N° d'ordre : .....

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU (UMMTO)

FACULTE DES SCIENCES / DEPARTEMENT DE CHIMIE

THESE DE DOCTORAT LMD EN CHIMIE

SPECIALITE : CHIMIE DES MATERIAUX ET DE L'ENVIRONNEMENT

OPTION : CHIMIE PHARMACEUTIQUE

THEME

CONCEPTION PAR BIOINFORMATIQUE ET MODELISATION MOLECULAIRE DE NOUVELLES MOLECULES BIOACTIVES DANS LE DOMAINE DU CANCER

Présentée Par

M^me^ HAMMAD Sara

Soutenu publiquement le / /2020, devant le jury composé de :

Nom et Prénom	Grade	Lieu d'exercice	Qualité
Mr ELKECHAI Aziz	Professeur	UMMTO	Président
Mme BOUAZIZ-TERRACHET Souhila	Professeur	UMBB	Rapportrice
Mme MEZIANE Dalila	Professeur	UMMTO	Co-rapportrice
Mr REKIS Mâammar	Professeur	USTHB	Examinateur
Mr SAAL Amar	MCA	UMMTO	Examinateur
Mme KHEMILI-TALBI Souad	MCA	UMBB	Examinatrice

Dédicaces

Je Dédie cette présente thèse à ...

A tous mes enseignants depuis le primaire jusqu'à l'université

A Mes Très Chers Parents Nadia** et **Slimane

Source constante de confiance, d'amour, d'encouragement, de soutien et d'affection. Aucune dédicace n'est assez forte, aucun mot n'est assez éloquent pour exprimer ce que je ressens. Mais à travers ce modeste travail, je voudrais vous dire mon estime et mon respect inégale, ma gratitude infinie et surtout mon immense amour filial.

"Puisse dieu vous accorde longue vie, santé, bonheur et sérénité"

A Mon Cher Mari Hamid

Pour ta bonté, ton respect, ta gentillesse, ta confiance, ton soutien, ta compréhension, ton amour, tes conseilles et encouragements.

A Ma Très Chère Petite Sœur Adorable Romaissa

Source de joie, d'innocence, d'énergie et d'espoir dans ma vie.

A Ma grande Sœur Tassadit et Son Epoux

A Ma Chère Belle Famille et Mes Beaux Parents

A Tous Mes Proches, amies, collègues et tous ceux qui me sont chères

A toute personne qui va lire cette présente thèse

HAMMAD Sara, épouse ZOUAD

A votre bénédiction la plus sainte, Je présente cette thèse .....

Remerciements

Nous remercions Dieu, le tout puissant de nous avoir donné la santé le courage la force et la patience afin de munir ce modeste travail à sa fin.

Je tiens à exprimer ma profonde gratitude et mes remerciements chaleureux à ma directrice de thèse, madame la professeure ***S. BOUAZIZ-TERRACHET****, à l'UMBB et membre du laboratoire de physico chimie théorique et chimie informatique (LPCTCI) de l'USTHB, d'avoir dirigé ce travail, de m'avoir fait partager son ambition à la recherche scientifique et son expertise en modélisation moléculaire et chimie informatique qui sont et resteront la base de mon travail de recherche. Je la remercie infiniment pour son encadrement riche, sa rigueur et son apport scientifique, son aide précieuse, sa disponibilité, et sa patience infinie. Je suis très reconnaissante envers elle de m'avoir formée, assistée et orientée tout au le long de ce travail, et surtout pour tous ses conseils, recommandations et remarques pertinentes, qui m'ont été d'une grande utilité.*

Je tiens à exprimer mes plus sincères remerciements, et ma plus profonde gratitude à ma co-directrice de thèse, madame la professeure ***D. MEZIAN****, responsable de la formation doctorale et présidente du conseil scientifique du département chimie de l'UMMTO, pour sa disponibilité, son aide précieuse, sa rigueur scientifique, sa patience et bienveillance. Je la remercie infiniment pour ses recommandations pertinentes, ses remarques et conseils, qui ont fortement contribuer à la finalisation de cette thèse, je la remercie profondément et je suis très reconnaissante envers elle.*

Un grand merci spécial est adressé au président de jury, monsieur ***A. ELKECHAI*** *professeur et responsable du laboratoire de physique et chimie théorique (LPCQ) de l'UMMT, d'avoir accepté de lire et corrigé ce manuscrit et pour l'intérêt qu'il a bien voulu porter à ce travail, en me faisant l'honneur de présider ce jury.*

Je suis très honorée par la présence de monsieur ***M. REKIS*** *professeur à USTHB. Je le remercie profondément d'avoir accepté d'examiner et de juger ce présent travail.*

Je tiens à exprimer mes remerciements sincères et ma profonde gratitude à monsieur ***A. SAAL****, Maître de Conférences Classe A à UMMTO, pour l'intérêt qu'il a accordé à ce travail et d'avoir accepté d'être examinateur de cette thèse.*

J’adresse mes sincères remerciements et ma profonde gratitude à Madame ***S. KHEMILI-TALBI****, Maître de Conférences Classe A à l'UMBB, d’avoir accepté d'être examinatrice de ce présent travail.*

Mes sincères remerciements et ma profonde reconnaissance à tous les membres du laboratoire LCPCQ de l'USTHB, en tout particulier la directrice du laboratoire madame la professeure ***S. Kellou*** et madame le docteur ***S. Zaater***.

Je remercie profondément et très chaleureusement mon ***cher époux*** *qui m'a toujours encouragé et soutenu, je le remercie pour son aide morale, sa confiance et sa compréhension.*

Je remercie tout particulièrement mes ***chers parents*** *et* ***mes chères sœurs*** *pour leurs aide, amour et leurs soutiens.*

A toutes les personnes qui m'ont assisté de près ou de loin, qu'ils trouvent dans ce modeste travail l'expression de mes sincères remerciements.

Liste des abréviations

ADT :	Autodock tools.
AGI :	Apigenin
AGL :	Algorithme génétique lamarkien.
Ala :	Alanine
AM1 :	Austin model 1
aPK :	Atypical protein kinase
Ar :	Cycle aromatique
Arg :	Arginine
Asp :	Asparagine
AUC :	Area under the curve
BBB :	Blood-brainbarrier (barriere hemato-encephalique)
BUW :	Block unscaled weighting
CHARMM :	Chemistry at harvard macromolecular mechanics
CID :	Compound identification
CK2 :	Caseine kinase 2
COMFA :	Comparative moleculare field analysis
CSNK2A1 :	Caseine kinase 2, alpha polypeptide 1
CSNK2A2 :	Caseine kinase 2, alpha polypeptide 2
CSNK2B :	Caseine kinase 2, beta polypeptide
Cys :	Cysteine
DFT:	Density functional theory
DL :	Dose letale moyenne
EF:	Enrichment factor
EMT:	Epithelial mesenchymal transition
ePK :	Eukaryotic protein kinase
FN :	Faux negatifs
FP :	Faux positifs
Flv :	Flavonoïde
Glu:	Glutamine
Gly:	Glycine
GTP :	Guanosine triphosphate
HBA :	Liaison hydrogène accepteur
HBD :	Liaison hydrogène donneur
HF :	Hartree Fock
His :	Histidine
HY :	Zones hydrophobes
IBAC :	Interactions-Based Accuracy Classification
IC_{50} :	50% de la Concentration de l'activité enzymatique de la substance inhibitrice
Leu :	Leucine
CV-LMO:	Leave Many Out Cross Validation
CV-LOO:	Leave One Out Cross Validation
CV-LTO:	Leave Two Out Cross Validation
Lys :	Lysine
Met :	Méthionine
MIF :	Molecular interaction field
MLR :	Multiple linear regression
MMFF94 :	Merck Molecular Force Field
MQ :	Mécanique quantique
NCBI:	National Center for Biotechnology Information
NI :	Zone ionisable négative

Liste des abréviations

PBC:	Periodic Boundary Conditions
PCR :	Principal component regression
PCs:	Principal components
PDB :	Protein Data Bank (banque de donnée protéique)
Phe :	Phénylalanine
PI :	Zone ionisable positive
PLS:	Partial least squares regression
PM3:	Parametric Method 3
PME:	Particle Mesh Ewald
POM :	Polyoxométallates
Pro :	Proline
Pyr:	Pyrrolysine
QED:	quantitative estimate of drug-likeness
QSAR :	Quantitative de relation structure-activité
RDE :	Relative Displacement Error
RMSD :	Root Mean Square Deviation (déviation quadratique moyenne)
ROC:	Receiver Operating Characteristic
RSR:	Real Space R-factor
SDEC:	Standard Deviation of Error Calculator
SDEP:	Standard Deviation of Error Predicted
Se :	Sensibilité
SE :	Semi-empiriques
Ser :	Sérine
Sp :	Spécificité
TBB :	Tetra bromo benzimidazole
Thr :	Thréonine
Try:	Tryptophane
Tyr:	Tyrosine
Val :	Valine
VN :	Vrais négatifs
VP :	Vrais positifs

Liste des figures

Figure I-1 : Arbre phylogénétique de la superfamille des protéines kinases « ePKs », du kinome humain.....**05**

Figure I-2 : Structure tridimensionnelle de l'holoenzyme tétramérique $CK2\alpha_2\beta_2$ (Code PDB : 1JWH) **06**

Figure I-3 : Représentation tridimensionnelle de la structure de la CK2, résolue à 1.06 Å (Code PDB : 1JWH) ……..... **07**

Figure I-4 : Représentation bidimensionnelle de la structure de la sous-unité CK2β (Code PDB : 1JWH) …….....**08**

Figure I-5 : Représentations bidimensionnelle de la structure du monomère CK2β et l'atome de zinc (Code PDB : 1JWH) ..……....**08**

Figure I-6 : Représentations tridimensionnelle de la structure de la sous-unité CK2β (Code PDB : 1JWH) **08**

Figure I-7 : Représentation tridimensionnelle de la structure de la sous unité CK2α (Code PDB : 3EED) **09**

Figure I-8 : Représentation tridimensionnelle de la structure de la sous unité CK2β (Code PDB : 3EED) **09**

Figure I-9 : Alignement multiple des séquences en acides aminés des sous-unités CK2α et CK2β (Code PDB : 3EED*)* **10**

Figure I-10 : Schéma d'activation de la protéine Kinase CK2 via sa phosphorylation par une protéine kinase depuis une molécule d'ATP.....**12**

Figure I-11 : Structure tridimensionnel et bi dimensionnelle du ligand AGI dans l'emplacement du site actif pour la protéine CK2 (Code PDB : 4DGM) **13**

Figure I-12 : Représentation des différents types d'inhibiteurs de la CK2.....**14**

Figure I-13 : Structure de base des flavonoïdes "Enchainement C6-C3-C6' ".....**18**

Figure I-14 : Mécanisme de formation de Benzimidazole……..**22**

Figure II-1 : Représentation de la sélection de "n" molécules dans une chimiothèque de "N" composés **45**

Figure II-2: Courbe ROC (Receiver Operating Characteristic curve) **47**

Figure III-1 : Structure chimique de l'Apigenine (AGI).....**61**

Figure III-2 : Représentation 3D des caractéristiques pharmacophoriques des trois modèles (a, b et c) générés par LigandScout. **70**

Liste des figures

Figure III-3 : Représentation des courbes ROC obtenues par le criblage de 26 composés actifs et 1000 decoys, en utilisant les trois modèles : A- modèle 1, B- modèle 2, C- modèle 3.....**74-75**

Figure III-4 : Organigramme systématique du criblage virtuel utilisée dans notre étude **76**

Figure III-5 : Comparaison entre la structure cristallographique X-ray (représentée en boules et tiges vertes) et la structure issue du docking de ligand « AGI » (représentée en tiges grises)…….....**78**

Figure III-6 : Représentation graphique des énergies prédites par le docking des composés de l'ensemble de test, en fonction des valeurs expérimentales (pIC50) **79**

Figure III-7 : Superposition de la conformation dockée du composé prédit (boules et bâtonnets grises) et de la structure cristallographique de l'AGI (bâtonnets vertes) **81**

Figure III-8 : Vue bidimensionnelle du mode de liaison du composé prédit hit dans le site actif de la CK2.....**82**

Figure III-9 : Représentation tridimensionnelle de la superposition de la molécule prédite sur le modèle pharmacophore 3.....**83**

Figure III-10 : Superposition de la conformation issue du docking (boules et batonnêts grises) et celle obtenue après 100 ns de dynamique molécualire (boules et batonnêts de colorés selon la nature des atomes) du complexe prédit (hit-CK2) **85**

Figure III-11 : Nombre de liaisons hydrogène entre le ligand « hit » et la protéine CK2.....**86**

Figure III-12 : Nombre de liaisons hydrogène entre le ligand AGI et la protéine CK2.....**86**

Figure III-13 : RMSD des atomes Cα du backbone des complexes hit - CK2 (rouge) et AGI -CK2 (bleu).....**87**

Figure III-14 : RMSD du ligand prédit « hit » (rouge) et expérimental « AGI » (bleu) durant les 100 ns de dynamique moléculaire.....**88**

Figure III-15 : Courbes de RMSF des complexes étudiés : complexe prédit « hit–CK2 » (rouge), complexe cristallographique « AGI–CK2 » (vert) et le B factor (bleu).....**89**

Figure III-16 : Courbe de propriétés « SASA total » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant les 100 ns de simulations de dynamique moléculaire.....**90**

Figure III-17 : Courbe des propriétés « SASA Hydrophiles » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant les 100 ns de simulations de dynamique moléculaire.....**91**

Liste des figures

Figure II-18 : Diagramme des propriétés « SASA Hydrophobes » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant 100 ns de simulations de dynamique moléculaire.....**92**

Figure III-19 : Courbe des valeurs de (rGyr) calculées pour les complexes étudiés « hit - CK2 » (rouge) et « AGI–CK2 » (bleu) pendant les 100 ns de simulation de dynamique moléculaire.....**93**

Figure III-20 : Visualisation de l'espace réel de la formation de liaisons hydrogène au sein d'une macromolécule biologique (8-hydroxyquinoléine) **136**

Figure III-21 : Formation de quatre liaisons hydrogène d'une molécule d'eau avec les molécules voisines.....**136**

Figure VI-1 : Structure commune de dérivés de flavonoïdes.....**100**

Figure VI-2 : La superposition entre la conformation de la quercitrine (Flv26) issue du docking moléculaire (en vert) et celle du ligand expérimental AGI (en rouge) dans le site actif de la CK2.....**108**

Figure VI-3 : Interactions moléculaires clés entre la Quercitrine et les résidus du site actif de la CK2.....**109**

Figure VI-4 : Alignement des 83 molécules des dérivés des flavonoïdes dans la boite cubique de 1 $Å^3$.....**110**

Figure VI-5 : Visualisation du recouvrement du champ électrostatique, les régions favorables et défavorables du champ électrostatique, sont illustrées en bleu et en rouge, respectivement.....**111**

Figure VI-6 : Visualisation du recouvrement du champ stérique, les régions favorables et défavorables du champ stérique, sont illustrées en vert et en jaune, respectivement.....**111**

Figure VI-7 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 1), en fonction des activités expérimentales PIC50.....**112**

Figure VI-8 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2), en fonction des activités expérimentales PIC50.....**112**

Figure VI-9 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=3), en fonction des activités expérimentales PIC50.....**113**

Figure VI-10 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=4), en fonction des activités expérimentales PIC50.....**113**

Figure VI-11 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=5), en fonction des activités expérimentales PIC50.....**114**

Liste des figures

Figure VI-12 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.....**116**

Figure VI-13 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.....**116**

Figure VI-14 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.....**117**

Figure VI-15 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.....**117**

Figure VI-16 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**118**

Figure VI-17 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**118**

Figure VI-18 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**119**

Figure VI-19 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**119**

Figure VI-20 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**120**

Figure VI-21 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**120**

Figure VI-22 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**121**

Figure VI-23 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.....**121**

Liste des tableaux

Tableau I-1 : Différents type d'inhibiteurs, classification moléculaire et exemples correspondants.....**15**

Tableau I-2 : Différentes classes des flavonoïdes, structures, nomenclature IUPC et leurs distributions dans les fruits et légumes (source naturelle) **19**

Tableau III-1 : Les noms chimiques et les structures 2D des ligands utilisés dans le lot d'entrainement **61-62**

Tableau III-2 : Fragment commun des flavonoïdes, IC50 (µM), pIC50 = -log (IC50), fit score, énergies du Docking (kJ/mol), et structures 2D des composés actifs utilisés comme échantillon de test « set test » **64**

Tableau III-3 : position des caractéristiques pharmacophoriques des trois Modèles pharmacophore générés (1,2 et 3) **71**

Tableau III-4 : Paramètres de validation calculés à l'aide de l'échantillon de test constitué de 26 composés actifs et de 1000 decoys, pour les trois modèles pharmacophore générés (1, 2 et 3 .....**75**

Tableau III-5 : Structure chimique, propriétés physicochimiques (drug-likeness), résultats d'ajustement (fit score) et énergies du docking (Docking Score) obtenus pour le meilleur composé hit, ainsi que pour l'AGI.....**80**

Tableau III-6 : Résultats de toxicité (DL50) des composés étudiés et classifications correspondantes.....**84**

Tableau VI-1 : Structures moléculaires, activités expérimentales IC50 (µM) / pIC50 = -log IC50 et énergies prédites par le docking moléculaire **101-103**

Tableau VI-2 : Résultats de la validation non croisée des cinq modèles QSAR générés, PC = 1, jusqu'à PC = 5.....**114**

Tableau VI-3 : Résultats de validation croisée par les méthodes statistiques LOO, LTO et LMO, pour les quatre modèle QSAR, PC = 2, PC = 3, PC = 4 et PC = 5, générés par COMFA du logiciel Open 3D QSAR.....**122**

Liste des communications et publications scientifiques

Liste des communications scientifiques nationales

✓ Le 24-25 Mars 2015 : 6 èmesJournées de Chimie (JCH'06) : ECOLE MILITAIRE POLYTECHNIQUE de Bordj -El-Bahri, Alger (Algérie)

Titre de la communication poster: "MOLECULAR DOCKING STUDY OF BENZOPYRAN DERIVATIVES AS CASEIN KINASE 2 INHIBITORS".

Auteurs : Hammad Sara, Souhila Bouaziz-Terrachet

✓ Le 28-30 Mars 2017 : 12 èmesJournées de Chimie Théoriques et Computationnelle (LPCQ/JCTC12) : Université MOULOUD MAMMERI Tizi-Ouzou (Algérie)

Titre de la communication poster : "THREE DIMENSIONAL QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS OF EIGHTY-THREE FLAVONOÏDES DERIVATIVES BASED ON MOLECULAR INTERACTION FIELDS DESCRIPTORS".

Auteurs : Hammad Sara, Souhila Bouaziz-Terrachet

✓ Le 16 Avril 2017 : 1ère Journée Nationale de l'Agro-alimentaire, Engineering et Environnement (JNAEE) : Université M'HAMED BOUGARA, Boumerdes (Algérie)

Titre de la communication poster: "EXPLORING THE BINDING MECHANISM OF FLAVONOÏDES DERIVATIVES TO CK2 PROTEIN BY TWO DIFFERENT SEARCH ALGORITHMS: LAMARCKIAN GENETIC ALGORITHM (LGA) AND MONTE CARLO SIMULATION ANNEALING (SA)".

Auteurs : Hammad Sara, Souhila Bouaziz-Terrachet

Publication scientifique

Hammad, S., Bouaziz-Terrachet, S., Meghnem, R. et Meziane D. Pharmacophore development, drug-likeness analysis, molecular docking, and molecular dynamics simulations for identification of new CK2 inhibitors. J Mol Model 26, 160 (2020). https://doi.org/10.1007/s00894-020-04408-2

Table des Matières

Dédicaces

Remerciements

Liste des abréviations

Liste de figures

Liste des tableaux

Liste de communication et de publication

Introduction Générale……..…….....1

Chapitre I : Notions bibliographiques sur la protéine Caséine Kinase II (CK2)

I. Introduction 4
II. Généralités sur la famille des protéines kinase 4
III. Protéine caséine kinase II "CK2" 5
III.1 Structure de la CK2 5
III.2 Structure de la sous-unité CK2β 7
III.3 Isoformes de la protéine Kinase CK2 8
III.3.1 CK2α 8
III.3.2 CK2α' 9
III.3.3 CK2β 9
III.4 Structures et séquences en acides aminés des sous-unités CK2α et CK2β 9
IV. Rôle de la protéine CK2 10
IV.1 Activité catalytique et mécanisme d'activation de la CK2 11
IV.2 Site actif de la CK2 13
V. Inhibiteurs de la CK2 14
V.1 Les différents types d'inhibiteurs de la CK2 14
V.1.1 Inhibiteurs concurrentiels (compétitifs) d'ATP 16
V.1.2 Inhibiteurs allostériques 16
V.1.3 Inhibiteurs inorganiques de la protéine CK2 (polyoxométallates) 16
V.1.4 Inhibiteurs concurrentiels du substrat 17
V.1.5 Inhibiteurs visés par la CK2β 17
V.1.6 Inhibiteurs d'interaction entre les deux sous - unités CK2α et CK2β 17
VI. Flavonoïdes, nouvelle famille d'inhibiteurs de la CK2 17
VI.1 Structure générale des flavonoïdes 18
VI.2 Classification des flavonoïdes 18
VI.3 Activité anti cancéreuse des flavonoïdes et exemples 21
VII. Benzomidazoles 22
VIII. Références 23

I. Introduction 27
II. Méthodes de la modélisation moléculaire 27
II.1 Mécanique Quantique 27
II.2 Mécanique moléculaire 28
II.2.1 Principe de la mécanique moléculaire 28
II.2.2 Champ de force 29
II.2.3 Optimisation de l'énergie d'une molécule et/ou d'un système moléculaire 29
II.2.4 Méthode de minimisation énergétique 30
II.2.5 Algorithmes de minimisation 30
II.2.5.1 Méthode de "Steepest Descent" 30
II.2.5.2 Méthode du gradient conjugué 31
II.2.5.3 Méthode Newton-Raphson 31
II.2.5.4 Méthode de recuit simulé 31
II.3 Dynamique moléculaire 32
II.3.1 Dynamique moléculaire Newtonienne (dynamique classique) 32
II.3.2 Dynamique de Langevin 32
II.3.3 Etapes de la dynamique moléculaire 33
II.3.3.1 Préparation du système moléculaire et optimisation énergétique 33
II.3.3.2 Thermalisation 33
II.3.3.3 Equilibration 34
II.3.3.4 Production 34
II.3.3.5 Interprétation des résultats / Analyses usuelles en dynamique moléculaire 34
II.3.3.5.1 L'écart quadratique moyen (Root Mean Square Deviation : RMSD) 34
II.3.3.5.2 Fluctuations atomiques (Root Mean Square Fluctuation: RMSF) 35
II.3.3.5.3 Facteur d'agitation thermique (B Factor) 35
II.3.3.5.4 Liaisons hydrogènes 35
II.3.3.5.5 La surface accessible au solvant (Solvent Accessible Surface Area : SASA) 35
II.3.3.5.6 Le rayon de giration (Radius of gyration : Rg) 36
II.3.4 Technique numériques et outils de la dynamique moléculaire 36
II.3.4.1 Méthode du maillage particulier d'Ewald (particular mesh Ewald : PME) 36
II.3.4.2 Conditions périodiques aux limites (Periodic boundary conditions : PBC) 36
II.3.4.3 Barostat de Langevin 37
II.3.4.4 Thermostat de Langevin 37
II.3.4.5 Le rayon de troncature (cut-off radius) 37
II.3.4.6 Le rayon de switch (switch radius/smooth cut-off) 38

III. Techniques de conception de médicament "Drug design" 38
III.1 Docking Moléculaire 38
III.1.1 Docking rigide 38
III.1.2 Docking semi-flexible 39
III.1.3 Docking flexible 39
III.1.4 Insuffisance (limites) des méthodes de Docking moléculaire 39
III.2 Criblage (Screening) 40
III.2.1 Criblage réel 40
III.2.2 Criblage virtuel (Virtual Screening) 40
III.2.3 Approches du criblage virtuel 41
III.2.3.1 Approche "Structure-Based Virtual Screening" 41
III.2.3.2 Approche "Ligand-Based Virtual Screening" 41
III.2.4 Outils de criblage virtuel 41
III.2.4.1 Recherche de similarité moléculaire (paradigme de similitude) 41
III.2.4.2 Descripteurs de similarité 42
III.2.4.3 Modèles pharmacophores (ligand-based / structure-based) 42
III.2.5 Evaluation des méthodes de criblage virtuel 42
III.2.5.1 Précision du positionnement 43
III.2.5.1.1 Ecart quadratique moyen (Root Mean Square Deviation : RMSD) 43
III.2.5.1.2 Erreur relative de déplacement (Relative Displacement Error : RDE) 44
III.2.5.1.3 Espace réel du facteur RSR (Real Space R-factor) 44
III.2.5.2 Classification de la précision basée sur les interactions (IBAC) 45
III.2.5.2.1 Enrichissement de la chimiothèque 45
III.2.5.2.2 Précision 46
III.2.5.2.3 Sensibilité 46
III.2.5.2.4 Spécificité 46
III.2.5.2.5 Facteurs d'enrichissement (FE) 46
III.2.5.2.6 Courbe ROC (Receiver Operating Characteristic curve) 47
III.3 Méthode quantitative de relation structure-activité (QSAR) 48
III.3.1 Paramètres biologiques (base de données QSAR) 49
III.3.2 Descripteurs moléculaires 49
III.3.3 Méthodes mathématiques de la modélisation QSAR 50
III.3.3.1 La régression linéaire multivariée (Multiple Linear Regression : MLR) 50
III.3.3.2 La régression en composantes principales (Principal Component Regression PCR) .. 51
III.3.3.3 Méthodes des moindres carrés partiels (partial least squares : PLS) 51
III.3.4 Méthodes de validation des modèles QSAR 51

III.3.4.1 Validation non-croisée 51
III.3.4.1.1 Leave-n-out 51
III.3.4.2 Validation croisée 52
III.3.4.2.1 Leave-one-out (LOO) 52
III.3.4.2.2 Leave-two-out (LTO) 52
III.3.4.2.3 Leave-many-out (LMO) 52
III.3.4.2.4 Paramètres statistiques de la validation croisée 52
III.3.5 Paramètres de calculs 53
III.3.5.1 Coefficient de détermination (r^2) 53
III.3.5.2 Racine carrée de coefficient de détermination (R) 53
III.3.5.3 Test Fischer (F) 53
III.3.5.4 Ecart type (s) 54
III.3.5.5 Coefficient de prédiction (q^2) 54
III.3.5.6 Coefficient de corrélation prédit Q 54
IV.Références 55

Chapitre III : Identification de nouveaux inhibiteurs de la CK2 par les approches in silico

I. Introduction 59
II. Matériel et méthodes 60
II.1 Génération du modèle pharmacophore basé sur la structure de la cible «Structure-Based» 60
II.2 Génération du modèle pharmacophore « Ligand-Based » 61
II.3 Validation et optimisation des modèles pharmacophoriques 62
II.4 Criblage virtuel 65
II.5 Calcul et analyse des propriétés « Drug-Likeness » 65
II.6 Etude de docking moléculaire 66
II.7 Analyse in silico de la toxicité 66
II.8 Simulation de dynamique moléculaire 67
III. Résultats et discussion 68
III.1 Génération du modèle pharmacophore 1 « Structure-Based » 68
III.2 Génération du modèle pharmacophore 2 « ligand-Based » 69
III.3 Génération du modèle pharmacophore 3 69
III.4 Validation des modèles pharmacophores 71
III.4.1 Cas du modèle 1 : 71
III.4.2 Cas du Modèle 2 72
III.4.3 Cas du Modèle 3 73
III.5 Criblage Virtuel 75
III.6 Analyse Drug-likeness 77

III.7 Docking moléculaire 77
III.7.1 Génération des paramètres du docking 77
III.7.2 Validation du modèle de Docking.....78
III.7.3 Mode de liaison des composés criblés 79
III.8 Analyse *in silico* de la toxicité 83
III.9 Simulation de Dynamique Moléculaire.....84
III.9.1 Analyse des liaisons Hydrogène 85
III.9.2 Analyse RMSD 86
III.9.3 Analyse RMSF 88
III.9.4 Analyse des propriétés du ligand 90
IV. Conclusion 93
V. Références 95

Chapitre IV : Etude des relations quantitatives structure-activité d'une série de Flavonoïdes

I. Introduction..... 98
II. Matériel et Méthodes 98
II.1 Docking Moléculaire 98
II.1.1 Préparation de la protéine (récepteur)..... 98
II.1.2 Préparation des ligands..... 98
II.1.3 Protocol du docking moléculaire 99
II.1.4 Evaluations des résultats de Docking 99
II.2 Etude QSAR 99
II.2.1Création de la base de données « structure - activité »..... 100
II.2.2 Alignement des molécules..... 100
II.2.3 Génération des champs d'interactions moléculaires (MIFs) 104
II.2.4 Répartition de la base de données..... 104
II.2.5 Méthode de régression linéaire par les moindres carrées partiels (PLS)..... 105
II.2.6 Validation des modèles QSAR 105
II.2.6.1 Validation non-Croisée (non-Cross Validation)..... 105
II.2.6.2 Validation Croisée (CV)..... 106
II.2.6.2.1 Validation croisée LOO..... 106
II.2.6.2.2 Validations croisées LTO, LMO 106
III. Résultats et discusion 108
III.1 Docking moléculaire 108
III.2 Etude QSAR..... 110
III.2.1 Alignement des molécules 110

III.2.2 Analyse des MIFs 110
III.2.3 Analyse des matrices PLS des descripteurs moléculaires MIFs 111
III.2.4 Validation non Croisée (non-Cross Validation) 111
III.2.5 Validation Croisée (Cross Validation) 115
III.2.5.1 Validation croisée CV-LOO 115
III.2.5.2 Validation croisée CV-LTO 118
III.2.5.3 Validation croisée CV-LMO 120
III.2.5.4 Comparaison des résultats des trois méthodes de Validation croisée 121
III.2.5.5 Sélection du meilleur modèle QSAR 122
VI. Conclusion 123
IV. Références ……....124

Conclusion générale & perspectives, Annexes et résumé

Conclusion générale et perspectives…….. 126
Annexe 01 ……130
Annexe 02 ……....133
Résumé ……....139
Abstract ……..140

La bio-informatique est une science à l'interface de plusieurs disciplines numériques (l'informatique et les mathématiques) et des sciences de la vie comme la biochimie, la génétique, la génomique structurale, la protéomique, la biologie structurale (structure spatiale des molécules biologiques, modélisation moléculaire ...).

Cette science nous aide à visualiser des structures invisibles à l'œil nu tels que les protéines et d'en apprendre davantage sur leur fonction. Cela conduit à comprendre les questions essentielles de la vie : Comment les organismes fonctionnent-ils ? Comment la vie s'est-elle développée ? Comment peuvent se développer de nouveaux traitements contre des maladies. Cette science fournit des outils pour la modélisation moléculaire, la visualisation, l'exploration et l'interprétation des données.

La modélisation moléculaire implique l'utilisation des méthodes de calcul théoriques (mécanique moléculaire, dynamique moléculaire et mécanique quantique) permettant de déterminer la représentation graphique de la géométrie ou de la configuration des atomes d'une molécule et d'évaluer les propriétés physico-chimiques de la molécule étudiée. Ainsi, la modélisation moléculaire permet d'interpréter des phénomènes physico-chimiques, de suggérer de nouvelles expériences et d'analyser des résultats d'une façon plus critique que les expériences.

Le comportement des molécules peut être simulé de façon statique ou dynamique. Cette simulation inclut notamment le comportement et la flexibilité des molécules en solution, les mécanismes réactionnels et interactions mises en jeu dans la catalyse enzymatique, la reconnaissance ou la complexation, la prédiction de propriétés et de fonctionnalités à partir d'analogies de structure.

Dans ce travail, nous avons utilisé des approches de modélisation moléculaire et de bioinformatique pour identifier des molécules bioactives d'intérêt thérapeutiques dans le domaine du cancer. Ce dernier est une prolifération incontrôlée de cellules au sein de l'organisme. Des études récentes ont démontré le rôle déterminant d'une enzyme nommée la caséine kinase 2 (CK2) dans l'apparition de cette maladie. En effet, cette protéine joue un rôle dans la plasticité cellulaire épithéliale. En effet, une altération de son activité multifonctionnelle, implique des variations importantes dans la transformation des cellules épithéliales en cellules cancéreuses plus actives à l'origine des métastases.

Cette protéine se produit principalement en tant qu'hétéro-tétramérique holoenzyme, formé par l'association de deux sous-unités catalytiques **(CK2α et CK2α')** et un dimère composé de deux sous-unités de normalisation ou régulatrices identiques (non catalytiques :

CK2β et CK2β'). Les sous-unités catalytiques CK2α et CK2α', ne sont pas liées et elles se diffèrent par leurs tailles de séquences peptidiques.

Actuellement, la majorité des inhibiteurs de la protéine CK2 développés sont des compétitifs de l'un des substrats de l'enzyme : l'adénosine triphosphate (ATP). Ces inhibiteurs se fixent dans le site catalytique de l'ATP, situé entre les sous unités **CK2α** et **CK2β**, empêchant ainsi la protéine de progresser dans son activité catalytique.

D'après la littérature, plusieurs classes d'inhibiteurs puissants de la CK2 ont été développés ces dernières années. Parmi ces inhibiteurs nous pouvons citer l'Apigénine et le lutéoline, le tétrabromobenzotriazole (TBB), les dérivés du pyrazolotriazine (IQA), ainsi que la nouvelle molécule CX-4945. Parmi tous ces inhibiteurs, seulement un composé, le CX-4945, est arrivé en phase des études cliniques.

Les objectifs principaux de nos travaux de recherches consistent, en un criblage virtuel d'une chimiothèque issue de la base de données Pubchem, afin d'identifier des composés pouvant être des inhibiteurs puissants de la protéine CK2. Le pouvoir d'inhibition de ces composés a été évalué en utilisant plusieurs approches de la bioinformatique et de modélisation moléculaire comme les études in silico de la toxicité, les études des propriétés *druglikeness*, les études de docking et de dynamique moléculaire pour prendre en considération la flexibilité du complexe ligand-récepteur.

Après une introduction générale où nous présentons le but du travail, ce manuscrit est structuré de la manière suivante : Le premier chapitre intitulé « Notions bibliographiques sur la protéine Caséine Kinase II (CK2) **»** présente une revue bibliographique, menée pour mieux connaitre la structure, le rôle biologique, les propriétés et les caractéristiques de la protéine caséine kinase 2 (CK2), ainsi que les différentes classes de ses inhibiteurs spécifiques.

Le deuxième chapitre intitulé « Méthodes de la bio-informatique et de la modélisation moléculaire » expose les différentes approches de la bio-informatique et de la modélisation moléculaire et, en particulier celles utilisées dans nos travaux de recherche.

Dans le troisième chapitre intitulé « Identification de nouveaux inhibiteurs de la CK2 par des approches *in silico* », nous reportons une étude in silico réalisée dans le but d'identifier de nouveaux inhibiteurs de la CK2. Cette simulation a été réalisée en utilisant plusieurs approches bio-informatiques notamment le screening virtuel, la modélisation pharmacophore, le docking moléculaire et la dynamique moléculaire.

Dans le quatrième et dernier chapitre nous exposons une étude quantitative de relation structure activité (QSAR), effectuée sur une série de molécules présentant une activité

inhibitrice de l'enzyme CK2, à savoir les dérivés des flavonoïdes, dans le but de mettre en évidence les éléments structuraux responsables de la sélectivité d'une molécule pour la CK2.

Nous terminons ce manuscrit par une conclusion générale qui récapitule les différentes démarches empruntées dans cette investigation avec un résumé de l'essentiel des résultats obtenus et quelques perspectives qui peuvent être un prolongement de cette étude.

I. Introduction

Le cancer est une maladie caractérisée par une aberration dans la croissance, la prolifération et la survie cellulaire. Elle entraine des multiplications anarchiques de cellules anormales (tumeurs malignes) dans l'organisme, des troubles de contrôle des hémostasies cellulaires et dérèglements du processus d'apoptose (mort cellulaire programmée). Cela est dû à plusieurs facteurs environnementaux (radiations ionisantes, rayonnements ultraviolet, virus, bactéries...etc.), mais aussi à des prédispositions héréditaires.

Selon la société algérienne d'oncologie médicale (**SAOM**) et le centre spécialisé dans la lutte contre le cancer "Pierre et Marie Curie" (Alger), plus de 50.000 nouveaux cas de cancer et 20.000 décès sont enregistrés chaque année en Algérie.

En effet, il existe plusieurs types de cancer, dont on trouve le cancer du sein (cancer des glandes mammaires), qui est classé en 2020, parmi les 4 principaux cancers (à savoir : cancers du sein, de la prostate, des poumons et colorectaux) [1].

Dans le but d'explorer cette maladie mortelle, une protéine nommée : Caséine Kinase 2 "*CK2*", a été identifiée comme une cible thérapeutique pertinente pour lutter contre la progression des tumeurs mammaires [2-3]. En effet, des études ont démontrés que la CK2, est impliquée dans de nombreux cancers, notamment le cancer du sein [3], de la prostate [4], des poumons et de la leucémie myéloïde aiguë [2-8].

Dans ce chapitre nous résumons l'état actuel des connaissances concernant la protéine CK2 et de son inhibition. Nous y reportons également les classes d'inhibiteurs actuellement développées.

II. Généralités sur la famille des protéines kinases

Les protéines kinases sont des enzymes qui catalysent des réactions de phosphorylation d'acides aminés spécifiques (sérine, thréonine ou tyrosine) d'une autre molécule cible, appelée "substrat". Cette dernière peut être une protéine, un lipide, un acide nucléique , un sucre ou encore une autre kinase. Le kinome humain est constitué de 518 protéines kinases [9-10], dont :

- ✓ 478 protéines eucaryotique kinases (ePK), qui possèdent au moins un domaine catalytique.
- ✓ 40 protéines kinases atypiques (aPK), qui sont caractérisées par leurs divergences substantielles au niveau de leurs séquences, en les comparant à la superfamille des ePK.

La figure I-1 représente l'arbre phylogénétique du kinome humain, répartis-en 8 groupes majeurs et distincts pour la super famille des protéines kinase [9-10], à savoir : **TK** (*Tyrosine-*

Kinase), **TKL**(*Tyrosine-Kinase-Like*), **STE** (*stérile 20 serine / threonine kinase*), **CK1**(*Casein Kinase 1*), **AGC** (*protein A, G and C kinase*), **CAMK** (*Calcium /calmoduline dependent kinase*), **GYC** (*receptor guanylate cyclase*) et **CMGC** (*cycline dependent kinase*). Cette dernière famille contient les protéines kinases atypiques CK2 alpha qui feront objet de ce travail de thèse.

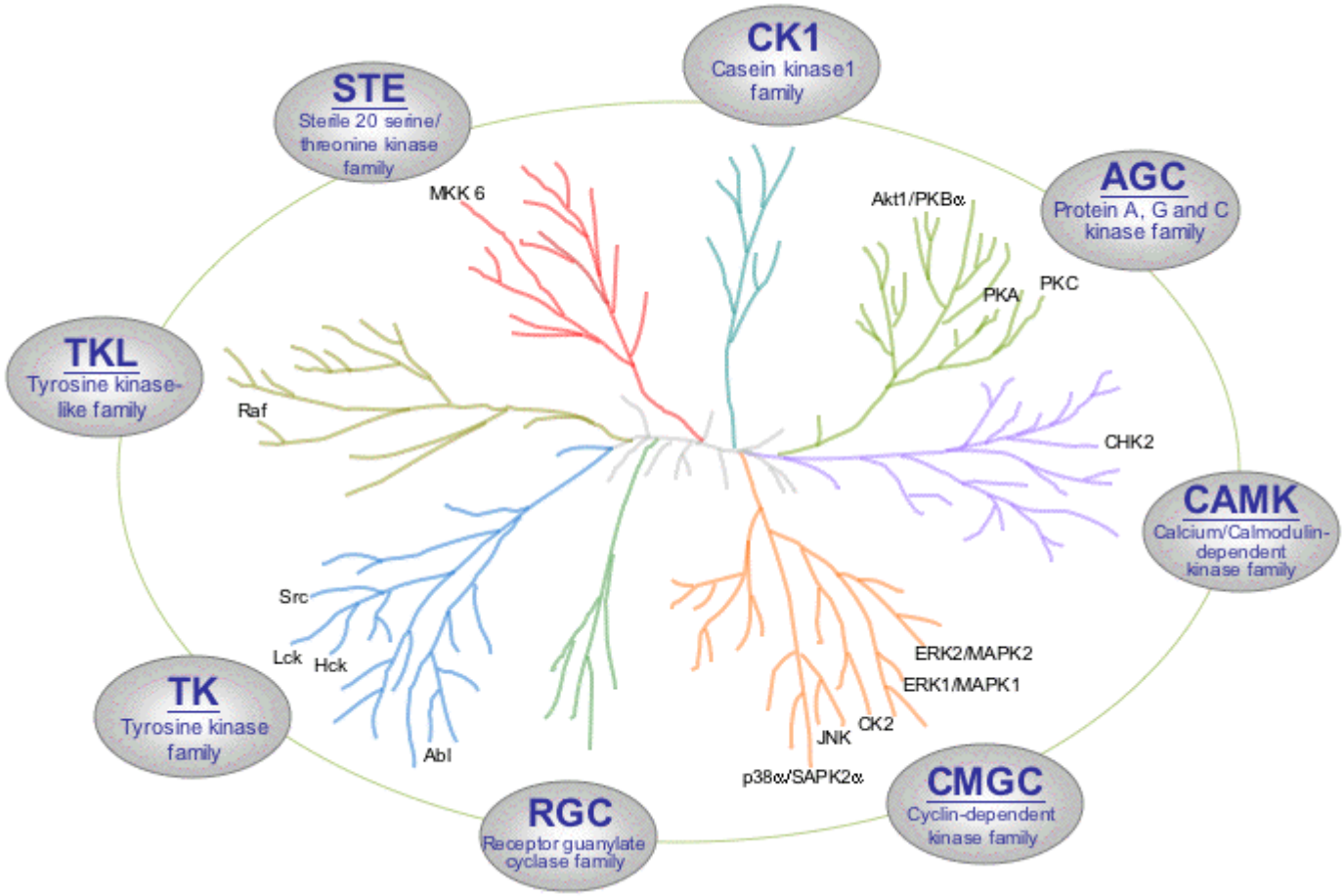


Figure I-1 : Arbre phylogénétique de la superfamille des protéines kinases « ePKs », du kinome humain [10].

III. Protéine caséine kinase 2 "*CK2*"

III.1 Structure de la CK2

La protéine kinase CK2 (acronyme de *Caséine Kinase 2*), est une sérine-thréonine kinase ubiquitaire, qui se produit principalement en tant qu'hétéro-tétramérique holoenzyme, formé par l'association de deux sous-unités catalytiques **(CK2α et CK2α')** et un dimère composé de deux sous-unités de normalisation ou régulatrices identiques (non catalytiques : **CK2β et CK2β'***)* qui subissent une autophosphorylation [11-18].

La structure cristallographique de la CK2, révèle que chaque sous-unité CK2β interagit avec les deux sous-unités catalytiques, tandis que les deux sous-unités CK2α, n'ont aucun

contact entre elles. L'holoenzyme peut donc se retrouver sous la forme de complexes de type : **$\alpha_2\beta_2$, $\alpha\alpha'\beta_2$**, ou bien **$\alpha'_2\beta_2$**[11-18].

Le terme "Hétéro-tétramérique", indique que la protéine CK2 est composée de quatre monomères (sous-unités) différents (hétéro) comme illustré sur la figure I-2 [15].

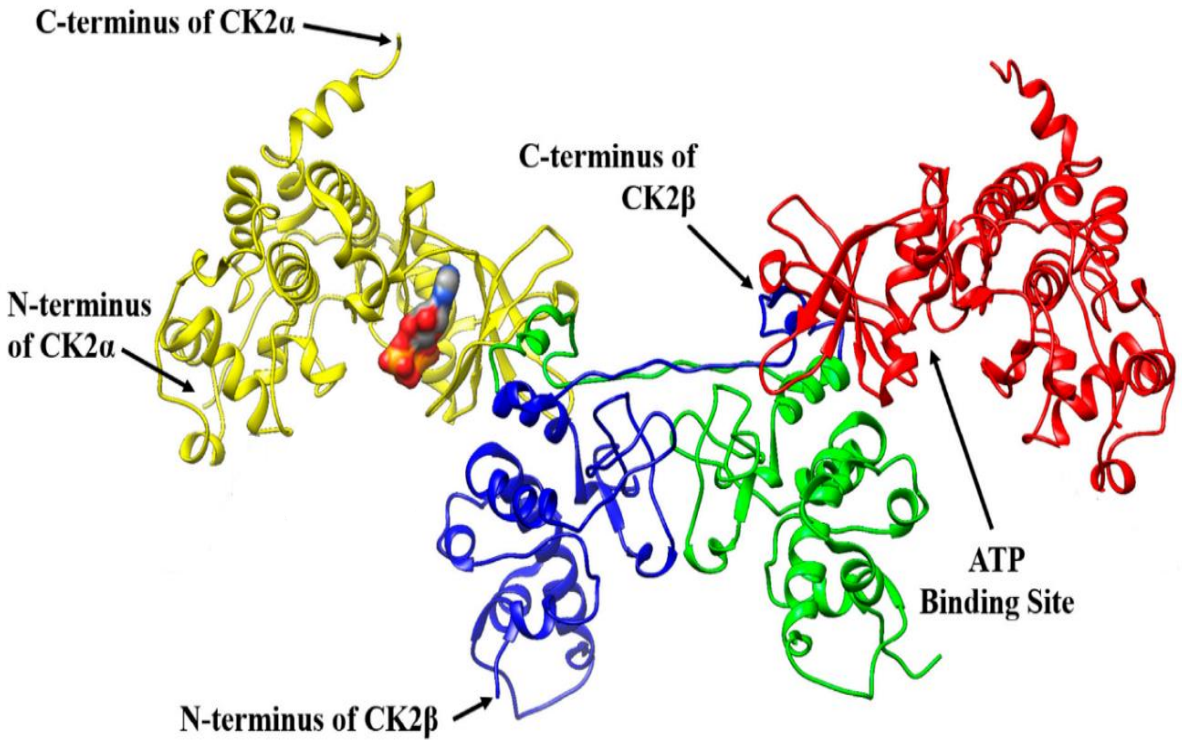


Figure I-2 : Structure tridimensionnelle de l'holoenzyme tétramérique CK2$\alpha_2\beta_2$ (Code PDB : 1JWH) [15].

Comme illustrée sur la figure précédente, la CK2 est composée de 04 monomères :

- ✓ C-Terminal de la **CK2α** jusqu'au C-Terminal de la **CK2β** ;
- ✓ C-Terminal de la **CK2β** jusqu'au N-Terminal de la **CK2β** (deux monomères CK2β, appelé "**Dimère CK2β**")
- ✓ Le site récepteur d'ATP (**ATP Binding Site**).

L'holoenzyme tétramérique "**CK2α2β2**" a la forme d'un papillon, dans lequel les deux sous-unités **CK2β** forment un dimère stable, auquel les monomères **CK2α** et **CK2α'** sont attachés indépendamment par l'intermédiaire d'une queue prolongée de **C-terminal** de la **CK2β**, une molécule d'ATP est présente dans l'un des sites catalytiques.

La figure I-3 représente la structure globale tridimensionnelle d'un seul monomère de la CK2 avec une résolution de 1.06 Å [18].

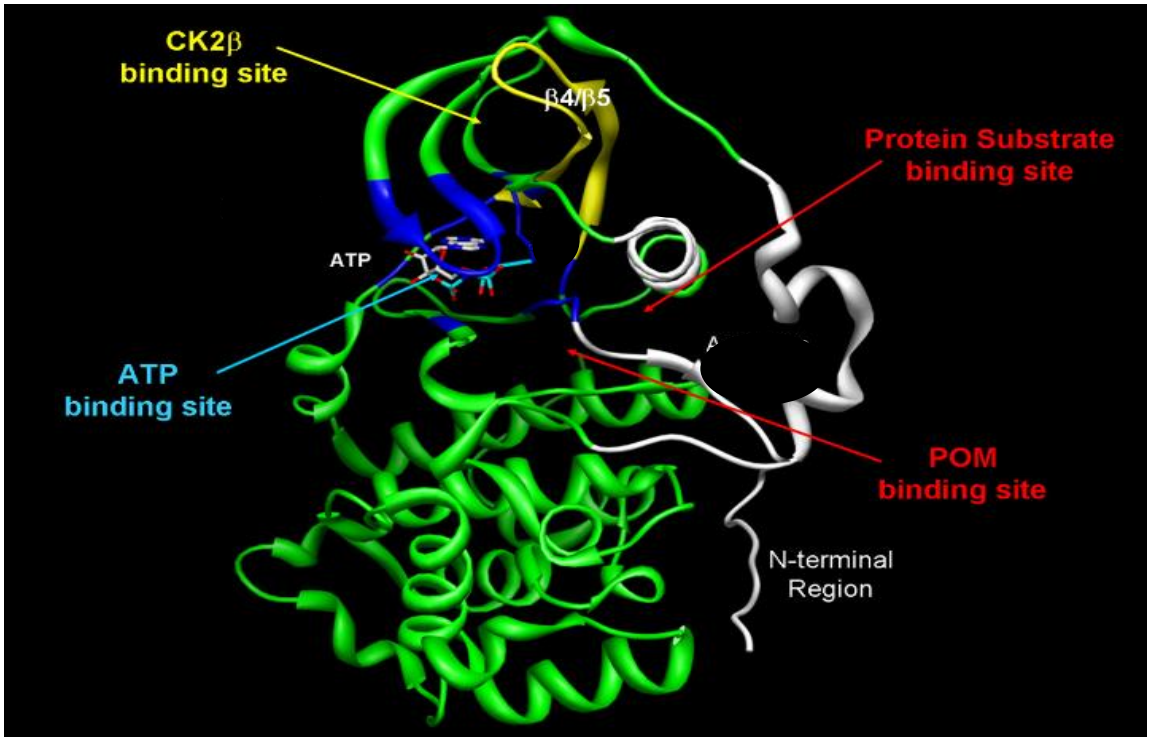


Figure I-3 : Représentation tridimensionnelle de la structure de la CK2, résolue à 1.06 Å (Code PDB : 1JWH) [18].

- ✓ Les groupements phosphorylés, sont représentés par la couleur bleu.
- ✓ Les structures clefs, tel que la lysine (Lys), les segments d'activation, et la région de N-terminal, par la couleur grise.
- ✓ La CK2βpar la couleur jaune.
- ✓ Le site accepteur des POM (Polyoxométallates) en rouge.

III.2 Structure de la sous-unité CK2β

La sous-unité **CK2β** est une protéine régulatrice dépourvue d'activité catalytique. Sa structure cristallographique réduite sous forme d'un dimère, est stabilisée grâce à des interactions hydrophobes entre les résidus acides et basiques au sein d'un motif original en brin de zinc [19].

En effet, les résidus basiques se concentrent dans la partie C-terminal "**brin de zinc**", tandis que les résidus acides se distribuent largement du côté N-terminal. Cette répartition asymétrique des résidus aux extrémités de la protéine génère un potentiel électrostatique et une force de répulsion importante à l'origine de la forme atypique du dimère CK2β [19-22].

Le domaine d'interaction de la sous-unité CK2β avec l'une des sous-unités catalytiques CK2α ou CK2α', est caractérisé par une interface hydrophobe constituée d'un domaine Amino-terminal riche en feuillets β et d'un domaine carboxy- terminal riche en hélices α [19-23].

La figure I-4 représente l'interface du dimère CK2β. Les atomes de zinc sont schématisés par des sphères bleues, le domaine de dimérisation "brin de zinc", est représenté en jaune et les résidus essentiels à la structure du brin de zinc "cysteines" sont représentés en noir.

La figure I-5 représente la structure du monomère **CK2β** (du N-terminal vers C-terminal), en illustrant l'atome de Zinc "Zn^{2+}".

La figure I-6 illustre la structure tridimensionnelle de la surface du dimère CK2β coloré selon le potentiel électrostatique, en bleu les charges positives (résidus basiques) et en rouge les charges négatives (résidus acides).

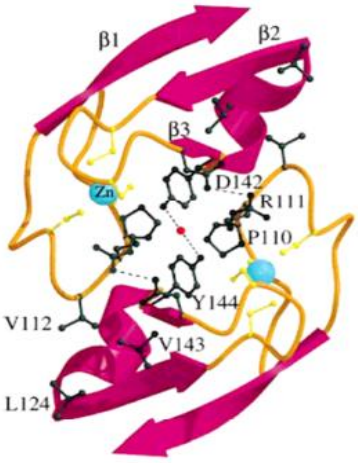


Figure I-4 : Représentation bidimensionnelle de la structure de la sous-unité CK2β (Code PDB : 1JWH) [22].

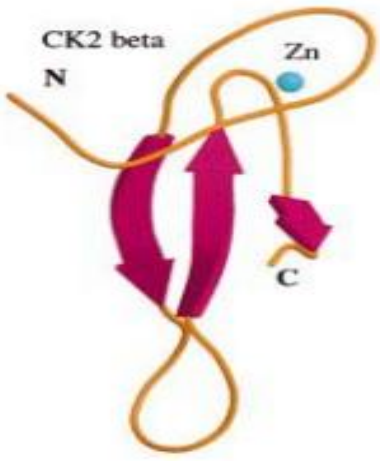


Figure I-5 : Représentation bidimensionnelle de la structure du monomère CK2β (Code PDB : 1JWH) et l'atome de zinc [22].

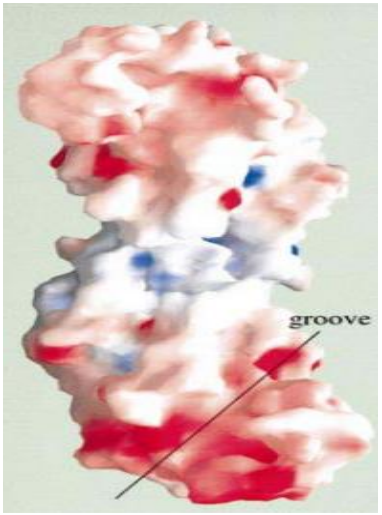


Figure I-6 : Représentation tridimensionnelle de la structure de la sous-unité CK2β (Code PDB : 1JWH) [22].

III.3 Isoformes de la protéine Kinase CK2

La protéine CK2, possède les isoformes suivants :

III.3.1 CK2α

La **CK2α**, est une protéine monomérique (42 KD), présentant une séquence primaire extrêmement conservée au cours de l'évolution. Elle possède une activité constitutive liée au caractère rigide de son segment d'activation qui est maintenu en permanence en position ouverte par de nombreuses interactions électrostatiques et des liaisons hydrogènes avec le lobe N-terminal [21-22].

III.3.2 CK2α'

La sous-unité **CK2α'** de la protéine kinase II, est une isoforme particulière, présente chez tous les eucaryotes, possède une séquence plus courte (37 KD) et un domaine C-terminal différent de celui de la **CK2α** [21-22].

III.3.3 CK2β

La sous-unité CK2β de la protéine kinase II, qui joue un rôle important en morphologie neural et la transmission synaptique [19,22-23].

III.4 Structures et séquences en acides aminés des sous-unités CK2α et CK2β

Les séquences en acides aminés des sous unités CK2α et CK2β, ainsi que leurs structures tridimensionnelles correspondantes, sont représentées sur les figures I-7 ; I-8 et I-9, respectivement [16].

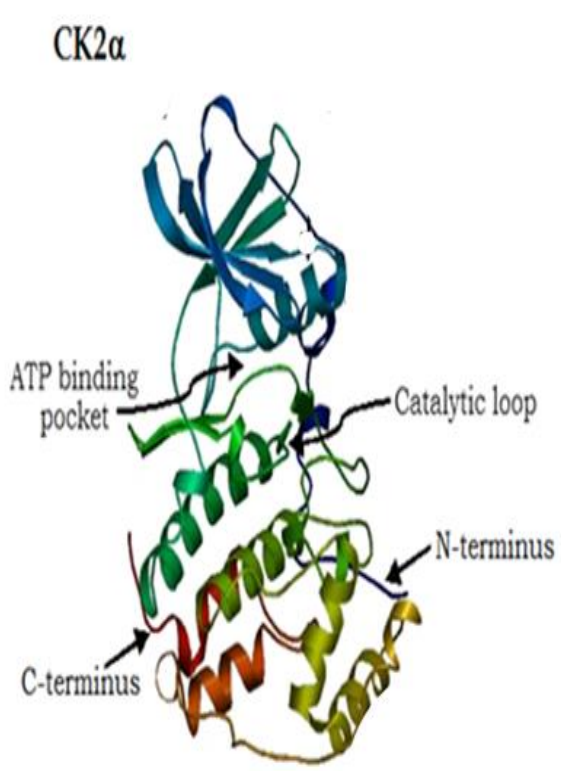


Figure I-7 : Représentation tridimensionnelle de la structure de la sous unité CK2α (Code PDB : 3EED) [16].

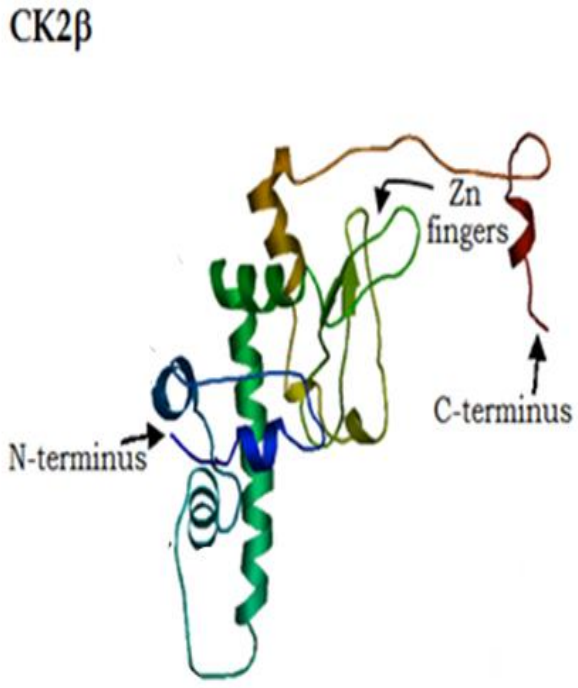


Figure I-8 : Représentation tridimensionnelle de la structure de la sous unité CK2β (Code PDB : 3EED) [16].

```
CK2 beta          VSWISWFCG----------LRGNEFFCEVDEDYIQDKFNLTGLNEQ---VPHYRQA------LDMI----
CK2 alpha         GPVPSRARVYTDVNTHRPREYWDYESHVVEWGNQDDYQLV-RKLGRGKYSEVFEAINITNNEKVVVKILK
CK2 alpha prime   ----SRARVYAEVNSLRSREYWDYEAHVPSWGNQDDYQLV-RKLGRGKYSEVFEAINITNNERVVVKILK

CK2 beta          ----------------------LDLEQSDLI---EQAAEMLYGLIHAR------------YILTNRG
CK2 alpha         PVKKKKIKREIKILENLRGGPNIITLADIVKDPVSRTPALVFEHVNNTDFKQLYQTLTDYDIRF-YMYEI
CK2 alpha prime   PVKKKKIKREVKILENLRGGTNIIKLIDTVKDPVSKTPALVFEYINNTDFKQLYQILTDFDIRF-YMYEL

CK2 beta          IAQMLEKYQQGDFGYCPRVYCENQPMLPIGLSDI--------------PG-----------EAMVK---
CK2 alpha         LKALDYCHSMGIM------------HRDVKPHNVMIDHEHRKLRLIDWGLAEFYHPGQEYNVRVASRYF
CK2 alpha prime   LKALDYCHSKGIM------------HRDVKPHNVMIDHQQKKLRLIDWGLAEFYHPAQEYNVRVASRYF

CK2 beta          LYCPKCMDVY---TPKSSRHHHTDGAYFGTGFPHMLF--------------------------------
CK2 alpha         KG-----PELLVDYQMY--DYSL----DMWSLGCMLASMIFRKEPFFHGHDNYDQLVRIAKVLGTEDLYD
CK2 alpha prime   KG-----PELLVDYQMY--DYSL----DMWSLGCMLASMIFRREPFFHGQDNYDQLVRIAKVLGTEELYG

CK2 beta          ---------------------------------------MVHPEYRP-KRPANQFVPRLYGFKI-----
CK2 alpha         YIDKYNIELDPRFNDILGRHSRKRWERFVHSENQHLVSPEALDFLDKLLRYD-----------HQSRLT
CK2 alpha prime   YLKKYHIDLDPHFNDILGQHSRKRWENFIHSENRHLVSPEALDLLDKLLRYD-----------HQQRLT

CK2 beta          -----------------
CK2 alpha         AREAMEHPYFYTVVKD-
CK2 alpha prime   AKEAMEHPYFYPVVKEQ
```

Figure I-9 : Alignement multiple des séquences en acides aminés des sous unités CK2β, CK2α et CK2α' (Code PDB : 3EED) [16].

IV. Rôle de la protéine CK2

La CK2 est une protéine omniprésente dans toutes les cellules des mammifères. Elle agit sur une protéine en l'activant ou en la désactivant. Cette protéine, exerce un rôle critique pour la survie cellulaire, la régulation des protéines, les processus sécrétoires et pathologies cellulaires, notamment le cancer du sein, ce qui la rend très importante pour la physiologie humaine [12-15].

Des études, ont mis en évidence le rôle déterminant de la CK2 dans la plasticité cellulaire épithéliale [12-15]. En effet, une altération de son activité multifonctionnelle, implique des variations importantes dans la transformation des cellules épithéliales en cellules cancéreuses plus actives (mobiles) à l'origine des métastases.

Le sein, est constitué d'un ensemble de cellules épithéliales cohésives et polarisées. Au cours du développement du cancer, certaines de ces cellules, présentes dans la tumeur maligne, peuvent réactiver un mécanisme cellulaire appelé "transition épithélio- mésenchymateuse"/ *"Epithelial-to-Mesenchymal Transition* : EMT" [15,24].

Ce processus « EMT », est responsable de l'évolution des cellules d'un état épithélial vers un état mésenchymateux caractérisé par une mobilité cellulaire accrue et une forte résistance à certains anticancéreux [24].

Le dérèglement de l'activité de la CK2, observée dans certains cancers du sein, est due à l'expression des marqueurs du mécanisme **« EMT »**. En effet, la diminution de l'expression de la sous-unité régulatrice **CK2β** dans des cellules épithéliales mammaires, conduit à des changements caractéristiques de l'EMT, induisant l'activation et la migration des cellules cancéreuses dans l'organisme [15,24].

Des analyses cellulaires, biochimiques et transcriptomiques, ont permis d'identifier trois familles de facteurs de transcriptions de la CK2 comme étant des cibles de régulation des bio marqueurs de l'EMT (à savoir : "SNAIL/SNAI" ; "ZEB" et "bHLH"). Par conséquent, en induisant la phosphorylation de ces facteurs, l'activité de la protéine-kinase CK2 est nécessaire pour empêcher l'activation du mécanisme "EMT", et ainsi de diminuer le risque de développement de métastases. Cette kinase est donc essentielle au maintien de l'état épithélial des cellules [12,24].

Des études ont démontré l'importance de la sous-unité **CK2β** dans le contrôle de la plasticité cellulaire en supprimant l'**EMT** et en maintenant l'intégrité des cellules épithéliales [24-27].

Par conséquent, la protéine-kinase CK2 est un régulateur clé de l'homéostasie des cellules épithéliales, grâce à ses propriétés de marqueur dans plusieurs types de cancers et à ses qualités en tant que cible thérapeutique.

Des évidences récentes suggèrent que la protéine CK2, puisse exercer un rôle anti-apoptotique, en protégeant les protéines de normalisation contre la dégradation. Cette fonction anti-apoptotique de la CK2 contribuent à sa capacité de participer à la transformation et au tumorigènes (Carcinogenèses) [24,27].

IV.1 Activité catalytique et mécanisme d'activation de la CK2

La protéine kinase CK2 est une enzyme du groupe des transférases. Elle possède une sous-unité catalytique CK2α, dont le rôle est de catalyser le transfert d'un groupement "Phosphate" d'une molécule (ATP/GTP) à un groupement Hydroxyde accepteur "OH", situé sur la chaine latérale d'une serine, d'une thréonine, ou bien d'un groupement phénolique "alcool" d'une tyrosine du substrat [13,16-18,24].

La **"Fonction catalytique"** de la protéine kinase CK2 est primordiale, puisqu'elle entraîne des modifications dans la conformation, la charge et la thermodynamie de la fonction de la protéine cible et de son activité enzymatique (activité biologique) [24,28].

La figure I-10 montre le mécanisme d'activation détaillé de la protéine Kinase CK2 via sa phosphorylation par une autre protéine kinase depuis une molécule d'ATP [24].

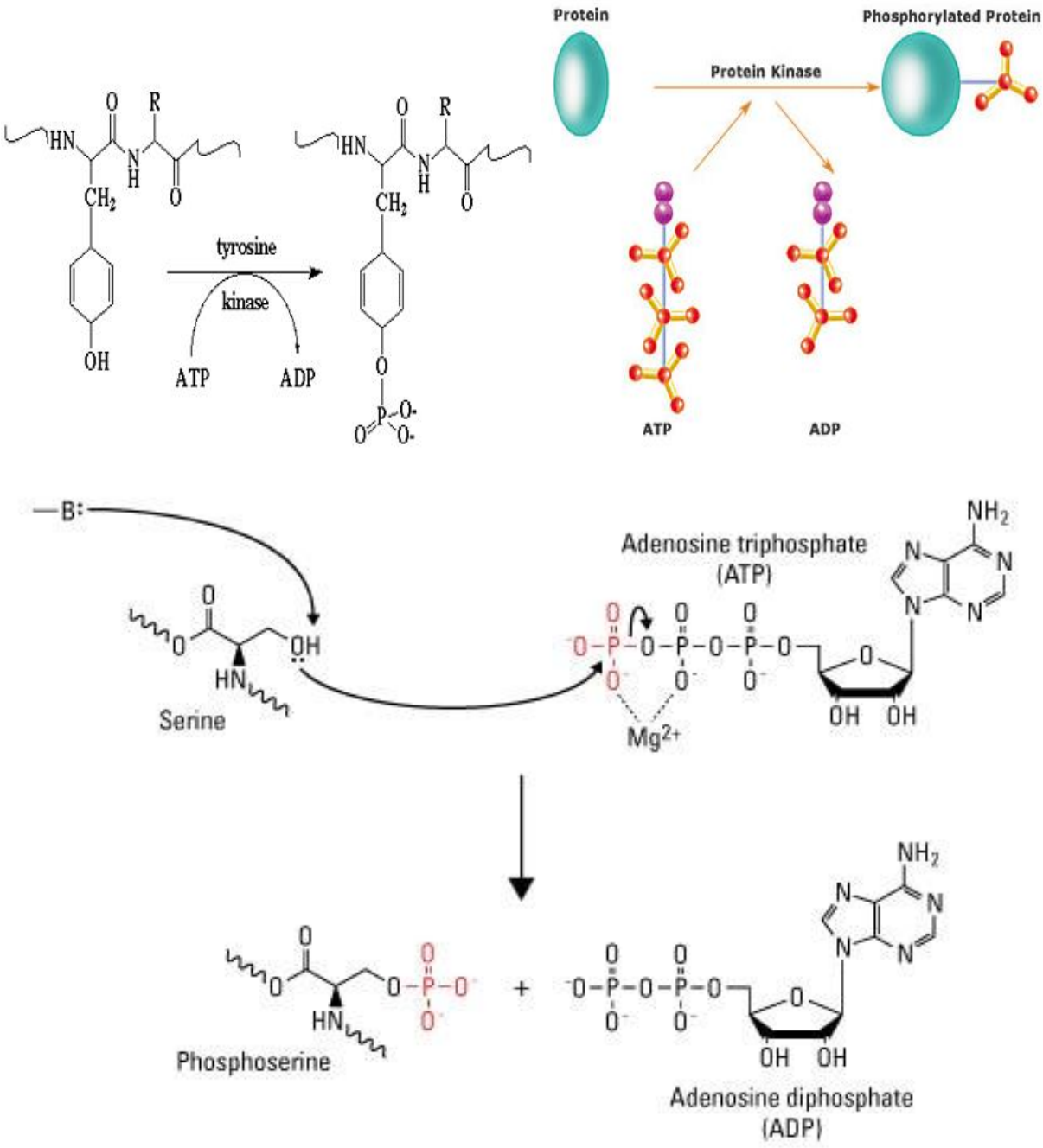


Figure I-10 : Schéma d'activation de la protéine Kinase CK2 via sa phosphorylation par une protéine kinase depuis une molécule d'ATP [24].

IV.2 Site actif de la CK2

Le site catalytique est une partie protéique possédant des propriétés spécifiques pour la reconnaissance du substrat par l'enzyme, qui entre en interactions avec les résidus actifs de cette dernière. En effet, le site actif de la CK2, est situé entre le domaine « N-terminal » riche en feuillets β et le domaine « C- terminal » riche en hélices α [24,28].

La figure I-11 représente la région d'interaction « **Ligand - Site actif** » sur la surface électrostatique du complexe pour la protéine kinase **CK2**. A noter que, le complexe a été obtenu, en utilisant la protéine 4DGM (résolue à 1.65Å), composée de 325 acides aminés, avec son ligand expérimental Apigenine "AGI" de formule brute ($C_{15}H_{10}O_5$) et ayant le poids moléculaire de 270.24 g/mol, à l'aide de logiciel UCSF [29].

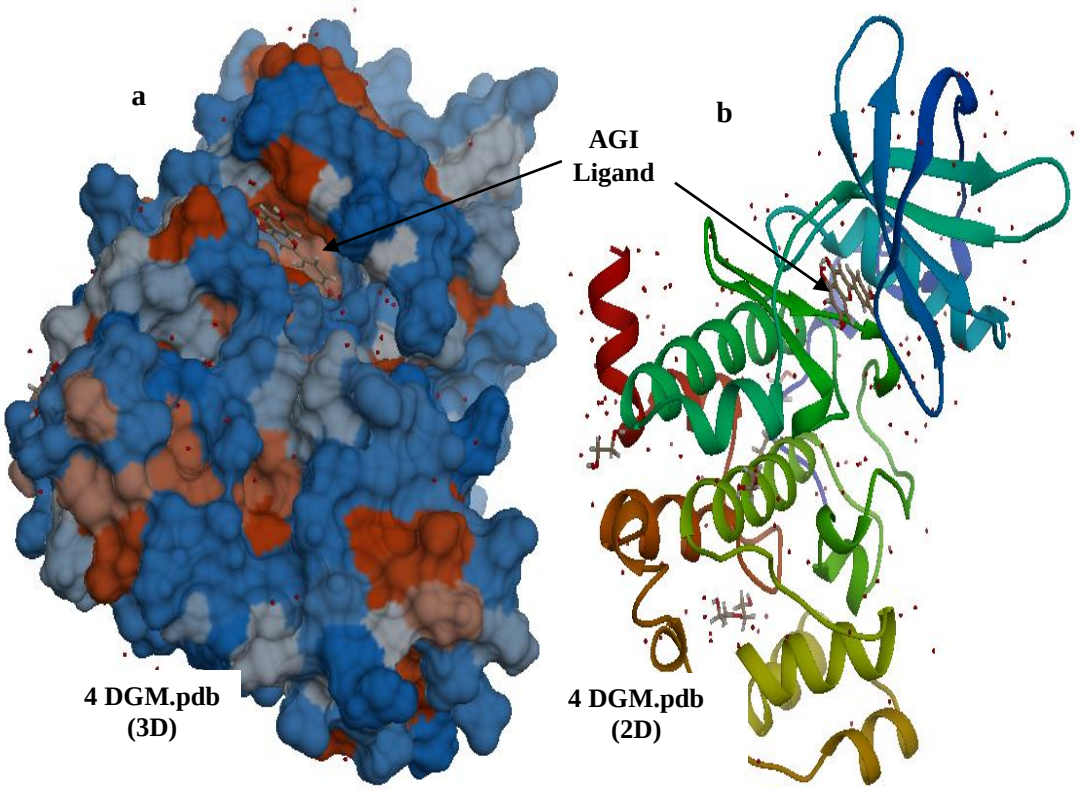


Figure I-11 : Structure tridimensionnel et bi dimensionnelle du ligand AGI dans l'emplacement du site actif pour la protéine CK2(Code PDB : 4DGM) [29].

La surface électrostatique présentée sur la figure, est colorée selon le potentiel électrostatique du bleu profond (charge positive) au rouge (charge négative). Elle illustre les différentes forces de distributions de charge sur la surface externe du complexe (AGI-CK2). L'intérieur du site actif est fortement positif, non seulement en raison du faisceau de base du complexe AGI-CK2, mais aussi par la présence de la boucle de fixation d'ATP.

V. Inhibiteurs de la CK2

L'inhibition de l'association des sous-unités CK2α/CK2β, perturbe la signalisation de la protéine CK2, conduisant à une inhibition de sa prolifération et une induction de l'apoptose de cellules cancéreuses mammaires [17,20]. De ce fait, la recherche de nouveaux inhibiteurs de la CK2, représente une approche innovante qui offre la possibilité de tester pharmacologiquement l'importance de l'inhibition dans les processus de croissance tumorale [17,20].

Le développement des inhibiteurs de la protéine CK2, est le but principal de la recherche thérapeutique sur le cancer. Il existe plusieurs inhibiteurs dont le mode d'action est lié à la parfaite adéquation (haute-affinité) entre leurs structures et le site actif de la protéine CK2 dans lequel ils viennent se placer. La figure I-12, résume les différents types d'inhibiteurs de la protéine kinase CK2.

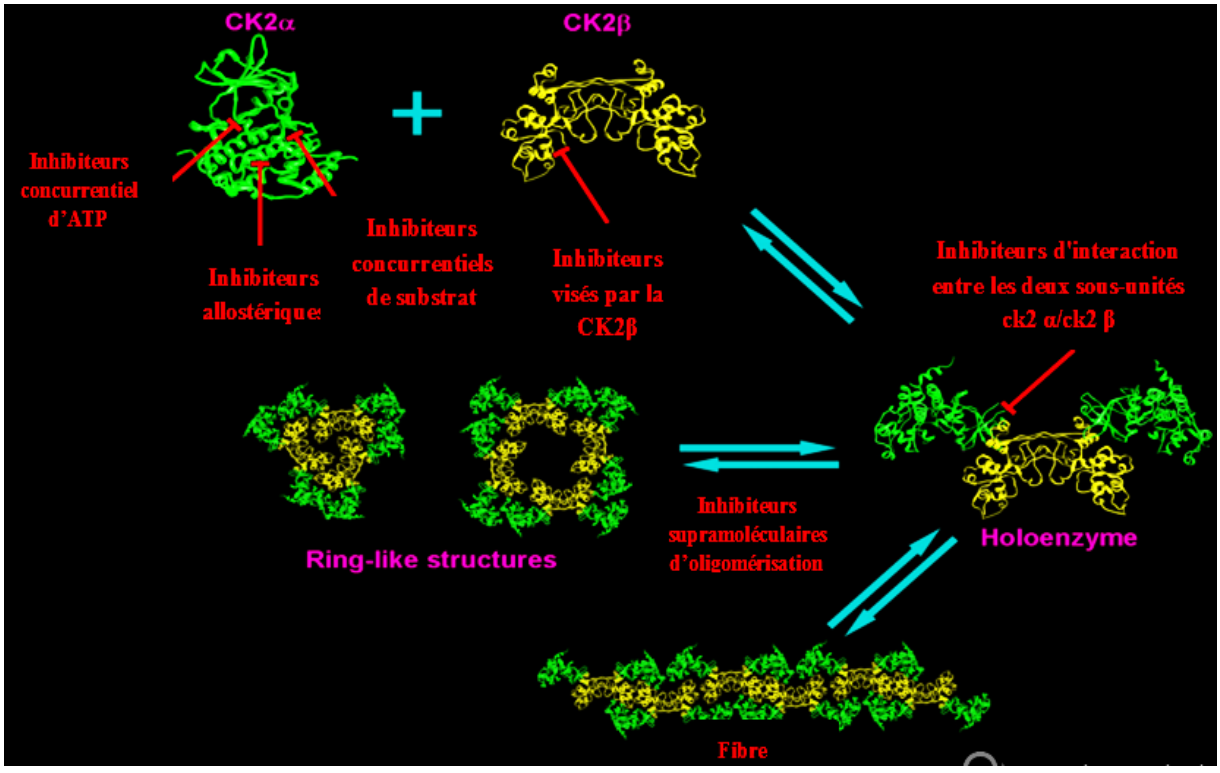


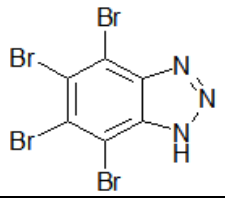
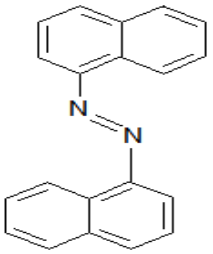
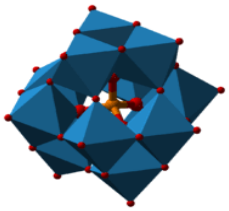
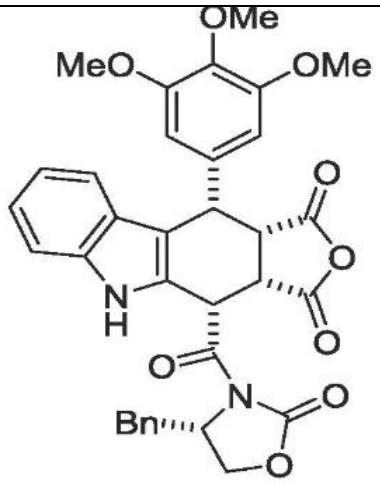
Figure I-12 : Représentation des différents types d'inhibiteurs de la CK2 [20].

Cette organisation moléculaire dynamique, donne plusieurs occasions de viser la superficie de chaque sous-unité avec des inhibiteurs spécifiques [17].

V.1 Les différents types d'inhibiteurs de la CK2

Le tableau I-1, résume les différents types d'inhibiteurs de la CK2, leurs classifications moléculaires, ainsi que les exemples correspondants [20,30-32].

Tableau I-1 : Différents type d'inhibiteurs, classification moléculaire et exemples correspondants [20,30-32].

Types d'inhibiteur	Classification moléculaire	Exemple
Inhibiteur concurrentiel d'ATP (Type I)	**Molécule organique** (pyrido carbazoles et benzopyridoindoles)	**4,5,6,7-Tetra bromo benzotriazole (TBB : $C_6HBr_4N_3$)** 
Inhibiteurs allostériques	**Molécule organique** Azonaphtalènes (Quinobène)	**1,1'-Azonaphthalene ($C_{20}H_{14}N_2$)** 
Inhibiteurs inorganiques	**Molécules inorganiques** (**POM**)	**Poly-oxo-métallates (POMs)** 
Inhibiteurs visés par les substrats	**Peptides cycliques P15**	**PeptideP15** **(CWMSPRHLGTC)**
Inhibiteur visé par CK2β	**Peptide P1**	**PeptideP1** **(GKMNGVLPLAWPSLYLRL)**
Inhibiteurs d'interaction des sous-unités ck2α-ck2β	**Peptides cycliques et Molécules organiques** **W16** (podophyllotoxine indolo-analogue)	

V.1.1 Inhibiteurs concurrentiels (compétitifs) d'ATP

Ce sont des inhibiteurs de Type I développés dans l'objectif d'imiter l'ATP afin d'entrer en compétition avec celui-ci, et se mettre dans le site catalytique de la CK2 en conformation active. Toutes les molécules développées s'inspirent de la partie adénine de l'ATP. Ainsi, tous les inhibiteurs de ce type ont un centre aromatique, faisant une interaction favorable π-π avec les résidus hydrophobes du brin 2 dans le lobe C-terminal et du brin 6 dans le lobe N-terminal, auquel s'ajoute une combinaison parmi deux donneurs et un accepteur de liaisons hydrogènes [20,30].

V.1.2 Inhibiteurs allostériques

Ce sont des Inhibiteurs de **type II** du fait qu'ils reconnaissent une conformation différente de la forme Active, à savoir la conformation inactive du motif Asp-Phe-Gly, nommée « *DFG out* ».

Le déplacement du motif **DFG** change considérablement la morphologie du site catalytique, en donnant accès à une grande poche hydrophobe, exploitée par ce type d'inhibiteurs.

Il est nécessaire que le résidu appelé "***gatekeeper***" soit petit pour permettre à l'inhibiteur de s'étendre du site de liaison de l'ATP jusqu'à la poche hydrophobe.

Il est bien de rappeler que le résidu **gatekeeper** est le segment qui relie les lobes supérieur et inférieur de la poche catalytique, qui bloque partiellement ou totalement une région hydrophobe profondément dans la poche de liaison d'ATP et contribue à la sélectivité des kinases pour les inhibiteurs de petites molécules. Contrairement aux autres types d'inhibiteurs, les inhibiteurs de type II n'occupent pas pleinement la poche de liaison de l'ATP, du fait de la présence de la phénylalanine du motif DFG qui réduit le site [20,30].

V.1.3 Inhibiteurs inorganiques de la protéine CK2 (polyoxométallates)

Les polyoxométallates "POM" est une nouvelle famille d'inhibiteurs de la CK2. Les POMs sont des molécules, inorganiques, agrégats nanométriques anioniques de métaux de transition (vanadium, molybdène, tungstène) et d'oxygène. Les POMs sont caractérisées par leurs structures différentes de celles des inhibiteurs organiques (Absence du squelette en carbone). Elles possèdent des propriétés intéressantes en catalyse chimique ainsi que des activités antivirales, antibiotiques et anticancéreuses remarquables. Elles sont considérées,

comme les inhibiteurs les plus puissants, vu leurs grandes affinités et spécificités pour la CK2 [20, 30-32].

Les POMs agissent sur l'enzyme selon un mode d'inhibition totalement inédit, en se fixant sur un site en position distale par rapport au site catalytique, nommé : "*exo site*" [20, 30-32].

V.1.4 Inhibiteurs concurrentiels du substrat

Il s'agit d'inhibiteurs de **type III**. Ces molécules se positionneraient soit sur le site de fixation du substrat, soit dans une poche allostérique. L'inhibiteur peut ainsi s'étendre jusqu'à la poche hydrophobe sans être gêné par le résidu gatekeeper [20,30].

V.1.5 Inhibiteurs visés par la CK2β

La CK2β possède une activité de base pour la fixation du substrat. Cependant, une perte fonctionnelle de cette activité est mortelle, puisqu'elle conduit à des dérèglements phénotypiques des interactions dans lesquelles cette protéine est impliquée. Les inhibiteurs de ce type ayant une structure chimique d'un peptide, se lient avec une grande affinité dans le N-terminal de CK2β sans perturber la formation de l'holoenzyme CK2 [20,30].

V.1.6 Inhibiteurs d'interaction entre les deux sous - unités CK2α et CK2β

Les interactions protéine-protéine jouent un rôle clé dans le signal cellulaire. La sous-unité CK2β améliore la signalisation de la protéine CK2, en apportant une meilleure flexibilité pour la fixation aux substrats. Trois composés de ce type ont été développés afin de bloquer l'interaction entre les deux sous-unités de la CK2. **W16** est le plus puissant inhibiteur de cette classe, il se lie au site de liaison entre la **CK2α**et la **CK2β** ou sa dérivé **CK2β'**. Ces composés provoquent une inhibition non compétitive pour la protéine kinase CK2 [20,30].

VI. Flavonoïdes, nouvelle famille d'inhibiteurs de la CK2

La CK2 est impliquée dans de nombreux processus de phosphorylation physiologique de plus de 300 substrats. Il existe des preuves scientifiques abondantes, montrant que les flavonoïdes sont considérés comme des inhibiteurs de la CK2, qui pourraient être utilisés comme des agents pharmaceutiques afin de traiter les maladies cancéreuses, les infections virales et les maladies inflammatoires [33-36].

Les flavonoïdes sont des composés phénoliques "*polyphénols*", qui appartiennent à la classe d'inhibiteurs du type I "Inhibiteurs Concurrentiels d'ATP". Ce sont des pigments naturels répandus dans tout le règne végétal où ils interviennent comme des filtres UV et comme agents de protection contre des organismes pathogènes. Ils sont apportés au niveau de l'organisme grâce à la consommation quotidienne des fruits, des légumes, des céréales et de boissons (Thé vert, café, jus…etc.) [33-36].

Les flavonoïdes couvrent une large gamme de couleur, allant du rouge à l'ultraviolet en passant par le jaune. Leur couleur dépend de leur structure mais aussi de l'acidité du milieu (pH).

Les jaunes viennent des Chalcones, Aurones et Flavonols jaunes, les rouges et les mauves des Anthocyanosides et les bleus, trouvent leurs origines dans les Co-pigments Flavones-Anthocyanosides [37-42].

VI.1 Structure générale des flavonoïdes

Les flavonoïdes possèdent tous un même squelette de base à quinze atomes de carbones, constitué de deux cycles benzéniques C6 (A et B) reliés par une chaîne propyle C3 qui peut être complétée par une fonction éther formant un hétérocycle (cycle C).

La structure d'un flavonoïde s'organise toujours autour d'un squelette : C6-C3-C6 (1,3-diphénylpropane), comme illustré sur la figure I-13 [42- 44].

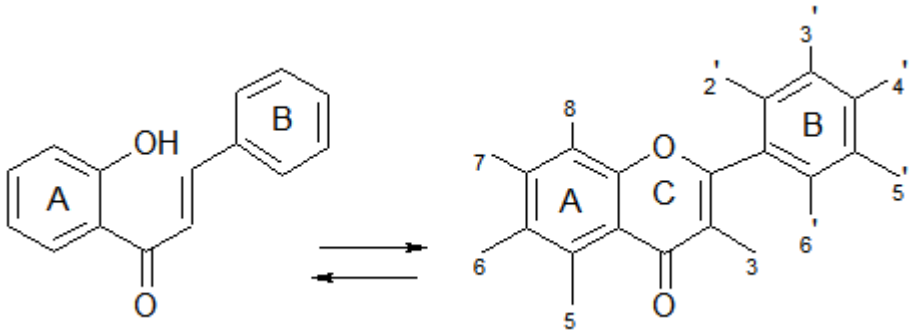


Figure I-13 : Structure de base des flavonoïdes "Enchainement C6-C3-C6'" [42].

VI.2 Classification des flavonoïdes

Il existe six classes de flavonoïdes : **flavones**, **flavonols** , **dihydroflavonols** ou **flavanonol** , **flavanones** , **Aurones** (2-benzylidène-coumaranones) , et les **Chalcones** [44-45].

- ✓ Les **flavones** et les **flavonols**, représentent environ **80%** des flavonoïdes.
- ✓ Les **flavonols** , sont les **antioxydants** les plus efficaces des flavonoïdes. Le cycle A de ces deux types de molécules est substitué par deux hydroxyles phénoliques en C5 et en C7.

Ces hydroxyles peuvent être libres ou estérifiés. Alors que, le cycle B est substitué en C4' ou di-substitué en C3' et C4', par des groupements OH ou méthoxyles (OCH_3). Les flavonols se distinguent des flavones par un **OH** en **C3** [42].

- ✓ Les **Flavanones** et les **dihydroflavonols**, sont caractérisés par l'absence de la double liaison entre les carbones **C2** et **C3** et par la présence de centre d'asymétrie. Les dihydroflavonols se distinguent des flavanones par l'hydroxylation de la position **C3** [42].
- ✓ Les **Aurones** sont caractérisées par une structure de 2-benzylidène coumarone [42-43].
- ✓ Les **chalcones**, ont leur noyau pyranique central ouvert, ils sont constitués par deux unités aromatiques reliées par une chaîne tri-carbonée, cétonique et insaturée. Le cycle B est fréquemment non substitué, alors que les substitutions sur le cycle A sont plus souvent identiques à celles des autres flavonoïdes [42,46].

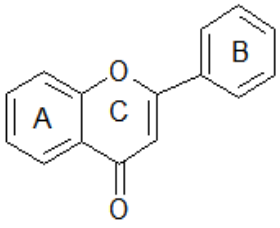
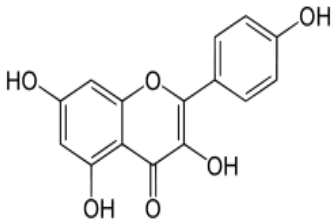
Les flavonoïdes se divisent en plusieurs sous-classes, se distinguant par une diversité fonctionnelle au niveau des positions **2**, **3** et **4** du cycle **C**, comme illustré sur la figure I-13.

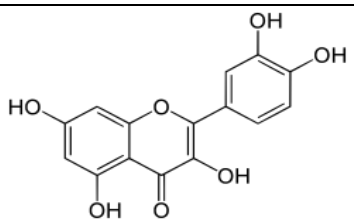
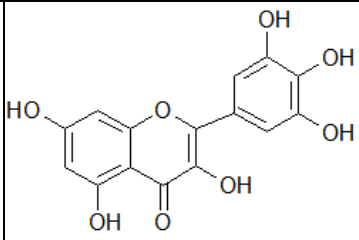
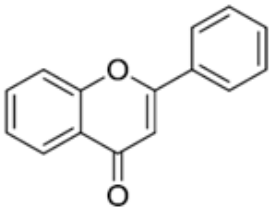
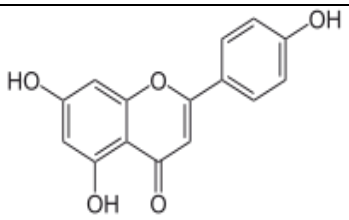
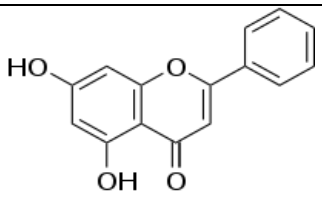
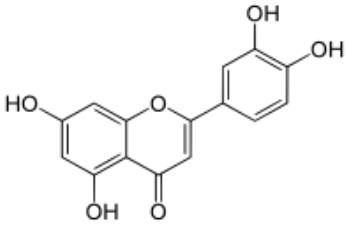
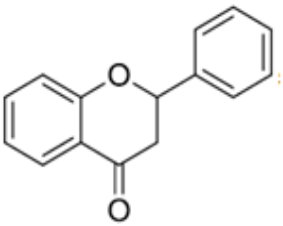
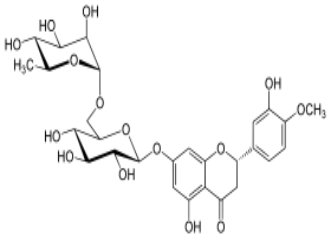
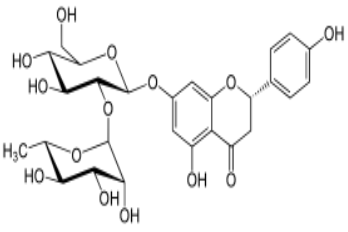
Pour chaque classe de flavonoïde les molécules se diffèrent entre elles par le nombre et la position des groupements hydroxyle, la présence de substituant sur la **Génine** (flavonoïde non substitué) et le degré de polymérisation.

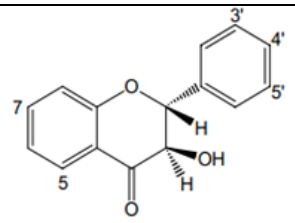
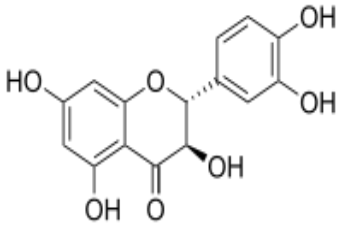
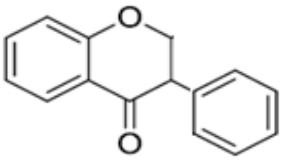
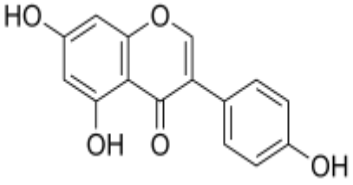
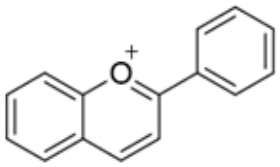
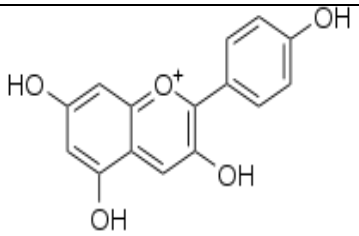
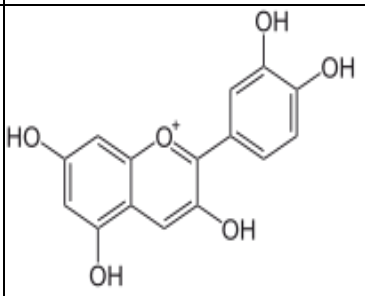
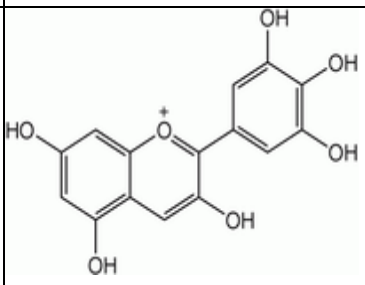
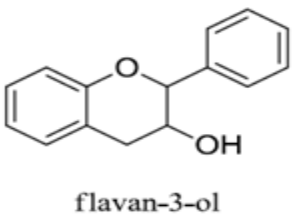
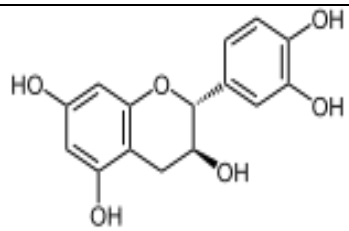
Les formes les plus rencontrées sont issues d'hydroxylation, alkylation (méthylation ou prénylation), glycosylation et acylation soit sur la génine (aglycone), soit sur un groupe hydroxyle d'un ose (monosaccharide) [42-46].

Le tableau I-2, regroupe les différentes classes, caractéristiques, structures et sources des flavonoïdes [42, 46-49].

Tableau I-2 : Différentes classes des flavonoïdes, structures, nomenclature IUPAC et leurs distributions dans les fruits et légumes (source naturelle) [42, 46-49].

Structure des différentes classes de flavonoïdes	**Exemples**	**Nomenclature IUPAC**	**Structure moléculaire Substitutions en positions (5 6 7 3' 4' 5')**	**Source naturelle**
 **Flavonol**	**Kaempferol ($C_{15}H_{10}O_6$)**	**3,4′,5,7-tetrahydroxyflavone**		Pomme, pamplemousse, radis, endive, brocoli, oignon, chou, laitue, raisin, olive, thé noir

	Quercetin ($C_{15}H_{10}O_7$)	**2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-4H-chromen-4-one**		
	Myricetin ($C_{15}H_{10}O_8$)	**3,5,7-Trihydroxy-2-(3,4,5-trihydroxyphenyl) -4-chromenone**		
 flavone	**Apigenin ($C_{15}H_{10}O_5$)**	**5,7-dihydroxy-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one**		Pomme, céleri, grains de céréale, herbes aromatiques (persil, romarin,thym).
	Chrysine ($C_{15}H_{10}O_4$)	**5,7-dihydroxy-2-phénylchromèn-4-one**		
	Lutéoline ($C_{15}H_{10}O_6$)	**5,7-dihydroxy-2-(3,4-dihydroxyphényl) -chromèn-4-one**		
 flavanone	**Hespéridine ($C_{28}H_{34}O_{15}$)**	**(2S) -5-hydroxy-2-(3-hydroxy-4-methoxyphenyl) -4-oxo-3,4-dihydro-2H-chromen-7-yl 6-O-(6-deoxy-α-L-mannopyranosyl) -β-D-glucopyranoside**		Citrus et agrumes
	Naringénine ($C_{27}H_{32}O_{14}$)	**7-[[2-O-(6-Deoxy-α-L-mannopyranosyl) -β-D-glucopyranosyl]] oxy] -2,3-dihydro-5-hydroxy-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one**		

 Dihydroflavonols	**Taxifoline ($C_{15}H_{12}O_7$)**	**2R,3R) -2-(3,4-dihydroxyphenyl) -3,5,7-trihydroxy-2, 3-dihydrochromen-4-one**		Les agrumes
 isoflavone	**Génistéine ($C_{15}H_{10}O_5$)**	**5,7-dihydroxy-3-(4-hydroxyphényl) -4H-chromén-4-one**		Soja
 anthocyane	**Pélargonidine ($C_{15}H_{11}O_5$)**	**2-(4-hydroxyphényl) chroménylium-3,5,7-triol**		Orage, pomme, fleurs et raisin
	Cyanidine ($C_{15}H_{11}O_6$)	**2-(3,4-dihydroxyphényl) chroménylium-3,5,7-triol**		
	Delphinidine ($C_{15}H_{11}ClO_7$)	**3,5,7-Trihydroxy-2-(3,4,5-trihydroxyphényl) chrométium**		
 flavan-3-ol	**Catéchine ($C_{15}H_{14}O_6$)**	**(2R,3S) -2-(3,4-dihydroxyphenyl) -3,4-dihydro-2H-chromene-3,5,7-triol**		(Même source que les flavanols)

VI.3 Activité anti-cancéreuse des flavonoïdes et exemples

L'effet anticancéreux des flavonoïdes observé in vitro, a été confirmé par plusieurs études épidémiologiques. En effet, la consommation des flavonoïdes diminue le risque et

l'incidence de plusieurs types de cancers. Ceci est confirmé par leurs capacité d'activer les mécanismes naturels de la défense anticancéreuse, en bloquant les premières phases d'initiation cancéreuse avec l'interception et le métabolisme des agents mutagènes, par la synthèse des enzymes spécifiques qui peuvent oxyder et convertir les substrats mutagènes en espèces hydrolysables facilement excrétées hors des cellules [46,50-53].

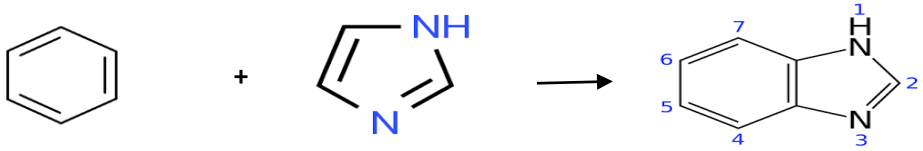
L'**Apigénine** ou Apigénol "UICPA : 5,7-dihydroxy-2-(4-hydroxyphenyl) -4H-1-benzopyran-4-one", est un inhibiteur sélectif de la CK2, qui appartient à la famille des flavones. Il possède des propriétés anti-inflammatoires , anticancéreuses et anti oxydantes , de fait qu'il inhibe l'angiogenèse anormale des cellules cancéreuses , prévient le cancer de la prostate et empêche sa croissance (induction l'apoptose) [50-51].

La **quercétine** ou quercéto "UICPA : 3,3',4',5,7-pentahydroxy-2-phénylchromén-4-one ou 2-(3,4-dihydroxyphényl) -3,5,7-trihydroxy-4H-chromén-4-one" : C'est un inhibiteur de type flavonol, considéré comme le plus actif des flavonoïdes et de nombreuses plantes médicinales, par son efficacité due à sa forte concentration en quercétol. En effet, cet inhibiteur, prévient la cancérogenèse, surtout le cancer de la peau et du colon. La présence de 20 % de quercétine dans l'alimentation chez les animaux diminue le cancer du côlon et prévient l'apparition des cryptes anormales [41- 44,49].

La **lutéoléine** "UICPA : 5,7-dihydroxy-2-(4-hydroxyphenyl) 4H chromen-4-one)" : Un flavonoïde commun qui existe dans de nombreuses plantes utilisées en médecine traditionnelle chinoise vu ses effets biologiques anti-inflammatoires, anti-allergiques, anticancéreux et antioxydant biochimique. En effet, la propriété anticancéreuse de la lutéoline est associée à l'induction de l'apoptose et à l'inhibition de la prolifération des cellules cancéreuses, des métastases et de l'angiogenèse [54].

VII. Benzomidazoles

Les benzimidazoles sont des composés organiques aromatiques bicycliques (hétérocyclique), formés par la fusion des noyaux benzène et imidazole, comme illustré sur la Figure I-14 [54].



Benzène **Imidazole** **Benzimidazole**

Figure I-14 : Mécanisme de formation de Benzimidazole [55].

Les dérivés du benzimidazole ont révélé des activités : anti-inflammatoires, analgésiques, antivirales, antibactériennes, antifongiques et antitumoral. Ces molécules possèdent des structures privilégiées en chimie des médicaments, et une grande attention particulière en tant que cible thérapeutique pour le développement de pharmacophores [55-57].

En conclusion, nous pouvons dire que les nombreuses propriétés thérapeutiques des flavonoïdes et des benzomidazoles nous ont poussé à l'exploitation de ces deux familles de molécules dans le développement de nouveaux inhibiteurs de la protéine CK2, cible thérapeutique potentielle dans le traitement du cancer.

VIII. Références Bibliographiques

[1] Siegel R L., Miller K D., Jemal A., (2020). Cancer statistics, 2020. CA Cancer J Clin, 70 (1) :7-30. http://doi.org/10.3322/caac.21590

[2] Gowda C., Sachdev M., Muthusami S., Kapadia M., Petrovic-Dovat L., Hartman M., Ding Y., Song C., Payne J L., Tan B H., Dovat S., (2017). Curr Pharm Des., 23(1):95-107.

[3] Laramas M., Pasquier D., Filhol O., Ringeisen F., Descotes J L., Cochet C., (2007). Nuclear localization of protein kinase CK2 catalytic subunit (CK2α) is associated with poor prognostic factors in human prostate cancer. Eur J Cancer, 43(5):928-34. http://doi.org/10.1016/j.ejca.2006.11.021

[4] Yenice S., Davis A T., Goueli S A., Akdas A., Limas C., Ahmed K., (1994). Prostate 24(1):11–16. https://doi.org/10.1002/pros.2990240105

[5] Charoenrat P O., Rusch R., Talbot S G., (2004). Nuclear casein kinase 2 (CK-2) activity in human normal, benign hyperplastic, and cancerous prostate. InderpalSarkaria, Agnes Viale, Nicholas Socci, Ivan Ngai, Pulivarthi Rao and Bhuvanesh Singh, Clinical Cancer Research, 10(17):5792-5803. http://doi.org/1078-0432.CCR-03-0317

[6] Kim J S., Eom J I., Cheong J W., Choi A J., Lee J K., Yang W I., and Min Y H., (2007). Protein kinase CK2α as an unfavorable prognostic marker and novel therapeutic target in acute myeloid leukemia. Clinical Cancer Research, 13(3):1019-1028. http://doi.org/1158/1078-0432.CCR-06-1602

[7] Stalter G., Siemer S., Becht E., Ziegler M., Remberger K., Issinger O G., (1994). Asymmetric Expression of Protein Kinase CK2 Subunits in Human Kidney Tumors. BiochemBiophys Res Commun 202(1):141–147. https://doi.org/10.1006/bbrc.1994.1904

[8] Scaglioni P P., Yung T M., Cai L F., Erdjument-Bromage H., Kaufman A J., Singh B., Teruya-Feldstein J., Tempst P., Pandolfi P P., (2006). A CK2-Dependent Mechanism for Degradation of the PML Tumor Suppressor. Cell 126(2):269–283. https://doi.org/10.1016/j.cell.2006.05.041

[9] Hanks S. K., et Hunter T., (1995). The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. The FASEB Journal, 9(8), 576- 596. http://doi.org/10.1096/fasebj.9.8.7768349

[10] Manning G., Whyte D B., Martinez R., Hunter T., Sudarsanam S., (2002). The Protein Kinase Complement of the Human Genome. Science, 298 (5600):1912-34. http://doi.org/10.1126/science.1075762

[11] Pietsch M., Viht K., Schnitzler A., Ekambaram R., Steinkrüger M., Enkvist E., Nienberg C., Nickelsen A., Lauwers M., Jose J., Uri A., Niefind K., (2020). Unexpected CK2β-antagonistic functionality of bisubstrate inhibitors targeting protein kinase CK2. Bioorganic Chemistry, 96:103-608. http://doi.org/10.1016/j.bioorg.2020.103608

[12] Ziercher L, (2007). Étude fonctionnelle de la sous-unité régulatrice de la protéine kinase CK2 dans les cellules souches embryonnaires murines : survie cellulaire, Oligo d'endogenèse et cancérogenèse. Thèse de doctorat : Biologie Cellulaire et Moléculaire. Université Joseph Fourier – Grenoble 1, p.75-101.

[13] Raaf J., Brunstein E., Issinger G O., Niefind K., (2008). The interaction of CK2α and CK2β, the subunits of protein kinase CK2, requires CK2β in a preformed conformation and is enthalpically driven. Protein Science, 17(12): 2180–2186. http://doi.org/10.1110/ps.037770.108

[14] Golub A G., Bdzhola V G., Briukhovetska N V., Balanda A O., Kukharenko O P., Kotey I M., Ostrynska O V., Yarmoluk S M., (2011). Synthesis and biological evaluation of substituted (thieno[2,3-d]pyrimidin-4-ylthio)carboxylic acids as inhibitors of human protein kinase CK2. European Journal of Medicinal Chemistry, 45(3): 870-876. http://doi.org/10.1016/j.ejmech.2010.12.025

[15] Litchfield D W., (2003). Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. Biochem journal, 369 (1):1-15. http://doi.org/10.1042/BJ20021469

[16] Olsten M E K., Litchfield D W., (2004). Order or chaos? An evaluation of the regulation of protein kinase CK2. Biochem Cell Biol., 82(6):681-93. http://doi.org/10.1139/o04-116

[17] Lozeman F J., Litchfield D W., Piening C., TakioK.,Walsh K A, and Krebs E G., (1990). Isolation and characterization of human cDNA clones encoding the alpha. and the alpha' subunits of casein kinase II. Biochemistry, 29 (36) : 8436 – 8447. http://doi.org/10.1021/bi00488a034

[18] Kinoshita T., Nakaniwa T., Sekiguchi Y., Sogabe Y., Sakurai A., Nakamura S. and Nakanishi I., (2013). Crystal structure of human CK2[alpha] at 1.06 Å resolution. Journal of Synchrotron Radiation, 20 (6): 974-979. http://doi.org/10.1107/S0909049513020785

[19] Raaf J, Brunstein E, Issinger O G., Niefind K., (2008). The CK2α/CK2β Interface of Human Protein Kinase CK2 Harbors a Binding Pocket for Small Molecules. Chemistry Biology, 15(2):111-7. http://doi.org/10.1016/j.chembiol.2007.12.012

[20] Prudent R., et Cochet C., (2009). New Protein Kinase CK2 Inhibitors: Jumping out of the Catalytic Box. Chemistry & Biology, 16(2) :112-120. http://doi.org/10.1016/j.chembiol.2009.01.004

[21] Chantalat L., Leroy D., Filhol O., Nueda A., Benitez M J., Chambaz E M.,Claude Cochet C., Dideberg O., (1999). Crystal structure of the human protein kinase CK2 regulatory subunit reveals its zinc finger-mediated dimerization. The EMBO Journal, 18 (11) :2939-2940. http://doi.org/10.1093/emboj/18.11.2930

[22] Canton D. A., Zhang C., et Litchfield D. W., (2001). Assembly of protein kinase CK2: investigation of complex formation between catalytic and regulatory subunits using a zinc-finger-deficient mutant of CK2beta. Biochemical Journal, 358(1): 87–94. http://doi.org/10.1042/0264-6021:3580087

[23] Mingzi C., et Jonathan C A., (1997). The β subunit of CKII negatively regulates Xenopus oocyte maturation. Proc NatlAcadSci U S A, 94(17): 9136–9140. http://doi.org/10.1073/pnas.94.17.9136

[24] DESHIERE A., (2010). Régulation du facteur de transcription SNAIL1 au cours de la Transition Épithélio-Mésenchymateuse dans les cellules mammaires MCF10A : implication de la protéine-kinase CK2.Thèse doctorale : Biodiversité du Développement et Oncogenèse. Université de GRENOBLE, 216 p. Régulation de facteur de transcription Snail 1 EMT

[25] Schittek B., et Tobias Sinnberg T., (2014). Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. Molecular Cancer, 13(1):231. http://doi.org/10.1186/1476-4598-13-231

[26] Li J., Gao K., Cai S., Liu Y., Wang Y., Huang S., Zha J., Hu W., Yu S., Yang S., Xie H., Yan H., Wang J., Wu Y.& Jiang Y., (2019). Germline de novo variants in CSNK2B in Chinese patients with epilepsy. Germline de novo variants in CSNK2B in Chinese patients with epilepsy. Journal Scientific reports, 9 (1):17909. http://doi.org/10.1038/s41598-019-53484-9

[27] Deshiere A., Duchemin-Pelletier E., Spreux E., Ciais D., Combes F., Vandenbrouck Y., Couté Y., Mikaelian I., Giusiano S., Charpin C., Cochet C., Filhol O., (2012). Unbalanced expression of CK2 kinase subunits is sufficient to drive epithelial-to-mesenchymal transition by Snail1 induction. Oncogene, 32(11) :1373-83. http://doi.org/10.1038/onc.2012.165

[28] Lemmon M A., et Schlessinger J., (2010). Cell Signaling by Receptor Tyrosine Kinases. Cell, 141(7) : 1117-1134. http://doi.org/10.1016/j.cell.2010.06.011

[29] Eric F., Pettersen Thomas D., Goddard Conrad C., Huang Gregory S., Couch Daniel M., Greenblatt Elaine C., Meng T.,Ferrin E., (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry, 25 (3): 1605–1612 http://doi.org/10.1002/jcc.20084

[30] Leproult É., (2011). Thèse de doctorat : Aspect Moléculaires et Cellulaires de la Biologie. Université de Strasbourg, 228 p.

[31] Hasenknopf B., (2005). POLYOXOMETALATES: INTRODUCTION TO A CLASS OF INORGANIC COMPOUNDS AND THEIR BIOMEDICAL APPLICATIONS. Frontiers in Bioscience review, 10 (1):275-287. http://doi.org/10.2741/1527

[32] Toshihiro Yamase., (2005). Anti-tumor, -viral, and -bacterial activities of polyoxometalates for realizing an inorganic drug Journal of Materials. Chemistry,15(45): 4773-4782 http://doi.org/10.1039/B504585A

[33] Harborne, J. B., (1995). The Flavonoids: Advances in Research Since 1986 (Harborne, J. B.). Journal of Chemical Education, 72 (3): A73. http://doi.org/10.1021/ed072pA73.11

[34] Kris Morreel, Geert Goeminne, VéroniqueStorme, Lieven Sterck, John Ralph, WouterCoppieters, Peter Breyne, Marijke Steenackers, Michel Georges, Eric Messens, WoutBoerjan, (2006). Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. The plant journal, 47(2):224-37. http://doi.org/10.1111/j.1365-313X.2006.02786.x

[35] Middleton Jr E, Kandaswami C &Theoharides TC., (2001). The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. Pharmacological Reviews, 52(4):673-751. PMID: 11121513

[36] DimitriosI.Tsimogiannis, VassilikiOreopoulou., (2006). The contribution of flavonoid C-ring on the DPPH free radical scavenging efficiency. A kinetic approach for the 3′,4′-hydroxy substituted members. Innovative Food Science & Emerging Technologies, 7(1-2): 140-146. http://doi.org/10.1016/j.ifset.2005.09.001

[37] Havsteen B., (2002). The biochemistry and medical significance of the flavonoids. Pharmacology & Therapeutics, 96(2-3):67-202. http://doi.org/10.1016/S0163-7258(02)00298-X

[38] Edenharder R.,Grünhage D., (2003). Free radical scavenging abilities of flavonoids as mechanism of protection against mutagenicity induced by tert-butyl hydroperoxide or cumene hydroperoxide in Salmonella typhimurium TA102. Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 540 (1):1-18. http://doi.org/10.1016/s1383-5718(03)00114-1

[39] Yao L H., Jiang Y M.,Shi J., Tomás-Barberán F A., Datta N., Singanusong R., Chen SS., (2004). Flavonoids in Food and Their Health Benefits.Plant Foods for Human Nutrition, 59(3):113-22.http://doi.org/10.1007/s11130-004-0049-7

[40] Ono E., Hatayama M., Isono Y., Sato T., Watanabe R., Yonekura-Sakakibara K., Fukuchi-Mizutani M., Tanaka Y., Kusumi T., Nishino T., Nakayama T., (2006).Localization of a flavonoid biosynthetic polyphenol oxidase in vacuoles. The planet journal, 45(2):133-143. http://doi.org/10.1111/j.1365-313X.2005.02625.x

[41] Yasuo Mukohata, Shoji Nakabayashi, Mitsuhiro Higashida, (1978). Quercetin, an energy transfer inhibitor in photophosphorylation. FEBS Letters, 85 (2): 215-218. http://doi.org/10.1016/0014-5793(78)80458-X

[42] Bruneton J., (1999). In Pharmacognosie, Phytochimie, Plantes Medicinales, 3ème édition; Tec et Doc: Paris, France. Pp369-404.

[43] Pier-Giorgio P., (2000). Flavonoids as Antioxidants. Journal of Natural Products, 63(7): 1035-1042. http://doi.org/10.1021/np9904509

[44] Kris Morreel, Geert Goeminne, VéroniqueStorme, Lieven Sterck, John Ralph, WouterCoppieters, Peter Breyne, Marijke Steenackers, Michel Georges, Eric Messens, WoutBoerjan. (2006). Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. the plant journal, 47(2):224-237. http://doi.org/10.1111/j.1365-313X.2006.02786.x

[45]N.C.Cook, S.Samman, (1996). Flavonoids—Chemistry, metabolism, cardioprotective effects, and dietary sources. The Journal of Nutritional Biochemistry, 7(2): 66-76.http://doi.org/10.1016/S0955-2863(95)00168-9

[46] Di C G., Mascolo N., Izzo A A., Capasso F., (1999). Flavonoids: Old and new aspects of a class of natural therapeutic drugs. Life Sciences, 65 (4): 337-353. https://doi.org/10.1016/S0024-3205(99)00120-4

[47] Hollman P C H., et Katan M B., (1999). Dietary Flavonoids: Intake, Health Effects and Bioavailability. Food and Chemical Toxicology 1999, 37, (9-10), 937-942. https://doi.org/10.1016/s0278-6915(99)00079-4

[48] Liu H., Zhang L., et Lu S., (2012). Evaluation of Antioxidant and Immunity Activities of Quercetin in Isoproterenol-Treated Rats. Molecules, 17 (4) : 4281–4291. https://doi.org/10.3390/molecules17044281

[49] Gonzalez-Paramas A M., Santos-Buelga C., Duenas M.,et Gonzalez-Manzano S., (2011). Analysis of Flavonoids in Foods and Biological Samples. Mini reviews in Medicinal Chemistry, 11 (14): 1239-1255. https://doi.org/10.2174/13895575111091239

[50] Shukla S., Shankar E., Fu P., Lennan M G T., Gupta S., (2015). Suppression of NF-κB and NF-κB-Regulated Gene Expression by Apigenin through IκBα and IKK Pathway in TRAMP Mice. PLoS ONE. 10(9): e0138710. https://doi.org/10.1371/journal.pone.0138710

[51] Tong X., Pelling J C., (2013). Targeting the PI3K/Akt/mTOR Axis by Apigenin for Cancer Prevention. Anticancer Agents Med Chem. 13(7):971-978. https://doi.org/10.2174/18715206113139990119

[52] ShuklaS., et Gupta S., (2011). The Impact of Aspect Ratio on the Biodistribution and Tumor Homing of Rigid Soft-Matter Nanorods. BiochemBiophysRes Commun, 287(4): 914-920. https://doi.org/10.1134/S1990750811030164

[53] Si Dayonget, Wang Y., Zhou Y H., Guo Y., Wang J., Zhou H., Li A S., (2009). Mechanism of CYP2C9 Inhibition by Flavones and Flavonols. (3):629-634. https://doi.org/doi:10.1124/dmd.108.023416

[54] Lin Y.,ShiR.,WangX.,et Ming Shen H., (2008). Luteolin, a Flavonoid with Potential for Cancer Prevention and Therapy. Curr Cancer Drug Targets,8(7): 634–646. https://doi.org/10.2174/156800908786241050

[55] Saber A., Kheira Sebbar N., El Mokhtar E., (2019). SYNTHESE, REACTIVITE ET ACTIVITES BIOLOGIQUES DES DERIVES DU BENZIMIDAZOLE. J. Mar. Chim. Heterocycl., 18 (3) : 1-50

[56] Salahuddin N., Shaharyar M., Mazumder A., (2017). Arabian Journal of Chemistry. 10 (1): 157-173. Benzimidazoles: A biologically active compounds. https://doi.org/10.1016/j.arabjc.2012.07.017

[57] Efouako S K., (1993). LES DERIVES DU BENZIMIDAZOLE, UN INTERET CROISSANT EN CHIMIE THERAPEUTIQUE : PERSPECTIVES D'AVENIR. Thèse de doctorat : Pharmacie. UNIVERSITE MOHAMMED V- RABAT. P3.

I. Introduction

Ce chapitre porte sur les principales méthodes de modélisation moléculaire et bio-informatiques, que nous avons utilisé pour la recherche de nouveaux inhibiteurs de la protéine Caséine Kinase 2 (CK2), impliquée dans de nombreuses pathologies cancéreuses.

La **bio-informatique** est une science multidisciplinaire. Elle se situe au carrefour entre la biologie, l'informatique, la chimie et beaucoup d'autre disciplines. Elle est constituée par l'ensemble des concepts et des techniques nécessaires à l'interprétation informatique et la prédiction des informations issues des données de la biologie expérimentale, afin d'établir les liens entre la structure des macromolécules biologiques, leurs fonctions et leurs activités cellulaires dans l'organisme [1].

La **Modélisation Moléculaire**, est un outil complémentaire à l'expérimentation. Cette technique a pour but de représenter, de prévoir la structure et la réactivité des molécules ou d'un système moléculaire. Elle consiste à préciser, à l'aide des calculs adéquats, les propriétés et le positionnement des atomes constituant ce système dans l'espace, ainsi de calculer l'énergie de la structure engendrée. En effet, les méthodes de la modélisation moléculaire comprennent : les méthodes quantiques, la mécanique moléculaire et la dynamique moléculaire [2].

II. Méthodes de la modélisation moléculaire

II.1 Mécanique Quantique

La mécanique quantique **(MQ)**, est la prolongation de la théorie du **quanta**, issue des travaux de **Planck**, de leurs interprétations par **Einstein** et de leurs applications à la théorie atomique par **Bohr** et **Sommerfeld**. Elle explique la quantification de certaines grandeurs physiques (énergie, moment cinétique) et fait émerger le principe d'exclusion de **Pauli**.

La mécanique quantique a pour objectif la détermination de l'énergie, ainsi que la distribution électronique d'un système moléculaire [2-4]. Elle consiste à résoudre l'équation de Schrödinger, qui impliquent un opérateur linéaire appelé hamiltonien, qui décrit l'évolution des états du système physique dans le temps [4-6], comme suit :

$$\hat{\mathbf{H}}_{\mathbf{tot}}\boldsymbol{\Psi}_{\mathbf{tot}}(\mathbf{r},\mathbf{R}) = \mathbf{E}_{\mathbf{tot}}\boldsymbol{\Psi}_{\mathbf{tot}}(\mathbf{r},\mathbf{R}) \quad \text{..... (1)}$$

Où :

r : est le vecteur position pour les électrons "**e**"

R : est le vecteur position pour les noyaux "**N**"

Ψ: Fonction d'onde de la molécule.

E : Energie totale d'une molécule associée à la fonction d'onde **Ψ**

$\hat{\mathbf{H}}_{\mathbf{tot}}$**:** L'opérateur Hamiltonien total de la molécule, est définit ci-dessous comme la somme des cinq énergies : Energie cinétique des noyaux, Energie cinétique des électrons, Energie potentielle d'interaction (répulsion) entre les noyaux, Energie potentielle d'attraction noyaux-électrons et Energie potentielle de répulsion entre les électrons.

$$\hat{\mathbf{H}}_{\mathbf{tot}} = \hat{\mathbf{T}}_{\mathbf{N}} + \hat{\mathbf{T}}_{\mathbf{e}} + \hat{\mathbf{V}}_{\mathbf{NN}} + \hat{\mathbf{V}}_{\mathbf{Ne}} + \hat{\mathbf{V}}_{\mathbf{ee}} \quad \text{(2)}$$

Néanmoins, en raison de la complexité de l'opérateur hamiltonien, qui couple le mouvement des électrons, l'équation de Schrödinger ne se résout pas de manière exacte que pour des systèmes particuliers (1e- ou 2e-).

Pour obtenir l'énergie d'un système, deux grandes philosophies sont présentes pour décrire son état fondamental : Une approche dans laquelle la fonction d'onde est déterminée (Hartree-Fock : HT, post HT…etc.) et une autre basée sur la densité électronique calculée (théorie de la fonctionnelle de la densité : DFT et la DFT dépendante du temps : TD-DFT) ; Cette dernière approche est fortement utilisée.

Pour ne pas alourdir le manuscrit, nous nous contenterons de détailler uniquement les méthodes théoriques utilisées dans le cadre de ce travail de thèse. De plus, les méthodes quantiques ont été largement définies dans plusieurs manuscrits de thèses soutenus à l'UMMTO [7-8].

II.2 Mécanique moléculaire

La mécanique moléculaire **(MM)** ou mécanique newtonienne, est une méthode empirique qui permet d'obtenir des résultats de géométries et d'énergies moléculaires, en se basant sur la mécanique classique, en utilisant comme outil, le "**Champ de force**".

II.2.1 Principe de la mécanique moléculaire

La mécanique moléculaire, considère une molécule (dans le vide) comme un ensemble d'atomes sur lesquels s'exercent des forces élastiques et harmoniques. Cette méthode, ne tient pas compte des électrons car elle ne s'intéresse qu'aux noyaux (atomes), qui sont représentés par des masses ponctuelles chargées, et les liaisons par des ressorts de différentes forces. Chacune de ces forces est décrite par une fonction d'énergie potentielle et la combinaison de toutes ces fonctions est appelée **"champ de force"** [9-10].

II.2.2 Champ de force

Un champ de force, est la somme d'une équation régissant l'énergie stérique d'un système moléculaire et de ses paramètres associés aux écarts par rapport à la structure de référence.

Cependant, l'équation d'énergie de la molécule est exprimée sous la forme suivante :

$$E = E_r + E_\theta + E_\Phi + E_\omega + E_{vanderWalls} + E_{électrostatique} + E_{Liaisonshydrogène} + E_{r\theta} \text{..... (3)}$$

Chaque terme, représente la variation de l'énergie avec une variable (coordonnée interne) du système :

"r" : Longueur de la liaison ;

"θ" : Angles de valence ;

"Φ" : Angles dièdres ;

"ω" : Angles dièdres impropres ;

"$E_{vanderWalls}$" : Potentiel de **Van der Waals** qui représente l'énergie d'interaction de liaisons des distances entre les atomes non liés, le plus souvent de type « LENNARD JONES ».

"$E_{électrostatique}$" **:** Energie électrostatique (charge ou dipôle), le plus souvent de type coulombien.

Il existe plusieurs types de champ de force, qui sont plus ou moins adaptés aux différents problèmes étudiés [9]. Ci-dessous, les principaux champs de forces utilisés :

- ✓ **AMBER** : Assisted Model Building and Energy Refinement (Logiciel : UCSFChimera) Champ de forces spécialisé dans l'alignement et la minimisation énergétique des molécules biologiques (protéines, ADN, ARN, etc.)
- ✓ **CHARMM :** (Chemistry at HARvard Macromolecular Mechanics) : Champ de forces spécialisé dans le traitement des molécules biologiques (protéines, ADN, ARN).
- ✓ **MM2, MM3 et MM4 :** Allinger Molecular Mechanics : Champ de forces spécialisé dans le traitement (Alignement) des petites molécules organiques.
- ✓ **MMFF94 :(**Merck Molecular Force Field) Champ de forces spécialisé dans le traitement des petites molécules organiques.
- ✓ **UFF** :(united force field) Champ de forces adapté aux molécules organométalliques (Logiciel : Cerius).

II.2.3 Optimisation de l'énergie d'une molécule et/ou d'un système moléculaire

L'optimisation de la géométrie d'une molécule où d'un système moléculaire consiste à trouver la géométrie qui correspond à l'énergie la plus basse du système, c'est-à-dire la

prédiction des conformations stables d'une structure moléculaire correspondant aux minimas de son énergie intramoléculaire [11].

II.2.4 Méthode de minimisation énergétique

La minimisation des structures chimiques, donne parfois des hydrogènes et des doublets d'électrons autour des atomes, dans des positions structurales impossibles. Ceci, est souvent dû à un mouvement initial inapproprié de ces atomes, qui rendent la structure fortement déformée.

Pour un système moléculaire (ex : complexe "protéine-ligand"), les possibilités d'angles de torsion, de translations, d'orientations sont quasi-infinies. C'est pour cela qu'il faudrait atteindre la valeur du minimum d'énergie global, qui représente l'état le plus stable de ce complexe. Le plus pratique est de trouver un minimum d'énergie local, qui s'approche le plus du minimum global. Toutes les méthodes de minimisation, commencent l'opération d'optimisation de l'hyper-surface de la molécule, en descendant vers le minimum le plus proche, sans savoir si ce minimum est local ou absolu.

Les conformations de départ doivent être présentées sous forme de coordonnées internes, en s'inspirant des modèles moléculaires [11].

II.2.5 Algorithmes de minimisation

Pour une molécule donnée, contenant **N** atomes, sa fonction de minimisation comprend 3**N** variables (un minimum global et des minima locaux). Donc, à partir de la géométrie initiale, on recherche le jeu de coordonnées cartésiennes, qui réduit à son minimum la somme de toutes les contributions énergétiques. En principe, il suffit de prendre la dérivée première de l'énergie stérique par rapport à chacun des degrés de liberté de la molécule et de trouver l'endroit sur l'hyper-surface énergétique où pour chaque coordonnée « **ri** », on aura :

$$(\mathbf{dE/dri}) = \mathbf{0} \quad \text{..... (4)}$$

Les procédures pour atteindre ce but sont de deux types : Les premières utilisent uniquement la pente de la surface (**dérivée première**), les deuxièmes utilisent à la fois cette pente et la courbure de la surface (**les dérivées première et seconde**) [12].

II.2.5.1 Méthode de "Steepest-Descent"

Le premier programme de minimisation, pouvant effectuer une optimisation de géométrie est due à Wiberg. En utilisant la méthode de la plus grande pente (**steepest-descent**).

En effet, après avoir calculé l'énergie correspondante à une géométrie initiale, chaque atome est déplacé individuellement selon ses trois coordonnées cartésiennes (X, Y et Z). Par la suite, l'énergie de cette géométrie est recalculée systématiquement après chaque déplacement, en calculant la dérivée première uniquement [12]. Ensuite tous les atomes seront déplacés sur une distance qui dépend de (**dE/dri**), et ainsi de suite.

Cet algorithme, suivra donc la direction imposée par les forces interatomiques les plus dominantes. C'est pourquoi, il se révèle très efficace pour supprimer les mauvais contacts ou les principaux problèmes stéréochimiques, qui existent dans les coordonnées brutes d'une structure cristalline ou modélisée, tout en perturbant très peu cette dernière.

En effet, la direction suivie par cette méthode, sera celle indiquée par l'opposé du gradient d'énergie (la direction de la plus grande pente de la fonction d'énergie, qui est la direction dans laquelle l'énergie diminue le plus vite) [12].

II.2.5.2 Méthode du gradient conjugué

Cette méthode, est aussi fondée sur le principe de la direction opposée au gradient d'énergie. Mais, elle prend également en compte les étapes précédentes, afin de déterminer plus intelligemment la direction et le pas. En effet, le "**Pas**" est ajusté à chaque cycle pour l'obtention d'une meilleure diminution d'énergie, en conservant son efficacité au cours du calcul [13].

En revanche, il reste l'inconvénient, que cette méthode soit plus coûteuse en temps de calcul par rapport à la méthode "**steepest descent**" [13].

II.2.5.3 Méthode Newton-Raphson

L'optimisation complète selon la méthode **Newton-Raphson**, nécessite le calcul de la matrice complète des dérivées secondes (matrice des constantes de force). Cette matrice contient **3N*3N** éléments pour une molécule de **N** atomes. Les dimensions **3N** correspondent aux trois degrés de mouvement pour chaque atome. Cependant, l'optimisation de la géométrie nécessite seulement 3**N-6** degrés de liberté, puisque les trois translations et les trois rotations, ne sont pas accompagnées par des changements géométriques. Les constantes de forces, sont alors manipulées sous la forme d'une matrice de dérivées secondes [11].

II.2.5.4 Méthode de recuit simulé

Toutes les méthodes décrites ci-dessus, ont la particularité de faire décroître à chaque « **pas** » la fonction d'énergie. Ces méthodes, ne peuvent donc pas échapper au minimum local proche de la structure de départ.

La méthode de recuit simulé « **anneal** », développée par Kirk patrick, autorise la fonction d'énergie à augmenter momentanément, afin de franchir des barrières énergétiques pour retomber dans un minimum plus profond, ce qui permet d'aller au-delà des minima locaux, au voisinage de la structure initiale pour explorer de façon plus extensive l'espace conformationnel accessible, et afin de découvrir des minima plus profonds et plus éloignés de la structure initiale que les minima locaux [5].

En effet, un "**minimiseur**", est une boîte noire mathématique qui demeure un outil important pour la recherche d'un minimum d'énergie. En général, plusieurs minimiseurs sont utilisés, tout dépend du nombre de variables ou du nombre de variations d'angle de liaisons introduites [5].

II.3 Dynamique moléculaire

La dynamique moléculaire (DM), est une technique de modélisation moléculaire, réalisée par un ordinateur, durant laquelle le système subit des changements conformationnels et cinétiques, décrivant l'évolution au cours du temps des positions (trajectoires) atomiques [2,9].

Cette méthode est basée sur la simulation d'une trajectoire réelle dans l'espace des interactions intramoléculaires, et forces provoquées par l'ensemble des particules du système moléculaire qui correspondent à des vibrations autour d'un minimum, lors du passage d'un minimum d'énergie à un autre. En effet, lorsqu'un minimum local est atteint, ces forces sont nulles [14-16].

II.3.1 Dynamique moléculaire Newtonienne (dynamique classique)

Dans cette approche, chaque atome, considéré comme une masse ponctuelle, ne peuvent pas se déplacer librement au cours d'un intervalle de temps suffisamment court, du fait des diverses forces auxquelles ils sont soumis [17]. L'équation fondamentale de newton du mouvement des atomes est donc exprimée comme suit :

$$\sum_j \mathbf{F_{ij}} = \mathbf{m_i a_j} \quad \text{..... (5)}$$

Où :

$\mathbf{F_{ij}}$: Force exercée par les atomes **j** sur l'atome **i**, de masse $\mathbf{m_i}$ et d'accélération $\mathbf{a_i}$

II.3.2 Dynamique de Langevin

La dynamique de Langevin, est une approche de la modélisation mathématique de la dynamique moléculaire, caractérisée par l'utilisation de modèles simplifiés basés sur des

équations différentielles stochastiques du mouvement brownien [17]. En effet, la dynamique de Langevin est exprimée par l'équation suivante :

$$\mathbf{m}\frac{\mathbf{d}\vec{\mathbf{v}}(\mathbf{t})}{\mathbf{dt}} = -\mathbf{k}\,\vec{\mathbf{v}}(\mathbf{t}) + \vec{\boldsymbol{\eta}}(\mathbf{t}) \quad \text{..... (6)}$$

Où :

m : masse de la particule , $\vec{\mathbf{v}}(\mathbf{t})$: vitesse de la particule à l'instant « t » .

$\vec{f} = -\boldsymbol{k}\,\vec{\boldsymbol{v}}(\boldsymbol{t})$: force de frottement, $\vec{\boldsymbol{\eta}}(\mathbf{t})$: force complémentaire

k : constante positive, pour une particule sphérique de rayon « a » , $\mathbf{k} = \mathbf{6}\,\boldsymbol{\pi\eta a}$ (loi de Stokes).

L'équation de Langevin, permet d'associer le mouvement des atomes à un bain thermique, par l'inclusion de termes stochastiques dans l'équation de Newton de la dynamique classique, comme suit :

$$\mathbf{m_i}\dot{\mathbf{v}}_\mathbf{i} = \mathbf{m_i a_i} = -\nabla_\mathbf{r}\mathbf{U_p}(\mathbf{r_i}, \mathbf{t}) - \mathbf{m_i}\boldsymbol{\gamma_i}\mathbf{v_i}(\mathbf{t}) + \mathbf{R_i}(\mathbf{t}) \quad \text{..... (7)}$$

$\boldsymbol{\gamma_i}$: Terme de friction qui s'oppose au mouvement de l'atome i

$\dot{\mathbf{v}}_\mathbf{i} = \mathbf{a_i} = {}^{\mathbf{dv_i}}/_{\mathbf{dt}} \cdot \boldsymbol{\gamma_i}$: Fréquence de collisions de l'atome i

$\mathbf{R_i\,(t)}$: Force aléatoire due aux forces de collisions associées aux mouvement browniens des particules.

II.3.3 Etapes de la dynamique moléculaire

II.3.3.1 Préparation du système moléculaire et optimisation énergétique

Pour lancer un calcul de dynamique moléculaire, il est nécessaire de connaitre les coordonnées cartésiennes (X, Y, Z) et les vitesses initiales à attribuer selon la statistique de Boltzmann [15]. La structure de la molécule peut être obtenue expérimentalement ou générée, en utilisant le champ de force adéquat. Des algorithmes de minimisation énergétique, sont utilisés afin de relaxer la structure et d'éliminer tous les mauvais contacts et gênes stériques entre les atomes qui la constituent.

II.3.3.2 Thermalisation

La structure moléculaire déjà minimisée sera chauffée, jusqu'à la température souhaitée. Un jeu de vitesses initiales correspondant à quelques dizaines de Kelvin (k), sera tiré aléatoirement et les équations du mouvement seront intégrées, en augmentant la température jusqu'à atteindre la température souhaitée [16].

II.3.3.3 Equilibration

Dans cette phase, la température du système est stabilisée, par un équilibre parfait entre l'énergie potentielle et l'énergie cinétique. En effet, la température de déviation est calculée, en suivant un processus de relaxation, où la température est recalculée après chaque étape, afin de la ramenée à la température de référence [15], par la formule suivante :

$$\frac{d}{dt}T = \tau^{-1}(T_0 - T) \quad \text{.....} \quad (8)$$

Où :

"τ" : temps de relaxation

"T_0": Température de référence.

"**T**": Température du système.

Durant la phase d'équilibration, la valeur de "τ", doit être assez petite pour permettre un couplage efficace (généralement "τ = 10 fs") [15].

II.3.3.4 Production

Dans cette étape qui suit la phase d'équilibration, les conformations et les énergies du système, sont sauvegardées à des intervalles réguliers et prédéfinies, qui seront utilisées par la suite, pour calculer les caractéristiques relatives au système moléculaire [15].

II.3.3.5 Interprétation des résultats / Analyses usuelles en dynamique moléculaire

Différents outils, sont utilisés pour analyser les résultats obtenus lors de la dynamique moléculaire, afin d'évaluer la pertinence des mouvements et interactions intramoléculaire, ainsi que de statuer sur la stabilité du système moléculaire [15], en utilisant les paramètres suivants :

II.3.3.5.1 L'écart quadratique moyen (*Root Mean Square Deviation* : RMSD)

Le RMSD, est utilisé pour quantifier la modification moyenne, que va subir la protéine au cours de la simulation de dynamique moléculaire. Il est calculé par l'équation suivante :

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\delta_i^2} \quad \text{.....} \quad (9)$$

Où :

"δi" : La distance entre N paires d'atomes équivalents à comparer.

Le RMSD, est utilisé pour comparer la structure des protéines. En effet, plus il est proche de 0, plus la structure d'intérêt est semblable à celle de référence. Pour obtenir une vue d'ensemble du mouvement protéique, la mesure n'est réalisée que sur la chaine peptidique principale des protéines, ou simplement sur les carbones alphas (Cα).

La valeur du RMSD augmente progressivement au cours de la production, jusqu'à la stabilisation du système [15].

II.3.3.5.2 Fluctuations atomiques (*Root Mean Square Fluctuation*: RMSF)

Le RMSF, concerne l'évolution de l'écart des distances atomiques, au cours du temps par rapport à la structure de référence. Il représente, la moyenne de la déviation pour chaque résidu (le Cα de chaque résidu) comparativement aux mêmes atomes de la structure de référence.

Le RMSF, est calculé par la formule suivante :

$$\mathbf{RMSF_i} = \sqrt{\frac{1}{T}\sum_{t=0}^{t=T}(x_i^t - \bar{x}_i)} \quad \text{..... (10)}$$

Où :

$RMSF_i$: La fluctuation atomique de l'atome i, calculée pour une durée de trajectoire de « T » pas, avec la position de l'atome au temps « t » et « xi » : la position moyenne sur le même intervalle.

II.3.3.5.3 Facteur d'agitation thermique (B Factor)

Le facteur d'agitation thermique (*B Factor*, en anglais), est calculé à partir de la valeur de (RMSF), en utilisant l'équation suivante :

$$\mathbf{B_{factor}} = \frac{8\pi^2}{3}\mathbf{RMSF^2} \quad \text{..... (11)}$$

IL est à noter, qu'une valeur nulle pour la fluctuation atomique, signifie que la position moyenne du résidu, ne diffère pas de celle de la structure de référence. Cette grandeur permet aussi bien de localiser les régions charnières de la protéine, ainsi que les résidus responsables des mouvements de grandes amplitudes [15-16].

II.3.3.5.4 Liaisons hydrogènes

Les liaisons hydrogènes, sont les liaisons les plus importantes en biologie structurelle, puisqu'elles jouent un rôle primordial et fondamental dans le phénomène de solvatation et dans la catalyse enzymatique. Nous avons détaillé, en annexe 1 de cette présente thèse, tous les types de liaisons hydrogènes.

II.3.3.5.5 La surface accessible au solvant (*Solvent Accessible Surface Area* : SASA)

C'est la surface d'une biomolécule accessible au solvant, mesurée en unité d'Angströms au carré ($Å^2$), parfois appelée surface moléculaire Lee-Richards [17]. Cette

surface, est généralement calculée à l'aide de l'algorithme « *rollingball* » développé par Shrake et Rupley en 1970 [18].

II.3.3.5.6 Le rayon de giration (*Radius of gyration* : Rg)

Le rayon de giration « Rg », est utilisé pour statuer sur la stabilité du système, au cours de simulation de dynamique moléculaire. En effet, une élévation substantielle de plus de 5Å par exemple, signifiée un dépliement de la protéine.

Ce paramètre, est utilisé aussi pour suivre la dénaturation lors du processus de chauffage à haute température [19] et il est calculé par la formule suivante :

$$Rg = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\overrightarrow{r_i} - \overrightarrow{r_{cm}}\|^2} \quad \text{..... (12)}$$

II.3.4 Technique numériques et outils de la dynamique moléculaire

II.3.4.1 Méthode du maillage particulier d'Ewald (*particule mesh Ewald* : PME)

La sommation d'Ewald, est une méthode de calcul des énergies d'interaction de systèmes biomoléculaires, en particulier les énergies électrostatiques. Cette méthode est applicable à de nombreux domaines, y compris la dynamique moléculaire, puisqu'elle permet le traitement rapide et approprié des forces de coulomb à long et à courte portée [19].

Les énergies électrostatiques, comprennent à la fois des interactions de courtes et de longues distances. De ce fait, il est plus judicieux de décomposer le potentiel d'interaction en termes de courte portée (dont la sommation se fait dans l'espace réel) et de longue portée (dont la sommation se fait dans l'espace de Fourier). Par conséquent, la sommation (maillage) d'Ewald, réécrit le potentiel d'interaction comme la somme des deux termes [19], comme suit :

$$\emptyset(\boldsymbol{r}) \stackrel{\text{def}}{=} \emptyset_{\boldsymbol{sr}}(\boldsymbol{r}) + \emptyset_{\boldsymbol{lr}}(\boldsymbol{r}) \quad \text{..... (13)}$$

Où :

$\emptyset_{\boldsymbol{sr}}$ **:** Le terme de courte portée dont la sommation est rapide dans l'espace réel

$\emptyset_{\boldsymbol{lr}}$: Le terme de longue portée dont la sommation est rapide dans l'espace de Fourier

II.3.4.2 Conditions périodiques aux limites (*Periodic boundary conditions* : PBC)

Dans une simulation de dynamique moléculaire, les conditions périodiques aux limites (*periodic boundary conditions*, en anglais), sont utilisées particulièrement en conjonction avec

les méthodes de sommation d'Ewald. Ces conditions, consistent à répliquer un ensemble fini et relativement restreint de particules dans une boite de simulation centrale dans les trois directions, de telle sorte que les forces agissant sur les particules, soient équivalentes à celles d'un système réel [20].

En effet, lors de l'application des conditions périodiques aux limites, la charge électrostatique du système doit être nulle afin d'éviter une somme infinie, cela est assuré par l'addition des contre ions (Na^+, Cl^-, etc.) en quantités appropriés (si les molécules étudiées sont chargées) [20].

II.3.4.3 Barostat de Langevin

Le barostat de Langevin (*berendsen*), est une méthode de contrôle de la pression dans une simulation de dynamique moléculaire. En effet, cette méthode ajoute un terme supplémentaire aux équations de mouvement, qui génère le changement de pression [21], comme exprimé par l'équation suivante :

$$\frac{dP}{dt} \Big| = \frac{P_0 - P}{\tau_P} \quad \text{..... (14)}$$

A noter que:

P_0 : Pression de référence, **P** : pression instantanée et **τ_p** : constante du temps.

II.3.4.4 Thermostat de Langevin

Lors d'une simulation de dynamique moléculaire, le thermostat de Langevin est utilisé pour contrôler la température du système moléculaire. En effet, chaque atome est couplé à un bain thermique **(γ_i)** de température **T_0** (thermalisation des atomes) [21-22]. Par conséquent, l'équation du mouvement résultante est exprimée comme suit :

$$m_i \frac{dv_i}{dt} = F_i + m_{i\gamma}\left(\frac{T_0}{T} - 1\right) v_i \quad \text{..... (15)}$$

Avec : **F_i** : C'est la force exercée sur l'atome i

II.3.4.5 Le rayon de troncature (*cut-off radius*)

Dans une simulation de dynamique moléculaire, les interactions moléculaires doivent être limitées à une distance, qui est égale à la moitié de la taille de la boite de simulation, appelée : « rayon de troncature » ou « rayon de coupure » (*cut-off*, en anglais). En effet, toutes

les interactions avec les molécules se situant à une distance supérieure au rayon de troncature sont négligées [21-22].

II.3.4.6 Le rayon de switch (*switch radius/smooth cut-off*)

Cette méthode est utilisée pour réduire le potentiel d'interaction sur une portée prédéfinie, dans laquelle, il sera ensuite commuté à zéro entre la première et la dernière coupure [21-22].

III. Techniques de conception de médicament "*Drug design*"

La conception de médicaments, souvent appelée "***Drug Design***", est l'ensemble de techniques informatiques utilisées pour la découverte d'un nouveau médicament [23]. Nous présentons ci-dessous, les principales techniques utilisées.

III.1 Docking Moléculaire

Le Docking moléculaire (*Molecular Docking*, en anglais), ou encore (Ancrage ou Amarrage moléculaire), est une technique qui consiste à faire interagir une petite molécule organique, appelée "**ligand**", dans le site actif d'une macromolécule (Protéine), appelée "**récepteur**".

L'objectif de cette technique est de trouver le meilleur mode de fixation du complexe "**Ligand-Récepteur**", ainsi de prédire la conformation dite "bioactive" du ligand au sein de son récepteur. En effet, chaque conformation relative issue du docking est associée à une énergie, appelée "**Score**". La conformation bioactive correspond à l'énergie la plus minime [14,24-25].

Il existe de nombreux programmes de docking moléculaire, tel que Autodock [26] et "FlexX" [27]. Cependant, l'évaluation de ces outils, se fait par le choix du meilleur compromis entre leurs rapidités, coûts et précisions. Cela, en comparant les prédictions des résultats expérimentaux, d'un ensemble représentatif de complexes "protéines-ligand" issus des bases de données de protéines (*Protein Data Bank* : PDB, en anglais).

La simulation de doking moléculaire, repose principalement sur l'association ligand - protéine qui peuvent être considérés rigides, flexibles ou bien semi-flexibles.

III.1.1 Docking rigide

Le ligand et le récepteur (protéine), sont considérés comme deux entités rigides, la recherche de la conformation optimale (stable) se limite au positionnement, en énumérant toutes les rotations et les translations possibles pour le ligand à l'intérieur du site d'interaction, et en attribuant les valeurs d'énergie appropriées [28].

Toutes les conformations incompatibles avec le site actif sont éliminées, afin d'optimiser les meilleures conformations compatibles sélectionnées [28]. Cette approche est peu coûteuse en termes de temps de calcul, et elle constitue un filtre permettant de ne pas retenir les molécules aberrantes de grande taille et de mauvaise complémentarité avec le site de récepteur [28-29].

III.1.2 Docking semi-flexible

Le système moléculaire est séparé en deux parties, une partie flexible contenant le ligand et les résidus flexibles du site actif et une partie rigide contenant le reste de la protéine sans le ligand. Les programmes de Docking semi-flexible, sont les plus efficaces. Ils emploient des méthodes de recherche aléatoires, connues sous le nom "**Stochastiques**", tel que les méthodes de Monte Carlo et les méthodes évolutionnaires basées sur des algorithmes génétiques [28].

III.1.3 Docking flexible

Cette méthode, considère la totale flexibilité de la protéine de manière indirecte, en diminuant certains termes de répulsion. Ce qui permet, au ligand d'infiltrer légèrement dans la surface des protéines, qui se déforment et changent de conformation au cours des réactions enzymatiques pour s'adapter au ligand, en formant des interactions spécifiques et sans gènes stériques, afin d'améliorer la complémentarité de surface autorisant la formation de liaisons hydrogènes [29].

L'évaluation de l'affinité du complexe protéine-ligand se fait, en se basant sur les effets d'entropie et en calculant la fonction d'énergie "***Scoring Fonction***". Il convient de noter que ces modélisations sont plus rapides à calculer, bien qu'elles soient moins précises [30].

III.1.4 Insuffisance (limites) des méthodes de Docking moléculaire

Les méthodes de Docking moléculaire, possèdent les insuffisances suivantes :

- ✓ Limitation à l'expérimentation animale, puisque ces méthodes ne peuvent pas prétendre garantir l'absence d'effets indésirables sur un organisme.
- ✓ L'étude de grand nombre d'atomes, impose un temps de calcul très long pour simuler l'évolution du système sur une durée permettant d'observer des mouvements importants de la molécule.
- ✓ Le fait de considérer la molécule en plusieurs parties rigides, ne permet pas une description correcte, car la flexibilité de la molécule entière est à prendre en compte pour la décrire en réalité [31-32].

III.2 Criblage (Screening)

Le processus de conception et de synthèse de nouvelles molécules thérapeutiques (médicaments), est long et très coûteux. Il se fait souvent, en testant un ensemble très important de molécules candidates sur une cible biologique bien identifiée, selon une méthode appelée : "**Criblage**" (*Screening*, en anglais) [33]. Cette technique d'investigation computationnelle, permet de réduire le coût de la synthèse et des tests (*in silico*, *in vitro* et *in vivo*), ainsi que d'effectuer un tri parmi les composés testés.

Le criblage distingue les composés possédant les propriétés pharmacologiques recherchées, qui sont susceptibles d'interagir avec la cible thérapeutique visée, et il permet d'éliminer les composés indésirables (ex : molécules inactives, de grande taille, ...etc.) [33].

III.2.1 Criblage réel

Une technique expérimentale, basée sur la chimie combinatoire, qui permet de générer de grandes bibliothèques de composés chimiques, appelées "**Chimiothèque**", ainsi que de tester leurs puissances thérapeutiques sur des cellules biologiques (*in vitro*) [34-35]. En effet, ce procédé est utilisé fréquemment en chimie médicinale, pour l'identification des molécules actives, par leurs mises en contact avec la cible biologique (ex : protéines dont l'implication dans un processus pathologique a été identifié expérimentalement). Cependant, cette technique est limitée par le nombre de composés à tester en un temps raisonnable et par le coût très cher des tests appropriés [36].

III.2.2 Criblage virtuel (*Virtual Screening*)

Une méthode théorique réalisée in-silico, qui permet d'explorer des chimiothèques de petites molécules appelées "**ligands**", afin d'identifier de nouveaux composés bioactifs ainsi que leurs liaisons et interactions vis-à-vis d'une cible thérapeutique particulière [37]. Le criblage virtuel, est fréquemment utilisé pour accélérer les découvertes scientifiques, puisqu'il permet de réaliser de manière rapide et à moindre coût des prédictions de l'activité de nouvelles molécules, ce qui constitue un complément primordial au criblage biologique à haut débit [37]. En effet, les techniques de criblage virtuel sont basées sur des modèles mathématiques et utilisent un matériel informatique sophistiqué. Cependant, cette méthode est parfois limitée par l'expérimentation animale, car elle ne peut pas prétendre garantir l'absence d'effets indésirables sur un organisme humain [37].

III.2.3 Approches du criblage virtuel

Selon l'information expérimentale disponible, on distingue deux approches distinctes pour le criblage virtuel :

III.2.3.1 Approche "*Structure-Based Virtual Screening*"

Cette approche est fondée, essentiellement sur la connaissance, ainsi que la disponibilité de la structure cristallographique et tridimensionnelle de la cible biologique. Cette dernière peut être obtenue expérimentalement, aux moyens des rayons X et/ou RM.

Cette méthode consiste, à estimer la complémentarité structurale, de chaque molécule cible avec son site actif approprié [38]. En effet, le criblage virtuel « *structure-based* », est l'équivalent *in silico* d'un test expérimental, étudiant la liaison ligand-cible biologique.

En revanche, cette approche est très coûteuse, en puissance de calcul et en termes d'expertise [39].

III.2.3.2 Approche "*Ligand-Based Virtual Screening*"

Cette approche est basée sur la connaissance, d'un nombre suffisant d'informations, concernant une ou plusieurs molécules actives de référence. Elle a pour objectif, la recherche d'un ensemble de molécules "ligands", ayant tendance à présenter des profils d'activités très proches, ainsi qu'une forme similaire à un ligand actif de référence, connu pour une cible thérapeutique.

Des méthodes d'analyse de similarité, en bi et tridimensionnelle sont utilisées, afin de construire un modèle de prédiction d'affinité appelé "**Modèle pharmacophore**" [40].

En effet, chaque ligand candidat, sera ensuite comparé au modèle pharmacophore, pour la détermination de son aptitude et compatibilité de se lier à la cible spécifique.

Cette approche rapide et simple à mettre en œuvre, dispose d'un inconvénient majeur d'interdépendance, envers les informations de référence, utilisées pour construire le modèle pharmacophore de référence [40-41].

III.2.4 Outils de criblage virtuel

III.2.4.1 Recherche de similarité moléculaire (paradigme de similitude)

La méthode de recherche de similarité, également appelée "**métriques de similarité**", est applicable dans le cas d'obtention d'un petit nombre de ligands actifs, pour une cible biologique spécifique. Elle repose sur l'utilisation de descripteurs de similarité, en comparant les molécules à cribler à un ligand actif connu de référence, pour prédire leur profil d'activité vis à vis de la cible biologique [40].

III.2.4.2 Descripteurs de similarité

Les descripteurs de similarité sont obtenus, en utilisant des procédures mathématiques et expérimentales appropriées, classées selon leurs dimensions. Ils représentent des caractéristiques structurales importantes de composés, pour déterminer au sein d'une base de données, quels sont les composés les plus ressemblants aux ligands actifs de référence [40-42].

En effet, ils permettent de transformer l'information chimique, d'une molécule encodée par une représentation symbolique, en un nombre ou vecteur utile, dont le résultat est utilisé comme descripteur [40-42].

III.2.4.3 Modèles pharmacophores (*ligand-based* / *structure-based*)

Un modèle pharmacophore de type (*ligand-based* / *structure-based*), est déterminé exclusivement à partir de la structure de composés actifs de référence seul ou en interaction avec sa cible biologique. Selon l'IUPAC (***I****nternational* ***U****nion of* ***P****ure and* ***A****pplied* ***C****hemistry*), le pharmacophore constitue l'ensemble des propriétés stériques et électroniques d'une molécule, essentielles pour assurer les interactions supramoléculaires optimales avec une cible spécifique, afin d'engendrer ou inhiber une réponse biologique [43].

Les molécules ayant le même modèle pharmacophore, pour une cible donnée, devraient se lier de manière identique, à ce récepteur et présenter des profils d'activité similaires. Par conséquent, le modèle pharmacophore résultant, est utilisé comme référence dans le criblage virtuel.

En effet, les différentes caractéristiques pharmacophoriques recherchées, dans un modèle phramcophore sont : les liaisons hydrogènes accepteur HBA, liaisons hydrogènes donneur HBD, les zones ionisables chargées positivement (PI), ou négativement (NI), qui forment des interactions électrostatiques avec ceux de charge opposée, les groupements hydrophobes (Hy) et cycles aromatiques (Ar) [44].

III.2.5 Evaluation des méthodes de criblage virtuel

Cette étape du criblage virtuel, consiste à évaluer la capacité d'une méthode de docking, à prédire de manière prospective de nouveaux ligands, ainsi que l'affinité de leurs liaisons vis à vis du récepteur.

En générale, elle se fait d'une manière rétrospective, en se basant sur deux critères : la capacité de la méthode, à retrouver le meilleur mode d'emplacement du ligand au sein du site actif (précision de positionnement), ainsi que sa capacité à distinguer les ligands actifs, de ceux inactifs (Enrichissement de la chimiothèque) [45].

III.2.5.1Précision du positionnement

Consiste à trouver la meilleure position (conformation) adéquate, du complexe ligand-récepteur. En effet, la meilleure conformation correspond, à la meilleure fonction énergétique empirique (*Scoring Function*, en anglais), obtenue avec les interactions moléculaires [46], en utilisant les métriques suivantes :

III.2.5.1.1 Ecart quadratique moyen (*Root Mean Square Deviation* : RMSD)

L'écart RMSD, permet de mesurer la distance entre les paires d'atomes, correspondants aux différentes conformations, observées entre la structure expérimentale du ligand et celle de la structure prédite du complexe ligand-récepteur.

En effet, les conformations comparées, sont considérées identiques lorsqu'elles sont séparées par une valeur d'RMSD inférieur ou égal à **0.3Å**. Alors qu'elles ne sont pas, lorsqu'elles sont séparées par un RMSD d'une valeur supérieure à **2Å** [47].

Le calcul du RMSD est donné par l'équation suivante :

$$\mathbf{RMSD}_{(v,w)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)} \quad \text{..... (16)}$$

A noter que :

n : est le nombre total d'atomes, v_i et w_i : sont les atomes identiques de la structure expérimentale et prédite, respectivement des cordonnées cartésiennes : **x**, **y** et **z**

Le calcul de RMSD, est utilisé pour les raisons de son objectivité, sensibilité et la simplicité d'automatisation. Néanmoins, cette métrique n'est pas considérée comme un critère d'évaluation parfait, vu qu'elle présente certaines limites, puisqu'elle ne reflète pas vraiment les interactions existantes entre le ligand et le récepteur, ce qui s'explique comme suit :

- ✓ Le RMSD, permet d'évaluer les distances entre les positions atomiques, mais il ne fournit aucune information sur la conservation des interactions, entre la structure prédite et celle cristallisée.
- ✓ La dépendance vis-à-vis du poids moléculaire (les petits composés sont généralement associés à de faibles valeurs de RMSD).
- ✓ Une grande sensibilité aux variations des conformations de la structure prédite, bien que le mode de liaison soit identique (correct).
- ✓ Une augmentation de la valeur de RMSD, alors que les molécules sont presque symétriques et les interactions clés sont conservées.

Pour toutes ces raisons, l'évaluation des méthodes de criblage fait appel à d'autres éléments de mesure (métriques), telles que : l'erreur de déplacement relative (**RDE**), l'espace réel du facteur R (**RSR**), ainsi que l'évaluation visuelle du positionnement par la classification basée sur la précision des interactions (**IBAC**) [47].

III.2.5.1.2 Erreur relative de déplacement (*Relative Displacement Error* : RDE)

L'utilisation de RDE, permet de réduire l'impact de la grande divergence, par rapport à la valeur moyenne à analyser, ainsi que les erreurs d'interprétation [47-48]. Cette erreur est donnée par l'équation suivante :

$$\mathbf{RDE} = \mathbf{100}\left(\mathbf{1} - \frac{\mathbf{L}}{\mathbf{N}}\left(\sum_{\mathbf{i=1,N}} \frac{\mathbf{1}}{\mathbf{L}+\mathbf{D_{ii'}}}\right)\right) \quad \textbf{(17)}$$

A noter que :

N : est le nombre d'atomes **i** du ligand prédit (**i'** pour le ligand cristallisé), **L** : le paramètre d'échelle de précision, dont la valeur est couramment située entre **1,5** et **3 Å**, et « $\mathbf{D_{ii'}}$ » : la déviation de l'atome **i** prédit par rapport à l'atome **i'** de référence (cristallisé) [48].

- Lorsque le positionnement prédit est correct, les valeurs de déviations sont nulles ou très proches de zéro. Par conséquent, la valeur de **RDE** correspondante est également presque de **0%**.
- Cependant, la valeur résultante de RDE est de **50%**, lorsque :
  - ✓ Le positionnement des atomes, est tel que la valeur de la déviation "**D**" et que celle-là est égale à la valeur du paramètre d'échelle "**L**". Ou bien ;
  - ✓ Le positionnement de la moitié des atomes est correct, c'est à dire : (**D << L**) et celui de l'autre moitié est incorrect (**D >>L**), ce qui ne permet pas d'y remédier aux limitations du calcul d'**RMSD** [49].

III.2.5.1.3 Espace réel du facteur RSR (*Real Space R-factor*)

La valeur de RSR, est calculé à l'aide de la densité électronique observée expérimentalement ($\boldsymbol{\rho_{obs}}$) et celle calculée($\boldsymbol{\rho_{calc}}$) à partir des coordonnées atomiques du modèle pharmacophore (le ligand cristallisé et la conformation prédite) [48], selon l'équation suivante :

$$\mathbf{RSR} = \frac{\sum|\boldsymbol{\rho_{obs}} - \boldsymbol{\rho_{calc}}|}{\sum|\boldsymbol{\rho_{obs}} + \boldsymbol{\rho_{calc}}|} \quad \textbf{(18)}$$

En effet, la valeur de RSR est faible, quand les densités électroniques observées expérimentalement et celles calculées à partir d'un modèle pharmacophore fiable (correct) sont similaires. Dans le cas contraire, si le modèle n'est pas correct, la valeur de RSR est élevée.

III.2.5.2 Classification de la précision basée sur les interactions (IBAC)

La classification IBAC (***I**nteractions-**B**ased **A**ccuracy **C**lassification*, en anglais), est utilisée afin d'affranchir les autres limitations du RMSD. Cette procédure, repose principalement sur une inspection visuelle, des structures cristallisées et prédites. Dans cette procédure, le ligand est divisé en deux parties :

La première est constituée d'un noyant, comprenant les caractéristiques essentielles à sa liaison avec la protéine et la deuxième partie, est considérée comme un recouvrement superficiel, moins important pour la liaison. Par conséquent, la conformation résultante est considérée comme identique [50-51].

III.2.5.2.1 Enrichissement de la chimiothèque

Le deuxième critère d'évaluation, d'une méthode de criblage *in silico*, est la capacité de distinguer les composants actifs des inactifs. Cela, dans le but de différencier les molécules réellement capables, de se lier au récepteur pour provoquer un effet biologique.

Ces composés sont appelés "**ligands bioactifs**", par rapport au reste des molécules inactives de la chimiothèque.

Par conséquent, la méthode de criblage est considérée comme étant efficace, lorsque le pourcentage des composants actifs, retrouvé grâce aux moyens de cette dernière est largement supérieur, au pourcentage retrouvé à l'aide de la sélection, au hasard dans la chimiothèque [49].

La figure 1, représente le flux de la sélection des "**n**" molécules, au sein d'une chimiothèque de "**N**" composés.

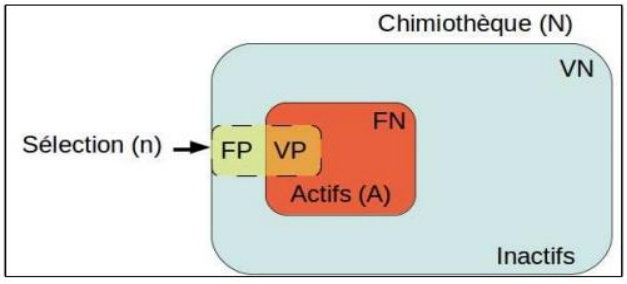


Figure 1 : Représentation de la sélection de "n" molécules dans une chimiothèque de "N" composés [49].

Avec :

- A : Nombre de composés actifs
- N : Nombre total de composés
- Les vrais positifs (VP) : Composés actifs appartenant aux "n" molécules sélectionnées.
- Les faux négatifs (FN) : Composés actifs non sélectionnés.
- Les faux positifs (FP) : Composés inactifs sélectionnés.
- Les vrais négatifs (VN) : Composés actifs non sélectionnés.

III.2.5.2.2 Précision

La précision (p), décrivant le pourcentage de composés actifs retrouvés, par rapport au nombre total de composés présents dans la chimiothèque, est donnée par l'équation suivante :

$$\textbf{Précision} = \frac{\mathbf{N_{actifs\ sélectionnés}}}{\mathbf{N_{actifs\ totaux}}} = \frac{\mathbf{VP}}{\mathbf{VP+FN}} \quad \textbf{(19)}$$

III.2.5.2.3 Sensibilité

La **sensibilité** (**Se**), décrivant le pourcentage de composés actifs retrouvés, par rapport au nombre total d'actifs présents dans la chimiothèque, est donnée par l'équation suivante :

$$\textbf{Sensibilité} = \frac{\mathbf{N_{actifs\ sélectionnés}}}{\mathbf{N_{composés\ actifs}}} = \frac{\mathbf{VP}}{\mathbf{A}} \quad \textbf{(20)}$$

III.2.5.2.4 Spécificité

La **spécificité** (**Sp**), décrivant le pourcentage de composés inactifs non sélectionnés, par rapport au nombre totale des inactifs dans la chimiothèque, est donnée par l'équation suivante :

$$\textbf{Spécificté} = \frac{\mathbf{N_{inactifs\ non\ sélectionnés}}}{\mathbf{N_{inactifs\ totaux}}} = \frac{\mathbf{VN}}{\mathbf{N-A}} = \frac{\mathbf{VN}}{\mathbf{VN+FP}} \quad \textbf{(21)}$$

III.2.5.2.5 Facteurs d'enrichissement (FE)

Le facteur d'enrichissement (Enfrichement factor : EF, en anglais), est une métrique de performance, calculée pour des parties choisies de la chimiothèque. Elle compare la capacité de la méthode, à retrouver les composants actifs dans une fraction donnée, de la chimiothèque par rapport à une sélection aléatoire [49], selon l'équation suivante :

$$\mathbf{EF_{100*(n/N)\%}} = \frac{\mathbf{VP/n}}{\mathbf{A/N}} \quad \textbf{(22)}$$

A noter que :

"n" : nombre de composés dans la fraction de la chimiothèque étudiée

"N" : nombre total de composés dans la chimiothèque ;

"A" : Nombre de composés actifs ;

"VP" : vrais positifs et **"FN"** : faux négatifs.

Cependant, la mesure du facteur d'enrichissement "EF", présente deux limites majeures :

- ✓ La dépendance de "EF" du pourcentage d'actifs de la chimiothèque, surtout lorsqu'il s'agit de comparer les performances obtenues, sur des jeux de données présentant des pourcentages d'actifs /inactifs différents.
- ✓ Le facteur d'enrichissement "EF", est commun à toutes les métriques classiques d'évaluation des performances. En effet, il chiffre le nombre d'actifs présents dans la fraction étudiée de la chimiothèque, sans tenir compte de leur distribution.

III.2.5.2.6 La courbe ROC (*Receiver Operating Characteristic curve*)

La courbe **ROC**, représente l'évolution de la sensibilité (**Se**) pour chaque changement possible de « un moins la spécificité (**1-Sp**) » [52]. En outre, la courbe ROC exprime le pourcentage de composés actifs, en fonction des inactifs (decoys) pour chaque partie de la chimiothèque (Figure 2).

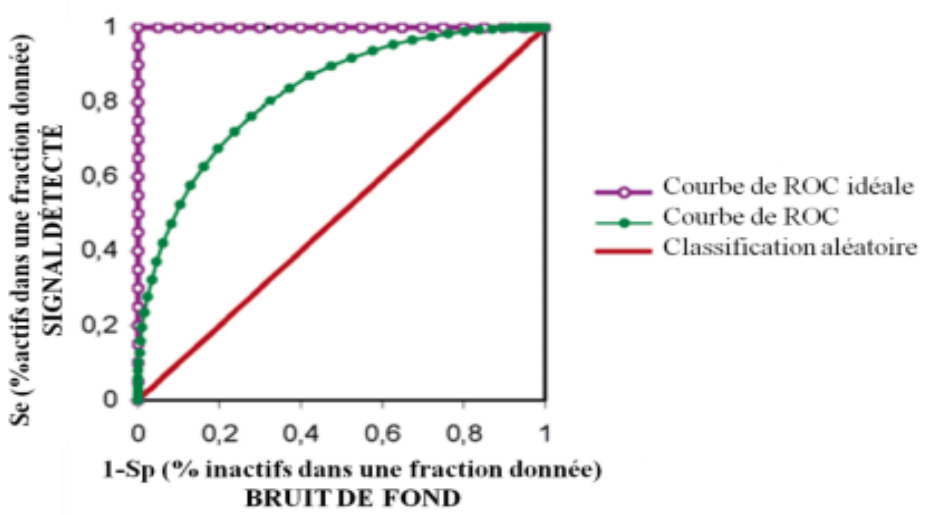


Figure 2: Courbe ROC (***R**eceiver **O**perating **C**haracteristic curve*) [52].

La diagonale allant du point (**0,0**) au point (**1,1**), colorée en rouge sur la figure 2, représente une **classification aléatoire** des composés de la chimiothèque. Si, une méthode permet de distinguer les actifs et les inactifs de manière plus efficace que le hasard, la majorité des points de la courbe ROC, est située au-dessus de la diagonale (courbe verte). Pour une **distribution idéale**, les actifs sont tous classés en premier. C'est à dire, la courbe monte en ligne droite, jusqu'au point **(0,1**), Où la (**Se = Sp =1**, pour tous les actifs) [52], puis viennent ensuite les inactifs **(Se = 1** et **Sp = 0**, pour tous les inactifs) représentés par une ligne horizontale entre les points **(0,1**) et **(1,1**) de la figure 2 (courbe représente en violet) [52].

La quantification de la performance globale d'enrichissement, de la méthode de criblage étudiée à partir de la courbe ROC, se fait à l'aide de calcul de l'aire sous la courbe, généralement notée **AUC** (**A**rea Under the **C**urve), qui est exprimée par l'équation suivante :

$$\mathbf{AUC} = \sum[\mathbf{Se_{i+1}}(\mathbf{Sp_{i+1}} - \mathbf{Sp_i})] \quad \text{.....} \quad \mathbf{(23)}$$

La valeur maximale de **AUC** est de **1**. Elle est obtenue pour les cas idéaux, où tous les actifs sont associés à un score supérieur, à ceux de tous les composés inactifs (decoys).

Cependant, plus la valeur de **AUC** associée à une méthode est élevée, plus la capacité de la méthode à distinguer les actifs des inactifs est forte. En effet, les valeurs de **AUC** peuvent donc être utilisées, pour comparer les performances de plusieurs méthodes de docking, ce qui est aussi possible par une inspection visuelle des courbes **ROC**.

La valeur de **AUC** est calculée, en sommant tous les rectangles formés par les valeurs de "**Se**", et "**1-Sp**" de chaque point de la courbe ROC représentée sur la figure 2, correspondant à la molécule classée en position **i**.

En effet, pour une classification aléatoire des composés de la chimiothèque, la valeur de AUC est égale à **0,5**. En effet, les performances d'une méthode résultantes à une AUC inférieure ou égale à **0,5** sont mauvaises, puisque la méthode ne permet pas d'obtenir une meilleure distinction entre les actifs (hit) et les inactifs (decoys) que celle générée par le hasard [49].

III.3 Méthode quantitative de relation structure-activité (QSAR)

Le **QSAR** (***Q**uantitative **S**tructure **A**ctivity **R**elationship*, en anglais), est une technique de corrélation basée sur des calculs statistiques, qui consiste à établir une relation mathématique à l'aide de méthodes d'analyse de données structurelles, entre les propriétés moléculaires microscopiques, d'une série de composés chimiques similaires, appelée

« descripteurs moléculaires » et leurs effets expérimentaux (activité biologique, toxicité ou affinité vis-à-vis d'un récepteur) [53-54].

III.3.1 Paramètres biologiques (base de données QSAR)

Une base de données QSAR, doit contenir des résultats fiables et homogènes, d'un ensemble de composés chimiques testés expérimentalement. Cette base est répartie approximativement, en deux ensembles : un lot d'essai ou d'entrainement (training set), contenant ¾ de l'ensemble de molécules et un lot de test (test set), contenant ¼ des molécules restantes [53-54].

Les données biologiques ou biochimiques des composés étudiés (ex : concentration inhibitrice IC50), sont généralement exprimées sur une échelle logarithmique, en raison de la relation linéaire entre la réponse et le logarithme de la concentration du composé. En outre, les logarithmes inverses de l'activité (log 1/IC50), sont également utilisés pour obtenir des valeurs mathématiques plus élevées, lorsque les structures sont fortement bioactives [55-57].

III.3.2 Descripteurs moléculaires

L'activité biologique est codée par des grandeurs, qui représentent d'une manière quantitative les informations contenues, dans une structure moléculaire, appelées : « descripteurs moléculaires » [55-57], qui sont définis selon les catégories suivantes :

- **Descripteurs constitutionnelsD1 :** Ils sont faciles et rapides à calculer, à partir de la composition atomique de la structure moléculaire brute (ex : nombre d'atome, nombre de liaison, nombres de cycles) [58].

- **Descripteurs topologiques D2 :** Ils utilisent des représentations graphiques bidimensionnelles des molécules, contenant des informations sur la connectivité des fragments moléculaires, et des estimations de propriétés physicochimiques. En effet, les indices topologiques sont généralement utilisés, pour l'interprétation des équations QSAR résultantes, afin de comprendre les liaisons et interactions physicochimiques.

Cependant, les descripteurs 2D sont insuffisants, pour expliquer convenablement certaines propriétés ou activités biologiques [59-60].

- **Descripteurs géométriques D3 :** Ils décrivent les caractéristiques moléculaires complexes, tel que les propriétés :

- ✓ **Géométriques :** Le volume, la surface moléculaire, le moment d'inertie, les distances et angles dièdres particuliers entre les atomes dans une molécule, le nombre de liaisons hydrogènes, la surface, volume et liaisons de Van Der Waals, etc. [61].
- ✓ **Physicochimiques :** Propriétés hydrophiles ou lipophiles (coefficient de partage, log P (Octanol/eau), la polarisabilité, le nombre de liaisons hydrogènes donneurs et le nombre de liaisons hydrogènes accepteurs [61].
- ✓ **Thermodynamiques :** La température, la pression et l'enthalpie [61].
- ✓ **Electroniques :** Les interactions inter et intramoléculaires, données énergétiques et vibrationnelles [61].
- ✓ **Structurelles :** Les forces d'interactions inter et intramoléculaires.

Le calcul des descripteurs moléculaire 3D, nécessite de connaître la géométrie moléculaire optimisée ou les coordonnées cristallographiques de la molécule et demandent un temps de calcul important. Cependant, ils apportent toutes les informations nécessaires à la modélisation de l'activité biologique [61].

III.3.3 Méthodes mathématiques de la modélisation QSAR

III.3.3.1 La régression linéaire multivariée (*Multiple Linear Regression : MLR*)

La MLR, est la plus simple des méthodes mathématiques, utilisées dans les études QSAR. Elle, repose sur l'hypothèse qu'il existe une relation linéaire, entre une variable dépendante Y (activité biologique) et une série de n variables indépendantes notés Xi (Descripteurs moléculaires) [62].

En effet, le nombre de variables doit être inférieur, ou égal au nombre de molécules, pour décrire le système par la formule suivante :

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \cdots .. + a_nX_n \text{.....(24)}$$

Où :

a_i (i=0, 1, …… n) : Sont les coefficients de régression

X_1, …..,X_n : Sont des descripteurs moléculaires, relatifs aux coefficients a_1,....a_n.

Cependant, l'application de la régression multiple linéaire est limitée, vu qu'elle nécessite l'utilisation d'une base de données complète et idéale, contenant des informations fiables issues de l'expérimentation [62-63].

III.3.3.2 La régression en composantes principales (*Principal Component Regression : PCR*)

La PCR, est une méthode statistique très puissante, qui permet d'analyser, de visualiser et d'explorer la structure des données, afin de déterminer les corrélations entre la variable indépendante et les variables quantitatives d'entrée [64].

III.3.3.3 Méthodes des moindres carrés partiels (*partial least squares : PLS*)

La régression PLS, est une méthode rapide, efficace et optimale pour une parfaite minimisation des covariances. Cette méthode est très utilisée lorsqu'il y'a un grand nombre de variables prédictives, ou lorsqu'il y a de fortes colinéarités entre ces dernières [65].

En premier temps, la PLS réduit les variables prédictives, à un plus petit ensemble de composantes non corrélées. Ensuite, elle effectue la régression par les moindres carrés, sur ces composantes plutôt que sur les données initiales [66].

Contrairement à la régression linéaire multiple (MLR), la PLS ne suppose pas, que les variables prédictives sont fixes. Ainsi, les mesures de ces variables tolèrent des erreurs, en incluant des incertitudes des mesures [67].

III.3.4 Méthodes de validation des modèles QSAR

III.3.4.1 Validation non-croisée

La validation non-croisée (*Non-Cross Validation*), est appliquée sur le lot d'entrainement, afin de vérifier et de déterminer la stabilité du modèle QSAR et de tester l'impact et l'influence de chaque molécule sur l'ensemble d'essai, en utilisant la technique Leave-n-out [68-70], suivante :

III.3.4.1.1 Leave-n-out

Cette méthode, consiste à extraire à chaque fois, un certain nombre « n » de molécules de lot d'essai initial à « k » molécules et à construire un nouveau modèle, avec les « k-n » molécules restantes, en utilisant les mêmes descripteurs moléculaires, choisis initialement pour construire le modèle d'origine [68-70].

Ensuite, ce processus est répété « p » fois, pour retirer et prédire les valeurs de toutes les molécules de lot d'essai, par le modèle généré [68-70].

III.3.4.2 Validation croisée

La validation croisée, est utilisée afin de valider la fiabilité et la performance prédictive du modèle, dans la détermination de la corrélation existante, entre les activités prédites et celles expérimentales des molécules externes (lot de test).

En effet, plus cette corrélation est importante, plus le modèle est performant et valide, pour prédire les activités biologiques des molécules, qui n'appartiennent pas au lot d'essai [71-75]. Les méthodes de validation croisée les plus populaires, sont les suivantes :

III.3.4.2.1 *Leave-one-out* (LOO)

Pour un nombre N de molécules de lot d'essai, une molécule i est retirée à chaque itération, de lot d'essai initial. Donc, un nouvel ensemble d'essai est construit, pour les N-1 molécules restantes, et par conséquent la molécule retirée est prédite par le modèle formé [71].

III.3.4.2.2 *Leave-two-out* (LTO)

La validation croisée en double (LTO), consiste à supprimer deux molécules, de la base de données initiale (en deux étapes consécutives), sur lesquels l'interpolation doit être effectuée, afin de déterminer le meilleur paramètre optimal [72-73].

III.3.4.2.3 *Leave-many-out* (LMO)

La validation croisée dite LMO « leave-n-out », est appliquée pour éviter une surestimation de la capacité prédictive du modèle QSAR résultant, en excluant 50 % des composés à chaque étape. Cette méthode, consiste à subdiviser le lot d'essai en « n » parties, et de sélectionner un échantillon parmi les « n » échantillons générés, comme un lot de validation.

En effet, les « n-1 » échantillons restants, sont considérés comme un nouveau lot d'essai (ensemble d'entrainement) et le coefficient de corrélation Q^2 (r^2cv), entre les activités biologiques et celles prédites est calculé systématiquement [74].

III.3.4.2.4 Paramètres statistiques de la validation croisée

La performance et la fiabilité des modèles de régression, sont évaluées par les paramètres statistiques [71-74], de la validation croisée suivants :

- ✓ L'écart type de la validation croisée S_{CV}, est calculé par la formule suivante :

$$S_{CV} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_{icv})^2} \quad \text{.....(25)}$$

- ✓ Coefficient de prédiction de la validation croisée Q^2_{cv} , est calculé par la formule suivante :

$$Q^2_{cv} = 1 - \frac{\sum_{i=1}^{N}(Y_{prédite} - Y_{réelle})^2}{\sum_{i=1}^{N}(Y_{réelle} - Y_{moyenne})^2} \text{..... (26)}$$

Où : $Y_{Prédite}$, $Y_{réelle}$ et $Y_{moyenne}$: sont les valeurs prédites, réelles et moyennes de l'activité biologique (pIC_{50}) de la molécule i, respectivement.

III.3.5 Paramètres de calculs

III.3.5.1 Coefficient de détermination (r^2)

Le coefficient de détermination r^2, est la mesure de degré de liaison entre la variante dépendante (activité biologique) et la variante indépendante (Descripteur moléculaire). En effet, une valeur de r^2 proche de 1, indique un bon ajustement du modèle de régression [53-54,75].

Le coefficient R^2, est calculé selon l'équation suivante :

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \text{..... (27)}$$

A noter que :

TSS : Somme des carrées totaux
ESS : Somme des carrées exprimés par le modèle de régression
RSS : Somme des carrées résiduels.

III.3.5.2 Racine carrée de coefficient de détermination (R)

La racine carrée de coefficient de détermination R, appelée également le coefficient de corrélation, est calculé entre les valeurs prédites et celles expérimentales, par le modèle de régression, sa valeur variée entre -1 et 1 [53-54,73-75].

En effet, plus la valeur absolue du coefficient de corrélation R est proche de 1, plus la régression linéaire est idéale et le modèle linéaire est fiable et valide [75].

III.3.5.3 Test Fischer (F)

Le test Fischer, permet de justifier la liaison entre l'activité biologique et le descripteur moléculaire, en mesurant le rapport entre la variance de l'activité biologique expliquée et non expliquée par le modèle de régression.

En effet, le test de Fischer permet de tester l'hypothèse nulle, lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, si la valeur de F observée dépasse le seuil (F > seuil) l'hypothèse sera rejetée, et dans le cas contraire (F < seuil) l'hypothèse sera conservée [53-54,75]. Le paramètre « F », est calculé comme suit :

$$\mathbf{F = \frac{ESS}{P} = \frac{n-p-1}{RSS}} \quad \textbf{(28)}$$

Où : ESS, est associé avec p degrés de liberté et RSS associé avec n-p-1 degrés de liberté.

III.3.5.4 Ecart type (s)

La valeur de l'écart type (s), indique dans quelle mesure la fonction de régression prédit les données observées [53-54,75]. Ce paramètre est calculé par l'équation suivante :

$$\mathbf{S = \sqrt{\frac{RSS}{n-p-1}}} \quad \textbf{(29)}$$

Où : **P** : est le nombre de variables indépendants (descripteurs moléculaires).

III.3.5.5 Coefficient de prédiction (q^2)

Le coefficient de prédiction q^2, est une mesure de la justesse de la prédiction, puisqu'il mesure la capacité prédictive d'un modèle de régression [53-54,75]. Ce paramètre est exprimé comme suit :

$$\mathbf{Q^2 = 1 - \frac{RSS}{TSS}} \quad \textbf{(30)}$$

III.3.5.6 Coefficient de corrélation prédit Q

La racine carrée du coefficient de prédiction Q, appelée également le coefficient de corrélation prédit, est calculé entre les valeurs prédites et celles expérimentales, lors de la validation croisée du modèle de régression. En effet, plus la valeur absolue du coefficient de corrélation prédit Q est proche de 1, plus la modèle de régression est fiable et valide [53-54,73-75].

IV. Références du chapitre II

[1] Bergon A., (2012). Thèse de doctorat. Université de la méditerranée -AIX Marseille II, 308p.
[2] Wilfred F., Gunsteren V., Herman Berendsen J. C., (1990). Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angewandte Chimie International Edition, 29 (9) : 939-1076. https://doi.org/10.1002/anie.199009921
[3] Rivail J. L., (1994). ELÉMENTS DE CHIMIE QUANTIQUE, À L'USAGE DES CHIMISTES. Inter Editions/CNRS Editions, 2éme édition, paris. ISBN 10 : 2271052416ISBN 13 : 9782271052414
[4] Vulliermet N., (2000). Thèse de doctorat N° 3188. Investigation théorique du mécanisme de physisorption : application d'une méthode de partition fondée sur la fonctionnelle de la densité. Université de Genève (Suisse), 117p. https://doi.org/10.13097/archive-ouverte/unige:98
[5] Schrödinger E., (1926). Quantisierung als Eigenwertproblem. ANNALEN DER PHYSIK, 385(13) : 437-490p. https://doi.org/10.1002/andp.19263851302
[6] Schrödinger E., (1926). An Undulatory Theory of the Mechanics of Atoms and Molecules. Physical Review Journals Archive, 28(6), 1049p. https://doi.org/10.1103/PhysRev.28.1049
[7] Ayache H., (2018). Thèse de doctorat : Propriétés structurales et électroniques des complexes organométalliques : étude DFT, 2018, UMMTO, Algérie
[8] Merzoud L., (2018). Thèse de doctorat : Etude théorique de l'évolution statique et dynamique des réactions péricycliques : cas de la réaction de Diels-Alder, 2018, UMMTO, Algérie.
[9] : LOMAS J.S, (1986). L'actualité chimique, la mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l'énergie d'entités moléculaire 7-22p.
[10] Burkert U., Allinger N. L., (1982). Molecular Mechanics. Journal of the American Chemical Society, monograph, Washington, 339p. https://doi.org/10.1002/jcc.540040420|
[11] Dugas H., (1996). Méthodes de La modélisation moléculaire. Chapitre 3. 4éme édition, Librairie université de Montréal.
[12] Tristram F., Durier V., Vergoten G., (1996). The structures and vibrational frequencies of a series of aliphatic alcohols as obtained using the spectroscopic potential SPASIBA. Journal of Molecular Structure, 378 (3): 249-256.https://doi.org/10.1016/0022-2860(95)09183-1
[13] Leontidis E., Pablo J., Laso M., Suter U.W., (1994). A critical evaluation of novel algorithms for the off-lattice Monte Carlo simulation of condensed polymer phases. Advances in Polymer Science, 116: 283-318.https://doi.org/10.1007/BFb0080202
[14] Chen J., Houk K N., (1996). Molecular Modeling: Principles and Applications. J. Chem. Inf. Comput. Sci. 38 (5): 939p. https://doi.org/10.1021/ci9804241
[15] Diu B., Guthmann C., Lederer D., Roulet B., (1989). Livre de Physique statistique, Hermann, paris.
[16] Hercend C., (2012). Thèse de doctorat : Contribution de la modélisation moléculaire à l'étude de Pathologies humaines : Application au transporteur ATP7B et au récepteur 5HT2B. Paris : Université paris, 34-37 p.
[17] Poger D., (2005). Thèse de doctorat : Biophysique. Biologie Structurale et Nanobiologie, Université Joseph-Fourier-Grenoble I (paris), 201 P.
[18] Shrake A., Rupley J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. Journal of Molecular Biology, 79 (2): 351- 71. https://doi.org/10.1016/0022-2836(73)90011-9
[19] Lee B., Richards F. M., (1971). The interpretation of protein structures: Estimation of static accessibility. Journal of Molecular Biology, 55 (3): 379 - 400. https://doi.org/10.1016/0022-2836(71)90324-X
[20] Essman U., Perera L., Berkowitz ML., Darden T., Lee H., Pedersen LG., (1995). A smooth particle mesh Ewald method. Journal of Chemical Physics 103:8577–8593. https://doi.org/10.1063/1.470117
[21] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983). Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics,79:926–935. https://doi.org/10.1063/1.445869
[22] Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., and Haak J. R, (1984). Molecular dynamics with coupling to an external bath. Journal of Chemical Physics 81(8) 3684-3690, https://doi.org/10.1063/1.448118
[23] Wolber G., Seidel T., Bendix F., Langer T., (2008). Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discovery Today, 13, 23-29. https://doi.org/10.1016/j.drudis.2007.09.007
[24] Taylor R. D., Jewsbury P. J., Essex J. W., (2002). A review of protein-small molecule docking methods. Journal of Computer-Aided Molecular Design. 16,151-166. https://doi.org/10.1023/A:1020155510718

[25] Leach A R,. (2001). Molecular Modelling: Principles and Applications. Prentice Hall (Pearson Education), Modelling: Principles and Applications, 2(2): 744p. https://doi.org/10.1093/bib/2.2.199
[26] Morris G M., Huey R., Lindstrom W., Sanner M F., Belew R K., Goodsell D S., et Olson A J., (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J. Computational Chemistry, 30(16): 2785-91. https://doi.org/10.1002/jcc.21256
[27]FlexX version 2.3.2; BioSolveIT GmbH, Sankt Augustin, Germany, 2017, www.biosolveit.de/FlexX
[28] Mitchell M., (1996). An introduction to genetic algorithms. First MIT Press paperback edition, Massachusetts Institute of Technology, London, England, 143p.
[29] Vogt A. D., Cera E. D., (2013). Conformational selection is a dominant mechanism of ligand binding. Biochemistry, 52: 5723-5729. https://doi.org/10.1021/bi400929b
[30] Ferrari A. M., Wei B. Q., Costantino L., Shoichet B. K., (2004). Soft Docking and Multiple Receptor Conformations in Virtual Screening. Journal of Medicinal Chemistry, 47 (21): 5076-5084. https://doi.org/10.1021/jm049756p
[31] Bonvin A., (2006). Flexible protein–protein docking. Current Opinion in Structural Biology, 16(2):194-200. https://doi.org/10.1016/j.sbi.2006.02.002
[32] Zhao Y., Stoffler D., Sanner M., (2006). Hierarchical and multi-resolution representation of protein flexibility. Bioinformatics. 22(22): 2768-2774. https://doi.org/10.1093/bioinformatics/btl481
[33] Walters W. P., Stahl M. T., et Murcko M. A., (1998). Virtual screening—an overview. Drug Discovery Today, 3(4): 160-178. https://doi.org/10.1016/S1359-6446(97)01163-X
[34] Tanrikulu Y., Kruger B., et Proschak E., (2013). The holistic integration of virtual screening in drug discovery. Pub Med. 18 (7-8): 358-364. https://doi.org/10.1016/j.drudis.2013.01.007
[35] Battersby B J., Trau M., (2002). Novel miniaturized systems in high-throughput screening. Trends in Biotechnology, 20(4) :167-173. https://doi.org/10.1016/S0167-7799(01)01898-4
[36] Rollinger J M., Stuppner H., Langer T., (2008). Virtual screening for the discovery of bioactive natural products. 65 (211) : 211-213. https://doi.org/10.1007 / 978-3-7643-8117-2-6
[37] Lagarde N., (2014). 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 du Conservatoire National des Arts et Métiers : bio-informatique. École Doctorale Abbé Grégoire, Paris, France, 284p.
[38] Sun H., (2008). Pharmacophore-Based Virtual Screening. CurrentMedicinal Chemistry. 15 (10): 1018-1024. https://doi.org/10.2174/092986708784049630
[39] Cummings B. S., (2007). Phospholipase A2 as targets for anti-cancer drugs. BiochemPharmacol. 74 (7): 949-959. https://doi.org/10.1016/j.bcp.2007.04.021
[40] Koeppen H., Kriegl J., Lessel U., Tautermann S., Wellenzohn B., (2011). Ligand-Based Virtual Screening. 536p. https://doi.org/10.1002/9783527633326.ch3
[41] Hristozov D P., Oprea T I., Gasteiger J., (2007). Virtual screening applications: a study of ligand-based methods and different structure representations in four different scenarios. Journal of Computer Aided Molecular Design. 21(10-11) :617-40. https://doi.org/10.1007/s10822-007-9145-8
[42] Bender A., Glen R.C., (2004). Molecular similarity: a key technique in molecular informatics. Organic Biomolecular Chemistry, 2(22): 3204-3218. https://doi.org/10.1039/B409813G
[43] Wermuth C.-G., Gannelin C. R., Lindberg P., et Mitscher L A., (1998). Glossary of terms used in medicinal chemistry. Pure and Applied Chemistry. 70 (5): 1129-1143. https://doi.org/10.1351/pac199870051129
[44] Horvath D., Marcou G., et Varnek A., (2009). Chemical Domain of QSAR Models from Atom-Centered Fragments. Journal of Chemical Information and Modeling. 49(12): 1762-1776. https://doi.org/10.1021/ci900313u
[45] Leach A.R., Shoichet B.K., Peishoff C. E., (2006). Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps. Journal of Medicinal Chemistry. 49(20): 5851-5855. https://doi.org/10.1021/jm060999m
[46] Irwin J., (2008). Community benchmarks for virtual screening. Journal of Computer-Aided Molecular Design. 22(3-4): 193-199. https://doi.org/10.1007/s10822-008-9189-4
[47] Hawkins P.C., Warren G.L., Skillman A.G., (2008). Recommendations for evaluation of computational methods. Journal of Computer Aided Molecular Design. 22(3-4): 179-190. https://doi.org/10.1007/s10822-008-9196-5
[48] Abagyan R A., Totrov M.M., (1997). Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. Journal of Molecular Biology. 268(3): 678-685. https://doi.org/10.1006/jmbi.1997.0994

[49]Kirchmair J., Markt P., Distinto S., Wolber G., Langer T., (2008). 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. 22(3-4): 213-228. https://doi.org/10.1007/s10822-007-9163-6

[50] Kroemer R.T., Vulpetti A., McDonald J.J., Rohrer D C., Trosset J Y., Giordanetto F., Cotesta S., McMartin C., Kihlén M., Stouten P., (2004). Assessment of Docking Poses: Interactions-Based Accuracy Classification (IBAC) versus Crystal Structure Deviations . J Chem Inf Comput Sci. 44(3): 871-81. https://doi.org//10.1021/ci049970m

[51] Thomas IV B E., Joseph-McCarthy D., Alvarez J C., (2000). Combining Structure-Based Drug Design and Pharmacophores. University Line, La Jolla, California,353-367.

[52] Triballeau N., Acher F., Brabet I., Pin J P., Bertrand H O., (2005). Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. Journal of Medicinal Chemistry. 48(7): 2534- 2547. https://doi.org/10.1021/jm049092j

[53] Mark T D Cronin., Schultz T.W., (2001). Development of Quantitative Structure−Activity Relationships for the Toxicity of Aromatic Compounds to Tetrahymena pyriformis: Comparative Assessment of the Methodologies. Chemical Research in Toxicology. 14, 1284-1295. https://doi.org/10.1021/tx0155202

[54] Balaza S., Sturdik E., Durkova E., Antalik M., Sulo P., (1986). Quantitative structure-activity relationship of carbonylcyanide phenylhydrazones as uncouplers of mitochondrial oxidative phosphorylation. Biochimica et Biophysica Acta (BBA) – Bioenergetics. 851(1): 93-98. https://doi.org/10.1016/0005-2728(86)90252-5.

[55] Bouaziz-Terrachet S., Terrachet R., Tairi-Kellou S., (2013). Receptor and ligand-based 3D-QSAR study on a series of nonsteroidal anti-inflammatory drugs. Med ChemRes 22(4) :1529–1537. https://doi.org/10.1007/s00044-012-0174-z

[56] Bouaziz-Terrachet S., (2006). Conception d'inhibiteurs sélectifs de l'activité enzymatique de la cyclo-oxygénase-2" COX-2". Mémoire de Magister, Alger, USTHB, N° d'ordre :14/2006-M/CH

[57] Tairi-Kellou S., Bouaziz Terrachet S., Maouche B., et Moreau G., (2008). 2D–QSAR Autocorrelation Study on Selective COX–2 Inhibitors. Internet Electron. J.Mol. Des, 7 : 161-185

[58] Benson S W., Buss J H., (1958). Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. J. Chem. Phys., 29(3): 546-572. https://doi.org/10.1063/1.1744539

[59] Geerlings P F., Proft D., Langenaeker W., (2003). Conceptual density Functional Theory. Chem. Rev.,103 (5): 1793-1874. https://doi.org/10.1021/cr990029p

[60] Chermette H., (1999). Chemical reactivity indexes in density functional theory. J. Comput. Chem., 20(1): 129-154. https://doi.org/10.1002/(SICI)1096-987X(19990115)20:1<129::AID-JCC13>3.0.CO;2-A

[61] Labute P., (2000). A widely applicable set of descriptors. J Mol Graph Model, 18(4-5):464-77. https://doi.org/10.1016/s1093-3263(00)00068-1

[62] Montgomery D C., Elizabeth A P., G. Geoffrey V G., (2013). Introduction to Linear Regression Analysis, Fifth Edition by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. International Statistical Review, 81(2):318-319. https://doi.org/10.1111/insr.12020_10

[63] Croux C., Joossens K., (2005). Influence of observations on the misclassification probability in quadratic discriminant analysis. J Multivar Anal, 96(3):348 –403. https://doi.org/10.1016/j.jmva.2004.11.001

[64] Hotelling H., (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6): 417–441.https://doi.org/10.1037/h0071325

[65] Wold S., (1994). Methods and Principles in Medicinal Chemistry, Van de Waterbeemd, H. (Ed.). Verlag-Chemie, Weinheim, Germany.

[66] Clark M., Cramer R D., (1993). The Probability of Chance Correlation Using Partial Least Squares (PLS). Quantitative Structure-Activity Relationships, 12(2), 137-145. https://doi.org/10.1002/qsar.19930120205

[67] Sjöström M., Eriksson L., (1995). Chemometric methods in molecular design, han van de Waterbeemd, 2: 63- 90.

[68] Yao Y., Cheng X Z., Wang L. X., Wang S. H., Ren G. X., (2011). A Determination of Potential α-Glucosidase Inhibitors from Azuki Beans (Vigna angularis). Inter. J. Mol. Sci.,12(10): 6445-6451. https://doi.org/10.3390/ijms12106445

[69] Hawkins D M., (2004). The Problem of Overfitting. Introduction The Purpose of Quantitative Models. J. Chem. Inf. Comput. Sci. 44 (1): 1–12. https://doi.org/10.1021/ci0342472

[70] Tropsha A., Gramatica P., Gombar V K., (2003). The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR & Comb. Sci., 22 (1): 69–77. https://doi.org/10.1002/qsar.200390007
[71] Gramatica P., (2013). On the Development and Validation of QSAR Models. Methods in molecular biology (Clifton, N.J.) 930:499-526 https://doi.org/10.1007/978-1-62703-059-5_21
[72] Kunal R., Indrani M., (2012). On the Use of the Metric rm 2 as an Effective Tool for Validation of QSAR Models in Computational Drug Design and Predictive Toxicology. Mini Reviews in Medicinal Chemistry,12(6):491-504. https://doi.org/10.2174/138955712800493861
[73] Azarboni R H., Keyanpour M., Yaghouti M., (2019). Leave-Two-Out Cross Validation to optimal shape parameter in radial basis functions. Engineering Analysis with Boundary Elements,100: 204-210. https://doi.org/10.1016/j.enganabound.2018.06.011
[74] Shameera T K A., Vijisha K., Rajan K M., (2019). QSAR modeling of benzoquinone derivatives as 5-lipoxygenase inhibitors. Food Science and Human Wellness, 8(1): 53-62. https://doi.org/10.1016/j.fshw.2019.02.001
[75] Luco J M., Ferretti F H., (1997). QSAR Based on Multiple Linear Regression and PLS Methods for the Anti-HIV Activity of a Large Group of HEPT Derivatives. J. Chem. Inf. Comput. Sci., 37(2): 392-401. https://doi.org/10.1021/ci960487o
[76] Pearson B., (2008). Pearson's Correlation Coefficient. Correlation Coefficient. In: Kirch W. (eds) Encyclopedia of Public Health. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5614-7_2569

Référence de l'annexe 01

[77] Israelachvili J N., (2011). Intermolecular and Surface Forces - 3rd Edition – Elsevier. Academic Press, 3e édition, Londres, 704p.
[78] Martin T W., Derewenda Zygmunt S., (1999). The name is bond — H bond. Nature Structural Biology. 6 (5): 403 – 406. https://doi.org/10.1038/8195
[79] Gerschel A., (2012). Liaisons intermoléculaires. EDP Sciences, 2e éd., Paris, 288 p.
[80] : Van derVegt N., Nayar D., (2017). The Hydrophobic Effect and the Role of Cosolvents. J. Phys. Chem., 121(43) : 9986–9998. https://doi.org/10.1021/acs.jpcb.7b06453.
[81] Böhm H J., (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. Journal of Computer-Aided Molecular Design. 8(3): 243-256. https://doi.org/10.1007/BF00126743
[82] Klebe G., (2006). Drug Discovery Today. Virtual ligand screening: strategies, perspectives and limitations. Pub Med. 11(13-14): 580-594. https://doi.org/10.1016/j.drudis.2006.05.012
[83] Bonnet P., Barriol J., (2017). Détermination de la polarisabilité moléculaire à partir de mesures de permittivité de liquides purs non polaires. J. Chim. Phys., 83(1986) : 519–525. https://doi.org/10.1051/jcp/1986830519

I. Introduction

L'objectif de ce chapitre est d'identifier de nouveaux inhibiteurs sélectifs de la protéine caséine kinase (CK2), protéine impliquée dans de nombreuses pathogènes cancéreuses, en utilisant les approches de la modélisation moléculaire et de la bio-informatique.

Dans la première étape de notre travail, nous avons utilisé les simulations de screening virtuel basées sur la modélisation de pharmacophore, qui consistent à simuler l'affinité d'un très grand nombre de ligands pour une cible thérapeutique donnée, en vue d'identifier des ligands potentiellement actifs pour la protéine étudiée [1-5] à l'aide de logiciel Ligand Scout [6].Les composés retenus dans la première étape, sont soumis par la suite à une analyse de propriétés « *Drug likeness* », afin d'éliminer les composés ayant les propriétés physico-chimiques les moins similaires avec les médicaments disponibles sur le marché en utilisant le logiciel DruLiTo[7].

Dans le but de sélectionner uniquement les composés prometteurs qui réagissent plus favorablement avec la cible d'intérêt, nous avons effectué des simulations de docking moléculaire sur les composés obtenus avec le logiciel FlexX implémenté dans le programme lead IT [8]. De plus, l'analyse de toxicité *in silico* effectuée à l'aide de serveur Web ProTox-II [9], a montré l'absence d'effets indésirables dangereux liés à la toxicité des composés proposés par notre étude.

Dans la dernière étape de notre étude, nous avons réalisé des simulations de dynamique moléculaire pour confirmer la stabilité des composés issus de la présente étude dans le site actif de la protéine étudiée en moyen du logiciel NAMD [10]. L'évaluation des résultats a été effectuée en calculant les valeurs des métriques suivantes : Déviation quadratique moyenne (RMSD), le nombre de liaisons hydrogène entre les ligands et la protéine, la fluctuation quadratique moyenne (RMSF), les changements des surfaces accessibles aux solvants (SASA) ainsi que le rayon de giration (rGyr). Les résultats de cette étude doivent permettre de guider notre recherche de nouvelles molécules inhibitrices et sélectives de la CK2.

II. Matériel et méthodes

II.1 Génération du modèle pharmacophore basé sur la structure de la cible « Structure-Based »

Dans un premier temps, nous avons exploré le mode de liaison des flavonoïdes dans le site actif de la CK2 pour identifier les interactions importantes pouvant expliquer l'inhibition de cette protéine par cette famille de composés. A cet effet, nous avons sélectionné, la structure de la protéine cristallisée avec un inhibiteur dérivant de la famille des flavonoïdes. Cette structure, codée 3AMY ayant une résolution égale à 2.30 Å, correspond au complexe : CK2α humaine / Apigenin (AGI). Le ligand AGI, dérivant de la famille des flavonoïdes, est un inhibiteur sélectif de la CK2 (Figure III-1). La structure cristallographique 3AMY est constituée d'une seule chaine constituant le monomère. Le choix de la structure 3AMY résulte d'un compromis entre une bonne résolution et le fait de disposer d'informations d'origine expérimentales quant au mode de liaison des flavonoïdes à la protéine CK2 humaine. Ainsi, les chaines latérales des acides aminés constituant la protéine ont été optimisées avec le logiciel LigandScout [6], en utilisant le champ de force « MMFF94 ». Le but de cette simulation est de supprimer les mauvais contactes et de relaxer la structure de la protéine étudiée.

Le modèle pharmacophore basé sur la structure du récepteur (*structure-based*), permettant d'avoir des informations directes sur les interactions s'établissant entre le ligand et la protéine CK2, a été construit par le logiciel LigandScout [6]. La visualisation en 3D du modèle ainsi généré nous a permis d'identifier et de sélectionner les points pharmacophoriques favorables. Les interactions entre le ligand et la protéine sont définies comme des points pharmacophoriques qui peuvent être du type donneurs et accepteurs de liaison hydrogène (HBD et HBA), les zones chargées positivement et les zones chargées négativement (PI et NI, respectivement) réalisant des interactions ioniques, les groupements aromatiques conduisant à du π-stacking (Ar) et les groupements hydrophobes donnant des interactions hydrophobes (Hy). Pour optimiser davantage le modèle pharmacophorique ainsi généré (modèle 1), nous avons ajouté des sphères de volume d'exclusion représentant ainsi la forme du site actif. Il est bien de noter que, pour les liaisons hydrogène, la distance est comprise entre 2.5 et 3.8 Å par défaut et l'angle entre les donneurs et les accepteurs est de 180° ± 34° [6]. De plus, la flexibilité entre les atomes sp^2 et sp^3 et les atomes d'hydrogènes adjacents est prise en compte.

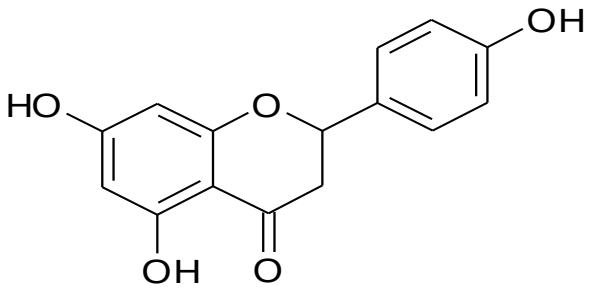
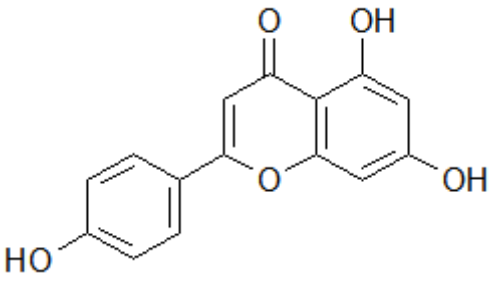
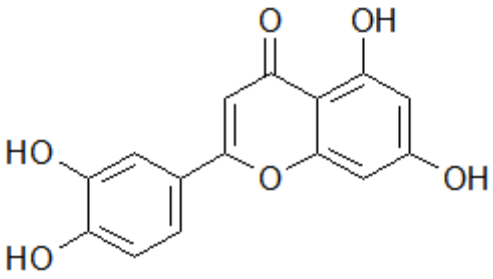


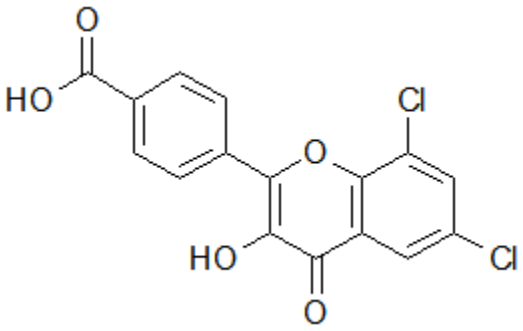
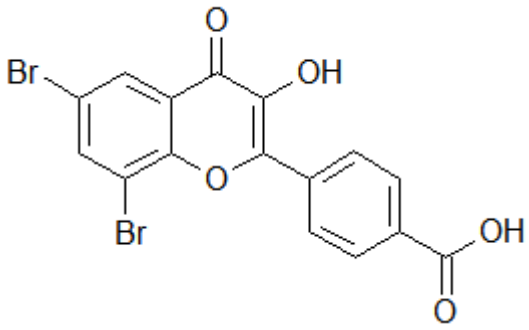
Figure III-1 : Structure chimique de l'Apigenine (AGI)

II.2 Génération du modèle pharmacophore « *Ligand-Based* »

La construction des modèles pharmacophoriques *ligand-based* a été effectuée à l'aide du logiciel LigandScout [6]. En effet, nous avons sélectionné dans la littérature, quatre composés potentiellement actifs pour la CK2, dérivant de la famille des flavonoïdes. Dans le Tableau 1, nous y reportons les noms chimiques et les structures 2D de ces composés, que nous avons utilisé dans le lot d'entrainement (Training-Set) et qui nous ont servi dans la génération du modèle pharmacophore *ligand-based*.

Tableau III-1 : Les noms chimiques et les structures 2D des ligands utilisés dans le lot d'entrainement

Ligand	Nom chimique	Structure
Apigenin	5,7-dihydroxy-2-(4-hydroxyphenyl) -4Hchromen- 4-one)	
Luteolin	5,7-dihydroxy-2-(4-hydroxyphenyl) 4H chromen-4-one)	

FLC21	(4-[6,8-bis(chloranyl)-3-oxidanyl-4-oxidanylidenechromen-2-yl] benzoicacid)	
FLC26	(4-(6,8-dibromo-3-hydroxy-4-oxo-4Hchromen-2-yl) benzoicacid)	

Ainsi, nous avons généré un set composé de 200 conformations pour chaque ligand du lot d'entrainement en utilisant les meilleurs paramètres du programme « iCon » implémenté dans le logiciel LigandScout [6]. Chaque conformation est analysée et les points pharmacophoriques sont placés dans toutes les conformations. Par la suite, les différentes conformations obtenues pour ces composés sont alignées et les propriétés pharmacophoriques présentant des caractéristiques chimiques identiques et géométriques similaires sont identifiées formant ainsi des modèles pharmacophoriques intermédiaires. Ces derniers sont ensuite ordonnés par rapport à une fonction score appelée « pharmacophore fit », qui considère uniquement les propriétés pharmacophoriques et le *RMSD (root mean-square deviation*) de ces propriétés. Le modèle présentant le meilleur score correspondant à l'alignement ayant le plus grand nombre de points pharmacophoriques présents dans toutes les molécules, sont sélectionnées pour construire le modèle pharmacophorique ligand-based final, que nous avons noté modèle 2.

II.3 Validation et optimisation des modèles pharmacophoriques

Les modèles développés ont été testés sur une série constituée de 26 inhibiteurs sélectifs de la CK2, présentant des activités inhibitrices « IC_{50} » inférieures à 10 µM, comme illustré dans le tableau 2 [11].

Pour tester le pouvoir de notre modèle pharmacophore à discriminer les composés actifs des non actifs, nous avons inséré dans le lot d'essai un ensemble de 1000 composés non actifs sélectionnés dans la base de données ZINC [12].

En effet, les vingt-six composés actifs ont été importés dans la partie « Decoy Finder 2.0 » et les identifiants « IDs ZINC » des 1000 composés non actifs « Decoys » ont été obtenus. Le but de cette étape est de vérifier la fiabilité des modèles pharmacophore que nous avons développé.

Nous avons obtenu les valeurs « pIC_{50} » par la conversion en logarithme négatif des valeurs d'activités « IC_{50} » des composés constituant le lot d'entrainement [11]. En effet, la valeur IC_{50} correspond à la concentration du composé nécessaire pour une inhiber à 50 % in vitro l'activité de la cible biologique.

Les structures 2D des composés actifs ont été dessinées avec Marvin Sketch [13] et le lot de composés non actifs « Decoys set » a été construit à l'aide du logiciel Decoy Finder 2.0 [14].

La validation des modèles pharmacophoriques générés, a été effectuée sur la base de plusieurs paramètres comme le « ROC » (*Receiver Operating Characteristic curve*), l'aire sous la courbe ROC (AUC : *Area Under the Curve*) [15-16], la précision de la classification, la sensibilité, la spécificité et le facteur d'enrichissement (EF). Il est bien de noter que nous avons détaillé tous ces paramètres dans le chapitre 2.

Il est bien de noter que la précision de la classification, la sensibilité, la spécificité et le facteur d'enrichissement (EF), ont été calculés en utilisant les équations suivantes :

$$\text{Précision} = \frac{VP}{VP+FN} \quad \text{..... (1)}$$

$$\text{Sensibilité} = \frac{VP}{A} \quad \text{..... (2)}$$

$$\text{Spécificté} = \frac{VN}{N-A} \quad \text{..... (3)}$$

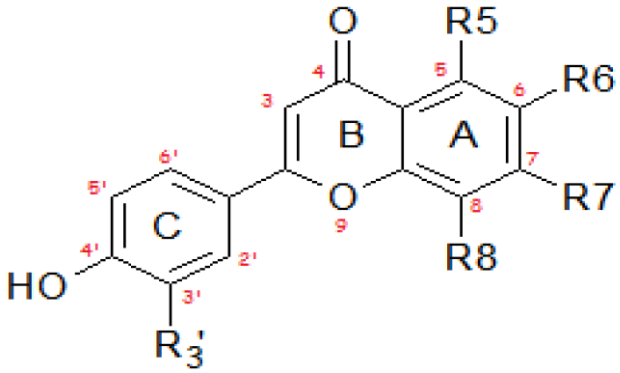
$$EF_{100*(n/N)\%} = \frac{VP/n}{A/N} \quad \text{..... (4)}$$

Où :

« A » : Nombre de composés actifs, « N » : Nombre total de composés, « n » : molécules sélectionnées, « VP » : Les vrais positifs, « FN » : Les faux négatifs, « FP » : Les faux positifs et « VN » : Les vrais négatifs.

Dans ce processus, la fonction d'énergie « *Scoring Function* » a été considérée comme l'énergie de l'alignement du pharmacophore « *pharmacophore fit score* » sans aucune omission de points pharmacophorique.

Tableau III-2 : Fragment commun des flavonoïdes, IC50 (µM), pIC50 = -log (IC50), fit score, énergies du Docking (kJ/mol), et les structures 2D des composés actifs utilisés comme échantillon de test « test set »



Molécule	R3'	R5	R6	R7	R8	IC_{50} (µM)	pIC_{50}	Fit score	Docking energy
1	OMe	H	H	H	H	0.800	-0.10	58.60	-16.84
2	OMe	H	H	Me	H	0.090	-1.05	58.39	-23.24
3	OMe	H	H	Me	Me	0.300	-0.52	58.60	-15.75
4	OMe	H	H	C_4H_4	H	0.028	-1.55	58.48	-24.91
5	OMe	H	Me	H	Me	0.100	-1.00	59.61	-16.60
6	OMe	H	Me	H	H	0.320	-0.49	58.59	-17.28
7	OMe	H	Me	Me	H	3.000	-0.48	58,59	-13.31
8	OMe	H	OMe	H	H	0.200	-0.70	58.61	-18.80
9	OMe	H	Et	H	H	0.120	-0.92	58.59	-20.38
10	OMe	H	Cl	H	H	0.240	-0.62	58.60	-18.66
11	OMe	H	Cl	H	Cl	0.010	2.00	58,30	-19.32
12	OMe	H	Cl	H	Me	0.040	-1.40	58.21	-18.52
13	OMe	H	Cl	Me	H	0.790	-0.10	58.53	-15.58
14	OMe	H	Br	H	H	0.120	-0.92	58.59	-18.70
15	OMe	H	Br	H	Br	0.004	2.40	58,21	-19.98
16	OMe	H	NAc	H	H	0.150	-0.82	58.53	-19.97
17	OMe	H	OH	H	H	0.250	-0.60	58.61	-17.14
18	OMe	Me	H	Me	H	0.030	-1.52	58.28	-18.75
19	OMe	Me	Cl	Me	H	2.000	-0.30	58,56	-13.50
20	OEt	H	H	H	H	7.500	-0.88	58,59	-12.97
21	OEt	H	Me	H	H	8.000	-0.90	58.59	-11.02
22	OH	H	Me	H	Me	0.450	-0.35	58.52	-17.03
23	NO_2	H	Me	H	Me	0.012	1.92	58,15	-22.10
24	Br	H	Me	H	Me	0.050	-1.30	58.20	-18.57
25	H	H	Me	H	H	8.000	0.90	58.76	-11.98
26	H	H	Me	H	Me	1.600	-0.20	58.72	-16.22

II.4 Criblage virtuel

Dans cette partie de notre travail, nous avons exploité le meilleur modèle pharmacophorique validé, pour cribler une grande chimiothèque de molécules, construite manuellement, afin d'obtenir de nouveaux inhibiteurs puissants pour la CK2. En effet, plusieurs centaines de milliers de composés ont été téléchargés à partir de la base de données Pubchem Drug Bank. Ces composés dérivent de différentes familles chimiques, telles que les flavonoïdes, les coumarines, les benzimidazoles et les pyrazolo-triazines.

Au début du processus du criblage virtuel, une chimiothèque de 737 920 composés a été générée, en utilisant le programme LigandScout [6]. Cette dernière a été criblée par la suite avec notre meilleur modèle pharmacophorique validé. De la même manière que lors des étapes de construction des pharmacophores, la flexibilité des molécules de la chimiothèque a été prise en considération par la génération de 200 conformations pour chaque composé. Les valeurs du score d'ajustement du pharmacophore « *Fit Score* », ont été calculées pour chaque composé, en superposant leurs structures chimiques sur le modèle pharmacophorique. En effet, les composés ayant les valeurs de Fit Score les plus élevées indiquant une bonne superposition, ont été sélectionnés et soumis à un autre filtre qui est le « *Drug-Likeness* ».

II.5 Calcul et analyse des propriétés « *Drug-Likeness* »

Dans le but de déterminer les molécules dont les propriétés sont le plus '*drug-like*' possible, nous avons soumis les composés obtenus par le screening virtuel à d'autres filtres basés sur des propriétés physico-chimiques. Parmi les règles utilisées dans ces filtres nous citons les règles de : Lipinski [17], Veber [18], ***Ghose filter*** [19], *Quantitative Estimate of Drug-likeness* (*QED*) [20], *MACCS-II Drug Data Report* (*MDDR*)*-like*, *Comprehensive Medicinal Chemistry (CMC)-50-like*, et *Blood-Brain Barrier* (*BBB*) *likeness* [21], implémentés dans le logiciel DruLiTo [7].

Nous avons choisi l'application du filtre « *BBB likeness* » pour l'analyse « *Drug-Likeness* », afin de déterminer les composés ayant de bons profils de perméabilité pour franchir la barrière hémato-encéphalique (*blood-brain barrier*, en anglais). Cette dernière protège le milieu cérébral des agents pathogènes et des toxines exogènes et endogènes [21]. Il est bien de noter que, des études récentes ont démontré que la protéine CK2 est surexprimée dans certains types de cancer du cerveau et que son inhibition devrait diminuer la taille de la tumeur ainsi que le taux de mortalité animale [22].

De plus, selon les règles de Lipinski (règle de 5) [17], un médicament ayant une bonne absorption orale, ne compte pas plus de 5 liaisons hydrogène donneur, moins de 10 liaisons hydrogène accepteur, moins de 500 Dalton (Da) du poids moléculaire et le coefficient de partage dans l'octanol-eau (log P) ne doit pas dépasser 5 [17].

En effet, tous les composés satisfaisant à ces règles, ont été sélectionnées pour la suite de notre investigation.

II.6 Etude de docking moléculaire

Pour analyser les interactions ligand-récepteur, nous avons soumis tous les composés retenus par l'analyse « *Drug-Likeness* » à une simulation du docking moléculaire, en utilisant le logiciel « FlexX » implémenté dans le programme LeadIT [8]. Comme structure du récepteur de départ, nous avons recherché des structures de la CK2 co-cristalisée avec un ligand dérivant de la famille des flavonoïdes dont la structure est similaire à celle de l'apigenine (AGI). En effet, la structure cristallographique codée 4DGM, correspondant à la protéine CK2 co-cristallisée avec son ligand AGI ayant une résolution égale à 1.65 Å, a été sélectionnée.

La protéine 4DGM, est composée d'une seule chaîne protéique contenant 326 résidus, et ayant un nombre total de 5435 d'atomes. Son site actif, a été identifié à une distance de 8 Å du centre de liaison du ligand (AGI), qui a été défini comme ligand de référence pour l'étude du docking moléculaire. Les résidus englobant le site sont : Val45, Met163, Val174, Lys68, Asp175, Val95, Ile66 et Val116.

FlexX, nous a permis de réaliser un docking semi-flexible (flexibilité totale pour le ligand, et une rigidité pour le récepteur), en appliquant les paramètres par défaut pour toutes les simulations de docking. En effet, dix conformations résultantes ont été générées pour chaque composé. Cependant, les conformations ayant les plus faibles énergies de docking correspondant aux conformations stables ou bioactives, ont été comparés au ligand de référence « AGI ». Seuls, les composés ayant des énergies de docking inférieures à celle du ligand de référence, ont été retenus pour la suite de notre étude.

II.7 Analyse *in silico* de la toxicité

Afin de prédire la toxicité des composés obtenus par la simulation du docking moléculaire, nous avons évalué leurs éventuels effets néfastes sur le corps, en utilisant le

serveur Web « ProTox-II » [9]. Ce dernier est considéré comme un laboratoire virtuel pour la prédiction de la toxicité de petits composés moléculaire. Il est bien de noter que, ProTox-II utilise des descripteurs moléculaires et des indices de similarité pour la prédiction de divers paramètres de toxicité tels que la toxicité aiguë, l'hépatotoxicité, la cancérogénicité… Il convient de noter que, la dose létale moyenne (DL50) est un indicateur quantitatif de la toxicité d'une substance à laquelle 50% des sujets testés meurent lors d'exposition.

Cependant, les résultats des calculs de LD50, obtenus pour les composés issus des simulations de docking moléculaire, sont comparés à ceux obtenus pour deux autres inhibiteurs potentiels de la CK2, à savoir : Silmitasertib (CX-4945) et Apigenin (AGI).

II.8 Simulation de dynamique moléculaire

Dans le but de vérifier la stabilité des composés retenus dans le site actif de la protéine CK2, nous avons réalisé des simulations de dynamique moléculaire (DM), sur le complexe (protéine CK2 + hit) et sur la structure cristallographique 4DGM (CK2 + AGI). Comme structures de départ, nous avons utilisé la structure du complexe (protéine CK2 + hit) issue du docking et la structure expérimentale dépourvue de molécules d'eau issues de la RX. Les deux types de systèmes (hit+CK2 et AGI+CK2) ont été immergés dans des boites cubiques de côté 7 Å, en ajoutant 9261 molécules d'eau. Ensuite, les deux systèmes sont neutralisés, en ajoutant un contre ion « Cl- », pour compenser la charge nette positive de la protéine. Nous avons fixé la concentration de NaCl à la valeur physiologique (0.15 mol/L) et protonés (PH = 7) [23] pour se mimer au mieux les caractéristiques expérimentales et physiologiques d'un système biologique.

Les simulations de dynamique moléculaire ont été réalisées en utilisant le logiciel Nano Molecular Dynamics software (NAMD version 2.12) [10] avec le champ de force CHARMM 36. Les paramètres topologiques des ligands (hit et AGI), ont été générées à l'aide du serveur Web CHARMM General Force Field (CGenFF).

Afin de parfaire encore le calcul des interactions entre les particules non-liées de courte et longue distance, nous avons utilisé l'algorithm *Particle Mesh Ewald* (PME) [24], basée sur la somme d'Ewald [25] avec une distance de troncature (***cut-off***, en Anglais) de 12 Å et une distance séparant les points constituant la grille de 1 Å.

Pour maintenir les propriétés biophysiques du système, nous avons appliqué des conditions aux limites (*Periodic Boundary Conditions* notée PBC) [26], qui consiste à copier le système étudié dans les trois directions de l'espace. Ainsi, une particule qui sortirait par un

coté de la boite arriverait directement par le côté opposé. En utilisant cette approche, le système simulé parait avoir des bornes infinies et garde une densité constante.

Afin de réguler les fluctuations de pression et de température, nous avons utilisé un thermostat (température constante) et barostat (pression constante) réglés à 1 atm et 310 K, respectivement, en effectuant une dynamique de Langevin, avec un coefficient d'amortissement égale à 1,0 picosecondes.

Les deux systèmes ont subi 100000 étapes de minimisation d'énergie en utilisant l'algorithme de minimisation steepest-descent, pour éliminer les mauvais contactes et relaxer les géométries des deux systèmes étudiés. Ensuite, les deux systèmes ont été équilibrés pour 50000 itérations en 100 picosecondes, en maintenant le nombre de particules, de pression et de température constant (NPT).

Enfin, les deux systèmes obtenus ont été soumis à 100 nanosecondes de dynamique, en utilisant un pas de temps égale à 2 femtoseconde. Les énergies de Van der Waals ont été calculées, en utilisant un rayon de switch de 10 Å et un rayon de troncature de 13,5 Å. L'analyse des résultats a été réalisé à l'aide du logiciel VMD [27] et le programme Ligplus [28].

III. Résultats et discussion

III.1 Génération du modèle pharmacophore 1 « *Structure-Based* »

Dans cette étape, nous avons utilisé la technique de modélisation du pharmacophore, pour identifier les régions favorables aux interactions CK2-Inhibiteurs. Cette méthode, prend en considération les structures 3D et les modes de liaisons d'inhibiteurs dans le site actif de la CK2.

La figure III-2-a, représente le premier modèle pharmacophore Structure-Based (modèle 1), que nous avons généré.

Comme indiqué dans le tableau III-3, et rapporté sur la figure III-2-a, le modèle pharmacophore généré pour la structure 3AMY (Modèle 1), est composé d'une combinaison de deux liaisons hydrogène donneur (HBD), situées sur les atomes d'oxygène des groupements hydroxyle en positions 3' et 7 (Tableau III-2). Une liaison hydrogène accepteur (HBA) sur l'atome d'oxygène du groupement hydroxyle en position 3'. Et une surface hydrophobe (Hy) située sur le cycle phényle C.

Il est à noter que, plusieurs modèles pharmacophore ont été développés pour les inhibiteurs de la CK2, et la plupart d'entre eux sont composés des mêmes types interactions et caractéristiques, à savoir : HBA, HBD et Hy [1-3].

III.2 Génération du modèle pharmacophore 2 « *ligand-Based* »

Le deuxième modèle pharmacophore Ligand-based (modèle 2), a été généré à partir des quatre ligands : Apigenine, lutéoline, FLC21 et FLC26, en utilisant le logiciel LigandScout [6].

Le modèle 2, se compose de deux liaisons HBA, une surface hydrophobe (Hy) et deux noyaux aromatiques (Ar) (Figure III-2-b). Cependant, la superposition des structures 2D des composés actifs sur le modèle pharmacophore 2, a révélé que : les deux liaisons HBA sont assurées par les groupements en positions 4 et 3', la surface hydrophobe positionnée sur le phényle C et les deux cycles aromatiques (Ar) sur les noyaux A et B, respectivement.

Des études antérieures ont reporté l'importance des cycles aromatiques « Ar », dans la conception des inhibiteurs de CK2 [2,29-31].

III.3 Génération du modèle pharmacophore 3

Toutes les caractéristiques des deux modèles 1 et 2 obtenues précédemment, ont été combinées et fusionnées dans un troisième modèle pharmacophore que nous avons noté modèle 3, cela afin d'améliorer les capacités et les caractéristiques pharmacophoriques prédites dans les deux modèles 1 et 2, pour la reconnaissance des composés actifs.

En effet, le modèle 3 (Figure III-2-c), est constitué de sept caractéristiques : une surface hydrophobe (Hy), deux cycles aromatiques (Ar), deux HBA et deux HBD. Il convient de noter que l'interaction HBD, assurée par le groupement en position 7, a été définie comme optionnelle.

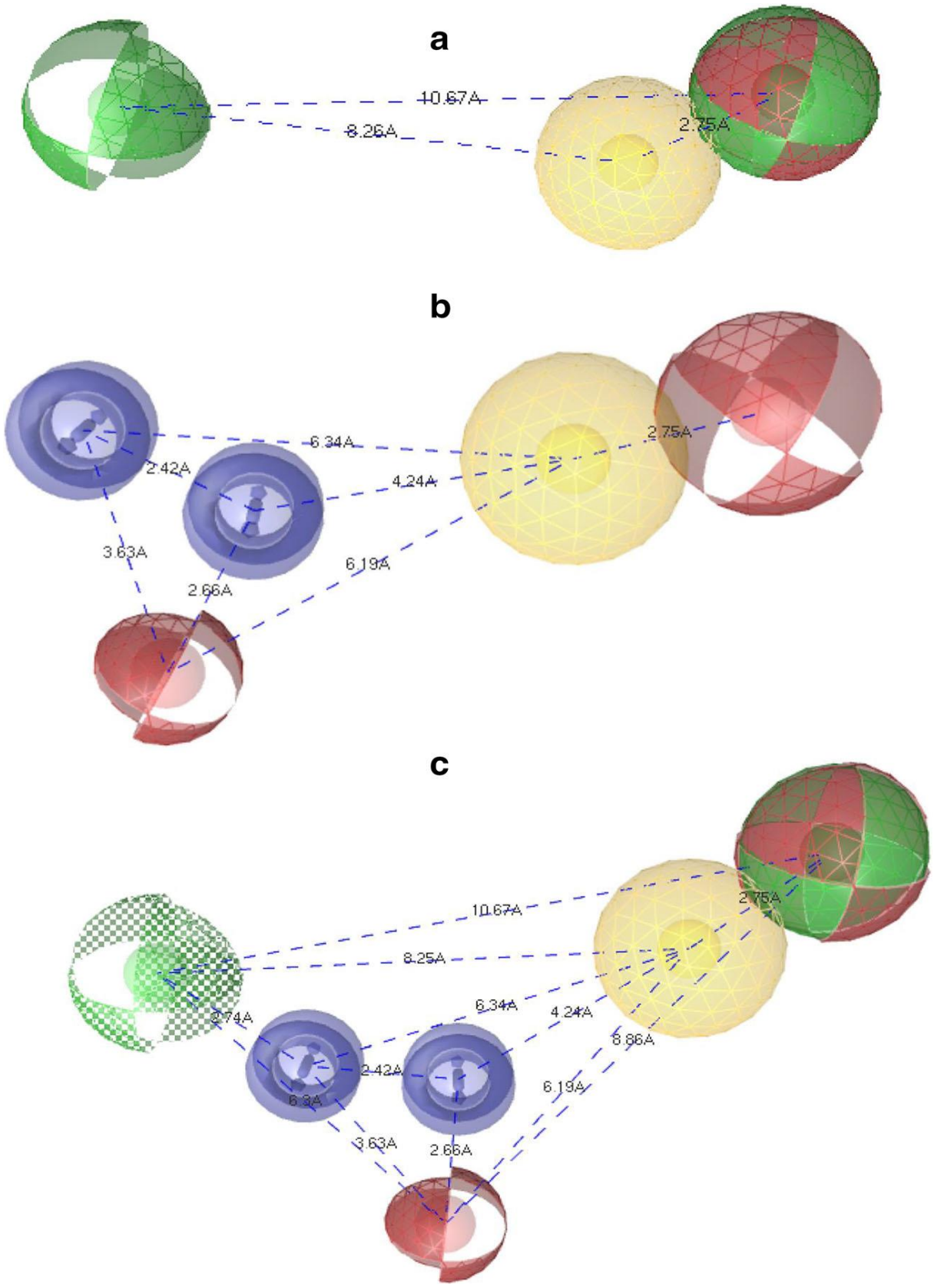


Figure III-2 : Représentation 3D des caractéristiques pharmacophoriques des trois modèles générés par LigandScout, a- Modèle 1, b- Modèle 2, et c- Modèle 3. Les interactions HBA, sont représentées en sphères Rouge, HY en sphères jaunes, HBD en vert, les noyaux aromatiques (Ar) en bleu et les longueurs des liaisons sont données en Å.

Tableau III-3 : position des caractéristiques pharmacophoriques des trois Modèles pharmacophores générés par notre étude (1,2 et 3).

Model	HBD		HBA		Hydrophobic		Ar
	Nber	Position	Nber	Position	Nber	Position	
Model 1	2	7, 3'	1	3'	1	Ring C	0
Model 2	0	-	2	4, 3'	1	Ring C	Ring A Ring B
Model 3	2	3', 7(Optional)	2	4, 3'	1	Ring C	Ring A Ring B

III.4 Validation des modèles pharmacophores

Pour évaluer la robustesse des trois modèles pharmacophoriques que nous avons développé, nous les avons soumis à plusieurs métriques de performance. Dans un premier temps la validation a été effectuée en utilisant un lot d'essai contenant 26 composés actifs sélectionnés de la littérature [11] et de 1000 composés non actifs « Decoys », collectés à partir de la base de données ZINC [12]. Par la suite les courbes de ROC (*Receiver Operating Characteristics*), permettant de visualiser les fractions de vrais positifs et de faux positifs, ont été générées pour chaque modèle pharmacophoriques. Les trois modèles sont évalués aussi sur la base de l'aire sous la courbe de ROC (*Area Under the Curve ou AUC)*, résumant la capacité de discrimination globale d'un modèle. Beaucoup d'autres métriques de validation ont été au cœur de ce travail notamment les facteurs d'enrichissement, la sensibilité, la spécificité, la précision, le nombre de faux positifs et le nombre de faux négatifs.

III.4.1 Cas du modèle 1 :

La courbe de ROC calculée pour le modèle 1 est représentée sur la figure III-3-a. L'analyse de cette figure montre que cette courbe tend vers la classification aléatoire (la diagonale tracée entre les points (0,0) et (1,1) du graphique). Cela signifie que le modèle 1 n'est capable de distinguer les vrais positifs des faux positifs.

D'un autre côté, les facteurs d'enrichissement calculés pour les fractions de la chimiothèque classée 1%, 5%, 10% et 100%, par rapport à une sélection aléatoire des composés, sont 7.9 ; 6.6 ; 6.6 et 6.6, respectivement. De plus, l'aire sous la courbe de ROC (*Area Under the Curve : AUC*), calculé pour ce modèle est égale à 0.53 (figure 3-a et tableau 4), une valeur proche de la classification aléatoire. De ce fait, les performances du modèle 1 sont mauvaises puisqu'il ne permet pas d'obtenir une discrimination entre les actifs et les

decoys. Il est bien de noter que plus la valeur de AUC est proche de 1 plus la classification réalisée est discriminante et plus cette valeur est proche ou inférieure à 0.5 plus les performances de ce modèle sont mauvaises.

En effet, parmi les 26 composés actifs, le modèle 1 a pu distinguer seulement deux composés comme actifs (vrai positif, VP). Tandis que le reste de composés actifs (24), sont prédits comme inactifs (faux négatifs, FN). Par conséquent les classifications obtenues sont comme suit :

- ✓ La sensibilité obtenue pour ce modèle, déterminant la fraction des vrais positifs (FVP), est faible du moment que sa valeur avoisine les 8%.
- ✓ La spécificité calculée pour ce modèle, déterminant la fraction des vrais négatifs (FVN), est de 99% sachant que le nombre de faux positifs est de 10. Il est bien de noter que ce pourcentage est en rapport avec le nombre de composés inactifs utilisés dans la validation (1000 composés)
- ✓ La précision, qui représente la part des vrais positifs parmi tous les composés hits, est de 17%, ce qui est très insuffisant.

III.4.2 Cas du Modèle 2

La courbe de ROC calculée pour le modèle 2 est représentée sur la figure III-3-b. L'analyse des résultats montre que cette courbe atteint directement le point (0,1) avant de rejoindre le point (1,1) du graphique. Cela signifie que le modèle 2 permet de discriminer parfaitement les ligands actifs des inactifs (decoys).

D'un autre côté, les facteurs d'enrichissement calculés pour les fractions de la chimiothèque classée 1%, 5%, 10% et 100%, par rapport à une sélection aléatoire des composés, sont 39.5 ; 33.1 ; 33.1 et 33.1, respectivement. De plus, l'aire sous la courbe de ROC (AUC : Area Under the Curve), calculé pour ce modèle est égale à 1.00 (figure III-3-b et tableau 4), une valeur qui caractérise une discrimination parfaite.

L'analyse visuelle des résultats obtenus pour le modèle 2, montre que ce modèle a pu distinguer tous les 26 composés actifs (VP). En effet, les calculs statistiques obtenues sont :

- ✓ La sensibilité, est de 100%.
- ✓ La spécificité, est de 99,5%, sachant que le nombre de faux positifs est de 5.
- ✓ La précision de ce modèle est de 84%.

L'analyse des résultats de validation des deux modèles 1 et 2, a montré que quatre Faux positifs (FP) sélectionnés par le modèle 2, ne sont pas prédits par le modèle 1.

III.4.3 Cas du Modèle 3

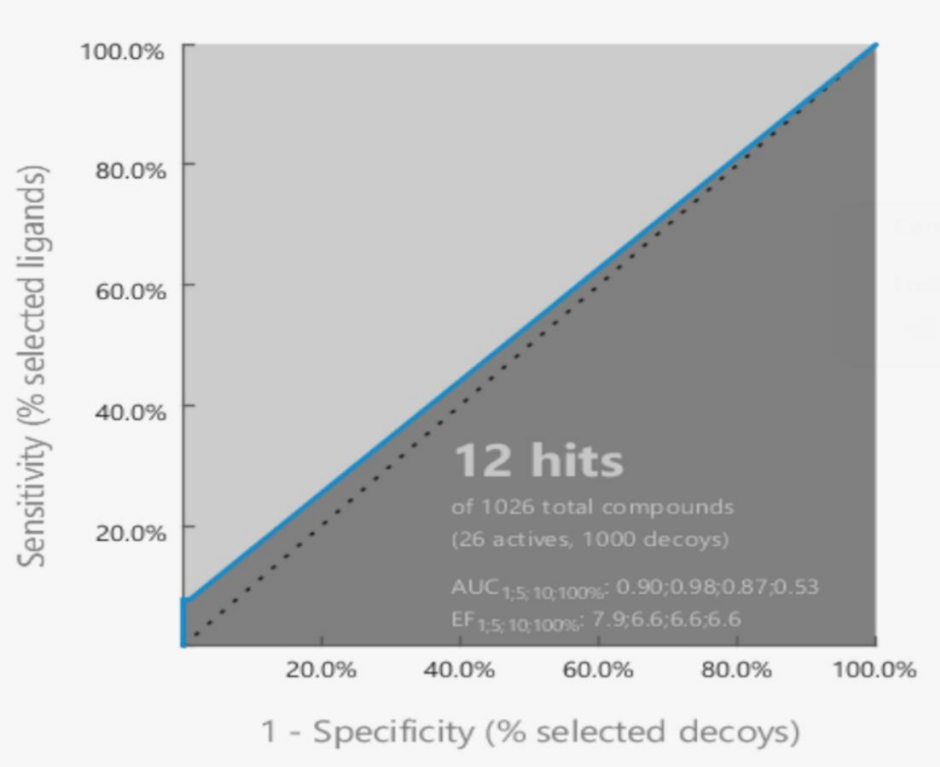
La combinaison des caractéristiques des deux modèles, 1 et 2, nous a permis d'obtenir un autre modèle, noté modèle 3. En effet, la courbe de ROC calculée pour ce modèle est représentée sur la figure III-3-c. La forme de cette courbe ressemble à celle obtenue pour le modèle 2, c'est-à-dire qu'elle atteint directement le point (0,1) avant de rejoindre le point (1,1) du graphique. Cela signifie que, le modèle 3, tout comme le modèle 2, permet de discriminer parfaitement les ligands actifs des inactifs (decoys).

D'un autre côté, nous observons une amélioration dans les facteurs d'enrichissement calculés pour les fractions de la chimiothèque classée 1%, 5%, 10% et 100%, par rapport à une sélection aléatoire des composés, qui sont égales à 39,5, 38,0, 38,0 et 38,0, respectivement. De plus, l'aire sous la courbe de ROC (*Area Under the Curve* : AUC), calculé pour ce modèle est égale à 1.00 (figure 3-c et tableau 4), une valeur qui caractérise une discrimination parfaite.

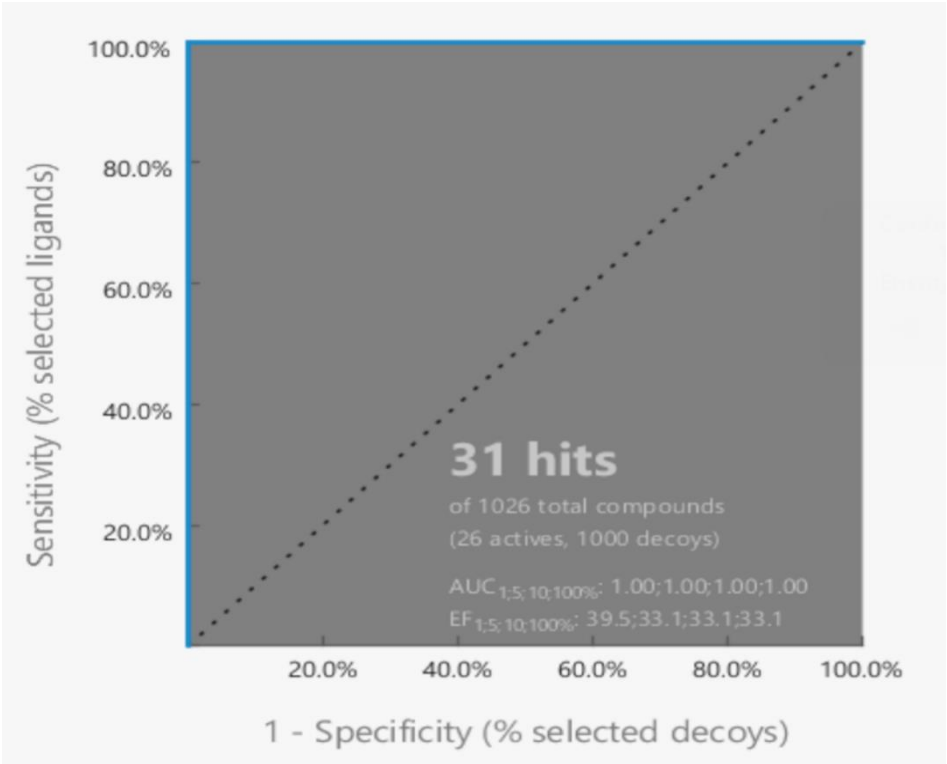
Le modèle 3 a pu distinguer tous les 26 composés actifs (VP). Par conséquent, nous avons obtenu les valeurs de classifications suivantes :

- ✓ La sensibilité calculée pour ce modèle est supérieure à 99%.
- ✓ La spécificité calculée pour le modèle 3 est de 99,9%, sachant qu'un seul composé faux positif (FP) a été détecté.
- ✓ La précision du modèle 3 est supérieur à 96%.

A


100.0%
80.0%
60.0%
40.0%
20.0%
Sensitivity (% selected ligands)
12 hits
of 1026 total compounds
(26 actives, 1000 decoys)
AUC1;5;10;100%: 0.90;0.98;0.87;0.53
EF1;5;10;100%: 7.9;6.6;6.6;6.6
20.0%
40.0%
60.0%
80.0%
100.0%
1 - Specificity (% selected decoys)

B


100.0%
80.0%
60.0%
40.0%
20.0%
Sensitivity (% selected ligands)
31 hits
of 1026 total compounds
(26 actives, 1000 decoys)
AUC1;5;10;100%: 1.00;1.00;1.00;1.00
EF1;5;10;100%: 39.5;33.1;33.1;33.1
20.0%
40.0%
60.0%
80.0%
100.0%
1 - Specificity (% selected decoys)

C

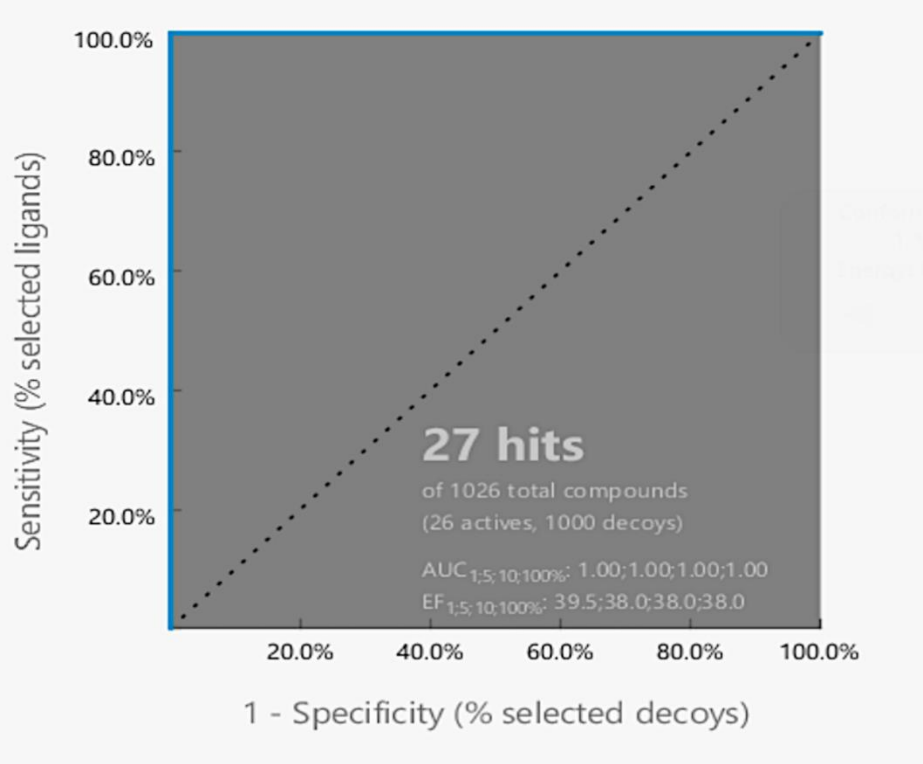


Figure III-3 : Représentation des courbes ROC, obtenues par le criblage de 26 composés actifs et 1000 decoys, en utilisant les trois modèles : A- modèle 1, B- modèle 2, C- modèle 3. La courbe en bleu représente la courbe de ROC obtenue pour chaque modèle, celle en pointillée noir est la classification aléatoire

Tableau III-4 : Paramètres de validation calculés à l'aide de l'échantillon de test constitué de 26 composés actifs et de 1000 decoys, pour les trois modèles pharmacophore générés (1, 2 et 3)

Model	TP	TN	FP	Précision	Sensitivité	Spécificité	AUC	EF
Model 1	2	990	10	0.17	0.080	0.990	0.530	6.600
Model 2	26	995	5	0.84	1.000	0.995	1.000	33.100
Model 3	26	999	1	0.96	1.000	0.999	1.000	38.100

L'analyse des résultats de validation des trois modèles pharmacophoriques (1, 2 et 3), montre bien que le « modèle 3 » a plus de possibilités de distinguer un composé actif dans la base de données avec moins de Faux Positifs. Par conséquent, ce dernier sera utilisé dans la prochaine étude de criblage virtuel de la base de données Pubchem Drug Bank.

III.5 Criblage Virtuel

Après avoir obtenu un pharmacophore suffisamment discriminant et sélectif de la protéine CK2, ce modèle peut être utilisé pour la recherche de nouveaux hits par le criblage de chimiothèques. Ainsi, le modèle 3 a été utilisé pour cribler 737 920 composés issus de la base

de données Pubchem Drug Bank , vu son habilité à distinguer les ligands actifs des inactifs (Decoys), ainsi que sa capacité de récupérer le plus grand nombre de composés actifs (VP) avec moins de faux positifs (FP).

L'organigramme systématique suivi pendant les simulations du criblage virtuel est représenté sur la figure III-4. Il est à noter qu'à la fin de la première étape, la comparaison du nombre et du type des points pharmacophoriques entre chaque molécule et le pharmacophore a permet de sélectionner 5984 composés avec une valeur de pharmacophore-fit supérieur à 57,0.

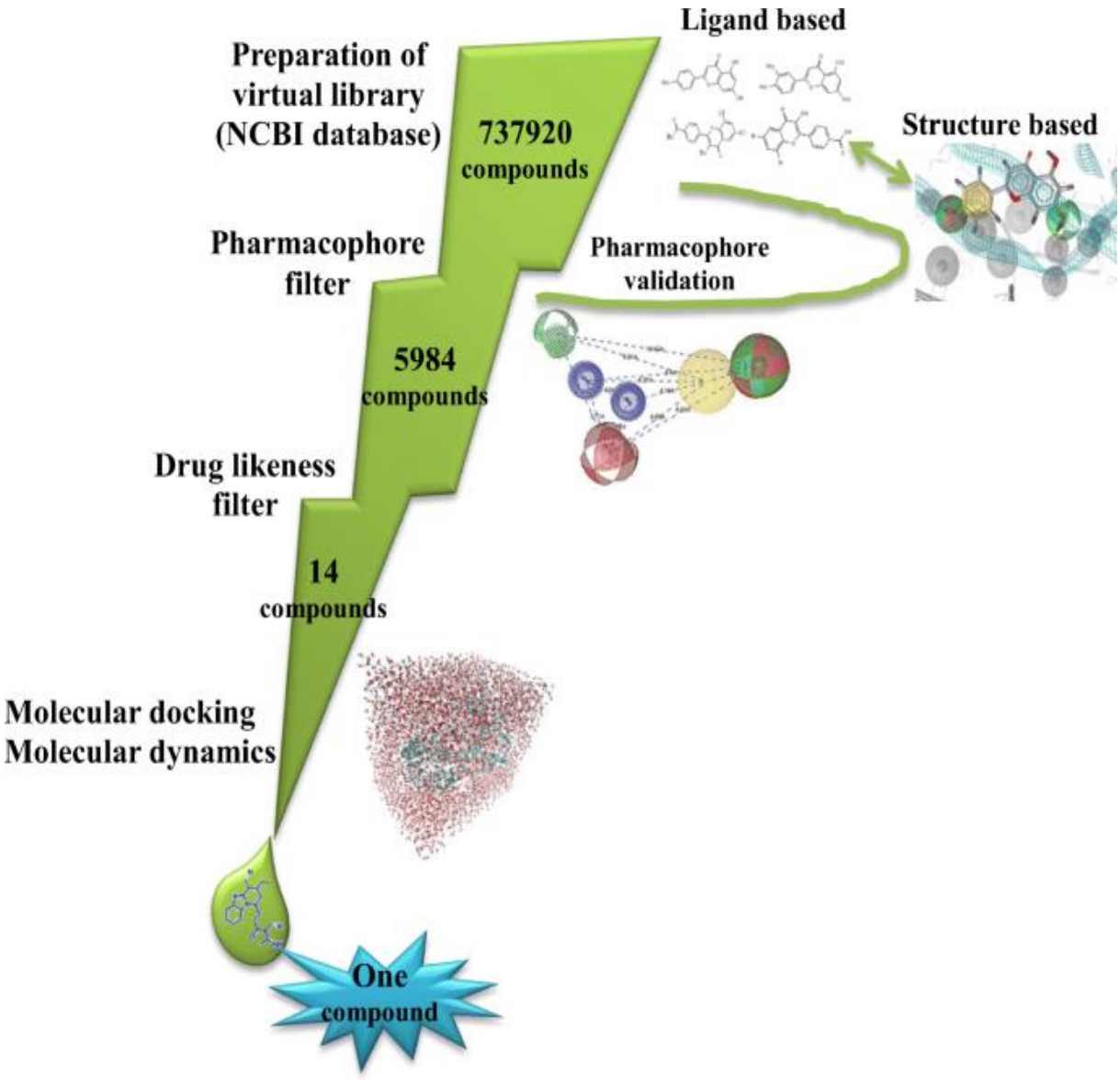


Figure III-4 : Organigramme systématique du criblage virtuel utilisée dans notre étude.

III.6 Analyse Drug-likeness

Dans le but de filtrer encore plus les 5984 hits obtenus dans l'étape de criblage virtuel et d'éliminer toutes les molécules possédant des propriétés incompatibles avec celles de composés d'intérêt pharmaceutique (*drug-like*), nous les avons soumis à un filtrage pharmacocinétique, chimique et de profil de médicament en utilisant le programme DruLiTO [7]. En effet, pour qu'une molécule devient un médicament, elle doit satisfaire certaines règles, notamment la règle de 5 de Lipinski [17], Veber [18], Ghose filter [19], QED (*quantitative estimate of drug-likeness*) [20], BBB likeness (*blood-brain barrier*) [21], CMC-50-like et MDDR-like, (voir l'annexe 02).

Le but de cette étape est d'éliminer toute molécule potentiellement toxique et celles comportant des groupements jugés trop réactifs.

Au terme de cette étape, 14 composés, appartenant aux familles chimiques des flavonoïdes et des benzimidazole, satisfont l'ensembles des critères constituant les différentes règles utilisées. Ces composés ont été retenus pour la suite de notre investigation à savoir les études de docking moléculaire.

III.7 Docking moléculaire

Dans le but de prévoir la capacité ou non des molécules obtenues, suite aux filtres appliqués à notre chimiothèque à se lier au site actif de la protéine CK2, nous les avons soumises aux études de docking moléculaire.

II.7.1 Génération des paramètres du docking

Avant de docker les molécules étudiées dans le site actif de la CK2, il est nécessaire de tester les performances de prédiction du modèle de docking que nous avons développé avec le logiciel FlexX [8]. Aussi, la structure du ligand de référence Apigenine (AGI), extraite du complexe cristallographique 4DGM a été redockée dans le site actif de la CK2.

La figure III-5 montre la superposition de la structure cristallographique et la conformation du ligand AGI issue de Docking. Ainsi, la conformation dockée de l'AGI est très proche de celle de la structure expérimentale, puisque la déviation quadratique moyenne (*root mean square deviation* ou RMSD) obtenue entre les deux conformations n'était que de 0,75.

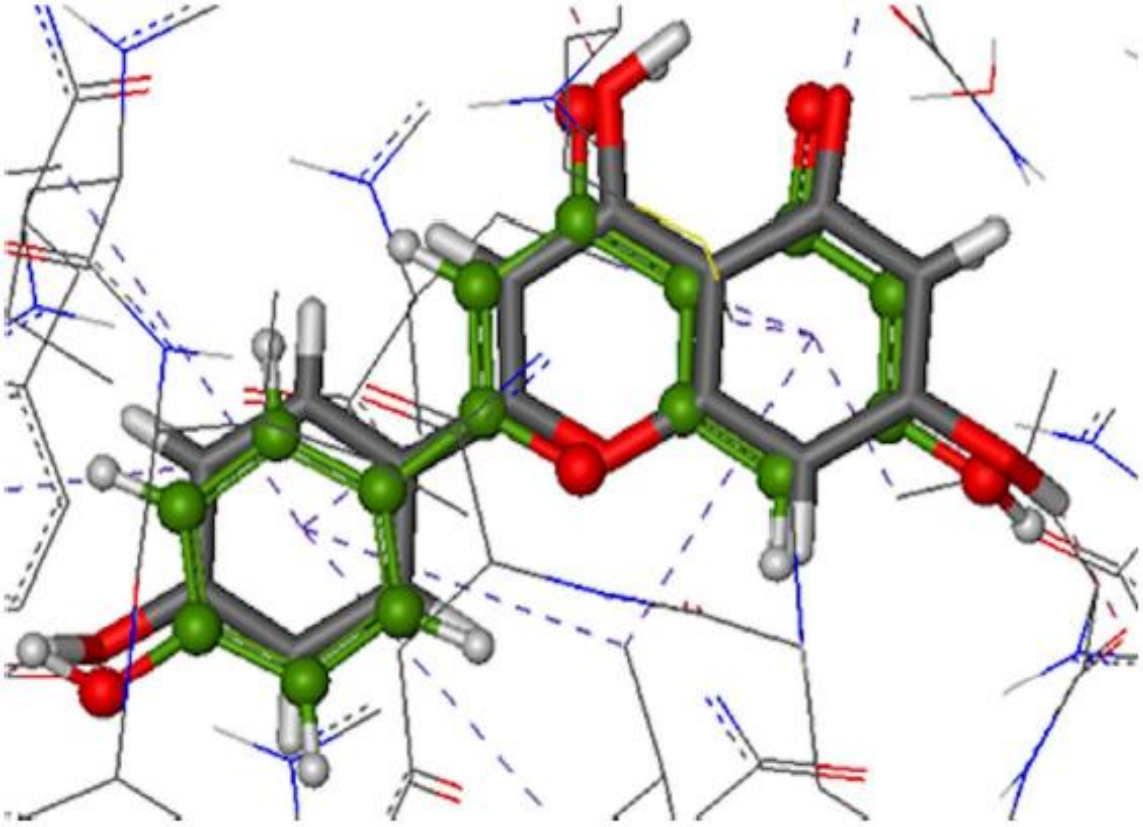


Figure III-5 : Comparaison entre la structure cristallographique X-ray (représentée en boules et tiges vertes) et la structure issue du docking de ligand « AGI » (représentée en tiges grises)

III.7.2 Validation du modèle de Docking

Pour confirmer l'efficacité du modèle de docking moléculaire, nous avons procédé au calcul de la corrélation entre les énergies de liaisons prédites, obtenues par le programme FlexX pour les 26 composés de l'ensemble de test (tableau 2) et les activités expérimentales correspondantes (valeurs pIC50 = - log (IC50)).

En effet, d'après la figure III-6, le coefficient de corrélation obtenu est R ≈ 0,84). En se basant sur les résultats, ainsi que les résultats des études de docking moléculaire précédentes [32], nous pouvons confirmer que le modèle de docking moléculaire proposé, pourrait décrire de manière satisfaisante les interactions pouvant avoir lieu entre les inhibiteurs et les acides aminés du site actif de la CK2.

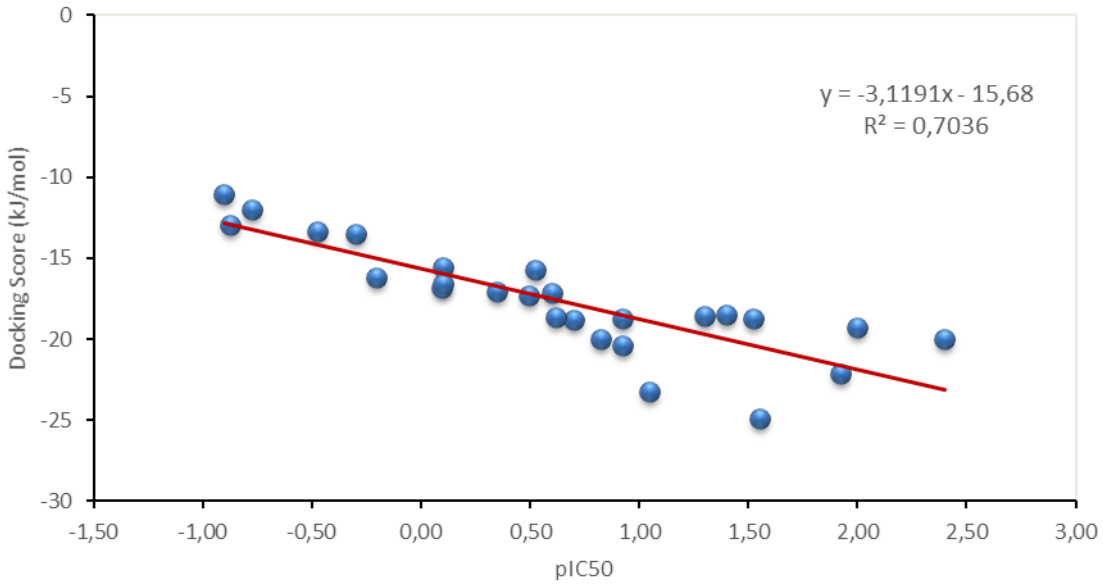


Figure III-6 : Représentation graphique des énergies prédites par le docking des composés de l'ensemble de test, en fonction des valeurs expérimentales (pIC50).

III.7.3 Mode de liaison des composés criblés

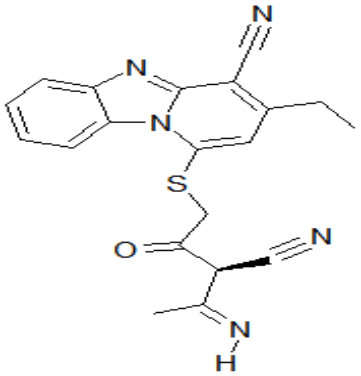
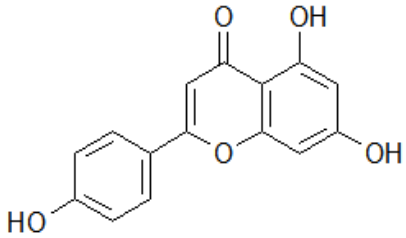
Pour sélectionner les composés potentiellement actifs (*drug-like*), les 14 molécules obtenues ont été dockées dans le site actif de la protéine CK2, en utilisant le logiciel LeadIT [8]. En effet, le mode de liaison prédit, les énergies de docking moléculaire et les interactions des molécules hits ont été comparés à celles du ligand de référence (AGI).

L'énergie de docking calculée pour le ligand de référence AGI est égale à -29,1 kJ/mol. Ainsi, tous les composés dont les énergies de docking inférieures à -30 kJ/mol ont été sélectionnés et leurs interactions avec les résidus du site actif ont été analysées.

L'analyse des résultats a révélé qu'un seul composé, n'appartenant pas à la famille des flavonoïdes, présente une énergie de docking inférieure à celle du ligand AGI. Ce composé (hit), ayant un PubChem CID : 7250636, correspond à la structure chimique $C_{20}H_{17}N_5OS$. Son nom chimique selon la nomenclature IUPAC est : 1-[(3S) -3-cyano-4-imino-2-oxopentyl] sulfanyl-3-ethylpyrido[1,2-a] benzimidazole-4-carbonitrile.

En effet, le composé hit, est un dérivé de pyrido[1,2-a] benzimidazoles, contenant une fonction nitrile et un groupement éthyle en positions 3 et 4 du cycle pyridine, respectivement, et un groupement 3-cyano-4-imino-2-oxopentylsulfanyl en position 6, comme illustré dans le Tableau III-5.

Tableau III-5 : Structure chimique, propriétés physicochimiques (drug-likeness), pharmacophore fit (fit score) et les énergies du docking (Docking Score) obtenus pour le composé hit, ainsi que pour l'AGI

Composé	Structure	HBD	TPSA	logP	HBA	MW	Fit Score	Docking Score
Hit		1	129.40	1.43	6	375.12	66.71	-36.41
AGI		4	94.06	1.39	0	271.06	58.51	-29.10

Il convient de noter que, l'analyse de la base de données ChEMBL (https ://www.ebi.ac.uk/chembl/), a montré que ce composé n'a jamais été étudié (exploré) pour l'inhibition de la CK2. De plus, les chercheurs du monde entier accordent une plus grande attention aux dérivés de benzimidazole substituées, en raison de leurs activités biologiques intéressantes vis-à-vis de nombreuses maladies [33].

En se basant sur l'énergie de docking calculée pour le composé criblé hit, qui est de – 36,03 kJ/mol, nous avons estimé que ce composé prédit par notre étude, montre une affinité plus élevée pour la protéine CK2, que celle du ligand expérimental « AGI ». Par conséquent, ce composé a été sélectionné pour la suite de notre investigation.

La conformation du composé prédit avec la meilleure énergie de docking, est illustrée avec l'AGI sur la Figure III-7. Ce composé, se lie à la protéine CK2 dans le site de fixation d'ATP (Adénosine Triphosphate), de ce fait il a une disposition similaire à celle du ligand expérimental.

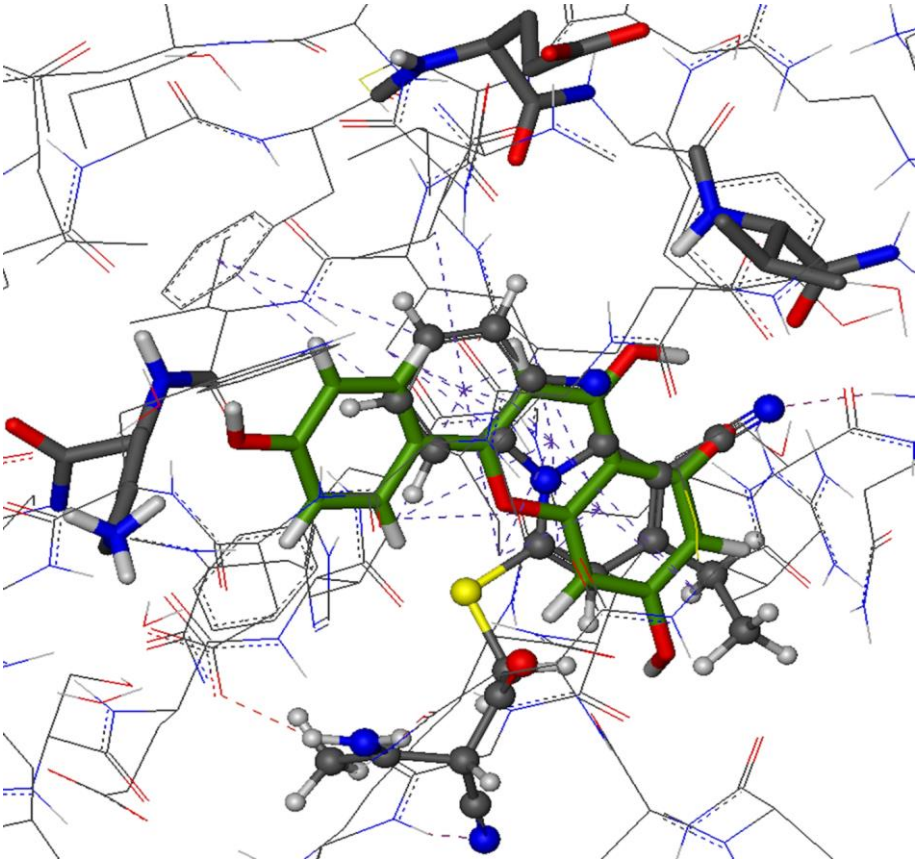


Figure III-7 : Superposition de la conformation dockée du composé prédit (boules et bâtonnets grises) et de la structure cristallographique de l'AGI (bâtonnets vertes). Les acides aminés Val116, Glu114 et Lys68 du site actif de la CK2, sont représentés en bâtonnets grises.

La figure III-8 illustre les interactions entre le composé prédit, hit, et la protéine CK2. En effet, quatre résidus clés sont impliqués dans des interactions de type liaisons hydrogène dont : l'arginine (Arg47), l'acide aspartique (Asp175), deux asparagines (Asn161 et Asn118). De plus, huit acides aminés sont impliqués dans des interactions hydrophobes dont deux isoleucines (Ile66, Ile174), deux valines (Val45, Val53), phénylalanine (Phe113), glycine (Gly46), méthionine (Met163) et l'asparagine (Asn118).

Il convient de noter que, des études antérieures théoriques et expérimentales sur la recherche d'inhibiteurs de la CK2, ont également montrés que la plupart des résidus d'acides aminés cités ci-dessus, contribuent à la formation de liaisons hydrogène et aux interactions hydrophobes entre la CK2 et ses inhibiteurs potentiels [34-36].

En effet, Le groupement nitrile (–CN) en position *meta* du cycle pyridine du hit, forme deux liaisons hydrogène avec l'amine terminale du backbone de l'Asn118 et la molécule d'eau HOH581. Le groupement protoné imine (C=NH2+) du groupement « oxo pentylsulfanyl » est engagé dans deux liaisons hydrogènes avec les résidus Asn161 et

Asp175. De plus, le groupement fonctionnel cyano (–CN) du groupement « oxo pentylsulfanyl » forme une liaison hydrogène avec le résidu Arg47.

Ajouter à cela, l'atome de soufre du résidu Met163, est situé à une distance de 3.47 Å du cycle pyridine du composé hit. Ceci, cette disposition 3D peut conduire à des interactions favorables du type « π-soufre ». D'autre part, le cycle « phényle » du composé hit est entouré par les résidus Val53, Phe113, Ile66, et Ile174, conduisant à la formation d'interactions hydrophobes favorables.

En effet, les interactions hydrophobes formées avec le résidu Phe113, jouent un rôle important dans la fixation des inhibiteurs dans le site de liaison d'ATP [3,37]. Ajouter à cela, il a été démontré que l'Arg47 et Asp175, sont aussi très importants pour la fixation, dans le site actif, de certains inhibiteurs de la protéine CK2 [38].

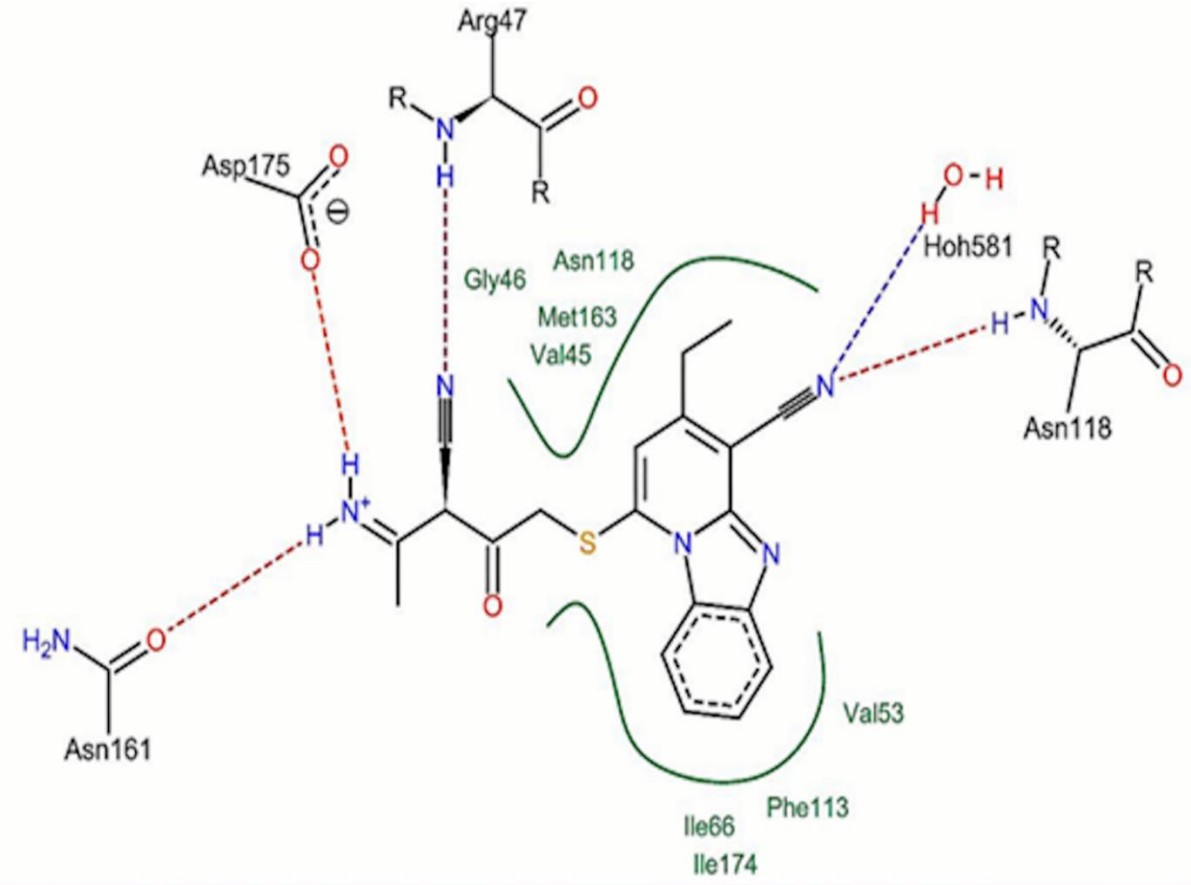


Figure III-8 : Vue bidimensionnelle du mode de liaison du composé prédit hit dans le site actif de la CK2.

La superposition du hit sur le meilleur modèle pharmacophore (modèle 3) montre une parfaite adéquation entre la position des caractéristiques pharmacophoriques et les groupements fonctionnels de ce composé (Figure III-9).

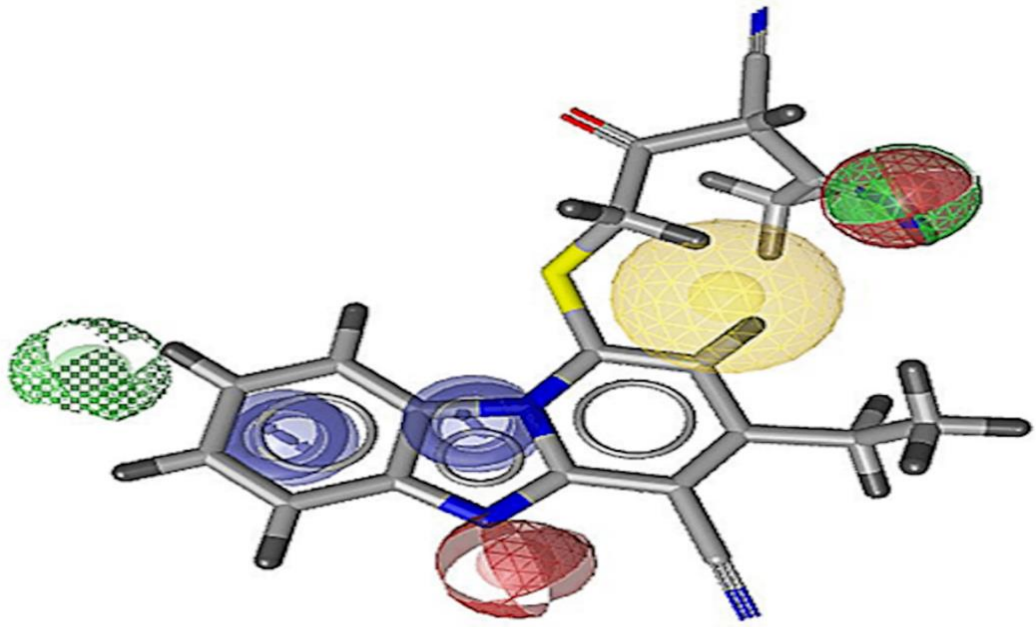


Figure III-9 : Représentation tridimensionnelle de la superposition de la molécule prédite sur le modèle pharmacophore (modèle 3).

La figure III-9 montre que le composé hit correspond parfaitement à toutes les caractéristiques du pharmacophore 3, à savoir : deux HBA (en position 4 et 3'), une HBD (3'), une HY (phényle cycle C) et deux caractéristiques Ar, à l'exception de la HBD (en position 7) qui a été définie initialement comme caractéristique facultative.

Il a été démontré qu'un modèle pharmacophorique d'inhibiteurs de la CK2, doit contenir des groupements appropriés pour former des liaisons hydrogène, entre les chaînes latérales des résidus du site actif de la CK2, correspondant au site de liaison de l'ATP, et l'inhibiteur [39].

III.8 Analyse *in silico* de la toxicité

La toxicité d'un composé drug-like est l'un des paramètres les plus redoutés par les développeurs de médicaments. En effet, un composé drug-like présentant des effets indésirables graves lors des essais cliniques sera immédiatement abandonné de manière définitive. Il est donc important de prédire sa toxicité à partir de sa structure lors de la phase de recherche et développement. Dans ce contexte, nous avons réalisé une analyse *in silico* de la toxicité à l'aide du serveur Web ProTox-II [9], afin de démontrer l'absence d'effets secondaires (toxiques) dangereux du composé prédit par notre étude.

En effet, les valeurs de la toxicité calculée en termes de la dose létale moyenne (DL50) ainsi que leurs classifications, sont résumées dans le tableau III-6.

Tableau III-6 : Résultats de toxicité (DL50) des composés étudiés et classifications correspondantes.

Composé	DL50	Classe de toxicité
Hit	900 mg/kg	4
CX-4945	729 mg/kg	4
AGI	2500 mg/kg	5

Il convient de noter que plus la valeur DL50 est élevée, plus la toxicité est faible. En revanche, aucun événement indésirable lié à la toxicité du composé prédit par notre étude ou à l'un de ses dérivés, n'a été rapporté à ce jour dans la littérature scientifique.

III.9 Simulation de Dynamique Moléculaire

La dynamique moléculaire permet d'identifier les régions de la protéine influencées par la présence du ligand. Ces méthodes considèrent explicitement l'ensemble du système protéine-ligand comme flexible et peuvent donc être utilisées pour la validation des poses obtenues par le docking. A cet effet, nous avons réalisé des simulations de dynamique moléculaire, pour confirmer le mode de liaison prédit par nos simulations de docking pour le composé hit dans le site actif de la CK2 et comparer les résultats ainsi obtenus à ceux du ligand AGI dans la structure cristallographique 4DGM. Ainsi, nous avons évalué la stabilité des systèmes par : les variations des valeurs de RMSD, le nombre de liaisons hydrogène entre les ligands et la protéine, la fluctuation quadratique moyenne (RMSF), les changements des surfaces accessibles aux solvants (SASA) ainsi que le rayon de giration (rGyr).

La figure III-10 montre la superposition de la conformation issue du docking du complexe (hit-CK2) avant la dynamique moléculaire (DM) et celle après les 100 ns de DM. L'analyse des résultats montre que le ligand hit a conservé le même site de liaison (avant et après les simulations de dynamique moléculaire). De plus, les liaisons hydrogène avec les résidus Arg47 et His160, ainsi que les interactions hydrophobes avec les résidus Gly46, Val53, Ile174, Val45, Met163, Asn118 et Val116, sont également conservées.

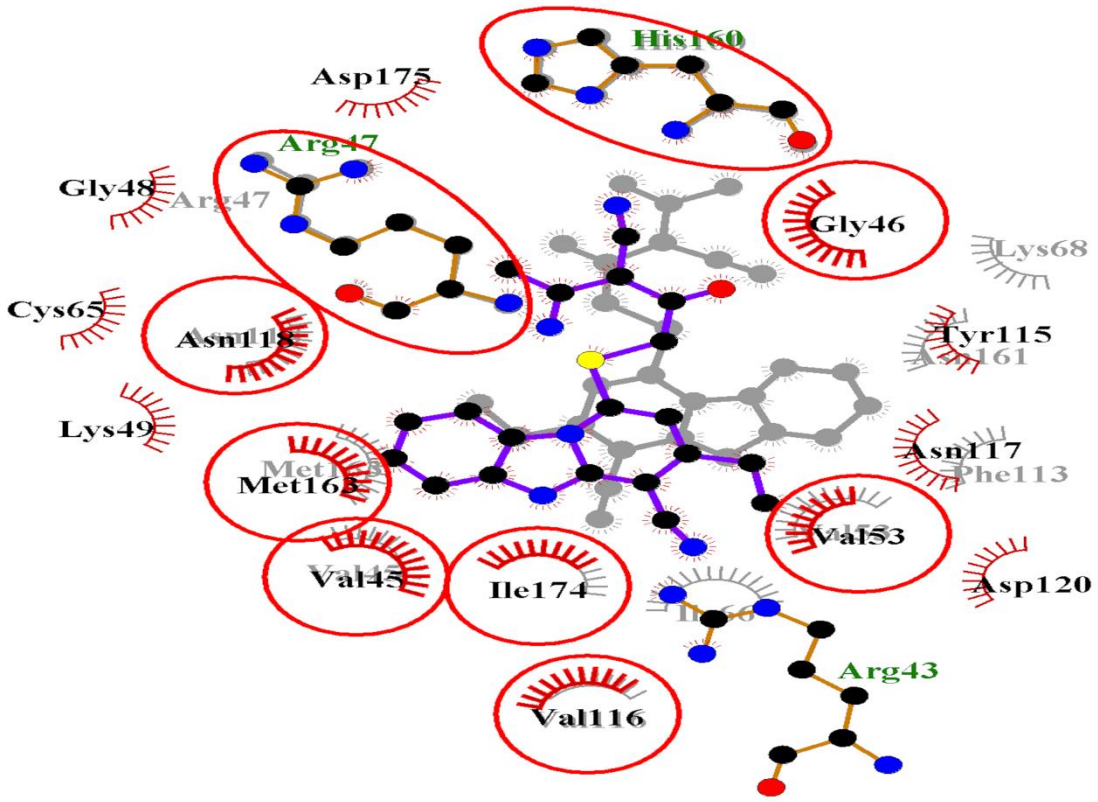


Figure III-10 : Superposition de la conformation issue du docking (boules et batonnêts grises) et celle obtenue après 100 ns de dynamique molécualire (boules et batonnêts colorés selon la nature des atomes) du complexe prédit (hit-CK2).

III.9.1 Analyse des liaisons Hydrogène

Le nombre de liaisons Hydrogène formées entre les ligands hit, AGI et la protéine CK2 durant les 100 ns de dynamique moléculaire, sont représentées sur les figures III-11 et III-12 respectivement.

L'analyse des résultats, montre que le ligand hit a formé 3 à 5 liaisons hydrogène en moyenne, avec les résidus du site actif de la protéine CK2 durant les 100 ns de DM. Ce résultat confirme la forte inhibition de la protéine CK2, par le composé criblé appuyant ainsi les résultats déjà obtenus par les simulations de docking, ayant prédit quatre liaisons hydrogènes avec les résidus du site actif de la CK2.

D'autre part, le ligand AGI forme 2 à 4 liaisons hydrogène en moyenne avec les résidus du site actif de la CK2, le long des 100 ns de DM.

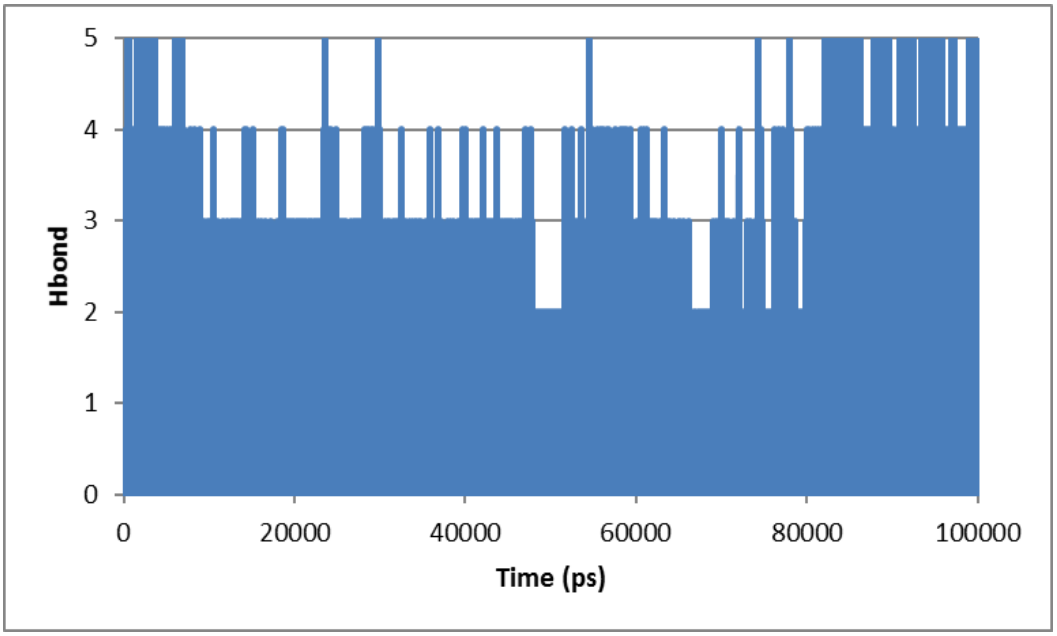


Figure III-11 : Nombre de liaisons hydrogène entre le ligand « hit » et la protéine CK2.

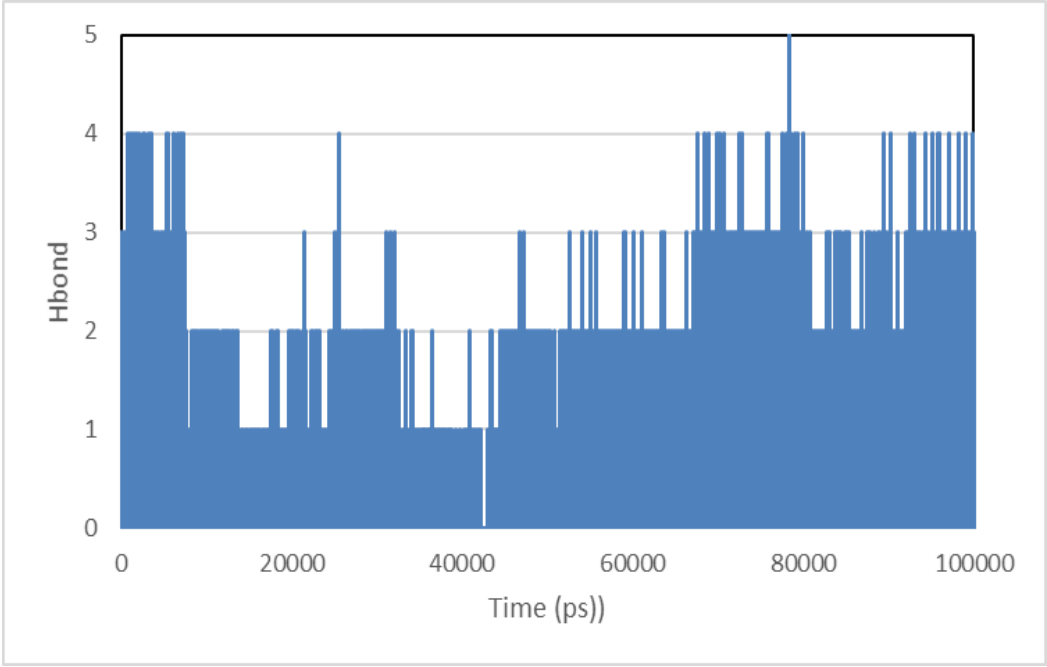


Figure III-12 : Nombre de liaisons hydrogène entre le ligand AGI et la protéine CK2.

III.9.2 Analyse RMSD

Les diagrammes RMSD calculés pour les atomes Cα du backbone de la protéine CK2, pour les complexes, hit-CK2 et cristallographique AGI-CK2, le long des 100 ns de la dynamique moléculaire, sont reportés sur la figure 13.

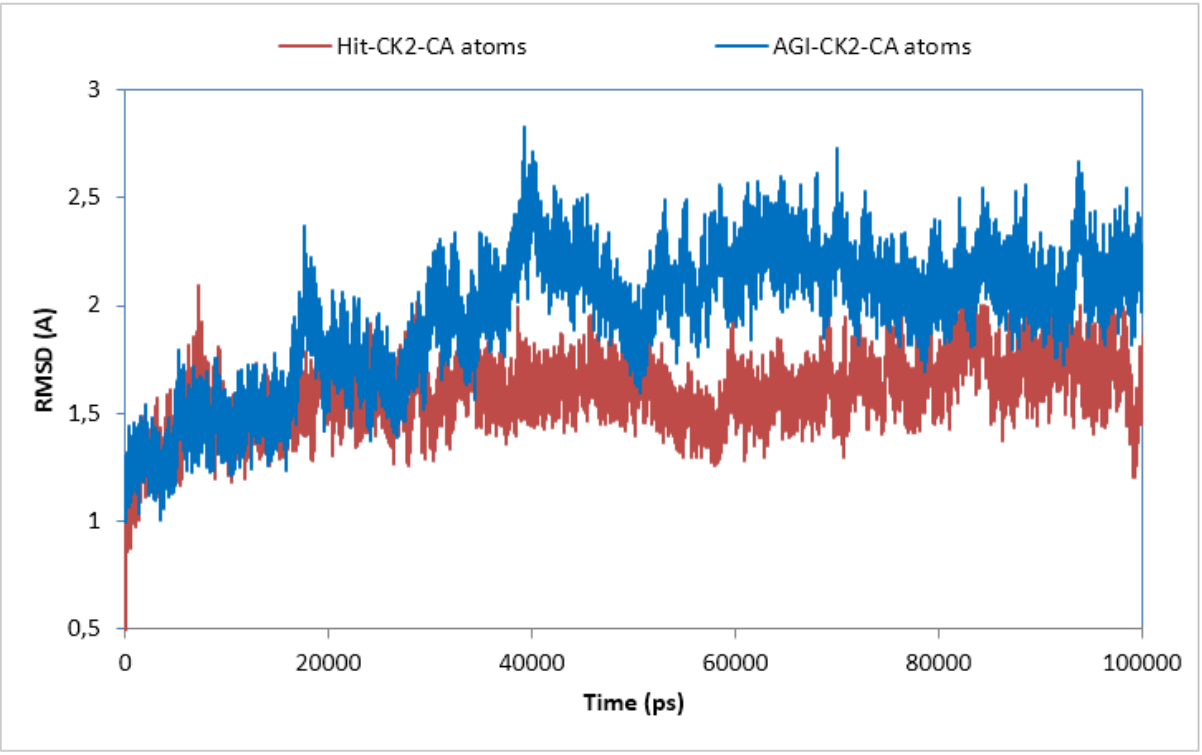


Figure III-13 : RMSD des atomes Cα du backbone des complexes hit – CK2 (rouge) et AGI – CK2 (bleu).

Le complexe hit-CK2 a atteint la stabilité, autour de 10 ns. Au-delà de ce point aucune déviation significative n'a été observée, pour les atomes Cα jusqu'à la fin de la simulation. La valeur de RMSD résultante varie entre 1,4 Å et 1,6 Å.

Cela signifie que, la structure du complexe hit-CK2, adopte une conformation plus stable que celle de départ. En revanche, certaines fluctuations ont été observées pour le complexe cristallographique « AGI-CK2 » jusqu'à 80 ns, où le système commence à se stabiliser autour d'une valeur de RMSD de 2 Å.

La figure III-14, illustre les valeurs de RMSD calculées pour les ligands hit et AGI dans les deux complexes étudiés, par rapport à la conformation de référence de chaque système, au cours du temps de simulation. On voit bien, qu'il existe une différence entre les deux courbes, ce qui indique la présence de mouvements modérés des deux ligands dans le site actif de la protéine CK2.

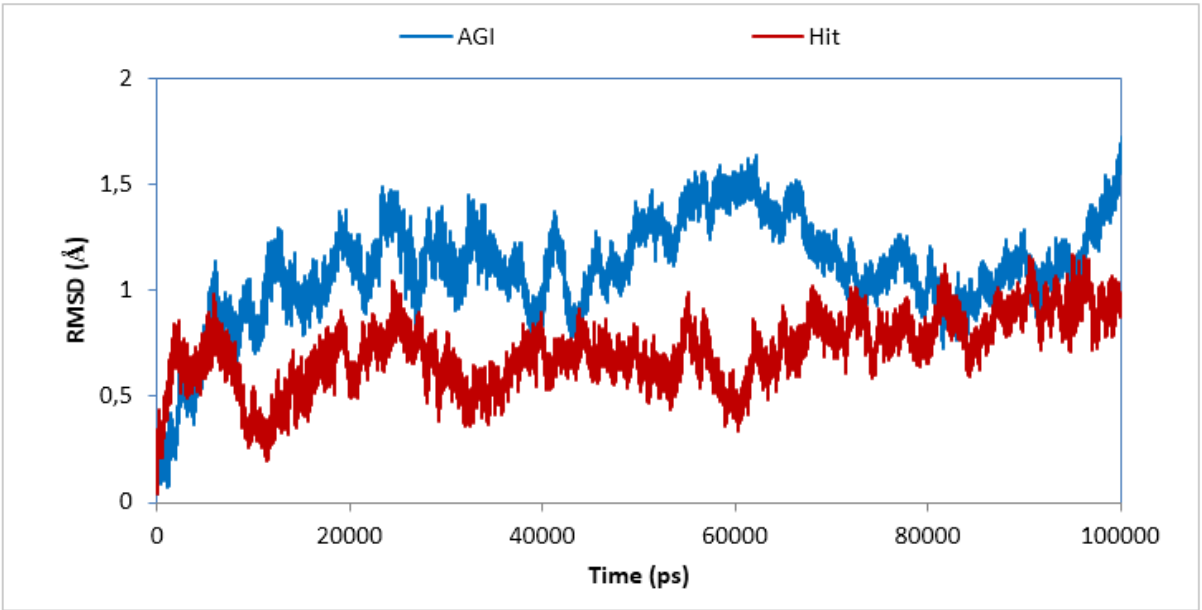


Figure III-14 : RMSD du ligand prédit « hit » (rouge) et expérimental « AGI » (bleu) durant les 100 ns de dynamique moléculaire.

Cette figure montre également, que le ligand prédit « hit » est caractérisé par de faibles fluctuations de l'ordre de 0,5 ± 0,3 Å, jusqu'à 70 ns. De ce point jusqu'à la fin du temps de simulation, le système a atteint une stabilité à une valeur de RMSD avoisinant de 0,8 ± 0,1 Å. Ce résultat montre que le ligand hit oscille entre plusieurs conformations dans le site actif de la CK2 jusqu'au 70 ns, après lesquelles il commence à se stabiliser.

En revanche, les valeurs de RMSD du ligand expérimental AGI étaient plutôt élevées, conduisant à des changements conformationnels et des écarts par rapport à la conformation de référence. Ceci se traduit par des fluctuations importantes de RMSD, dans les 40 premières nanosecondes (RMSD ≈ 1,1 ± 0,4 Å). En effet, à partir de 70 ns le système a atteint une stabilité autour de RMSD ≈ 1 ± 0,2 ns jusqu'à 95 ns, où le RMSD a de nouveau augmenté.

La comparaison des valeurs de RMSD calculées, pour les deux systèmes étudiés, montre que le complexe prédit « hit-CK2 » est plus stable que le système expérimental.

III.9.3 Analyse RMSF

Le RMSF, est un outil indispensable pour mesurer l'écart par rapport à une position de référence d'un atome de ligand pendant le temps de simulation. Ce paramètre représente la moyenne de la déviation pour chaque résidu comparativement aux mêmes atomes de la structure de référence. En effet, la figure III-15 montre le diagramme RMSF des acides aminés de la protéine CK2 dans les deux systèmes étudiés hit-CK2 et AGI-CK2 (4DGM). Les

résultats obtenus, ont été comparés au facteur de déplacement atomique « B factor » des atomes Cα de chaque résidu de la protéine étudiée.

La figure III-15, montre de grandes fluctuations des résidus, qui se produisent principalement dans le complexe cristallographique « AGI-CK2 ».

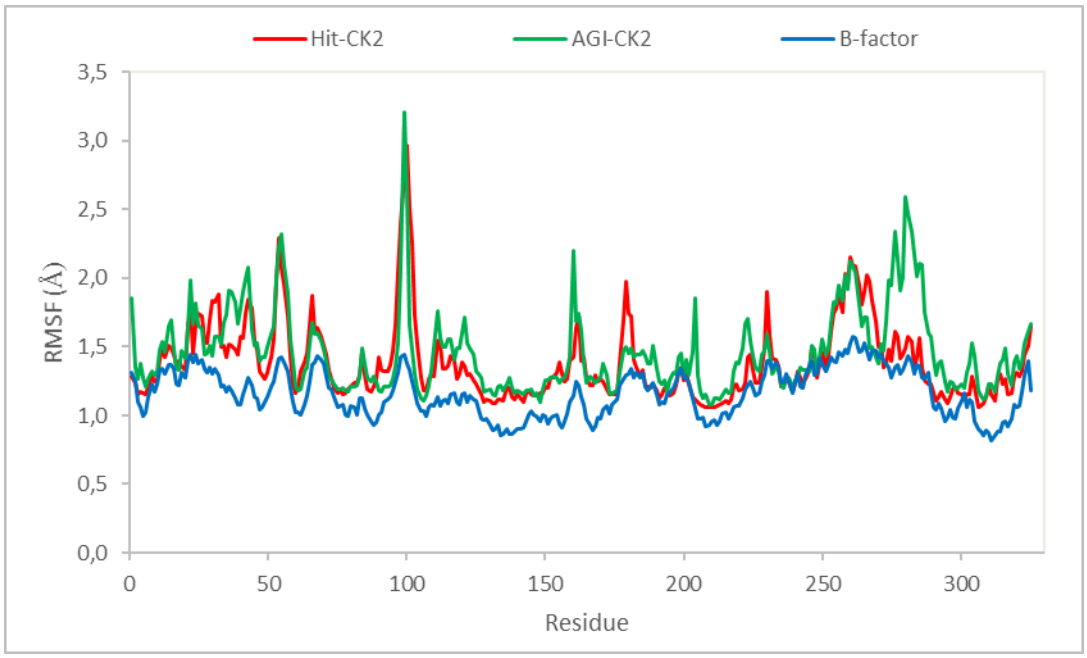


Figure III-15 : Courbes de RMSF des complexes étudiés : complexe prédit « hit–CK2 » (rouge), complexe cristallographique « AGI–CK2 » (vert) et le B factor (bleu)**.**

En effet, les résidus : Ala28 (1,98 Å), Lys49 (2,08 Å), Asn61 (2,32 Å), His105 (3,20 Å), His166 (2,2 Å), Asp210 (1,85 Å), Leu282 (2,34 Å) et Asn286 (2,60 Å), ont des valeurs de RMSF des carbones alpha (Cα) plus élevées dans le complexe « AGI-CK2 » en les comparant à ceux dans le complexe prédit « hit – CK2 ». Dans ce dernier, les valeurs de RMSF des résidus cités plus haut sont : Ala18 (1,77 Å), Lys49 (1,84 Å), Asn61 (2,07 Å), His105 (2,76 Å), His166 (1,42 Å), Asp210 (1,11 Å), Leu282 (1,61 Å) et Asn286 (1,49 Å). Cette différence dans les valeurs de RMSF est provoquée par une grande mobilité de ces résidus au cours de la simulation de dynamique moléculaire.

Cependant, les valeurs de RMSF des carbones alpha (Cα) des résidus Pro72, Gly185, His236 et Gln268, dans le complexe prédit hit – CK2 sont 1,88 Å, 1,98 Å, 1,89 Å et 2,09 Å, respectivement. Les valeurs des mêmes résidus dans le complexe cristallographique AGI-CK2 sont 1,68 Å, 1,50 Å, 1,59 Å et 2,04 Å), respectivement. Ainsi, ces acides aminés ont des valeurs de RMSF plus élevées dans le complexe prédit comparativement à ceux dans le

complexe cristallographique. Par conséquent, la comparaison des deux courbes de RMSF des carbones alpha (Cα) des deux complexes avec la courbe du facteur de déplacement atomique « B-factor », indique une bonne simulation, puisque les deux courbes sont parallèles à celle de « B-factor », tout au long de la simulation de dynamique moléculaire.

III.9.4 Analyse des propriétés du ligand

Pour décrire la surface des deux complexes étudiés « AGI – CK2 » et « hit – CK2 » et leurs interactions potentielles avec les molécules d'eau au cours du temps, nous avons calculé la surface accessible aux solvants (SASA), SASA hydrophobes et SASA hydrophiles pendant les 100 ns de dynamique moléculaire. Un autre critère que nous avons utilisé pour évaluer la stabilité des complexes étudias est le rayon de giration (*radius of giration* ou rGyr). Ce paramètre permet de donner une idée sur l'extension spatiale d'une molécule à partir de son centre de masse.

Comme nous pouvons voir sur les figures III-16 et III-17, les deux systèmes montrent des fluctuations (augmentation /diminution en fonction du temps de simulation) pour les surfaces « SASA hydrophile » et « SASA totale ».

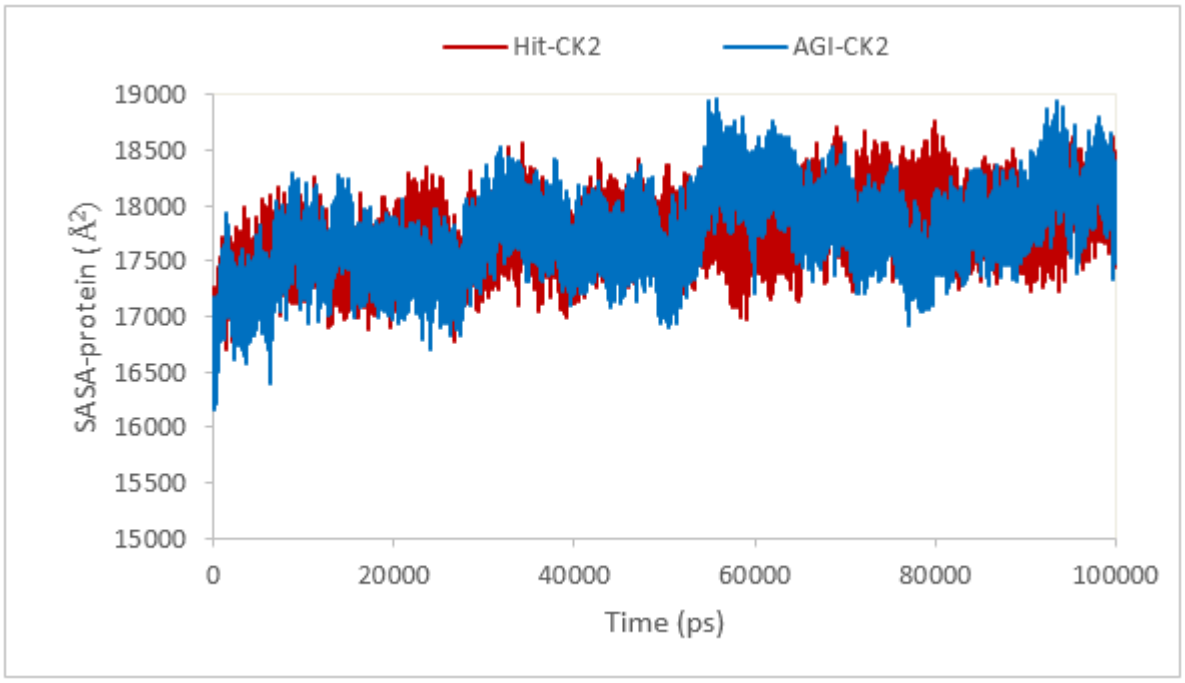


Figure III-16 : Courbe de propriétés « SASA total » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant les 100 ns de simulations de dynamique moléculaire

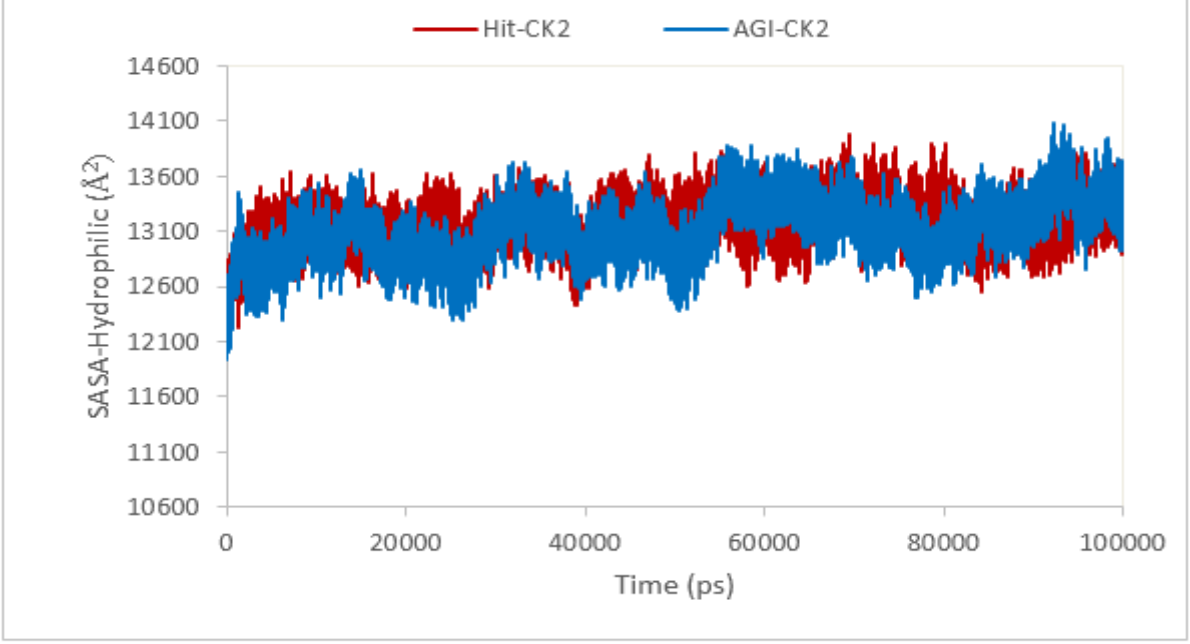


Figure III-17 : courbe des propriétés « SASA Hydrophiles » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant les 100 ns de simulations de dynamique moléculaire.

Il est à noter que, les valeurs moyennes de SASA hydrophile pour les complexes « hit-CK2 » et « AGI-CK2 » sont : 13 165,60 Å2 et 13 109,59 Å2, respectivement, et celles de la SASA totale, sont : 17 740,50 $Å^2$ et 17 749,09 $Å^2$, respectivement. D'après ces résultats, on peut déduire que l'interaction hydrophile a augmenté dans le complexe prédit « hit-CK2 ».

En revanche, la SASA hydrophobe est restée relativement constante jusqu'à 52 ns pour « AGI – CK2 ») et 82 ns pour « hit – CK2 »). A ces deux moments de simulation, on a observé une légère augmentation de la surface « SASA hydrophobe » des deux complexes (Figure III-18).

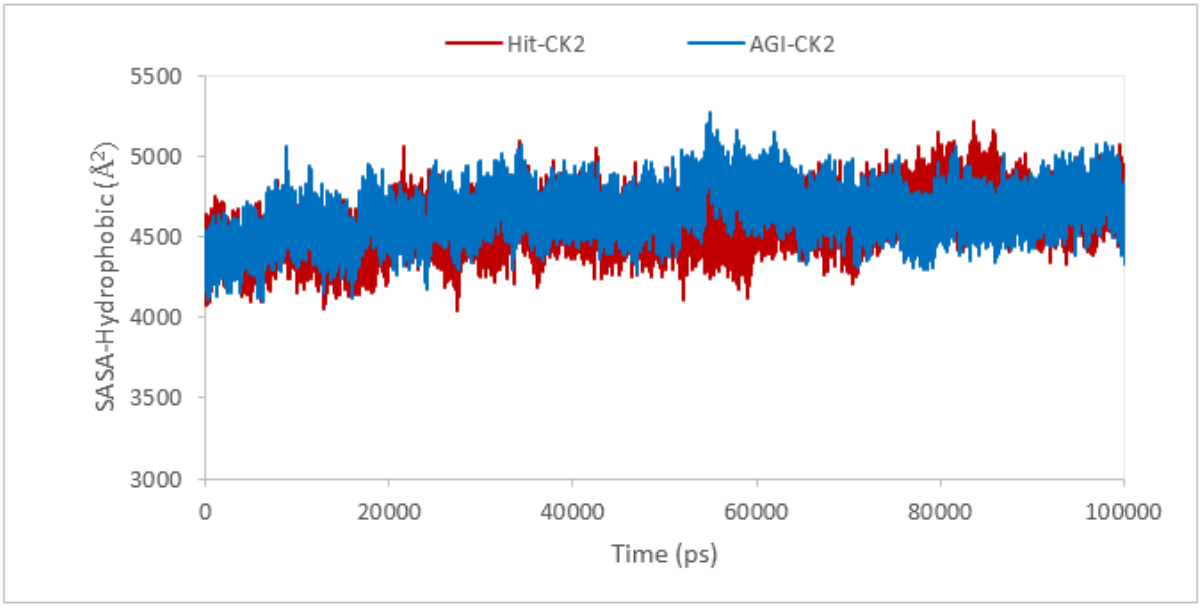


Figure III-18 : Diagramme des propriétés « SASA Hydrophobes » calculées, pour les complexes prédit « hit-CK2 » (rouge) et « AGI-CK2 » (bleu), pendant 100 ns de simulations de dynamique moléculaire.

La figure III-19 illustre les valeurs de rGyr calculées pour les deux complexes, cristallographique et prédit. En effet, la valeur moyenne obtenue pour ces deux derniers, sont : 21,08 Å et 21,12 Å, respectivement.

Il est bien de noter qu'aucune élévation substantielle (> 5 Å) n'a été repérée au cours de toutes les simulations de dynamique et cela dans les deux cas. Cela indique l'absence de repliement de la protéine et de désorganisation des structures secondaires des complexes étudiés.

L'analyse des résultats montre que le complexe prédit « hit-CK2 », est stabilisé à 25 ns et a atteint l'équilibre après 60 ns à environ 21,10 Å. Tandis que le complexe cristallographique « AGI-CK2 » est passé de 20,5 Å à 20,9 Å à 7 ns, pour atteindre un équilibre à 29 ns. Après cet instant, les valeurs de rGyr ont de nouveau augmenté pour atteindre l'équilibre autour de 21,3 ± 0,2 Å après 60 ns.

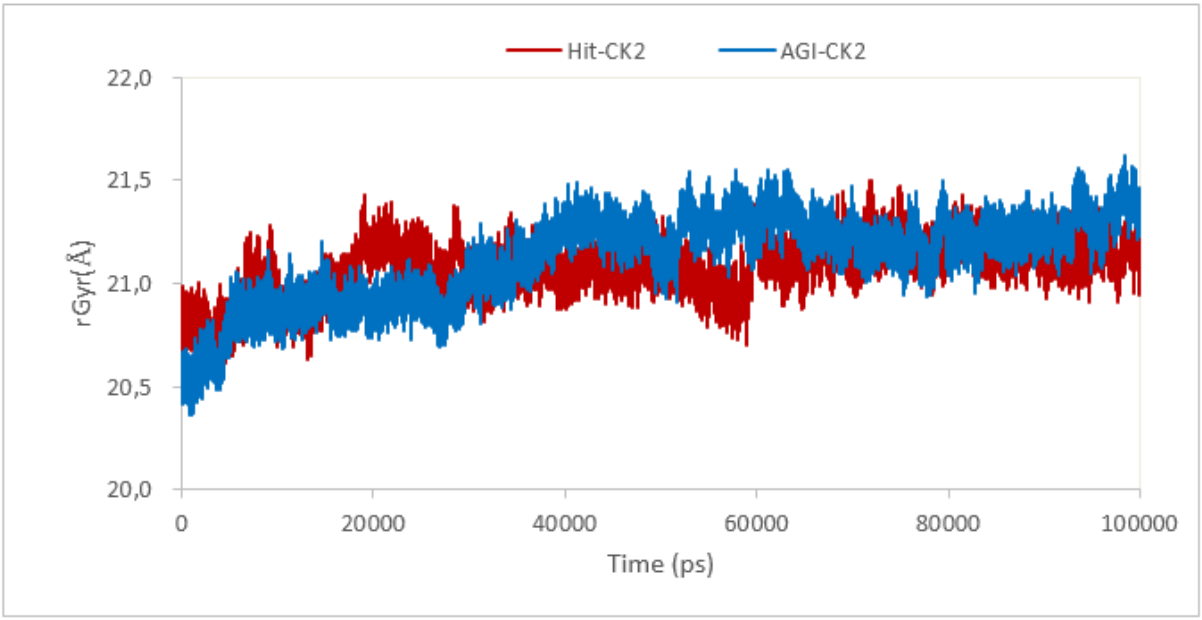


Figure III-19 : Courbe des valeurs de (rGyr) calculées pour les complexes étudiés « hit – CK2 » (rouge) et « AGI–CK2 » (bleu) pendant les 100 ns de simulation de dynamique moléculaire.

D'après les résultats reportés plus haut, nous pouvons alors conclure que les complexes étudiés, expérimental « CK2-AGI » et prédit « CK2-hit », sont stables, puisque les valeurs de rGyr obtenues varient entre 20,4 et 21,5 Å (< 5 Å).

IV. Conclusion

Dans cette étude, nous avons généré des modèles pharmacophores, en utilisant et en combinant les caractéristiques des pharmacophores « *Structure-based* » et « *Ligand-based* ».

En effet, les trois modèles pharmacophore générés, ont été évalués et validés pour leurs qualités de pouvoir identifier et de discriminer des composés chimiques actifs des inactifs, en utilisant plusieurs paramètres, notamment ROC, AUC et un lot d'essai.

Le meilleur modèle pharmacophore, obtenu en fusionnant les caractéristiques pharmacophoriques basées sur la structure du ligand « *ligand-based* » et celle basées sur la structure de la protéine « *structure-based* », consiste en sept caractéristiques, à savoir : deux liaison HBA, deux HBD, une hydrophobe (Hy) et deux caractéristiques aromatiques (Ars). Ce modèle a été alors utilisé pour cribler virtuellement une chimiothèque issue de la banque de données Pubchem Drug Bank dans le but de rechercher de nouveaux inhibiteurs de la CK2.

Les 5984 composés retenus dans l'étape précédente, ont été analysés, testés et criblés à l'aide de plusieurs techniques et filtres, telles que la règle de Lipinski et la simulation du docking moléculaire.

Les modes de liaison favorable et l'énergie de docking des quatorze composés soumis à la simulation de docking, ont été comparés à celui du ligand expérimental « AGI » et analysés par la suite, pour identifier les interactions essentielles dans le site actif de la CK2.

L'analyse des résultats de docking montre que les groupements « oxo pentyl sulfanyle » et « cyano » (CN) du composé criblé (hit), forment des interactions favorables avec les acides aminés Asp175, Asn161, Arg47 et Asn118 du site actif de la CK2.

De plus, le cycle phényle et le groupement éthyle en position meta du cycle pyridine sont entourés des résidus : Val53, Phe113, Ile66, et Ile174 et Val45, Gly46, Met163 et Asn118, engendrant des interactions hydrophobes favorables.

Les 100 ns de simulation de dynamique moléculaire, nous ont fourni un aperçu sur la stabilité moléculaire du ligand prédit « hit » dans le site actif de la CK2. Ces résultats ont été analysés et validés à l'aide de plusieurs paramètres tels que les RMSD (du ligand et des carbone Cα), les liaisons hydrogène formées entre le ligand et la protéine, la fluctuation quadratique moyenne des carbone Cα (RMSF), la surface accessible aux solvants (SASA totale, SASA hydrophobe et SASA hydrophile) et le rayon de giration du complexe (rGyr). Ainsi, les résultats des simulations de dynamique moléculaire sont en parfait accord avec les résultats du docking moléculaire. Par conséquent, un nouvel candidat potentiel « candidat médicament », a été découvert, qui appartient à la famille des pyrido [1,2-a] benzimidazoles, et qui n'a pas encore été caractérisé dans la littérature scientifique comme inhibiteur de la CK2.

V. Références du chapitre III

[1] Di-wu L, Li LL, WangWJ, Xie HZ, Yang J, Zhang CH, Huang Q, Zhong L, Feng S, Yang SY (2012). Identification of CK2 inhibitors with new scaffolds by a hybrid virtual screening approach based on Bayesian model; pharmacophore hypothesis and molecular docking. J Mol Graph Model 36:42–47. https://doi.org/10.1016/j.jmgm.2012.03.004

[2] Morshed MN,MuddassarM, Pasha FA, Cho SJ (2009). Pharmacophore Identification and Validation Study of CK2 Inhibitors Using CoMFA/CoMSIA. Chem Biol Drug Des 74(2):148–158. https://doi.org/10.1111/j.1747-0285.2009.00841.x

[3] Sun HP, Zhu J, Chen FH, You QD (2011). Structure-Based Pharmacophore Modeling from Multicomplex: a Comprehensive Pharmacophore Generation of Protein Kinase CK2 and Virtual Screening Based on it for Novel Inhibitors. Mol Inform 30(6–7): 579–592. https://doi.org/10.1002/minf.201000178

[4] Prudent R, Sautel CF, Cochet C (2010). Structure-based discovery of small molecules targeting different surfaces of protein-kinase CK2. Biochim Biophys Acta 1804(3):493–498. https://doi.org/10.1016/j.bbapap.2009.09.003

[5] Yang SY (2010). Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15(11–12):444–450.https://doi.org/10.1016/j.drudis.2010.03.013

[6] Wolber G, Langer T (2005). LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 45(1):160–169. [22]https://doi.org/10.1021/ci049885e

[7] Baurin N, Baker R, Richardson C, Chen I, Foloppe N, Potter A, Jordan A, Roughley S, Parratt M, Greaney P, Morley D, Hubbard RE (2004). Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. J Chem Inf Comput Sci 44(2):643–651. https://doi.org/10.1021/ci034260m

[8]FlexX version 2.3.2; BioSolveIT GmbH, Sankt Augustin, Germany, 2017, www.biosolveit.de/FlexX

[9] Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R (2014). ProTox: a web server for the insilico prediction of rodent oral toxicity. Nucleic Acids Res 42: W53–W58. https://doi.org/10.1093/nar/gku401

[10] Phillips JC, Braun R, WangW, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005). Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802. https://doi.org/10.1002/jcc.20289

[11] Golub AG, Bdzhola VG, Ostrynska OV, Kyshenia IV, Sapelkin VM, Prykhod'ko AO, Kukharenko OP, Yarmoluk SM (2013). Animal Communication: When I'm Calling You, Will You Answer Too. Bioorg Med Chem 21(21):6681–6689. https://doi.org/10.1016/j. bmc.2013.08.013

[12] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG, (2012). ZINC: A Free Tool to Discover Chemistry for Biology. J Chem Inf Model 52(7):1757–1768.https://doi.org/10.1021/ci3001277

[13] Marvin Sketch (version 18.28.0), (2014). calculation module developed by ChemAxon, http://www.chemaxon.com/products/marvin/marvinsketch/

[14] Cereto-Massagué A, Guasch L, Valls C, Mulero M, Pujadas G, Garcia-Vallvé S (2012). DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. Bioinformatics. 28(12):1661–1662.https://doi.org/10.1093/bioinformatics/bts249

[15] Truchon JF, Bayly CI., (2007). Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. J Chem Inf Model 47:488–508.https://doi.org/10.1021/ci600426e

[16] Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005). Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. J Med Chem 48(7):2534–2547.https://doi.org/10.1021/jm049092j

[17] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1–3):3–26.https://doi.org/10.1016/S0169-409X(00)00129-0

[18] Verber DF, Johnson SR, Cheng HY, Smith BR,WardKW, Kopple KD (2002). Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. J Med Chem 45(12):2615–2623.https://doi.org/10.1021/jm020017n

[19] Ghose AK, Viswanadhan VN, Wendoloski JJ (1999). A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. J Comb Chem 1(1):55–68.https://doi.org/10.1021/cc9800071

[20] Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL, (2012). Quantifying the chemical beauty of drugs. Nat Chem 4(2):90–98.https://doi.org/10.1038/nchem.1243
[21] Daneman R., Prat A., (2015). Cold Spring Harbor Perspectives in Biology. 7 (1).https://doi.org/10.1101/cshperspect.a020412
[22] Castello J, Ragnauth A, Friedman E, Rebholz H (2017). CK2—An Emerging Target for Neurological and Psychiatric Disorders. Pharmaceuticals (Basel) 10(1): E7.https://doi.org/10.3390/ph10010007
[23] Ibragimova GT, Wade RC (1998). Drug-like Annotation and Duplicate Analysis of a 23-Supplier Chemical Database Totalling 2.7 Million Compounds. Biophys J 74(6):2906–2911.https://doi.org/10.1021/ci034260m
[24] Essman U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995). A smooth particle mesh Ewald methodJ Chem Phys 103:8577–8593.https://doi.org/10.1063/1.470117
[25] Ewald P., (1921) « *Die Berechnungoptischer und elektrostatischerGitterpotentiale* », *Ann. Phys.* **64**, pp: 253-287.
[26] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983). Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935.https://doi.org/10.1063/1.445869
[27] Humphrey W, Dalke A, Schulten K (1996). VMD: visual molecular dynamics. J Mol Graph 14 (1): 33-38. https://doi.org/10.1016/0263-7855(96)00018-5
[28] Wallace AC, Laskowski RA, Thornton JM (1996). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8: 127–134. https://doi.org/10.1093/protein/8.2.127
[29] Protopopov MV, Starosyla SA, Borovykov OV, Sapelkin VN, Bilokin YV, Bdzhola VG, Yarmoluk SM (2017). Hit identification of CK2 inhibitors by virtual screening. Biopolymers and Cell 33(4):291–301.https://doi.org/10.7124/bc.00095B
[30] Haidar S, Bouaziz Z, Marminon C, Laitinen T, Poso A, Le Borgne M, Jose J (2017). Development of Pharmacophore Model for Indeno[1,2-b] indoles as Human Protein Kinase CK2 Inhibitors and Database Mining. Pharmaceuticals (Basel) 10(1):8.https://doi.org/10.3390/ph10010008
[31] Pardhi T, Vasu K (2018). Identification of dual kinase inhibitors of CK2 and GSK3β: combined qualitative and quantitative pharmacophore modeling approach. J Biomol Struct Dyn 36(1):177–194.https://doi.org/10.1080/07391102.2016.1270856
[32] Bouaziz-Terrachet S, Toumi-Maouche A, Maouche B, Taïri-Kellou S., (2010). Modeling the binding modes of stilbene analogs to cyclooxygenase 2: a molecular docking study. J Mol Model 16(12):1919–1929.https://doi.org/10.1007/s00894-010-0679-7
[33] Shiri F, Pirhadi S, Rahmani A (2018). Identification of new potential HIV-1 reverse transcriptase inhibitors by QSAR modeling and structure-based virtual screening. J Recept Signal Transduct Res 38(1):37–47.https://doi.org/10.1080/10799893.2017.1414844
[34] Fei F, Zhou Z (2013). New substituted benzimidazole derivatives: a patent review (2010 – 2012). Expert OpinTher Pat 23(9):1157–1179.https://doi.org/10.1517/13543776.2013.800857
[35] Kinoshita T, Sekiguchi Y, Fukada H, Nakaniwa T, Tada T, Nakamura S, Kitaura K, Ohno H, Suzuki Y, Hirasawa A, Nakanishi I, Tsujimoto G (2011). A detailed thermodynamic profile of cyclopentyl and isopropyl derivatives binding to CK2 kinase. Mol Cell Biochem 356(1–2): 97–105.https://doi.org/10.1007/s11010-011-0960-9
[36] Yde CW, Ermakova I, Issinger OG, Niefind K (2005). Inclining the Purine Base Binding Plane in Protein Kinase CK2 by Exchanging the Flanking Side-chains Generates a Preference for ATP as a Cosubstrate. J Mol Biol 347(2):399–414.https://doi.org/10.1016/j.jmb.2005.01.003
[37] Battistutta R, Negro A, Zanotti G (2000). Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. Proteins 41(4):429–437. https://doi.org/10.1002/1097-0134(20001201)41:4<429::aid-prot10>3.0.co;2-d
[38] ProtopopovMV, Ostrynska OV, Starosyla SA, VodolazhenkoMA, Sirko SM, Gorobets NY, Bdzhola V, Desenko SM, Yarmoluk SM (2018). Dihydrobenzo[4,5]imidazo[1,2-a]pyrimidine-4-ones as a new class of CK2 inhibitors. Mol Divers 22(4):991–998.https://doi.org/10.1007/s11030-018-9836-1
[39] Liu H, WangX, Wang J, Wang J, Li Y, Yang L, Li G (2011). Structural Determinants of CX-4945 Derivatives as Protein Kinase CK2 Inhibitors: A Computational Study. Int J Mol Sci 12(10):7004–7021.https://doi.org/10.3390/ijms12107004
[40] Chekanov MO, Ostrynska OV, Synyugin AR, Bdzhola VG, Yarmoluk SM (2014). Design, synthesis and evaluation of 2-phenylisothiazolidin-3-one-1,1-dioxides as a new class of human protein kinase CK2 inhibitors. J Enzyme Inhib Med Chem 29(3) : 338–343.https://doi.org/10.3109/14756366.2013.782024

Références de l'annexe 02

[41] Zhang1 J., Chen P., Yuan B., Ji W., Cheng Z., Qiu X., (2013). Real-Space Identification of Intermolecular Bonding with Atomic Force MicroscopyReal-Space Identification of Intermolecular Bonding with Atomic Force Microscopy". Science 342(6158): 611-614. https://doi.org/10.1126/science.1242603

[42] Horton H R., Moran L A., Ochs R S., Rawn J D., Scrimgeour K G., (1994). Principes de biochimie. Sciences Techniques et Médecine, Sciences de la vie- Biologie- Génétique. Principes de Biochimie, Ed. DeBoeck Universités.

[43] Seve M., (2012). UE1 : Biomolécules (1) : Acides aminés et protéines : Les acides aminés : propriétés physico-chimiques, Université Joseph Fourier de Grenoble, France.

[44] Rarey M., Kramer B., Lengauer T., (1999). The particle concept: placing discrete water molecules during protein-ligand docking predictions. Drug Discovery Today. 34(1): 17-28. https://doi.org/10.1002/(SICI)1097-0134(19990101)34:1<17::AID-PROT3>3.0.CO;2-1

I. Introduction

Comme nous l'avons noté dans le chapitre I, les flavonoïdes constituent une nouvelle classe d'inhibiteurs de la protéine Caséine Kinase 2 (CK2). Ainsi, la liaison produite entre cette protéine et ces composés naturels est gouvernée par les structures moléculaires de ces derniers.

L'objectif principale de ce travail, est de développer des inhibiteurs sélectifs de la CK2 mieux tolérés par l'organisme. A cet égard, nous avons élaboré des modèles OSAR (*Quantitatives Structure Activity Relationships*) en se basant sur les champs d'interaction moléculaire (*molecular interaction field* ou MIFs). Par la suite, nous avons testé le pouvoir prédictif de ces modèles en utilisant plusieurs paramètres statistiques tel qu'un lot d'essai (*test set*), une validation non croisée (*Non-Cross Validation*) et une validation croisée : CV-LOO (*Leave-One-Out Cross validation*) [1], CV-LTO (*Leave-Two-Out Cross Validation*) [2-3] et CV-LMO (*Leave-Many-Out-Cross Validation*) [4].

II. Matériel et Méthodes

II.1 Docking Moléculaire

Nous avons débuté notre travail par une étude de docking moléculaire, afin d'exploiter aussi efficacement possible, les propriétés moléculaires des structures cristallographiques de dérivés de flavonoïdes, et leurs modes de fixation sur le site actif de la protéine CK2, en se basant sur l'approche "*Structure-based*".

II.1.1 Préparation de la protéine (récepteur)

Dans cette étape, nous avons choisi la structure cristallographique de code PDB ID : 4DGM, cristallisée avec son ligand expérimental AGI (Apigenine) et résolue à 1.65 Å [5].

En utilisant le programme « Autodock1.5.4 » [6], nous avons sélectionné tous les acides aminés du site actif de cette protéine se trouvant à une distance inférieure à 3.5Å du centre de ligand AGI.

II.1.2 Préparation des ligands

Nous avons sélectionné une série de 83 molécules dérivées des flavonoïdes et présentant des activités inhibitrices pour la CK2 [7-8]. Nous avons optimisé les géométries de ces molécules, dessinées à l'aide du logiciel ISIS Draw 2.3 [9], en utilisant la méthode steepest-descent implémentée dans le logiciel Chimera UCSF [10].

II.1.3 Protocol du docking moléculaire

Afin de déterminer les modes de liaison ainsi que les interactions pertinentes entre les molécules étudiées et la protéine CK2, nous avons réalisé des simulations de docking moléculaire en utilisant le programme AutoDock [6], avec son algorithme génétique Lamarckien (*Algorithm-Lamarckian-Genetic :* AGL) [6].

En effet, nous avons opté pour un docking semi-flexible [11], en séparant le système en une partie flexible, contenant les résidus flexibles du site actif à savoir : Asp175, Asn 118, Phe 113, Ile 174, Val 53, Val 45, Ile66, Lys 68, Glu114 et Met163, ainsi que le ligand expérimental AGI, et une partie rigide contenant le reste de la protéine CK2.

II.1.4 Evaluations des résultats de Docking

Dans cette étape nous avons évalué les énergies prédites, par le docking moléculaire des complexes obtenus par le docking (Flavonoïde$_n$-CK2), en se basant sur leurs meilleures énergies libres de docking *(Δ_G binding*) [12], et leurs superpositions avec le ligand cristallographique expérimental "AGI".

Il est bien de rappeler que les énergies prédites par le docking moléculaire, sont calculées par l'équation suivante [12] :

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{polar} + \Delta G_{apolar} + \Delta G_{solv} + \Delta G_{flexi} \text{..... (1)}$$

A noter que :

ΔG_0 : La perte d'entropie de translation et de rotation du ligand lors de sa liaison.

$\Delta G_{polaire}$ et **$\Delta G_{apolaire}$** : Les interactions polaires et apolaires, respectivement.

ΔG_{solv} et **ΔG_{flexi}** : Sont les énergies de solvatation et de flexibilité de la structure du ligand.

En effet, toutes les conformations bioactives résultantes, ont été sélectionnées et utilisées dans l'étude QSAR.

II.2 Etude QSAR

Afin de déterminer les éléments structuraux responsables de l'activité des composés étudiés pour la protéine CK2, nous avons effectué une étude QSAR (*Quantitative Structure Activity Relationship*). Cette étude résulte d'une analyse statistique reliant d'une manière quantitative, entre les propriétés physico-chimiques des structures moléculaires (descripteurs moléculaires) et les activités biologiques des molécules organiques. Ceci revient à développer

des modèles mathématiques prédictifs, qui traduisent la relation entre la structure et l'activité, sous la forme suivante :

Activité = *ƒ* (Descripteurs moléculaire) [13-17].

II.2.1 Création de la base de données « structure - activité »

Dans un premier temps, nous avons généré notre base de données, à partir des 83 molécules actives et de leurs activités inhibitrices expérimentales IC_{50} (comprises entre 0.1 et 30 µM) [7-8], Tableau IV-1.

II.2.2 Alignement des molécules

Dans cette étape, nous avons appliqué l'approche « *atom-based alignment method* », implémenté dans le logiciel Open3dallig tools [18], pour aligner les fragments communs, des 83 molécules dans une boite cubique de côté 1Å (Grid box $1Å^3$). Le but de cette étape est d'obtenir un profil d'alignement multiple de la conformation bioactive, en se basant sur la structure commune des dérivés des flavonoïdes, représentée sur la figure 1.

Par la suite, toutes les molécules ont été réalignées encore une fois, selon leurs énergies de docking minimales, par rapport au ligand expérimental de référence AGI (Apigenine) et visualisées à l'aide de logiciel PyMOL [19].

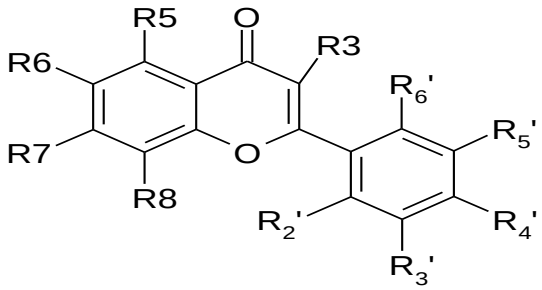


Figure IV-1 : Structure commune de dérivés de flavonoïdes [7].

Tableau IV-1 : Structures moléculaires, activités expérimentales IC_{50} (µM) / pIC_{50} = -log IC_{50} et énergies prédites par le docking moléculaire

Flv N^{br}	Substitution en positions (R2' R3' R4' R5' R3 R5 R6 R7 R8)										Activités expérimentales		Energie prédite par le docking
	R1'	R2'	R3'	R4'	R5'	R3	R5	R6	R7	R8	IC50 (µM)	PIC50	
1	H	H	OMe	H	H	OAc	H	H	H	H	> 30	- 1,48	-11.50
2	H	H	OMe	H	H	H	H	Me	H	H	> 30	- 1,48	-14.81
3	H	H	OMe	H	H	H	H	OH	H	H	30	- 1,48	-13.61
4	H	H	OMe	H	H	H	H	OAc	H	H	> 30	- 1,48	-14.93
5	H	H	Br	H	H	OH	H	Me	H	H	> 30	- 1,48	-14.39
6	H	OMe	OMe	OMe	H	H	H	Me	H	H	> 30	- 1,48	-14.28
7	H	H	NMe_2	H	H	H	Me	H	Me	H	> 30	- 1,48	-14.58
8	H	H	NMe_2	H	H	OH	H	H	H	H	> 30	- 1,48	-12.93
9	H	H	OMe	H	H	OH	H	H	H	H	> 30	- 1,48	-13.98
10	H	OMe	OMe	H	H	OH	H	H	H	H	> 30	- 1,48	-15.46
11	H	H	F	H	H	H	H	H	H	H	> 30	- 1,48	-13.93
12	H	OMe	OH	H	H	H	H	H	H	H	0.8	+ 0,10	-14.37
13	H	OEt	OH	H	H	H	H	H	H	H	7.5	- 0,88	-14.48
14	H	H	H	H	H	OH	H	Me	H	H	> 30	- 1,48	-13.34
15	H	OMe	OMe	H	H	H	H	Me	H	H	> 30	- 1,48	-14.31
16	H	OEt	OH	H	H	H	H	Me	H	H	8	- 0,90	-14.48
17	H	H	Cl	H	H	OH	H	Me	H	H	23	- 1,36	-14.13
18	H	H	F	H	H	OH	H	Me	H	H	> 30	- 1,48	-13.88
19	H	H	Me	H	H	OH	H	Me	H	H	> 30	- 1,48	-14.43
20	H	H	i-Pr	H	H	OH	H	Me	H	H	26	- 1,41	-14.37
21	OH	H	H	H	H	OH	H	Me	H	H	28	- 1,45	-14.92
22	H	H	OMe	H	H	OH	H	Me	H	H	28	- 1,45	-14.62
23	H	OMe	H	H	H	OH	H	Me	H	H	25	- 1,40	-13.55
24	H	H	OEt	H	H	OH	H	Me	H	H	> 30	- 1,48	-14.74
25	H	OMe	OMe	H	H	OH	H	Me	H	H	> 30	- 1,48	-15.70

26	H	H	OH	OH	H	OH	OH	Me	H	OH	> 30	- 1,48	-14.41
27	H	H	Me	H	H	H	H	Me	H	Me	21	- 1,32	-13.97
28	H	OMe	OH	H	H	H	H	Me	H	Me	0.1	+ 1,00	-15.03
29	H	OMe	OMe	H	H	H	H	Me	H	Me	> 30	- 1,48	-14.95
30	H	Br	H	H	H	H	H	OH	H	H	19	- 1,28	-14.02
31	H	OMe	OMe	OMe	H	H	H	OH	H	H	26	- 1,41	-13.89
32	H	OMe	OMe	H	H	OMe	H	H	H	H	> 30	- 1,48	-15.10
33	H	H	H	H	H	OCH_2CO	H	Me	H	H	> 30	- 1,48	-12.16
34	H	H	Me	H	H	OCH_2CO_2H	H	Me	H	H	> 30	- 1,48	-07.83
35	H	H	OMe	H	H	OCH_2CO_2H	H	Me	H	H	> 30	- 1,48	-14.05
36	H	H	Me	H	H	H	H	OH	H	H	19	- 1,28	-13.96
37	H	H	Me	H	H	OCH_2CO	H	H	H	H	> 30	- 1,48	-10.66
38	Cl	H	H	H	H	OCH_2CO	H	H	H	H	> 30	- 1,48	-14.61
39	H	H	Me	H	H	OCH_2CO_2H	H	Me	H	Me	28	- 1,45	-14.11
40	H	H	OMe	H	H	OSO_2Me	H	Me	H	Me	28	- 1,45	-14.94
41	H	H	OMe	H	H	OCO (1.4 Benzo dioxixane-2-yl)	Me	H	Me	H	21	- 1,32	-04.88
42	H	H	OMe	H	H	H	CH2NMe Benzyl	OH	H	H	> 30	- 1,48	-09.37
43	H	H	OMe	H	H	H	CH2(4-Me-1-Pyperidyl)	OH	H	H	> 30	- 1,48	-16.59
44	H	OMe	OMe	OMe	H	H	CH2(4-Et-1-Pyperazinyl)	OH	H	H	> 30	- 1,48	02.77
45	H	OMe	OMe	OMe	H	H	CH2NMe Benzyl	OH	H	H	> 30	- 1,48	-10.28
46	H	OMe	OMe	OMe	H	H	CH2(1-Pyperidyl)	OH	H	H	> 30	- 1,48	-12.33
47	H	H	OMe	H	H	H	CH2Morfoline	OH	H	H	> 30	- 1,48	-14.45
48	H	H	OMe	H	H	H	CH2N(2-furfuryl) CH2O	H	H	H	12	- 1,08	-14.31
49	H	H	OMe	H	H	H	CH2N(3-trifluoromethyle) CH2O	H	H	H	15	- 1,18	-16.35
50	H	OH	OMe	H	H	H	H	Me	H	Me	> 30	- 1,48	-15.27
51	OH	OMe	H	H	H	H	H	Me	H	Me	> 30	- 1,48	-14.74
52	H	OH	H	H	H	H	H	Me	H	Me	2.2	- 0,34	-13.55
53	H	H	OH	H	H	H	H	Me	H	Me	1.6	- 0,20	-15.10
54	H	H	OH	H	H	H	H	Me	H	H	8	- 0,90	-15.31
55	H	OH	OH	H	H	H	H	Me	H	Me	0.45	+ 0,35	-13.95

56	H	OMe	OH	H	H	H	H	OH	H	H	0.25	+ 0,60	-13.70
57	H	OMe	OH	H	H	H	H	Me	H	H	0.32	+ 0,49	-14.59
58	H	OMe	OH	H	H	H	H	OMe	H	H	0.20	+ 0,70	-14.31
59	H	OMe	OH	H	H	H	H	Cl	H	H	0.24	+ 0,62	-14.65
60	H	OMe	OH	H	H	H	H	Et	H	H	0.12	+ 0,92	-14.49
61	H	OMe	OH	H	H	H	H	NAc	H	H	0.15	+ 0,82	-12.39
62	H	OMe	OH	H	H	H	H	Br	H	H	0.12	+ 0,92	-14.85
63	H	OMe	OH	H	H	H	H	H	Me	H	0.09	+ 1,05	-14.09
64	H	OMe	OH	H	H	H	H	Me	H	Me	0.09	+ 1,05	-14.58
65	H	OMe	OH	H	H	H	H	Me	H	Me	0.65	+ 0,19	-15.11
66	H	OMe	OH	H	H	H	H	Cl	Me	H	0.79	+ 0,10	-14.85
67	H	OMe	OH	H	H	H	H	Me	Me	H	3	+ 0,48	-14.46
68	H	OMe	OH	H	H	H	H	Cl	H	Cl	0.01	+ 2,00	-15.14
69	H	OMe	OH	H	H	H	H	Cl	H	Cl	0.02	+ 1,70	-15.67
70	H	NO_2	OH	H	H	H	H	Me	H	Me	0.012	+ 1,92	-15.45
71	H	OMe	OH	H	H	H	Me	H	Me	H	0.03	+ 1,52	-14.99
72	H	OMe	OH	NO_2	H	H	H	Me	H	Me	0.066	+ 1,18	-15.88
73	H	Cl	OH	Cl	H	H	H	Me	H	Me	0.078	+ 1,11	-15.48
74	H	Br	OH	H	H	H	H	Me	H	Me	0.05	+ 1,30	-14.75
75	H	Br	OH	Br	H	H	H	Me	H	Me	0.18	+ 0,74	-15.72
76	H	OMe	OH	H	H	H	H	Cl	H	Me	0.04	+ 1,40	-15.30
77	H	OMe	OH	H	H	H	H	H	C_4H_4	H	0.028	+ 1,55	-15.44
78	H	OMe	OH	Cl	H	H	H	Br	H	Br	0.018	+ 1,74	-15.42
79	H	OMe	OH	H	H	H	H	Br	H	Br	0.004	+ 2,40	-15.48
80	H	OMe	OH	Cl	H	H	H	H	C_4H_4	H	0.087	+ 1,06	-14.79
81	H	OMe	OH	H	H	H	Me	Cl	Me	H	2	- 0,30	-15.48
82	H	OMe	OH	H	H	H	H	H	Me	Me	0.3	+ 0,52	-13.99
83	H	OMe	OH	H	H	H	H	OMe	H	H	0.175	+ 0,76	-14.36

II.2.3 Génération des champs d'interactions moléculaires (MIFs)

Dans cette étape nous avons créé et choisi les descripteurs moléculaires, qui sont les champs d'interactions moléculaires notés : MIFs (*Molecular Interaction fields*, en anglais), en utilisant l'outil COMFA (*Comparative molecular field approach*), implémenté dans le logiciel Open 3D QSAR [18].

En effet, les MIFs sont calculés en déplaçant un atome de chaque molécule alignée, appelé atome sonde (*probe atome*, en anglais) à travers tous les nœuds de la boite cubique ($1Å^3$), en évaluant par conséquent son énergie d'interaction avec sa molécule.

Ensuite, nous avons fixé des seuils d'alerte, par des rayons de switch minimal et maximal (*cutoffs*, en anglais), réglés à des valeurs d'énergies -30,0 kcal / mol et +30,0 kcal / mol, respectivement. Cela, afin de protéger les modèles QSAR des valeurs élevées, qui peuvent fortement détruire leurs matrices PLS (*partial least squares*) correspondantes.

Toutes les interactions interatomiques, des structures tridimensionnelles des molécules étudiées, sont utilisées pour établir les équations QSAR.

En effet, les énergies d'interaction stérique $\mathbf{E_{Vdw}}$ (*van der Waals*), électrostatique $\mathbf{E_{ele}}$ (*coulomb*), ont été calculées également pour chaque molécule de la base de données, à chaque nœud de la boite cubique (*Grid box*, en anglais), en utilisant les champs de force MMFF94 et MMFF94s [20] de l'outil COMFA [18], par les équations suivantes :

$$E_{Vdw} = \sum_{i=1}^{n} \left(\frac{A_i}{r_i^{12}} - \frac{B_i}{r_i^{6}} \right) \quad \text{..... (2)}$$

$$E_{ele} = \sum_{i=1}^{n} \frac{Qq_i}{r_i^{d}} \quad \text{..... (3)}$$

A noter que :

n: le nombre d'atomes ; $\mathbf{r_i}$: la distance entre l'atome sonde et l'atome i

$\mathbf{A_i}$ et $\mathbf{B_i}$: sont les paramètres de Van der Waals pour l'atome i

Q :la charge de l'atome sonde ; $\mathbf{q_i}$: la charge de l'atome i

d: la distance diélectrique (d = 1, constant et d = 2, distance dépendante)

II.2.4 Répartition de la base de données

Nous avons divisé notre base de données, composée initialement de 83 dérivés de flavonoïdes, en deux ensembles :

- **Un lot d'entrainement** : appelé aussi lot d'apprentissage (*training set*, en anglais), contenant 61 molécules de flavonoïdes, utilisées afin de générer les modèles QSAR.
- **Un lot de test** : appelé aussi lot d'essai (*test set*, en anglais), contenant les 22 molécules restantes, pour évaluer et valider les capacités et les performances prédictives des modèles QSAR résultants, sur des molécules externes.

II.2.5 Méthode de régression linéaire par les moindres carrées partiels (PLS)

Nous avons utilisé la méthode de régression linéaire par les moindres carrés partiels PLS (*partial least squares*) pour l'analyse des modèles QSAR obtenus [21-24].

En effet, la corrélation statistique entre les valeurs d'énergie des MIFs à chaque nœud de la boite cubique (x) avec les activités biologiques inhibitrice (y), est exprimée par la matrice des pseudo-coefficients de PLS, suivante :

$$\boldsymbol{Y = XA + B} \text{..... (4)}$$

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_N \end{pmatrix} \begin{pmatrix} X_{11} & \cdots & X_{1K} \\ \vdots & \ddots & \vdots \\ X_N & \cdots & X_{NK} \end{pmatrix}$$

Où :

Y : le vecteur de l'activité biologique (variable dépendante), à savoir : Y = 1,2, ..., n, N

X : la matrice des attributs ou descripteurs moléculaires (variables indépendantes), à savoir : X = 1,2,3,4, ..., n, N

A : la matrice des pseudo-coefficients

B : la matrice des erreurs de régression

En outre, les paramètres optimaux de l'algorithme PLS implémenté dans le logiciel Open 3D QSAR, ont été réglés à extraire jusqu'à 5 composantes principales (*Principal components* : PCs, en anglais) des descripteurs MIFs [18].

II.2.6 Validation des modèles QSAR

II.2.6.1 Validation non-Croisée (*non-Cross Validation*)

En premier temps, nous avons réalisé une validation non croisée, par la méthode de régression linéaire PLS, afin de déterminer :

- ✓ Le nombre optimal de composantes principales (PCs) de logarithme PLS, à extraire des descripteurs MIFs.

- ✓ Les coefficients de détermination (r^2) entre les activités biologiques et les descripteurs moléculaires [25].
- ✓ Les coefficients de corrélation (R), ou les racines carrées des coefficients de détermination (r^2), issues de la validation non croisée [25].
- ✓ Les coefficients de pondération BUW (*Block Unscaled Weighting*) attribués pour équilibrer la répartition des champs stérique et électrostatique lors du calcul des matrices PLS des modèles QSAR.
- ✓ Le standard déviation de l'erreur d'estimation calculée (*SDEC : Standard deviation of Error Calculated*, en anglais), exprimé par la formule suivante :

$$\mathbf{SDEC} = \sqrt{\frac{\sum(\mathbf{Y_{training,i}} - \hat{\mathbf{Y}}_{\mathbf{training,i}})^2}{\mathbf{n}}} \quad \textbf{..... (5)}$$

Avec :

$Y_{training,i}$: c'est la valeur de l'activité biologique de la molécule i (composante de l'ensemble d'essai), pour les différentes pc=1,2,....5.

$\hat{Y}_{training,i}$: C'est la valeur moyenne de l'activité biologique de la molécule i (composante de l'ensemble d'essai) pour les différentes pc=1, 2, ...5.

n : C'est nombre de molécules de l'ensemble d'essai.

II.2.6.2 Validation Croisée (CV)

II.2.6.2.1 Validation croisée LOO

En premier temps, nous avons réalisé une validation croisée interne, sur le lot d'entrainement, afin de déterminer la stabilité du modèle QSAR, et de tester l'impact et l'influence de chaque molécule de l'ensemble d'entrainement. Cela, en utilisant la technique de validation croisée Leave one out (CV-LOO) [1], détaillée dans le chapitre 2 de cette thèse.

II.2.6.2.2 Validations croisées LTO, LMO

Par la suite, nous avons réalisé une validation croisée externe, en appliquant les modèles QSAR obtenus sur les 22 molécules du lot de test, en utilisant les méthodes de validation croisée Leave-two-out (LTO) [2-3] et leave-many-out (LMO) [4].

Cela, afin de valider la fiabilité et la performance prédictive des modèles QSAR, pour mesurer la robustesse de prédiction des activités pour des molécules externes.

Selon Bravais-Pearson [25], plus le coefficient de corrélation est supérieur à 0.75 et proche de 1, plus la corrélation est idéale, et notre modèle QSAR est performant et fiable, pour prédire les activités biologiques des molécules, qui n'appartiennent pas au lot d'entrainement.

En dernière étape, nous avons évalué la performance et la fiabilité des modèles QSAR obtenus, en comparant entre les coefficients de prédiction q^2, les coefficients de corrélation prédite Q, et les standards déviation de l'erreur prédite (SDEP) des modèles QSAR, obtenus par les trois méthodes décrites ci-dessus : LOO, LMO et LTO, à savoir les paramètres statistiques suivants :

- ✓ $q^2_{CV\text{-}LOO}$, $q^2_{CV\text{-}LTO}$ et $q^2_{CV\text{-}LMO}$
- ✓ $Q_{CV\text{-}LOO}$, $Q_{CV\text{-}LTO}$ et $Q_{CV\text{-}LMO}$
- ✓ SDEP $_{CV\text{-}LOO}$, SDEP $_{CV\text{-}LTO}$ et SDEP $_{CV\text{-}LMO}$

Il convient de noter que :

- ✓ Le standard déviation de l'erreur prédite (SDEP : *Standard Deviation Error Predected*, en anglais), est calculé par la formule suivante :

$$\boldsymbol{SDEP_{test} = \sqrt{\frac{\sum_i \left(Y_{test,i} - \hat{Y}_{test,i}\right)^2}{n}}} \quad \textbf{(7)}$$

Avec :

$Y_{test,i}$: c'est la valeur de l'activité biologique de la molécule i (composante de l'ensemble de test), pour les différentes pc=1,2,....5.

$\hat{Y}_{test,i}$: C'est la valeur moyenne de l'activité biologique de la molécule i (composante de l'ensemble de test) pour les différents pc=1, 2, ...5.

n : C'est nombre de molécules de l'ensemble de test.

- ✓ Le Coefficient de prédiction q^2, est calculé par la formule suivante :

$$\mathbf{q^2 = 1 - \frac{\sum_{i=1}^{N} \left(Y_{prédite} - Y_{réelle}\right)^2}{\sum_{i=1}^{N} \left(Y_{réelle} - Y_{moyenne}\right)^2}} \quad \textbf{(8)}$$

- ✓ Le coefficient de corrélation prédit **Q**, est la racine carrée du coefficient de prédiction q^2.

Où :

$Y_{Prédite}$, $Y_{réelle}$ et $Y_{moyenne}$, sont les valeurs prédites, réelles et moyennes de l'activité biologique (pIC_{50}) de la molécule i, respectivement.

III. Résultats et discusion

III.1 Docking moléculaire

Pour chaque complexe obtenu par le docking : « Flavonoïde- CK2 », nous avons généré 50 conformations de l'inhibiteur (dérivé de flavonoïde) dans le site actif de la protéine CK2.

En se basant sur les trois critères d'évaluation : (meilleure énergie d'interaction, meilleure énergie libre de docking et meilleure superposition avec le ligand cristallographique expérimental "AGI"), nous avons choisi pour chaque complexe, une meilleure conformation dite : "bioactive", que nous avons énuméré dans le tableau 1. En effet, nous avons reporté dans le tableau IV-1, les énergies prédites par le docking moléculaire, qui correspondent aux conformations bioactives des 83 dérivés de flavonoïdes étudiées.

En outre, la figure IV-2 montre un exemple de la parfaite superposition, entre la quercitrine (Flv26) et le ligand expérimentale Apigenine (AGI).

Nous avons conclu par l'analyse de tous les complexes "Flavonoïde-CK2", que leurs conformations bioactives relatives, sont stabilisées dans le site actif de la CK2, par l'intermédiaires de liaisons hydrogènes et d'interactions hydrophobes avec les résidus du site actif, notamment : Asp175, Asn 118, Phe 113, Ile 174, Val 53, Val 45, Ile66, Lys 68, Glu114 et Met163.

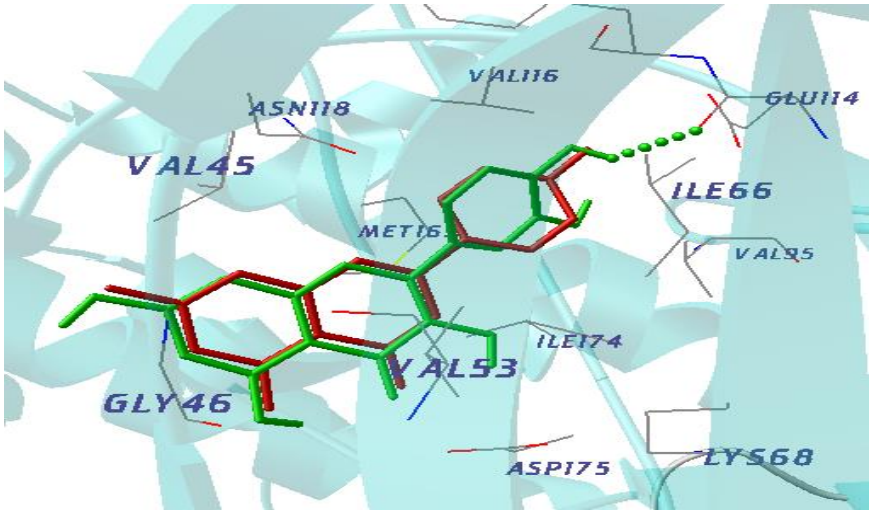


Figure IV-2 : La superposition entre la conformation de la quercitrine (Flv26) issue du docking moléculaire (en vert) et celle du ligand expérimental AGI (en rouge) dans le site actif de la CK2.

La figure IV-3 représente les interactions moléculaire clés, du complexe obtenu par le docking "quercitrine-CK2" dans le site actif de la CK2.

Nous avons observé, que la molécule d'eau Hoh581 et l'acide aminé Asn118, forment des liaisons hydrogènes avec le groupement carbonyle (C=O) de la quercitrine. Aussi, la molécule d'eau Hoh501 et les acides aminés : Lys 68, Asn118 et Val 116, forment des liaisons hydrogène avec le groupement OH de la quercitrine (Flv26).

Ajouter à cela, le groupement OH en position R'_4 de la quercitrine, forme une liaison hydrogène avec le groupement NH_3^+ chargé positivement, de la fonction "Amine" (**NH_3**) de la chaîne latérale de l'acide aminé Lys 68.

En outre, nous avons constaté lors de l'analyse des résultats, que les acides aminés :"Lys68, Val45, Asn118, Val116, et Asp175, Phe113, Ile 174, val 53, Ile 66 et Met163", sont les principaux résidus clés, impliqués dans la formation de liaisons hydrogènes avec les flavonoïdes ayant de basses activités inhibitrices IC50 envers la CK2.

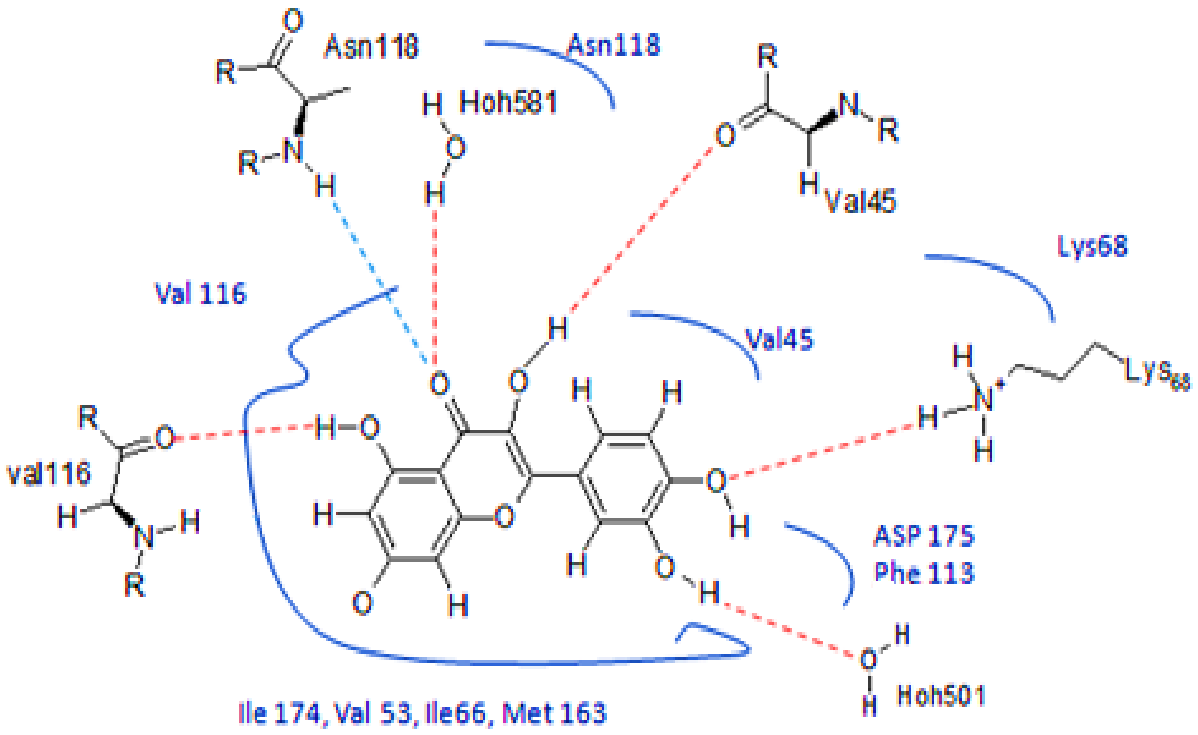


Figure IV-3 : Interactions moléculaires clés entre la Quercitrine et les résidus du site actif de la CK2.

III.2 Etude QSAR

III.2.1 Alignement des molécules

La figure IV-4, illustre la superposition des 83 molécules dérivés de flavonoïdes, alignées dans la boite cubique de 1 $Å^3$, en utilisant la méthode d'alignement "*atom-based*", du logiciel Open3dallig [18].

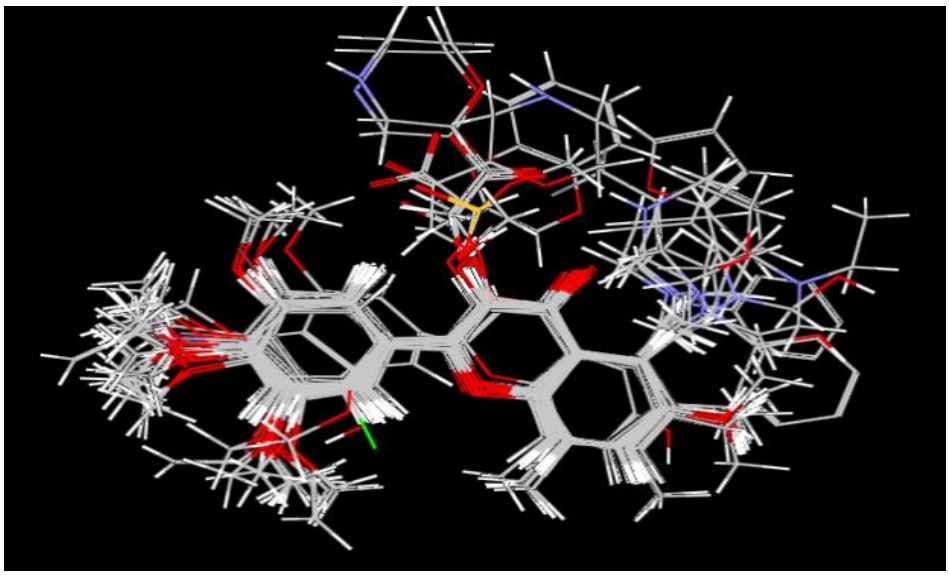


Figure IV-4 : Alignement des 83 molécules des dérivés des flavonoïdes dans la boite cubique de 1$Å^3$.

III.2.2 Analyse des MIFs

Nous avons effectué l'analyse des descripteurs moléculaires MIFs (*Molecular Interaction Fields*) par l'outil COMFIA implémenté dans le logiciel Open3DQSAR, en appliquant la méthode de régression linéaire des moindre carrés partiels (PLS).

La répartition des valeurs des énergies des descripteurs MIFs dans les matrices PLS, a été équilibrée par l'attribution des coefficients de pondération, nommés : BUW (*Block Unscaled Weighting*) [18], qui était de l'ordre de **1.0656** et **0.9344**, pour les champs stérique et électrostatique, respectivement.

Les recouvrements des champs stériques et électrostatiques, calculés pour les divers emplacements des atomes sondes (*probe atome*) à travers la boite cubique de 1 $Å^3$, pour les 83 molécules alignées sont représentés sur les figures IV-5 et IV-6, respectivement.

Il est bien de noter que la charge du groupement carbone alkyle "Alkyl Carbon : R-AC", a été appliquée par défaut lors du calcul des énergies des champs stérique et

électrostatique, en utilisant les rayons de switch (*cutoffs*) minimal et maximal, réglés à -30 et+30 kcal/mol, respectivement.

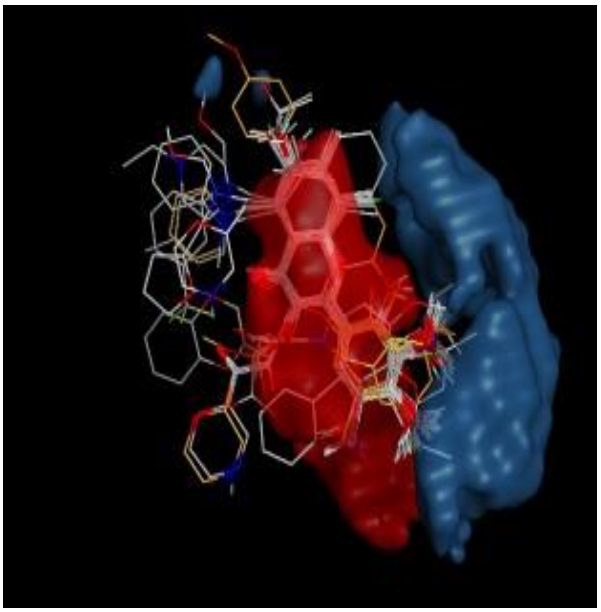


Figure IV-5 : Visualisation du recouvrement du champ électrostatique, les régions favorables et défavorables du champ électrostatique, sont illustrées en bleu et en rouge, respectivement.

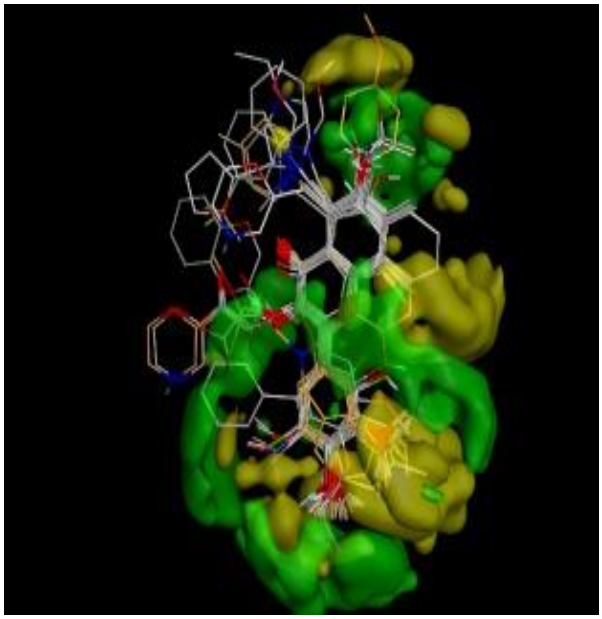


Figure IV-6 : Visualisation du recouvrement du champ stérique, les régions favorables et défavorables du champ stérique, sont illustrées en vert et en jaune, respectivement.

III.2.3 Analyse des matrices PLS des descripteurs moléculaires MIFs

Les matrices PLS résultantes, ont été analysées par la méthode de régression linéaire des moindres carrés partiels (PLS), en combinant les champs stérique et électrostatique uniquement.

Afin d'améliorer considérablement les performances de calcul, nous avons réduit le nombre de variables actives des matrices PLS, par l'extraction jusqu'à 5 composantes principales (*Principal components : PCs*, en anglais) de l'algorithme PLS, en recherchant les paramètres optimaux. En effet, nous avons obtenus cinq modèles QSAR, générées pour les différents paramètres optimaux de composantes principales PCs, à savoir : (PC = 1, PC = 2, PC = 3, PC = 4 et PC = 5).

III.2.4 Validation non Croisée (*non-Cross Validation*)

En premier temps, nous avons vérifié l'aptitude prédictive des cinq modèles QSAR, générés pour les différents paramètres optimaux de composantes principales (PCSs), en réalisant une validation non croisée par la méthode PLS.

Les figures IV-7, IV-8, IV-9, IV-10 et IV-11, représentent les courbes de régression linéaire des moindres carrées partiels (PLS) de la validation non croisée des modèles QSAR, générés pour les : pc=1, jusqu'à pc=5, respectivement.

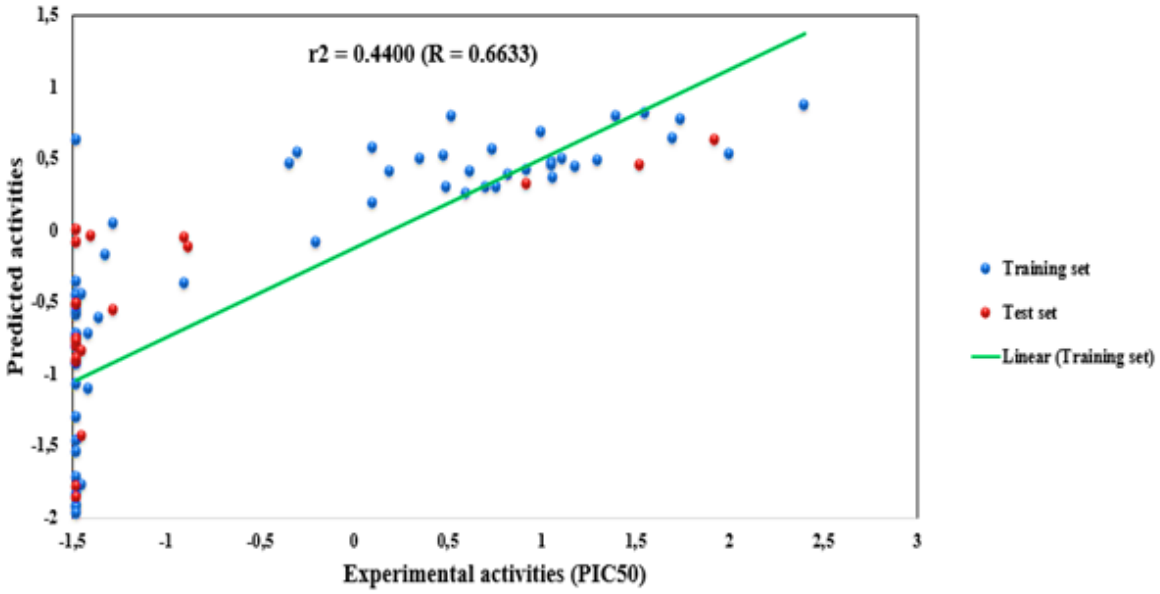


Figure IV-7 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 1), en fonction des activités expérimentales PIC50.

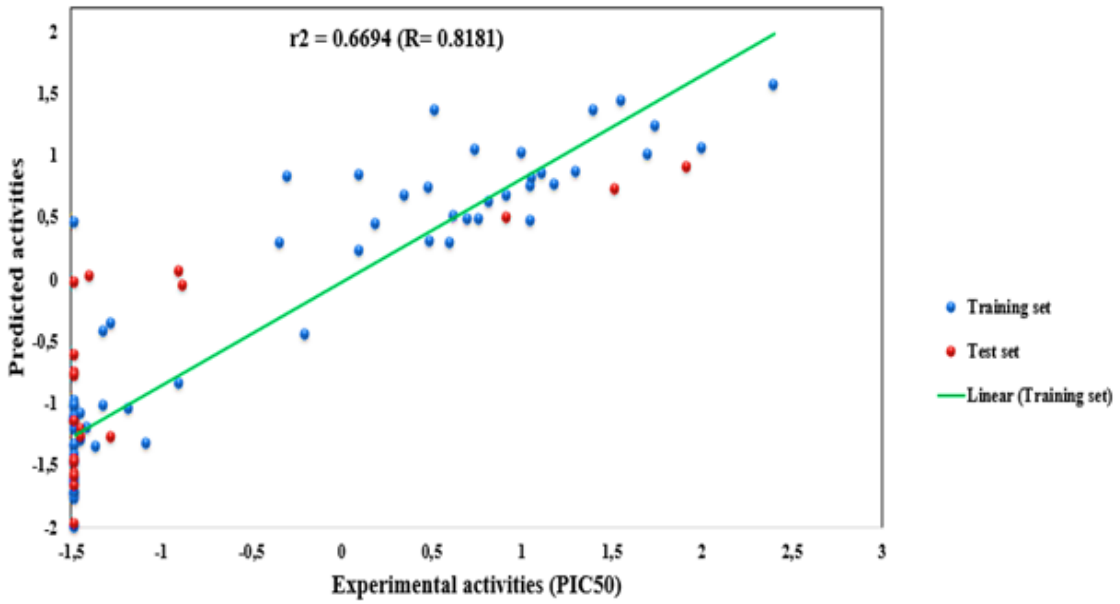


Figure IV-8 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2), en fonction des activités expérimentales PIC50.

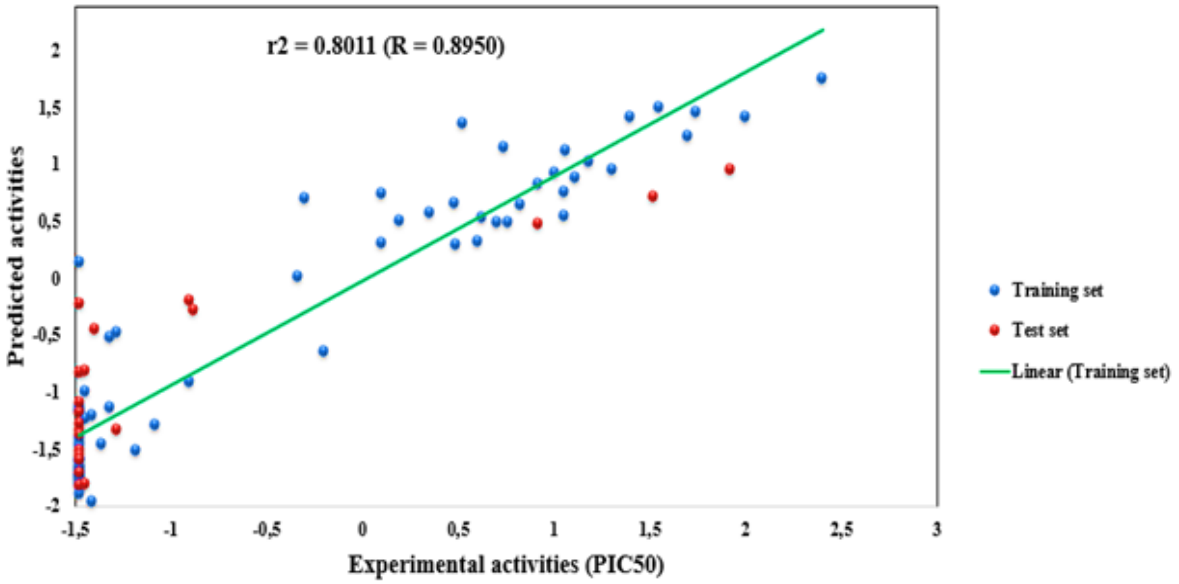


Figure IV-9 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=3), en fonction des activités expérimentales PIC50.

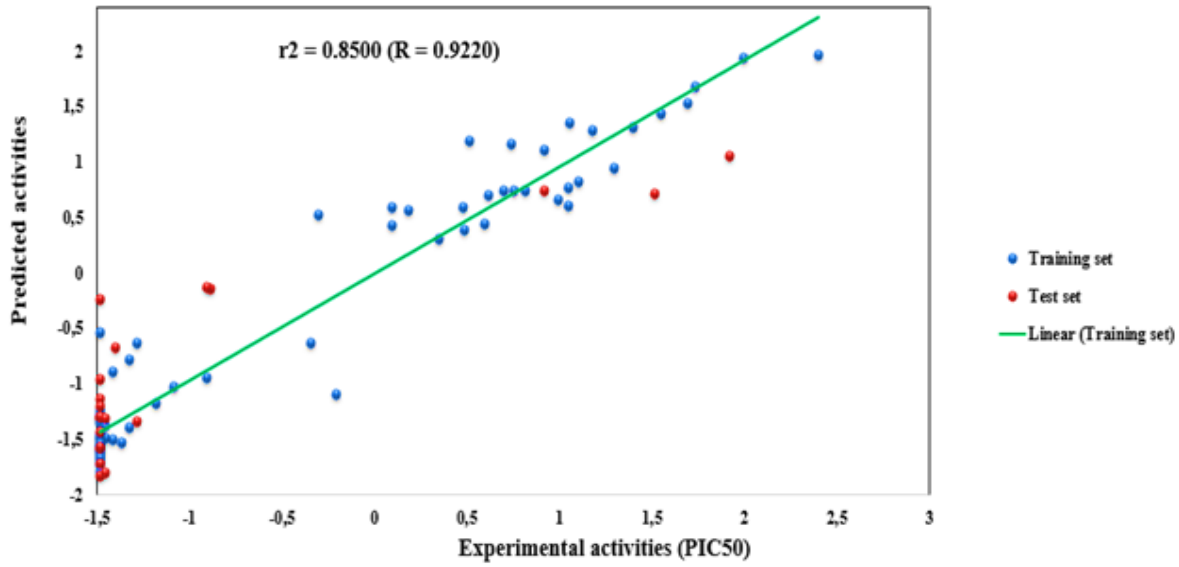


Figure IV-10 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=4), en fonction des activités expérimentales PIC50.

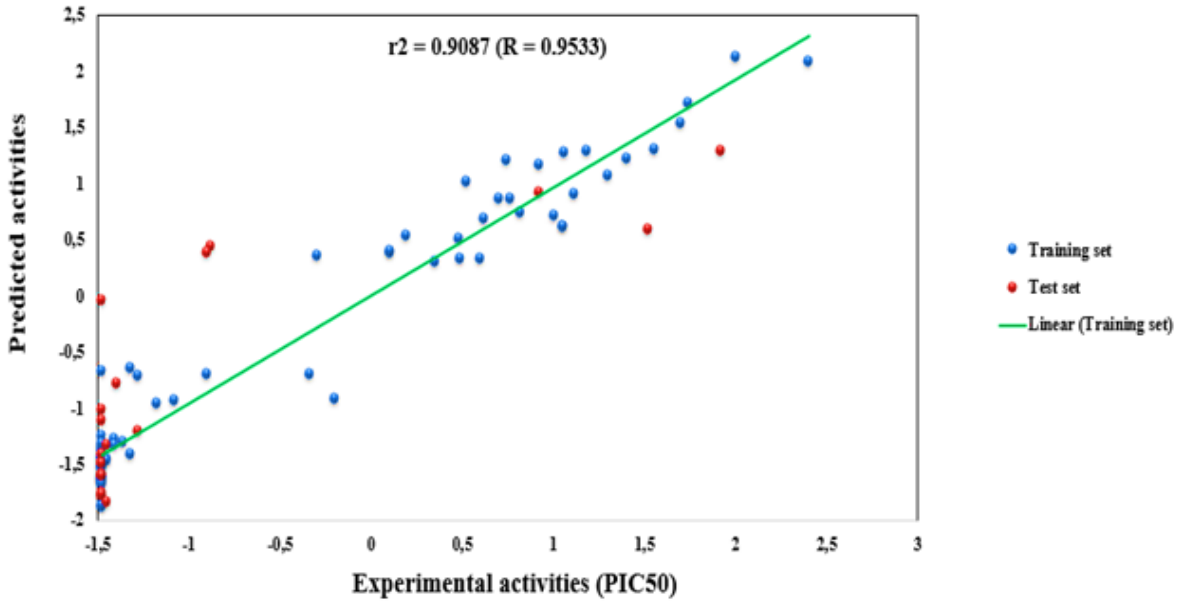


Figure IV-11 : Représentation graphique des activités inhibitrices prédites par la PLS (pc=5), en fonction des activités expérimentales PIC50.

En effet, nous avons résumé dans le tableau IV-2, les valeurs des paramètres statistiques obtenues pour les cinq modèles QSAR générés, à savoir : le standard déviation de l'erreur d'estimation calculée (SDEC), le coefficient de détermination r^2, la racine carrée du coefficient de détermination R (Coefficient de corrélation) et les coefficients de pondération attribués pour les champs stérique et électrostatique.

Tableau IV-2 : Résultats de la validation non croisée des cinq modèles QSAR générés, PC = 1, jusqu'à PC = 5.

PC	SDEC %	Coefficient de détermination (r^2)	Coefficient de corrélation (R)	Coefficient de pondération (BUW)	
				Champ stérique (Vdw)	Champ électrostatique (Coulomb)
1	0.9256	0.4400	0.6633	1.0656	0.9344
2	0.7112	0.6694	0.8181	1.0657	0.9345
3	0.5517	0.8011	0.8950	1.0658	0.9346
4	0.4790	0.8500	0.9220	1.0659	0.9347
5	0.3737	0.9087	0.9533	1.0660	0.9348

L'analyse des résultats portés dans le tableau IV-2, et illustrés sur les figures IV-7, IV-8, IV-9, IV-10 et IV-11, montre que quatre sur cinq (4/5) modèles QSAR, sont très

satisfaisants et acceptables par rapport aux limites de spécifications requises définies par Bravais-Pearson [25].

En outre, les modèles PLS retenus correspondent aux modèles générés avec les paramètres optimaux suivants : PC = 2, PC = 3, PC = 4 et PC = 5.

En effet, nous avons obtenus des valeurs de coefficients de corrélation R pour les modèles PC = 2, PC = 3, PC = 4 et PC = 5, largement supérieures à la limite de spécification **0.75**, à savoir 0.8181, 0.8950, 0.9220 et 0.9533, respectivement.

Ajouter à cela, les valeurs de standard déviation de l'erreur calculée (SDEC), relatives aux quatre modèles, sont largement inférieures à la valeur de SEDC maximale tolérée, qui est égale à 5%, à savoir : SDEC= 0.7112 % (PC = 2), 0.5517 % (PC = 3), 0.4790 % (PC = 4) et 0.3737 % (PC = 5).

Nous concluons que les paramètres optimaux calculés par la méthode de validation PLS non croisée (Non-Cross Validation) pour les quatre modèles générés, PC = 2, PC = 3, PC = 4 et PC = 5 sont acceptables et peuvent être soumis aux tests de validation croisée (Cross Validation).

III.2.5 Validation Croisée (*Cross Validation*)

Par la suite de notre étude QSAR, nous avons validé les quatre modèles QSAR vérifiés précedament, par la validation non croisée, en utilisant les trois méthodes de validation croisée LOO, LTO et LMO, détailées dans le chapitre 2 de ce manuscrit.

III.2.5.1 Validation croisée CV-LOO

Nous représentons sur les figures IV-12, IV-13, IV-14 et IV-15 les courbes de régression linéaire des moindres carrées partiels (PLS) de la validation croisée Leave-One-Out (CV-LOO), des quatre modèles QSAR, générés, PC = 2, PC = 3, PC = 4 et PC = 5.

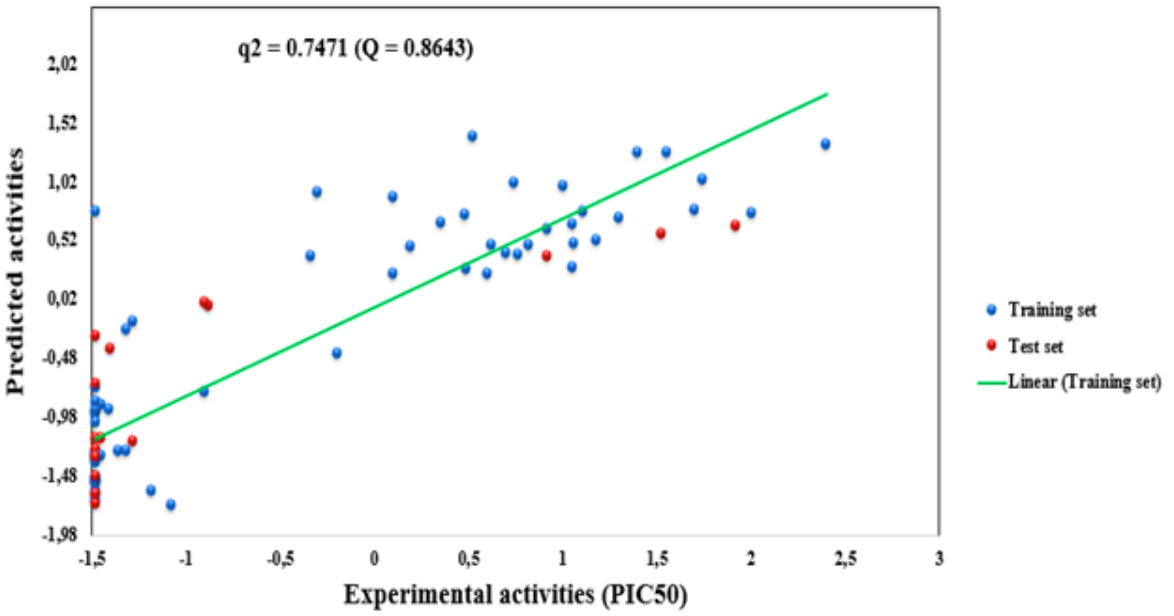


Figure IV-12 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.

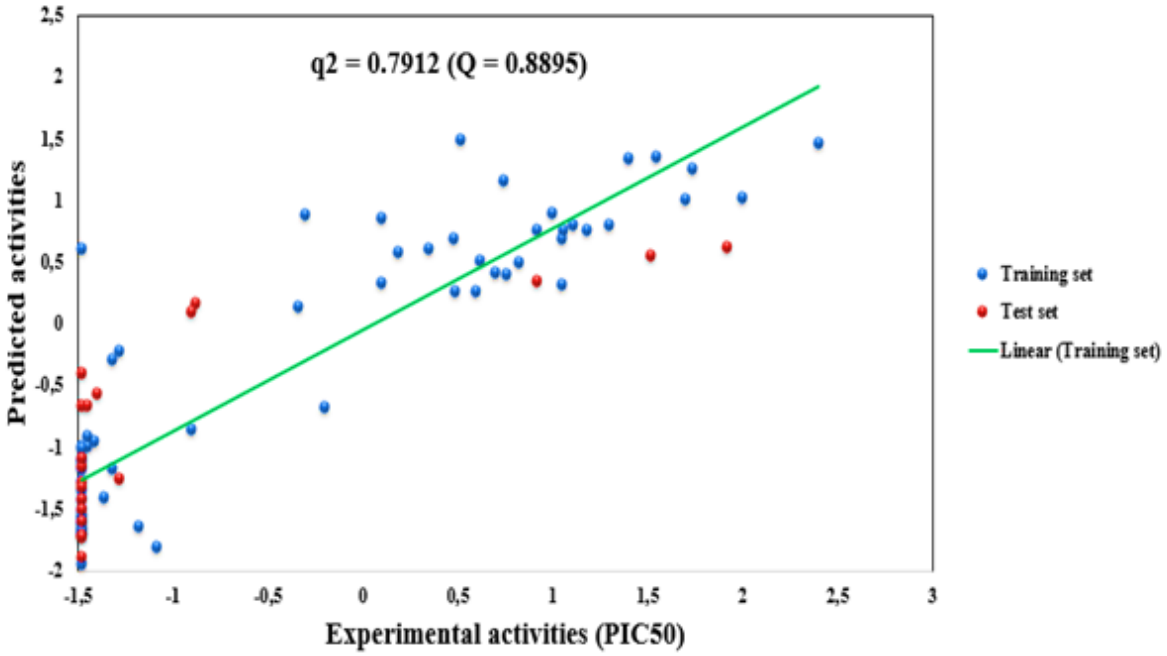


Figure IV-13 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.

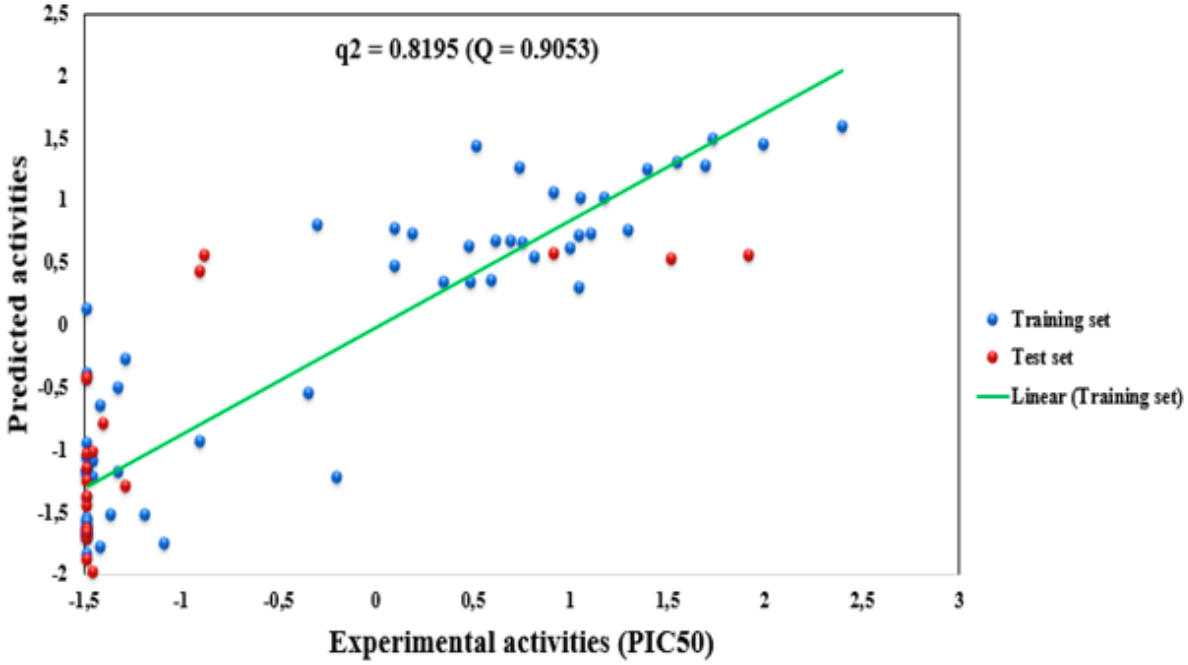


Figure IV-14 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.

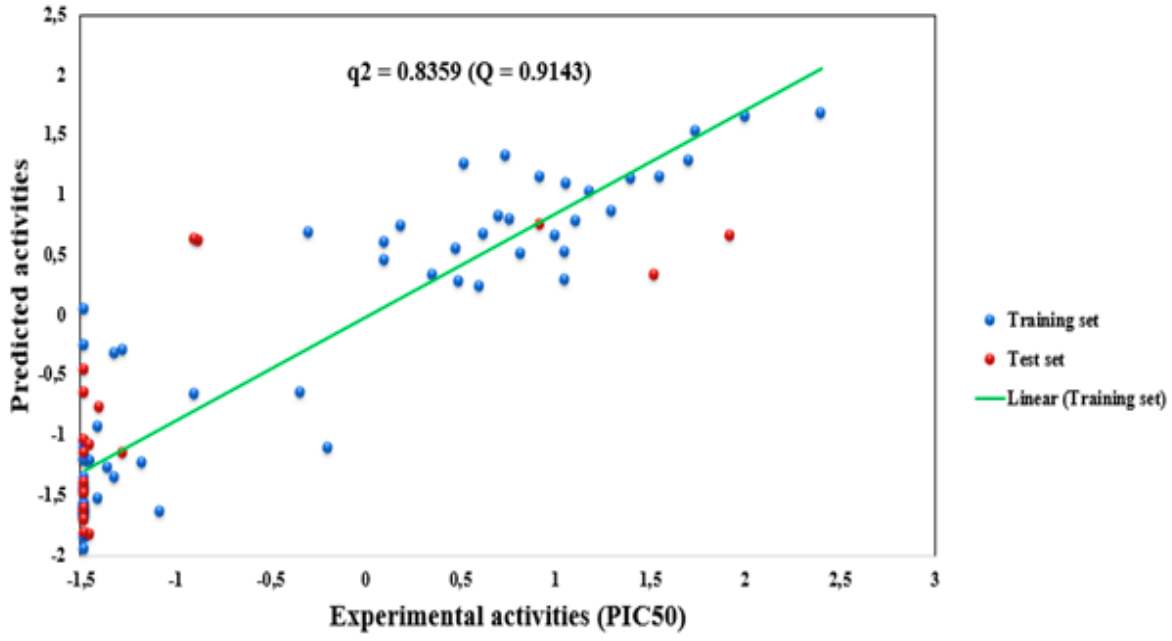


Figure IV-15 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LOO, en fonction des activités expérimentales PIC50.

III.2.5.2 Validation croisée CV-LTO

Les figures IV-16, IV-17, IV-18 et IV-19 , représentent les courbes de régression linéaire des moindres carrées partiels (PLS) de la validation croisée Leave-Two-Out (CV-LTO), des quatre modèles QSAR, notés PC = 2, PC = 3, PC = 4 et PC = 5.

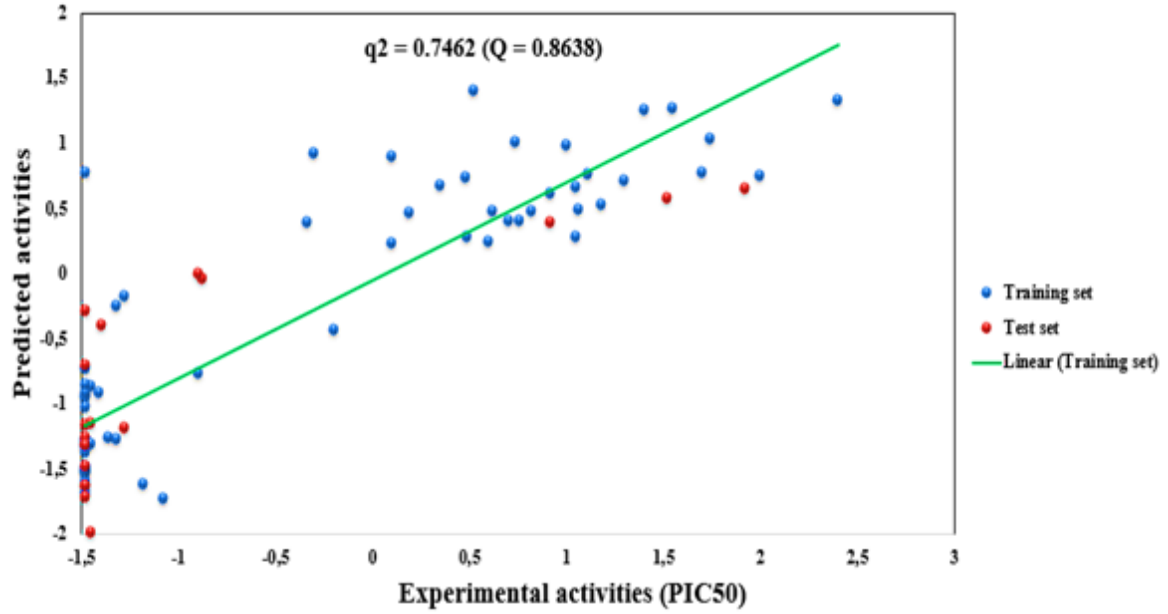


Figure IV-16 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

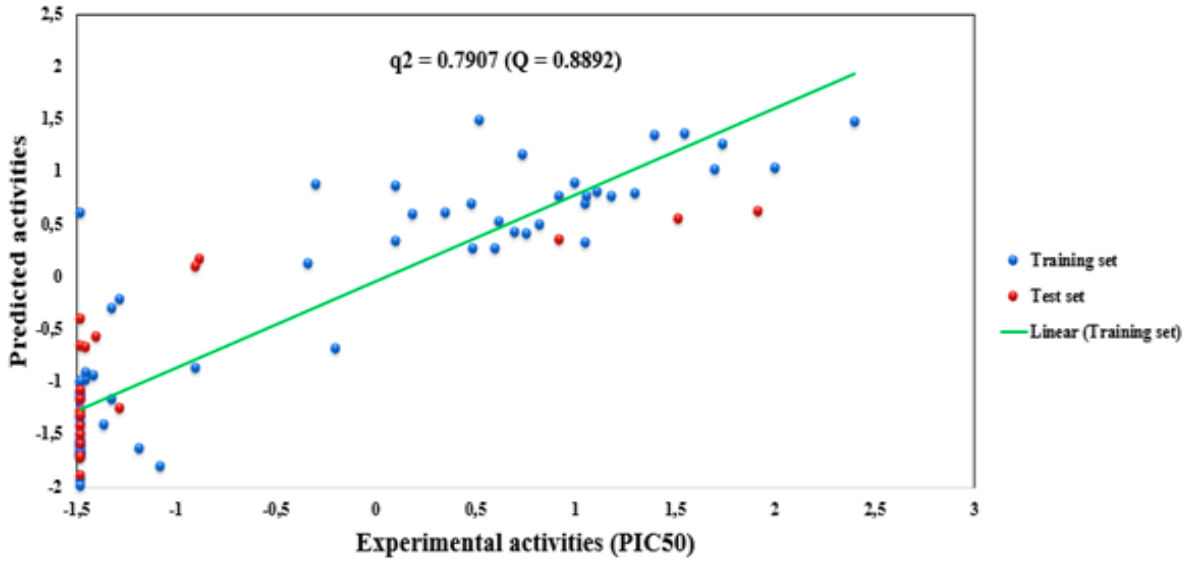


Figure IV-17 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

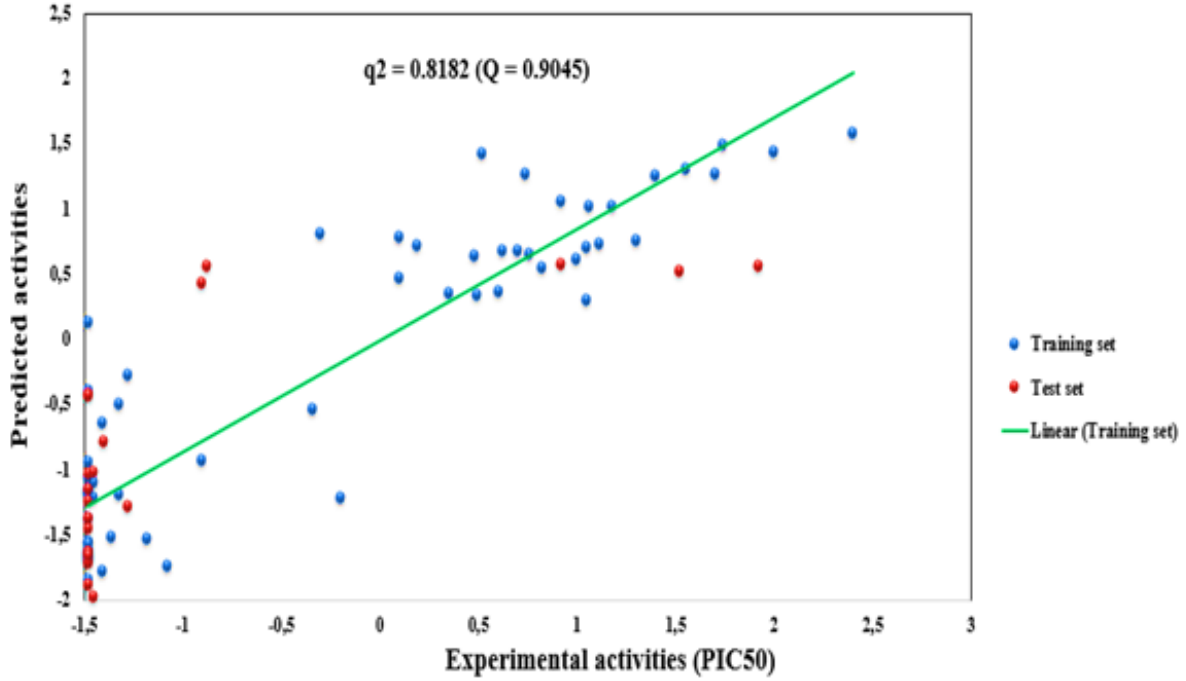


Figure IV-18 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

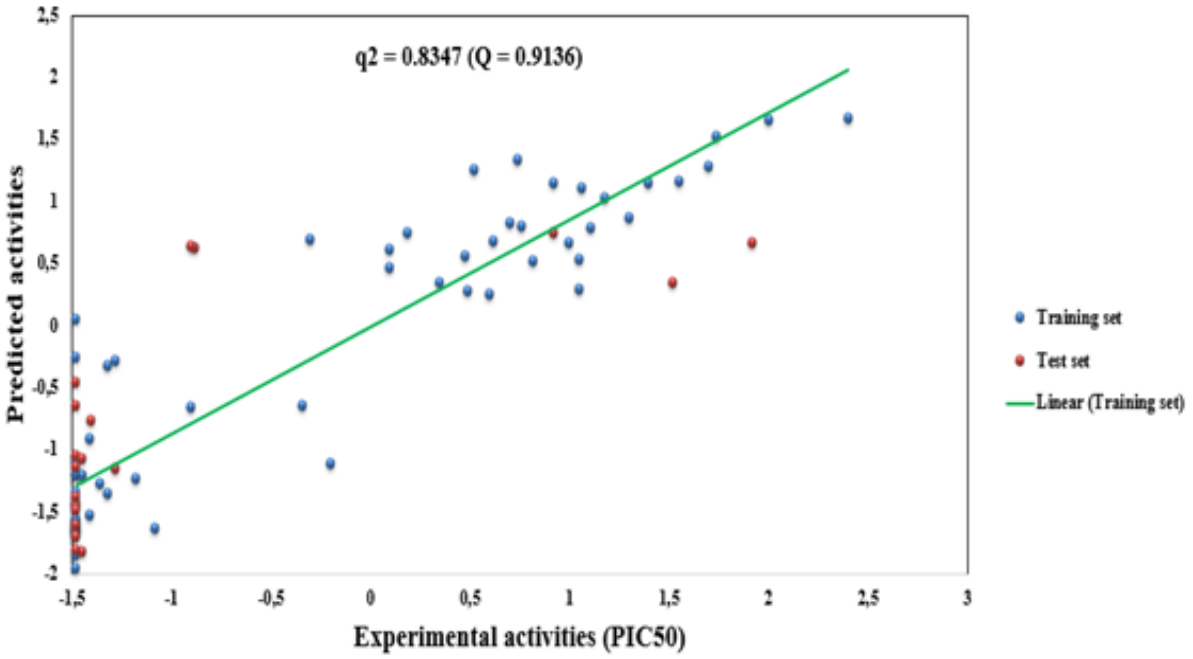


Figure IV-19 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

III.2.5.3 Validation croisée CV-LMO

Les figures IV-20, IV-21, IV-22 et IV-23 , représentent les courbes de régression linéaire des moindres carrées partiels (PLS) de la validation croisée Leave-Many-Out (CV-LMO), des quatre modèles QSAR notés PC = 2, PC = 3, PC = 4 et PC = 5.

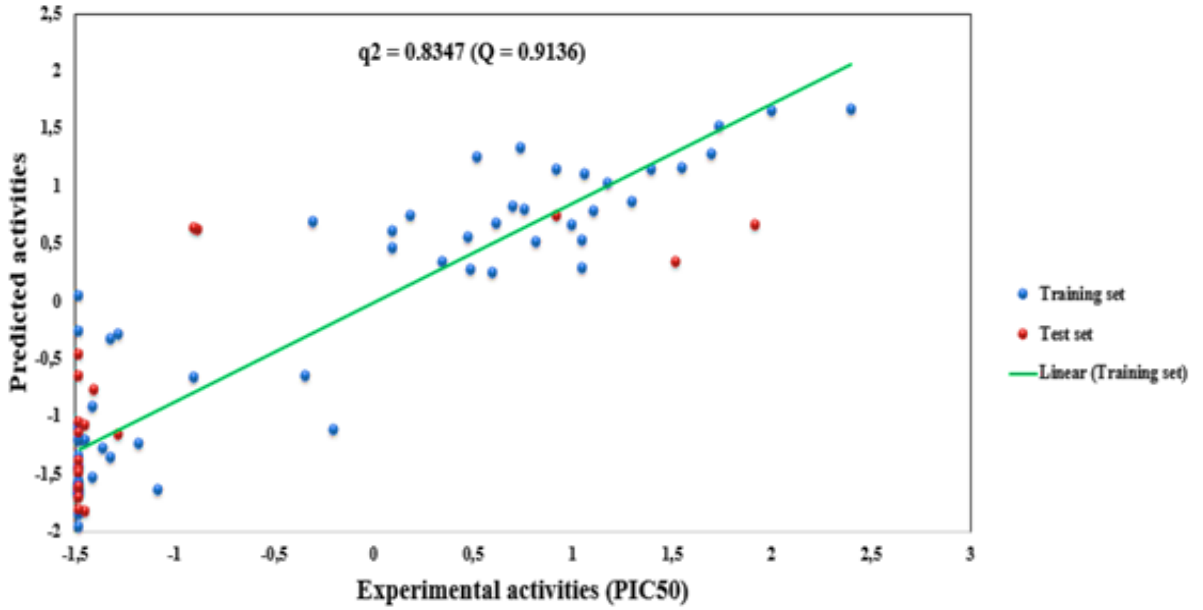


Figure IV-20 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 2) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

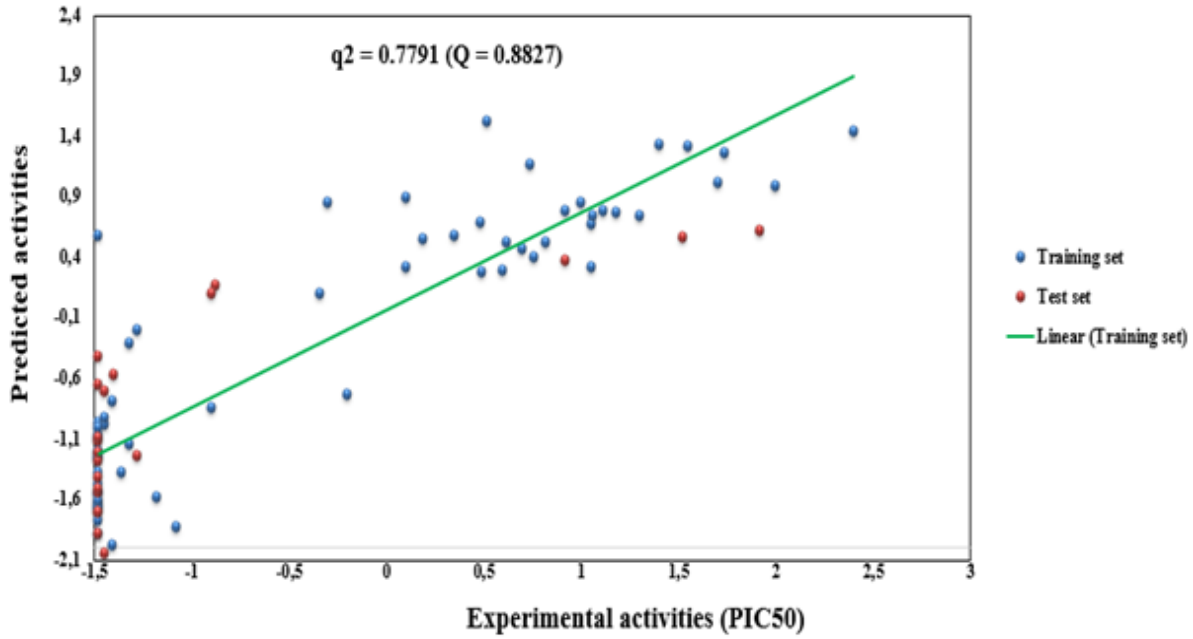


Figure IV-21 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 3) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

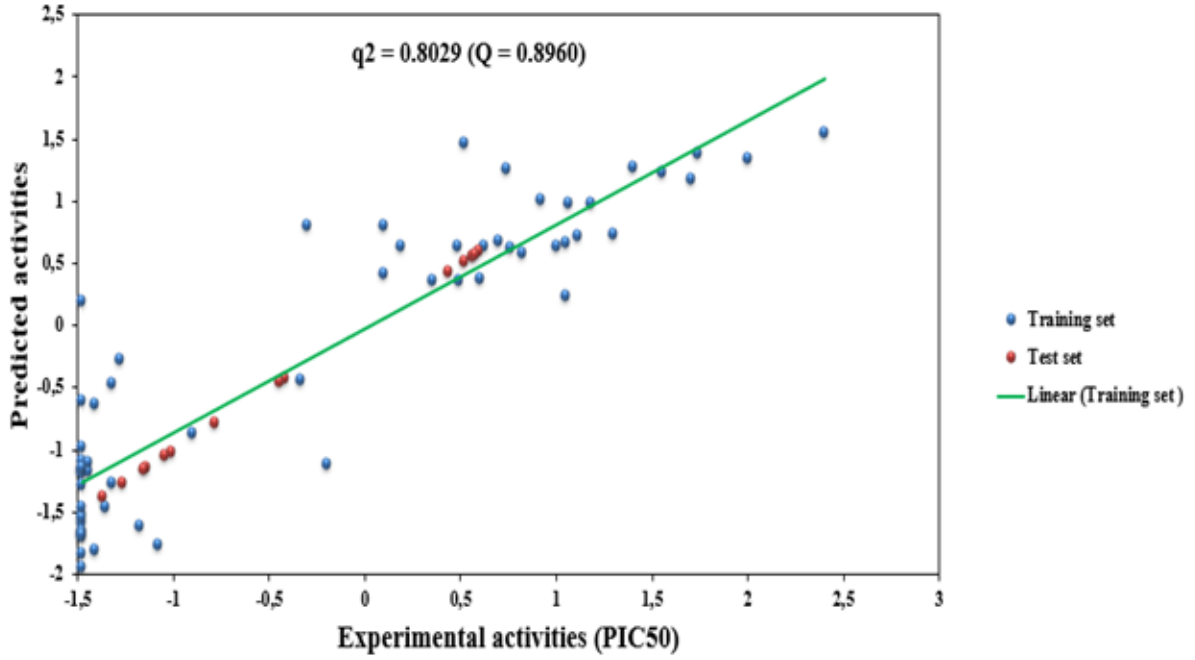


Figure IV-22 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 4) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

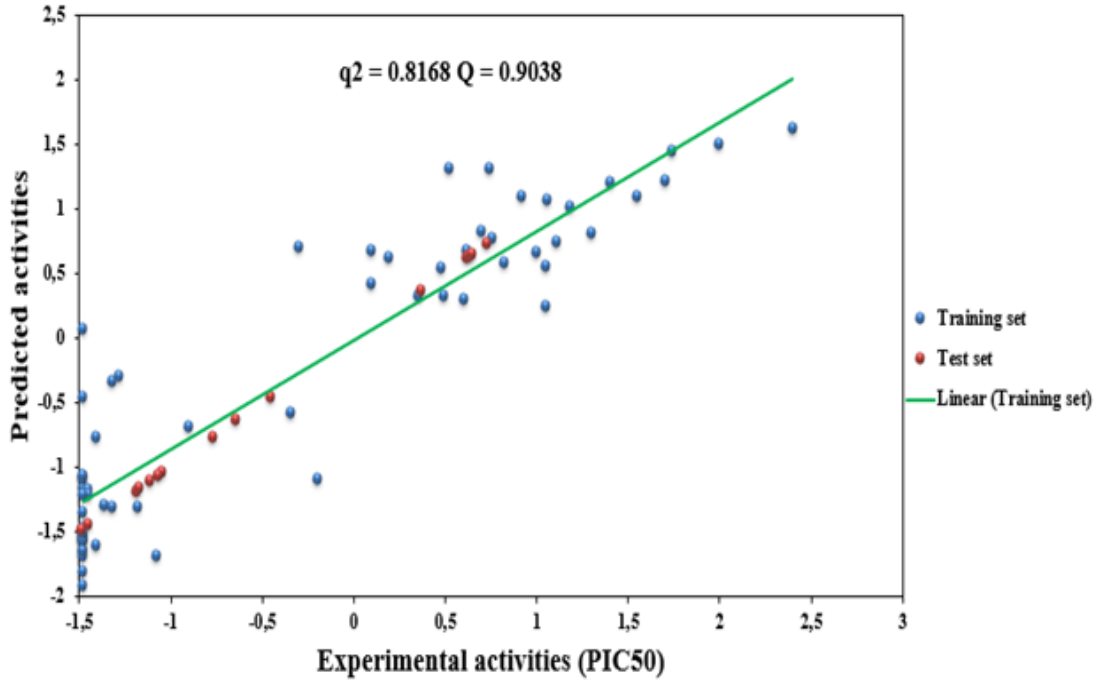


Figure IV-23 : Représentation graphique des activités inhibitrices prédites par la PLS (PC = 5) de la validation croisée CV-LTO, en fonction des activités expérimentales PIC50.

III.2.5.4 Comparaison des résultats des trois méthodes de Validation croisée

Les résultats obtenus, par les trois méthodes de validations croisée décrites ci-dessus (LOO, LTO et LMO), sont regroupés dans le tableau IV-3 suivant :

Tableau IV-3 : Résultats de validation croisée par les méthodes statistiques LOO, LTO et LMO, pour les quatre modèle QSAR, PC = 2, PC = 3, PC = 4 et PC = 5, générés par COMFA du logiciel Open 3D QSAR

pc	LOO			LTO			LMO		
	q^2	Q	SDEP %	q^2	Q	SDEP %	q^2	Q	SDEP %
2	0.7471	0.8643	0.6220	0.7462	0.8638	0.6232	0.7343	0.8569	0.0168
3	0.7912	0.8895	0.5651	0.7907	0.8892	0.5660	0.7791	0.8827	0.0223
4	0.8195	0.9053	0.5256	0.8182	0.9045	0.5274	0.8029	0.8960	0.0271
5	0.8359	0.9143	0.5011	0.8347	0.9136	0.5029	0.8168	0.9038	0.0320

Nous constatons par l'analyse des résultats résumés dans le tableau 3, que les résultats obtenus par les trois méthodes de validations croisée (LOO, LTO et LMO), sont similaires et sur la même concordance avec les activités expérimentales PIC50.

En effet, nous avons obtenu des coefficients de prédiction q^2 très satisfaisants pour les quatre modèles QSAR validé avec les trois méthodes de validation croisée (LOO, LTO et LMO), puisque tous les coefficients de corrélation prédit Q, sont largement supérieurs à limite de spécification **0.75** et très proches de 1 [25], à savoir :

- **CV- LOO** : (**0.75**< Q = 0.8643, 0.8895, 0.9053 et 0.9143 <**1**).
- **CV- LTO** : (**0.75**< Q = 0.8638, 0.8892, 0.9045 et 0.9136 <**1**).
- **CV- LMO** : (**0.75**< Q = 0.8569, 0.8827, 0.8960 et 0.9038 <**1**).

De plus, les valeurs de standard déviation de l'erreur prédite (SDEP) relatives aux quatre modèles QSAR, générés avec les trois méthodes de validation croisée, sont largement inférieures à la valeur de SDEP maximale tolérée, qui est égale à 5%, à savoir :

- **CV-LOO : (SDEP** = 0.6220% ; 0.5651% ; 0.5256% et 0.5011% << **5%)**
- **CV-LTO : (SDEP** = 0.6232% ; 0.5660% ; 0.5274% et 0.5029% << **5%)**
- **CV-LMO : (SDEP** = 0.0168% ; 0.0223% ; 0.0271% et 0.0320% << **5%)**

III.2.5.5 Sélection du meilleur modèle QSAR

Nous avons observé, au cours des précédentes analyses des résultats, que la meilleure méthode de validation croisée, ayant le meilleur coefficient de corrélation, et un plus faible standard déviation de l'erreur prédire SDEP, est la méthode **LMO** (*Leave-Many-Out*). Cette méthode est appliquée aux paramètres optimaux **pc=5** ; C'est-à-dire avec l'extraction de cinq

composantes principales (PCs), des descripteurs moléculaires MIFs de la matrice PLS résultante.

En effet, nous avons obtenu un coefficient de prédiction **q^2= 0.8168**, qui correspond à un meilleur coefficient de corrélation prédit **Q=0.9038**, avec le plus faible SDEP, qui est égale à : **0.0320%**.

VI. Conclusion

Au cours de cette présente étude, nous avons générés des modèles QSAR, entre la structure et l'activité inhibitrice de 83 dérivés de flavonoïdes envers la protéine CK2. Le docking moléculaire a été utilisé dans un premier temps, pour préparer les molécules de flavonoïdes, constituants la base de données QSAR. Cette base est répartie en deux ensembles : un lot d'entrainement pour générer les modèles QSAR et un lot de test pour valider les modèles QSAR résultants.

Ensuite, les conformations bioactives des flavonoïdes, issues du docking, ont été alignées dans une boite cubique de $1Å^3$, en utilisant l'approche « *atom-based* » alignement method', implémenté dans le logiciel Open3dallig. Par conséquent, les champs d'interactions moléculaires MIFs, combinant les champs stérique et électrostatique, ont été créés et calculés par l'outil COMFA implémenté dans le logiciel Open 3D QSAR.

Tous les modèles QSAR générés pour les différents paramètres optimaux, ont été analysés par la méthode de régression linéaire des moindres carrées partiels PLS.

En effet, une vérification initiale de la fiabilité des modèles QSAR, a été menée par une validation non croisée (*Non-cross Validation*). Cette étape est effectuée par comparaison des paramètres statistiques (r^2, R et SDEC) avec les critères d'acceptation définis par Pavais-Pearson. Cette validation conduit à des résultats acceptables pour quatre modèles parmi les cinq prédit par notre étude.

Ensuite, les quatre modèles retenus, ont été soumis aux tests de validation croisée, en utilisant plusieurs méthodes statistiques comme LOO (*Leave-One-Out*), LTO (*Leave-Two-Out*) et LMO (*Leave-Many-Out*). De la même manière que la validation non croisée, la validation croisée est effectuée en comparant les paramètres statistiques q^2, Q et SDEP calculés avec les limites de spécifications requises. Les trois méthodes ont démontré une

bonne concordance, des modèles QSAR, avec les activités biologiques expérimentales des dérivés de flavonoïdes.

La comparaison des résultats obtenus par les différentes méthodes de validation utilisées nous a permis de choisir le meilleur modèle QSAR prédictif obtenu avec le plus faible taux d'erreur. Il s'agit du modèle généré par l'extraction de 5 composantes principales de descripteurs moléculaires MIFs (PC = 5), ayant un coefficient de corrélation de prédiction q^2= 0.8168 (Q=0.9038).

Finalement, les résultats encourageants obtenus au cours de cette étude nous permettent de dire que les modèles QSAR générés peuvent être utilisés dans des calculs de screening virtuel à haut débit pour l'identification de nouveaux inhibiteurs potentiels de la CK2.

IV. Références bibliographiques du chapitre IV

[1] Gramatica P., (2013). Methods in molecular biology (Clifton, N.J.) 930:499-526. On the Development and Validation of QSAR Models. https://doi.org/10.1007/978-1-62703-059-5_21
[2] Kunal R., Indrani M., (2012). On the Use of the Metric rm 2 as an Effective Tool for Validation of QSAR Models in Computational Drug Design and Predictive Toxicology. Mini Reviews in Medicinal Chemistry,12(6):491-504. https://doi.org/10.2174/138955712800493861
[3] Azarboni R H., Keyanpour M., Yaghouti M., (2019). Leave-Two-Out Cross Validation to optimal shape parameter in radial basis functions. Engineering Analysis with Boundary Elements,100: 204-210. https://doi.org/10.1016/j.enganabound.2018.06.011
[4] Shameera T K A., Vijisha K., Rajan K M., (2019). QSAR modeling of benzoquinone derivatives as 5-lipoxygenase inhibitors. Food Science and Human Wellness, 8(1): 53-62. https://doi.org/10.1016/j.fshw.2019.02.001
[5] Lolli G C., MazzoranaM., Tibaldi M., Cesaro E., DonellaD L., MeggioA., Venerando F., Franchin A., Sarno C., Battistutta S., Pinna R LA., (2012). Inhibition of Protein Kinase CK2 by Flavonoids and Tyrphostins. A Structural Insight. Biochemistry, 51:6097-6107. https://doi.org/10.1021/bi300531c
[6]Goodsell D S., Morris G M., Olson A J., (1996). Automated Docking of Flexible Ligands: Applications of AutoDock J. Mol. Recognition, 9(1): 1-5. https://doi.org/10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6
[7] Golub A G., Bdzhola V G., Ostrynska O V., Kyshenia I V., Sapelkin V M., Prykhod'ko A O., Kukharenko O P., Yarmoluk S M., (2013). Discovery and characterization of synthetic 4′-hydroxyflavones—New CK2 inhibitors from flavone family. Bioorganic&MedicinalChemestry, 21(21) :6681-6689.https://doi.org/10.1016/j.bmc.2013.08.013
[8] Li N., Sun C., Zhou B., Xing H., Ma D., Chen G, et Weng D., (2014). Low Concentration of Quercetin Antagonizes the Cytotoxic Effects of Anti-Neoplastic Drugs in Ovarian Cancer. PLOS One. 9(7): e100314.https://doi.org/10.1371/journal.pone.0100314
[9] JohnJE., McNulty L A M., Magers J., Morgan P., et WilsonA M., (2010). Technology for the Organic Chemist: Three Exploratory Modules. Journal of Chemical Education, 87 (10): 1074-1077. https://doi.org/10.1021/ed100362d

[10] UCSF Chimera-a visualization system for exploratory research and analysis. Pettersen EF., Goddard TD., Huang CC., Couch GS., Greenblatt DM., Meng EC., Ferrin TE., (2004) J Comput Chem. 25(13):1605-12.
[11] Bouaziz-Terrachet S., Toumi-Maouche A., Maouche B., Taïri-Kellou S., (2010). Modeling the binding modes of stilbene analogs to cyclooxygenase-2: a molecular docking study. J Mol Model 16(12) :1919–1929. https://doi.org/10.1007/s00894-010-0679-7
[12] Cheng Y., Prusoff W H., (1973). Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. BiochemPharmacol. 22(23): 3099-3108. https://doi.org/10.1016/0006-2952(73)90196-2
[13] Bouaziz-Terrachet S., Terrachet R., Tairi-Kellou S., (2013). Receptor and ligand-based 3D-QSAR study on a series of nonsteroidal anti-inflammatory drugs. Med ChemRes 22(4):1529–1537. https://doi.org/10.1007/s00044-012-0174-z
[14] Bouaziz-Terrachet S., (2006). Conception d'inhibiteurs sélectifs de l'activité enzymatique de la cyclo-oxygénase-2" COX-2". Mémoire de Magister, Alger, USTHB, N° d'ordre :14/2006-M/CH
[15]Tairi-Kellou S., Bouaziz Terrachet S., Maouche B., et Moreau G., (2008). 2D–QSAR Autocorrelation Study on Selective COX–2 Inhibitors. Internet Electron. J.Mol. Des, 7 : 161-185
[16] Puzyn T., Leszczynski J., Cronin M T., (2010). Recent Advances in QSAR Studies, 8: 3-11.
[17] Schultz T W., Netzeva T I., Cronin MTD., (2003). Selection of data sets for qsars: Analyses of tetrahymena toxicity from aromatic compounds. SAR QSAR Environ Res,14 :59–81.https://doi.org/10.1080/1062936021000058782
[18] Tosco P., BalleT., (2010). Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields Journal of Molecular Modeling/ PubMed 17(1):201-8. https://doi.org/10.1007/s00894-010-0684-x
[19] DeLano, W. L. (2009). The PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, CA, 2002.
[20] Halgren T A., (1999). MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. J. Comput. Chem., 20(7): 730-748. https://doi.org/10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2T
[21] WoldS., (1994). Methods and Principles in MedicinalChemistry, Van de Waterbeemd, H. (Ed.). Verlag-Chemie, Weinheim, Germany
[22] Clark M., Cramer R D., (1993).The Probability of Chance Correlation Using Partial Least Squares (PLS). Quantitative Structure-Activity Relationships, 12(2), 137-145.https://doi.org/10.1002/qsar.19930120205
[23] Sjöström M., Eriksson L., (1995). Chemometric methods in molecular design,hanvan de Waterbeemd, 2: 63-90
[24] Wold S., Sjöström M., Eriksson L., (2001). PLS-regression: a basic tool of chemometrics, Chemometrics Intell. Lab. Syst., 58: 109-130
[25] Pearson B., (2008). Pearson's Correlation Coefficient. In: Kirch W. (eds) Encyclopedia of Public Health. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5614-7_2569

Les travaux de recherches présentés dans ce manuscrit sont orientés vers la conception de nouvelles molécules bioactives dans le domaine du cancer, en utilisant les approches de la modélisation moléculaire et de la bio-informatique. Ils sont réalisés en deux étapes distinctes. La première est orientée vers la conception de nouveaux composés anticancéreux. La deuxième consiste en une étude quantitative des relations structure-activité (QSAR), pour déterminer les éléments structuraux responsables de l'activité anticancéreuse.

Le cancer est une maladie mortelle caractérisée par une prolifération incontrôlée de cellules au sein de l'organisme. Des études récentes ont démontré le rôle déterminant d'une enzyme nommée la caséine kinase 2 (CK2) dans l'apparition de cette maladie. Cette protéine joue un rôle dans la plasticité cellulaire épithéliale. En effet, une altération de son activité multifonctionnelle, implique des variations importantes dans la transformation des cellules épithéliales en cellules cancéreuses plus actives à l'origine des métastases.

Cette protéine se produit principalement en tant qu'hétéro-tétramérique holoenzyme, formé par l'association de deux sous-unités catalytiques **(CK2α et CK2α')** et un dimère composé de deux sous-unités de normalisation ou régulatrices identiques (non catalytiques : **CK2β et CK2β'**). Les sous-unités catalytiques CK2α et CK2α', ne sont pas liées et elles se diffèrent par leurs tailles de séquences peptidiques.

Actuellement, la majorité des inhibiteurs développés de la protéine CK2 sont des compétitifs de l'un des substrats de l'enzyme : l'adénosine triphosphate (ATP). Ces inhibiteurs se fixent dans le site catalytique de l'ATP, situé entre les sous unités **CK2α et CK2β,** empêchant ainsi la protéine de progresser dans son activité catalytique.

D'après la littérature, plusieurs classes d'inhibiteurs puissants de la CK2 ont été développés ces dernières années. Parmi ces inhibiteurs nous pouvons citer l'Apigenin et le lutéoline, le tétrabromobenzotriazole (TBB), les dérivés du pyrazolotriazine (IQA), ainsi que la nouvelle molécule CX-4945. Parmi tous ces inhibiteurs, seulement un composé, le CX-4945, est arrivé en phase des études cliniques.

Après avoir résumé les résultats de notre recherche bibliographique concernant l'état actuel des connaissances sur les protéines caséines kinases 2 (CK2), leur ligand endogène ainsi que les différentes classes d'inhibiteurs, nous avons présenté les méthodes de la modélisation moléculaire et de bioinformatique.

Ce travail de thèse s'est divisé en deux grande parties distinctes. La première, consiste en l'identification de nouvelles molécules anticancéreuses au moyen d'un criblage virtuel d'une chimiothèque issue d'une base de données Pubchem, afin d'identifier des composés pouvant

être des inhibiteurs puissants de la protéine CK2. Pour cela, nous avons généré des modèles pharmacophore, en fusionnant et en combinant les caractéristiques des pharmacophores « Structure-based » et « Ligand-based ». En effet, les trois modèles pharmacophore générés, ont été évalués et validés pour leurs qualités de pouvoir identifier et de discriminer des composés chimiques actifs des inactifs, en utilisant plusieurs paramètres, notamment ROC, AUC et un lot d'essai.

Le meilleur modèle pharmacophore, obtenu en fusionnant les caractéristiques pharmacophoriques basées sur la structure du ligand « *ligand-based* » et celles basées sur la structure de la protéine « *structure-based* », consiste en sept caractéristiques, à savoir : deux liaison HBA, deux HBD, une hydrophobe (Hy) et deux caractéristiques aromatiques (Ars). Ce modèle a été alors utilisé pour cribler virtuellement une chimiothèque issue de la banque de données Pubchem dans le but de rechercher de nouveaux inhibiteurs de la CK2.

Les 5984 composés retenus dans l'étape précédente, ont été analysés, testés et criblés à l'aide de plusieurs techniques et filtres, telles que la règle de Lipinski et la simulation du docking moléculaire. Les modes de liaison favorable et l'énergie de docking des quatorze composés soumis à la simulation de docking, ont été comparés à celui du ligand expérimental « AGI » et analysés par la suite, pour identifier les interactions essentielles dans le site actif de la CK2.

L'analyse des résultats de docking montre que les groupements « oxo pentylsulfanyle » et « cyano » (CN) du composé criblé (hit), forment des interactions favorables avec les acides aminés Asp175, Asn161, Arg47 et Asn118 du site actif de la CK2. De plus, le cycle phényle et le groupement éthyle en position *meta* du cycle pyridine sont entourés des résidus : Val53, Phe113, Ile66, et Ile174 et Val45, Gly46, Met163 et Asn118, engendrant des interactions hydrophobes favorables.

Afin de confirmer la stabilité des complexes ligand-CK2 obtenus par le docking moléculaire, nous les avons soumis à des simulations de dynamique moléculaire pour prendre en considération la flexibilité de ces systèmes. Les 100 ns de simulation de dynamique moléculaire, nous ont fourni un aperçu sur la stabilité moléculaire du ligand prédit « hit » dans le site actif de la CK2.

Ces résultats ont été analysés et validés à l'aide de plusieurs paramètres tels que les RMSD (du ligand et des carbones Cα), les liaisons hydrogène formées entre le ligand et la protéine, la fluctuation quadratique moyenne des carbone Cα (RMSF), la surface accessible aux solvants (SASA totale, SASA hydrophobe et SASA hydrophile) et le rayon de giration (rGyr) des complexes étudiés.

Ainsi, les résultats des simulations de dynamique moléculaire sont en parfait accord avec les résultats du docking moléculaire. En résumé, nous pouvons dire que dans cette partie de notre travail, deux objectifs ont été atteints. Le premier, était le développement d'un nouvel modèle pharmacophorique capable de discriminer les composés actifs des inactifs. Le deuxième, était de proposer un nouvel candidat potentiel « candidat médicament », qui appartient à la famille des pyrido [1,2-a] benzimidazoles, et qui n'a pas encore été caractérisé dans la littérature scientifique comme inhibiteur de la CK2.

La deuxième partie de ce travail, consiste en une étude quantitative de relation structure activité (QSAR). Pour cela, nous avons générés des modèles QSAR, entre la structure et l'activité inhibitrice de 83 dérivés de flavonoïdes envers la protéine CK2. Dans un premier temps, nous avons docké les molécules étudiées dans le site actif de la CK2, pour générer les conformations bioactives constituants la base de données QSAR. Cette base est répartie en deux ensembles : un lot d'entrainement pour générer les modèles QSAR et un lot de test pour valider les modèles QSAR résultants. Les conformations bioactives ont été alignées dans une boite cubique de $1Å^3$, puis utilisées pour générer les champs d'interactions moléculaires MIFs, combinant les champs stérique et électrostatique.

Par la suite, les modèles QSAR ont été générés par la méthode de régression linéaire des moindres carrées partiels PLS puis validé, dans un premier temps, par une validation non croisée (Non-cross Validation). Cette validation conduit à des résultats acceptables pour quatre modèles parmi les cinq prédit par notre étude. Ensuite, les quatre modèles retenus, ont été soumis aux tests de validation croisée, en utilisant plusieurs méthodes statistiques comme LOO (Leave-One-Out), LTO (Leave-Two-Out) et LMO (Leave-Many-Out). Les trois méthodes ont démontré une bonne concordance, des modèles QSAR, avec les activités biologiques expérimentales des dérivés de flavonoïdes.

La comparaison des résultats obtenus par les différentes méthodes de validation utilisées nous a permis de choisir le meilleur modèle QSAR prédictif obtenu avec le plus faible taux d'erreur. Il s'agit du modèle généré par l'extraction de 5 composantes principales de descripteurs moléculaires MIFs (PC = 5), ayant un coefficient de corrélation de prédiction q^2 = 0.8168 (Q=0.9038).

En conclusion les recherches réalisés au cours de la première étape de notre travail, ont permis l'identification d'un nouveau candidat médicament ayant une grande affinité pour la CK2 et un profil de toxicité plus faible. Ainsi, nous comptons poursuivre ce travail en développant des analogues structuraux du nouveau candidat médicament obtenu (pyrido [1,2-

a] benzimidazoles), et les proposer aux expérimentateurs pour des synthèses organiques et des tests biologiques.

D'un autre côté, les études de dynamique moléculaire effectuées sur les complexes obtenus par le docking nous ont permis de confirmer leurs stabilités en prenant en considération la flexibilité de la protéine. Néanmoins, ces études ne nous permettent pas d'étudier la réaction qui peut avoir lieu entre le ligand et les acides aminés du site actif.

Cependant, nous souhaitons poursuivre ce travail au moyen de méthodes hybrides mixtes QM/MM qui consistent à traiter, de manière quantique, le ligand et les résidus du site actif de l'enzyme et, de manière classique, le reste de l'enzyme voire son environnement.

De plus, nous comptons améliorer la robustesse des modèles QSAR que nous avons développé dans la deuxième partie de ce travail en enrichissant notre base de données par la recherche, dans la littérature, de nouvelles molécules de la famille des flavonoïdes ayant une activité inhibitrice pour la CK2.

I. Principales interactions Ligand-Récepteur

I.1 Forces, liaisons ou interactions de Van der Waals (VdW)

Les forces de Van der Waals, est la combinaison de trois termes distincts : Effets d'orientation (forces de Keeson), d'induction (forces de Debye) et de dispersion (forces de London), qui correspondent aux différentes forces interactions électriques, entre les atomes et/ou les molécules. En effet, la caractéristique commune de ces trois termes est que leur énergie est proportionnelle à -1/r^6 [77], comme suit :

$$\mathbf{E_{VdW} = E_{keeson} + E_{Debye} + E_{London} \propto -\frac{1}{r^6}} \quad \text{..... (27)}$$

Les forces de Van der Waals, ont été nommées en l'honneur du physicien néerlandais, Johannes Diderik van der Waals (1837-1923), lauréat du prix Nobel de physique en 1910, qui fut le premier à introduire leurs effets, dans les équations d'état des gaz en 1873, permettant d'aller au-delà de la description simpliste du gaz parfait [77].

I.2 Liaisons hydrogène

C'est une interaction intermoléculaire ou intramoléculaire, de type dipôle-dipôle, qui se forme entre un atome d'hydrogène, portant une forte charge partielle positive et un autre atome très électronégatif (**O**, **N**, **F**). Cette liaison est stabilisée, en présence de groupements donneurs « D » qui sont généralement : **F-H**, **O-H**, **N-H**, **Cl-H**, **P-H**, et parfois **C-H** et de groupements accepteurs « A », qui doivent obligatoirement porter un doublet d'électrons libre : **F**, **O**, **N** (Voir aussi : **Cl**, **S** et **P** et les entités chargées telles que l'anion carboxylate) [78].

Les liaisons hydrogène peuvent être classées dans les trois catégories suivantes :

I.2.1 Liaisons hydrogène fortes (Enthalpies situées entre 15 et 40 kJ·mol^{-1})

Elles sont formées, quand il y'a déficience en électrons au niveau du donneur, ou excès d'électrons au niveau de l'accepteur [78-79].

I.2.2 Liaisons hydrogène modérées (Enthalpies situées entre 4 et 15 kJ·mol^{-1})

Ce sont les liaisons les plus communes en chimie, elles sont formées entre un donneur et un accepteur généralement neutre. L'atome donneur est relativement plus électronégatif, que l'atome d'hydrogène et l'atome accepteur possède un doublet libre d'électrons [78-79].

I.2.3 Liaisons hydrogène faibles (Enthalpies de liaison situées entre 1 et 4 kJ·mol^{-1})

Elles sont comparables aux interactions de Van Der Waals, et distinguées par l'électronégativité de l'atome donneur, par rapport à celle de l'hydrogène engagé dans la liaison covalente avec ce dernier, tel que la liaison (C-H) [78-79].

I.3 Interactions hydrophobes

Une molécule hydrophobe, est une chaine organique qui ne se dissout pas dans l'eau et qui n'a pas d'affinité polaire avec elle. L'interaction de type hydrophobe, résulte des forces d'attraction et de répulsion entre les molécules polaires et les molécules apolaires, qui forment entre elles des interactions de Van der Waals du type dipôle-dipôle. Cette interaction est à l'origine de la séparation des phases aqueuses et lipidiques d'une émulsion acide gras/eau.

Les acides aminés, possédant un radical hydrophobe et apolaire (aliphatiques, aromatiques), ont la propriété d'empêcher la formation de liaisons hydrogène, entre les molécules d'eau, en formant dans la structure des protéines des « **zones hydrophobes** », où les molécules d'eau ne peuvent échanger aucune liaison, avec les radicaux libres d'acides aminés [78-79].

I.4 Interactions hydrophiles

Une molécule hydrophile, est un composé ionique qui a tendance à se dissoudre dans l'eau (hydrosoluble) et les solvants polaires. En effet, la polarité des composés hydrophiles leurs permet, de créer des liaisons hydrogène avec les molécules d'eau et celles des solvants polaires, en incluant les groupements fonctionnels tel que, la fonction amine, un alcool ou une cétone [78-79].

I.5 Interactions Lipophiles

La lipophilie, est l'affinité d'une substance pour les solvants apolaires (lipides). En effet, une substance lipophile (ou hydrophobe) est soluble dans un milieu gras, mais insoluble dans les solvants polaires (eau). Cette caractéristique, interprète les propriétés biologiques d'une molécule (solubilité, métabolisme, distribution dans les tissus, liaison avec le site récepteur..., etc.) et elle constitue une variable importante (descripteur moléculaire), dans les équations de corrélation utilisées pour prédire l'activité biologique (QSAR) [78-79].

I.6 Liaisons ioniques (liaison électrovalence ou pont salin)

Une liaison ionique, se forme entre des atomes possédant une grande différence d'électronégativité, à l'aide des attractions électrostatique dans une protéine, entre les résidus chargés positivement et ceux chargés négativement. Cette liaison, joue un rôle primordial dans les liaisons entre la protéine et d'autres molécules [78-81].

I.7 Ionisation / Protonation des acides aminés

Les acides aminés, sont des molécules amphotères qui peuvent agir comme des acides et comme des bases, grâce à leurs fonctions carboxyle « COOH » et amine « NH2 » ionisables,

qui peuvent être à la fois donneur ou accepteur d'électrons. En effet, selon le pH de la solution, un acide aminé peut exister, sous une de ces trois formes suivantes [78-81] :

- ✓ Zone de pH très acide : NH_3^+–C_α–COOH (cation)
- ✓ Milieu Tampon (Zwitterion) : NH_3^+–C_α–COO^- (Ion dipolaire)
- ✓ Zone de pH très alcalin : NH_2–C_α–COO^- (anion)

En générale, l'énergie d'ionisation fournie à un atome neutre, ou à une molécule pour arracher un électron, forme un ion positif (zone ionisable positive), et l'énergie dégagée lorsqu'un atome neutre capte un électron, forme un ion négatif (zone ionisable négative) [78-81].

I.8 Liaisons aromatique

L'aromaticité, est essentiellement liée à la délocalisation des électrons, sur toute la molécule sous forme d'un nuage électronique. En effet, les composés aromatiques sont impliqués dans des réactions de substitution électrophile, qui préservent leurs cycles aromatiques et permettent d'introduire les groupes fonctionnels sur les cycles benzéniques.

Les acides aminés aromatiques (l'histidine, la phénylalanine, le tryptophane et la tyrosine), sont caractérisés par l'hydrophobie de leurs radicaux, qui forment des liaisons hydrophobes faibles avec d'autres acides aminés hydrophobes (Ile, Leu, etc.) qui contribuent à la structure tertiaire et quaternaire des protéines [79-82].

I.9 Interactions électrostatique (polarisabilité)

La polarisabilité, se définit par la facilité avec laquelle un nuage électronique, se déforme sous l'influence d'un champ électrique **E**. Par conséquent, la molécule subit certaines déformations et acquiert un moment électrique dipolaire induit, proportionnel au champ électrique **E** [81-83], comme exprimé par l'équation suivante :

$$\mathbf{P_{(e)} = \varepsilon_0 \alpha E} \quad \text{..... (28)}$$

A noter que : $P_{(e)}$: Coefficient de polarisabilité, **ε_0**: Constante diélectrique et **α** : Moment électrique dipolaire induit.

Il existe, des acides aminés chargés positivement ou négativement, ce qui donne lieu à des interactions électrostatiques, qui agissent à longues distances (autour de 10 Å), avec des groupes d'atomes électriquement chargés [83].

I. Notions et terminologie

I.1 La règle de Lipinski

La règle de Lipinski (appelée aussi règle de 5), est appliquée en analyse chimio-informatique, pour estimer la biodisponibilité d'un composé par voie orale à partir de sa structure bidimensionnelle, selon les critères suivants [7,17] :

- Un poids moléculaire inférieur à 500 Dalton
- Logarithme décimale LogP calculé inférieur à 5
- Nombre de liaison hydrogène accepteur inférieur ou égal à 10
- Nombre de liaisons hydrogène donneur inférieur ou égal à 5
- Nombre de liaisons de rotables inférieur à 15

I.2 La règle de veber

En plus des propriétés moléculaires définies par Lipinski, d'autres propriétés ont été étudiées par veber, en ce qui concerne la biodisponibilité orale [7,18], à savoir :

- La surface spécifique polaire inférieure à 140 $Å^2$
- Le nombre de liaisons de rotation (rotatable bonds en anglais) inférieur à 10.

I.3 Le filtre de Ghose

Le filtre Ghose, définit les critères d'acceptation de l'analyse « Drug-Likness » comme suit [7,19] :

- Logarithme décimale Log P calculé est compris entre - 0.4 et 5.6
- Le poids moléculaire est compris entre 160 et 480
- La réfractivité molaire est comprise entre 40 et 130
- Le nombre total d'atomes est compris entre 20 et 70
- La surface spécifique polaire inférieure à 140 $Å^2$

I.4 Estimation quantitative de la similarité au médicament (QED)

L'estimation quantitative de la similarité au médicament (Quantitative Estimate of Drug-likeness : QED, en anglais), est une analyse intuitive, transparente et simple à mettre en œuvre, basée sur des mesures empiriques et explicites qui reflètent la distribution des propriétés moléculaires implicites de composés bioactifs, permettant leurs classements relatifs [7,20].

I.5 La perméabilité de la barrière hémato-encéphalique (BBB)

Le système nerveux central est bien protégé par la barrière hémato-encéphalique (Blood- Brain Barrier : BBB, en anglais) qui maintient son homéostasie. En raison de cette barrière, de nombreux médicaments potentiels, ne peuvent pas atteindre le cerveau à des concentrations suffisantes [7,21].

De ce fait, la simulation in silico à travers le filtre BBB Likeness, est utilisée en première phase de découverte de médicament « drug design » pour calculer quelques propriétés physico-chimiques, tels que la perméabilité, la solubilité, la charge, la masse molaire et le nombre de liaisons hydrogène [7,21].

I.5 La règle CMC-50-like

La règle CMC-50-like, définit les critères d'acceptation de l'analyse « Drug-Likness » comme suit [7] :

- Logarithme décimale Log P calculé est compris entre 1.3 et 4.1
- Le poids moléculaire est compris entre 230 et 390
- La réfractivité molaire est comprise entre 70 et 110
- Le nombre total d'atomes est compris entre 30 et 55

I.6 MDDR-like

Le règle MDDR-like, définit les critères d'acceptation de l'analyse « Drug-Likness » comme suit [7] :

- Nombre de cycles aromatiques supérieur ou égale à 3
- Nombre de liaisons rigides supérieur ou égale à 18
- Nombre de liaisons de rotables supérieur ou égale à 6

II. Liaisons hydrogène au sein d'une macromolécule biologique

Une liaison hydrogène est une conséquence fondamentale de la polarité de la molécule d'eau, exerçant une force d'attraction sur les molécules d'eau avoisinantes. En outre, l'atome d'hydrogène d'une molécule subit l'attraction de l'atome d'oxygène d'une autre molécule ; Et reste lié par covalence à son atome d'oxygène, et une distance presque double le sépare de l'atome d'oxygène accepteur [41].

La figure 20, représente les premiers clichés de liaisons hydrogènes obtenus au sein d'une macromolécule biologique (8-hydroxyquinoléine), au moyen de la microscopie à force atomique sans contact (Non-Contact Atomic Force Microscopy : NC-AFM) [41].

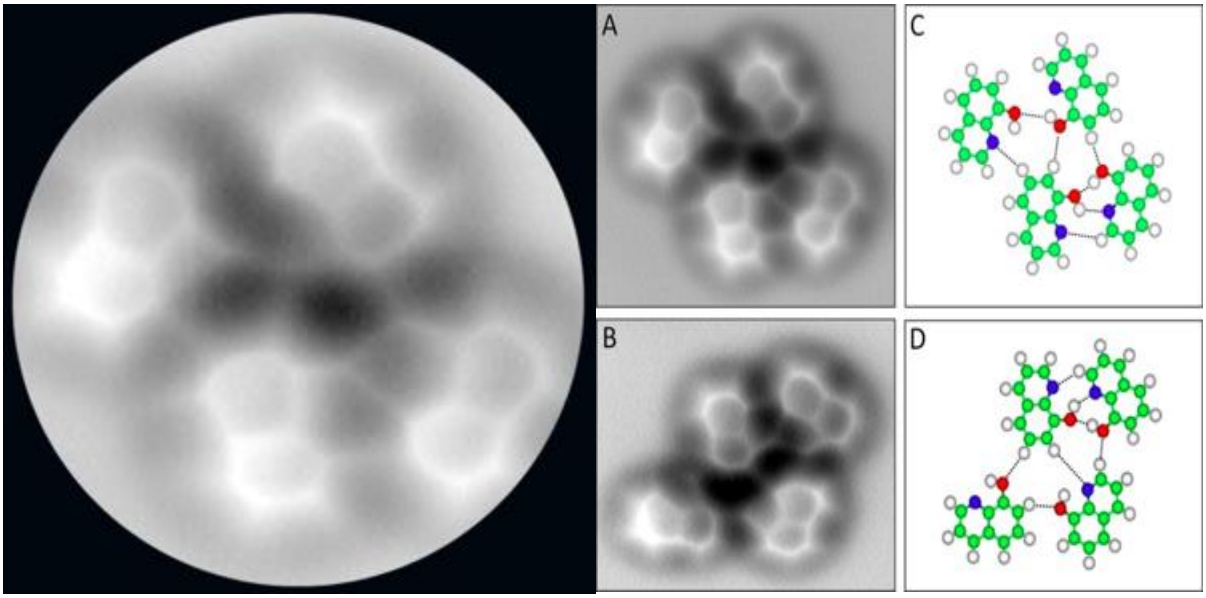


Figure 20 : Visualisation de l'espace réel de la formation de liaisons hydrogène au sein d'une macromolécule biologique (8-hydroxyquinoléine) [41].

Il est bien de noter que la stabilité de la liaison hydrogène dépend de sa géométrie ; En effet, elle sera d'autant plus stable lorsque l'oxygène et l'hydrogène de la liaison « O-H », ainsi que l'oxygène accepteur seront colinéaires [41-42].

Cependant, dans certains cas une molécule d'eau peut former jusqu'à 4 liaisons hydrogène, en formant des structures complexes [42] comme le montre la figure 21.

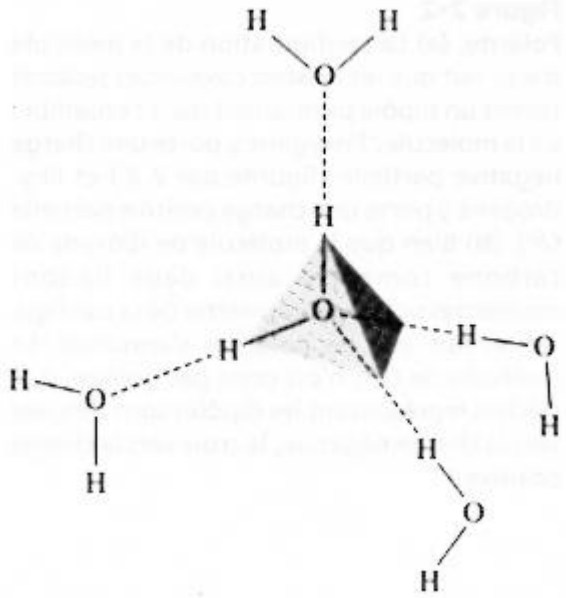


Figure 21 : Formation de quatre liaisons hydrogène d'une molécule d'eau avec les molécules voisines [42]

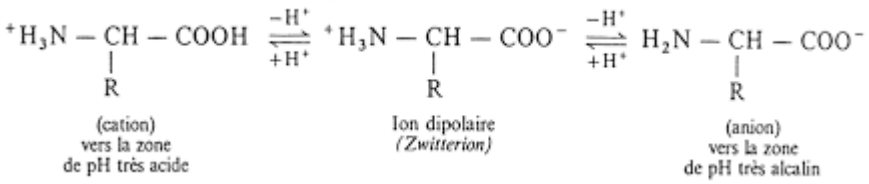
- ✓ L'atome d'oxygène d'une molécule d'eau, accepte deux hydrogènes d'autres molécules (HBA) et chaque groupe « O-H » sert de donneur d'hydrogène (HBD) [41-42].
- ✓ Généralement, environ 15% de molécules d'eau à 37°C (température du corps humain), forment quatre liaisons intermoléculaires dont la durée de vie est très courte [42].

III. Propriétés des acides aminés

La constante d'équilibre isoélectrique (pI) est la valeur à laquelle la charge nette sur une molécule est égale à zéro, ce qui correspond à la forme zwitterionic d'une protéine. En effet, un acide aminé zwitterionic se compose d'un ion négatif de carboxylate et d'un ion ammonium positif sur leur chaîne latérale, qui varie d'un acide aminé à un autre.

Le PKa d'un acide aminé est une constante de dissociation acide, qui se résulte de l'équilibre entre les formes protonées (ajout d'un H+) et déprotonées (enlèvement d'un H+) de la fonction amine (-NH2), la fonction carboxylique acide (-COOH) de la chaine latérale, et n'importe quelle composante potentielle du radical variable (-R). Par conséquent, la charge nette de la protéine est déterminée en additionnant la charge de différents acides aminés qui la constituent [43].

Les acides aminés sont des molécules amphotères, vu qu'ils peuvent agir comme des acides et comme des bases selon leurs valeurs de « pka » et « pI » relatives [43].



Les acides aminés peuvent être classés en fonction de la nature de leurs chaînes latérales comme suit [43] :

- ✓ Aliphatiques : Glycine, Alanine, Valine, Leucine, Isoleucine
- ✓ Hydroxylés/soufrés : Sérine, Cystéine, Thréonine, Méthionine
- ✓ Cycliques : Proline
- ✓ Aromatiques : Phénylalanine, Tyrosine, Tryptophane
- ✓ Basiques : Histidine, Lysine, Arginine
- ✓ Acides: Acide Aspartique, Acide Glutamique

IV. Aperçu général sur les logiciels et programmes utilisés

IV.1 FlexX version 2.3.2 du programme LeadIT

Un logiciel conçu pour la réalisation, des simulations de docking moléculaire. Il se base principalement sur les interactions "protéine-ligand", ainsi que la fonction d'énergie de Böhm (Scoring Function), utilisée dans le classement et l'analyse des conformations issues de docking moléculaire. En effet, FlexX est capable de simuler la présence de molécules d'eau, dans les cavités des protéines et de site actif, qui sont primordiales pour assurer les interactions entre le ligand et les résidus de site actif de la protéine [8,44].

La fonction d'énergie résultante, pour chaque conformation, est calculée selon les propriétés physico-chimiques du ligand, ainsi que la géométrie de son interaction dans le site actif. En d'autres termes, cette fonction considère la baisse d'entropie, lors de la fixation du ligand dans le site actif, en tenant compte de toutes les interactions (liaisons hydrogène, ioniques, aromatiques, etc.) lors du calcul.

En effet, lors de la première itération, le ligand est placé dans toutes ses conformations possibles dans le site actif, et seules les meilleures conformations sont prises, en considération dans l'itération suivante [8,44].

IV.2 Logiciel LigandScout

Un logiciel informatique, qui permet de créer des modèles pharmacophores tridimensionnels, pour la prédiction des activités biologiques de nouveaux médicaments, à partir des données structurelles de complexes « macromolécule-ligand », ou à partir de molécules organiques, réparties en un lot d'entrainement (training set) et un lot de test (test set).

LigandScout, intègre une définition complète des caractéristiques chimiques (liaisons hydrogène donneurs et accepteurs, zones lipophiles, groupements chimiques ionisables positivement et négativement) décrivant l'interaction du ligand dans le site actif de la macromolécule (protéine).

En effet, les pharmacophores peuvent être superposés, en utilisant un algorithme d'alignement, basé sur la correspondance des caractéristiques communes, qui peuvent être interpolées, pour créer un modèle pharmacophore dit à caractéristiques fusionnées, ayant toutes les caractéristiques et interactions communes de plusieurs sites de liaison du ligand [6].

IV.3 NAMD (Nano Molecular Dynamics software)

NAMD 2.12, est un code de dynamique moléculaire parallèle, conçu pour la simulation haute performance de grands systèmes biomoléculaires. NAMD s'adapte à des processeurs sur les plates-formes de haut-gamme, ainsi que sur les simples ordinateurs, en utilisant les formats de fichier AMBER et CHARMM. Les méthodes utilisées par le programme NAMD, décrivent le champ de force de la dynamique moléculaire classique, les équations de mouvement et les méthodes d'intégration, ainsi que les algorithmes d'évaluation électrostatique, utilisés pour le contrôle de température et de pression. Ces fonctionnalités permettent de diriger la simulation de dynamique moléculaire, à travers des barrières énergétiques et de calculer les différences d'énergie libre alchimique et conformationnelle [10].

Résumé

L'objectif principal de cette thèse est de concevoir de nouveaux inhibiteurs de la protéine caséine kinase 2 (CK2), identifiée comme une cible pathologique pertinente dans le traitement du cancer. Pour se faire, nous avons envisagé d'utiliser plusieurs méthodes de modélisation moléculaire et de bio-informatique pour développer des modèles mathématiques nous permettant la prédiction des structures de nouveaux inhibiteurs de la CK2 et de prédire les éléments structuraux responsables de l'activité biologique envers cette protéine. Ainsi, ce travail a été réalisé en deux parties distinctes :

La première partie, a été dédiée à la conception et à la validation d'un modèle pharmacophorique basé sur les approches structure-based et ligand-based. Par la suite, ce modèle a été utilisé comme filtre d'une chimiothèque issue de base de données NCBI. Plusieurs autres méthodes ont été utilisées pour affiner les résultats et obtenir la structure du meilleur inhibiteur, parmi lesquelles nous pouvons citer : l'analyse in silico des propriétés druglikeness, le docking moléculaire, l'analyse in silico de la toxicité et la dynamique moléculaire. Par conséquent, cette étude nous a permis de proposer un nouvel candidat potentiel « candidat médicament », qui appartient à la famille des pyrido [1,2-a] benzimidazoles, et qui n'a pas encore été caractérisé dans la littérature scientifique comme inhibiteur de la CK2.

La deuxième partie, a été consacrée à une étude QSAR, afin de déterminer la relation quantitative structure-activité d'une série de quatre-vingt-trois flavonoïdes, présentant une activité inhibitrice envers la protéine CK2. L'objectif de cette partie est de développer un modèle QSAR capable de prédire la propriété physico-chimique et géométrique responsable de l'activité biologique des molécules étudiées. En effet, le meilleur modèle généré par l'extraction de cinq composantes principales de descripteurs moléculaires MIFs ayant un coefficient de corrélation de prédiction Q = 0.9038.

En conclusion, les résultats encourageants obtenus au cours de cette étude nous permettent de dire que le modèle QSAR généré peut être utilisé dans des calculs de screening virtuel à haut débit pour l'identification de nouveaux inhibiteurs potentiels de la CK2. De plus, la visualisation des champs MIFs peuvent aider dans la proposition de modifications structurelles pouvant améliorer l'activité biologique.

Mots clés : Modélisation moléculaire, Bio-informatique, CK2, Cancer, Flavonoïdes, Inhibiteurs, Pharmacophore, Docking, Criblage virtuel, Dynamique moléculaire, QSAR

Abstract

The main objective of this thesis is to develop new inhibitors of the casein kinase 2 (CK2) protein, considered an interesting target for cancer treatments. For this, Different molecular modeling such as pharmacophore modeling, molecular docking and molecular dynamics simulations approaches have been used to develop new CK2 inhibitors. Thus, this work was carried out in two distinct but complementary parts:

The first part was devoted to presents a pharmacophore model that was generated by combining and merging the structure-based and ligand-based pharmacophore features and validated using Receiver Operating Characteristic (ROC). Based on validation results revealing good predictive ability, this pharmacophore model was used as a three-dimensional query in a virtual screening simulation. Several compounds with different chemical scaffold were retrieved as hits, which were further analyzed and refined using several molecular property filters. The obtained compounds were then filtered and compared to the crystallographic ligand on the basis of their predicted docking energies, binding mode and interactions with CK2 active site residues. This step resulted in a compound with a high pharmacophore fit value and better docking energy. Molecular dynamics simulation indicated a stable binding of the predicted compound to CK2 protein, characterized by Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) and hydrogen bond.The second part was dedicated to QSAR study to determine the quantitative structure-activity relationship of a series of eighty-three inhibitors belonging to flavonoids family. The objective of this part is to develop a QSAR model able to predict the physico-chemical and geometric property's responsible for the biological activity of the studied molecules. However, the best QSAR model, obtained by extraction of five component analysis of molecular interaction fields (MIFs) descriptors, was validated by an external set and several statistical parameters. Thus, the statistical results show a predicted correlation coefficient of Q = 0.9038 indicating a satisfied predictive ability.

The developed QSAR model can be used as three-dimensional query in a virtual screening simulation to identify new CK2 inhibitors. In addition, the analysis of QSAR model, through the MIFs field helped in understanding the possible structural modifications of molecules to improve the inhibitory potency.

Keywords: Molecular Modeling, Bioinformatics, CK2, Cancer, Flavonoïdes, Inhibitors, Pharmacophore, Docking, Virtual screening, Molecular Dynamics, QSAR