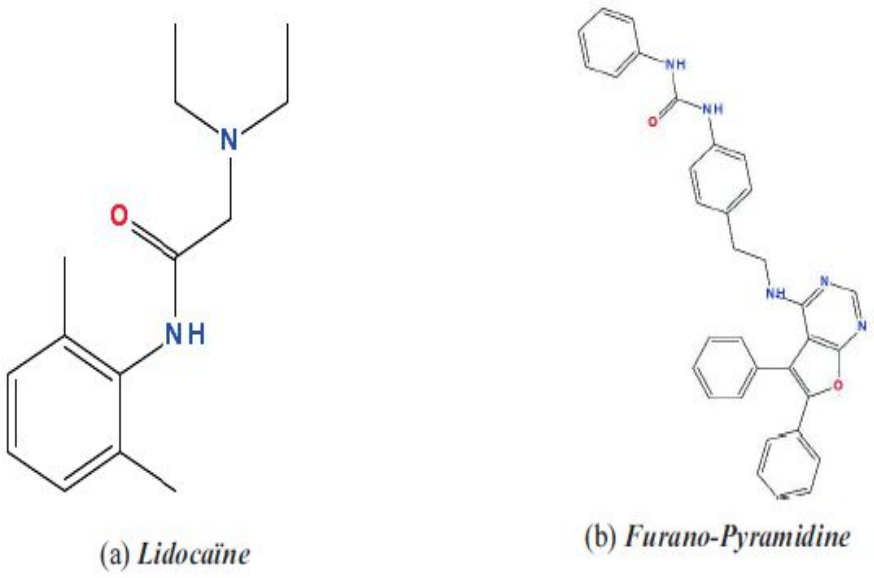
Figure 5.1 : visualisation 2D d'un fragment par guhautil

1.2. Les molécules de référence :

2 cas d'études sont effectués pour mesurer l'efficacité d'outil proposé et la qualité des résultats obtenues :

- **Cas d'étude 1 :** la conception des molécules a été basé sur la molécule de référence lidocaine, c'est l'anesthésique de type amide le plus utilisé par les dentistes.
- **Cas d'étude 2 :** la conception a été basé sur la molécule de référence furano-pyrimidine, c'est une molécule anti-cancer expérimental déclarée dans l'année 2010.

Les deux molécules sont téléchargé en format .mol a partir de la base de données de la chimoinfomatique Chemspider, et voila une figure qui représente la structure 2D de ces deux molécules.



(a) *Lidocaïne* (b) *Furano-Pyramidine*

Figure 5.2 : la structure 2D des 2 molécules de référence

2. La configuration des paramètres :

Comme tous les algorithmes évolutionnaire un algorithme génétique nécessite la configuration de certains paramètres comme la taille de la population et le nombre d'itération, nous proposons pour cela une interface (figure 3) qui permet de saisir tous les informations nécessaire pour l'exécution de l'algorithme proposé, durant les expérimentations fait, l'influence de ces paramètres sur le fonctionnement d'outil proposé a été étudiée.

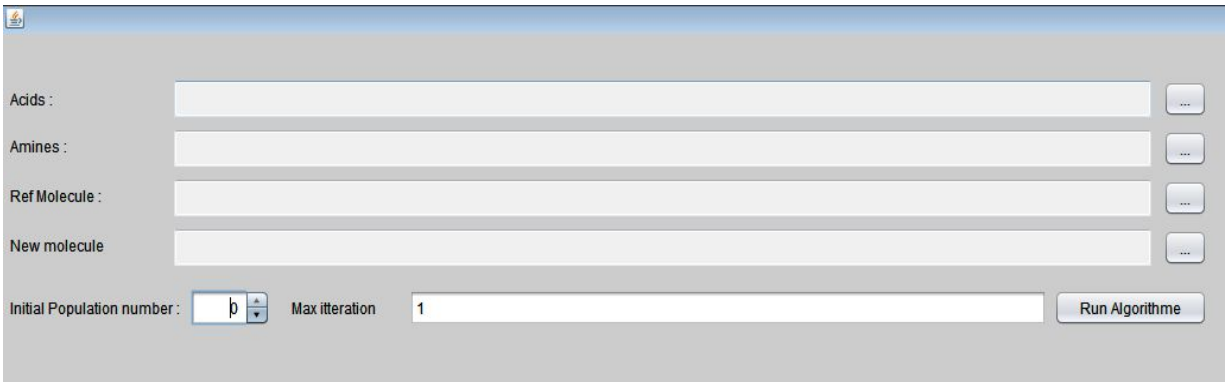


Figure 5.3 : l'interface pour la configuration des paramètres pour l'algorithme

Les champs acids, amines, RefMolecule, et new molecule représente le chemin d'emplacement des Library de fragments acide, amine et des molécules de référence, le champ new molecule représente l'emplacement pour sauvegarder la molécule de référence , le champ initialePopulation number représente la taille de la population et le champ Max itteration le nombre de génération :

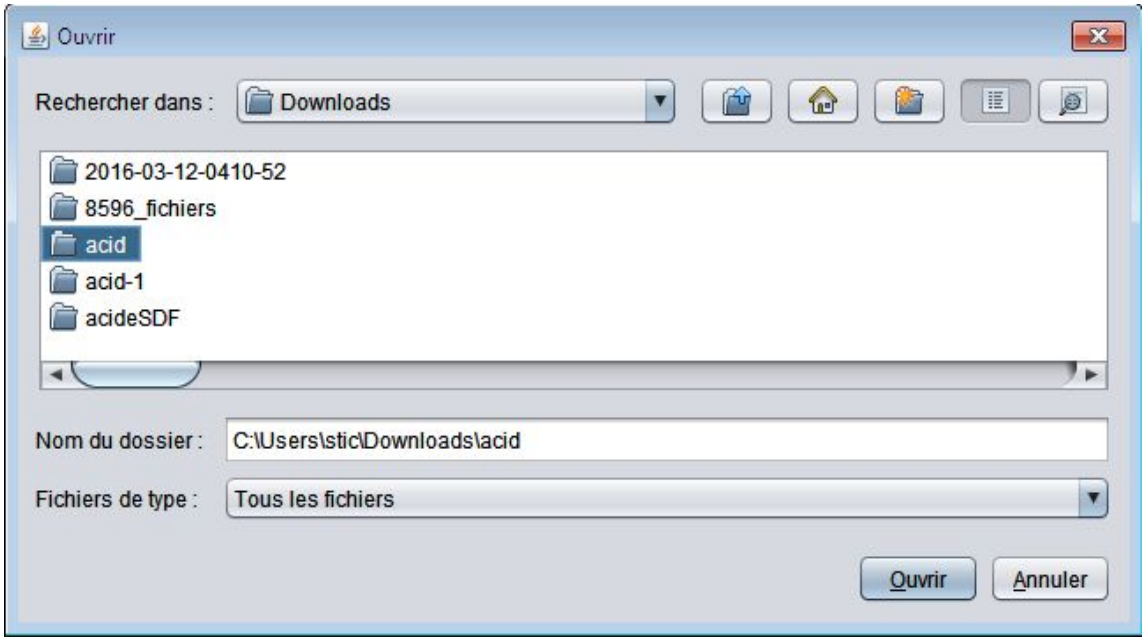


Figure 5.4 : interface pour ajouter l'emplacement de la Library acide.

2.1. L'effet de la taille de population P :

La taille de la population P est l'un des paramètres les plus influents sur le fonctionnement d'un algorithme évolutionnaire pour cela nous choisissons de l'étudier, le nombre d'itération est fixé a 50, le facteur de mutation a 0.5 et de

croisement a 0.5, le tableau ci-dessus représente les résultats obtenue avec une population de taille 10,15,25,50 chaque cas est exécuté 4 fois .

Nb exécution/taille population	10	15	25	50	100	150
Exécution 1	0.9252	0.9328	0.93442	0.7088	0.93156	0.93605
Exécution 2	0.9334	0.7029	0.93605	0.9335	0.9344	0.9327
Exécution 3	0.7047	0.9252	0.7050	0.9366	0.93661	0.9366
Exécution 4	0.9265	0.9257	0.9273	0.9312	0.93522	0.9360

Tableau 5.1 : les valeurs de la fonction de finesse obtenue on modifiant la taille de la population sur 4 exécutions

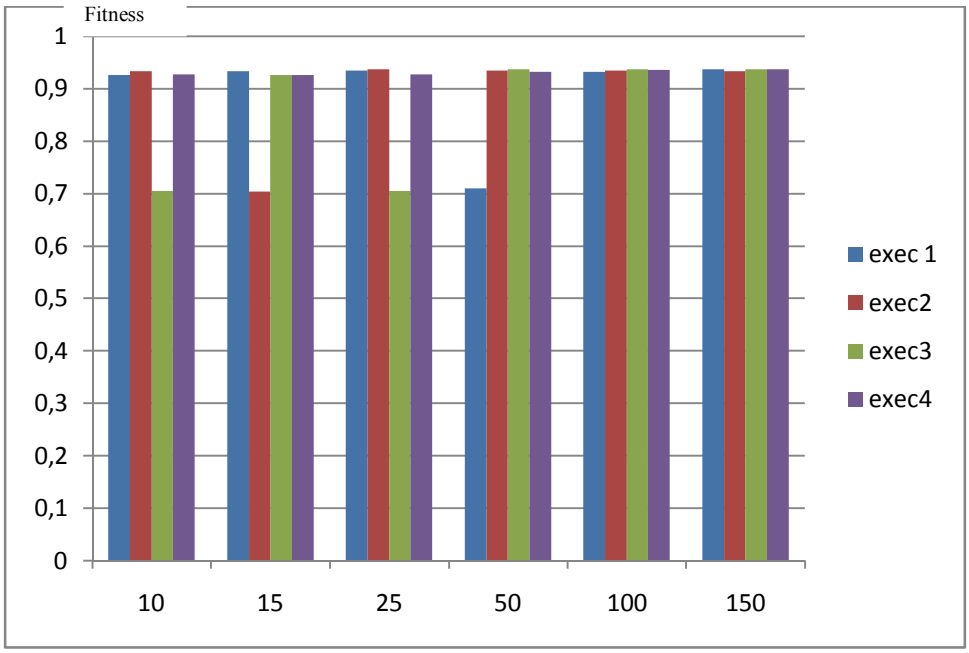


Figure 5.5 : évaluation de la fonction de fitness pour chaque taille de population

On peut voire que pour une taille de population =10, 15, et 25 la valeur de la fonction fitness n'attient pas la valeur maximale de la fonction de fitness, pour une taille de population=50 la fonction de fitness attient la valeur maximale mais pour certaine exécution la valeur de la fonction de fitness est faible par rapport a la valeur maximale, pour une taille de population= 100 et 150 en peut voir que la valeur de la fonction de fitness converge vers la valeur maximale pour tous les exécutions, on peut déduire que

le résultat obtenu dépend de la taille de population et l'augmentation de celui-ci donne des résultats plus performants.

2.2. L'effet de nombre d'itération :

Le nombre d'itération est aussi un paramètre important et influant sur le fonctionnement d'un algorithme évolutionnaire pour cela nous choisissons d'étudier cet influence pour un nombre d'itération= 50,100 et 150 pour chaque cas l'étude se fait sur 4 exécution, et pour le faire les valeurs de la taille de population, de facteur de mutation et de facteur de croisement sont fixé a 100,0.5, 0.5, les résultats obtenus sont représentés dans la figure 5.6 et le tableau 5.2.

	Nb génération=50	Nb génération=100	Nb génération=150
Exécution 1	0.9366	0.9344	0.9344
Exécution 2	0.7088	0.9339	0.9348
Exécution 3	0.9344	0.9344	0.9339
Exécution 4	0.7063	0.7084	0.9366

Tableau 5.2 : les valeurs de la fonction d'évaluation pour chaque nombre d'itération

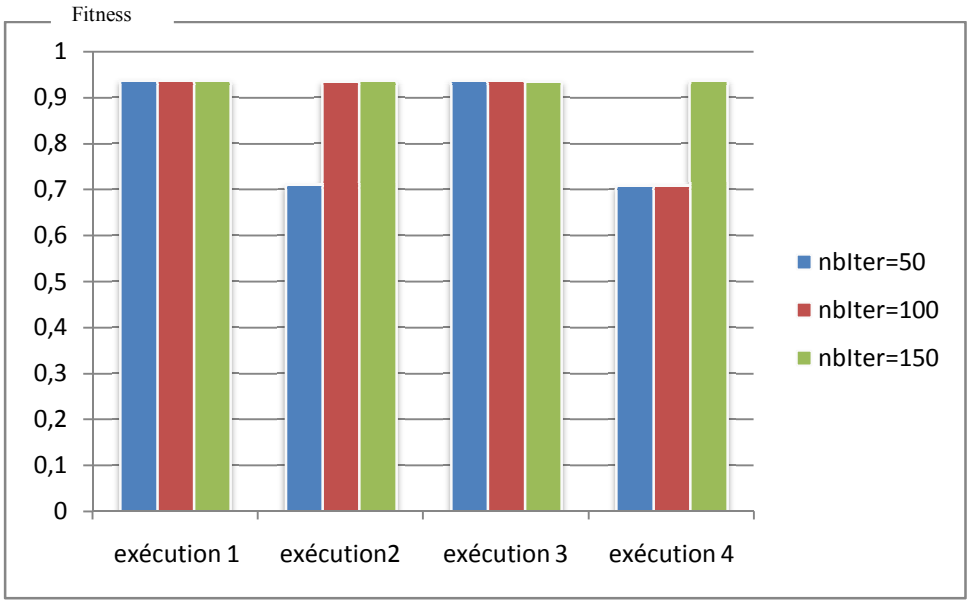


Figure 5.6 : valeur de la fonction de fitness pour chaque modification dans le nombre d'itération

On peut voir que pour un nombre d'itération =50 la valeur de la fonction de fitness est faible dans 2 exécution par rapport à la valeur maximale obtenu, la valeur augmente et converge vers la valeur maximale de la fonction de fitness pour un nombre d'itération =100 et 150, en peut déduire que l'augmentation de nombre d'itération améliore la performance des résultats obtenue.

Les études expérimentaux et les résultats obtenue montre bien la relation entre les paramètres de l'algorithme et la qualité des résultats obtenu, en peut conclure que l'algorithme donne des bonne résultats avec :

- Une taille de population >=50
- Un nombre d'itération>= 100

Ces paramètres seront utilisé dans la section qui suite.

3. Les testes et les résultats obtenus :

Dans cette section nous commençons d'appliquer l'algorithme proposé et de présenter les résultats obtenus.

L'objectif est d'étudier le comportement de l'algorithme proposé dans le cas mono et multi objectif par rapport aux 2 molécules de référence lidocaine et furano_pyrimidine:

3.1. Résultat obtenus dans les 2 contextes mono et multi objectif pour les deux cas de teste :

Notre algorithme est implémenté en java et testé sous Windows 7, sur un processeur core i4, cadencé a 2.20 GHz, avec 4 Go de RAM, pour évaluer la performance de l algorithme proposé nous effectuons les deux cas de teste suivant :

3.1.1. Cas de teste 1(Lidocaine) :

La première cas de teste se base sur la molécule de référence lidocaine , le comportement de l'algorithme proposé a été suivie en enregistrant les meilleures valeur de la fonction objectif, 2 cas on été considéré dans cette étude, le premier cas c'est l'étude mono objectif qui se base sur la valeur de similarité(Tcoef), la valeur de ce coefficient comprise entre 0 et 1, une valeur de 0 signifie l'absence de similarité entre la molécule conçue et la molécule de référence, plus la valeur converge vers 1 plus la molécule est similaire au molécule de référence, le deuxième cas c'est le cas multi objectif qui se focalise sur la valeur globale de la fonction d'évaluation la régle de 5 de lipinski et la valeur de coefficient de Tanimoto, la valeur globale est compris entre 0 et 1 plus la valeur converge vers 1

plus la molécule conçue est efficace, la capture écran ci-dessus donne un exemple sur l'interface de notre outil qui visualise la structure 2D et les valeurs de la fonction de fitness obtenu :

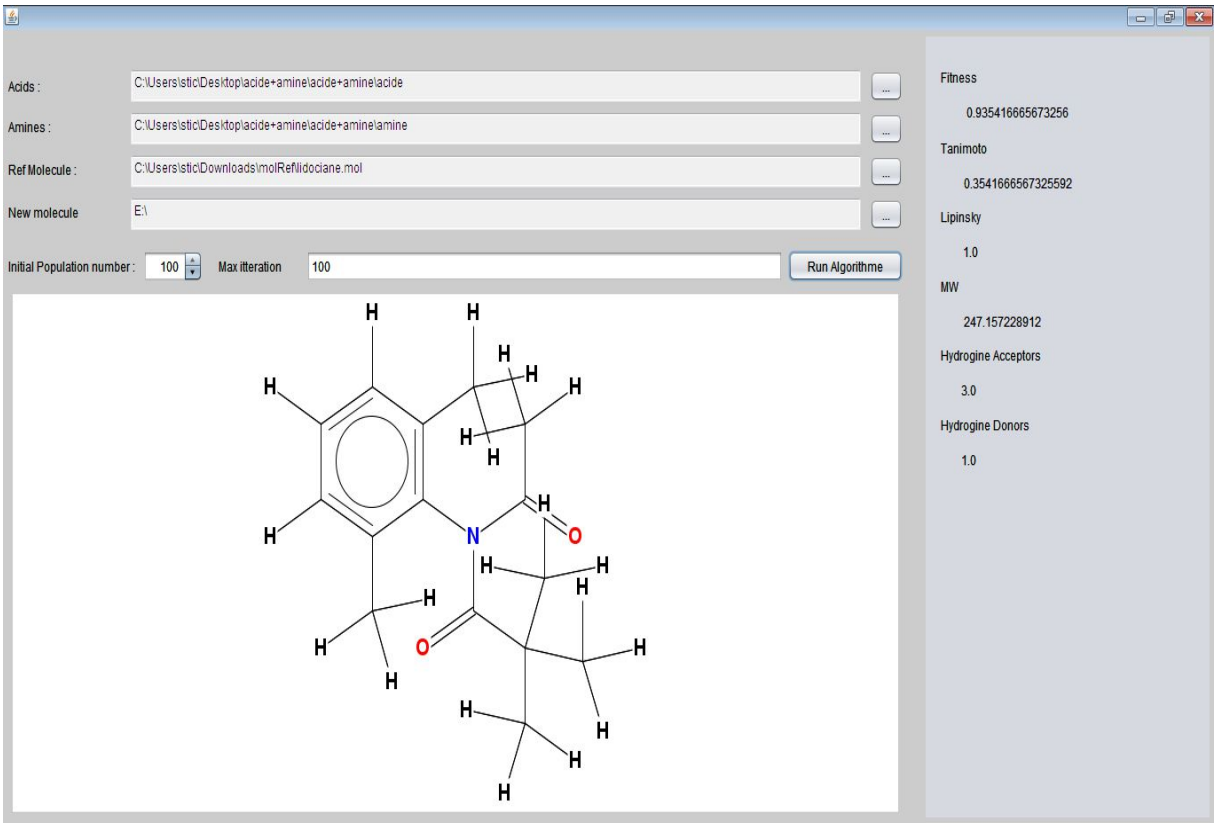


Figure 5.7 : interface qui affiche la meilleure solution

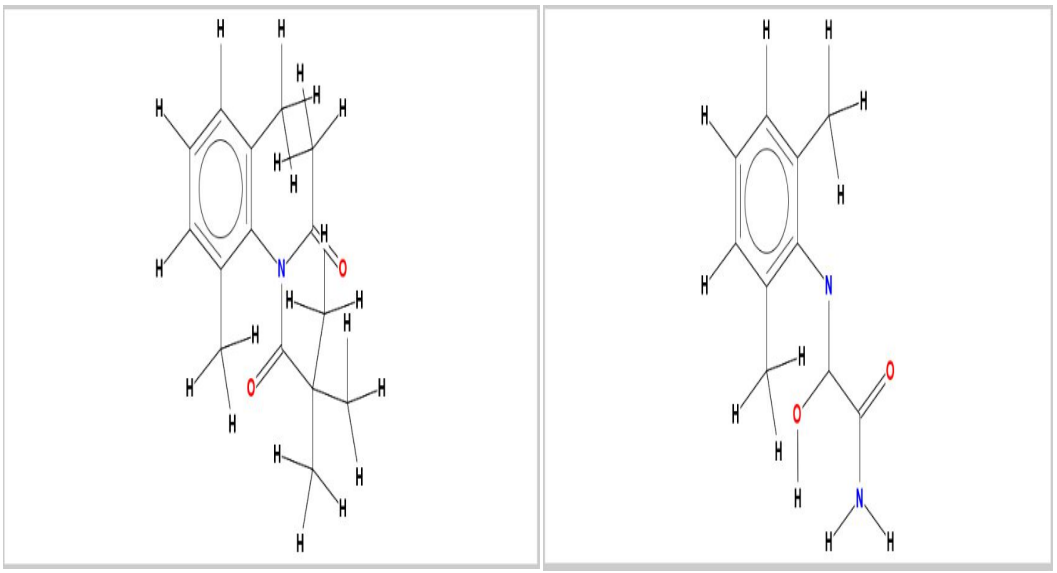
Et voila un tableau qui présente les résultats apres plusieur exécution de l'algorithme :

molécules	Cas d'etude	Tcoeff	OBA	A1*OBA+a2*Tcoeff
A	Multi objectif	0.3541	1.0	0.93541
B	Multi objectif	0.3378	1.0	0.93378
C	Multi objectif	0.3246	1.0	0.93246
D	Multi objectif	0.3203	1.0	0.93203
E	Multi objectif	0.31603	1.0	0.931603
F	Mono objectif	0.25	0.75	0.6945

Tableau 5.3 : les résultats obtenus et les valeurs de la fonction objectif pour le cas d'etude 1(lidociane)

Figure 5.8 : un graphe qui représente les résultats obtenue

A partir de ce graphe et de tableau on peut voir que la meilleure solution est obtenue par l'application multi objectif (OBA=1, Tcoeff= 0.3541 et fitness= 0.93541), cette solution est obtenu dans la 38 eme génération, par contre une application mono objectif donne des résultats qui reste loin de la valeur maximale (Tcoeff= 0.25, OBA=0.75) ce résultat est obtenu dans presque tous les exécutions, la majorité des résultats obtenu a une bon valeur de lipinski et une valeur de Tanimoto un peut diminué(noter que la valeur de lipinski est plus important que la similarité car il représente la vérification de l'absorption l'une des propriété favorable d'un médicament), et voila une figure qui représente les 6 meilleurs résultats obtenu :



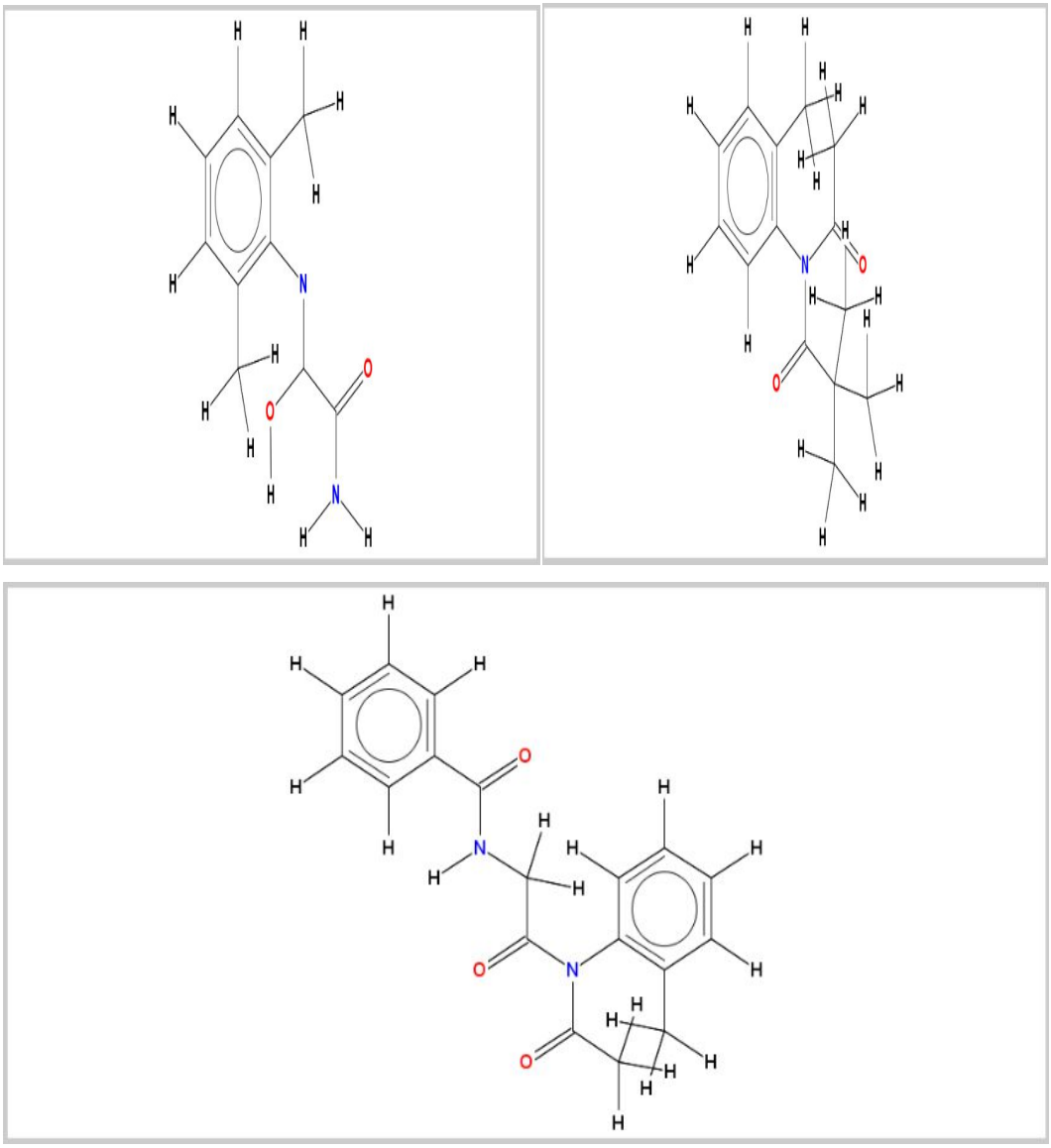


Figure 5.9 : les structures 2D des meilleures molécules obtenues

3.1.2. Etude de cas 2(furano_pyrimidine) :

Ce deuxième cas d'étude se focalise sur la molécule de référence furano_pyrimidine, deux scénario sont considéré dans cette étude, le premier scénario c'est le cas mono objectif qui se focalise seulement sur la valeur de coefficient de Tanimoto comprise entre 0 et 1, plus la valeur converge vers 1 plus la molécule conçue est similaire a la molécule de référence, le deuxième scénario se focalise sur la valeur globale de la fonction objectif(OBA qui se base sur les règles de 5 de Lipinki et le coefficient de Tanimoto) la valeur est comprise entre le 0 et 1 plus la valeur converge vers 1 plus la molécule conçue est efficace , les résultats obtenus sont présenté dans le tableau et le graphe ci-dessus :

molécule	Cas d'étude	Tanimoto	OBA	Fitness
A	Multi objectif	0.4031	1.0	0.94031
B	Multi objectif	0.3949	1.0	0.93949
C	Multi objectif	0.3658	1.0	0.93658
D	Mono objectif	0.4033	0.75	0.7153
E	Mono objectif	0.3989	0.75	0.7148

Tableau 5.4 : tableau des résultats obtenu pour le cas d'étude 2 (furano_pyrimidine)

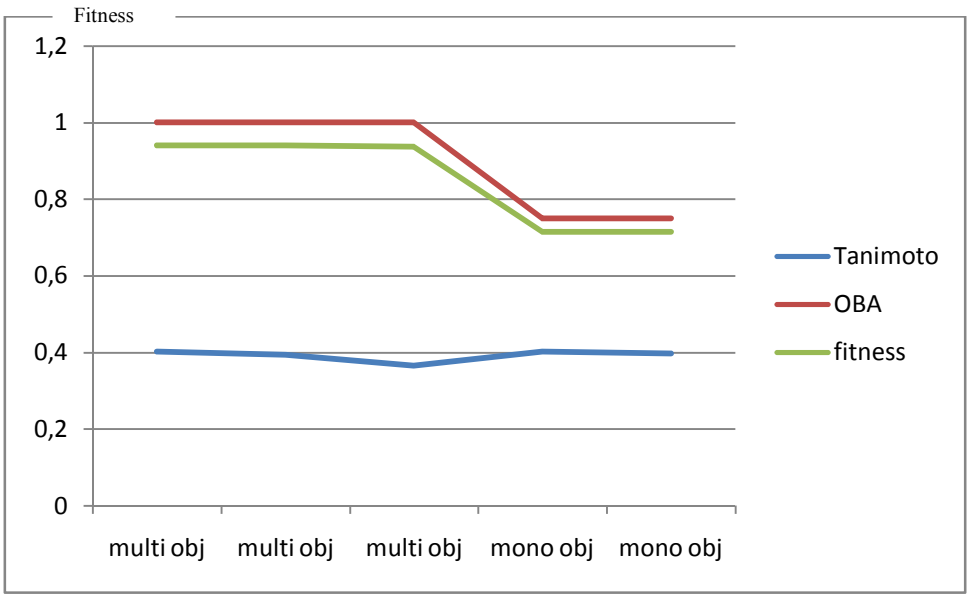


Figure 5.10 : graphe qui représente les résultats obtenu

À partir de graphe et de tableau en peut voir que le meilleur résultat obtenu qui attient la valeur maximale de fitness (fitness= 0.94031, Tanimoto= 0.4031, OBA=1.0) est obtenu dans le cas multi objectif par contre une étude mono objectif donne des résultats qui reste loin de la valeur maximal (Tanimoto=0.4033, OBA=0.75),ces résultats voile une des règles de lipinski pour cela il prend une valeur = 0.75. On peut marquer que le multi objectif donne des bons résultats par rapport au mono objectif dans presque tous les exécutions. L'interface qui visualise le meilleur résultat obtenu se présente dans la (figure 5.11) et les structures 2D des bons résultats (figure 5.12) :

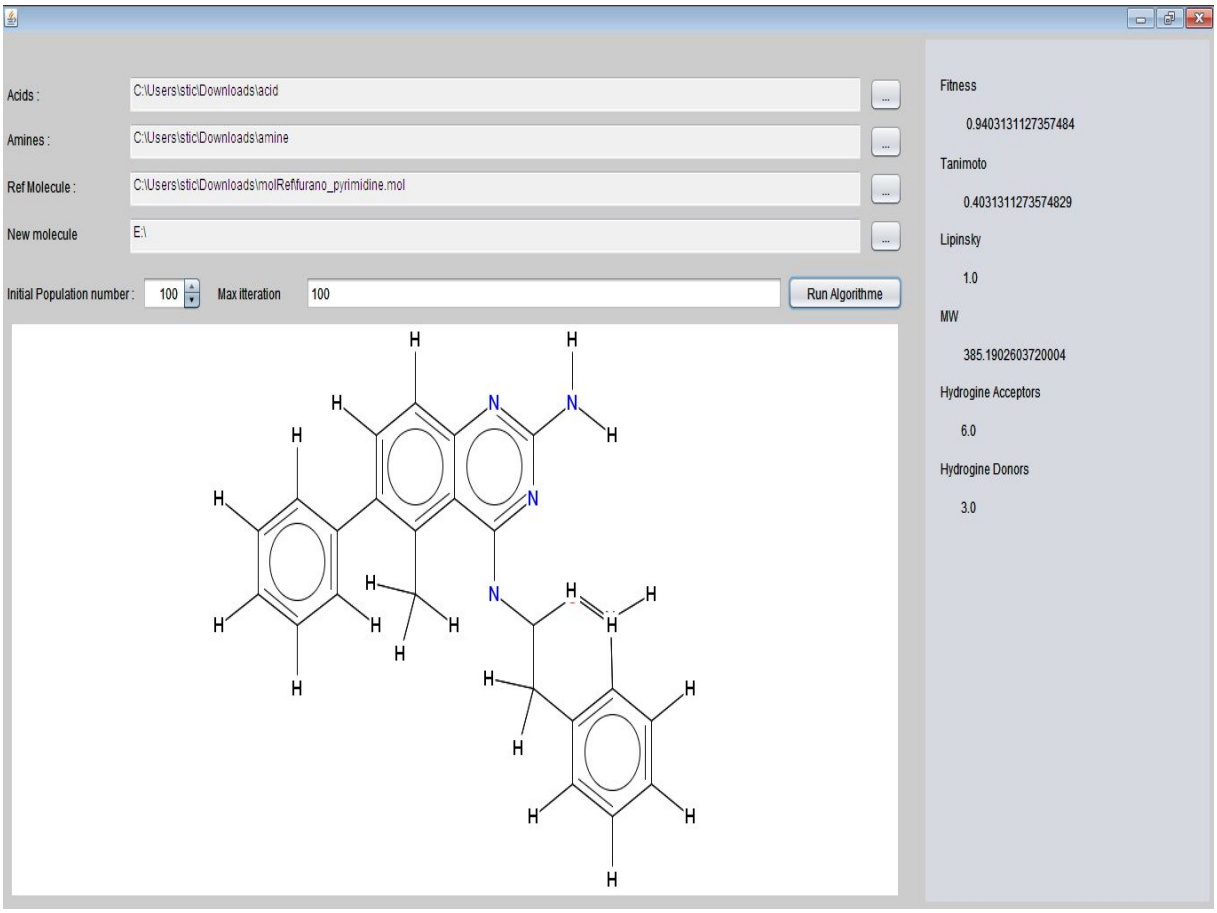
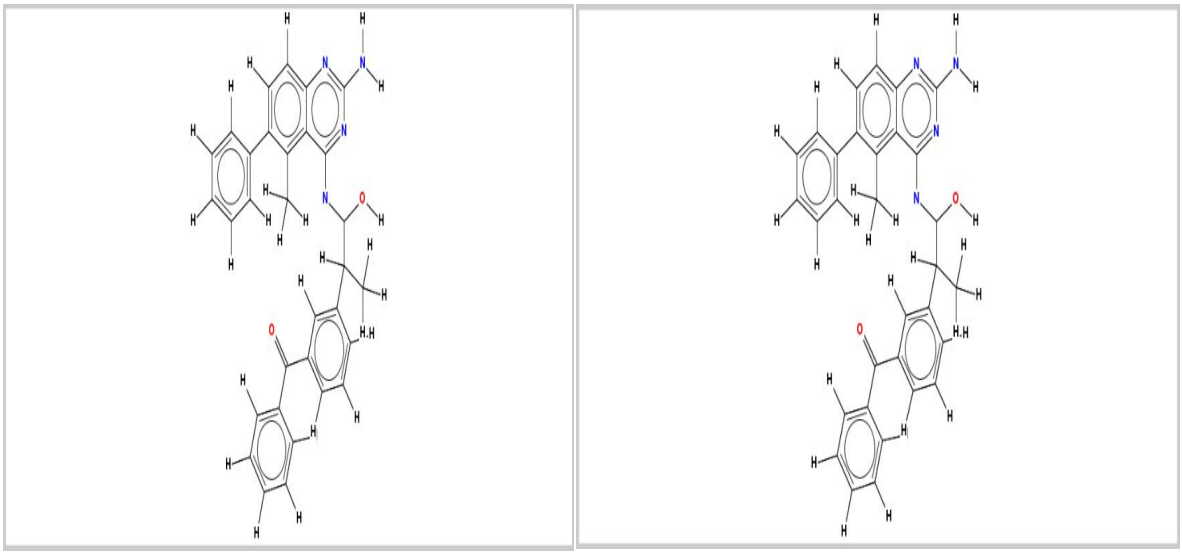


Figure 5.11 : l'interface qui visualise le meilleur résultat et les valeurs de la fonction de fitness



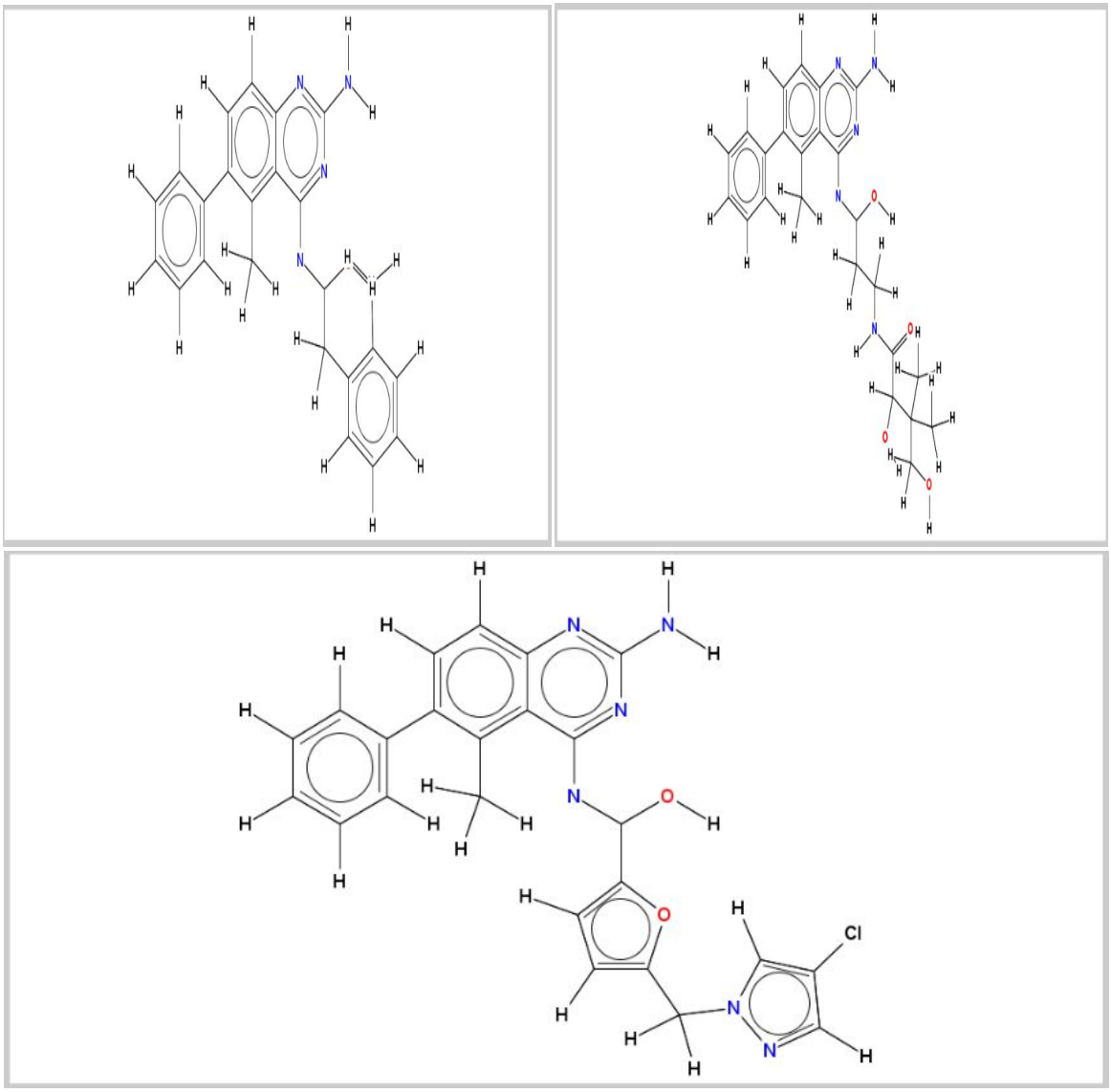


Figure 5.12 : les structures 2D des meilleurs résultats

Conclusion :

A partir des résultats obtenus après les étude expérimentales dans les deux cas d'étude (avec lidocaine comme molécule de référence et avec furano_pyrimidine comme molécule de référence) ,dans les deux scenario mono et multi objectif .On peut conclure que : notre outil fonction de mieux dans le cas multi objectif par rapport au cas mono objectif, la valeur de la fonction de fitness avec furano_pyrimidine est meilleur que dans le cas de lidocaine, le temps d'exécution est raisonnable par rapport au autre algorithme. Les résultats obtenus dépend des molécule acide et amine téléchargé sous le format (.mol). Notre but principale est d'offrir un outil qui permet de générer des molécules candidat d'être médicament qui satisfait plusieurs propriété dans un temps raisonnable. Les travaux future peut baser sur l'amélioration de la qualité des molécule de référence sur le coté de similarité surtout, et d'appliquer d'autre méthode pour optimiser de plus le temps d'exécution.

Conclusion générale

Les investissements dans le domaine de conception de médicament sont en augmentation continue, mais ce processus est un processus très long et couteux, pour cela l'utilisation d'un outil informatique pour la génération plus rapide des médicaments prend une place très importante afin de résoudre le problème de durcissement et de diminuer le temps de découvert, C'est dans ce cadre que nous avons proposé ce travail. A l'issue de ce projet une exploration du domaine de la chemoinformatique a été effectuée et a été concrétisée par une revue du matériel de base et des méthodes inhérentes. A côté de cela, nous proposons un outil informatique pour la conception de novo des nouveaux médicaments, qui permet de générer des médicaments et de calculer certaines propriétés, le noyaux de cette outil consiste a un optimiseur basé sur l'algorithme génétique multi objectif , l'outil proposé comporte également un ensemble de Library de fragments acide, une library d' amine, une Library des molécule de référence, et un outil open source java le CDK qui permet de faire des réaction chimique entre les fragments acide et amine et de produire un évaluateur de similarité et un évaluateur de la bio disponibilité orale pour évaluer les molécules conçu a base d'une fonction objectif choisi par agrégation des deux propriété OBA et le Tcoef , la bio disponibilité orale est une propriété très important pour la sélection des médicaments, cette dernière est évalué par les règles de 5 de lipinski , le deuxième évaluateur c'est la similarité qui s'évalue par le coefficient de Tanimoto ce dernier permet d'évaluer la similarité entre la nouvelle molécule et une molécule de référence connu, l'évaluation de ces deux propriétés permet a l'algorithme de générer des molécules synthétisables susceptible d'être médicament efficace avec une valeur élevé de bio disponibilité .

Afin de mesurer la performance de l'outil proposé une étude expérimentale est effectuée sur deux cas, le premier cas se focalise sur la molécule de référence lidocaine, et le deuxième cas se focalise sur la molécule de référence furano-pyrimidine, les deux cas sont étudié dans le scénario mono et multi objectif.

Les résultats obtenus montrent bien que l'algorithme proposé travailler de mieux dans le scénario multi objectif , l'application de l'algorithme proposé dans le scénario multi objectif permet de générer des molécules avec des bons valeur de biodisponibilité orale a l'aide de l'outil CDK qui permet de relier n'importe quelle acide de Library acide avec n'importe quelle amine dans la Library amine.

Une amélioration de temps d'exécution a été marquée aussi lors d'exécution de l'outil proposé, les travaux future oriente vers l'utilisation de parallélisassions afin d'améliorer le temps d'exécution, d'améliorer la valeur de similarité soit par la recherche des bons fragments afin d'améliorer la diversification ou de les modifier, d'appliquer des algorithmes plus efficace et innovant.

Les références bibliographiques

Les références de chapitre 1

- *[1]http://www.ecole.ensicaen.fr/~mbrun/1A_MCF_PROJETS/HEMELAERE_CASTRO/conteneur.php?page=pageb&ban=1*
- *[2] H.J. Huang, H.W. Yu, C.Y. Chen et al., "Current developments of computer-aided drug design," J. TAIWAN. INST. CHEM. E., vol. 41, pp. 623-635, 2010.*
- *[3] F. K. Brown, "Annual. Reports in Medicinal Chemistry," Chem., 1998.*
- *[4] T. I. Oprea ."Chemoinformatics in Drug Discovery," John Wiley & Sons, 6 mars 2006, 515 pages.*
- *[5] W.A. Warr, "Balancing the needs of the recruiters and the aims of the educators,» Presented at 218 th ACS National Meeting. New Orleans, Louisiana, August 22-26, 1999.*
- *[6] J. Gasteiger, "The central role of chemoinformatics," Chemom. Intell. Lab. Syst., vol. 82, pp.200-209, 2006.*
- *[7] A. Varnek et I. I. Baskin, "Chemoinformatics as a Theoretical Chemistry Discipline," Mol. Inf., vol. 30, pp. 20-32, 2011*
- *[8] mémoire « IDE-DD: Un outil d'aide à la conception De novo de médicaments en chimioinformatique à base d'évolution différentielle multiobjectif en nombres entiers » , Mahdadi Abla, Université Constantine 2 - Abdelhamid Mehri.*
- *[9] A. Bhalerao et al, "Chemoinformatics: The Application of Informatics Methods to Solve Chemical Problems,"*
- *[10] J. B. O. Mitchell, "Machine learning methods in chemoinformatics," WIREs Comput.Mol.Sci., vol. 4, pp. 468-481,2014.*
- *[11] http://www.iaf.inrs.ca/evenements/chimie-computationnelle-developpement-molecules-bioactives*
- *[12] fr.wikipedia.org/wiki/Chemoinformatique.*
- *[13] Stuart K. Card, Jock D.Mackinlay and Ben Shneiderman (1999). Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers.*
- *[14] Hann, M. M.; Oprea, T. I. Pursuing the leadlikeness concept in pharmaceutical research. Curr Opin Chem Biol 2004, 8, 255-263.*
- *[15] Bohacek, R.S.; McMartin, C.; Guida, W.C. The Art and Practiceof Structure-based Drug Design: a Molecular Modelling Perspective. Med. Res. Rev. 1996, 16, 3-50;*

- *[16] Ertl, P. Cheminformatics Analysis of Organic Substituents: Identification of the Most Common. J. Chem. Inf. Comput. Sci. 2003, 43, 374-380*
- *[17] Schuffenhauer, A.; Popov, M.; Schopfer, U.; Acklin, P.; Stanek, J. Jacoby, E. Molecular Diversity Management Strategies for Building and Enhancement of Diverse and Focused Lead Discovery Compound*
- [18] http://www.didiersvt.com/cd_1s/html/c7/c7a4p1a2.htm
- *[19] IUPAC Combinatorial Chemistry*
- *[20] Kutchukian, Peter; Lou, David; Shakhnovich, Eugene (2009). "FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules occupying Druglike Chemical".*
- *[21] Rollinger JM, Stuppner H, Langer T (2008). "Virtual screening for the discovery of bioactive natural products". Prog Drug Res. Progress in Drug Research 65 (211): 213–49.*
- *[22]http://www.ecole.ensicaen.fr/~mbrun/1A_MCF_PROJETS/HEMELAERE_CASTRO/conteneur.php?page=1&id=1&ban=1*
- *[23]Madsen U, Krogsgaard-Larsen P, Liljefors T (2002). Textbook of Drug Design and Discovery. Washington, DC: Taylor & Francis.*
- *[24] Reynolds CH, Merz KM, Ringe D, eds. (2010). Drug Design: Structure- and Ligand-Based Approaches (1 ed.). Cambridge, UK: Cambridge University Press. .*
- *[25]Shirai H, Prades C, Vita R, Marcatili P, Popovic B, Xu J, Overington JP, Hirayama K, Soga S, Tsunoyama K, Clark D, Lefranc MP, Ikeda K (Nov 2014). "Antibody informatics for drug discovery". Biochimica et Biophysica Acta 1844 (11): 2002–2015.*
- *[26]A.T. Balaban, "Topological indices based on topological distances in molecular graphs," Pure and Applied Chemistry, vol. 55, pp. 199-206, 1983.*
- *[27] Steinbeck, Christoph et al, "Recent Developments of the Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics," Current Pharmaceutical Design, vol. 12, pp. 2111-2120, June 2006.*
- *[28] H. Hong et al., "Molecular Descriptors from 2D Structures for Chemoinformatics and Toxicoinformatics," Journal of Chemical Information and Modeling, vol. 48, pp. 1337-1344, 2008.*
- *[29] A. Dalby et al., "Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited (MDL)," J. Chem. Inf. Comput. Sci., vol. 32, pp.244-255, 1992.*

- *[30] D. Weininger, "SMILES, a chemical language and information system: Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci, vol. 28, pp. 31-36, 1988.*
- *[31] "CTFile Formats". Elsevier MDL, 14600 Catalina St., San Leandro, CA 94577, 2005.*
- *[32] D. Weininger, "SMILES, a chemical language and information system: Introduction to methodology and encodingrules," J. Chem. Inf. Comput. Sci, vol. 28, pp. 31-36, 1988.*
- *[33] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug. Deliv. Rev. 2001, 46, 3-26.*
- *[34] Lipinski, C.A. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov. Today 2004, 1, 337-341.*
- *[35]Wild DJ (2009) Grand Challenges for Cheminformatics. J Cheminform 1: 1.*
- *[36]LI et AP, "Screening for human ADME/Tox drug properties in drug discovery," Drug Discovery Totay, vol. 6, pp. 357-366, 2001.*
- *[37] LI et AP, "Screening for human ADME/Tox drug properties in drug discovery," Drug Discovery Totay, vol. 6, pp.357-366, 2001.*
- *[38]* https://jfaulon.wikispaces.com/
- *[39]M.M. Hann and T.I. Oprea, "Pursuing the leadlikeness concept in pharmaceutical research," Curr. Opin. Chem.Biol.*
- *[40]http://www.ecole.ensicaen.fr/~mbrun/1A_MCF_PROJETS/HEMELAERE_CASTRO/conteneur.php?page=paged2&ban=1*
- *[41] W.T. Tutte, "Graph theory," Cambridge University Press, Cambridge [Cambridgeshire]; New York, NY, USA,1984.*
- *[42] Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. J. Med. Chem. 2000, 43, 3714-3717.*
- [43] https://www.info2.uqam.ca
- [44]Glen, R. C.; Payne, A. W. A genetic algorithm for the automated generation

Les références de chapitre 2

- *[1] Madsen U, Krogsgaard-Larsen P, Liljefors T (2002). Textbook of Drug Design and Discovery. Washington, DC: Taylor & Francis*
- *[2] Reynolds CH, Merz KM, Ringe D, eds. (2010). Drug Design: Structure- and Ligand-Based Approaches (1 ed.). Cambridge, UK: Cambridge University Press*
- *[3] https://www.info2.uqam.ca*
- *[4] www.intechopen.com*
- *[5] Dixon SJ, Stockwell BR (Dec 2009). "Identifying druggable disease-modifying gene products". Current Opinion in Chemical Biology 13 (5-6): 549–55.*
- *[6] Imming P, Sinning C, Meyer A (Oct 2006). "Drugs, their targets and the nature and number of drug targets". Nature Reviews. Drug Discovery 5 (10): 821–34.*
- *[7] Anderson AC (Sep 2003). "The process of structure-based drug design". Chemistry & Biology 10 (9): 787–97.*
- *[8]Ganellin CR, Jefferis R, Roberts SM (2013). "The small molecule drug discovery process — from target selection to candidate selection". Introduction to Biological and Small Molecule Drug Research and Development: theory and case studies. Elsevier.*
- *[9]Yuan Y, Pei J, Lai L (Dec 2013). "Binding site detection and druggability prediction of protein targets for structure-based drug design". Current Pharmaceutical Design 19 (12): 2326–33.*
- *[10] Nicolaou CA, Brown N (Sep 2013). "Multi-objective optimization methods in drug design". Drug Discovery Today. Technologies 10 (3): 427–35.*
- *[11] Ban TA (2006). "The role of serendipity in drug discovery". Dialogues in Clinical Neuroscience 8 (3): 335–44.*
- *[12]Leach, Andrew R.; Harren, Jhoti (2007). Structure-based Drug Discovery. Berlin: Springer.*
- *[13] Mauser H, Guba W (May 2008). "Recent developments in de novo design and scaffold hopping". Current Opinion in Drug Discovery & Development 11 (3): 365–74.*
- *[14]Klebe G (2000). "Recent developments in structure-based drug design". Journal of Molecular Medicine 78 (5): 269–81.*

- *[12]http://esilrch1.esi.umontreal.ca/~syguschj/cours/BCM6200/bcm6200_Structure-based%20Drug%20design.pdf*
- *[13] J. John, M. Frech and A. Wittinghofer, "Biochemical Pro- perties of Haras Encoded p21 Mutants and Mechanism of the Autophosphorylation Reaction," The Journal of Bio-logical Chemistry, Vol. 263, 1988, pp. 11792-11799.*
- *[14]http://biochem.wustl.edu/wildmans/Methods/de_novo_Design.html*
- *[6] van de Waterbeemd, H. et coll., Glossary of Terms Used in Computational Drug Design, http://www.iupac.org/reports/1997/6905vandewaterbeemd/glossary.html.*
- *[8] http://fr.slideshare.net/prasanthperceptron/de-novo-drug-design*
- *[15] B. Hammer, "Recurrent Networks for Structured Data— A Unifying Approach and its Properties," Cognitive Sys-tems Research, Vol. 3, No. 2, 2002, pp. 145-165. doi:10.1016/S1389-0417(01)00056-0*
- *[42]K. R. Oldenburg, "Annual Report in Medicinal Chemis-try," J. A. Bristol, Ed., Academic Press, London, Vol. 33, 1998, pp. 301-307.*
- *[16]W. H. Moos, G. D. Green and M. R. Pavia, "Chapter 33. Recent Advances in the Generation of Molecular Diver-sity," Annual Reports in Medicinal Chemistry, Vol. 28, 1993, pp. 315-324.*
- *[17] F. Ooms, "Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medici-nal Chemistry," Current Medicinal Chemistry, Vol. 7, No. 2, 2000, pp. 141-158.*
- *[18] J. Kuhlman, International Journal of Clinical Pharma-cology and Therapeutics, Vol. 35, 1997, pp. 541-552.*
- *[19] R. G. Halliday, S. R. Walker and C. E. Lumley, Journal of Pharmaceutical Medicine, Vol. 2, 1992, pp. 139-154.*
- *[20] A. K. Ghose and J. J. Wendoloski, "Perspective in Drug Discovery and Design," Kluwer/Escom, Vol. 9-11, 1998, pp. 253-271.*
- *[21] D. J. Abraham and G. E. Kellogg, "3D-QSAR in Drug Design," H. Kubinyi, Ed., Escom, Leiden, Vol. 1, 1993, pp. 506-522.*
- *[22] http://www.molfunction.com*
- *[23] MA.Murcko "reviews in the computational chimestry", Lipkowitz KB,Boyd DB editors.Hoboken,NJ,USA:JhonWilly &Sons,Inc,vol.11,pp.1-67,2007*
- *[24] Kitchen DB, Decornez H, Furr JR, Bajorath J (Nov 2004). "Docking and scoring in virtual screening for drug discovery: methods and applications". Nature Reviews. Drug Discovery 3 (11): 935–49.*

- *[25] Cerqueira NM, Bras NF, Fernandes PA, Ramos MJ (Jan 2009). "MADAMM: a multistaged docking with an automated molecular modeling protocol". Proteins 74 (1): 192–206*
- *[32] Acharya C, Coop A, Polli JE, Mackerell AD., Jr Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr. Comput. Aided Drug Des. 2011;7(1):10–22. [33]http://strbio.biochem.nchu.edu.tw/classes/special%20topics%20biochem/course%20ppts/rational%20drug%20design-2014.pdf*
- *[34] http://fr.slideshare.net/prasanthperceptron/de-novo-drug-design*
- *[35] Nicolaou, C. A.; Apostolakis, J.; Pattichis, C. S. De Novo Drug Design Using Multiobjective Evolutionary Graphs. J. Chem. Inf. Model., 2009, 49, 295-307.*
- *[37] http://www.pharmainventor.com/structligand.html*
- *[41] https://www.perso.ibcp.fr*
- *[42] http://fr.slideshare.net/rahulbs89/molecular-docking-28000661*
- *[43] Rational drug design 2014.pdf*

Les références de chapitre 3

- *[1] M.Samir,Optimisation Multiobjectif Par Un nouveau Schéma De Coopération Méta/Exacte," Mémoire de Magister, Université Mentouri de Constantine, 2005.*
- *[2]http://www.medinfo.cs.ucy.ac.cy/doc/Publications/Masters/C.Kannas/Master_Christos_Kannas.pdf*
- *[3] E.G.Talbi, Métaheuristiques pour l'optimisation combinatoire multi objectif: Etat de l'art,". CENT, vol. 33, pp. 98-757, 1999*
- *[4] S.Meshoul, H. AlBaity and A. Kaban, "On Extending Quantum Behaved Particle Swarm Optimization to MultiObjective Context," In IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane,Australia.*
- *[5] A. Osyczka, "Multicriteria optimization for engineering design," In J. S. Gero, editor, Design Optimization, pp.*
- *193-227, 1985.*
- *[6]http://www.gauvainmarquet.fr/wpcontent/uploads/2014/04/rapport_projet_dolphin.pdf*
- *[7] C.A. Nicolaou, C. Kannas and E. Loizidou, "Multi-Objective Optimization*

Methods in De Novo Drug Design",Mini-Reviews in Medicinal Chemistry, 2012, Vol. 12, No. 5

- *[8] Nicolaou, C. A.; Apostolakis, J.; Pattichis, C. S. De Novo Drug Design Using Multiobjective Evolutionary Graphs. J. Chem.Model., 2009, 49, 295-307.*
- *[9] https://www.info2.uqam.ca*
- *[10] Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drugderived building blocks. J. Comput. Aided Mol. Des, 2000, 14, 487-494.*
- *[11] http://link.springer.com/article/10.1023/A%3A1008184403558#page-1*
- *[12] Krovat, E. M.; Steindl, T.; Langer, T. Recent Advances in Docking*
- *and Scoring. Current Computer - Aided Drug Design, 2005, 1, 93-102.*
- *[14] Pegg, S. C.-H.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based de novo design. Journal of Computer-Aided Molecular Design, 2001, 15, 911-933.*
- *[15] Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. J Chem Inf Model, 2008, 48, 679-690.*
- *[16]http://www.optibrium.com/downloads/Advances%20in%20MPO%20for%20de%20Novo%20Drug%20Design%20preprint.pdf*
- *[17] Ching-Lai Hwang; Abu Syed Md Masud (1979). Multiple objective decision making, methods and applications: a state-of-the-art survey. Springer-Verlag. ISBN 978-0-387-09111-2. Retrieved 29 May 2012.*
- *[18] Kaisa Miettinen (1999). Nonlinear Multiobjective Optimization. Springer. ISBN 978-0-7923-8278-2. Retrieved 29 May 2012.*
- *[23]http://www.optibrium.com/downloads/Advances%20in%20MPO%20for%20de%20Novo%20Drug%20Design%20preprint.pdf*
- *[24] Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. A Graph-Based Genetic Algorithm and Its Application to the Multiobjective Evolution of Median Molecules. J. Chem. Inf. Model., 2004, 44, 1079-1087.*
- *[25] Nicolaou, C. A.; Apostolakis, J.; Pattichis, C. S. De Novo Drug Design Using Multiobjective Evolutionary Graphs. J. Chem. Inf. Model., 2009, 49, 295-307.*
- *[26] Ekins, S.; Honeycutt, J. D.; Metz, J. T. Evolving molecules using multi-objective optimization: applying to ADME/Tox. Drug Discov. Today, 2010, 15, 451-460.*

- *[27] Lameijer, E.-W.; Kok, J. N.; Bäck, T.; Ijzerman, A. P. The molecule evoluator. An interactive evolutionary algorithm for the design of drug-like molecules. J Chem Inf Model, 2006, 46, 545-552.*
- *[28] Ecemis, M. I.; Wikel, J.; Bingham, C.; Bonabeau, E. A Drug Candidate Design Environment Using Evolutionary Computation. IEEE Trans. Evol. Computat., 2008, 12, 591-603.*
- *[29] Christos A. Nicolaou,*,†,‡ Joannis Apostolakis,§ and Costas S. Pattichis† , 'De Novo Drug Design Using Multiobjective Evolutionary Graphs', J. Chem. Inf. Model. vol. 49(2), pp. 295-307, 2009*
- *[30] N. Brown et al. "Graph-based genetic algorithm and its application to the multi-objective evolution of median molecules," J. Chem. Inf. Model. vol. 44, pp. 1079-1087, 2004.*
- *[31] J.W. Kruisselbrink et al, "Enhancing search space diversity in multiobjective Evolutionary Drug Molecule Design using Niching," ACM 978-1-60558-325-9/09/07, 2009.*
- *[32] Christos Kannas, "A PARALLEL IMPLEMENTATION OF A MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM ", A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science at the University of Cyprus*
- *[33] Qingfu Zhang, Senior Member, IEEE, and Hui Li," MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 11, NO. 6, DECEMBER 2007*
- *[34] Gauvain Marquet, " Stratégies scalaires et parallèles en optimisation multi-objectifs ", Master informatique — Université Lille 1 Centre de recherche INRIA Lille - Nord Europe, Octobre 2013 – Février 2014*
- *[35] A. Jaszkiewicz, "On the performance of multiple-objective genetic local search on the 0/1 knapsack problem – A comparative experiment,"IEEE Trans. Evol. Comput., vol. 6, no. 4, pp. 402–412, Aug.2002.*
- *[36] Siwei Jiang1;2, Student Member, IEEE, Zhihua Cai2, Jie Zhang1, Yew-Soon Ong1," Pareto-adaptive Weight Vectors", School of Computer Engineering, Nanyang Technology University, Singapore1 School of Computer Science, China University of Geosciences, Wuhan, China*
- *[37] http://www.hindawi.com/journals/jam/2014/906147/*

Les références de chapitre 4

- *[1] Recent Advances in the Open Access Cheminformatics Toolkits, Software Tools, Workflow Environments,and Databases livre*
- *[2] Landrum G (2013) RDKit: cheminformatics and machine learning software. rdkit.org*
- *[3] http://ggasoftware.com/opensource/indigo*
- *[4] Guha R, Howard MT, Hutchison GR, Murray-Rust P, Rzepa H, Steinbeck C,Wegner J, Willighagen EL (2006) The blueobelisk interoperability in chemical informatics.J Chem Inf Model 46:991–998*
- *[5] Cao D-S, Xu Q-S, Hu Q-N, Liang Y-Z (2013) ChemoPy: freely available python package for computational biology and chemoinformatics. Bioinformatics 29:1092–1094*
- *[6] Cao Y, Charisi A, Cheng L-C, Jiang T, Girke T (2008) ChemmineR: a compound mining framework for R. Bioinformatics 24:1733–1734*
- *[7] Cao D-S, Xiao N, Xu Q-S, Chen AF (2014)Rcpi: R/Bioconductor package to generate*
- *various descriptors of proteins, compounds, and their interactions. Bioinformatics. doi:10.1093/bioinformatics/btu1624*
- *[8] http://wiki.chemkit.org/Main_Page*
- *[9] Herraez A (2006) Biomolecules in the computer:Jmol to the rescue. Biochem Mol BiolEduc 34:255–261*
- *[10] Krause S, Willighagen E, Steinbeck C (2000) JChemPaint—using the collaborative forces of the internet to develop a free editor for 2D chemical structures. Molecules 5:93–98*
- *[11] P. Ambure et al., "Recent Advances in the Open Access Cheminformatics Toolkits, Software Tools, Workflow Environments, and Databases," Methods in Pharmacology and Toxicology, DOI 10.1007/7653, 2014.*
- *[12] http://www.taverna.org.uk/*
- *[13] T. Kuhn, EL.Willighagen, A. Zielesny and C. Steinbeck, "CDK-Taverna: an open workflow environment for cheminformatics," BMC Bioinformatics, vol. 11, p. 159, 2010.*
- *[14] KNIME: The Konstanz Information Miner,Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias K¨otter, Thorsten Meinl, Peter Ohl, Christoph*

Sieb, Kilian Thiel, and Bernd Wiswedel,ALTANA Chair for Bioinformatics and Information Mining,Department of Computer and Information Science, University of Konstanz,

- *[15] S. Beisken, T. Meinl, B. Wiswedel, LF. ,de Figueiredo, MR. Berthold and C. Steinbeck, "KNIME-CDK: workflow-driven cheminformatics," BMC Bioinformatics, pp. 214-257, 2013.*
- *[16] O'boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33*
- *[17] Muthukumarasamy Karthikeyan • Renu Vyas, "Practical Chemoinformatics", Division of Chemical Engineering and process development National Chemical Laboratory Pune India.*
- *[18] R. Vasundhara Devi, S. Siva Sathya, Mohane Selvaraj Coumar, " Multi-Objective Genetic Algorithm for De Novo Drug Design", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4, Issue-2, May 2014.*
- *[19] http://www.chups.jussieu.fr/polys/pharmaco/poly/cinetique.html*
- *[20] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (March 2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". Adv. Drug Deliv. Rev. 46 (1-3): 3–26.*
- *[21] Lipinski CA (December 2004). "Lead- and drug-like compounds: the rule-of-five revolution". Drug Discovery Today: Technologies 1 (4): 337–341. doi:10.1016/j.ddtec.2004.11.007.*
- *[22] N. Brown, Chemoinformatics - An introduction for Computer Scientists. ACM Computing Surveys, 41, 2009, 8.*
- *[23] T. Tanimoto. An Elementary Mathematical theory of Classification and Prediction. "IBM Internal Report," IBM technical report series, 1957.*
- *[24] http://sis.univ-tln.fr/~tollari/TER/AlgoGen1/node5.html*
- *[25] C. Bertelle, « Chap. 2 : Algorithmes génétiques », LIH - Laboratoire d'Informatique du Havre M1-Math-Info Le Havre.*