

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED KHIDER BISKRA

Faculté des sciences exactes et
des sciences de la nature et de la vie

Département des Sciences de la matière

THESE

Présentée par

ZEKRI Afaf

En vue de l'obtention du diplôme de :

DOCTORAT EN SCIENCES

Option :

Chimie informatique et pharmaceutique

Intitulée :

*Une étude computationnelle des relations structure-
activité d'un ensemble de composés à visées
thérapeutiques*

Soutenue le 20 Mai 2021

Devant la commission d'Examen

M. BOUMEDJANE Youcef	Prof.	Université de Biskra	Président
M. TOUAFRI Lasnoui	MC/A	Université de Khemis Miliana	Examineur
M. HELLAL Abdelkader	MC/A	Université de Sétif 1	Examineur
M. BOUHLALEG Lazhar	MC/A	Université de Batna 2	Examineur
M ^{me} . HARKATI Dalal	MC/A	Université de Biskra	Directrice de thèse
M. BELAIDI Salah	Prof.	Université de Biskra	Co-Directeur

Remerciements

Ce travail a été réalisé à l'Université Med Khider Biskra, Département des sciences de la matière, au Laboratoire de Chimie Moléculaire et Environnement (LCME) au sein de l'équipe de Chimie Informatique et Pharmaceutique.

Avant tout, je remercie Dieu tout puissant pour m'avoir aidé à réaliser ce modeste travail et de m'avoir permis d'arriver à ce stade.

Je tiens à remercier très sincèrement ma directrice de thèse Madame HARKATI Dalal, Maître de conférences « A » à l'Université Mohamed Khider Biskra, pour avoir dirigé mes travaux et pour toute l'aide qu'elle m'a apportée au cours de ce travail. Son intérêt, sa disponibilité, son suivi quotidien de mes travaux et ses précieux conseils m'ont été d'un grand profit. Je la remercie infiniment.

Je remercie Monsieur BELAIDI Salah, Professeur à l'Université de Biskra pour son aide et son intérêt à mon travail.

J'exprime ma profonde et respectueuse gratitude à Monsieur BOUMEDJANE Youcef, Professeur à l'Université de Biskra, qui m'a fait l'honneur d'accepter de présider le jury de cette thèse.

Je tiens à adresser mes vifs remerciements et l'expression de mon profond respect aux membres de jury : à Mr. TOUAFRI Lasnoui, Maître de conférences « A » à l'Université de Khemis Miliana, Mr HELLAL Abdelkader, Maître de conférences « A » à l'Université de Sétif 1 et Mr BOUHLALEG Lazhar, Maître de conférences « A » à l'Université de Batna 2, pour avoir accepté d'examiner et de juger ce travail.

Je remercie KENOUCHE Samir, Maître de conférences « A » à l'Université Mohamed Khider Biskra pour l'intérêt porté à ce travail de thèse, les nouvelles idées qu'il m'a présentées et pour sa contribution fructueuse à une publication liée à ce travail.

J'adresse mes sincères remerciements à Mr Daoud Ismail, Maître de conférences « A » à l'Université Mohamed Khider Biskra pour m'avoir fait don de quelques idées lumineuses et ses précieux conseils.

Un grand merci à tous mes collègues de l'équipe de « chimie informatique et pharmaceutique » du laboratoire de recherche LMCE et à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Enfin, je remercie toute ma famille, particulièrement ma mère, pour le soutien, l'affection, et les encouragements qu'elle m'a apportés durant toute ma vie.

À mes très chers parents

À mon fils Temim

À mes sœurs et à mes frères,

À toute ma famille

À tous ceux qui sont chers.

Table des matières

Remerciements.....	I
Table des matières.....	IV
Liste des tableaux.....	VIII
Liste des figures.....	IX
Liste des abréviations.....	XI
Introduction générale.....	1
Références bibliographiques.....	6

Chapitre I

Aperçue bibliographique sur les récepteurs aux estrogènes et les anti-estrogènes.

1	Introduction.....	10
1.1	Récepteurs hormonaux.....	11
1.1.1	Récepteurs à la progestérone.....	11
1.1.2	Récepteurs aux estrogènes.....	11
1.2	Traitement par hormonothérapie.....	12
2	Récepteurs aux estrogènes.....	14
3	Anti-estrogènes.....	20
3.1	Tamoxifène.....	20
3.2	Torémifène.....	21
3.3	Fulvestrant.....	22
4	Conclusion.....	23
	Références bibliographiques.....	24

Chapitre II

Étude quantitative des relations structure-activité d'une série de dérivés de l'indazole.

1	Introduction.....	28
2	Aperçue générale sur la relation quantitative structure-activité.....	29
2.1	Préparation d'un ensemble des molécules.....	30
2.2	Génération des descripteurs moléculaires.....	31
2.3	Sélection des descripteurs moléculaires.....	31
2.3.1	Sélection objective.....	32

2.3.2	Sélection subjective.....	32
2.3.2.1	Introduction progressive.....	32
2.3.2.2	Elimination progressive.....	33
2.3.2.3	Sélection pas à pas.....	33
2.4	Construction du modèle de QSAR.....	33
2.4.1	Critères de choix de modèle (Cp de Mallows et le BIC).....	34
2.4.1.1	Cp de Mallows.....	35
2.4.1.2	BIC.....	35
2.4.2	Algorithmes de sélection.....	36
2.5	Validation du modèle.....	37
2.5.1	Validation interne.....	37
2.5.2	Validation externe.....	38
2.5.3	Coefficient de détermination.....	38
2.5.4	Test de Fisher.....	38
2.6	Critères de validation.....	39
3	Matériels et Méthodes.....	40
3.1	Préparation d'un ensemble des molécules de dérivés de l'indazole	40
3.2	Génération des descripteurs moléculaires de dérivés de l'indazole	45
3.3	Sélection des descripteurs moléculaires de dérivés de l'indazole.....	46
3.4	Construction du modèle de dérivés de l'indazole.....	49
3.5	Validation du modèle QSAR de dérivés de l'indazole.....	49
4	Résultats et discussions.....	50
5	Conclusion.....	56
	Références bibliographiques.....	57

Chapitre III

Étude de docking moléculaire d'une série de dérivés de l'indazole.

1	Introduction.....	61
2	Aperçue générale sur le docking moléculaire.....	62
2.1	Différentes techniques du docking moléculaire	64
2.1.1	Docking rigide.....	64
2.1.2	Docking semi-flexible.....	65
2.1.3	Docking flexible.....	65
2.2	Algorithmes de docking moléculaire	66
2.3	Fonction de score.....	67

2.4	Interactions entre le ligand et le récepteur.....	68
2.4.1	Interaction ionique.....	68
2.4.2	Interaction hydrogène.....	69
2.4.3	Interaction de Van der Waals.....	70
2.4.4	Interactions hydrophobes.....	71
2.5	Étapes typiques du docking moléculaire.....	72
2.5.1	Détermination des structures.....	73
2.5.2	Préparation des structures.....	73
2.5.3	Docking moléculaire.....	73
2.5.4	Prédiction et évaluation.....	73
3	Aperçue générale sur les propriétés ADME.....	73
4	Matériels et Méthodes.....	74
4.1	Préparation du récepteur (ID : 3ERT).....	74
4.2	Préparation des structures de dérivés de l'indazole.....	75
4.3	Etude de docking moléculaire de dérivés de l'indazole.....	77
4.4	Etude des propriétés ADME de dérivés de l'indazole.....	77
5	Résultats et discussions.....	77
5.1	Résultats de docking moléculaire.....	77
5.2	Evaluation des propriétés ADME in silico de dérivés de l'indazole.....	88
6	Conclusion.....	90
	Références bibliographiques.....	91

Chapitre IV

Étude de la réactivité chimique d'une série de dérivés de l'indazole.

1	Introduction.....	96
2	Aperçue générale sur la réactivité chimique.....	96
2.1	Théorie des orbitales moléculaires frontières.....	96
2.2	DFT conceptuelle.....	97
2.3	Fondements de la DFT.....	97
2.3.1	Premier théorème de Hohenberg-Kohn.....	98
2.3.2	Deuxième théorème de Hohenberg-Kohn.....	98
2.4	Descripteurs de la DFT Conceptuelle.....	99
2.4.1	Descripteurs globaux.....	101
2.4.1.1	Potentiel chimique.....	101
2.4.1.2	Dureté chimique.....	101

2.4.1.3	Mollesse.....	102
2.4.1.4	Indice d'électrophilie globale (Indice d'électrophilicité).....	102
2.4.1.5	Indice de nucléophilie globale (Indice de nucléophilicité).....	103
2.4.2	Gap énergétique.....	103
2.4.3	Énergie totale.....	104
3	Matériels et Méthodes	104
4	Résultats et discussions.....	105
5	Conclusion.....	108
	Références bibliographiques.....	109
	Conclusion générale	113

Liste des tableaux

Chapitre II

Tableau 1	Structures chimiques développées et les activités biologiques expérimentales des molécules utilisées dans l'étude QSAR.....	40
Tableau 2	Descripteurs moléculaires sélectionnés.....	47
Tableau 3	Diagnostic statistique pour la régression.....	52
Tableau 4	Matrice de corrélation entre pIC ₅₀ et les descripteurs impliqués dans le modèle QSAR.....	54

Chapitre III

Tableau 1	Données cristallographiques.....	75
Tableau 2	Résidus de la première cavité détectée par MOE du RE α	78
Tableau 3	Scores et interactions entre les dérivés de l'indazole sélectionnés et les résidus du site actif de la protéine (ID : 3ERT).)	79
Tableau 4	Résultats de RMSD de docking moléculaire pour les 6 composés possédant de meilleurs scores.....	88
Tableau 5	Propriétés ADME pour les six composés ayant les meilleures énergies de scores.....	89

Chapitre IV

Tableau 1	Formats SMILES des composés utilisés dans l'étude des grandeurs globales de la réactivité.....	105
Tableau 2	Énergies HOMO et LUMO, gap énergétique, énergie totale, et indices de réactivité globaux μ , η , ω et N de dérivés de l'indazole sélectionnés.....	106

Liste des figures

Introduction générale

Figure 1.	Structures chimiques des anti-estrogènes.	2
------------------	--	---

Chapitre I

Figure 1.	Structure du sein.....	10
Figure 2.	Stratégies du traitement hormonal du cancer du sein.....	13
Figure 3.	Rôle des estrogènes dans la cancérogenèse mammaire.....	15
Figure 4.	Distribution tissulaire des RE α et RE β	16
Figure 5.	Relation structure fonction des récepteurs aux estrogènes.....	16
Figure 6.	Voies de signalisation de RE.....	18
Figure 7.	Structures de tamoxifène	20
Figure 8.	Structure de torémifène.....	21
Figure 9.	Structure du fulvestrant.....	22

Chapitre II

Figure 1.	Procédure suivie dans la construction du modèle de QSAR.....	30
Figure 2.	Boîte à moustaches (Box plots) de la distribution de la déviation standard d'erreur de prédiction (SDEP) par le nombre de descripteurs moléculaires..	50
Figure 3.	Proportions des descripteurs moléculaires sélectionnés.....	51
Figure 4.	Corrélation entre les activités biologiques prédites et expérimentales pour le modèle obtenu.....	54

Chapitre III

Figure 1.	Illustration schématique de l'encrage d'une petite molécule à une cible protéique produisant un complexe stable	62
Figure 2.	Principe général du docking moléculaire	64
Figure 3.	Schéma illustrant le principe de « lock and key » d'Emile Fisher.....	65
Figure 4.	Interaction entre deux molécules de charges différentes.....	69
Figure 5.	Exemples d'une liaison hydrogène.....	70
Figure 6.	Patch hydrophobe.....	71

Figure 7.	Étapes typiques d'un docking.....	72
Figure 8.	Structure du récepteur aux estrogènes de type α (ID : 3ERT).....	75
Figure 9.	Dérivés de l'indazole utilisés dans l'étude de docking moléculaire.....	76
Figure 10.	Modèle simplifié et la cavité choisie de RE α	78
Figure 11.	Interactions entre le ligand 14 et les résidus de site actif de la protéine (ID : 3ERT) (Images (a) 2D et (b) 3D).....	81
Figure 12.	Interactions entre le ligand 35 et les résidus de site actif de la protéine (ID : 3ERT). (Images (a) 2D et (b) 3D).....	82
Figure 13.	Interactions entre le ligand 43 et les résidus de site actif de la protéine (ID : 3ERT). (Images (a) 2D et (b) 3D).....	83
Figure 14.	Interactions entre le ligand 8 et les résidus de site actif de la protéine (ID : 3ERT) (Images (a) 2D et (b) 3D).....	84
Figure 15.	Interactions entre le ligand 45 et les résidus de site actif de la protéine (ID : 3ERT). (Images (a) 2D et (b) 3D).....	85
Figure 16.	Interactions entre le ligand 44 et les résidus de site actif de la protéine (ID : 3ERT). (Images (a) 2D et (b) 3D).....	86
Figure 17.	Interactions entre le ligand de référence (4-Hydroxytamoxifen) et les résidus de site actif de la protéine (ID : 3ERT).....	87

Chapitre IV

Figure 1.	Dérivées de l'énergie aux 1 ^{er} et 2 ^{ème} ordres	100
------------------	--	-----

Liste des abréviations

ABS	absorption
ADME	Absorption, Distribution, Metabolism and Elimination
ADN	Acide DésoxyriboNucléique
AM1	Austrin Model 1
ATP	Adénosine triphosphate
AVK	Anti-vitamines K
B3LYP	Becke 3-Parameter Lee-Yang-Parr
BIC	Bayesian Information Criterion
CDFT	Conceptual Density Functionnal Theory
D	Density
DFT	Density Functional Theory
DM	Dipole Moment
E	Énergie
E₂	estradiol
ERE	Éléments de réponse aux estrogènes
E_t	Énergie totale
F	Fisher
F_{crit}	F critique
FMO	Frontier Molecular Orbital
HE	Hydration Energy
HBA	Hydrogen-Bond Acceptors
HBD	Hydrogen-Bond Donors
HOMO	Highest Occupied Molecular Orbital
IA	Inhibiteur d'aromatase
IC₅₀	la concentration molaire en inhibiteur, nécessaire pour inhiber 50 % de l'activité enzymatique
LMO	Leave-Many-Out
log P	coefficient de partition eau / octanol
LOO	Leave-One-Out
LUMO	Lowest Unoccupied Molecular Orbital
MAPK	Mitogen Activated Protein Kinase
Me	Méthyle

MLR	Multiple Linear Regression
MM	Mécanique Moléculaire
MOE	Molecular Operating Environment
MR	Molar refractivity
MV	Molecular Volume
MW	Molecular weight
n-ROTB	Number of Rotatable Bonds
OHT	4-Hydroxytamoxifen
OM	Orbitale Moléculaire
PDB	Protein Data Bank
Pc	Parachor
Polar	Polarisabilité
PSA	Polar Surface Area
QSAR	Quantitative structure-activity relationship
R ind	Refractive index
RE	Récepteurs aux estrogènes
RE+	Récepteurs aux estrogènes positifs
REα	Récepteurs aux estrogènes de type α
REβ	Récepteurs aux estrogènes de type β
RMN	Résonance Magnétique Nucléaire
RMSD	Root-Mean-Square Deviation
RP	Récepteur à la progestérone
S	Surface moléculaire
SDEP	Standard Deviation Error in Prediction
SERDs	Selective Estrogen Receptor Degradors
SERM	Selective Estrogen Receptor Modulator
SMILES	Simplified Molecular Input Line Entry Specification
surf ten	surface tension
Tam	tamoxifène
TCT	Tétracyanoéthylène
TPSA	Topological Polar Surface Area
2D et 3D	Deux et trois dimensions
ΔE	Écart énergétique
ΔG	Énergie libre

Introduction générale

Introduction générale

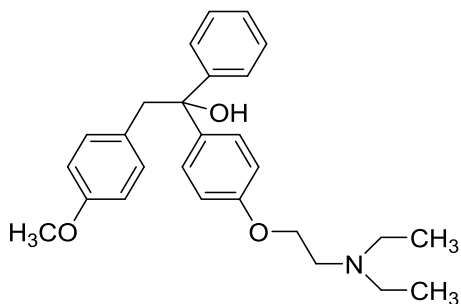
Le cancer est un problème de santé publique qui est en nette progression à travers le monde. ¹ Il touche toutes les catégories de la population mondiale quels que soient leurs âges, sexes, ou encore leurs niveaux socioéconomiques. ²

Le cancer du sein est le cancer féminin le plus fréquent ³ avec près de 1.7 million de nouveaux cas diagnostiqués chaque années à travers le monde. ⁴ En Afrique du Nord et au Moyen-Orient, le cancer du sein est également le premier cancer de la femme. Il représente 14 à 42 % de tous les cancers féminins avec une augmentation exponentielle. ⁵ Son incidence standardisée selon l'âge (ISA) varie entre 9,5 et 54 pour 105 femmes. Ces éléments font que l'Organisation mondiale de la santé (OMS) considère le cancer du sein comme une priorité de santé publique et un problème majeur chez la femme dans cette région du monde. ⁶

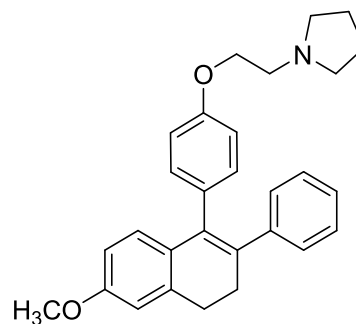
En Algérie la forte progression du cancer du sein prend des proportions épidémiques inquiétantes avec plus de 9 000 nouveaux cas en 2009, soit 54 nouveaux cas pour 100 000 femmes ; actuellement ce chiffre est estimé à 11 000, soit une augmentation de 500 nouveaux cas par an. De plus, le cancer du sein touche autant la femme jeune que la femme ménopause, ce qui complique son dépistage : l'âge médian est de 47 ans c'est-à-dire que 50% des cancers du sein ont déjà eu lieu avant cet âge. ⁷⁻⁸

La grande majorité des cancers du sein, soit 70 à 75%, expriment les récepteurs hormonaux à l'estrogène (RE) et/ou à la progestérone (RP). ⁹ En fait, les estrogènes interviennent dans le développement et la régulation des tissus gynécologiques (sein, utérus, ovaire) mais aussi dans celui d'un grand nombre d'autres tissus. Leur action biologique au niveau des cellules cibles passe par leur liaison à des protéines nucléaires réceptrices, les récepteurs aux estrogènes (RE). ¹⁰⁻¹¹ Les RE sont des régulateurs transcriptionnels capables de moduler la production de facteurs intervenant dans la prolifération et la différenciation cellulaire. Leur participation dans la cancérogenèse de certains tissus cibles est bien démontrée (sein, utérus). Les travaux de Lacassagne dans les années 1930 ont mis en évidence une relation entre les estrogènes et le développement des cancers du sein. ¹² Il a fallu attendre la découverte du premier récepteur des estrogènes (RE α) ¹¹ pour développer les premières molécules anti-estrogéniques (**figure 1**), (MER 25, nafoxidine, clomiphène, tamoxifène...).

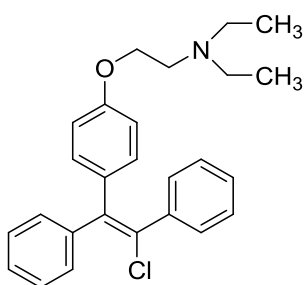
Introduction générale



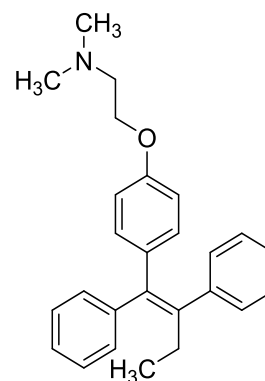
MER 25



Nafoxidine



Clomiphène



Tamoxifène

Figure 1. Structures chimiques des anti-estrogènes.

Il est important de noter que le tamoxifène a été, pendant plus de 30 ans, le chef de file des anti-estrogènes utilisés dans le traitement des cancers du sein hormono-dépendants, à la fois dans les formes avancées et comme traitement adjuvant. Sa place a changé depuis la mise sur le marché dans les années 2000 des IA de troisième génération. Le développement clinique du tamoxifène a débuté à la fin des années 1970 et c'est l'anti-estrogène pour lequel on dispose le plus de données cliniques et biologiques.¹³

La découverte et l'évolution d'un nouveau médicament présentent un processus très long, coûteux et difficile qui peut prendre entre 10 à 15 ans et plus. Malgré les progrès technologiques et une très bonne compréhension des systèmes biologiques, le développement des coûts a augmenté tandis que le nombre absolu de médicaments récemment sanctionnés a continuellement réduit. Traditionnellement, de nouveaux médicaments étaient découverts soit en identifiant le médicament biologiquement actif à partir de traitements traditionnels, soit par hasard arbitraire.¹⁴ Récemment, le

Introduction générale

développement de la chimie computationnelle a provoqué une révolution inhabituelle dans la chimie médicinale. Le modèle de découverte de molécules pharmacologiquement intéressantes a été changé au cours des dernières décennies, ce qui a considérablement amélioré l'efficacité de la découverte de médicaments entraînant une accélération significative du processus global.¹⁵

Les méthodes *in silico* telles que les relations quantitatives de structure-activité (QSAR) et le docking moléculaire peuvent réduire les échecs coûteux des candidats-médicaments en identifiant les principaux composés prometteurs et en réduisant ainsi le nombre d'expériences coûteuses. Ces méthodes font partie des principaux outils de découverte de médicaments. La relation quantitative structure-activité (QSAR), en tant que domaine important de la chimiométrie, joue un rôle majeur dans le développement d'un nouveau médicament. Conceptuellement, le QSAR est un outil qui permet la prédiction des activités des composés nouvellement désignés avant qu'une décision ne soit prise si ces composés doivent être réellement synthétisés et testés. L'étude QSAR s'intéresse essentiellement à la corrélation des activités biologiques spécifiques et de divers paramètres physico-chimiques (en termes de descripteurs) mesurés ou calculés d'une série de composés.¹⁶⁻¹⁸

L'utilisation des méthodes de docking dans le processus de conception de médicaments a débuté il y a plus de 30 ans.¹⁹ L'objectif était de prévoir la capacité ou non d'une molécule à se lier au site actif d'une protéine en se basant sur la prédiction de la conformation et de l'orientation de la molécule lors de sa liaison au récepteur.²⁰

Les propriétés Absorption, Distribution, Métabolisme, et Élimination (ADME) d'un médicament sont très importantes pour le succès clinique final d'un candidat médicament. Elles jouent un rôle très important au tout début de la découverte de médicaments pour réduire le taux d'échec des candidats médicaments dans les essais cliniques.²¹⁻²²

La prédiction de la réactivité des espèces chimiques est l'un des principaux objets de la chimie théorique, de nombreux travaux ont été réalisés sur cette ligne.²³ La théorie fonctionnelle de la densité est un outil très important pour calculer de nombreuses propriétés moléculaires dans les systèmes chimiques et biologiques.

Ainsi, dans ce travail une étude QSAR a été réalisée sur une série de dérivés d'indazole, qui sont de puissants antagonistes.²⁴ Notre choix de cette molécule est à cause

Introduction générale

de ses multiples propriétés intéressantes. En fait, l'indazole a fait l'objet de plusieurs travaux de recherche, d'ailleurs il est incorporé dans de nombreux produits présentant un intérêt majeur pour l'industrie pharmaceutique. A titre d'exemple, les dérivés de l'indazole sont connus en tant qu'agents antidépresseurs, anti-inflammatoires²⁵, analgésiques et antipyrétiques, antagonistes de la dopamine, antimicrobiens²⁶⁻²⁷, anticancéreux²⁸⁻²⁹ et anti-VIH³⁰. La structure de l'indazole est également présente dans d'autres composés tels que les herbicides et les colorants. Grâce à cette variété de propriétés, la chimie de l'indazole a suscité beaucoup d'intérêt non seulement sur le plan académique mais également sur le plan industriel, bien que de nombreux efforts aient été fournis pour l'élaboration de nouvelles voies de synthèse plus faciles, peu coûteuses et surtout régiosélectives.³¹

Notre étude tente de prédire des modèles QSAR de ces composés, les dérivés d'indazole, en utilisant la régression linéaire multiple (MLR). Pour tester les performances et la stabilité du modèle obtenu, une méthode de validation a été choisie.

Cette étude a été accompagnée d'un docking moléculaire sur une série de **21** composés de Selective Estrogen Receptor Degradables (SERD) qui présentent les composés les plus actifs de l'ensemble des données. La réactivité moléculaire de dérivés d'indazole a été analysée par les descripteurs globaux de réactivité pour bien comprendre la relation entre la structure, la stabilité et la réactivité de ces molécules. L'étude des propriétés ADME in silico a montré la ressemblance avec les médicaments (drug-likeness) de six composés ayant les meilleurs scores dans l'étude de docking moléculaire pour identifier les médicaments sans danger.

Donc, les résultats de notre étude ont été développés selon quatre chapitres :

Dans le premier chapitre nous présentons des généralités sur les récepteurs aux œstrogènes et les anti-œstrogènes.

Dans le deuxième chapitre nous présentons l'étude quantitative des relations structure-activité d'une série de 54 molécules de dérivés de l'indazole

Le troisième chapitre concerne l'étude de docking moléculaire suivie par l'étude des propriétés ADME sur les meilleurs composés obtenus par l'étude de docking moléculaire.

Introduction générale

Le quatrième chapitre est consacré à l'étude de la réactivité chimique de dérivés de l'indazole en se basant sur les descripteurs globaux de réactivité issus de la DFT conceptuelle. (Travail publié dans : Journal of Molecular Structure, Vol 1217, 2020, 128442).

Introduction générale

Références bibliographiques

1. Dangou, J.; Sambo, B.-H.; Moeti, M.; Diarra-Nama, A.-J., Prévention et lutte contre le cancer dans la région africaine de l’OMS: un appel à l’action. *Journal africain du cancer/African Journal of Cancer* **2009**, *1* (1), 56-60.
2. Maamri, A., Données épidémiologiques sur le cancer dans le monde et au Maroc Revue bibliographique. *Annales des sciences de la santé* **2015**, *1* (1), 20-29.
3. Belkacémi, Y.; Boussen, H.; Hamdi-Cherif, M.; Benider, A.; Errihani, H.; Mrabti, H.; Bouzid, K.; Bensalem, A.; Fettouki, S.; Ben Abdalah, M.; Abid, L.; Gligorov, J., Épidémiologie des cancers du sein de la femme jeune en Afrique du Nord. *32° Journées de la Société Française de Sénologie et de Pathologie Mammaire. Journées, Strasbourg, FRA, 2010-11-03: La femme jeune face au cancer du sein (Strasbourg, 3-5 novembre 2010)/Young woman facing breast cancer* **2010**.
4. Ghita, F. B. Cancer du sein chez la femme jeune de moins de 35 ans. Thèse de doctorat Université Sidi Mohamed Ben Abdellah, **2019**.
5. El Saghir, N. S.; Khalil, M. K.; Eid, T.; El Kinge, A. R.; Charafeddine, M.; Geara, F.; Seoud, M.; Shamseddine, A. I., Trends in epidemiology and management of breast cancer in developing Arab countries: a literature and registry analysis. *International journal of surgery* **2007**, *5* (4), 225-233.
6. Hance, K. W.; Anderson, W. F.; Devesa, S. S.; Young, H. A.; Levine, P. H., Trends in inflammatory breast carcinoma incidence and survival: the surveillance, epidemiology, and end results program at the National Cancer Institute. *Journal of the National Cancer Institute* **2005**, *97* (13), 966-975.
7. Comité national charge de suivi de la lutte contre le cancer en Algérie, Plan cancer 2014-2019 (mise à jour Avril 2017) consulté en **14/ 04/2017**.
8. Derouache, N. E. H.; Dib, N.; Hachim, A. Cancer du sein chez la femme « Etude descriptive et rétrospective de cas de cancer du sein» Prise en charge au niveau du service de la gynécologie obstétrique du CHU de Tlemcen de 2014 - 2016. Thèse de doctorat en médecine, Université Abou Bakr Belkaid– Tlemcen, **2016**.
9. Odermatt, R.; Wolfer, A.; Zaman, K., Hormonothérapie dans le cancer du sein: efficacité et effets adverses. *Rev Med Suisse* **2013**, *9*, 1090-1094.
10. Jensen, E. V.; De Sombre, E. R., Estrogen-receptor interaction: Estrogenic hormones effect transformation of specific receptor proteins to a biochemically functional form. *Science* **1973**, *182* (4108), 126-134.

Introduction générale

11. Jensen, E. V.; Jacobson, H. I., Basic guides to the mechanism of estrogen action. *Recent Progr Hormone Res* **1962**, *18*, 387-414.
12. Lacassagne, A., Hormonal pathogenesis of adenocarcinoma of the breast. *The American Journal of Cancer* **1936**, *27* (2), 217-225.
13. De Cremoux, P., Hormonothérapie des cancers du sein. *Bulletin du cancer* **2011**, *98* (11), 1311-1319.
14. Agatonovic-Kustrin, S.; Morton, D., Data Mining in Drug Discovery and Design. In *Artificial Neural Network for Drug Design, Delivery and Disposition*, Elsevier: **2016**; pp 181-193.
15. Wu, G.; Zhao, T.; Kang, D.; Zhang, J.; Song, Y.; Namasivayam, V.; Kongsted, J.; Pannecouque, C.; De Clercq, E.; Poongavanam, V.; Liu, X.; Zhan, P., Overview of recent strategic advances in medicinal chemistry. *Journal of Medicinal Chemistry* **2019**, *62* (21), 9375-9414.
16. Hansch, C.; Leo, A., *Substituent constants for correlation analysis in chemistry and biology*. Wiley: **1979**.
17. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A., QSAR modeling: where have you been? Where are you going to? *Journal of Medicinal Chemistry* **2014**, *57* (12), 4977-5010.
18. Eroglu, E.; Türkmen, H., A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. *Journal of Molecular Graphics and Modelling* **2007**, *26* (4), 701-708.
19. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **1982**, *161* (2), 269-288.
20. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **2004**, *3* (11), 935-949.
21. Li, A. P., Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today* **2001**, *6* (7), 357-366.
22. Zhang, Y.-H.; Xia, Z.-N.; Qin, L.-T.; Liu, S.-S., Prediction of blood-brain partitioning: A model based on molecular electronegativity distance vector descriptors. *Journal of Molecular Graphics and Modelling* **2010**, *29* (2), 214-220.

Introduction générale

23. Babu, N. S.; Jayaprakash, D., Global and reactivity descriptors studies of cyanuric acid tautomers in different solvents by using of density functional theory (DFT). *HOMO* **2015**, *1*, 1.
24. Lai, A.; Kahraman, M.; Govek, S.; Nagasawa, J.; Bonnefous, C.; Julien, J.; Douglas, K.; Sensintaffar, J.; Lu, N.; Lee, K.-j.; Aparicio, A.; Kaufman, J.; Qian, J.; Shao, G.; Prudente, R.; Moon, M. J.; Joseph, J. D.; Darimont, B.; Brigham, D.; Grillot, K.; Heyman, R.; Rix, P. J.; Hager, J. H.; Smith, N. D., Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. *Journal of Medicinal Chemistry* **2015**, *58* (12), 4888-4904.
25. Schenone, S.; Bruno, O.; Ranise, A.; Brullo, C.; Bondavalli, F.; Filippelli, W.; Mazzeo, F.; Capuano, A.; Falcone, G., 2-Aryl-3-phenylamino-4, 5-dihydro-2h-benz [g] indazoles with analgesic activity. *Il Farmaco* **2003**, *58* (9), 845-849.
26. Park, J. S.; Yu, K. A.; Kang, T. H.; Kim, S.; Suh, Y.-G., Discovery of novel indazole-linked triazoles as antifungal agents. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17* (12), 3486-3490.
27. Yakaiah, T.; Lingaiah, B. P. V.; Narsaiah, B.; Kumar, K. P.; Murthy, U. S. N., GdCl₃ catalysed Grieco condensation: a facile approach for the synthesis of novel pyrimidine and annulated pyrimidine fused indazole derivatives in single pot under mild conditions and their anti-microbial activity. *European Journal of Medicinal chemistry* **2008**, *43* (2), 341-347.
28. Bouissane, L.; El Kazzouli, S.; Léonce, S.; Pfeiffer, B.; Rakib, E. M.; Khouili, M.; Guillaumet, G., Synthesis and biological evaluation of N-(7-indazolyl) benzenesulfonamide derivatives as potent cell cycle inhibitors. *Bioorganic & Medicinal Chemistry* **2006**, *14* (4), 1078-1088.
29. Daidone, G.; Raffa, D.; Maggio, B.; Raimondi, M. V.; Plescia, F.; Schillaci, D., Synthesis and antiproliferative activity of triazenoindazoles and triazenopyrazoles: a comparative study. *European Journal of Medicinal Chemistry* **2004**, *39* (3), 219-224.
30. Rodgers, J. D.; Johnson, B. L.; Wang, H.; Erickson-Viitanen, S.; Klabe, R. M.; Bacheler, L.; Cordova, B. C.; Chang, C.-H., Potent cyclic urea HIV protease inhibitors with 3-aminoindazole P2/P2' groups. *Bioorganic & Medicinal Chemistry Letters* **1998**, *8* (7), 715-720.

Introduction générale

31. Ben-Yahia, A.; BENCHIDMI, M.; ESSASSI, E. M., L'indazole et ses dérivés: synthèses, réactivités et propriétés biologiques. *Moroccan Journal of Heterocyclic Chemistry* **2018**, *17* (1), 127-163.

Chapitre I

Aperçue bibliographique sur les récepteurs aux estrogènes et les anti-estrogènes.

1 Introduction

Le cancer du sein représente un problème majeur de santé publique, il constitue la première cause de décès chez les femmes entre 40 et 50 ans. Environ 60 % des patientes non ménopausées et 75% des patientes ménopausées présentent un cancer du sein. ¹ C'est une maladie hétérogène la plus fréquemment diagnostiquée et la cause majeure de décès chez les femmes à travers le monde. Le cancer du sein est une tumeur maligne du sein, où les cellules mammaires se multiplient de façon anarchique suite à une défaillance du mécanisme de régulation de la croissance cellulaire. ² Plusieurs types de cancers du sein existent en fonction de leur localisation au niveau du sein. Le sein possède une structure assez complexe composée de tissus graisseux dans lesquels se trouvent des compartiments comprenant des lobules et des canaux (**figure 1**). Les lobules servent à produire le lait en période d'allaitement et les canaux servent à transporter le lait vers le mamelon.

Le sein comporte également un important système de drainage lymphatique, ce qui explique la possible dissémination via ce réseau.

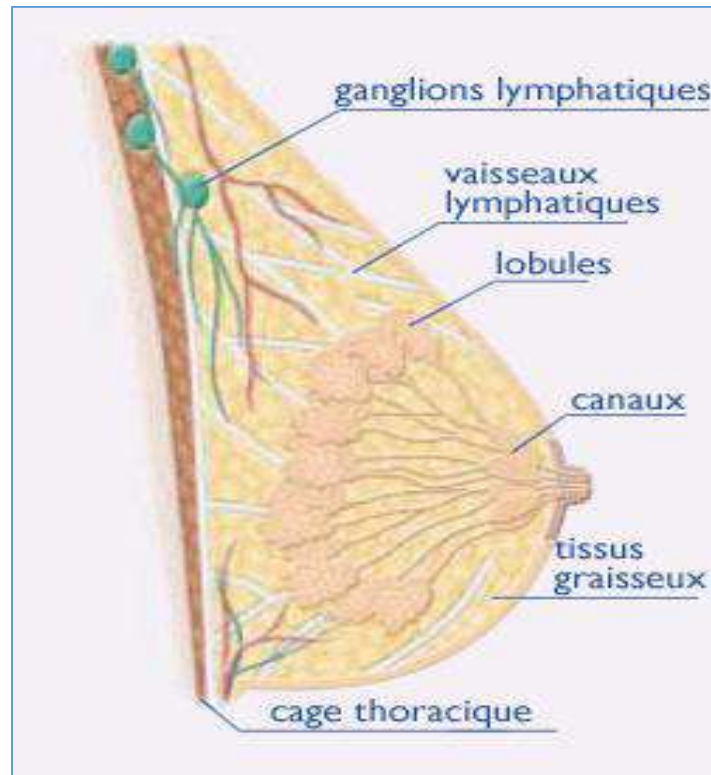


Figure 1. Structure du sein

Ces différentes structures peuvent être atteintes par le cancer. Le type le plus fréquemment rencontré est l'adénocarcinome qui se développe au dépend de la glande mammaire.

Le cancer peut se développer soit au dépend des cellules des canaux, on parle alors de cancer canalaire (le plus fréquent), soit à partir des lobules, on parle de cancer lobulaire. Il existe aussi des cancers in situ ou infiltrant. ³

1.1 Récepteurs hormonaux

Le rôle des hormones sexuelles dans les cancers du sein a été supposé dès le XIX^{ème} siècle par sir Beatson. ⁴ Les récepteurs hormonaux sont des protéines intracellulaires. L'hormone va se fixer de manière sélective et spécifique sur ces récepteurs. Il existe deux types de récepteurs hormonaux : ⁵

- ❖ Récepteurs à la progestérone
- ❖ Récepteurs aux estrogènes

1.1.1 Récepteurs à la progestérone

Le récepteur à la progestérone (RP) est un récepteur stéroïde, car il lie de façon spécifique la progestérone, hormone stéroïde importante pour le fonctionnement de l'organisme, notamment en période de grossesse et de cycles menstruels chez la femme. Le RP est exprimé dans les tissus des organes reproducteurs, en particulier dans les ovaires, mais aussi dans le tissu osseux, la prostate, les tissus adipeux ainsi que dans les membranes des vaisseaux sanguins. ⁶⁻⁸

1.1.2 Récepteurs aux estrogènes

Les récepteurs aux estrogènes (RE) se divisent en 2 groupes, les RE α et les RE β . Leurs gènes sont situés sur 2 chromosomes différents : le chromosome **6** pour le RE α et le chromosome **14** pour le gène du RE β . ⁹

Ils appartiennent à la grande famille des récepteurs nucléaires, qui comprend les récepteurs des différentes hormones stéroïdes, et des hormones thyroïdiennes. ⁹⁻¹¹

1.2 Traitement par hormonothérapie

L'hormonothérapie fait partie des traitements utilisés dans le cancer du sein, de même que la chirurgie, la chimiothérapie et la radiothérapie. Toutefois, afin de pouvoir bénéficier de ce type de traitement, il faut que la tumeur soit hormonosensible.⁵ L'hormonothérapie a pour but d'arrêter la prolifération cellulaire ; elle revêt une grande importance du fait du caractère hormono-dépendant de nombreux cancers du sein. Elle repose sur deux stratégies : la première utilise des antiestrogènes pour bloquer l'action d'estrogène au niveau de leurs récepteurs, la deuxième consiste en l'inhibition de la biosynthèse de ces hormones par les inhibiteurs de l'aromatase.

Il existe trois types d'hormonothérapie :¹²

- ❖ L'hormonothérapie inhibitrice avec les anti-aromatases. Elle consiste en l'inhibition de l'aromatase, enzyme responsable de la conversion des androgènes en estrogènes (**figure 2, voie a**). L'inhibition de la biosynthèse des estrogènes prévient non seulement une prolifération cellulaire liée à la fixation des estrogènes sur leurs récepteurs mais également les effets génotoxiques des métabolites de ces estrogènes.
- ❖ L'hormonothérapie compétitive avec les anti-estrogènes. Ces derniers bloquent l'action des estrogènes en se fixant aux récepteurs estrogéniques des cellules tumorales (**figure 2, voie b**).
- ❖ L'hormonothérapie additive avec la prise de progestatifs de synthèse (la médroxyprogestérone, l'acétate de mégestrol). A forte dose, ils se comportent comme les anti-estrogènes. Cependant, certains progestatifs inhibent également des enzymes impliquées dans la formation de l'estradiol (sulfatases et 17 β -hydroxystéroïde déshydrogénases).¹³ Peu d'informations sur les effets de ces progestatifs vis-à-vis du cancer du sein sont pour l'instant disponibles, aussi leur utilisation reste très limitée.

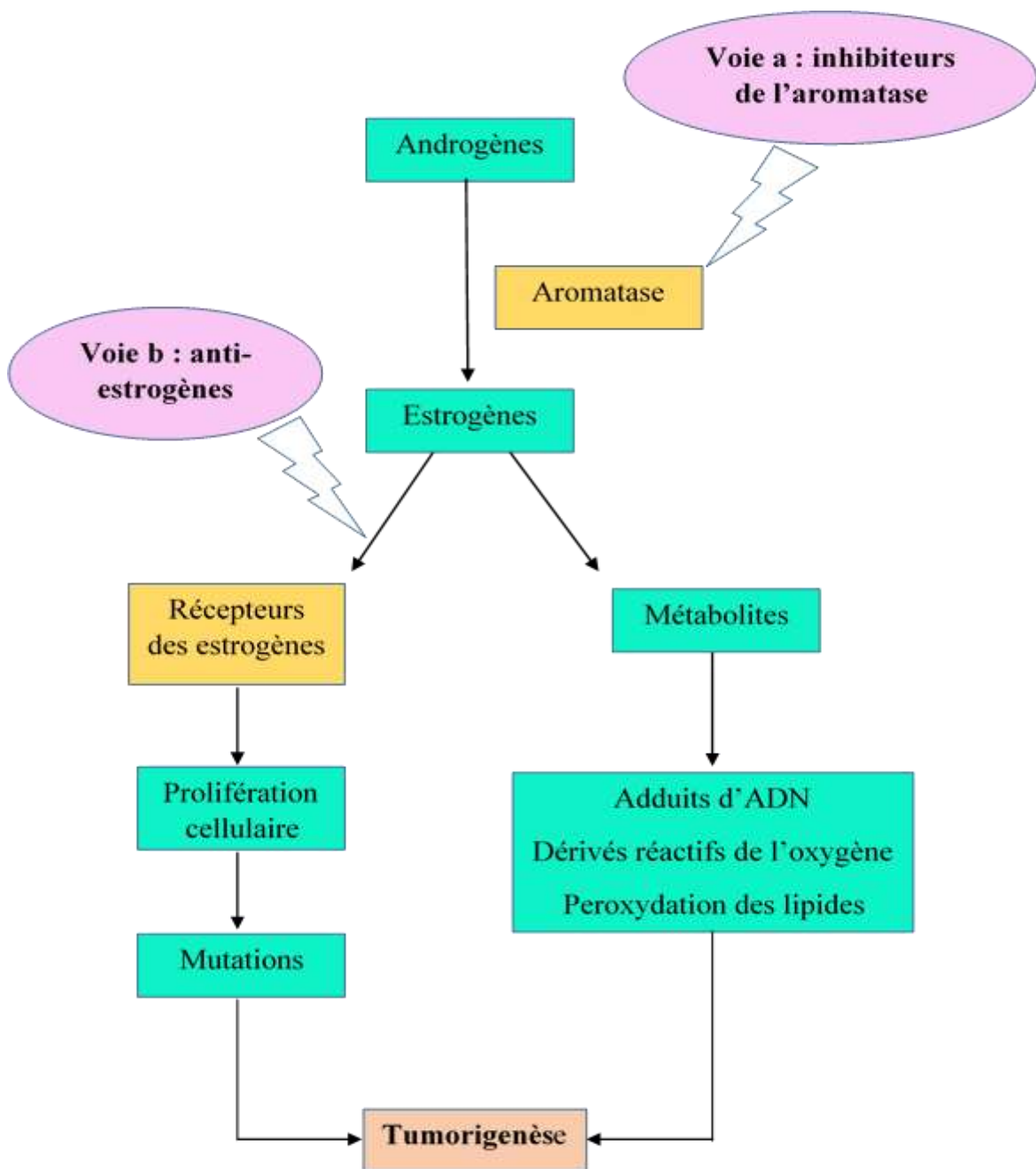


Figure 2. Stratégies du traitement hormonal du cancer du sein

Dans la **figure 2** Voie a présente inhibition de la biosynthèse des estrogènes par les anti-aromatase et Voie b : blocage de l'action des estrogènes au niveau de leurs récepteurs.

Le niveau d'action des inhibiteurs de l'aromatase, permet d'expliquer l'efficacité supérieure de ces molécules par rapport aux anti-estrogènes, supériorité démontrée par plusieurs études récentes.¹⁴⁻¹⁵

2 Récepteurs aux estrogènes

Les estrogènes sont des hormones à **18** atomes de carbone secrétées naturellement par les ovaires lors de chaque cycle menstruel avant la ménopause. Cette synthèse est en constante modulation, elle est plus importante lors de la première partie du cycle menstruel durant la phase folliculaire. Cette sécrétion augmente de façon croissante jusqu'à l'ovulation. Après l'ovulation, le taux d'estrogène diminue. ³

Plusieurs études montrent que ces hormones augmentent le risque de développer un cancer du sein chez la femme pré et post-ménopausée. Ces hormones favorisent également la croissance et la progression continues des tumeurs cancéreuses. ¹² Elles interviennent également dans des effets biologiques au niveau cardiovasculaire, musculosquelettique et du système nerveux. Une fois produites, ces hormones passent dans le sang, agissent au niveau de leurs récepteurs et sont éliminées par les urines. ³

Deux mécanismes par lesquels les estrogènes pourraient être à l'origine du développement du cancer du sein ont été décrits (**figure 3**). ¹⁴⁻¹⁶

- ❖ 1^{er} mécanisme : il est démontré que l'estradiol peut promouvoir la prolifération de cellules cancéreuses. En effet, l'estradiol ou 17- β -estradiol (E2) peut se lier aux récepteurs des estrogènes et stimuler ainsi la transcription de gènes impliqués dans la division cellulaire (**figure 3, voie a**). Lors de ces divisions, les cellules sont plus sensibles aux erreurs de réplication et à l'effet d'agents cancérogènes extérieurs. Les estrogènes constituent donc des promoteurs de cancérogenèse.
- ❖ 2^{ème} mécanisme : certaines études décrivent une action initiatrice de la cancérogenèse qui passe cette fois par la conversion de l'estradiol en métabolites génotoxiques (**figure 3, voie b**). Différentes voies métaboliques peuvent conduire à la synthèse de composés génotoxiques. La voie principale conduit à la synthèse de dérivés hydroxylés de l'estradiol, les catéchols (2-hydroxy-estradiol et 4-hydroxy-estradiol principalement) qui peuvent aboutir à des semi-quinones et enfin à des quinones. Ces dernières, fortement réactives, sont capables d'une part de former des adduits d'ADN (liaison par covalence à la guanine ou à l'adénine), et d'autre part, de provoquer l'oxydation des lipides cellulaires, via la formation de dérivés réactifs de l'oxygène comme l'ion superoxyde O₂. ¹⁴

Ces deux mécanismes peuvent agir de concert pour stimuler la prolifération cellulaire, favoriser l'accumulation de mutations et augmenter ainsi la tumorigenèse.

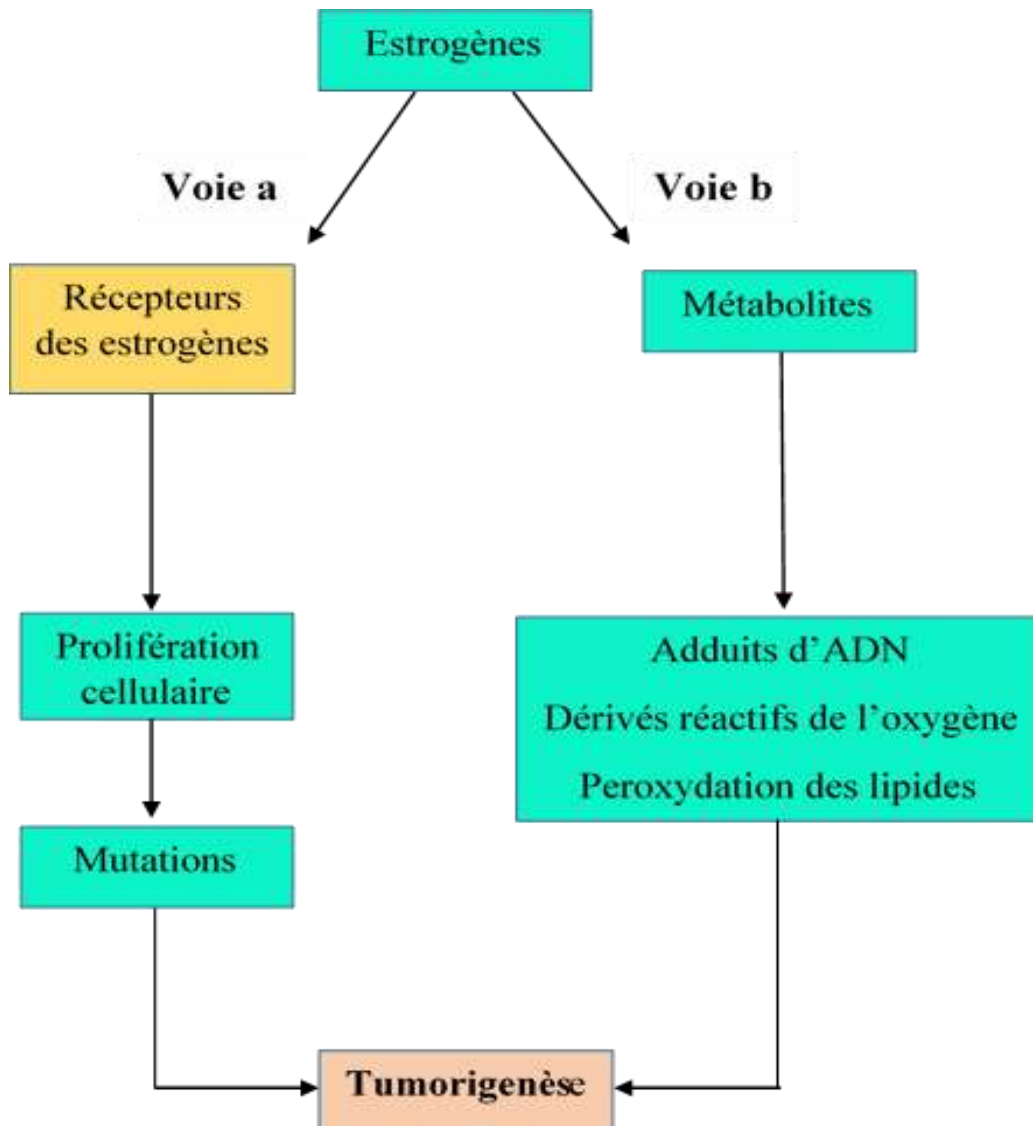


Figure 3. Rôle des estrogènes dans la cancérogenèse mammaire

❖ Localisation tissulaire

Les deux formes des récepteurs aux estrogènes $RE\alpha$ et $RE\beta$ ont une distribution non uniforme dans les différents tissus cibles (**figure 4**).¹⁷ Certains organes, comme le sein et l'utérus sont plus riches en $RE\alpha$. Pour d'autres, comme les cellules endothéliales sont plus riches en $RE\beta$.¹⁸⁻²³

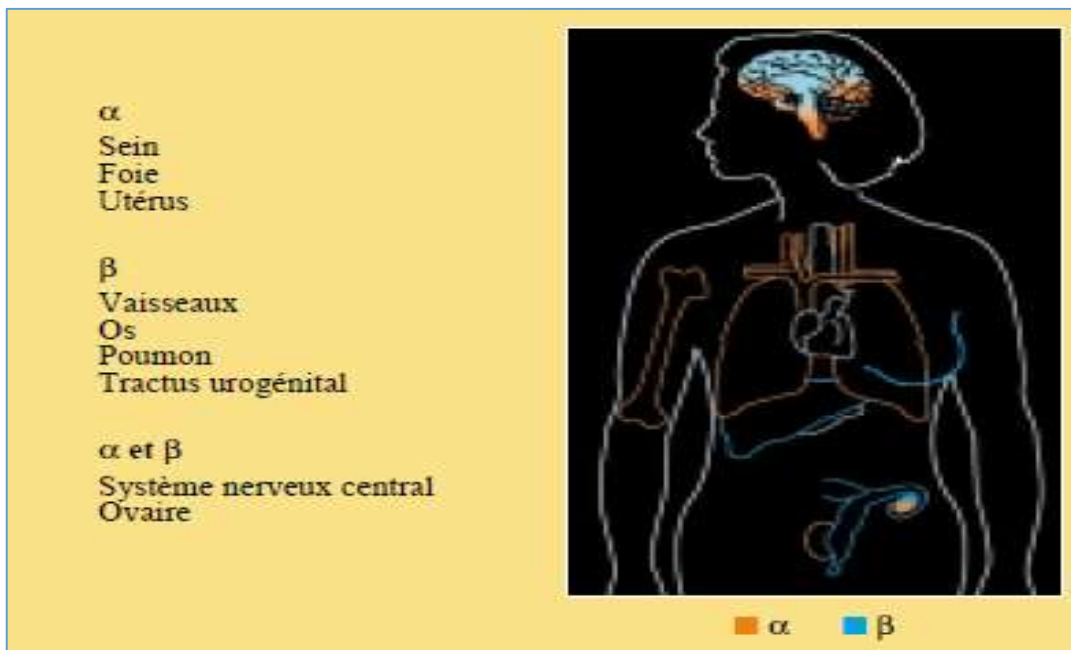


Figure 4. Distribution tissulaire des RE α et RE β .

❖ Structure

La structure primaire des RE est organisée en 6 domaines (de A à F) (figure 5) dont les fonctions sont bien définies.²⁴

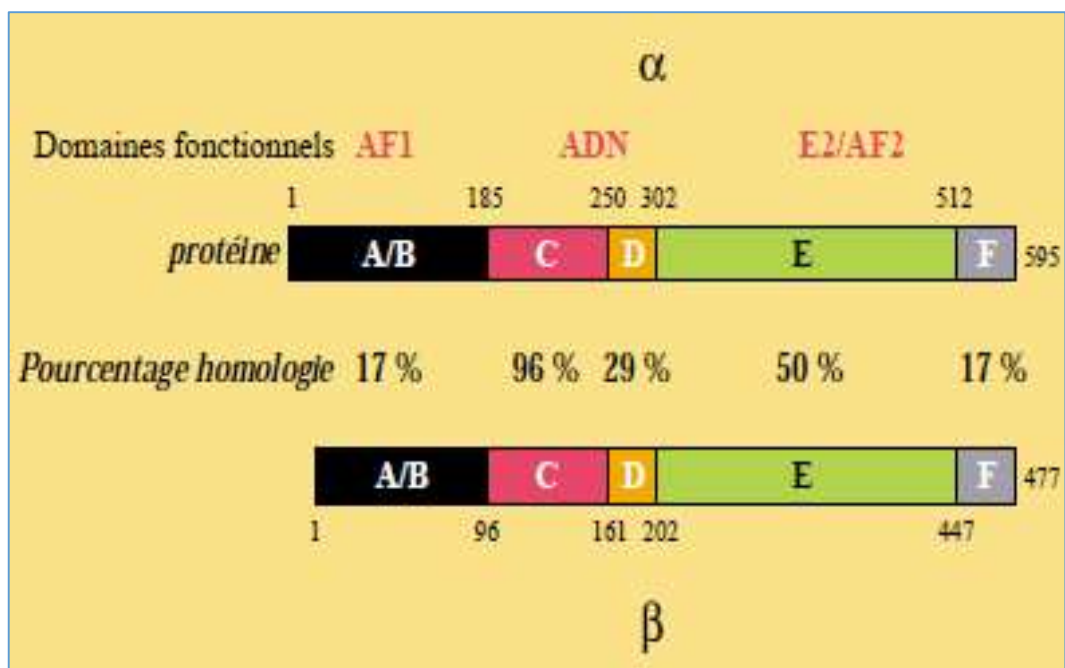


Figure 5. Relation structure fonction des récepteurs aux estrogènes.

Deux régions sont responsables de l'activité régulatrice de la transcription du RE :

- le domaine A/B a une fonction régulatrice, dite AF1, qui serait dépendante du tissu et serait activée dès que le RE se fixe sur l'ADN quel que soit le ligand (composé qui lie le RE) ;
- la région E/F est la région de liaison des ligands physiologiques (estradiol) et par extension des SERM. Elle est aussi responsable d'une activité régulatrice AF2 qui va dépendre du composé qui est lié au récepteur. La faible homologie de séquence des 2 isoformes du RE dans les régions AF1 et AF2 suggère que leurs fonctions biologiques pourraient être différentes ;
- la région C est le domaine de liaison à l'ADN. Il se lie à des régions spécifiques du RE appelées "éléments de réponse aux estrogènes" (ERE) et qui sont situées dans les régions régulatrices des gènes dont la transcription dépend des estrogènes.²⁵

❖ Rôles

Les récepteurs des estrogènes influencent de nombreux processus physiologiques chez les mammifères comme ceux impliqués dans la reproduction, les maladies cardio-vasculaires, le développement osseux, la cognition, le comportement. Etant donné leurs nombreux rôles, il n'est pas surprenant que les récepteurs des estrogènes soient aussi impliqués dans le développement de nombreux cancer comme le cancer du sein, des ovaires, de l'utérus ou de la prostate, ou dans d'autres maladies comme l'ostéoporose, les maladies cardio-vasculaires ou neurodégénératives, ou l'obésité. Les récepteurs aux estrogènes (REs), en particulier RE α , sont fortement impliqués dans le développement tumoral mammaire.²⁶

❖ Signalisation des REs

Les mécanismes de la cancérogénèse hormonale sont complexes. En absence de ligand, les REs sont séquestrés dans le cytosol et associés à des complexes multiprotéiques comprenant la protéine chaperonne Hsp-90²⁷ qui stabilise les récepteurs dans une conformation non active.²⁸ L'inhibition de la liaison de la Hsp-90 entraîne une dégradation des REs par le protéasome.²⁹⁻³¹

La signalisation des REs passe par quatre voies majeures (**figure 6**) : la voie génomique Estrogen Response Element (ERE)-dépendante classique, la voie génomique ERE-indépendante, la voie non-génomique : effet rapide des estrogènes, et la voie non-génomique à génomique : voie ligand-indépendante.

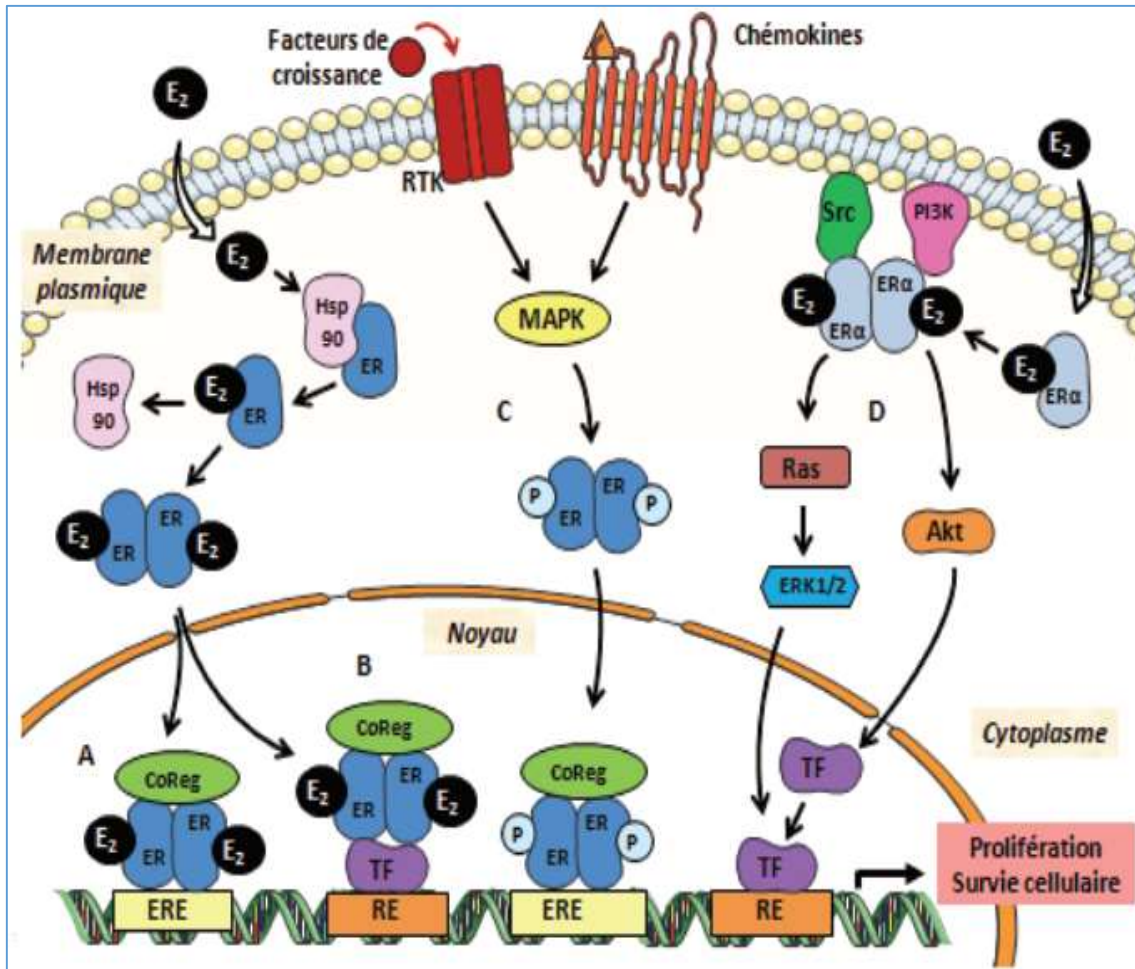


Figure 6. Voies de signalisation de RE

Quand les voies sont communes pour RE α et RE β , on parle de RE. Voie génomique (A), non classique (B), indépendante des estrogènes (C) et Voies non génomiques (D)

La voie classique, génomique ERE-dépendante : les estrogènes étant des hormones lipophiles, elles diffusent passivement à travers les membranes cellulaires. Leur liaison au LBD des récepteurs aux estrogènes va provoquer un changement de

conformation permettant la libération du récepteur du complexe inactivateur et sa dimérisation. Les dimères de récepteurs des estrogènes vont alors interagir spécifiquement avec l'ADN, au niveau de leurs éléments de réponse appelés ERE. Ces derniers sont situés dans les promoteurs des gènes cibles des estrogènes, tels que TFF1 ou *cmyc*.³²⁻³⁴

La voie génomique indirecte implique des interactions protéines-protéines entre le RE α et d'autres facteurs de transcription, en réponse à la liaison des estrogènes. En effet, le RE α est capable d'interagir avec les éléments de réponse d'autres facteurs de transcription, tels Fos/Jun (sites AP-1)³⁵ ou SP-1³⁶. Ces interactions vont permettre la régulation de l'expression des gènes par une interaction indirecte entre l'ADN et le récepteur des estrogènes en conformation agoniste.

Les effets rapides des estrogènes résultent de la voie non-génomique activée dans de nombreux tissus, mais pas encore bien comprise. Elle pourrait impliquer l'activation d'un récepteur à la membrane plasmique ou nucléaire. L'induction de la cascade d'effecteurs pourrait être activée directement par le récepteur ou mettre en jeu un second messenger, menant à la réponse physiologique rapide n'impliquant pas la régulation de gènes. Les estrogènes provoquent, par exemple, une mobilisation rapide du calcium intracellulaire, la production d'adénosine triphosphate (ATP), active la voie de signalisation des Mitogen Activated Protein Kinase (MAPK).³⁷ Plusieurs études suggèrent que ces réponses pourraient être médiées par une sous-population des récepteurs aux estrogènes RE α et RE β localisée au niveau de la membrane cellulaire.³⁷

Finalement, la voie de signalisation non-génomique à génomique est ligand - indépendante : elle est activée par d'autres voies de signalisation comme celle des facteurs de croissance, par exemple. Les récepteurs aux estrogènes peuvent subir des modifications post-traductionnelles, telles que leur phosphorylation par la voie des MAPK en réponse à des signaux de croissance reçus par les cellules. Le récepteur aux estrogènes RE α , ainsi activé, va se lier à ses EREs sur l'ADN et moduler la transcription de ses gènes cibles.³⁸

3 Anti-estrogènes

Les anti-estrogènes sont des médicaments qui entrent en compétition avec les estrogènes. Ils ne suppriment pas la sécrétion des estrogènes, mais ils prennent leur place et bloquent ainsi leurs effets de stimulation sur les cellules cancéreuses. Ils peuvent être classés en deux grands groupes ; les Selective Estrogen Receptor Degradation (SERD) et les Selective Estrogen Receptor Modulator (SERM) :

- ❖ Les SERD agissent sur les récepteurs hormonaux des cellules en s'y fixant avec une affinité comparable à celle de l'estradiol et en accélèrent la dégradation.
- ❖ Les SERM représentent un groupe de médicaments de structure chimique non stéroïdienne qui exercent des effets antagonistes et/ou agonistes des estrogènes selon le tissu cible. ⁵

Parmi les molécules anti-estrogènes, trois sont actuellement commercialisées : le tamoxifène (Nolvadex®), le torémifène (Fareston®) et le fulvestrant (Faslodex®).

3.1 Tamoxifène

Le tamoxifène est un composé dont la structure chimique est non stéroïdienne. Il s'agit d'un dérivé triphenyléthylénique (**figure 7**). Il appartient à la famille des modulateurs sélectifs des récepteurs aux estrogènes (SERM). ¹¹ Ainsi, il présente des propriétés antagonistes sur les récepteurs mammaires et agonistes sur les récepteurs localisés au niveau endométrial et osseux.

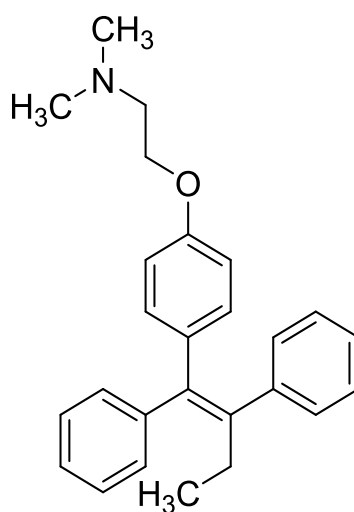


Figure 7. Structure de tamoxifène

La posologie habituelle est de 20 mg par jour en une à deux prises ; dans les formes évoluées, elle peut être augmentée à 40 mg par jour. La durée de prescription maximale est de cinq ans et il a été rapporté que le bénéfice du tamoxifène est d'autant plus important que le taux de récepteurs est élevé.³⁹ Associé aux anticoagulants oraux et plus particulièrement aux anti-vitamines K (AVK), il entraîne une augmentation du risque hémorragique par réduction de leur métabolisme.

La posologie des AVK peut, dans certains cas, être diminuée. Métabolisé par le CYP2D6, le tamoxifène ne devra être associé à des inducteurs ou des inhibiteurs de ce même cytochrome. Les principaux effets indésirables rencontrés sont des nausées et vomissements, des bouffées de chaleur (40 %), des oedèmes périphériques et des accidents thromboemboliques par diminution de l'activité antithrombine. En raison de son activité agoniste sur les récepteurs endométriaux, une augmentation du risque de cancer de l'endomètre a été constatée chez des patientes traitées sur de longues périodes par le tamoxifène.⁴⁰

3.2 Torémifène

Le torémifène est un dérivé du tamoxifène (**figure 8**). C'est un antagoniste compétitif des récepteurs aux estrogènes indiqué dans le cancer du sein métastatique hormonosensible chez la femme ménopausée.

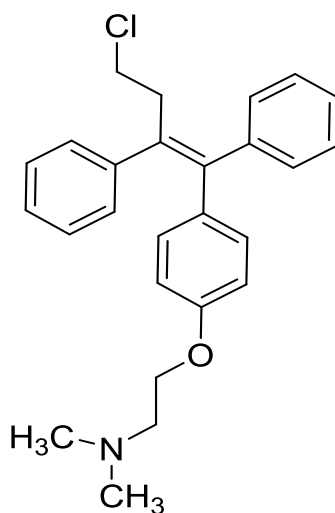


Figure 8. Structure de torémifène

Il est administré à raison de 60 mg par jour en une prise et est contre-indiqué en cas de troubles du rythme cardiaque, d'insuffisance cardiaque gauche ou d'allongement de l'espace QT (les associations médicamenteuses responsables de ce trouble sont déconseillées). Les principaux effets indésirables rencontrés avec le torémifène sont des nausées et vomissements, des bouffées de chaleur, de la fatigue, des oedèmes des membres, des étourdissements et des vertiges.⁴⁰

3.3 Fulvestrant

Le fulvestrant est un anti-estrogène de structure stéroïdienne qui est analogue structuralement à des estrogènes (**figure 9**). Il appartient à la famille des régulateurs de diminution sélective des récepteurs d'estrogènes (SERD).¹¹

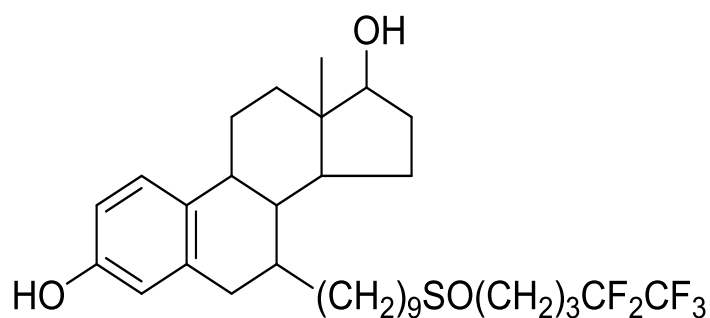


Figure 9. Structure du fulvestrant

Le fulvestrant est un antagoniste pur des récepteurs aux estrogènes qui présente également la capacité d'inhiber leur expression. Cette molécule est indiquée dans le traitement du cancer du sein localement avancé ou métastasé chez la femme ménopausée présentant des récepteurs aux estrogènes positifs, en cas de récurrence pendant ou après un traitement adjuvant par un anti-estrogène. Administré à raison de 250 mg par mois, en une injection intramusculaire (IM) unique, le fulvestrant est contre-indiqué en cas de grossesse et d'allaitement. Les principaux effets indésirables rencontrés sont des bouffées de chaleur, des nausées et vomissements, des accidents thromboemboliques, des éruptions cutanées et des réactions au point d'injection.⁴⁰

4 Conclusion

Le cancer du sein est une maladie hétérogène la plus fréquemment diagnostiquée et la cause majeure de décès chez les femmes à travers le monde. Plusieurs études montrent que les estrogènes augmentent le risque de développer un cancer du sein chez la femme pré et post-ménopausée. Elles favorisent également la croissance et la progression continues des tumeurs cancéreuses.

Les traitements hormonaux des cancers du sein visent à arrêter la prolifération cellulaire en bloquant l'activation des récepteurs aux estrogènes par les estrogènes endogènes, soit par inhibition compétitive tels que les anti-estrogènes (SERM ou SERD), soit en inhibant la biosynthèse de ces hormones par les inhibiteurs de l'aromatase. Les anti-estrogènes sont des médicaments qui entrent en compétition avec les estrogènes car ils prennent leur place et bloquent ainsi leurs effets de stimulation sur les cellules cancéreuses. Parmi les molécules anti-estrogènes, trois sont actuellement commercialisées : le tamoxifène, le torémifène et le fulvestrant.

C'est dans ce contexte que nous avons choisi d'étudier, dans notre travail de thèse, une série de dérivés de l'indazole qui sont des puissants antagonistes de type SERD. Notre but est d'identifier, par l'application de différentes méthodes de calcul, de nouveaux agents puissants qui peuvent être utiles pour désigner des médicaments anticancéreux.

Références bibliographiques

1. Vilquin, P.; Cohen, P.; Maudelonde, T.; Tredan, O.; Treilleux, I.; Bachelot, T.; Heudel, P.-E., Nouvelles stratégies thérapeutiques dans le cancer du sein hormono-dépendant métastatique. *Bulletin du cancer* **2015**, *102* (4), 367-380.
2. ElAmrani, A.; Khyatti, M.; Benhassou, M.; Attaleb, M.; Ennaji Moulay, M., Cancer du sein: Données Actuelles. *ScienceLib Editions Mersenne* **2013**, 5.
3. Rousselle, V. Place et intérêt de l'hormonothérapie dans les cancers du sein chez la femme ménopausée: guide d'information destiné aux pharmaciens d'officine. Thèse de doctorat, Université de Lorraine, **2013**.
4. Beatson, G., On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment, with illustrative cases. *The Lancet* **1896**, *148*, 162-165.
5. Frodeau, C. L'hormonothérapie dans le cancer du sein et l'observance des traitements. Quel est le rôle du pharmacien officinal. Thèse de doctorat, Université Joseph Fourier, **2012**.
6. Misrahi, M.; Atger, M.; d'Auriol, L.; Loosfelt, H.; Meriel, C.; Fridlansky, F.; Guiochon-Mantel, A.; Galibert, F.; Milgrom, E., Complete amino acid sequence of the human progesterone receptor deduced from cloned cDNA. *Biochemical and Biophysical Research Communications* **1987**, *143* (2), 740-748.
7. Rousseau-Merck, M.; Misrahi, M.; Loosfelt, H.; Milgrom, E.; Berger, R., Localization of the human progesterone receptor gene to chromosome 11q22-q23. *Human Genetics* **1987**, *77* (3), 280-282.
8. Kuiper, G. G.; Gustafsson, J.-Å., The novel estrogen receptor- β subtype: potential role in the cell-and promoter-specific actions of estrogens and anti-estrogens. *FEBS Letters* **1997**, *410* (1), 87-90.
9. Enmark, E.; Pelto-Huikko, M.; Grandien, K.; Lagercrantz, S.; Lagercrantz, J.; Fried, G.; Nordenskjöld, M.; Gustafsson, J.-A. k., Human estrogen receptor β -gene structure, chromosomal localization, and expression pattern. *The Journal of Clinical Endocrinology & Metabolism* **1997**, *82* (12), 4258-4265.
10. Menasce, L. P.; White, G. R.; Harrison, C. J.; Boyle, J. M., Localization of the estrogen receptor locus (ESR) to chromosome 6q25. 1 by FISH and a simple post-FISH banding technique. *Genomics;(United States)* **1993**, *17* (1), 263-265.

11. De Cremoux, P., Hormonothérapie des cancers du sein. *Bulletin du cancer* **2011**, 98 (11), 1311-1319.
12. Yahiaoui, S. Cancer du sein. Synthèse et évaluation de dérivés flavonoïdiques inhibiteurs de l'aromatase. Thèse de doctorat, Université de Limoges, **2005**.
13. Pasqualini, J. R.; Paris, J.; Sitruk-Ware, R.; Chetrite, G.; Botella, J., Progestins and breast cancer. *The Journal of Steroid Biochemistry and Molecular Biology* **1998**, 65 (1-6), 225-235.
14. Santen, R. J., Inhibition of aromatase: insights from recent studies. *Steroids* **2003**, 68 (7-8), 559-567.
15. Miller, W. R., Biological rationale for endocrine therapy in breast cancer. *Best Practice & Research Clinical Endocrinology & Metabolism* **2004**, 18 (1), 1-32.
16. Coumoul, X.; Barouki, R., Génotoxicité des métabolites des estrogènes et cancers. *Médecine/Sciences* **2002**, 18 (1), 86-90.
17. Enmark, E.; Gustafsson, J. A., Estrogen receptor β -a novel receptor opens up new possibilities for cancer diagnosis and treatment. *Endocrine-Related Cancer* **1998**, 5 (3), 213-222.
18. Cowley, S. M.; Hoare, S.; Mosselman, S.; Parker, M. G., Estrogen receptors α and β form heterodimers on DNA. *Journal of Biological Chemistry* **1997**, 272 (32), 19858-19862.
19. Jensen, E. V.; Jordan, V. C., The estrogen receptor: a model for molecular medicine. *Clinical Cancer Research* **2003**, 9 (6), 1980-1989.
20. Pettersson, K.; Grandien, K.; Kuiper, G. G.; Gustafsson, J.-A. k., Mouse estrogen receptor β forms estrogen response element-binding heterodimers with estrogen receptor α . *Molecular Endocrinology* **1997**, 11 (10), 1486-1496.
21. Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Öhman, L.; Greene, G. L.; Gustafsson, J.-Å.; Carlquist, M., Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, 389 (6652), 753-758.
22. McKenna, N. J.; Lanz, R. B.; O'Malley, B. W., Nuclear receptor coregulators: cellular and molecular biology. *Endocrine Reviews* **1999**, 20 (3), 321-344.
23. Norris, J. D.; Paige, L. A.; Christensen, D. J.; Chang, C.-Y.; Huacani, M. R.; Fan, D.; Hamilton, P. T.; Fowlkes, D. M.; McDonnell, D. P., Peptide antagonists of the human estrogen receptor. *Science* **1999**, 285 (5428), 744-746.

24. White, R.; Parker, M., Molecular mechanisms of steroid hormone action. *Endocrine-Related Cancer* **1998**, 5 (1), 1-14.
25. Pujol, P.; Maudelonde, T., SERM: concept et mécanismes d'action: Prévention de l'ostéoporose. *La Lettre du gynécologue* **2001**, (258), 29-36.
26. Lecomte, J. Effets des ligands de PPAR [gamma] sur la voie de signalisation des oestrogènes dans les cellules cancéreuses mammaires. Thèse de doctorat, Université Henri Poincaré-Nancy 1, **2009**.
27. Segnitz, B.; Gehring, U., Subunit structure of the nonactivated human estrogen receptor. *Proceedings of the National Academy of Sciences* **1995**, 92 (6), 2179-2183.
28. Johnson, J. L.; Toft, D. O., A novel chaperone complex for steroid receptors involving heat shock proteins, immunophilins, and p23. *Journal of Biological Chemistry* **1994**, 269 (40), 24989-24993.
29. Roe, S. M.; Prodromou, C.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H., Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. *Journal of medicinal chemistry* **1999**, 42 (2), 260-266.
30. Maloney, A.; Workman, P., HSP90 as a new therapeutic target for cancer therapy: the story unfolds. *Expert opinion on biological therapy* **2002**, 2 (1), 3-24.
31. Gougelet, A.; Bouclier, C.; Marsaud, V.; Maillard, S.; Mueller, S. O.; Korach, K. S.; Renoir, J.-M., Estrogen receptor α and β subtype expression and transactivation capacity are differentially affected by receptor-, hsp90-and immunophilin-ligands in human breast cancer cells. *The Journal of steroid biochemistry and molecular biology* **2005**, 94 (1-3), 71-81.
32. Klein-Hitpaß, L.; Schorpp, M.; Wagner, U.; Ryffel, G. U., An estrogen-responsive element derived from the 5' flanking region of the *Xenopus vitellogenin A2* gene functions in transfected human cells. *Cell* **1986**, 46 (7), 1053-1061.
33. Klinge, C. M., Estrogen receptor interaction with estrogen response elements. *Nucleic acids research* **2001**, 29 (14), 2905-2919.
34. Shang, Y.; Hu, X.; DiRenzo, J.; Lazar, M. A.; Brown, M., Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **2000**, 103 (6), 843-852.
35. Kushner, P. J.; Agard, D. A.; Greene, G. L.; Scanlan, T. S.; Shiau, A. K.; Uht, R. M.; Webb, P., Estrogen receptor pathways to AP-1. *The Journal of steroid biochemistry and molecular biology* **2000**, 74 (5), 311-317.

36. Saville, B.; Wormke, M.; Wang, F.; Nguyen, T.; Enmark, E.; Kuiper, G.; Gustafsson, J.-Å.; Safe, S., Ligand-, cell-, and estrogen receptor subtype (α/β)-dependent activation at GC-rich (Sp1) promoter elements. *Journal of Biological Chemistry* **2000**, *275* (8), 5379-5387.
37. Bjornstrom, L.; Sjoberg, M., Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Molecular Endocrinology* **2005**, *19* (4), 833-842.
38. Safi, R. m. Etude de la férutinine et de ses analogues: hémissynthèse et activité anticancéreuse vis-à-vis des lignées cellulaires hormono-dépendantes. Thèse de doctorat, Université Toulouse 3 Paul Sabatier, **2015**.
39. Delozier, T., Le tamoxifène en situation adjuvante dans le cancer du sein: trente ans d'expérience. 26^{es} journées de la SFSPM, Nancy: **2004**.
40. Clere, N., Les traitements du cancer du sein. *Actualités Pharmaceutiques* **2016**, *55* (558), 20-25.

Chapitre II

**Étude quantitative des relations
structure-activité d'une série de
dérivés de l'indazole.**

1 Introduction

Le développement d'outils informatiques fiables couplée à la croissance de la puissance informatique a permis la mise en place des techniques de modélisation moléculaire, qui devenues, actuellement des outils indispensables dans le domaine de la conception des médicaments. Elle a connu ces dernières décennies un essor très important dans de nombreux branches d'applications à savoir la structure électronique de l'atome, des molécules ou des complexes organométalliques, l'évaluation de leurs propriétés magnétiques et spectroscopiques ou encore la structuration de molécules d'intérêts biologiques. Il s'agit d'un ensemble des techniques basées sur des méthodes de chimie théorique et les données expérimentales permettant d'étudier et de traiter les problèmes chimiques sur un ordinateur sans avoir besoin d'aller dans la salle de manipulation pour monter des expériences. ¹

Durant les dernières années, la relation quantitative structure-activité (QSAR) est devenue un alternatif outil à la chimie théorique, pour la démonstration et la prédiction des propriétés et des activités de l'ensemble des molécules. ² Les relations quantitatives structure-activité, en tant que méthodes de calcul efficaces, peuvent aider à capturer les informations à partir des données expérimentales et à accélérer la découverte de médicaments. ³ L'analyse quantitative des relations existantes entre les structures d'un ensemble de composés et leurs activités permet d'identifier et d'évaluer l'impact des propriétés influençant l'activité biologique. Cette relation peut être décrite par une équation qui corrèle mathématiquement les influences réciproques des paramètres concernés. L'extrapolation de ces résultats peut servir de base à la prédiction de l'activité de nouveaux composés. ⁴ La QSAR est devenue un outil avantageux et précieux, notamment dans la recherche pharmaceutique industrielle, plus particulièrement dans le cas où la disponibilité des échantillons est limitée ou les mesures expérimentales sont dangereuses, longues et chères. Les prévisions fournies par QSAR permettent de proposer une bonne stratégie pour le développement d'un bon pharmacophore, elle peut être ainsi un support adéquat pour valider ou rejeter une hypothèse.

La dérivation de la relation directe avec la structure moléculaire n'est pas facile. Il est cependant possible d'identifier plusieurs facteurs structurels connus comme descripteurs moléculaires qui influent sur la propriété moléculaire choisie.⁵

Les informations extraites à partir des résultats d'étude de QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les propriétés physicochimiques et les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures.⁶

Dans cette partie de travail, nous avons réalisé une étude QSAR sur un ensemble de données contenant **54** molécules de dérivés de l'indazole. Notre objectif est de modéliser l'activité anticancéreuse et de former un modèle QSAR présentant de bons paramètres statistiques en utilisant la régression linéaire multiple (MLR).

2 Aperçue générale sur la relation quantitative structure-activité

Le principe des méthodes QSAR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative la structure moléculaire, codée par des propriétés moléculaires appelées descripteurs, avec une activité en utilisant des méthodes d'analyse de données.⁷ Par ces relations nous pouvons développer des modèles prédictifs de la forme générale suivante :

$$\text{Activité} = f(\text{descripteurs moléculaires})$$

L'objectif de ces méthodes est alors d'analyser les données structurales afin de déterminer les facteurs influençant l'activité étudiée.⁸

Le développement d'un modèle commence par la sélection de la base de données. En effet, sans valeurs expérimentales fiables et homogènes, il est difficile d'obtenir un modèle prédictif. L'homogénéité des données étant importante⁹, les données doivent être mesurées, autant que possible, par le même protocole dans les mêmes conditions expérimentales. La taille de la base de données est aussi importante : une base de données trop petite rend difficile l'obtention de modèles prédictifs et parfois impossible leur validation.¹⁰ Ensuite, il est nécessaire de générer un nombre de descripteurs, caractérisant les structures moléculaires, utilisés dans le développement du modèle QSAR.¹¹

Les données collectées sont divisées aléatoirement en deux parties :

- ❖ Les données d'apprentissage « Training Set » : utilisées pour la construction des modèles finaux.
- ❖ Les données de prédiction « Test Set » : utilisées pour la validation des modèles formés. Ces données ne subissent en revanche aucun traitement durant la construction des modèles et elles sont réservées uniquement pour tester le pouvoir prédictif des modèles.

Une fois le modèle construit, il doit être évalué par des méthodes de validation (validations interne et externe) afin d'estimer sa robustesse et son pouvoir prédictif.¹²⁻¹⁴

La **figure 1** résume la procédure suivie dans la construction du modèle QSAR.

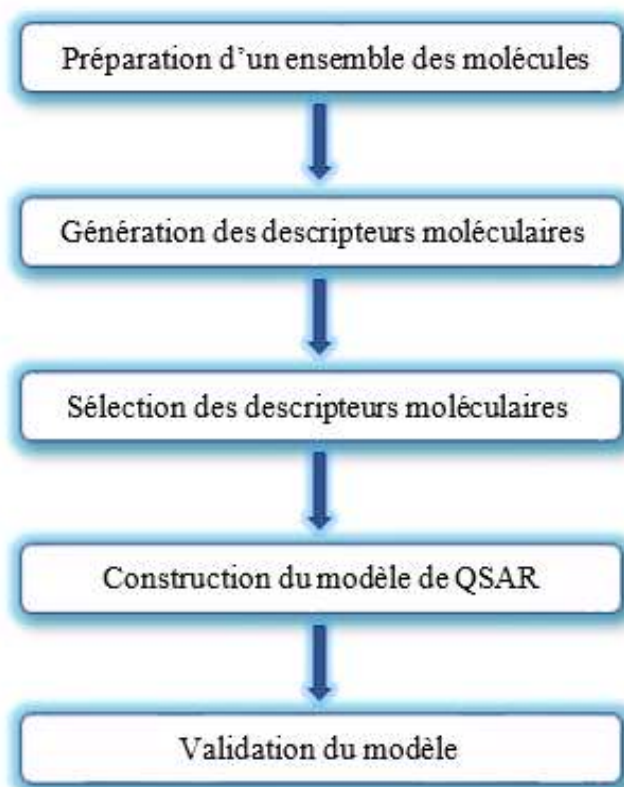


Figure 1. Procédure suivie dans la construction du modèle de QSAR

2.1 Préparation d'un ensemble des molécules

La préparation de l'ensemble de molécules se fait comme suit :

Au début, la représentation structurale de tous les composés, existant dans l'ensemble de données, sous forme de croquis bidimensionnel. Plusieurs logiciels peuvent être utilisés pour le dessin des structures moléculaires des composés organiques on cite à titre d'exemple : Hyperchem, ChemDraw, MarvinSketch...etc. ⁵ Ensuite, une structure tridimensionnelle (3D) est générée par une optimisation rapide à partir d'une méthode quantique qui permet d'identifier les conformations stables par rotation autour des différentes liaisons de la molécule et la géométrie de la conformation la plus stable obtenue pour chaque molécule soit retenue dans l'analyse. ^{10, 15} Enfin, la division de l'ensemble de données en des groupes d'apprentissage et des groupes test.

2.2 Génération des descripteurs moléculaires

Un descripteur moléculaire est un paramètre propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule. Les descripteurs calculés, permettent d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est l'un des objectifs de la modélisation moléculaire. Ils jouent un rôle fondamental dans les études quantitatives des relations structure-activité (QSAR). ²

Les descripteurs moléculaires sont généralement classés en trois catégories ; les descripteurs physicochimiques, topologiques et électroniques. Ces descripteurs sont caractéristiques de la structure bidimensionnelle ou tridimensionnelle de la molécule. ¹

2.3 Sélection des descripteurs moléculaires

Lorsqu'une grande quantité de descripteurs est introduite, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. De plus, les descripteurs considérés comme pertinents sont ceux ayant une grande corrélation avec la propriété et ayant une variance significative sans laquelle le descripteur ne permet pas la distinction des différentes données entre elles. Les descripteurs employés doivent être, autant que possible, porteurs de sens et facilement interprétables d'un point de vue chimique, aussi les modèles QSAR devraient être simples, transparents et compréhensibles d'un point de vue phénoménologique. Il est nécessaire donc d'éliminer

les descripteurs dont l'influence est inférieure à celle de l'erreur, et de sélectionner uniquement les plus pertinents d'entre eux. De manière générale, pour qu'un descripteur soit retenu, il faut que son retrait entraîne une décroissance significative de performance du modèle. Il faut donc être attentif à ne pas perdre de l'information essentielle.

Finalement, le sens chimique des descripteurs utilisés doit être pris en considération puisque, plus les descripteurs sont reliés chimiquement au phénomène, plus la probabilité de faire intervenir des descripteurs par hasard est réduite.¹⁶

La procédure de sélection et de réduction des descripteurs peut être effectuée en deux étapes :

- ❖ Sélection objective.
- ❖ Sélection subjective.

2.3.1 Sélection objective

Elle consiste en la sélection des variables en réduisant le nombre de descripteurs sans faire participer la variable dépendante (l'activité biologique). La première étape de cette procédure consiste à exclure tous les descripteurs ayant un pourcentage élevé de valeurs identiques pour l'ensemble des composés (variance non significative). Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final. De même, lorsque deux descripteurs sont fortement corrélés et leur combinaison possède un coefficient de détermination supérieur au seuil requis ($R^2 > 0,90$), seul celui présentant la plus grande variance est retenu. Non seulement ces procédures évitent l'introduction, dans le modèle, de descripteurs inappropriés mais elles rendent la suite de l'analyse moins coûteuse en terme de temps de calcul, puisqu'elles réduisent le nombre de descripteurs restant à traiter.¹⁶

2.3.2 Sélection subjective

2.3.2.1 Introduction progressive

Cette méthode consiste à incorporer, une à une, les variables au modèle en sélectionnant à chaque étape la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée.

2.3.2.2 Elimination progressive

Cette méthode consiste en l'établissement du modèle avec l'ensemble des descripteurs pour ensuite ne garder que ceux qui permettent l'obtention d'un modèle ayant une bonne corrélation. ¹

2.3.2.3 Sélection pas à pas

C'est la combinaison des deux méthodes citées précédemment. Les variables sont incorporées une à une dans le modèle par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives. ¹

2.4 Construction du modèle de QSAR

Pour construire un modèle QSAR nous avons besoin d'une méthode d'analyse de données, cette méthode permet de quantifier la relation qui existe entre l'activité et la Structure.

Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, certaines sont linéaires telles que la régression linéaire multiple (MLR), la régression aux moindres carrés partiels (PLS), d'autres sont non linéaires comme la régression non-linéaire multiple (MNLR), les réseaux de neurones artificiels (RNA)... Ces méthodes sont disponibles dans des logiciels tels que, Excel, Systat, Exstat, Minitab, Statistica, SPSS, R,... ¹ La méthode utilisée dans notre étude est la Régression Linéaire Multiple (MLR). Cette dernière est la plus simple méthode statistique de modélisation et la plus employée pour le développement de modèles prédictifs. ¹⁷ Elle consiste à rechercher une équation linéaire par rapport à ses paramètres, reliant la variable dépendante Y et une série de p variables indépendantes Xi. ¹⁸

L'équation de la régression MLR s'écrit :

$$Y_i = \alpha_0 + \alpha_1 X_{i,1} + \dots + \alpha_p X_{i,p} + b_i \quad (1)$$

Où Y_i est la i -ème observation de la variable Y ; $X_{i,j}$ est la i -ème observation de la j -ème variable ; b_i est l'erreur du modèle, elle résume les informations manquantes qui permettrait d'expliquer linéairement les valeurs de Y à l'aide des p variables X_j .

Au final, la méthode MLR résout l'équation :

$$Y = X.A + B \quad (2)$$

Où Y , X , A et B représentent respectivement le vecteur de propriétés, la matrice des attributs (descripteurs), la matrice des coefficients et la matrice des erreurs de régression.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{31} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad A = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ a_n \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_n \end{pmatrix}$$

La méthode MLR consiste alors à choisir les coefficients a_i de sorte à minimiser la somme des carrés des écarts entre les valeurs calculées de la propriété et celles expérimentales, l'équation du modèle devient donc sous forme matricielle :

$$X.A = Y \quad (3)$$

Connaissant X et Y , les coefficients peuvent être obtenus à partir l'équation matricielle suivante :

$$A = (X^T \cdot X)^{-1} X^T \cdot Y \quad (4)$$

2.4.1 Critères de choix de modèle (Cp de Mallows et le BIC)

De nombreux critères de choix de modèle sont présentés dans la littérature sur la régression linéaire multiple. Citons dans ce travail le critère de Cp de Mallows et le critère d'information de Schwartz (BIC)¹⁹. Ils sont équivalents lorsque le nombre de variables à

sélectionner, ou niveau du modèle, est fixé. Le choix du critère est déterminant lorsqu'il s'agit de comparer des modèles de niveaux différents.

2.4.1.1 Cp de Mallows

Cet indicateur est une estimation de l'erreur quadratique moyenne de prévision. En supposant que les estimations du modèle complet sont sans biais, l'expression de l'erreur quadratique moyenne totale standardisée (ou réduite) pour un modèle à j variables explicatives s'écrit :

$$C_p = (n - q - 1) \frac{MSE_j}{MSE} - [n - 2(q + 1)] \quad (5)$$

La valeur du C_p de Mallows est définie pour les q variables considérées. MSE et MSE_j désignant respectivement l'erreur quadratique moyenne du modèle complet et celle d'un modèle à j variables explicatives. Il est alors d'usage de rechercher un modèle qui minimise le C_p tout en fournissant une valeur inférieure et proche de $(q + 1)$. Ceci revient à considérer que le « vrai » modèle complet est moins fiable qu'un modèle réduit donc biaisé mais d'estimation plus précise.¹⁹

2.4.1.2 BIC

Le critère Bayesian information criterion (BIC) cherche, approximativement (asymptotiquement), le modèle associé à la plus grande probabilité a posteriori. Dans le cas d'un modèle issu de la maximisation d'une log-vraisemblance \mathcal{L} , il se met sous la forme :

$$BIC = -2\mathcal{L} + \log(n) \frac{d}{n} \quad (6)$$

Où d est le nombre de paramètres du modèle (nombre de variables plus un) et n le nombre d'observations. BIC tend à pénaliser plus lourdement les modèles complexes. Ce n'est pas le cas du C_p qui tend à choisir des modèles trop complexes. Néanmoins à taille finie, petite, BIC risque de se limiter à des modèles trop simples. Par conséquent, il est facile de choisir le modèle présentant le plus faible BIC parmi ceux considérés.

2.4.2 Algorithmes de sélection

Lorsque p est grand, il n'est pas raisonnable de penser explorer les 2^p modèles possibles afin de sélectionner le "meilleur" au sens de l'un des critères de choix de modèle.

Trois types d'algorithmes sont résumés ci-dessous ¹⁹ par ordre croissant de temps de calcul nécessaire c'est-à-dire par nombre croissant de modèles considérés parmi les 2^p et donc par capacité croissante d'optimalité.

Pas à pas

- ❖ **Sélection** (forward) à chaque pas, une variable est ajoutée au modèle. C'est celle dont la valeur p ("prob-value"), associée à la statistique partielle du test de Fisher qui compare les deux modèles, est minimum. La procédure s'arrête lorsque toutes les variables sont introduites ou lorsque p reste plus grande qu'une valeur seuil fixée par défaut à 0, 50.
- ❖ **Élimination** (backward) l'algorithme démarre cette fois du modèle complet. À chaque étape, la variable associée à la plus grande valeur p est éliminée du modèle. La procédure s'arrête lorsque les variables restant dans le modèle ont des valeurs p plus petites qu'un seuil fixé par défaut à 0, 10.
- ❖ **Mixte** (stepwise) cet algorithme introduit une étape d'élimination de variable après chaque étape de sélection afin de retirer du modèle d'éventuels variables qui seraient devenues moins indispensables du fait de la présence de celles nouvellement introduites.

Global

L'algorithme de Furnival et Wilson est utilisé pour comparer tous les modèles possibles en cherchant à optimiser l'un des critères : R^2 , R^2 ajusté, ou C_p de Mallows choisi par l'utilisateur. Par souci d'économie, cet algorithme évite de considérer des modèles de certaines sous-branches de l'arborescence dont on peut savoir a priori qu'ils ne sont pas compétitifs. En général les logiciels exécutant cet algorithme affichent le (best=1) ou les meilleurs modèles de chaque niveau.

2.5 Validation du modèle

La validation des modèles est une partie sensible, intégrante et importante du développement d'un modèle. Elle est nécessaire pour estimer la fiabilité d'un modèle, de déterminer la reproductibilité des résultats obtenus et la pertinence d'un modèle développé pour une application donnée.²⁰ Il existe plusieurs types de validation qui se complètent : la validation dite interne et la validation externe.

2.5.1 Validation interne

La validation interne ou validation croisée est la plus utilisée dans la plupart du temps dans les études QSAR afin de mesurer la robustesse des modèles formés. Il existe plusieurs méthodes de validation croisée : Leave-One -Out (Q^2_{LOO}), Leave-Many-Out (Q^2_{LMO}) et le bootstrap (Q^2_{boot}).²¹

Dans le cas du Leave-One -Out (LOO), une seule molécule de l'ensemble d'apprentissage (Training Set) est retirée et les coefficients de la régression sont optimisés sur les N-1 autres données. La propriété Y_{pred} est recalculée à partir de l'équation du modèle formé pour la molécule isolée. Cette manipulation est réalisée pour les N molécules de l'ensemble d'apprentissage, puis le coefficient de corrélation de validation croisée noté Q^2 ^{10, 14} est calculé avec l'équation suivante :

$$Q^2_{\text{LOO}} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (7)$$

Où \hat{Y}_i est la valeur prédite en excluant l'i-ème élément dans le développement du modèle. La valeur seuil de Q^2 est de 0,5.

Dans le cas du Leave-Many-Out (LMO), un groupe de molécules de l'ensemble d'apprentissage est retiré au lieu d'une seule molécule. Une faible Q^2 implique que le modèle n'est pas robuste et ne sera pas prédictif, mais la réciproque n'est pas nécessairement vraie.¹³

Le bootstrap²²⁻²⁴ ou la technique de bootstrap Q^2_{boot} est une autre méthode de validation qui a été utilisée dans cette thèse. Il s'agit d'une technique de rééchantillonnage avec « remise » d'un grand nombre d'itérations (de 50 à 2000) qui permet d'évaluer l'intervalle de confiance et des estimateurs statistiques tels que la

variance. Comme dans la validation de LMO, une moyenne élevée du Q^2 dans la validation du bootstrap est une démonstration de la robustesse du modèle.¹⁴

2.5.2 Validation externe

Cette méthode consiste à prédire l'activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle¹, cette validation est caractérisée par le paramètre R^2_{pred} qui présente l'indice de corrélation et qui décrit le pouvoir de prévision externe.^{14, 25} R^2_{pred} a été calculé selon la formule ci-dessous :

$$R^2_{\text{pred}} = 1 - \frac{\sum(\hat{Y}_{\text{test}} - Y_{\text{test}})^2}{\sum(Y_{\text{test}} - \bar{Y}_{\text{tr}})^2} \quad (8)$$

Avec Y_{test} et \hat{Y}_{test} les valeurs d'activités observées et prédites pour l'ensemble de test et \bar{Y}_{tr} la valeur moyenne de l'activité des composés de l'ensemble d'apprentissage.

2.5.3 Coefficient de détermination

Le coefficient de détermination R^2 est un indicateur de qualité. Bien évidemment ($0 \leq R^2 \leq 1$), plus il tend vers 1 meilleur sera le modèle. Il est calculé selon la formule

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2} \quad (9)$$

Avec \bar{Y}_i est la moyenne des valeurs observées pour l'ensemble du calibrage et \hat{Y}_i et la valeur prédite.

2.5.4 Test de Fisher

Le test F de Fisher permet de savoir si un modèle de régression est globalement significatif ou non. Il mesure le rapport entre la variance de la variable dépendante expliquée et non expliquée par le modèle de régression. F est calculé par la formule suivante :

$$F = \frac{\sum(\hat{Y}_i - \bar{Y}_i)^2}{p} \frac{n - p - 1}{\sum(Y_i - \hat{Y}_i)^2} \quad (10)$$

L'équation (10) représente le nombre de degrés de liberté associé avec chaque paramètre : $\sum(\hat{Y}_i - \bar{Y}_i)^2$ associé avec p degrés de liberté et $\sum(Y_i - \hat{Y}_i)^2$ associé avec n-p-1 degrés de liberté.

Un modèle sera significatif si F est plus grand à 95 % que le seuil F critique (F_{crit}) du tableau, on rejette l'hypothèse H_0 et dans ce cas on peut dire que d'un point de vue statistique les variances sont significativement différentes.

La p-value est le niveau de significativité du test de Fisher-Snedecor c'est-à-dire la probabilité de dépasser le F observé si l'hypothèse nulle est vraie. On compare la p-value au risque α choisi (par exemple $\alpha = 0.05$). Si p-value $\leq \alpha$, on rejette l'hypothèse nulle. ²⁶

2.6 Critères de validation

Tropsha ¹⁴ a proposé des critères de validation afin d'accéder à la prédictivité du modèle en mesurant les coefficients de détermination lorsque la ligne de régression passe par zéro, R_0^2 (valeurs prédites vs valeurs expérimentales) et $R_0'^2$ (valeurs expérimentales vs valeurs prédites) ainsi que les pentes k et k' de ces lignes de régressions :

- ❖ $Q^2 > 0.5$
- ❖ $R^2 > 0.6$
- ❖ $\frac{R^2 - R_0^2}{R^2} < 0.1$ ou $\frac{R^2 - R_0'^2}{R^2} < 0.1$
- ❖ $0.85 < k$ ou $k' < 1.15$

De manière générale, les coefficients R^2 et Q^2 doivent avoir des valeurs proches de 1 (de préférence supérieures à 0,6) et leur différence doit être faible pour considérer le modèle comme robuste. Cependant, l'évaluation des coefficients doit se faire au regard de la taille de la base de données (notamment pour R^2) et de l'ordre de grandeur de l'incertitude expérimentale. Mais d'autres paramètres sont pris en considération pour le choix du modèle comme la possibilité d'interprétation des descripteurs. ¹⁰

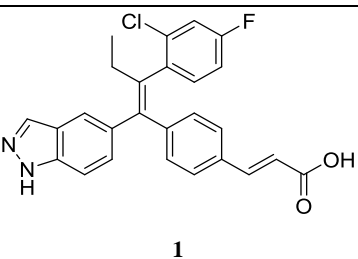
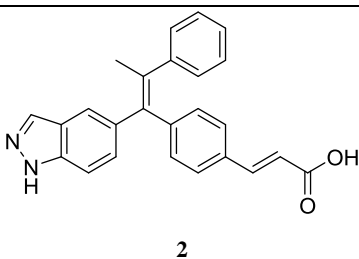
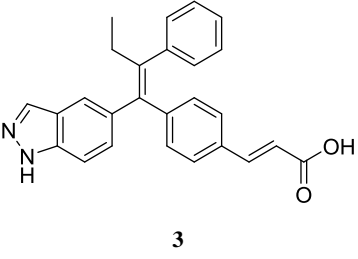
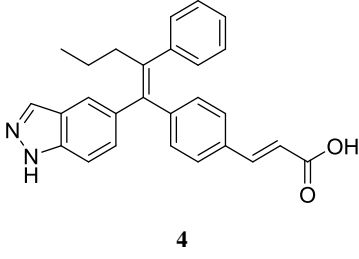
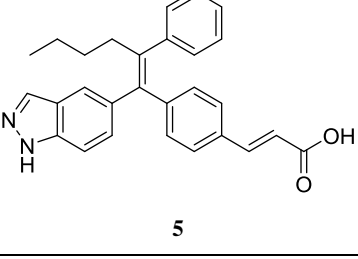
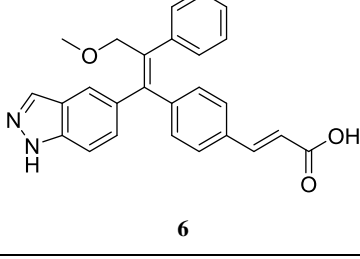
3 Matériels et Méthodes

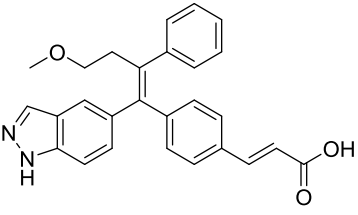
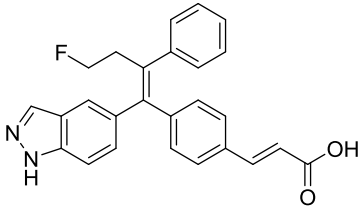
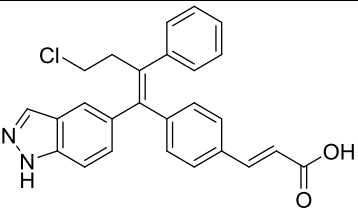
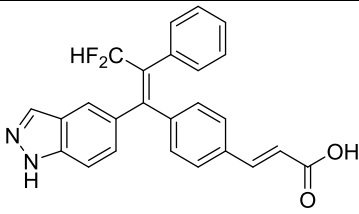
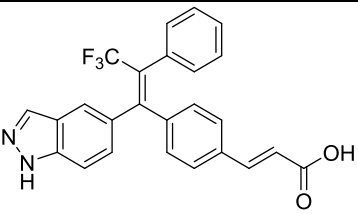
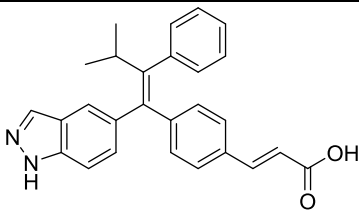
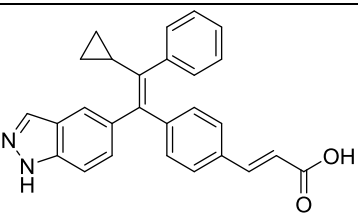
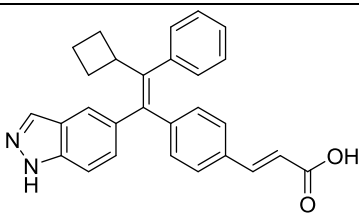
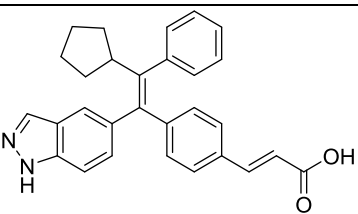
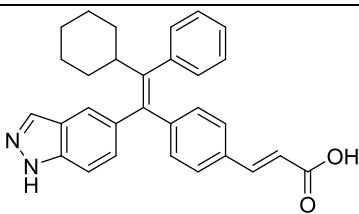
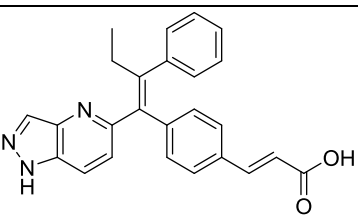
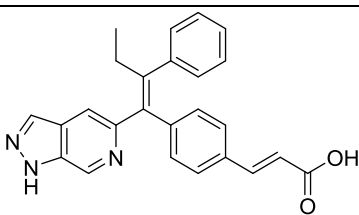
3.1 Préparation d'un ensemble des molécules de dérivés de l'indazole

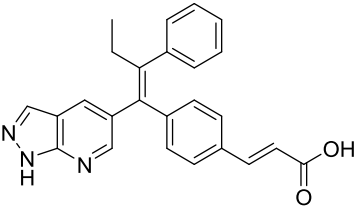
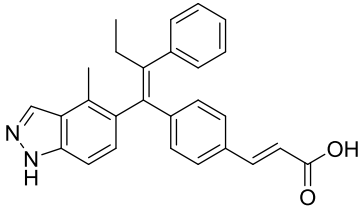
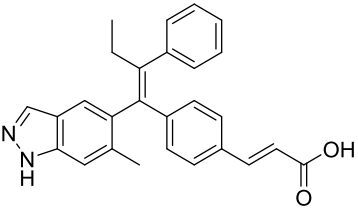
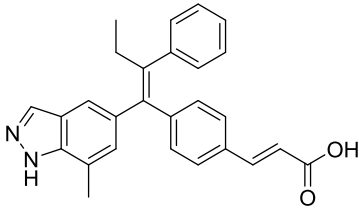
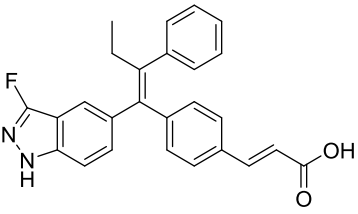
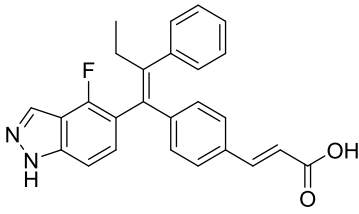
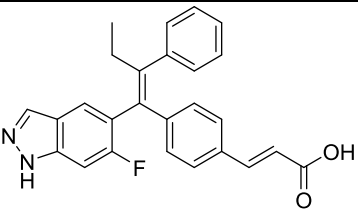
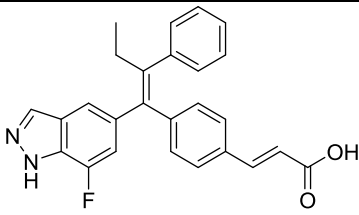
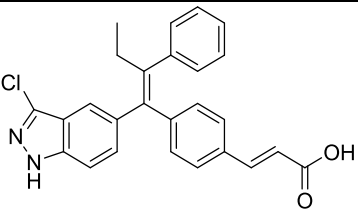
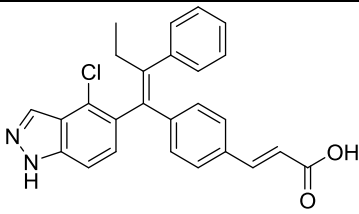
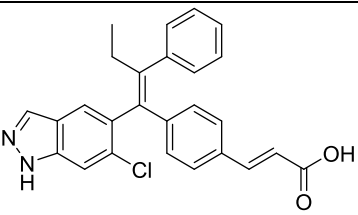
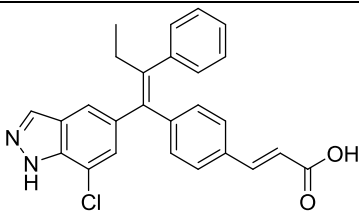
Dans cette étude, l'ensemble à modéliser est constitué de 54 molécules dérivées de l'indazole, dont les activités pour le traitement de la tumeur RE+ du cancer du sein sont rapportées dans la littérature par Andiliy Lai et al ²⁷.

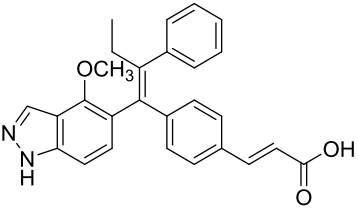
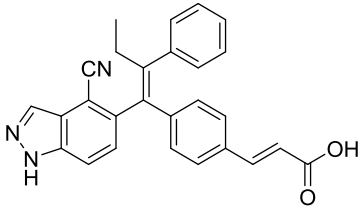
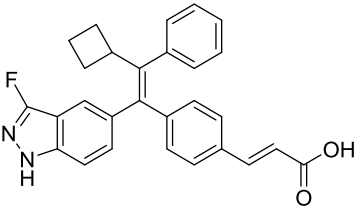
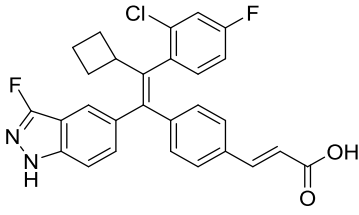
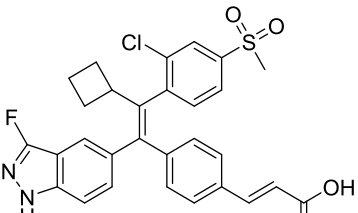
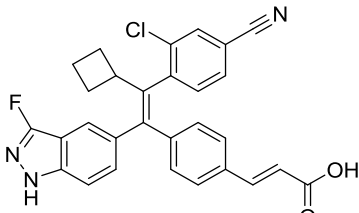
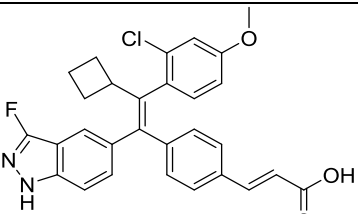
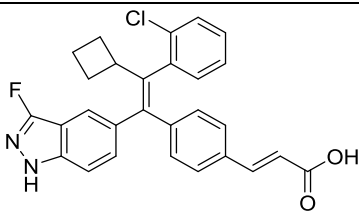
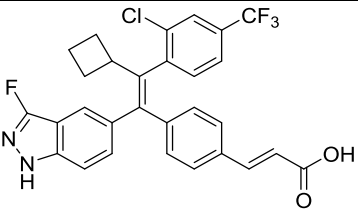
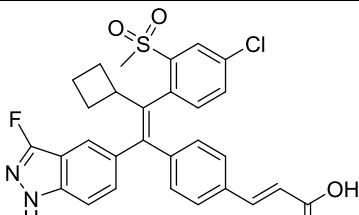
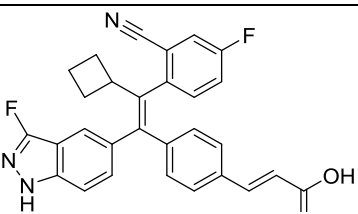
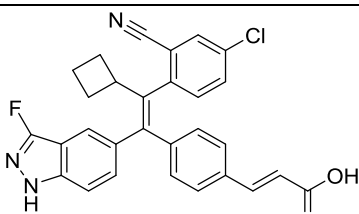
Les activités biologiques sont exprimées par le rapport logarithmique pIC_{50} ou $\text{Log}(1/IC_{50})$ où les valeurs de IC_{50} en M présente la concentration molaire en inhibiteur, nécessaire pour inhiber 50 % de l'activité enzymatique (croissance stimulée par l'estrogène). Les structures moléculaires et leurs activités anticancéreuses sont illustrées dans le **tableau 1**.

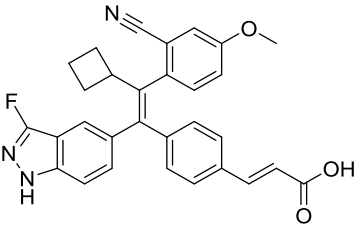
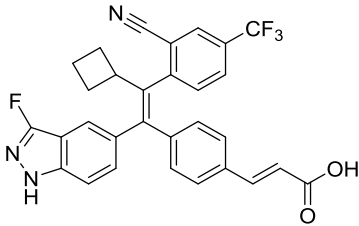
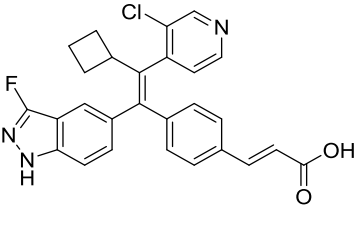
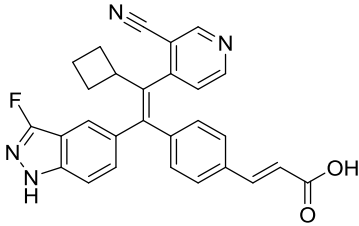
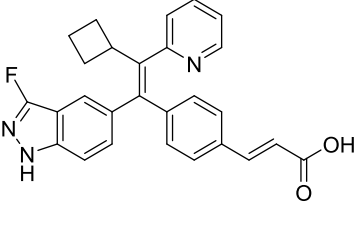
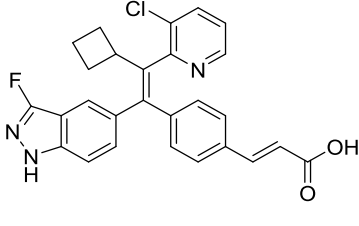
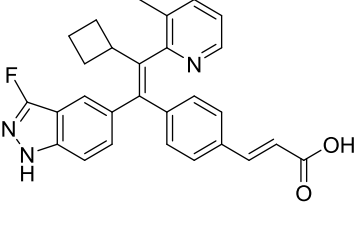
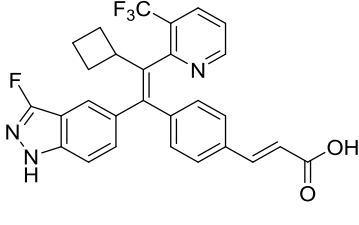
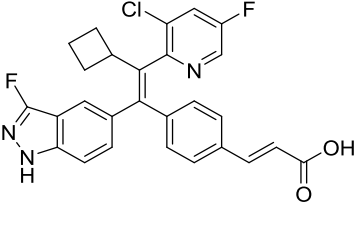
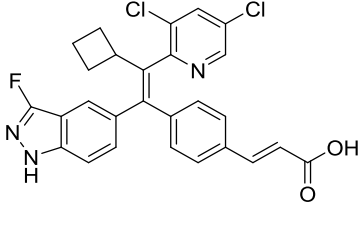
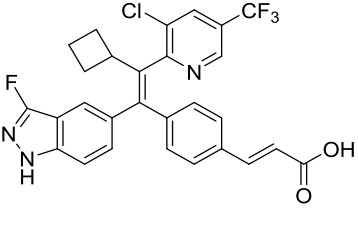
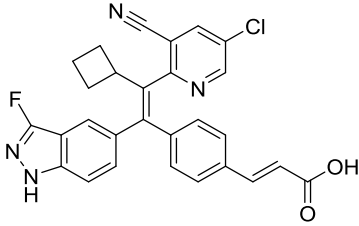
Tableau 1: Structures chimiques développées et les activités biologiques expérimentales des molécules utilisées dans l'étude QSAR.

Composé	pIC_{50}	Composé	pIC_{50}
 1	-0.3979	 2	-1.6021
 3	-0.8573	 4	-1.0414
 5	-1.0000	 6	-1.9445

 7	-1.7993	 8	-0.3222
 9	-0.9542	 10	-1.1461
 11	-0.7404	 12	-0.6021
 13	-1.1139	 14	-0.1461
 15	-1.0000	 16	-1.8062
 17	-1.5185	 18	-2.6454

 <p style="text-align: center;">19</p>	-2.4800	 <p style="text-align: center;">20</p>	-1.6990
 <p style="text-align: center;">21</p>	-2.6703	 <p style="text-align: center;">22</p>	-1.7709
 <p style="text-align: center;">23</p>	-0.6990	 <p style="text-align: center;">24</p>	-0.4771
 <p style="text-align: center;">25</p>	-1.6721	 <p style="text-align: center;">26</p>	-0.8751
 <p style="text-align: center;">27</p>	-2.4533	 <p style="text-align: center;">28</p>	-0.2041
 <p style="text-align: center;">29</p>	-2.1367	 <p style="text-align: center;">30</p>	-1.7482

 <p>31</p>	-1.7404	 <p>32</p>	-0.1761
 <p>33</p>	-0.1461	 <p>34</p>	-0.3010
 <p>35</p>	-0.3617	 <p>36</p>	0.1549
 <p>37</p>	-0.3802	 <p>38</p>	-0.0414
 <p>39</p>	-0.9138	 <p>40</p>	0.5229
 <p>41</p>	-0.0792	 <p>42</p>	0.1549

 <p style="text-align: center;">43</p>	0.3010	 <p style="text-align: center;">44</p>	-0.0792
 <p style="text-align: center;">45</p>	0.0969	 <p style="text-align: center;">46</p>	-0.5798
 <p style="text-align: center;">47</p>	-1.3424	 <p style="text-align: center;">48</p>	-0.0792
 <p style="text-align: center;">49</p>	-0.6233	 <p style="text-align: center;">50</p>	-0.0414
 <p style="text-align: center;">51</p>	-0.2305	 <p style="text-align: center;">52</p>	-0.5441
 <p style="text-align: center;">53</p>	-0.6435	 <p style="text-align: center;">54</p>	-0.0414

Pour vérifier la capacité prédictive du modèle QSAR, l'ensemble des molécules a été divisé aléatoirement en des sous-ensembles d'apprentissages et des sous-ensembles test pour 2000 fois. Un total de **43** molécules (environ 79,63%) a été placé dans chaque sous-ensemble d'apprentissage (Training Set) utilisé dans la construction des modèles QSAR, tandis que les **11** molécules restantes (environ 20,37%) ont composé l'ensemble de test (Test Set) utilisé pour la validation externe et évalué le pouvoir prédictif des modèles formés.

3.2 Génération des descripteurs moléculaires de dérivés de l'indazole

Les descripteurs sont des propriétés numériques qui encodent les caractéristiques physicochimiques des molécules à partir de leurs représentations structurales. La génération des descripteurs constitue une étape importante et cruciale lors de l'analyse de QSAR car leur choix conditionne la qualité du modèle.

Avant toute modélisation, il est nécessaire de calculer ou de mesurer un grand nombre de descripteurs différents, afin d'obtenir un modèle QSAR significatif. Il faut ensuite sélectionner parmi ces variables celles qui sont les plus pertinentes pour la modélisation.

Tout d'abord, les **54** molécules de dérivées de l'indazole ont été dessinées et pré-optimisées au moyen de la mécanique moléculaire, champ de force (MM+) en utilisant le logiciel HyperChem 8.0.8²⁸. Ensuite, les structures obtenues ont été optimisée en utilisant la méthode semi-empirique AM1²⁹ par le même logiciel.

Un ensemble de **19** descripteurs moléculaires ont été calculé à l'aide de trois logiciels comme suit :

Le module « propriétés QSAR » de l'HyperChem 8.0.8 a été utilisé pour calculer : le volume moléculaire (**MV**), la surface moléculaire (**S**), l'énergie d'hydratation (**HE**), le coefficient de partage octanol/eau (**log P**), et le poids moléculaire (**MW**).

La théorie de la fonctionnelle de densité (DFT) utilisant la fonction B3LYP couplée avec un ensemble de base 6-311G³⁰⁻³¹ implémentée par le logiciel Gaussian 09³²⁻³³ a été utilisé pour calculer les descripteurs électroniques tels que les énergies des orbitales frontières (**E_{HOMO}**, **E_{LUMO}**), le moment dipolaire (**DM**), la différence entre les énergies des orbitales

HOMO et LUMO (le gap énergétique ΔE), la dureté chimique (η), le potentiel chimique (μ), l'énergie totale (E_t), et l'indice d'électrophilicité ou l'électrophilicité globale (ω).

Le logiciel ChemSketch ³⁴ a été utilisé pour calculer les paramètres suivants : La réfractivité moléculaire (**MR**), le parachor (**Pc**), la densité, (**D**), l'indice de réfraction (**R ind**), la surface de tension (**surf ten**) et la polarisabilité (**polar**).

3.3 Sélection des descripteurs moléculaires de dérivés de l'indazole

La sélection des meilleurs descripteurs, hautement significatifs et indicatifs de l'activité biologique concernée, est l'une des étapes les plus inévitables et cruciales dans la modélisation QSAR. L'avantage tiré de cette sélection dans l'étude QSAR n'est pas seulement la stabilité du modèle, mais aussi la facilité d'interprétation de la relation entre l'activité biologique étudiée et les descripteurs calculés. ³⁵ De plus, les descripteurs calculés n'ont pas nécessairement une influence sur l'activité à modéliser.

Les descripteurs ont été sélectionnés comme suit :

- ❖ La matrice de corrélation de Pearson a été effectuée sur tous les descripteurs ; car elle fournit des informations sur la corrélation positive ou négative entre les variables. En utilisant un test de significativité de corrélation, les p-values ont été calculées pour tester l'hypothèse nulle par rapport à l'alternative qu'il y avait une corrélation non nulle.
- ❖ Les descripteurs très faiblement corrélés à la réponse (activité) avec un $r < 0,1$ ont été éliminés et pour diminuer la redondance existante dans la matrice de données, les descripteurs fortement corrélés ($r > 0,94$) ont été exclus. Ensuite les p-values ont été classées par ordre croissant et 12 descripteurs moléculaires ont été sélectionnés (**tableau 2**).

Afin d'obtenir le meilleur modèle en termes de capacité prédictive avec le meilleur sous-ensemble de descripteurs, nous avons utilisé, par le logiciel Matlab ³⁶⁻³⁷, un algorithme de recherche par la régression pas à pas (Stepwise regression). Toutes les combinaisons possibles de descripteurs de $x_j = x_1, \dots, x_p$ ont été calculées et la combinaison la plus excellente a été effectuée avec la valeur la plus faible du paramètre statistique du C_p de Mallows ou le critère d'information de Schwartz (BIC).

Tableau 2: Descripteurs moléculaires sélectionnés.

Composé	E_{HOMO} (a.u)	E_{LUMO} (a.u)	HE (Kcal/mol)	MW (amu)	ΔE (a.u)	η	μ	E_t (a.u)	Pc (cm ³)	D (g/cm ³)	Surf ten (dyne/cm)	ω
1	-0.2321	-0.0871	-13.1	446.91	0.1450	0.0725	-0.1596	-1822.747	921.7	1.342	60.9	0.1735
2	-0.2154	-0.0852	-13.86	380.45	0.1302	0.0651	-0.1503	-1224.591	838.9	1.267	59.2	0.1725
3	-0.2156	-0.0848	-13.49	394.47	0.1308	0.0654	-0.1502	-1263.902	878.7	1.245	57.7	0.1725
4	-0.2153	-0.0848	-13.05	408.5	0.1305	0.0653	-0.1501	-1303.214	918.5	1.225	56.3	0.1725
5	-0.2153	-0.0848	-12.62	422.53	0.1305	0.0653	-0.1501	-1342.526	958.3	1.208	59.8	0.1740
6	-0.2169	-0.0855	-14.33	410.47	0.1314	0.0657	-0.1512	-1339.096	898.4	1.27	58.2	0.1725
7	-0.2166	-0.0849	-15.09	424.5	0.1317	0.0659	-0.1508	-1378.41	938.2	1.249	65.9	0.1723
8	-0.2219	-0.0883	-13.66	412.46	0.1336	0.0668	-0.1551	-1363.147	888.2	1.280	57.7	0.1801
9	-0.2237	-0.0894	-13.73	428.92	0.1343	0.0672	-0.1566	-1723.511	917.2	1.307	56.8	0.1822
10	-0.2293	-0.0897	-14.01	416.43	0.1396	0.0698	-0.1595	-1423.084	855.3	1.336	55.1	0.1926
11	-0.2355	-0.0943	-14.13	434.42	0.1412	0.0706	-0.1649	-1522.338	862.6	1.372	56.8	0.1706
12	-0.2198	-0.0843	-13.13	408.5	0.1355	0.0678	-0.1521	-1303.209	915.9	1.224	67.2	0.1720
13	-0.2139	-0.0845	-13.46	406.48	0.1294	0.0647	-0.1492	-1301.965	886.3	1.313	64.2	0.1713
14	-0.2197	-0.0846	-13.19	420.51	0.1351	0.0676	-0.1522	-1341.279	926.3	1.284	61.5	0.1707
15	-0.2195	-0.0843	-12.92	434.54	0.1352	0.0676	-0.1519	-1380.619	966.4	1.259	59.2	0.1708
16	-0.2185	-0.0843	-12.62	448.56	0.1342	0.0671	-0.1514	-1419.929	1006.4	1.236	62.9	0.1732
17	-0.2221	-0.0855	-13.76	395.46	0.1366	0.0683	-0.1538	-1279.925	872.9	1.275	60.1	0.1599
18	-0.2221	-0.0797	-13.9	395.46	0.1424	0.0712	-0.1509	-1279.926	910	1.252	62.9	0.1772
19	-0.2241	-0.0873	-14.54	395.46	0.1368	0.0684	-0.1557	-1279.93	872.9	1.275	57.3	0.1717
20	-0.2172	-0.0848	-12.67	408.5	0.1324	0.0662	-0.1510	-1303.216	916.4	1.226	57.3	0.1733
21	-0.2179	-0.0853	-12.76	408.5	0.1326	0.0663	-0.1516	-1303.215	916.4	1.226	57.3	0.1719
22	-0.2133	-0.0844	-12.03	408.5	0.1289	0.0645	-0.1489	-1303.219	916.4	1.226	58	0.1772
23	-0.2208	-0.0871	-12.85	412.46	0.1337	0.0669	-0.1540	1363.142	885.9	1.285	58	0.1734
24	-0.2223	-0.0856	-13.21	412.46	0.1367	0.0684	-0.1540	-1363.148	885.9	1.285	58	0.1668
25	-0.2247	-0.0829	-13.3	412.46	0.1418	0.0709	-0.1538	-1363.149	885.9	1.285	58	0.1764
26	-0.2216	-0.0868	-12.93	412.46	0.1348	0.0674	-0.1542	-1363.146	885.9	1.285	59.9	0.1772
27	-0.2213	-0.0871	-12.66	428.92	0.1342	0.0671	-0.1542	-1723.501	914.6	1.304	59.9	0.1740

28	-0.2245	-0.0860	-13.16	428.92	0.1385	0.0693	-0.1553	-1723.5	914.6	1.304	59.9	0.1690
29	-0.2252	-0.0839	-13.24	428.92	0.1413	0.0707	-0.1546	-1723.503	914.6	1.304	59.9	0.1773
30	-0.2232	-0.0873	-12.77	428.92	0.1359	0.0680	-0.1553	-1723.506	914.6	1.304	56.8	0.1633
31	-0.2091	-0.0806	-12.97	424.5	0.1285	0.0643	-0.1449	-1378.411	935.4	1.245	71.9	0.1794
32	-0.2288	-0.0885	-15.5	419.48	0.1403	0.0702	-0.1587	-1356.134	926.3	1.31	62.9	0.1766
33	-0.2251	-0.0871	-12.56	438.5	0.1380	0.0690	-0.1561	-1440.519	933.4	1.322	62.2	0.1828
34	-0.2335	-0.0902	-12.19	490.94	0.1433	0.0717	-0.1619	-1999.367	976.4	1.412	64.7	0.1905
35	-0.2387	-0.0937	-14.85	551.03	0.1450	0.0725	-0.1662	-2487.88	1092.7	1.43	74.2	0.1906
36	-0.2380	-0.0937	-16.75	497.96	0.1443	0.0722	-0.1659	-1992.353	1022.4	1.42	60.8	0.1751
37	-0.2257	-0.0865	-14.17	502.97	0.1392	0.0696	-0.1561	-2014.637	1026	1.368	63.4	0.1769
38	-0.2295	-0.0875	-12.41	472.95	0.1420	0.0710	-0.1585	-1900.123	969.3	1.377	55	0.1889
39	-0.2379	-0.0930	-12	540.94	0.1449	0.0725	-0.1655	-2237.188	1026.5	1.435	64.7	0.2051
40	-0.2445	-0.0999	-12.33	551.03	0.1446	0.0723	-0.1722	-2487.838	1092.7	1.43	71	0.1827
41	-0.2361	-0.0903	-14.79	481.5	0.1458	0.0729	-0.1632	-1631.998	992.7	1.4	74.2	0.1812
42	-0.2332	-0.0895	-14.77	497.96	0.1437	0.0719	-0.1614	-1992.348	1022.4	1.42	71.4	0.1720
43	-0.2301	-0.0854	-16.73	493.54	0.1447	0.0724	-0.1578	-1647.262	1043.9	1.37	65.5	0.1912
44	-0.2401	-0.0941	-14.59	531.51	0.1460	0.0730	-0.1671	-1869.818	1047.3	1.44	67.1	0.1865
45	-0.2363	-0.0919	-13.67	473.93	0.1444	0.0722	-0.1641	-1916.144	963.5	1.407	77.6	0.1873
46	-0.2388	-0.0924	-16.28	464.5	0.1464	0.0732	-0.1656	-1548.774	980.9	1.4	66.6	0.1768
47	-0.2251	-0.0872	-12.87	439.49	0.1379	0.0690	-0.1562	-1456.548	927.6	1.353	67.1	0.1759
48	-0.2285	-0.0870	-12.68	473.93	0.1415	0.0708	-0.1578	-1916.149	963.5	1.407	64.2	0.1748
49	-0.2256	-0.0864	-12.29	453.52	0.1392	0.0696	-0.1560	-1459.862	965.2	1.33	57.1	0.1752
50	-0.2299	-0.0868	-12.54	507.49	0.1431	0.0716	-0.1584	-1793.61	984.8	1.416	65.7	0.1817
51	-0.2328	-0.0897	-12.45	491.92	0.1431	0.0716	-0.1613	-2015.39	970.6	1.443	67.5	0.1830
52	-0.2334	-0.0903	-12.42	508.38	0.1431	0.0716	-0.1619	-2375.749	999.3	1.458	57.8	0.1913
53	-0.2367	-0.0939	-12.23	541.93	0.1428	0.0714	-0.1653	-2253.211	1020.7	1.464	79.2	0.1899
54	-0.2363	-0.0933	-15.04	498.94	0.1430	0.0715	-0.1648	-2008.379	1018	1.46	57.3	0.1733

3.4 Construction du modèle de dérivés de l'indazole

Pour expliquer la relation structure-activité, plusieurs méthodes statistiques ont été utilisées. L'objectif principal est de dériver des modèles empiriques qui relient l'activité biologique de certains dérivés de l'indazole à leurs structures chimiques et bien comprendre la corrélation quantitative entre elles. Toutes les méthodes statistiques présentées dans cette étude ont été implémentées par le logiciel MATLAB. La régression linéaire multiple est la méthode statistique utilisée afin d'étudier la relation entre la variable dépendante, appelée aussi la réponse, (dans notre cas l'activité biologique) et les variables indépendantes, les prédicteurs, (ici, les descripteurs moléculaires). Il s'agit d'une technique mathématique qui réduit la différence entre les valeurs expérimentales et prédites d'un ensemble d'activités biologiques.³⁸⁻³⁹

Dans ce travail, les diagnostics de la régression pour détecter les points atypiques ont été effectués en calculant les valeurs de levier (h_{ii}) et les résidus r^*_i . Les éléments diagonaux de la matrice H (dite Hat Matrix) nommés les valeurs du levier, ont été utilisés pour identifier les observations X éloignées. Une valeur de h_{ii} supérieur à $2(p+1/n)$ a été considéré en indiquant des cas éloignés par rapport à leurs valeurs X.⁴⁰⁻⁴¹ Cependant, la magnitude des résidus r^*_i a été utilisée pour identifier les observations Y éloignées (aberrantes).

3.5 Validation du modèle QSAR de dérivés de l'indazole

Tester la stabilité, la capacité de prédiction et la capacité de généraliser les modèles sont des étapes très importantes dans l'étude QSAR. Pour valider le pouvoir prédictif du modèle QSAR, le coefficient de détermination (R^2) et le test de Fisher ont été utilisés pour évaluer la qualité de l'équation de régression. La validation par le bootstrap ou le coefficient multiple de corrélation de bootstrapping (R^2_{Boots}) et la validation externe (R^2_{pred})^{14, 25} ont été utilisées pour évaluer la capacité de prédiction du modèle.

4 Résultats et discussions

La sélection de l'ensemble d'apprentissage ou d'entraînement est l'une des étapes les plus importantes dans les études QSAR, puisque l'applicabilité et la prédictibilité d'un modèle QSAR dépendent de la sélection de cet ensemble.

Afin de construire le modèle QSAR, l'ensemble à modéliser (la base de données) a été divisé en 2000 divisions. Chaque division contient deux séries, une série d'apprentissage (training set) de **43** composés (**79,63%**) et une série de test (test set) pour la validation externe qui contient **11** composés (**20,37%**). Ensuite, la déviation standard d'erreur de prédiction/ standard deviation error in prediction (SDEP) pour tous les sous-ensembles de test a été calculée et tracée en fonction du nombre de descripteurs moléculaires (**figure 2**).

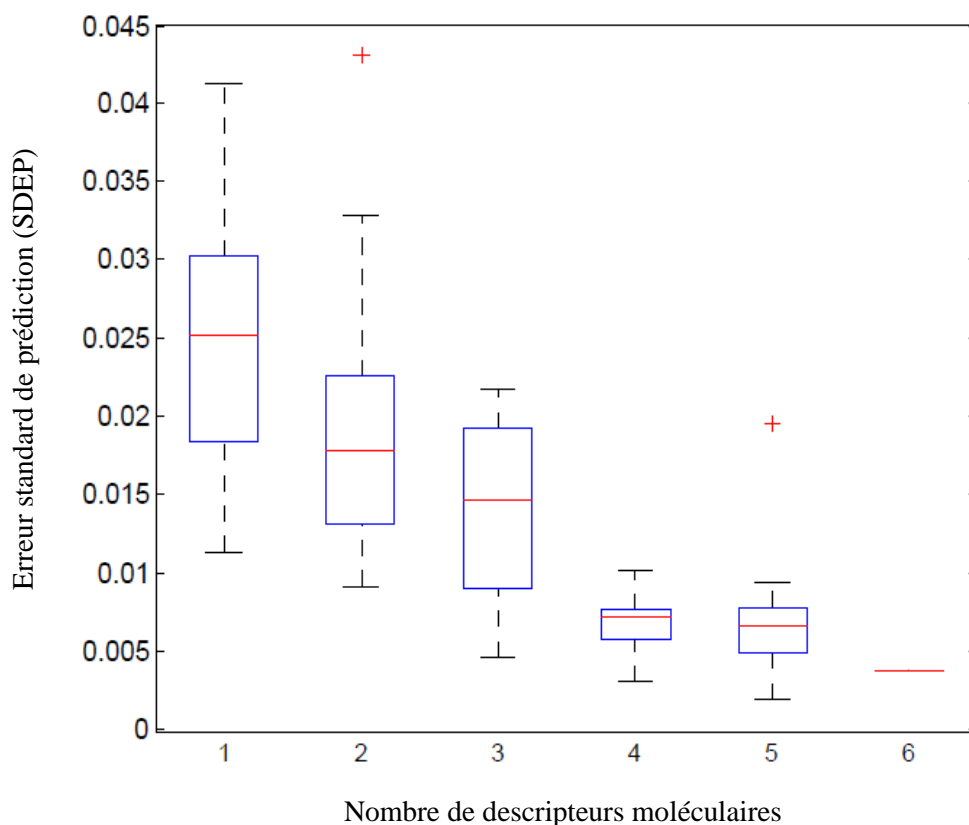


Figure 2. Boîte à moustaches (Box plots) de la distribution de la déviation standard d'erreur de prédiction (SDEP) par le nombre de descripteurs moléculaires.

D'après le diagramme en boîte, les boîtes de **4** et **5** descripteurs moléculaires ont la SDEP la plus basse. De plus, le diagramme en boîte avec quatre variables est relativement court ; cela suggère que les descripteurs globaux ont un niveau élevé d'accord les uns avec les autres. Un bon modèle QSAR a une petite SDEP ; elle indique la précision de l'ajustement statistique, donc le meilleur modèle possède quatre descripteurs moléculaires.

Afin d'identifier le meilleur sous-ensemble de descripteurs, un algorithme de recherche a été utilisé en suivant la sélection par la méthode pas à pas. Pour **2000** ensembles d'apprentissage, toutes les combinaisons possibles avec **4** variables ont été calculées. Pour chaque sous-ensemble, la meilleure combinaison a la valeur la plus faible de Bic. Afin de bien comprendre les proportions des descripteurs moléculaires sélectionnés, les proportions des variables ont été tracées par rapport aux descripteurs moléculaires sélectionnés, comme indiqué sur la **figure 3**.

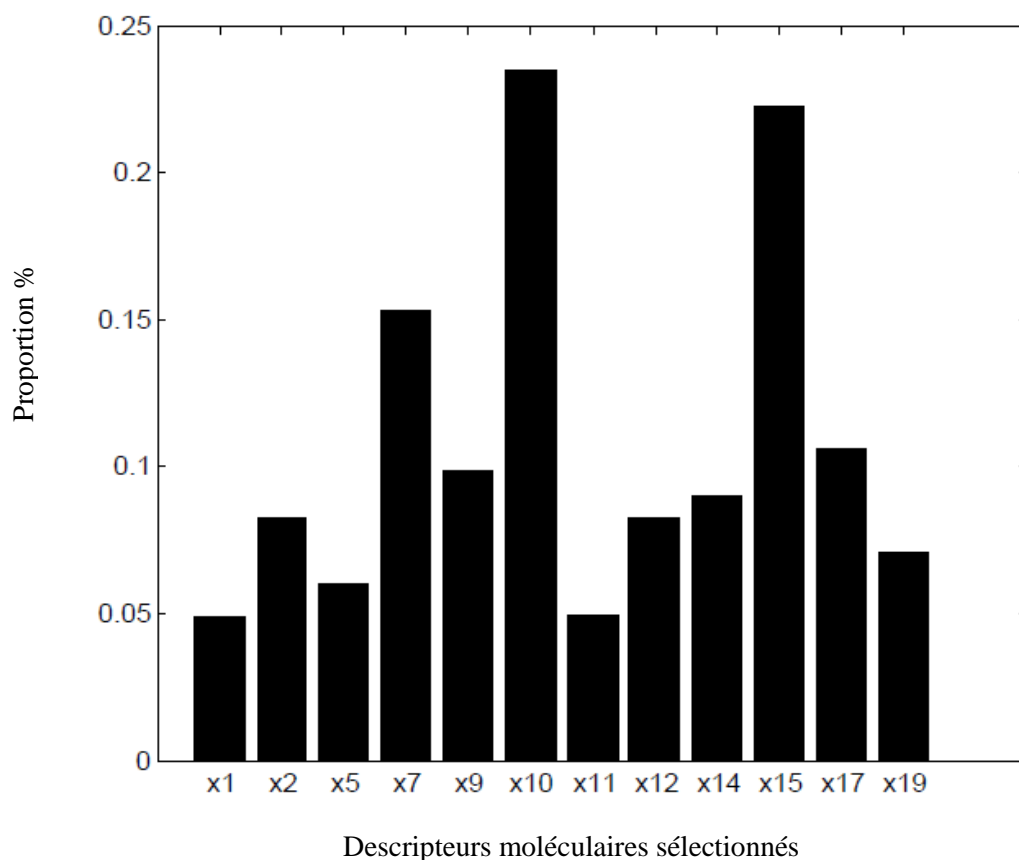


Figure 3. Proportions des descripteurs moléculaires sélectionnés.

Chapitre II Étude quantitative des relations structure-activité d'une série de dérivés de l'indazole.

N.B.: $x1(E_{HOMO})$, $x2(E_{LUMO})$, $x5(HE)$, $x7(MW)$, $x9(\Delta E)$, $x10(\eta)$, $x11(\mu)$, $x12(E_t)$, $x14(Pc)$, $x15(D)$, $x17(Surf\ ten)$, and $x19(\omega)$.

La **figure 3** présente un graphique à barres (histogramme) qui constitue un bon moyen de montrer les tailles relatives et de voir quels types de variables sont les plus importants dans la construction d'un modèle QSAR prédictif. À partir de ce graphique, on remarque quatre variables avec les proportions les plus élevées. Ces variables sont respectivement : X10 (η), X15 (**D**), X7 (**MW**) et X17 (**Surf ten**).

Les diagnostics de régression ainsi que les valeurs des activités anticancéreuses observées $pIC_{50} / Sy (Y)$ et calculées par les équations MLR sont répertoriées dans le **tableau 3**.

Tableau 3: Diagnostic statistique pour la régression.

Training set			
Y	Y _{FIT}	h_{ii}	r^*_i
8.383151162	8.011968156	0.155448551	0.695527766
7.659518963	7.553000213	0.125994853	0.221536977
8.017246432	7.527383124	0.090852045	0.870580208
7.766589276	7.558090371	0.110065852	0.410558289
7.550323298	7.647522593	0.098021479	-0.103597701
7.706402927	7.733546121	0.073659845	0.007292763
8.079391446	7.804061704	0.090014372	0.516075389
8.152508738	7.858069869	0.128060131	0.55151737
8.147784695	7.86876132	0.16595159	0.530877177
8.219237883	7.846153545	0.165144553	0.684667638
8.146773976	7.858078609	0.14633001	0.529432026
7.886002186	7.842524662	0.130216945	0.097492891
7.483114129	7.779418199	0.128244622	-0.501403372
7.087810741	7.70118192	0.134986914	-1.060787702
6.870078842	7.637362842	0.126944095	-1.330863186
6.931063993	7.589643318	0.105236561	-1.11726169
7.162919329	7.578321879	0.090137947	-0.679213832
7.507020081	7.615608723	0.085948262	-0.147266698
7.745083333	7.677206104	0.082593659	0.145267833
7.776199774	7.742849683	0.081938855	0.077522609
7.700198772	7.798915877	0.074977848	-0.146492411
7.559190703	7.848876825	0.068755052	-0.466635274
7.5468532	7.874667603	0.072470438	-0.518418558
7.695746432	7.93626183	0.083373821	-0.354719802
7.991028592	8.046374287	0.088901484	-0.031310641
8.346517963	8.248852994	0.094813593	0.217796844
8.631205747	8.515070828	0.103886389	0.238096249
8.769953285	8.685325733	0.108891194	0.18041416

8.778025896	8.720027723	0.115767573	0.131157861
8.783036029	8.712813552	0.124683175	0.147148002
8.821065348	8.767869771	0.142740899	0.120598072
8.962012013	8.848570991	0.145754659	0.234982585
9.06730731	8.909165869	0.143186727	0.322792597
9.100020375	8.912840246	0.131570807	0.374353474
9.017907905	8.88554343	0.128274292	0.269254258
8.69943602	8.782194791	0.128487141	-0.108780421
8.443638158	8.657223268	0.113445579	-0.326085485
8.430609487	8.524389661	0.080076968	-0.116685456
8.613083288	8.473593476	0.079742283	0.284154272
8.754910362	8.616774449	0.100165071	0.28532989
8.616565372	8.858131261	0.111345389	-0.356670016
8.420452692	8.905093956	0.194472948	-0.812811566
8.849670622	9.224840105	0.248425527	-0.629576628
Test set			
7.963909897	8.21928635	0.486735529	0.358447841
7.833748126	8.153003496	0.438389086	0.202326716
7.781345835	8.189953031	0.452642481	-0.02280633
7.602830421	8.12658816	0.450472063	-0.290262131
7.616902377	8.197475331	0.407037346	-0.436500589
7.835812162	8.392252052	0.384574112	-0.341609101
8.204284606	8.683301441	0.399445259	-0.130373571
8.555715657	8.974182845	0.43729022	0.034118935
8.725910473	9.12989633	0.478763019	0.062020843
8.689722437	9.13086483	0.483256118	-0.040488241
8.675851394	9.217044095	0.581394767	-0.295170222

Toutes les observations des sous-ensembles d'apprentissage ont des valeurs de levier inférieures au seuil de **0,23 (10/43)**, sauf la dernière qui est légèrement supérieure au seuil. À partir des valeurs résidus r^*_i , toutes les observations des sous-ensembles d'apprentissage sont inférieures au seuil $|2|$. Alors que toutes les observations des sous-ensembles de tests ont des valeurs de levier inférieures au seuil **0,90 (10/11)** et leurs valeurs résidus r^*_i sont comprises entre **-2** et **2**.

Le **tableau 3** présent la moyenne de tous les diagnostics statistiques pour la régression et il montre un bon accord entre les valeurs expérimentales et prédites. La **figure 4** présente la corrélation entre les valeurs prédites et expérimentales pour l'activité biologique.

La valeur normalisée du coefficient de régression de chaque descripteur met en évidence l'importance relative des descripteurs dans la détermination de l'activité biologique des composés.

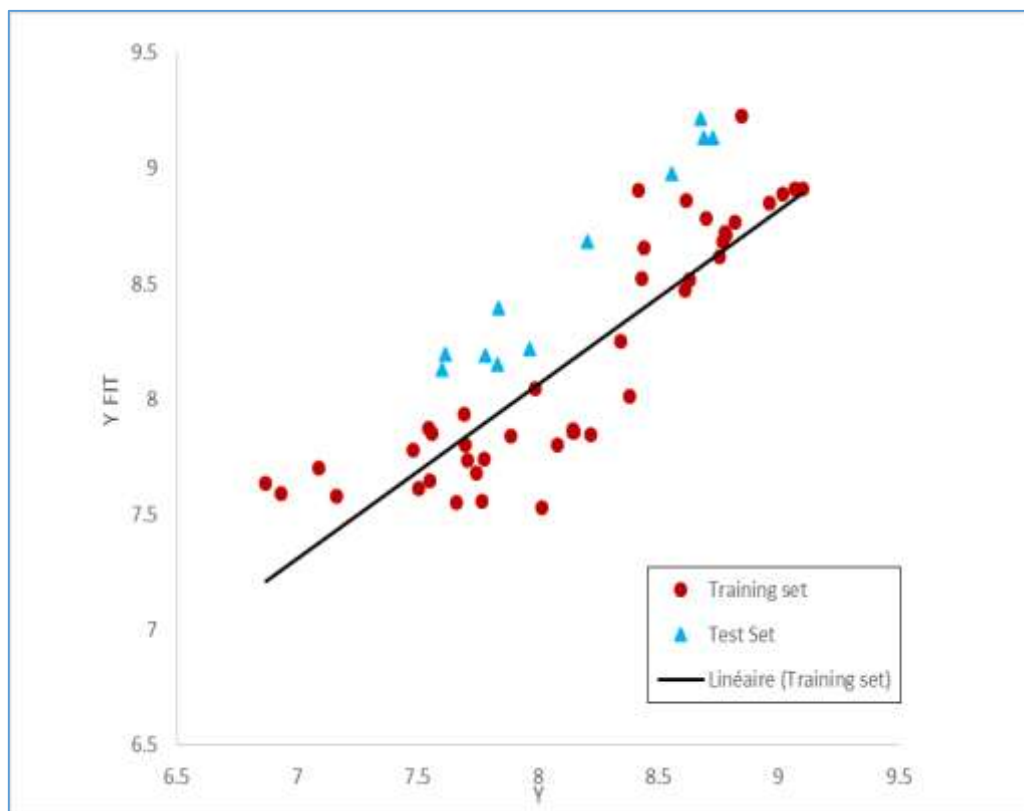


Figure 4. Corrélation entre les activités biologiques prédites et expérimentales ²⁷ pour le modèle obtenu.

La construction du modèle QSAR a été établie sur la base de quatre descripteurs ; X10 (η), X15 (**D**), X7 (**MW**) et X17 (**Surf ten**). La matrice de corrélation entre ces descripteurs et l'activité biologique étudiée est reportée dans le **tableau 4**.

Tableau 4: Matrice de corrélation entre pIC₅₀ et les descripteurs impliqués dans le modèle QSAR

	$\log(1/IC_{50})$	MW	η	D	Surf ten
$\log(1/IC_{50})$	1				
MW	0.615	1			
η	0.562	0.748	1		
D	0.649	0.879	0.838	1	
Surf ten	0.482	0.389	0.506	0.554	1

Le **tableau 4** montre que la densité et le poids moléculaire sont des paramètres importants dans la corrélation entre les descripteurs sélectionnés et l'activité anticancéreuse pIC₅₀. De plus, la matrice de corrélation confirme l'absence du problème de colinéarité entre les descripteurs du modèle. Ceci est expliqué par les valeurs des coefficients de corrélation qui sont inférieures à 0.9 ($r < 0,9$).

La corrélation entre l'activité biologique et les descripteurs sélectionnés est exprimée par l'équation suivante :

$$Y = 9,96(\pm 0,45) + 0,28 (\pm 0,10) X_7' + 0,04 (\pm 0,02) X_{10}' + 0,26 (\pm 0,14) X_{15}' + 0,21 (\pm 0,05) X_{17}' \quad (1)$$

$$R^2_m = 0,59 \quad R^2_{adj, m} = 0,55 \quad R^2_{Boots} = 0,66 \quad B = 2000 \quad SDEP = 0,395 \quad R^2_{pred, m} = 0,57 \\ F = 8,46 \quad p\text{-value} = 1,26 \times 10^{-4}$$

Où $Y = pIC_{50}/S_y$ et $X'_j = X_j/S_{X_j}$. Le terme S est l'écart-type et le paramètre B est le nombre de répliques bootstrap. Le descripteur moléculaire X10 (η) présente un rôle négligeable sur l'activité anticancéreuse. Les descripteurs moléculaires X17 (**Surf ten**), X7 (**MW**) et X15 (**D**) présentent relativement le même effet sur l'activité anticancéreuse. Dans le modèle obtenu on remarque que toute augmentation des valeurs de tous ces descripteurs révèle une augmentation de la valeur de l'activité biologique.

Le coefficient de corrélation ($R = 0,768$) nous indique qu'il existe une forte corrélation entre les valeurs observées de l'activité biologique et celles prédites par le modèle de régression. En outre, un coefficient de corrélation élevé et une faible valeur de SDEP indiquent que le modèle est fiable.

La valeur du coefficient de détermination ($R^2 = 0,59$) signifie que **59%** de la variabilité de l'activité anticancéreuse peut être expliquée par les **4** descripteurs sélectionnés.

La valeur F est jugée la signification statistique au niveau de **95%**. Dans notre cas la valeur F calculée est supérieure à la valeur critique ($F_{(0,05 ; 4 ; 38)}$) lus dans la table de Fisher. L'équation de la régression est significative car la valeur p-value est plus inférieure à 0.05 (rejetant l'hypothèse nulle).

Afin de tester la validité de pouvoir prédictif du modèle QSAR, nous avons utilisé 20.37% des données dans l'ensemble de validation externe. Une valeur moyenne élevée

du R^2_{Boots} supérieure à 0.5 est une démonstration que le modèle obtenu a un bon pouvoir prédictif et une bonne robustesse.

De plus, une valeur moyenne de R^2_{pred} (**0.57**) approximativement égale à 0.6 peut confirmer la bonne capacité prédictive du modèle pour les composés n'ayant pas participé aux calculs. Elle indique que ce modèle peut être appliqué avec succès pour prédire l'activité anticancéreuse pIC_{50} pour le traitement de la tumeur RE+ du cancer du sein de dérivés de l'indazole du SERDs.

5 Conclusion

Ce chapitre a été consacré à l'application de la méthode de QSAR, pour modéliser l'activité anticancéreuse à partir d'un ensemble de **54** dérivés de l'indazole. La méthode de régression linéaire multiple a été utilisée dans le développement du modèle pour déterminer quantitativement l'effet de la structure moléculaire des composés étudiés sur leur activité biologique.

Le meilleur modèle est celui à quatre descripteurs moléculaires qui ont les proportions les plus élevées. La validité de ce modèle a été établie par la détermination du coefficient de détermination, le test de Fisher, la validation par le bootstrap (R^2_{Boots}) et la validation externe. La moyenne de tous les diagnostics statistiques pour la régression montre un bon accord entre les valeurs expérimentales et prédites.

Les résultats obtenus de la validation confirment la bonne capacité prédictive du modèle et indiquent que l'activité anticancéreuse pour la série étudiée de dérivés de l'indazole peut être modélisée avec succès en utilisant les quatre descripteurs moléculaires sélectionnés.

Références bibliographiques

1. Ghamali, M.; Chtita, S.; Bouachrine, M.; Lakhli, T., Méthodologie générale d'une étude RQSA/RQSP. *Revue Interdisciplinaire* **2016**, *1* (1).
2. Karelson, M., *Molecular descriptors in QSAR/QSPR*. Wiley-Interscience New York: **2000**; Vol. 230.
3. Fang, J.; Pang, X.; Wu, P.; Yan, R.; Gao, L.; Li, C.; Lian, W.; Wang, Q.; Liu, A. I.; Du, G. h., Molecular Modeling on Berberine Derivatives toward BuChE: An Integrated Study with Quantitative Structure–Activity Relationships Models, Molecular Docking, and Molecular Dynamics Simulations. *Chemical Biology & Drug Design* **2016**, *87* (5), 649-663.
4. Lagarde, N. Méthodes de criblage virtuel in silico: importance de l'évaluation et application à la recherche de nouveaux inhibiteurs de l'interleukine 6. Thèse de doctorat, Paris, CNAM, **2014**.
5. Kraim, K. Etudes de QSAR sur des activités biologiques utilisant des produits d'origines naturels. Thèse de doctorat, Université de Guelma, **2009**.
6. Huynh, T. N. P. Synthèse et études des relations structure/activité quantitatives (QSAR/2D) d'analyse benzo [c] phénanthridiniques. Thèse de doctorat, Université d'Angers, **2007**.
7. Li, F.; Wu, H.; Li, L.; Li, X.; Zhao, J.; Peijnenburg, W. J., Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. *Ecotoxicology and Environmental Safety* **2012**, *80*, 273-279.
8. Reddy, A. S.; Kumar, S.; Garg, R., Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of Tipranavir analogs for HIV protease inhibition. *Journal of Molecular Graphics and Modelling* **2010**, *28* (8), 852-862.
9. Dearden, J.; Cronin, M. T.; Kaiser, K. L., How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research* **2009**, *20* (3-4), 241-266.
10. Prana, V. Approches structure-propriété pour la prédiction des propriétés physico-chimiques des substances chimiques. Thèse de doctorat, Université Pierre et Marie Curie, **2013**.
11. Zhao, M.; Li, Z.; Peng, L.; Tang, Y.-R.; Wang, C.; Zhang, Z.; Peng, S., Novel 1-oxyl-2-substitutedphenyl-4, 4, 5, 5-tetramethylimidazolines: Synthesis, selectively

analgesic action, and QSAR analysis. *Bioorganic & Medicinal Chemistry* **2007**, *15* (8), 2815-2826.

12. Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J.; De Nucci, G., QSAR models for inhibitors of physiological impact of *Escherichia coli* that leads to diarrhea. *Biochemical and Biophysical Research Communications* **2013**, *432* (2), 214-225.

13. Golbraikh, A.; Tropsha, A., Beware of q^2 ! *Journal of Molecular Graphics and Modelling* **2002**, *20* (4), 269-276.

14. Tropsha, A.; Gramatica, P.; Gombar, V. K., The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science* **2003**, *22* (1), 69-77.

15. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American chemical society* **1985**, *107* (13), 3902-3909.

16. Saihi, Y. Etude de la relation quantitative structure-activité inhibitrice des enzymes hydrolytiques : cas des alpha-glucosidases. Thèse de doctorat, Université Badji Mokhtar-Annaba, **2015**.

17. Rakotomalala, R., Pratique de la regression lineaire multiple. *Diagnostic et selection de variables* **2011**.

18. Roy, K.; Roy, P. P., Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *European Journal of Medicinal Chemistry* **2009**, *44* (7), 2913-2922.

19. Besse, P., Apprentissage statistique & data mining. *Toulouse: INSA* **2009**.

20. Varnek, A.; Tropsha, A., *Chemoinformatics approaches to virtual screening*. Royal Society of Chemistry: **2008**.

21. Gramatica, P., Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science* **2007**, *26* (5), 694-701.

22. Efron, B.; Tibshirani, R. J., *An introduction to the bootstrap*. CRC press: **1994**.

23. Wehrens, R.; Putter, H.; Buydens, L. M., The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems* **2000**, *54* (1), 35-52.

24. Efron, B., Bootstrap methods : another look at the jackknife. *The Annals of Statistics* **1979**, 7 (1), 1-26.
25. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K., Validation of QSAR models-strategies and importance. *International Journal of Drug Design and Discovery* **2011**, 2 (3), 511-519.
26. Melkemi, N. Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques. Thèse de doctorat, Université Mohamed Khider-Biskra, **2013**.
27. Lai, A.; Kahraman, M.; Govek, S.; Nagasawa, J.; Bonnefous, C.; Julien, J.; Douglas, K.; Sensintaffar, J.; Lu, N.; Lee, K.-j.; Aparicio, A.; Kaufman, J.; Qian, J.; Shao, G.; Prudente, R.; Moon, M. J.; Joseph, J. D.; Darimont, B.; Brigham, D.; Grillot, K.; Heyman, R.; Rix, P. J.; Hager, J. H.; Smith, N. D., Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. *Journal of Medicinal Chemistry* **2015**, 58 (12), 4888-4904.
28. Hyperchem v8, molecular modeling system, Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA, **2009**.
29. Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007**, 13 (12), 1173-1213.
30. Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics* **1971**, 54 (2), 724-728.
31. Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics* **1972**, 56 (5), 2257-2261.
32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloin, o. J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, F.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari,

- K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J., Gaussian 09, Revision A. 02, Gaussian, Inc. *Wallingford, CT* **2009**.
33. Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A., Gaussian-1 theory of molecular energies for second-row compounds. *The Journal of Chemical Physics* **1990**, *93* (4), 2537-2545.
34. Advanced Chemistry Development, Inc., Toronto, Canada, **2009**, <http://www.acdlabs.com/resources/freeware/chemsketch/>.
35. Hemmateenejad, B.; Miri, R.; Akhond, M.; Shamsipur, M., QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. *Chemometrics and Intelligent Laboratory Systems* **2002**, *64* (1), 91-99.
36. <https://www.mathworks.com/products/matlab/whatsnew.html>
37. James, G.; Witten, D.; Hastie, T.; Tibshirani, R., *An introduction to statistical learning*. Springer: **2013**; Vol. 112.
38. Hadaji, E. G.; Bourass, M.; Ouammou, A.; Bouachrine, M., Organic Compounds Based on (E)-N-Aryl-2-ethene-sulfonamide as Microtubule Targeted Agents in Prostate Cancer: QSAR Study. *Advances in Physical Chemistry* **2017**, *2017*.
39. Mata, J., Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. *Engineering Structures* **2011**, *33* (3), 903-910.
40. Kenouch, S.; Harkati, D.; Ghamri, M.; Chikhaoui, A. R.; Melkemi, N., Predictive QSAR model and clustering analysis of some Benzothiazole derivatives as cytotoxic inhibitors. *SDRP Journal of Computational Chemistry & Molecular Modelling* **2018**, *2* (3), 1-8.
41. Rawlings, J. O.; Pantula, S. G.; Dickey, D. A., *Applied regression analysis: a research tool*. Springer Science & Business Media: **2001**.

Chapitre III

Étude de docking moléculaire d'une série de dérivés de l'indazole.

1 Introduction

La conception de médicaments repose fréquemment sur les techniques de modélisation sur ordinateur dites aussi « in silico ». Ce type de modélisation est connu sous le nom de conception de médicaments assistée par ordinateur. Les techniques de modélisation tentent de prédire l'affinité entre le ligand et sa cible et, à ce jour, les résultats sont globalement satisfaisants. Cependant, il existe d'autres propriétés qui doivent être optimisées pour que le ligand devienne sûr et efficace autant que puisse l'être un médicament. Ces caractéristiques moléculaires sont l'absence d'effets secondaires, la biodisponibilité, la demi vie métabolique, la toxicité...etc. sont par ailleurs assez difficiles à optimiser avec les techniques de conception de médicaments rationnelle. ¹

Le mécanisme d'action de nombreux médicaments consiste à agir comme inhibiteur/ activateur et une récepteur impliquée dans le développement des médicaments. L'élucidation de la structure tridimensionnelle de protéines impliquées dans plusieurs pathologies a permis, via des simulations informatiques, la découverte d'inhibiteurs puissants pour ces protéines, en réduisant considérablement le nombre d'essais de screening nécessaires pour aboutir à un nouveau médicament. ²

Les interactions entre molécules sont à la base de la plupart des mécanismes biologiques. Les détails de ces interactions, au niveau moléculaire, sont donc d'un très grand intérêt et peuvent être étudiés par cristallographie aux rayons X ou résonance magnétique nucléaire (RMN). Malheureusement, ces techniques ne permettent pas de détailler la totalité de ces interactions étant donné le nombre incalculable de molécules différentes au sein d'une seule cellule.

Le docking moléculaire in silico vise à prédire la structure d'un complexe moléculaire à partir des molécules isolées, ce qui est considérablement plus facile à mettre en oeuvre, moins cher et plus rapide que l'utilisation d'une des méthodes expérimentales mentionnées ci-dessus. ³

Nous présentons dans ce chapitre l'étude des interactions entre **21** inhibiteurs de dérivés de l'indazole et la protéine (**ID : 3 ERT**) par les méthodes de modélisation moléculaire. Nous avons rassemblé les résultats et leurs discussions de calcul par docking que nous avons engagé en se basant sur les énergies de scores et les distances entre les

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

différents inhibiteurs et les résidus du site actif de la protéine (**ID : 3 ERT**). Une étude des propriétés ADME a été ajoutée sur les meilleurs composés obtenus par l'étude de docking afin d'évaluer leurs propriétés et de sélectionner ceux qui ne posent pas de problème avec la biodisponibilité orale.

2 Aperçue générale sur le docking moléculaire

Docking (ancrage/ amarrage) est le nom donné aux simulations moléculaires dans lesquelles différentes approches sont combinées pour étudier les modes d'interaction entre deux molécules. Les logiciels de docking sont des outils très utiles en biologie, pharmacie et médecine, car la plupart des principes actifs sont de petites molécules qui interagissent avec une cible biologique d'intérêt thérapeutique. Le récepteur macromoléculaire étant le plus souvent une protéine, le terme docking seul est couramment employé pour désigner un «docking protéine ligand » (**figure 1**). ⁴

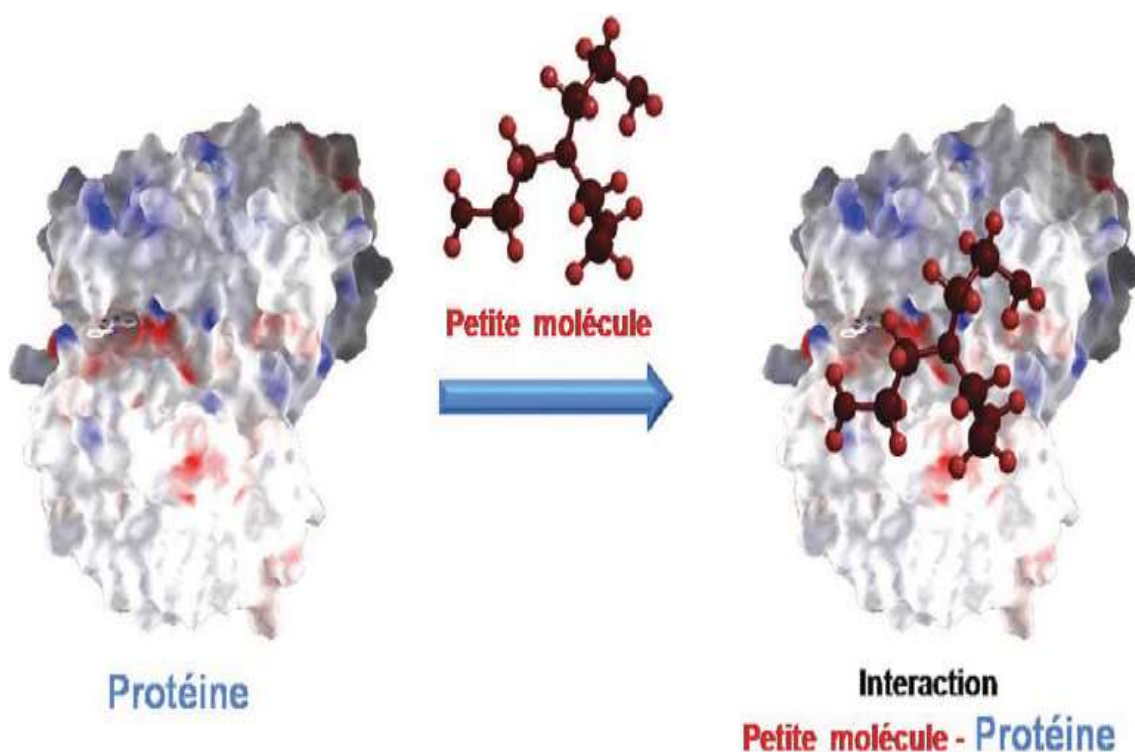


Figure 1. Illustration schématique de l'encrage d'une petite molécule à une cible protéique produisant un complexe stable ⁵

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Les méthodes de docking combinent l'utilisation d'un algorithme de recherche, permettant de générer des modes de liaisons putatifs du ligand dans le récepteur, ou « poses », et d'une fonction de score, employée pour classer les différentes poses selon un score prédit d'affinité.⁶ Les méthodes de docking s'attachent donc, d'une part, à identifier les molécules qui sont des ligands véritables du récepteur parmi toutes celles étudiées, et aussi d'autre part, à déterminer les poses correctes soit les conformations adoptées par les ligands lors de la liaison au récepteur.⁷

Le docking comprend essentiellement deux étapes : le docking proprement dit et le scoring (**figure 2**) :

- ❖ La première étape (le docking) : c'est l'étape de sélection, elle consiste à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celles qui représentent les modes d'interaction les plus favorables. Bien que pouvant être faite manuellement, cette étape est le plus souvent exécutée de manière automatisée à l'aide d'algorithmes de docking, ce qui améliore la vitesse et la précision des simulations.⁴
- ❖ La deuxième étape (scoring) : les fonctions de score sont utilisées pour estimer mathématiquement l'affinité de liaison entre un récepteur et chacune des poses générées pendant le docking.⁸ L'efficacité de ces fonctions de score est au moins tout aussi importante que celle des algorithmes de recherche conformationnelle. En effet, même si la conformation bioactive du ligand a été obtenue lors docking, si les fonctions de score ne permettent pas de différencier les poses correctes de celles incorrectes, les composés les plus prometteurs pour la cible ne pourront pas être identifiés.⁹

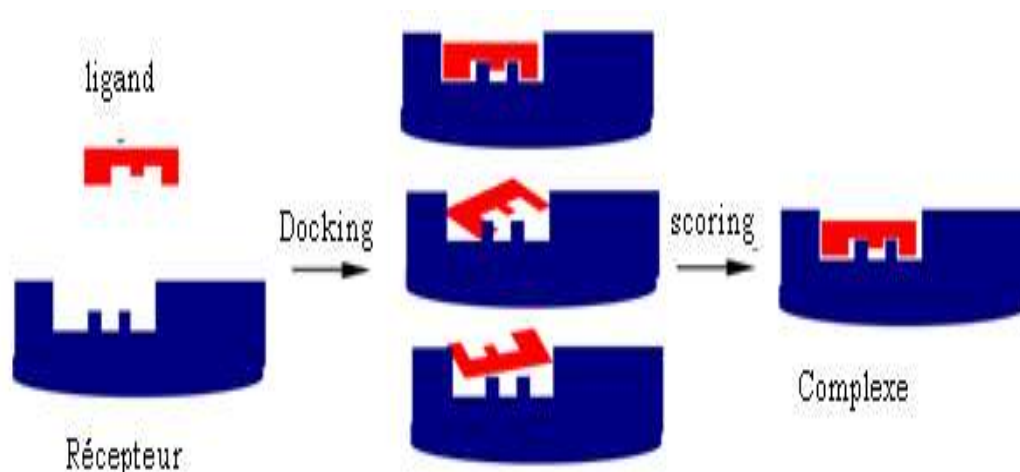


Figure 2. Principe général du docking moléculaire

2.1 Différentes techniques du docking moléculaire

Les techniques de docking moléculaire peuvent être classées en trois catégories selon la prise en compte de la flexibilité du ligand, les algorithmes utilisés pour rechercher les poses possibles d'un ligand et les fonctions de score permettant d'évaluer l'affinité de liaison de chacune de ces poses.¹⁰ Ainsi, nous pouvons avoir trois types de docking :

2.1.1 Docking rigide

Le docking rigide, au cours duquel la protéine et le ligand sont tous deux traités comme entièrement rigides. Cette technique reflète le principe de la clé et de la serrure « lock and key » énoncé par Emile Fisher en 1890 où seul le ligand (représentant la clé) possédant la bonne taille et la bonne forme, peut ouvrir une serrure : la protéine (**figure 3**).¹¹⁻¹²

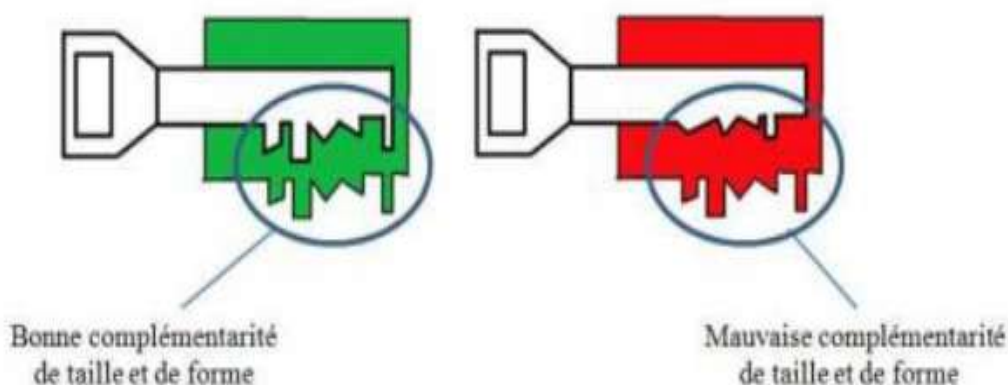


Figure 3. Schéma illustrant le principe de « lock and key » d'Emile Fisher.

Dans cette technique ni le ligand ni le récepteur ne subissent de réarrangements conformationnels suite à leur interaction. Elle est généralement employée pour l'amarrage entre deux macromolécules (docking protéine-protéine ou protéine/acide nucléique structuré) pour lesquelles les degrés de liberté sont trop importants pour réaliser un échantillonnage conformationnel efficace dans des temps de calculs raisonnables.¹³

2.1.2 Docking semi-flexible

Dans cette technique, le ligand est flexible afin d'explorer toutes ses conformations possibles, alors que la cible est maintenue rigide durant l'assemblage. Ce type de docking présente, certes, l'avantage de nécessiter un temps de calcul relativement court mais il ne tient pas compte des changements conformationnels, requis par la cible.

14

2.1.3 Docking flexible

Cette technique traite la flexibilité du ligand et du récepteur à la fois. L'un des grands enjeux des méthodes de docking flexible est d'explorer l'espace conformationnel des ligands dans le but de déterminer d'une part les bonnes poses du ligand dans le site actif de la protéine et d'autre part de trouver les modes de liaison correspondant à une énergie libre de liaison basse.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Les deux derniers types de docking sont appliqués, plus généralement, à la modélisation d'interaction entre les protéines et les petites molécules.¹⁵

2.2 Algorithmes de docking moléculaire

En principe, un docking peut être fait de façon « manuelle », directement par le modélisateur, en plaçant le ligand dans le site actif de la protéine à l'aide d'une interface graphique. Ensuite, la géométrie de l'ensemble est optimisée de manière à corriger les problèmes stériques et obtenir un complexe énergétiquement stable. Cette approche est appliquée quand on a une idée précise du mode d'interaction réel du ligand.¹⁶ Néanmoins, le plus souvent, le mode d'interaction réel n'est pas connu. Dans ce cas, tester manuellement toutes les conformations et orientations des ligands s'avère impossible d'un point de vue pratique, même en considérant la protéine comme un corps rigide. Pour contourner cette difficulté, les algorithmes de docking ont été conçus pour rechercher de façon objective, rapide et efficace les modes d'association protéine ligand les plus favorables.¹⁷

L'aptitude d'un algorithme à trouver l'emplacement correct du ligand par rapport à son récepteur est habituellement déterminée au moyen de la déviation quadratique moyenne ou RMSD (Root-Mean-Square Deviation) du modèle conçu par le logiciel. Le RMSD entre deux poses est une mesure géométrique de la distance entre les positions atomiques de la structure expérimentale et celles de la structure prédite du complexe ligand /site de liaison.¹⁸ L'écart RMSD est donné par l'équation suivante :

$$\text{RMSD}(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2 \right)}$$

(1)

Avec v_i et w_i sont respectivement les atomes identiques de la structure expérimentale et de la structure prédite. x, y et z les coordonnées cartésiennes et n le nombre total d'atomes. Plus la prédiction du positionnement est précise, plus les différences entre les deux structures sont faibles, plus la valeur du RMSD est basse. La valeur admise est une

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

différence maximale de deux angströms au-delà de laquelle la prédiction est considérée comme non adéquate.¹⁹⁻²⁰

Pour réaliser un calcul de RMSD, le ligand co-cristallisé d'une cible donnée est extrait, et repositionné par docking dans le site actif. C'est ce qu'on appelle en anglais le « cognate redocking » ou encore « self docking ». ¹⁸ Pour évaluer la précision du positionnement, seuls les atomes lourds sont généralement considérés. Les avantages présentés par le calcul du RMSD, tels que l'objectivité, la sensibilité et la simplicité d'automatisation ²¹, ont permis à cette métrique de s'imposer comme la référence dans l'évaluation de la précision du positionnement.

2.3 Fonction de score

La fonction de score est une donnée numérique utile pour quantifier le degré avec lequel un ligand se complexe à un récepteur. C'est globalement une approximation de l'énergie libre résultant du passage de la forme libre de la protéine et du ligand à l'association sous forme de complexe. Le principe thermodynamique est le suivant ²² :

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{proteine}} \quad (2)$$

L'établissement d'une bonne fonction de score est un important problème du docking. Il arrive souvent que la solution évaluée comme étant la plus probable ne soit pas la forme native attendue. Ceci peut être dû au fait que le complexe natif n'est pas forcément celui qui présente la plus grande surface d'accès, ou encore le plus grand nombre de liaisons hydrogène disponibles. Pour cela il existe différents types de fonctions de score selon les critères sur lesquels celles-ci sont basées. ²³ L'affinité chimique peut être calculée par l'énergie libre de Gibbs ΔG_1 pour une température T ²⁴⁻²⁵ :

$$\Delta G_l = -RT \ln K_{eq} \quad (3)$$

Où, R est la constante des gaz parfaits et K_{eq} est la constante d'équilibre. Le calcul de l'énergie libre étant cependant lourd, son application devient impossible lorsqu'il s'agit d'évaluer un grand nombre de complexes. L'estimation de l'affinité repose généralement sur un calcul d'énergie contenant un certain nombre d'approximations. Pour cela il existe

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

différentes fonctions de score qui sont ajustées dans les programmes de docking et ont fait l'objet de nombreuses publications.²⁶⁻²⁸

2.4 Interactions entre le ligand et le récepteur

Les interactions entre une protéine et un ligand sont en général de nature non covalente. L'énergie libre de Gibbs se calcule selon l'équation suivante :

$$\Delta G = -RT \ln K_i = \Delta H - T \Delta S \quad (4)$$

R: constante des gaz parfaits (8,314 J. K⁻¹ mol⁻¹)

T: Température (K)

K_i: constante d'inhibition

La grandeur du ΔG en solution aqueuse est comprise entre -10 et -70 KJ/mol. Les composantes de l'énergie libre sont²⁹ :

- ❖ la complémentarité géométrique, stérique et de surface entre le ligand et le récepteur
- ❖ les contacts présents entre deux régions lipophiles des deux entités.
- ❖ les liaisons hydrogène

Les molécules d'eau dans les cavités des protéines peuvent parfois être un élément fondamental.³⁰ Elles sont capables d'assurer le relais entre le récepteur et le ligand et ainsi créer des réseaux de liaisons hydrogène.

Parmi les interactions qui peuvent exister entre le ligand, qui est considéré comme une petite molécule organique, et la protéine on cite celles de type polaire (liaisons hydrogène, liaisons ioniques) et celles de type hydrophobe qui résultent d'un contact entre des groupements hydrophobe.³¹

2.4.1 Interaction ionique

Les interactions entre molécules portant des charges électriques sont régies par la loi de Coulomb. L'énergie potentielle d'interaction de deux charges électriques Q_1 et Q_2 séparées par une distance d est obtenue en calculant le travail nécessaire à la séparation

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

de ces deux charges à une distance infinie (**figure 4**). L'énergie potentielle (V) de deux molécules de charges différentes est exprimée par l'équation 3.

$$V = \frac{Q_1 Q_2}{4 \pi \epsilon \epsilon_0 d} \quad (5)$$

Avec :

ϵ_0 : permittivité du vide.

ϵ : permittivité du milieu dans lequel évoluent les charges.

d : distance entre les deux charge.

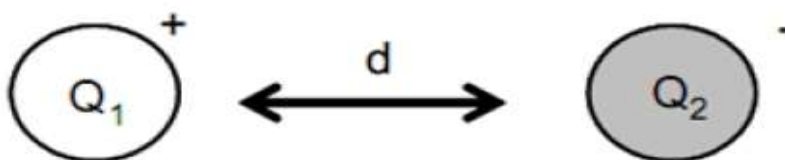


Figure 4. Interaction entre deux molécules de charges différentes.

En milieu aqueux, les acides aminés chargés sont majoritairement exposés à la surface de la protéine et sont donc très solvatés, ce qui réduit considérablement l'interaction entre les charges. ³¹

2.4.2 Interaction hydrogène

La liaison hydrogène est une force attractive qui s'opère entre deux groupes d'atomes impliquant un atome d'hydrogène. C'est une liaison physique qui se situe entre la liaison covalente et l'interaction électrostatique. Le caractère covalent de la liaison hydrogène a été démontré à la fin des années 1990, mais la force d'interaction prédominante reste électrostatique. ³²

Cette liaison ne s'effectue pas avec n'importe quel atome, il faut un donneur et un accepteur (sous-entendu de proton) :

- ❖ Le donneur est l'atome d'hydrogène lié de manière covalente à un atome électronégatif ; typiquement azote, oxygène, fluor.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

- ❖ L'accepteur est un autre atome électronégatif ; uniquement azote, oxygène ou fluor.

La **figure 5** montre des exemples de liaison hydrogène.

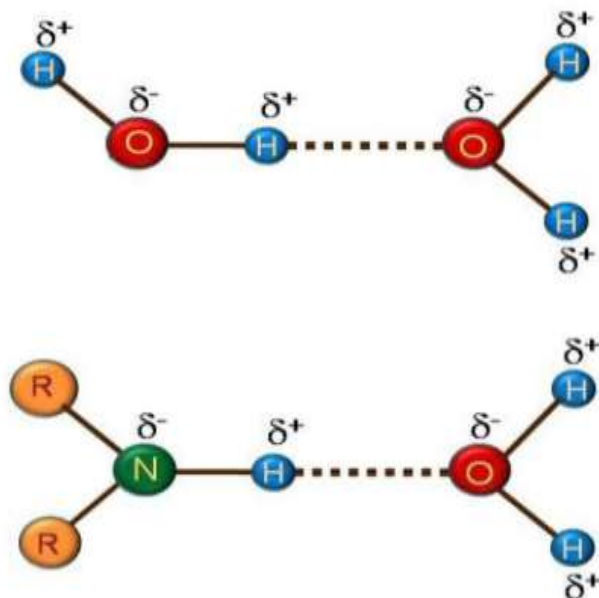


Figure 5. Exemples d'une liaison hydrogène.

2.4.3 Interaction de Van der Waals

Ces interactions ont été étudiées par J.D. van der Waals, physicien hollandais (prix Nobel de physique 1910). Elles reposent sur les interactions entre les dipôles constitués par les molécules. On distingue trois types de ces interactions :

- ❖ Interaction dipôle permanent – dipôle permanent ou effet d'orientation de Keesom : Ce type d'interaction se développe entre deux molécules polaires.
- ❖ Interaction dipôle permanent – dipôle instantané ou effet d'induction de Debye : Ce type d'interaction se développe entre une molécule polaire et une molécule quelconque (polaire ou apolaire).
- ❖ Interaction dipôle instantané – dipôle instantané ou dispersion de London : Ce type d'interaction se développe entre deux molécules quelconques (polaires ou apolaires).

Ces interactions sont très faibles mais dans le cas des macromolécules, leur nombre élevé va produire au totale une force importante. ³³

2.4.4 Interactions hydrophobes

Les molécules ou groupes d'atomes non polaires ne sont pas capables de former des liaisons hydrogènes et ne peuvent donc pas s'hydrater : pour cette raison, on les nomme substances hydrophobes. L'effet hydrophobe est la tendance qu'ont ces groupes à se rassembler par coalescence de façon à minimiser les contacts avec l'eau. Les liaisons hydrophobes possèdent une énergie de l'ordre de 7 kcal/mol.³⁴

Une molécule hydrophobe rompt localement le réseau de liaisons hydrogènes du solvant et crée un état énergétiquement défavorable (perte d'entropie). Les molécules d'eau se réorganisent alors autour des substances hydrophobes de manière à minimiser l'énergie. D'un point de vue thermodynamique, la diminution du nombre de molécules d'eau autour des groupes non-polaires et l'agrégation de ces derniers sous l'effet des forces de dispersion, c'est-à-dire la séparation de deux phases distinctes, organique et aqueuse, s'accompagne d'un gain d'entropie de l'ensemble du solvant et de la molécule non polaires. Pour cette raison, les régions non polaires sont souvent enfouies à l'intérieur des molécules. Cet effet a d'importantes répercussions sur la conformation et la stabilité des molécules. Elle est à l'origine du processus de repliement des protéines par exemple, ou encore de l'auto-assemblage de biomolécules.³⁵⁻³⁶

Plusieurs acides aminés hydrophobes se trouvent en contact à l'interface par complémentarité hydrophobe. Ces derniers entraînent une interaction stabilisante pour beaucoup de complexes. Les acides aminés non polaires tendent à créer entre elles des liaisons de type interactions "patch" hydrophobe³⁷ (**figure 6**).

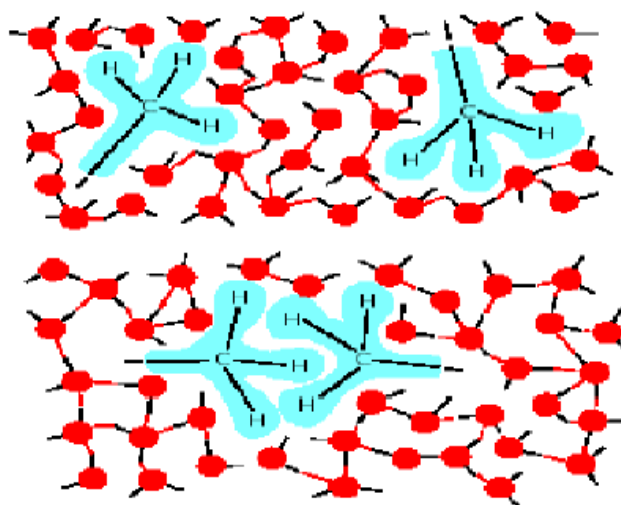


Figure 6. Patch hydrophobe

2.5 Étapes typiques du docking moléculaire

La technique de docking comprend quatre étapes principales (figure 7)³ :

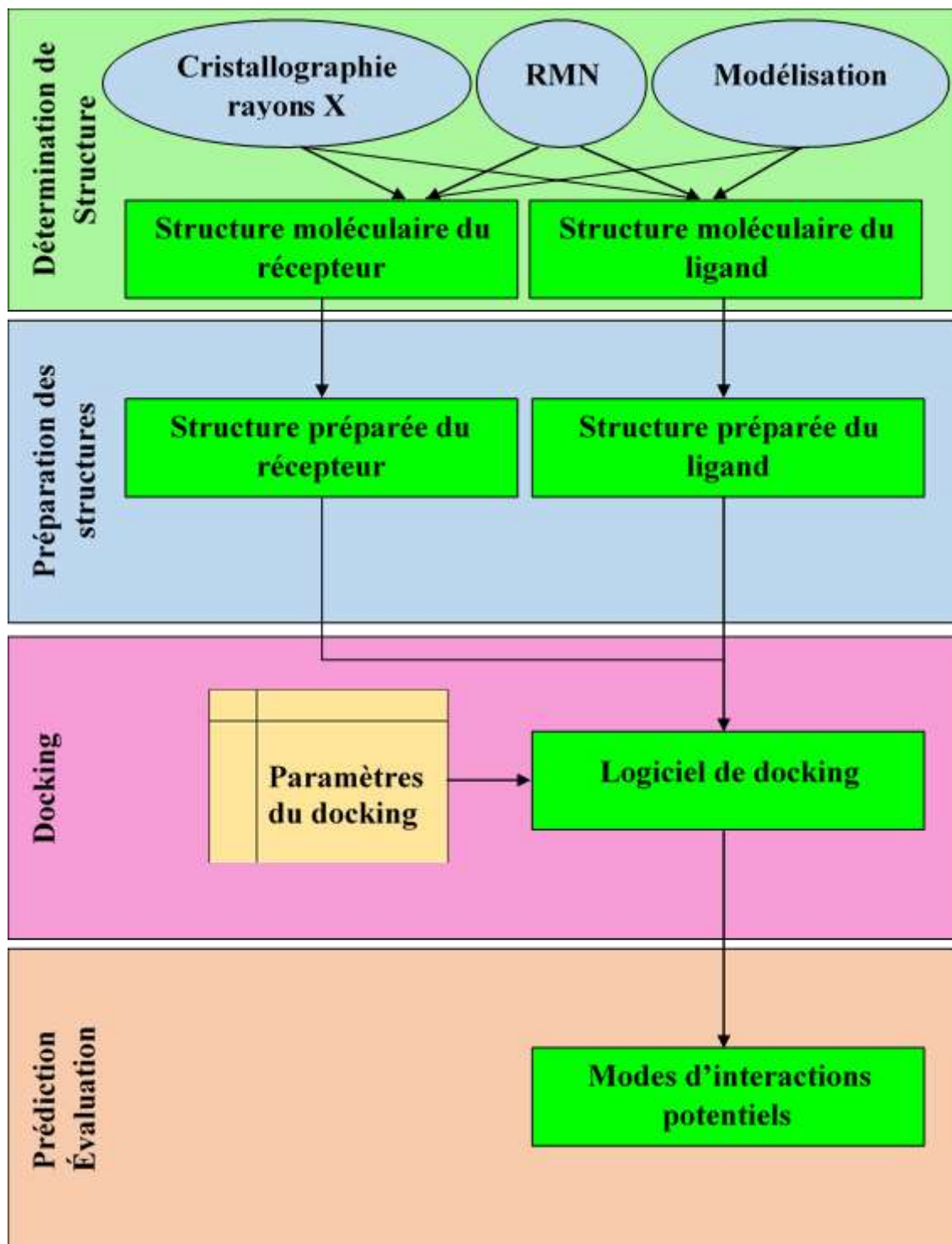


Figure 7. Étapes typiques d'un docking

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

2.5.1 Détermination des structures

La première étape est l'obtention d'une structure pour le récepteur par cristallographie aux rayons X, par RMN ou par modélisation. Ces méthodes expérimentales (**figure 7**) permettent aujourd'hui de déterminer la structure des protéines.

³⁸ La majorité des structures protéiques sont disponibles via la «Protein Data Bank ». ³⁹

2.5.2 Préparation des structures

La préparation du récepteur en vue du docking ne peut être réalisée qu'une fois les structures mises en jeu sont identifiées. Pour cela il faut veiller à résoudre les problèmes de clashes stériques, ainsi que ceux des états de protonation. La conformation initiale du ligand est généralement modifiée et optimisée durant le docking et n'est donc, en principe, pas très importante. ³

2.5.3 Docking moléculaire

Dans le domaine de la modélisation moléculaire, le docking est une méthode qui calcule l'orientation préférée d'une molécule vers une seconde lorsqu'elles sont liées pour former un complexe stable. ⁴⁰

2.5.4 Prédiction et évaluation

Une fois les paramètres du docking moléculaire sont établis, le logiciel de docking propose un ou plusieurs modes d'interactions potentiels qui peuvent être ensuite utilisés.

³

3 Aperçue générale sur les propriétés ADME

Les propriétés Absorption Distribution Métabolisme et Elimination- Toxicité (ADME/Tox) des médicaments présentent des paramètres importants habituellement utilisés pour atteindre les essais cliniques et la sélection des médicaments. ⁴¹⁻⁴²

Dans cette étude nous avons appliqué les règles de base les plus couramment utilisés : les règles de Lipinski ⁴³ et de Veber ⁴⁴. Les conditions empiriques pour satisfaire la règle de

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Lipinski et de manifester une bonne biodisponibilité orale impliquent un équilibre entre la solubilité aqueuse d'un composé et sa capacité à diffuser passivement à travers les différentes barrières biologiques. D'après cette règle, les composés sont plus susceptibles de posséder une mauvaise biodisponibilité orale lorsqu'ils présentent plus d'une violation de cette règle. ⁴⁵⁻⁴⁶

- ❖ Un poids moléculaire (MW) ≤ 500 Da.
- ❖ Log P ≤ 5
- ❖ Nombre de donneurs de liaisons hydrogène (HBD) (OH ou NH) ≤ 5
- ❖ Nombre d'accepteurs de liaisons hydrogène (HBA) (O ou N) ≤ 10

Les règles de Veber suggèrent que la flexibilité moléculaire et la surface polaire (TPSA) sont des déterminants importants de la biodisponibilité orale. ⁴⁴ Il est bien connu qu'une biodisponibilité orale élevée est un facteur important pour le développement de molécules bioactives en tant qu'agents thérapeutiques. La réduction de la flexibilité moléculaire (mesurée par le nombre de liaisons rotatives n-ROTB) et la faible surface polaire sont des prédicateurs importants de la bonne biodisponibilité orale. ⁴⁷

4 Matériels et Méthodes

4.1 Préparation du récepteur (ID : 3ERT)

Le téléchargement de l'enzyme a été fait à partir de la base de données Bookhaven Protein Data Bank (www.rcsb.org/pdb) ³⁹ :

Le récepteur aux estrogènes RE α (ID : 3ERT) ⁴⁸ est Co-cristallisé avec l'inhibiteur 4-Hydroxytamoxifen OHT (C₂₆H₂₉NO₂). Le RE α cristallise sous forme de monomère, avec 247 résidus et 1962 atomes (**figure 8**).

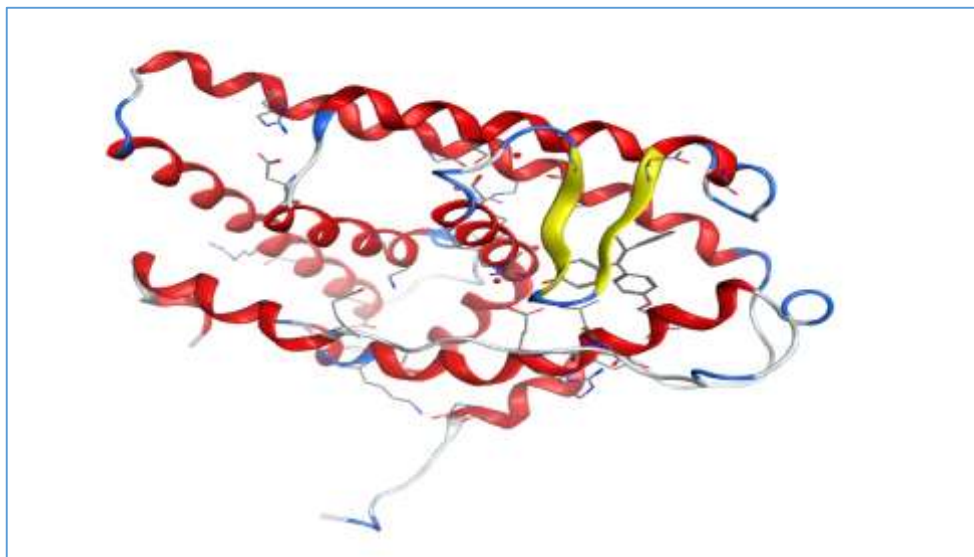


Figure 8. Structure du récepteur aux estrogènes de type α (ID : 3ERT)

La structure tridimensionnelle de l'enzyme a été obtenue par diffraction en rayon X. Les données de la Co-cristallisation sont résumées dans le **tableau 1**.

Tableau 1: Données cristallographiques.

Enzyme	Inhibiteur de Co-cristallisation	Classification	Polymère	Chaîne	Résolution (Å)
RE α (ID : 3ERT)		Récepteur nucléaire	1	A	1.9

4.2 Préparation des structures de dérivés de l'indazole

Les composés les plus actifs ayant de meilleures valeurs d'activité d'IC₅₀ dans la gamme de 0.3 - 2.5 nM dans l'ensemble de données⁴⁹ ont été utilisés dans cette étude de docking moléculaire (**figure 9**).

Tous les composés utilisés ont été dessinés et optimisés par le logiciel HyperChem 8.08 en utilisant à la fois la mécanique moléculaire, champ de force MM+ et la méthode semi-empirique AM1. Après cela, les composés ont été optimisés par la méthode DFT en utilisant la base B3LYP (6-311G) implémentée dans le programme Gaussian 09. Les molécules ainsi obtenues sont enregistrées au format pdb ou mol 2.

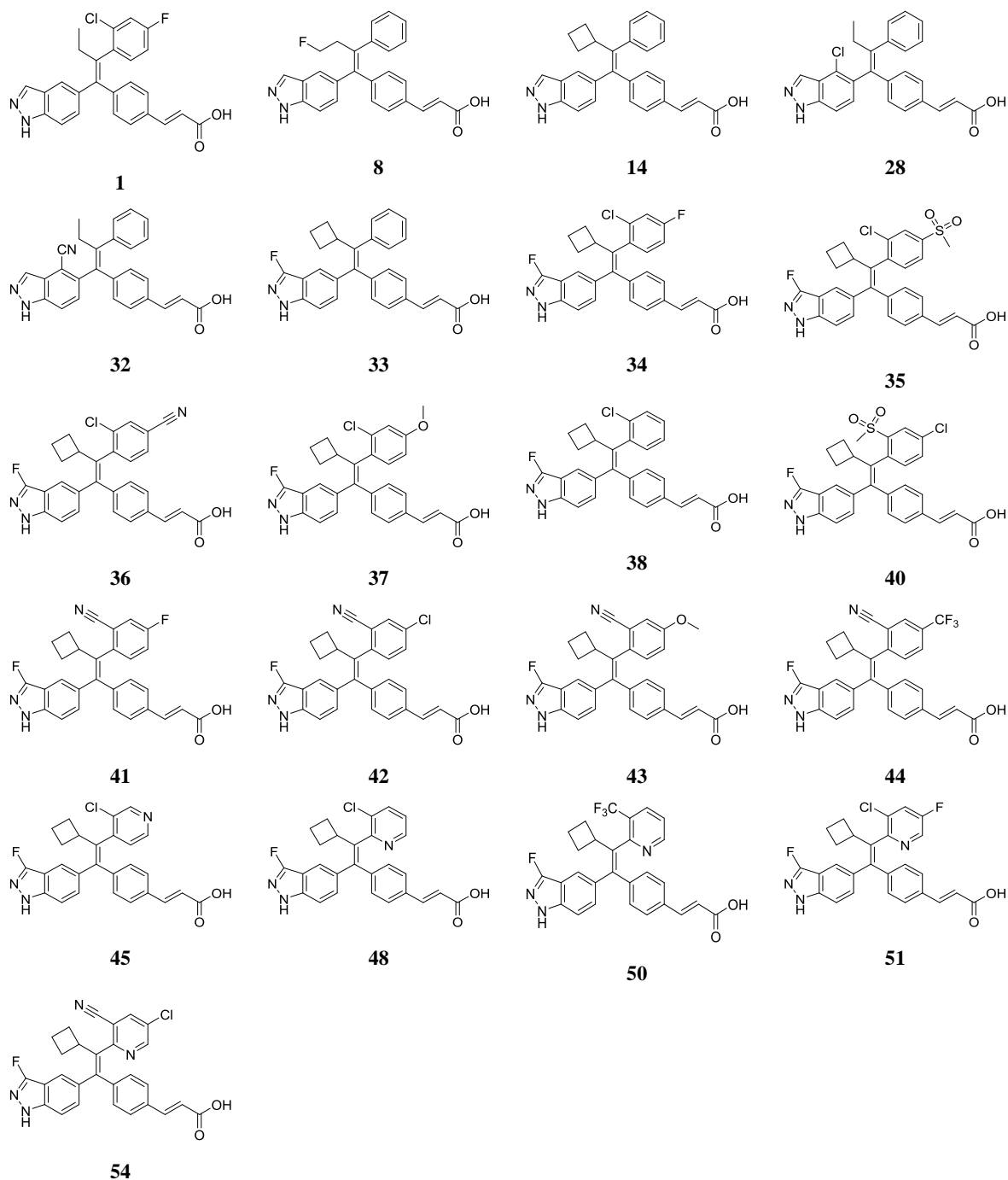


Figure 9. Dérivés de l'indazole utilisés dans l'étude de docking moléculaire.

4.3 Etude de docking moléculaire de dérivés de l'indazole

L'étude de l'interaction entre le site actif de l'enzyme de RE α et les inhibiteurs pour former un complexe stable et de mieux comprendre le mécanisme moléculaire de l'inhibition est réalisée à l'aide du logiciel MOE (Molecular Operating Environment).⁵⁰ Les molécules d'eau ont été laissées dans le modèle afin d'assurer un relais entre les ligands et le site actif de récepteur et de créer des réseaux de liaisons hydrogène.⁵¹

4.4 Etude des propriétés ADME de dérivés de l'indazole

Les règles de Lipinski, les deux descripteurs identifiés par Veber (n-ROTB et TPSA) et la propriété pharmacocinétique d'absorption, GI absorption (ABS), ont été calculés en utilisant le calcul de propriété en ligne SwissADME⁵². Ces propriétés sont utilisées pour évaluer la biodisponibilité orale de six dérivés d'indazole les plus performants obtenus par l'étude de docking moléculaire.

5 Résultats et discussions

5.1 Résultats de docking moléculaire

Le docking moléculaire est une étape importante et nécessaire dans la compréhension des réactions biologiques et dans la conception de médicaments. Dans cette étude, il a été utilisé pour étudier et identifier les interactions entre les différents inhibiteurs et la protéine (**ID : 3ERT**). Ce dernier était considéré comme rigide, tandis que les dérivés de l'indazole étaient considérés comme flexibles.

Afin d'obtenir de meilleurs sites de liaisons potentiels dans la protéine (**ID : 3ERT**), un maximum de 24 cavités ont été détectés en utilisant le module « Site Finder » de MOE qui comprend un outil pour la détection de la cavité enzymatique, ainsi que le site le plus favorable « en couleur bleue ». ^{3, 53} Nous avons choisi la cavité 1 (**figure 10**) pour faire notre étude car elle possède le volume le plus important et le ligand qui est Co-cristallisé. Les résidus de la première cavité détectée par le logiciel MOE ont été donnés dans le **tableau 2**.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

⁵⁴ les interactions comprises entre 2.5 Å et 3.1 Å sont considérées comme fortes, celles comprises entre 3.1 Å et 3.55 Å sont supposées moyennes et les interactions supérieures à 3.55 Å sont faibles.

Tableau 3: Scores et interactions entre les dérivés de l'indazole sélectionnés et les résidus du site actif de la protéine (ID : 3ERT).

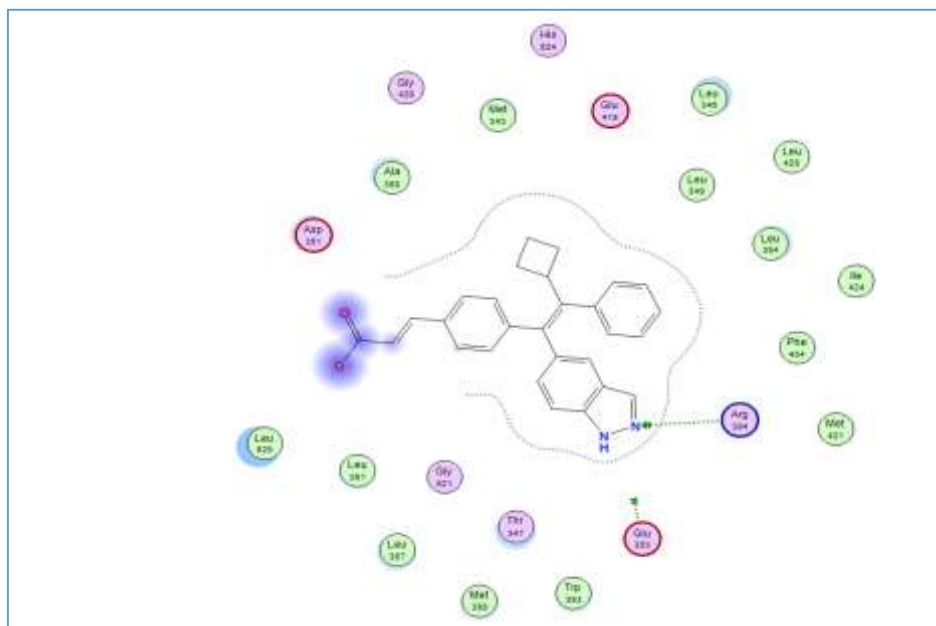
Ligand	IC ₅₀ (nM)	S-score (Kcal/mol)	Liaisons entre les atomes de ligands et les résidus du site actif					
			Atome de ligand	Atome de récepteur	Résidus impliqués	Type interaction	Distance (Å)	Energies (kcal/mol)
1	2.5	-5.4957	N1 1	O	GLY 521	H-donor	3.23	-1.2
8	2.1	-6.9592	N2 2	NH2	ARG 394	H-acceptor	2.87	-2.4
14	1.4	-7.4142	N2 2	NH2	ARG 394	H-acceptor	2.90	-2.6
28	1.6	-6.2294	N2 2	NH2	ARG 394	H-acceptor	2.66	-1.4
			6-ring	CA	THR 347	pi-H	4.46	-0.6
32	1.5	-5.8466	N2 2	NH2	ARG 394	H-acceptor	2.63	-0.8
33	1.4	-6.5222	N1 1	O	GLY 521	H-donor	3.23	-2.1
34	2.0	-6.0913	N1 1	O	HOH 20	H-donor	3.40	-0.6
35	2.3	-7.2145	O21 21	OH	TYR 526	H-acceptor	3.02	-1.9
			6-ring	CB	LEU 525	pi-H	4.46	-0.7
			6-ring	CD1	LEU 525	pi-H	4.30	-0.7
36	0.7	-6.5414	/	/	/	/	/	/
		-6.5410	/	/	/	/	/	/
37	2.4	-6.3147	N1 1	O	HOH 20	H-donor	3.39	-0.7
38	1.1	-5.8776	N1 1	O	GLY 521	H-donor	3.19	-2.0
40	0.3	-6.5721	/	/	/	/	/	/
		-6.5635	/	/	/	/	/	/
41	1.2	-6.0827	N35 35	CA	LYS 529	H-acceptor	3.72	-0.7
42	0.7	-6.6114	O20 20	SD	MET 343	H-donor	3.46	-0.2
			N35 35	N	CYS 530	H-acceptor	2.88	-4.8
			6-ring	N	LEU 536	pi-H	4.69	-0.7
43	0.5	-7.1970	N35 35	N	CYS 530	H-acceptor	2.95	-4.8
			6-ring	N	LEU 536	pi-H	4.42	-0.6
44	1.2	-6.6486	N1 1	O	HOH 73	H-donor	3.11	0.4
45	0.8	-6.7885	N2 2	NH2	ARG 394	H-acceptor	2.99	-2.2

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

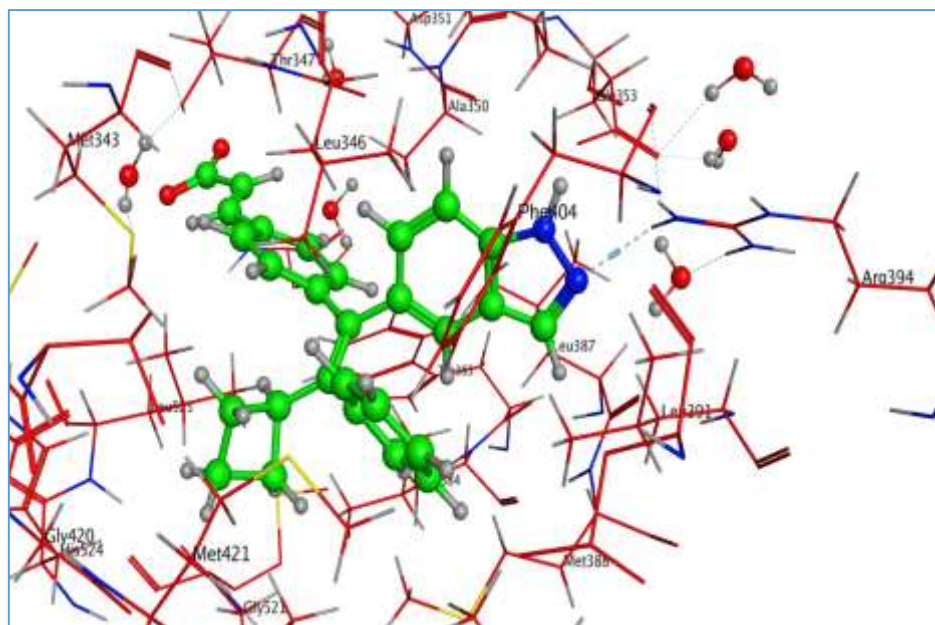
48	1.2	-6.6227	N1 1	O	GLY 521	H-donor	3.32	-1.1
50	1.1	-6.5070	6-ring	CA	THR 347	pi-H	4.74	-0.6
51	1.7	-6.5202	N1 1	O	GLY 521	H-donor	3.26	-1.7
			C30 30	SD	MET 421	H-donor	3.68	-0.9
54	1.1	-6.2656	N35 35	CA	LYS 529	H-acceptor	3.64	-0.6
			N35 35	N	CYS 530	H-acceptor	3.33	-1.1
4-Hydroxy tamoxifen	/	-10.8988	O4 27	OE2	GLU 353	H-donor	2.42	1.4
			N24 46	OD1	ASP 351	ionic	3.82	-0.9

D'après les résultats obtenus dans le **tableau 3**, on remarque que les scores de la plupart des composés sont compris entre -5.4957 et -7.4142 Kcal/mol. On observe aussi, sur ce tableau, que les six composés **14**, **35**, **43**, **8**, **45** et **44** forment des complexes qui possèdent les énergies les plus faibles par rapport aux autres complexes ce qui leur confèrent les conformations les plus stables. Ainsi les résultats obtenus pour ces six composés sont comme suit :

- ✓ Le complexe formé par le composé **14** (**figure 11**) a une faible énergie de Score (-7.4142 Kcal/mol). Ce composé a le meilleur score de docking moléculaire par rapport aux autres composés. Une forte liaison hydrogène a été observée entre le N₂ de ce composé et le groupe NH₂ du résidu **ARG 394** avec une distance de 2.9 Å.



(a)



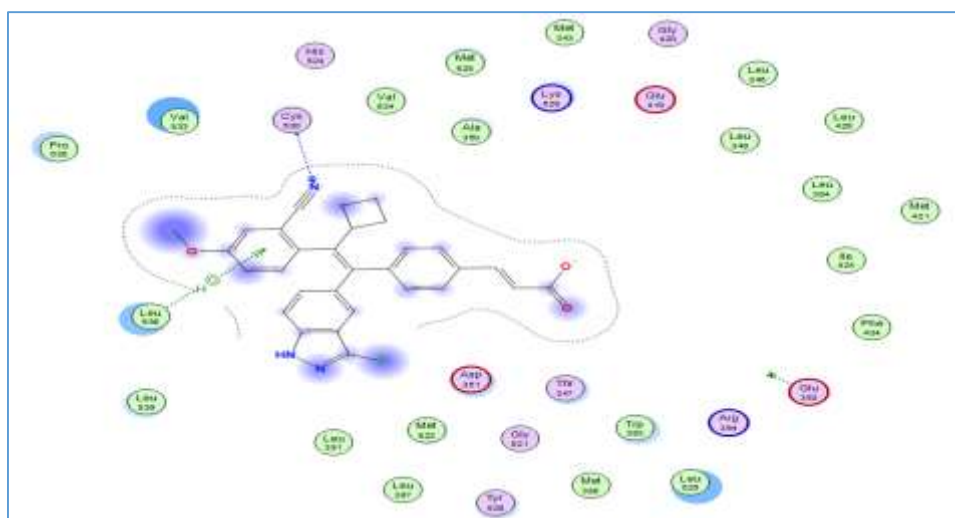
(b)

Figure 11. Interactions entre le ligand 14 et les résidus de site actif de la protéine (ID : 3ERT) (Images (a) 2D et (b) 3D).

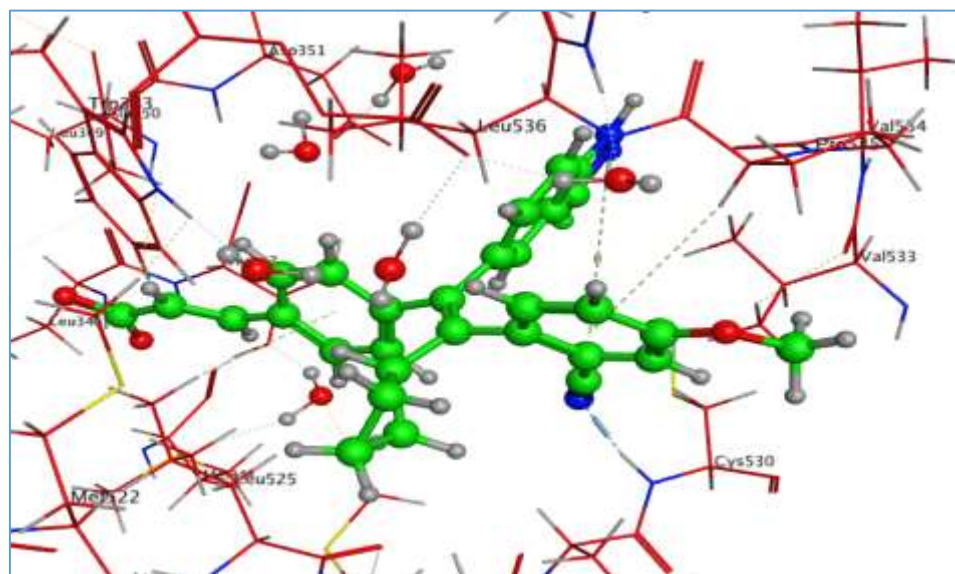
- ✓ Le composé **35** (figure 12) forme trois interactions : une forte liaison hydrogène avec une distance 3.02 Å entre O21 et le groupe hydroxyle OH du résidu **TYR 526**, une

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

- ✓ Les composés **43** et **8** ont de faibles valeurs énergétiques -7.1970 et -6.9592 kcal/mol respectivement. Ils sont impliqués dans la formation des interactions avec les résidus du site actif de RE α . Le composé **43** (**figure 13**) forme deux interactions avec les résidus **CYS 530** et **LEU 536**, une forte liaison hydrogène de distance de 2.95 Å avec le résidu **CYS 530** et une faible interaction de type pi-H (4.42 Å) avec **LEU 536**. Tandis que le composé **8** (**figure 14**) forme une interaction unique avec le résidu **ARG 394** en formant une forte liaison d'hydrogène de distance de 2.87 Å.



(a)

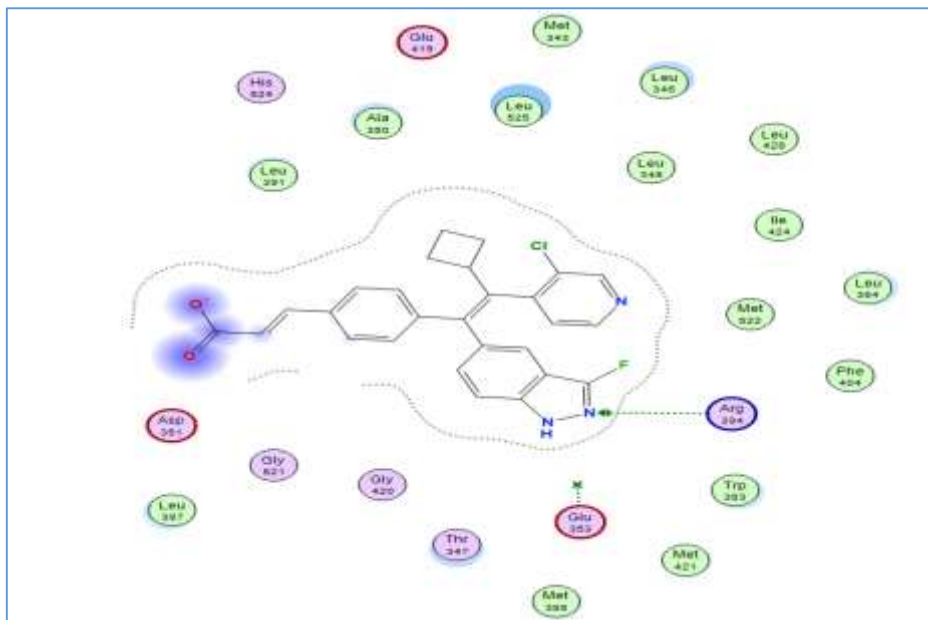


(b)

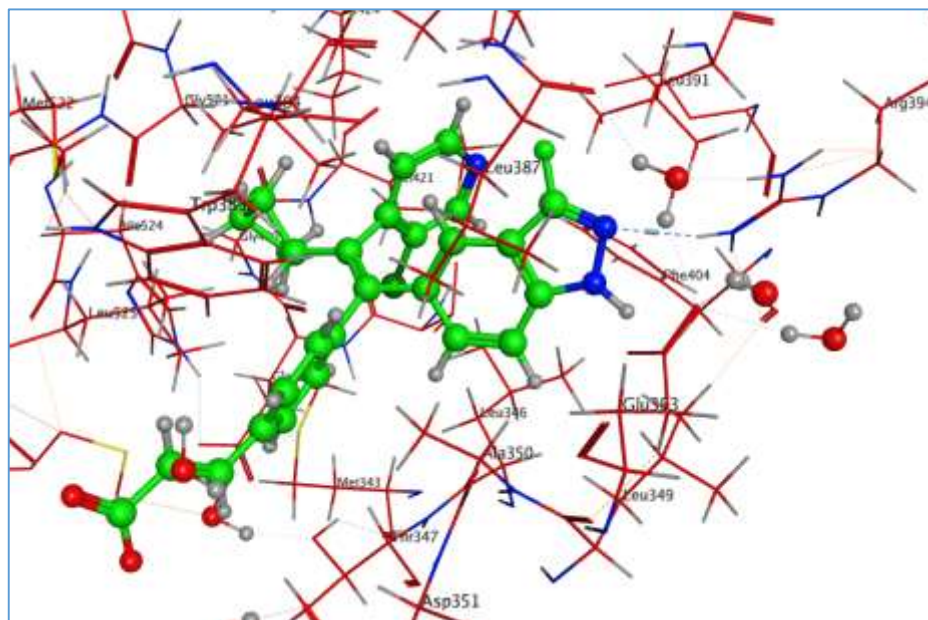
Figure 13. Interactions entre le ligand 43 et les résidus de site actif de la protéine (**ID : 3ERT**). (Images (a) 2D et (b) 3D).

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

forte liaison hydrogène (2.99 Å) avec le résidu **ARG 394**. Alors que le composé **44** (**figure 16**) interagit avec **HOH 73** en formant une liaison hydrogène modérée (3.11 Å).



(a)



(b)

Figure 15. Interactions entre le ligand 45 et les résidus de site actif de la protéine (**ID : 3ERT**). (Images (a) 2D et (b) 3D)

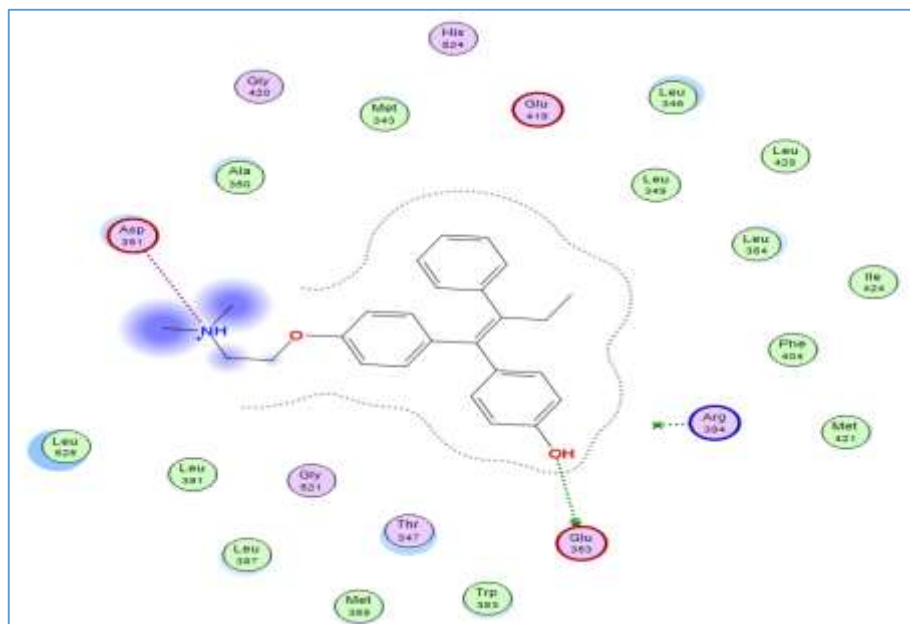


Figure 17. Interactions entre le ligand de référence (4-Hydroxytamoxifène) et les résidus de site actif de la protéine (ID : 3ERT).

La prédiction du mode d'interaction consiste à déterminer le positionnement correct du ligand par rapport à son récepteur. La capacité d'un programme à réussir ce travail est habituellement jugé au moyen de la déviation quadratique moyenne.⁵⁵ Afin de bien analyser les résultats on utilise l'écart quadratique moyen (Root mean square deviation ou RMSD) entre les différentes positions des inhibiteurs au cours du docking et la structure de départ. La prédiction du ligand est acceptable si sa valeur ne dépasse pas 2 angströms.^{20, 56}

Les valeurs de RMSD de 6 ligands obtenus par l'étude de docking moléculaire sont reportées dans le **tableau 4**.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Tableau 4: Résultats de RMSD de docking moléculaire pour les 6 composés possédant de meilleurs scores.

No. Composés	RMSD(Å)
14	1.4323
35	2.3530
43	1.7633
8	1.3547
45	1.3178
44	1.9966
4-Hydroxytamoxifen	0.7758

Selon les résultats du **tableau 4**, on peut dire que les valeurs de RMSD pour presque tous les ligands sont considérées comme acceptables car elles sont inférieures à 2Å. Notons que le composé **35** présente une valeur de RMSD de 2.3530 Å qui est légèrement supérieure à celle du seuil (2 Å). Tous les ligands possèdent une valeur de RMSD supérieure à celle du 4-Hydroxytamoxifen (0.7758 Å). Les ligands **8**, **45** et **14** ont des valeurs de RMSD très proches de celle du ligand de référence, ce qui nous ramène à déduire que ces ligands peuvent inhiber la protéine (**ID : 3 ERT**).

5.2 Evaluation des propriétés ADME in silico de dérivés de l'indazole

Une étude computationnelle des six composés qui ont les meilleures énergies de scores (les affinités les plus élevées avec la cible) a été réalisée pour l'évaluation des propriétés ADME. Le **tableau 5** présent les valeurs obtenues.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Tableau 5: Propriétés ADME pour les six composés ayant les meilleures énergies de scores

No. Composés	ABS	TPSA (Å ²)	n-ROTB	MW <500	MLog ≤5	n-ON accepteurs (HBA) < 10	n-OHND donneurs (HBD) < 5	violations de Lipinski ≤1
14	Haute	65.98	6	420.50	4.72	3	2	0
35	Faible	108.50	7	551.03	5.31	6	2	2
43	Faible	99.00	7	493.53	4.30	6	2	0
8	Haute	65.98	7	412.46	4.42	4	2	0
45	Faible	78.87	6	473.93	4.79	5	2	0
44	Faible	89.77	7	531.50	5.39	8	2	2

Les résultats de **tableau 5** ont révélé que :

- ✓ Seuls les composés **14** et **8** qui ont une haute absorption.
- ✓ Les composés **14**, **43**, **8** et **45** sont conformes aux règles de Lipinski de cinq : une masse moléculaire inférieure à 500 DA, le coefficient de partition log P est inférieur à cinq et le nombre de liaisons hydrogènes dans l'intervalle approprié de Lipinski (HBD <5 et HBA <10), suggérant que ces composés théoriquement n'auraient pas de problème avec la biodisponibilité orale.
- ✓ Les composés **35** et **44** ont deux violations de Lipinski (des poids moléculaires supérieurs à 500 DA et des coefficients de partitions supérieurs à 5).
- ✓ Tous les composés sélectionnés sont en accord aux règles de Veber avec des valeurs de PSA dans la gamme de 65-110 Å, ce qui leur permet de très bien se corréler avec l'absorption intestinale humaine. De plus, ces composés sont flexibles avec un certain nombre de liaisons rotatives (n-ROTB) dans la gamme de 6-7.

Les dérivés de l'indazole les plus actifs contre l'activité anticancéreuse étaient les composés **43** et **45**. Ces composés contiennent du nitrile et de la chloro pyridine au niveau du cycle aryle, ce qui les rend très puissants. Le remplacement du *para*-F (composé **34**) par du SO₂Me polaire dans le composé **35** et par le CF₃ lipophile dans le composé **44** diminue l'exposition orale.

6 Conclusion

L'étude de docking moléculaire a été effectuée à partir du logiciel MOE pour former les complexes ligand-récepteur et identifier les interactions existantes. Les résultats de cette étude ont montré que les composés **14**, **35**, **43**, **8**, **45** et **44** forment des complexes qui possèdent les énergies les plus basses par rapport aux autres complexes. Parmi ces ligands, nous avons trouvé que les composés **8**, **45** et **14** ont les valeurs de RMSD les plus proches de celle du ligand de référence. Ce qui nous ramène à déduire que ces ligands peuvent inhiber la protéine (**ID : 3 ERT**).

Le calcul des propriétés ADME montre que les composés **14**, **43**, **8** et **45** répondent aux règles de Lipinski et de Veber, ce qui suggère que ces composés n'auront pas, théoriquement, de problème avec la biodisponibilité orale. Les dérivés de l'indazole les plus actifs contre l'activité anticancéreuse étaient les composés **43** et **45**. Ces composés contiennent du nitrile et de la chloro pyridine au niveau du cycle aryle, ce qui les rend très puissants.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

Références bibliographiques

1. Asses, Y. Conception par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met. Thèse de doctorat, Université Henri Poincaré – Nancy I, **2011**.
2. Geldenhuys, W. J.; Gaasch, K. E.; Watson, M.; Allen, D. D.; Van der Schyf, C. J., Optimizing the use of open-source software applications in drug discovery. *Drug Discovery Today* **2006**, *11* (3-4), 127-132.
3. Grosdidier, A. Conception d'un logiciel de docking et applications dans la recherche de nouvelles molécules actives. Thèse de doctorat, Université Joseph Fourier, **2007**.
4. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* **2006**, *49* (20), 5912-5931.
5. Bach, J.-F., *La chimie et la santé au service de l'homme*. EDP sciences: **2012**.
6. Barril, X.; Soliva, R., Molecular modelling. *Molecular BioSystems* **2006**, *2* (12), 660-681.
7. Lagarde, N. Méthodes de criblage virtuel in silico: importance de l'évaluation et application à la recherche de nouveaux inhibiteurs de l'interleukine 6. Thèse de doctorat, Paris, CNAM, **2014**.
8. Dias, R.; de Azevedo Jr., W. F., Molecular docking algorithms. *Current Drug Targets* **2008**, *9* (12), 1040-1047.
9. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **2004**, *3* (11), 935-949.
10. FeddaL-Bouchebaba, S. Etude Structurale et Conception in silico de Molécules présentant un profil Anti-inflammatoire. Thèse de doctorat, Université des Sciences et de la Technologie Houari Boumediene, **2018**.
11. Lichtenthaler, F. W., 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? *Angewandte Chemie International Edition in English* **1995**, *33*, 2364-2374.
12. Vuorinen, A.; Schuster, D., Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* **2015**, *71*, 113-134.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

13. Chevrollier, N. Développement et application d'une approche de docking par fragments pour modéliser les interactions entre protéines et ARN simple-brin. Thèse de doctorat, Université Paris-Saclay, **2019**.
14. Alvarez, J. C., High-throughput docking as a source of novel drug leads. *Current opinion in chemical biology* **2004**, 8 (4), 365-370.
15. Sotriffer, C.; Klebe, G.; Stahl, M.; Böhm, H. J., Docking and scoring functions/virtual screening. *Burger's Medicinal Chemistry and Drug Discovery* **2003**, 281-331.
16. Coupez, B.; Lewis, R., Docking and scoring-Theoretically easy, practically impossible? *Current Medicinal Chemistry* **2006**, 13 (25), 2995-3003.
17. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E., Prediction of protein– ligand interactions. Docking and scoring: successes and gaps. *Journal of Medicinal Chemistry* **2006**, 49 (20), 5851-5855.
18. Hawkins, P. C.; Warren, G. L.; Skillman, A. G.; Nicholls, A., How to do an evaluation: pitfalls and traps. *Journal of Computer-Aided Molecular Design* **2008**, 22 (3-4), 179-190.
19. Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks III, C. L., Assessing energy functions for flexible docking. *Journal of Computational Chemistry* **1998**, 19 (14), 1612-1622.
20. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J., Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology* **1997**, 272 (1), 106-120.
21. Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T., Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *Journal of Computer-Aided Molecular Design* **2008**, 22 (3-4), 213-228.
22. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research* **2000**, 33 (12), 889-897.
23. Brut, M. Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques: les Modes Statiques. Thèse de doctorat, Université Toulouse III - Paul Sabatier, **2009**.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

24. Kollman, P., Free energy calculations: applications to chemical and biochemical phenomena. *Chemical Reviews* **1993**, 93 (7), 2395-2417.
25. Simonson, T.; Archontis, G.; Karplus, M., Free energy simulations come of age: protein– ligand recognition. *Accounts of Chemical Research* **2002**, 35 (6), 430-437.
26. Gohlke, H.; Klebe, G., Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie International Edition* **2002**, 41 (15), 2644-2676.
27. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., Assessing scoring functions for protein– ligand interactions. *Journal of Medicinal Chemistry* **2004**, 47 (12), 3032-3047.
28. Wang, R.; Lu, Y.; Fang, X.; Wang, S., An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein– ligand complexes. *Journal of Chemical Information and Computer Sciences* **2004**, 44 (6), 2114-2125.
29. Arrault, A. Stratégies de docking-scoring assistées par analyse de données. Application au criblage virtuel des cibles thérapeutiques COX-2 et PPAR gamma. Thèse de doctorat, Université d'Orleans, **2007**.
30. Maréchal, Y., *The hydrogen bond and the water molecule: The physics and chemistry of water, aqueous and bio-media*. Elsevier: **2006**.
31. Bouchagra, S. Modélisation des interactions protéine-petites molécules: étude de la relation structure-fonction dans le cas des lipases. Thèse de doctorat, Université Badji Mokhtar-Annaba, **2018**.
32. Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A., Covalency of the hydrogen bond in ice: A direct X-ray measurement. *Physical Review Letters* **1999**, 82 (3), 600.
33. Gerschel, A., *Liaisons intermoléculaires-Les forces en jeu dans la matière condensée*. 2^{ème} ed.; EDP Sciences: Paris, **2012**.
34. Schulz, G. E.; Schirmer, R. H., *Principles of protein structure*. s.l. : Springer-Verlag: **1978**.
35. Lin, M.-S.; Chen, L.-Y.; Tsai, H.-T.; Wang, S. S.-S.; Chang, Y.; Higuchi, A.; Chen, W.-Y., Investigation of the mechanism of β -amyloid fibril formation by kinetic and thermodynamic analyses. *Langmuir* **2008**, 24 (11), 5802-5808.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

36. Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W., Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **1992**, 255 (5041), 178-183.
37. Israelachvili, J., *Intermolecular and Surface Forces*. Academic Press 2^{ème} ed.; Londres, **1997**.
38. Howard, G. C.; Brown, W. E., *Modern protein chemistry: practical aspects*. CRC press: **2001**; p 272.
39. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic Acids Research* **2000**, 28 (1), 235-242.
40. Lengauer, T.; Rarey, M., Computational methods for biomolecular docking. *Current Opinion in Structural Biology* **1996**, 6 (3), 402-406.
41. Li, A. P., Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today* **2001**, 6 (7), 357-366.
42. Ekins, S.; Rose, J., In silico ADME/Tox: the state of the art. *Journal of Molecular Graphics and Modelling* **2002**, 20 (4), 305-309.
43. 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. *Advanced Drug Delivery Reviews* **2012**, 64, 4-17.
44. Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* **2002**, 45 (12), 2615-2623.
45. 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. *Advanced Drug Delivery Reviews* **1997**, 23 (1-3), 3-25.
46. Vistoli, G.; Pedretti, A.; Testa, B., Assessing drug-likeness—what are we missing? *Drug Discovery Today* **2008**, 13 (7-8), 285-294.
47. Remko, M.; Swart, M.; Bickelhaupt, F. M., Theoretical study of structure, pKa, lipophilicity, solubility, absorption, and polar surface area of some centrally acting antihypertensives. *Bioorganic & Medicinal Chemistry* **2006**, 14 (6), 1715-1728.
48. Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L., The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, 95 (7), 927-937.

Chapitre III Étude de docking moléculaire d'une série de dérivés de l'indazole.

49. Lai, A.; Kahraman, M.; Govek, S.; Nagasawa, J.; Bonnefous, C.; Julien, J.; Douglas, K.; Sensintaffar, J.; Lu, N.; Lee, K.-j.; Aparicio, A.; Kaufman, J.; Qian, J.; Shao, G.; Prudente, R.; Moon, M. J.; Joseph, J. D.; Darimont, B.; Brigham, D.; Grillot, K.; Heyman, R.; Rix, P. J.; Hager, J. H.; Smith, N. D., Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. *Journal of Medicinal Chemistry* **2015**, *58* (12), 4888-4904.
50. Molecular Operationg Environment (MOE), 2013.08, chemical Computing Group Inc., 1010 Sherbooke St, West, Suite 910., Montreal, QC, Canada, H3A2R77, (2014).
51. Klebe, G., Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11* (13-14), 580-594.
52. Daina, A.; Michielin, O.; Zoete, V., SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7*, 42717.
53. Yamaguchi, H.; Akitaya, T.; Yu, T.; Kidachi, Y.; Kamiie, K.; Noshita, T.; Umetsu, H.; Ryoyama, K., Homology modeling and structural analysis of 11 β -hydroxysteroid dehydrogenase type 2. *European Journal of Medicinal Chemistry* **2011**, *46* (4), 1325-1330.
54. Imberty, A.; Hardman, K. D.; Carver, J. P.; Perez, S., Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. *Glycobiology* **1991**, *1* (6), 631-642.
55. Bouchareb, L. F. Etude par modélisation moléculaire des mécanismes de complexation. Thèse de doctorat, Université Abou Bekr Belkaid de Tlemcen, **2016**.
56. Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004**, *303* (5665), 1813-1818.

Chapitre IV

Étude de la réactivité chimique d'une série de dérivés de l'indazole.

1 Introduction

L'étude de la réactivité chimique a fait l'ambition de plusieurs chimistes expérimentateurs et théoriciens, pour cela ils ont utilisés différentes théories et des recherches afin de répondre et de traiter les questions principales suivantes :

Pourquoi certaines molécules sont stables tandis que d'autres sont réactives ? Est-il possible de prédire la sélectivité d'une réaction chimique ?

Le besoin pour les chimistes d'avoir à disposition des outils reliant structure électronique et réactivité chimique a conduit à la mise au point de plusieurs modèles théoriques de réactivité. Certains sont basés sur la théorie des orbitales moléculaires, telle que la célèbre Théorie des Orbitales Frontières (FMO pour « Frontier Molecular Orbital ») initiée par K. Fukui en 1952.¹⁻² D'autres modèles de réactivité se fondent plutôt sur la DFT, donc sur la densité électronique comme quantité principale. Ces derniers font partie de ce qu'on appelle la DFT conceptuelle (CDFT, pour « Conceptual Density Functionnal Theory »).³

La DFT conceptuelle, initiée par Parr et Pearson⁴, peut être définie comme étant une théorie interprétative dérivée de la DFT calculatoire. Elle vise à extraire de la densité électronique une quantité d'informations « chimiques » afin de mieux comprendre et prédire la réactivité des molécules.

Dans cette partie de travail, nous avons utilisé les descripteurs globaux de réactivité issus de la DFT conceptuelle sur un ensemble de données contenant **21** molécules de dérivés de l'indazole ayant les meilleures activités. Notre objectif est de bien comprendre la relation entre la structure, la stabilité et la réactivité chimique des composés étudiés.

2 Aperçue générale sur la réactivité chimique

2.1 Théorie des orbitales moléculaires frontières

La théorie des orbitales moléculaires frontières FMO (Frontier Molecular Orbital)⁵⁻⁸ a été développée par K. Fukui¹ afin d'expliquer la régiosélectivité observée lors de

réactions mettant en jeu des composés aromatiques. Son champ d'application était initialement établi pour l'étude des systèmes π , puis a été étendu à l'étude des systèmes σ . L'idée principale de Fukui est que dans une réaction chimique à contrôle frontalier, seules deux orbitales moléculaires présentent un réel intérêt : la plus haute orbitale moléculaire occupée (HOMO) et la plus basse orbitale moléculaire vacante (LUMO). Sa théorie est fondée sur les approximations suivantes :

- ❖ Les interactions entre OM occupées sont négligées.
- ❖ Les interactions entre OM occupée et OM vacante, à l'exception des interactions frontalières HOMO-LUMO, sont négligées.
- ❖ Le système est correctement décrit par une seule configuration électronique.
- ❖ Les orbitales frontières utilisées dans le traitement standard sont celles des réactifs de départ.

2.2 DFT conceptuelle

La DFT classique est utilisée pour la détermination des propriétés électroniques d'un système, c'est une méthode de calcul qui s'est imposée comme une méthode crédible aux méthodes quantique ab initio. Mais il existe un autre aspect qui n'a pas été mentionné. La DFT est aussi pratique pour la détermination de concept. Plusieurs concepts chimiques de base (dureté, électronégativité) apparaissent naturellement. Ainsi la DFT est utile non seulement pour leur calcul mais aussi de les mieux comprendre cet aspect de la DFT est la cause d'une nouvelle branche qu'on a nommé DFT conceptuelle.⁹

La DFT conceptuelle¹⁰⁻¹¹, développée à partir des années 1980, permet de calculer des propriétés liées à la réactivité. Elle a pour but de donner un cadre théorique et des définitions précises à des concepts largement utilisés par les chimistes expérimentateurs.

2.3 Fondements de la DFT

Hohenberg et Kohn¹² ont formulé et démontré deux théorèmes qui ont donné les bases mathématiques solides pour le développement des méthodes DFT modernes. Leur contribution a consisté à formuler la DFT comme une théorie exacte pour des systèmes à N particules.

2.3.1 Premier théorème de Hohenberg-Kohn

Le premier théorème de Hohenberg et Kohn énonce que toute observable d'un état fondamental stationnaire non-dégénéré peut être calculée, exactement en théorie, par le biais de la densité électronique de cet état fondamental. En effet, l'Hamiltonien d'un système ne dépend que du nombre d'électrons considéré et du potentiel extérieur V_{ne} , qui représente l'interaction entre les électrons et les noyaux. Il est possible de démontrer qu'à une densité $\rho(\mathbf{r})$ donnée correspond un unique potentiel extérieur (à une constante près). La connaissance de la densité entraîne ainsi celle du potentiel, qui entraîne celle du Hamiltonien, et donc des valeurs moyennes associées.

Ce théorème est valide pour les cas où la densité est N-représentable, c'est-à-dire quand l'intégration de la densité sur la coordonnée spatiale restante permet de retrouver le nombre d'électrons N et quand la densité est non-négative, et V-représentable (il existe un potentiel extérieur à partir duquel la densité peut être dérivée).

Dans les conditions de validité de ce théorème, la fonctionnelle d'énergie s'exprime :

$$E[\rho] = T[\rho] + V_{ee}[\rho] + v_{\text{ext}}[\rho] \quad (1)$$

Avec $T[\rho]$ la fonctionnelle d'énergie cinétique, $V_{ee}[\rho]$ le potentiel d'interaction et $v_{\text{ext}}[\rho]$ le potentiel externe.

Les deux premières quantités sont rassemblées en une fonctionnelle universelle F_{HK} , indépendante du potentiel extérieur :

$$E[\rho] = \int \rho(\mathbf{r}_1) v_{\text{ext}}(\mathbf{r}_1) d\mathbf{r}_1 + F_{HK}[\rho] \quad (2)$$

2.3.2 Deuxième théorème de Hohenberg-Kohn

Le second théorème de Hohenberg et Kohn énonce que l'énergie d'un état fondamental non dégénéré peut être déterminée, exactement en théorie, par la densité minimisant l'énergie de l'état fondamental. C'est-à-dire que pour la densité ρ , N-représentable et V-représentable, on a :

$$E_0 \leq E[\rho] \quad (3)$$

Ainsi, pour obtenir l'énergie de l'état fondamental, on va chercher à minimiser la fonctionnelle énergie :

$$\left[\frac{\delta E[\rho]}{\delta \rho} \right] = 0 \quad (4)$$

La minimisation de l'énergie passe par l'intermédiaire de l'utilisation du formalisme de Lagrange, c'est-à-dire par l'établissement d'un Lagrangien avec comme restriction la N-représentabilité de la densité qui doit remplir la condition (5) :

$$\int \rho(r_1) dr_1 = N \quad (5)$$

$$L[\rho] = E[\rho] - \mu \left[\int \rho(r) dr - N \right] \quad (6)$$

Où μ est le multiplicateur de Lagrange inconnu. La minimisation du Lagrangien (6) implique :

$$\frac{\delta}{\delta \rho(r)} \{ E[\rho] - \mu \left[\int \rho(r) dr - N \right] \} = 0 \quad (7)$$

Comme la différentielle d'une fonctionnelle s'exprime sous la forme

$$\delta F = \int \frac{\delta F}{\delta f(x)} \delta f(x) dx \quad (8)$$

2.4 Descripteurs de la DFT Conceptuelle

La théorie de la DFT, utilisée pour la détermination des propriétés électroniques d'un système, a été fondée par Fermi, Dirac, Hohenberg et Kohn.¹³ Ils ont démontré que toutes les propriétés d'un système électronique à l'état fondamental peuvent être déduites de la densité électronique.¹⁴

Dans les années 80, une seconde branche s'est développée dont le but consiste à donner un cadre théorique précis à des concepts souvent mal définis caractérisant la réactivité chimique. Il s'agit du modèle de « DFT conceptuelle », selon l'appellation de R. G.Parr.⁴ La DFT conceptuelle¹¹ donne accès à des descripteurs caractérisant la réactivité des molécules. Elle a été introduite dans l'objectif de définir physiquement des concepts empiriques issus de la chimie, comme l'électronégativité ou la dureté. A travers les dérivées successives de l'énergie dans l'ensemble canonique $[N, v(r)]$ de nombreux

descripteurs de réactivité ont pu être proposés et des significations physiques ont émergé.

¹⁵ La **figure 1** présente comment ils peuvent être calculés.

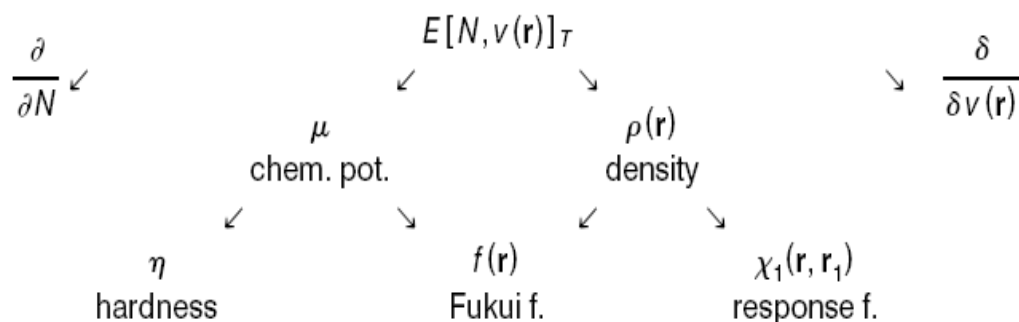


Figure 1. Dérivées de l'énergie aux 1^{er} et 2^{ème} ordres.

Ces dérivées de l'énergie peuvent être classées en trois catégories ^{10, 16} (illustrées en **figure 1**):

- ❖ **Grandeurs globales** : dérivées par rapport à N , le nombre d'électrons, (tel que le potentiel chimique μ et l'électronégativité χ , la dureté η , la mollesse S et l'indice d'électrophilie ω). Ces grandeurs sont indépendantes de l'espace. Elles correspondent à la réponse d'une grandeur globale à une perturbation globale. À ce titre elles peuvent permettre de décrire des concepts chimiques tels que la stabilité ou encore la réactivité globale d'une molécule.
- ❖ **Locales** : dérivées qui dépendent d'une coordonnée spatiale uniquement, (tel que les indices de Fukui, la mollesse locale, l'électrophilie locale et la nucléophilie locale). Ces grandeurs dépendent de la position dans l'espace. Leurs valeurs diffèrent en chaque point du système qu'ils caractérisent. Ces descripteurs de réactivité sont appelés descripteurs locaux. Ils correspondent à la réponse soit d'une grandeur globale à une perturbation locale, soit d'une grandeur locale à une perturbation globale. Ils peuvent renseigner sur la différence de réactivité de deux sites intra- ou intermoléculaire. Ce sont donc des descripteurs adaptés pour décrire la réactivité locale.
- ❖ **Non locales** : dérivées qui dépendent de deux coordonnées spatiales. Ce sont des grandeurs non-locales dont la valeur dépend de deux points de l'espace. Leur

signification physique est difficile à appréhender. Elles caractérisent la réponse d'une grandeur locale à une perturbation locale.

2.4.1 Descripteurs globaux

2.4.1.1 Potentiel chimique

En 1983, Parr a défini le potentiel chimique électronique (μ) comme réponse aux changements de l'énergie du système par rapport au nombre des électrons N à potentiel externe fixe créé par les noyaux.¹⁷⁻¹⁸ Il mesure la tendance du nuage électronique à s'échapper de la molécule. L'électronégativité est l'opposé du potentiel chimique et mesure donc la tendance à ne pas laisser s'échapper les électrons du système. Cette définition de l'électronégativité est celle selon la DFT dont la différence finie amène à retrouver celle de Mulliken.

$$\mu = -\chi = \left(\frac{\partial E}{\partial N} \right)_{V(r)} = -\frac{1}{2} (I + A) \quad (9)$$

Où I est le potentiel d'ionisation et A l'affinité électronique.¹⁹ L'utilisation du théorème de Koopman²⁰⁻²³ et le formalisme de Kohn-Sham²⁴⁻²⁵ dans la DFT permet à ces énergies d'être approchées par celles des orbitales moléculaires frontières HOMO et LUMO (I par $-E_{\text{HOMO}}$ et A par $-E_{\text{LUMO}}$). Par conséquent, le potentiel chimique électronique μ peut être exprimé comme suit :

$$\mu = \frac{(E_{\text{HOMO}} + E_{\text{LUMO}})}{2} \quad (10)$$

2.4.1.2 Dureté chimique

En 1963, Pearson²⁶ a établi une classification des acides et des bases de Lewis en deux catégories : les substances dites « dures » et celles dites « molles ». Il a proposé que dans une réaction acide / base, les interactions les plus favorables aient lieu entre les acides durs et les bases dures ou entre les acides mous et les bases molles.

En 1983, Parr et Pearson ont défini une expression quantitative pour la dureté chimique η , qui peut être exprimée comme la dérivée seconde de l'énergie par rapport au nombre d'électrons N à un potentiel externe fixe $v(r)$:

$$\eta = \left(\frac{\partial \mu}{\partial N} \right)_{v(r)} = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(r)} \quad (11)$$

La dureté η est une valeur positive qui peut être vue comme la résistance d'un système moléculaire au transfert d'électron, il s'agit d'une mesure de la stabilité d'une molécule. L'application de l'approximation des différences finies à l'équation (3) conduit à une expression approximative de la dureté :

$$\eta = I - A \quad (12)$$

Dans le travail original de Parr et Pearson, un facteur $\frac{1}{2}$ a été artificiellement ajouté pour créer une cohérence avec la définition opératoire du potentiel chimique.¹⁷ La dureté peut être calculée à partir des énergies des orbitales moléculaires frontières de la manière suivante :

$$\eta = \frac{E_{LUMO} - E_{HOMO}}{2} \quad (13)$$

2.4.1.3 Mollesse

La mollesse S (Softness) est la propriété inverse de la dureté chimique.²⁷ Elle correspond à la capacité d'un système à conserver une charge acquise.²⁸

$$S = \frac{1}{\eta} \quad (14)$$

2.4.1.4 Indice d'électrophilie globale (Indice d'électrophilicité)

En 1999, Parr a défini l'indice d'électrophilie globale ω ²⁹, qui donne une mesure de l'énergie de stabilisation d'une molécule lorsqu'elle acquiert une quantité supplémentaire de densité électronique, ΔN , de l'environnement. L'indice d'électrophilicité mesure le caractère électrophile d'une molécule, c'est-à-dire sa capacité à attirer les électrons. Il est défini comme la stabilisation énergétique due au transfert de charge¹⁹ et il est donné par l'expression simple suivante :

$$\omega = \frac{\mu^2}{2\eta} \quad (15)$$

L'indice d'électrophilicité est devenu un outil puissant pour l'étude de la réactivité des molécules qui participent aux réactions polaires.³⁰⁻³² Une étude détaillée réalisée, en 2002 par l'équipe de Domingo, sur l'électrophilicité d'une série de réactifs impliqués dans les réactions de Diels-Alder a permis d'établir une échelle d'électrophilie en fonction de la valeur de ω .³³ Cette échelle a permis la classification des molécules organiques comme suit :

- Des électrophiles forts pour $\omega > 1.5$ eV
- Des électrophiles modérés pour $0.8 < \omega < 1.5$ eV
- Des électrophiles faibles pour $\omega < 0.8$ eV.

2.4.1.5 Indice de nucléophilie globale (Indice de nucléophilicité)

En 2008, Domingo et al³⁴ ont proposé un indice de nucléophilie N empirique (relatif) pour les molécules basées sur les énergies HOMO (E_{HOMO}) obtenues par la méthode de Kohn-Sham. La nucléophilie globale est une grandeur qui reflète la capacité d'un site réactif à céder des électrons, elle est définie comme suit :

$$N = E_{\text{HOMO}} (\text{nucléophile}) - E_{\text{HOMO}} (\text{TCT}) \quad (16)$$

On note que l'échelle de la nucléophilie³⁵ est référenciée par rapport à la molécule du Tétracyanoéthylène (TCT), car elle présente la plus basse énergie HOMO dans une grande série de molécules déjà étudiées dans le cadre des cyclo-additions polaires de Diels-Alder. Ce choix permet de gérer facilement une échelle nucléophilique des valeurs positives. L'analyse d'une série de molécules qui participent à des réactions polaires a permis de classer les molécules organiques comme suit:

- Des nucléophiles forts pour $N > 3.0$ eV
- Des molécules dites modérées pour $2.0 < N < 3.0$ eV
- Des nucléophiles faibles pour $N < 2.0$ eV.

2.4.2 Gap énergétique

Le Gap énergétique, ou l'écart HOMO-LUMO, noté E_{gap} , traduit l'énergie comprise entre l'orbitale moléculaire la plus haute occupée et celle la plus basse vacante, c'est un indice de stabilité important qui caractérise l'activité chimique d'une molécule.

Cette différence d'énergie sert à mesurer l'excitabilité d'une molécule. Ainsi, plus l'intervalle d'énergie sera faible et plus la molécule pourra interagir avec l'environnement.

Un grand écart HOMO-LUMO implique une grande stabilité pour la molécule dans le sens de sa faible réactivité dans les réactions chimiques, et de même, un faible écart implique une grande réactivité de la molécule.³⁶⁻³⁷

2.4.3 Énergie totale

Pour une molécule isolée à l'état fondamental, l'énergie totale calculée, notée E_t , peut être utilisée comme descripteur moléculaire quantique. Cette énergie approximative a été calculée pour une conformation optimisée de la géométrie la plus stable dont la structure d'énergie est minimale.³⁸

3 Matériels et Méthodes

Les descripteurs globaux de la réactivité dérivant de la DFT conceptuelle, sont utilisés pour comprendre la relation entre la structure, la stabilité et la réactivité chimique globale ainsi que pour le développement de QSAR.^{11, 35, 39} Cette étude a été réalisée sur un sous-ensemble de composés qui ont montré une meilleure activité anticancéreuse à partir de l'ensemble de données⁴⁰ (**tableau 1**).

Les géométries des molécules ont été optimisées au niveau de calcul B3LYP/ 6-311 G en utilisant le programme Gaussian 09. Parmi les indices globaux utilisés: le potentiel chimique μ (équation (10)), la dureté chimique (équation (13)), l'indice d'électrophilie globale ω (équation(15)), l'indice de nucléophilie N (équation(16)), le gap énergétique ΔE ($\Delta E = E_{LUMO} - E_{HOMO}$) et l'énergie totale.

Tableau 1: Formats SMILES des composés utilisés dans l'étude des grandeurs globales de la réactivité.

Composé	IC ₅₀ (nM)	SMILES
1	2.5	<chem>CC/C(=C(\c1ccc2c(c1)cn[nH]2)/c1ccc(cc1)/C=C/C(=O)O)/c1ccc(cc1Cl)F</chem>
8	2.1	<chem>FCC/C(=C(\c1ccc2c(c1)cn[nH]2)/c1ccc(cc1)/C=C/C(=O)O)/c1ccccc1</chem>
14	1.4	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccccc1)/C1CCC1)/c1ccc2c(c1)cn[nH]2</chem>
28	1.6	<chem>CC/C(=C(\c1ccc2c(c1Cl)cn[nH]2)/c1ccc(cc1)/C=C/C(=O)O)/c1ccccc1</chem>
32	1.5	<chem>CC/C(=C(\c1ccc2c(c1C#N)cn[nH]2)/c1ccc(cc1)/C=C/C(=O)O)/c1ccccc1</chem>
33	1.4	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccccc1)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
34	2.0	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccc(cc1Cl)F)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
35	2.3	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccc(cc1Cl)S(=O)(=O)C)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2)C=C1</chem>
36	0.7	<chem>N#Cc1ccc(c(c1)Cl)/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>
37	2.4	<chem>COc1ccc(c(c1)Cl)/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>
38	1.1	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccccc1Cl)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
40	0.3	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccc(cc1S(=O)(=O)C)Cl)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
41	1.2	<chem>N#Cc1cc(F)ccc1/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>
42	0.7	<chem>N#Cc1cc(Cl)ccc1/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>
43	0.5	<chem>N#Cc1cc(OC)ccc1/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>
44	1.2	<chem>N#Cc1cc(ccc1)/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1)C(F)(F)F</chem>
45	0.8	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ccncc1Cl)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
48	1.2	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ncccc1Cl)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
50	1.1	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1ncccc1C(F)(F)F)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
51	1.7	<chem>OC(=O)/C=C/c1ccc(cc1)/C(=C(\c1nccc(cc1Cl)F)/C1CCC1)/c1ccc2c(c1)c(F)n[nH]2</chem>
54	1.1	<chem>N#Cc1cc(Cl)cn1/C(=C(/c1ccc2c(c1)c(F)n[nH]2)\c1ccc(cc1)/C=C/C(=O)O)/C1CCC1</chem>

4 Résultats et discussions

Les indices de réactivité calculés par la DFT de 21 composés ayant les valeurs d'activité les plus faibles ont été analysés à l'aide des indices globaux : le potentiel chimique électronique (μ), l'écart énergétique (ΔE), l'énergie totale (E_t), la dureté

absolue (η), l'indice d'électrophilie globale (ω) et l'indice de nucléophilie (N). Les valeurs de ces indices sont rapportées dans le **tableau 2**.

Tableau 2: Énergies HOMO et LUMO, gap énergétique, énergie totale, et indices de réactivité globaux μ , η , ω et N de dérivés de l'indazole sélectionnés.

Composé	E_{HOMO} (au)	E_{LUMO} (au)	ΔE (au)	E_t (au)	μ (au)	η (au)	ω (eV)	N (eV)
1	-0.2321	-0.0871	0.1450	-1822.7470	-0.1596	0.0651	4.7810	3.0607
8	-0.2219	-0.0883	0.1336	-1363.1472	-0.1551	0.0668	4.9008	3.3383
14	-0.2197	-0.0846	0.1351	-1341.2790	-0.1522	0.0676	4.6613	3.3982
28	-0.2245	-0.0860	0.1385	-1723.5000	-0.1553	0.0707	4.7348	3.2675
32	-0.2288	-0.0885	0.1403	-1356.1340	-0.1587	0.0690	4.8817	3.1505
33	-0.2251	-0.0871	0.1380	-1440.5190	-0.1561	0.0717	4.8055	3.2512
34	-0.2335	-0.0902	0.1433	-1999.3670	-0.1619	0.0725	4.9742	3.0226
35	-0.2387	-0.0937	0.1450	-2487.8800	-0.1662	0.0722	5.1838	2.8811
36	-0.2380	-0.0937	0.1443	-1992.3530	-0.1659	0.0696	5.1865	2.9002
37	-0.2257	-0.0865	0.1392	-2014.6370	-0.1561	0.0710	4.7647	3.2349
38	-0.2295	-0.0875	0.1420	-1900.1230	-0.1585	0.0725	4.8137	3.1315
40	-0.2445	-0.0999	0.1446	-2487.8380	-0.1722	0.0729	5.5811	2.7233
41	-0.2361	-0.0903	0.1458	-1631.9980	-0.1632	0.0719	4.9715	2.9519
42	-0.2332	-0.0895	0.1437	-1992.3480	-0.1614	0.0724	4.9307	3.0308
43	-0.2301	-0.0854	0.1447	-1647.2620	-0.1578	0.0730	4.6804	3.1152
44	-0.2401	-0.0941	0.1460	-1869.8180	-0.1671	0.0722	5.2028	2.8430
45	-0.2363	-0.0919	0.1444	-1916.1440	-0.1641	0.0732	5.0749	2.9464
48	-0.2285	-0.0870	0.1415	-1916.1490	-0.1578	0.0696	4.7865	3.1587
50	-0.2299	-0.0868	0.1431	-1793.6100	-0.1584	0.0716	4.7674	3.1206
51	-0.2328	-0.0897	0.1431	-2015.3900	-0.1613	0.0716	4.9443	3.0417
54	-0.2363	-0.0933	0.1430	-2008.3790	-0.1648	0.0715	5.1674	2.9464

L'énergie HOMO -0.34458 au du système de référence (TCT) a été calculée avec la même méthode.

D'après les valeurs de l'écart énergétique ΔE rapportés dans le **tableau 2**, les composés sélectionnés sont stables dans le sens absolu car ils ont des valeurs faibles et approximatives. Le composé **8** a la plus petite valeur de l'écart énergétique ce qui rend le système plus réactif comparait aux autres composés.

La valeur d'énergie totale la plus faible a été enregistrée pour le composé **35** suivi des composés **40** et **51**, ce qui signifie que le composé **35** est plus stable que les autres composés. De plus, ces composés ont des valeurs d'énergies totales inférieures à celle de l'ARN-810 (composé **1**), ce qui a été démontré par Andiliy Lai et al⁴⁰ comme une activité robuste dans les modèles sensibles au tamoxifène et résistants au tamoxifène de cancer du sein.

Le potentiel chimique électronique (μ) du composé **14** ($\mu = -0.1522\text{au}$) est plus élevé que celui des autres composés. Il est suivi de celui des composés : **8** ($\mu = -0.1551\text{au}$), **28** ($\mu = -0.1553\text{au}$), **33** ($\mu = -0.1561\text{au}$) et **37** ($\mu = -0.1561\text{au}$), ceci indique que le composé **14** peut avoir une meilleure densité électronique d'échange avec l'environnement.

Les résultats présentés dans le **tableau 2** indiquent que le composé **40** est le nucléophile le plus pauvre de cette série et le composé **14** est l'une des meilleures molécules nucléophiles car il se caractérise par une valeur élevée de N ($N = 3.3982\text{ eV}$) et une faible valeur de ω ($\omega = 4.6613\text{ eV}$).

L'indice d'électrophilie globale ω est devenu un outil puissant pour l'étude de la réactivité des composés organiques pouvant participer aux réactions polaires. Tous les composés sont des électrophiles forts présentant un indice d'électrophilie ω supérieur à 4,00 eV et donc montrant une réactivité élevée dans les réactions polaires.³⁵

Ainsi, le composé **40** ($\omega = 5.5811\text{ eV}$, **tableau 2**) est caractérisé par le plus haut indice d'électrophilie, suivi des composés **44** ($\omega = 5.2028\text{ eV}$), **36** ($\omega = 5.1865\text{ eV}$), **35** ($\omega = 5.1838\text{ eV}$), **54** ($\omega = 5.1674\text{ eV}$), puis **45** ($\omega = 5.0749\text{ eV}$). De plus, le composé **40** a une faible valeur d' $IC_{50} = 0.3\text{ nM}$, ce qui le rend, d'après la littérature⁴⁰ comme le premier composé actif de cette série. De même, les composés **44**, **36**, **35** et **45** ont de faibles valeurs d' IC_{50} : 1.2, 0.7, 2.3, 1.1 et 0.8 nM respectivement. Par conséquent, le composé **40** est le système le plus réactif de cette série suivi par les composés **44**, **36**, **35**, **54** et **45**. Ces résultats concordent avec ceux obtenus dans l'étude de docking moléculaire.

5 Conclusion

Dans cette partie de travail, nous avons étudié la réactivité de **21** dérivés de l'indazole ayant les meilleures activités à l'aide des indices globaux issus de la DFT conceptuelle dont le but est de comprendre la relation entre la structure, la stabilité et la réactivité chimique globale des composés étudiés. Les résultats de cette étude montrent que :

- le composé **8** a le plus petit écart énergétique ΔE donc le système le plus réactif par rapport aux autres.
- le composé **14** possède une meilleure densité électronique d'échange avec l'environnement. De plus, il est considéré comme l'une des meilleures molécules nucléophiles grâce à une valeur élevée de l'indice de nucléophilie **N** et une faible valeur de l'indice d'électrophile ω .
- tous les composés sont des électrophiles forts présentant un indice d'électrophilie ω supérieur à 4,00 eV montrant une réactivité élevée dans les réactions polaires. Par conséquent, le composé **40** est le système le plus réactif de cette série suivi par les composés **44**, **36**, **35**, **54** et **45**.

Références bibliographiques

1. Fukui, K.; Yonezawa, T.; Shingu, H., A molecular orbital theory of reactivity in aromatic hydrocarbons. *The Journal of Chemical Physics* **1952**, *20* (4), 722-725.
2. Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H., Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *The Journal of Chemical Physics* **1954**, *22* (8), 1433-1442.
3. Parr, R.; Yang, W., Density-functional Theory of Atoms and Molecules clarendon. *Oxford* **1989**.
4. Parr, R. G.; Yang, W., Density-functional theory of the electronic structure of molecules. *Annual Review of Physical Chemistry* **1995**, *46* (1), 701-728.
5. Alston, P.; Ottenbrite, R.; Guner, O.; Shillady, D., A transition state FMO approach for prediction of the regioselectivity of the Diels-Alder reaction. *Tetrahedron* **1986**, *42* (16), 4403-4408.
6. Bradley, J.; Gerrans, G., Frontier molecular orbitals. A link between kinetics and bonding theory. *Journal of Chemical Education* **1973**, *50* (7), 463.
7. Olbrich, C.; Strümpfer, J.; Schulten, K.; Kleinekathöfer, U., Theory and simulation of the environmental effects on FMO electronic transitions. *The Journal of Physical Chemistry Letters* **2011**, *2* (14), 1771-1776.
8. Houk, K. N., Frontier molecular orbital theory of cycloaddition reactions. *Accounts of Chemical Research* **1975**, *8* (11), 361-369.
9. Medjdoub-née-Ghomri, A. Contribution à l'étude de la réactivité chimique à l'aide de la DFT conceptuelle. Applications à la chimie hétérocyclique. Thèse de doctorat, Université Abou Bekr Belkaid de Tlemcen, **2012**.
10. Chermette, H., Chemical reactivity indexes in density functional theory. *Journal of Computational Chemistry* **1999**, *20* (1), 129-154.
11. Geerlings, P.; De Proft, F.; Langenaeker, W., Conceptual density functional theory. *Chemical Reviews* **2003**, *103* (5), 1793-1874.
12. Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Physical Review* **1964**, *136* (3B), B864.
13. Kohn, W., Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Reviews of Modern Physics* **1999**, *71* (5), 1253.
14. Salah, M. Applications des méthodes ab initio et DFT à différentes problématiques chimiques: Spectroscopie (MgS/MgS²⁺), interaction intermoléculaire

(Imidazole@ CO₂) & Réactivité chimique (1, 2, 3-Triazole). Thèse de doctorat, Université Mohammed V-Rabat, **2018**.

15. Morell, C. Un nouveau descripteur de la réactivité chimique: étude théorique et applications à la sélectivité de quelques réaction chimiques. Thèse de doctorat, Université Joseph-Fourier - Grenoble I, **2006**.

16. Labet, V. Etude théorique de quelques aspects de la réactivité des bases de l'ADN- Définition de nouveaux outils théoriques d'étude de la réactivité chimique. Thèse de doctorat, Université Joseph Fourier-Grenoble 1, **2009**.

17. Parr, R. G.; Pearson, R. G., Absolute hardness: companion parameter to absolute electronegativity. *Journal of the American Chemical Society* **1983**, *105* (26), 7512-7516.

18. Politzer, P.; Parr, R. G.; Murphy, D. R., Relationships between atomic chemical potentials, electrostatic potentials, and covalent radii. *The Journal of Chemical Physics* **1983**, *79* (8), 3859-3861.

19. Prana, V. Approches structure-propriété pour la prédiction des propriétés physico-chimiques des substances chimiques. Thèse de doctorat, Université Pierre et Marie Curie, **2013**.

20. Luo, J.; Xue, Z. Q.; Liu, W. M.; Wu, J. L.; Yang, Z. Q., Koopmans' theorem for large molecular systems within density functional theory. *The Journal of Physical Chemistry A* **2006**, *110* (43), 12005-12009.

21. Levy, M.; Nagy, A., Excited-state Koopmans theorem for ensembles. *Physical Review A* **1999**, *59* (2), 1687.

22. Tsuneda, T.; Song, J.-W.; Suzuki, S.; Hirao, K., On Koopmans' theorem in density functional theory. *The Journal of Chemical Physics* **2010**, *133* (17), 174101.

23. Bellafont, N. P.; Illas, F.; Bagus, P. S., Validation of Koopmans' theorem for density functional theory binding energies. *Physical Chemistry Chemical Physics* **2015**, *17* (6), 4015-4019.

24. Perdew, J. P.; Levy, M., Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. *Physical Review Letters* **1983**, *51* (20), **1884**.

25. Casida, M. E.; Wesolowski, T. A., Generalization of the Kohn-Sham equations with constrained electron density formalism and its time-dependent response theory formulation. *International Journal of Quantum Chemistry* **2004**, *96* (6), 577-588.

26. Pearson, R. G., Hard and soft acids and bases. *Journal of the American Chemical Society* **1963**, *85* (22), 3533-3539.

27. Gázquez, J. L., Hardness and softness in density functional theory. In *Chemical hardness*, Springer: **1993**; pp 27-43.
28. Politzer, P., A relationship between the charge capacity and the hardness of neutral atoms and groups. *The Journal of Chemical Physics* **1987**, *86* (2), 1072-1073.
29. Parr, R. G.; Szentpaly, L. v.; Liu, S., Electrophilicity index. *Journal of the American Chemical Society* **1999**, *121* (9), 1922-1924.
30. Chattaraj, P. K.; Giri, S.; Duley, S., Update 2 of: Electrophilicity Index. *Chemical Reviews* **2011**, *111* (2), PR43-PR75.
31. Chamorro, E.; Chattaraj, P. K.; Fuentealba, P., Variation of the electrophilicity index along the reaction path. *The Journal of Physical Chemistry A* **2003**, *107* (36), 7068-7072.
32. Pérez, P.; Domingo, L. R.; Aizman, A.; Contreras, R., The electrophilicity index in organic chemistry. In *Theoretical and Computational Chemistry*, Elsevier: **2007**; Vol. 19, pp 139-201.
33. Domingo, L. R.; Aurell, M. J.; Pérez, P.; Contreras, R., Quantitative characterization of the global electrophilicity power of common diene/dienophile pairs in Diels–Alder reactions. *Tetrahedron* **2002**, *58* (22), 4417-4423.
34. Domingo, L. R.; Chamorro, E.; Pérez, P., Understanding the reactivity of captodative ethylenes in polar cycloaddition reactions. A theoretical study. *The Journal of Organic Chemistry* **2008**, *73* (12), 4615-4624.
35. Domingo, L. R.; Ríos-Gutiérrez, M.; Pérez, P., Applications of the conceptual density functional theory indices to organic chemistry reactivity. *Molecules* **2016**, *21* (6), 748.
36. Lewis, D.; Ioannides, C.; Parke, D., Interaction of a series of nitriles with the alcohol-inducible isoform of P450: computer analysis of structure—activity relationships. *Xenobiotica* **1994**, *24* (5), 401-408.
37. Zhou, Z.; Parr, R. G., Activation hardness: new index for describing the orientation of electrophilic aromatic substitution. *Journal of the American Chemical Society* **1990**, *112* (15), 5720-5724.
38. Chtita, S. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. Thèse de doctorat, Université Moulay Ismail, **2017**.

39. Vijayaraj, R.; Subramanian, V.; Chattaraj, P. K., Comparison of global reactivity descriptors calculated using various density functionals: a QSAR perspective. *Journal of Chemical Theory and Computation* **2009**, 5 (10), 2744-2753.
40. Lai, A.; Kahraman, M.; Govek, S.; Nagasawa, J.; Bonnefous, C.; Julien, J.; Douglas, K.; Sensintaffar, J.; Lu, N.; Lee, K.-j.; Aparicio, A.; Kaufman, J.; Qian, J.; Shao, G.; Prudente, R.; Moon, M. J.; Joseph, J. D.; Darimont, B.; Brigham, D.; Grillot, K.; Heyman, R.; Rix, P. J.; Hager, J. H.; Smith, N. D., Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. *Journal of Medicinal Chemistry* **2015**, 58 (12), 4888-4904.

Conclusion générale

Conclusion générale

Cette étude vise à concevoir un modèle QSAR robuste et fiable pour prédire l'activité anticancéreuse sur la base d'un ensemble de données de **54** dérivés de l'indazole de SERDs. Dans ce travail, la sélection du modèle par l'algorithme de recherche pas à pas (stepwise) a été utilisé pour identifier le meilleur sous-ensemble de descripteur et la régression linéaire multiple MLR a été utilisé afin d'établir le modèle QSAR. Le meilleur modèle obtenu contient quatre descripteurs moléculaires qui ont les proportions les plus élevées.

Selon la qualité d'ajustement statistique, l'équation de la régression MLR a montré un bon accord entre les activités biologiques prédites et observées. Les valeurs du coefficient de corrélation de boots trapping R^2_{Boots} et du coefficient de corrélation au carré R^2_{pred} , entre les activités biologiques expérimentales et prédites des composés de l'ensemble de données, ont révélé que le modèle obtenu a une bonne capacité prédictive et une bonne robustesse. De plus, l'équation MLR résultante était significative en raison de leurs p-values inférieures au niveau de signification 0,05. Ces résultats indiquent que l'équation MLR développée sur les dérivés de l'indazole peut être appliquée avec succès pour prédire l'activité anticancéreuse (cellule de cancer du sein MCF-7) contre le cancer du sein RE+.

L'étude de docking moléculaire a été réalisé sur un ensemble d'inhibiteurs RE α avec de meilleures valeurs d'activité d'IC₅₀ dans la gamme de 0.3 - 2.5 nM à partir de l'ensemble de données contre la protéine (**ID : 3 ERT**) complexé avec le 4-hydroxytamoxifène (10,8988 kcal / mol). Les quatre principaux composés de cette série ont des affinités de liaison élevées et de nombreuses interactions avec la cible dans la simulation de docking. Sur la base de l'étude des propriétés ADME, les composés **14**, **43**, **8** et **45** répondent aux règles de Lipinski de cinq et de Veber, ce qui suggère que ces composés théoriquement n'auraient pas de problème de biodisponibilité orale.

Les dérivés d'indazole les plus actifs contre l'activité anticancéreuse étaient les composés **43** et **45**. Ces composés contiennent du nitrile et du chloro pyridine, qui pourrait être des agents très puissants, au niveau du cycle aryle.

Les paramètres de la réactivité globale dans cette étude ont été calculés par la théorie de la fonctionnelle de densité conceptuelle (DFT). Les résultats ont montré que le composé **14** est considéré comme l'une des meilleures molécules nucléophiles grâce à

Conclusion générale

une valeur élevée de l'indice de nucléophilie N et une faible valeur de l'indice d'électrophile ω . Le composé **40** est caractérisé par l'indice d'électrophilie le plus élevé ($\omega = 5,5811$ eV), ceci indique que ce système est le plus réactif, suivi par les composés **44**, **36**, **35**, **54** et **45**. La plupart des résultats de la réactivité globale sont en accord avec ceux obtenus par l'étude de docking moléculaire.

Les résultats de cette étude confirment que la combinaison de différentes méthodes de calcul fournit plus d'informations sur les SERD pour identifier de nouveaux agents puissants qui peuvent être utiles pour désigner des médicaments anticancéreux puissants.

ملخص

تم إجراء تحليل كمي للعلاقة بين البنية والنشاط على 54 من مشتقات الإندازول باستخدام الانحدار الخطي المتعدد MLR لتطوير نموذج QSAR ذو معنى.

تم استعمال وتنفيذ طريقة DFT-B3LYP مع مجموعة الأساس 6-311G لتحديد البنية، الفعالية الكيميائية، وخصائص المركبات. رافق هذه الدراسة الالتحام الجزيئي *docking moléculaire*، الخصائص العامة للفعالية، خصائص الامتصاص التوزيع، التمثيل الغذائي والحذف ADME على سلسلة مكونة من 21 من محولات مستقبلات الاستروجين الانتقائية SERD والتي تعتبر من أكثر المركبات نشاطاً من بين مجموعة البيانات. بينت معادلة الانحدار MLR توافق جيد بين الخصائص البيولوجية المعدلة والملاحظة. متوسط عالي R^2_{Boots} أكبر من 0.5 يدل على ان النموذج المحصل عليه ذو قوة تنبؤية جيدة. بالإضافة إلى ذلك يمكن اعتبار متوسط قيمة R^2_{pred} (0.57) المقاربة لـ 0.6 مؤشراً على القدرة على التنبؤ الخارجي الجيد. الروابط بين المركبات الأكثر نشاطاً من مجموعة سلسلة الإندازول وانزيم مستقبلات الاستروجين من نوع الفا تم اكتشافها عن طريق الالتحام الجزيئي وذلك لتوضيح نوع الروابط بين المركبات المدروسة والبروتين المعتمد $RE\alpha$. بناء على هذه الدراسة المركبات 14، 43، 8 و 45 لم تنتهك قاعدتي ليبينسكي وفيبير Lipinski et de Veber مما يدل على ان هذه المركبات لا تحدث مشاكل عن طريق الاستهلاك او التوافر البيولوجي عن طريق الفم. معظم نتائج الفعالية العامة global reactivity تتفق مع النتائج المحصل عليها من خلال دراسة الالتحام الجزيئي.

الكلمات الأساسية: ADME ، DFT ، مستقبلات الاستروجين، الإندازول، الالتحام الجزيئي *docking moléculaire*، تحليل QSAR، الفعالية *Réactivité*.

ABSTRACT

A quantitative structure-activity relationship analysis was carried out on 54 indazole derivatives using multiple linear regression (MLR) in order to develop a meaningful QSAR model. The DFT-B3LYP method, with the basis set 6-311G, was performed to define the structure, chemical reactivity and properties of the compounds. This study was accompanied by molecular docking, global reactivity descriptors, absorption, distribution, metabolism and elimination (ADME) properties on a series of 21 compounds of selective estrogen receptor degraders (SERDs) that present the most active compounds from the dataset. The MLR regression equation displayed a good agreement between fitted and observed biological activities. A high average R^2_{Boots} greater than 0.5 indicates that the obtained model has good predictive power and robustness. In addition, an average value of R^2_{pred} (0.57) approximately equal to 0.6 may be taken as an indicator of good external predictability. The interactions between the most active compounds from the indazole series and ER α targets were explored through a molecular docking to elucidate the binding mode between the studied compounds and corresponding protein. Based on this study compounds **14**, **43**, **8** and **45** have no violated Lipinski's rule of five and Veber's rule, suggesting that these compounds would not have problems with oral bioavailability. Most of global reactivity results have an agreement with those obtained in this docking study.

Keywords: ADME, DFT, Estrogen Receptor, Indazole, Molecular Docking, QSAR Analysis, Reactivity.

Résumé

Une analyse quantitative de la relation structure-activité a été effectuée sur 54 dérivés de l'indazoles utilisant la régression linéaire multiple (MLR) afin de développer un modèle QSAR significatif. La méthode DFT-B3LYP, avec la base 6-311G, a été effectuée pour définir la structure, la réactivité chimique et les propriétés des composés. Cette étude a été accompagnée par le docking moléculaire, de descripteurs globaux de la réactivité, et les propriétés Absorption, Distribution, Métabolisme et Élimination (ADME) sur une série de 21 composés de dégradateurs sélectifs de récepteurs d'estrogènes (SERD) qui présentent les composés les plus actifs de l'ensemble de données. L'équation de régression MLR a montré un bon accord entre les activités biologiques ajustées et observées. Une moyenne élevée de R^2_{Boots} supérieure à 0,5 indique que le modèle obtenu a une bonne puissance prédictive et une bonne robustesse. De plus, une valeur moyenne de R^2_{pred} (0,57) approximativement égale à 0,6 peut être considérée comme un indicateur de bonne prédictibilité externe. Les interactions entre les composés les plus actifs de la série d'indazole et la cible du RE α ont été explorées par une docking moléculaire pour élucider le mode de liaison entre les composés étudiés et la protéine correspondante. Selon cette étude, les composés **14**, **43**, **8** et **45** n'ont pas violé la règle des cinq de Lipinski et la règle de Veber, ce qui suggère que ces composés n'auraient pas de problème de biodisponibilité orale. La plupart des résultats de la réactivité globale ont un accord avec ceux obtenus dans cette étude de docking.

Mots Clés : ADME, DFT, Récepteur aux estrogènes, Indazole, Docking Moléculaire, Analyse QSAR, Réactivité.